UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Process development for the obtention and use of recombinant glycosidases: expression, modelling and immobilisation

PhD Dissertation by

Marta Tortajada Serra

Supervisors: Prof. Dr. Jesús Picó i Marco
Prof. Dr. Daniel Ramón Vidal

May, 2012

Instituto Universitario de Automática e Informática Industrial
Universitat Politècnica de València

“There is a driving force more powerful
than steam, electricity and nuclear power: the will.”
Albert Einstein

Agradecimientos

Quiero dar las gracias a todas las personas que me han acompañado estos años y a las que debo, por distintas razones pero todas importantes, haber conseguido llevar a cabo esta tesis.

A mis directores, Daniel, por darme su confianza desde el principio y ejemplo, día tras día, y Jesús, por descubrirme la biología de sistemas y su respaldo incondicional y permanente. A los dos, por pensar que es buena idea mezclar a biólogos e ingenieros, por ofrecerme la oportunidad de trabajar con ellos, y haber encontrado siempre tiempo para ayudarme.

A la empresa Biopolis, por financiar mis cursos de formación y permitirme dedicar parte de mi tiempo a realizar este trabajo de investigación.

A todo el Departamento de Ingeniería de Sistemas y Automática, por su profesionalidad, y por facilitarme siempre compatibilizar mi trabajo con la realización de la tesis doctoral, y en particular a Jose Luis y Quique por decidir un buen día aplicar su trabajo a estos bichos tan complejos.

Muchas gracias a Kiko, por todas las tardes de trabajo y artículos compartidos, por su rigor y generosidad en la colaboración, sin la que gran parte de esta tesis no habría existido.

A Pedro, Daniel y Jamal, del Instituto de Ciencia de los Materiales de la Universitat de València, por proporcionarme su ayuda, materiales para la inmovilización e ideas para el análisis y la discusión a lo largo de todos estos años.

A todos los compañeros de Biopolis que han estado a mi lado cada día. Muchas gracias a mis más queridos moleculares Verónica, Alex, Antonia, Bea, Dani, Diego – con su ayuda salió el mejor transformante! – Dunja y Silbia, por demostrarme que lo imposible puede ser, y que vale la pena. Gracias a Esther, por animarme más y mejor que nadie. Gracias a María por sus consejos y HPLC inestimables. Gracias a Boro, Empar, Bea, Angela, Roseta, Josep, Patri, Aida, Pepa, Fernando, Silvia, Nuria, Rosa, Maria, Marta, Vicente, Bruno, Nico, Jose y Javi, y a Iryna, Ana, Lola, Jose, Vicky, Carlos, Nacho, Montse, Daniel y Violeta. Gracias a Emma, Chari, David y Mercedes por acogerme en su día en la planta piloto. Gracias a los chicos de Life, y especialmente a Cristina, Juan y Paco por tener siempre una sonrisa fuera de horario.

A Jaime Primo y Mª Angeles Lluch, por abrirme las puertas de la biotecnología. A Chelo, Rosa y Miguel Ángel, por iniciarme en la investigación. A Jose Luis y Auxi y todas sus chicas, por acogerme con absoluta generosidad en su laboratorio y enseñarme cómo clonan los profesionales. Gracias también a los investigadores del IATA, Andrew, Marga, Luisa, Adela, Oscar, Juanan, Jose Vicente, Salvador y Paloma, por su apoyo imprescindible en los comienzos.

A mis padres y mi hermana Carmela, por quererme y apoyarme siempre, y a toda mi familia, y especialmente, a mis abuelos, por creer en mí. A ellos, y los amigos que ya son de casa, por darme fuerzas y perspectiva. A Alfonso, porque sin él nada sería importante.

¡Por fin! Gracias a todos.

Table of contents

Abstract

The general goal of this doctoral thesis is the development of tools for the production and application of two glycosidic enzymes: α-L-arabinofuranosidase from the filamentous fungi *Aspergillus niger* (Abf) and β-D-glucosidase from the yeast *Candida molischiana* (Bgl). These hydrolases are used to release sugars in biomass conversion processes and food industry, and to synthesise aminoglycosides, glycoconjugates and oligosaccharides, high-added value compounds for chemical and pharmaceutical industries. The genes coding for these enzymes have been expressed in the methylotrophic yeast *Pichia pastoris*, and the produced enzymes purified in order to characterise their biochemical properties. Their capacity to catalyse transglycosylation reactions in high yield has been validated. In view of production process assessment and monitorisation, a constraint-based model of *P. pastoris* metabolism has been devised and validated using possibilistic metabolic flux analysis, in order to evaluate model consistency and estimate biomass growth rate and intracellular flux distributions using only a few extracellular measured rates. The model has been extended to estimate recombinant protein productivity, and used to analyse different growth conditions of the recombinant strains overproducing Abf and Bgl enzymes. Finally, the recombinant enzymes have been immobilised on bimodal organosilicas of the UVM-7 family. The resulting biocatalysts have been biochemically and physicochemically characterised and evaluated in different applications of biotechnological interest.

Resumen

El objetivo general de la presente tesis doctoral es el desarrollo de herramientas para la producción y aplicación de dos enzimas glicosídicas: una α-L-arabinofuranosidasa proveniente del hongo *Aspergillus niger* (Abf) y una β-D-glucosidasa (Bgl), proveniente de la levadura *Candida molischiana*. Estas hidrolasas se utilizan tanto para liberar azúcares en procesos de conversión de biomasa y en la industria alimentaria, como en la síntesis de aminoglicósidos, glicoconjugados y oligosacáridos, compuestos de alto valor añadido para la industria químico-farmacéutica. Los genes que codifican estas enzimas se han expresado en la levadura metilotrófica *Pichia pastoris,* y las enzimas producidas se han purificado para caracterizar sus propiedades bioquímicas. Asimismo, se ha comprobado su capacidad para catalizar reacciones de transglicosilación con alto rendimiento. Con el fin de estudiar y monitorizar su producción, se ha planteado y validado un modelo basado en restricciones del metabolismo de *P. pastoris*, evaluando su consistencia mediante análisis posibilístico de flujos metabólicos. El modelo permite estimar la tasa de crecimiento y la distribución de flujos intracelulares a partir de unos pocos flujos extracelulares medidos experimentalmente. Adicionalmente, el modelo se ha extendido para estimar la productividad de proteína recombinante, y se ha empleado para analizar diferentes condiciones de cultivo de las cepas transgénicas que sobreproducen las enzimas Abf y Bgl. Finalmente, las enzimas se han inmobilizado en organosilicas bimodales de la familia UVM-7. Los biocatalizadores resultantes se han caracterizado bioquímica y físico-químicamente y se han evaluado en diferentes aplicaciones de interés biotecnológico.

Resum

L'objectiu general de la present tesi doctoral és el desenvolupament d´eines per a la producció i la aplicació de dos enzims glicosídics: α-L-arabinofuranosidasa, provinent del fong *Aspergillus niger* (Abf), i β-D-glucosidasa (Bgl), provinent del llevat *Candida molischiana*. Aquestes hidrolases s'empren en l'alliberament de sucres en processos de conversió de biomasa i a la indústria alimentària, tanmateix com a la síntesi d'aminoglicòsids, glicoconjugats i oligosacàrids, compostos de gran valor afegit a la indústria químico-farmacèutica. Els gens que codifiquen aquests enzims s'han expressat al llevat metilotròfic *Pichia pastoris*, i els enzims produïts s'han purificat per a caracteritzar les seves propietats bioquímiques. S'ha comprovat també la seua capacitat per catalitzar reaccions de transglicosilació amb elevats rendiments. En relació a la seua producció, s'ha establert i validat un model basat en restriccions del metabolisme de *P. pastoris*, avaluant la seua consistència mitjançant anàlisis de fluxes metabòlics probabilístic. El model obtingut permet estimar la taxa de creixement i la distribució de fluxes intracel·lulars a partir de un reduït nombre de fluxes extracel·lulars mesurats experimentalment. Addicionalment, el model s'ha extès per a estimar la productivitat de proteïna recombinant, i s'ha utilitzat per a validar diverses condicions de cultiu de les soques transgèniques que sobreprodueixen els enzims Abf i Bgl. Finalment, els enzims recombinants obtinguts s'han immobilitzat en materials mesoporosos, organosíliques bimodals de la família UVM-7. Els biocatalitzadors resultants s'han caracteritzat bioquímica i físico-químicament, i s'han avaluat en diferents aplicacions d'interès biotecnològic.

Index of abbreviations

α-MF	alpha mating factor
Abf	α-L-arabinofuranosidase
ACCOA	Acetyl-coenzyme-A
ACD	Acetaldehyde
ACE	Acetate
AKG	α-ketoglutarate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
Bgl	β-glucosidase
Biom	Biomass
Cit	Citric acid
CPR	Carbon dioxide production rate
Cyt	Cytosolic metabolite
DHA	Dihydroxyacetone
DHAP	Dihydroxyacetone phosphate
DNA	Desoxiribonucleic acid
DoE	Design of experiments
DTT	Dithiothreitol
(E)	Extracellular flux
E4P	Erythrose-4-phosphate
EDC	Ethyl dimethylaminopropyl carbodiimide
EM	Elementary modes
EndoH	Endoglycosidase H
EtOH	Ethanol
F6P	Fructose-6-phosphate
FBA	Flux balance analysis
FBP	Fructose-1,6-biphosphate
FS-MFA	Flux spectrum metabolic flux analysis
G6P	Glucose-6-phosphate
GAME	Growth-associated maintenance energy
GAP	Glyceraldehyde-3-phosphate
GLC	Glucose
GLY	Glycerol
HCHO	Formaldehyde
HPLC	High performance liquid chromatography
HPNO	Hierachical porous nanosized organosilicas
HSA	Human serum albumin
ICIT	Isocitrate
iCO2	Carbon dioxyde (intracellular)
iO2	Oxygen (intracellular)
MAL	Malate
MCM	Mobil composition of matter
MET	Methanol
MFA	Metabolic flux analysis
mit	Mitochondrial metabolite
MOC	Model of constraints
MεC	Model of error constraints
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NGAME	Non growth-associated maintenance energy
NMR	Nuclear magnetic resonance
OAA	Oxaloacetate
OD	Optical density
OLS	Ordinary least squares
o/n	Overnight
OTR	Oxygen transfer rate
OUR	Oxygen uptake rate
PAGE	Polyacrylamide gel electrophoresis
PEP	Phosphoenolpyruvate
PG3	3-phosphoglycerate
pNP	p-nitrophenol
pNPA	p-nitrophenyl-α-L-arabinofuranoside
pNPG	p-nitrophenyl-β-D-glucopyranoside
pNPX	p-nitrophenyl-β- D-xylofuranoside
PS-MFA	Possibilistic metabolic flux analysis
PYR	Pyruvate
R5P	Ribose-5-phosphate
RMSE	Root mean square error
RU5P	Ribulose-5-phosphate
S7P	Sedoheptulose-7-phosphate
SN	Culture supernatant
SUC	Succinate
TME	Transmission electron microscopy
vvm	Volume per culture volume per minute
XU5P	Xylulose-5-phosphate

I. Justification, objectives and contributions

Enzymes are proteins capable to catalyse biochemical reactions with high specificity. Isolated enzymes are active *in vitro* in suitable conditions and thus can be used as natural catalysts in many industrial applications, such as colour brightening, fat removal and protein degradation in detergent and textile industries, bleaching agents in pulp and paper industry, and fructose and glucose release and juice clarification in baking, brewing and dairy industries. Besides large volume applications, enzymes are also employed in specialty uses such as analytics and biosensors, DNA-technology, flavour production, personal care products, fine chemical and biopharmaceuticals production. All of them correspond to a world market of more than US $ 2.5 billion (Sharma *et al*., 2009).

The vast majority of enzymes are produced by microbial submerged cultures in large biological reactors. Complex operations in biochemical engineering are required for their obtention with optimal yields. As a consequence, enzymes usually represent large contributions in the estimated cost of bioprocesses in which they are involved. For this reason, substantial efforts are devoted to improve their production processes in all steps. Microbial strains are selected and genetically engineered to overproduce the enzyme of choice. Monitoring and control strategies are developed to cope with complexity and reduce the inherent variability associated to the production of biologicals. In some cases, enzymes can also be immobilised to solid supports to improve biocatalyst performance in high-added value applications.

In this context, the general aim of this thesis is to develop tools for the obtention and industrial production of two glycosidic enzymes, α-L-arabinofuranosidase and β-D-glucosidase. Glycosidases are hydrolytic enzymes that catalyse the release of sugars from polymeric substrates or from compounds in which they are bonded to other molecules such as alcohols or terpenes. They are used in biomass conversion

processes such as bioethanol production, and also in food industry, for detoxification treatments and to promote aroma release. In particular conditions, glycosidases can catalyse the formation of unions between sugars, and can be used to synthesise alkyl and aminoglycosides, glycoconjugates and oligosaccharides, compounds of high interest for chemical, food and pharmaceutical industries.

In turn, *Pichia pastoris* has been selected as a host for the heterologous expression of these proteins. This microorganism is a world-wide recognised platform for the production of recombinant proteins, thanks to the possibility to grow cultures to very high cell densities, its ability to introduce post-translational modifications and the good protein yield/cost ratio. Thus, the first objective in this thesis has been to develop and characterise *P. pastoris* strains able to overproduce the enzymes α-L-arabinofuranosidase from *Aspergillus niger* and β-D-glucosidase from *Candida molischiana*.

In relation to the production process, heuristic, trial and error, approaches are generally applied for the optimisation of fermentation protocols. In contrast, rational modelling strategies can shorten experimental timing and increase recombinant protein productivity. Generally, black-box models are satisfactory for process control purposes, but the increasing availability of large genomics, metabolomics and transcriptomics databases has made it possible to compile more structured, mechanistic-based models of cellular behaviour. Among them, constraint-based models arise as valuable tools to analyse physiological responses of microbial cultures under a systemic approach. Constraint-based models combine stoichiometric description of biochemical pathways with additional restrictions such as reaction irreversibilities, thermodynamics or enzyme capacities. As a result, intracellular mass balances can be established, determining the physiological state of the cell in relation to particular extracellular exchange rates and circumventing the kinetics of the system. In this way, macroscopic mass balances of the components (biomass, substrates and products) of the bioreactor can be linked to metabolic models of the intracellular behaviour. Such “grey models” can provide important qualitative understanding of the system and support advanced monitoring and control strategies. In this framework, the second objective of this thesis has been to establish a constraint-based model of *P. pastoris* metabolism.

Protein productivity is an essential variable in the obtention of recombinant enzymes in *P. pastoris* cultures as it correlates with the variable costs of the bioprocess. This information is also important for monitoring, optimisation and control schemes. However, whereas biomass monitoring strategies are readily available, the estimation of protein productivity has been seldom addressed. Empiric models based on direct correlations with specific growth rate have been established for different *P. pastoris* strains, but they lack generality and are usually non-linear. To overcome these limitations, the third objective of the thesis has been to design an estimator of protein production rate based on the previously developed constraint-based model.

With regard of their technological application, glycosidases have been immobilised on different supports to enable reuse and increase productivity. In addition, covalent binding to solid supports can improve the stability of the enzyme and promote changes in operating conditions. Mesoporous sieves have been used as promising host materials because of their convenient pore sizes and high chemical and mechanical resistance. In view of their use as catalysts in biosynthetic reactions, the fourth objective of this thesis has been to immobilise Abf and Bgl enzymes on bimodal organosilicas.

Finally, it must be mentioned that the present work results from the collaboration between the Institut d´Automàtica i Informàtica Industrial at the Universitat Politècnica of Valencia (ai2-UPV) and the biotechnological company Biopolis S.L, devoted to the selection and manufacturing of microorganisms and cellular metabolites. For this reason, an underlying objective of the thesis has been that the results, tools and methodologies developed herein may be generalised or potentially applicable to other bioprocesses of industrial interest.

Thesis outline

The present document has been structured as follows. Chapter II provides a general introduction to the main topics addressed, that are (i) mathematical modeling and particularly, constraint-based modelling of microbial cultures, (ii) *P. pastoris* as a host for recombinant protein expression, (iii) general features of glycosidases and (iv) immobilisation techniques. Chapter III describes in detail the particular materials, methods and protocols used throughout this work. Chapter IV addresses the expression and characterisation of Abf and Bgl enzymes in *P. pastoris*. The recombinant enzymes have been purified, biochemically characterised and compared to the ones produced in *S. cerevisiae.* As a first approach to their use in biocatalytic applications, their capacity to catalyse transglycosylation reactions has been analysed. Chapter V describes and validates a constraint-based model of *P. pastoris* metabolism. The model and working strategy have been selected to be amenable in industrial environments, where available measurements are scarce and imprecise. For this reason, a variant of classical MFA, possibilistic MFA, developed by Llaneras and coworkers (2009) has been selected and applied to evaluate model consistency and estimate flux distributions using only a few extracellular measured rates in different *P. pastoris* culture conditions. Chapter VI extends the previously developed model to predict recombinant protein productivity. In a first-principles based approach, the estimation has been based on the correlation of energetic cofactors devoted to biomass growth and protein expression. The resulting model has been used to analyse different growth conditions for the recombinant *P. pastoris* strains overproducing Abf and Bgl enzymes. Chapter VII is devoted to the immobilisation of Abf and Bgl enzymes on mesoporous sieves. With this purpose, functionalised materials of the UVM-7 family, designed by El Haskouri and coworkers (2005) have been used as supports. The resulting biocatalysts have been biochemically characterised and preliminary evaluated for different applications of biotechnological relevance. In relation to the obtained results, Chapter VIII introduces future perspectives and potential applications. Finally, Chapter IX summarises the general conclusions of this work.

Contributions

The results of this thesis have been published in:

Refereed journal papers

- Tortajada M, Ramón D, Beltrán D, Amorós P. (2005) Hierarchical bimodal porous silicas and organosilicas for enzyme immobilisation. J. Mater. Chem., 15: 3859-3868.
- Tortajada M, Llaneras F, Picó J. (2010) Validation of a constraint-based model of *Pichia pastoris* metabolism under data scarcity. BMC Syst. Biol., 4:115-125.
- Tortajada M, Llaneras F, Ramón D, Picó J. (2012) Estimation of recombinant protein production in *Pichia pastoris* based on a constraint-based model. Accepted for publication in J. Proc. Cont.
- Pérez-Cabero M, Hungría AB, Morales JM, Tortajada M, Ramon D, Moragues A, El Haskouri J, Beltrán A, Beltrán D, Amorós P. (2012) Interconnected mesopores and high accessibility in UVM-7-like silicas. Under review in J. Nanopart. Res.

Communications and conference papers

- Tortajada M, Llaneras F, Picó J. (2007) Structural analysis of metabolic pathways applied to heterologous protein production in *P. pastoris*. European Congress on Biotechnology. Barcelona, Spain.
- Tortajada M, Llaneras F, Picó J. (2008) Constraint-based modelling applied to heterologous protein production with *P. pastoris*. Reunión de la red Española de Biología de Sistemas. Valencia, Spain.
- Tortajada M, Ramón D, Beltrán D, Amorós P. (2009) Glycosidase enzymes stabilisation through immobilisation onto nanoparticulated bimodal organosilicas European Congress on Biotechnology. Barcelona, Spain.

- Tortajada M, Llaneras F, Picó J. (2010) Possibilistic validation of a constraint based model for *Pichia pastoris* under data scarcity. Proceedings of the 11th International Symposium on Computer Applications in Biotechnology. WeMT1.4. 1-24. Leuven, Belgium.
- Pérez-Cabero M, Hungría AB, Morales JM, Tortajada M, Ramón D, Moragues A, El Haskouri J, Beltrán D, Amorós P. (2011) Interconnected mesopores and high accessibility in UVM-7-like silicas. 5th International FEZA Conference. ISBN 978-84-8363-722-7. Valencia, Spain.
- Llaneras F, Tortajada M, Ramón D, Picó J. (2012) Dynamic metabolic flux analysis for online estimation of recombinant protein productivity in *Pichia pastoris* cultures. MATHMOD 2012 - 7th Vienna International Conference on Mathematical Modelling. Vienna, Austria.

II. Introduction

2.1 Mathematical modelling of microbial cultures

Models are selectively simplified representations or abstractions of reality that describe, typically in mathematical terms, the behaviour of a system. The purpose of a model is to understand a system by creating a workable structure to investigate its properties and usually to predict its outcomes, with the ultimate goal of building a control mechanism for improvement or regulation (Bailey, 1998).

A plethora of mathematical models have been developed in order to describe biological systems and, in particular, microbial cultures (Nielsen and Villadsen, 1992; Picó-Marco, 2004). Due to their characteristics (high complexity, lack of full deterministic knowledge and non-linearity) the purpose of these models is, generally, to address the study of the system as a whole, to explain the interactions among its components, and to explore questions not possible or expensive to be conducted experimentally (Shimizu, 2002). In this context, this section will review some relevant concepts related to mathematical modelling of microbial cultures and focus on constraint-based models as the selected tool to address the representation of microbial metabolism.

2.1.1 Bioreactors as dynamic systems

From a biochemical engineering point of view, microbial cell cultures can be considered heterogenous (solid-liquid) systems in which substrates dissolved in the culture medium (carbon, nitrogen and minor nutrients) are consumed by a growing organism (the biomass). In addition, aerobic organisms require oxygen and produce carbon dioxide both found dissolved in the liquid phase and in the overhead space. The biomass can also generate products such as polymers or small biochemicals. At the same time, cells will grow at a rate dependent of their own concentration (these are autocatalytic

systems) but also of the surrounding conditions, such as substrate concentration and environmental variables such as dissolved oxygen, pH and temperature.

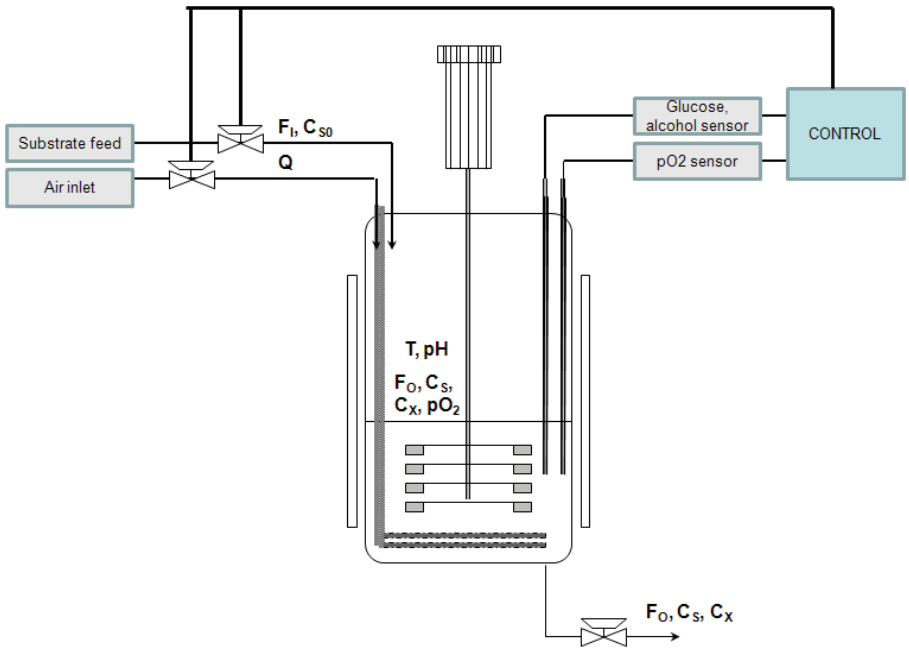


Figure 1. Scheme of a fermentation system.

Figure 1 depicts a scheme of a fermentation system. The volume will vary according to the rates of substrate and outlet feeding, depending on the operating mode. When mass balances are applied to cells, substrates and products, this dynamical system is mathematically described attending to the set of Ordinary Differential Equations (ODE) shown in (1-4), in which losses of products in exhaust gases have been disregarded.

$$\frac{d(C_X \cdot V)}{dt} = r_X = \mu \cdot V \cdot C_X - F_{OUT} \cdot C_X \quad \textbf{(1)}$$

$$\frac{d(C_S \cdot V)}{dt} = r_S = F_{IN} \cdot C_{S0} - \sigma \cdot V \cdot C_X - F_{OUT} \cdot C_S \quad \textbf{(2)}$$

$$\frac{d(C_P \cdot V)}{dt} = r_P = \pi \cdot V \cdot C_X - F_{OUT} \cdot C_P \quad \textbf{(3)}$$

$$\frac{dV}{dt} = F_{IN} - F_{OUT} \quad \textbf{(4)}$$

The rates of substrates into the culture/cell are the elements of the vector r_S, and the rates of the metabolic products are elements of the vector r_P. The formation of new biomass, or accumulation, happens within the "black-box" at a specific growth rate, μ (Stephanopoulos, 1998). These rates are correlated to biomass growth rate through the substrate consumption and production formation yields shown in (5):

$$Y_{S,i/X} = \frac{r_{S,i}}{\mu}$$
$$Y_{P,i/X} = \frac{r_{P,i}}{\mu} \quad \textbf{(5)}$$

The specific rates μ, σ and π are functions that depend on the physiological state of the culture and will vary according to process conditions. They are important targets for optimisation since they will determine the productivity of the process. The unknown parameters of such a black-box model (specific yields and specific rates) can be obtained adjusting experimental data, thus providing a heuristic model, whereas in first-principles models, mechanistic knowledge of the system is used to *a priori* estimate such values.

2.1.2 Constraint-based models of microbial metabolism

In a black-box model, the rates of substrates, or eventually products, together with microbial growth, are the only variables considered (Fredrickson *et al.*, 1970; Bastin and Dochain, 1990). Conversely, structured models make use of several defined compartments that can either relate to real subcellular entities such as organelles or

biochemical compounds or simply represent useful partitions for the purpose of study (Provost and Bastin, 2004).

Generally, black-box models are satisfactory for process control purposes, as more complex models imply additional analysis and computational costs that do not result in an enhanced control performance (Picó-Marco, 2004). However, the increasing availability of large genomics (*e.g.,* GenBank, EMBL, DDBJ), transcriptomics (*e.g.,* TRANSFAC), proteomics (*e.g.,* SWISS-Prot, PIR, BRENDA), metabolomics (*e.g.,* KEGG, WIT/MPW) and cell signalling (*e.g.,* CSNDB, TRANSPATH, GeneNet) datasets has boosted the number of references describing more structured models of cellular behaviour that try to link molecular and biochemical mechanisms to physiological responses under a systemic approach (Palsson, 2000). This insight into the intracellular state of the system can provide important qualitative information, potentially lead intervention within cells and in the framework of biochemical engineering, support monitoring and control strategies (Mahadevan *et al.*, 2005; Mo *et al.*, 2009).

In order to exploit this knowledge for modelling purposes, a collection of biochemical reactions involved in cell metabolism can be connected in a network and related to extracellular substrate consumption and biomass growth rates. This results in a stoichiometric model.

Stoichiometric models are represented by graph-oriented networks including different metabolites of a living organism. The metabolites are linked by arrows and each arrow stands for a particular reaction. Generally speaking, each network branch corresponds to a specific enzymatic reaction, however simplified pathways in which several enzymes (enzyme sets) are lumped can also be considered. Mathematically, the system is described by a matrix N in which metabolites are distributed along files, and reactions, along columns. Each a_{ij} element of the matrix is the stoichiometric coefficient of the i^{th} metabolite in the j^{th} reaction. The fluxes (ν_j) of these reactions are then defined as the *in vivo* rates of the individual reactions in the cellular system (Llaneras and Picó, 2008).

In order to relate this representation to the dynamic model of the bioreactor, the cell is assimilated to the whole culture and understood as a single node (Figure 2).

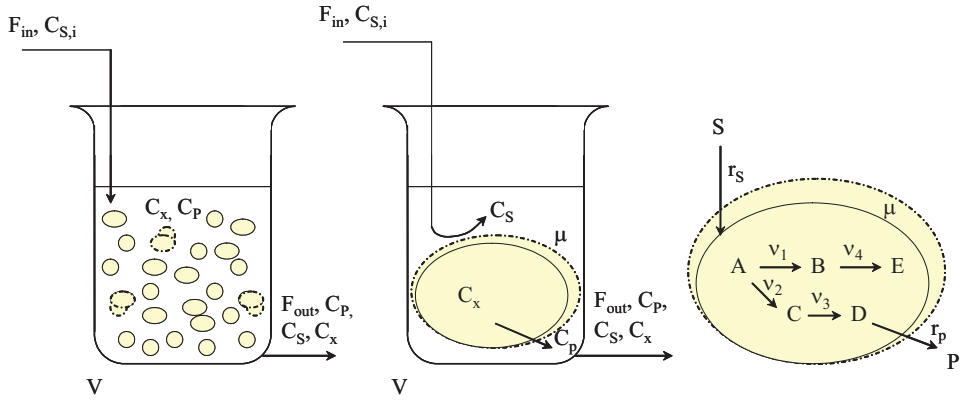


Figure 2. System representation of microbial cells and microbial cell cultures.

In a constant volume system, mass balances result:

$$\begin{aligned}
\frac{dC_{S,i}}{dt} &= -r_{S,i}\cdot C_X + D\cdot(C_{S,0} - C_{S,i}) \\
\frac{dC_{P,i}}{dt} &= r_{P,i}\cdot C_X - D\cdot C_{S,i} \\
\frac{dC_X}{dt} &= (\mu - D)\cdot C_X \\
D &= \frac{F}{V}
\end{aligned} \qquad \textbf{(6)}$$

where D represents dilution.

In chemostats, where concentration of cells is kept constant the dynamics of specific concentration are directly related to consumption/production rates and specific growth rates:

$$\begin{aligned}
\frac{dc_{S,i}}{dt} &= -r_{S,i} + D\cdot(c_{S,0} - c_{S,i}) \\
\frac{dc_{P,i}}{dt} &= r_{P,i} - D\cdot c_{S,i} \\
\mu &= D
\end{aligned} \qquad \textbf{(7)}$$

Material balances for each internal metabolite can be then described by (8) in an analogous expression to (7):

$$\frac{dc_{met,i}}{dt} = \sum_j a_{i,j} \cdot \nu_j - \mu \cdot c_{met,i} \quad \textbf{(8)}$$

and in matricial terms:

$$\frac{dc_{met}}{dt} = r_{met} = N \cdot \nu - \mu \cdot c_{met} \quad \textbf{(9)}$$

An arbitrary selection of the boundaries of the system that usually accounts for extracellular/intracellular environment distinguishes external from internal metabolites. Some of the internal metabolites are considered to fulfil much faster dynamics, so they will not accumulate inside the cell and can be supposed to be at permanent equilibrium, or at pseudo-steady-state. This pseudo-state assumption is the key to solve the matricial system (9).

$$0 = N \cdot \nu - \mu \cdot c_{met} \quad \textbf{(10)}$$

In addition, the second term, representing the effect of dilution due to growth, is usually disregarded as concentrations of subcellular compounds are much smaller than flux rates.

$$0 = N \cdot \nu \quad \textbf{(11)}$$

The main advantage of this representation is that subcellular dynamics are totally disregarded. More complex mechanistic models, in which detailed mechanisms of biochemical interactions, gene regulation or signal processing are represented, require the estimation of a high number of parameters, including kinetic information of all modules involved. The simplicity of stoichiometric models makes them highly suitable to extract valuable information of the feasible states of a living organism while the kind of parameters required (stoichiometric coefficients of biochemical reactions) are largely available (Gombert and Nielsen, 2000).

Stoichiometric models can be extended by introducing additional constraints, such as the irreversibility of certain fluxes, the maximal capacity of some enzymes, regulatory constraints, thermodynamic information or kinetic constants (Rizzi *et al.*, 1997; Visser *et al.*, 2000; Feist *et al.*, 2007). These constraints, together with the initial stoichiometric reactions, reduce the number of functional states that can be achieved by the *in silico* model.

$$\mathrm{MOC} = \begin{cases} \mathrm{N}\cdot\nu = 0 & \text{(a)} \\ \mathrm{D}\cdot\nu \geq 0 & \text{(b)} \end{cases} \qquad \textbf{(12)}$$

Geometrically, the space of feasible flux distributions will be a subspace of R^n in a simple stoichiometric model, whereas if irreversibility constraints are incorporated it will be converted into a convex polyhedral cone. If additional constraints of capacity are considered, the convex polyhedral cone will be bounded. The system of equations of (12) must now be solved through convex algebra techniques as linear algebra is no longer applicable (Heijden *et al.*, 1994).

In this way, constraint-based modeling is based on (i) a network representation of biochemical reactions that is translated into a stoichiometric matrix, (ii) the pseudo-steady state assumption, generally valid for intracellular metabolites, and (iii) additional constraints such as irreversibilities (Llaneras and Picó, 2008).

Constraint based models have been used extensively to simulate and understand the behaviour of microbial cultures under dynamic conditions (Herwig and von Stockar, 2002; Sainz *et al.*, 2003), to apply metabolic control analysis, in order to redirect particular target metabolic fluxes (Cortassa *et al.*, 1995) and to study how the change in environmental conditions (external variables) may modify metabolism (Mo *et al.*, 2009).

2.1.3 Representation of biomass in constraint-based models

Several approaches are considered to integrate the cellular concentration or biomass in constraint-based models. Since this variable is as interesting target for industrial fermentation monitoring and control, they will be reviewed shortly in the following lines. As the pseudo-state hypothesis is in principle applied to intracellular species, expressed in specific units (relative to total cellular mass), the simplest approach is to fully avoid introducing biomass in the metabolic model, and relate it afterwards to the internal metabolites. However this makes it necessary to introduce artificial "external" variables, as sinks for mass accumulation. Some examples can be found in Carlson and coworkers (2002) and Provost and Bastin (2004), usually in small-scale networks for methodological validation or control purposes.

A second approach is to directly relate biomass formation to a set of selected key metabolic precursors, such as AcCoA, ATP, G3P, G6P and NADH. This pseudo-stoichiometric relationship is defined taking into account literature information with regard to each biomass macroscopic, accumulative component (*e.g.* nucleic acids, protein, polysaccharides, among others) requirements of precursors and energy and the elemental composition of biomass (Stephanopoulos, 1998). This simplification is frequently used in practice in medium-scale metabolic models developed for ^{13}C analysis and phenotype study in chemostat cultures and is particularly useful for control purposes if the specific growth rate is explicit in the model (Jin and Jeffries, 2004; Dragosits *et al.*, 2009). If a reference biomass equation is used, the sensitivity of the model towards biomass can be calculated (Cortassa *et al.*, 1995).

Finally, a third approach in which specific equations are introduced for each biomass component is also possible. The biomass can be calculated in a summary equation accounting for its macromolecular composition or disregarded if each component is measured separately (Rizzi *et al.*, 1997; Pfeiffer *et al.*, 1999; Pitkänen *et al.*, 2003). The method can be more accurate if a specific biomass elemental composition is experimentally derived and used for the particular condition studied.

2.1.4 Structural analysis of constraint-based models

With independence of the particular state of the culture (determined by the specific flux distributions of external and internal fluxes), the analysis of the structure of the constraint-based model and the underlying metabolic network provides information related to capacity of the modelled organism. This structural analysis is based on the study of the shape and configuration of the solution space formed by all feasible flux distributions.

In an approach analogue to the concept of basis in linear algebra, sets of relevant simple pathways can be used to represent the whole solution space in constraint based models. Elementary modes (EM) are usually preferred for their nondecomposability and biological significance (Llaneras and Picó, 2010). EM are flux vectors, representing valid flux distributions (fulfilling the model of constraints), that cannot be expressed as a combination of two other solution vectors containing additional zero components. In this way, EM are the minimal sets of equations (of enzymes) that can operate at steady-state (that fulfil the model of constraints), that is, the simplest flux distributions possible for the cell and every feasible flux distribution can be represented by their linear combination. In general, this expression will not be unique as several combinations of different EM can represent the same flux distribution. However, the set of elementary vectors for a metabolic network is univocally defined by the stoichiometric matrix (Schuster *et al.*, 2000; Gagneur and Klamt, 2004; Klamt *et al.*, 2005). EM can be used to understand the capacities of the overall metabolic network, to locate essential reactions for the production of a particular metabolite, infer the effect of deletions, and estimate the optimal yield to produce a product or biomass from a particular substrate.

2.1.5 Solving constraint-based models

In a given constraint-based model, a number of fluxes, m, will be defined. Some of them will correspond to internal fluxes and some others will be exchange rates with the extracellular medium. Depending on the available measurements, a number of fluxes, u, will be unknown. Note that in general, in biochemical networks, the majority of metabolites participate at least in two biochemical reactions, so m>n. There will be n balancing equations (one per internal metabolite) and m fluxes to determine. As a consequence the system will have m-n degrees of freedom (Stephanopoulos, 1998).

The determinacy and redundancy of constraint-based systems (the network in relation to available measurements) can be established in accordance to Klamt and coworkers (2002) as follows:

(i) A system is determined if and only if rank(N_n) = u, and underdetermined if and only if rank(N_n) > u. In this late case, the number of linearly independent constraints is insufficient to solve all rates of rn uniquely. Note that a given number of m-n known fluxes do neither ensure full determinacy, as some of these could be dependent.

(ii) A system is redundant if rank(N_n) < m. This means that some of the rows in N_n are linear combinations of others, and thus the system may be inconsistent if the known rates do not satisfy simultaneously all the constraints.

The first condition implies that a sufficient and independent number of fluxes must be available to solve the system. Clasically, Metabolic Flux Analysis (MFA) has been successfully applied in this case to estimate the unknown fluxes for a given stoichiometric network.

Firstly gross-measurements errors are detected analysing redundancy, and then the system is solved as a weighted least squares problem. In addition to extracellular rates, intracellular measurements from tracer experiments can be incorporated to increase the number of constraints (Szyperski, 1998; Schmidt *et al.*, 1999; Fischer *et al.*, 2004). In underdetermined systems, although the flux distribution cannot be fully solved through MFA uniquely, some rates can still be calculated (Klamt *et al.*, 2002). However, the formulation has limitations, as it only considers equality constraints and it requires a large number of reliable fluxes to be applied. Several alternatives are possible in order to face under-determinacy that rely on different assumptions for pathway selection, such as optimal cell behaviour in Flux Balance Analysis (FBA). FBA makes use of a fixed optimality criterion (*e.g.* maximal growth, minimal energy consumption) to select a particular flux distribution among the infinite solutions (Varma and Palsson, 1994; Edwards *et al.*, 2001 and 2002).

Intervalar approaches, such as Flux Spectrum-MFA (FS-MFA), provide the range in which the calculated fluxes should stand, for a given combination of measured fluxes (Llaneras and Picó, 2007a, b). FS-MFA can be applied to solve the intervalar distribution of the unknown fluxes, accounting for uncertainty of the measured ones through the set of equations shown in (13).

$$\begin{gathered} S_U \cdot v_U = -S_M \cdot v_M \\ If\ v_M\ C\ [v_{min}, v_{max}] \\ \forall v_{Uj}, j = 1 \ldots n_U \\ Min\{v_{Uj}\}, Max\{v_{Uj}\} \\ S_U \cdot v_U = -\langle S_M \cdot v_M \rangle^{min} \\ S_U \cdot v_U = -\langle S_M \cdot v_M \rangle^{max}, v_i \geq 0 \end{gathered} \quad \textbf{(13)}$$

The distance between the interval extremes will be reduced, and this estimation will be more accurate, as the number of measured (known) fluxes increases. In relation to MFA and FBA, the FS-MFA can tackle reaction irreversibilities, together with other inequality constraints, such as maximal flux values, and also provide more reliable and richer estimates as these are intervals instead of point-wise estimates dependent of a particular optimisation criterion. In addition, the uncertainty of measurements is explicit in the formulation of the problem.

This intervalar approach can be linked to possibility theory as described by Llaneras and coworkers (2009). This methodology is known as possibilistic MFA (PS-MFA). PS-MFA is based on (i) representing the available knowledge with constraints, and (ii) adding slack variables penalised in a cost index to deal with uncertainty.

With this purpose, a set of measurement constraints is introduced to complement the model of constraints and to take into account imprecision. The error between actual measurements and flux values is split into two non-negative decision variables in order

to formulate the problem in linear-programming terms. A linear cost is defined that penalises the error among measurements.

In this way, a particular vector v that simultaneously fulfills the model of constraints (MOC) and the measurement constraints has an associated possibility index that can be calculated on the basis of the corresponding linear cost of the error. The main formulation of PS-MFA approach is described in Section 3.5.3. Further detail can be found in Llaneras and coworkers (2009) and Llaneras (2010). The main advantages of PS-MFA are:

(i) PS-MFA is applicable to stoichiometric balances, but also to constraint-based models that consider additional restrictions, such as irreversibilities.

(ii) Formulation of uncertainty is flexible, and non-symmetric error can be introduced.

(iii) Intervalar estimations are produced, which is more reliable than point-wise estimations when multiple flux values are reasonably possible.

(iv) It can be achieved even when only a few fluxes are measured, providing a wider but reliable interval embracing the feasible solutions for a given possibility.

(v) Inconsistencies can be detected and handled both in the model and the measurements.

(vi) High computational efficiency.

2.2 Recombinant protein production in *Pichia pastoris*

2.2.1 General features

The yeast *Pichia pastoris* is a well known platform for the obtention of recombinant proteins as an alternative system to *Saccharomyces cerevisiae*. *P. pastoris* is a methylotrophic yeast and thus can grow using methanol as sole carbon source. It is world-wide recognised as a reference platform for the heterologous expression of recombinant proteins in eukaryotes, due to the possibility to grow cultures to very high cell densities, its ability to produce post-translational modifications, and the generally appealing protein yield/cost ratio (Daly and Milton, 2005).

These cultures were initially developed during the 1970s by the Phillips Petroleum Company for the obtention of the so-called “Single Cell Protein”, an alternative protein source for human and animal feed to be generated by microbial biomass from a cheap carbon source (methanol). With the advent of recombinant DNA technology and the regular cost increase of fuel derived products such as methanol, a proprietary expression system based upon the generated knowledge was developed in collaboration with the Salk Institute (La Jolla, California). *P. pastoris* expression system, now widely spread, relies on two key features of the yeast: (i) the production of a high density culture growth and (ii) the use of its strong, tightly regulated, methanol-inducible promoter (Higgins and Cregg, 1998).

Since *P. pastoris* is a Crabtree negative yeast, respiratory metabolism will be featured even under glucose surplus (Porro *et al.,* 2005). This limits the formation of ethanol, increases the yield in biomass from glycerol or glucose and enables the obtention of highly concentrated cell broths that can be then induced.

The growth of *P. pastoris* on methanol is controlled by the gene encoding the enzyme alcohol oxidase (AOX). This enzyme catalyses the oxidation of methanol according to the scheme illustrated in Figure 3 and generates hydrogen peroxide and formaldehyde. Formaldehyde is toxic to the cell, and is eliminated by two different pathways. The oxidative pathway degrades formaldehyde to formic acid, coupled to the reduction of glutathione, and then further to carbon dioxide. An alternative pathway transforms

formaldehyde into dihydroxiacetone that enters the glycolitic pathway of the cell. All these reactions take place inside peroxisomes, whose proliferation is induced by methanol (Cereghino and Cregg, 2000).

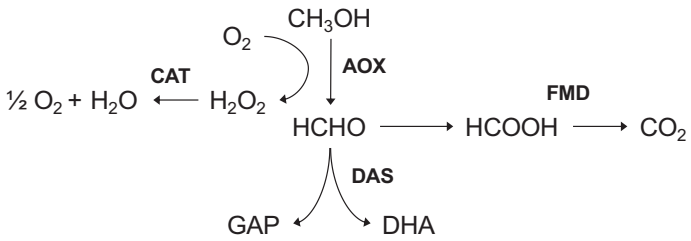


Figure 3. Scheme of methanol metabolism in *P. pastoris*. AOX: alcohol oxidase, CAT: catalase, FMD: formiate dehydrogenase, DAS: dihidroxiacetone synthase.

Orthologous genes to the *P. pastoris* AOX gene have been identified in other methylotrophic yeast such as *Hansenula polymorpha*. The regulation of AOX promoter is not yet fully understood, but it is generally accepted that it occurs at transcriptional level in response to a mechanism of catabolite repression in the presence of glucose, glycerol and ethanol. The promoter is derepressed in absence of such carbon sources and strongly induced in the presence of methanol (Cereghino and Cregg, 2000). There are two codifying genes for the alcohol oxidase activity in *P. pastoris*, named AOX1 and AOX2. A strain that contains both is able to grow quickly on methanol and this phenotype is known as Mut^+. The AOX1 gene encodes a more active enzyme; the strains in which AOX1 has been deleted grow slowly on methanol showing a Mut^S phenotype. If both genes are deleted, the mutant is viable but uncapable to grow on methanol, corresponding to a Mut^- phenotype (Cregg *et al.*, 1989).

When induced by methanol, the amount of alcohol oxidase in a Mut^+ strain can reach up to 30% of the total protein content. When a heterologous gene is cloned under the control of the AOX promoter in *P. pastoris,* its recombinant product is synthesised by the cell at a growth associated rate when the cells grow on methanol as sole or combined carbon source (Higgins and Cregg, 1998). Other characteristics of *P. pastoris* that make it convenient for the obtention of recombinant protein are:

(i) The possibility to achieve high cell density cultures, as even with low productivities per cell, a good protein concentration will be produced in the culture,

(ii) In contrast to prokariotic systems such as *Escherichia coli*, *P. pastoris* does not produce endotoxins and this lowers purification cost for proteins for pharmaceutical applications,

(iii) As a eukaryotic cell, it can introduce post-translational modifications required for the correct functional activity of proteins coming from higher organisms, but in contrast to *S. cerevisiae*, *P. pastoris* will not hyperglycosylate the protein.

(iv) Operation and maintenance cost are lower than those required by cell culture systems such as CHO cells or insect cells.

(v) Recombinant enzymes produced in *P. pastoris* have obtained the Generally Regarded As Safe (GRAS) consideration by the American Food and Drug Administration (Pais et al., 2003)

The expression system was licenced and improved by the company Invitrogen that provides different sets of vectors for cloning, fusion to secretion signals such as *S. cerevisiae* α-mating factor (α–MF), epitope recognition sites and affinity chromatography ligands. Recombinant strains are usually obtained by integration of the expression cassette. This ensures the stability of the strain and avoids the use of antibiotic for selective pressure.

2.2.2 Fermentation guidelines for *P. pastoris* cultures

An important know-how on the use of *P. pastoris* for the obtention of recombinant proteins has been generated. Most of it derives from manufacturer's guidelines, and multiple examples of culture medium design, environmental conditions, set-up and feeding strategy development for *P. pastoris* cultures can be found in the literature (Cos *et al.*, 2006a).

This yeast is usually grown in a basal medium proposed by Invitrogen (Brierley *et al.*, 1990) rich in salts and basic and trace elements to support high cell density growth, although some alternatives have been formulated (Stratton *et al.*, 1998; Boze *et al.*, 2001). The standard operating protocol for *P. pastoris* cultures (Invitrogen Corp., 2000) is based on three sequential carbon-limited growth and induction phases. The culture starts as a batch on glycerol, followed by a transition fed-batch phase on glycerol to achieve high-cell density, and finally a fed-batch induction phase on methanol to induce protein expression. A mixed transition feed of glycerol and methanol can be inserted for improved adaptation on methanol. Glycerol has also been supplemented during the whole induction phase, but its consumption rate must be controlled as excess glycerol can partially repress AOX1 promoter. Methanol concentration must also be kept below inhibitory levels (3-4 g/L). Alternative non-repressing carbon sources such as sorbitol have been successfully used to improve protein productivity in mixed methanol feeds (Jungo *et al.*, 2007).

The pH value of the culture medium is usually set to slightly acidic values (5-6) to reduce proteolytic activity and avoid salt precipitation. Nitrogen limitation has also been related to increased proteolytic activity but this can be limited if NH_4OH is used to control pH or supplemented during growth (Cos *et al.*, 2006b). *P. pastoris* optimal growth temperature is 30ºC. Protein synthesis stops above 32ºC and considerable heat is generated during fed-batch stage on methanol, due to its high heat of combustion (Jungo *et al.*, 2007). For this reason, cultures can be incubated at 28ºC if efficient cooling systems cannot be ensured. Combined glycerol/sorbitol feeding during induction phase on methanol is also useful to reduce overheating.

At present, around 430 proteins have been successfully expressed in *P. pastoris*, more than 300 extracellularly. The yields vary from a few mg/L up to several g/L for some proteins, depending on the characteristics of the strain and the protein, and the operating conditions and feeding strategy during fermentation. A significant number of proteins synthesised in *P. pastoris* are being tested for their use in clinical trials. Modified *P. pastoris* strains are now available that mimic human glycosylation patterns for an improved therapeutic efficacy of the recombinant proteins produced in the yeast (Hamilton *et al.*, 2006).

2.2.3 Process optimisation of *P. pastoris* cultures

Exhaustive optimisation of the production of particular recombinant proteins in *P. pastoris* is frequently pursued beyond standard operating protocols, either to produce higher amounts of a given protein, or to enhance the performance of the process for industrial implementation and scale-up. Usual targets for process optimisation are protein concentration, measured in terms of specific production per cell, together with biomass concentration, whose combination provides volumetric productivity, which is known to correlate with the variable costs of a bioprocess (Thiry and Cingolani, 2002; Maurer *et al.*, 2006).

Other qualitative factors such as solubility or presence of degradation products must also be taken into account (Table 1). In the case of *P. pastoris*, high cell densities are achieved by means of combined substrates feeding policies. Since protein production is growth related in recombinant strains using AOX promoter, optimisation is directed towards improved specific productivity per cell (resulting in increased final protein concentration). Efficient control of proteolytic activity is also important (Jahic *et al.*, 2003; Cos *et al.*, 2006a).

Table 1. Target variables for optimisation of recombinant protein production.

Target variable	Units	Goal
Glycosylation pattern	-	Improved product quality
Protein concentration	g/L	Increased purification yield, diminished purification steps
Protein solubility	-	Easier purification, purification reduction cost
Protein specific productivity	mg/(g·h)	Improved process performance
Protein volumetric productivity	mg/(L·h)	Improved process performance
Protein yield on substrate	g/g	Decrease in cost of materials
Size distribution	-	Improved product quality – improved product titer

Heuristic optimisation of heterologous protein production involves statistical culture media design together with fine tuning of environmental variables such as pH, temperature and dissolved oxygen following combinatorial experimental designs. Several examples of design of experiments applied to *P. pastoris* cultures have been published. Holmes and coworkers (2009) used a multi-well mini-bioreactor for the establishment of high-yielding environmental conditions that were transferred to the bench-top bioreactor cultivation of a GFP producing *P. pastoris* strain. Optimal temperature, dissolved oxygen concentration and pH conditions were determined. Kupcsulik and Sevella (2005) applied a similar scheme to the optimisation of pH and oxygen conditions of HSA production.

Although biomass concentration can be easily measured, this is mainly achieved offline by regular sampling. Since product concentration cannot be determined online either, dissolved oxygen and methanol concentration in the broth are the most usually measured variables for monitoring and control purposes. In this way, glycerol and methanol feeding rates, and aeration controls (stirring and airflow) are the most frequent operating variables in order to maintain optimal, preset μ (Cos *et al.*, 2006a).

2.2.4 Modelling and operating strategies in *P. pastoris* cultures

Besides empirical approaches, different strategies based on mathematical models have been applied in order to improve the performance of this system and define the most suitable feeding policy. In general, simple, kinetic, black-box models are used to describe biomass specific growth and substrate consumption rates. Growth kinetics of *P. pastoris* cultures are commonly described by models based on the Monod equation or inhibition growth models, such as Haldane model. Both kinetic expressions are shown in (14-15).

Monod model

$$\mu = \frac{\mu_{max} \cdot C_{GLYC}}{K_M + C_{GLYC}} \quad (14)$$

Haldane model

$$\mu = \frac{\mu_{max} \cdot C_{MET}}{K_S + C_{MET} + C_{MET}^2 / K_I} \quad (15)$$

Monod model is suitable to describe growth on glycerol. A typical inhibition profile is well documented at high methanol concentrations, and many authors fit growth profiles on methanol to uncompetitive inhibition models such as Haldane model (Ren *et al.*, 2003; Zhang *et al.*, 2000; 2004; Mendoza-Muñoz *et al.*, 2008). Other references use Monod model also for growth on methanol, but in a limited range of concentrations (D'Anjou and Daugulis, 1997; Curvers *et al.*, 2002; Jahic *et al.*, 2002; Oliveira *et al.*, 2005; Jungo *et al.*, 2007). Typical growth parameters of different *P. pastoris* strains are detailed in Table 2.

Table 2. Parameters values for kinetic models of *P. pastoris* growing on glycerol and methanol.

$\mu_{max,gly}$ 1/h	$Y_{X/GLY}$ g/g	$\mu_{max,met}$ 1/h	$K_{M,met}$ g/L	$Y_{X/MET}$ g/g	References
0.26	0.42	-	0.005	0.61-1.73	D'Anjou and Daugulis, 2001
0.26	0.40	-	0.10	-	Jahic *et al.*, 2002
-	-	0.05	-	-	Zhou and Zhang, 2002
0.09	0.62		-	1.02	Pais *et al.*, 2003
-	-	0.01	-	-	Ren *et al.*, 2003
-	-	0.07	-	1.20	Zhang *et al.*, 2004
-	0.24-0.33	-	-	0.29-0.37	Charoenrat *et al.*, 2005
0.18	0.50	0.05	0.22-0.40	-	Cos *et al.*, 2006b

$\mu_{max,met}$ values provided correspond to Mut^+ strains for all references.

Different control strategies are then applied depending on whether methanol concentration can be determined online or not. If methanol concentration cannot be measured, a predefined feeding rate profile can derived from mass balance equations and set in an open-loop structure to maintain a particular specific growth rate (Sinha *et al.*, 2003). Substrate consumption rate can also be related to oxygen consumption rate, and then the feeding is tuned to achieve a constant dissolved oxygen concentration (D'Anjou and Daugulis *et al.*, 2001; Charoenrat *et al.*, 2005) or a specific consumption rate (Chen *et al.*, 1996; Zhang *et al.*, 2000; Ren, *et al.*, 2003).

If methanol can be determined online, then constant methanol concentration or a more direct control of specific growth rate can also be implemented. In these cases, different control schemes have been employed, from Proportional–Integral–Derivative control algorithms (Cos *et al.*, 2006b), to more sophisticated strategies such as dynamic programming for optimal specific growth rate (Kobayashi *et al.*, 2000), or a minimal-variance-controller and a semi-continuous Kalman-filter and online methanol consumption rate-based feeding profile (Hang *et al.*, 2008).

In any case, the desired specific growth rate, μ, must be either previously determined experimentally, or estimated to optimise protein production or instant protein productivity, π. The following section will revise the different approaches used in this second case.

2.2.5 Modelling recombinant protein production

In equation (4), π stands for recombinant protein production specific rate (specific productivity), measured in terms of mg of product per biomass and time. In order to solve product concentration in this system, the identification of specific productivity rate, π, is required. Although π has been assumed to be constant and determined experimentally in very specific process conditions (D'Anjou and Daugulis, 1997), it is generally understood that in systems driven under constitutive promoters such as pAOX, protein productivity depends on the specific growth rate (Zhang *et al.* 2000; Cunha *et al.* 2004; Jungo *et al.* 2007). Protein production is considered to be growth-associated in *P. pastoris* cultures and hence,

$$\pi = f(\mu) \qquad \textbf{(16)}$$

However, the exact function $f(\mu)$ of this dependence and its biological basis for recombinant protein production is still subject to discussion (Maurer *et al.*, 2006). In this way, several strategies have been employed for the description of the relationship between specific growth rate and protein productivity.

The simplest expression is to establish a linear function with μ.

$$\pi = a \cdot \mu \quad \textbf{(17)}$$

This approach was used by Mendoza-Muñoz and coworkers (2008) in a partially structured model that divides total biomass into substrate, peroxisome and remaining intracellular and non-viable fractions. Losses of cell viability and generation of proteases are correlated to methanol uptake rate. The model is able to simulate the production of different proteins in *P. pastoris* although through fitting of more than 16 parameters to experimental data. Also Zhang and coworkers (2000) used this equation to establish a growth model in order to select a suitable methanol feeding strategy to maintain the optimal μ for the production of the *Clostridium* botulinum neurotoxin heavy chain fragment C. The authors suggest that this relationship might be protein specific.

Expressions such as Luedeking-Piret model have been used to approximate protein production (Ren *et al.*, 2003; Sinha *et al.*, 2003, Zhang *et al.*, 2004).

$$\pi_{LP} = a \cdot \mu + b \quad \textbf{(18)}$$

This expression represents a model of partial growth-dependence, as it introduces a non-growth related contribution, b (Picó-Marco, 2004). This simple kinetic model of protein expression was combined by Ren and coworkers (2003) with a small metabolic network to predict HSA production by *P. pastoris*. The model was based on stoichiometric balances of carbon source, biomass, ATP and NADH. After parameter fitting, the model achieved correct prediction of biomass growth and substrate consumption rates. Protein production rate was also accurately estimated, but after 140 h cultivation, a metabolic shift was detected and protein concentration and productivity decayed. Sinha and coworkers (2003) and Zhang and coworkers (2004) refer a similar expression for interferon-τ production (a = 0.75, b=0.003-0.005), though valid only in a range of dilution rates.

Some authors provide evidence that secreted protein productivity is saturated at high μ, and different mathematical strategies are used in order to represent this relationship. Examples can be found in which it has been approximated by a several step linear function (Ohya *et al.*, 2005). The parameters of this function are experimentally

determined in continuous cultures growing at different specific growth rates. Different non-linear models for f(μ) have also been used. Zhang and coworkers (2004, 2005) used different polynomial regressions on μ and C_x (cell concentration) to describe and optimise protein productivity:

$$\pi = a + b\cdot\mu + c\cdot\mu^2 + d\cdot\mu^3 + e\cdot e^{-\mu} \quad \textbf{(19)}$$

$$\pi = a + b\cdot C_X + c\cdot\mu - d\cdot C_X^2 - e\cdot\mu^2 - f\cdot C_X\cdot\mu \quad \textbf{(20)}$$

Curvers and coworkers (2002) and Maurer and coworkers (2006) approximated this function to a Monod type model to capture saturation at higher μ.

$$\pi = \pi_{max}\cdot\frac{\mu(t)}{K_q + \mu(t)} \quad \textbf{(21)}$$

Finally, other authors have reported monotonously (Kobayashi *et al.*, 2000) and non-monotonously (Potgieter *et al.*, 2010) decreasing protein productivities with increasing specific growth rates. The former approximated protein productivity to the following expression:

$$\pi = a\cdot\mu^3 - b\cdot\mu^2 + c\cdot\mu + d \quad \textbf{(22)}$$

All these strategies find considerable limitations. As previously reported (Picó-Marco, 2004), the kinetic function π may be monotonously increasing, decreasing, or non-monotonous. This largely compromises the applicability of optimization and control strategies to different recombinant *P. pastoris* strains. When constant specific productivities are calculated, the resulting model is only valid for a very limited range of operating conditions. On the other hand, non-linear models require advanced methods to solve the optimisation problem for control purposes, and empirical fitting of multiple parameters. The non-applicability of data obtained from steady-state conditions (continuous) to define the relationship between μ and π in fed-batch fermentations has also been reported (Kobayashi *et al.*, 2000).

Furthermore, complete decoupling from growth can occur and steep descents in protein productivity have been detected even during apparent steady-state growth in chemostat conditions. With this regard, Plantz and coworkers (2005) described a steep

descent in protein productivity is detected during growth in chemostat conditions. The authors show that these changes are associated to a change in the energetic state of the yeast and relate them to variations in the ratio ATP/ADP. Such phenomenon has also been related to an increase in proteolytic processes (Zhou and Zhang, 2002).

2.3 Glycosidases

2.3.1 Definition

Glycosides are compounds in which a sugar (the glycon) is linked to a non-sugar moiety (the aglycon or genin) such as an aliphatic alcohol, a phenol group or a carboxylic acid. (See Figure 4 for some examples).

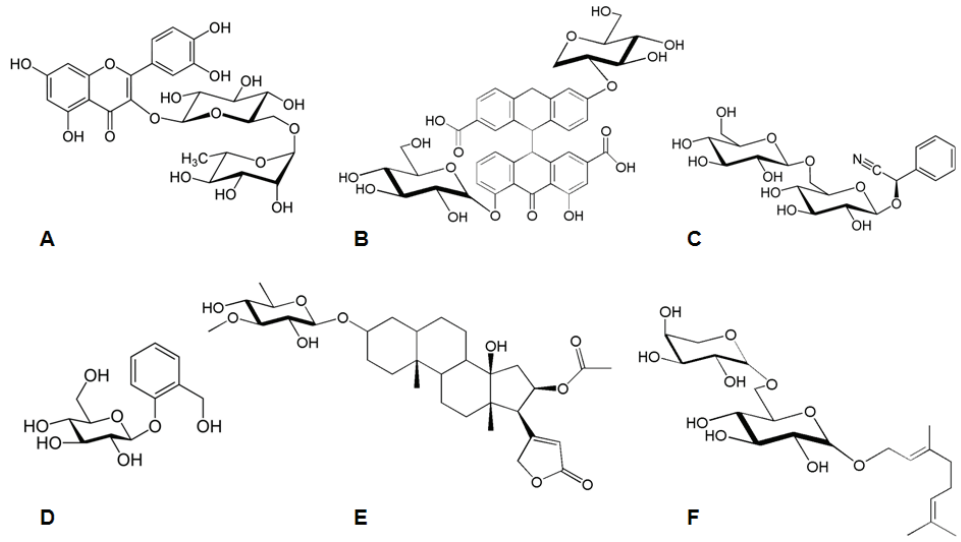


Figure 4. Some examples of O-glycosides found in nature: A – rutin (flavonoid glycoside), B – sennoside (anthraquinone glycoside), C – amygdalin (cyanogenic glycoside), D – salicin (phenolic glycoside), E – oleandrin (cardiac glycoside), F- geraniol glycoside (terpenic glycoside).

These compounds are stored in living organisms and can be activated by enzymatic hydrolisis when the chemical is selectively released from the glycon, as its solubility in water decreases and the volatility is enhanced (Berger, 2009). Glycosidic bonds link the anomeric carbon of the sugar to the aglycon through an ether group (O-glycosides), a secondary amine (N-glycosides) or a thiol (S-glycosides).

Glycosidases, also named glycoside hydrolases (Enzyme Class 3.2.1.–), are enzymes able to catalyse the hydrolysis of O- and S-glycosidic bonds between carbohydrates or between a carbohydrate and a non-carbohydrate moiety. These enzymes are widespread in nature and can be found in microorganisms, plants and animals. Their functions relate to biomass conversion processes (Singhania, 2009), cell signaling (Zhao *et al.*, 2008b), fruit maturation and aroma release (Sarry and Günata, 2004), nutrient acquisition and pathogenic mechanisms (Varghese *et al.*, 1992).

2.3.2 Mechanism of action

Glycosidases are classified as inverting or retaining glycosyl hydrolases, depending on the anomeric configuration of the released glycon, which can be modified or conserved in relation to the substrate.

Inverting glycosidases catalyse the hydrolysis through the direct displacement of the aglycon by water, whereas retaining glycosidases generate a covalent glycosyl-enzyme intermediate in a double displacement mechanism that results in the net retention of the stereochemistry (Figure 5). In both cases, the enzyme active groups are some of its side-chain carboxylic residues (Wan *et al.*, 2007).

