ANA RAQUEL SANTOS FERREIRA

Licenciada em Engenharia Química

A Systems Biology framework for pathway level culture media engineering: application to *Pichia pastoris* cultures

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

Orientador: Rui Oliveira, Doutor, DQ - FCT/UNL

Co-orientador: António Cunha, Engenheiro, IBET - Unidade Piloto

Co-orientador: João Dias, Doutor, DQ - FCT/UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2012

Agradecimentos

A realização desta dissertação não teria sido possível sem a contribuição de várias pessoas e entidades, às quais desejo expressar o meu agradecimento.

Em primeiro lugar ao meu orientador principal Professor Doutor Rui Oliveira quero agradecer a possibilidade que me deu de integrar na sua equipa de Biologia de Sistemas. A sua excelente competência e orientação bem como o seu contínuo estímulo e interesse pelo trabalho que fui efectuando foi um contributo essencial para tornar possível a realização desta Tese.

Ao meu co-orientador Engenheiro António Cunha pelo seu conhecimento, apoio e pelas valiosas sugestões ao longo desta Tese de Doutoramento. Quero também agradecer as condições de trabalho que me proporcionou, nomeadamente, as instalações do laboratório sem as quais esta Tese não teria sido possível ser realizada.

Ao meu co-orientador Doutor João Dias pela sua disponibilidade e ajuda contínua ao longo de várias etapas desta Tese. Quero também expressar o meu agradecimento ao Engenheiro João Clemente pelo seu tempo dispensado e ajuda na implementação e execução do sistema de controlo do bioreactor piloto que utilizei para as minhas experiências.

Um agradecimento especial à Doutora Lídia, pela sua explicação e formação da técnica ELISA para a medição do meu produto (scFv).

A todos os meus colegas da Unidade Piloto do IBET pela sua boa disposição, amizade e convívio. Aos meus colegas do antigo grupo Bioeng, actual grupo Biochemical and Process Engineering (BPEG), e em especial aos meus colegas e amigos do grupo Systems Biology and Engineering (SBE) no qual eu estou inserida.

Às minhas amigas que não são da área de bioquímica e tiveram de ouvir os meus desabafos sem saberem do que eu estava realmente a falar.

Agradeço também o apoio financeiro, tal como a concessão da minha bolsa de doutoramento SFRH/BD/36285/2007, financiado pela Fundação para a Ciência e a Tecnologia sem os quais não teria sido possível a realização desta Tese.

O meu profundo agradecimento aos meus Pais, pela sua compreensão e carinho nos momentos menos bons, por me darem força para seguir sempre em frente e nunca desistir, pela ajuda imprescindível a todos os níveis, monetários e emocionais, o meu Muito Obrigada por tudo.

À minha família por compreender a minha ausência em certos momentos importantes por ter de ficar a trabalhar no laboratório ao fim de semana a cuidar das minhas fermentações.

Por último, mas igualmente importante, ao meu namorado César pela sua presença, apoio e paciência em me ouvir a falar de problemas que sugeriam e pelo constante encorajamento ao longo destes últimos 4 anos para a concretização da minha Tese de doutoramento.

E a todos aqueles que de algum modo contribuíram para o desenvolvimento desta Tese.

A todos, MUITO OBRIGADA.

"Somos aquilo que fazemos consistentemente.
Assim, a excelência não é um acto mas sim um hábito."

[Aristóteles]

Abstract

Culture media (CM) formulations contain hundreds of ingredients in aqueous solutions that may be involved in complex interactions in the same or competing pathways within the cell. This thesis proposes a new methodology for determining the optimal composition of CM that migrates from an empirical to a mechanistic or hybrid mechanistic CM development approach. A framework consisting in the execution of an array of cell cultures, endpoint exometabolomic assays and bioinformatics algorithm were brought together into a platform for CM engineering called Cell Functional Enviromics. This technology consists of a large-scale reverse engineering approach that reconstructs cellular function on the basis of measured dynamic exometabolome data. To support this concept, a computational algorithm, called "envirome-guided Projection to Latent Pathways", was developed. This method yields envirome-wide Functional Enviromics Maps (FEM), with rows representing medium factors, columns representing elementary (orthogonal) cellular functions and color intensity values, the strength of up-/down- regulation of cellular functions by medium factors.

This method was applied to optimize Pichia Trace Metal salts for the yeast *Pichia pastoris* to improve the expression of heterologous proteins. An array of shake flasks experiments of the *P. pastoris* X33 strain were performed and used to build a FEM. Then, optimized CM formulations were calculated targeting predefined single-chain Fragment variable antibody (scFv) production improvements. Experimental validation shows a scFv productivity increase of approximately twofold, in relation to the control BSM recipe proposed by Invitrogen. These results were further validated in 2 L bioreactor experiments. Thereafter, scale-up to 50 L bioreactors was developed a mathematical model for further optimization of BSM salts in experiments of *P. pastoris* GS115. Direct adaptive (DO)-stat feeding controller that maximizes glycerol feeding through the regulation of DO concentration at 5% of saturation was developed and applied to the 50 L bioreactor, with the fully optimized CM composition.

Keywords: *Pichia pastoris*, single-chain Fragment variable antibody, Cell Functional Enviromics, chemically defined medium optimization, process scale-up, adaptive process control.

Resumo

As formulações para o meio de cultura (MC) contêm centenas de ingredientes em soluções aquosas que podem participar em inúmeras interacções complexas no interior da célula. Nesta tese é proposta uma nova metodologia para determinar a formulação ideal de MC que permite utilizar uma abordagem empírica em alternativa a uma abordagem mecanicista. Esta metodologia consiste na execução dum conjunto de experiências de cultura celular, análise exometabolómica e desenvolvimento dum algoritmo, "envirome-guided Projection to Latent Pathways", que foram integrados numa plataforma para desenho de MC, Cell Functional Enviromics. Esta tecnologia permite reconstituir em grande escala e de forma reversa as funções celulares com base na medição dinâmica de dados de exometabolómica. Este método permite construir um Functional Enviromics Maps (FEM), onde as linhas representam factores de meio e as colunas representam funções celulares elementares.

Este método foi aplicado na optimização de sais metálicos para a levedura *Pichia pastoris* X33 com o objectivo de melhorar a expressão de proteínas. Foi realizado um conjunto de ensaios em *shake flasks* cujos resultados foram utilizados para construir o FEM. De seguida, as formulações optimizadas de MC foram calculadas com o objectivo de aumentar a produção de um Fragmento variável de anticorpo de cadeia simples. A validação experimental executada em experiências de bioreactores de 2L demostrou um aumento até ao dobro de produtividade em relação à formulação de BSM proposta pela Invitrogen. Posteriormente, foi desenvolvido um modelo matemático para optimizar os sais BSM utilizando o bioreactor de 50L em experiências de *P. pastoris* GS115. Adicionalmente, foi desenvolvido um controlo adaptativo de alimentação do substrato a partir do oxigénio dissolvido que permitiu maximizar a alimentação de glicerol através da regulação do oxigénio a 5% da concentração de saturação em ar e aplicado ao bioreactor de 50 L com a composição de MC totalmente optimizado.

Palavras-chave: *Pichia pastoris*, Fragmento variável de anticorpo em cadeia simples, Ambientómica funcional, meio quimicamente definido, aumento de escala do processo, Controle de processo adaptável.

Contents

List of figures

Appendix

List of tables

Appendix

List of abbreviations

AOX	Alcohol oxidase enzyme;
AOX1	Alcohol oxidase 1 gene;
AOX2	Alcohol oxidase 2 gene;
ATP	Adenosine triphosphate;
BCA	Bicinchoninic acid;
BSA	Bovine serum albumin;
BSM	Basal salts medium;
Ca	Calcium;
CFE	Cell functional enviromics;
cGMP	current good manufacturing practices;
C_{H1}, C_{H2}, C_{H3}	Constant domains of Heavy chain;
C_L	Constant domain of Light chain;
CM	Culture media;
CO_2	Carbon dioxide;
DCW	Dry cell weight (g/L);
DO	Dissolved oxygen (% of air saturation);
DoE	Design-of-experiments;
ED-B	Extracellular domain B;
EFM	Elementary flux modes;
ELISA	Enzyme-linked immunosorbent assay;
F	Feeding rate (g/h);
F_0	Initial feeding rate (g/h);
Fab	Fragment antigen binding;
Fc	Fragment crystallizing;
FEM	Functional enviromics map;
g	Gram;
GAP	Glyceraldehyde-3-phosphate dehydrogenase;
GB	Glycerol batch;
GC	Gas chromatography;
GFB	Glycerol fed-batch;

Gly	Glycerol;
h	Hour;
HPICE	High performance ion chromatography exclusion;
HPLC	High performance liquid chromatography;
ICP-AES	Inductively coupled plasma-atomic emission spectroscopy;
IgG	Imunoglobulin G;
K	Potassium;
L	Liter;
LC	Liquid chromatography;
m/z	mass-to-charge ratio;
MALDI	Matrix-assisted laser desorption/ionization;
MeOH	Methanol;
MFB	Methanol fed-batch;
mg	Milligram;
Mg	Magnesium;
min	Minute;
mL	Milliliter;
MOS	Metal oxide semiconductor;
MS	Mass spectrometry;
Mut^{-}	Methanol utilization minus phenotype;
Mut^{+}	Methanol utilization plus phenotype;
Mut^{s}	Methanol utilization slow phenotype;
$NADH_2$	Reduced nicotinamide adenine dinucleotide;
NMR	Nuclear magnetic resonance;
O_2	Oxygen;
OD_{600nm}	Optical density at 600 nm;
OTL	Oxygen transfer limitation;
P	Phosphorus;
PBS	Phosphate buffer saline;
PFEM1	*Pichia* functional enviromics medium 1;
PI	Proportional-integral controller;

PID	Proportional-integral-derivative controller;
PLP	Projection to latent pathways;
PLS	Projection to latent structures;
PMS	*Pichia* main salts;
PSM	Protein size marker;
PTM1	*Pichia* trace metal salts supplements;
RI	Refractive index;
rpm	Revolution per minute;
S	Sulfate;
scFv	single-chain fragment variable;
SDS – PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis;
std	Standard;
t	Duration of the cultivation (h);
TCA	Tricarboxylic acid;
TOF	Time-of-flight;
UBICON	Universal BIo-process CONtrol system;
V_H	Variable domain of Heavy chain;
V_L	Variable domain of Light chain;
WCW	Wet cell weight (g/L);
μ	Specific growth rate (h^{-1});
μL	Microliter.

List of publications

This dissertation is based on the following publications:

I) Oliveira R, Dias JML, Ferreira AR. PAT 42751/10. A Functional Enviromics method for cell culture media engineering. 2010. *(Chapter 4).*

II) Ferreira AR, Teixeira AP, Carinhas N, Portela RMC, Isidro IA, von Stosch M, Dias JML, Oliveira R. 2011. Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC Systems Biology 5:181. *(Chapter 3).*

III) Ferreira AR, Ataíde F, von Stosch M, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2012. Application of adaptive DO-stat feeding control to *Pichia pastoris* X33 cultures expressing a single chain antibody fragment (scFv). Bioprocess biosyst Eng. *(Chapter 6).*

IV) Ferreira AR, Marques A, Dias JML, Clemente J, Cunha AE, Oliveira R. Dynamic modeling and optimization of main elements consumption in *Pichia Pastoris* cultures. *(Submitted)*. *(Chapter 5).*

V) Ferreira AR, von Stosch M, Dias JML, Clemente JJ, Cunha AE, Oliveira R. Optimization of scFv expression in a *Pichia pastoris Mut*$^{+}$ strain using hybrid modeling methods. *(Submitted)*. (In this article are described in detail the strategies of glycerol and methanol feeding for all the experiments used in *Chapter 5).*

Poster and Oral presentations

I) Ataíde F, Ferreira AR, Clemente JJ, Cunha AE, Oliveira R. 2008. Adaptive control of glycerol e methanol feeding in recombinant *Pichia pastoris* cultures: Impact on antibody titer. 10th International Chemical Enginering Conference (CHEMPOR), Braga, Portugal, 4-6 September.

II) Ferreira AR, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2009. Improving titer and glycosylation homogeneity of an antibody fragment in a *Pichia pastoris* *Mut*$^+$ process. Mass Spec Europe (MSE), Barcelona, Spain, 5-6 November.

III) Ferreira AR, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2010. Projection to latent pathways: Application to the monitoring of *Pichia pastoris* cultures. 4th Conference on Physiology of yeast and Filamentous Fungi (PYFF4), Rotterdam, Netherlands, 1-4 June.

IV) Ferreira AR, Ataíde F, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2011. Systems Biology and Engineering, Envirome guided metabolic engineering. Encontro de Rede de Química e Tecnologia, REQUIMTE, Fátima, Portugal, 13-14 January.

V) Ferreira AR, Ataíde F, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2011. Industry Liaison II, FuncMed – Functional Enviromics Media. Encontro de Rede de Química e Tecnologia, REQUIMTE, Fátima, Portugal, 13-14 January.

VI) Ferreira AR, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2011. A Systems Biology framework for culture media engineering: application to *Pichia pastoris* cultures. 12th International Conference on Systems Biology (ICSB), Heidelberg/Mannheim, Germany, 28 Agosto - 1 September.

VII) Ferreira AR, Dias JML, Clemente JJ, Cunha AE, Oliveira R. 2011. A Systems Biology Method for Top-Down Gene-to-Environment Culture Media Engineering: Application to *Pichia pastoris* cultures. 11th International Chemical and Biological Engineering Conference (CHEMPOR), Caparica, Portugal, 5-7 September.

CHAPTER 1

Introduction of the yeast *Pichia pastoris*

1.1 *Pichia pastoris* expression system

During the 70s, *Pichia Pastoris* was considered a potential expression system for single-cell protein production due to its ability to utilize methanol as sole carbon source (van Dijken *et al.* 1974). The Philips Petroleum Company was a pioneer in developing culture protocols for this methylotrophic yeast. Since then, *P. pastoris* has become a successful host to express many different heterologous proteins over the past two decades. In 2000, Cereghino and Cregg published a list of heterologous proteins that have been expressed in this methylotrophic yeast.

Pichia pastoris is a unique system for production high levels of functionally active recombinant proteins (Cregg *et al.* 1993) because it enables high expression levels comparable to the *Escherichia coli,* is easy to scale-up and uses inexpensive culture media (CM). On the other hand, since *P. pastoris* is a eukaryotic system and integrates machinery for protein processing and secretion it is capable of performing many post-translational modifications such as proteolytic processing, folding, disulphide bond formation and glycosylation (Cereghino and Cregg 2000). In comparison with other yeasts, *P. pastoris* are not subjected to the extensive hyperglycosylation that commonly occurs in proteins secreted from *Saccharomyces cerevisiae* (8-9 mannose residues for *P. pastoris*, which compares to 50-150 mannose residues for *S. cerevisiae*) (Montesino *et al.* 1998). Other advantage is that *P. pastoris* prefers a respiratory rather than a fermentative growth pathway. Nevertheless, small quantities of byproducts such as ethanol and acetic acid, can quickly build up to toxic levels in large-scale cultivation (high cell density) and can repress heterologous protein production (Cereghino *et al.* 2002). All in all, the high productivity and flexibility of the *Pichia* expression system makes it an ideal expression system for laboratory research as well as for industrial recombinant proteins production.

In Table 1.1 a summary of the main advantages and disadvantages of the microorganism *P. pastoris* are presented (Cregg 1999, Daly and Hearn 2005, Macauley-Patric *et al.* 2005).

Table 1.1 Advantages and disadvantages of *P. pastoris.*

Advantages	Disadvantages
• High yield and productivity; • Strong promoters (AOX1 and GAP); • Stable production strains; • Chemically defined media, inexpensive formulation; • High levels of expression of intracellular and secreted proteins; • Eukaryotic post-translational modifications; • Broad pH range: 3-7; • Preference for respiratory growth rather than fermentative (a major advantage over *S. cerevisiae*); • No hyperglycosylation as in *S. cerevisiae*; • Low purification cost; • No endotoxin production; • Non-pathogenic; • Ability of utilizing methanol as inductor and/or carbon source.	• Potential of proteolysis, non-native glycosylation; • Long time for cell cultivation compared to bacteria; • Risks associated to storage of high volumes of methanol (fire hazard).

The *P. Pastoris* strains more commonly used are X33 (genotype: wild-type) and GS115 (genotype: his4), which are wild type with regard to the alcohol oxidase (AOX) enzyme growing on methanol (Cregg *et al.* 1985). AOX is encoded by two genes in *P. pastoris* namely, AOX1 and AOX2. The AOX1 promoter regulates 90 % of AOX activity in the cell while the AOX2 promoter is responsible for the rest (Koutz *et al.* 1989). The *P. pastoris* expression system uses the methanol induced AOX1 promoter which controls the gene that codes for the expression of alcohol oxidase, the enzyme which catalyzes the first step in the metabolism of methanol. This step involves oxidation of methanol catalyzed, generating hydrogen peroxide and formaldehyde in the process by alcohol oxidase (Couderc and Baratti 1980).

In order to avoid, hydrogen peroxide and formaldehyde toxicity, methanol metabolism take place within a specialized cell organelle called the peroxisome, which sequesters toxic by-products from the rest of the cell (Veenhuis *et al.* 1983). The main advantage of this promoter (AOX) is that one can starve cells of methanol in the first stage of bioreaction to focus on building up cell density. Once the cell density has reached the desired concentration, methanol can be added to induce protein production.

Three different phenotypes of *P. pastoris* are generated, depending on the locus of insertion: methanol utilization plus (Mut^+) that contains both AOX1 and AOX2 genes, methanol utilization slow (Mut^s) that only contain AOX2 gene, and the methanol utilization minus phenotype (Mut^-) result of the disruption of both genes (Cregg and Russell 1998). This last strain is unable to grow on methanol as the only carbon source (Chiruvolu et al. 1997).

Meanwhile, other strong promoters that do not require methanol induction have been discovered. The glyceraldehyde-3-phosphate dehydrogenase (GAP) is a constitutive promoter that enables high level of expression on glucose or glycerol (Delroisse *et al.* 2005, Waterham *et al.* 1997). The use of a constitutive promoter leads to simultaneous biomass and protein production and eliminates the need for induction by a second substrate. As such, the main advantage for the GAP promoter avoiding methanol utilization (Cereghino and Cregg 2000). For this reason the GAP promoter is more suitable for large scale production because of the hazard and cost associated with the storage and delivery of large volumes of methanol are eliminated (Goodrick *et al.* 2001, Khasa *et al.* 2007).

1.2 Protein expression levels

The methylotrophic *P. pastoris* has been shown to be a suitable host for high level expression of various heterologous proteins. This yeast is able to produce large amounts of soluble and active protein into the CM in range of gram/Liter. In Table 1.2 presents different heterologous protein produced by recombinant methylotrophic *P. pastoris* for 1987 to 2012.

Table 1.2 Heterologous protein produced by recombinant methylotrophic *P. pastoris*.

Protein	Mode*	Amount (g/L)	Reference
Hepatitis B surface antigen	I	0.30	Cregg *et al.* 1987
Invertase	S	2.50	Tschopp *et al.* 1987
Bovine lysozyme	S	0.30	Digan *et al.* 1989
Human tumor necrosis (P69)	I	8.00	Sreekrishna *et al.* 1989
Human epidermal growth factor	S	0.50	Siegel *et al.* 1990
Pertactin (P69)	I	3.00	Romanos *et al.* 1991
Aprotinin analog	S	0.80	Vedvick *et al.* 1991
Tetanus toxin fragment C	I	12.00	Clare *et al.* 1991
Kunitz protease inhibitor	S	1.00	Wagner *et al.* 1992
rTAP	S	1.70	Laroche *et al.* 1994
β_2-Glycoprotein I	S	0.30	Katakura *et al.* 1998
Human serum albumin (HSA)	S	1.40	Kobayashi *et al.* 2000
mAb4813	S	0.04	Hellwing *et al.* 2001
scFv	S	0.04	Cunha *et al.* 2004
scFv	S	3.50	Khatri *et al.* 2006
Rhizopus oryzae lipase	S	0.50	Cos *et al.* 2006
angiostatin (AS)	S	0.15	Zhang *et al.* 2009
recombinant human growth hormone (rhGH)	S	0.27	Calik *et al.* 2010
recombinant human erythropoietin (rHuEPO)	S	0.16	Soyaslan and Calik 2011
Limit dextrinase inhibitor (LDI)	S	0.20	Jensen *et al.* 2011
Trichoderma reesei cellobiohydrolase 2 (TrCBH2)	S	6.55	Mellitzer *et al.* 2012
Human serum albumin (HSA)	S	18.00	Plank *et al.* 2012

* I = Intracellular; S = Secreted

Regarding to *P. pastoris* protein expression kinetics they may vary with the expression vector and also with the product itself. Table 1.3 shows some examples of cell growth associated, partially cell growth associated and cell growth dissociated expression kinetics reported in the literature.

Table 1.3 List of examples for cell growth associated, partially cell growth associated and cell growth dissociated expression kinetics reported in the literature for *P. pastoris*.

Product		
Growth not associated		
BoNT/A(Hc)	GS115 (*Mut*$^+$)	(Zhang *et al.* 2000)
r-oIFN-τ	GS115 (*Mut*$^+$)	(Sinha *et al.* 2003)
R-galactosidase	GS115 (*Mut*$^+$)	(Zhang *et al.* 2005)
scFv	GS115 (*Mut*$^+$)	(Cunha *et al.* 2004)
Antibody	YGLY4140	(Potgieter *et al.* 2010)
Partially growth associated		
rHSA	HA2 (from GS115)	(Kobayashi *et al.* 2000)
Growth associated		
human trypsinogen	X33	(Hohenblum *et al.* 2003)
MPI	GS115 (*Mut*^{S})	(Pais *et al.* 2003)
rHSA	HA2 (from GS115)	(Ohya *et al.* 2005)
r-oIFN-τ	X33 (from GS115)	(Plantz *et al.* 2006)
recGAvi	GS115 (*Mut*$^+$)	(Jungo *et al.* 2006)
hGM-CSF	GS115 (*Mut*$^+$)	(Pal *et al.* 2006)

1.3 Protein of interest

It is of paramount importance to find antibodies that can identify and accumulate into tumor tissues and that can be eliminated from the blood stream, without causing toxicity into healthy tissues (Berndorff 2005). Such antibodies can block tumor growth or mediate the killing of tumor cell by several different mechanisms (Hoogenboom 2002, Houghton and Scheinberg 2000).