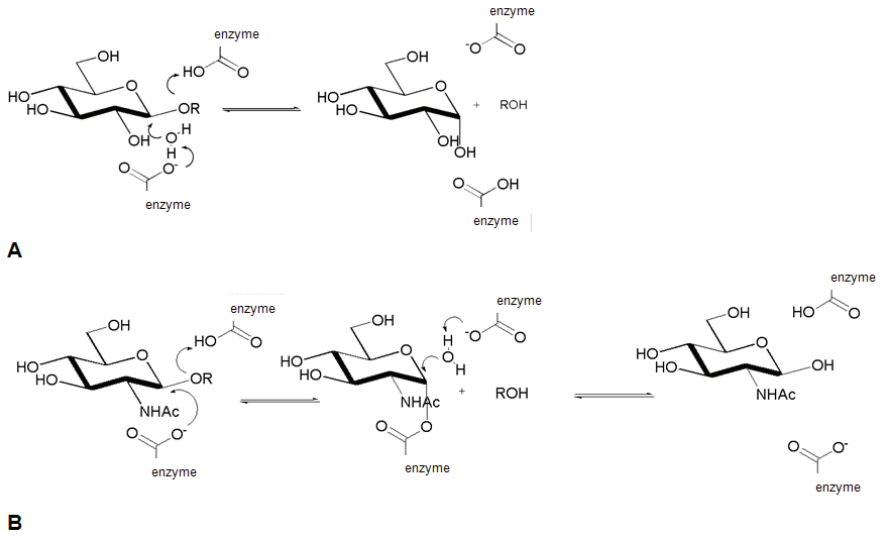


Figure 5. Mechanism of action of inverting (higher panel) and retaining (lower panel) glycosidases.

Besides hydrolysis, retaining glycosidases can under suitable conditions catalyse the formation of glycosidic bonds and can be used to synthesise oligosaccharides and different glycoconjugates including alkylated glycosides (Rémond *et al.*, 2002). In these reactions, the incubation of the glycoside with the enzyme generates a glycosyl-enzyme intermediate that can be intercepted by an acceptor instead of being hydrolysed by water, forming a new glycoside or oligosaccharide (Crout and Vic, 1998) as depicted in Figure 6.

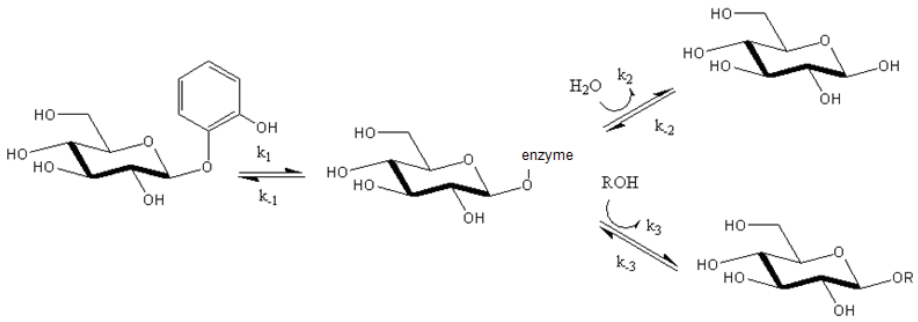


Figure 6. Glycosidase catalysed glycoside transfer reaction scheme (adapted from Crout and Vic, 1998).

With activated substrates that are irreversibly cleaved ($k_{-1}\approx0$), such as p-nitrophenol (pNP) derivatives, the reaction will proceed under kinetic control, and the proportion between hydrolysis and transfer product will be determined by the ratio $k_2(H_2O)/k_3(ROH)$. This mechanism of action is also known as transglycosylation (Crout and Vic, 1998).

2.3.3 Applications of glycosidases

Besides their biological importance, glycoside hydrolases have been extensively studied for their application in (i) biomass conversion processes such as bioethanol production, (ii) flavour enhancement in beverages such as musts and wines (Gil *et al.*, 2005; Sánchez-Torres *et al.*, 1998) and (iii) food detoxification processes (Birk *et al.*, 1996).

The degradation of cellulose and its conversion to glucose are limiting steps for obtaining fuels and chemicals from polysaccharides found in disposed biomass. Cellulase complex, formed by an endoglucanase, an exoglucanase and a β-glucosidase, is responsible for the enzymatic hydrolysis of cellulose in which β-glucosidase activity is usually the rate limiting step.

Many plant glycosidic flavor precursors of interest for food processing and quality improvement have been identified so far. A number of reports have been published that demonstrate that flavor precursors can be exploited as a potential reserve aroma in different fresh and processed vegetables and fruits as well as in wines by the action of different glycosidases. The precursors, found in plants and fruits as O-β-D-glucosides or O-diglicosides, can be released with the use of the suitable combination of glycosidases (Pérez-González *et al.*, 1993; González-Candelas *et al.*, 1995; Berger, 2009). The scheme of terpene (in this case, geraniol) and sugar release lead by the sequential action of α-L-arabinofuranosidase and β-D-glucosidase is shown in Figure 7.

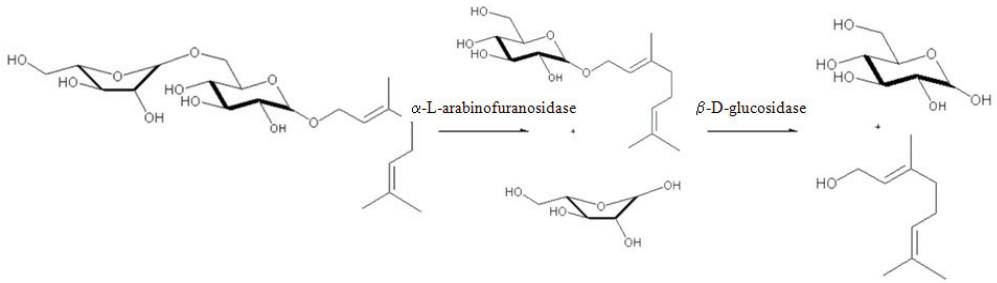


Figure 7. Release of terpenes through the action of glycosidases.

In addition, glycosidases can be used to catalyse the formation of glycosides and oligosaccharides, through transglycosylation and reverse hydrolysis reactions as detailed above. These compounds have wide applications in chemical, food and pharmaceutical industries, as biodegradable surfactants (Rantwick *et al.*, 1998), prebiotics (van den Broek *et al.*, 1999) and therapeutic agents (Zopf and Roth, 1996). Chemical synthesis for the obtention of glycoconjugates typically involves acid catalysis and protection-deprotection strategies in multistep reaction sequences, resulting in a mixture of isomers and low yields (Crout and Vic, 1998).

Conversely, enzymatic synthesis is a good substitute as it provides absolute control of anomeric configuration with lower number of reaction steps. Glycosyl transferases (EC 3.2.4.–) catalyse regioselective glycosylation reactions with high transfer yields. However, these enzymes require nucleotide-sugars (*e.g.* UDP-glucose) to be used as substrates. This, together with the low availability of these enzymes, has hindered their application. In contrast to glycosyl-transferases, glycosidases of microbial origin are readily available and can be easily overproduced in recombinant hosts, and are thus a good alternative for the synthesis of glycosidic bonds.

2.3.4 α-L-arabinofuranosidase and β-N-glucosidase

α-L-arabinofuranosidase (Abf, EC 3.2.1.55) and β-N-glucosidase (Bgl, EC 3.2.1.21) are glycosidases able to release terminal non-reducing sugar residues from their bound aglycons. Whereas Abf is able to hydrolyse α-L-1,2-, α-L-1,3-, and α-L-1,5-arabinofuranosyl moieties and is active on small synthetic substrates, arabino-oligosaccharides and polymers such as arabinans, arabinogalactans, and arabinoglucuronoxylans (Margolles-Clark *et al.*, 1996), Bgl acts on terminal β(1-4) bonds linking two glucose or glucose-substituted molecules such as cellobiose, showing specificity for a variety of β-D-glycoside substrates (Janbon *et al.*, 1994).

Applications

Both enzymes have received much attention for their application in the bioethanol industry, as supplemented Bgl to cellulase preparations increases the rate of cellulose degradation through cellobiose hydrolysis, and Abf acts synergistically with other hemicellulases and pectic enzymes in the hydrolysis of lignocelluloses from agricultural residues (Numan and Bhosle, 2005). Besides, Abf and Bgl have found broad applications in other agro-industrial subsectors, such as flavour boost in food industry by monoterpene release from aroma precursors (Sánchez-Torres *et al.*, 1996; Manzanares *et al.,* 2003), staling prevention in bread storage by pentosan liberation (Lavermicocca *et al.,* 2000), fruit juice clarification and enhancement of pulp delignification in paper industry (Margolles-Clark *et al.*, 1996).

More recently, a growing interest has been paid to their use in biosynthetic applications for the obtention of medicinal compounds, such as metastatic and anticarcinogenic drugs (Bae *et al.*, 2000) and in the synthesis of oligosaccharides, glycoconjugates and alkyl and aminoglucosides by stereoselective transglycosylation (Bhatia *et al.*, 2002; Wallecha and Mishra, 2003). Among others, the synthesis of multiple oligosaccharides using the transglycosylation activity of β-glucosidase from *Agrobacterium tumefaciens* has been successfully achieved (Prade *et al.*, 1998). Recombinant arabinofuranosidase from *Thermobacillus xylanilyticus* has also been reported to produce different alkylated arabinosides in high yield (Rémond *et al.*, 2002).

Origin

In the last decades, numerous arabinofuranosidases and β-glucosidases of microbial and plant origin have been identified, isolated and characterised. The properties of different cloned β-glucosidases are reviewed by Bhatia and coworkers (2002). The majority of purified Bgl are glycoproteins with molecular weights ranging from 40 to 140 kDa and heterogeneous pH, temperature and substrate specificity. Abf activity is known to be produced, mostly extracellularly, by plants and microorganisms from the genera *Aspergillus*, *Bacillus*, *Ruminococcus*, *Streptomyces* and *Trichoderma* (Flipphi *et al.*, 1993; Sánchez-Torres *et al*.,1996; Vincent *et al.*, 1997). The specificity of each Abf for natural substrates varies greatly. The purification and characterisation of several Abf has been accomplished from *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus sojae*, *Aspergillus terreus and Trichoderma harzianum* (Numan and Bhosle, 2005).

Commercial preparations for Abf and Bgl are usually obtained from yeasts and filamentous fungi such as *A. niger*. In general, fungal glycosidases are able to process a wider range of substrates than the ones isolated from plants (Günata *et al.*, 1990). Fungal glycosidases are produced in high yields, usually secreted in complex mixtures together with other related enzymatic activities.

2.4 Immobilisation of enzymes

2.4.1 Definition and advantages

An immobilised enzyme can be defined as an enzyme that is physically confined to a defined region in space, is insoluble in its surrounding medium, retains its catalytic activity, and can be repeatedly and continuously used. Immobilisation techniques are applied not only to enzymes, but also to microbial, vegetal and mammalian cells among others. The purpose of immobilisation is to increase the stability and catalytic activity of the enzymes in denaturing conditions, such as extreme pH, high temperature, or organic solvents (Fernández-Lafuente *et al.,* 2000). This strategy also allows increasing the productivity as the catalyst can be concentrated and reused in continuous operation

(Arroyo, 1998). In general, immobilisation involves the physical or chemical binding of the enzyme to a macroscopic, insoluble and inert support. Biocompatible and inert materials such as organic polymers and silicas are usually preferred for this purpose (Schuleit and Luisi, 2003).

The main advantages of enzyme immobilisation are:

(i) The ease of separation of the reaction broth from catalyst, by filtration or centrifugation, with an improved control of the reaction time, avoidance of the need for inactivation of the catalyst, and reduction of enzyme losses into product.

(ii) Repeated use of the enzyme over production cycles, diminishing the total production cost in enzymatic reactions.

(iii) Ease of use in continuous reactors, in packing beds, when the support shows an adequate compressibility.

(iv) More suitable conditions for enzymatic activity, as immobilised enzymes are found in reactor configurations that enhance diffusion and concentration, in comparison to more diluted systems in which soluble enzymes are used.

(v) Enhancement of the enzyme resistance and stability, as the support can be used to improve its resistance over time to severe work conditions, such as extreme pH, high ionic strength and temperature, organic solvents and detergents. These conditions also reduce the risk of bacterial contamination.

The latest is one of the most interesting points, as it implies an improvement in the catalytic behaviour of the enzyme. By means of the union to a support, the tertiary configuration of the enzyme becomes more rigid, and conformational changes will be hindered. As a consequence, the irreversible deployments that conduce to losses in the catalytic activity will be reduced (Blanco, 1979).

The main disadvantage of immobilisation is that a third of the enzymes require cofactors such as NADH, NADPH or acetyl-CoA for their activity. Cofactor recycling is difficult to achieve *in vitro*. Also enzymatic immobilisation is not applicable to all enzymes, as the effect of attachment to a support can reduce enzymatic activity depending on the three dimensional architecture of the enzyme and how the active site is exposed.

In general, the use of immobilised enzymes is recommended when the cost of the enzyme is high, continuous operation is envisaged, higher stability of the enzyme is required, due to reuse, reaction medium or conditions, and also when product yield is a concern.

Some examples of the commercial use of immobilised enzymes are:

(i) Glucose-isomerase to produce fructose from sugar, in more than 8 million tons per year.

(ii) Penicillin-acylase for the synthesis of 6-amino-penicillanic acid with a production of more than 7500 tons per year.

(iii) Lipases for the interesterification of fats.

(iv) Aminoacylase or the purification of L-amino acids by optical resolution of the racemic.

(v) Sensors and electrodes for cholesterol, ethanol, galactose, glucose, lactose or urea determination in analytical and clinical diagnosis (Scouten et al., 1995).

(vi) Supports for HPLC, therapeutic agents, in organic synthesis and classical enzymology applications (Liu et al., 1997).

2.4.2 Immobilisation methods

Four methods are possible to immobilise enzymes: physical adsorption, cross-linking, covalent binding and entrapment (Arroyo, 1998).

Adsorption is a quick and simple method to fix an enzyme to a support, by weak physicochemical interactions such as Van der Waals interactions or hydrogen bonds. Inorganic supports can be used, such as alumina, ceramics, diatomeous earth, silica and also ionic exchange resins, such as Amberlite, Dowex or Sephadex. Common organic supports are activated charcoal, cellulose (carboximethylcellulose, diethyl-amino-ethyl cellulose) or starch. The enzyme will be weakly bonded to the support, and can be easily desorbed depending on pH and ionic strength of the surrounding medium.

Cross-linking is accomplished when chemical reagents (such as bis-diazobenzidine, 2,2-disulfonic acid or glutaraldehyde) are used to induce the formation of covalent bonds directly among the enzyme in solution. In this way, insoluble but active enzyme aggregates similar to gelatin are generated. This application is cheap as no solid support is required but diffusion problems can appear and the enzymatic activity can be damaged.

Direct entrapment can be carried out in semi-permeable membranes with pores in which the enzyme is trapped, microcapsules, as calcium alginate, carrageenan or polyacrylamide and porous fibers, made from cellulose, nylon or polysulphone. In this case, the enzyme is not chemically but physically immobilised in the support. The enzyme is mixed with the monomer and then polymerisation is induced. As the enzyme is physically included in the support, the main hurdles of this technique is desorption by leakage and diffusion limitations.

Covalent binding generates a permanent carbon-carbon bond between the enzyme and the support when some functional groups of the enzyme, such as amino, carboxyl, hydroxyl and sulfhydril groups react with a functional group found in the support (López-Gallego *et al.*, 2005). The support can be further activated with chemical agents such as BrCN, carbodiimides or glutaraldehyde. Covalent binding minimises losses per leakage or desorption but can affect the activity of the enzyme if the active site is

concerned. Covalent binding can also be performed by multipoint attachment when the support presents a wide contact surface for interaction with the enzyme and multiple groups that can be activated, such as hydroxyl or amine groups (Guisán *et al.,* 1993; Guisán, 1998).

2.4.3 Molecular sieves as supports for enzyme immobilisation

Molecular sieves, such as mesoporous materials or zeolites, show interesting properties as chemical catalysts, supports and matrixes, due to their high specific surface (around 1000 m^2/g) together with their mechanical and chemical resistance, their capacity to show hydrophobic or hydrophilic behaviour and to distinguish molecules according to their size. They are particularly interesting for enzyme immobilisation as they can establish physical and chemical interactions and show good mechanical properties (Ciesla and Schüth, 1999). However, the inclusion of enzymes into the most common molecular sieves, zeolites, is limited to small enzymes such as lysozyme due to the small pore diameter of these materials, around 20 Å (Corma *et al.*, 2001, 2002)

Mesoporous sieves, with pores of around 30-70 Å, are more suitable for enzyme immobilisation as their mean pore (around 100 nm) is in the range of the average size of enzymes. MCM-41 materials were synthesised for first type by Mobil, are used as catalysers of reactions of electronic transfer, with substrates of high volume, conductant polymers and metal complexes, and can be characterised by electronic transmission microscopy, N_2 adsorption-desorption techniques and X-ray diffraction. Molecular sieves are characterised by an irregular pore distribution with a very shallow size range, which can be tuned from 20 to 200 Å distributed along channels that can be several nanometers long. Particularly, MCM-41 structure is characterised by a hexagonal channel distribution. These materials have been used as supports for the inclusion of many different substances, such as antibiotics or sugars, and also in separation by size exclusion of proteins and other biomolecules in agri-food and pharmaceutical industry (Han *et al.,* 1999).

This type of materials exhibit additional advantages, such as the possibility to modify the opening pore size of the channels that can be reduced after the retention of the hosted molecule for enzyme entrapment. In addition, the uniform pore distribution enables the study of enzyme theoretical kinetic models, which requires the isotropy of the support (Balkus and Diaz, 1996). Finally, they display higher physical, chemical and mechanical resistance and lower sensitivity to microbial contamination, fouling and degradation in comparison to organic supports.

UVM-7 materials for enzyme immobilisation

Bimodal organosilicas present a hierarchical pore architecture constructed from soldered small mesoporous (ca. 3 nm) nanoparticles that generates a non-ordered system of interparticle large pores (ca. 40-60 nm). UVM-7 materials are a variant of tradicional mesoporous sieves type MCM-41. They show an unusual architecture, formed by a network of nanoparticles creating a bimodal system of macro and mesopores as depicted in Figure 8.

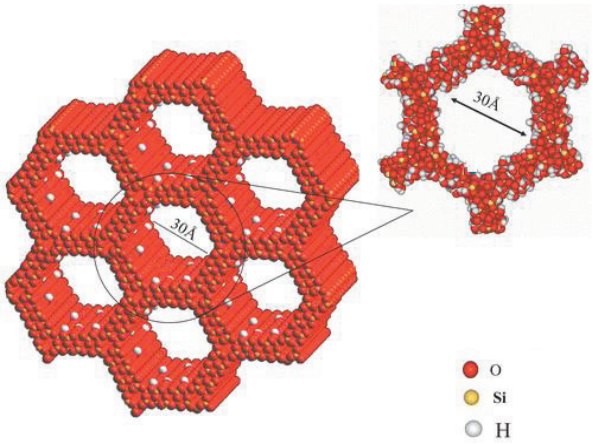


Figure 8. Scheme of the structure of the materials UVM-7.

The isothermal curves show two well-defined, characteristic steps (El Haskouri *et al.*, 2001, 2002a) illustrated in Figure 9 and typical of these hierarchical porous nanosized organosilicas (HNPO).

The standard size of the particles is around 12-17 nm and the mesopores are cylindrical and uniform. The specific surface is high, comparable to aerogels and mesocellular foams. By extrusion or molding, monoliths can be produced, in which UVM-7 and derivatives keep their surface and adsorptive properties (El Haskouri *et al.*, 2002b).

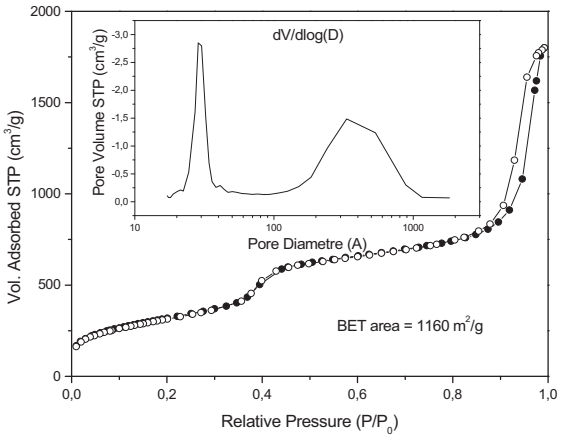


Figure 9. N_2 isothermal curve and pore size distribution for UVM-7 material. Two distribution of pores exist, mesopores (30 Å), and macropores (500 Å). The specific surface and volume are 1160 m^2/g and 2.5 cm^3/g, respectively.

Some electronic transmission micrographs are shown for these materials in Figure 10 in which the shape and pore distribution can be observed.

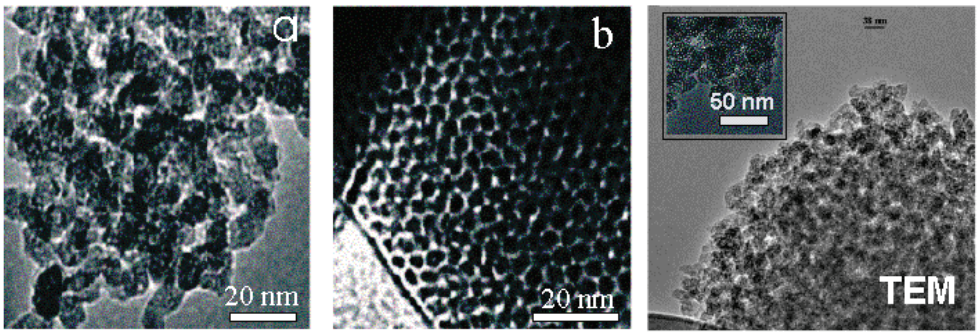


Figure 10. Electronic transmission micrographs. Panel a. UVM-7, panel b. MCM-41 and panel c. UVM-7. Source: Dr. P. Amorós, ICMUV.

In this way, UVM-7 materials are attractive as immobilisation supports, as they share the advantages of traditional mesoporous materials, but show shorter channels. This can improve the accessibility of enzyme for an increased concentration of the immobilised catalyst and also favour the diffusion of substrates (Lei *et al.*, 2002).

III. Materials and Methods

3.1 Materials

3.1.1 Genetic materials

Oligonucleotides

The oligonucleotides shown in Table 3 were used as primers for PCR reactions and were purchased from Thermo-Fisher Scientific (Ulm, Germany) as concentrated HPLC purified freeze-dryed stocks.

Table 3. Oligonucleotides used in this study.

Name	Sequence
3-AOX	5´-GCA AAT GGC ATT CTG ACA TCC-3´
5-AOX	5´-GAC TGG TTC CAA TTG ACA AGC-3´
Abf-EcoRI	5´-AAG AAT TCA AAA TGT TCT CCC GCC GAA ACC TCG-3´
Abf-XhoI	5´-TTT CTC GAG TTA CGA AGC AAA CGC CGT CTC-3´
Bgl-800	5´-TGG TCT CAA GTC CGA GC-3´
Bgl-aMF-PmlI	5´-CTC ACG TGA CCA AGA ATA TCT CCA AAG CGG -3´
Bgl-PmlI	5´-CTC ACG TGA CCA TGG AAT CAA C-3´
Bgl-SacII	5´-CTG GCC GCG GAG TTT GTG AAG ACC-3´
SP6	5´-ATT TAG GTG ACA CTA TAG AA-3'
T7	5´-TAA TAC GAC TCA CTA TAG GG-3'
α-factor	5´-TAC TAT TGC CAG CAT TGC TGC-3´

Genes

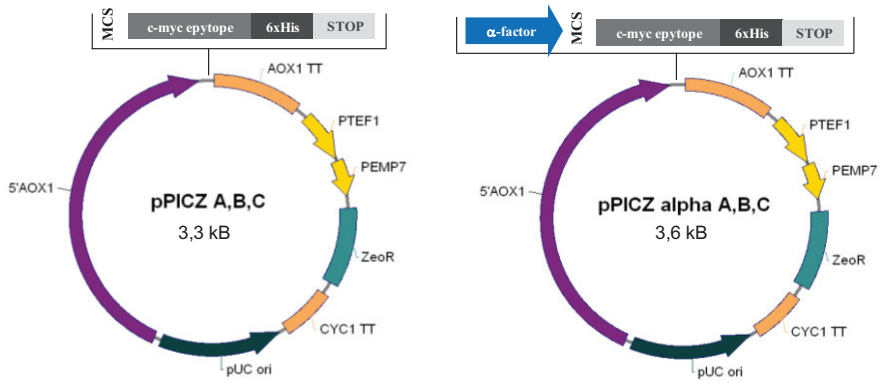
Abf and Bgl full cDNA sequences are available in GenBank database as L23502 (Flipphi *et al.*, 1993) and U16259.1 (Janbon *et al.*, 1995), respectively. Both proteins were analysed for signal peptides cleavage sites with SignalP 3.0 Server (Bendtsen *et al*., 2004) and two targeting signals were identified as shown below:

Abf_sp: MFSRRNLVALGLAATVSA-

Bgl_sp: MESTIIILSVLAAATA-

Vectors

The commercial vector pGEM-T Easy (A1360, Promega Corp, Madison, USA) was used to subclone PCR fragments. pPICz shuttle expression system for *E. coli*/*P. pastoris* (EasySelect^{R} *Pichia* Expression Kit, Invitrogen, Carlslab, USA) was used to integrate the genes under control of AOX promoter in the yeast. Precisely, the vectors pPICzA, pPICzB and pPICzαB, this later including the native *S. cerevisiae* α-factor secretion signal, were used. Their most significant features are shown in Figure 11.



5'AOX1 – Alcohol oxidase 1 (AOX1) promoter region

AOX1 TT – AOX1 transcription termination region

PTEF1 – promoter of the gene encoding the elongation factor 1 α for bleomycin resistance gene expression in yeast

PEMP7 – synthetic promoter for bleomycin resistance gene expression in bacteria

ZeoR – bleomycin resistance gene

CYC1 TT – bleomycin gene transcription termination region

pUC ori – replication origin in *E. coli*

c-myc – epitope tag for affinity purificationand detection

6xHis – epitope tag for affinitity purification and detection

α-factor – secretion signal, *S. cerevisiae* yeast α–mating factor pheromone (α-MF)

Figure 11. pPICz (left panel) and pPICzα (right panel) plasmid series map and their relevant features (Figures extracted from EasySelect^{R} *Pichia* Expression Kit User Manual, Invitrogen, Carlslab, USA).

3.1.2 Microbial strains

The strains described in Table 4 were used throughout this work.

Table 4. Strains used for genetic modification and background expression controls.

Code	Species	Genotype	Reference
DH10B	*E. coli*	F$^-$ *end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*L Δ*lac*X74 Φ80*lac*ZΔM15 *ara*D139 Δ(*ara*,*leu*)7697 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) λ$^-$	Casdaban and Cohen, 1980
DH5α	*E. coli*	F- *end*A1 *gln*V44 *thi*-1 *rec*A1 *rel*A1 *gyr*A96 *deo*R *nup*G Φ80*dlac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169, *hsd*R17(rK- mK+), λ–	Meselson and Yuan, 1968
GS115	*P. pastoris*	*his4*	Invitrogen
GS115/His+ Mut^S/HSA	*P. pastoris*	HSA	Invitrogen
GS115/pPICZ /lacZ	*P. pastoris*	*lacZ*	Invitrogen
pGEMT-Abf	*E. coli*	DH5α bearing Abf PCR amplicon	This study
pGEMT-Bgl	*E. coli*	DH5α bearing Bgl PCR amplicon	This study
pGEMT-αBgl	*E. coli*	DH5α bearing αBgl PCR amplicon	This study
pPICz-Abf	*E. coli*	DH10B bearing pPICz-Abf construct	This study
pPICzB	*E. coli*	TOP10 bearing pPICzB plasmid	Invitrogen
pPICz-Bgl	*E. coli*	DH10B bearing pPICz-Bgl construct	This study
pPICzαB	*E. coli*	TOP10 bearing pPICzαB plasmid	Invitrogen
pPICzα-Bgl	*E. coli*	DH10B bearing pPICzα-Bgl construct	This study
T73-Abf	*S. cerevisiae*	T73 strain transformed with the plasmid YCAbfB (YCA_1)	Sánchez-Torres *et al.*, 1996
T73-Bgl	*S. cerevisiae*	T73 strain transformed with the plasmid YCBgl ($YCB3_5$)	Sánchez-Torres *et al.*, 1998
TOP10	*E. coli*	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rps*L(Str^R) *end*A1 λ$^-$	Casdaban and Cohen, 1980
X-33	*P. pastoris*	Wild type (phenotype: mut+)	Invitrogen
X33-Abf	*P. pastoris*	X33 yeast strain transformed with plasmid pPICz-Abf	This study
X33-Bgl	*P. pastoris*	X33 yeast strain transformed with plasmid pPICz-Bgl	This study
X33-αBgl	*P. pastoris*	X33 yeast strain transformed with plasmid pPICzα-Bgl	This study
X33-pPICz	*P. pastoris*	X33 yeast strain transformed with plasmid pPICzB	This study
X33-pPICzα	*P. pastoris*	X33 yeast strain transformed with plasmid pPICzαB	This study

3.1.3 Culture media

Recipes for culture media used in this work are shown in Table 5.

Table 5. Culture media recipes.

Short name	Composition
BMGH	100 mM K_2HPO_4/KH_2PO_4 buffer at pH 6.0, 1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 1% glycerol
BMGY	100 mM K_2HPO_4/KH_2PO_4 buffer at pH 6.0, 1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 1% glycerol, 1% yeast extract, 1% bacteriological peptone
BMMH	100 mM KH_2PO_4 pH 6.0, 1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 0.5% methanol
BMMY	100 mM KH_2PO_4/KH_2PO_4 buffer at pH 6.0, 1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 0.5% methanol, 1% yeast extract, 1% bacteriological peptone
LB	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, pH 7.0 (NaOH)
LBZ	10 g/L tryptone, 2.5 g/L NaCl, 5 g/L yeast extract, pH 7.0 (NaOH), 25 μg/mL zeocin,
Low salt LB	10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract, pH 7.0 (NaOH)
MDH	1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 2% dextrose
MMH	1.34% yeast nitrogen base, $4{\cdot}10^{-5}$% biotin, 0.5% methanol
YPD	1% yeast extract, 2% bacteriological peptone, 2% dextrose
YPDS	1% yeast extract, 2% bacteriological peptone, 2% dextrose, 1M sorbitol
YPM	1% yeast extract, 2% bacteriological peptone, 0.5% methanol

- Yeast nitrogen base contained ammonium sulphate but did not include amino acids.
- When required, liquid media compositions were formulated with 2% bacteriologic agar to obtain solid growth media.
- All media and buffers were autoclaved at 121ºC, 20 minutes except biotin, methanol and yeast nitrogen base that were filter-sterilised.

3.1.4 Mesoporous sieves

Abf and Bgl immobilisation was accomplished by using several bimodal mesoporous organosilica of the UVM-7 family containing different proportions of aminopropyl groups on the silica surface kindly provided by the group of Dr. Amorós and Dr. Beltrán (Instituto de Ciencia de los Materiales, Universitat de València). These NH_2-UVM-7 (designated NUVM-7 from now onwards) were prepared as described by El Haskouri and coworkers (1999, 2000) by using different amino silane derivatives as organic functional groups as shown in Table 6.

For UVM-7 obtention, tetraorthosilicate (TEOS) was reacted with triethanolamine at 145ºC. Surfactant agent cethyltrimethylammonium bromide (CTAB) was added and after 24 h, the precipitated solid was filtered, washed, extracted with ethanol and dried. NUVM-7 materials were produced either by co-condensation or post-synthetic impregnation of mesoporous materials. For co-condensation, amino silane derivatives were added in substitution of TEOS in an analogous procedure. For post-synthetic impregnation a procedure previously described was used (El Haskouri *et al.*, 2000).

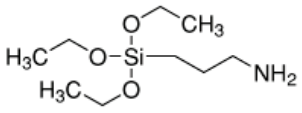
In all cases, the materials were synthesised using a proportion of 5% w/w of amino radicals in relation to TEOS. The different aminated radicals and their corresponding chemical structures are shown in Table 6 and Figure 12 respectively.

Table 6. NUVM7 supports obtained with different types of amino silane derivatives

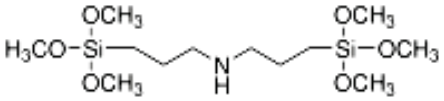
Code	Synthesis	Reference*	Aminated radical
NUVM7-1	Impregnation	09324	3-aminopropyl-triethoxysilane
NUVM7-2	Co-condensation	413356	Bis(3-trimethoxysilyl)-propylamine
NUVM7-3	Impregnation	413348	3(trimethoxysilylpropyl)-diethylenetriamine
NUVM7-4	Co-condensation	104884	N-3(trimethoxysilylpropyl)-ethylenediamine

*Sigma-Aldrich

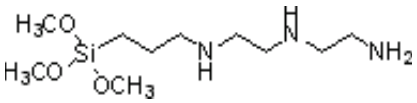
NUVM7-1



NUVM7-2



NUVM7-3



NUVM7-4

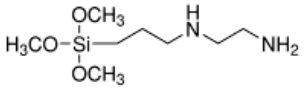


Figure 12. Amino silane derivatives used for functionalisation.

3.2 Microbial culture and transformation

3.2.1 General procedures

Unless otherwise stated, general standard molecular biology procedures were followed as described by Sambrook and Russell (2001) and Ausubel *et al.* (2010). The strains *E. coli* DH5α/DH10B were used as recipients for plasmid DNA and were transformed as described in (Hanahan, 1983). The extraction of plasmid DNA from *E. coli* was performed using the kit Qiaprep SPIN miniprep kit (Qiagen, Hilden, Germany). The extraction of total DNA from *P. pastoris* was performed as described by Querol and coworkers (1992). The extraction of plasmid DNA from *S. cerevisiae* was performed according to Brozmanova and Holinova (1988).

PCR reactions for gene cloning were performed using “Expand High Fidelity PCR System” (Roche, Penzberg, Germany) according to the manufacturer instructions. PCR reactions for screening of transformants were performed using DNA polymerase Netzyme (Need, Valencia, Spain) according to the manufacturer instructions. The purification of PCR fragments was performed using a commercial kit (PCR Cleanup, Qiagen, Hilden, Germany) according to the manufacturer instructions. Capilar DNA sequencing was performed in Secugen S.L. (Madrid, Spain). The resulting chromatograms were analysed using the program Chromas 1.3 (Griffith University, Brisbane, Australia).

Cloning and manipulation of *P. pastoris* was carried out following the manufacturers instructions (User Manual, cat no KT1740-01, MAN000042, Invitrogen, LifeSciences, Carlsbad, USA) unless stated otherwise. Protein concentration was determined as described in Bradford (1976) using bovine serum albumin as the standard. Enzyme kinetic constants were determined as described in Nelson and Cox (2000) and fitted using the software Sigma-Plot and non-linear regression analysis.

3.2.2 Microbial transformation

For *E. coli* transformation, a preparation of 10 μg of plasmid DNA was produced and introduced into competent cells by heat shock or electroporation as detailed in Sambrook and Russell (2001). The cells were recovered by incubation at 37ºC, for 45 minutes, in low salt LB medium. 100 mL were plated onto LBZ medium and incubated overnight.

For *P. pastoris* integrative transformation, 5-10 μg of plasmid DNA were digested with the restriction enzymes *Bst*XI, *Pme*I or *Sac*I, depending on the construct. After linearisation, the digests were extracted with phenol/chloroform once, precipitated with ethanol in the presence of 1/10 volume of 3 M sodium acetate and 2.5 volumes of 100% v/v cold ethanol. The solution was centrifuged to pellet plasmid DNA, washed with 80% v/v ethanol, air-dried and resuspended in 10 μL of sterile, deionised water. *P. pastoris* cells were transformed by electroporation. For this purpose, the selected strain (X-33) was cultured in 5 mL of YPD medium overnight and used to seed 500 mL of fresh YPD medium in a 2 L flask. The culture was grown overnight until stationary phase. The cells were recovered by centrifugation, in sterile conditions, and resuspended in 500 mL ice-cold sterile, deionised water. This procedure was repeated and the cells were resuspended in 250 mL ice-cold, deionised water. The cells were centrifuged once again, and resuspended in 20 mL of ice-cold, sterile 1 M sorbitol. The cells were centrifuged again, and resuspended in 1 mL ice-cold, sterile 1 M sorbitol for a final volume of 1.5 mL.

For electroporation, 80 μL of the former cells were mixed with 5 μg of linealised plasmid DNA at approximately 1 μg/μL and transferred to an ice-cold 2 mm gap electroporation cuvette (BTX, Harvard Biosciences, Holliston, USA). Without further incubation, the mix was pulsed at 1500 V for 5 ms. Immediately after pulsing, 1 mL of ice-cold 1 M sorbitol was added to the cuvette, and its contents were transferred to a sterile 15 mL tube. The cells were incubated at 30ºC without shaking for 2 hours. 100 mL were spread onto YPDS plates containing 100 μg/mL zeocin. The plates were incubated at 30ºC for 2-3 days and the colonies were isolated for further validation and characterisation of the transformants.

3.2.3 Transformant validation and characterisation

Zeo^R transformants were cultured in 5 mL of YPDZ medium and its genomic DNA was isolated as referred above. PCR amplification was carried out using either 5AOX-3AOX primers or α-factor-3AOX primer for constructs bearing α-factor signal with an initial denaturation step at 94ºC, for 2 minutes, followed by 25 cycles of 94ºC for 1 minute, 55ºC for 1 minute, and 72ºC for 1 minute, and a final extension step of 7 minutes at 72ºC.

For Mut phenotype testing, the Zeo^R transformants were picked and patched on both a MMH and a MDH plate in this order. The strains GS115/HSA (Mut^S) and GS115/lacZ (Mut^+) were used as controls. The plates were incubated at 30ºC for 2 days. Transformants were considered Mut^+ if they could grow normally on both plates, whereas they were typed Mut^S if they failed to grow or showed little growth on the MMH plate. In order to select multicopy transformants, YPDS plates containing 500, 1000 and 2000 μg/mL zeocin were prepared and used for selection. The maximal resistance to zeocin was used as indirect indicator of the copy number.

3.2.4 Activity plate screening

Activity screening was performed on solid agar plates in order to preliminarly select the transformants. For this purpose, chromogenic and fluorescent substrates upon hydrolisis by glycosidases were adapted from previous protocols (Hernández *et al.*, 2001; Fia *et al.*, 2005). For detection of Bgl activity, 4-methylumbelliferyl-β-D-glucoside (MU-Bgl) was dissolved from a stock at 50 mg/mL in dimethylformamide (DMF) at a final concentration of 10 μg/mL in McIlvaine buffer 50 mM pH 5.0. MDH and MMH plates in which the transformants had previously been grown were overlaid with this substrate containing agarose at a final concentration of 0.75%. Alternatively, 5-bromo-4-chloro-3-indolyl-β-D-glucoside (X-glu) was used to detect intracellular expression by adding 50 μL per plate from a 25 mg/mL stock in DMF. The overlayed plates were incubated at 40ºC for 15 minutes and observed in a UV-transilluminator. For detection of Abf activity, the same protocol was used but 4-methylumbelliferyl-α-L-arabinoside (MU-Abf) was used instead.

3.2.5 Expression assay in liquid flask cultures

To test expression levels, 25 mL of BMGY medium were inoculated with a single colony of the selected strain and cultured at 30ºC in a shaking incubator at 250 rpm overnight. The cells were harvested by centrifugation at room temperature, and the cell pellet was resuspended in sterile BMMY medium at a final cell density of 1.0 absorbance units at 600 nm (typically in 150-200 mL). The culture was placed in a 1 L baffled flask and cultured at 28ºC for up to 5 days. Every 24 hours, cell density was monitored and methanol was added at 0.5% v/v final concentration assuming that all previously supplied methanol had been consumed. Aliquots of 1 mL were extracted, centrifuged at 10.000·g for 2 minutes and stored for functional assay of the expression levels and protein quantitation.

3.2.6 Recombinant *S. cerevisiae* cultures

Genetically modified *S. cerevisiae* strains able to express the genes encoding the enzymatic activities β-glucosidase from *C. molischiana* (Sánchez-Torres *et al.*, 1998) and α-L-arabinofuranosidase from *A. niger* (Sánchez-Torres *et al.*, 1996) were used to produce these enzymes as controls. The strains were kindly provided by Dr. Paloma Manzanares from the Oenological Enzymes Group at the Institute of Agrochemistry and Food Technology of the Spanish National Research Council (IATA-CSIC from the Spanish title “Instituto de Agroquímica y Tecnología de Alimentos; Consejo Superior de Investigaciones Científicas”). Their names and genetic background are detailed in Table 4.

The growth of the modified *S. cerevisiae* strains (T73/YCA1, T73/YCB3$_5$) was carried out in YPD medium (see Table 5) in a 50 L bioreactor. For this purpose, a preinoculum was prepared in 50 mL of YPD medium in a 250 mL flask containing 1 µg/mL cycloheximide (Sigma-Aldrich, St. Louis, MO) and was cultured for 24 h at 30ºC in a rotary incubator at 200 rpm. The preinoculum was used to seed a 500 mL inoculum of the same medium in a 2 L inoculum flask that was cultured for 24 hours at 30ºC and 200 rpm. The inoculum was used to seed 30 L of culture medium supplemented with 1 µg/mL of cycloheximide that had been previously sterilised in a 50 L bioreactor (Biostat

C, Braun-Biotech, Sartorius, Goettingen, Germany). The culture was grown at 30ºC, pO_2 30% and pH 6.0 for 48 hours. Microbial growth was controlled by measurement of absorbance at 600 nm and dry cell weight.

3.2.7 *μ24*-microreactor cultures

A microbioreactor (μ-24 system, MicroReactor Technologies, Pall Corporation, New York, USA) was used to carry out batch cultures of X33-Abf and X33-Bgl strains (i) to identify optimal environmental conditions and (ii) for modelling purposes. This device is a 6x4 wells micro-bioreactor where each individual well has a temperature sensor, a thermal heat conductor, optical pH and dissolved oxygen sensors, and a sparge membrane for gas blending (air, O_2, CO_2, and N_2) and pH control (NH_3 and CO_2). Environmental variables can be measured and controlled individually in each well. A single use sterile cassette holding 24 baffled, 10 mL cylinders was used in each run.

A factorial design of experiments (DoE) was performed in this μ-24 system in order to identify the optimal environmental conditions for recombinant α-L-arabinofuranosidase and β-glucosidase expression.

The wells were filled with 3 mL of BMMY medium and inoculated at an initial optical density of 0.2 with an overnight culture of the selected clones in BMGY medium. The μ-24 cassette was sealed with a sterile gas permeable membrane. Agitation was set at 500 rpm. Batch fermentations were monitored for dissolved oxygen, pH, and temperature at 1 minute increments for 48 hours. Process controls were modified by MicroReactor software from a labtop computer. Methanol was added at 1% v/v final concentration whenever exhausted attending to dissolved oxygen consumption. In parallel, cell growth was measured offline by absorbance readings at 600 nm. The factorial design and the conditions applied in each well are shown in Table 7 and 8. Cultures contained in each well were recovered and assayed for Abf and Bgl activities as detailed in Section 3.3.4.

Table 7. Factors and levels considered for DoE.

Factor	Level 1	Level 2	Level 3
Temperature (ºC)	26	28	30
Airflow	0	1	2
pH	5	5.5	6

Table 8. Matrix of the functional design.

Experiment	Factors			Variables		
	X_1	X_2	X_3	Temperature	Airflow	pH
1	-1	-1	-1	26	0	5.0
2	-1	0	0	26	1	5.5
3	-1	1	1	26	2	6.0
4	0	-1	-1	28	0	5.0
5	0	0	1	28	1	6.0
6	0	1	0	28	2	5.5
7	1	-1	1	30	0	6.0
8	1	0	-1	30	1	5.0
9	1	1	0	30	2	5.5

A series of batch cultures were performed in order to compare the behaviour of different recombinant *P. pastoris* strains and growth conditions. For each culture series, preinocula were grown o/n in 3 mL YPDZ from a single colony taken from a solid culture in the same medium. From these, inocula were prepared by seeding 20 mL BMGH cultures that were incubated o/n in a rotary shaker at 28ºC. The μ24 wells were filled with 4 mL of BMGH and/or BMMH as indicated in the following section and inoculated to achieve an initial absorbance at 600 nm of around 0.5. Antifoam 204 (Sigma-Aldrich) was added in each well at a final concentration of 1 μL/mL. The μ-24 cassette was sealed with a sterile gas permeable membrane. All these operations were conducted in sterility in a laminar hood cabinet.

Orbital agitation was set-up at 500 rpm and temperature controlled at 28ºC. No CO_2 or ammonia was required to maintain pH of buffered media that was stable at 6.0. Air was bubbled in all wells at 5 vvm until saturation was reached. No further oxygen was supplied to the growing cultures.

Batch cultures were continuously monitored for dissolved oxygen saturation, pH, and temperature at 1 minute intervals for up to 40 hours. Process controls were modified if required by MicroReactor software from a labtop computer. Offline analysis was performed by extracting 200 μL per well in sterile conditions that were used to determine cell, protein and substrate concentration. Whenever necessary, replica wells were used for convenient sampling.

3.3 Enzyme procedures and biochemical analysis

3.3.1 Enzyme recovery and purification

Abf was purified from 500 mL culture supernatant collected after centrifugation at 4000·g for 10 minutes and filtration through 0.45 μm. The supernatant was diluted 6 times in Tris-HCl 10 mM pH 7.5, adsorbed onto 50 mL of Sephadex A-50 resin (GE-Healthcare) previously equilibrated and washed with 200 mL of the same buffer and eluted in 10 mL fractions with a gradient of NaCl between 0 and 1 M in Tris-HCl 10 mM pH 7.5. The fractions containing Abf activity were collected and diafiltered against McIlvaine buffer pH 4.0 using a disposable crossflow cassette (Vivaflow, 50 kDa, Sartorius AG, Göttingen, Germany). The concentrate was lyophilised for long-term storage in a laboratory freeze-dryer (Beta 1-8, Martin Christ, Osterode am Harz, Germany).

Bgl was extracted from the cell-bound fraction of 1 L cultures with Tris-HCl 10 mM pH 7.5, 1 M NaCl, dialysed and purified by ionic exchange adsorption onto 50 mL of Q-Sepharose resin (GE-Healthcare) following the same protocol described above for Abf. The fractions containing Bgl activity were collected and the purified extract was concentrated and diafiltered against McIlvaine buffer pH 5.0 using a disposable crossflow cassette (Vivaflow, 100 kDa, Sartorius AG, Göttingen, Germany) and freeze-dried for long-term storage.

Lysozyme (mucopeptide N-acetylmuramoylhydrolase) was purchased as a lyophilised powder from Sigma-Aldrich (50.000 units/mg protein, Ref. L7001, Sigma-Aldrich, St Louis, USA).

3.3.2 Protein gel electrophoresis

Crude extracts and purified proteins were analysed by polyacrylamide gel electrophoresis (PAGE). Sodium dodecyl sulphate (SDS) denaturing, reducing PAGE was used to determine the molecular weight of proteins present in the sample. For this purpose, 8% acylamide gels containing 0.1% SDS were prepared, run and stained with Coomassie Blue as described by Ausubel and coworkers (2010). Samples were resuspended and heated at 95ºC for 10 minutes in buffer containing β-mercaptoethanol (5%), bromophenol blue (0.02%), glycerol (10%), SDS (2%) in Tris-HCl pH 6.8 (0.065 M).

Non-reducing gel electrophoresis was used to evaluate enzymatic activity in semi-native conditions. Samples were resuspended in buffer containing bromophenol blue (0.02%), glycerol (10%), SDS (1%) in Tris-HCl pH 6.8 (0.065 M) and run as described previously. After electrophoresis, gels were incubated for 2 h in a solution containing Triton X-100 2.5% to eliminate SDS and then equilibrated in McIlvaine buffer 50 mM (pH 4.0 in the case of Abf and pH 5.0 in the case of Bgl) for 1 hour. MU-Abf (for Abf) or MU-Bgl (for Bgl) was added at 10 μg/mL final concentration and gels were incubated at 37ºC for 2-3 hours. Activity was measured by detecting the luminescence under 365-nm UV-light.

3.3.3 Deglycosylation

Endoglycosidase H (EndoH, NewEngland Biolabs, Ipswich, Massachussets, USA) was used as indicated by the supplier to cleave high mannose residues from protein samples. 1-20 μg protein were suspended in buffer containing SDS (5%) and dithiothreitol (DTT, 0.4 M) and incubated at 100ºC for 10 minutes. Denatured, reduced samples were suspended in buffer containing sodium citrate pH 5.5 (0.5 M) and incubated at 37ºC with 500 mU of EndoH for 1-3 hours. Native proteins were deglycosylated without pretreatment in buffer containing sodium citrate pH 5.5 (0.5 M) and incubated at 37ºC with 500 mU of EndoH o/n.

3.3.4 Enzymatic reactions

Abf and Bgl activities were determined spectrophotometrically at 405 nm, by reaction with p-nitrophenyl-arabinofuranoside (pNPA) and p-nitrophenyl-β-glucoside (pNPG). A mixture of 100 μL of sample, 50 μL of pNPA/pNPG 0.2%, and 150 μL McIlvaine buffer, pH 4.0/pH 5.0 respectively, were incubated at 30ºC for 15 minutes. Reactions were stopped by addition of Na_2CO_3 0.25 M. When solid biocatalyst was present, absorbance was measured after centrifugation of solid residue. One activity unit is defined as the amount of pNP in μmol released per minute in these conditions.

Determinations of optimal pH and temperature, inhibition and specificity assays were carried out in McIlvaine buffer. Thermal stability assays were carried out incubating the enzyme in McIlvaine buffer, at pH 4.0 (for Abf) or pH 5.0 (for Bgl) for 15 minutes at 30ºC, 40ºC, 50ºC, 60ºC, 70ºC and 80ºC, followed by a standard activity with pNPG 0.2% assay after cooling.

3.3.5 Subcellular location

The cellular location of Abf and Bgl in the recombinant strains was investigated following the protocol described previously by González-Candelas and co-workers (1995). Briefly, cells and culture supernatant were separed by centrifugation. Extracellular enzymatic activity was directly measured in supernatant samples. Cells were washed with saline solution 0.9% and resuspended in McIlvaine buffer pH 4.0 (for Abf) and pH 5.0 (for Bgl) to calculate cell-bound activity. An equal volume of 0.45 μm diameter zirconia/silica beads was added to the sample and the cells were disrupted in a vortex apparatus in cycles alternating shaking and ice-cooling phases. Cell extracts were separated from debris by centrifugation at 13.000·g for 15 minutes at 4ºC. Intracellular activity was calculated as the difference between the activity of this fraction and cell-bound activity.

3.3.6 Substrate specificity

The specificity of Abf and Bgl towards different substrates was analysed as described by Sánchez-Torres and coworkers (1998). Substrate concentration used in each case is detailed in Table 9. The reactions were prepared using 150 μL of McIlvaine buffer at pH 4.0 for Abf and pH 5.0 for Bgl, 50 μL of 6 times concentrated substrate and 100 μL of sample and incubated at 30ºC, for 15 minutes. In the case of Bgl, the release of glucose was measured using a commercial kit (D-glucose hexokinase, ref K-GLUHKR, Megazyme International, Wirklow, Ireland). In the case of Abf, the release of arabinose was measured by HPLC as detailed in Section 3.3.9.

Table 9. Substrates and concentration used to determine substrate specificities.

Enzyme	Linkage	Substrate	Concentration
Abf	α(1,6)	pNPA	5 mM
Abf	α(1,6)	Arabinogalactan (oat spelts)	5 g/L
Bgl	β(1,4)	pNPG	5 mM
Bgl	β(1,4)	pNPX	5 mM
Bgl	α(1,6)	pNPA	5 mM
Bgl	β(1,4)	Salicin	20 mM
Bgl	β(1,4)	Cellobiose	50 mM
Bgl	β(1,4)	Methyl-β-glucoside	20 mM
Bgl	α(1,4)	Maltose	50 mM
Bgl	β(1,4)	Lactose	50 mM
Bgl	β(1,2)	Saccharose	50 mM

3.3.7 Transglycosylation assays

Transglycosylation reactions were carried out as described by Rémond and coworkers (2002) and related conditions for optimisation. Briefly, 6 mM pNPG/pNPA (from a stock at 30 mM) was incubated in the presence of 25% methanol for 1 hour at 45ºC in McIlvaine buffer, pH 4.0/pH 5.0 respectively, in a final volume of 500 μL unless otherwise stated. The reactions were stopped by heating at 100ºC for 5 minutes in a

dry bath, cooled on ice, centrifuged for 1 minutes at 10.000·g and filtered by 0.2 μm for further analysis.

In the case of immobilised enzymes, 100 μL-250 μL aliquots of homogeneous suspensions of the different biocatalysts were used in McIlvaine buffer at pH 4.0 for Abf activity and pH 5.0 for Bgl activity. The reactions were heat-inactivated by boiling at 100ºC for 5 minutes and cooling on ice. The products were evaluated by HPLC as described below, after centrifugation and filtration of supernatant through 0.22 μm.

3.3.8 Isoflavone release

Soy isoflavones were extracted from soy milk following the protocol described in Ortiz (2008). Soy milk was freezed at -20ºC to induce coagulation of protein, centrifuged and filtered through 0.45 μm. The filtrate was fractionated in C18 reverse phase cartridges (C18-RP, Sep-Pack 500 mg, Waters Corp, Milford, Mass., USA). 40 mL of filtrate were adsorbed in a single cartridge previously activated with 10 mL methanol and 20 mL of deionised water. After washing with 10 mL of water and 10 mL of pentane, the glycosylated isoflavones were eluted with 10 mL methanol and dried by rotavaporation at low pressure and 45ºC. The resulting oil was resuspended in 3 mL of McIlvaine buffer 75 mM pH 4.5. The determination of free and glycosylated isoflavones was performed by HPLC as detailed below. Daidzein, daidzin, genistein. genistin, glicitein and glicitin (Sigma Aldrich) at suitable concentrations were used as internal standards.

The enzymatic release of free isoflavones was assayed by incubating 400 μL of isoflavone extract with 100 μL of enzyme sample and incubating at 40ºC for 4 h. The reactions were quenched by boiling at 100ºC for 5 minutes and cooled on ice. The supernatant was extracted with 500 μL of acetonitrile and evaluated by HPLC after filtration through 0.22 μm.

3.3.9 HPLC analysis

The products of enzymatic reactions were analysed in the Laboratory of Analytical Techniques at Biopolis S.L. Transglycosylation products and substrate specificity assays were determined using a Rezex ROA-Organic Acid H+ column (8%, Phenomenex), at 0.5 mL/min, in isocratic mode, using H_2SO_4 2.5 mM as unique eluent. The column was kept at 60ºC. A refraction index detector was used, at a temperature of 40ºC. Samples of 50 μL were injected in all cases. Arabinose, glucose, methanol, methyl-β-glucoside, pNP-arabinoside and pNP-glucoside at suitable concentrations were used as standards in transglycosylation reactions. In the case of glucose release, the reactions were analysed in parallel using a commercial kit (D-glucose assay kit, ref. K-GLUC, Megazyme, Wiklow, Ireland).

The products of isoflavone release assays were run on a Sunfire C18 column, 5 μm, 4.6x150 mm (Waters) at 1 mL/min, using double milliQ water + 0.1% TFA (eluent A) and acetonitrile + 0.1% TFA (eluent B) in gradient mode. The gradient established was for 5 min, 90% A; for 20 min, a linear decrease from 90% A to 60% A; for 10 min, a linear decrease from 60% A 0% A; for 2 min, constant flow of 100% B; finally for 5 min, a linear increase from 0% A to 90% A. The oven was kept at 30ºC. A photodiode detector was used and the absorbance was measured at 254 nm. 30-50 μL of samples were injected. Daidzein, daidzin, genistein. genistin, glicitein and glicitin at suitable concentrations were used for standard curve calibration.

3.4 Constraint-based model definition and analysis

3.4.1 Constraint-based model definition

The constraint-based model is mathematically defined by the following set of equations:

$$\mathrm{MOC} = \begin{cases} \mathrm{N}\cdot v = 0 & \text{(a)} \\ \mathrm{D}\cdot v \geq 0 & \text{(b)} \end{cases} \qquad \textbf{(23)}$$

In this way, the metabolic network is defined by *m* metabolites and *n* reactions. The reactions are assembled in the stoichiometric matrix, N, in which the a_{ij} element is the stoichiometric coefficient of the i^{th} metabolite in the j^{th} reaction. The matrix equation (23a), where v is the vector of reaction rates or fluxes, implies that internal metabolites are at steady-state and that the effect of dilution is disregarded. The inequality (23b) considers the irreversibility of some reactions. D is a diagonal matrix with $D_{ij} = 1$ if the flux is irreversible and null otherwise. Only flux vectors that fulfill equations (23a) and (23b) are considered valid cellular states.

3.4.2 Structural analysis

Structural analysis refers to the intrinsic properties of the metabolic network. In order to determine if the metabolic network is well-conditioned, the calculation of the condition number corresponding to the stoichiometric matrix must be performed. The condition number is given by the expression:

$$C(G) = \left\|G\right\| \cdot \left\|G^{\#}\right\| \qquad \textbf{(24)}$$

Where ||G|| stands for the matrix norm and $(G^T)^{\#}$ is the pseudo-inverse of the stoichiometric matrix (Stephanopoulos, 1998). A requirement for well-conditioned system is that the condition number is comprised between 1 and 100 (Stephanopoulos, 1998; Cakir *et al.*, 2007).

The calculation of the EM of the system has been carried out with the programs Metatool 5.0 (Pfeiffer *et al.*, 1999; Schuster *et al.*, 2000) and FluxAnalyser 4.0 (Klamt *et al.*, 2003) running under MATLAB (The MathWorks Inc., Natick, USA). Both programs are based on the enumeration of the extreme rays of the polyhedral cone represented by the solution space of the constraint-based model (23). Every elementary mode can be related to a particular flux distribution and to a macroscopic reaction equivalent. The reaction equivalents are obtained through the expressions (25) and (26):

$$\begin{aligned} v_S &= N_S \cdot v \\ v_P &= N_P \cdot v \end{aligned} \qquad \textbf{(25)}$$

$$K = \begin{pmatrix} -N_S \\ N_P \end{pmatrix} \cdot E \qquad \textbf{(26)}$$

Each of these macroscopic equivalents provides a theoretical yield value, corresponding to the stoichiometric ratio between a particular product (*e.g.* biomass) and a substrate.

3.5 Flux calculations

3.5.1 Flux units

The flux through each reaction of the network is defined as the mass flow exchanged in each of these reactions (amount of substrate that is being transformed per biomass unit and per hour). Thus, the flux of a particular metabolic reaction is the specific rate at which this reaction occurs and the flux distribution of the system is the vector formed by all reaction rates. All fluxes are expressed in mmol/(g·h) units with the exception of biomass growth rate, μ, which is calculated in Cmmol/(g·h). μ is converted from h^{-1} units (g/(g·h) taking into account the apparent molecular weight of yeast biomass calculated in Cmol/g from its elementary composition (*e.g.* a specific growth rate of 0.1 h^{-1}, for a culture with an elementary composition of biomass of $CH_{1.83}N_{0.11}O_{0.65}$ – this is, 25.77 g/Cmol, corresponds to 3.88 Cmmol/(g·h)).

3.5.2 External flux calculation

Cell concentration was calculated from absorbance readings at 600 nm using the following correlation (Solà *et al.*, 2004):

$$C_x \text{ (g/L, dcw)} = 0.22 \cdot OD_{600} \qquad \textbf{(27)}$$

Protein concentration was measured by denaturing polyacrylamide gel electrophoresis gel (SDS-PAGE), Coomassie blue staining and densitometry analysis, and correlated to enzymatic activity that was measured as detailed previously in Section 3.3.4.

Substrates (glycerol, methanol) and products (ethanol, pyruvate, citrate) were determined by HPLC by the Laboratory of Analytical Techniques at Biopolis S.L. HPLC analysis were run on a Rezex ROA-Organic Acid H+ column (8%, Phenomenex), at 0.5 mL/min, in isocratic mode, using H_2SO_4 2.5 mM as unique eluent. The column was kept at 60ºC. A refraction index detector was used, at a temperature of 40ºC. 50 μL of sample were injected in all cases. Methanol, glycerol and ethanol at suitable concentrations were used as standards for calibration. Alternatively, glycerol was measured using a biochemical kit (Glycerol Assay Kit, Abcam, Cambridge, UK).

Cumulative consumption rates of glycerol and methanol, and production of biomass were calculated from measured concentrations. Volumetric consumption and production rates were determined by linear regression fitting through the data points within the selected regions and normalised to the corresponding biomass concentration.

Oxygen uptake rate was calculated from dissolved oxygen saturation index in accordance to the following equations:

$$\frac{dC_{O2}}{dt} = OTR - OUR$$
$$\frac{dC_{O2}}{dt} = k_{La} \cdot (C_{O2}^{sat} - C_{O2}) - OUR \quad \textbf{(28)}$$
$$OUR = \frac{dC_{O2}}{dt} - k_{La} \cdot (C_{O2}^{sat} - C_{O2})$$

Where:

C_{O2}: dissolved oxygen concentration in the liquid phase, in mol $O_2 \cdot L^{-1} \cdot h^{-1}$.
OTR: oxygen transfer rate from the gas phase to the liquid phase, in mol $O_2 \cdot L^{-1} \cdot h^{-1}$.
OUR: oxygen consumption rate by the microbial culture, in mol $O_2 \cdot L^{-1} \cdot h^{-1}$.
K_{La}: global coefficient of oxygen transfer in the liquid phase, in h^{-1}. A k_{La} value of 32.4 h^{-1} was used, corresponding to μ24 device operating with no oxygen or air supply at 500 rpm (Isett *et al.*, 2007).
C_{O2}^{sat}: dissolved oxygen concentration in saturating conditions, in mg $O_2 \cdot L^{-1}$. This parameter is determined by temperature and pressure (Lewis *et al.*, 2011).

3.5.3 Flux calculation methods: MFA, FS-MFA and PS-MFA

MFA has been performed in the case of determined systems following the standard procedures described in Stephanopoulos (1998). For this purpose, the programs Metatool 5.0 (Pfeiffer *et al.*, 1999; Schuster *et al.*, 2000) and Fluxanalyser 4.0 (Klamt *et al.*, 2003) have been used.

FS-MFA and PS-MFA were carried out both for determined and underdetermined systems using MATLAB routines as described by Llaneras and coworkers (2009, 2010).

PS-MFA has been used to estimate the measured and non-measured metabolic fluxes in several scenarios. Some details on this formulation are given below. All the computations have been performed with MATLAB, and YALMIP toolbox (Lofberg, 2004) was used to perform the optimisation steps required by PS-MFA.

Model and measurements constraints

The stoichiometric and irreversibility constraints forming the model (MOC) given in (28) are considered together with measures (MεC) of some extracellular fluxes that are incorporated as additional constraints in a constraint-satisfaction problem (29):

$$\mathrm{MOC}=\begin{cases}\mathrm{N}\cdot\nu=0 & \text{(a)}\\ \mathrm{D}\cdot\nu\geq 0 & \text{(b)}\end{cases}$$

$$\mathrm{M}\varepsilon\mathrm{C}=\begin{cases}v_m=w+\varepsilon_1-\mu_1+\varepsilon_2-\mu_2\\ 0\leq\varepsilon_1,\mu_1\\ 0\leq\varepsilon_2\leq\varepsilon_{2,\max}\\ 0\leq\mu_2\leq\mu_{2,\max}\end{cases} \qquad \textbf{(29)}$$

In (29), the vector v_m represents the actual metabolite concentrations and w the measured values that may differ due to errors and imprecision (uncertainty). This uncertainty is represented by the vectors of slack variables ε and μ. These parameters are used as non-negative decision variables in order to formulate the problem in linear-programming terms.

Possibility

Each candidate solution of (29) can be denoted as δ={v, w, *ε's*, *μ's*} in a decision space Δ. The basic building block of possibility theory is a user-defined possibility distribution $\pi(\delta)$: $\Delta \rightarrow (0,1)$. This function defines the possibility of each solution δ in Δ, ranging between impossible (π=0) and fully possible (π=1). The solution is the space of fluxes and associated uncertainty satisfying both constraints, model and measuraments (30). Among different possible choices, a simple way to define possibility is using a linear cost index such as (31):

$$\delta = M\varepsilon C \cap MOC \tag{30}$$

$$J(\delta) = \alpha \cdot \varepsilon_1 + \beta \cdot \mu_1 \tag{31}$$

The possibility of each solution δ can be then defined as follows:

$$\pi(\delta) = \exp(-J(\delta)) \qquad \delta \subset \Delta \tag{32}$$

where α and ß are row vectors of user-defined, sensor accuracy coefficients. In this way, the results can be interpreted as "v_m = w is fully possible; the more v_m and w differ, the less possible such situation is".

Representation of uncertainty

Two pairs of slack variables have been chosen to represent each measurement: ε_2 and μ_2 define an interval of fully possible values, and ε_1 and μ_1 penalise values out of it (with weights α and β). Hence, in all computations the uncertainty of each measurement has been represented as follows:

(i) Full possibility (π=1) is assigned to values with less than ±5% of deviation.

(ii) Larger deviations are penalised, so values with a deviation equal to ±20% have a possibility of π=0.1, and those with a deviation equal to ±10% have a π≈0.5.

(iii) Uncertainty has been considered symmetric, and thus α=β.

This representation is achieved choosing particular bounds ($\varepsilon_{2,max}$ and $\mu_{2,max}$) and weights (α and β) for each measurement: (i) implies that $\varepsilon_{2,max}=\mu_{2,max}=0.05\cdot w$, and (ii) defines α, noticing that $0.2\cdot w=\mu_{1,20\%}+\mu_{2,max}$, then $\alpha=-\log(0.1)/(0.2-0.05)/w$.

Note that the joint possibility of two variables $\delta_1,\delta_2\in\Delta$ can be expressed as the product of two univariate ones, $\pi(\delta_1,\delta_2)=\pi(\delta_1)\pi(\delta_2)$, and equivalently, $J(\delta_1,\delta_2)=J(\delta_1)+J(\delta_2)$. In other words, the joint possibility of two non-fully possible event is lower than the independent possibilities, *e.g.*, if $\pi(v_4=3)$ is 0.9 and $\pi(v_7=2)$ is 0.5, the joint $\pi(v_4=3\cap v_7=2)$ is 0.45.

PS-MFA flux estimation

The simplest estimate is given by the solution of the maximum possibility (minimum-cost) solution $\mathbf{v_{mp}}$ in δ_{mp} of the constraint satisfaction problem (29), which can be obtained solving a simple linear programming problem (LP):

$$J^{min}=\min_{\varepsilon,\mu,v} J \quad \text{s.t. } \{\mathcal{MOC}\cap\mathcal{MEC} \tag{33}$$

However, a point-wise estimate is insufficient if multiple solutions are reasonably possible, so an estimated interval is in general better. The interval of values with conditional possibility higher than γ for a given flux or metabolite can be computed solving two LP's:

$$v_{i,\gamma}^{m}=\min_{\varepsilon,\mu,v} v_i \quad \text{s.t.}\begin{cases}\mathcal{MOC}\cap\mathcal{MEC}\\ J-J^{min}<-\log\gamma\end{cases} \tag{34}$$

The upper bound would be obtained by replacing minimum by maximum. Intervals of conditional possibilities higher than a given value can also be computed by solving two linear programming problems (*e.g.*, γ=0.8, 0.5 and 0.1).

3.6 Consistency analysis based on PS-MFA

PS-MFA can also be applied to evaluate the degree of consistency between a given model and a set of experimental measurements. The most possible solution of the constraint-satisfaction problem (29) has an associated degree of possibility:

$$\pi^{mp} = \exp(-J^{min}) \quad \textbf{(35)}$$

This value, π^{mp} in (0,1), grades the consistency between model and measurements. A possibility equal to one must be interpreted as complete agreement, while lower values imply that there is certain degree of error in measurements, model or both, which severity depends on how the uncertainty has been defined.

In MFA problems the consistency between model predictions and measurements can be estimated calculating the vector v that fulfils MOC and minimises the sum of errors, accounting for small deviations e_m. Assuming that e_m is a normally distributed variable with zero mean, with a variance-covariance matrix F, and only linear equality constraints are considered, then the residual is a stochastic variable that follows a χ^2 distribution. Otherwise, the χ^2 test cannot be used, although the residual will still provide a qualitative indication of consistency. Consistency can also be approached as a possibilistic constraint satisfaction problem. A flux distribution that fulfils the constraints and is compatible with the measurements will be fully possible. As measurements errors have to be considered for the flux distribution to fulfil the model of constraints, the possibility of this flux distribution will be lower (Llaneras *et al.*, 2009).

For every possible flux distribution, three different indicators of consistency have been used:

(i) The variance-weighted sum of residuals (φ was used as a qualitative indicator of consistency as the assumptions of the χ^2 test are not applicable due to the inequality constraints introduced in MOC.

(ii) The possibility of the most possible flux vector was used as defined in and is preferred due to the limitations of the φ calculation (Llaneras *et al.*, 2009).

(iii) A complementary error index ($\varepsilon_{\pi=1}$) was used, calculated as the percentage of error which has to be considered to achieve a possibility equal to 1.

3.7 Ordinary least-squares regression and cross-validation

3.7.1 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) was used for parameter fitting in the model developed in Chapter VI. The goodness-of-fit of the OLS regression was assessed with the coefficient of determination R^2 and the normalised root-mean-square error (RMSEn). The statistical significance of each parameter was evaluated with p-values for p=0.01 and p=0.05.

3.7.2 Cross validation

A cross-validation analysis was performed to evaluate the predictive capacity of the model with independent data sets. Two types of cross-validation were used: 8-fold and leave-one-out. In both cases, the original dataset was partitioned in folds, the R^2 and RMSEn indicators were computed to assess goodness-of-fit, and the results for each fold were combined by average to produce single indicators. In 8-fold the original sample was randomly partitioned into 8 subsamples; a single subsample was retained as the validation data for testing the model, and the remaining 7 subsamples were used as training data. The cross-validation process was then repeated 8 times, with each of the 8 subsamples used exactly once as the validation data. The process can be repeated multiple times with new 8-fold partitions. In leave-one-out cross validation the original sample data was partitioned in 34 subsamples, using one observation as validation, and the 33 remaining ones as training data. This was repeated such that each observation in the sample was used once as the validation data. Notice that leave-one-out is equivalent to a 34-fold cross validation.

3.8 Immobilisation methods

3.8.1 Material and biocatalyst characterisation

Materials and biocatalyst samples were characterised by the group of Dr. Amorós and Dr. Beltrán by X-ray powder diffraction techniques in a Seifert 3000TT h–h diffractometer using Cu-Kα radiation.

For TME, samples were gently ground in dodecane, and microparticles were deposited on a holey carbon film supported on a Cu grid and analysed in a JEOL JEM-1010 instrument operating at 100 kV. 29Si MAS NMR and N_2 adsorption–desorption isotherms (Micromeritics, ASAP2010 analyser) were obtained from porous samples degassed for 5 h at 120ºC and 1026 Torr prior to analysis.

3.8.2 Immobilisation procedures

Adsorption and covalent binding of glycosidic enzymes was performed following the two step protocol described previously (Blanco *et al.*, 1989; Fernández-Lafuente *et al.*, 2000). Briefly, Abf and Bgl were solved and gently stirred in 10 mM McIlvaine buffer pH 5 containing 100 mg of NUVM-7 for 1 hour. When less than 5% initial activity was detected in supernatant, 1-ethyl-3-(3-dimethyl-aminopropyl) carbodiimide (EDC) was added to 10 mM final concentration and the coupling reaction was developed under gentle agitation for up to 90 min. The support was collected by centrifugation and washed several times with 1 M NaCl, McIlvaine buffer, pH 5 until no remaining activity was detected in the supernatant. For co-immobilisation, Abf and Bgl were simultaneously incubated under saturating conditions for the selected NUVM-7 supports. Immobilisation yield and relative activity were estimated as follows:

- Immobilisation yield (%) = 100 x (total enzyme activity - non-immobilised activity) / total enzyme activity
- Activity yield (%) = 100 x immobilised activity / (total enzyme activity - non-immobilised activity)

IV. Expression of two glycosidases in *Pichia pastoris*

This chapter describes the cloning and heterologous expression in the methylotrophic yeast *P. pastoris* of two genes encoding glycosidic enzymes of biotechnological relevance, the *abfB* gene from *Aspergillus niger* encoding an α-L-arabinofuranosidase and the *bglN* gene from *Candida molischiana* encoding a β-glucosidase. This expression yeast system has been chosen in order to compare the production of these two enzymes with other microbial models previously established. The recombinant enzymes have been purified, characterised and used to catalyse transglycosylation reactions in high yield. The modified strains overproducing both enzymes are available for further improvement and validation of fermentation models.

4.1 Background

Heterologous expression of glycosidases has been carried out in bacteria, yeast, fungi and cell cultures. Whereas bacterial β-glucosidases are usually expressed intracellularly, cloned fungal Bgl are mostly extracellular, with a few exceptions, such as BglA from *Aspergillus kawachii* found in the periplasmic space when expressed in *Saccharomyces cerevisiae* (Bhatia *et al.*, 2002).

In particular, β-glucosidase from the yeast *C. molischiana* and α-L-arabinofuranosidase from the fungus *A. niger* have been cloned and expressed in *S. cerevisiae* (Crous *et al.*, 1996; Sánchez-Torres *et al.*, 1996, 1998). Both enzymes have attracted considerable biotechnological interest for their application in aroma release from terpene-bound precursors in juices, musts and wines and are also potentially applicable in the synthesis of modified monosaccharides and oligosaccharides.

The enzymes have been previously overexpressed in T73, an industrial *S. cerevisiae* yeast for oenological applications (Ramón *et al.*, 1993; Sánchez-Torres *et al.*, 1996,

1998). The aim in this case was to bring together the hydrolytic activity of these enzymes with the fermentative capability of the yeast, in order to increase the aroma of wine. The recombinant enzymes have been also purified and used to release the aroma in tomato juice (Ortiz, 2008). However, although the expression levels were high and the use of the modified strains was validated in a technological application, the recombinant constructs were based on multicopy plasmids and required the addition of antibiotics to maintain selective pressure. In addition, recombinant Bgl was found to be hyperglycosylated when expressed in *S. cerevisiae* and was poorly secreted to the supernatant while the majority of the activity was retained bound to cell walls. The expression in filamentous fungi was also accomplished although in this case, the recombinant enzymes co-purified with similar endogenous hydrolases. Attemps to express the enzymes in bacterial systems such as *Escherichia coli* have been unsuccessful (Dr. Andrew McCabe, personal communication).

In comparison to *S. cerevisiae*, *Pichia* strains frequently provide higher protein yields (Morawski *et al*, 2000; Morton and Potter, 2000), although this effect is protein-dependent and in some cases *Saccharomyces* has proven to be a more suitable host (Zhu *et al.*, 1998). As *Pichia* can be grown in cell suspensions in reasonably strong methanol solutions, cross-contamination is minimised. Integration in AOX locus provides stable transformants, and thus the system is easy to set-up and maintain industrially. Very high cell densities can be achieved, and this increases the total protein concentration in the broth for a given amount of protein expressed per cell. Finally, although *Pichia* shows particular glycosylation patterns, hyperglycosylation is less usual in recombinant proteins expressed in the methylotrophic yeast (Kato *et al.*, 2001, Montesino *et al.*, 1998). For these reasons, *P. pastoris* is expected to be a good system for the expression of the fungal glycosidases Abf and Bgl.

As shown in Table 10, *P. pastoris* has been successfully used as a host for the heterologous expression of several β-glucosidases from fungi (*A. niger* and *Phanerochaete chrysosporium)* and plants, such as barley and Thai Rosewood (Cairns *et al.*, 2000). Arabinofuranosidases from the filamentous fungi *A. kawachii* and *Trichoderma koningii* have also been expressed in *Pichia*. Most of them were produced in secreted active forms with varying yields. An endogenous intracellular β-glucosidase has also been identified and cloned in *P. pastoris* (Turan and Zheng, 2005).

Table 10. Recombinant glycosidases expressed in *P. pastoris.*

Protein	Mode	Details	References
Aspergillus awamori glucoamylase	S	400 mg/L	Fierobe *et al.*, 1997
A. awamori glucoamylase	S	400 mg/L	Heimo *et al.*, 1997
Aspergillus kawachii α-L-arabinofuranosidase	S	α-MF	Miyanaga *et al.*, 2004
Aspergillus niger endo-β-1,4-xylanase	S	60 mg/L	Berrin *et al.*, 2000
A. niger β-glucosidase	S	0.5 g/L	Dan *et al.*, 2000
Bacillus licheniformis α-amylase	S	2.5 g/L	Montesino *et al.*, 1998
Barley aleurone tissue α-glucosidase	S	α-MF	Tibbot *et al.*, 1998
Barley α-amylase 1 and α-amylase 2	S	50 mg/L, 1 mg/L	Juge *et al.*, 1996
Barley β-glucosidase	S		Muslin *et al.*, 2000
Bovine β-1,4-galactosyltransferase I	S	11mU/L, α-MF	Bencúrová *et al.*, 2003
Candida albicans MNT1 mannosyltransferase	S	250 mg/L, α-MF	Thomson *et al.*, 2000
Coffee bean α-galactosidase	S		Zhang *et al.*, 2002
Coffee bean α-galactosidase	S	400 mg/L, α-MF	Zhu *et al.,* 1995
Escherichia coli β-D-glucuronidase	I		Ayra-Pardo *et al.*, 1999
E. coli β-Galactosidase	I	2 g/L, 103 U/mg	Tchopp *et al.*, 1987
Fusarium solani pectate lyase	S		Guo *et al.*, 1996
Hevea brasiliensis hydroxynitrile lyase	I	22 g/L	Hasslacher *et al.*, 1997
Human chitinase	S	400 mg/L, α-MF	Goodrick *et al.*, 2001
Human α-1,2-mannosidase	S	α-MF	Tremblay *et al.*, 1998
Human α-N-Acetylgalactosaminidase	S	11.6 mg/L, α-MF	Zhu *et al.*, 1998
Human β-1,4-galactosyltransferase	S	α-MF, 0.6 U/L	Malissard *et al.*, 2000
Mouse salivary α-amylase	S	240 mg/L	Kato *et al.*, 2001
Olea europaea β-1,3-glucanase	S	100 mg/L, α-MF	Palomares *et al.*, 2003
Pepper endo-β-1,4-glucanase	S	α-MF	Ferrarese *et al.*, 1998
Penicillium chrysosporium β-glucosidase	S	28,500 U/L	Kawai *et al.*, 2003
Streptomyces viridosporus T7A endoglucanase	S	2.47 g/L, α-MF	Thomas and Crawford, 1998
Thai Rosewood β-glucosidase	S	6000 U/L	Cairns *et al.*, 2000
Trichoderma harzianum β-(1-6)-glucanase	S	9.3 mg/L	Bom *et al.,* 1998
Trichoderma koningii α-L-arabinofuranosidase	S	α-MF	Wan *et al.*, 2007
Trichoderma reesei 1,2-α-mannosidase	S	17.5 mg/L, α-MF	Maras *et al.,* 2000
T. reesei cellobiohydrolyase Cel7A	S		Boer *et al.*, 2000
T. reesei cellobiohydrolyase I	S	α-MF	Godbole *et al.,* 1999
Trypanosoma cruzi acid α-mannosidase	S	11.5 mg/L	Vandersall-Nairn *et al.*, 1998

4.2 Objectives

The general goal of this chapter is to obtain transformants of *P. pastoris* able to express the activities β-glucosidase and α-L-arabinofuranosidase encoded by the genes *abfB* from *A. niger* and *bglN* from *C. molischiana*. With this purpose, the following specific objectives will be addressed:

(i) To select transformants of *P. pastoris* in which the genes encoding the Abf activity of *A. niger* and the Bgl activity of *C. molischiana* have been expressed under control of the AOX promoter. These constructs should express Abf with its own signal peptide, and Bgl both with its own signal peptide, and α-MF of *S. cerevisiae*,

(ii) To quantify in liquid cultures assays the expression level of these enzymes for each of the selected transformants, and to optimise the production in batch controlled cultures.

(iii) To purify and characterise the recombinant enzymes produced in *P. pastoris*, and compare them to the native enzymes produced by *A. niger* and *C. molischiana*, and the recombinant enzymes produced in *S. cerevisiae*.

(iv) To assess the capacity of these enzymes to catalyse model transglycosylation reactions as the methylation of glucose and arabinose using pNP activated substrates.

4.3 Results

4.3.1 Obtention of recombinant *P. pastoris* strains expressing *A. niger abfB* and *C. molischiana bglN* genes

Construction of recombinant expression plasmids

As shown in Figure 13, *abfB* and *bglN* genes were placed under control of AOX gene promoter inside an integrative cassette in different vectors of the pPICz series. *abfB* contains a native signal peptide (Sánchez-Torres, 1996) that can be efficiently processed by *S. cerevisiae*, and it was conserved to create an extracellular expression construct bearing the original secretion signal from *A. niger*. On the contrary, Bgl is poorly secreted when expressed in *S. cerevisiae* strains in which the protein is retained in the cell wall. Bgl also contains a native signal peptide (see Materials and Methods) that was removed in a first construct in which the protein was cloned in translational fusion with the α-mating factor of *S. cerevisiae*. In an alternative construct, the protein was cloned with its own signal peptide in a pPICzB vector to compare the secretion efficiency of the two systems.

For this, *abfB* and *bglN* genes were recovered from *S. cerevisiae* T73-Abf and T73-Bgl by plasmid DNA extraction (from YCAbf and YCBglN plasmids, respectively) and PCR following the protocols described in Section 3.2.

AbfB gene was obtained by PCR reaction using the primers Abf-*Eco*RI and Abf-*Xho*I whose sequences are shown in Table 3, digested with the restriction enzymes *Eco*RI and *Xho*I and subcloned in vector pPICzB. The construction was linearised with *Pme*I, and used to transform *P. pastoris* X33 strain as refered previously.

BglN gene was amplified using the primers Bgl-*Pml*I and Bgl-*Sac*II, and Bgl-αMF*Pml*I and Bgl-*Sac*II. These fragments were subcloned into pPICzB and pPIC$z\alpha$B vectors, respectively, linearised with *Bst*XI and used to transform *P. pastoris* X33 strain.

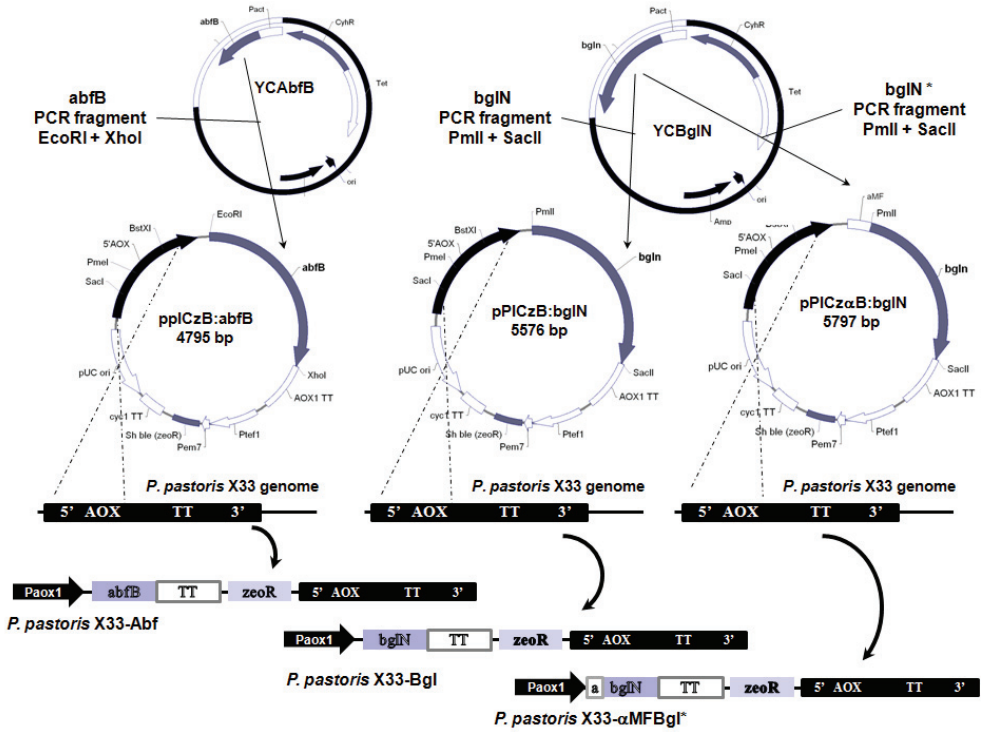


Figure 13. Cloning and integration scheme for X33-Abf, X33-Bgl and X33-αMFBgl strains.

Selection and screening of transformants in solid media cultures

Pichia transformants are usually selected in order to recover multicopy integrants for improved expression. This is accomplished by plating in the presence of successively increasing antibiotic (zeocin) concentrations. However, direct evaluation of the actual activity of each multicopy integrant must be performed in parallel, as it has been shown that copy number and activity correlation varies depending on the gene and the particular construct (Hohenblum *et al.*, 2004). In the case of glycosidase activities, the selection can be directly conducted on the basis of enzymatic activity, since direct screening with chromogenic and fluorescent substrates in solid plates is possible as described in Section 3.2.4.

Pichia transformants overexpressing Abf and Bgl activities were thus selected for their resistance to zeocin and then evaluated by direct activity screening of Zeo^R transformants on solid plates. The transformants were further characterised evaluating their Mut phenotype.

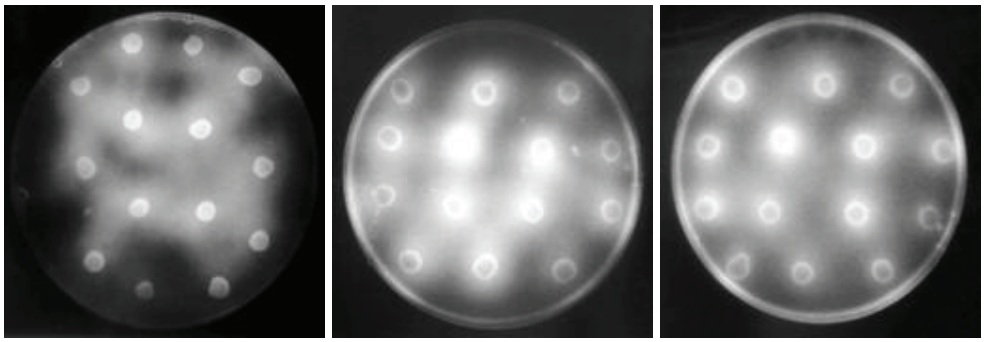


Figure 14. Plate screening for X33-Abf (left panel), X33-Bgl (center panel) and X33-αBgl (right panel) transformants.

Preliminary results of the plate activity screening showed efficient secretion of Abf (see Figure 14, left panel), in accordance to the results described by Sánchez-Torres and coworkers (1998) for recombinant *S. cerevisiae* strains, that were able to process efficiently endogeneous *A. niger* signal peptide. Both the Bgl construct bearing its own signal peptide, and the α-factor fusion, showed faint secretion halos after week-long incubation in methanol solid growth medium although stronger Bgl activity was detected bounded to the cells (Figure 14, center and right panel). All transformants exhibited Mut^+ phenotype when grown on MMH plates.

Characterisation of transformants in liquid cultures

Four transformants were selected for each construct and their expression levels were evaluated in BMGY and BMMY liquid cultures in shaking baffled flasks for several days according to the manufacturers' protocol. Zeocin resistance was assessed as an indirect marker of copy number (data not shown).

As can be seen in Figure 15, significant variability was observed in the activity of the selected transformants in the three cases. X33-Abf constructs reached 0.2-0.4 U/mL after 5 days cultivation in flask cultures with pulsed addition of methanol, in the range of similar proteins produced at this scale in *P. pastoris* AOX expression system (Cereghino and Cregg, 2000). Most of the activity was found located in the supernatant (Table 11) although the transformant (X33-Abf12) with the greatest levels of activity showed increasing cell-bound activity in time, possibly due to limitations in secretion or folding at higher protein concentrations.

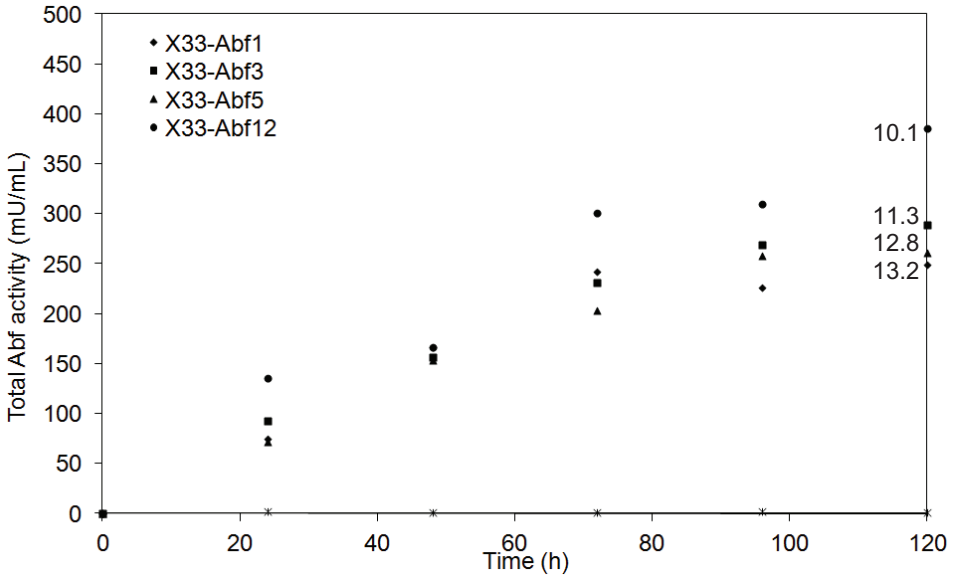


Figure 15. Expression patterns of X33-Abf selected strains in liquid cultures.OD_{600} reached at the end of cultivation is given for each strain.

Table 11. Subcellular location of enzymatic activity in transformants X33-Abf.

	Enzymatic activity (%)							
	X33-Abf1		X33-Abf3		X33-Abf5		X33-Abf12	
Time	SN	Cell wall	SN	Cell wall	SN	Cell wall	SN	Cell wall
24	79,8%	20,2%	76,2%	23,8%	76,2%	23,8%	93,9%	6,1%
48	83,1%	16,9%	81,4%	18,6%	81,4%	18,6%	85,1%	14,9%
72	84,8%	15,2%	85,1%	14,9%	85,1%	14,9%	74,9%	25,1%
96	85,0%	15,0%	82,1%	17,9%	85,5%	14,5%	73,5%	26,5%
120	86,4%	13,6%	83,3%	16,7%	85,6%	14,4%	65,0%	35,0%

This transformant (X33-Abf12) showed also an increased resistance to zeocin (data not shown) probably corresponding to the integration of several copies of the gene. Note that increased activity levels also correlate to lower growth due to the metabolic burden caused by the expression of the recombinant protein.

X33-Bgl strains showed also variable but lower expression levels, below 0.1 U/mL after 5 days pulsed methanol addition in complex medium (Figure 16). At the same time, stress or toxicity effects due to overexpression resulted in lower final biomass growth. The low specific activity obtained may also be an indication of limitations in secretion or folding. As can be seen in Table 12, most of Bgl activity was in fact detected bound to the cell wall. In this first approach, intracellular activity was disregarded. Background expression of endogenous *P. pastoris* β–glucosidase was detected in the cell-bound fraction in the control strain X33-pPICzB.

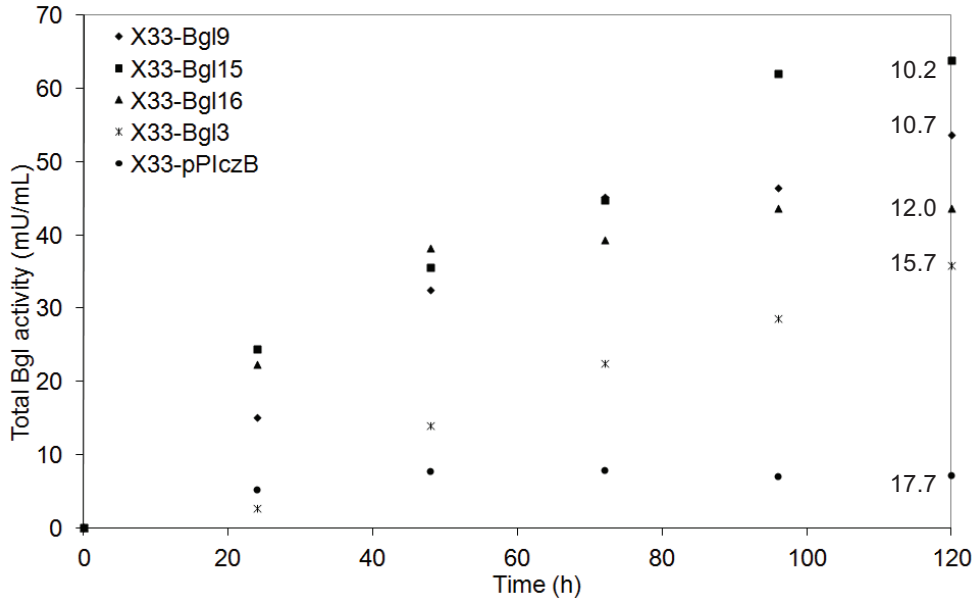


Figure 16. Expression patterns of X33-Bgl selected strains in liquid cultures. OD_{600} reached at the end of cultivation is given for each strain.

Table 12. Subcellular location of enzymatic activity in transformants X33-Bgl.

	Enzymatic activity (%)							
	X33-Bgl3		X33-Bgl9		X33-Bgl15		X33-Bgl16	
Time	SN	Cell wall	SN	Cell wall	SN	Cell wall	SN	Cell wall
24	14,6%	85,4%	0,0%	100,0%	10,3%	89,7%	0,8%	99,2%
48	16,0%	84,0%	13,5%	86,5%	18,1%	81,9%	4,8%	95,2%
72	12,7%	87,3%	18,7%	81,3%	23,4%	76,6%	9,4%	90,6%
96	13,0%	87,0%	26,8%	73,2%	22,4%	77,6%	14,3%	85,7%
120	11,2%	88,8%	27,2%	72,8%	28,5%	71,5%	14,7%	85,3%

A similar result was obtained when X33-αBgl strains, expressing α-factor fusionated Bgl, were analysed (Figure 17, Table 13). Even lower levels of Bgl activity were detected, although slightly higher relative values of secretion (30-50%) were found in comparison to the construct bearing the native signal peptide.

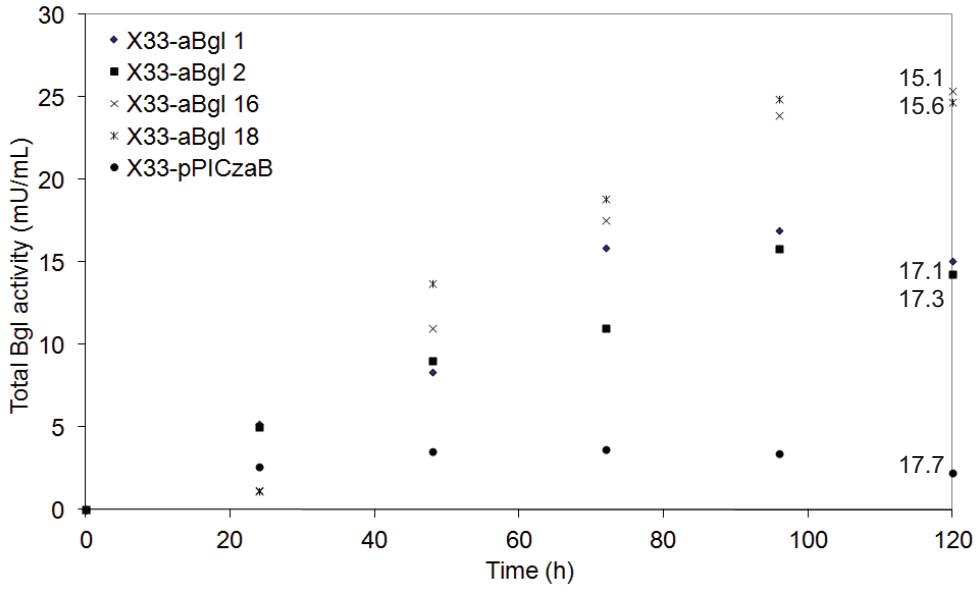


Figure 17. Expression patterns of X33-αBgl selected strains in liquid cultures. OD_{600} reached at the end of cultivation is given for each strain.

Table 13. Subcellular location of enzymatic activity in transformants X33-αBgl.

	Enzymatic activity (%)							
	X33-αBgl1		X33-αBgl2		X33-αBgl16		X33-αBgl18	
Time	SN	Cell wall	SN	Cell wall	SN	Cell wall	SN	Cell wall
24	6,4%	93,6%	9,3%	90,7%	12,1%	87,9%	33,3%	66,7%
48	19,1%	80,9%	14,9%	85,1%	33,3%	66,7%	37,5%	62,5%
72	22,6%	77,4%	24,8%	75,2%	50,2%	49,8%	36,8%	63,2%
96	23,0%	77,0%	23,1%	76,9%	52,6%	47,4%	33,7%	66,3%
120	24,9%	75,1%	29,1%	70,9%	51,8%	48,2%	35,9%	64,1%

On the basis of these results, transformants X33-Abf12 and X33-Bgl15 were selected for the following experiments. Due to their low levels of activity, and the still high percentage of activity found in the cell-wall fraction, X33-αBgl constructs were disregarded.

4.3.2 Characterisation of the selected transformants

The objective of these assays was to establish improved growth and expression conditions in discontinuous cultures for the obtention of Abf and Bgl enzymes from the selected transformants. With this purpose, the kinetic growth curves in minimal and complex media were calculated for both strains, and a factorial design of experiments was used to define optimal environmental conditions for expression in the medium of choice.

Growth and expression kinetics

It is known that *P. pastoris* growth and expression is strongly affected by culture medium composition (Cos *et al.*, 2006a). Growth curves in glycerol and methanol (0.5% p/v) were obtained for the selected transformants in buffered minimal (Figure 18, BMGH medium; Figure 20, BMMH medium) and complex media (Figure 19, BMGY medium; Figure 21, BMMY medium) in order to calculate the corresponding kinetic parameters (specific growth rate, overall yield and specific productivity) in flask cultures.

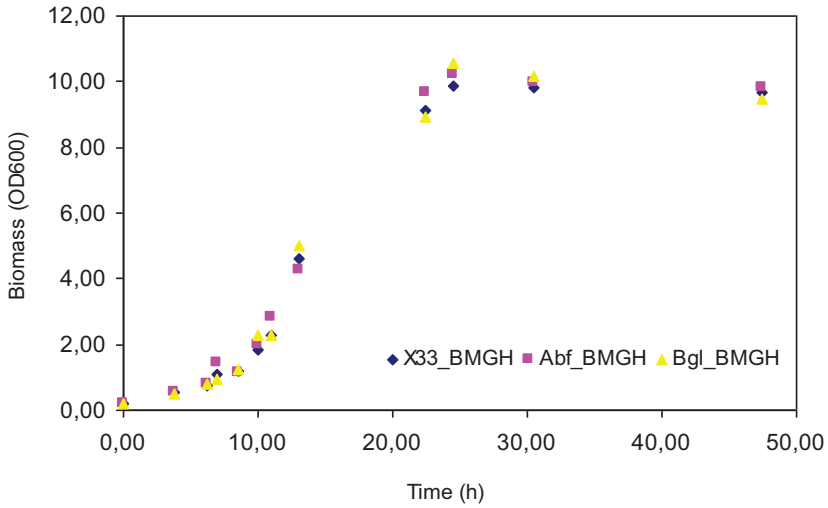


Figure 18. Growth curve of selected transformants and wild type strain in BMGH medium.

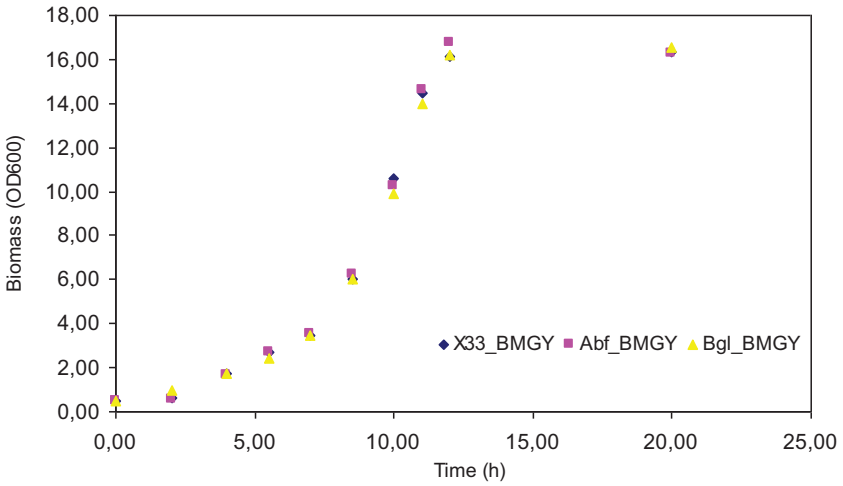


Figure 19. Growth curve of selected transformants and wild type strain in BMGY medium.

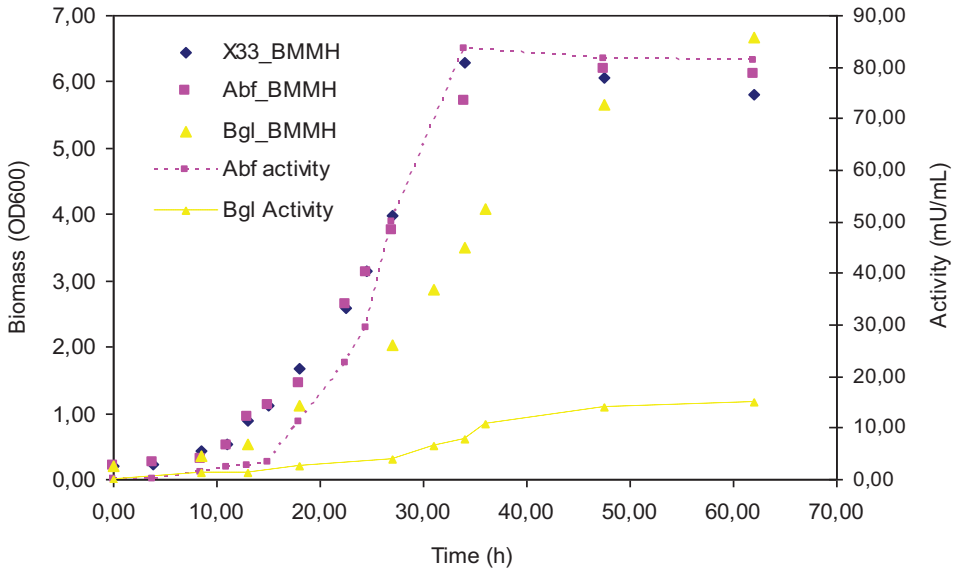


Figure 20. Growth curve of selected transformants and wild type strain in BMMH medium. Activity levels are shown in the right-hand side axis.

The subcellular location of activity in buffered minimal methanol medium was also determined together with intracellular activity. As can be seen in the following tables (Tables 14 and 15), the results were in accordance to the previous experiments performed in complex medium. Abf activity was mostly located in culture supernatant whereas Bgl activity was found bound to the cell wall.

Table 14. Subcellular location of enzymatic activity in transformant X33-Abf12, BMMH medium.

	Enzymatic activity (%)		
Time (h)	**Extracellular**	**Cell wall**	**Intracellular**
18,0	77,2	11,6	11,2
24,5	72,0	21,7	6,3
47,5	71,9	19,6	8,5
62,0	63,7	22,7	13,6

Table 15. Subcellular location of enzymatic activity in transformant X33-Bgl15, BMMH medium.

	Enzymatic activity (%)		
Time (h)	**Extracellular**	**Cell wall**	**Intracellular**
18,0	4,1	71,9	24,0
24,5	6,3	80,8	12,9
47,5	5,9	75,4	18,7
62,0	4,8	74,7	20,5

Supplementary experiments were performed in complex medium in order to calculate kinetic constants and overall yields on substrate in each system. The results, shown in Figure 21 and Tables 16-17, show that higher levels of activity can be produced in complex medium in both Abf and Bgl but subcellular location is not modified. The higher enzymatic activity correlates to higher growth levels that derive from the supplemented carbon source in nitrogen complex components such as yeast extract and peptone.

If specific activity levels are compared, it can be seen how Abf specific protein production remains approximately constant in both media, reflecting methanol growth-association of the expression. This is also the case for Bgl, although once again lower levels of expression are attained. Tables 16 and 17 show the subcellular location of enzymatic activity in complex methanol medium, confirming the results previously obtained in both cases.

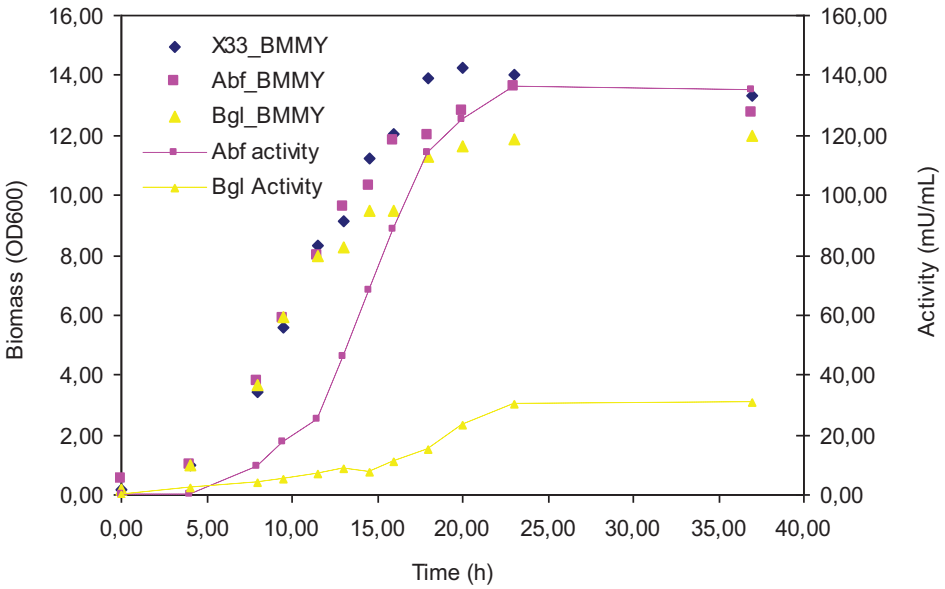


Figure 21. Growth curve of selected transformants and wild type strain in BMMY medium. Activity levels are shown in the right-hand side axis.

Table 16. Subcellular location of enzymatic activity in transformant X33-Abf12, BMMY medium.

	Enzymatic activity (%)		
Time (h)	**Extracellular**	**Cell wall**	**Intracellular**
14,5	73,1	11,9	15,0
18,0	68,9	13,9	17,2
23,0	72,6	16,8	10,6
37,0	64,2	15,1	20,8

Table 17. Subcellular location of enzymatic activity in transformant X33-Bgl15, BMMY medium.

	Enzymatic activity (%)		
Time (h)	**Extracellular**	**Cell wall**	**Intracellular**
14,5	8,2	71,4	20,4
18,0	9,1	75,1	15,8
23,0	15,1	72,2	14,7
37,0	12,3	67,9	19,8

The kinetic parameters are summarised in Table 18 for each of the conditions assayed.

Table 18. Kinetic parameters calculated for the different growth media assayed.

Medium	Carbon source	Strain	μ (h-1)	$Y_{X/S}$ (g/g)	η (UA/g)
BMGH	Glycerol, 0.5%	X33	0.23	0.43	-
BMGH	Glycerol, 0.5%	X33-Abf12	0.23	0.44	-
BMGH	Glycerol, 0.5%	X33-Bgl15	0.24	0.46	-
BMGY	Glycerol, 0.5%	X33	0.31	0.71	-
BMGY	Glycerol, 0.5%	X33-Abf12	0.32	0.73	-
BMGY	Glycerol, 0.5%	X33-Bgl15	0.29	0.72	-
BMMH	Methanol, 0.5%	X33	0.11	0.27	-
BMMH	Methanol, 0.5%	X33-Abf12	0.10	0.27	2.93
BMMH	Methanol, 0.5%	X33-Bgl15	0.07	0.29	0.48
BMMY	Methanol, 0.5%	X33	0.10	0.62	-
BMMY	Methanol, 0.5%	X33-Abf12	0.08	0.55	2.17
BMMY	Methanol, 0.5%	X33-Bgl15	0.06	0.50	0.55

Specific growth rates in glycerol and methanol are within the range previously described for Mut+ *P. pastoris* strains, together with the yields on biomass both on glycerol and methanol in minimal media. Please note that the increase in the overall yield results from the contribution of complex nitrogen sources to growth in BMGY and BMMY media. As can be seen in Table 18, lower specific growth rates are found in the transformants in comparison to the wild type strain. Growth limitation is more severe in the case of X33-Bgl in both media assayed.

Optimisation of environmental conditions

The objective of these assays was to establish the optimal environmental conditions for the expression of Abf and Bgl activities of the selected transformants in contrast to the previous experiments performed in flask cultures. For this purpose, X33-Abf12 and X33-Bgl15 strains were cultured in BMGY medium, recovered and resuspended in MMY medium, non-buffered.

The experimental conditions were set as described in Section 3.2.7 in duplicate (9x2 wells) for each transformants. Biomass growth was determined offtline by regular sampling and optical density monitoring. Enzymatic activity was determined after 24 hours. The results are shown in Tables 19 and 20.

Table 19. Results of the functional design for X33-Abf12.

Condition	Variables			Abf activity	
	Temperature (ºC)	Airflow (vvm)	pH	mU/mL	
1	26	0	5.0	116.1	129.6
2	26	1	5.5	675.1	688.7
3	26	2	6.0	612.3	617.1
4	28	0	5.0	345.3	241.8
5	28	1	6.0	786.4	882.1
6	28	2	5.5	657.7	592.0
7	30	0	6.0	108.3	147.0
8	30	1	5.0	712.9	754.5
9	30	2	5.5	591.0	651.9

Table 20. Results of the functional design for X33-Bgl15.

Condition	Variables			Bgl activity	
	Temperature (ºC)	Airflow (vvm)	pH	mU/mL	
1	26	0	5.0	10.9	11.9
2	26	1	5.5	88.4	87.4
3	26	2	6.0	81.2	88.2
4	28	0	5.0	13.3	10.5
5	28	1	6.0	53.0	59.0
6	28	2	5.5	44.3	49.1
7	30	0	6.0	16.2	10.9
8	30	1	5.0	38.9	39.9
9	30	2	5.5	39.9	43.8

As can be seen in Tables 19 and 20 notorious differences appear in relation depending on the environmental conditions. In order to establish the influence of each variable, an analysis of variance (ANOVA) was carried out. The statistical significance of these effects is shown in Table 21 and Figure 22. In the case of Abf, only airflow and temperature have a statistically significant effect on enzymatic activity. In the case of Bgl, all three variables we found to be relevant to recombinant expression.

Table 21. Variance analysis results of the multifactorial design for X33-Abf and X33-Bgl.

Abf activity	Factors	Sum of squares	F quotient	P value
	Airflow	$1.06 \cdot 10^6$	190.04	0.000
	pH	$5.07 \cdot 10^3$	0.90	0.433
	Temperature	$4.17 \cdot 10^4$	7.44	0.009
Bgl activity	**Factors**	**Sum of squares**	**F quotient**	**P value**
	Airflow	$8.92 \cdot 10^3$	80.56	0.000
	pH	$1.18 \cdot 10^3$	10.69	0.003
	Temperature	$2.92 \cdot 10^3$	26.41	0.000

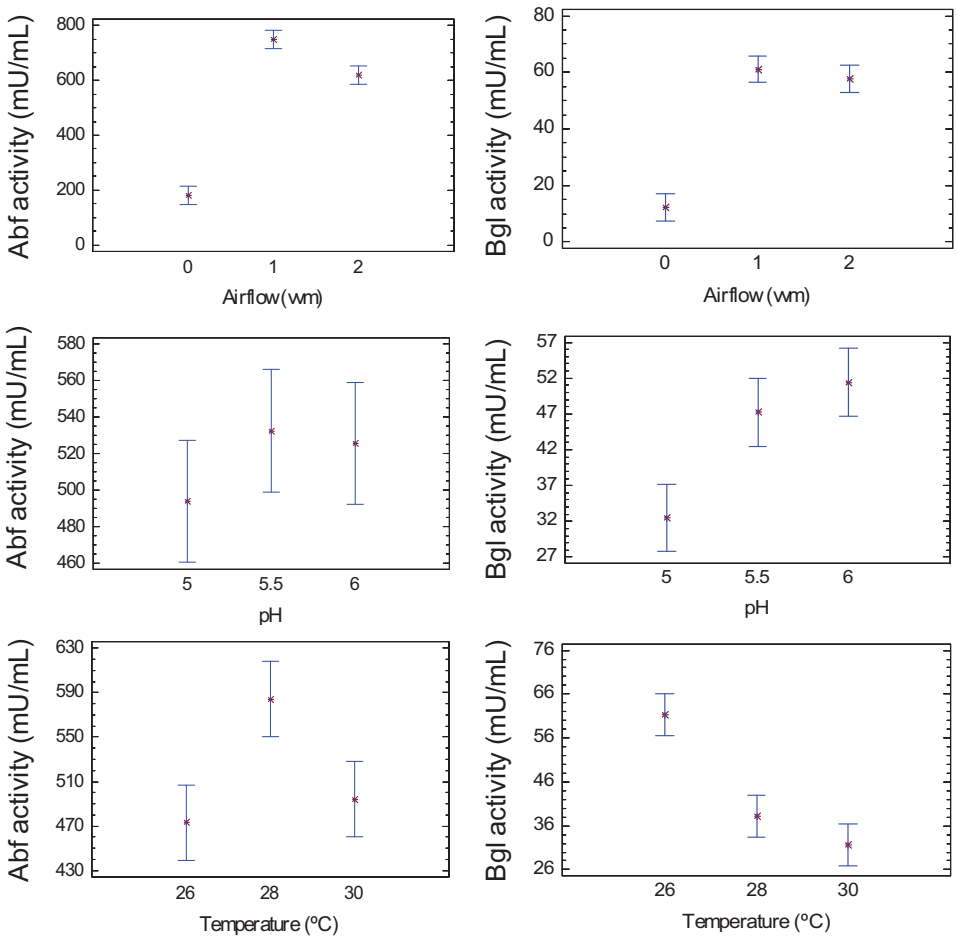


Figure 22. Effect of airflow, pH and temperature on Abf activity of X33-Abf and X33-Bgl.

Abf activity can be increased when temperature is slightly lowered to 28ºC and airflow is kept at 1 vvm. Further aireation does not seem to improve expression and in fact, has a minor negative effect. Optimal expression when temperature is maintained at 28ºC is consistent to previous reports of recombinant protein expression in *P. pastoris*, that suggest avoiding overheating with lower setpoints, since protein synthesis is completely stopped above 32ºC. This could be a favourable condition in this particular system, as the microreactor device does not support active cooling, which can result in punctual temperature deviations of +1-2ºC.

In the case of Bgl activity, optimal conditions correspond to pH 6.0, 26ºC and 1 vvm airflow. The effect of airflow, shared in both systems, must be strongly related to methanol consumption. When no air is provided, growth is limited by oxygen transport and this causes lower methanol consumption rates. These cultures (conditions 1, 4 and 7) also show, for both enzymes, lowered growth rates in relation to the aireated conditions. Again, 1 vvm is sufficient to enhance growth and protein expression and no significant improvement is achieved with 2 vvm.

Finally, a remarkable effect of lowering temperature to 26ºC can be ascertained in the case of Bgl, where enzymatic activity is doubled in comparison to 28ºC and 30ºC. Since 28ºC are sufficient to avoid overheating, this benefit must be related to a specific physiological effect. Other authors have reported that lower temperatures releave stress in *P. pastoris* cultures affected by unfolded protein response, which results in improved recombinant protein expression (Dragosits *et al.*, 2009). This fact further supports the hypothesis that either folding or secretion are bottlenecks for Bgl expression.

4.3.3 Characterisation of the recombinant enzymes

Culture broths produced during the previous experiments (Section 3.2.5) were used to recover, purify and characterise both enzymes as described in Section 3.3.1 with the objective to compare the biochemical properties of the enzymes produced in *P. pastoris* to the ones produced in *S. cerevisiae* and the native ones produced by *A. niger* and *C. molischiana* respectively.

Identification by PAGE

In order to identify the molecular weight of the recombinant proteins, SDS-PAGE gels were run as described in Section 3.3.2. As shown in Figure 23, a clear band at around 70 kDa could be identified in supernatant of X33-Abf cultures whereas a smear of proteins above 90 kDa could be seen in X33-Bgl extracts.

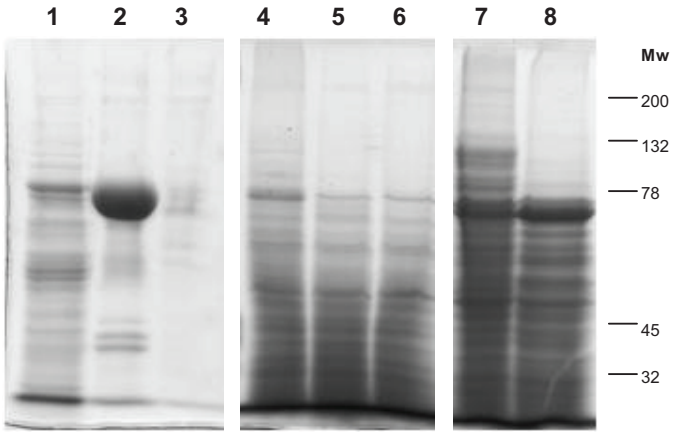


Figure 23. Denaturing and reducing gel electrophoresis of cell extracts and culture supernatant of X33, X33-Abf and X33-Bgl strains. 1: X33-Bgl supernatant, 2: X33-Abf supernatant, 3: X33 supernatant, 4: X33-Bgl cell extract, BMMH, 5: X33-Abf cell extract BMMH medium, 6: X33 cell extract BMMH medium, 7: X33-Bgl cell extract, BMMY medium, 8: X33 cell extract, BMMY medium.

Non-reducing PAGE gels were incubated with chromogenic substrates in order to identify the bands corresponding to Abf and Bgl proteins (Figure 24).

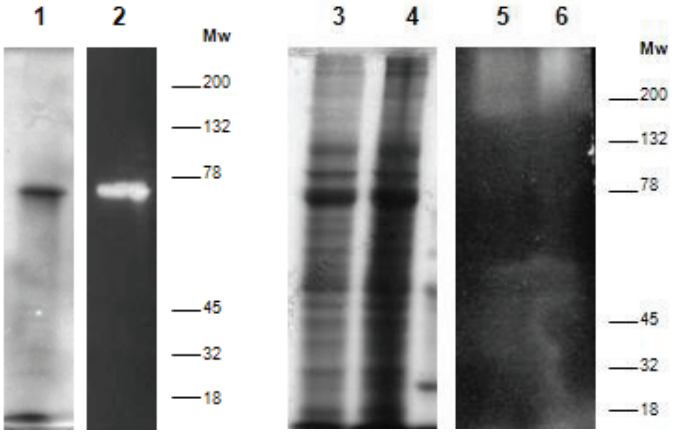


Figure 24. Semi-native, non-reducing gel electrophoresis and corresponding UV-transilluminated zymograms of 1-2: X33-Abf supernatant, 3-6: X33-Bgl cell extract.

As can be seen in Figure 24, Abf activity was associated to a single band of around 70 kDa, whereas Bgl activity was found to correspond to a smear located above 200 kDa, as usually happens with glycosilated proteins. Residual Bgl activity was found below 50 kDa possibly due to partial degradation of the protein.

Enzyme purification

Abf and Bgl enzymes were isolated from culture broths from the corresponding transformants. Abf was isolated directly from the culture supernatant in a single purification step. Bgl was located both in the cells and the culture supernatant and was purified from the cell bound fraction as described in Section 3.3.1. Details are given in Table 22.

Table 22. Purification of Abf and Bgl from *Pichia* cultures.

Purification step	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Purification (Fold)	Yield (%)
Culture supernatant Abf	242.2	98	2.50	1	100
Anionic exchange	154.8	10.39	14.90	5.9	64
Nanofiltration (50 kDa)	147.2	9.60	15.30	6.2	61
Culture supernatant Bgl	13.9	96	0.20	1	-
Cell fraction Bgl	48.1	-	-	1	100
1 N NaCl extract	34.8	69.6	0.47	1	72
Anionic exchange	25.4	19.8	1.28	2.72	53
Nanofiltration (100 kDa)	24.8	17.9	1.38	2.93	51

As can be seen in Figure 25, purified Abf from *P. pastoris* X33-Abf shows a similar molecular weight to the protein purified from *S. cerevisiae*, with an apparent weight of 67 kDa.

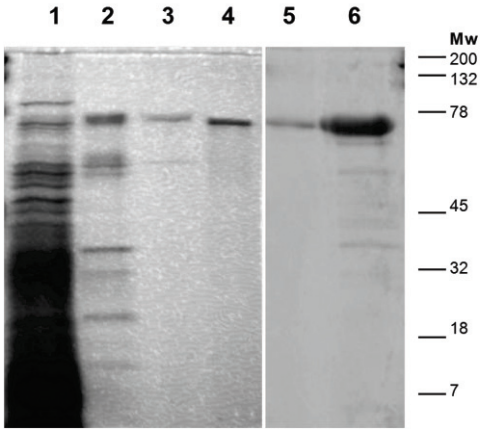


Figure 25. Gel electrophoresis of purified Abf from X33-Abf12. Lanes from left to right: 1: cell extract 1 NaCl, 2: culture supernatant, 3: anionic exchange, 4: purified protein after nanofiltration (freeze-dried), 5: purified Abf from X33-Abf, 6: purified Abf from T73-Abf.

Conversely, purified Bgl seems to be heavily glycosylated, as once again suggested by the smeared, diffuse band observed with a high apparent molecular weight (more than 200 KDa) that could be reduced by treatment with EndoH (Figure 26). Similar results were obtained by Sánchez and coworkers (1998) when the enzyme was expressed in *S. cerevisiae*.

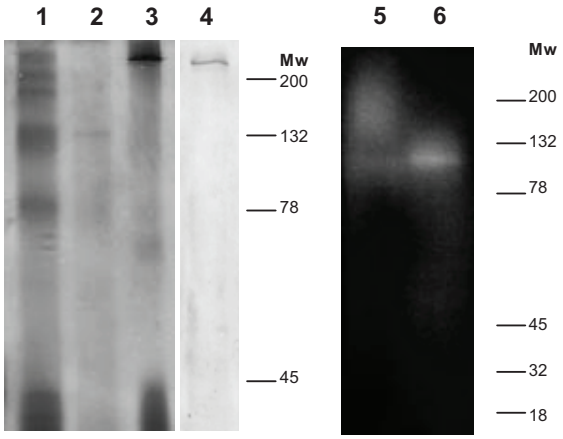


Figure 26. Gel electrophoresis of purified Bgl from X33-Bgl15. 1: cell extract 1 NaCl, 2: culture supernatant, 3: anionic exchange, 4: purified protein after nanofiltration (freeze-dried), 5: purified Bgl after anionic exchange, 6: Bgl after treatment with EndoH.

Biochemical properties

Optimal pH and temperature values of 3.8 and 60ºC, and 4 and 55ºC were found for recombinant Abf and Bgl respectively, similar to what has been described by other authors for the native (Flipphi *et al.*, 1993, Janbon *et al.*, 1995) and the recombinant enzymes produced in T73 *S. cerevisiae* recombinant strains (Sánchez-Torres *et al.*, 1996, 1998), as can be seen in Figures 27-30.

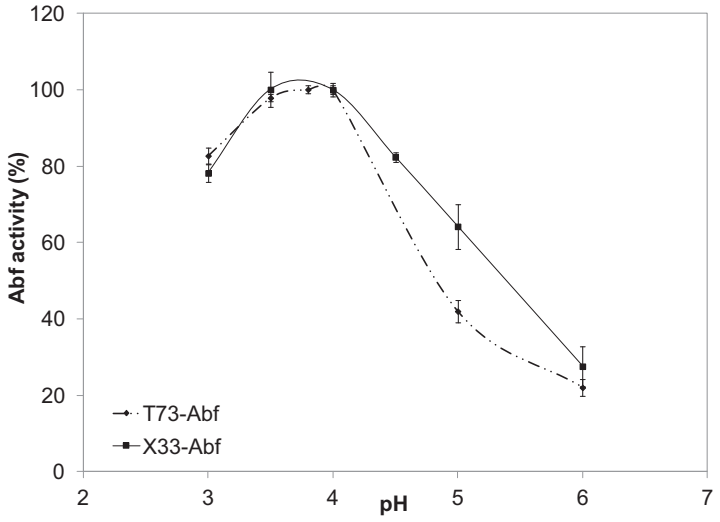


Figure 27. Optimal pH for recombinant Abf produced in *S. cerevisiae* and *P. pastoris*.

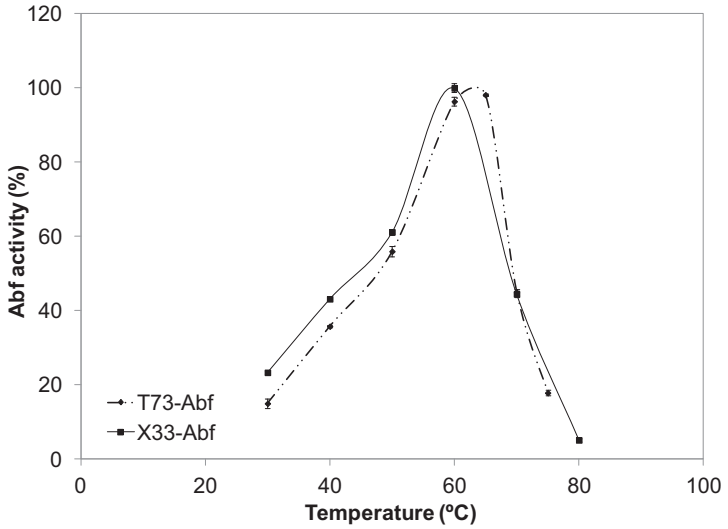


Figure 28. Optimal temperature for recombinant Abf produced in *S. cerevisiae* and *P. pastoris*.

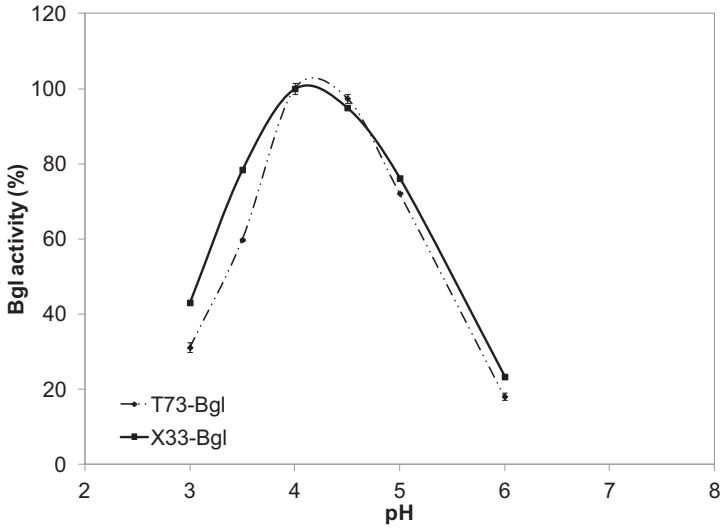


Figure 29. Optimal pH for recombinant Bgl produced in *S. cerevisiae* and *P. pastoris*.