In the present thesis the protein produced by methylotrophic *P. pastoris* is a single-chain Fragment variable (scFv) antibody against extracellular domain B (anti-ED-B) of Fibronectin. The extracellular domain B (ED-B) of Fibronectin is associated with tissue remodeling during a wound healing and with neovasculature solid tumor (such as invasive ductal carcinoma, aggressive brain tumor and ocular angiogenesis) (Berndorff *et al.* 2006). This protein is a promising marker for angiogenesis in growing solid tumors, as result, it is very important to develop a target to ED-B fibronectin.

An antibody consists of two identical light (L) chains and two identical heavy (H) chains. Each light chain is bound to a heavy chain by a dissulfide bond and a combination of noncovalent interactions such as salt bridges, hydrogen bonds and hydrophobic interaction. The amino-terminal regions of light and heavy chains, which vary greatly among antibodies with different specificities, are called variable (V) regions, V_L for light chain and V_H for heavy chain. The regions of relatively constant sequences beyond the variable regions are called constant (C) regions, C_L for light chain and C_H for heavy chain (Kunkel 1998).

Schematic representation of one Imunoglobulina G (IgG) with domain structure is represented in Figure 1.1a. The recombinant scFv antibody produced by *P. pastoris* consists of a V_H and V_L variable regions connected by a short peptide linker (Berndorff 2005) represented in Figure 1.1b.

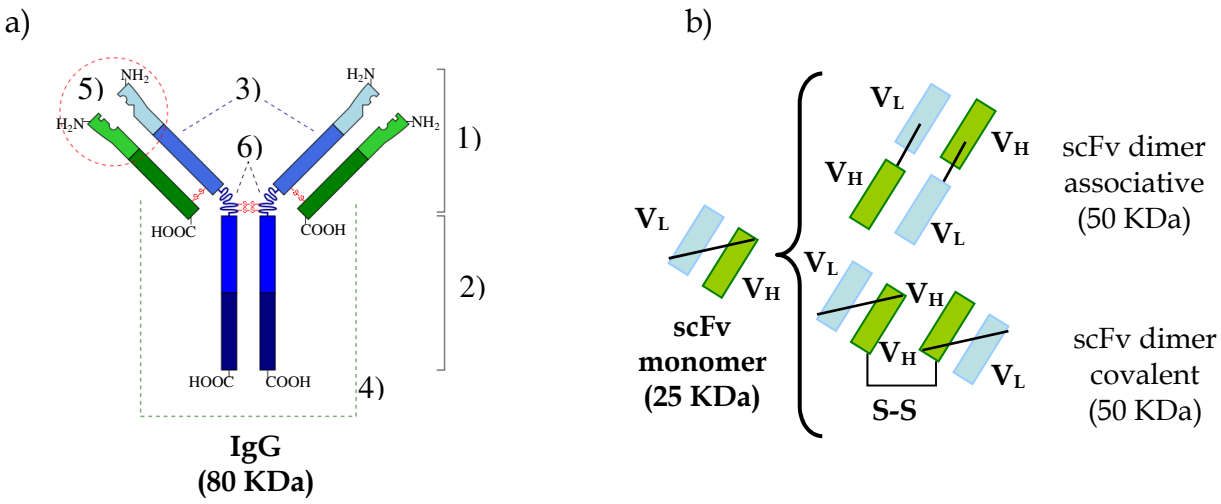


Figure 1.1 (a) Schematic representation of IgG showing the domain structure. 1) Fab; 2) Fc; 3) Light chain, consisting of V_L and C_L regions; 4) Heavy chain, consisting of V_H, C_{H1}, C_{H2} and C_{H3} regions; 5) antigen binding site, scFv; 6) hinge regions. (b) Schematic representation of scFv.

The scFv is found in three different forms: covalent homodimer, associative homodimer and monomer. Both homodimeric forms can be converted to the monomeric form, under reducing conditions and be efficiently radiolabeled, unlike the monomeric form of scFv which cannot be converted to homodimer form due to its different folding structure (Cunha *et al.* 2004).

Among the different possible antibody structures, scFv are predicted to be more effective at penetrating solid tumors (Robinson *et al.* 2004) because they are smaller molecules that still retain binding properties. Due to their size (Marty *et al.* 2001), they penetrate faster and deeper into tumor tissues. Consequently scFv have a reduced tumor retention time and they also are cleared more rapidly from the blood. This evidence have been demonstrated by fluorescent techniques have been shown that anti-ED-B scFv accumulate more selectively around tumoral blood vessels (Neri *et al.* 1997).

1.4 Strategies to improve expression

The production of recombinant protein in *P. pastoris* can be carried out under different operation conditions (Rosenfeld 1999). Many feeding strategies of one or more substrate, dissolved oxygen (DO) controller or integration of methanol sensor have been implemented in order to achieve high titer and high and reproducible product quality.

Empirical methanol feeding strategies are commonly used by many researchers. One strategy is based on the DO spike (Jimenez *et al.* 1997, Lim *et al.* 2003), another uses preprogrammed linear feed rates that are designed to maintain very low methanol concentrations in the growth medium (Brierley *et al.* 1994, *Pichia* fermentation process guidelines 2000). One more approach to increase the productivity is the use of an exponential feeding with mixed glycerol/methanol substrate (d'Anjou and Daugulis 2001, Files *et al.* 2001, Zhang *et al.* 2003). A proportional-integral-derivative (PID) controller of DO was also implemented by manipulating the methanol feeding rate (Chung 2000, Oliveira *et al.* 2005, Zhang *et al.* 2002). Alternatively, methanol concentration in bioreactor can be kept constant by closed-loop control using a methanol sensor and a feeding controller (Guarna *et al.* 1997, Katakura *et al.* 1998, Zhang *et al.* 2000).

Alternative, to obtain large amounts of the target protein is to optimize the composition of CM. Medium desired can be optimized by adjusting the concentrations of salts, vitamins and trace elements (Boze *et al.* 2001, Oehler *et al.* 1998). Current methodologies to optimize CM formulations are based on statistical design-of-experiments (DoE) (Ghosalkar *et al.* 2008, Mandenius and Brundin 2008, Ooijkaas *et al.* 1999). The most common DoE method is called reduced factorial design with two levels (Montgomery 2005), which allows evaluating more than one independent variable at the same time. However this methodology is time-consuming and costly. For this reason a new methodology to optimize cell CM composition based on Cell Functional Enviromics was applied in this thesis.

This method allows building a Functional Enviromics Map (FEM) through the execution of several cell culture experiments using preferably high throughput analytical methods (MS, NMR). This FEM enables to understand how medium factors control the relative intensity of cell functions.

The main objective of this thesis is to combine some aforementioned strategies in order to optimize scFv production by *P. pastoris*. For this purpose, first the composition of CM was optimized and second the most important cultivation parameters, such as pH, temperature, glycerol feeding strategy and methanol feeding strategy, application of the methanol sensor and DO control design were investigated.

1.5 Motivation for this Thesis

P. pastoris is currently an industrial workhorse for the expression of heterologous proteins. It is to data the unique expression system with the capability to express all types of proteins, from low added value catalytic proteins to full length monoclonal antibodies with human glycosylation. Cell culture technologies suffer from several drawbacks such as the inefficiency of therapeutic proteins development (long development time for new products) and inefficient manufacturing (low yield and low productivity). These two problems have major implications, namely very high production costs and high requirements of expensive current Good Manufacturing Practices (cGMP) compliant cultivation capacity. With the introduction of the biosimilars business, the optimization of cultivation conditions is an extremely important vehicle to ensure competitiveness. The development of specialized cell CM formulations, which enable high productivity gains of the right quality therapeutic proteins, is thus an invaluable tool to enforce that path.

Cell CM is comprised of over 50 active ingredients that can have strong interactions making the ratios of component concentrations to each other a critical aspect of optimization.

One of the concerns for the development of new cell CM is the time investment required to optimize a new CM formulation. It is not atypical for the development of a new CM to take between 6 to 12 months of resource intensive laboratorial work. A very high number of CM formulations need to be screened in cultivations for differences in protein expression levels and protein quality.

The current technologies, such as DoE, are eminently empirical and very inefficient. For this reason, a new approach for CM development called Cell Functional Enviromics was developed in this thesis. This methodology may be defined as

"*cell functional enviromics is the systematic identification of the effect of the entirety of medium factors on the entirety of cellular functions.*"

The main difference of this methodology in relation to others is that mechanistic knowledge is extracted at each screening step and saved in a database, this has the unique capability of engineering cellular metabolism towards a desired state. This knowledge, combined with that gained from previous screening steps, allows to understand the metabolic bottlenecks of the cells and to formulate improved media that delivers higher protein productivity. This approach enables to optimize CM to reach higher yields much faster and better than by conventional DoE approaches. The proof-of-concept of the functional enviromics method was achieved in this work by the application to the previously described production of the scFv antibody by *P. pastoris*. In summary, the main objectives were the following:

1) To develop a functional enviromics methodology for rational CM engineering;
2) To develop new chemically defined CM formulations with increased productivity of scFv;
3) To validate the optimized CM formulations in bioreactor experiments up to 50 L scale;

4) To optimize complementary cultivation parameters such as pH, temperature and substrate feeding strategies targeting a highly efficient scFv production process by *P. pastoris*.

1.6 Thesis outline

The present PhD thesis has been structured in the following 7 chapters:

Chapter 1 (current chapter) frames the current thesis in the current state of the art. It starts by providing an introduction to the methylotrophic yeast *P. pastoris* and the expression systems used (X33 and GS115), reviewing the main advantages and disadvantages of this microorganism and some examples of heterologous protein produced by this recombinant yeast. It also includes a brief description of the protein of interest, the single-chain fragment variable antibody produce by *P. pastoris,* and the main strategies to improve the production of this protein. Finally, the motivation, objectives and structure of the thesis are outlined.

Chapter 2 provides detailed information about the materials and methods employed in this thesis. A specification of the BSM medium (*Pichia* fermentation process guidelines proposed by Invitrogen, 2000) is provided, which has been used in many experiments for all growth steps (cell stock, pre-inoculum, inoculum and bioreactor cultivation). Also the incubation conditions employed, such as temperature, agitation rate and time are described. The BSM medium formulation is the baseline formulation for the optimization experiments in proceeding chapters. Then, the bioreactor operation is described, including on-line sensors (O_2, CO_2 and MeOH), process monitoring and control system employed in this thesis. Off-line analytical techniques used in this thesis are also described namely OD_{600nm}, WCW, DCW, ELISA, SDS-PAGE, Western Blot, MALDI-TOF/TOF MS, ICP-AES, HPLC and HPICE. Finally, the *P. pastoris* metabolic network and the respective elementary flux modes adopted in this work are presented, which were basic elements of the computation algorithms employed in this thesis for process optimization.

Chapter 3 presents a new algorithm developed in this thesis (the PLP algorithm) for identification of medium composition control of cellular functions. The metabolism of the cell is decomposed into elementary cellular functions using elementary flux modes analysis. The PLP algorithm then maximizes the covariance between medium composition data (required measurements) and respective measure flux data (required measurements) under the constraint of elementary cellular functions. In the end the algorithm ranks the elementary cellular functions according to the degree of correlation with the medium factors. In this chapter the PLP is applied to one case study: a mammalian case.

Chapter 4 starts by presenting a new framework for CM optimization, called Cell Functional Enviromics. The use of Cell Functional Enviromics to map medium factors to cellular functions is the "breakthrough" idea proposed in this PhD thesis. All the steps required for applying this framework to map trace elements to *P. pastoris* cellular functions and then to optimize the trace element solution PTM1 for increasing scFv productivity are presented. This optimized PTM1 medium was then tested in protein expression experiments at 2 L bioreactor scale. This chapter presents the results of validation experiments, showing a twofold increase in scFv productivity.

Chapter 5 investigates the dynamics of the main elements consumption in cultures of *P. pastoris* GS115 expressing a scFv. Nine experiments were performed under different operational conditions in a pilot 50 L bioreactor. A mathematical model describing the concentration profiles of biomass and of the five main elements (Mg, K, Ca, P and S) of the medium used was developed. These elements were measured by ICP-AES. A biomass growth model was derived from the Monod equation while the elements consumption rates were defined by the respective element/biomass yields. The model describes accurately the measured concentration profiles of both calibration and validation experiments, while the parameters confidence bounds are generally very low denoting high statistical confidence. The resulting yield coefficients were used to design a new formulation

of the main salts contained in BSM for a target biomass concentration. This new formulation was incorporated in the cell CM developed in Chapter 4 resulting in the new medium called *Pichia* Functional Enviromics Medium 1 (PFEM1).

Chapter 6 presents an adaptive dissolved oxygen–stat feeding controller that maximizes glycerol feeding under the constraint of available oxygen transfer capacity with the goal of maximizing productivity. The algorithm was applied to a pilot 50 L bioreactor for the cultivation of a *P. pastoris* X33 strain constitutively expressing a scFv. A protein expression increase was observed when running the process at DO concentrations as low as 5% of saturation in optimized CM. Finally, this DO-stat feeding controller was tested in two experiments at 50 L bioreaction scale with different medium, one experiment with the PFEM1 medium and the other with the BSM medium. The results were very positive, namely a 37.0 % increase of the average productivity (mass per unit time) with the PFEM1 medium in comparison to the DO-stat feeding controller with the BSM medium.

Chapter 7 presents the main conclusions derived from this thesis and elaborates on perspectives for future directions.

Bibliography

Berndorff D, Borkowski S, Moosmayer D, Viti F, Müller-Tiemann B, Sieger S, Friebe M, Hilger CS, Zardi L, Neri D, Dinkelborg LM. 2006. Imaging of tumor angiogenesis using 99mTc-labeled human recombinant anti-ED-B fibronectin antibody fragments. J Nucl Med. 47(10): 1707-16.

Berndorff D. 2005. Radioimmunotherapy of solid tumors by targeting extra domain B fibronectin: Identification of the best-suited radioimmunoconjugate. Clinical Cancer Reseach 11(19): 7053S-7063S.

Boze H, Laborde C, Chemardin P, Fabien R, Christine V, Yves C, Guy M. 2001. High-level secretory production of recombinant porcine follicle-stimulating hormone by *Pichia pastoris*. Process Biochem 36: 907–913.

Brierley RA, Davis GR and Holtz GC. 1994. Production of insulin-like growth factor-1 in methylotrophic yeast cells. US Patent 5,324,639.

Calik P, Bayraktar E, Inankur B, Soyaslan ES, Sahin M, Taspinar H, Acik E, Yilmaz R, Ozdamar TH. 2010. Influence of pH on recombinant human growth hormone production by *Pichia pastoris*. Journal of Chemical Techonology and Biotechnology 85(12): 1628-1635.

Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM. 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Current Opinion in Biotechnology 13(4): 329-332.

Cereghino JL and Cregg JM. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiology Reviews 24, 45-66.

Chiruvolu V, Cregg JM, Meagher MM. 1997. Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. Enzyme and microbial technology 21(4): 277-283.

Chung JD. 2000. Design of metabolic feed controllers: Application to high-density fermentations of *Pichia pastoris*. Biotechnol Bioeng 68(3): 298-307.

Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K and Romanos MA. 1991. High-level espression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Bio Technology 9: 455-460.

Cos O, Ramon R, Montesinos JL, Francisco Valero F. 2006. A simple model-based control for *Pichia pastoris* allows a more efficient heterologous protein production bioprocess. Biotechnology and Bioengineering 95(1): 145-154.

Couderc R and Baratti J. 1980. Oxidation of methanol by the yeast, *Pichia pastoris*. Purification and properties of the alcohol oxidase. Agric Biol Chem 44: 2279–2289.

Cregg JM, Barringer KJ, Hessler AY and Madden KR. 1985. *Pichia pastoris* as host system for transformations. Molecular Biotechnology 5: 3376-3385.

Cregg JM and Russell KA. 1998. Transformation. Methods in molecular biology. *Pichia* protocols. 103: 27-39.

Cregg JM, Tschopp JF, Stillman C, Siegel R, Akong M, Craig WS, Buckholz RG, Madden KR, Kellaris PA, Davis GR, Smiley BL, Cruze J, Torregrossa R, Velicelebi G and Thill GP. 1987. High level expression and efficient assembly of hepatitis-B surface-antigen in the methylotrophic yeast. *Pichia pastoris*. Molecular Biotechnology 5(5): 479-485.

Cregg JM, Vedvick TS and Raschike WC. 1993. Recent advances in the expression of foreign genes in *Pichia pastoris*. Molecular Biotechnology 11(8): 905-910.

Cregg JM. 1999. Expression in the methylotrophic yeast *Pichia pastoris*. Gene expression systems 157-191.

Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto R, Moosmayer D, Donner P and Carrondo MJT. 2004. Methanol induction optimization for scFv antibody fragment production in *Pichia pastoris*. Biotechnology and Bioengineering 86(4): 458-467.

Daly R and Hearn MTW. 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J. Mol. Recognit. 18: 119–138.

d'Anjou MC and Daugulis AJ. 2001. A rational approach to improving productivity in recombinant *Pichia pastoris* fermentation. Biotechnol Bioeng 72:1-11.

Delroisse JM, Dannau M, Gilsoul JJ, El Mejdoub T, Destain J, Portetelle D, Thonart P, Haubruge E, Vandenbol M. 2005. Expression of a synthetic gene encoding a Tribolium castaueum carboxylesterase in *Pichia pastoris*. Protein Expr Purif 42: 286–294.

Digan ME, Lair SV, Brierley RA, Siegel RS, Williams ME, Ellis SB, Kellaris PA, Provow SA, Craig WS, Veliçelebi G, Harpold MM, Thill GP. 1989. Continuous Production of a Novel Lysozyme via Secretion from the Yeast, *Pichia pastoris*. Nature Biotechnology 7, 160 – 164.

Files D, Ogawa M, Scaman CH, Baldwin SA. 2001. A *Pichia pastoris* fermentation process for producing high-levels of recombinant human cystatin-C. Enzyme Microb Technol 29: 335-340.

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

Goodrick JC, Xu M, Finnegan R, Schilling BM, Schiavi S, Hoppe H, Wan NC. 2001. High-level expression and stabilization of recombinant human chitinase produced in a continuous constitutive *Pichia pastoris* expression system. Biotechnol Bioeng 74: 492–497.

Guarna MM, Lesnicki GJ, Tam BM, Robinson J, Radziminski CZ, Hasenwinkle D, Boraston A, Jervis E, Macgillivray RTA, Turner RFB, Kilburn DG. 1997. On-line monitoring and control of methanol concentration in shake-flask cultures of *Pichia pastoris*. Biotechnol. Bioeng. 56: 279-286.

Hellwig S, Emde F, Raven NPG, Henke M, van der Logt P, Fischer R. 2001. Analysis of single-chain antibody production in *Pichia pastoris* using on-line methanol control in fed-batch and mixed-feed fermentations. Biotechnology and Bioengineering 74(4): 344-352.

Hohenblum H, Borth N, Mattanovich D. 2003. Assessing viability and cell-associated product of recombinant protein producing *Pichia pastoris* with flow cytometry. J Biotechnol 102 (3): 281-290.

Hoogenboom HR. 2002. Overview of antibody phage-display technology and its applications. Methods Mol Biol 178: 1–37.

Houghton AN and Scheinberg DA. 2000. Monoclonal antibody therapies - a 'constant' threat to cancer. Nature Medicine 6(4): 373-374.

Jensen JM, Vester-Christensen MB, Moller MS, Bonsager BC, Christensen HEM, Abou Hachem M, Svensson B. 2011. Efficient secretory expression of functional barley limit dextrinase inhibitor by high cell-density fermentation of *Pichia pastoris*. Protein Expression and Purification 79(2): 217-222.

Jimenez ER, Sanchez K, Roca H, Delgado JM. 1997. Different methanol feeding strategies to recombinant *Pichia pastoris* cultures producing high level of dextranase. Biotechnol Technique11: 461-466.

Jungo C, Rerat C, Marison IW, von Stockar U. 2006. Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of recombinant avidin in a *Pichia pastoris Mut*$^{(+)}$ strain. Enzyme Microb Technol 39 (4): 936-944.

Katakura Y, Zhang W, Zhuang G, Omasa T, Kishimoto M, Goto Y, and Suga K. 1998. Effect of methanol concentration on the production of human β2-glycoprotein I domain V by a recombinant *Pichia pastoris*: A simple system for the control of methanol concentration using a semiconductor gas sensor. J. Ferm. Bioeng. 86: 482-487.

Khasa YP, Khushoo A, Srivastava L, Mukherjee KJ. 2007. Kinetic studies of constitutive human granulocyte-macrophage colony stimulating factor (hGM-CSF) expression in continuous culture of *Pichia pastoris*. Biotechnol Lett 29(12): 1903–1908.

Khatri NK and Hoffmann F. 2006. Oxygen-limited control of methanol uptake for improved production of a single-chain antibody fragment with recombinant *Pichia pastoris*. Microbiology and Biotechnology 72(3): 492-498.

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 (3): 280-288.

Koutz P, Davis GR, Stillman C, Barringer K, Cregg J, Thill G. 1989. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast, 5(3): 167–177.

Kunkel, JP. 1998. Dissolved oxygen concentration in serum-free continuous culture affects n-linked glycosylation of a monoclonal antibody. Journal of Biotechnology 62 (1): 55-71.

Laroche Y, Storme V, De Meutter J, Messens J and Lauwereys M. 1994. High-level secretion and very efficient isotopic labeling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast *Pichia pastoris*. Bio Technology 12: 1119-1124.

Lim HK, Choi SJ, Kim KY, Jung KH. 2003. Dissolved-oxygen-stat controlling two variables for methanol induction of rGuamerin in *Pichia pastoris* and its application to repeated fed-batch. Appl Microbiol Biotechnol 62: 342-348.

Macauley-Patrick S, Fazenda ML, McNeil B and Harvey LM. 2005. Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22: 249–270.

Mandenius CF and Brundin A. 2008. Bioprocess Optimization Using Design-of-Experiments Methodology. Biotechnology Progress 24(6): 1191-1203.

Marty C, Scheidegger P, Ballmer-Hofer K, Klemenz R, Schwendener RA. 2001. Production of functionalized single-chain Fv antibody fragments binding to the ED-B domain of the B-isoform of fibronectin in *Pichia pastoris*. Protein expression and purification 21(1): 156-164.

Mellitzer A, Weis R, Glieder A, Flicker K. 2012. Expression of lignocellulolytic enzymes in *Pichia pastoris*. Microbial Cell Factories, 11:61.

Montesino R, Garcia R, Quintero O and Cremata JA. 1998. Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. Protein expression and purification 14(2): 197-207.

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

Neri D, Carnemolla B, Nissim A, Leprini A, Querze G, Balza E, Pini A, Tarli L, Halin C, Neri P, Zardi L, Winter G. 1997. Targeting by affinity-matured recombinant antibody fragments of an angiogenesis associated fibronectin isoform. Nature Biotechnology 15(12): 1271-1275.