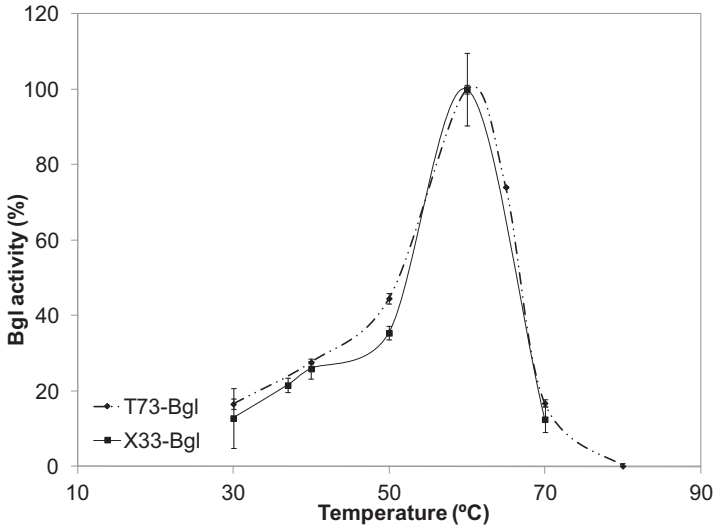


Figure 30. Optimal temperature for recombinant Bgl produced in *S. cerevisiae* and *P. pastoris*.

The determination of kinetic constants was carried out by using pNPA and pNPG in a substrate range comprised between 0.01 and 2 mM. Enzyme kinetics were adjusted to Michaelis-Menten kinetics using non-linear regression as shown in Figure 31.

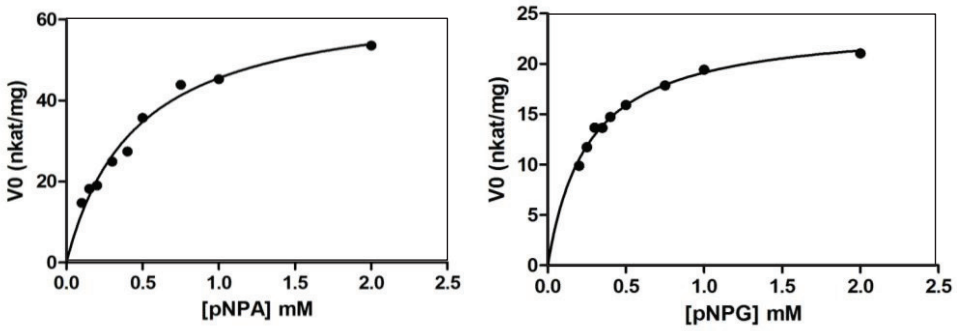


Figure 31. Kinetic constant estimation on pNPA for recombinant Abf (X33-Abf12, left panel) and on pNPG for recombinant Bgl (X33-Bgl15, right panel).

The results are summarised in Table 23.

Table 23. Kinetic constants determined for recombinant and native Abf and Bgl.

Enzyme	Producer	v_{max} nkat/mg	K_m mM	Reference
Abf	*A. niger*	23.5	0.52	Veen *et al*, 1991
Abf	*S. cerevisiae*	43.4	0.37	Sánchez-Torres *et al*, 1996
Abf	*P. pastoris*	66.2	0.45	This work
Bgl	*C. molischiana*	1667	0.24	Vasserot *et al*, 1991
Bgl	*C. molischiana*	717	0.10	Gondé *et al*, 1985
Bgl	*S. cerevisiae*	25.89	0.25	Sánchez-Torres *et al*, 1998
Bgl	*P. pastoris*	24.69	0.28	This work

Some differences can be observed in relation to the kinetic constants calculated for the native enzymes and the recombinant forms produced in *S. cerevisiae*. In the case of Abf, although the specificity for the substrate is maintained, a significant increase in the maximum reaction rate can be observed. In the case of Bgl, kinetic constants

correspond to those obtained for the recombinant enzyme produced by *S. cerevisiae* (Sánchez-Torres *et al.*, 1998) with a high specificity for pNPG as substrate.

Finally the effect of ethanol and glucose as the principal inhibitors of Abf and Bgl activities, respectively, was studied with the two heterologous enzymes expressed in *P. pastoris* and *S. cerevisiae* (Figures 32 and 33).

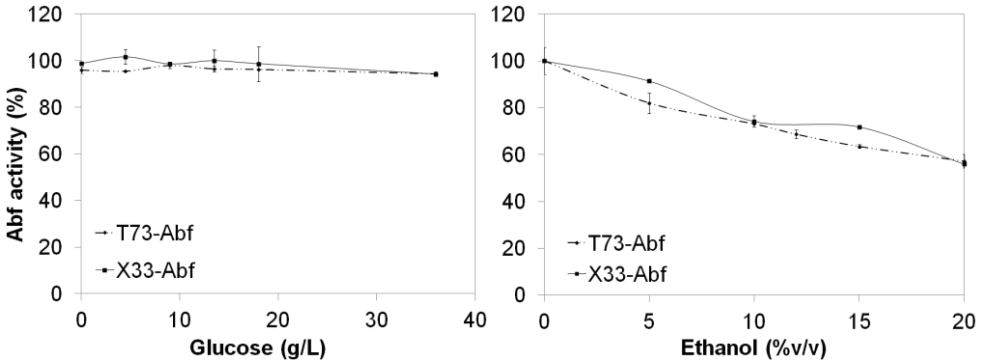


Figure 32. Effect of glucose (right panel) and ethanol (left panel) on recombinant Abf activity expressed in *S. cerevisiae* and *P. pastoris*.

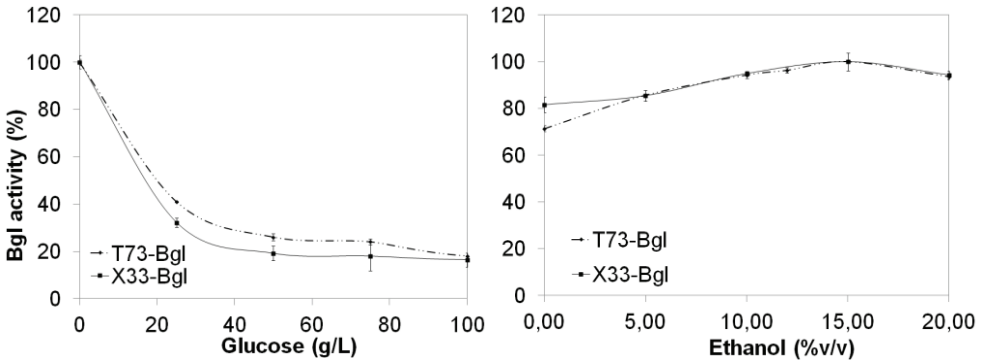


Figure 33. Effect of glucose (right panel) and ethanol (left panel) on recombinant Bgl activity expressed in *S. cerevisiae* and *P. pastoris*.

As expected, an important inhibition of the enzymatic activity was found in the case of ethanol (up to 20% at 20 g/L ethanol for Abf), and glucose (reduction to less than 60% at 25 g/L glucose, with a complete depletion of activity at 40 g/L glucose for Bgl). The effect was similar regardless of the expression system used. Conversely, ethanol was an inducer of the enzymatic activity of Bgl, as found previously by Sánchez-Torres and coworkers (1998) and is common in β-glucosidases from fungal origin.

Substrate specificity was determined for Abf and Bgl (Table 24).

Table 24. Substrate specificity of recombinant Abf and Bgl enzymes.

			Abf	Bgl
Substrate	**Linkage**	**Concentration**	**% Activity**	**% Activity**
pNPG	β(1,4)	5 mM	n.d.	100
Cellobiose	β(1,4)	50 mM	n.a.	12
Salicin	β(1,4)	20 mM	n.a.	48
Methyl-β-glucoside	β(1,4)	20 mM	n.a.	19
Maltose	α(1,4)	50 mM	n.a.	n.d.
Lactose	β(1,4)	50 mM	n.a.	n.d.
Saccharose	β(1,2)	50 mM	n.a.	n.d.
pNPX	β(1,4)	5 mM	0.8	7
pNPA	α(1,6)	5 mM	100	0.9
Arabinogalactan	α(1,6)	0.2%	42	n.a.

n.a. not analysed, n.d. not detected

As can be seen in Table 24, Abf showed activity mostly on α(1,6) linked substrates whereas Bgl hydrolysed preferably β(1,4) linkages. Abf activity on arabinogalactan has already been reported (Margolles-Clark *et al.*, 1996) and Bgl activity on cellobiose, salicin and methyl-β-glucoside is also known. In accordance to our results, several β-glucosidases of fungal origin have also been described to have associated β-xylosidase activity (Kohchi *et al.*, 1985; McCleary and Harrington, 1988; Vasserot *et al.*, 1989).

Unexpectedly, side activities of Abf on pNPX and Bgl on pNPA, respectively, were detected. Although Sánchez-Torres and coworkers (1998) reported that the recombinant Bgl expressed in *S. cerevisiae* was inactive on this substrate, our results

are in accordance with previous finding by Vasserot and coworkers (1991) suggesting that β-glucosidase from *C. molischiana* may hydrolyse other linkages than β(1,4) type. Side arabinofuranosidase activity has also been reported for other β-glucosidases (Michlmayr *et al.*, 2010). No activity could be detected, however, on maltose or other disaccharides. Interestingly, β-xylosidase activity of Abf from *Trichoderma reesei* has also been reported previously (Margolles-Clark *et al.*, 1996).

Use in transglycosylation reactions

In order to verify the capacity of recombinant Abf and Bgl to catalyse alkylation reactions by kinetic controlled transglycosylation, pNPA and pNPG were used as activated substrates and methanol as donor in a standard reaction system as detailed in Section 3.3.7, to produce methyl-arabinoside and methyl-glucoside respectively. Unless stated otherwise, the reaction was carried out at 45ºC for 45 minutes in McIlvaine buffer at pH 4.5. The yield is expressed in terms of molar concentration of formed product (methylated arabinose or glucose) *versus* initial substrate (pNPA and pNPG, at initial concentration of 6 mM in each reaction).

The reaction was firstly optimised in terms of methanol concentration and kinetics. The results of the evolution of alkylation ratio in relation to methanol concentration are shown in Figures 34 and 35 after 10 minutes reaction. As expected, the ratio of alkylated product *versus* the hydrolysed sugar is directly dependent on methanol concentration, and increasing yields of the alkylated product are obtained when methanol concentration increases from 5 to 25%.

The optimal methanol concentration is established for both enzymes in a ratio of 25% v/v (6.25 M methanol in reaction mixture). Methanol concentration higher than 25% seems to inhibit enzymatic activity, particularly in the case of Bgl (Figure 35). In this case, methylation yield of is as high as 83% when methanol is used at 25% v/v. Abf methylation ratio is lower; however a maximal value of 43% is achieved.

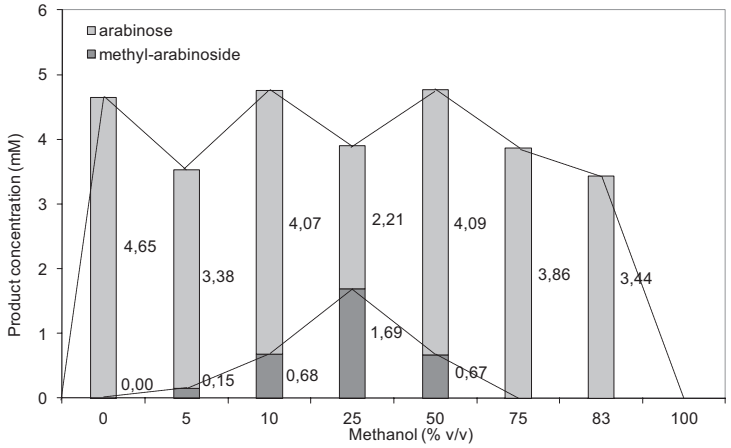


Figure 34. Transglycosylation yield in relation to methanol concentration for recombinant Abf.

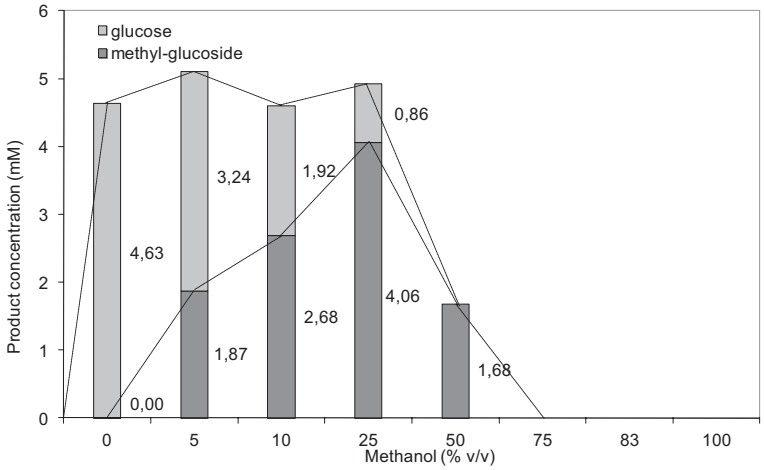


Figure 35. Transglycosylation yield in relation to methanol concentration for recombinant Bgl.

The kinetics of these transformations were studied evaluating the final product concentration in the presence of methanol at 25% v/v at different reaction times. As can be seen in Figures 36 and 37, the reactions proceed quickly and after only 10 minutes the majority of products (either hydrolysed or alkylated sugars) have been formed.

Again, the ratio of Bgl seems to be diverted towards alkylation, whereas Abf enzyme produces preferably the hydrolysed product. The maximal yield is attained after 30-45 minutes. In these conditions, all pNP substrate has been consumed (data not shown). These results are consistent with the previous of other authors that found that alkylation reactions carried out with recombinant Abf cloned from *Thermobacillus xylanilyticus* were fulfilled in less than 30 minutes (Rémond *et al.*, 2002).

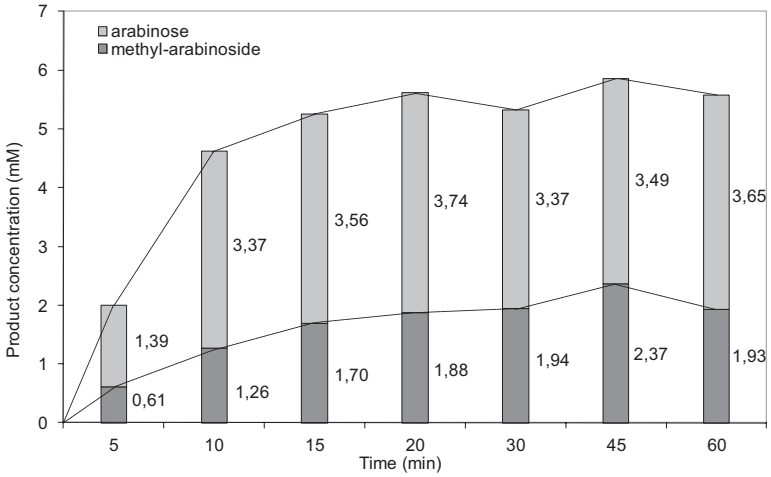


Figure 36. Kinetics of transglycosylation for recombinant Abf.

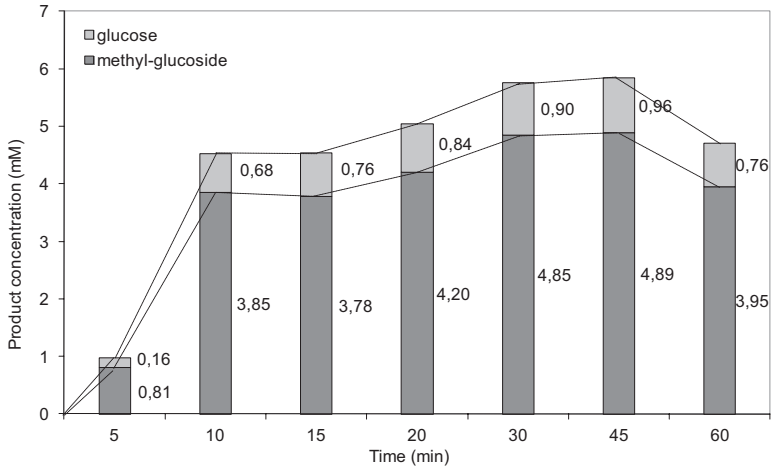


Figure 37. Kinetics of transglycosylation for recombinant Bgl.

Finally, the alkylation yield was evaluated in relation to the enzyme units used in the optimal conditions established previously. As can be seen in Figure 38, transglycosylation occurs even at low concentrations of the enzyme for both Abf and Bgl and 25 mU of the enzyme are sufficient to achieve yields close to the maximum.

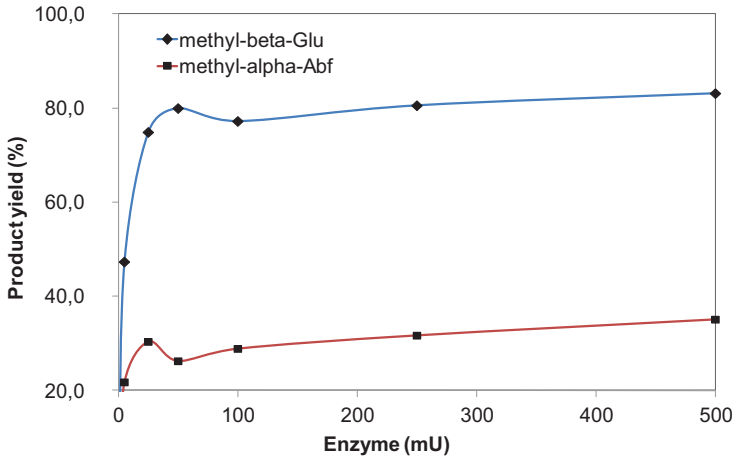


Figure 38. Transglycosylation yield with varying concentration of Abf and Bgl, measured as ratio of methyl-α-arabinoside (methyl-β-glucoside) concentration produced *versus* initial pNPA (pNPG) moles introduced.

4.4 Discussion

This chapter describes the results obtained in the expression in the methylotrophic yeast *P. pastoris* of two glycosidases of fungal origin, encoded by the genes *abfB* from *A. niger* and *bglN* from *C. molischiana*. Both have been cloned, overproduced, purified and characterised. Abf and Bgl enzymes had been previously expressed in *S. cerevisiae* and Bgl, in *A. nidulans* (Sánchez-Torres, 1996). In the case of Abf, the physicochemical and biochemical properties of the native enzyme (Veen *et al.*, 1991) correspond to ones found for the enzyme purified from recombinant *S. cerevisiae,* T73-Abf. Both native and T73-Abf show an apparent molecular weight of 67 kDa, no hyperglycosylation and optimal pH and temperature of 4.0 and 60ºC, respectively (Sánchez-Torres *et al.*, 1996), in accordance to our results for X33-Abf.

In the case of Bgl, the properties of the recombinant enzyme purified from *P. pastoris* were also in general agreement with those described previously for the enzyme purified from *S. cerevisiae*. Recombinant T73-Bgl produced in *S. cerevisiae* showed an apparent molecular weight of more than 200 kDa, due to its high degree of hyperglycosylation. Indeed, Bgl has eleven putative sites of N-glycosylation (Janbon *et al.*, 1995, Sánchez-Torres *et al.*, 1998). Although the expression in *P. pastoris* is known to minimize the number of mannose residues added to recombinant glycoproteins (Montesino *et al.*, 1999), in the case of X33-Bgl a similar behavior was found and the protein is recovered as a glycoprotein. Mostly infrequent, hyperglysosilation has also been described in some proteins expressed in *P. pastoris* (Scorer *et al.*, 1993). In the case of X33-Bgl, the high number of N-glycosylation sites seems to favour the addition of sugar residues, although the quantity and type of glycosylation could be different in the proteins expressed in *P. pastoris* and *S. cerevisiae*: due to the high number of N-glycosylation sites, shorter mannose chains typically found in proteins expressed in *P. pastoris* system (Montesino *et al.*, 1998) could still result in a heavily glycosylated protein.

Whereas X33-Abf is efficiently secreted to culture supernatant, this has not been achieved in the case of X33-Bgl, and the protein is cell-bound regardless of the signal peptide used. The protein is released in high-osmolarity conditions, as was previously observed in other glycosylated glycosidases (Akiyama *et al.*, 1998). Indeed,

glycosylation is strongly related to subcellular location: glycosylated proteins are generally produced extracellularly or bound to cell wall whereas intracellular protein is not glycosylated, in *P. pastoris* (Tschopp *et al.*, 1987) and other fungi (Freer, 1993).

In addition, the toxic effect of the expression of Bgl on *P. pastoris* growth is manifested as lower specific growth rates in the presence of methanol both in minimal and complex culture media. This could be caused by the recombinant expression of the protein itself, as a consequence of folding or secretional bottlenecks, possibly related to the high number of N-glycosylation sites. This is evidenced by the fact that lowering temperature to 26ºC enhances Bgl expression as has been previously reported for other proteins that cause unfolded protein response when expressed in *P. pastoris* (Dragosits *et al.*, 2009). Moreover, the transformants in which the secretion is enforced by using *S. cerevisiae* α-MF as a signal peptide produce even lower amounts of activity. However, it is known that *S. cerevisiae* strains expressing Bgl are capable to produce higher amounts of the enzyme under control of a constitutive promoter (Sánchez-Torres *et al.*, 1998). Interestingly, this strain, contrary to the wild type *S. cerevisiae* T73 and T73 transformants expressing different glycosidases, is unable to grow on ethanol as sole carbon source (Biopolis, unpublished data). The toxicity of Bgl expression alone or in combination with methanol may be analysed and eventually corrected in *P. pastoris* using different promoters available for the yeast, such as pGAP (constitutive expression, inducible by glucose) or pFLD (nitrogen source dependent, inducible by methylamine).

Regarding expression levels, Abf can be recovered from culture supernatant at around 0.7 U/mL under optimised conditions. This value is comparable to the yield described in flask cultures for recombinant *S. cerevisiae* strains expressing this enzyme in multicopy plasmids (Sánchez-Torres *et al.*, 1996) and also to other glycosidases expressed in *P. pastoris* when assayed in flask cultures (Cairns *et al*, 2000). The enzyme yield in relation to biomass concentration is also promising with values close to 1 mg/g. Enzyme titer may be increased further in high-cell density cultures applying standard fermentation protocols (Cos *et al.*, 2006a).

Unfortunately this is not the case for recombinant Bgl, whose expression is limited to 100 mU/mL in flask cultures under optimised conditions. Different strategies can be envisaged in order to increase these levels, related to improved recombinant constructs and optimised fermentation strategies. As previously stated, other promoters such as pGAP o pFLD may be assayed for recombinant expression in *P. pastoris*, as notable differences have been observed in the expression levels achieved for various β-glucosidases. Others authors have also reported site-directed mutagenesis in order to reduce the number of putative N-glycosylation sites although this may affect enzymatic activity and thermostability (Chen *et al.*, 1994). Two-stage feeding policies have routinely been applied in order to increase cell concentration before induction, to achieve higher enzyme concentration even when productivity per cell is low (Zhao *et al.*, 2008a).

Regarding the biotechnological application of the enzymes, both have been previously validated in food applications such as flavor release in wine-making (Sánchez-Torres *et al.*, 1998). This contribution demonstrates that both glycosidases are also capable to catalyse transglycosylation reactions in high yields, comparable to the highest values reported in both cases (Rémond *et al.*, 2002; Mladenoska *et al.*, 2007). In the case of Bgl, this is possibly a consequence of the low specific activity found on the methylated substrate in contrast to the high affinity for activated glucose in the form of pNPG, together with the convenient activation of the enzyme in the presence of methanol. Further work will relate to the use and specificity of these enzymes for the obtention of alkyl-glycosides and other glycoconjugates for biosynthetic applications.

V. Validation of a constraint-based model of *Pichia pastoris* metabolism

The aim of this chapter is twofold. On one hand, to propose a constraint-based model of the metabolism of the yeast *P. pastoris*. Besides its intrinsic interest, this model will be used in the following chapter as a basis for the prediction of recombinant protein production. On the other hand, the second goal of the chapter is to apply the possibilistic approach developed by Llaneras and coworkers (2009) for the validation and exploitation of the proposed model. For this purpose, the consistency of the model has been evaluated against several literature datasets corresponding to chemostat cultures of different recombinant *Pichia* strains. Its predictive capacity has been confirmed with the calculation of theoretical yields and the estimation of biomass growth rate. Furthermore, the model is capable to predict the qualitative behaviour of significant intracellular fluxes on the basis of measurable extracellular rates. This work sets up a methodology for the design and application of constraint-based models of microbial systems of use in scenarios where measurements are scarce.

5.1 Background

Constraint-based models have been widely used to describe and analyse the behaviour of different model microorganisms. Such analysis have been done with large, even genome-scale, reconstructions of well-characterised organisms such as *Escherichia coli* (Feist *et al.*, 2007), *Pseudomonas putida* (Nogales *et al.*, 2008) or *Saccharomyces cerevisiae* (Förster *et al.*, 2003), and also with simpler networks that consider only a few key metabolites (Lei *et al.*, 2001; Provost and Bastin, 2004).

Among yeast, *S. cerevisiae* arises as a model for eukaryotic organisms. Nielsen and Villadsen (1992) established the first stoichiometric model describing the anaerobic

metabolism of *S. cerevisiae* during growth on glucose, consisting of 37 reactions and 43 compounds. Since then, modelling of *S. cerevisiae* metabolism has been widely addressed (Cortassa *et al*., 1995; Carlson *et al.*, 2002; Jin and Jeffries, 2004; Cakir *et al.*, 2007). In 2003, Pitkänen and coworkers developed a network of 71 metabolites and 63 reactions to represent the central carbon metabolism of *S. cerevisiae* and examine cofactor requirements in a recombinant strain modified for enhanced growth on xylose. Metabolic models have also been used to enlighten differences in the metabolic flux distribution between *S. cerevisiae* and pentose fermenting yeasts, such as *Pichia stipitis* (Fiaux *et al.* 2003) or *Candida utilis* (vanGulik and Heijden*,* 1995). The models were used to accurately describe biomass yields on each substrate and also to predict flux distribution in transitional regimes from glucose to ethanol growth. Using a simple metabolic network, a comparison of the flux distribution of glucose metabolism in fourteen hemiascomycetous yeasts has also been achieved (Blank and Sauer*,* 2004).

Although *P. pastoris* is widely used as model yeast for the expression of heterologous proteins, studies related to its molecular and biochemical fundamentals are rare in comparison to *S. cerevisiae*. In the later years, several metabolic models have been published with different objectives. Ren and coworkers (2003) presented a simple stoichiometric model (14 reactions and 14 metabolites) that was combined with a bioreactor model to build a macrokinetic model of *P. pastoris* expressing recombinant human serum albumin. Other authors used a metabolic network based on a *S. cerevisiae* stoichiometric model to describe *P. pastoris* growth on glucose that included 39 reactions and 41 metabolites (Dragosits *et al.,* 2009). The model was used to perform metabolic flux calculations in cultures of *P. pastoris* producing a fragment of the recombinant antibody Fab under control of the constitutive GAP promoter at different temperatures. Compartmentation between cytoplasm and mitochondria was taken into account but ATP balance was not considered. A similar sized network was modified and validated against ^{13}C tracer experiments for *Pichia* growth on glycerol and methanol (Santos, 2008).

Çelik and coworkers (2007, 2010) presented a medium-sized network consisting of 102 metabolites and 141 reactions. The model included nitrogen and phosphate transport and metabolism. The biomass equation was established on the basis of *S. cerevisiae*

macromolecular composition. Lactate accumulation and secretion were considered. Particular biosynthesis equations were included for each amino acid. Recombinant protein production was represented depending exclusively on amino acid availability and ATP cost related to peptide bond formation. The network was used to investigate the effect of the combined used of methanol and sorbitol on the intracellular distribution of fluxes in chemostat cultures.

Two genome-scale models have also been released with the recent availability of the full genome sequence of *P. pastoris*. The first reconstruction of the genome scale metabolic model consists of 540 genes, 1254 metabolic reactions and 1147 metabolites compartmentalised into eight organelles (Sohn *et al.*, 2010). The model includes the production of heterologous protein through a set of equations that represent translation and transcription stoichiometrically. The contribution confirms the effect of oxygen limitation that had been previously described to enhance recombinant protein productivity in batch cultures. The second genome-scale model was available shortly afterwards including 668 genes, 1361 reactions and 1177 metabolites. *In silico* analysis was applied to corroborate the use of sorbitol as a promising combined carbon source with methanol. Recombinant protein production was studied in this case on the basis of amino acid biosynthetic rates (Chung *et al.*, 2010).

The established methodology to design and validate metabolic models such as the ones described involves a consistency analysis that requires the system to be overdetermined (Stephanopoulos, 1998). This becomes difficult in medium-sized systems (*e.g.* with more than 30-50 reactions) as the number of degrees of freedom increases with the complexity of the network making necessary a larger amount of measurements. ^{13}C tracer experiments can be used to solve the system, but this not always be possible or convenient, particularly when substrates that are different from glucose are used, for which isotopic tracer reactants are expensive or non-available. Besides, the strategy is sensititive to experimental errors in measurements and data reconciliation is usually required to avoid inconsistencies. Yet these models are established *a priori* upon reductionist hypothesis in comparison to genome scale networks that typically present more than 1000 reactions.

In this context, a procedure to validate a constraint-based model of *P. pastoris* metabolism must be developed that makes it possible to define when the selected model is suitable (correct for the purposes of the study) integrating uncertainty (under-determinancy in the model and errors in measurements).

5.2 Objectives

The main goal of this chapter is to select and exploit a constraint-based model of *P. pastoris* metabolism. This model must be useful for control and monitorisation applications. The validation of this model and the definition of how much information can be obtained from extracellular, measurable rates, such as substrate uptake and gas exchange are also important goals. In view that the model cannot be solved by means of classical MFA, possibilistic MFA (PS-MFA) is selected to address this problem.

The specific objectives of this chapter are:

(i) To obtain a constraint-based model of *P. pastoris* metabolism.

(ii) To set up a methodology useful to validate and design constraint-based models suitable for monitorisation and control purposes in lacking-data scenarios.

(iii) To evaluate the consistency of the model in relation with experimental datasets of measurable, external rates of substrate consumption, product formation and biomass growth, and the use of PS-MFA.

(iv) To prove the predictive capacity of the model to provide valuable interval estimates of biomass growth rate.

(v) To verify if such a model can provide valuable interval estimates of the internal flux distribution, in comparison with ^{13}C conditioned MFA results.

5.3 Results

5.3.1 Constraint-based metabolic model

In order to validate the consistency of experimental datasets, for monitorisation purposes, and to study the underlying physiology of the microbial culture at a given state, a constraint-based model of *P. pastoris* metabolism is required. For this purpose, a stoichiometric model was assembled taking into account the published networks of *P. pastoris* central carbon metabolism, and particularly, the reaction set for *P. pastoris* growth on glucose (Dragosits *et al.*, 2009). The stoichiometric model was extended to account for growth on other relevant substrates, such as glycerol and methanol (Santos, 2008). A scheme of the metabolic network used is depicted in Figure 39. This is a simplified representation whose objective is not to accurately describe the full biochemical pathways of the yeast but to generate a framework in which to apply methodologies of interest aimed to process analysis, monitoring and control.

The network provides a schematic representation of the main catabolic pathways of the yeast, including Embden-Meyerhoff-Parnas pathway, tricarboxylic acid cycle, pentose phosphate and fermentative pathways. Despite not being a fermentative yeast, *P. pastoris* is known to produce ethanol and acetate under hypoxemic conditions and this has been shown to worsen recombinant protein production both in shake flasks and bioreactors (Zhang *et al.*, 2004).

Key branchpoint metabolites such as NAD, NADP, AcCoA, oxalacetate and pyruvate have been considered in distinct cytosolic (cyt) and mitochondrial (mit) pools (Dragosits *et al.*, 2009). No other cellular compartments (such as peroxisomes) were introduced. Notice that ATP equivalents formed or required in each of the reactions are not considered at this point (see Chapter VI for further information). $FADH_2$ is considered to be equivalent to NADH to all effects and particularly in equation 23. Equation 9 represents gluconeogenesis. Transport from cytosolic to mitochondrial AcCoA is not considered (Dragosits *et al.*, 2009). Due to the fact that the activity of the malic enzyme is not significant, glyoxilate shunt is not considered either (Solà *et al.,* 2004).

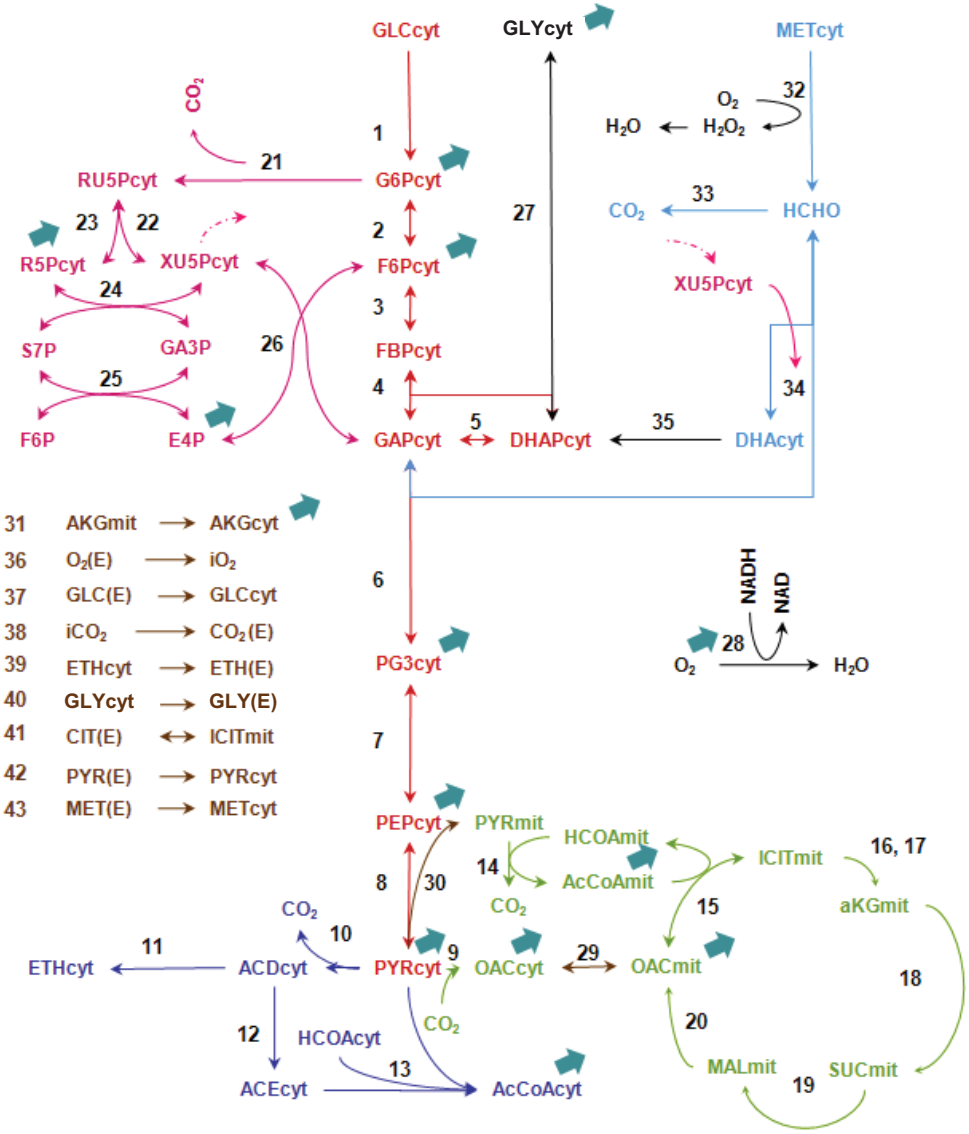


Figure 39. Metabolic network of *P. pastoris* central carbon metabolism. Biomass precursors are marked by arrows (see Table 27).

With regard to anabolism, a mean biomass equation was used that is derived from the composition of the yeast and the equations for the generation of each of its constituents from pool biochemical precursors (Pitkänen *et al.*, 2003; Dragosits *et al.*, 2009). The equation summarises the generation of macromolecular compounds such as carbohydrates, lipids, nucleic acids and proteins.

The model considers 45 compounds and 44 metabolic reactions that are shown in Tables 25-27. The balanced growth condition can be applied to 36 metabolites, resulting in 8 degrees of freedom. The corresponding 36x44 stoichiometric matrix N and the vector of reactions reversibilities—the diagonal of matrix D—are given in Table 28. With regard to constraints irreversibility is assumed for all reactions except for reaction number 2-8; 15, 22-27, 29 34 and 41 in order to account for glycerol uptake (Dragosits *et al.*, 2009). The matrices N and D define the constraint-based model that will be used in the following sections.

Table 25. Extracellular, non balanced, metabolites.

Biom	Biomass
Cit(E)	Citric acid
CO2(E)	Carbon dioxide
EtOH(E)	Ethanol
GLU(E)	Glucose
GLY(E)	Glycerol
Met(E)	Methanol
O2(E)	Oxygen
Pyr(E)	Pyruvic acid

Table 26. List of intracellular, balanced metabolites.

ACCOAcyt	acetyl-coenzyme-A
ACCOAmit	acetyl-coenzyme-A
ACDcyt	acetaldehyde
ACEcyt	acetate
AKGcyt	alpha-ketoglutarate
AKGmit	alpha-ketoglutarate
DHAcyt	dihydroxyacetone
DHAPcyt	dihydroxyacetone phosphate
E4Pcyt	erythrose-4-phosphate
EtOH cyt	ethanol
F6Pcyt	fructose-6-phosphate
FBPcyt	fructose-1,6-biphosphate
G6Pcyt	glucose-6-phosphate
GAPcyt	glyceraldehyde-3-phosphate
GLCcyt	glucose
GLYcyt	glycerol
HCHOcyt	formaldehyde
ICITmit	isocitrate
iCO2	carbon dioxyde
iO2	oxygen
MALmit	malate
MeOHcyt	metanol
NAD	nicotinamide adenine dinucleotide, oxidised
NADP	nicotinamide adenine dinucleotide phosphate, oxidised
NADH	nicotinamide adenine dinucleotide, reduced
NADPHcyt	nicotinamide adenine dinucleotide phosphate, reduced
NADPHmit	nicotinamide adenine dinucleotide phosphate, reduced
OAAcyt	oxaloacetate
OAAmit	oxaloacetate
PEPcyt	phosphoenolpyruvate
PG3cyt	3-phosphoglycerate
PYRcyt	pyruvate
PYRmit	pyruvate
R5Pcyt	ribose-5-phosphate
RU5Pcyt	ribulose-5-phosphate
S7Pcyt	sedoheptulose-7-phosphate
SUCmit	succinate
XU5Pcyt	xylulose-5-phosphate

mit- mitochondrial, cyt- cytosolic

Table 27. List of reactions included in the model.

Glycolysis	**1.** 1 GLCcyt → 1 G6Pcyt
	2. 1 G6Pcyt ↔ 1 F6Pcyt
	3. 1 F6Pcyt ↔ 1 FBPcyt
	4. 1 FBPcyt ↔ 1 DHAPcyt + 1 GAPcyt
	5. 1 DHAPcyt ↔ 1 GAPcyt
	6. 1 GAPcyt + 1 NADcyt ↔ 1 PG3cyt + 1 NADHcyt
	7. 1 PG3cyt ↔ 1 PEPcyt + 1 H_2O
	8. 1 PEPcyt ↔ 1 PYRcyt
Pyruvate branch point	**9.** 1 PYRcyt + 1 iCO_2 → 1 OAAcyt
	10. 1 PYRcyt → 1 ACDcyt + 1 iCO_2
Fermentative patways	**11.** 1 ACDcyt + 1 NADHcyt → 1 ETHcyt + 1 NADcyt
	12. 1 ACDcyt + 1 NADPcyt → 1 ACEcyt + 1 NADPHcyt
	13. 1 ACEcyt + 1 HCOAcyt → 1 ACCOAcyt
TCA cycle	**14.** 1 PYRmit + 1 HCOAmit + 1 NADmit → 1 ACCOAmit + 1 iCO_2 + 1 NADHmit
	15. 1 ACCOAmit + 1 OAAmit ↔1 ICITmit + 1 HCOAmit
	16. 1 ICITmit + 1 NADmit →1 AKGmit + 1 iCO_2 + 1 NADHmit
	17. 1 ICITmit + 1 NADPmit → 1 AKGmit + 1 iCO_2 + 1 NADPHmit
	18. 1 AKGmit + 1 NADmit → 1 SUCmit + 1 iCO2 + 1 NADHmit
	19. 1 SUCmit + 1 NADmit → 1 MALmit + 1 NADHmit
	20. 1 MALmit + 1 NADmit → 1 OAAmit+ 1 NADHmit
Pentose phosphate pathway	**21.** 1 G6Pcyt + 2 NADPcyt → 1 RU5Pcyt + 1 iCO_2 + 2 NADPHcyt
	22. 1 RU5Pcyt ↔1 XU5Pcyt
	23. 1 RU5Pcyt ↔ 1 R5Pcyt
	24. 1 R5Pcyt + 1 XU5Pcyt ↔ 1 S7Pcyt + 1 GAPcyt
	25. 1 S7Pcyt + 1 GAPcyt ↔ 1 E4Pcyt + 1 F6Pcyt
	26. 1 E4Pcyt + 1 XU5Pcyt↔1 F6Pcyt + 1 GAPcyt
Glycerol formation	**27.** 1 DHAPcyt + 1 NADHcyt ↔ 1 GLYcyt + 1 NADcyt
Oxidative phosphorylation	**28.** 1 NADH + 0.5 iO_2 → 1 NAD
Transport (intracellular)	**29.** 1 OAAcyt ↔ 1 OAAmit
	30. 1 PYRcyt →1 PYRmit
	31. 1 AKGmit →1 AKGcyt
Methanol metabolism	**32.** 1 METcyt + 1/2 O2 → 1 HCHOcyt + 1 H2O
	33. 1 HCHOcyt + 2 NADcyt → 2 NADHcyt + iCO2
	34. 1 HCHOcyt + XU5Pcyt ↔ 1 DHAcyt + 1 GAPcyt
	35. 1 DHAcyt → 1 DHAPcyt
Transport (extracellular)	**36.** 1 O_2(E) → 1 iO_2
	37. 1 GLC(E) → 1 GLCcyt
	38. 1 iCO_2 →1 CO_2(E)
	39. 1 ETHcyt → 1 ETH(E)
	40. 1 GLY(E)→ 1 GLYcyt
	41. 1 CIT(E) ↔ 1 ICITmit
	42. 1 PYR(E) →1 PYR cyt
	43. 1 MET(E) → 1 METcyt
Biomass formation	**44.** 0,0033 ACCOAcyt + 0,008 ACCOAmit + 0,0266 AKGcyt + 0,0146 E4Pcyt + 0,0363 F6Pcyt + 0,0165 PG3cyt + 0,0363 G6Pcyt + 0,0000003 GLYcyt + 0,000002 iO2 + 0,0242 OAAcyt + 0,00079 OAAmit + 0,0252 PEPcyt + 0,0294 PYRmit + 0,011 R5Pcyt + 0,199 NADPHcyt + 0,056 NADPHmit + 0,0626 NAD → 1 BIOM + 0,0127 iCO2 + 0,0626 NADH + 0,0033 HCCOAcyt + 0,008 HCCOAmit + 0,199 NADPcyt + 0,056 NADPmit

Table 28. Stoichiometric matrix.

Reaction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	O2	Glc	CO2	Et	Gly	Cit	Pyr	Met	BIO
Reversible	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
GLCcyt	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
G6Pcyt	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,036
F6Pcyt	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,036
FBPcyt	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DHAPcyt	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GAPcyt	0	0	0	1	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
PG3cyt	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,017
PEPcyt	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,025
PYRcyt	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
GLYcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
NADPHcyt	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,199
iCO2	0	0	0	0	0	0	0	0	-1	1	0	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0,018
RU5Pcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R5Pcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,011
XU5Pcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
S7Pcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E4Pcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,015
OAAcyt	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,024
PYRmit	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,029
ACCOAmit	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,008
OAAmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,001
ICITmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0	-1	0	0	1	0	1	0	1	1	1	0	0	0	0	0	0	-1	-1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0,063
AKGmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
NADPHmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,056
AKGcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-0,027
SUCmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MALmit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACDcyt	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACEcyt	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACCOAcyt	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,003
iO2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,5	0	0	0	-0,5	0	0	0	1	0	0	0	0	0	0	0	-2E-5
EtOH cyt	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
MeOHcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0
HCHOcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0
DHAcyt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0

5.3.2 Structural analysis

The condition number for the stoichiometric matrix shown in Table 28 is 13, and thus the system can be considered to be well-conditioned (Stephanopoulos, 1998). The model considers 45 compounds and 44 metabolic reactions that are shown in Tables 25-27. The balanced growth condition can be applied to 36 metabolites, resulting in 8 degrees of freedom (44 fluxes and 36 linear equations). Note that although there are 9 potentially measurable fluxes, these measurements introduce only 7 linear independent additional constraints. In this way, even if all external fluxes were known, the system remains under-determined with 1 degree of freedom according to Klamt and coworkers (2002).

5.3.3 Validation against literature datasets

In order to validate the former network, the following strategy was devised. Several data subsets coming from independent chemostat studies of recombinant protein production in *P. pastoris* were compiled from the literature. The datasets correspond to external, measurable fluxes of substrates (glucose, glycerol and methanol), gas exchange (oxygen and carbon dioxide), byproduct formation (acetic acid, citric acid, ethanol and pyruvic acid), biomass growth and protein productivity. The set of literature scenarios is shown in Table 29.

For those available, the actual experimental elemental composition of biomass and ash content was taken into account. For the others, mean values resulting from the former were used. A general elemental composition for recombinant protein was taken from Stephanopoulos (1998). For datasets 1-19, all external rates were explicit in the referenced work, whereas for datasets 20-37 they were calculated taking into account the available information, usually biomass growth rate, experimental substrate consumption yields and substrate feeding rates.

Table 29. Literature datasets.

Ref	Code	μ	Q_{Glu}	Q_{Gly}	Q_{Met}	Q_{EtOH}	OUR	CPR	Q_p
[1]	1	3,72	0,96	0,00	0,00	0,016	2,00	2,09	0,020
[2]	2	1,88	0,00	1,09	0,00	0,00	2,16	1,56	0,000
[2]	3	2,07	0,00	0,95	0,63	0,00	2,70	1,70	0,001
[2]	4	1,72	0,00	0,74	1,48	0,00	3,90	2,10	0,014
[2]	5	2,02	0,00	0,57	2,33	0,00	4,85	2,21	0,024
[2]	6	6,17	0,00	2,75	0,00	0,00	3,62	2,35	0,000
[2]	7	6,18	0,00	2,77	1,87	0,00	7,19	4,18	0,001
[2]	8	6,24	0,00	2,23	2,73	0,00	7,20	3,60	0,012
[3]	9	5,74	1,51	0,00	0,00	0,00	2,71	3,18	0,000
[3]	10	1,60	0,00	0,00	6,31	0,00	7,56	3,44	0,613
[4]	11	2,32	0,00	0,67	2,01	0,00	3,21	1,77	0,012
[4]	12	2,32	0,00	0,51	2,49	0,00	3,46	1,89	0,020
[4]	13	2,32	0,00	0,43	2,73	0,00	3,58	1,97	0,019
[4]	14	2,32	0,00	0,31	3,09	0,00	3,76	2,09	0,021
[4]	15	2,32	0,00	0,28	3,18	0,00	3,79	2,09	0,021
[4]	16	2,32	0,00	0,18	3,49	0,00	3,96	2,17	0,021
[4]	17	2,32	0,00	0,13	3,62	0,00	4,02	2,21	0,022
[4]	18	2,32	0,00	0,11	3,69	0,00	4,06	2,25	0,020
[4]	19	2,32	0,00	0,09	3,74	0,00	4,08	2,25	0,021
[4]	20	2,32	0,00	0,00	4,02	0,00	4,22	2,33	0,022
[5]	21	0,39	0,00	0,27	0,38	0,00	n.a.	n.a.	0,015
[5]	22	0,77	0,00	0,54	0,50	0,00	n.a.	n.a.	0,020
[5]	23	1,16	0,00	0,82	0,63	0,00	n.a.	n.a.	0,040
[5]	24	1,93	0,00	1,09	0,66	0,00	n.a.	n.a.	0,025
[5]	25	2,71	0,00	1,36	0,94	0,00	n.a.	n.a.	0,050
[5]	26	3,09	0,00	1,90	0,55	0,00	n.a.	n.a.	0,060
[5]	27	3,48	0,00	2,45	0,44	0,00	n.a.	n.a.	0,065
[6]	28	0,31	0,00	0,00	0,99	0,00	n.a.	n.a.	0,006
[6]	29	1,39	0,00	0,00	4,66	0,00	n.a.	n.a.	0,037
[6]	30	1,62	0,00	0,00	5,64	0,00	n.a.	n.a.	0,036
[6]	31	1,04	0,00	0,00	3,02	0,00	n.a.	n.a.	0,019
[6]	32	1,20	0,00	0,00	3,39	0,00	n.a.	n.a.	0,030
[6]	33	1,93	0,00	0,00	4,34	0,00	n.a.	n.a.	0,021
[6]	34	1,31	0,00	0,00	2,23	0,00	n.a.	n.a.	0,032
[6]	35	1,66	0,00	0,00	4,02	0,00	n.a.	n.a.	0,025
[6]	36	0,54	0,00	0,00	1,73	0,00	n.a.	n.a.	0,025
[6]	37	0,66	0,00	0,00	1,67	0,00	n.a.	n.a.	0,012
[7]	38	1,35	3,24	0,00	0,00	0,00	n.a.	n.a.	0,125

[1] Dragosits *et al.*, 2009, [2] Solà *et al.*, 2007, [3] Solà, 2004, [4] Jungo *et al.*, 2007, [5] D'Anjou and Daugulis, 2001, [6] Zhang *et al.*, 2004, [7] Schilling *et al.*, 2001

μ in C_{mmol}/(g·h), Q_{glu}, Q_{gly}, Q_{met}, Q_{EtOH}, OUR, CPR in mmol/(g·h), QP in mg/(g·h)

n.a. not available

The consistency of each complete dataset (for some of them gas exchange fluxes were not available) was firstly checked on the basis of C-mol balance of substrates and products including biomass. Recombinant protein was only considered as an independent product from biomass when secreted to the medium. However, its corresponding mass flux was not significant in comparison to biomass or carbon dioxide. As shown in Table 30, carbon balance is mostly consistent in the majority of the scenarios. Only in datasets 15-19, a deviation higher than 10% is found between carbon fluxes for substrates and products, although in this case, experimental specific growth rate values are not available for each of the conditions analysed. Scenarios lacking carbon dioxide production rate have been excluded from this analysis.

Table 30. Carbon balance for selected datasets.

Code	Q_{Glu}	Q_{Glyc}	Q_{Met}	ΣQsubstrates	μ	Q_{EtOH}	CPR	Q_p	ΣQproducts
1	0,96	0,00	0,00	5,78	3,72	0,016	2,088	0,020	5,86
2	0,00	1,09	0,00	3,27	1,88	0,00	1,56	0,000	3,44
3	0,00	0,95	0,63	3,48	2,07	0,00	1,70	0,001	3,77
4	0,00	0,74	1,48	3,70	1,72	0,00	2,10	0,014	3,82
5	0,00	0,57	2,33	4,04	2,02	0,00	2,21	0,024	4,23
6	0,00	2,75	0,00	8,25	6,17	0,00	2,35	0,000	8,52
7	0,00	2,77	1,87	10,18	6,18	0,00	4,18	0,001	10,36
8	0,00	2,23	2,73	9,42	6,24	0,00	3,60	0,012	9,84
9	1,51	0,00	0,00	9,06	5,74	0,00	3,18	0,000	8,92
10	0,00	0,00	6,31	6,31	1,60	0,00	3,44	0,613	5,06
11	0,00	0,67	2,01	4,02	2,32	0,00	1,77	0,012	4,09
12	0,00	0,51	2,49	4,02	2,32	0,00	1,89	0,020	4,21
13	0,00	0,43	2,73	4,02	2,32	0,00	1,97	0,019	4,29
14	0,00	0,31	3,09	4,02	2,32	0,00	2,09	0,021	4,41
15	0,00	0,28	3,18	4,02	2,32	0,00	2,09	0,021	4,41
16	0,00	0,18	3,49	4,03	2,32	0,00	2,17	0,021	4,49
17	0,00	0,13	3,62	4,01	2,32	0,00	2,21	0,022	4,53
18	0,00	0,11	3,69	4,02	2,32	0,00	2,25	0,020	4,57
19	0,00	0,09	3,74	4,01	2,32	0,00	2,25	0,021	4,57
20	0,00	0,00	4,02	4,02	2,32	0,00	2,33	0,022	4,65

Q_{Glu}, Q_{Glyc}, Q_{met}, Q_{EtOH} and CPR are given in mmol/(g·h), μ is given in C_{mmol}/(g·h), Q_P is given in mg/(g·h). ΣQsubstrates and ΣQproducts are given in C_{mmol}/(g·h).

For some of the datasets, reconciliated values of the external fluxes have been provided by the authors fully consistent with carbon balance (Santos, 2008). Nevertheless, these corrected values were not considered for validation as one of the objectives is to assess the consistency of the model without disregarding uncertainty related to experimental data.

As explained in Section 3.5, a deviation of 10% was considered in all measurements. Although experimental standard deviations were available for some of the datasets, 10% error stands for intrinsic uncertainty for these values (*e.g.* instrument calibration, offset, etc) whereas standard deviation accounts only for repeatability of measurements.

The methodology referred in Section 3.5.4 was used for further validation of each dataset. The results are shown in Table 31. In datasets 1 (corresponding to growth on glucose), 2 and 6 corresponding to growth on glycerol) and 11-15 (corresponding to growth on mixtures of glycerol and methanol), the measured data have full possibility, this is, a flux vector compatible with model and measurement exists and is enclosed by a band of 1%. Their residuals are also very low ($\phi < 0.5$). Notice that these conditions produce relatively small amounts of protein (1, 11-15), or none (2, 6). Datasets 20-37 for which oxygen and carbon dioxide exchange rates are not available are also fully possible.

Scenarios 3, 7-9 and 16-20 corresponding to cultures growing on mixtures of glycerol and methanol also show a good agreement with the model, and errors below 15% are enough to enclose full possibility values.

Datasets 4, 5 and 10 correspond to cultures growing exclusively or mainly on methanol, and producing larger amounts of protein. Their discrepancy is larger and in fact scenarios 5 and 10 are fully inconsistent. Notice that in the case of scenario 15, the carbon balance is not consistent, revealing contradictions within experimental data.

As can be seen, the majority of the scenarios analysed are in full accordance with the model (70% if only the datasets with all extracellular values available are considered), 15% seem to be highly consistent (an error lower than 20% is enough to enclose full possibility), whereas 15% of the considered datasets are rejected by the model. The datasets seem to be in larger disagreement with the model when higher methanol consumption rates are used, resulting in larger protein productivities.

Table 31. Validation of experimental datasets.

Code	μ Cmmol/(g·h)	Q_{Met} mmol/(g·h)	Q_p mg/(g·h)	Φ	P	ε for P=1
1	3,72	0,00	0,020	0,086	1,000	2%
2	1,88	0,00	0,000	0,278	1,000	7%
3	2,07	0,63	0,001	1,201	0,726	12%
4	1,72	1,48	0,014	2,815	0,253	20%
5	2,02	2,33	0,024	5,359	0,087	29%
6	6,17	0,00	0,000	0,070	1,000	1%
7	6,18	1,87	0,001	0,880	0,821	12%
8	6,24	2,73	0,012	2,340	0,323	19%
9	5,74	0,00	0,000	0,966	0,660	13%
10	1,60	6,31	0,613	4,760	0,054	24%
11	2,32	2,01	0,012	0,154	1,000	5%
12	2,32	2,49	0,020	0,268	1,000	6%
13	2,32	2,73	0,019	0,416	1,000	7%
14	2,32	3,09	0,021	0,754	1,000	10%
15	2,32	3,18	0,021	0,740	1,000	10%
16	2,32	3,49	0,021	0,945	0,887	11%
17	2,32	3,62	0,022	1,170	0,697	13%
18	2,32	3,69	0,020	1,290	0,628	13%
19	2,32	3,74	0,021	1,318	0,606	13%
20	2,32	4,02	0,022	1,555	0,494	15%
21	0,39	0,38	0,015	1E-23	1,000	<0.1%
22	0,77	0,50	0,020	1E-21	1,000	<0.1%
23	1,16	0,63	0,040	1E-69	1,000	<0.1%
24	1,93	0,66	0,025	1E-24	1,000	<0.1%
25	2,71	0,94	0,050	1E-24	1,000	<0.1%
26	3,09	0,55	0,060	1E-24	1,000	<0.1%
27	3,48	0,44	0,065	1E-24	1,000	<0.1%
28	0,31	0,99	0,006	1E-25	1,000	<0.1%
29	1,39	4,66	0,037	1E-24	1,000	<0.1%
30	1,62	5,64	0,036	1E-23	1,000	<0.1%
31	1,04	3,02	0,019	1E-24	1,000	<0.1%
32	1,20	3,39	0,030	1E-24	1,000	<0.1%
33	1,93	4,34	0,021	1E-24	1,000	<0.1%
34	1,31	2,23	0,032	1E-21	1,000	<0.1%
35	1,66	4,02	0,025	1E-24	1,000	<0.1%
36	0,54	1,73	0,025	1E-24	1,000	<0.1%
37	0,66	1,67	0,012	1E-24	1,000	<0.1%
38	1,35	0,00	0,125	1E-108	1,000	<0.1%

Additionally, one hundred supplementary scenarios with random values taken within the range of the experimental data were generated and used to assess their consistency with the model.

Table 32. Validation results of random datasets.

	μ	Q_{Glu}	Q_{Glyc}	Q_{Met}	Q_{EtOH}	OUR	CPR	Φ	P
Random 1	0-10	0-10	0-10	0-10	0-10	0-10	0-10	>10 for 99%	<0,1 for 99%
Random 2	1,5-6	0-2	0-2,7	0-2,7	0-0,1	2,1-7,2	1,5-4	>10 for 86%	<0,1 for 95%

Q_{Glu}, Q_{Glyc}, Q_{met}, Q_{EtOH}, OUR and CPR are given in mmol/(g·h), μ is given in C_{mmol}/(g·h).

As can be seen in Table 32, the combination of random values for the different variables used does not fulfil the validation, as 99% of them provide a possibility value lower than 0.1. When the values are restricted to the ranges covered by experimental data, more than 95% of the random datasets have possibility values lower than 0.1.

5.3.4 Structural analysis

Metatool (Pfeiffer *et al.*, 1999) was used to calculate the elementary modes (EM) corresponding to the described network. A set of 98 EM was obtained that were classified as shown in Figure 40, depending firstly on the predicted biomass yield and secondly on the carbon source used (glucose, methanol or glycerol). 17 EMs were found that did not result in biomass production, whereas 9 generated ethanol. No ethanol could be produced using a single substrate while growing.

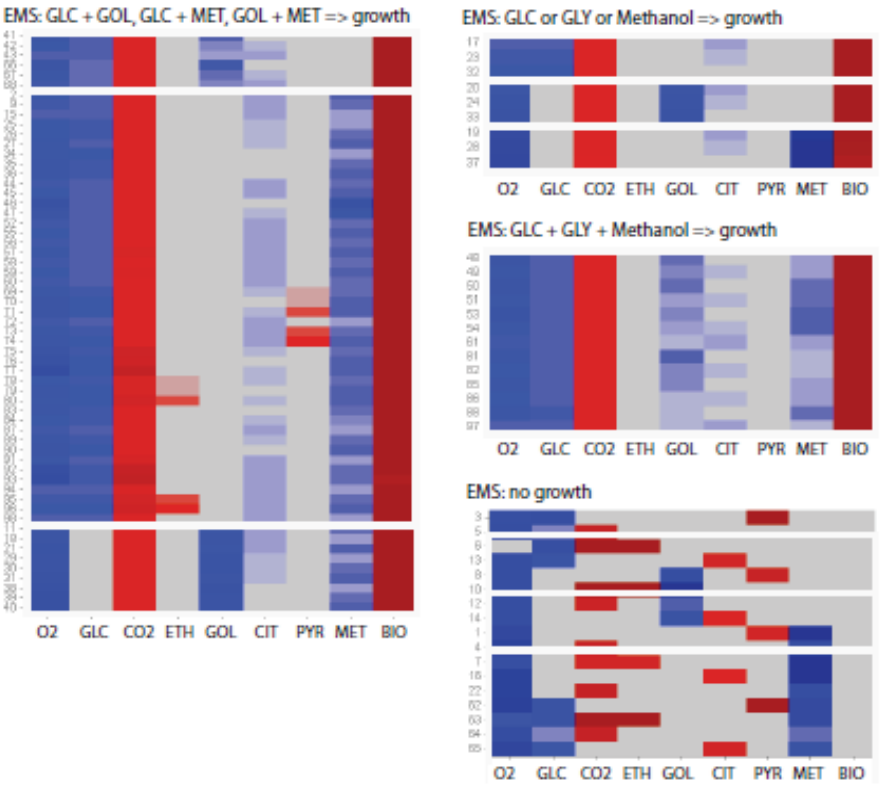


Figure 40. Schematic representation of the macroscopic equivalents of the EM (Llaneras, 2010). Substrates and products in each reaction are depicted in blue and red respectively, non participating metabolites appear in grey. Darker tones correspond to higher stoichiometric coefficients.

Carbon yield on each substrate and their combination were calculated for each EM and are shown in Figure 41 and Table 33. A maximum yield of 4.93 Cmol dcw/Cmol in the presence of glucose was found. Glucose is thus the most efficient substrate for growth, also in combination with glycerol or methanol. Methanol is the worst biomass yielding substrate.

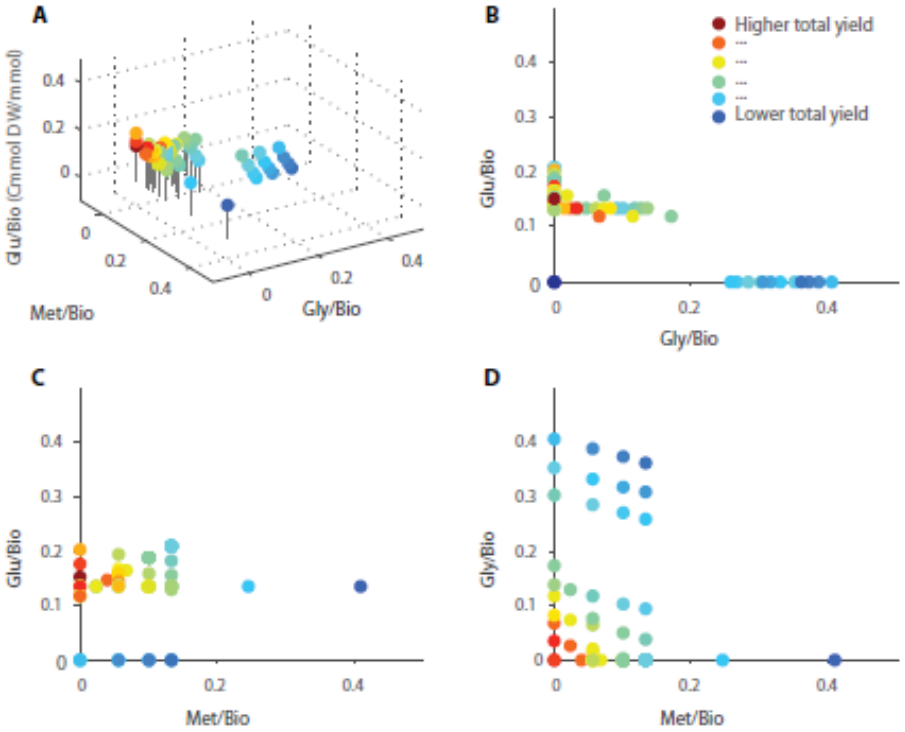


Figure 41. Biomass yields corresponding to each EM (Llaneras, 2010). Projections show the yields when one of the substrates is absent.

Table 33. Maximal theoretical growth yield predicted by the model and corresponding EM.