Oehler R, Lesnicki G and Galleno M. 1998. High cell density fermentation of *Pichia pastoris* using nonphosphate precipitate forming sodium hexametaphosphate as a phosphate source. Current topics in gene expression annual meeting. San Diego, CA, USA.

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 (7): 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.

Ooijkaas LP, Wilkinson EC, Tramper J, Buitelaar RM. 1999. Medium optimization for spore production of Coniothyrium minitans using statistically-based experimental designs. Biotechnology and Bioengineering 64(1): 92-100.

Pais JM, Varas L, Valdes J, Cabello C, Rodriguez L, Mansur M. 2003. Modeling of mini-proinsulin production in *Pichia pastoris* using the AOX promoter. Biotechnol Lett 25 (3): 251-255.

Pal Y, Khushoo A, Mukherjee KJ. 2006. Process optimization of constitutive human granulocyte-macrophage colony-stimulating factor (hGM-CSF) expression in *Pichia pastoris* fed-batch culture. Appl Microbiol Biotechnol 69 (6): 650-657.

Pichia fermentation process guidelines. 2000. Invitrogen Corporation.

Plank H, Bona R, Dib I, Gruber U, Purkarthofer T, Weis R. 2012. *Pichia* and HAS (fusions) – a love story. *Pichia* 2012 Conference, p. 89.

Plantz BA, Sinha J, Villarete L, Nickerson KW, Schlegel VL. 2006. *Pichia pastoris* fermentation optimization: energy state and testing a growth-associated model. Appl Microbiol Biotechnol 72 (2): 297-305.

Potgieter TI, Kersey SD, Mallem MR, Nylen AC, d'Anjou M. 2010. Antibody Expression Kinetics in Glycoengineered *Pichia Pastoris*. Biotechnol Bioeng 106 (6): 918-927.

Robinson MK, Weiner LM, Adams GP. 2004. Improving Monoclonal Antibodies for Cancer Therapy. Drug Development Research 61: 172–187.

Romanos MA, Clare JJ, Beesley KM, Rayment FB, Ballantine SP, Makoff AJ, Dougan G, Fairweather NF and Charles IG. 1991. Recombinant Bordetella pertussis pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties. Vaccine 9: 901-906.

Rosenfeld SA. 1999. Use of *Pichia pastoris* for expression of recombinant proteins. Expression of recombinant genes in eukaryotic systems 306: 154-169.

Siegel RS, Buckholz RG, Thill GP, Wondrack LM. 1990. New DNA fragment for expressing epidermal growth factor - in methylotrophic yeast, contg. host derived promoter and terminator and *S. cerevisiae* pre-pro sequence. WO9010697-A1.

Sinha J, Plantz BA, Zhang WH, Gouthro M, Schlegel V, Liu CP, Meagher MM. 2003. Improved production of recombinant ovine interferon-tau by *Mut*$^{(+)}$ strain of *Pichia pastoris* using an optimized methanol feed profile. Biotechnol Prog 19 (3): 794-802.

Soyaslan ES and Calik P. 2011. Enhanced recombinant human erythropoietin production by *Pichia pastoris* in methanol fed-batch/sorbitol batch fermentation through pH optimization. Biochemical Engineering Journal 55(1): 59-65.

Sreekrishna K, Nelles L, Potenz R, Cruze J, Mazzaferro P, Fish W, Motohiro F, Holden K, Phelps D, Wood P and Parker K. 1989. High-level expression, purification, and characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast *Pichia pastoris*. Biochemistry 28: 4117-4125.

Tschopp JF, Sverlow G, Kosson R, Craig W, Grinna L. 1987. High-Level Secretion of Glycosylated Invertase in the Methylotrophic Yeast, *Pichia Pastoris*. Nature Biotechnology 5: 1305 – 1308.

van Dijken JP and Harder W. 1974. Optimal conditions for enrichment and isolation of methanol assimilating yeast. Journal of general microbiology 84: 409-411.

Vedvick T, Buckholz RG, Engel M, Urcan M, Kinney J, Provow S, Siegel RS, Thill GP. 1991. High-level secretion of biologically active aprotinin from the yeast *Pichia pastoris*. J Ind Microbiol. 7(3): 197-201.

Veenhuis M, VanDijken JP, Harder W. 1983. The significance of peroxisomes in the metabolism of one-carbon compounds in yeasts. Advances in microbial physiology 24: 1-82.

Wagner SL, Siegel RS, Vedvick TS, Raschke WC, Van Nostrand Wevannostrand WE. 1992. High level expression, purification, and characterization of the Kunitz-type protease inhibitor domain of protease nexin-2/amyloid β-protein precursor. Biochemical and biophysical research communications 186(2): 1138-1145.

Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. 1997. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186: 37-44.

Zhang AL, Zhang TY, Luo JX, Fu CY, Qu Z, Yi GH, Su DX, Tu FZ, Pan YW. 2009. Inducible expression of human angiostatin by AOXI promoter in *P. pastoris* using high-density cell culture. Molecular Biology reports 36(8): 2265-2270.

Zhang W, Sinha J, Smith LA, Inan M, Meagher MM. 2005. Maximization of production of secreted recombinant proteins in *Pichia pastoris* fed-batch fermentation. Biotechnol Prog 21 (2): 386-393.

Zhang WH, Bevins MA, Plantz BA, Smith LA and 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, Potter KJ, Plantz BA, Schlegel VL, Smith LA, Meagher MM. 2003. *Pichia pastoris* fermentation with mixed-feeds of glycerol and methanol: growth kinetics and production improvement. J Ind Microbiol Biotechnol 30: 210-215.

Zhang WH, 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.

CHAPTER 2

Materials and methods

Abstract

Over this thesis it was necessary to develop, implement and validate computational and experimental techniques specific to the cultivation of *P. pastoris* and to the analysis of compounds including biomass and the scFv antibody studied in this work. Cultivations were carried in 250 mL shake flasks and in 2 L and 50 L bioreactors with different monitoring and control systems. In this chapter we describe in detail the *P. pastoris* strains used, the composition of the chemically defined medium, the incubation conditions (temperature, pH, agitation rate and incubation time) of the pre-inoculum step and of the bioreactor cultivation steps. The standard on-line sensors used in the pilot 50 L bioreactor, namely the dissolve oxygen probe, the gas analyzer and the metal oxide semiconductor gas sensor are briefly described. The qualitative and quantitative methods of analysis employed in this work are described in detail. For measuring the cell concentration, optical density at 600 nm (OD_{600nm}) and wet cell weight (WCW) techniques were used. The concentration of the antibody fragment was determined by ELISA, while identification and characterization was performed by SDS-PAGE, Western Blot and MALDI-TOF/TOF MS. Concentration of selected elements (P, K, Mg, Ca and S) present in the chemical defined medium was determined using the ICP-AES technique. HPLC and HPICE were used for identification and quantification of individual components of the supernatant, such as, glycerol, methanol and some organic acids. Finally, we present the *P. pastoris* metabolic network and the respective elementary flux modes, which are basic network structures employed in several chapters of this thesis.

2.1 *P. pastoris* strains, culture medium and inoculum preparation

Two *P. pastoris* strains were used in this work: a stable *P. pastoris* X33 strain constitutively expressing the scFv under the control of GAP promoter, and a stable *P. pastoris* GS115 (*Mut*$^+$) strain expressing the same scFv under the control of AOX1 promoter. The cultivation techniques were the same for both strains. The cryogenic vials containing the cells stock were stored at -80 ^oC. The Basal Salts Medium (BSM) proposed by Invitrogen (*Pichia* fermentation process guidelines, 2000) is the reference medium formulation and was used very frequently in all growth steps (cell stock, pre-inoculum, inoculum and bioreactor cultivation). The composition of the BSM is the following:

Basal Salts Medium (BSM) (baseline formulation): *Pichia* Main Salts (PMS) solution 995.65 mL/L, *Pichia* Trace Metal 1 (PTM1) salts supplements solution 4.35 mL/L and glycerol 40.00 g/L.

Pichia Main Salts (PMS): H_3PO_4 85 % 26.70 mL/L, $CaSO_4.2H_2O$ 0.93 g/L, K_2SO_4 18.20 g/L, $MgSO_4.7H_2O$ 14.90 g/L and KOH 4.13 g/L.

Pichia Trace Metals 1 (PTM1): $CuSO_4.5H_2O$ 6.00 g/L, NaI 0.08 g/L, $MnSO_4.H_2O$ 3.00 g/L, $Na_2MoO_4.2H_2O$ 0.20 g/L, H_3BO_3 0.02 g/L, $CoCl_2.6H_2O$ 0.50 g/L, $ZnCl_2$ 20.00 g/L, $FeSO_4.7H_2O$ 65.00 g/L, H_2SO_4 5.00 mL/L and biotin 0.20 g/L.

The pH of the PMS solution is approximately 1.5 and must be adjusted to the working pH of 5.0 with addition of 25 % NH_4OH before medium sterilization, which also served as the sole nitrogen source. BSM was sterilized at 121 ^oC for 30 minutes. The PTM1 salts supplements solution was filter sterilized with 0.2 mm pore size filter and then added aseptically to the PMS.

A pre-inoculum was prepared in a 250 mL shake flask by inoculating 40 mL of BSM, pH=5.0, with one cryovial, containing 1 mL of *P. pastoris* cell stock, and incubated at 30 ^oC and at 150 rpm for 3 days.

For the growth of the inoculum, 10 mL of the pre-inoculum was used to inoculate 750 mL of BSM, pH=5.0, then incubated for 3 days at 30 ºC and at 150 rpm. The inoculum, 750 mL of the previous cells growth with OD_{600nm}= 2-6, was used to inoculate the 50 L bioreactor, with an initial volume of 15 L of BSM, pH=5.0.

2.2 Bioreactor operation

2.2.1 *P. pastoris* X33 strain

Cultivations were carried out in 1.5 L and 42 L working volume bioreactors. The 2 L bioreactor (Biostat B, Braun Biotech, Germany) was operated in batch mode using glycerol as carbon source. Temperature, pH and dissolved oxygen (DO) were monitored and controlled. The pilot bioreactor (Lab Pilot Fermenter Type LP351, 50 L, Bioengineering, Switzerland) was operated in two phase: glycerol batch (GB) phase and glycerol fed-batch (GFB) phase. The initial operational conditions for volume, pressure, agitation rate, and airflow rate were 15 L, 100 mbar, 300 rpm and 1800 L/h, respectively. The temperature at kept constant 30 ºC and pH was controlled to 5.0 using a 25 % NH_4OH. DO in the medium begin at 100 % air saturation concentration and was allowed to decrease down to 50 %. DO concentration was then controlled by varying the agitation rate between 300 and 1000 rpm using a PID controller. As the cells grow, oxygen consumption increases requiring higher agitation rates to be used for keeping DO concentration at 50 %. An agitation rate of 440 rpm under the used conditions corresponds to a cell concentration reached when all glycerol initially added to the CM was consumed, being this initiated the GFB phase by feeding a glycerol rich solution containing glycerol 99 % w/v and 12 mL/L of PTM1 solution over time. After reaching the maximum agitation rate, DO concentration was allowed to decrease to the required value for the experiment to be performed (typically between 5 and 10 %), and then controlled by varying the pressure between 100 and 800 mbar using a PID controller. The DO is a critical parameter for the cultivation with *P. pastoris* due the high cell density cultivation (WCW>450 g/L). The oxygen consumption varies and depends on the amount of carbon source added in the medium and the biomass concentration achieved.

The 50 L bioreactor was equipped with a standard polarographic DO probe from Mettler Toledo (Inpro 6820) and a pH probe from Mettler Toledo (405-DPAS-SC-K8S/120).

2.2.2 *P. pastoris* GS115 (*Mut*$^+$) strain

Cultivations of this strain were carried out in 42 L working volume bioreactor. The pilot bioreactor was operated in three phase: glycerol batch (GB) phase, glycerol fed-batch (GFB) phase and methanol fed-batch (MFB) phase. The initial operational conditions and the details of the bioreactor operation were the same described above. After the GFB phase finished, the MFB phase begins by feeding a methanol rich solution containing methanol and 24 mL/L of PTM1 solution over time.

2.2.3 Gas analyzer

The metabolism of the yeast is reflected on the ratio between carbon dioxide production and oxygen consumption. The respiratory quotient, RQ= rCO_2/rO_2 indicates the efficiency of the energetic metabolism. For this reason the monitoring of these gases is very important, as the ratio between CO_2 produced and O_2 consumed indicates the efficiency of the production process and the status of the bioreactor. The O_2 consumption and CO_2 production in the bioreaction was determined by mass balance using measured concentration of O_2 and CO_2 in the outlet gas flow, analyzed using a Gas Analyzer from Hach Ultra Orbisphere, serie 3621. This sensor is a combination of a Thermal Conductivity (TC) sensor and a Electrochemical Sensor (EC). The basic elements of the system are the Microprocessor-controlled Analyzer and the Thermal Conductivity Sensor(s). If one of these elements does not function properly, the whole system is affected. The analyzer measures the thermal conductivity signal and converts it into a digital signal for processing. The analyzer controls the measuring mode of the sensor and provides an analogue or digital output signal for the process monitoring and control system UBICON.

2.2.4 Methanol sensor

The methanol concentration in the liquid phase was determined indirectly by measuring the concentration of methanol in the outlet gas flow, using a Metal Oxide Semiconductor (MOS) gas sensor TGS 2620 from Figaro Engineering (Osaka, Japan). The sensing element of this type of sensor comprises of a metal oxide semiconductor layer (most typically SnO_2) formed on an alumina substrate of a sensing chip together with an integrated heater. In the presence of a detectable gas, the sensor conductivity increases with its concentration in the exhausting gas. A simple electrical circuit converts the changes in conductivity to an output signal which corresponds to the gas concentration. The MOS gas sensor TGS 2620 has high sensitivity to the vapors of organic solvents as well as other volatile vapors. This sensor was very important for the control of methanol feeding in the *P. pastoris* GS115 (*Mut*$^+$) experiments. In our case, this sensor detects methanol concentration in the liquid above toxic levels, 3.7 g/L to 20.0 g/L (Potvin *et al.* 2012) to enable the control of methanol feeding which is interrupted at methanol concentrations below that level.

2.2.5 Bioreactor supervisory system

The bioreactor operation was supervised and controlled by the Universal BIo-process CONtrol system (UBICON, Electronic System Design, Hannover, Germany). This supervisory system, based on a VME architecture (industrial standard IEC 1014), includes hardware and software interfaces, connecting equipment via a CAN field-bus (ISO 11898), and enables real-time and multi-tasking operation. UBICON is an open control system enabling the implementation of tailored control algorithms based on discrete–time or ordinary differential equations. All the above instruments were connected to the UBICON.

2.3 Off-line analytic methods implemented

2.3.1 Biomass concentration

The measurement of cell concentration in culture was routinely performed by measuring light absorbance at 600 nm of cells in suspension (OD_{600nm}). The OD_{600nm} value is directly proportional to the cell density in culture in diluted suspensions. Absorbance of samples was also measured against distilled water using an UltraViolet/Visible-1101 spectrophotometer (Biotech Photometer, WPA, UK) at 600 nm, after appropriate dilution ensuring a value within the linear range (0.1-0.6). Each sample was measured in triplicate.

Cell concentration was also quantified routinely as grams of wet cell weight per Liter of broth, g-WCW/L. The samples were prepared by centrifuging 14 mL of broth at 15000 rpm for 10 minutes to remove the supernatant. Each sample was measured in duplicate.

For calibration purpose, cell concentration was also assayed as grams of dry cell weight per Liter of broth, g-DCW/L, by centrifuging 14 mL of broth at 15000 rpm for 10 minutes to remove the supernatant and then washed twice to remove soluble components (salts of the medium). Cells were washed again by resuspension of the cell pellet in deionized water followed by centrifuging and supernatant removal. The washed cell pellet is then dried in a drying oven at 80 - 100 ^{o}C until complete drying of biomass, indicated by constant mass. One gram of wet cell weight (WCW) is equivalent to approximately 0.28 gram of dry cell weight (DCW).

2.3.2 Gram's method

Selected samples were also analyzed visually under a light microscope (OLYMPUS, Japan) to check for possible contaminants. The Gram's method is almost always the first step in the identification of a bacterial organism. This empirical method allows differentiating bacterial species into two large groups: Gram positive (which purple color) and Gram negative (which pink color), based

on the chemical and physical properties of their cell walls. In our case, the yeast is a Gram positive that is represented in the Figure 2.1 with the purple color.

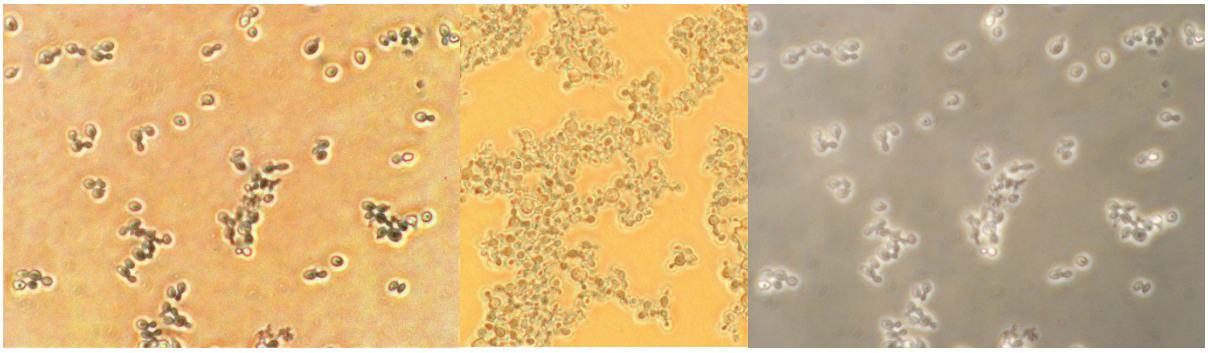


Figure 2.1 *P. pastoris* visualized by optic microscopy (100x magnification).

2.3.3 Quantification of the standard scFv

Bicinchoninic acid (BCA) Protein Assay kit from Pierce was used to quantify total protein concentration (scFv) of a pure protein sample to be used as standard for all analytical methods requiring a standard with known protein concentration, namely the ELISA. This colorimetric method is based on reduction of Cu^{+2} to Cu^{+1} in an alkaline medium with a selective colorimetric detection of the cuprous cation (Cu^{+1}) using a unique reagent containing bicinchoninic acid. The protein standard used for curve calibration was a BSA (Bovine serum albumin), a standard provided in the BCA assay kit. The final volume was 250 µL (50 µL of sample plus for 200 µL of BCA reagent).

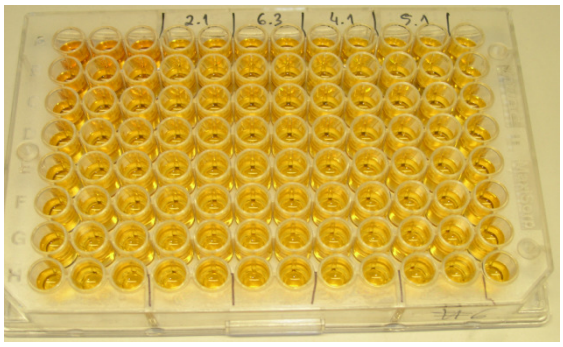
The assay was conducted at 37 ^{o}C for 30 min and the absorbance was measured at λ = 562 nm. The protein identification is made by the transition of a colorless solution to an intense purple color. The BSA was the reference protein and the dilution curve consisted of eight points in duplicate with the concentration between 0.1 – 2.0 mg/mL. The scFv was made in duplicate and the concentration was calculated by linear regression. An illustrative example of this assay is shown in Appendix A. This purified scFv solution was aliquoted and stored frozen (-80 ^{o}C) and used as standard scFv for the sandwich ELISA and positive control for SDS-PAGE and Western blot in all analysis performed.

2.3.4 scFv quantification by ELISA

The Enzyme-Linked Immunosorbent Assay (ELISA) was used to quantify the scFv of antibody fragment in different samples. A "sandwich" ELISA with two antibodies was used for scFv quantification. The main advantage of this technique is given by the specificity of the antibodies, permitting the quantification of scFv protein in non purified samples (containing other proteins).

The sandwich ELISA method used for this work was performed as follows: 96-well microplates (F96 MaxiSorp, Nunc Immunoplate), shown in Figure 2.2a), were coated with 1 g/L of Protein A (step 1) in 10 mL of PBS buffer (10 mM phosphate buffer saline, pH=7.4), and allowed to incubate overnight at 4 ^{o}C or 1 hour at 37 ^{o}C. The plates were washed three times in PBS-Tween (10 mM phosphate buffer saline contains tween 20 pH 7.4), and blocked in PBS-Tween with 5 % (w/v) skim milk (step 2) for 1 hour at 37 ^{o}C. After washed three times with PBS-Tween, standard purified scFv, samples to be analyzed and PBS negative control (step 3), were applied after appropriate dilution with skim milk (samples were analyzed in duplicate) for 1 hour at room temperature, and then washed three times with PBS-Tween. Then the plates were incubated with anti-human IgG-Peroxidase (step 4) for 30 minutes at 37 ^{o}C and washed again with PBS-Tween. The plates were then incubated in the dark with the coloring solution, o-Phenylenediamine dihydrochloride (SIGMA FAST™ OPD) (step 5) for approximately 10 minutes. After sufficient color development, the stop solution 2.5 M H_2SO_4 (step 6) were added to each well. Absorbance was measured at a reading wavelength of 450 nm using a microplate reader (PowerWave HT Microplate Spectrophotometer, BioTek, USA). Quantification of protein samples was carried out by known standards of purified scFv. The concentration was calculated by linear regression. In Appendix A an example of this technique is presented. The several steps of ELISA are represented in Figure 2.2b.

a)



b)

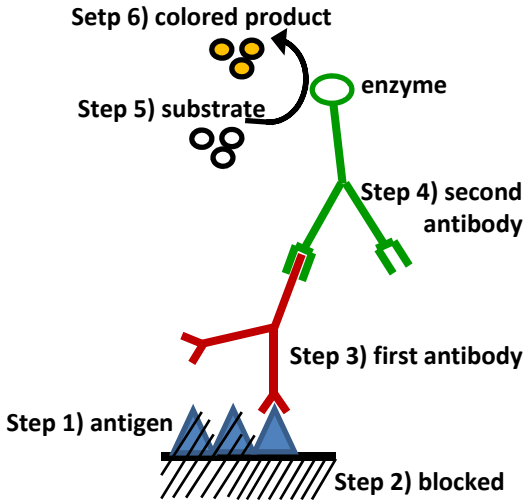


Figure 2.2 a) ELISA 96-well microplate after color development. b) Steps of sandwich ELISA.

2.3.5 SDS-PAGE

Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE) is a technique widely used to separate proteins according to their electrophoretic mobility, a function of polypeptide chain length or molecular weight.

For SDS-PAGE analysis, 5-10 µL of protein solution was mixed with 10 µL of 2x sample buffer solution (2% v/v Sodium Dodecyl Sulfate (SDS), 0.1 M dithiothreitol (DTT), 10 % v/v Glycerol, 0.001 % Bromophenol Blue, 0.08 M Tris-HCl pH=6.8). The mixture was heated (approximately 70 ^{o}C) for 3 min. A NuPAGE 4-12 % Bis-Tris-HCl polyacrylamide mini-gel (Invitrogen) in MES SDS running buffer (Invitrogen) was used. The gel was run for 35 minutes under reducing condition, and then stained with Coomassie Blue staining reagent for 1 hour at room temperature and washed with water.

2.3.6 Western blot

The Western blot is a technique extremely useful to detect specific proteins. After SDS-PAGE analysis, the gel was transferred to polyvinylidene difluouride (PVDF) membrane for 1 hour at 45 mA. The membrane was blocked for 1 hour at room temperature with 10 % Skim Milk power in PBS-Tween (140 mM NaCl, 0.05 % (w/v) tween 20, 10 mM phosphate buffer, pH=7.4 at 25 ^{o}C). After this step, the

membrane was incubated for 1 hour at room temperature with the first antibody (rabbit, anti scFv) in blocking solution. The membrane was then incubated for 1 hour at room temperature with the second antibody (anti-rabbit - alkaline phosphatase conjugated) in blocking solution. The bands were visualized by incubation with BCIP/NBT alkaline phosphatase substrate. Three washing steps with PBS-Tween were carried out between all incubation steps.

2.3.7 MALDI-TOF/TOF MS

The Matrix-Assisted Laser Desorption/Ionization (MALDI) technique is a soft ionization technique used in mass spectrometry to allow identification, verification and quantification of proteins isolated from natural sources, recombinant proteins, metabolites, oligonucleotides, peptides, polymers, and organic compounds. The mass spectrometer provides high sensitivity and at same time structural and chemical details. The basic function of a mass spectrometer is to determine the mass-to-charge ratio (**m/z**) of charged compounds, where **m** is the ion mass measured in daltons (Da) and **z** is the number of elementary charges. This technique consists of three main components: the ion source that will ionize the molecules of the sample (MALDI); the mass analyzer, which separates the ions based in the ratio **m/z**, that can be Time-of-flight (TOF), quadruple or an ion trap; and lastly, the detector, which records the ions that arrive and measure the ratio **m/z** (quantifying the ion abundance). The ionization is triggered by a laser beam (normally a nitrogen laser). A matrix is used to protect the biomolecule from being destroyed by direct laser beam and to facilitate vaporization and ionization. The MALDI technique has several advantages over other mass spectrometry techniques such as soft ionization, it can analyze intact biomolecules and synthetic polymers; it can analyze a large variety of compounds, with masses over 300 kDa; it can be more tolerant to buffers and salts than either Fast-atom bombardment (FAB) or Electrospray-ionization (ESI); MALDI is 10-100 times more sensitive than FAB for detection of underivatized oligosaccharides and it is faster.

MALDI-TOF was applied in this work to quantify the amounts of scFv in monomeric and dimeric forms. The protein was eluted directly onto a MALDI plate with Sinapinic Acid (SA) as the matrix solution. Mass spectra of the peptide mixtures were acquired in the linear positive MS mode using a 4800 plus MALDI-TOF/TOF MS analyzer. The samples were irradiated with ultraviolet (UV) light, a 337 nm, and nitrogen laser at 200 Hz. The acceleration voltage was 25 KV. Each mass spectrum was generated by averaging 3200 laser shots and given molecular masses represent the average masses of the H^+ adducts ($[M + H]^+$). The goal of this analysis was to obtain an exact mass of the protein, the monomeric form of the scFv is approximately 25 KDa. Figure 2.3 shows the three different forms of scFv identified.

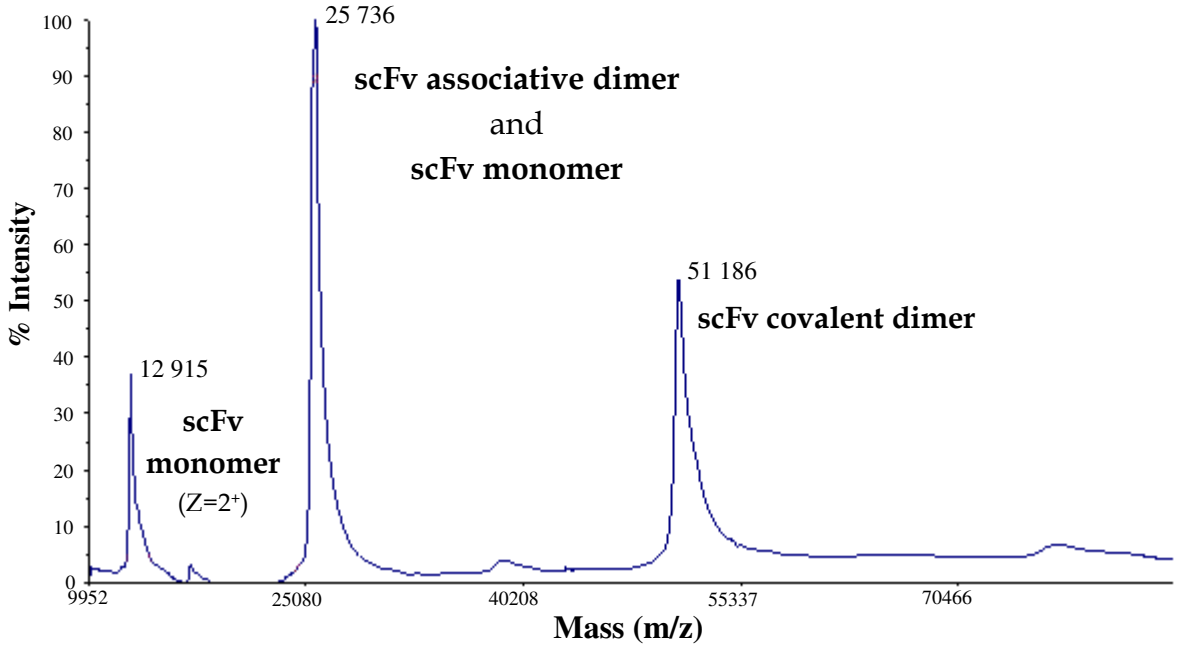


Figure 2.3 MALDI-TOF/TOF MS spectrum of purified scFv sample.

2.3.8 ICP-AES

Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES) is an analytical technique used for the detection of main elements present in the CM. It is a type of emission spectroscopy that uses the inductively coupled plasma to produce excited atoms and ions that emit electromagnetic radiation at

wavelengths characteristic of a particular element. The intensity of this emission is indicative of the concentration of the element in the sample.

ICP-AES was used to determine the concentration of some elements, such as P, K, Mg, Ca and S present in the chemical medium. The system was operated at Argon flow rate of 15 L/min, temperature between 5700 - 10000 ^{o}C, 3 bar of pressure and plasma potency of 1 KW.

2.3.9 HPLC

High Performance Liquid Chromatography (HPLC) is a chromatographic technique that enables to separate a mixture of compounds and is used to identify, quantify and purify the individual components in a mixture. HPLC typically utilizes different types of stationary phases, a pump that moves the mobile phase(s) and the compound through the column, and a detector that identify each compound at a characteristic retention time. Compound retention time may vary with the strength of its interactions with the stationary phase, the ratio/composition of solvent(s) used, and the flow rate of the mobile phase. The Aminex HPX-87H column was used for this analysis and eluent composition can be manipulated to adjust the relative retention times of the compounds to be resolved.

HPLC analysis was performed using an AMINEX HPX-87H (Biorad, Germany) column at the following conditions. The temperature used was 50 ^{o}C and the pressure was 30-50 mbar, the mobile phase was a H_2SO_4 0.01 M solution with a flow of 0.6 mL/min and the washing solvent a H_2SO_4 0.05 M solution. Detection was performed using a refractive index (RI).

The feeding of the glycerol is controlled to avoid accumulation in the medium. Glycerol and organic acids concentrations were determined by HPLC. Analysis of samples in the induction phase confirmed low glycerol concentration in the medium is low (0.22 g/L on average) indicating a correct feeding strategy.

Some organic acids were also detected, such as malic acid, succinic acid, lactic acid, formic acid and fumaric acid. An example is shown in Chapter 4, Table 4.4.

2.3.10 HPICE

The organic acids present in the samples were also analyzed by High Performance Ion Chromatography Exclusion (HPICE) in Ion Chromatographic System Dionex, ICS 3000 SP with ICS 3000 conductivity detector, AD 25 UltraViolet/Visible detector, sample changer (AS) and Chromeleon 6.80 control evaluation software. The separation mechanism in ion-exclusion chromatography is governed by Donnan exclusion, steric exclusion, sorption processes and, depending on the type of separation column, by hydrogen bonding. Ion-exclusion chromatography is primarily employed for the separation of weak inorganic and organic acids. Due to Donnan exclusion, fully dissociated acids are not retained at the stationary phase, eluting therefore within the void volume as a single peak. Undissociated compounds, however, can diffuse into the pores of the resin, since they are not subject to Donnan exclusion. In this case, separations are based on non-ionic interactions between the solute and the stationary phase. In combination with suitable detection systems, this separation method is also useful for determining amino acids, aldehydes, and alcohols. In this work, the analysis was performed by HPICE with an IonPac ICE-AS1 column (Dionex). The selectivity of the IonPac ICE-AS1 column is designed to separate an extensive group of low molecular weight organics acids in less than 20 minutes. The ICE-AS1 column consists of a cross-linked (8 %), microporous, hydrophilic resin that has been sulfonated. The nature of the cross-linked polymeric structure of the packing material makes the ICE-AS1 columns compatible with pH 0 - 7 eluents. A strong acid eluent facilitates protonation of weak organic acids. Temperature can affect the retention time and the selectivity of organic acids.

The HPICE analysis with an IonPac ICE-AS1 column was coupled to an UV detector, using sulphuric acid (H_2SO_4 0.01 N) as eluent, at a flow rate of 0.6 mL/min and a temperature of 30 ^{o}C. The detection was performed at 210 nm.

2.4 *P. pastoris* metabolic network

Here we describe the *P. pastoris* metabolic network adopted in this thesis. A *P. pastoris* metabolic network was built based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa Laboratories 2010) and papers by Chung *et al.* 2010 and Çelik *et al.* 2010. The genes associated to each reaction are in most cases known and can be found in Chung *et al.* 2010. The network included the following processes/pathways:

- Uptake reactions (glycerol, methanol, sulphate, phosphate);
- Glycolysis/gluconeogenesis;
- Pentose phosphate pathway;
- Tricarboxylic acid (TCA) cycle;
- Biosynthesis of macromolecular components of biomass (proteins, amino acids, lipids, carbohydrates and nucleotides);
- Biosynthesis of biomass;
- Biosynthesis of product (scFv);
- Oxidative phosphorylation;
- Interconversion of folate compounds;
- Energy interconversions.

The metabolic network was further simplified by lumping together in single reactions the consecutive reactions in the pathways for the synthesis and degradation of the biomass and product precursors and assuming a fixed P/O ratio of 2 mol-ATP/mol-$NADH_2$. The stoichiometry of adenosine triphosphate (ATP), reduced nicotinamide adenine dinucleotide ($NADH_2$) and water (H_2O) was also accounted for the metabolic reactions in order to close the balance of oxygen, hydrogen and phosphorus. The resulting metabolic network has 102 reactions (thus 102 fluxes), 90 intracellular metabolites and 11 extracellular metabolites (12 % of all metabolites). The complete list of metabolic reactions is provided in the Appendix B, Table B.1.

2.4.1 Elementary flux modes

Elementary flux modes (EFM) are basic structural properties of metabolic networks, which were used several times in the following chapters of this thesis. An EFM can be defined as a minimal set of enzymes able to operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function (Schuster *et al.* 2000). Here we adopted EFM as function descriptors because they enable systematic, large-scale computational analysis of biochemical networks from a functional viewpoint. Mathematically, EFM form the convex basis of the null space solution of a metabolic network stoichiometric matrix. Biologically, they represent elementary network topologies defining all possible independent operational modes of a cell. Motivated by these unique properties, EFM analysis has become a widespread technique for systems level metabolic pathways analysis (Klamt and Stelling 2003, Palsson *et al.* 2003, Schuster *et al.* 1999, Trinh *et al.* 2009).

In particular, the phenotype of a cell, as defined by its fluxome, r, can be expressed as a weighted sum of the contribution of EFM:

$$\mathbf{r} = \lambda_1 \mathbf{em}_1 + \lambda_2 \mathbf{em}_2 + \ldots + \lambda_K \mathbf{em}_K = \cdot\sum_{i=1}^{K} \lambda_i \times em_i \qquad (2.1)$$

with λ_i the weighting factor of EFM em_i, K the number of EFM and $\dim(r)=\dim(em_i)=q$ the number of metabolic reactions of the cell. The universe of EFM is primarily determined by the genome of the cells.

The METATOOL 5.0 software (von Kamp and Schuster 2006) was used to calculate the *P. pastoris* metabolic network EFM. The total number of EFM for glycerol (X33 strain) feeding are 3368, namely, for biomass growth, scFv synthesis, simultaneous biomass growth and scFv synthesis and catabolism are 2048, 768, 512, 40 respectively. The total number of EFM for methanol (GS115 strain) feeding are 1368, namely, for biomass growth, scFv synthesis, simultaneous biomass growth and scFv synthesis and catabolism are 880, 436, 8, 44 respectively.

Bibliography

Chung BKS, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY. 2010. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microbial Cell Factories 9.

Çelik E, Calik P, Oliver SG. 2010. Metabolic Flux Analysis for Recombinant Protein Production by *Pichia pastoris* Using Dual Carbon Sources: Effects of Methanol Feeding Rate. Biotechnology and Bioengineering 105(2): 317-329.

Klamt S and Stelling J. 2003. Two approaches for metabolic pathway analysis? Trends Biotechnol, 21(2): 64-69.

Kyoto Encyclopedia of Genes and Genomes (KEGG) developed by Kanehisa Laboratories, 2010. http://www.genome.jp/kegg/

Palsson BO, Price ND, Papin JA. 2003. Development of network-based pathway definitions: the need to analyze real metabolic networks. Trends Biotechnol 21(5): 195-198.

Pichia fermentation process guidelines. 2000. Invitrogen Corporation.

Potvin G, Ahmad A and Zhang Z. 2012. Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: a review. Biochemical Engineering Journal 64:91-105.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17(2): 53-60.

Schuster S, Fell DA, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18(3): 326-332.

Trinh CT, Wlaschin A, Srienc F. 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81(5): 813-826.

von Kamp A and Schuster S. 2006. METATOOL 5.0: fast and flexible elementary modes analysis. Bioinformatics 22(15): 1930-1931.

CHAPTER 3

Projection to Latent Pathways (PLP): a computational method for the identification of environmental control over core cellular functions

Adapted from:

Ferreira AR, Teixeira AP, Carinhas N, Portela RMC, Isidro IA, von Stosch M, Dias JML, Oliveira R. 2011. Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC Systems Biology 5:181.

Abstract

Elementary flux modes (EFM) are unique and non-decomposable sets of metabolic reactions able to operate coherently in steady-state. A metabolic network has in general a very high number of EFM reflecting the typical functional redundancy of biological systems. However, most of these EFM are either thermodynamically unfeasible or inactive at pre-set environmental conditions. This thesis presents a new algorithm that discriminates the "active" set of EFM on the basis of dynamic envirome data. The algorithm merges together two well-known methods: Projection to Latent Structures (PLS) and EFM analysis, and is therefore termed Projection to Latent Pathways (PLP). PLP has two concomitant goals: (1) maximization of correlation between EFM weighting factors and measured envirome data and (2) minimization of redundancy by eliminating EFM with low correlation with the envirome. Overall, our results demonstrate that PLP slightly outperforms PLS in terms of predictive power. But more importantly, PLP is able to discriminate the subset of EFM with highest correlation with the envirome, thus providing in-depth knowledge of how the environment controls core cellular functions. This offers a significant advantage over PLS since its abstract structure cannot be associated with the underlying biological structure. In this chapter the PLP is applied to one case study: a mammalian case.

3.1 Introduction

As described in Chapter 2.4.1, an elementary flux mode can be defined as a minimal set of enzymes able to operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function (Schuster *et al.* 2000). The universe of EFM of a given metabolic network defines the full set of non-decomposable steady-state flux distributions that the network can support. Any particular steady-state flux distribution can be expressed as a non-negative linear combination of EFM. Motivated by these unique properties, EFM analysis has become a widespread technique for systems level metabolic pathways analysis (Klamt and Stelling 2003, Palsson *et al.* 2003, Schuster *et al.* 1999, Trinh *et al.* 2009).

The number of EFM of a metabolic network is in general very high, denoting the innate adaptability and robustness of biological systems. As illustrative example, the central carbon metabolism of a genome-scale reconstructed *E. coli* metabolic network has approximately 26 million EFM (Terzer and Stelling 2008). Over the last decade several methods were proposed to reduce the number of EFM founded on different principles (Table 3.1).

Some of the proposed methods reduce EFM based solely on structural information of the metabolic network. De Figueiredo *et al.* 2009 presented a method to enumerate the EFM in increasing order of number of reactions. This approach enabled to identify the K-shortest EFM in *E. coli* and *Corynebacterium glutamicum* metabolic networks, which are in principle energetically more efficient. Song and Ramkrishna 2009 proposed a reduction algorithm based on the effect of EFM on the convex hull volume. This allowed the a priori reduction, without any experimental data, from the initial 369 to 35 EFM for a yeast metabolic network fermenting both glucose and xylose.

Table 3.1 Classification of methods for EFM reduction.

Principle	Method	Data required	Ref.
Network connectivity and stoichiometry	**K-shortest EFM:** Enumerates the EFM in increasing order of number of reactions. **Yield Analysis**: Excludes EFM with negligible contribution to convex hull in yield space.	Parameter free	De Figueiredo *et al.* 2009, Song and Ramkrishna 2009.
Thermodynamics	**Fractional contributions of EFM:** Estimates the EFM Coefficients based on calculated EFM thermodynamic properties**.** **Maximum Entropy Principle:** Calculates the EFM Coefficient by maximizing Shannon's entropy, which is an indirect measure of system complexity.	Thermodynamic data	Wlaschin *et al.* 2006, Zhao and Kurata 2009.
(Non)linear programming	**α-spectrum:** Uses linear optimization to maximize and minimize the weightings of each metabolic pathway that produces steady state flux distributions. **Flux regulation coefficients**: Estimates the EFM coefficients that optimize a given performance function (e.g. minimum error in flux or yield prediction). **Quadratic program:** Calculates the weights for a large set of EFM by using quadratic program to reconstruct flux distributions from subsets of EFM.	'-omics' data can be used to shrink the α-spectrum. Fluxomics and possibly other omic datasets	Wiback *et al.* 2003, Wiback *et al.* 2004, Llaneras and Pico 2007, Nookaew *et al.* 2007, Wang *et al.* 2007.
Enzyme kinetics	**Quantitative elementary mode analysis of metabolic pathways:** Combines structural and kinetic modeling to assess the effect of changes in enzyme kinetics on the usage of EFM.	Enzyme kinetic parameters	Schwartz and Kanehisa 2006.

EFM can also be discriminated on the basis of reaction thermodynamics. Wlaschin *et al.* 2006 demonstrated with experimentally determined intracellular fluxes that EFM weights are inversely correlated with the entropy generated by the involved metabolic reactions. This suggests that evolution induced cellular regulatory patterns to favor efficient pathways with low entropy generation. Zhao and Kurata 2009 proposed a method for correlating enzyme activity and flux distribution which uses the Shannon's maximum entropy principle, a measure of system complexity, as an objective function to estimate the enzyme control flux.

Several methods have been proposed that merge linear programming and experimental data. Palsson and coauthors (Wiback *et al.* 2003, Wiback *et al.* 2004) suggested linear optimization to determine how extreme pathways (the systemically independent subset of EFM) contribute to a given (measured) steady-state flux distribution. There is a range of possible nonnegative weighting values associated to extreme pathways that produce a given steady-state flux distribution. This range was calculated by maximizing and minimizing the extreme pathway weighting factors, resulting in the so called a-spectrum. Wang *et al.* 2007 presented a method to calculate the EFM coefficients for a large set of EFM by devising a quadratic program to explore the possibility and performance of using a subset of the EFM to reconstruct flux distributions. Alternatively, a framework based on EFM analysis and the convex properties of EFM was developed to calculate EFM flux regulation coefficients (FRC) corresponding to an appropriate fractional operation of this mode within the complete set of EFM (Nookaew *et al.* 2007). Schwartz and Kanehisa 2006 showed that a combination of structural and kinetic modeling in yeast glycolysis significantly constraints the range of possible behaviors of a metabolic system.

All EFM are not equal contributors to physiological cellular states, and this approach may open a direction towards a broader identification of physiologically relevant EFM among the very large number of stoichiometrically possible modes.

Then is study in detail the computational algorithm to reduce EFM based on the degree of correlation of EFM weighting factors with measured envirome factors, which it call Projection to Latent Pathways (PLP). The underlying principles are:

(i) only a moderate number of EFM are active at given environmental conditions;
(ii) the envirome plays a critical role in their regulation;
(iii) active EFM deliver a characteristic environmental footprint that can be used for their identification.

In what follows it is presented all mathematical details underlying PLP and compares it with PLS in relation to a case study.

3.2 Materials and methods

3.2.1 Projection to Latent Pathways (PLP) algorithm – problem statement

By applying steady-state material balance equations to a metabolic network with m metabolites and q metabolic reactions, the following system of linear algebraic equations is obtained:

$$\mathbf{N} \cdot \mathbf{r} = 0 \tag{3.1}$$

$$\mathbf{r_k} > 0 \tag{3.2}$$

with $\mathbf{r}$ a vector of q metabolic fluxes, $\mathbf{r_k}$ the subset of fluxes associated to irreversible reactions and $\mathbf{N}$ a $m \times q$ stoichiometric matrix. It is a well-known property of system (Eq. 3.1) that its null space solution takes the form of a polyhedral cone (Schuster and Hilgetag 1994). Furthermore, the convex basis of system (Eq. 3.1) is formed by a large number of base vectors, which are the elementary flux modes studied in this paper:

$$\mathbf{r} = \lambda_1 \mathbf{em}_1 + \lambda_2 \mathbf{em}_2 + \ldots + \lambda_K \mathbf{em}_K = \cdot \sum_{i=1}^{K} \lambda_i \times em_i \tag{3.3}$$

with $\mathbf{em_i}$ a $q \times 1$ vector of reaction weight factors that define elementary flux mode i and λ_i a scalar variable defining the partial contribution of $\mathbf{em_i}$ to the overall flux phenotype, $\mathbf{r}$, and $\mathbf{n_{em}}$ the number of EFM.