Glucose	Glycerol	Methanol	Y_T	EM
X			4.93	32
	X		2.46	33
		X	0.82	37
X	X		3.68	41
	X	X	2.25	38
X		X	3.98	34
X	X	X	3.47	85

5.3.5 Using the model to predict growth

Theoretical yields* versus *experimental yields

For the model to be valid, the experimental yields cannot exceed the theoretical growth yields calculated on the basis of the elementary modes. Lower values are possible, as the actual behaviour of the culture does not necessarily correspond to the optimal in growth, whereas values higher than the theoretical maxima would imply that the model does not properly represent the full potential of the yeast and that probably essential pathways are missing.

The comparison of the experimental yields for each of the scenarios is shown in Table 34. As can be seen, none of the calculated experimental values exceed the theoretical ones. For instance, the maximum yield on combined glycerol and methanol is 2.25, whereas the experimental yields are comprised between 1.51 and 0.25.

It can also be observed that the experimental yields tend to be lower than the theoretical ones. This can be attributed to several reasons:

(i) Further restrictions exist on the flux distribution than the ones considered in the model, such as thermodynamic constraints, that could lower the maximal “allowable” capabilities of the microorganism,

(ii) Recombinant protein production is not considered in the model, and thus the resources devoted to it are not represented, which can possibly lead to an increased theoretical maximum,

(iii) The cells do not necessarily follow optimal pathways in their actual behaviour.

Table 34. Comparison of experimental and theoretical yields.

Code	Q_{Glu} mmol/(g·h)	Q_{Glyc} mmol/(g·h)	Q_{Met} mmol/(g·h)	μ Cmmol/(g·h)	$Y_{Biomass/Substrates}$ Experimental	$Y_{Biomass/Substrates}$ Max. Theoretical
1	0,96	0,00	0,00	3,72	3,87	6,620
2	0,00	1,09	0,00	1,88	1,72	2,460
3	0,00	0,95	0,63	2,07	1,31	2,250
4	0,00	0,74	1,48	1,72	0,77	2,250
5	0,00	0,57	2,33	2,02	0,70	2,250
6	0,00	2,75	0,00	6,17	2,24	2,460
7	0,00	2,77	1,87	6,18	1,33	2,250
8	0,00	2,23	2,73	6,24	1,26	2,250
9	1,51	0,00	0,00	5,74	3,80	4,930
10	0,00	0,00	6,31	1,60	0,25	2,250
11	0,00	0,67	2,01	2,32	0,87	2,250
12	0,00	0,51	2,49	2,32	0,77	2,250
13	0,00	0,43	2,73	2,32	0,73	2,250
14	0,00	0,31	3,09	2,32	0,68	2,250
15	0,00	0,28	3,18	2,32	0,67	2,250
16	0,00	0,18	3,49	2,32	0,63	2,250
17	0,00	0,13	3,62	2,32	0,62	2,250
18	0,00	0,11	3,69	2,32	0,61	2,250
19	0,00	0,09	3,74	2,32	0,61	2,250
20	0,00	0,00	4,02	2,32	0,58	2,250
21	0,00	0,27	0,38	0,39	0,60	2,250
22	0,00	0,54	0,50	0,77	0,74	2,250
23	0,00	0,82	0,63	1,16	0,81	2,250
24	0,00	1,09	0,66	1,93	1,11	2,250
25	0,00	1,36	0,94	2,71	1,18	2,250
26	0,00	1,90	0,55	3,09	1,26	2,250
27	0,00	2,45	0,44	3,48	1,21	2,250
28	0,00	0,00	0,99	0,31	0,31	0,820
29	0,00	0,00	4,66	1,39	0,30	0,820
30	0,00	0,00	5,64	1,62	0,29	0,820
31	0,00	0,00	3,02	1,04	0,35	0,820
32	0,00	0,00	3,39	1,20	0,35	0,820
33	0,00	0,00	4,34	1,93	0,45	0,820
34	0,00	0,00	2,23	1,31	0,59	0,820
35	0,00	0,00	4,02	1,66	0,41	0,820
36	0,00	0,00	1,73	0,54	0,31	0,820
37	0,00	0,00	1,67	0,66	0,39	0,820
38	3,24	0,00	0,00	1,35	0,42	4,930

The calculation of maximal theoretical yields can be used to infer which substrates are potentially more efficient for biomass generation. Table 35 shows maximal theoretical yields obtained if citric acid is considered to act as a substrate. Interestingly, the yields appear to be higher in all cases if citric acid is supplied as a combined carbon source together with glucose, glycerol, or methanol. Citric acid, or citrate, is frequently used as component of batch growth medium in *P. pastoris* cultures. Some authors also refer the addition of this carbon source in the feed solution during fed-batch growth or adaptation stage in methanol (Baumann *et al*., 2008).

Table 35. Maximal theoretical yields obtained when citric acid is used as a substrate.

Glucose	Glycerol	Methanol	Y_T
X			6.62
	X		3.30
		X	1.10
X	**X**		5.95
	X	**X**	5.40
X		**X**	2.93
X	**X**	**X**	5.40

Biomass growth rate prediction

For each of the presented datasets, PS-MFA can be applied in order to estimate the biomass growth rate. This can be done by excluding the measured growth rate from the flux calculation through PS-MFA. As can be seen in Figure 42, the estimation for the growth rate in the majority of the scenarios is found to be in very good agreement with the measured one. Indeed, the estimates for the scenarios 1-9 and 11-37 corresponding to cultures growing at different growth rates, on different substrates, and coming from different literature references, are highly accurate.

In some particular scenarios, such as 10 and 38, the experimental value lies outside the bounds of the 80% possibility interval of the estimate. It is worth noting that this scenario corresponds to cultures that are producing both a significant amount of

recombinant protein and/or growing at high rate on methanol. The overestimation of biomass growth rate could be thus related to the metabolic burden caused by the production of recombinant protein which is not considered in the model.

The predictive capacity of the model is particularly noteworthy in those cases in which the uncertainty of the scenarios is high, as from 21 to 37, where the gas exchange rates are not available. Though the minimal possible value is set to zero, as non-growth is theoretically possible, since no gases may be consumed, or produced, the best possible estimate for biomass growth rate is clearly close to the experimental value.

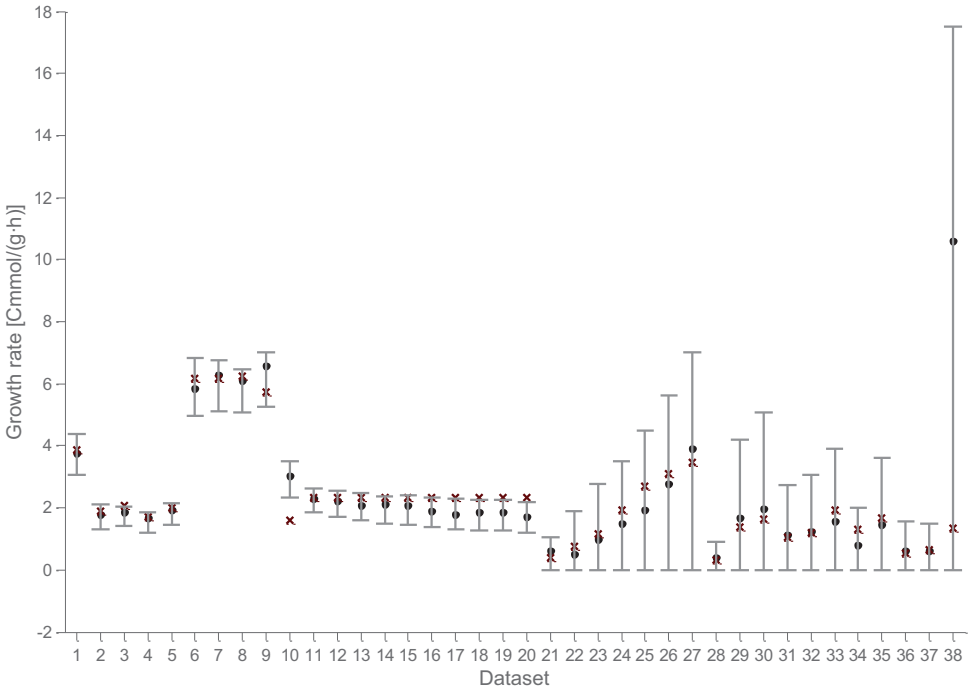


Figure 42. Comparison between estimated (black dot) and experimental (red cross) biomass growth rates.

The predictive capacity of the model for biomass specific growth rate provides further validation for this constraint-based representation. Also, we must consider that the growth rate is highly interconnected in the network structure, as the biomass stoichiometric reaction is dependent of several metabolic precursors that are, in turn, produced by different catabolic and anabolic pathways. This implies that the correspondence between biomass growth rate and substrate uptake and respiratory fluxes cannot be inferred in a straight forward way from the simple examination of the network.

5.3.6 Using the model to predict the intracellular flux distribution

In the previous sections, PS-MFA was used to exploit the constraint-based model and estimate the specific growth rate. The same procedure can be applied to estimate every non-measured flux in the network using all the extracellular, known fluxes (including biomass) as input variables of the MOC. This knowledge on the intracellular flux distributions at given conditions provides a valuable insight on the metabolism of the culture. As stated in the introductory section, this estimation cannot be addressed by means of traditional MFA, since the available measurements would be insufficient to calculate the unknown rates whereas PS-MFA is able to provide estimates is able to get an estimate partly thanks to the optimal use of the model irreversibility constraints.

As an example of the kind of results that can be obtained, all the intracellular fluxes estimated in three different cases are depicted for illustration purposes in Figures 43, 44 and 45. A complimentary estimation based on ^{13}C tracer experiments (Santos, 2008; Dragosits *et al.*, 2009) is also shown for comparison. The different analysed datasets correspond to:

- Dataset 2: growth on glycerol, fully consistent (P=1.00)
- Dataset 5: growth on glycerol and methanol, inconsistent (P=0.09)
- Dataset 9: growth on glucose, partially consistent (P=0.67)

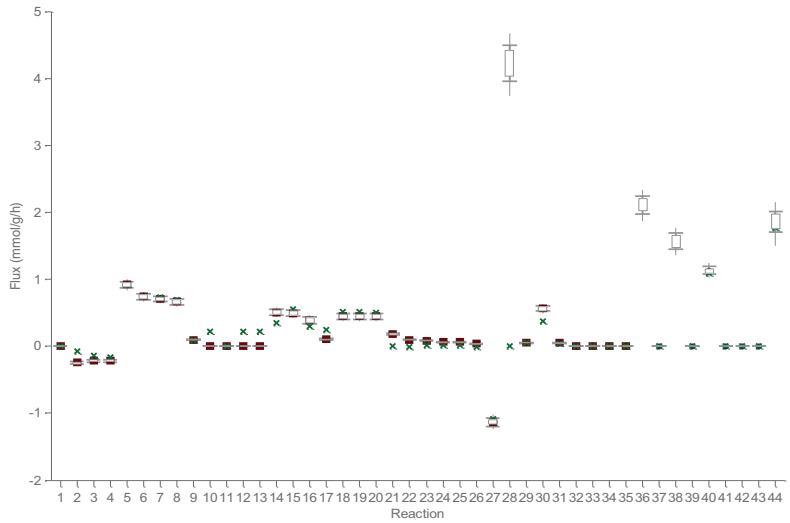


Figure 43. Intervals of 100% and 80% possibility for intracellular flux estimations and scenario 2 in Table 1. Measured, external rates are shown in red. (Santos, 2008) are shown in green.

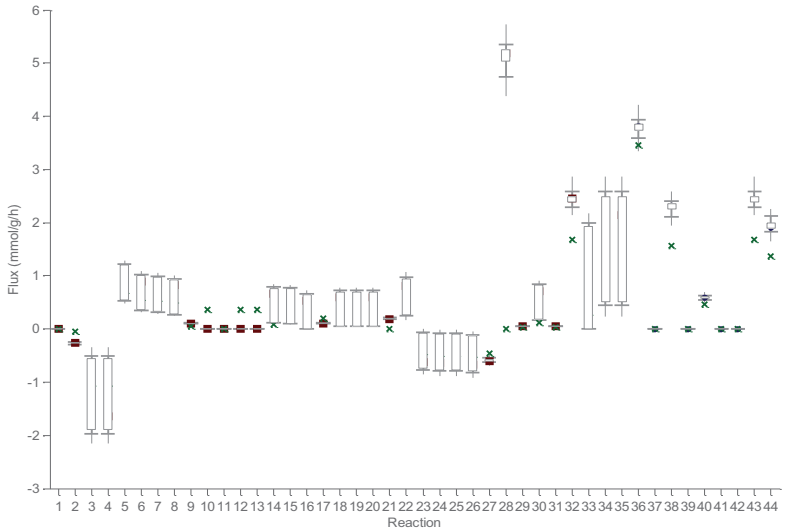


Figure 44. Intervals of 100% and 80% possibility for intracellular flux estimations and scenario 5 in Table 29. Measured, external rates are shown in red. (Santos, 2008) are shown in green.

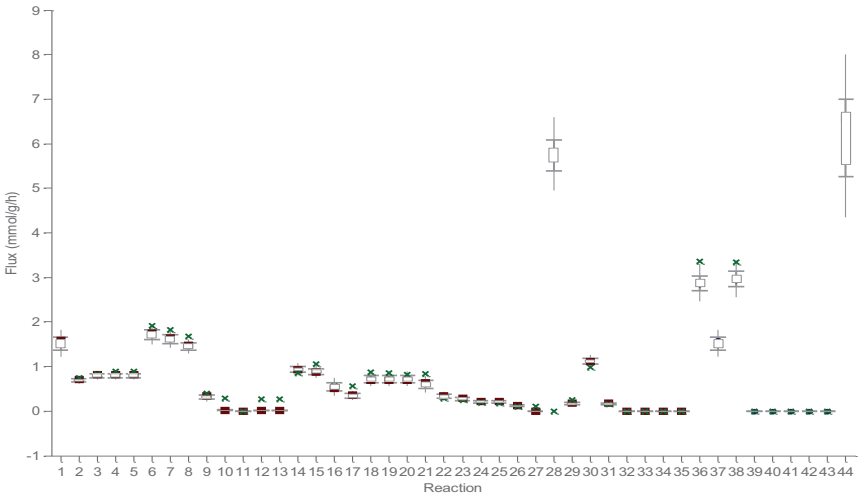


Figure 45. Intervals of 100% and 80% possibility for intracellular flux estimations and scenario 9 in Table 29. Measured, external rates are shown in red. ^{13}C values (Santos, 2008) are shown in green.

Figures 43 to 45 show how the constraint-based model provided for *P. pastoris* metabolism can be used in combination with PS-MFA to illustrate the metabolic state of the yeast in different growth conditions, using only the measurable extracellular fluxes. The quality and precision of the resulting estimates depend on the information provided.

Whereas closer estimates are produced in highly consistent, single-substrate scenarios such as dataset 2 and 9, wider bounds are found when highly inconsistent scenarios such as dataset 5 are used. This indicates that there is a larger error than expected in the measurements or in the model constraints. Another reason to this could be the increased capabilities of the network when two substrates are available. In addition, the larger uncertainty —wider estimates— is probably responsible of the larger deviation between ^{13}C-MFA and PS-MFA estimates in this case. However, the first approach is forced to provide point-wise estimates, even balancing the original measurements in a significant degree, which could be unreliable given the high degree of uncertainty for this particular dataset.

As previously stated, this estimation could not have been performed with standard MFA because the problem is underdetermined (Klamt *et al.*,2002). The stoichiometric matrix N is a 37x45 matrix with eight degrees of freedom. These particular datasets have nine known extracellular fluxes at most (see Table 29). However, these measured fluxes impose only seven linearly independent constraints, so the resultant system is redundant and underdetermined (23 of 37 unknown fluxes remain non-calculable). In this case, the estimation problem has not a unique solution, so point-wise estimates would be unreliable.

Besides semi-quantitative estimation of unknown intracellular fluxes, PS-MFA can be used to capture significant changes in relevant metabolic branch-points. The qualitative changes between Figures 43-45 in reference pathways highlight the capability of the approach to capture the metabolic state of the yeast. For example, fluxes 3-4 and 22-26 correspond to the pentose phosphate pathway. As it can be seen in Figure 46, flux 22 is distinct from zero in dataset 5, whereas it is close to zero in dataset 2. This is also the case for fluxes 23-26. Fluxes 3 and 4 are also higher in dataset 5. This implies that when methanol and glycerol are fed, reaction 22 is providing Xu5P (required for HCHO metabolism), and reactions 23-26 are reverted, whereas 3 and 4 are increased. In summary, a higher activity of the pentose phosphate pathway is predicted for dataset 5 than in dataset 2, in accordance with the data provided by Santos (2008).

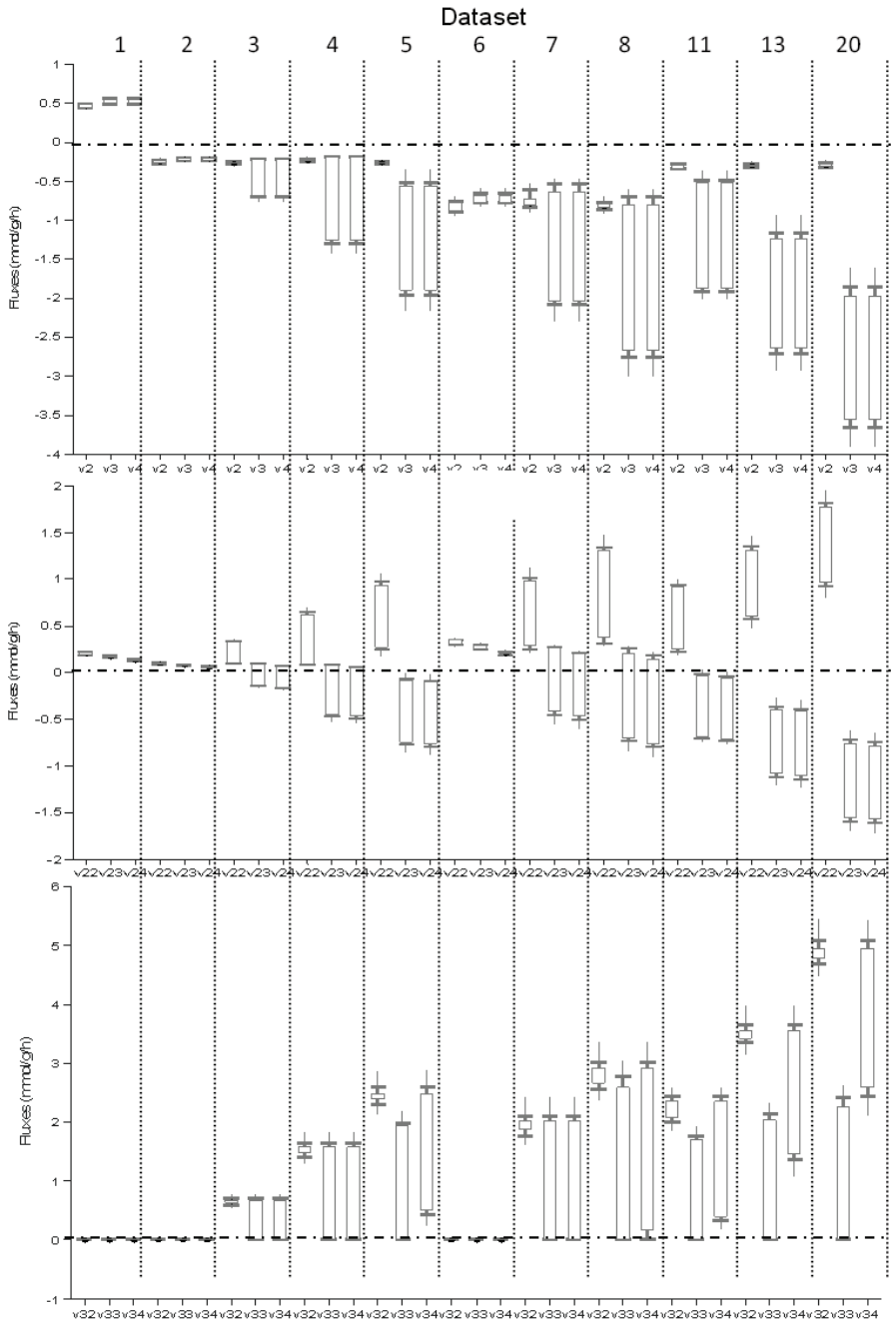


Figure 46. Estimations of relevant sets of fluxes at branchpoints for different datasets. showing intervals of 0.8, 0.5 and 0.1 possibility.

Some complimentary examples are highlighted in Figure 46, illustrating this approach lets us solve the reversibility of particular target reactions. Obvious changes, such as the inversion of fluxes v2, v3 and v4 (positive when cultures grown on glucose, and negative when they grow on methanol or glycerol), which belong to glycolysis pathway, appear in the estimation (Figure 46, upper panel). The isomerisation of R5P into Ru5P and Xu5P is also relevant: v23 inverts its direction for growing methanol fluxes, as the increasing methanol consumption requires higher amounts of Xu5P. Note how the best possible estimate is correct in sign for all scenarios except for scenarios number 8 and 12, in which the 90% band encloses the experimental value, close to zero (Figure 46, central panel). Finally fluxes v32, v33 and v34 are represented, as this branchpoint relates to use of methanol, and how methanol is split between direct oxidation and catabolic pathways. High methanol fluxes are necessarily conducted via CO_2 generation. This is captured as flux v34 is distinct from zero in scenarios number 13 and 20 (Figure 46, lower panel).

5.4 Discussion

In this chapter, a constraint-based model of *P. pastoris* metabolism has been built and validated. The goal of this representation was to provide a framework to simulate growth on different substrates, in order to assess the most efficient in biomass production, to verify if it is possible to provide estimates of biomass growth rate, and finally to correlate the pattern of extracellular fluxes to the intracellular flux distribution, when only a few measurements are available, such as substrate uptake and gas exchange rates.

The model has been derived from partial models available in the literature and is based upon first-principles and sensible hypotheses. For this, different metabolic networks previously established for the yeast have been taken into account, dealing with growth on glucose (Dragosits *et al.*, 2009) or glycerol and methanol mixtures (Santos, 2008). Genomic scale models have also been developed (Chung *et al.*, 2010; Sohn *et al.*, 2010), but their high complexity makes it difficult to use them for control and monitorisation purposes.

In this case, a novel methodology has been applied in order to validate the constraint-based model, using PS-MFA. The purpose was to set a procedure to evaluate if the model was consistent with experimental data, without assuming that model, or data, are correct, when only few measurements are available. In this way, the constraint-based model was confronted against several experimental datasets and a possibility index was calculated. Most of the experimental scenarios evaluated showed good agreement with the model, whereas complimentary scenarios built on random values for the different extracellular fluxes were rejected. This could be accomplished without circumventing the under-determinancy of the network, and with a limited set of (uncertain) extracellular measurements. It must be stressed, though, that in networks with a high number of degrees of freedom, the datasets are expected to be highly consistent, as many flux vectors will be compatible with the few available measurements. In this case, the validity of the approach will depend on whether the artificial, randomly generated scenarios are rejected by the model.

Interestingly, higher deviations were found in the cases in which protein productivity was high, which may be an indication of non-modelled phenomena, and particularly of metabolic resources devoted to heterologous production. This phenomenon has also been reported by Chung and coworkers (2010), who analysed scenarios 2-7 using a genome-scale model, and is also verified in this case in scenarios 11 to 20. The authors relate this discrepancy to possible changes in biomass composition and non-modelled phenomena, such as glycerol repression on AOX promoter. For this reason, further improvement and exploitation of the model can relate to protein productivity prediction, as this parameter, together with biomass growth, will determine the final yield and product concentration in the broth. With this regard, energetic equivalents (ATP), strongly related to protein expression, have not been yet considered and the following chapter will address this issue.

Once the model has been validated, PS-MFA can be used to estimate the specific growth rate in each of the experimental datasets. For this purpose, only the remaining extracellular known rates were used. In this way, consistent predictions were generated in all the scenarios analysed. Note that this is achieved even when an additional degree of freedom has been introduced (as one of the measurements is no longer available). This provides further validation of the model, as the biomass growth rate is only

indirectly related to extracellular measurements (substrates and exhaust gases), through different precursors in the corresponding stoichiometric reaction. In addition, the model predictive capacity for cell growth rate is an attractive target for industrial fermentation monitoring and control. Additional structural analysis, such as the reconstruction of elementary modes, enables to estimate maximal theoretical growth yields, which are also in accordance with the experimentally calculated ones.

Finally, the same approach is applied to calculate intracellular fluxes, which are found in reasonable agreement to the ones obtained including constraints derived from tracer experiments (Santos, 2008). Notice that these estimations cannot be done by means of traditional MFA because once again the measurements would be insufficient to get a determined system.

In this way, this work develops a small-sized network of *P. pastoris* metabolism that can be validated through a rational, quantitative procedure even when measurements are not sufficient to produce a determined system. This systematic, yet simple method can be used to validate and exploit constraint-based models using only extracellular measurements, also with other microorganisms of industrial interest for which substrates different than glucose are used. Finally, the model could be used in monitorisation environments (*e.g* to check liability of experimental measurements in continuous experiments), and also as a framework for optimisation.

VI. Constraint-model based estimation of recombinant protein production in *Pichia pastoris* cultures

This chapter is devoted to the estimation of protein productivity in *Pichia pastoris* cultures. This parameter is essential for the exploitation of the system based on the use of *P. pastoris* as a host for the overproduction of recombinant proteins. For this purpose, the constraint-based model of *P. pastoris* metabolism previously validated and PS-MFA have been used to design an estimator of protein production rate. The use of PS-MFA makes it possible to estimate ATP dissipation rate, which is shown to correlate both to biomass growth and protein formation rate. Cross-validation is then applied for parameter fitting, on the basis of experimental datasets corresponding to chemostat runs. The resulting system can be exploited to predict protein productivity even if only a few extracellular fluxes are known, such as substrate consumption rates, and oxygen and carbon dioxide uptake and production rates. Complimentary estimation of biomass growth has also been shown in different lacking-data conditions, a frequent situation in industry. Finally, the estimation has been applied in cultures of the previously developed recombinant strains X33-Abf and X33-Bgl. In different working conditions, the space of substrate uptake and growth rates has been explored and used to analyse this datasets and successfully predict protein productivity.

6.1 Background

Whereas biomass concentration can be often efficiently monitored online, and models and control algorithms have been developed for this purpose (Fehrenbach *et al.*, 1992; Dabros *et al.*, 2010), protein production directed monitoring approaches are rare. In particular, online measurements of protein production are typically not available;

however, this information is highly valuable to monitor the process, optimise production and implement feedback control strategies. Some protein estimators have been developed in the literature, but most of them are heuristic and case-by-case parameter fitting is required (Chae *et al.*, 2000; Nadri *et al.*, 2006).

As previously explained in Chapter II, protein productivity estimation has been frequently addressed through simple black-box models. For example, factorial design has been applied to predict protein production rate on the basis of environmental culture conditions (Holmes *et al.*, 2009). Kinetic expressions of protein productivity as a function of specific growth rate (μ) have also been broadly used (Cunha *et al*., 2004; Zhang *et al*., 2004; Jungo *et al*., 2007). However, the exact function $\pi = f(\mu)$ of this dependence is still subject to discussion (Maurer *et al*., 2006; Dietzsch *et al*., 2010) and may be highly nonlinear, monotonously increasing, decreasing, or even non-monotonous (Curvers *et al.,* 2002*;* Sinha *et al*., 2003; Ohya *et al.*, 2005; Potgieter *et al.*, 2010). It would seem that specific growth rate is only an indirect estimator for protein productivity that can only be adjusted locally to trustworthy functions. In addition, this complexity largely compromises the general applicability of this approach to different recombinant *P. pastoris* strains. Moreover, decoupling from growth can occur and steep descents in protein productivity are detected during apparent steady-state growth in chemostat conditions. This phenomenon has been related to changes in the energetic state of the yeast (Plantz *et al*., 2005) and also to proteolytic processes (Zhou *et al*., 2002).

In order to introduce a more structured representation of microbial physiology, constraint-based models can be used. Constraint-based models have been used to describe the accumulation of different metabolic products such as ethanol (Pitkänen *et al.*, 2003), penicillin (vanGulik *et al.*, 2001), polyhydroxyalkanoates (Dias *et al.*, 2008), riboflavin (Sauer *et al.*, 1997) or tryptophan (Schmid *et al.*, 2004).

However, examples of the use of constraint-based models as predictive tools for protein productivity are scarce. A straightforward approach consists in directly introducing protein productivity as an external flux in the matrix representation (as it is typically done with biomass synthesis). This can be achieved by means of lumped biochemical

equations describing precursor and energetic requirements for protein synthesis (González *et al*., 2003; Pitkänen *et al*., 2003) although quantitative prediction may be inaccurate due to its large sensitivity to the measured fluxes (Carinhas *et al*., 2011). In this case, MFA is mostly used to calculate the intracellular flux distribution and to analyse the metabolic changes induced by genetic modification (Jin *et al.*, 1997; Tsai *et al.*, 1995; Jin and Jeffries, 2004; Dragosits *et al.*, 2009).

More detailed models considering induction, proteolysis, transcription and translation phenomena have also been used in this context, but these are necessarily protein-specific (Chung *et al*., 2010; Çelik *et al*., 2010; Sohn *et al*., 2010) or require extensive parameter fitting (Mendoza-Muñoz *et al.*, 2008).

To overcome these limitations, constraints-based models have been combined with data-driven expressions for protein productivity, such as kinetic, growth-rate associated, functions described above (Ren *et al.,* 2003; Mendoza-Muñoz *et al*., 2008) or more sophisticated relationships built through projection to latent structures (Carinhas *et al*., 2011).

Metabolic burden: energetic requirements for protein overproduction

The negative impact of the overproduction of desired metabolites in microorganisms is known as “metabolic burden”. The term was introduced to describe the adverse effect in growth of protein overproduction in *E. coli* that results from the redirection of available precursors and energy from biomass synthesis to plasmid replication and recombinant gene expression (Glick, 1995). Indeed, for significant product formation, increased amounts of carbon precursors, reducing equivalents and energy have to be supplied. This drain is usually manifested as a decreased growth rate and has been captured through different modelling approaches including stoichiometric networks (vanGulik *et al.*, 2001). Other effects of metabolic burden are metabolic by-product secretion, expression of stress-response proteins and changes in the energetic state of the cell (Özkan *et al.*, 2005).

Although both biosynthetic precursors and energetic cofactors are required for product formation, microbial metabolite overproduction can be dramatically limited in terms of energetic cost. The calculation of supplementary energetic requirements has been used to successfully predict maximal theoretical product yield (vanGulik *et al.*, 2001).

ATP costs related to substrate catabolism are usually well defined; however, intracellular transport, turnover of macromolecules and regulatory mechanisms require large but not easily determined ATP quantities that are known as maintenance. ATP maintenance cost is typically divided into growth-related (Growth-Associated Maintenance Energy, GAME) and non-growth related terms (NGAME). The corresponding mathematical expression is shown below (van der Beek *et al.*, 1973)

$$r_{ATP} = Y_{ATP} \cdot \mu + m_{ATP} \quad \textbf{(36)}$$

The non-growth related maintenance term, m_{ATP}, is consumed to keep concentration and electrical potential gradients, futile cycles, and turnover of macromolecules. ATP consumption parameters can be calculated using constraint-based models and experimental data from series of chemostat cultures on different substrates together with ATP production parameters (particularly P/O ratio) corresponding to the electronic transport chain and used for estimation of biomass and product formation (vanGulik *et al.* 2001). Values found in the literature for these parameters are shown in Tables 36 and 37.

Table 36. P/O ratio values used in metabolic models for yeasts described in the literature.

P/O	Organism	Reference
2.00	*P. pastoris*, from (Nielsen *et al.*, 2003) for *S. cerevisiae*	Çelik *et al.*, 2010
1.48	*P. pastoris*	Chung *et al.*, 2010
1.50	*P. pastoris*, from (Yuan *et al.*, 1994) for *S. cerevisiae*	Ren *et al.*, 2003
1.00	*P. pastoris*	Santos, 2008
1.20	*S. cerevisiae*	vanGulik and Heijden, 1995
1.09	*S. cerevisiae*	Vanrolleghem *et al.*, 1996

Table 37. Energetic metabolism parameters for *P. pastoris* described in the literature.

m_{ATP} g/(g·h)	Y_{ATP} $mmol_{ATP}$/g dcw	Organism	Limiting substrate	Reference
0.033	14.7	*P. chrysogenum*	Glucose	vanGulik and Heijden, 1995
0.760	95.2	*P. pastoris*	Glycerol	Ren *et al.*, 2003
0.051	95.2	*P. pastoris*	Methanol	Ren *et al.*, 2003
0.035	n.a.	*P. pastoris*	Methanol	Charoenrat *et al.*, 2005
1.146	20.4	*P. pastoris*	Glucose	Chung *et al.*, 2010
0.009	n.a.	*P. pastoris*	Glycerol	Jungo *et al.*, 2007
0.014	n.a.	*P. pastoris*	Methanol	Jungo *et al.*, 2007
0.013	n.a.	*P. pastoris*	Methanol	Jahic *et al.*, 2002

n.a. not available

6.2 Objectives

In this framework, the purpose of this chapter is to expand the previously developed constraint-based model to include protein production rate, as a simple, but structured and generalised approach to protein estimation in *P. pastoris* cultures. The following particular objectives will be pursued:

(i) To define an accurate but simple estimation for protein productivity, based on PS-MFA,

(ii) To validate this estimation strategy using only a few extracellular values in literature scenarios, and to accordingly define the parameters of the model,

(iii) To apply the developed model to analyse growth, substrate consumption and recombinant protein production in cultures of the modified *P. pastoris* strains generated in Chapter IV growing on glycerol and methanol.

6.3 Results

6.3.1 Constraint-based metabolic model

The constraint-based model whose corresponding metabolic scheme is shown in Figure 47 has been used throughout this chapter. The model consistency against experimental data has been previously validated, and it has been successfully used to predict biomass growth rate on the basis of a few known external fluxes in different experimental scenarios (see Section 3.5).

In this way, the metabolic network previously described in Chapter II was extended to include ATP consumption and formation in each reaction and a flux (v_{36}) has been added, that aggregates the total specific production of ATP (vanGulik *et al.*, 2001). A P/O ratio of 1.48 was assumed for ATP generation by electronic transport chain (Chung *et al.*, 2010).

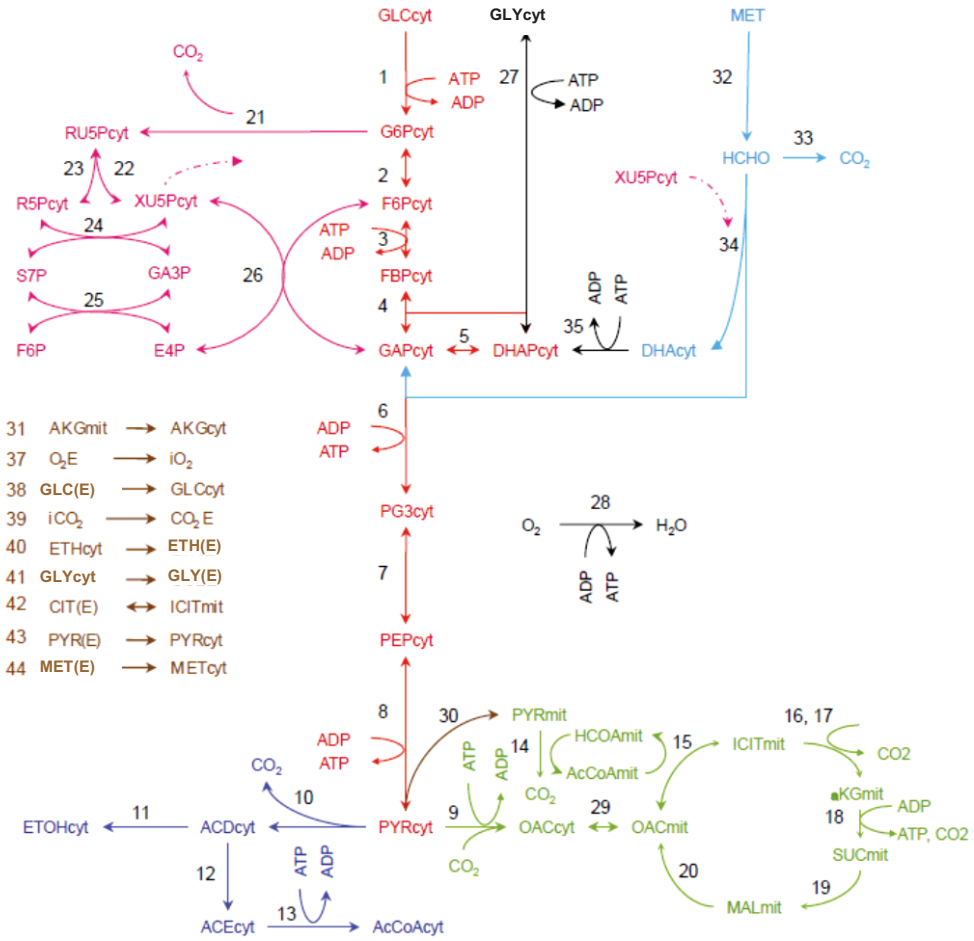


Figure 47. Metabolic network for *P. pastoris* growth on glucose, glycerol and methanol.

The network was then translated into a mathematical form as a constraint-based metabolic model (Palsson, 2006; Llaneras and Picó, 2008). In accordance to the previous section, two types of constraints were defined: (a) mass balances around the internal metabolites (assumed to be at pseudo-steady state and disregarding the effect of dilution), and (b) reversibility constraints for those reactions considered as irreversible.

Once again, the model is described with a set of model constraints (MOC), as follows:

$$\mathcal{MOC} = \begin{cases} \mathbf{N}\cdot\mathbf{v} = \mathbf{0} & \text{(a)} \\ \mathbf{D}\cdot\mathbf{v} \geq \mathbf{0} & \text{(b)} \end{cases} \quad \textbf{(37)}$$

The model has 46 compounds and 45 metabolic reactions. The reactions are assembled in the stoichiometric matrix, N, in which the a_{ij} element is the stoichiometric coefficient of the i^{th} metabolite in the j^{th} reaction, and where v is the vector of reaction rates, or fluxes, representing the mass flow through each of the *n* reactions in the network. Matrix D is a diagonal matrix with D_{ij} = 1 if the flux is irreversible and null otherwise.

The model (37) defines a space of feasible steady-state flux distributions or flux states: only flux vectors that fulfil both equations are considered valid cellular (steady) states. All internal reactions are assumed irreversible except for {2-8; 15; 22-27; 29; 34} and the exchange reaction 42 (citrate can be both consumed and produced). The balanced growth condition can be applied to 37 internal metabolites, resulting in a 37×45 stoichiometric matrix with 8 degrees of freedom.

6.3.2. Literature datasets

The set of experimental scenarios corresponding to *P. pastoris* chemostat cultures grown on methanol and glycerol mixtures has already been described in Chapter V (Table 38). Each scenario is defined by specific values of the measured fluxes: “substrates” (citrate, glycerol, methanol and oxygen) and “products” (biomass, carbon dioxide, ethanol and pyruvate) specific consumption and production rates, respectively. In this case, datasets corresponding to growth on glucose have been excluded to avoid variability associated to different gene promoters or respiratory parameters. As can be seen in Table 38, the majority of datasets are highly consistent with the model (70% are fully possible, and 91% lead a possibility value higher than 0.9), whereas only 3 of 34 datasets have a possibility lower than 0.5. Note that, at this point, resources devoted to protein production have not been included in the model.

Table 38. Selected literature datasets for model development.

Code	μ Cmmol/(g·h)	Q_{Glyc} mmol/(g·h)	Q_{Met} "	OUR "	CPR "	Q_p mg/(g·h)	v_{atp} "	Poss. (0,1)	Ref
1	1,88	1,09	0,00	2,16	1,56	0,000	7,18	1,00	[1]
2	2,07	0,95	0,63	2,70	1,70	0,001	7,40	0,74	[1]
3	1,72	0,74	1,48	3,90	2,10	0,014	8,46	0,25	[1]
4	2,02	0,57	2,33	4,85	2,21	0,024	8,46	0,08	[1]
5	6,17	2,75	0,00	3,62	2,35	0,000	11,78	1,00	[1]
6	6,18	2,77	1,87	7,19	4,18	0,001	18,77	0,82	[1]
7	6,24	2,23	2,73	7,20	3,60	0,012	15,55	0,32	[1]
8	2,32	0,67	2,01	3,21	1,77	0,012	6,77	1,00	[2]
9	2,32	0,51	2,49	3,46	1,89	0,020	6,89	1,00	[2]
10	2,32	0,43	2,73	3,58	1,97	0,019	6,96	1,00	[2]
11	2,32	0,31	3,09	3,76	2,09	0,021	6,96	1,00	[2]
12	2,32	0,28	3,18	3,79	2,09	0,021	6,92	1,00	[2]
13	2,32	0,18	3,49	3,96	2,17	0,021	6,94	0,91	[2]
14	2,32	0,13	3,62	4,02	2,21	0,022	6,99	0,71	[2]
15	2,32	0,11	3,69	4,06	2,25	0,020	7,08	0,64	[2]
16	2,32	0,09	3,74	4,08	2,25	0,021	7,05	0,61	[2]
17	2,32	0,00	4,02	4,22	2,33	0,022	7,17	0,50	[2]
18	0,39	0,27	0,38	n.a.	n.a.	0,015	3,18	1,00	[3]
19	0,77	0,54	0,50	n.a.	n.a.	0,020	5,50	1,00	[3]
20	1,16	0,82	0,63	n.a.	n.a.	0,040	7,84	1,00	[3]
21	1,93	1,09	0,66	n.a.	n.a.	0,025	8,38	1,00	[3]
22	2,71	1,36	0,94	n.a.	n.a.	0,050	9,75	1,00	[3]
23	3,09	1,90	0,55	n.a.	n.a.	0,060	13,75	1,00	[3]
24	3,48	2,45	0,44	n.a.	n.a.	0,065	18,71	1,00	[3]
25	0,31	0,00	0,99	n.a.	n.a.	0,006	2,27	1,00	[4]
26	1,39	0,00	4,66	n.a.	n.a.	0,037	10,85	1,00	[4]
27	1,62	0,00	5,64	n.a.	n.a.	0,036	13,29	1,00	[4]
28	1,04	0,00	3,02	n.a.	n.a.	0,019	6,57	1,00	[4]
29	1,20	0,00	3,39	n.a.	n.a.	0,030	7,25	1,00	[4]
30	1,93	0,00	4,34	n.a.	n.a.	0,021	7,92	1,00	[4]
31	1,31	0,00	3,62	n.a.	n.a.	0,032	7,58	1,00	[4]
32	1,66	0,00	4,02	n.a.	n.a.	0,025	7,74	1,00	[4]
33	0,54	0,00	1,73	n.a.	n.a.	0,025	3,96	1,00	[4]
34	0,66	0,00	1,67	n.a.	n.a.	0,012	3,32	1,00	[4]

6.3.3 Stoichiometric estimation of protein

As described in the previous section, several authors have included stoichiometric equations in constraint-based models of metabolism to represent the production of recombinant protein. Three different approaches have been undertaken:

(i) to include an equation representing the consumption of key precursors and energetic equivalents in a single step equivalent to biomass formation (Stephanopoulos, 1998; Pitkänen *et al.*, 2003; Santos, 2008).

(ii) to include several equations representing the individual metabolism of amino acid formation and a final equation in which protein assembly from amino acids is used (González *et al.*, 2003; Santos, 2008; Çelik *et al.*, 2010).

(iii) to include stoichiometric equations representing transcription, amino acid formation and translation (Chung *et al.*, 2010; Sohn *et al.*, 2010).

In order to assess if any of these can be applied to estimate protein productivity in an analogous way as was done for biomass specific growth rate, the following sets of equations were included in the stoichiometric network presented in Chapter II:

(P1) $0{,}10 \cdot PYR_{mit}$ + 0,05·AKG + 0.05·OAC + $0.45 \cdot OAC_{mit}$ + 0.29·PG3 + 0.05·PEP + 0,70·R5P + 0,03·E4P + 0,01·AcCoA + $0{,}01 \cdot AcCoA_{mit}$ + 1,36·NADPH + 0,09·NADPH + 0,82·NADH + $0{,}03 \cdot NADH_{mit}$ + 8,27·ATP + $0{,}15 \cdot iCO_2$ → 1 Protein, deduced from (Stephanopoulos, 1998)

(P2) 0,060·PYRmit + 0,534·NADPH + 0,055·AKGmit + 1,238·ATP + 0,050·OAAcit + 0,033·PG3 + 0,019·R5P + 0,023·AcCoAcit + 0,044·PEP + 0,033·E4P → 1 Protein, from (Pitkänen *et al.*, 2003)

The corresponding constraint-based models were assembled by combining the previous equations sets and new metabolites with the stoichiometric matrix. The models were previously validated against several literature scenarios as shown in Chapter II (please refer to Table 29) excluding protein productivity values and using PS-MFA to predict protein production flux using the remaining known fluxes and the possibilistic approach introduced in the previous chapter. The resulting possibility values, biomass estimation and protein estimation are shown in Table 39.

Table 39. Validation of extended constraint-based models including protein production (specific growth rate, μ, is expressed in Cmmol/(g·h), protein productivity, P, in mg/(g·h)). Experimental values and estimates obtained with each constraint-based models are shown.

Code	(M1) π	(P1) π	(P2) π	Experimental mg/(g·h)	(P1) mg/(g·h)	(P2) mg/(g·h)
2	1,000	1,000	1,000	0,000	0,243	1,329
3	0,726	0,726	0,726	0,001	0,000	0,000
4	0,253	0,259	0,254	0,014	0,153	0,982
5	0,087	0,087	0,087	0,024	0,004	0,022
6	1,000	1,000	1,000	0,000	0,278	7,003
7	0,821	0,924	0,836	0,001	1,472	9,429
8	0,323	0,340	0,326	0,012	0,625	4,004
9	0,476	0,660	0,660	0,000	0,000	0,000
10	0,055	0,817	0,587	0,613	5,141	32,937
11	1,000	1,000	1,000	0,012	0,703	1,945
12	1,000	1,000	1,000	0,020	0,143	1,420
13	1,000	1,000	1,000	0,019	0,275	1,133
14	1,000	1,000	1,000	0,021	0,060	0,018
15	1,000	1,000	1,000	0,021	0,071	0,234
16	0,887	0,887	0,887	0,021	0,004	0,004
17	0,697	0,697	0,697	0,022	0,016	0,016
18	0,628	0,628	0,628	0,020	0,017	0,017
19	0,606	0,606	0,606	0,021	0,016	0,016
20	0,494	0,494	0,494	0,022	0,016	0,016
21	1,000	1,000	1,000	0,015	0,799	0,399
22	1,000	1,000	1,000	0,020	0,303	11,729
23	1,000	1,000	1,000	0,040	3,067	10,738
24	1,000	1,000	1,000	0,025	0,514	15,182
25	1,000	1,000	1,000	0,050	5,327	14,106
26	1,000	1,000	1,000	0,060	1,431	16,114
27	1,000	1,000	1,000	0,065	2,075	24,640
28	1,000	1,000	1,000	0,006	616,601	1018,110
29	1,000	1,000	1,000	0,037	0,848	29,546
30	1,000	1,000	1,000	0,036	2,284	33,878
31	1,000	1,000	1,000	0,019	0,632	4,994
32	1,000	1,000	1,000	0,030	1,990	14,987
33	1,000	1,000	1,000	0,021	0,161	12,990
34	1,000	1,000	1,000	0,032	0,850	4,388
35	1,000	1,000	1,000	0,025	0,017	29,197
36	1,000	1,000	1,000	0,025	0,816	8,524
37	1,000	1,000	1,000	0,012	0,004	4,165

Experimental protein values taken for datasets 2-10 from (Solà *et al.*, 2004), datasets 11-20 from (Jungo *et al.*, 2007), datasets 21-27 from (D'Anjou and Daugulis, 2001) and datasets 28-37 from (Zhang *et al.*, 2004).

As shown in Table 39, the majority of scenarios are consistent with the model, with values close to those obtained in the previous estimation. It is worth mentioning that in some cases (7-10), the possibility of the modified models including protein equation was higher than that obtained with the original model. In these cases it can be hypothesised that these particular scenarios benefit from an additional degree of freedom that relieves the inconsistency of the first model, possibly providing an output for excess carbon source (these scenarios show high values, at the same time, of produced protein). The estimation for biomass is generally lower, showing that the resources devoted to protein production are deprived from biomass precursor pool (data not shown).

Predicted protein values are, however, overestimated in relation to experimental ones, estimates at least one log order higher in most of the cases. The predicted values do not correspond either to the qualitative evolution of the experimental values, within each dataset. In fact, the minimal and maximal experimental values do not find their equivalent in the predicted ones.

The conclusion seems to be that this kind of models can neither be used to predict qualitative changes in protein productivity nor related variations in the distribution of metabolic fluxes, as the quantitative predictions are out of range. This is not surprising as the major costs of recombinant protein production are related to folding and secretion processes whereas the ATP cost considered herein relates only to gene expression, transcription and translation. Although stoichiometric models have been used previously to predict protein productivity, this requires the precise and particularised description of the metabolic precursors required in each case (Çelik *et al.*, 2010).

On the other hand, when protein productivity is depicted against specific growth rate (data not shown), no clear relationship is seen between these variables. This further supports the idea that it is not possible to represent this relationship by a general function $\pi = f(\mu)$.

6.3.4 Estimation of ATP dissipation rate

Taking into account the system represented in (37), ATP balance can be expressed as follows:

$$\frac{P}{O}\cdot v_{28} - \sum v_i^{ATP} - K_\mu \cdot \mu - K_P \cdot v_P - m_{ATP} = 0 \quad \textbf{(38)}$$

The terms of this equation represent ATP generation in oxidative phosphorylation, ATP generation and consumption in catabolic pathways (related mainly to substrate phosphorylation), ATP consumption for growth, ATP deployed in product formation and ATP maintenance costs respectively (vanGulik *et al.*, 2001).

If ATP consumption for biomass growth is excluded from biomass equation (v_{44}) in order to collect all ATP anabolic sinks in a single flux (v_{36}), the following expression results from pseudo-steady state condition applied to ATP node:

$$\frac{P}{O}\cdot v_{28} - \sum v_i^{ATP} = v_{36} = K_\mu \cdot \mu + K_P \cdot v_P + m_{ATP} \quad \textbf{(39)}$$

In this context, the intracellular fluxes can be estimated for each particular dataset. This includes the net ATP flux, v_{ATP}, represented in the network as v_{36}. ATP dissipation flux (v_{36}) can be then estimated in the constraint model (37) for each of the scenarios (Table 38) in possibilistic terms. Thus, although v_{ATP} cannot be measured, it can be directly inferred from the available external fluxes. This is shown in Figure 48 for each experimental dataset.

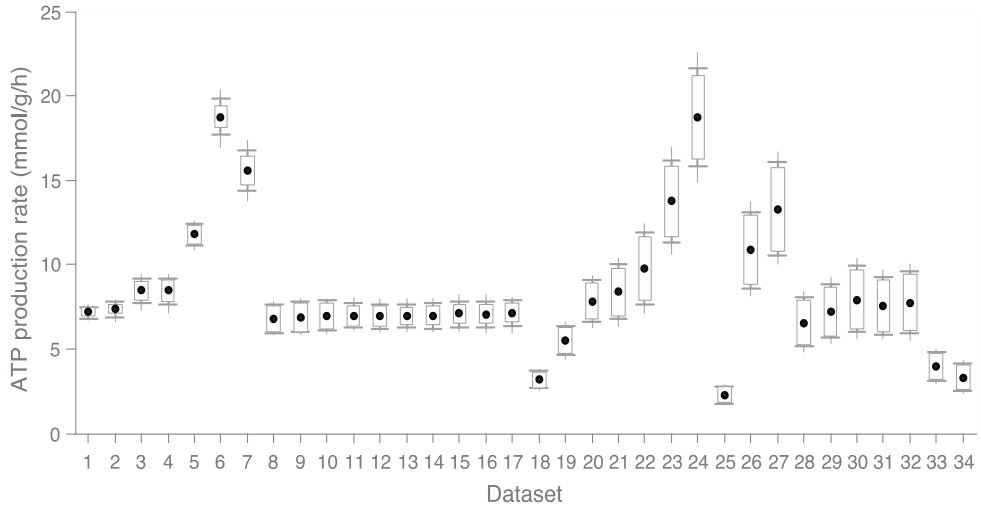


Figure 48. v_{ATP} possibilistic estimation for each dataset found in Table 38.

6.3.5 Protein estimator: definition

As already mentioned in Section 6.1, a linear mass balance is often established to represent ATP consumption in microbial systems, accounting for growth requirements and a basal need for maintenance (van der Beek *et al.*, 1973). This mass balance can be formulated as follows:

$$v_{ATP} = K_{\mu} \cdot \mu + m_{ATP} \quad \textbf{(40)}$$

Where v_{ATP} represents total ATP specific production in the network, the first term in the equality is the ATP consumption for growth, and the second one the ATP maintenance cost, assumed constant.

In addition, in recombinant hosts, heterologous protein production is also an energy demanding process (Glick, 1995). Thus, the previous equation can be extended to account for recombinant protein expression, as follows:

$$v_{ATP} = K_{\mu} \cdot \mu + K_P \cdot v_P + m_{ATP} \quad \textbf{(41)}$$

Notice that, as explained above, in our constraint-based model the total ATP specific production depends on the values of different intracellular fluxes and ATP balance in the cell can be then globally expressed as:

$$v_{ATP} = P/O \cdot v_{28} + \sum v_i^{ATP} \quad \textbf{(42)}$$

The first term in this equation represents ATP generation in oxidative phosphorylation and the second one the ATP generation and consumption in catabolic pathways (*e.g.*, substrate phosphorylation).

In this way, protein productivity can be linearly related to biomass growth, ATP production rate and maintenance. Equation (41) can then serve as a basis to build a simple, linear predictor of protein production given biomass growth rate and ATP generation as (43):

$$v_P = \beta_1 \cdot v_{ATP} + \beta_2 \cdot \mu + \beta_3 \quad \textbf{(43)}$$

where $\beta_1 = \frac{1}{K_p}$, $\beta_2 = -\frac{K_\mu}{K_p}$, $\beta_3 = -\frac{m_{ATP}}{K_p}$

Notice that in expression (43) all but v_p is known:

(i) the biomass growth is measured,

(ii) the flux of ATP can be estimated with PS-MFA from the constraint-based model and the available, measured fluxes. As previously shown, PS-MFA provides a bound estimate of v_{ATP} in most cases, although wider bounds are found when oxygen uptake rate and carbon dioxide production rate (CPR) are unknown such as in scenarios 18-34 (see Figure 48). No sinks other than ATP consumption in catabolic reactions have been introduced for this calculation. This is, ATP requirement for biomass formation has not been considered at this point in the stoichiometric matrix. This way, v_{ATP} represents “excess” ATP that should be available to cover growth, product formation and maintenance requirements.

(iii) ß$_1$, ß$_2$ and ß$_3$ are constant parameters that can be fitted as will be shown in the following section.

In this way protein productivity can be estimated through PS-MFA applied to the constraint-based model and the available measured extracellular rates.

6.3.6 Protein estimator: parameter fitting

The parameters in the previous model have been identified on the basis of the datasets previously introduced in Table 38. Then, $\{\beta_1, \beta_2, \beta_3\}$ and the equivalents $\{K_p, K_\mu, m_{ATP}\}$ are estimated by ordinary least squares (OLS) regression applied to the previous system and for the presented datasets. The results of the regression are summarised in Table 40.

To detect sample-in deviation, two different cross-validation procedures have been applied. As can be seen in Table 40, there are no qualitative changes in the parameter values and the R^2 indicator remains constant. Although only around 57% of dispersion is explained by the selected variables, the considered scenarios gather data from independent fermentation experiments and four different proteins, and both factors (inherent to experimental and strain-specific errors) are expected to introduce significant variability.

It should be pointed out that the estimation of m_{ATP}, corresponding to the maintenance rate, is not significant attending to the corresponding p-value. This should mean that the available datasets are insufficient to provide a significant estimation of this parameter. However, the estimation of ATP dissipation rate does not worsen if the constant is neglected, as shown in Table 40.

Table 40. Results of model fitting with Ordinary Least Squares (OLS) and cross validation.

		Model 1 (with constant)	Model 2 (no constant)	Model 2 Cross val. L1O	Model 2 Cross Val. 8-fold
Model fit	β_1	0.0047**	0.0050**	0.0050 ±0.0001	0.0050 ±0.0002
	(std. error)	*(0.0007)*	*(0.0005)*		
	β_2	-0.0085**	-0.0085**	-0.0085 ±0.0004	-0.0086 ±0.0008
	(std. error)	*(0.0019)*	*(0.0019)*		
	β_3	0.0033†			
	(std. error)	*(0.0045)*			
Equivalent	K_p	215.0	199.9	200.2	200.3
	K_μ	1.83	1.71	1.71	1.69
	m_{ATP}	-0.71			
Summary	Adjusted R^2	0.568	0.565	0.570±0.035	0.569±0.065
	RMSEn [a]	15.2%	15.1%	15.82% [b]	16.1% [b]

** Significant at p<0.01, * at p<0.05, † not significant at p=0.25.

[a] Normalised root mean squared error.

[b] RMSEn computed out-of-sample.

The previous results support that it is possible to correlate excess ATP rate to biomass growth and protein productivity. As expected, energetic requirements for protein formation are high in comparison to biomass growth, and a large amount of energy (around 200 mmoles ATP per mg recombinant protein, per g of cell dry weight), is required for product formation. This is in accordance with previous reports that demonstrate that protein production is limited by energetic metabolism in many organisms, and particularly in *P. pastoris* (Glick, 1995; Heyland *et al.*, 2011). It is also noticeable that ATP yield on biomass is consistent with the values reported in the literature for this coefficient in microbial cultures (Heijden *et al.*, 1994; Ren *et al.*, 2003).

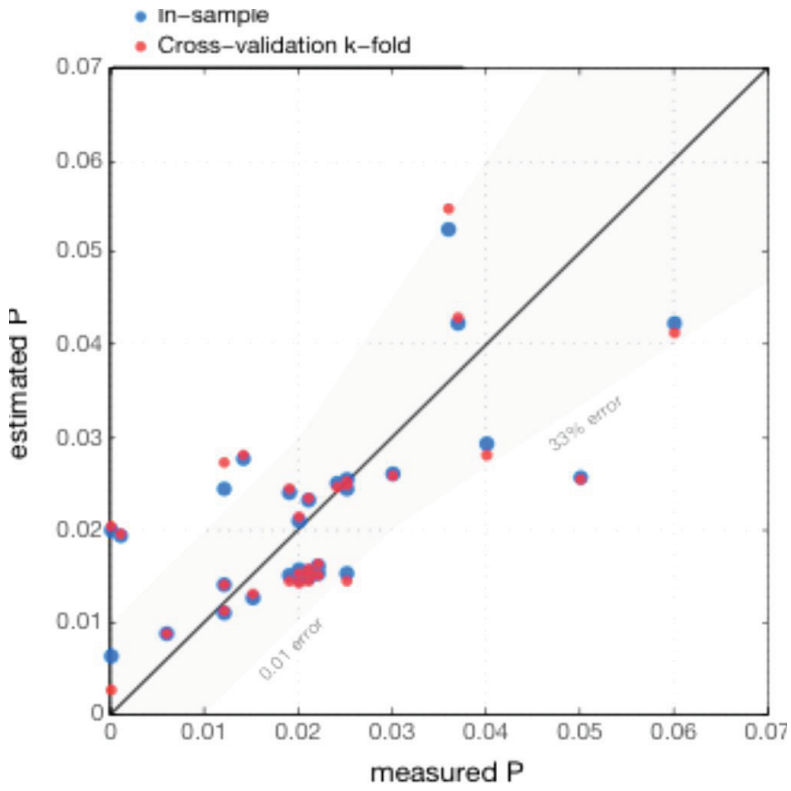


Figure 49. Estimation based on cross-validation (blue dots) is also consistent with the one obtained including the sample in the analysis (red dots).

Protein productivity can be estimated from known excess ATP rate and biomass growth. Figure 49 shows the agreement between protein estimates and experimental data. The absolute error is lower than 0.01 for all estimates but 7 (78% of the provided datasets), and all but 5 are enclosed by 33% deviation band. The lack of enough scenarios and corresponding experimental data above 0.04 mg/g·h makes it difficult to assess the suitability of the estimation for highly productive scenarios, as two of them (datasets 22 and 23) are overestimated, whereas productivities of datasets 26 and 27 are underestimated by the model. For small experimental productivities, the model seems to overestimate protein capabilities since all the outliers are found above 0.01 error band.

6.3.7 Applications of protein model

Constraint-model based protein productivity estimation

In the previous section we have seen how the equation (26) can be used to get a reasonable estimate for protein productivity based on ATP and biomass growth. Moreover, as the established relationship is linear, protein production can be easily integrated into the constraint-based model. For this purpose, the metabolic network is modified adding a flux (column in N) that represents protein production flux v_p, in mg/(g·h). Then, ATP balance is closed adding terms of ATP requirements for biomass synthesis ($K_\mu \cdot \mu$), protein production ($K_p \cdot v_p$), and maintenance ($m_{ATP}=v_{36}$). Notice that at this point the flux v_{36} represents maintenance rate, m_{ATP}. This value has been assumed a constant zero in the following computations because, as explained above, it could not be estimated with sufficient significance with the available data. Reasonable estimates of m_{ATP} can be equally incorporated if available. The following computations have been performed using literature values for m_{ATP} and the results remain unaffected (data not shown) with the current experimental data and considered uncertainty.

In this way, protein production can be estimated as any other flux in the model with MFA given some known external flux values (*e.g.*, glucose, glycerol, methanol, gases, etc.). In this case, PS-MFA was applied to estimate the protein production in each of the 34 scenarios using all the available measures except protein (see Table 38). Measures uncertainty was considered as explained in Section 3.5.4. Three interval estimates were computed, for conditional possibilities of 1, 0.5 and 0.1. Results are given in Figure 50.

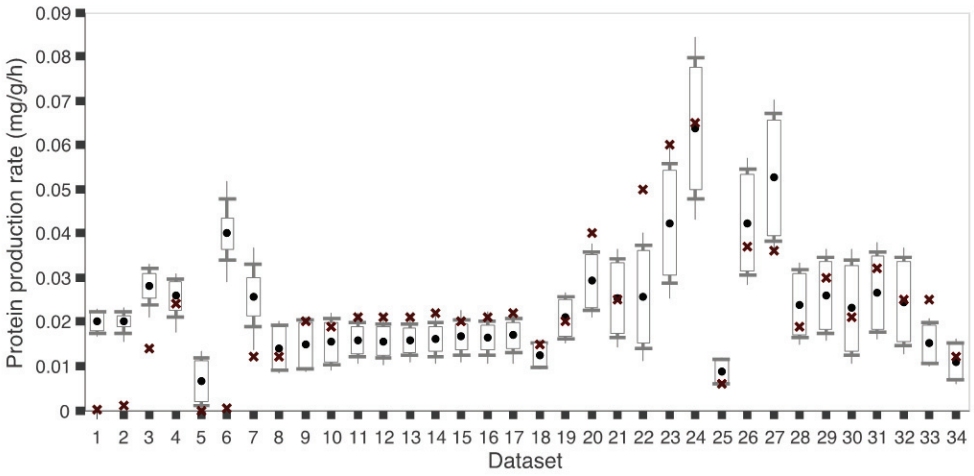


Figure 50. Protein estimation results for scenarios 1-34 found in Table 38. Crosses represent measured protein productivity values.