In this paper we study the reduction of EFM on the basis of dynamic envirome data sets. The basic premise is that measured fluxome vectors can be systematically deconvoluted into genetic dependent factors (the structure of elementary modes, $\mathbf{em_i}$) and envirome dependent factors (the partial contribution of each elementary mode to flux phenotype, λ_i).

To implement this method, we developed a discrimination algorithm that works according to the following criteria:

(1) Maximization of captured variance of flux data sets, $\mathbf{R} = \{\mathbf{r}(t)\}$;

(2) Maximization of correlation between elementary cellular functions weighting factors λ_i against envirome data (medium factors), $\mathbf{X} = \{\mathbf{x}(t)\}$;

(3) Minimization of the redundancy, i.e. minimization of the number of active EFM (elimination all EFM with weak correlation with the envirome).

with $\mathbf{X} = \{\mathbf{x}(t)\}$ a $np \times nx$ matrix of np independent observations of envirome vectors $\mathbf{x}(t)$ $(\dim(\mathbf{x}) = nx)$, $\mathbf{R} = \{\mathbf{r}(t)\}$ a $np \times nr$ matrix of np independent observations of reaction rates, $\mathbf{r}(t)$ $(\dim(\mathbf{r}) = q)$. These criteria are equivalent to a covariance maximization problem (covariance maximization implies maximization of correlation and minimization of redundancy) between envirome data, $\mathbf{X}$, and observed flux data, $\mathbf{R}$, under the constraint of a plausible set of EFM:

$$\begin{aligned} &\textit{Maximize} && \operatorname{cov}(\mathbf{X}, \mathbf{R}) \\ &\text{s.t.} && \begin{cases} \mathbf{R} = \mathbf{\Lambda} \times \mathbf{EM}^{\mathrm{T}} \\ \mathbf{\Lambda} = \mathbf{X} \times \mathbf{C}^{T} \end{cases} \end{aligned} \tag{3.4}$$

with $\mathbf{EM} = \{\mathbf{em}_i\}$ a $nr \times nem$ matrix of nem elementary flux modes, $\mathbf{em}_i$ $(\dim(\mathrm{em}_i) = q)$, $\mathbf{\Lambda} = \{\boldsymbol{\lambda}(t)\}$ a $np \times nem$ matrix of weight vectors $\boldsymbol{\lambda}(t)$ of elementary flux modes $(\dim(\boldsymbol{\lambda}) = nem)$ and $\mathbf{C}$ a $nem \times nx$ matrix of regression coefficients.

Unconstrained maximization of covariance can be performed by the popular method Projection to Latent Structures (PLS), also known as partial least squares. Figure 3.1 shows the structural differences between PLS and PLP. Since PLP is derived from PLS, in the lines below we first review PLS decomposition and then show how it can be extended to PLP.

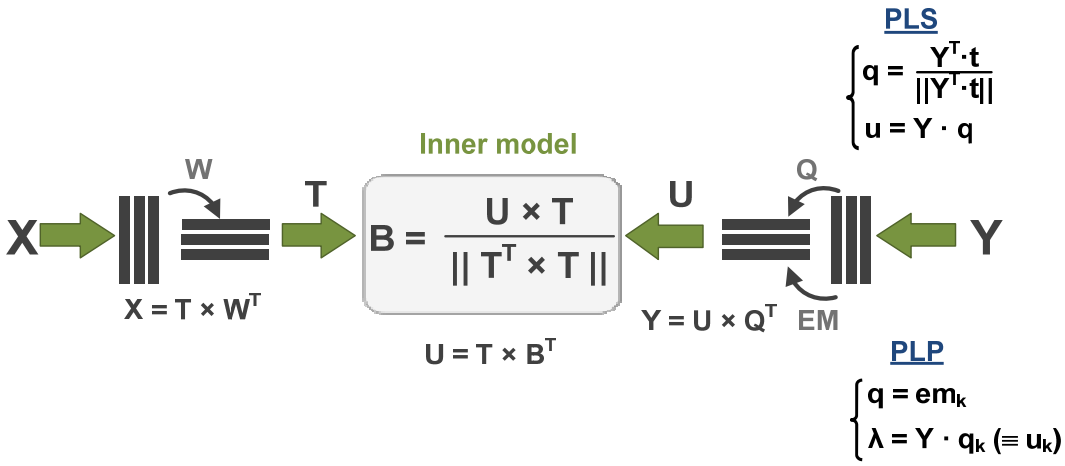


Figure 3.1 Schematic representations of decomposition operations performed by PLS and PLP algorithms. The main differences between PLS and PLP are related to the computation of Y-loadings. In PLS Q are abstract variables calculated to maximize correlation between X and Y, while in PLP Q comprises a subset of active EFM.

3.2.2 Projection to Latent Structures (PLS) algorithm

PLS is a multivariate linear regression technique between an input (predictor) matrix, $\mathbf{X}$, and an output response matrix, $\mathbf{Y}$. It differs from traditional multivariate linear regression in that it decomposes both the predictor and the response matrices into reduced sets of uncorrelated latent variables, which are then linearly regressed against each other. The most popular PLS algorithm is the NIPALS (non-iterative partial least squares) algorithm (Wold 1975), which provides the basis for PLP derivation. NIPALS proceeds according to the following steps:

1) Set the initial $^{ny\times 1}$ $\mathbf{Y}$-loading vector, $\mathbf{q}$, equal to an arbitrarily chosen nonzero row of $\mathbf{Y}$, $\mathbf{y}_t$

$$\mathbf{q} = \frac{\mathbf{y}_t^T}{\|\mathbf{y}_t\|} \tag{3.5}$$

in case of univariate PLS, $^{ny=1}$ and $\mathbf{q}=1$.

2) Compute the $^{np\times 1}$ $\mathbf{Y}$-score vector, $\mathbf{u}$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{q} \tag{3.6}$$

3) Compute the $nx\times 1$ weight vector, $\mathbf{w}$

$$\mathbf{w} = \frac{\mathbf{X}^{\mathrm{T}}\cdot\mathbf{u}}{\left\|\mathbf{X}^{\mathrm{T}}\cdot\mathbf{u}\right\|}, \tag{3.7}$$

4) Compute the $np\times 1$ $\mathbf{X}$-score vector, $\mathbf{t}$

$$\mathbf{t} = \mathbf{X}\cdot\mathbf{w} \tag{3.8}$$

5) Recalculate the $\mathbf{Y}$-loading vector, $\mathbf{q}$

$$\mathbf{q} = \frac{\mathbf{Y}^{\mathrm{T}}\cdot\mathbf{t}}{\left\|\mathbf{Y}^{\mathrm{T}}\cdot\mathbf{t}\right\|} \tag{3.9}$$

6) Repeat steps 1-5 until the convergence criterion $\left\|\mathbf{t}-\mathbf{t}_{\mathbf{old}}\right\| < eps$ is obeyed with, for instance, $eps = 1\times 10^{-8}$. In case of univariate PLS, Eq. 3.9 yields $\mathrm{q} = 1$ hence no iterations are performed.

7) Compute the $\mathbf{X}$ data block loadings, $\mathbf{p}$, and rescale accordingly:

$$\mathbf{p} = \frac{\mathbf{X}^{\mathrm{T}}\cdot\mathbf{t}}{\left\|\mathbf{t}^{\mathrm{T}}\cdot\mathbf{t}\right\|} \tag{3.10}$$

$$\mathbf{p}_{\mathbf{new}} = \frac{\mathbf{p}}{\left\|\mathbf{p}\right\|} \tag{3.11}$$

$$\mathbf{t} = \mathbf{t}\cdot\left\|\mathbf{p}\right\| \tag{3.12}$$

$$\mathbf{w} = \mathbf{w}\cdot\left\|\mathbf{p}\right\| \tag{3.13}$$

8) Compute the regression coefficient of the inner linear model

$$\mathbf{b} = \frac{\mathbf{u}^{\mathrm{T}}\cdot\mathbf{t}}{\mathbf{t}^{\mathrm{T}}\cdot\mathbf{t}} \tag{3.14}$$

9) Compute the $\mathbf{X}$ and $\mathbf{Y}$ residuals

$$\mathbf{E}_{\mathbf{X}} = \mathbf{X}-\mathbf{t}\cdot\mathbf{p}^{\mathrm{T}} \tag{3.15}$$

$$\mathbf{E}_{\mathbf{Y}} = \mathbf{Y}-\mathbf{b}\cdot\mathbf{t}\cdot\mathbf{p}^{\mathrm{T}} \tag{3.16}$$

10) Then go back to step 1 and repeat the procedure for the next latent variable after making

$$\mathbf{X} = \mathbf{E_X} \tag{3.17}$$

$$\mathbf{Y} = \mathbf{E_Y} \tag{3.18}$$

Steps 1-10 are repeated for $k = 1,...,\mathrm{Fac}$ latent variables resulting into the following overall decomposition:

$$\mathbf{X} = \mathbf{T}\cdot\mathbf{W}^{\mathrm{T}} + \mathbf{E_X} \tag{3.19}$$

$$\mathbf{Y} = \mathbf{U}\cdot\mathbf{Q}^{\mathrm{T}} + \mathbf{E_Y} \tag{3.20}$$

$$\mathbf{U} = \mathbf{T}\cdot\mathbf{B}^{\mathrm{T}} + \mathbf{E_U} \tag{3.21}$$

with $\mathbf{E}_i$ residuals matrices. Finally, the prediction of $\mathbf{Y}$ from $\mathbf{X}$ is given by

$$\hat{\mathbf{Y}} = \mathbf{X}\cdot\mathbf{RC}^{T} \tag{3.22}$$

with $\mathbf{RC}$ the $ny \times nx$ regression coefficients matrix given by

$$\mathbf{RC = B.W^T} \tag{3.23}$$

Solving system Eq. 3.19, Eq. 3.20 and Eq. 3.21, which imply minimizing residuals E_X, E_Y and E_U, can be effectively performed by the NIPALS algorithm. For more details about NIPALS and PLS see, for instance, Geladi and Kowalski 1986.

3.2.3 Projection to Latent Pathways (PLP) algorithm

PLP can be viewed as a constrained version of PLS that maximizes the covariance between X and R under the constraint of known EFM. PLP performs essentially the same decomposition described by Eqs. 3.19 - 3.23. The main difference resides in the computation of the output loadings, $\mathbf{Q}$. Since EFM are unique and non-decomposable fluxome solutions, any observed flux distribution can be expressed as a non-negative weighted sum of EFM (Eq. 3.3). Thus, EFM $\mathbf{em_i}$ can be interpreted as latent variables (or principle components of a metabolic network)

while the weights $\boldsymbol{\lambda_i}$ can be interpreted as score values of such latent variables. According to this analogy, PLS was modified as follows:

1) For each elementary flux mode k, set the loadings equal to $\mathbf{em_k}$ and compute the respective score vector, $\boldsymbol{\lambda_k}$:

$$\mathbf{q_k} = \mathbf{em_k} \tag{3.24}$$

$$\boldsymbol{\lambda_k} = \mathbf{R} \cdot \mathbf{q_k} \left(\equiv \mathbf{u_k} \right) \tag{3.25}$$

2) Perform a univariate PLS (with $\mathbf{q} = 1$) with input $\mathbf{X}$ and target $\mathbf{Y} = \boldsymbol{\lambda_k}$ for Fac latent variables as described in the previous section and compute the predicted $\boldsymbol{\lambda_k}$

$\hat{\boldsymbol{\lambda}}_\mathbf{k}$ predicted $\boldsymbol{\lambda_k}$ from univariate PLS (3.26)

3) Compute the predicted $\mathbf{R}$ by the k elementary mode and the respective explained variance

$$\hat{\mathbf{R}}_\mathbf{k} = \hat{\boldsymbol{\lambda}}_\mathbf{k} \cdot \mathbf{q}_\mathbf{k}^\mathrm{T} \tag{3.27}$$

$$var_k(\%) = 100 \cdot \left(1 - \frac{(\mathrm{R} - \hat{\mathrm{R}}_\mathrm{K})^T . (\mathrm{R} - \hat{\mathrm{R}}_\mathrm{K})}{R^T . R} \right) \tag{3.28}$$

4) Repeat steps 1-3 for every EFM $k = 1, \ldots, nem$ and choose the best, $kopt$, as the one that exhibits the highest variance value given by Eq. 3.28.

$kopt$: EFM with highest var_k value (3.29)

5) Remove $kopt$ from the list of EFM and make

$$\mathbf{R} = \mathbf{R} - \hat{\mathbf{R}}_\mathbf{kopt} \tag{3.30}$$

6) Go back to step 1 and repeat the procedure for a maximum number of elementary modes or until the explained variance of $\mathbf{R}$ does not increase any further.

With this procedure the PLS output loadings, $\mathbf{Q}$, hold a subset of EFM from matrix $\mathbf{EM}$ while the PLS output scores, $\mathbf{U}$, are equivalent to the elementary modes weights matrix, $\mathbf{\Lambda}$:

$$\mathbf{R} = \mathbf{\Lambda} \times \mathbf{EM}^{\mathrm{T}} + \mathbf{E_R} \quad . \tag{3.31}$$

An advantage of PLP over PLS is that both the latent variables and score values of the target matrix have in PLP a physical meaning. The latent variables are latent pathways while scores U are equivalent to Λ, i.e. they represent the relative weighting factors of latent pathways. For this reason, the algorithm is called Projection to Latent Pathways. The physical interpretation of the variables in PLP is:

1) The number of latent variables in PLS is analogous to the number of active EFM in PLP. Thus the subset of EFM that explain most of the variance of $\mathbf{R}$ are interpreted as the set of metabolic pathways activated by environmental factors.
2) The regression coefficients vector, $\mathbf{RC_{kopt}}$, of the inner univariate PLS, being directly associated with EFM, show the contribution of each environmental factor to the up- or down- regulation of EFM.

The PLS and PLP algorithms were coded in Matlab™ (Mathworks, Inc) as a modified version of the NIPALS algorithm, wherein the loadings of the outputs are fixed a priori according to the EFM structure. The calculation of the EFM is not automatically integrated in PLP. For that it was used the METATOOL 5.0 (von Kamp and Schuster 2006).

3.3 Results and discussion

3.3.1 Case study: recombinant BHK cell line

We have first tested the PLP algorithm with the data of a mammalian cell line published elsewhere (Teixeira *et al.* 2011). Data of a recombinant Baby Hamster Kidney (BHK) cell line expressing a fusion glycoprotein IgG1-IL2 was used to compare PLS and PLP. The data set comprises 134 observations acquired from 7 independent bioreactor experiments operated in batch and fed-batch modes.

The predictor matrix, $\mathbf{X}$ $(\dim(\mathbf{X}) = 134 \times 26)$, includes measured data of 26 environmental factors (pH, osmolarity and concentrations of viable cells, glucose, lactate, ammonia, IgG1-IL2 and 19 aminoacids) while the target matrix, $\mathbf{R}$ $(\dim(\mathbf{R}) = 134 \times 24)$, comprises 24 production or consumption fluxes of extracellular compounds.

A relatively small BHK metabolic network comprising 35 metabolites and 57 metabolic reactions was constructed. Its EFM were computed using METATOOL 5.0 resulting in 251 EFM. These 251 EFM were used as constraints to PLP decomposition.

Comparing PLP and PLS decomposition results

The full data set was divided into two partitions of randomly selected points with equal size for calibration and validation (with 67 points each). The results of a single run of PLS and PLP decomposition for the calibration data set are shown in Tables 3.2 and 3.3 respectively. PLS decomposition stops at latent variable 18, when the $\mathbf{X}$ variance reaches 100 %. The final explained $\mathbf{R}$ variance is 90.1 %. As for PLP, decomposition progresses up to the 17th EFM explaining 82.5 % of $\mathbf{R}$ variance, thus 7.5 % less than PLS. PLP decomposition stops when the threshold degree of correlation between λ_i and $\mathbf{X}$ can no longer be satisfied ($r^2 > 0.75$ and p-value < 0.05, see Table 3.3). This procedure ensures that the identified EFM are the ones with highest correlation with environmental state.

Table 3.2 PLS decomposition results in terms of % of explained variance (Var) over number of latent variables (LV). Var(**X**) and Var(**R**) are % of explained variance of envirome and fluxome data, respectively.

# Lv	Var X (%)	Var R (%)
1	48.9	32.4
2	59.6	51.8
3	79.0	58.0
4	84.6	64.3
5	89.8	67.4
6	92.2	70.9
7	94.5	74.1
8	96.2	76.4
9	97.7	78.6
10	98.3	82.1
11	98.9	83.0
12	99.4	84.1
13	99.6	85.8
14	99.8	86.7
15	99.9	87.9
16	99.9	89.0
17	99.9	89.6
18	100.0	90.1

Table 3.3 PLP decomposition results showing the subset of EFM with highest correlation with the envirome (as denoted by the r2 and p-value). Var(**λ**) and Var(**R**) are % of explained variance of EFM weighting factors and fluxome data, respectively.

EFM	# LV	r2	p-value	Var(λ)	Var(R)
179	4	0.95	1.14E-32	88.90	52.60
1	4	0.89	5.16E-23	79.90	57.30
210	4	0.87	1.56E-20	65.90	57.80
173	4	0.82	8.78E-17	62.40	58.30
116	4	0.82	2.49E-16	58.40	58.70
139	4	0.86	1.34E-19	52.50	58.90
206	4	0.92	2.14E-27	73.90	60.20
143	4	0.86	9.71E-20	66.70	60.60
69	4	0.82	1.04E-16	57.30	61.00
72	4	0.84	3.52E-18	57.80	61.30
4	4	0.92	1.96E-27	81.40	64.10
68	4	0.81	3.96E-16	60.20	64.80
11	4	0.91	4.16E-25	76.80	79.20
6	4	0.94	1.99E-30	84.60	80.90
7	4	0.82	1.72E-16	59.10	81.60
12	4	0.83	1.52E-17	58.30	82.10
2	4	0.85	7.26E-19	71.00	82.50

Figure 3.2 depicts predicted against "measured" λ_i illustrating the high degree of correlation with envirome variables for the discriminated set of EFM.

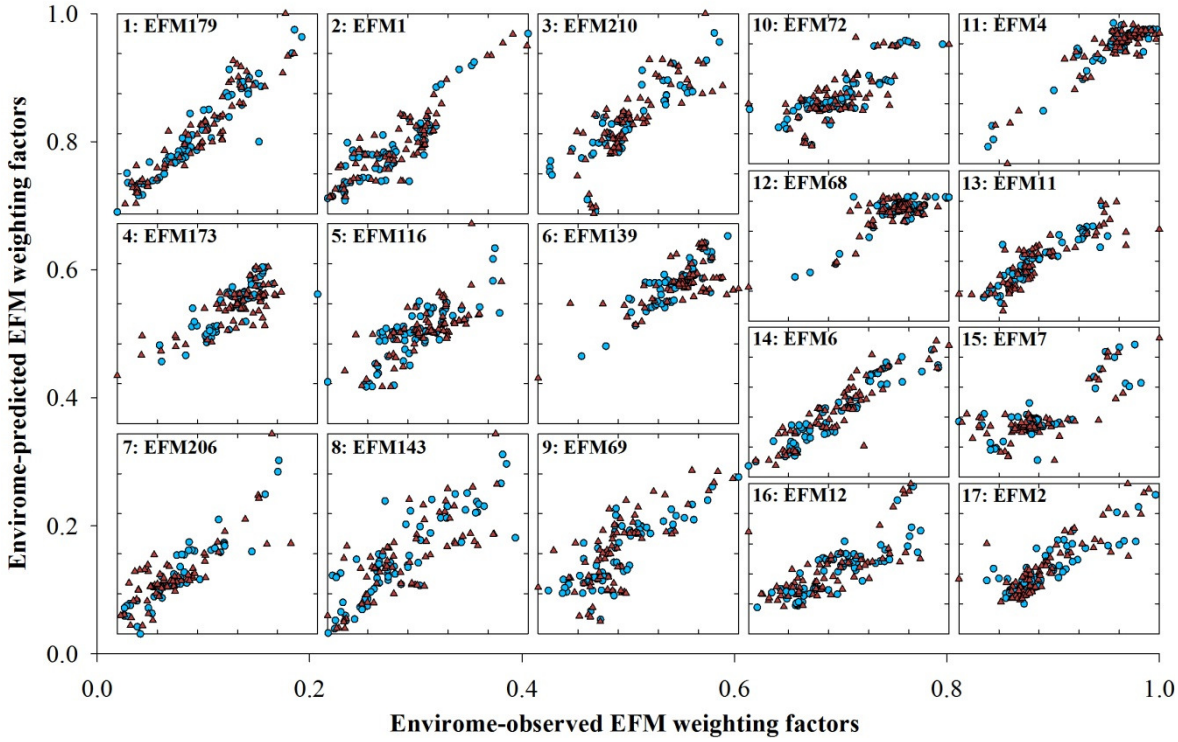


Figure 3.2 EFMs discrimination results. Observed and envirome-predicted weighting factors of discriminated EFMs by the PLP algorithm applied to the BHK data set. Blue circles and red triangles represent the calibration and validation data points, respectively, with 67 data points each.

Assessment of EFM reduction consistency

PLS belongs to a class of multivariate regression techniques that can be used to model high dimensional data sets with low number of sampling points (Boulesteix and Strimmer 2007). However, when the number of samples is too low, the partitioning into calibration and validation sets may have a high impact on the final model structure. Since stemming from PLS, the same problem does in principle apply to PLP. In order to assess EFM discrimination variability due to data partitioning, a bootstrapping technique was implemented, in which PLP and PLS were repeated 200 times with randomly selected calibration and validation partitions with 67 points each. Figure 3.3 shows the frequency of selection of EFM resulting from the bootstrapping analysis. These results evidence a subset of frequently selected EFM, which include EFM1, EFM2, EFM4, EFM6, EFM11,

EFM179 and EFM210 with frequency of selection higher than 75 % and EFM69, EFM72, EFM173 and EFM206 with frequency of selection higher than 50 %. Less frequently selected EFMs are very sensitive to the data partitioning and to experimental noise and thus less reliable to interpret.

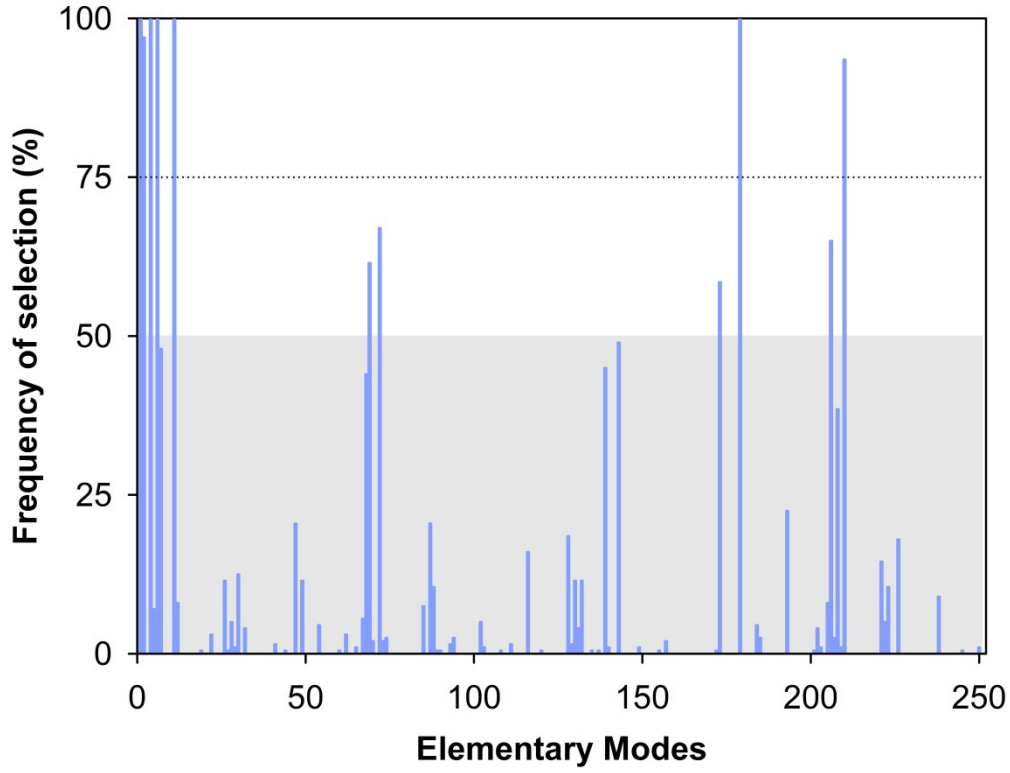


Figure 3.3 Frequency of selection of EFM. A bootstrapping technique was implemented in which 200 PLP runs are performed for randomly selected calibration and validation data sets with 67 points each. Frequency is calculated as the EFM selection count divided by the total number of runs.

Regression coefficients

While PLS regression coefficients are associated with latent variables lacking physical meaning, PLP regression coefficients are directly associated to the discriminated EFM (see Figure 3.4). Thus they provide information of how the envirome up- or down-regulates each EFM. This interpretation should however be done with care as regression coefficients cannot disclose between a cause and an effect. An EFM is per definition a non-decomposable sub-network. Most of them start and end in extracellular compounds. Each EFM produces a characteristic dynamic footprint in the environment in terms of consumed or produced metabolites, which is more an effect rather than a cause. Moreover, it is an

important feature of PLS and per inheritance of PLP that the X-loadings are computed in a way to maximize predictive power of Y in detriment of interpretability of the individual contribution of X variables. Although many papers have attempted to develop interpretation of PLS regression coefficients (de Alwis *et al.* 2007, Selvarasu *et al.* 2010), other techniques are in principle better suited for this purpose. Even so, main causal effects can be extracted from the analysis of regression coefficients. For this analysis it is however important to calculate the confidence intervals of the regression coefficients, which can be obtained from the previously described bootstrapping technique (Faber 2002). From the z = 200 PLP runs with randomly selected calibration and validation data sets, z = 200 vectors of regression coefficients are calculated. The respective mean and standard deviation can be estimated as follows:

$$\overline{\mathbf{B}} = \frac{\sum_{i=1}^{z} \mathbf{B}_i}{z} \tag{3.32}$$

$$\mathbf{S} = \sqrt{\frac{1}{z-1}\sum_{i=1}^{z} (\mathbf{B}_i - \overline{\mathbf{B}}_i)^2} \tag{3.33}$$

The 95 % confidence intervals can then be calculated from the t-student distribution with 0.975 half interval and z-Fac degrees of freedom

$$\mathbf{B} = \overline{\mathbf{B}} \pm \mathbf{S} \times \mathrm{t}_{0.975,\mathrm{z-EFM}} \tag{3.34}$$

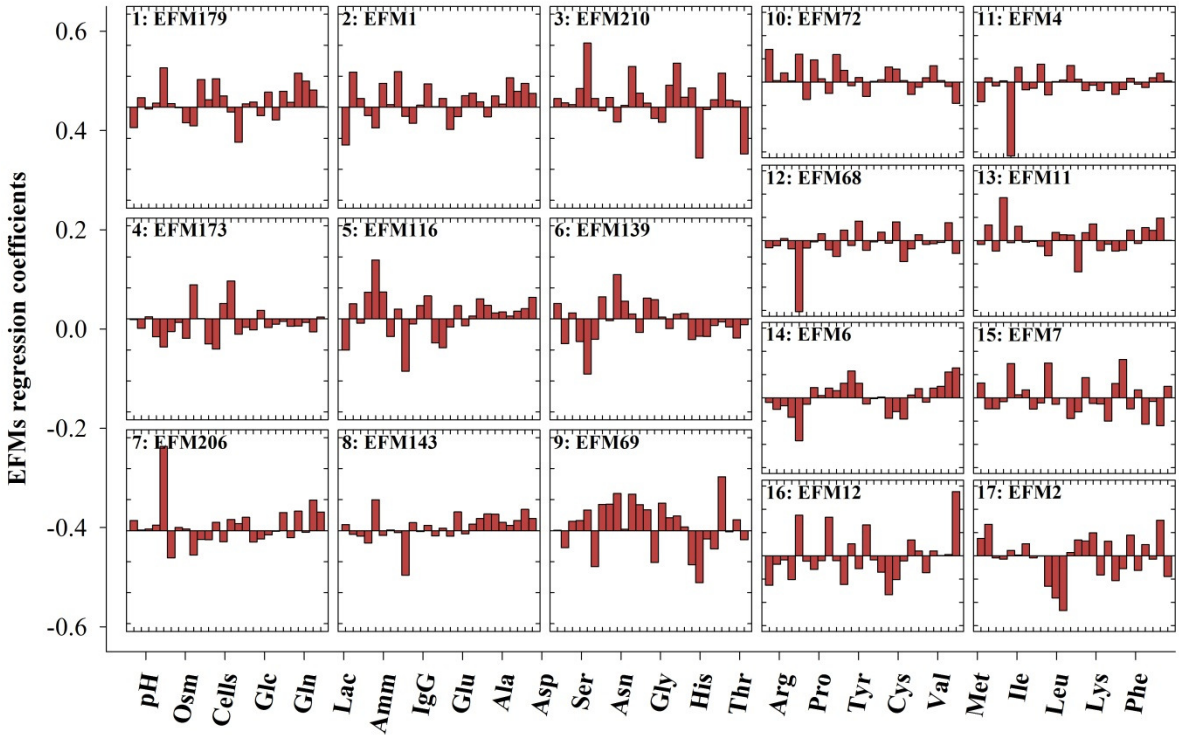


Figure 3.4 PLP regression coefficients. Regression coefficients of selected EFM quantify the contribution of each environmental factor in X to the respective EFM weighting factor.

As illustrative example, Figure 3.5 plots the confidence interval against the mean of the regression coefficients for the product formation EFM (EFM 1). It can be observed that only a subset of regression coefficients lay below the one half threshold line. These include the regression coefficients associated with pH, osmolality, glutamine, lactate, IgG, valine and lysine. These regression coefficients are the most statistically significant and thus more reliable interpretations can be withdrawn from them. As example, it is a rational result that the weighting factor of the product EFM 1 is highly correlated with the product concentration since the product results from EFM 1. All other identified environmental parameters are potential targets for manipulation in order to improve product synthesis. This analysis can be systematically extended to the full set of envirome components and full set of EFMs to support the concept of cell functional enviromics as defended in Teixeira *et al.* 2011.

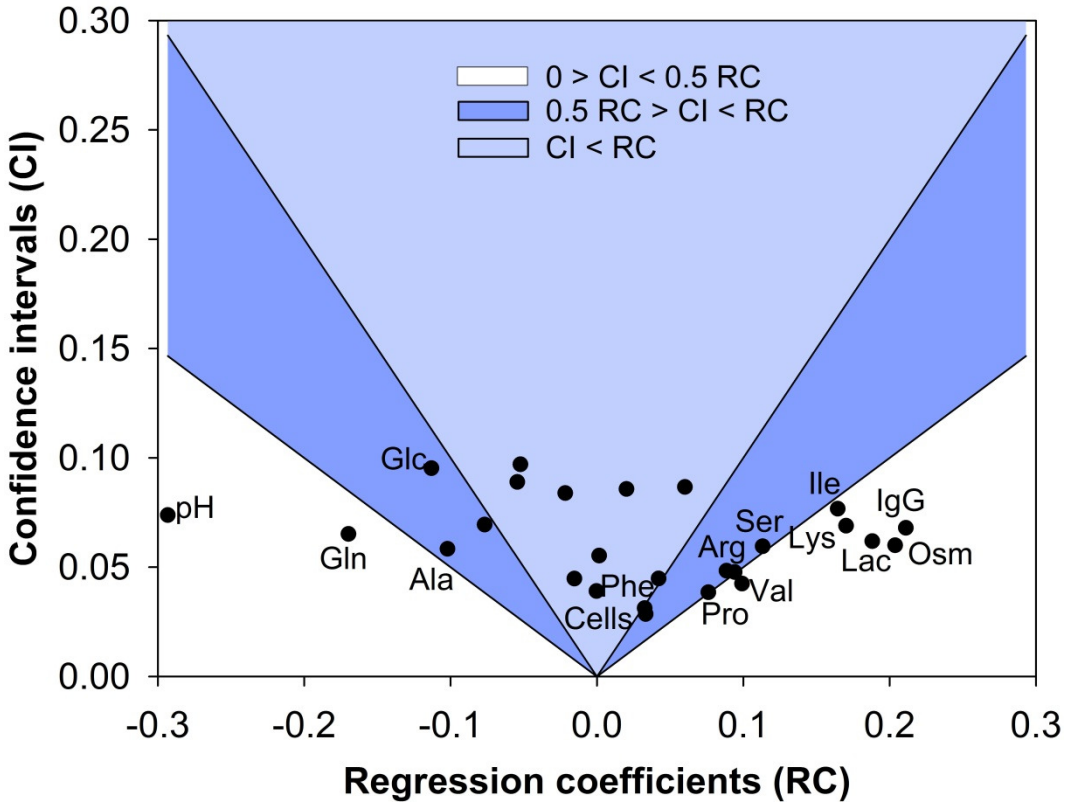


Figure 3.5 Regression coefficients confidence intervals for EFM 1. Confidence interval as function of regression coefficients obtained for the product formation EFM (EFM 1). Black full circles are envirome factors. The light and dark blue regions correspond to confidence intervals higher than 50% and 100% of the nominal value of the regression coefficient, respectively.

Predictive power

To test the predictive power, PLS and PLP models were calibrated with the calibration data set composed by 50 % of data points and then simulated on the validation data set composed by the remaining 50 % measured points. The PLS model with 18 latent variables explained 90.1 % of **R** variance in the calibration dataset but only 76.8 % of the validation dataset. The quality of the results can be visually inspected in Figure 3.6.

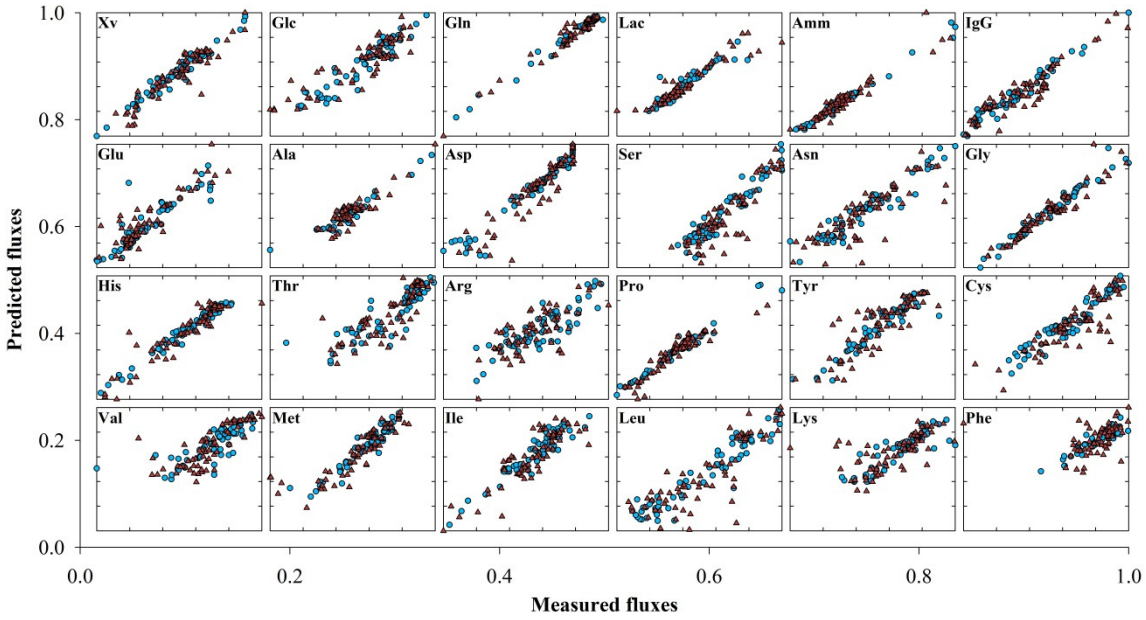


Figure 3.6 Predicted metabolic fluxes by PLS. Predicted against measured fluxes computed by the PLS model for the BHK data set. Blue circles and red triangles represent the calibration and validation data points, respectively.

The degradation of accuracy in the validation dataset is rational given that the model is requested to predict data points, which may lay outside of the domain of experience defined by the calibration data set. As for PLP it is a very interesting result to verify that despite explaining a lower variance in the calibration data set (83.2 % against 90.1 % for PLP and PLS respectively), the accuracy of the validation data set was higher than that of PLS (81.9 % against 76.8 % for PLP and PLS respectively). Moreover, the variance of the validation data set is almost equal to that of the calibration data set, denoting a more consistent model, with higher predictive power than the PLS one (Figure 3.7).

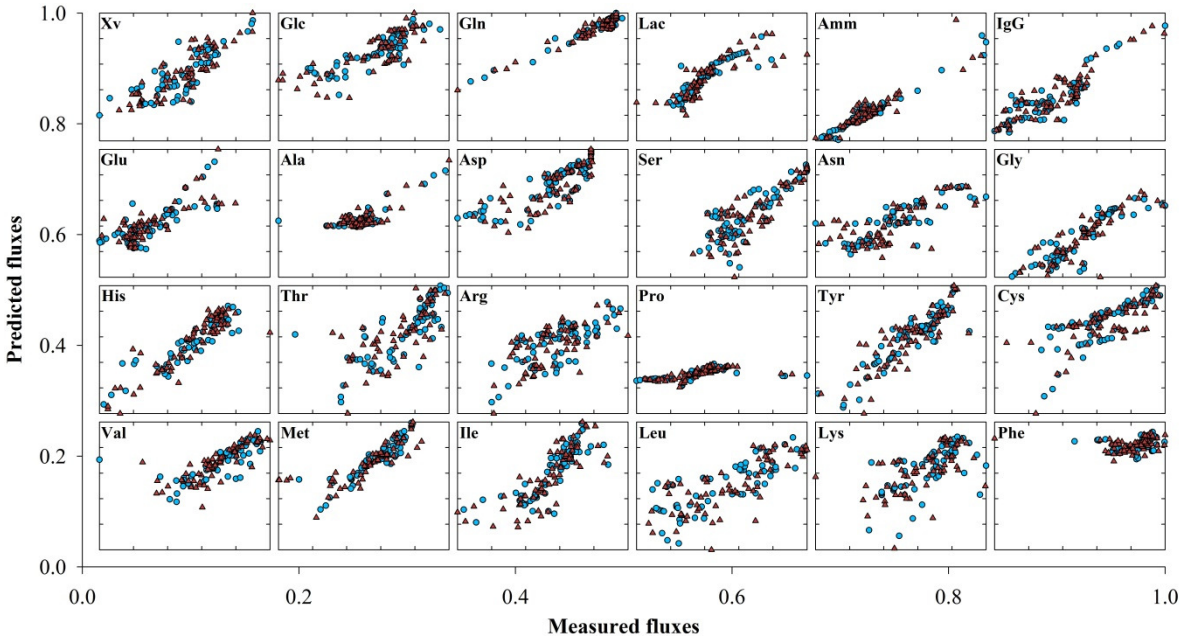


Figure 3.7 Predicted metabolic fluxes by PLP. Predicted against measured fluxes computed by the PLP model for the BHK data set. Blue circles and red triangles represent the calibration and validation data points, respectively.

In order to screen out the possibility of a casual better performance of PLP in relation to PLS due to the particular data partitioning employed, the same variance analysis was performed for the z = 200 PLP and PLS runs performed with randomly selected calibration and validation data points according to the bootstrapping technique previously described. The results show that the explained variance of the validation data set varied between 78.8 - 85.6 % for PLP and 50.4 - 82.7 % for PLS. In 194 out of 200 runs the PLP outperformed the PLS, thus confirming that while PLS is consistently more accurate in describing the calibration data than PLP, the latter is consistently more accurate at predicting the validation data than PLS.

Discussion

The key PLS feature is identifying independent **X** and **Y** scores so that the relationship between successive pairs of scores is as strong as possible. PLS may be thus viewed as a robust form of redundancy analysis, seeking directions in the factor space that are associated with high variation in the response **Y** but biasing them toward directions that are more accurately predicted.

Due to its advantages in handling highly redundant data sets, PLS has become a widely used regression analysis technique in systems biology. It has been applied as an inference tool for predicting metabolic fluxes using isotopomer flux data (Antoniewicz *et al.* 2006), analyzing genomic and proteomic data (Boulesteix and Strimmer 2007), identifying signaling networks by inducing cellular response to different stimuli (Ivakhno and Armstrong 2007, Janes *et al.* 2005, Miller-Jensen *et al.* 2007) and network structure using metabolomics data (Bundy *et al.* 2007). Moreover, PLS has also been applied for the identification of active cellular pathways as a function of the environment using metabolic and gene expression profiles (Li and Chan 2004), detection of gene-gene interactions from microarrays data (Pihur and Datta 2008, Tenenhaus *et al.* 2010) and CM optimization using nutritional profiling data (de Alwis *et al.* 2007, Selvarasu *et al.* 2010).

The main disadvantage of PLS lies in its empirical datadriven nature with limited added-value in terms of mechanistic knowledge generation. Although carrying some internal structure, this structure is not inspired by any a priori mechanistic knowledge of the system. PLP may be viewed as a constrained version of PLS, attuned to the structure of the biological system under study. While in PLS the loadings and score are abstract variables, in PLP loadings and scores refer to well defined metabolic structures. Specifically, PLP explores EFM as "principle components" of a metabolic network. Indeed, EFM obey to the principle of non-decomposability, meaning that any particular flux distribution can be expressed as a nonnegative weighted sum of EFM. Thus the ranking obtained in PLP refers to active pathways as inferred by their level of correlation with the environmental

state. In terms of data requirement, PLS belongs to the class of multivariate regression techniques particularly suitable to handle highly dimensional data sets even if the number of observations is limited (Boulesteix and Strimmer 2007).

PLS is typically used to model spectral data such as near infrared or 2D-fluorescence maps (Teixeira *et al.* 2009). A basic requirement is that the number of latent variables must be lower than the number of observations in the calibration data set. This means that reliable linear models can be identified from a moderate number of observations of highly dimensional datasets. The same properties apply to PLP. A basic constraint is that the number of discriminated EFM cannot be higher than the number of observations in the calibration data set. However the method offers no restriction in terms of the dimensionality of the input data set.

Finally it should be commented on the computational power requirements, which scales linearly with the number of EFM. In the present study with 251 EFM, computation requirements are in the order of seconds in a common PC. For a genome scale network with several million of EFM, computation power might easily rise to the scale of days in a common PC.

3.4 Conclusion

In this thesis an algorithm for the discrimination of active EFM on the basis of dynamical envirome data called projection to latent pathways (PLP) was developed. The algorithm is designed to maximize the covariance between envirome data and observed flux data under the constraint of universe of genes translated into a plausible set of EFM. In general lines, the algorithm discriminates a minimal set of envirome correlated EFM that maximize the variance of measured flux data.

Thus the algorithm may be viewed as a reverse, envirome-to-function metabolic reconstruction methodology as opposed to the generally accepted gene-to-function reconstruction approach. Although presented here as a method to analyze envirome data sets, PLP has broader scope. It is rather a general methodology for statistical elimination of redundant metabolic structures that, in a broader sense, has the potential to bring together all layers of 'omic' information under a common computational framework.

Bibliography

Antoniewicz MR, Stephanopoulos G, Kelleher JK. 2006. Evaluation of regression models in metabolic physiology: predicting fluxes from isotopic data without knowledge of the pathway. Metabolomics, 2(1): 41-52.

Boulesteix AL, Strimmer K. 2007. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform, 8(1): 32-44.

Bundy JG, Papp B, Harmston R, Browne RA, Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM. 2007. Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. Genome Res, 17(4): 510-519.

de Alwis DM, Dutton RL, Scharer J, Moo-Young M. 2007. Statistical methods in media optimization for batch and fed-batch animal cell culture. Bioprocess Biosyst Eng, 30(2): 107-113.

de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ. 2009. Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics, 25(23): 3158-3165.

Faber NM. 2002. Uncertainty estimation for multivariate regression coefficients. Chemometrics Intell Lab Syst 64(2): 169-179.

Geladi P, Kowalski BR. 1986. Partial Least-Squares regression - A tutorial. Anal Chim Acta, 185: 1-17.

Ivakhno S and Armstrong JD. 2007. Non-linear dimensionality reduction of signalling networks. BMC Syst Biol, 1:17.

Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. 2005. Systems model of signaling identifies a molecular basis set for cytokineinduced apoptosis. Science, 310(5754): 1646-1653.

Klamt S and Stelling J. 2003. Two approaches for metabolic pathway analysis? Trends Biotechnol, 21(2): 64-69.

Li Z and Chan C. 2004. Integrating gene expression and metabolic profiles. J Biol Chem, 279(26): 27124-27137.