As can be seen in Figure 50, the estimates show a good agreement with the experimental productivity values. In addition, the estimates seems to capture, in most cases, the qualitative behaviour of different experimental conditions for each particular strain: whereas varying substrate consumption rates are predicted for a particular scenario to result in constant protein productivities (see scenarios 9-17), in other cases (see scenarios 18-24 and 25-32) different experimental conditions generate different productivities. Furthermore, the optimality tendency (local optimum in relation to specific growth rate) is correctly predicted by the model. As can be seen in Figure 50 and Table 41, whereas in scenarios 18-24, the increase in specific growth rate (from 0.39 to 3.48 $C_{mmol}/(g \cdot h)$) results in superior protein productivity, to achieve a maximum at 3.48 $C_{mmol}/(g \cdot h)$, in scenarios 25 to 34, protein productivity optimum occurs for dataset 27, at 1.62 $C_{mmol}/(g \cdot h)$, followed by dataset 26, at 1.39 $C_{mmol}/(g \cdot h)$, both corresponding to high substrate consumption rates, but not to the fastest growing cultures.

Table 41 summarises the prediction accuracy in relation to the absolute value of productivity. Values lower to 0.015 mg/(g·h) tend to be overestimated, although 87.5% are enclosed below 0.035; values between 0.015 and 0.035 are correctly predicted and

88% are estimated within the interval; finally, a majority of values (66%) above 0.035 are well estimated.

The higher accuracy within 0.015-0.035 limits is not surprising because 62.5% of the available datasets correspond to this region, so this results in a better local fit. Regarding productivity overestimation, it may due to non-modelled phenomena, such as proteolysis and folding or secretion bottlenecks, both well documented in *P. pastoris* system (14, 39). In this case, the amount of protein synthesised by the cells would be higher than the actual detected in culture supernatant, and the devoted resources (*e.g.*, substrate and oxygen consumption) and resulting products would correspond to the original protein production. With reference to the large deviation found in the estimation of protein productivities below 0.015, this effect would be more easily detected in low-productivity scenarios, for which the relative error caused by protein degradation would be higher.

Table 41. Prediction accuracy of protein production (v_p).

	Estimated <0.015		**Estimated 0.015 to 0.035**		**Estimated >0.035**	
Measured <0.015*	**4**	**(44.5%)**	4	(44.5%)	**1**	**(11%)**
Measured 0.015 to 0.035	2	(10.5%)	**17**	**(89.5%)**	0	(0%)
Measured >0.035	**0**	**(0%)**	2	(33%)	**4**	**(66%)**

* Protein production (v_p) in mg/(g·h).

Lack of measurements

The experimental datasets in Table 38 include measures of many extracellular fluxes — substrates, by-products, gases, among others—, but in industrial practice the number of measurable fluxes is much lower. For this reason, it is considered of interest to assess the predictive performance of the model when not all the external fluxes are known. To investigate this issue, protein production was estimated for the 34 datasets in nine hypothetical scenarios of data scarcity, where different subsets of measurements are

account as available. All computations were performed with PS-MFA. Since OUR and CPR are not available in datasets 17-32, this results in a higher uncertainty and wider bounds in these cases.

The results in Table 42, and Figures 51 and 52 show that accurate prediction of protein productivity is still possible would OUR, ethanol or biomass measurements be non available. This is interesting because online measurements of substrates, OUR and CPR can be easily achieved in industrial environments without sophisticated equipment, whereas biomass and protein are target output variables. The estimation of protein production can be done even if both biomass and OUR are unknown, providing that CPR is properly measured. Similarly, the estimation can also be done when biomass and ethanol are unknown, if exhaust gases are measured. The estimates are poor only when both biomass and CPR are unknown (this is not surprising because this implies that the two main carbon sinks of the model being undefined). Accurate point-wise estimates are obtained in the remaining scenarios (S1, S3, S6 and S8) although wider bounds are found for all datasets.

Table 42. Goodness of protein estimation in nine scenarios lacking measurements.

Scenario	Substrates*	By-products*	Gases		Biomass	Protein
	gly, glu, met	et, cit, pyr	OUR	CPR	μ	vp
S0 All measured	o	o	(o)	(o)	o	GOOD
S1 All but gases	o	o			o	FAIR
S2 All but OUR	o	o		(o)	o	GOOD
S3 All but CPR	o	o	(o)		o	FAIR
S4 All but ethanol	o	cit, pyr	(o)	(o)	o	GOOD
S5 All but biomass	o	o	(o)	(o)		GOOD
S6 No biomass, no gases	o	o				Only bounds
S7 No biomass, no OUR	o	o		(o)		GOOD
S8 No biomass, no CPR	o	o	(o)			BAD
S9 No biomass, no ethanol	o	cit, pyr	(o)	(o)		GOOD

(o) Gases are only available for datasets 1 to 16; not in datasets 17-32.

* Either measured or assumed to be zero.

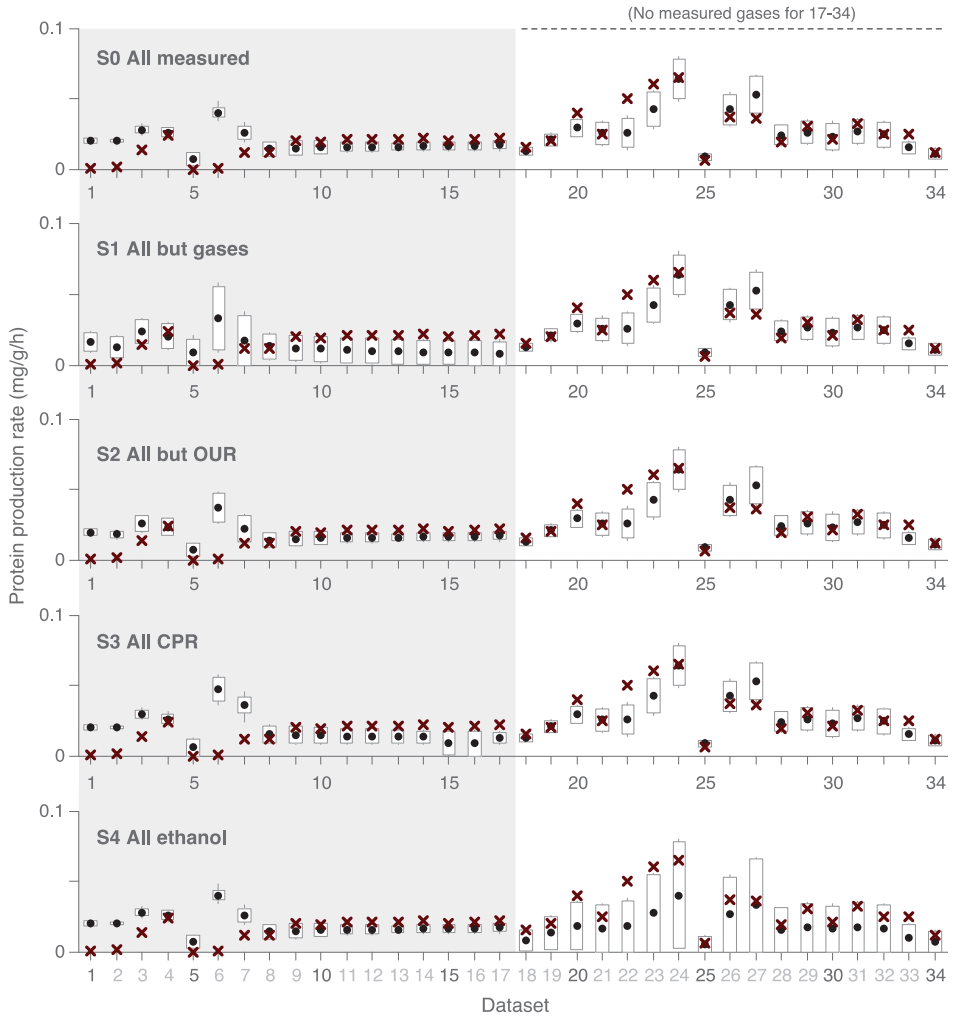


Figure 51. Possibilistic estimation of protein productivity for S0-S4 conditions found in Table 42. Bounds correspond to a possibility value of 80%.

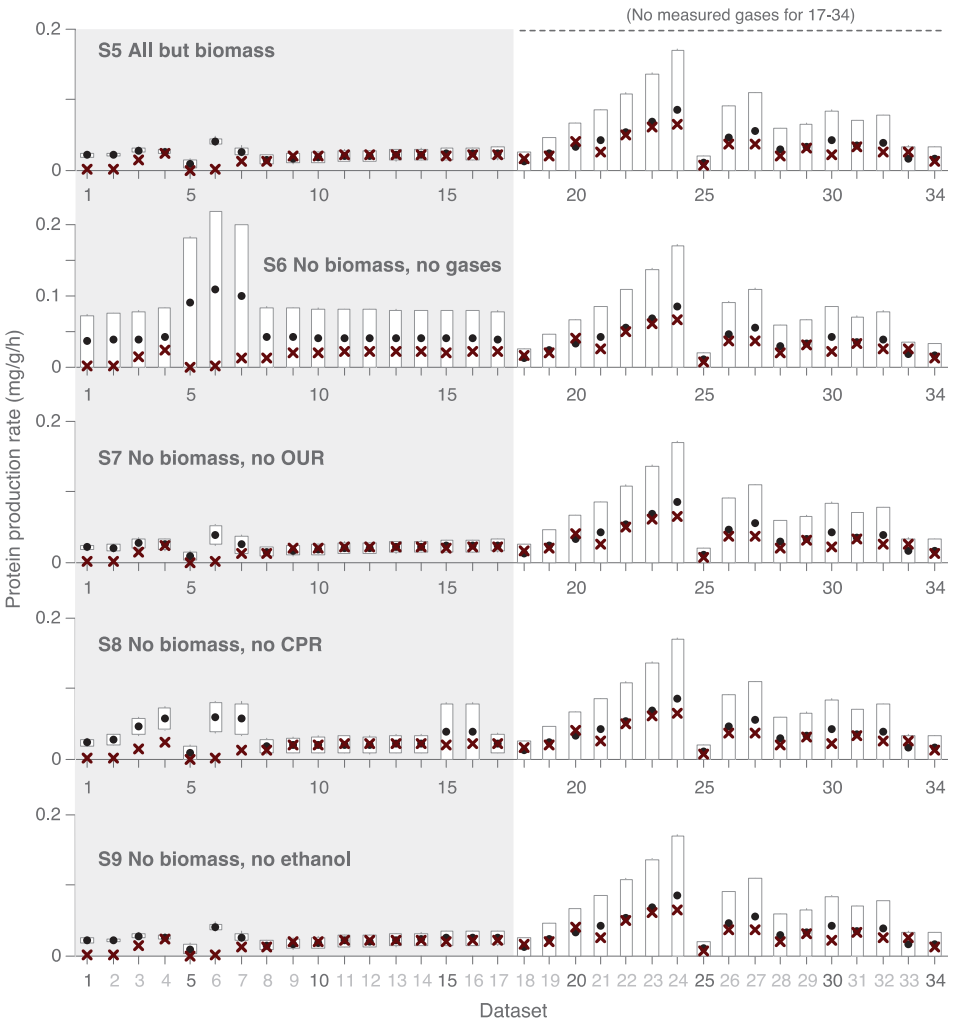


Figure 52. Possibilistic estimation of protein productivity for S5-S9 conditions found in Table 42. Bounds correspond to a possibility value of 80%.

Estimation of biomass growth rate

As stated in the previous section, online simultaneous prediction of protein and biomass productivity is interesting from an industrial standpoint. For this reason, we have analysed the accuracy of (i) biomass prediction, and (ii) simultaneous biomass and protein prediction in eight lacking data scenarios. The results are given in Table 43 and Figures 53 and 54. The most outstanding result is that useful estimates can be obtained for both protein and biomass provided that substrates, by-products, and CPR are known. In addition, the capability of the model to correctly estimate biomass when all measurements are used (scenario S10) provides a further validation of the model — simplified stoichiometry, assumed irreversible reactions, and biomass composition.

In most cases, after the experimental validation of a particular strain, the production of ethanol, pyruvate and citrate can be disregarded, particularly if aerobic conditions are established in bioreactors, this is, these fluxes could be assumed zero in most operating conditions.

Table 43. Goodness of biomass estimation lacking measurements.

Scenario	Substrates*	By-products*	Gases		Protein	Pred. biomass
	Gly, glu, met	et, pyr	OUR	v_m	v_p	μ
No biomass						
S10 All but biomass	o	o	(o)	(o)		GOOD
S11 No gases	o	o				Only bounds
S12 No OUR	o	o		(o)		GOOD
S13 No CPR	o	o	(o)			FAIR
S14 No ethanol	o	cit, pyr	(o)	(o)		FAIR
No biomass, but protein						
S20 All but biomass	o	o	(o)	(o)	o	GOOD
S21 No gases					o	GOOD
S22 No OUR	o	o		(o)	o	GOOD
S23 No CPR	o	o	(o)		o	GOOD
S24 No ethanol	o	cit, pyr	(o)	(o)	o	FAIR

(o) Gases are only available for datasets 1 to 16; not in datasets 17-32.

* Either measured or assumed to be zero.

According to Figures 53 and 54, the combined estimation of biomass growth and protein productivity is possible as long as carbon sources and gases are defined. This situation is particularly interesting in industrial fermentations, since carbon limited cultures in fed-batch mode are a standard for *Pichia* cultures, and oxygen and CO_2 measurement can be easily achieved by means of exhaust gas analysis. This would be the case in large-scale, pilot bioreactors in which a particular protein is being continuously produced.

On the contrary, bench-top systems at laboratory scale could benefit of rich protein productivity estimation if biomass growth rate can be estimated (*i.e.* by periodic sampling or absorbance-based biomass sensors) even if OUR and CPR cannot be calculated.

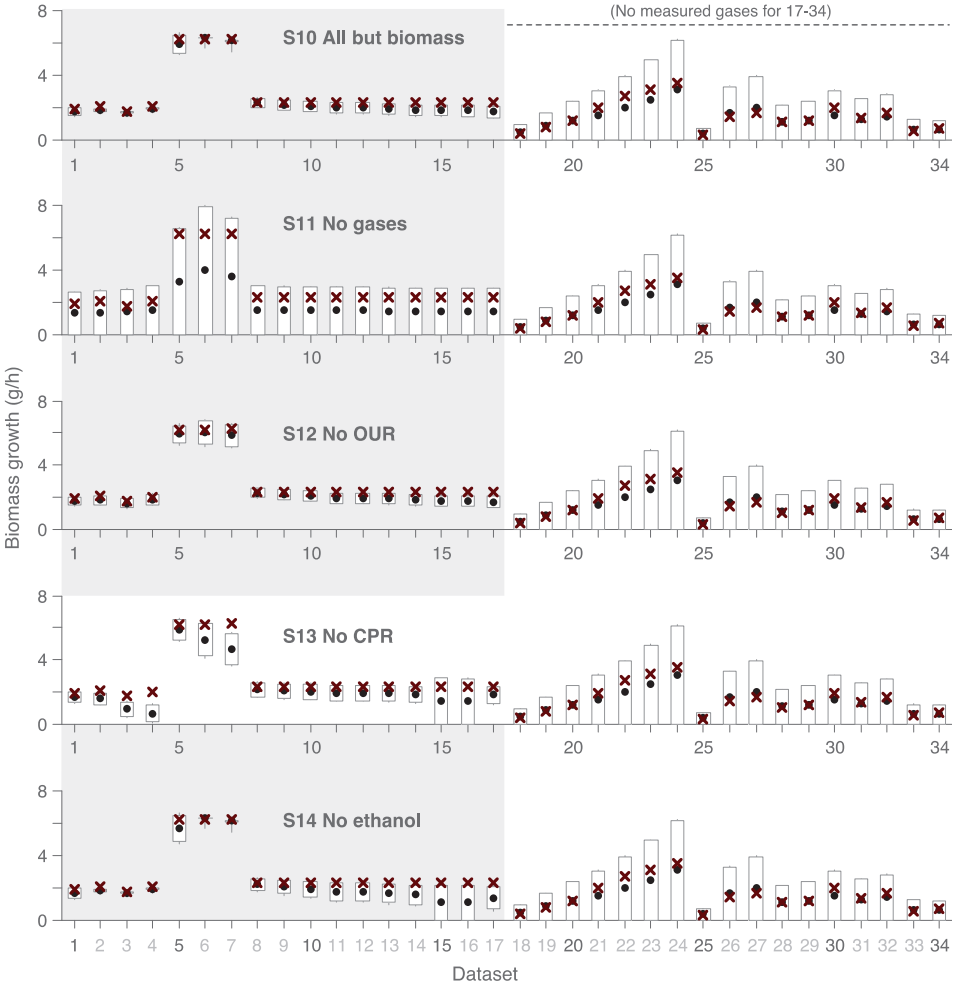


Figure 53. Possibilistic estimation of biomass growth rate for S10-S14 conditions found in Table 43. Bounds correspond to a possibility value of 80%.

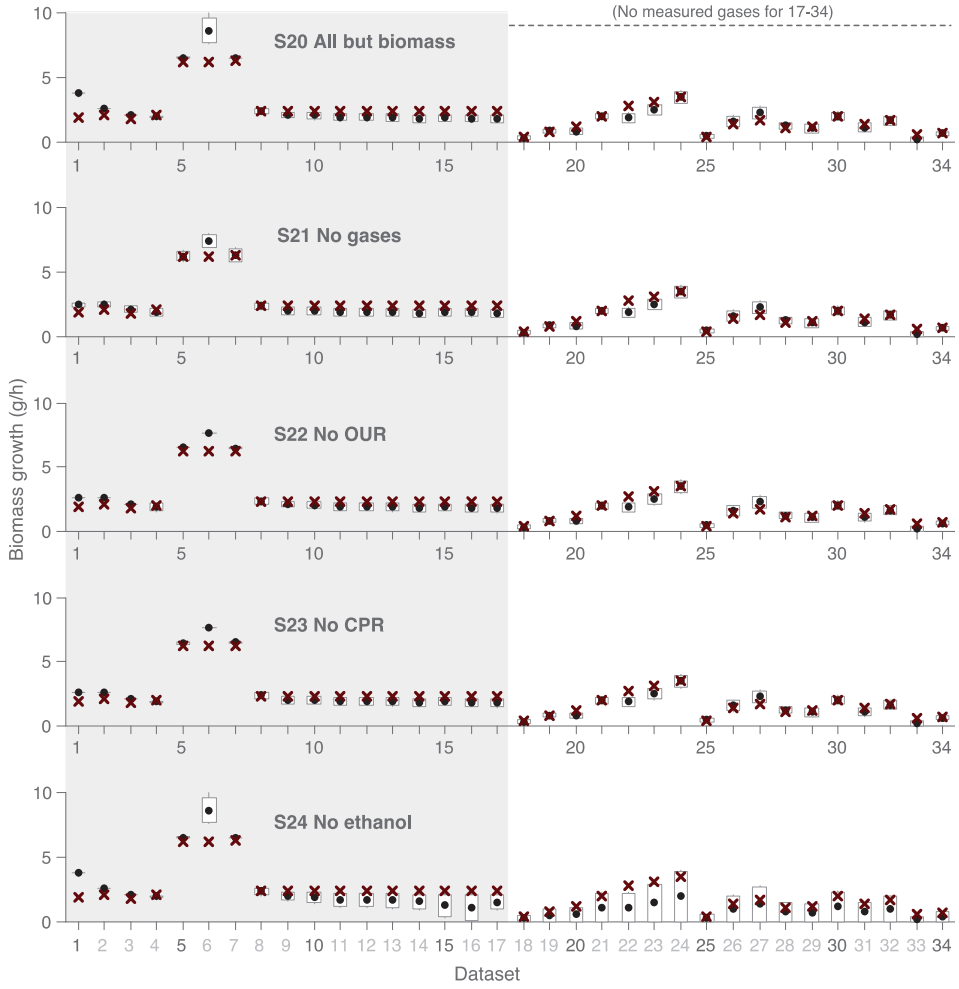


Figure 54. Possibilistic estimation of biomass growth rate for S20-S24 conditions found in Table 43. Bounds correspond to a possibility value of 80%.

6.3.8 Application of the model in X33-Abf, X33-Bgl cultures

As detailed in Section 3.2.7, several μ24 batch runs were carried out using the strains developed in Chapter IV. Different substrate concentrations were used in order to induce varying exponential phase growing conditions, in order to use the previously validated model to represent growth and recombinant protein production in X33-Abf and X33-Bgl strains. Glycerol and methanol initial concentrations used in each experiment are shown in Table 44. Three different experimental set-ups were used: (i) high initial glycerol and methanol concentrations, (ii) low glycerol initial concentration followed by high initial methanol concentration, (iii) mid initial glycerol concentration followed by high initial methanol concentration. A constant total amount carbon source was used in (ii) and (iii) set-ups.

Table 44. Experimental conditions for growing cultures analysed.

Reference	Strain	Glycerol (g/L)	Methanol (g/L)
1-1	X33-pPICzB	8.0	6.5
1-2	X33-Abf	8.0	6.5
1-3	X33-Bgl	8.0	6.5
2-1	X33-pPICzB	4.0	6.0
2-2	X33-Abf	4.0	6.0
2-3	X33-Bgl	4.0	6.0
3-1	X33-pPICzB	6.5	3.5
3-2	X33-Abf	6.5	3.5
3-3	X33-Bgl	6.5	3.5

Glycerol and methanol consumption, and biomass and protein production rates were determined by sampling and offline analysis as described in Section 3.5.2. Oxygen consumption was calculated from dissolved oxygen concentration, measured online. Temporal series of biomass and substrate concentration are shown in Figures 55, 57 and 59 for each experimental condition. Calculated volumetric oxygen uptake rate is shown in Figures 56, 58 and 60.

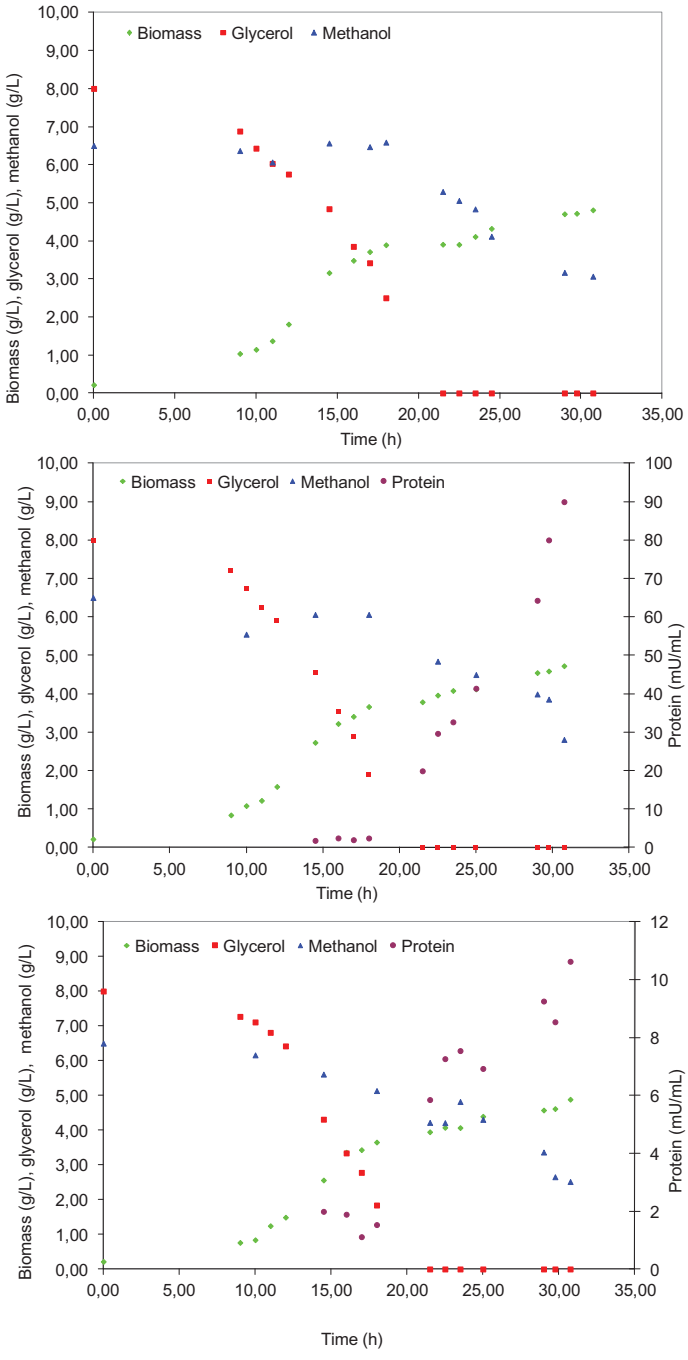


Figure 55. Batch cultures of X33-pPICzB, X33-Abf and X33-Bgl (conditions 1.1, 1.2, 1.3).

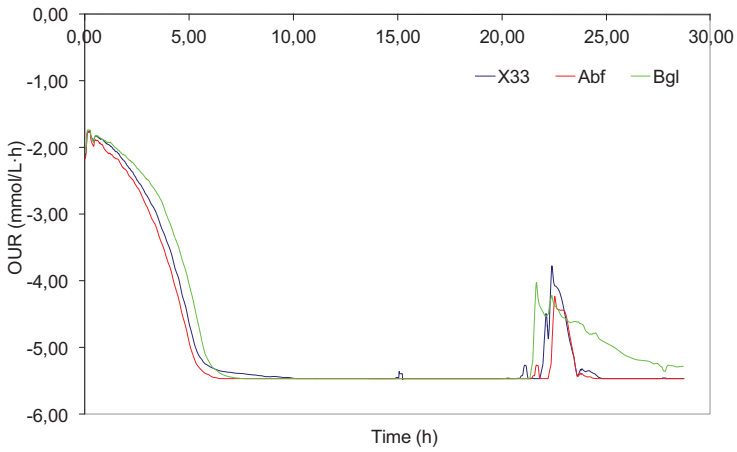


Figure 56. OUR evolution in X33-pPICzB, X33-Abf and X33-Bgl cultures (conditions 1.1, 1.2, 1.3).

As can be seen in Figure 55, all strains grow exponentially on glycerol and methanol. Glycerol is preferred over methanol, and is exhausted after approximately 20 hours. Oxygen is quickly consumed during glycerol early exponential phase and growth continues under oxygen limiting conditions. A pO_2 spike indicates complete glycerol depletion. Next, oxygen concentration decreases again as methanol is consumed (Figure 56).

Both specific growth rate and overall yield on substrate are lower on methanol than on glycerol for all strains, in accordance to previous studies (Cos *et al.*, 2006a).

Unlike X33-Abf, significant differences can be observed in the growth pattern of X33-Bgl in comparison to the wild-type strain, X33-pPICzB. Oxygen concentration decreases less sharply and oxygen consumption rate is, consequently, lower. This also correlates with a lower methanol consumption rate. A potential toxic effect of Bgl expression is thus detected when methanol is used as sole carbon source.

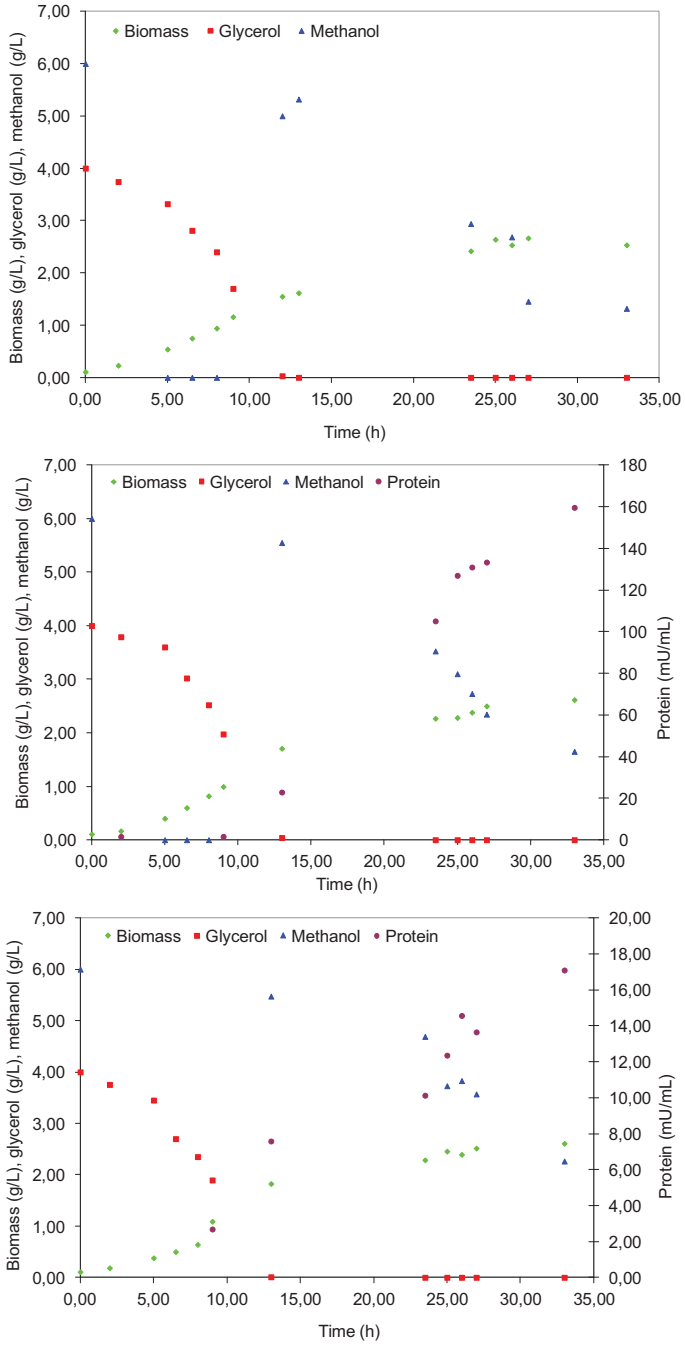


Figure 57. Batch cultures of X33-pPICzB, X33-Abf and X33-Bgl (conditions 2.1, 2.2, 2.3).

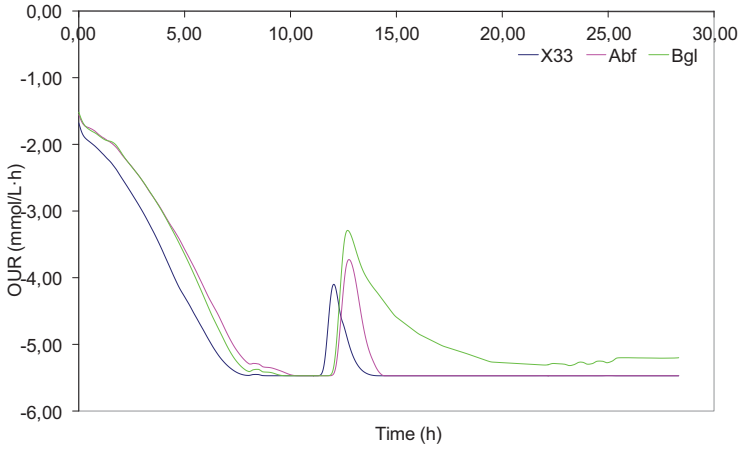


Figure 58. OUR evolution in X33-pPICzB, X33-Abf and X33-Bgl cultures (conditions 2.1, 2.2, 2.3).

As can be seen in Figure 57, methanol consumption can be detected before glycerol is fully exhausted. This non-strict diauxic behavior has been described in the literature and can be related to an only partial repression of AOX promoter at lower glycerol concentrations (Inan and Meagher, 2001). Indeed, combined glycerol and methanol feeding strategies can be used to improve *P. pastoris* cultures volumetric productivity. In this case, glycerol and methanol are simultaneously fed and consumed in semi-continuous mode (fed-batch mode) in order to maintain limiting concentrations of both substrates.

Once again, an anomalous oxygen consumption pattern is detected in the case of X33-Bgl (Figure 57), together with lower methanol consumption (Figure 58). Although growth is apparently unaffected, this could manifest the toxicity of Bgl production in this expression system.

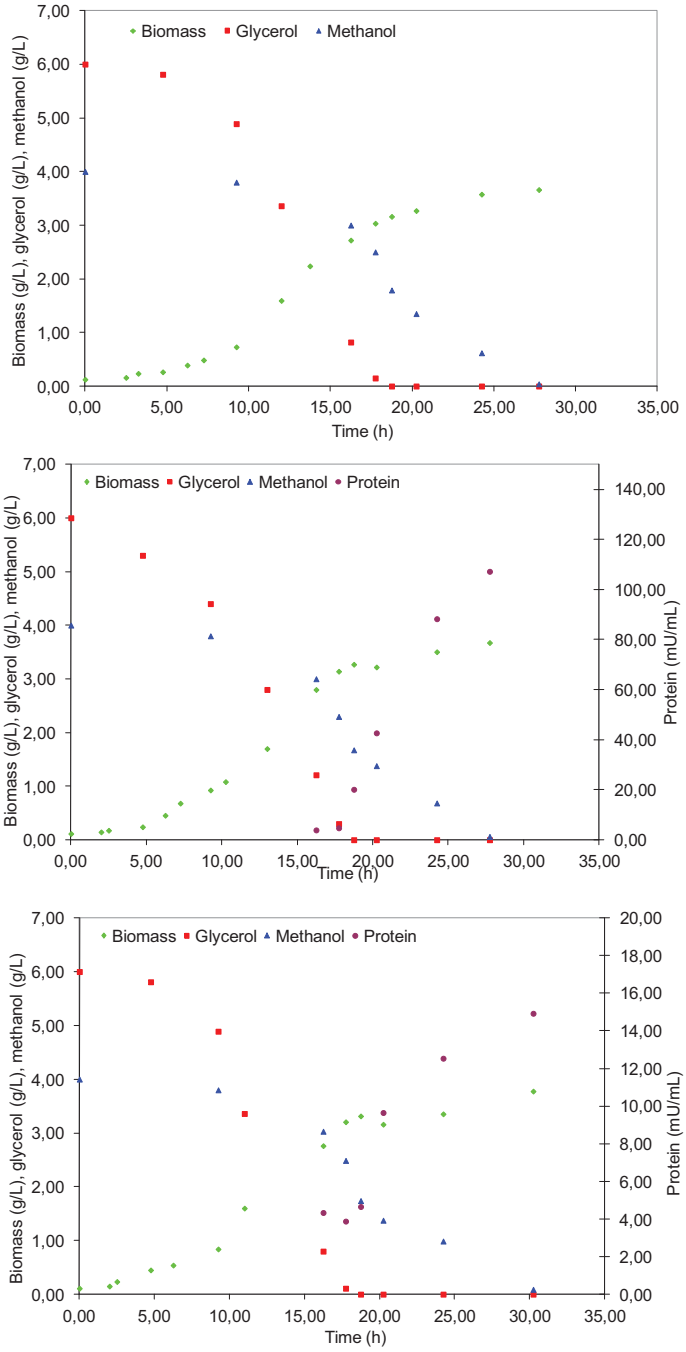


Figure 59. Batch cultures of X33-pPICzB, X33-Abf and X33-Bgl (conditions 3.1, 3.2, 3.3).

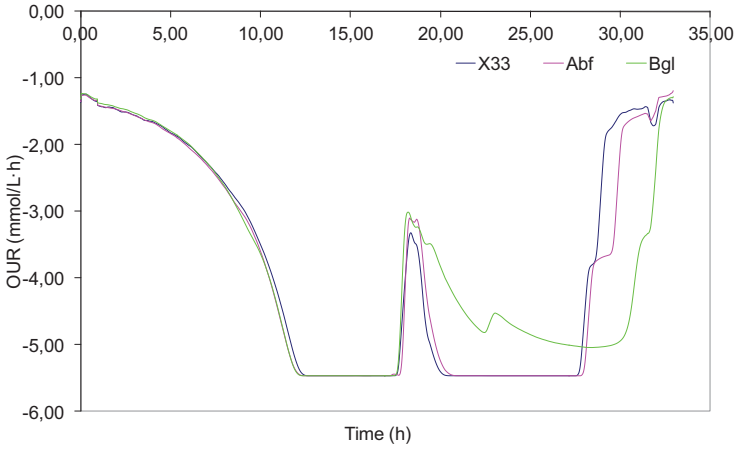


Figure 60. OUR evolution in X33-pPICzB, X33-Abf and X33-Bgl cultures (conditions 3.1, 3.2, 3.3).

Early methanol consumption can also be observed when 6 g/L glycerol are used (Figure 59). Regular sampling in the late exponential growth phase on glycerol clearly shows that methanol is consumed before glycerol is fully depleted, in accordance to previous reports by Inan and Meaguer (2001). Working points for model validation have been selected in regions where only glycerol or methanol is being consumed.

However, recombinant protein production can only be detected during strict methanol growth phase. In accordance to previous results obtained in flask cultures and the three tested scenarios, Abf expression levels are much higher than in the case of Bgl, which is only detected in trace amounts in the cultures.

For each condition, working time-points corresponding to exponential growth phases on glycerol and methanol were selected for local flux calculation as referred in Section 3.5.2. The results are shown in Table 45.

Table 45. Extracellular flux values for each experimental scenario.

Code	Strain	μ Cmmol/g·h	Gly mmol/g·h	Met mmol/g·h	OUR mmol/g·h	CO2* mmol/g·h
1	X33-pPICzB	8,07	3,68	0,00	3,99	2,96
2	X33-Abf	8,14	3,96	0,00	4,48	3,73
3	X33-Bgl	8,44	5,84	0,00	4,37	9,09
4	X33-pPICzB	0,90	0,00	1,89	1,26	0,99
5	X33-Abf	0,80	0,00	1,54	1,33	0,74
6	X33-Bgl	0,79	0,00	1,45	1,12	0,66
7	X33-pPICzB	8,79	4,50	0,00	4,71	4,70
8	X33-Abf	9,92	4,98	0,00	5,39	5,01
9	X33-Bgl	8,98	4,08	0,00	4,96	3,25
10	X33-pPICzB	1,40	0,00	2,55	2,16	1,15
11	X33-Abf	1,00	0,00	2,55	2,29	1,55
12	X33-Bgl	0,87	0,00	2,13	2,20	1,27
13	X33-pPICzB	8,66	4,06	0,00	4,19	3,53
14	X33-Abf	8,77	4,61	0,00	3,38	5,05
15	X33-Bgl	7,92	3,79	0,00	3,86	3,46
16	X33-pPICzB	0,94	0,00	1,87	1,67	0,93
17	X33-Abf	0,59	0,00	1,78	1,70	1,19
18	X33-Bgl	0,48	0,00	1,40	1,26	0,92

*CO_2 flux values have been estimated on the basis of carbon mass balances.

As expected, specific growth rates are much higher in glycerol growth phase than in methanol growth phase. For both substrates, specific growth rates correlate to specific substrate consumption rates. Slightly higher growth rates are achieved when higher concentrations of methanol are used (scenarios 1-12). Higher methanol specific consumption rates are also achieved in scenarios 7-12, in which a lower amount of glycerol has been used in the batch phase, since cell concentration at early exponential methanol phase is lower. Methanol growth phase shows also lower ratios of oxygen consumption with regard to substrate consumption, indicating a stronger effect of oxygen limitation in this phase.

Both recombinant strains show lower specific growth rates that the wild type strain, possibly as a consequence of the metabolic burden caused by the expression of foreign genes. Table 45 also shows how X33-Bgl has lower consumption methanol and oxygen rates than X33-Abf and the wild type strain in the three tested conditions. It can be hypothesised that recombinant protein expression alters methanol consumption profile and hence, diminishes oxygen consumption rate.

Using the obtained extracellular fluxes, PS-MFA-based consistency analysis was applied to each working point. The results of biomass and protein productivity prediction are shown in Table 46.

Table 46. PS-MFA-based consistency analysis and biomass and protein productivity prediction

Code	η	Π	Bio_{est_inf} Cmmol/(g·h)	Bio_{est_sup} Cmmol/(g·h)	$Prot_{est\_inf}$ mg/(g·h)	$Prot_{est\_sup}$ mg/(g·h)	Protein U/(g·h)
1	2,19	0,861	7,791	9,015	0,000	0,000	0,000
2	2,06	1,000	7,468	7,844	0,000	0,004	0,000
3	1,44	0,010	8,529	8,556	0,000	0,000	0,000
4	0,48	0,114	0,943	0,946	0,000	0,002	0,000
5	0,52	1,000	0,775	0,841	0,002	0,005	1,729
6	0,54	0,881	0,828	0,830	0,000	0,002	0,087
7	1,96	0,521	9,190	9,130	0,000	0,000	0,000
8	1,99	0,752	10,410	10,440	0,000	0,001	0,000
9	2,20	1,000	8,525	9,435	0,000	0,011	0,000
10	0,55	1,000	1,330	1,470	0,001	0,007	0,000
11	0,39	0,274	1,048	1,051	0,010	0,013	2,991
12	0,41	1,000	0,826	0,914	0,009	0,016	0,203
13	2,13	0,425	8,220	8,261	0,000	0,000	0,000
14	1,90	0,002	8,319	8,342	0,000	0,000	0,000
15	2,09	0,359	7,537	7,575	0,000	0,000	0,000
16	0,50	1,000	0,892	0,988	0,003	0,008	0,000
17	0,33	0,522	0,618	0,620	0,009	0,012	3,019
18	0,35	0,332	0,398	0,399	0,006	0,008	0,189

As can be seen in Table 46, the agreement between the calculated experimental fluxes and the previously validated model is high in most of the cases. Only two scenarios (3 and 14) show possibility values lower than 5%. These cases could in fact correspond to potentially inconsistent experimental external flux values, since carbon balance provides anomalously high CO_2 flux values. Only a few scenarios result in full agreement (possibility value equal to 1), thus demonstrating that PS-MFA enables analysis and prediction without circumventing data intrinsic variability. This is, flux reconciliation is not necessary to determine the degree of consistency of experimental values and to produce estimates of biomass growth and protein productivity.

Biomass estimation provides sensible boundaries for each experimental dataset. For most scenarios, biomass prediction is not enclosed by 90% boundaries, which is not surprising and possibly reflects data inconsistency (Table 46). However, 50% boundaries are sufficient to enclose most of biomass estimates (data not shown).

Biomass to substrate yield can be compared to theoretical maximal yields, as was done for literature datasets in Chapter V. None of the analysed scenarios violates theoretical maxima (0.82 in the case of methanol, 2.42 in the case of glycerol). Together with lower growth rates, lower yields are observed in most of the cases when recombinant strains are grown on methanol. Once more this is possibly a consequence of metabolic burden and the basis of the predictive model for protein productivity.

Protein productivity estimates are in general low in comparison to the scenarios analysed previously (Table 46). This is not unexpected since higher productivities can only be achieved when *P. pastoris* cultures are grown on dual-carbon limited conditions, using glycerol or sorbitol as supplementary substrates, or in fed-batch, methanol-limited conditions, with high volumetric supply rates are sustained together with low methanol concentration in the broth, to avoid inhibition. Nevertheless, higher values for protein productivity are obtained in scenarios where recombinant strains are grown on methanol, in comparison to the use of glycerol and the wild-type strain.

Basal levels of protein production are erroneously calculated in two cases (10, 16), in which the wild type strain is used and thus, recombinant protein production is null. This is not surprising since it was also shown, when literature scenarios were analysed, that

the model is prone to protein productivity overestimation, possibly due to the lack of protein unproductive scenarios used for parameter fitting. In the case of X33-Bgl, protein productivity seems to be overestimated. Taking into account the anomalous growth of this strain on methanol, that is clearly reflected on an abnormal pattern of oxygen consumption, this may again suggest a potential inhibitory effect of protein expression: the resources for enzyme synthesis are in fact employed, but no activity is detected due to secretional or folding bottlenecks.

However, as can be seen in Table 46, the scenarios in which protein productivity is higher are identified and in general, the datasets where no enzymatic activity is detected correspond to prediction for inexistent or low protein productivity values. This validates the use of this model as a qualitative approach to predict potential suitable conditions for protein expression and for process monitoring.

6.4 Discussion

This contribution introduces a simple but structured model of the yeast *P. pastoris* as a means to predict protein productivity from only a few extracellular measurements. The strategy is based on the previously established (see Chapter V) constraint-based model of *P. pastoris* metabolism. The stoichiometric network of *P. pastoris* main biochemical pathways is used to relate available measurements (mainly substrate consumption, gases exchange and biomass growth specific rates) by applying possibilistic metabolic flux analysis. This enables the estimation of unknown intracellular rates and particularly, the balance of key energetic equivalents such as ATP.

It has been shown that total ATP generation can be related to biomass growth and protein productivity specific rates by a simple linear regression. Parameter fitting has been accomplished using experimental chemostat data, by 8-fold cross-validation on 34 literature datasets corresponding to continuous cultures of *P. pastoris*. As expected, large amounts of ATP are required for protein formation. This represents the negative impact of the overproduction of target metabolites in microorganisms, known as metabolic burden. For significant product formation, increased amounts of carbon

precursors, reducing equivalents and energy have to be supplied. For equivalent substrate consumption, this drain is generally manifested as a decreased growth rate and has been captured through different modelling approaches, including stoichiometric networks (Glick, 1995), and has also been reported for *P. pastoris* (Baumann *et al.*, 2010; Heyland *et al.*, 2011).

Once the parameters have been estimated, and since the expression for protein productivity is linear, it can be integrated into the constraint-based model. In this way, the model couples a constraint-based model with a first-principles expression for the estimation of protein productivity in *P. pastoris* cultures, based on the distribution of energetic resources between growth and heterologous protein production. This working scheme considers independently stoichiometric coefficients coming from well-known biochemical pathways, whereas unknown ATP requirements for a specific anabolic pathway are estimated on the basis of experimental data and these constraints. This approach is particularly useful when cofactor equivalents cannot be precisely calculated only on the basis of enzymatic reactions, as complex cellular processes, such as folding, polymerisation or transport occur that make up most of the energetic cost.

Several constraint-based models have been described that include recombinant protein production, but examples of their use as predictive tools for protein productivity are scarce. Quantitative prediction based on the use of lumped biochemical equations has been shown to be inaccurate, possibly due to its large sensitivity to the measured fluxes (Carinhas *et al.*, 2011). More detailed models that consider induction, transcription and translation phenomena can be used, but these become necessarily protein-specific (Chung *et al.*, 2010; Çelik *et al.*, 2010; Sohn *et al.,* 2010).

Other hybrid approaches have been used to overcome the limitations of constraint-based models to produce valid estimates for protein productivity, for example combining them with data-driven expressions for protein productivity, such as kinetic, growth-rate associated, functions described above (Ren *et al.,* 2003; Mendoza-Muñoz *et al.*, 2008) or more sophisticated relationships built through projection to latent structures (Carinhas *et al.*, 2011). Note that in these cases, the structure and not only the value of parameters depends on the particular experimental system analysed. In contrast, this contribution introduces a physically based model, based on a first-

principles relationship (41) that distributes the available energetic resources between growth and protein production. The parameters in this equation are then fitted experimentally and integrated in the constraint-based model. Consequently, the results can be interpreted directly in relation with the energy resources, in the form of ATP, devoted to biomass growth and protein formation rate among others.

In addition, the relationship between ATP generation, growth and protein production is impervious enough to allow fitting one general model valid for a variety of *P. pastoris* cultures. The fitted model provides reasonably good predictions for different *P. pastoris* strains, expressing different heterologous proteins, and in wide operating conditions (*e.g.* different substrates and growth rates).

On the other hand, although protein productivity estimation has shown to be valid for different recombinant strains, it is also possible to particularise parameter fitting for a given protein if a sufficient number of experimental values are available, although this is not strictly required. The proposed approach could be also applied to other widely used expression systems, such as *Escherichia coli* or *Saccharomyces cerevisiae*, as well as other complex products of microbial metabolism, such as biopolymers, by modifying the underlying metabolic network.

At this point many standard methods can be used to exploit this model and analyse emergent properties, simulate genetic modifications, predict non-measured fluxes, *etc*. In particular, herein the model is used to predict biomass growth and protein productivity by applying an MFA-wise method in scenarios where only a few extracellular fluxes are known, such as substrate consumption, oxygen uptake and carbon dioxide production rates.

Finally, the advantage of this approach is than it can easily be implemented in industrial environments in which few on-line (at-line) measurements are available. Furthermore, combined estimation of biomass growth and protein productivity is possible as long as carbon sources and gases are known. This situation is amenable in industrial fermentations, since carbon limited cultures in fed-batch mode are a standard for *Pichia* cultures, and oxygen and CO_2 measurement can be easily achieved by means of exhaust gas analysis. This would be the case in large-scale, pilot bioreactors in which a

single protein is being continuously produced. On the contrary, bench-top systems at laboratory scale could benefit of rich protein productivity estimation if biomass growth rate can be estimated (*e.g.* by periodic sampling or absorbance-based biomass sensors) even if OUR and CPR cannot be calculated.

This makes it possible to represent a complex process such as recombinant protein production, in a workable format for industrial environments, with scarce and uncertain measurements and of reasonable complexity for practical compatibility with monitoring and control strategies, while generality is ensured by the first-principles based approach. Such “grey model”, developed in the framework of systems biology, enables a consistent link between cellular and bioreactor levels, with the ultimate goal of identifying the metabolic states corresponding to high productivity and developing monitoring and control methods to lead the system to the desired state.

VII. Glycosidase immobilisation on bimodal organosilicas

In this chapter, the enzymes Abf and Bgl that were overproduced and purified from recombinant yeast strains are immobilised onto conveniently functionalised bimodal organosilicas of the UVM-7 family by chemical adsorption and covalent binding. The biochemical and physicochemical properties of such biocatalysts are assayed along with their use in different applications of biotechnological interest. This strategy promotes a shift in the optimum enzymatic working conditions, sensitivity to inhibitors and specificity of the enzymes together with an improved stability. The capacity of the immobilised enzymes to release isoflavones from their bound aglycons and to catalyse transglycosylation reactions in high yield has been also verified.

7.1 Background

Immobilisation of glycosidases has been widely reported in the literature. Polymeric materials, such as polyhydroxilated natural matrices have been mainly used for this purpose, but also inorganic porous supports (Matthijs and Schacht, 1996; Karagulyan *et al.*, 2008; Singh *et al.*, 2011) and functionalised polymers (Hernaiz and Crout, 2000). Abf, together with other glycosidases such as Bgl, rhamnosidase and xylanases have been immobilised onto organic supports such as acrylic beads, agarose, chitin, chitosan and gelatine, and also on functionalised supports such as amino-agarose (Spagna *et al.*, 2000b; Martino *et al.,* 2003a, 2003b). Bgl has also been adsorbed onto duolite resins and xylanase has been immobilised on the synthetic polymer Eudragit based on methyl-metacrylate (Sardar *et al.,* 2000; Gawande, 2003).

Once immobilised, glycosidases have been applied to hydrolyse different types of natural compounds, such as glycosylated terpenes in wine (Gallifuoco *et al.*, 1997,

2003), isoflavones (Zhang *et al.*, 2008) and lactose (Pessela *et al.*, 2003). They have also been used in synthesis reactions in order to produce alkyl-glycosides (Kosary *et al.*, 1998; Yi *et al.*, 1998; Park *et al.*, 2000; Basso *et al.*, 2002; Gargouri *et al.*, 2004; Mladesnoska *et al.*, 2008) and oligosaccharides (Ravet *et al.*, 1993; Falkoski *et al.*, 2009; Mayer *et al.*, 2010).

The immobilisation of these enzymes has been performed mainly by adsorption, as a new recovery and purification method (Riccio *et al.*, 1999, Spagna *et al.,* 1998, 2000b), but also by gel entrapment and covalent binding with glutaraldehyde. This latter resulted in a diminishment of the catalytic activity of the enzyme (Spagna *et al.*, 2000a).

Whereas organic supports are economic, non-toxic and easy to find with different types of functionalisation, they show low chemical, mechanical and physical resistance and are prone to microbial contamination. Conversely, mesoporous silicas were proposed as promising host materials because of their large pore sizes and high chemical and mechanical resistance. Enzyme immobilisation onto mesoporous sieves has been widely described in the last decade (Yan *et al.*, 2002; Hudson *et al.*, 2005). Some of the most relevant examples are summarised in Table 47. In these supports, immobilisation is usually carried out by adsorption based on electrostatic affinity between the amino groups of the enzyme and the hydroxyl groups of the support. The support can be functionalised with amino groups or activated with glutaraldehyde to link them to carboxyl groups of the enzyme (Subramanian *et al.*, 1999; Casadonte *et al.*, 2010).

Unimodal mesopores structures such as pure and functionalised MCM-41, MCM-48 (MCM, Mobil Crystalline Material) and SBA-15 materials have been successfully used for enzyme immobilisation (Gimon-Kinsel *et al.*, 1998; He *et al.*, 2000; Yiu *et al.*, 2000; Takashaki *et al.*, 2001; Gomez *et al.*, 2010). The adsorption of proteins onto MCM-41 depends on the pore size distribution of the support (Balkus and Diaz, 1996). The dependence of the immobilisation efficiency also depends on the pore diameter in relation to the enzyme size (Balkus and Diaz*,* 1996, Deere *et al.*, 2001).

The immobilisation of enzymes has also been accomplished onto mesoporous sieves by covalent binding. α-amylase and trypsin have been covalently linked to MCM-41, MCM-48 and SBA-15 and also to aminated silicas activated with glutaraldehyde

(Humphrey *et al.*, 2001). The specific activity varies depending on the support but in general a reduction was observed after the chemical reaction (Balkus and Diaz*,* 1996). In general, covalent binding improves the stability of the enzyme, but can reduce its specific activity in comparison to the free or adsorbed enzyme (Jung *et al.*, 2010). The charge and proton distribution of the support will also influence the properties of the catalyst.

UVM-7 type materials, bimodal organosilicas produced by the research group of Dr. Amorós and Dr. Beltrán at the Institut de Ciència dels Materials from the Universitat de València (Valencia, Spain) seem promising supports for enzyme immobilisation. These materials are a particular class of mesoporous sieves with macro and mesopores. The larger size of the pore system is suitable for Abf and Bgl encapsulation, which is prevented in conventional surfactant assisted mesoporous silicas such as M41s or SBA15. The smaller size of the mesopore system enables the immobilisation of smaller enzymes and substrate accessibility. Their good physico-chemical and thermal resistance further supports their use in biocatalytic applications.

Table 47. Enzymes immobilised onto mesoporous hosts.

Enzymes	Origin	Molecular mass (kDa)	Support	Immobilisation method	Reference
α-amilase	Barley	48	FSM-16, MCM-41,SBA-15	Adsorption	Pandya *et al.*, 2005
β-galactosidase	*Aspergillus oryzae*	105	ITQ-6	Adsorption and covalent binding	Corma *et al.*, 2001
Cytochrome-c	Bovine heart	12	MCM-41	Adsorption	Balkus and Diaz, 1996
Glucose-oxidase	*Aspergillus niger*	65	zeolite	Adsorption	Lei *et al.*, 2002
Glucose-oxidase	*A. niger*	65	Mesoporous sieve	Entrapment	Wei *et al*, 2002
Papain	Papaya	39	MCM-41	Adsorption	Balkus and Diaz, 1996
Penicil acylase	*Escherichia coli*	86	ITQ-6	Adsorption and covalent binding	Corma *et al.*, 2001
Penicil-acylase	*E. coli*	86	MCM-41	Adsorption and covalent binding	He *et al.*, 2000
Peroxidase	Horseradish	44	MCM-41	Adsorption	Balkus and Diaz, 1996
Peroxidase	Horseradish	44	FSM-16, MCM-41,SBA-15	Adsorption	Takashaki *et al.*, 2001
Peroxidase	Horseradish	44	Mesoporous sieve	Entrapment	Wei *et al*, 2002
Subtilisin	*Bacillus subtilis*	38	FSM-16, MCM-41,SBA-15	Adsorption	Takashaki *et al.*, 2001
Trypsin	Bovine pancreas	46	Nanoporous silicate	Covalent binding	Wang *et al.*, 2001
Trypsin	Bovine pancreas	46	MCM-41	Adsorption	Balkus and Diaz, 1996
Trypsin	Bovine pancreas	46	MCM-41,48, SBA-15	Adsorption and covalent binding	Yiu *et al.*, 2000

7.2 Objectives

The main objective of the chapter is to show that Abf and Bgl can be immobilised onto UVM-7 type mesoporous materials and to characterise the properties of the biocatalysts obtained. The following particular objectives will be addressed:

(i) The suitability of UVM-7 materials for enzyme immobilisation will be firstly assessed using lysozyme and Abf as model enzymes, in order to evaluate the requirements, kinetics, maximum loading and characteristics of the resulting biocatalysts.

(ii) Several types of modified bimodal mesoporous hosts functionalised with different aminated radicals will be used to adsorb and immobilise by covalent binding Abf and Bgl.

(iii) The biochemical and physicochemical properties of such biocatalysts will be compared for Abf and Bgl immobilisation. The structure of the materials after enzyme immobilisation will be studied by N2 adsorption-desorption and Transmission Electronic Microscopy (TME). The influence of pH, temperature, glucose and ethanol on enzymatic activity will be evaluated and the kinetic parameters and substrate specificity of the biocatalysts will be determined.

(iv) The immobilised enzymes will be tested and compared to the free enzymes in two applications of biotechnological interest, the release of soy isoflavones from their glycosylated precursors and the obtention of alkylated sugars by transglycosylation reactions.

7.3 Results

7.3.1 Support and immobilisation method selection

In order to select the most suitable support and immobilisation method the following considerations were taken into account. Abf and Bgl are active at weakly acid pH and show isoelectric points of 3.5 and 4.0 respectively (Flipphi *et al.*, 1993; Janbon *et al.*, 1994). For this reason, positively charged supports were required and aminated materials (NUVM-7) were synthesised as described in Section 3.8.1. Covalent immobilisation can be induced with activation with glutaraldehyde and carbodiimides (Arroyo, 1998). In this case, carbodiimide-directed peptide group formation was preferred as previous studies had concluded that EDC induced covalent binding improved activity yield in a β-glucosidase isolated from a commercial product (Spagna *et al.*, 2000b). This procedure requires carboxylic acid residues from the enzyme to react with amino groups from the support. Amino acid sequences from Abf and Bgl were examined and found to contain 34 aspartic acid and 22 glutamic acid residues, and 49 aspartic acid and 39 glutamic acid, respectively. The significant number of acidic residues (11.2% for Abf, 11.5% for Bgl) further supports the selection of this immobilisation strategy.

7.3.2 Set-up of immobilisation procedure – adsorption assays

Preliminar immobilisation assays were performed in order to confirm this choice and to design the immobilisation protocol. For this, two glycosyl-hydrolases, lysozyme from chicken egg white (used as described in Section 3.3.1) and recombinant Abf were used as model enzymes. Lysozyme is a small globular enzyme (14.4 kDa) with a highly cationic isoelectric point (around 11), whereas Abf is a larger enzyme (67 kDa) with slightly acidic isoelectric point (3.8).

Lysozyme and recombinant Abf were immobilised by adsorption on different silica-based supports. The amount of retained enzyme was calculated in each case by measuring the remaining protein concentration or enzymatic activity in supernatant and

on the solid, pelleted and washed support as described in Section 3.8.2. The N_2 adsorption-desorption isotherms of such biocatalysts were obtained and the remaining surface area, and the size of macro and mesopores were calculated in each case as shown in Table 48.

Table 48. Physicochemical properties and distribution of the enzymes adsorbed on several silica-based porous materials.

Material	Enzyme load mg/g	BET surface area m^2/g	BJH small pore nm	BJH large pore nm	BJH large pore nm	Small pore volume cm^3/g	Large pore volume cm^3/g
UVM-7 (C16)	-	1160	3.2	36.9	27.6	1.12	1.78
Lys-UVM-7 (C16)	250		-	33.7	25.5	-	1.14
UVM-7 (C18)	-	1110	3.4	29.8	20.5	1.11	1.62
Lys-UVM-7 (C18)	350	230	-	25.3	19.5	-	1.00
Xerogel-10	-	190	-	47.5	30.8	-	1.72
Lys-Xerogel-10	125	145	-	42.0	29.5	-	1.19
Xerogel-25	-	350	-	15.1	12.8	-	1.33
Lys-Xerogel-25	125	270	-	13.6	10.8	-	0.85
NUVM7-1	-	1075	2.8	32.4	24.1	0.91	1.16
Abf-NUVM7-1	200	610	2.6	40.3	27.5	0.47	0.45

Figures 61 and 62 show the N_2 adsorption-desorption isotherms of UVM-7 materials and xerogels (right panel-10 nm higher panel, 25 nm lower panel) with adsorbed lysozyme (Figure 61) and NUVM-7 materials with adsorbed Abf (Figure 62). The N_2 isotherms confirm the porous nature of the inorganic supports as one (in the case of xerogel, right panel, Figure 61) or two (in the case of UVM-7 type materials) defined adsorption steps can be verified. After enzyme immobilisation (b), the lower pressure adsorption step is prevented in all cases as the pores are blocked. A slight hysteresis loop at high partial pressures can be observed in all samples and is characteristic of interconnected cage-like pores.

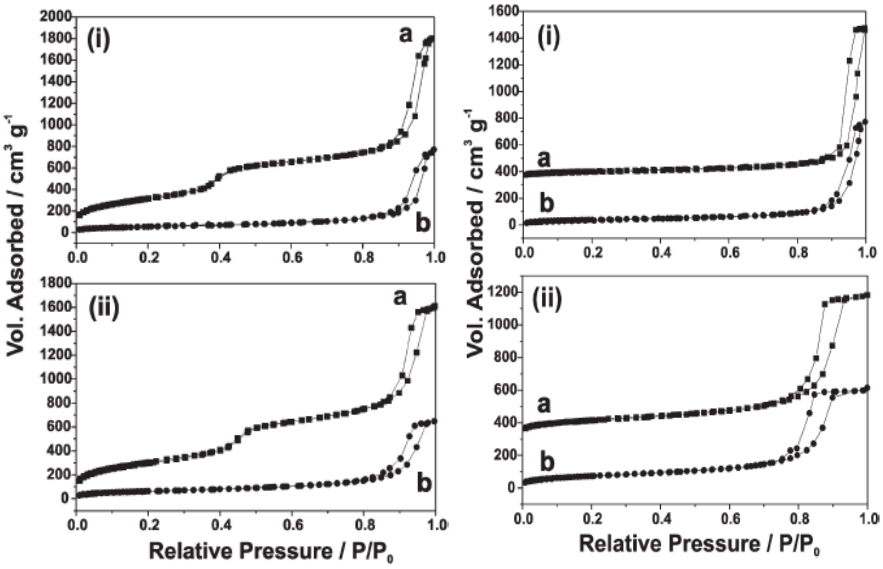


Figure 61. N_2 adsorption-desorption isotherms of UVM-7 materials (left panel, C16 higher panel, C18 lower panel) and xerogels (right panel-10 nm higher panel, 25 nm lower panel) with adsorbed lysozyme. (a) and (b) isotherms correspond to the materials before and after enzyme immobilisation, in each case.

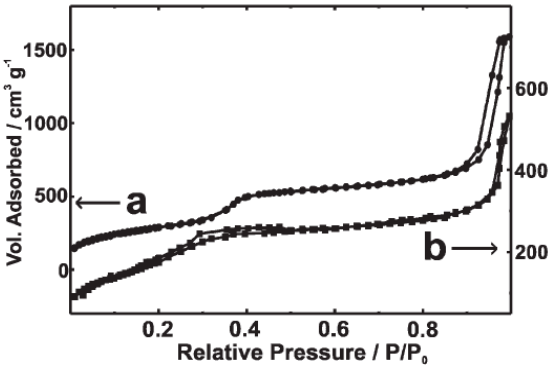


Figure 62. N_2 adsorption-desorption isotherms of NUVM-7 and NUVM7-1-Abf materials. (a) and (b) isotherms correspond to the materials before and after Abf immobilisation.

Two kinds of UVM-7 materials (soldered mesoporous nanoparticles, ca. 20-30 nm, describing a hierarchic large pore system) were assayed for lysozyme immobilisation, with slightly different cell parameters (C16 and C18), and compared to conventional xerogels (unimodal nanoparticulated xerogels constructed from small particles of 25 and 10 nm). In UVM-7 supports, the enzyme was precluded both in small and large pores as evidenced by the reduction in volume and pore size and more than 80% of the surface area is covered by the enzyme.