Llaneras F and Pico J. 2007. A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC Bioinformatics, 8.

Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. 2007. Common effector processing mediates cell-specific responses to stimuli. Nature, 448(7153):604-U611.

Nookaew I, Meechai A, Thammarongtham C, Laoteng K, Ruanglek V, Cheevadhanarak S, Nielsen J, Bhumiratana S. 2007. Identification of flux regulation coefficients from elementary flux modes: A systems biology tool for analysis of metabolic networks. Biotechnol Bioeng, 97(6): 1535-1549.

Palsson BO, Price ND, Papin JA. 2003. Development of network-based pathway definitions: the need to analyze real metabolic networks. Trends Biotechnol, 21(5): 195-198.

Pihur V and Datta S. 2008. Reconstruction of genetic association networks from microarray data: a partial least squares approach. Bioinformatics, 24(4): 561-568.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol, 17(2): 53-60.

Schuster S, Fell DA, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18(3): 326-332.

Schuster S and Hilgetag C. 1994. On Elementary Flux Modes in biochemical reaction systems at steady state. J Biol Syst 2(2): 165-182.

Schwartz JM, Kanehisa M. 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics, 7:20.

Selvarasu S, Kim DY, Karimi IA, Lee DY. 2010. Combined data preprocessing and multivariate statistical analysis characterizes fed-batch culture of mouse hybridoma cells for rational medium design. J Biotechnol, 150(1): 94-100.

Song HS, Ramkrishna D. 2009. Reduction of a Set of Elementary Modes Using Yield Analysis. Biotechnol Bioeng, 102(2): 554-568.

Teixeira AP, Dias JML, Carinhas N, Sousa M, Clemente JJ, Cunha AE, Stosch M von, Alves PM, Carrondo MJT, Oliveira R. 2011. Cell functional enviromics: Unravelling the function of environmental factors. Bmc Systems Biology 5.

Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. 2009. Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. Biotechnol Adv, 27(6): 726-732.

Tenenhaus A, Guillemot V, Gidrol X, Frouin V. 2010. Gene Association Networks from Microarray Data Using a Regularized Estimation of Partial Correlation Based on PLS Regression. IEEE-ACM Trans Comput Biol Bioinform, 7(2): 251-262.

Terzer M and Stelling J. 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24(19): 2229-2235.

Trinh CT, Wlaschin A, Srienc F. 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol, 81(5): 813-826.

von Kamp A and Schuster S. 2006. METATOOL 5.0: fast and flexible elementary modes analysis. Bioinformatics, 22(15):1930-1931.

Wang QZ, Yang YD, Ma HW, Zhao XM. 2007. Metabolic network properties help assign weights to elementary modes to understand physiological flux distributions. Bioinformatics, 23(9): 1049-1052.

Wiback SJ, Mahadevan R, Palsson BO. 2003. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J Theor Biol, 224(3): 313-324.

Wiback SJ, Mahadevan R, Palsson BO. 2004. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The Escherichia coli spectrum. Biotechnol Bioeng, 86(3): 317-331.

Wlaschin AP, Trinh CT, Carlson R, Srienc F. 2006. The fractional contributions of elementary modes to the metabolism of Escherichia coli and their estimation from reaction entropies. Metab Eng, 8(4): 338-352.

Wold H. 1975. Path models with latent variables: The NIPALS approach. In: Quantitative sociology: International perspectives on mathematical and statistical modeling. Edited by Blalock HM, Aganbegian A, Borodkin FM, Boudon R, Capecchi V. New York: Academic, 307-357.

Zhao QY, Kurata H. 2009. Maximum entropy decomposition of flux distribution at steady state to elementary modes. J Biosci Bioeng, 107(1): 84-89.

CHAPTER 4

Optimization of cell culture media composition by cell functional enviromics: application of *Pichia* Trace Metals (PTM1) composition

Adapted from:

Oliveira R, Dias JML, Ferreira AR. PAT 42751/10. A Functional Enviromics method for cell culture media engineering. 2010.

Abstract

This chapter presents a holistic approach for optimizing the composition of cell CM based on the concept of Functional Enviromics. Functional enviromics studies the interactions of the complete set of environmental factors on the complete set of cell functions. The basic premise is that the genome sets the universe of EFM while the environment of the cells, i.e. the medium factors, set the relative intensity of the EFM to support the phenotype of the cell. Thus the goal of the method is in a first stage to identify how medium factors control the relative intensity of EFM, i.e. functional enviromics analysis. This is accomplished through the execution of an experimental protocol, wherein an array of cell culture experiments is performed, wherein perturbations to baseline medium factors are introduced and response exometabolomic data is acquired and analyzed. The so acquired data is processed into the form of a functional enviromics map (FEM) of EFM against medium factors. In a second stage, from the FEM, optimized medium formulations are developed that either enhance or repress target EFM in order to enforce a desired phenotype. The main advantage of this method lies in enabling metabolic engineering through the CM composition manipulation, wherein an arbitrarily high number of cell functions are optimized through manipulation of medium factors, as opposed to previous methods, which are eminently empirical, are not cell function oriented, and require a much higher number of experiments. This method was applied for the development of a new chemically defined CM formulation for *P. pastoris*. Experiments were carried out in 250 mL shake flasks with a *P. pastoris* X33 expressing the scFv. The optimized CM formulation enables increasing by approximately twofold the scFv productivity in comparison to the BSM formulation proposed by Invitrogen. Furthermore, this result was confirmed in 2 L bioreactor experiments.

4.1 Introduction

Cell CM design is a very important step in bioprocess development (Dai *et al.* 2011, Kim *et al.* 2005). The current standard methodology for determining the optimal composition of cell CM for various microorganisms is design-of-experiments (DoE) (Mantha *et al.* 1998, Preetha *et al.* 2007, Zang and Greasham 1999). Numerous studies of CM optimization using DoE for bacterial, fungal and mammalian cultures supported by bioreactor (Ghosalkar *et al.* 2008) or shake flasks experiments (Deshpande *et al.* 2004, Ooijkaas *et al.* 1999) have been reported. The most common DoE method is the so called reduced factorial design with two levels of concentrations, which permits a preliminary screening of between five and ten medium factors in a limited number of experiments (Mandenius and Brundin 2008).These classical experimental designs require that only one variable be changed at a time to determine its effect.

Since chemically defined medium contains a fairly good number of components, it is not feasible to study independently each component for its effect on microbial growth. The main disadvantage of the statistical DoE method is that, due to its empirical nature, it is costly when applied to a large number of medium factors with potential interactions. To speed-up screening of high numbers of medium factors, costly high-throughput CM optimization equipment has been recently developed based on microbioreactor technology with the goal of enabling the screening of thousands of nutrient levels or combinations thereof to be run in parallel. Luan *et al.* 2008 discloses a rational method for cell CM design, wherein concentrations of aminoacids in the medium are calculated on the basis of protein content of cultured cells, aminoacids composition of expressed recombinant proteins and cellular maintenance needs. In alternative, Spaargaren 1996 used linear programming algorithms to design cell CM by approximate the average elemental composition of biological material to the composition of mixture of inorganic salts and glucose.

There are today more advanced Systems Biology tools that can be applied for less empirical and tendentially mechanistic CM design. In the worldwide patent nº WO2004101808 a method is described for the development of cell CM formulations using genomics and/or proteomics. Kell and co-workers (Kell *et al.* 2005) showed that there is a tight link between the exometabolome (the concentration of all extracellular metabolites) and the intracellular state of the cells. They showed that exometabolome dynamics provides an informative and accurate "footprint" of cellular metabolic activity and indirectly of genomic and proteomic states. Indeed, the medium transports not only the essential nutrients but also small molecules and proteins involved in gene expression regulation (Hunter 2005). The use of exometabolomics to improve the composition of CM has however never been described before.

An alternative to the modulation of particular biochemical transformations, is functional oriented metabolic engineering, which is the approach explored in this thesis. The metabolism of a cell can be decomposed into elementary cellular functions using EFM analysis (Schuster and Claus 1994). The decomposition of the metabolism of cells into EFM has been previously used to support genetic engineering (Trinh *et al.* 2006) and functional genomics (Forster *et al.* 2002). An elementary function oriented CM design method, which is the method described here, has however never been attempted before. The method is thus called Cell Functional Enviromics (CFE), because it is based on the systematic characterization of the effect of environmental variables (i.e. medium factors) on cellular function. This new method comprises two main stages.

In the first stage, a Functional Enviromics Map (FEM) is built through the joint screening of cell functions and medium factors by the execution of a specific cell culture protocol and exometabolome assays protocol. The FEM consists of a data array of intensity values of elementary cellular functions against medium factors.

In the second stage, from the FEM, optimized cell CM formulations are developed that either enhance or repress target elementary cellular functions in order to enforce a desired phenotype.

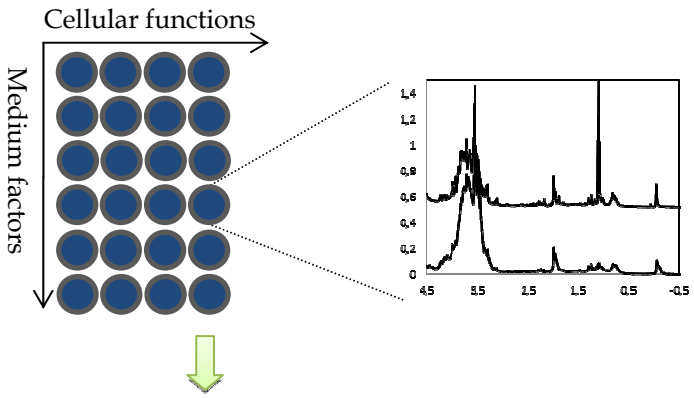
The main advantage of this method lies in enabling metabolic engineering through the CM composition manipulation, wherein an arbitrarily high number of cells functions are optimized through manipulation of medium factors, as opposed to previous methods (DoE) which are eminently empirical, are not cell function oriented, and require a much larger number of experiments.

4.2 A new method for culture medium optimization by cell functional enviromics

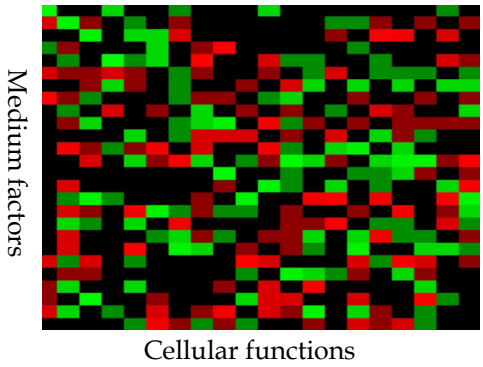
The method for cell CM development proposed here is primarily focused on the elucidation of the function of medium factors. A distinctive feature of the method is that medium factors and elementary cellular functions are joint screened through a particular experimental protocol to extract data, which is then processed into the form of a FEM.

Before the experimental protocol is executed, it is mandatory to clearly state the biological structure to which the medium will be designed, hereon referred to as "target biological structure". It can be a whole cell, an organelle, or a coherent set of metabolic reactions that represent a given cellular function. Once this general structure is known, a method comprising four steps is applied as follows (Figure 4.1).

1. Array of cell cultures



2. Functional enviromics map



3. Optimized culture medium formulation

$$x_j^{opt} = (1 + \eta_i c_{j,i}) x_j^{(0)}$$

4. Final validation step

Figure 4.1 Schematic representations of the main steps of the Cell Functional Enviromics method. Step 1 an array of cell cultures is executed and initial and endpoint exometabolome assays are performed. Step 2 the Functional Enviromics Map is computed, wherein in the xx axis elementary cellular functions set by the universe of genes are represented and in the yy axis medium factors are represented. Step 3 is the formulation of optimized culture medium and lastly step 4 represents the final validation in cell culture triplicates of the final optimized medium formulation.

4.2.1 Step 1: Array of cell cultures

First, at least N+1 (number of medium factors plus one) cell culture experiments are executed with varying CM composition. The different CM composition screened in each experiment is defined in a way to generate adequate experimental data for the purpose of linear regression of elementary cellular function weighting factors against CM factors values.

Then, initial and end-point exometabolome data for each cell culture is acquired using preferably fast and high-throughput analytical techniques such as NMR technique or chromatography coupled to mass spectrometry (GC-MS or LC-MS). This analysis can be complemented with more traditional metabolite specific analytical methods such as HPLC.

After this, a pre-screening of active elementary cellular functions by linear regression of elementary cellular functions intensity values (weighting factors) against medium factors values is performed. These intensity values are obtained from the initial and endpoint exometabolome data. Firstly, the rate of change of each compound (metabolite) is calculated by the following formula:

$$r_i^k = \cdot \frac{C_{Mi}^k(t_{batch}) - C_{Mi}^k(0)}{C_X^{av} t_{batch}} \tag{4.1}$$

With $C_{Mi}(0)$ the initial concentration, $C_{Mi}(t_{batch})$ the endpoint concentration, X_{av} the average cellular concentration, t_{batch} the duration of the batch experiment. Note that the superscript index denotes experiment while the subscript index denotes compound.

Then a statistical regression analysis is performed of ***r*** against medium factor values using the following linear model, according to Eq. 3.3:

$$\mathbf{r} = \cdot \sum_{i=1}^{k} \lambda_i \times em_i \tag{4.2}$$

$$\lambda_i = \sum_{j=1}^{N} x_j \times c_{j,i} \tag{4.3}$$

With $\mathbf{c_{j,i}}$ the regression coefficients (Eq. 3.4) which represent the intensity of activation of cellular function $\mathbf{em_i}$ by culture medium factor x_j. Then, linearly regressions of λ_i against medium factor values are performed and the elementary cellular functions are ranked according to the highest correlation with medium factors or maximum variance captured of **r**.

4.2.2 Step 2: Functiomal Enviromics Map

For the totality of experiments performed in step 1, the medium composition data is organized in a data matrix,

$$\mathbf{X} = \{x(t)\} \quad (4.4),$$

a $M \times N$ matrix of M medium formulations, and respective measured flux data,

$$\mathbf{R} = \{r(t)\} \quad (4.5),$$

a $M \times q$ matrix of measured fluxes. Then determine a subset of elementary cellular functions among the whole set of elementary cellular functions which is tightly linked to the medium factors and determine their weighting factor to the observed cellular phenotype, $\mathbf{R} = \{r(t)\}$, by regression analysis of $\mathbf{R} = \{r(t)\}$ against medium composition data $\mathbf{X} = \{x(t)\}$ satisfying the criteria of PLP method, described in Chapter 3, namely:

(1) Maximization of captured variance of flux data sets, $\mathbf{R} = \{\mathbf{r}(t)\}$;
(2) Maximization of correlation between elementary cellular functions weighting factors λ_i against envirome data (medium factors), $\mathbf{X} = \{\mathbf{x}(t)\}$;
(3) Minimization of the redundancy, i.e. minimization of the number of active EFM (elimination all EFM with weak correlation with the envirome).

These criteria can be fulfilled by maximizing the covariance between medium composition data, $\mathbf{X} = \{x(t)\}$, and respective measured flux data, $\mathbf{R} = \{\mathbf{r}(t)\}$. For further details, see Chapter 3.2.

The result of this procedure is the discrimination of the subset of elementary cellular functions that is tightly linked with medium composition. The information can finally be organized in a **N×K** data array, called Functional Enviromics Map:

Functional Enviromics Map (FEM)= $\{\mathbf{c}_{j,i}\}$ (4.6).

The rows represent **N**, medium factors, columns represent the universe of **K**, elementary cellular functions and $\mathbf{c}_{j,i}$ the relative "intensity" of up- or down-regulation of elementary cellular functions i by medium factor j.

In summary, in this step 2 a subset of **K** active elementary cellular functions are determined that show high correlation coefficients with medium factors values by regression analysis of exometabolome data or derived exometabolome data against medium composition data using the PLP algorithm.

4.2.3 Step 3: Optimized culture medium formulations

In CM design, ideally one should be able to fine tune cellular functionality according to user needs. Combining the information of desired functionality and of FEM it is possible to deduce aprioristic rules about how to tune medium composition in relation to some baseline formulation in a way to enforce the desired functionality. This allows reducing drastically the number of experiments for CM design. In limit, if the information in FEM is sufficiently accurate, a single design step results in a quasi-optimal CM formulation. More specifically, the procedure is as follows.

Each column of FEM matrix holds information of how the baseline medium factor values should be adjusted either to enhance or repress a particular elementary cellular function. More specifically, the new medium factor values should be changed according to the following formula:

$$x_j^{opt} = (1 + \eta_i c_{j,i}) x_j^{(0)} \quad (4.7)$$

With $x_j^{(0)}$ the baseline value of medium factor j, x_j^{opt} the optimized value of medium factor j, $\mathbf{c}_{j,i}$ the intensity value of the jth row and ith column of the FEM and η_i the desired enhancement factor of elementary cellular function i (design parameter). Cellular function specific medium supplementation formulations can be deduced from columns of the FEM. Alternatively, globally optimized medium formulations can be obtained by applying enhancement factors to several elementary cellular functions simultaneously.

In summary, in this step a second set of cultivation experiments comprising a number of cultivations equal or higher than the number of active cellular functions plus one (K+1) are executed, wherein each cultivation is performed in a different CM composition, wherein combinations of low and high intensity levels of elementary cellular functions are screened.

4.2.4 Step 4: Final validation

The newly optimized CM formulation is screened in additional culture experiments as describe in steps 1. The number of experiments is however now much lower, typically triplicates of a given optimized medium formulation. At the end of step 4 it is expected productivity and/or product quality or overall culture performance gains far beyond the baseline medium formulation.

4.3 Materials and methods

4.3.1 Strain and culture conditions

The yeast used was *P. pastoris* X33 strain constitutively expressing a scFv. The pre-inoculum containing 110 mL of BSM (baseline composition), pH=5.0, was inoculated with one cryovial, containing 1 mL of *P. pastoris* X33 cell stock, and incubated at 30 ^{o}C, 150 rpm for 26 hours. For each set of experiments, 52 (26 different conditions, duplicate shake flasks used for each condition) 250 mL shake flasks containing 40 mL of different medium composition were inoculated in with 2 mL of the same pre-inoculum described before.

These shake flasks were incubated at 30 ºC, 150 rpm for 110 hours on an orbital incubator. The pH was allowed to vary over time without adjustment in the shake flasks experiments. Samples were taken at 24 hours interval.

For 2 L bioreactor experiments the best CM formulations resulting from the shake flasks experiments was used for all growth steps: pre-inoculum, inoculum and cultivation. The 2 L bioreactor was inoculated with 100 mL inoculum, using the optimized medium at pH=5.0 and temperature 30 ºC. The bioreactor was operated in glycerol batch phase during approximately 90 hours. After this first phase, 100 mL of a solution containing glycerol 99 % w/v and 12 mL/L of optimized PTM1 was added for a second glycerol batch phase extending the culture for extra 60 hours.

4.3.2 Analytical techniques

Each sample was analyzed for biomass by optical density at 600nm (OD_{600nm}) and by the wet cell weight method, for scFv by ELISA and for metabolites by HPLC. For details see Chapter 2.3.

4.3.3 Metabolic network and elementary flux modes

The metabolic network of the *P. pastoris* X33 strain and respective EFMs are described in Chapter 2.4.

4.4 Results and discussion

The most commonly used medium for the high cell density *P. pastoris* cultivation is the BSM proposed by Invitrogen (*Pichia* fermentation process guidelines, 2000) with PTM1 salts supplements. Although is considered a standard CM, it may not be the optimum for production of heterologous proteins (Cos *et al.* 2006). As such, we have applied the previously described Cell Functional Enviromics method for the optimization of PTM1 composition in order to enhance the expression of heterologous proteins. Thus the "target biological structure" in this case is the lumped function of heterologous protein expression.

More specifically, the aim of the design is to enhance the EFM corresponding to the expression of the scFv.

4.4.1 Array of cell cultures

The 26 assays, 24=2×(N+1) cell cultures being N = 11 plus two control experiments, BSM (baseline formulation) and low concentration of all components, were performed in 250 mL shake flasks. The eleven medium factors selected for screening are listed in Table 4.1. Each medium factor has a baseline value taken from the *Pichia* fermentation process guidelines proposed by Invitrogen, a -1 level (10 times lower than the baseline value) and a +1 level, coincident to the baseline value.

Table 4.1 List of medium factors, respective baseline values and upper (+1) and lower (-1) values for cell culture experiments. The final medium formulation comprises mixtures of 1:200 (v/v) of the PTM1 and diluted PMS solutions respectively.

Factors	Components	+1 level	-1 level	Units
1	$CuSO_4$-$5H_2O$	6.000	0.600	g/L
2	NaI	0.080	0.008	g/L
3	$MnSO_4$-H_2O	3.000	0.300	g/L
4	Na_2MoO_4-$2H_2O$	0.200	0.020	g/L
5	H_3BO_3	0.020	0.002	g/L
6	$CoCl_2$-$6H_2O$	0.500	0.050	g/L
7	$ZnCl_2$	20.000	2.000	g/L
8	$FeSO_4$-$7H_2O$	65.000	6.500	g/L
9	Biotin	0.200	0.020	g/L
10	H_2SO_4	5.000	0.500	mL/L
11	PMS	40.000	20.000	mL

PSM includes glycerol 40 g/L in +1 level and 20 g/L in -1 level.

The combinations of medium factor values tested in each experiment are listed in Table 4.2. These were obtained by a D-optimal experimental design for linear function identification for 11 factors (columns) and 26 experiments (rows).

Table 4.2 D-optimal design, linear, (rowexch function) for 11 factors and 26 experiments, in duplicated.