In the case of aminated UVM-7 materials, Abf showed a similar behaviour. Around 50% of the surface area was covered, although in this case the enzyme was located in the larger pores. As can be seen in Table 48 the diameter of small pores remains unaltered whereas the reduction in the large pore volume is significant. Even if the intraparticle mesopore system is not effective for Abf immobilisation, its existence could eventually favour the mass transfer of substrate/product molecules to the Abf active sites. This would improve the biochemical activity of the enzyme justifying what these materials could be preferred against nanoparticulated xerogels of similar pore topology.

In the case of UVM-7 and NUVM7-1 materials, the kinetics of enzyme adsorption were determined for both, lysozyme and recombinant Abf. In both cases (UVM-7 and NUVM7-1 type materials) the adsorption was complete after 3 hours and more than 50% of the maximal loading could be achieved after only 20 minutes (Figure 63). As expected, recombinant Abf could not be adsorbed by UVM-7 materials at pH 5.0, and only 18% total activity was retained from supernatant at pH 3.0 (data not shown).

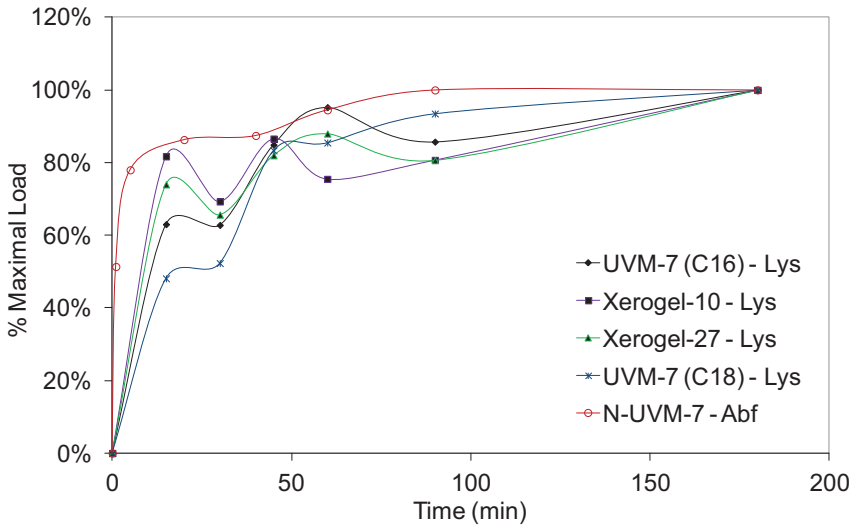


Figure 63. Immobilisation kinetics on UVM-7 and NUVM-7 materials of lysozyme and recombinant Abf.

These preliminary results support the use of UVM-7 type materials for quick and efficient enzyme loading. For acidic enzyme immobilisation such it is the case for recombinant Abf and Bgl, the presence of free amino groups on the surface is required but 5% functionalisation is sufficient to adsorb the enzyme at pH 5.0. N_2 adsorption-desorption isotherms reveal that small enzymes such as lysozyme are adsorbed both in small and mesopores, whereas recombinant Abf is preferably located inside the mesopores.

7.3.3 Biocatalyst obtention

Immobilisation tests were run with all the supports listed in Table 6 and Abf and Bgl dissolved in McIlvaine buffer at 5.0 as described in Section 3.3.4. The results are shown in Figure 64. The amount of enzymatic activity detracted from supernatant is represented for both enzymes and each support (immobilisation yield). The value of the remaining activity (activity yield) in the support is also shown.

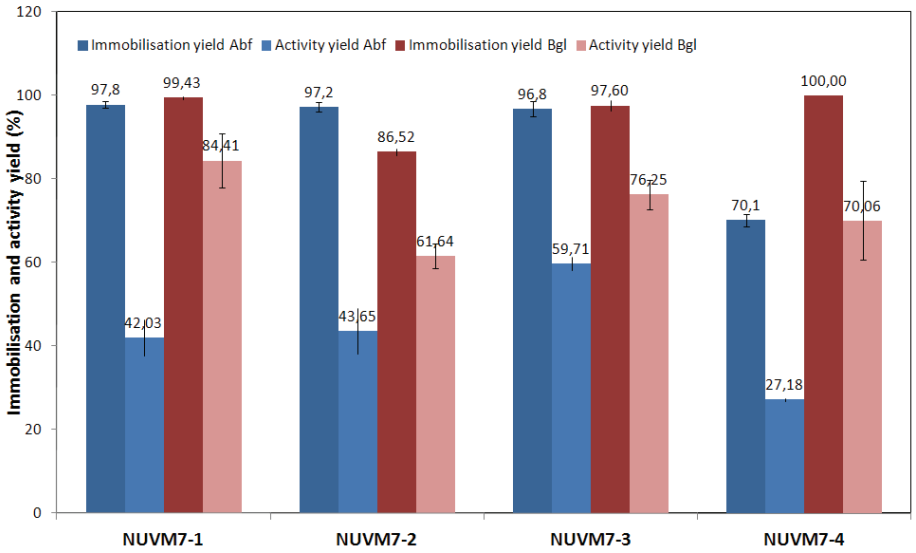


Figure 64. Charge and specific activity of the immobilised enzymes in each of the materials assayed.

As can be seen in Figure 64, whereas recombinant Bgl can be fully adsorbed on all NUVM-7 materials, only NUVM7-1, NUVM7-2 and NUVM7-3 could be efficiently used for recombinant Abf adsorption, whereas NUVM7-4 did not efficiently adsorb Abf activity (Figure 64). This could be caused by the different sizes of these enzymes and uneven distribution of the amino groups in the supports surface. Although all supports contain 5% of amino groups, these can be distributed more or less homogeneously as a consequence of the synthesis method. Attending to the ratio of functionalisation and the physical characteristics of these materials, supports produced by co-condensation show a uniform distribution of amino groups of around 0,5 radicals/nm^2. Conversely, amino groups in NUVM-7 supports produced by impregnation will be found mainly in the larger pores and the mesopores entrances, and the density of amino radicals in these materials will be higher on their external accessible surface.

As Bgl is a larger enzyme than Abf, a more scarce distribution of amino groups does not hinder immobilisation. However, Abf requires a larger degree of functionalisation of the surface of the external pores. NUVM7-2 is in this case an exception, as the amino

silane spacer is longer. The amino groups would be lodged preferably in the larger pores, even when the materials are produced by co-condensation, such as NUVM7-4, and both Abf and Bgl can be immobilised on this support. This means that, as expected, the enzymes are located, in all cases, in the external, larger pores of the supports.

With regard to the enzymatic activity retained by each support, it is worth noting that in the case of recombinant Abf, enzymatic activity is reduced in more than 50%. The best support seems to be NUVM7-3, for which around 60% of the enzymatic activity is retained. In the case of recombinant Bgl, all materials are capable to retain more than 70% of the enzymatic activity. Again, NUVM7-3 support (with a longer radical bearing 3 amino groups) showed higher activities (76%). This can be related to the larger chain of the aminated radical, as less rigid structures are known to favour enzymatic activity in covalent binding (Blanco *et al.,* 1989).

The materials NUVM7-1, NUVM7-2 and NUVM7-3 were selected for recombinant Abf immobilisation and the corresponding biocatalysts were produced in higher amounts for further analysis of their properties. Recombinant Bgl was immobilised on all NUVM-7 materials.

7.3.4 Physicochemical characterisation

The biocatalysts were further characterised by XRD, TEM and porosity curves. The results are similar in all cases and TEM and porosimetry curves obtained for the covalently immobilised enzymes on NUVM7-1 are presented in Figures 65 and 66 for illustration purposes. The NUVM-7 host presents the typical open nature of HPNO bimodal organosilicas (El Haskouri *et al.*, 2002a). According to XRD (data not shown) and TEM, the Abf and Bgl functionalised materials retain the HPNO characteristics. The existence of a hierarchical porous structure after Abf and Bgl immobilisation is also confirmed by porosimetry curves (Figure 66).

The nitrogen adsorption-desorption isotherms show that a significant accessibility to the two pore systems is preserved after enzyme immobilisation. However, the decrease in the large pore volume confirms that the covalently bound enzymes are located inside the larger pores, close to the silica surface. The small pore volume decrease could be

related to a partial blocking of the mesopore entrances by the enzyme, but also to the adsorption of small organic molecules of substrate or buffer. This intraparticle mesopore system could favour the mass transfer of substrate/product molecules to the Abf active sites improving the biochemical activity of the enzyme.

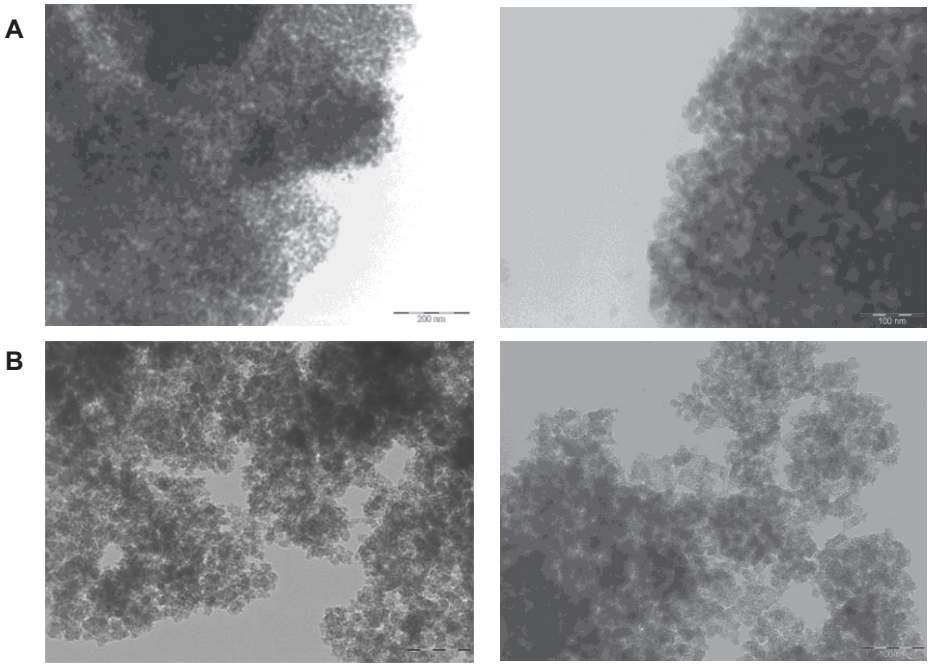


Figure 65. TEM images of (A) NUVM7-1-Abf and (B) NUVM7-1-Bgl biocatalysts.

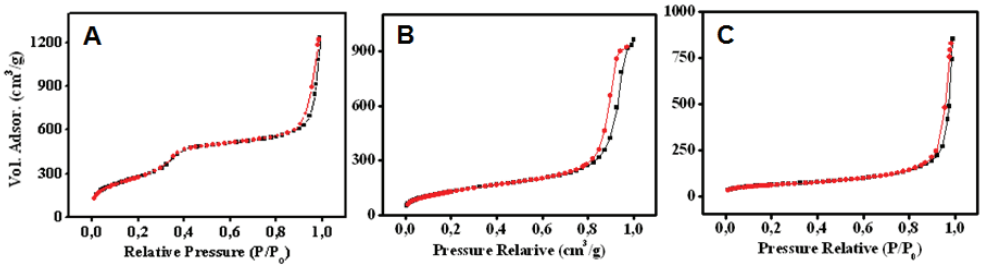


Figure 66. N_2 adsorption-desorption isotherms of (A) NUVM7, (B) NUVM7-1-Abf and (C) NUVM7-1-Bgl materials.

7.3.5 Biochemical characterisation of immobilised enzymes

The activity of the biocatalysts was assayed under different pH, temperatures and ethanol and glucose concentrations to assess if differences could be found between the free and immobilised enzymes. The results are shown in Figures 67 to 70 for recombinant Abf and Bgl enzymes.

Optimal environmental conditions

In the case of Abf activity, a decrease in the optimum pH range of the immobilised enzymes can be observed (Figure 67). The optimum pH is shifted towards more acid conditions, from ca. 4 for the free enzyme to values below 3. This is most likely due to the presence of basic amine groups at the HPNO surface that generate an increase of the local pH around the enzymes hindering the usual denaturing effect of acid conditions. It is worth mentioning that this effect does not occur in the case of the merely adsorbed enzyme (Tortajada *et al.*, 2005) and so a stabilising effect can also be attributed to covalent binding. This descent of optimum pH of glycosidases due to immobilisation has been previously reported for β-glucosidase and rhamnosidase linked to aminated chitosan (Spagna *et al.*, 2000a, b).

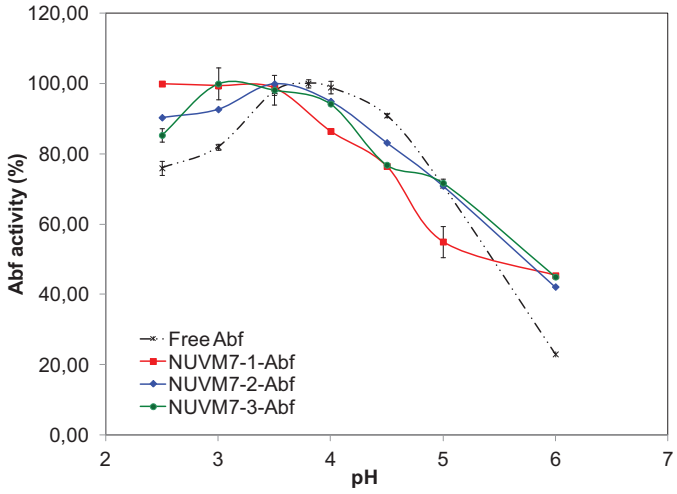


Figure 67. pH influence on Abf activity of the free and immobilised enzymes.

A general stabilising effect can also be observed in relation to temperature (Figure 68). This kind of effect has been widely described in covalently-bound enzymes, stabilised as thermal denaturation is hindered thanks to the chemical attachment to the support. In the particular case of mesoporous sieves, an effect of increased hydration due to the special water-structuring properties of the silanol groups at the silica surface has also been reported (Ravindra *et al.*, 2004).

The effect seems to be more acute in the case of NUVM7-1 support, which also favours the lowest optimum pH and is the one showing the shortest radical chain length, which may relate to more rigid configurations of the covalently attached enzyme (Figure 67).

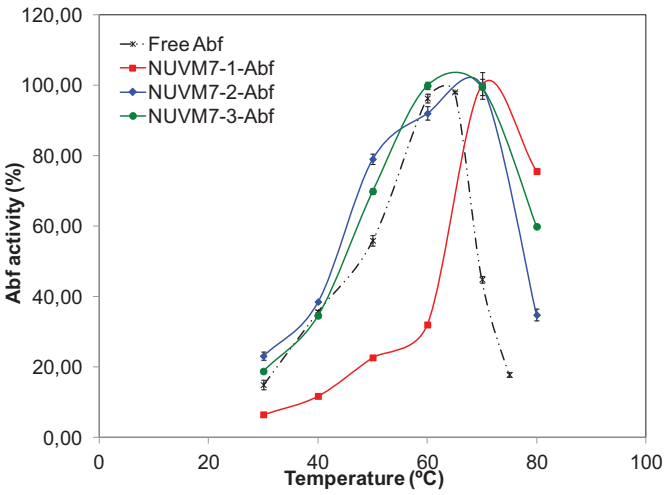


Figure 68. Temperature influence on Abf activity of the free enzyme and immobilised forms.

The effect of immobilisation on optimal pH of recombinant Bgl activity is shown in Figure 69. Interestingly, in this case, two homogenous groups are found. Supports NUVM7-1 and 4 favour lower pH optima (of around 3.5) than the free enzyme, whereas immobilisation on supports NUVM7-2 and 3 does not promote this shift. This behaviour correlates with the effect on immobilised Abf (for which NUVM7-1 produced the lowest optimum pH) and may be a consequence of the length of the radical, as the supports bearing shorter chains (NUVM7-1 and NUVM7-4) seem to share identical behaviour.

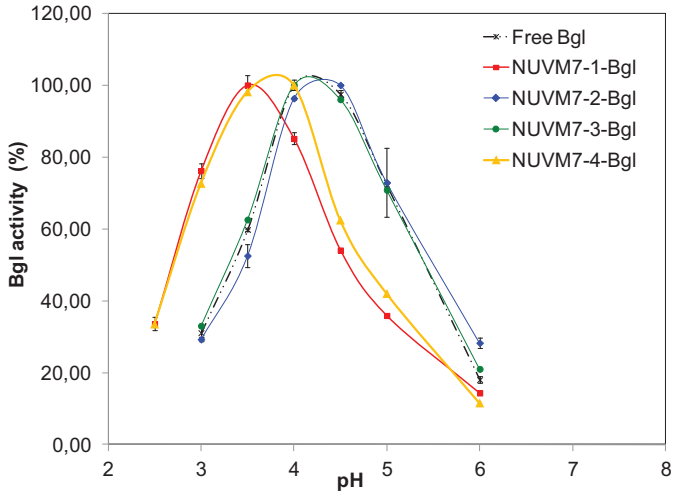


Figure 69. pH influence on Bgl activity of the free enzyme and immobilised forms.

In the case of Bgl, immobilised enzymes show similar optimum temperature than the free enzyme although the activity range appears to be wider, particularly in the case of NUVM7-3, for which at least 60% of the activity remains at 70ºC (Figure 70). Again, this is probably due to a more rigid spatial conformation of the covalently bound enzyme. The effect of the structure of the radicals and the amount of aminated residues on the stabilisation seems to be enzyme-dependent, possibly due to the particular size and tridimensional structure of the polypeptide.

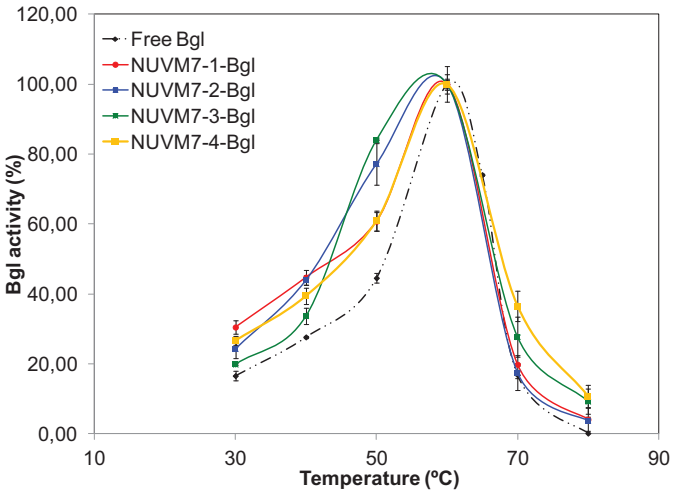


Figure 70. Temperature influence on Bgl activity of the free enzyme and immobilised forms.

Effect of inhibitors: ethanol and glucose

The effect of glucose and ethanol on Abf activity was also verified as shown in Figures 71 and 72. No significant differences can be observed in the case of glucose as Abf is not inhibited by the sugar (Sánchez-Torres *et al.*, 1996). A slight increase of the NUVM7-1-Abf biocatalyst activity can be observed.

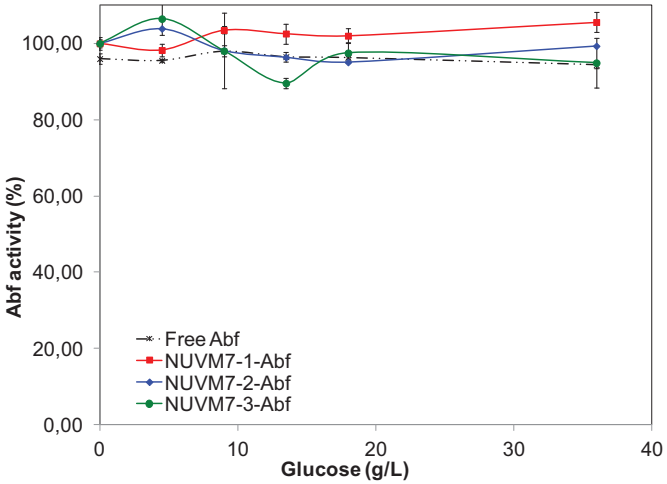


Figure 71. Glucose influence on Abf activity of the free enzyme and immobilised forms.

Conversely, free recombinant Abf is strongly inhibited by high ethanol concentrations and reduces its activity in 40% in the presence of 20 g/L ethanol. The immobilised enzyme seems to be more stable in all the studied ethanol range as the loss in activity is reduced to around 10-20% at 20 g/L ethanol (Figure 72). Again, this is possibly a consequence of multipoint attachment minimising enzymatic denaturation.

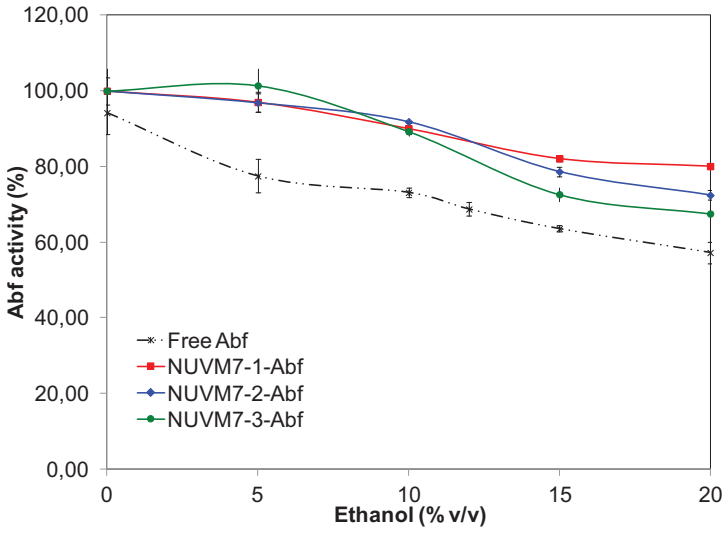


Figure 72. Ethanol influence on Abf activity of the free enzyme and immobilised forms.

Finally the effect of two potential inhibitors, glucose and ethanol, is represented in Figures 73 and 74 for free and immobilised Bgl. Free Bgl is characterised by a strong glucose inhibition (Sánchez-Torres *et al.*, 1998), whereas the enzyme is activated by increasing ethanol concentration. The situation is the same in the case of immobilised Bgl with respect to glucose and all NUVM7-Bgl biocatalysts show decreased activities (to 20-30% at only 20 g/L glucose) with growing glucose concentrations.

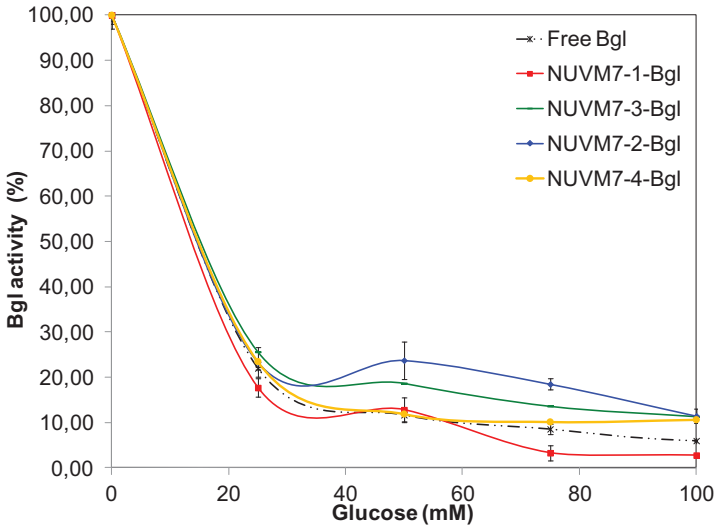


Figure 73. Glucose influence on Bgl activity of the free enzyme and immobilised forms.

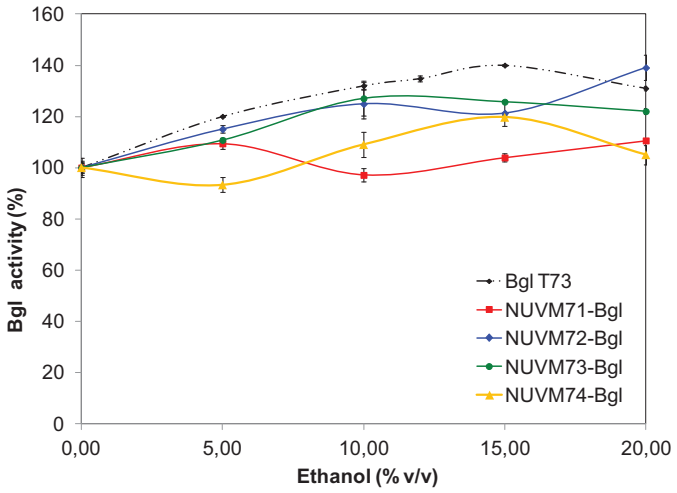


Figure 74. Ethanol influence on Bgl activity of the free enzyme and immobilised forms.

Thermal stability

Covalent immobilisation has been frequently used to stabilise enzymes against typically denaturing conditions such as high temperatures. In order to analyse if immobilised Abf and Bgl show improved thermal stability, samples of the free and immobilised enzymes were incubated at increasing temperatures as detailed in Section 3.3.4. The remaining enzymatic activity was then measured in a standard assay against pNP-activated substrates and compared with the initial activity. The results are shown in Figures 75 (for Abf) and 76 (for Bgl).

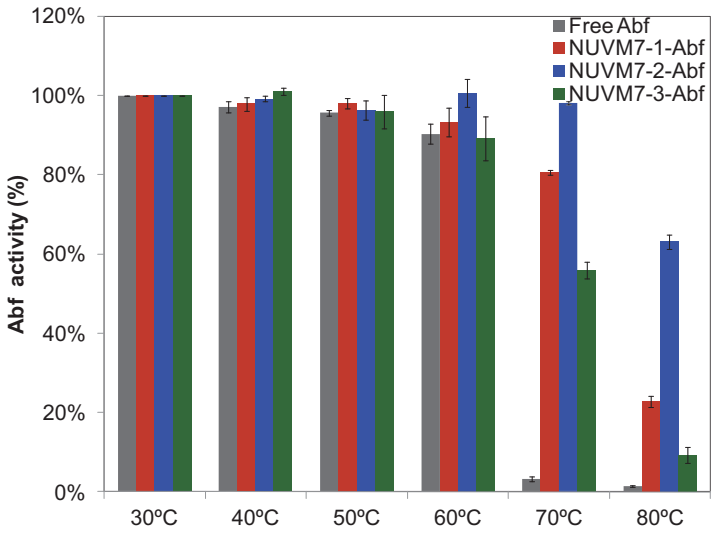


Figure 75. Thermal stability of Abf covalently immobilised on the different NUVM-7 supports.

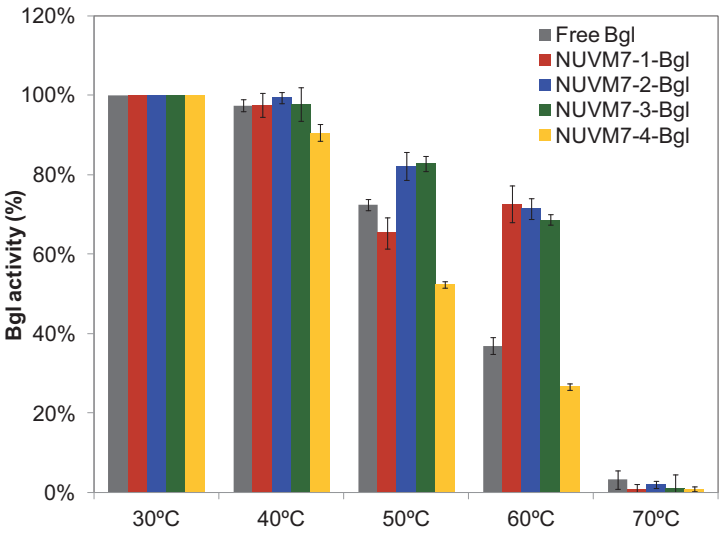


Figure 76. Thermal stability of Bgl covalently immobilised on the different NUVM-7 supports.

As can be seen in Figures 75 and 76, more enzymatic activity is preserved after incubation when Abf and Bgl are covalently immobilised. In both cases, covalent immobilisation seems to improve biocatalyst stability, possibly as an effect of prevented thermal denaturation due to multipoint attachment. In the case of Abf, the three assayed supports improve the stability of the enzyme, although NUVM7-2 has a greater effect. In the case of Bgl, the enzyme seems to be more thermally unstable than Abf, although NUVM7-1, NUVM7-2 and NUVM7-3 seem to prevent denaturation. Note that the supports able to stabilise the enzymes are also the ones for which higher optimal working temperatures (in the case of Abf) and higher activities at a given temperature (in the case of Bgl) are found (see Figures 68 and 70).

Substrate specificity

In order to assess if significant changes in substrate specificity of the immobilised enzymes could be found depending on the support used, the different biocatalysts were assayed in the presence of several hydrolisable compounds. The results are detailed in Table 49.

Table 49. Substrate specificity of Abf and Bgl immobilised in the different NUVM-7 materials.

Substrate	pNPA	Arabinan	pNPG	pNPX	Cellobiose	Salicin	Methyl-glucoside
Linkeage	α(1,6)	α(1,6)	β(1,4)	β(1,4)	β(1,4)	β(1,4)	β(1,4)
Free Abf	100	41.0±1.3	n.a.	2.9±0.1	n.a.	n.a.	n.a.
NUVM7-1-Abf	100	4.1±0.8	n.a.	1.0±0.3	n.a.	n.a.	n.a.
NUVM7-2-Abf	100	9.8±1.2	n.a.	4.1±0.2	n.a.	n.a.	n.a.
NUVM7-3-Abf	100	4.3±1.8	n.a.	1.2±0.1	n.a.	n.a.	n.a.
Free Bgl	n.d.	n.a.	100	7.1±0.5	12.9±3.0	48.4±1.4	19±1.2
NUVM7-1-Bgl	n.d.	n.a.	100	12.6±0.8	18.7±2.4	48.1±0.8	17±0.2
NUVM7-2-Bgl	n.d.	n.a.	100	1.9±1.6	n.d.	30.9±2.3	23±0.1
NUVM7-3-Bgl	n.d.	n.a.	100	1.3±0.5	n.d.	38.3±1.3	29±0.1
NUVM7-4-Bgl	n.d.	n.d.	100	16.5±0.5	n.d.	48.8±1.9	17±0.8

n.a. not analysed, n.d. not detected

In the case of Abf, the activity of the immobilised enzymes was assayed in comparison to the free enzyme. As can be seen in Table 49, covalently bound enzymes show diminished rates of arabinan hydrolysis, which may relate to a hindered accessibility of the enzyme to the substrate, a polymer which may not be diffusing properly towards the external pores where the enzyme is located. In addition, several peaks appear in HPLC chromatograms that could correspond to transglycosylation products generated from the available mono or disaccharides (data not shown).

With regard to recombinant Bgl, a wider range of substrates was assayed bearing different functional groups linked to glucose. As referred previously by Sánchez-Torres and coworkers (1998), Bgl shows a slight activity on pNP-linked xylose, and also on cellobiose, salicin and methyl-β-glucoside although the best processed substrate remains pNP-linked glucose.

The specificity of immobilised enzymes changes slightly and some of them show differences in relation to the free enzyme. In the case of NUVM7-2 and NUVM7-3 linked Bgl, the activity on pNPX and salicin is reduced, whereas the activity on methyl-β-glucoside is slightly increased. The opposite occurs in the case of NUVM7-1 and NUVM7-4. Since these two groups had been previously been identified in relation to their behaviour towards pH (see Figure 69), and hydrolysis proceeds through nucleophylic substitution, the differences in specificity may relate to changes in local pH/charge distribution in the vicinity of the enzyme, although it is unclear how to relate this effect to the structure of the aminated radical.

Determination of kinetic constants

Kinetic constants were determined for both supports produced by impregnation, NUVM7-1 and NUVM7-3. These supports provide maximum immobilisation and activity yields (Figure 64) and seem to induce different modifications in the behaviour of the immobilised enzymes.

Initial catalytic rates were determined at growing substrate concentrations (pNPA and pNPG, respectively) and adjusted to Michaelis-Menten kinetics using non-linear regression. The results are shown in Figures 77 to 78 and kinetic constants are summarised in Table 50 for comparison.

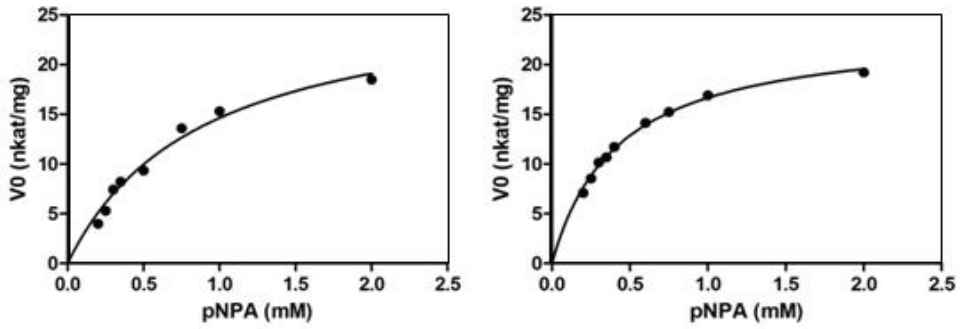


Figure 77. Kinetic constant determination for NUVM7-1-Abf (right panel) and NUVM7-3-Abf (left panel).

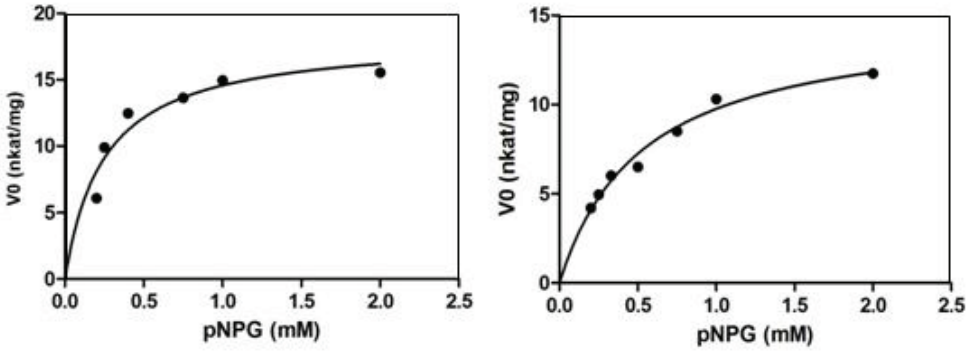


Figure 78. Kinetic constant determination for NUVM7-1-Bgl (right panel) and NUVM7-3-Bgl (left panel).

Table 50. Kinetic constants determined for free and immobilised Abf and Bgl.

Enzyme	Type	K_m mM	V_{max} nkat/mg
Abf	Free	0.45 ± 0.05	66.18 ± 3.6
Abf	NUVM7-1	0.88 ± 0.14	23.66 ± 0.50
Abf	NUVM7-3	0.42 ± 0.02	27.51 ± 2.28
Bgl	Free	0.26 ± 0.02	24.09 ± 0.49
Bgl	NUVM7-1	0.25 ± 0.07	18.23 ± 1.62
Bgl	NUVM7-3	0.53 ± 0.07	14.95 ± 0.07

An increase can be observed in K_m values in the case of NUVM7-1-Abf and NUVM7-3-Bgl which may be ascribable to conformational changes of the enzymes due to covalent binding, rather than diffusion difficulties found by the substrate as the porosity of the material is high and equal in both types of supports. In addition, these biocatalysts show the biggest differences with regard to biochemical properties in relation to the free enzyme (see Figures 77-78). Maximal catalytic rates also diminish both for covalently bound enzymes, as previously observed for Bgl immobilised on organic supports (Gallifuoco *et al.*, 1997, 2003).

7.3.6 Technological applications

Joint covalent immobilisation

As Abf and Bgl have been traditionally used in the release of terpenes bound to disaccharides of arabinose and glucose (Günata *et al.*, 1990), it was found of interest to assay if the joint immobilisation of the two enzymes was possible in NUVM7 type hosts. A similar approach has been previously attempted on chitosan derivatives (Spagna *et al.*, 2003). For this purpose, material NUVM7-1 was chosen, as wider differences between the free and the immobilised enzymes had been observed for this support.

The resulting biochemical properties of the biocatalysts were analysed in terms of pH, temperature, glucose and ethanol influence on Abf and Bgl relative activities. Both were found to behave similarly when immobilised together in comparison to the corresponding severally immobilised enzymes. The evolution of Abf activity with pH is shown below as an example (Figure 79).

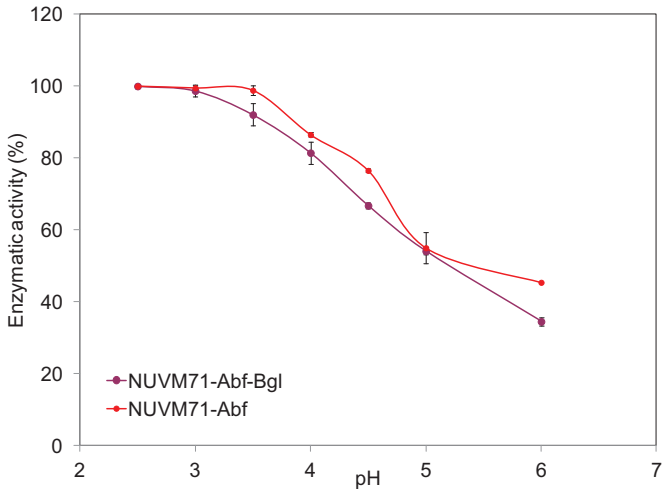


Figure 79. pH influence on Abf activity when Abf is jointly immobilised with Bgl.

The influence of pH in Abf activity is similar in NUVM7-Abf-Bgl and NUVM7-Abf biocatalysts, and in both cases, the optimum pH is shifted towards more acidic values. The situation is the same in the case of Bgl activity. The influence of temperature is shown in Figure 80 as an example.

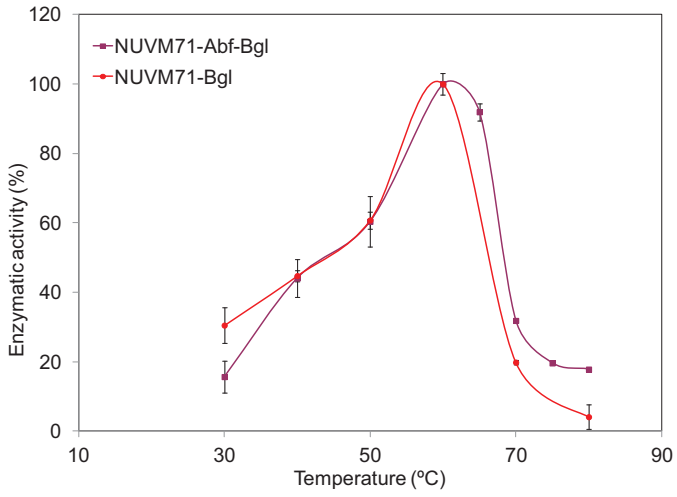


Figure 80. Temperature influence on Bgl activity when Bgl is jointly immobilised with Abf.

It is worth mentioning, however, that a reduction in Abf immobilisation yield was observed and only 23% of the activity was recovered on the support (data not shown). This may be due to a more efficient immobilisation of recombinant Bgl, which is competitive to Abf in relation to the available aminated radicals, because of higher enzyme size or number of acidic catalytic residues. It would be preferable then, even for a joint application, to severally immobilise both enzymes and combine the biocatalysts afterwards to reduce losses due to inefficient immobilisation.

Release of isoflavones

Isoflavones are found in nature in their glycosylated forms. Some examples of isoflavones that can be found in soy milk are shown in Figure 81. β-glycosidase is able to release the aglycon by hydrolysis.

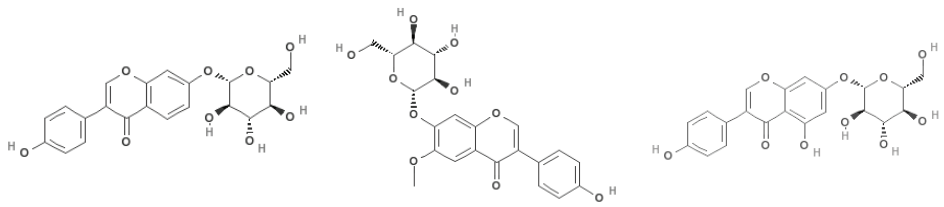


Figure 81. Glycosylated isoflavones found in soy milk. From left to right, daidzin, glycitin and genistin (corresponding to daidzein glycoside, glycitin glycoside and genistin glycoside).

Both the free and covalently bound enzymes were assayed in this application as described in Section 3.3.8. A solution containing a concentrated buffered extract of soy isoflavones was incubated with free and covalently bound Bgl and the commercial preparation AR2000 (DSM, Delft, The Netherlands). The amount of each released isoflavone was determined by HPLC and compared with the relative enzymatic activity on pNPG. In this case, NUVM7-1-Bgl, NUVM7-2-Bgl and NUVM7-3-Bgl biocatalysts were assayed.

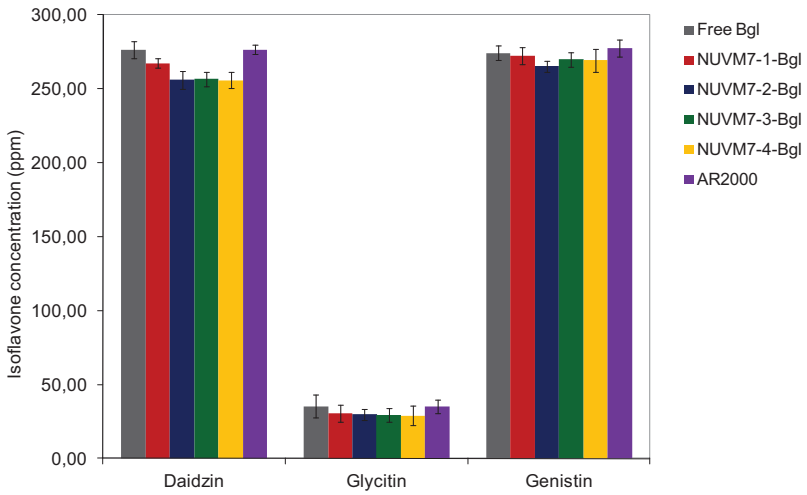


Figure 82. Isoflavone release with free and covalently bound β-glucosidase.

As can be seen in Figure 82, an efficient hydrolysis of all bound isoflavones is accomplished both with free Bgl and the commercial mixture AR2000. The covalently bound enzymes retain their capacity to release daidzin, genistein and genistin although immobilised enzymes show slightly incomplete releases. In relation to their hydrolytic activity on pNPG, the ratio of isoflavone release appears unaltered for the three forms studied (data not shown).

Transglycosylation

The use of α-L-arabinofuranosidase and β-glucosidase in transglycosylation reactions was described in Chapter IV. Abf and Bgl were able to catalyse the methylation reactions of arabinose and glucose, respectively. The same reactions were carried out, under optimised conditions, with the immobilised enzymes.

The ratio of methylated and hydrolysed sugar was calculated and compared between the free and immobilised enzymes (please note that a ratio equal to 1 corresponds to an equal amount of hydrolysed *versus* methylated product, and thus to a yield of 50% in the transglycosylated product).

In the case of Abf, the results are shown in Figures 83 and 84. In accordance to the results obtained in Chapter IV, both the free and immobilised enzyme preferably hydrolyse pNPA and varying yields ranging between 30-40% are obtained, although again slightly higher values are produced when the immobilised enzymes are used.

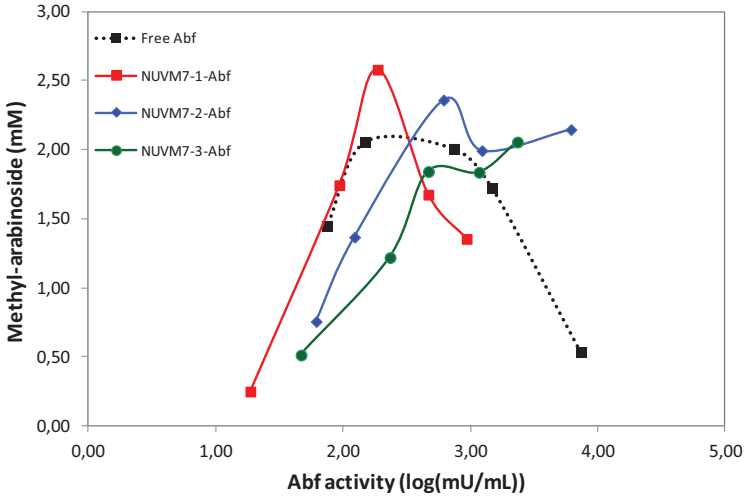


Figure 83. Transglycosylation yield achieved by the different biocatalysts NUVM7-Abf.

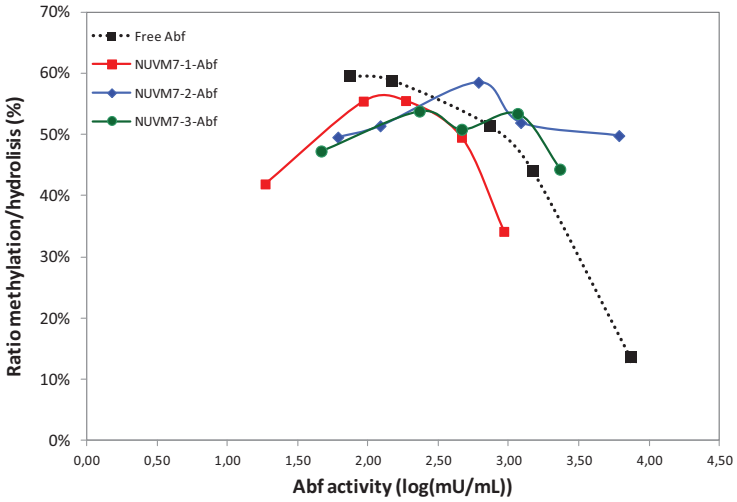


Figure 84. Transglycosylation selectivity achieved by the different biocatalysts NUVM7-Abf.

As seen in Chapter IV, methylation yield accomplished with Bgl is very high in all cases and the free enzyme achieves near 80% conversion. The ratio of methylated and hydrolysed sugar was calculated attending to the amounts of methyl-glucose or methyl-arabinose formed, and glucose or arabinose released, as determined by HPLC, and compared between the free and immobilised enzymes (Figures 85 and 86).

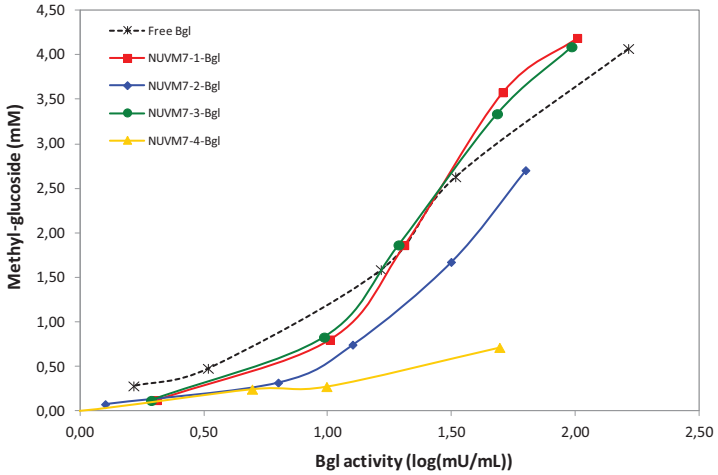


Figure 85. Transglycosylation yield achieved by the different biocatalysts NUVM7-Bgl.

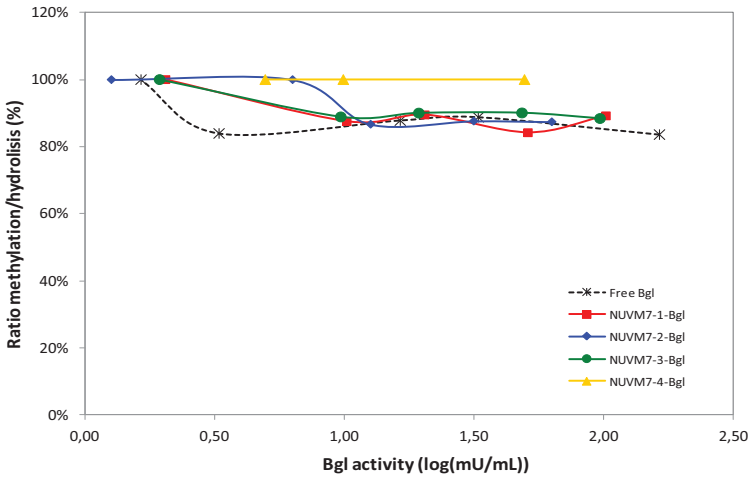


Figure 86. Transglycosylation selectivity achieved by the different biocatalysts NUVM7-Bgl.

As can be seen in Figure 86, the relative ratio of methylation seems to increase for all supports and the amount of methylated glucose is comparatively higher when some of the immobilised enzymes are used. This may be due to an altered ratio k_2/k_3 (kinetic constants of each reaction, see Chapter IV), as either the physicochemical environment of the support (altered charge profile, high concentration of silanol groups) or the covalent binding of the support can favour the addition of methanol in relation to water.

7.4 Discussion

In this chapter, several biocatalysts have been obtained by covalent immobilisation of the heterologous enzymes produced in the previous sections, and their properties and use in different biotechnological applications have been studied.

For this purpose, UVM-7 mesoporous materials have been firstly evaluated as supports for immobilisation. These organosilicas are characterised by a hierarchical, bimodal system with interconnected pores. Different UVM-7 structures were compared using lysozyme as a model enzyme. This particular architecture, together with their high specific surface, favours quick and high enzyme loading in conventional adsorption experiments. Improved adsorption kinetics in relation to conventional MCM-41 materials has also been shown (Pérez-Cabero *et al*., 2011). In order to immobilise acidic enzymes such as Abf, aminated UVM-7 derivatives (NUVM-7) were used. Rapid adsorption of Abf was also verified in NUVM-7 supports and the enzyme was found located in the external surface, in the mesopores.

Different aminated organosilicas were selected and used as enzyme carriers for Abf and Bgl immobilisation. Whereas all NUMV-7 materials tested could be used to adsorb Bgl, Abf could only be retained in the materials bearing a significant number of aminated radicals in the mesopores. This effect is possibly a consequence of the smaller size of the enzyme and also the reduced number of acidic residues available for immobilisation in comparison to Bgl. For this reason, impregnation seems a more suitable method for the obtention of aminated organosilicas destined to the immobilisation of larger glycosidases. The physicochemical properties of the obtained

biocatalysts were determined by N_2 adsorption-desorption and TEM. After immobilisation, NUVM-7 materials retain their characteristic structure and as expected, according to their size, both enzymes are preferably located in the external pores.

These experiments support the use of UVM-7 and NUVM-7 materials for fast and efficient adsorption of enzymes. Small globular proteins, such as lysozyme, may be retained both in nano and mesopores, and bigger enzymes, such as Abf or Bgl, which will be adsorbed in the external surface of the mesopores. In the case of acidic enzymes, the type of aminated radical employed for functionalisation will determine the activity and immobilisation yield for each of them.

Mesoporous silicas have been frequently used for the immobilisation of enzymes and covalent binding is usually achieved employing glutaraldehyde for support activation. The use of aminated organosilicas to adsorb acidic enzymes was recently reported (Casadonte *et al.*, 2010). In this contribution, a simple protocol based on the use of carbodiimides to couple amino groups with the free carboxyl groups of the enzymes enables fast and efficient covalent binding on the NUVM-7 supports. When Abf and Bgl are immobilised onto the selected NUVM-7 organosilicas, the biochemical properties of the resulting biocatalyst vary in relation to the free enzyme, depending on both the enzyme and the support. This may result from multipoint attachment (Blanco *et al.*, 1989), favoured in these enzymes in which a significant number of acidic residues are found.

In the case of Abf, the properties of the enzyme are modified with the three materials tested, and the differences are wider when the support bearing shorter radicals (NUVM7-1) is used. In this case, the lowest pH optimum is observed, although the other materials introduce similar patterns. This effect has been previously described for other glycosidases (Spagna *et al.*, 2000b) and is possibly ascribable to a higher local pH in the vicinity of the active site of the enzyme. A general stabilising effect can also be observed in relation to temperature and resistance to ethanol, also previously described for other covalently immobilised enzymes (Spagna *et al.*, 2000b; Wang *et al.*, 2001). Once again, the shift in relation to the behaviour of the free enzyme is sharper in the case of NUVM7-1 material. The effect of different radical lengths in the properties of the immobilised enzyme has been previously described. Whereas shorter radicals can

promote larger modifications in the behaviour of the immobilised enzyme, longer radicals enable better motility and substrate accessibility to the active site (Guisán, 2006). In this way, the changes introduced by immobilisation onto NUVM7-1 support correlate with an increase in the K_m constant (Table 50), possibly as a consequence of conformational changes, whereas NUVM7-3 material seems to preserve better Abf activity after adsorption (Figure 64) and the covalently immobilised enzyme shows similar kinetic properties to free Abf.

Regarding technological application, Abf biocatalysts also show differential properties in relation to the free enzyme. Whereas the hydrolysis of larger substrates such as arabinan seems to be hindered (Table 49), the conversion in transglycosylation reactions is increased in comparison to free Abf, for comparable specific hydrolytic activity. Interestingly, immobilisation onto NUVM-7 supports seems to tune transglycosylation properties of the enzyme and different patterns of conversion and selectivity are found in relation to hydrolytic activity. This effect may be a consequence of the modified biochemical properties of the immobilised enzyme: modified local pH (as transglycosylation involves various steps of nucleophylic substitution), improved activity at high temperatures (the reaction is carried out at 50ºC) and in the presence of alcohols such as the substrate, methanol. Since these supports also offer an improved chemical resistance to organic solvents, covalent immobilisation of Abf onto NUVM-7 organosilicas seems a promising tool for synthesis applications.

When Bgl is immobilised in the available NUVM-7 materials, several differences are observed in relation to biochemical properties of the resulting biocatalysts. Apparently, the supports are ranged in two different groups, and whereas silica with shorter radicals (NUVM7-1, NUVM7-4) modify their behaviour towards pH, supports with longer chains (NUVM7-2 and NUVM7-3) seem to promote thermal stabilisation and modified substrate specificity. It is unclear how Bgl immobilisation can generate these changes in specificity but the affinity of the supports for the different compounds may also induce local gradients of substrate concentration in the vicinity of the active site that in turn promote apparent affinity changes.

Technological application of immobilised Bgl is particularly relevant due to the biotechnological interest of this enzyme. The hydrolytic capacity of the biocatalysts has

been verified with different glycoconjugates and also naturally occurring glycosylated compounds, such as soy isoflavones. Immobilised Bgl enable similar conversions in relation to the free enzyme and also in comparison with the commercial benchmark. Since both enzymes (Abf and Bgl) have been used for terpene release in winemaking applications (Spagna *et al.*, 2003), joint immobilisation has also been attempted and successfully achieved, and the properties of the jointly immobilised enzymes are in accordance to their severally immobilised equivalents.

Bgl has previously been shown to be a particularly suitable enzyme for transglycosilation reactions, due to the high conversion rates found for the free enzyme at low concentrations (see Chapter IV). The selectivity towards methylation is also very high when the immobilised enzymes are used, reaching more than 90% in most of the cases.

Other examples of the use of immobilised glycosidases (particularly, β-glucosidases) to catalyse transglycosylation reactions can be found in the literature, although none of them refers to nanosilicas as enzyme carriers. Park and coworkers (2000) used almond β-glucosidase to produce alkyl-glucosides in high yield. Interestingly, transglycosylation activity was improved after immobilisation. Similarly, Nagatomo and coworkers (2005) immobilised β-glucosidase from *Pyrococcus furiosus* in gelatin-gels, and reported the improvement of the transglycosylating activity in relation to the free enzyme. However, Gargouri and coworkers (2004) immobilised fungal β-glucosidase and β-xylosidase in polyacrylamide gels and by adsorption, in diverse resins, but oligosaccharide formation reactions were impaired although the transglycosylating activity of the free enzyme was high. The combined use of mesoporous organosilicas such NUVM-7 and carbodiimides for covalent bond formation with glycosidases enables convenient loading, good mass transfer of substrates and products, and potential tuning of the biochemical properties of the resulting biocatalyst for biosynthetic reactions.

In conclusion, Abf and Bgl can be covalently immobilised onto different mesoporous materials. The structure of these biocatalysts, based on the architecture of UVM-7 materials, with large pores and pore-windows, provides high accessibility to the active site of the enzyme thanks to the hierarchical bimodal pore system. In both cases, the

properties of the immobilised biocatalysts are comparable to those of the free enzymes, although wider differences are found in the case of Abf. The improved resistance towards typically denaturing parameters, such as temperature or solvents, increases possibly due to a stabilising effect promoted by multipoint covalent binding. This, together with the shift of optimum pH towards more acidic conditions, favours the use of the enzymes in known biotechnological applications such as aroma release in winemaking as these (low pH, high ethanol concentration) are typical industrial process conditions. In addition, the resistance of these materials to solvents, in contrast to other natural supports such as agarose or chitosan supports its use in organic synthesis. Very likely, the strategy presented herein might be extended and adapted to the preparation of a diversity of hybrid materials of biotechnological interest.

VIII. Future Work

In relation to the results detailed previously, and enduring generality, future lines of work can be envisaged that are introduced in the following sections.

8.1 Glycosidase expression and use in biocatalysis

Abf enzyme has been successfully expressed by the selected *P. pastoris* constructs. These strains are now available for optimisation and scale-up of fermentation protocols and may be used to implement advanced monitoring and control strategies. As a first step, standard protocols described for *Pichia pastoris* growth and induction may be used to establish the maximal enzyme concentration that can be attained in high-cell density cultures in open-loop control.

On the other hand, it has not been possible to achieve high Bgl activity levels. This seems to be a consequence of the toxic effect of Bgl expression, caused either by folding or secretional bottlenecks. In order to establish whether Bgl toxic effect is increased by methanol consumption, strains expressing Bgl under control of alternative *P. pastoris* promoters such as the promoter of formaldehyde dehydrogenase (pFLD) may be used. FLD structural genes confer resistance to formaldehyde and are strongly but independently induced by methanol (as sole carbon source, in combination with ammonium sulphate as nitrogen source) or methylamine as sole nitrogen source (with glucose as carbon source). The promoter of the gene encoding glyceraldehyde phosphate dehydrogenase (pGAP) in *P. pastoris* has also been used successfully for recombinant protein expression.

Regarding glycosidation pattern, it has been established that Bgl is hyperglycosylated both in *P. pastoris* as in *Saccharomyces cerevisiae*. However, the number and type of

sugar residues may be different in each case. The modification of glycosylation pattern depending on the composition of the growth media has also been described in *P. pastoris*. Since glycosylation may affect the stability and secretability of the protein, it could be interesting to determine the glycosylation sites, oligosaccharide pattern, and monosaccharide composition in different culture conditions or strains.

Promising results have been obtained with both enzymes in transglycosylation reactions. These assays can be extended to evaluate the ability of these glycosidases to produce other alkyl-glycosides using alcohols other than methanol, with longer alkyl chains, phenyl groups, or secondary or terciary configurations. Also different glycoconjugates may be produced in combination with other sugar donors.

Site-directed mutagenesis has been used to reduce the number of glycosylation sites, but also to improve transglycosylation yield and tune specificity. Mutagenised glycosynthases have been produced to favour the introduction of particular functional groups, such as halogen groups, and a similar strategy may be applicable to Abf and Bgl enzymes.

Finally, inmobilisation of these glycosidases has proven to be a valuable tool to improve stability and modify the properties of the enzymes that is readily extendable to other proteins. The possibility of using these catalysts in organic reaction media may be interesting in order to explore if immobilisation by covalent binding can modify the selectivity, specificity or yield of different transglycosylation reactions.

8.2 Modelling extensions

The model has been developed and fitted in order to represent the basic metabolism of *P. pastoris* cultures together with the energetic resources devoted to recombinant protein production. Several extensions or modifications of the current system can be envisaged in order to increase the number of possible applications.

- Extensions of the current stoichiometric matrix - genome-scale modeling: as a first approximation, a reduced dimension for the constraint-based model has

been preferred in order to ensure adaptability to process monitoring in bioreactor operation. However, several genome scale models have been published (Chung *et al.*, 2010; Sohn *et al.*, 2010) and it should be expected the whole set of metabolic reactions for *P. pastoris* to be available soon. In this way, the current constraint based model can be extended to represent *P. pastoris* metabolism in the most accurate way possible. The performance of both approaches may be then compared and exploited to define normalised strategies for model reduction. In a more limited approach, the stoichiometric representation of recombinant protein may also include particular building blocks required, although it has also been stated that the amount of metabolic equivalents employed for protein synthesis is not significant in relation to energy demand or precursors used for biomass growth. Finally, other relevant substrates such as glucose and sorbitol may be considered as external compounds and adapt parameter fitting to these particular systems.

- Use with other model organisms: the strategy used to extend the constraint-based model to represent protein productivity can be readily adapted to other common microbial hosts for recombinant protein production, such as *Escherichia coli* or *S. cerevisiae*. Well defined constraint-based models have been reported in both systems (Förster *et al.*, 2003; Reed *et al.*, 2003), and a limited set of cultivation strategies (based on general in glucose feeding as the main substrate) are used in practice for each of them. Please note that no signaling approaches have been introduced to describe protein production. The amount of protein that is expected to be generated depends on the available resources and the amount of these that can be assigned. This means that the structure of the extended constraint-based model including protein productivity is in principle valid for any microbial system. On the other hand, protein productivity estimation has shown to be valid for different recombinant strains. Obviously, it is also possible to particularise parameter fitting for a given protein if a sufficient number of experimental values are available, although this is not strictly required.

- Adaptation to other metabolic products: the strategy developed to include protein productivity in the constraint-based model was initially described to represent penicillin production in a fungal host (vanGulik *et al.*, 2001). It stands to reason that this modeling strategy could be also applied to other microbial systems producing valuable metabolic products in high yields. Since no particular relationship has been established between growth and product, both growth and non-growth related products could be considered. Productivity is an essential parameter in most industrial biotechnology processes, such as biofuel production (Wyman, 1994) or biopolymer obtention (Choi and Lee, 1999). Again, detailed models are available for these systems (Pitkanen *et al.*, 2003; Nogales *et al.*, 2008). The developed strategy can be used as a simple and general approach to validate metabolic networks and to represent productivity in relation to devoted energy resources.

8.3 Monitoring, optimisation and control

This contribution introduces a simple but structured model of the yeast *P. pastoris* as a means to predict protein productivity using only a few extracellular measurements. This approach provides the basis for grey-model based process monitoring strategies, and may be used to develop rational process optimisation methodologies. These applications will be presented herein as examples of potential future uses of the model.

Online process monitoring

The system in which the proposed model will be exploited are stirred-tank bioreactors. Whereas continuous feeding strategies are frequent in process development, batch and fed-batch configurations are preferred for production purposes in *P. pastoris* cultures. The first two result in non-stationary systems for which several considerations must be taken into account:

- Sensorisation and external flux calculation: in order to perform consistent calculation of external fluxes in particular time-points during the whole duration of the process, frequent online determination of extracellular species must be ensured. In the case of the analysed system, temporal evolution of biomass (Navarro *et al.,* 2001), methanol (Mayson *et al.*, 2003) and offline gases for OUR and CPR calculation can be implemented online. In the case of glycerol, it should be determined atline by biochemical analysis or HPLC. Automated sampling will be possibly required in this case. Local values of external fluxes may be then calculated by linear regression fitting in the vicinity of the last available point. A simpler set-up will result in the case of fed-batch mode, carbon-limited conditions, where only residual substrate (glycerol and methanol) concentrations will be found in the broth. Exponential feeding strategies for constant specific growth rate maintenance may also be applied in this case.

- MFA in non-stationary conditions: although traditional MFA relies on the pseudo-steady state assumption, which is generally understood to be applicable in systems working at constant specific growth rate, some authors have already reported MFA applications in non-stationary systems. Antoniewicz and coworkers (2007) applied MFA to analyse 2,3-propanediol production by recombinant *E. coli* batch cultures on glycerol. Also Provost and Bastin (2004) studied the obtention of reduced-macroscopic models based on metabolic networks for CHO cells in different growth phases. Even in batch or fed-batch systems, the dynamics of intracellular metabolites can still be considered to be much quicker than those of non-balanceable compounds, such as biomass or extracellular products.

Three main applications can be then envisaged in relation to process analysis and monitoring:

- Validation of current process settings – failure control: if the complete set of extracellular fluxes (or a suitable combination, as detailed in Section 3.5) is available around a particular time-point, PS-MFA could be used to calculate a possibility index for this local experimental dataset. A low possibility value

would indicate the inconsistency between model and experimental data, and thus reveal either incorrect sample/sensor processing, or underlying gross process anomalies (i.e. cross-contamination, mutations, cell lysis) that may be early detected in the culture.

- Prediction of protein productivity: the same operation will provide a prediction for one of the process targets that cannot be easily measured and thus is the main, direct application of this contribution. This makes it possible to control possible deviations from the desired physiological state. This will be indirectly reflected in the combination of biomass growth, substrate and oxygen uptake and carbon dioxide production rates and will possibly result in irregular protein productivity predictions. This estimation could also guide end-of-run decisions, particularly in long-term inductions.
- Prediction of unknown, intracellular flux rates: a good agreement between ^{13}C-based calculation of intracellular fluxes and the model estimates has been shown in several literature scenarios. PS-MFA will provide a distribution of intracellular fluxes for each time-point that may be analysed as a valuable insight of the metabolic state of the culture. This is also particularly interesting in experimental fermentation runs, when substrates other than glucose are used, such as glycerol and methanol, expensive or unavailable in ^{13}C-isotope-labeled forms.

Process optimisation – Flux Balance Analysis

The current model defines which combinations of fluxes are feasible for the constraints provided. Since the manipulated variables (glycerol, methanol and oxygen uptake rate) are not sufficient to determine the system, the direct analysis of which combinations or valid flux vectors generate the highest protein productivity rate is not possible. This is, the steady-state flux distribution (flux vector) cannot be predicted using only the manipulated variables. However, this can be achieved with the use of an objective function within an optimality problem, by applying FBA.

FBA is an extension of MFA based on the definition of a linear programming problem, in order to select a set of particular solutions (a flux combination or vector) that fulfills the constraints and maximises a particular function. The most common objective is maximising biomass growth rate, assuming that the behavior of the modeled organism would correspond to flux distributions that enable maximal survival (growth). This hypothesis has been verified experimentally in different microbial systems (Sauer *et al.*, 1997) and other objective functions, such as minimal energy drain in relation to growth have also been validated (Schuetz *et al.*, 2007).

In our particular system, the application of FBA can be exploited:

- In order to analyse in terms of optimality the available experimental flux combinations. In this way, the actual biomass growth rate can be compared to the possible optimum for a given set of substrate consumption rates. This provides a complimentary analysis to the comparison with the maximal allowed biomass yield. Since it has been shown that higher protein productivities correlated to lower biomass yields, one should expect that experimental datasets producing higher amounts of protein correspond to biomass growth rates further away from the predicted optimum.

- In order to select the optimal combination of substrate consumption rates. For a given set of parameters describing the growth conditions (such as glycerol, methanol and oxygen consumption rates in our system), the optimal flux distribution can be calculated for all points in this (three-dimensional) space. This will result in a finite number of different flux distributions, with specific combinations of substrate uptake and biomass growth rates. If PS-MFA is applied in each of these points, the space of flux vectors can be screened for the highest protein production rate. Other optimisation criteria to select biomass growth rate, such as afore mentioned, could also be applied in this case.

IX. Conclusions

1. Two enzymes of biotechnological interest, Abf from *A. niger* and Bgl from *C. molischiana*, have been cloned and expressed in *P. pastoris* under control of the AOX gene promoter. Whereas Abf is efficiently expressed and secreted by the yeast, lower values of Bgl activity were obtained, possibly related to limitations in folding or secretional pathways. Both enzymes have been purified and characterised, showing similar properties to the recombinant enzymes produced by *S. cerevisiae* in terms of pH, temperature, resistance to ethanol and glucose and substrate specificity. The enzymes have been shown to catalyse transglycosylation reactions with fast kinetics and high yields at low enzyme concentrations.

2. A constraint-based model of *P. pastoris* metabolism has been devised and validated to represent the yeast growth on glucose, glycerol and methanol. For this purpose, a new method to assess consistency has been set, based on PS-MFA, taking into account scarcity and imprecision in the available measurements. The model has been used to calculate the maximal theoretical growth yields, and to quantitatively predict biomass growth rate on different substrates and their combinations using only a few extracellular measurements. The model has also been shown to provide a trustworthy profile of the intracellular flux distribution in *P. pastoris* cultures growing on glucose, glycerol and methanol.

3. The constraint-based model has been extended in order to relate the available measurements to protein productivity, through the balance of key energetic equivalents such as ATP. In this way, a simple, linear,

but structured model of the yeast *P. pastoris* is defined to predict protein productivity using only a few extracellular known rates and applying PS-MFA. The model is valid for a variety of *P. pastoris* cultures, expressing different heterologous proteins, and in a wide set of operating conditions including different substrates and growth rates. This extended model has been used to analyse the behaviour of the developed recombinant strains in batch cultures on glycerol and methanol. Several datasets have been generated corresponding to exponential growth states on each substrate. The calculated external fluxes have shown good agreement with the model, showing that this strategy can be used to produce valuable estimates of biomass and protein productivity, as a first step in the development of online monitoring and control strategies.

4. The developed methodology is fully applicable to other microbial systems and also adaptable, in computational terms, to larger scale networks, in which the constraint-based model can be validated and exploited without attending to its under-determinacy in a possibilistic framework. The approach can be also easily implemented in industrial environments, as a consistent prediction can still be obtained when only few online measurements are available. This sets up a suitable framework for fermentation monitoring as both biomass growth and protein production rate can be conveniently estimated.

5. A variant of classical mesoporous sieves, UVM-7 materials, characterised by a bimodal architecture, has been validated for enzyme immobilisation. The supports show high enzyme loading and fast adsorption rates. Immobilisation by covalent binding of acidic enzymes such as Abf and Bgl has been demonstrated using conveniently functionalised supports. The influence of immobilisation on optimum environmental conditions, effect of inhibitors and substrate specificity has been determined. The properties of the

resulting biocatalyst vary depending on the type of amino radical used for functionalisation.

6. A shift of optimum pH of Abf towards acidic conditions and higher optimum temperature were observed. The immobilised enzyme kept its resistance towards ethanol and glucose. In the case of Bgl, a stabilising effect was observed in relation to temperature, together with a shift of optimum pH towards acid values for some of the supports. Immobilised Bgl could efficiently release daidzin, glycitin and genistin, the main glycosylated isoflavones extracted from soy milk. Methylation of glucose and arabinose has been also achieved with immobilised Abf and Bgl, with high conversion yields and selectivity.

X. References

Akiyama T, Kaku H, Shibuya N. (1998) A cell wall-bound β-glucosidase from germinated rice: purification and properties. Phytochem. 48, 49-54.

Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, Stephanopoulos G. (2007) Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. Metabol. Eng. 9, 277–292.

Arroyo M. (1998) Inmovilización de enzimas. Fundamentos, métodos y aplicaciones. *Ars pharmaceutica.* 39, 23-39.

Ausubel FM. (2010) Current Protocols in Molecular Biology. (Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K eds.) Ed. John Wiley & Sons. ISBN 978-0-471-50338-5.

Ayra-Pardo C, Motejo-Sierra IL, Vasquez-Padron RI, Garcia-Martinez C. (1999) β-D-glucuronidase gene from *Escherichia coli* is a functional reporter in the methylotropic yeast *Pichia pastoris.* Lett. Appl. Microbiol. 28, 278-283.

Bae EA, Park SY, Kim DY. (2000) Constitutive β-glucosidases hydrolyzing ginsenoside Rb1 and Rb2 from human intestinal bacteria. Biol. Pharm. Bull. 23, 1481-1485.

Bailey JE. (1998) Mathematical modelling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol. Prog. 14, 8-20.

Balkus KJ, Diaz JF. (1996) Enzyme immobilization in MCM-41 molecular sieve. J. Mol. Cat. B: Enz. 2, 115-126.

Basso A, Ducret A, Gardossia L, Lortie R. (2002) Synthesis of octyl glucopyranoside by almond β-glucosidase adsorbed onto Celite R-640. Tetrahed. Lett., 43, 2005-2008.

Bastin G, Dochain D. (1990) On-line estimation and adaptative control of bioreactors. Ed. Elsevier; Amsterdam (The Netherlands). ISBN 0-444-88430-0.

Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, Maaheimo H, Gasser B, Albiol J, Mattanovich D, Ferrer P. (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions, BMC Syst. Biol. 1, 141-152.

Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P. (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. Biotechnol. Bioeng. 100, 177-183.

Bencúrová M, Rendi D, Fabini G, Kopecky EM, Altmann F and Wilson IBH. (2003) Expression of eukaryotic glycosyltransferases in the yeast *Pichia pastoris*. Biochimie 35, 413-422.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004) Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783-795.

Berger B. (2009) Biotechnology of flavours, the next generation. Biotechnol. Lett. 31, 1651-1659.

Berrin J, Williamson G, Puigserver A, Chaix JC, McLauchlan W, Juge N. (2000) High-level production of recombinantfungal endo-β-1,4-xylanase in the methylotropic yeast *Pichia pastoris*. Prot. Expr. Purif.19, 179-187.

Bhatia Y, Mishra S, Bisaria VS. (2002) Microbial β-glucosidases, cloning, properties and applications. Crit. Rev. Biotechnol. 22, 375-407.

Birk R, Bravdo B, Shoseyov O. (1996) Detoxification of cassava by *Aspergillus niger* B-1. Appl. Microbiol. Biotechnol. 45, 411–414.

Blanco RM, Calvete JJ, Guisán JM. (1989) Immobilization-stabilization of enzymes: variables that control the intensity of the trypsin (amine) agarose (aldehyde) multipoint attachment. Enz. Microb. Technol. 11, 353-359.

Blanco RM. (1979) PhD Thesis. Inmovilización de glutamato-racemasa sobre soportes sólidos. Univ. Complutense de Madrid.

Blank LM, Sauer U. (2004) TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. Microbiology 150, 1085-1093.

Boer H, Teeri TT, Koivula A. (2000) Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. Biotechnol. Bioeng. 69, 486-494.

Bom IJ, Dielbandhoesing SK, Harvey KN, Oomes SJ, Klis FM, Brul S. (1998) A new tool for studying the molecular architecture of the fungal cell wall: one-step purification of recombinant *Trichoderma* β-1-6-glucanase expressed in *Pichia pastoris*. Biochim. Biophys. Acta, 1425, 419-424.

Boze H, Celine L, Patrick C, Fabien R, Christine V, Yves C, Guy M. (2001) High-level secretory production of recombinant porcine follicle-stimulating hormone by *Pichia pastoris*. Proc. Biochem. 36, 907-913.

Bradford MM. (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72, 248–254.

Brierley RA, Bussineau C, Kosson R, Melton A, Siegel R. (1990) Fermentation development of recombinant *Pichia pastoris* expressing the heterologous gene: bovine lysozyme. Ann. NY Acad. Sci. 589, 350–362.

Brozmanova J and Holinova Z. (1988) A rapid preparation of plasmid DNA from *Saccharomyces cerevisiae*. Folia Microbiol. 33, 34–37.

Cairns JRK, Champattanachai V, Srisomsap C, Wittman-Liebold B, Thiede B, Svasti J. (2000) Sequence and expression of thai rosewood β-glucosidase/β-fucosidase, a family 1 glycosyl hydrolase glycoprotein. J. Biochem. 128, 999-1008.

Cakir T, Kirdar B, Onsan ZI, Ulgen KO, Nielsen J. (2007) Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. BMC Syst. Biol. 1, 18-27.

Carinhas N, Bernal V, Teixeira AP, Carrondo MJT, Alves PM, Oliveira R. (2011) Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products. BMC Syst. Biol. 5, 34-46.

Carlson R, Fell D, Srienc F. (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol. Bioeng. 79, 121–134.

Casadonte F, Pasqua L, Savino R, Terracciano R. (2010) Smart trypsin adsorption into n-(2-aminoethyl)-3-aminopropyl-modified mesoporous silica for ultra fast protein digestion. Chem. Eur. J. 16, 8998–9001.

Casdaban M, Cohen S. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138, 179-207.

Çelik E, Pinar C, Calik SG, Oliver SG. (2010) Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: effects of methanol feeding rate. Biotechnol. Bioeng. 105, 317-329.

Çelik EC, Calik P, Halloran SM, Oliver SG. (2007) Production of recombinant human erythropoietin from *Pichia pastoris* and its structural analysis. J. Appl. Microbiol. 103, 2074-2094.

Cereghino JL, Cregg JM. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. 24, 45-66.

Chae HJ, Delisa MP, Cha HJ, Weigand WA, Rao G, Bentley WE. (2000) Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. Biotechnol. Bioeng. 5, 275-285.

Charoenrat T, Cairns MK, Stendahl-Andersen H, Jahic M, Enfors SO. (2005) Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes, Bioproc. Biosyst. Eng. 27. 399–406.

Chen HM, Ford C, Reilly PJ. (1994) Substitution of asparagine residues in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation. J. Biochem. 301, 275-281.

Chen Y, Krol J, Cino J, Freedman D, White C, Komives E. (1996) Continuous production of thrombomodulin from a *Pichia pastoris* fermentation. J. Chem. Technol. Biotechnol. 67, 143-148.

Choi J, Lee SY. (1999) Factors affecting the economics of polyhydroxyalkanoate production by bacterial fermentation. Appl. Environ. Microbiol. 51, 13-21.

Chung B, Selvarasu S, Camattari A, Ryu J, Lee H, Ahn J, Lee H. (2010) Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb. Cell Fact. 9, 50-64.

Ciesla U, Schüth F. (1999) Ordered mesoporous materials. Micropor. Mesopor. Mat. 27, 131-149.

Corma A, Fornés V, Jordá JL, Rey F, Fernández-Lafuente R, Guisán JM, Mateo C. (2001) Electrostatic and covalent immobilization of enzymes on ITQ-6 delaminated zeolitic materials. Chem. Commun. 5, 419-420.

Corma A, Fornés V, Rey F. (2002) Delaminated zeolites: an efficient support for enzymes. Adv. Mat. 14, 71-74.

Cortassa S, Aon JC, Aon MA. (1995) Fluxes of carbon, phosphorylation, and redox intermediantes during growth of *Saccharomyces cerevisiae* on different carbon sources. Biotechnol. Bioeng. 4, 193-208.

Cos O, Ramón R, Montesinos JL and Valero F. (2006a) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. Microb. Cell Fact. 5, 17-23.

Cos O, Ramón R, Montesinos JL, Valero F. (2006b) A simple model-based control for *Pichia pastoris* allows a more efficient heterologous protein production bioprocess. Biotechnol. Bioeng. 5, 145-154.

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol. Cell. Biol. 9, 1316-1323.

Crous JM, Pretorius IS, Van Zyl WH. (1996) Cloning and expression of the α-L-arabinofuranosidase gene (ABF2) of *Aspergillus niger* in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 46, 256-260.

Crout DHG, Vic G. (1998) Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. Curr. Op. Chem. Biol. 2, 96-111.

Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto R, Moosmayer D, Donner P, Carrondo MJ. (2004) Methanol induction optimization for scFv antibody fragment production in *Pichia pastoris*. Biotechnol. Bioeng. 20, 458-467.

Curvers S, Linnemann J, Klauser T, Wandrey C, Takors R. (2002) Recombinant protein production with *Pichia pastoris* in continuous fermentation - kinetic analysis of growth and product formation. Eng Life Sci. 2, 229-235.

Dabros M, Schuler MM, Marison IW. (2010) Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. Bioproc. Biosyst. Eng. 33, 1109-1118.

Daly R, Milton TWH. (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. Rev. J. Mol. Recognit. 18, 119-138.

Dan S, Marton I, Dekel M, Bravdo BA, He S,Withers S, Soseyov O. (2000) Cloning, expression, characterization andnucleophile identification of family 3, *Aspergillus niger β*-glucosidase. J. Biol. Chem. 275, 4973-4980.

d'Anjou MC, Daugulis AJ. (1997) A model-based feeding strategy for fed-batch fermentation of recombinant Pichia pastoris. Biotechnol. Technol. 11, 865-868.

d'Anjou MC, Daugulis AJ. (2001) A rational approach to improving productivity in recombinant *Pichia pastoris* fermentation. Biotechnol. Bioeng. 72, 1-11.

Deere J, Magner E, Wall JG, Hodnett BK. (2001) Adsorption and activity of cytochrome C on mesoporous silicates. JCS Chem. Commun. 5, 465-466.

Dias JML, Oehmen A, Serafim LS, Lemos PC, Reis MAM, Oliveira R. (2008) Metabolic modelling of polyhydroxyalkanoate copolymers production by mixed microbial cultures. BMC Syst. Biol. 8, 59-78.

Dietzsch C, Spadiut O, Herwig C. (2010) A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for *Pichia pastoris*. Microb. Cell Fact. 10, 14-23.

Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D. (2009) The effect of temperature on the proteome of recombinant *Pichia pastoris*. J. Prot. Res. 8, 1380–1392.

Edwards JS, Covert M, Palsson B. (2002) Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. 4, 133–140.

Edwards JS, Ibarra RU, Palsson BO. (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nature Biotechnol. 19, 125–130.

El Haskouri J, Ortiz D, Guillem C, Beltrán-Porter A, Caldés M, Marcos M.D, Beltrán-Porter D, Latorre J, Amorós P. (1999) Surfactant assisted synthesis of mesoporous alumina showing continuously adjustable pore sizes. Adv. Mat. 11, 379-381.

El Haskouri J, Ortiz D, Guillem C, Beltrán-Porter A, Caldés M, Marcos M.D, Beltrán-Porter D, Latorre J, Amorós P. (2000) Generalised synthesis of ordered mesoporous oxides: the atrane route. Solid State Sci. 2, 405-420.

El Haskouri J, Ortiz D, Guillem C, Beltrán-Porter A, Caldés M, Marcos MD, Beltrán-Porter D, Latorre J, Amorós P. (2001) Ordered mesoporous materials: composition and topology control through chemistry. Int. J. Inorg. Mat. 1, 283-291.

El Haskouri J, Ortiz D, Guillem C, Beltrán-Porter A, Caldés M, Marcos MD, Beltrán-Porter D, Latorre J, Amorós P. (2002a) Hierarchical porous nanosized organosilicas. Chem. Mat. 14, 4502–4504.

El Haskouri, J, Ortiz D, Guillem C, Beltrán-Porter A, Caldés M, Marcos MD, Beltrán-Porter D, Latorre J, Amorós P. (2002b) Silica based powders and monoliths with bimodal pore systems. Chem. Commun. 4, 330-331.

Falkoski DL, Monteze V, Vieira M, Fernandes E, de Almeida MN, Gonçalves E, Tavares S. (2009) Covalent immobilization of α-galactosidase from *Penicillium griseoroseum* and its application in oligosaccharides hydrolysis. Appl. Biochem. Biotechnol. 158, 540–551.

Fehrenbach R, Comberbach M, Pêtre JO. (1992) On-line biomass monitoring by capacitance measurement. J. Biotechnol. 23, 303-314.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121-138.

Fernández-Lafuente R, Guisán JM, Ali S, Cowan D. (2000) Immobilization of functionally unstable catechol-2,3-dioxygenase greatly improves operational stability. Enz. Microb. Technol. 26, 568-573.

Ferrarese L, Trainotti L, Gattolin S, Casadoro G. (1998) Secretion, purification and activity of two recombinant pepper endo-β-1,4-glucanases expressed in the yeast *Pichia pastoris*. FEBS Lett. 422, 23-26.

Fia G, Giovani G, Rosi I. (2005) Study of β-glucosidase production by wine-related yeasts during alcoholic fermentation. A new rapid fluorimetric method to determine enzymatic activity. J. Appl. Microbiol. 99, 509-517.

Fiaux J, Çakar ZP, Sonderegger M, Wüthrich K, Szyperski T, Sauer U. (2003) Metabolic flux profiling of the yeasts *Sacharomyces cerevisiae* and *Pichia stipitis*. Eukar. Cell, 2, 170-180.

Fierobe HP, Mirgorodskaya E, Frandsen TP, Roepstorff P, Svensson B. (1997) Overexpression and characterization of *Aspergillus awamori* wild-type and mutant glucoamylase secreted by the methylotrophic yeast *Pichia pastoris*: comparison with wild-type recombinant glucoamylase produced using *Saccharomyces cerevisiae* and *Aspergillus niger* as hosts. Prot. Expr. Purif. 9, 159-170.

Fischer E, Zamboni N, Sauer U. (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ^{13}C constraints. Anal. Biochem. 325, 308–316.

Flipphi MJA, Van Heuvel M, Van der Veen P, Visser J, De Graaff LH. (1993) Cloning and characterization of the *abf* B gene coding for the major α-L-arabinofuranosidase (ABFB) of *Aspergillus niger*. Curr. Gen. 24, 525-532.

Förster J, Famili I, Fu P, Palsson B, Nielsen J. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res.13, 244-253.

Fredrickson AG, Megee RD, Tsuchiya HM. (1970) Mathematical models for fermentation processes. Adv. Appl. Microbiol. 13, 419–465.

Freer S. (1993) Kinetic characterization of a β-glucosidase from a yeast, *Candida wickerhamii*. J. Biol. Chem. 26, 9337-9342.

Gagneur J, Klamt S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinform. 5, 175-195.

Gallifuoco A, Alfani F, Cantarella M, Spagna G, Pifferi PG. (2003) Immobilized β□glucosidase for the wine-making industry: study of biocatalyst operational stability in laboratory-scale continuous reactors. Proc. Biochem. 35, 179-185.

Gallifuoco A, D'Ercole L, Alfani F, Cantarella M, Spagna G, Pifferi PG. (1997) On the use of chitosan-immobilized β-glucosidase in wine making: kinetics and enzyme inhibition. Proc. Biochem. 33, 163-168.

Gargouri M, Smaali I, Maugard T, Legoy MD, Marzouki N. (2004) Fungus β-glycosidases: immobilization and use in alkyl-α-glycoside synthesis. J. Mol. Cat. B: Enz. 29, 89–94.

Gawande PV. (2003) Preparation, characterization, and application of *Aspergillus sp*. xylanase immobilized on Eudragit S-100. J. Biotechnol. 66, 165-175.

Gil JV, Manzanares P, Genoves S, Valles S, González-Candelas L. (2005) Overproduction of the major exoglucanase of *Saccharomyces cerevisiae* led to an increase in the aroma of wine. Int. J. Food Microbiol. 103, 57-68.

Gimon-Kinsel ME, Jimenez VL, Washmon L, Balkus KJ. (1998) Mesoporous molecular sieve immobilized enzymes. Mesopor. Molec. Sieves, 117, 373-380.

Glick BR. (1995) Metabolic load and heterologous gene expression. Biotechnol. Adv. 2, 247-261.

Godbole S, Decker S, Nieves R, Adney W, Vinzant T, Baker J, Thomas S, Himmel M. (1999) Cloning and expression of *Trichoderma reesei* cellobiohydrolase I in *Pichia pastoris*. Biotechnol. Prog. 15, 828-833.

Gombert AK, Nielsen J. (2000) Mathematical modelling of metabolism. Curr. Op. Biotechnol. 11, 180–186.

Gomez, JM, Romero MD, Fernández TM, Garcia S. (2010) Immobilization and enzymatic activity of β-glucosidase on mesoporous SBA-15 silica. J Porous Mater. 17, 657–662.

Gondé P, Ratomahenina R, Arnaud A, Galzy P. (1985) Purification and properties of an exocellular β-glucosidase of *Candida molischiana* (Zykes) Meyer and Yarrow capable of hydrolyzing soluble cellodextrins. Can. J. Biochem. Cell. Biol. 63, 1160-1166.

González R, Andrews BA, Molitor J, Asenjo JA. (2003) Metabolic analysis of the synthesis of high levels of intracellular human SOD in *Saccharomyces cerevisiae* rhSOD. Biotechnol. Bioeng. 20, 152-169.

González-Candelas L, Cortell A, Ramón D. (1995) Construction of a recombinant wine yeast strain expressing a fungal pectate lyase gene. FEMS Microbiol. Lett. 126, 263-270.

Goodrick JC, Xu M, Finnegan R, Schilling BM, Schiavi S, Hoppe H, Wan NC. (2001) High-level expression andstabilization of recombinant human chitinase produced in a continuous constitutive *Pichia pastoris* expression system. Biotechnol. Bioeng. 74, 492-497.

Gueguen P, Chemardin P, Pien S, Arnaud A, Galzy P. (1997) Enhancement of aromatic quality of Muscat wine by the use of immobilized β-glucosidase. J. Biotechnol. 55, 151-156.

Guisán JM, Alvaro G, Fernández-Lafuente R, Rosell CM, Garcia JL, Tagliani A. (1993) Stabilization of heterodimeric enzyme by multipoint covalent immobilization: Penicillin G acylase from *Kluyvera citrophila*. Biotechnol. Bioeng. 42, 455-464.

Guisán JM. (1998) Aldehíde-agarose gels as activated supports for immobilization-stabilization of enzymes. Enz. Microb. Technol. 10, 375-382.

Guisán JM. (2006) Immobilization of enzymes and cells, in Methods in Biotechnology, vol 22. Ed. J.M Guisán. Humana Press Inc. Totowa, New Jersey.

Günata YZ, Bayonove CL, Baumes RL, Cordonnier RE. (1990) Hydrolysis of grape monoterpenyl-β-D-glucosides by various β-glucosidases. J. Agric. Food Chem. 38, 1232-1236.

Guo W, González-Candelas L, Kolattukudy PE. (1996) Identification of a novel pelD gene expressed uniquely in planta by *Fusarium solani* (*Nectria haematococca*, mating type VI) and characterization of its protein product as an endo-pectate lyase. Arch. Biochem. Biophys. 332, 305-312.

Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischnewski H, Roser J, Mitchell T,

Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. Science, 313, 1441-1443.

Han YJ, Stucky GD, Butler A. (1999) Mesoporous silicate sequestration and release of proteins. J. Am. Chem. Soc. 121, 9897–9898.

Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol., 166, 557-580.

Hang HF, Chen W, Guo MJ, Chu J, Zhuang JP, Zhang S. (2008) A simple unstructured model-based control for efficient expression of recombinant porcine insulin precursor by *Pichia pastoris*. Korean J. Chem. Eng. 25, 1065-1069.

Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Luck J, Griengl H, Kohlwein SD, Schwab H. (1997) High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts. Prot. Expr. Purif. 11, 61-71.

He J, Li X, Evans DG, Duan X, Li C. (2000) A new support for the immobilization of penicillin acylase. J. Mol. Cat. B: Enz. 11, 45-53.

Heijden RT, Romein B, Heijnen JJ, Hellinga C, Luyben KC. (1994) Linear constraint relations in biochemical reaction systems. I. Biotechnol. Bioeng. 43, 3–10.

Heimo H, Palmu K, Suominen I. (1997) Expression in *Pichia pastoris* and purification of *Aspergillus awamori* glucoamylase catalytic domain. Prot. Expr. Purif. 10, 70-79.

Hernaiz M, Crout D. (2000) Immobilization/stabilization on Eupergit C of the β-galactosidase forma *Bacillus circulans* and an α-galactosidase from *Aspergillus oryzae.* Enz. Microb Technol. 27, 26-32.

Hernández LF, Espinosa JC, Fernández-González M, Briones A. (2001) β-Glucosidase activity in a *Saccharomyces cerevisiae* wine strain. Int. J. Food Microbiol. 80, 171-176.

Herwig C, von Stockar U. (2002) A small metabolic flux model to identify transient metabolic regulations in *Saccharomyces cerevisiae*. Bioprocess Biosyst. Eng. 24, 395–403.

Heyland J, Fu J, Blank JM, Schmid A. (2011) Carbon metabolism limits recombinant protein production in *Pichia pastoris*. Biotechnol. Bioeng, 108, 1942-1953.

Higgins DR, Cregg JM. (1998) Methods in Molecular Biology, 103. *Pichia* Protocols Ed. Higgins DR, Cregg JM. Humana Press Inc. Totowa, NJ.

Hohenblum H, Gasser B, Maurer M, Borth M, Mattanovich D. (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotech. Bioeng. 85, 367-375.

Holmes WJ, Darby RAJ, Wilks MDS, Smith R, Bil RM. (2009) Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime. Microb. Cell Fact. 8, 35-48.

Hudson S, Magner E, Cooney J, Hodnett BK. (2005) Methodology for the immobilization of enzymes onto mesoporous materials. J. Phys. Chem. B, 109, 19496-19506.

Humphrey HP, Wright PA, Botting NP. (2001) Enzyme immobilisation using siliceous mesoporous molecular sieves. Micropor. Mesopor. Mater. 44/45, 763-768.

Inan M, Meagher MM. (2001) The effect of ethanol and acetate on protein expression in *Pichia pastoris*. Biosci Bioeng. 92, 337-341.

Invitrogen Corp. (2000) Version B 053002 Pichia Fermentation Process Guidelines. http://www.invitrogen.com/. Invitrogen Co. San Diego, CA, USA.

Isett K, George H, Herber W, Amanullah A. (2007) Twenty-four-well plate miniature bioreactor high-throughput system: assessment for microbial cultivations. Biotechnol. Bioeng. 98, 1017-1028.

Jahic M, Rotticci-Mulder JC, Martinelle M, Hult K, Enfors SO. (2002) Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess Biosyst. Eng. 24, 385–393.

Jahic M, Wallberg F, Bollok M, Garcia P, Enfors SO. (2003) Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. Microb. Cell Fact. 2, 6-11.

Janbon B, Arnaud A, Galzy P. (1994) Selection and study of a *Candida molischiana* derepressed for β-glucosidase production. FEMS Microbiol. Lett. 118, 207-212.

Janbon G, Magnet R, Arnaud A, Galzy P. (1995) Cloning and sequencing of the β-glucosidase-encoding gene from *Candida molischiana* strain 35M5N. Gene, 165, 109-113.

Jin S, Ye K, Shimizu KJ. (1997) Metabolic flux distributions in recombinant *Saccharomyces cerevisiae* during foreign protein production. J. Biotechnol. 9, 161-174.

Jin YS, Jeffries W. (2004) Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. Metabol. Eng. 6, 229–238.

Juge N, Andersen JS, Tull D, Roepstorff P, Svensson B. (1996) Overexpression, purification, and characterization of recombinant barley α-amylases 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*. Prot. Expr. Purif. 8, 204-214.

Jung D, Streb C, Hartmann M. (2010) Covalent anchoring of chloroperoxidase and glucose oxidase on the mesoporous molecular sieve SBA-15. Int. J. Mol. Sci. 11, 762-778.

Jungo C, Marison I, Stockar U. (2007) Mixed feeds of glycerol and methanol can improve the performance of *Pichia pastoris* cultures: a quantitative study based on concentration gradients in transient continuous cultures. J. Biotechnol. 128, 824–837.

Karagulyan HK, Gasparyan VK, Decker SR. (2008) Immobilization of fungal β-glucosidase on silica gel and kaolin carriers. Appl. Biochem. Biotechnol. 146, 39–47.

Kato S, Ishibashi M, Tatsuda D, Tokunaga H, Tokunaga M. (2001) Efficient expression, purification and characterization of mouse salivary α-amylase secreted from methylotrophic yeast, *Pichia pastoris*. Yeast, 18, 643-655.

Kawai R, Yoshida M, Tani T, Igarashi K, OhiraT, Nagasawa H, Samejima M. (2003) Production and characterization of recombinant *Phanerochaete chrysosporium* β-glucosidase in the methylotrophic yeast *Pichia pastoris*. Biosci. Biotechnol. Biochem. 67, 1-7.

Klamt S, Gagneur J, Kamp A. (2005) Algorithmic approaches for computing elementary modes in large biochemical reaction networks. BMC Syst. Biol. 152, 249–255.

Klamt S, Schuster S, Gilles ED. (2002) Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. Biotechnol. Bioeng. 77, 734–751.

Klamt S, Stelling J, Ginkel M, Gilles ED. (2003) FluxAnalyser: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. Bioinformatics, 19, 261–269.

Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Tomomitsu K. (2000) High level secretion of recombinant human serum albumin by fed-batch fermentation of the methylotrophic yeast, *Pichia pastoris*, based on optimal methanol feeding strategy. J Biosci. Bioeng. 90, 280-288.

Kohchi C, Hayashi M, Nagai S. (1985) Purification and properties of β-glucosidase from *Candida pelliculosa* var *acetaetherius*. J. Agric. Biol. Chem. 49, 779-784.

Kosary J, Stefanovits-Banyai E, Boross L. (1998) Reverse hydrolytic process for O-alkylation of glucose catalysed by immobilized α- and β-glucosidases. J. Biotechnol. 66, 83-86.

Kupcsulik B, Sevella B. (2005) Optimization of specific product formation rate by statistical and formal kinetic model descriptions of an HSA producing *Pichia pastoris* Mut^S strain. Chem. Biochem. Eng. 19, 99-108.

Lavermicocca P, Valerio F, Evidente A, Lazzaroni S, Corsetti A, Gobetti M. (2000) Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. Appl. Environ. Microbiol. 66, 4084–4090.

Lei C, Shin Y, Liu J, Ackerman EJ. (2002) Entrapping enzymes in a functionalized nanoporous support. J. Am. Chem. Soc. 124, 38, 11242–11243.

Lei F, Rotbøll M, Jørgensen SB. (2001) A biochemically structured model for *Saccharomyces cerevisiae*. J. Biotechnol. 88, 205-221.

Lewis ME. (2011) Correction factors for oxygen solubility and salinity U.S. Geological Survey TWRI Book 9. National Field Manual for the collection of water-quality data.

Liu B, Hu R, Deng J. (1997) Characterization of immobilization of an enzyme in a modified Y zeolite matrix and its application to an amperometric glucose biosensor. Anal. Chem. 69, 2343-2348.

Llaneras F (2010) Interval and possibilistic methods for constraint-based metabolic models. PhD Thesis. Universidad Politécnica de Valencia. Valencia (Spain).

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

Llaneras F, Picó J. (2007a) A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC Bioinform. 8, 421-445.

Llaneras F, Picó J. (2007b) An interval approach for dealing with flux distributions and elementary modes activity patterns. J. Theor. Biol. 246, 290-308.

Llaneras F, Picó J. (2008) Stoichiometric modelling of the cell metabolism. J. Biosci. Bioeng. 1, 1-12.

Llaneras F, Picó J. (2010) Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. J. Biomed. Biotechnol. 1, ID753904, 13 pg.

Llaneras F, Sala A, Picó J. (2009) A possibilistic framework for metabolic flux analysis. BMC Syst. Biol. 3, 79-100.

Lofberg J. (2004) YALMIP: a toolbox for modeling and optimization in MATLAB. Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design. pp. 284-289.

López-Gallego F, Betancor L, Mateo C, Hidalgo A, Alonso-Morales N, Dellamora-Ortiz G, Guisán JM, Fernández-Lafuente R. (2005) Enzyme stabilization by glutaraldehyde crosslinking of adsorbed proteins on aminated supports. J. Biotechnol,. 119, 70-75.

Mahadevan R, Burgard A, Famili I, Van Dien S, Schilling C. (2005) Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals. Biotechnol. Bioproc. Eng. 10, 408-417.

Malissard M, Zeng S, Berger E. (2000) Expression of functional suluble forms of human β-1,4-galactosyltransferase I, α-2,6 sialyltransferase and α-1,3-fucosyltransferase VI in the methylotrophic yeast *Pichia pastoris.* Biochem. Biophys. Res. Comm. 267, 169-173.

Manzanares P, Orejas M, Gil JV, de Graaff L, Visser J, Ramón D. (2003) Construction of a genetically modified wine yeast strain expressing the *Aspergillus aculeatus* rhaA gene, encoding an α-l-rhamnosidase of enological interest. Appl. Environ. Microbiol. 69, 7558-7562.

Maras M, Callewaert N, Piens K, Claeyssens M, Martinet W, Dewaele S, Contreras H, Dewerte I, Penttila M, Contreras R. (2000) Molecular cloning and enzymatic characterization of a *Trichoderma reesei* 1,2-α-D-mannosidase. J. Biotechnol. 77, 255-263.

Margolles-Clark E, Tenenen M, Nakari-Setäla N, Penttila M. (1996) Cloning of genes encoding α-L-arabinofuranosidase and β-xylosidase from *Trichoderma reesei* by expression in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 62, 3840-3846.

Martino A, Durante M, Pifferi PG, Spagna G, Bianchi G. (2003a) Immobilization of β-glucosidase from a commercial preparation, Part 1. A comparative study of natural supports Proc. Biochem. 31, 281-285.

Martino A, Pifferi PG, Spagna G. (2003b) Immobilization of β-glucosidase from a commercial preparation. Part 2. Optimization of the immobilization process on chitosan. Proc. Biochem. 31, 287-293.

Matthijs G, Schacht E. (1996) Comparative study of methodologies for obtaining β-glucosidase immobilized on dextran-modified silica. Enz. Microb. Technol. 19, 601-605.

Maurer M, Kühleitner M, Gasser B and Mattanovich D. (2006) Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with *Pichia pastoris.* Microb. Cell Fact. 5, 37-47.

Mayer J, Kranz B, Fischer L. (2010) Continuous production of lactulose by immobilized thermostable β-glycosidase from *Pyrococcus furiosus*. J. Biotechnol. 145, 387–393.

Mayson BE, Kilburn DG, Zamost BL, Raymond CK, Lesnick GJ. (2003) Effects of methanol concentration on expression levels of recombinant protein in fed-batch cultures of Pichia methanolica. Biotechnol. Bioeng. 81, 291-298.

McCleary BV, Harrington J. (1988) Purification of β-D-glucosidase from *Aspergillus niger*. Meth. Enzymol. 160, 575-583.

Mendoza-Muñoz DF, Barrera LA, Algecira Enciso NA, Cordoba HA. (2008) A simple structured model for recombinant IDShr protein production in *Pichia pastoris*, Biotechnol. Lett. 30, 1727-1734.

Meselson M, Yuan R. (1968) DNA restriction enzyme from *E. coli*. Nature 217, 1110-1114.

Michlmayr H, Schümann C, da Silva NM, Kulbe KD, del Hierro AM. (2010) Isolation and basic characterization of a β-glucosidase from a strain of *Lactobacillus brevis* isolated from a malolactic starter culture. J. Appl. Microbiol., 108, 550-559.

Miyanaga A, Koseki T, Matsuzawa H, Wakagi T, Shoun H, Fushinobu S. (2004) Expression, purification, crystallization and preliminary X-ray analysis of α-L-arabinofuranosidase B from *Aspergillus kawachii*, Acta Crystallography Section D, D60, 1286-1288.

Mladenoska I, Grey CE, Winkelhausen E, Kuzmanova S, Adlercreutz P. (2007) Competition between transglycosylation and hydrolysis in almond β-glucosidase-catalysed conversion of p-nitrophenyl-β-D-glucoside in monophasic water/alcohol mixtures. Biocatal. Biotransform. 25, 382-385.

Mladenoska I, Winkelhausen E, Kuzmanova S. (2008) Transgalactosylation/hydrolysis ratios of various β-galactosidases catalyzing alkyl-β-galactoside synthesis in single-phased alcohol media. Food Technol. Biotechnol. 46, 311–316.

Mo ML, Palsson BO, Herrgard MJ. (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst. Biol. 3, 37-53.

Montesino R, Garcia R, Quintero O, Cremata JA. (1998) Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. Protein Expr. Purif. 14, 197-207.

Montesino R, Nimtz M, Quintero O. (1999) Characterization of the oligosaccharides assembled on the *Pichia pastoris* expressed recombinant aspartic protease. Glycobiol. 9, 1037–1043.

Morawski B, Lin Z, Cirino P, Joo H, Bandara G, Arnold FH. (2000) Functional expression of horse radish peroxidase in *Pichia pastoris* and *Saccharomyces cerevisiae*. Prot. Eng. 13, 377-384.

Morton CL, Potter PM. (2000) Comparison of *Escherichia coli*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Spodoptera frugiperda*, and COS7 cells for recombinant gene expression. Application to a rabbit liver carboxylesterase. Mol Biotechnol. 16, 193-202.

Muslin E, Kanikula A, Clark S, Henson H. (2000) Overexpression, purification and characterization of a barley α-glucosidase secreted by *Pichia pastoris*. Prot. Expr. Purif. 18, 20-26.

Nadri M, Trezzani I, Hammouri H, Dhurjati P, Longin R, Lieto J. (2006) Modeling and observer design for recombinant *Escherichia coli* strain. Bioproc. Biosyst. Eng. 28, 217-225.

Nagatomo H, Matsushita YI, Sugamoto K, Matsui T. (2005) Preparation and properties of gelatin-immobilized β-glucosidase from *Pyrococcus furiosus*, Biosci. Biotechnol. Biochem. 69, 128-136.

Navarro, J. J. Picó, J. Bruno, E. Picó-Marco and S. Vallés. (2001) On-line method and equipment for detecting, determining the evolution and quantifying a microbial biomass and other substances that absorb light along the spectrum during the development of biotechnological processes. Patent ES20010001757, EP20020751179.

Nelson DL, Cox MM. (2000) Lehninger Principles of Biochemistry, 3rd Ed, Worth Publishers, USA.

Nielsen J, Villadsen J, Liden G. (2003) Bioreaction engineering principles. New York: Kluwer Academic Pub.

Nielsen J, Villadsen J. (1992) Modelling of microbial kinetics. Chem. Eng. Sci. 47, 4225–4270.

Nogales J, Palsson BO, Thiele I. (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst. Biol. 2, 79-98.

Numan M, Bhosle N. (2005) α-L-arabinofuranosidases: the potential applications in biotechnology. J. Ind. Microbiol. Biotechnol. 33, 247-260.

Ohya T, Ohyama M, Kobayashi K. (2005) Optimization of human serum albumin production in methylotrophic yeast *Pichia pastoris* by repeated fed-batch fermentation. Biotechnol. Bioeng. 90, 876-887.

Oliveira R, Clemente JJ, Cunha AE, Carrondo MJT. (2005) Adaptive dissolved oxygen control through the glycerol feeding in a recombinant *Pichia pastoris* cultivation in conditions of oxygen transfer limitation. J. Biotechnol. 116, 35–50.

Ortiz J. (2008) Influencia de la adición de glicosidasas sobre el potencial aromático del tomate (*Solanum lycopersicum* L.) PhD Thesis. Universitat de València.

Özkan P, Sariyar B, Utkur F, Akman U, Hortacsu A. (2005) Metabolic flux analysis of recombinant protein overproduction in *Escherichia coli*. J. Biochem. Eng. 22, 167–195.

Pais JM, Varas L, Valdés J, Cabello C, Rodríguez L, Mansur M. (2003) Source Modeling of mini-proinsulin production in *Pichia pastoris* using the AOX promoter. Biotechnol Lett. 25, 251-255.

Palomares O, Villalba M, Rodriguez R. (2003) The c-terminal segment of the 1,3-glucanaseole E9 from olive (*olea europaea*) pollen is an independent domain with allergenic activity: expression in *Pichia pastoris* and characterization. J. Biochem. 369, 593-601.

Palsson BO. (2000) The challenges of *in silico* biology. Nature Biotechnol. 18, 1147–1150.

Palsson BO. (2006) Systems biology: properties of reconstructed networks. New York, USA: Cambridge University Press New York

Pandya P, Jasra RV, Newalkar BL and Bhatt PN. (2005) Studies on the activity and stability of immobilized α-amylase in ordered mesoporous silicas. Micropor. Mesopor. Mat. 77, 67-77.

Park DW, Kim HS, Jung JK, Haam S, Kim WS. (2000) Enzymatic synthesis of alkylglucosides by amphiphilic phase enzyme reaction. Biotechnol. Lett. 22, 951–956.

Pérez-Cabero M, Hungría AB, Morales JM, Tortajada M, Ramón D, Moragues A, El Haskouri J, Beltrán D, Amorós P (2011) Interconnected mesopores and high accessibility in UVM-7-like silicas. 5th International FEZA Conference. ISBN 978-84-8363-722-7. Valencia, Spain.

Pérez-González JA, González R, Querol A, Sendra J, Ramón D. (1993) Construction of a recombinant wine yeast strain expressing β-(1,4)-endoglucanase and its use in microvinification processes. Appl. Environ. Microbiol. 59, 2801-2806.

Pessela BC, Mateo C, Fuentes M, Vián A, García JL, Carrascosa AV, Guisán JM, Fernández-Lafuente R. (2003) The immobilization of a thermophilic ß-galactosidase on Sepabeads supports decreases product inhibition: complete hydrolysis of lactose in dairy products. Enz. Microb. Technol. 33, 199-205.

Pfeiffer T, Sánchez-Valdenebro I, Nuno JC, Montero F, Schuster S. (1999) METATOOL: For studying metabolic networks. Bioinformatics 15, 251–257.

Picó-Marco E. (2004) Nonlinear robust control of biotechnological processes. PhD Thesis. Universidad Politécnica de Valencia, Valencia (Spain),

Pitkänen JP, Aristidou A, Salusjärvi L, Ruohonen L, Penttilä M. (2003) Metabolic flux analysis of xylose metabolism in recombinant *Saccharomyces cerevisiae* using continuous culture. Metabol. Eng. 5, 16-31.

Plantz BA, Sinha J, Villarete L, Nickerson KW, Schlegel VL. (2005) *Pichia pastoris* fermentation optimization: energy state and testing a growth-associated model. Appl. Microbiol. Biotechnol. 72, 297-305.

Porro D, Sauer M, Branduardi P, Mattanovich D. (2005) Recombinant protein production in yeasts. Mol. Biotechnol. 31, 245-259.

Potgieter TI, Kersey SD, Mallem MR, Nylen AC, d'Anjou M. (2010) Antibody expression kinetics in glycoengineered *Pichia pastoris*. Biotechnol. Bioeng. 106, 918-927.

Prade H, Mackenzie LF, Withers SG. (1998) Enzymatic synthesis of disaccharides using *Agrobacterium* sp. β-glucosidase. Carbohyd. Res. 305, 371-381.

Provost A, Bastin G. (2004) Dynamic metabolic modelling under the balanced growth condition. J. Proc. Cont. 14, 717-728.

Querol A, Barrio E, Ramón D. (1992) A comparative study of different methods of yeast strain characterization, Syst. Appl. Microbiol. 15, 439-446.

Ramón D, González R, Pérez-González JA, Querol A. (1993) Levadura vínica CECT1973, su método de obtención por técnicas de DNA recombinante y su aplicación como levadura vínica de uso industrial, útil para mejorar el aroma de los vinos. Patente de invención P9300789.

Rantwick F, Woudenberg-van Oosterom M, Sheldon RA. (1998) Glycosidase-catalysed synthesis of alkyl glycosides. J. Mol. Cat. B Enzym. 6, 511-532.

Ravet C, Thomas D, Legoy MD. (1993) Gluco-oligosaccharide synthesis by free and immobilized β-glucosidase. Biotechnol. Bioeng. 42, 303–308.

Ravindra R, Zhao S, Gies H and Winter R,. (2004) Protein encapsulation in mesoporous silicate: the effects of confinement on protein stability, hydration and volumetric properties. J. Am. Chem. Soc. 126, 12224-12225.

Reed JL, Vo TD, Schilling CH, Palsson BO. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR) Genome Biol. 4, R54.

Rémond C, Ferchichi M, Aubry N, Plantier-Royon R, Portella C, Donohue MJ. (2002) Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α-l-arabinofuranosidase. Tetrahedron Lett. 43, 9653-9655.

Ren HT, Yuan JQ, Bellgardt KH. (2003) Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance. J. Biotechnol. 106, 53–68.

Riccio P, Rossano R, Vinella M, Domizio P, Zito F, Sansevrino F, D'Elia A, Rosi I,. (1999) Extraction and immobilization in one step of two β-glucosidases released from a yeast strain of *Debaromyces hansenii*. Enz. Microb. Technol. 24, 123-129.

Rizzi M, Baltes M, Theobald U, Reuss M. (1997) *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*. II. Mathematical model. Biotechnol. Bioeng. 55, 592–608.

Sainz J, Pizarro F, Pérez-Correa JR, Agosin E. (2003) Modeling of yeast metabolism and process dynamics in batch fermentation. Biotechnol. Bioeng. 81, 818–828.

Sambrook J, Russell DW. (2001) Molecular cloning - a laboratory manual, 3rd ed. Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sánchez-Torres P. (1996) Construcción de levaduras y hongos recombinantes que expresen enzymas hidrolíticas implicadas en la liberación de aromas del vino. PhD Thesis. Universidad de Valencia, Valencia, Spain.

Sánchez-Torres P, González-Candelas L, Ramón D. (1996) Expression in a wine yeast strain of the *Aspergillus níger abfB* gene. FEMS Microbiol. Lett. 145, 189-194.

Sánchez-Torres P, González-Candelas L, Ramón D. (1998) Heterologous expression of a *Candida molischiana* anthocyanin β-glucosidase in a wine yeast strain. FEMS Microbiol. Lett. 46, 354-360.

Santos S. (2008) Análisis cuantitativo y modelización del metabolismo de la levadura *Pichia pastoris*. PhD Thesis. Universitat Autónoma de Barcelona, Bellaterra (Spain).

Sardar M, Roy I, Gupta M. (2000) Simultaneous purification and immobilization of *Aspergillus niger* xylanase on the reversibly soluble polymer Eudragit L-100. Enz. Microb. Technol. 27, 672-679.

Sarry JE, Gunata Z. (2004) Plant and microbial glycoside hydrolases: volatile release from glycosidic aroma precursors. Food Chem. 87, 509-521.

Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wüthrich K. (1997) Metabolic fluxes in riboflavin-producing *Bacillus subtilis.* Nature Biotechnol. 15, 448-452.

Schilling BM, Goodrick JC, Wan NC. (2001) Scale-up of a high cell density continuous culture with *Pichia pastoris* X-33 for the constitutive expression of rh-chitinase. Biotechnol. Prog. 17, 629-633.

Schmid JW, Mauch K, Reuss M, Gilles ED, Kremling A. (2004) Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*. Metabol. Eng. 4, 364-377.

Schmidt K, Nørregaard LC, Pedersen B, Meissner A, Duus JO, Nielsen JO, Villadsen J. (1999) Quantification of intracellular metabolic fluxes from fractional enrichment and ^{13}C-^{13}C coupling constraints on the isotopomer distribution in labeled biomass components. Metabol. Eng. 1, 166–179.

Schuetz R, Kuepfer L, Sauer U. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol. Syst. Biol. 3,119-134.

Schuleit M, Luisi PL. (2003) Enzyme immobilization on silica hardened organogels. Biotechnol. Bioeng. 72, 249-251.

Schuster S, Fell DA, Dandekar T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnol. 18, 326–332.

Scorer CA, Buckholz RG, Clare JJ, Romanos MA. (1993) The intracellular production and secretion of HIV-1 envelop protein in the methylotrophic yeast *Pichia pastoris*. Gene 136, 111–119.

Scouten WH, Luong JHT, Brown RS. (1995) Enzyme or protein immobilization techniques for applications in biosensor design. Trends Biotechnol. 13, 178-185.

Sharma R, Katoch M, Srivastava PS, Qazi GN (2009) Approaches for refining heterologous protein production in filamentous fungi. World J Microbiol. Biotechnol. 25, 2083-2094.

Shimizu H. (2002) Metabolic engineering-integrating methodologies of molecular breeding and bioprocess systems engineering. J. Biosci. Bioeng. 94, 563-573.

Singh RK, Zhang YW, Nguyen NPT, Jeya M, Lee JK. (2011) Covalent immobilization of β-1,4-glucosidase from *Agaricus arvensis* onto functionalized silicon oxide nanoparticles. Appl. Microbiol. Biotechnol. 89, 337-344.

Singhania RR. (2009) Cellulolytic Enzymes. In "Biotechnology for agro-industrial residues utilisation". (Singhnee' Nigarn P, Pandey A eds.), Ed. Springer Science, Business Media BV, Part IV, pp. 371-381.

Sinha J, Plantz BA, Zhang W, Gouthro M, Schlegel VL, Liu CP Meagher MM. (2003) Improved production of recombinant ovine interferon-t by Mut+ strain of *Pichia pastoris* using an optimized methanol feed profile. Biotechnol Prog. 19, 794-802.

Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY. (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. J. Biotechnol. 5, 705-715.

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. Microbiol. 153, 281-290.

Solà A, Maaheimo H, Ylönen K, Ferrer P, Szyperski T. (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. Eur. J. Biochem. 271, 2462-2470.

Solà A. (2004) Estudi del metabolisme central del carboni de Pichia pastoris. PhD Thesis. Escola Tècnica Superior d'Enginyeria de la Universitat Autònoma de Barceloana. Bellatera (Spain)

Spagna G, Andreani F, Salatelli E, Romagnoli D, Casarini D, Pifferi PG. (2003) Immobilization of the glycosidases α-L-arabinofuranosidase and β-D-glucopyranosidase from *Aspergillus niger* on a chitosan derivative to increase the aroma of wine. Enz. Microb. Technol. 23, 413-421.

Spagna G, Barbagallo R, Casarini D, Pifferi PG. (2000a) A novel chitosan derivative to immobilize α-L-rhamnopyranosidase from *Aspergillus niger* for application in beverage technologies. Enz. Microb. Technol. 28, 427-438.

Spagna G, Barbagallo R, Pifferi PG, Blanco RM, Guisán JM. (2000b) Stabilization of a β-glucosidase from *Aspergillus niger* by binding to an amine agarose gel. J. Mol. Cat. B: Enz. 11, 63-69.

Spagna G, Romagnoli D, Angela M, Bianchi G, Pifferi PG. (1998) A simple method for purifying glycosidases: α-L-arabinofuranosidase and β-D-glucopyranosidase form

Aspergillus niger to increase the aroma of wine. Part 1. Enz. Microb. Technol. 22, 298-304.

Stephanopoulos GN. (1998) Metabolic engineering: principles and Methodologies. (Stephanopoulos GN, Aristidou AA eds.), Ed. Academic Press. San Diego (USA)

Stratton J, Chiruvolu V, Meagher M. (1998) High cell-density fermentation. In "Methods in molecular biology: *Pichia* protocols". (Higgins DR, Cregg JM eds), Humana, Totowa, pp. 107– 120.

Subramanian A, Kennel ST, Oden PL, Jacobson KB, Woodward J, Doktycz MJ. (1999) Comparison of techniques for enzyme immobilization on silicon supports. Enz. Microb. Technol. 24, 26-34.

Szyperski T. (1998) ^{13}C-NMR, MS and metabolic flux balancing in biotechnology research. Quarter. Rev. Biophys. 31, 41–106.

Takashaki H, Li B, Sasaki T, Miyazaki C, Kajino T, Inagaki S. (2001) Immobilized enzymes in ordered mesoporous silica materials and improvement of their stability and catalytic activity in an organic solvent. Micropor. Mesopor. Mat. 44-45, 755-762.

Tchopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. (1987) Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris.* Nucl. Acids Res. 15, 3859-3876.

Thiry M, Cingolani D. (2002) Optimizing scale-up fermentation processes. Trends in Biotechnol. 203, 103-105.

Thomas L, Crawford DL. (1998) Cloning of clustered *Streptomyces viridosporus* T7A lignocellulose catabolism genes encoding peroxidase andendoglucanase and their extracellular expression in *Pichia pastoris*. Can. J.Microbiol. 44, 364-372.

Thomson L, Bates S, Yamazaki S, Arisawa M, Aoki Y, Gow N. (2000) Functional characterization of the *Candida albicans MNT1* mannosyltransferase expressed heterologously in *Saccharomyces cerevisiae* and *Pichia pastoris.* Prot. Eng. 13, 377-384.

Tibbot BK, Henson CA. Skadsen RW. (1998) Expression of enzymatically active, recombinant barley α-glucosidase in yeast and immunological detection of α-glucosidase from seed tissue. Plant Mol. Biol. 38, 379-391.

Tortajada M, Ramón D, Beltrán D, Amorós P. (2005) Hierarchical bimodal porous silicas and organosilicas for enzyme immobilization. J. Mater. Chem. 15, 3859-3868.

Tremblay LO, Campbell Dyke N, Herscovics A. (1998) Molecular cloning, chromosomal mapping and tissue-specific expression of a novel human α-1,2-mannosidase gene involved in N-glycan maturation. Glycobiol. 8, 585-595.

Tsai PS, Rao G, Bailey JE. (1995) Improvement of *Escherichia coli* microaerobic oxygen metabolism by itreoscilla hemoglobin: New insights from NAD(P)H fluorescence and culture redox potential. Biotechnol Bioeng. 5, 3, 347-354.

Tschopp JF , Sverlow G, Kosson R, Craig W, Grinna L. (1987) High-level secretion of glycosylated invertase in the methylotrophic yeast, *Pichia pastoris*. Nature Biotechnol. 5, 1305–1308.

Turan Y, Zheng M. (2005) Purification and characterization of an intracellular β-glucosidase from the methylotrophic yeast *Pichia pastoris.* Biochem. (Mosc), 70, 1363-1368.

van den Broek LAM, Ton J, Verdoes JC, Van Laere KMJ, Voragen AGJ, Beldman G. (1999) Synthesis of α-galacto-oligosaccharides by a cloned β-galactosidase from *Bifidobacterium adolescentis*. Biotechnol. Lett. 21, 441–445.

van der Beek G, Southeamer H. (1973) Oxidative phosphorylation in intact bacteria. Archiv Mikrobiol., 89, 327-339.

Vandersall-Nairn AS, Merkle RK, O'Brien K, Oeltmann TN, Morem KW. (1998) Cloning, expression, purification, and characterization ofthe acid α-mannosidase from *Trypanosoma cruzi*. Glycobiol. 8, 1183-1194.

vanGulik WM, Antoniewicz MR, deLaat WTAM, Vinke JL, Heijnen JJ. (2001) Energetics of growth and penicillin production in a high-producing strain of *Penicillium chrysogenum,* Biotechnol. Bioeng. 72, 185-193.

vanGulik WM, Heijnen JJ. (1995) A metabolic network stoichiometry analysis of microbial growth and product formation. Biotechnol. Bioeng. 20, 681-698.

Vanrolleghem PA, de Jong-Gubbels P, van Gulik WM, Pronk JT, van Dijken JP, Heijnen S. (1996) Validation of a metabolic network for *Saccharomyces cerevisiae* using mixed substrate studies. Biotechnol Prog. 12, 434-448.

Varghese JN, McKimm-Breschkin JL, Caldwell JB, Kortt AA, Colman PM. (1992) The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins 14, 327–332.

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

Vasserot Y, Chemardin P, Arnaud A, Galzy. (1991) Purification and properties of the β-glucosidase of a new strain of *Candida molischiana* able to work at low pH values: possible use in the liberation of bound terpenols. J. Basic Microbiol. 31, 301-312.

Vasserot YT, Christiaens H, Chemardin P, Arnaud A, Galzy P. (1989) Purification and properties of a β-glucosidase of *Hanseniospora vineae* with the view to its utilization in fruit aroma liberation. J. Appl. Bacteriol. 66, 271-279.

Veen P, Flipphi MJA, Voragen AGJ, Visser J. (1991) Induction, purification and characterization of arabinases produced by *Aspergillus niger*. Arch. Microbiol. 157, 23-28.

Vincent P, Shareck F, Dupont C, Moroscoli R, Kluepfel D. (1997) New α-L-arabinofuranosidase produced by *Streptomyces lividans*: cloning and DNA sequence of the *abf*B gene and characterization of the enzyme. J. Biochem. 322, 845-852.

Visser D, van der Heijden R, Mauch K, Reuss M, Heijnen S. (2000) Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. Metabol. Eng. 2, 252–275.

Wallecha A, Mishra S. (2003) Purification and characterization of two β-glucosidases from a thermo-tolerant yeast *Pichia etchellsii*. Biochim. Biophys. Acta, 1649, 74-84.

Wan CF, Chen WH, Chen CT, Chang MDT, Lo LC, Li YK,. (2007) Mutagenesis and mechanistic study of a glycoside hydrolase family 54 α-L-arabinofuranosidase from *Trichoderma koningii*. J. Biochem. 401, 551-558.

Wang P, Dai S,Waezsada SD, Tsao AY, Davison BH. (2001) Enzyme stabilization by covalent binding in nanoporous sol-gel glass for nonaqueous biocatalysis. Biotechnol. Bioeng. 74, 249-255.

Wei Y, Dong H, Xu J Feng Q. (2002) Simultaneous Immobilization of horseradish peroxidase and glucose oxidase in mesoporous sol-gel host materials. Chemphyschem, 3, 802-808.

Wyman CE. (1994) Ethanol from lignocellulosic biomass: technology, economics, and opportunities. Biores. Tech. 50, 3-15.

Yan AX, Li XW, Ye Y. (2002) Recent progress on immobilization of enzymes on molecular sieves for reactions in organic solvents. Appl. Biochem. Biotechnol. 101, 113-129.

Yi Q, Sarney DB, Khan JA, Vulfson EN. (1998) A novel approach to biotransformations in aqueous-organic two-phase systems: enzymatic synthesis of alkyl beta-D-glucosides using microencapsulated beta-glucosidase. Biotechnol. Bioeng. 5, 385-390.

Yiu HHP, Wright P, Botting N. (2000) Enzyme immobilization using siliceous mesoporous molecular sieves. Micropor. Mesopor. Mat. 44-45, 763-768.

Yuan JQ, Bellgardt KH. (1994) Investigations on the optimal control of storage stability of compressed baker's yeast *Saccharomyces cerevisiae*. J. Biotechnol. 32, 261–272.

Zhang T, Huang Z, Lin Z. (2008) Hydrolysis of soybean isoflavone by immobilized β-glucosidase in a two phase system. J. Chem. Ind. Eng. 59, 387-392.

Zhang W, Smith LA, Plantz BA, Schlegel VL, Meagher MM. (2002) Design of methanol feed control in *Pichia pastoris* fermentations based upon a growth model. Biotechnol. Prog. 18, 1392-1399.

Zhang WH, Bevins MA, Plantz BA, Smith LA, Meagher MM. (2000) Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of Botulinum neurotoxin, serotype A. Biotechnol. Bioeng. 70, 1-8.

Zhang WH, Liu CP, Inan M, Meagher MM. (2004) Optimization of cell density and dilution rate in *Pichia pastoris* continuous fermentations for production of recombinant proteins. J. Ind. Microbiol. Biotechnol. 31, 330-334.

Zhang WH, Sinha J, Smith L, Inan M, Meagher MM. (2005) Maximization of production of secreted recombinant proteins in *Pichia pastoris* fed-batch fermentation. Biotechnol. Prog. 21, 386-393.

Zhao HL, Xue C, Wang Y, Yao XQ, Liu ZM. (2008a) Increasing the cell viability and heterologous protein expression of *Pichia pastoris* mutant deficient in PMR1 gene by culture condition optimization. Appl. Microbiol. Biotechnol. 81, 235-241.

Zhao YY, Takahashi M, Gu JG, Miyoshi E, Matsumoto A, Kitazume S, Taniguchi N,. (2008b) Functional roles of N-glycans in cell signaling and cell adhesion in cancer. Cancer Sci. 99, 1304-1310.

Zhou XS, Zhang YX. (2002) Decrease of proteolytic degradation of recombinant hirudin produced by *Pichia pastoris* by controlling the specific growth rate. Biotechnol. Lett. 24, 1449-1453.

Zhu A, Monahan C, Zhang Z, Hurst R, Leng L, Goldstein J. (1995) High-level expression and purification of coffee bean α-galactosidase produced in the yeast *Pichia pastoris*. Arch. Biochem. Biophys. 324, 65-70.

Zhu A, Wang ZK, Beavis R. (1998) Structural studies of α-N-acetylgalactosaminidase: effect of glycosylation on the level of expression, secretion efficiency, and enzyme activity. Arch. Biochem. Biophys. 352, 1-8.

Zopf D, Roth S. (1996) Oligosaccharide anti-infective agents. Lancet 347, 1017-1021.