	$CuSO_4$ -$5H_2O$	NaI	$MnSO_4$-H_2O	Na_2MoO_4 -$2H_2O$	H_3BO_3	$CoCl_2$	$ZnCl_2$	$FeSO_4$ -$7H_2O$	Biotin	H_2SO_4	Vol. PMS
Exp. 1	1	1	-1	-1	1	-1	1	-1	-1	-1	1
Exp. 2	1	-1	-1	1	-1	-1	-1	1	-1	-1	1
Exp. 3	1	-1	1	1	1	-1	1	1	1	-1	-1
Exp. 4	1	1	-1	-1	1	1	-1	1	1	1	-1
Exp. 5	1	1	1	1	1	-1	-1	-1	1	1	1
Exp. 6	-1	1	-1	1	1	1	1	-1	1	-1	-1
Exp. 7	-1	1	1	1	1	1	1	1	1	1	1
Exp. 8	1	1	1	1	-1	1	1	1	-1	-1	-1
Exp. 9	1	-1	-1	1	-1	-1	1	1	1	1	-1
Exp. 10	-1	1	1	-1	-1	-1	-1	-1	1	-1	-1
Exp. 11	-1	1	-1	1	-1	1	-1	1	-1	-1	1
Exp. 12	-1	-1	-1	-1	1	-1	1	1	1	-1	1
Exp. 13	-1	-1	1	1	1	1	-1	-1	-1	-1	1
Exp. 14	1	1	-1	1	-1	1	1	-1	1	-1	1
Exp. 15	1	-1	-1	1	1	1	-1	-1	-1	1	-1
Exp. 16	1	1	1	-1	1	-1	-1	1	-1	-1	-1
Exp. 17	1	-1	1	-1	-1	1	-1	-1	1	-1	1
Exp. 18	-1	1	1	-1	-1	1	1	1	-1	1	1
Exp. 19	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1
Exp. 20	-1	-1	1	1	-1	-1	1	-1	-1	1	-1
Exp. 21	-1	1	-1	1	1	-1	-1	1	-1	1	1
Exp. 22	-1	-1	-1	-1	-1	1	-1	1	1	1	-1
Exp. 23	1	1	-1	-1	-1	-1	1	-1	-1	1	1
Exp. 24	1	-1	1	-1	1	1	1	1	-1	1	1
Exp. 25	1	1	1	1	1	1	1	1	1	1	1
Exp. 26	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Experiment nº 25 is the BSM (baseline formulation). Experiment nº 26 is an extra experiment with all the eleven factors taking the level -1.

The results of cell CM experiments with varying CM composition defined in Table 4.2 are presented below. These experiments were performed in duplicate for each condition tested. Figure 4.3 shows the array of shake flasks experiments inside one incubator (a second orbital incubator was used for the remaining shake flasks).

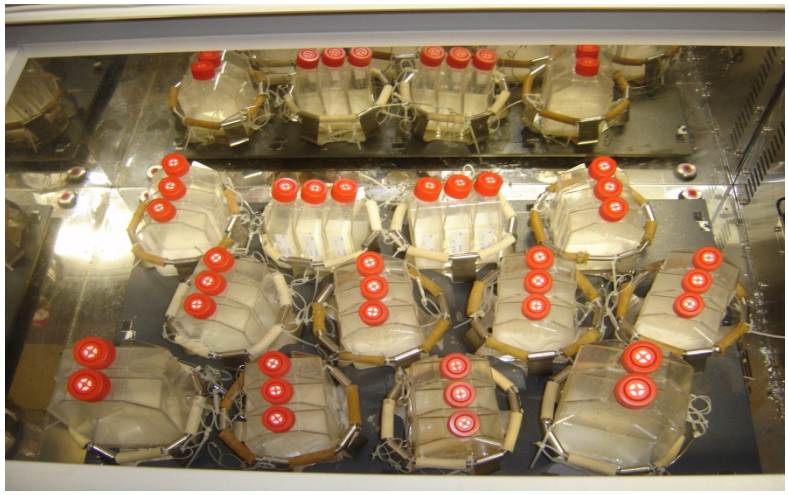


Figure 4.2 Cell culture experiments with varying culture medium composition in shake flasks incubated in an Innova 4300 orbital incubator at 30 ºC.

Biomass and product profile of the shake flask experiment are shown in Figure 4.3.

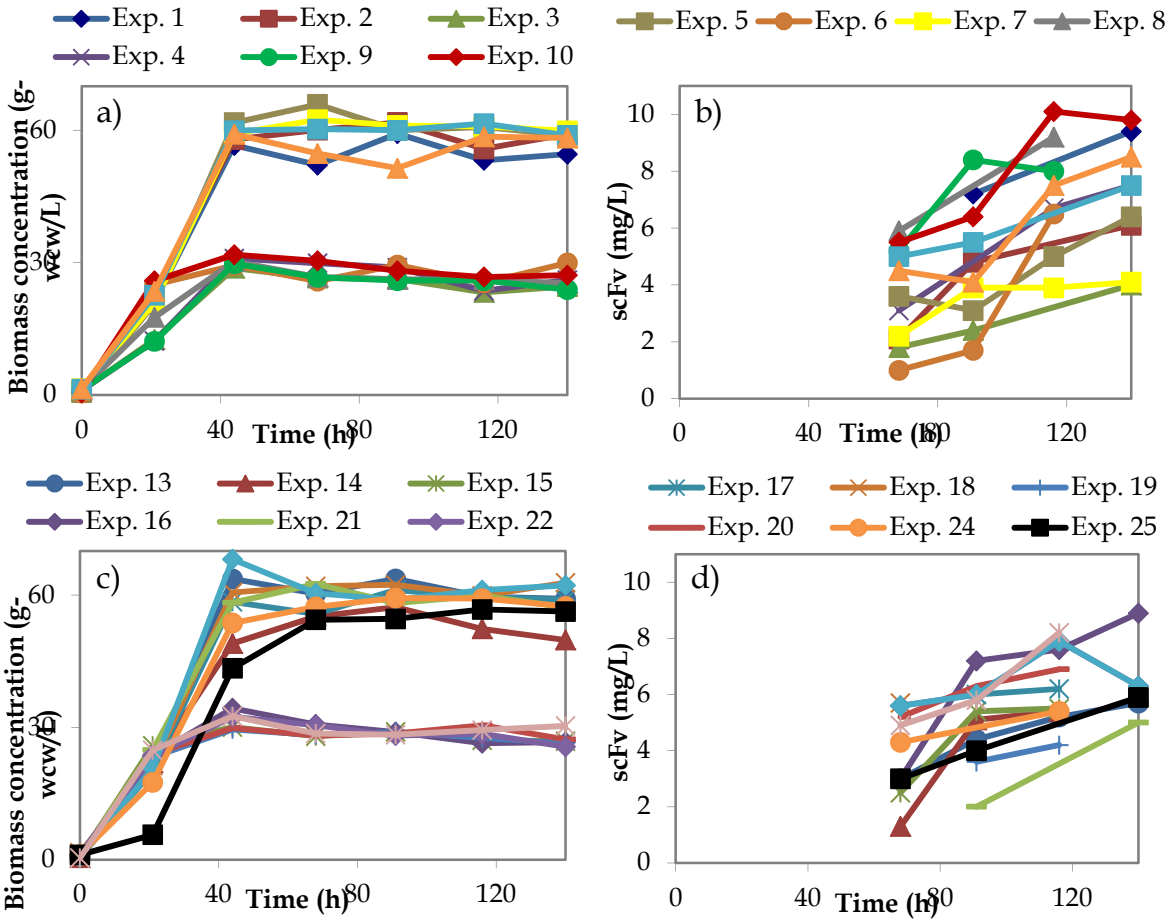


Figure 4.3 Off-line measured of biomass and scFv concentration profiles obtained in 26 experiments operated in shake flaskes. a) biomass concentration (g-WCW/L) over time of the exp.1 to exp. 12; b) scFv concentration (mg/L) over time of the exp.1 to exp. 12; c) biomass concentration (g-WCW/L) over time of the exp.13 to exp. 26; d) scFv concentration (mg/L) over time of the exp.13 to exp. 26.

From these dynamic data, the values of the biomass and scFv concentrations for each shake flasks experiments are shown in Table 4.3.

Table 4.3 Off-line measured of biomass (g-WCW/L) and scFv (mg/L) concentration obtained in 26 experiments operated in shake flasks for 140 hours.

	Biomass (g-WCW/L)				**scFv (mg/L)**			
	t = 68 h	t = 91 h	t = 116 h	t = 140 h	t = 68 h	t = 91 h	t = 116 h	t = 140 h
Exp. 1	52.13	59.27	53.20	54.60	--------	7.20	--------	9.40
Exp. 2	60.07	61.80	55.80	59.13	2.10	4.80	--------	6.10
Exp. 3	26.53	26.27	23.27	24.60	1.80	2.40	--------	4.00
Exp. 4	29.87	28.87	23.60	26.07	3.10	--------	6.70	7.50
Exp. 5	65.87	60.00	60.73	58.87	3.60	3.10	5.00	6.40
Exp. 6	25.80	29.47	25.40	29.93	1.00	1.70	6.50	--------
Exp. 7	62.40	61.07	60.87	59.87	2.20	3.90	3.90	4.10
Exp. 8	26.73	26.20	26.13	25.27	5.90	--------	9.20	--------
Exp. 9	26.67	25.93	25.93	23.87	5.20	8.40	8.00	--------
Exp. 10	30.40	28.13	26.73	27.13	5.50	6.40	10.10	9.80
Exp. 11	60.27	60.00	61.53	58.87	5.00	5.50	--------	7.50
Exp. 12	54.73	51.40	58.53	58.27	4.50	4.10	7.50	8.50
Exp. 13	60.33	63.67	59.60	59.13	3.00	4.40	5.20	5.70
Exp. 14	55.13	57.27	52.33	49.80	1.30	5.10	--------	5.40
Exp. 15	28.00	29.00	27.07	26.87	2.50	5.40	5.50	--------
Exp. 16	30.67	28.93	26.40	26.67	3.00	7.20	7.60	8.90
Exp. 17	55.73	61.13	59.53	58.93	--------	6.00	6.20	--------
Exp. 18	62.00	62.33	59.87	62.80	3.10	5.80	6.10	5.70
Exp. 19	28.27	28.93	27.60	26.40	--------	3.60	4.20	--------
Exp. 20	28.13	28.60	30.47	27.27	--------	6.60	7.60	--------
Exp. 21	62.53	58.20	59.87	57.20	5.20	6.30	6.90	--------
Exp. 22	30.47	28.13	28.53	25.60	--------	2.00	--------	5.00
Exp. 23	60.40	58.93	61.13	62.20	5.60	6.00	7.90	6.30
Exp. 24	57.33	59.33	59.20	57.47	4.30	--------	5.40	--------
Exp. 25	54.40	54.60	56.73	56.27	3.00	4.00	--------	5.90
Exp. 26	28.60	28.33	29.40	30.33	4.90	5.80	8.20	--------

Furthermore, more detailed metabolite profiling was performed for some of the experiments. More specifically, the concentration in the supernatant of glycerol and several organic acids, namely malic acid, succinic acid, lactic acid, formic acid, fumaric acid were determined by HPLC after 115 hours of growth.

Table 4.4 Exometabolome data, glycerol and organic acids concentrations determined by HPLC at time 115 hours.

	Gly (g/L)	Malic Acid (g/L)	Succinic Acid (g/L)	Lactic Acid (g/L)	Formic Acid (g/L)	Fumaric Acid (g/L)
Exp. 1	0.12	0.28	0.46	0.12	0.34	0.08
Exp. 8	0.01	0.15	0.11	0.10	0.11	0.07
Exp. 10	0.03	0.11	0.08	0.06	0.09	0.08
Exp. 20	0.01	0.07	0.12	0.10	0.12	0.13
Exp. 25	0.07	0.24	0.26	0.11	0.09	0.11
Exp. 26	0.01	0.19	0.09	0.09	0.15	0.07

The specific productivity was calculated from the initial and end-point exometabolome data for each cell CM experiments. Figure 4.4 shows the productivity results obtained in each experiment.

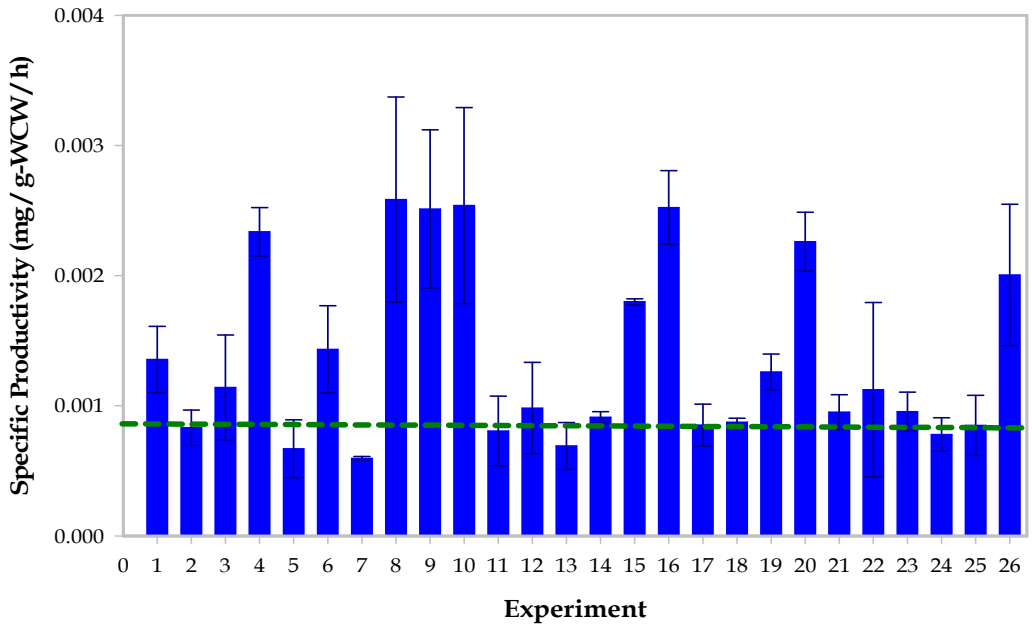


Figure 4.4 Productivity observed for tested medium composition. The dashed line green corresponds to the maximum achieved with the BSM formulation.

It should be noted that, remarkably, the specific scFv productivity in about half of the experiments is above that of the baseline BSM formulation, which opens up the possibility for a substantial increase in scFv productivity.

4.4.2 *P. pastoris* Functiomal Enviromics Map

Upon the application of the PLP algorithm to process the above data, a subset of elementary cellular functions was obtained, which correspond to the subset of EFM with higher correlation to the medium factors. The final regression coefficients $\mathbf{c_{j,i}}$ which represent the intensity of activation of cellular function $\mathbf{em_i}$ by culture medium factor x_j.

$$\lambda_i = \sum_{j=1}^{N} x_j \times c_{j,i} \qquad (4.3),$$

are shown in Table 4.5.

Table 4.5 FEM for the *P. pastoris* X33. Only the most significant elementary cellular functions are shown. All intensity values below 1 % of the standard deviation were removed.

	e10	**e12**	**e5**	**e9**	**e3**
$CuSO_4.5H_2O$	0.06	0.13	0.09	-0.06	-0.05
NaI	0.04	0.18	-0.16	0.00	0.04
$MnSO_4.H_2O$	-0.16	0.21	0.00	-0.03	0.00
$Na_2MoO_4.2H_2O$	0.07	0.04	0.05	-0.03	-0.01
H_3BO_3	0.01	-0.15	0.02	-0.02	-0.02
$CoCl_2.6H_2O$	-0.02	-0.21	0.06	-0.02	0.01
$ZnCl_2$	-0.02	0.15	0.11	-0.10	0.00
$FeSO_4.7H_2O$	-0.76	0.01	0.19	-0.32	-0.02
Biotin	-0.04	0.06	0.25	-0.07	-0.05
H_2SO_4	-0.20	0.06	0.24	0.05	0.00
PMS*	0.43	-0.24	-0.30	0.27	-0.02

This table defines the complete set of active EFM and as such it corresponds to the *P. pastoris* FEM.

$$\text{FEM} = \{\mathbf{c}_{j,i}\} \qquad (4.6).$$

Note that the dimension of each EFM vector is 102 and that the values within it represent the weight of a given metabolic reaction for that particular EFM.

Elementary mode 9 and 10 correspond to biomass synthesis, elementary mode 3 and 12 correspond to product synthesis and elementary mode 5 corresponds to catabolism. In Appendix C are presented these EFM of the metabolic network used that were found to be the most important.

The graphical representation of the *P. pastoris* FEM is shown in Figure 4.5.

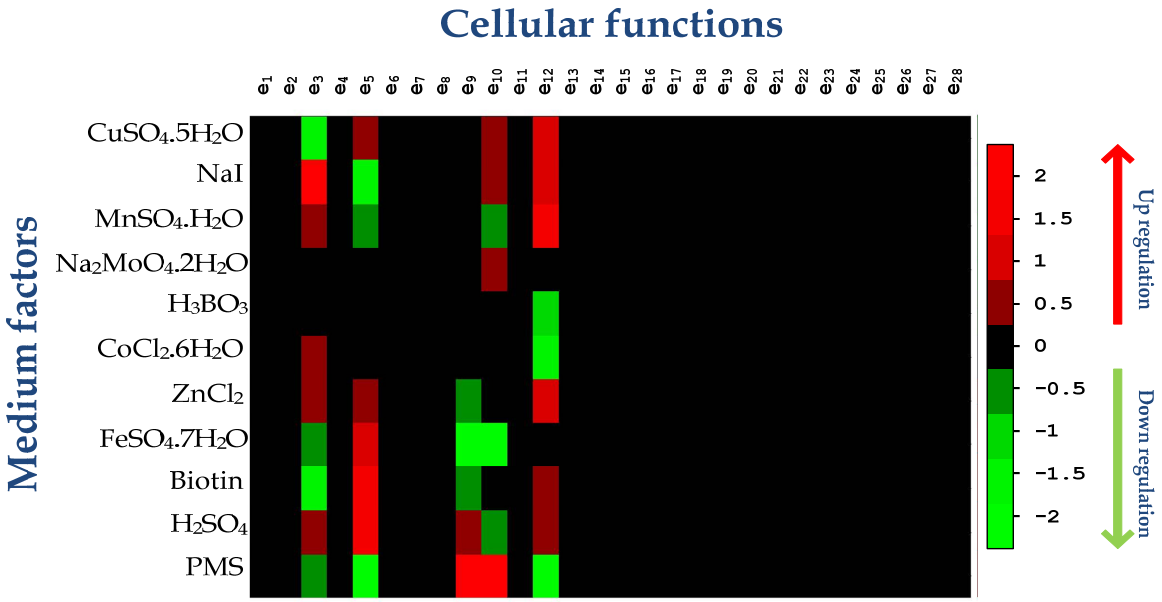


Figure 4.5 FEM for the *P. pastoris* X33 strain used in this work.

4.4.3 Optimized culture medium formulations

Using the data of the *P. pastoris* FEM shown above, optimized medium compositions were obtained in order to enhance the scFv EFM by applying the formula.

$$x_j^{opt} = (1+\eta_i c_{j,i})x_j^{(0)} \quad (4.7)$$

With $x_j^{(0)}$ the baseline BSM value of medium factor j and x_j^{opt} the optimized value of medium factor j. The η_i is the desired enhancement factor. The target EFM were both i=3 and i=12 EFM which are the EFM involved in the expression of the scFv.

In this second round of experiments, ten optimized formulations were tested in shake flasks (in triplicates). Table 4.6 shows the final optimized CM formulations.

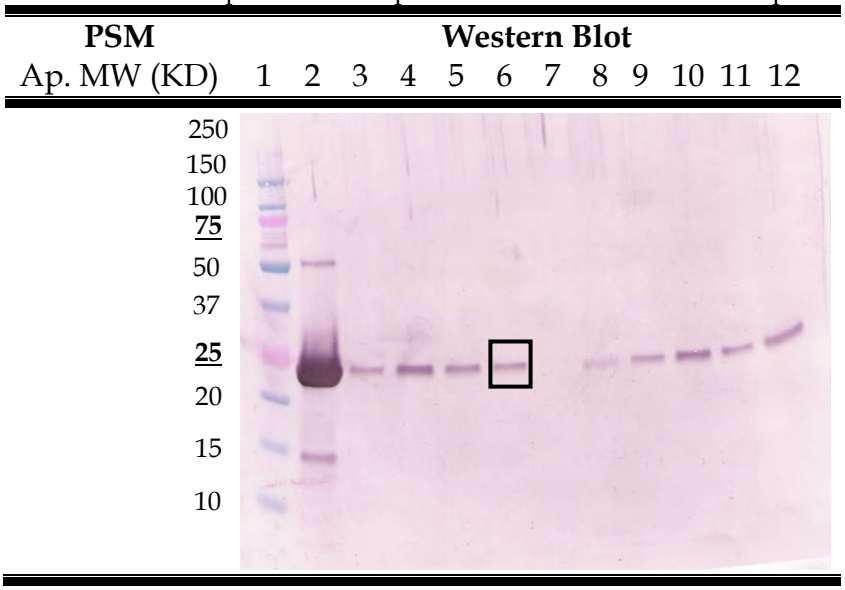
Table 4.6 Media formulations used in the second set of experiments (compositions based on the FEM obtained in the first set of experiments).

	$CuSO_4$-$5H_2O$ (g/L)	NaI (g/L)	$MnSO_4$-H_2O (g/L)	Na_2MoO_4-$2H_2O$ (g/L)	H_3BO_3 (g/L)	$CoCl_2$ (g/L)	$ZnCl_2$ (g/L)	$FeSO_4$-$7H_2O$ (g/L)	Biotin (g/L)	H_2SO_4 (mL/L)	PSM (v/v)
Exp. A	6.00	0.08	3.00	0.20	0.02	0.50	20.00	65.00	0.20	5.00	1.00
Exp. B	6.36	0.09	0.63	0.02	0.02	0.25	21.41	5.96	0.02	0.72	0.95
Exp. C	6.54	0.09	0.78	0.03	0.02	0.25	23.26	5.43	0.03	0.82	0.92
Exp. D	6.75	0.09	0.85	0.03	0.02	0.25	23.10	5.48	0.03	0.94	0.89
Exp. E	6.57	0.10	0.90	0.03	0.02	0.25	23.68	5.27	0.03	0.86	0.90
Exp. F	6.79	0.10	1.12	0.03	0.01	0.25	24.09	5.43	0.03	1.89	0.88
Exp. G	7.25	0.11	1.15	0.04	0.01	0.25	25.21	5.34	0.04	1.99	0.84
Exp. H	7.15	0.11	2.42	0.11	0.01	1.44	26.42	28.36	0.11	5.67	0.82
Exp. I	7.34	0.11	1.69	0.05	0.004	0.25	26.80	4.76	0.04	4.86	0.80
Exp. J	12.00	0.16	6.00	0.40	0.001	0.25	40.00	3.25	0.40	10.00	0.25

Exp. A – Baseline formulation medium; PSM includes 20 g/L glycerol.

The first analysis performed was by Western blot to quickly examine whether the product had increased or not (Table 4.7).

Table 4.7 Western Blot of final supernatant samples for the ten second round experiments.



Legend: Lane 1: PSM; Lane 2: Standard scFv; Lanes 3 to 12: samples from experiments A to H (Lane 3: A; Lane 4: B; Lane 5: C; Lane 6: D; Lane 7: E; Lane 8: F; Lane 9: G; Lane 10: H; Lane 11: I; lane 12: J). Experiment A (marked with a square) corresponds to baseline medium formulation.

As one can see the line in position 4 (a new formulation) is more intense than the line in position 6 (baseline formulation). The next step was to quantify this greater intensity by ELISA (Table 4.8).

Table 4.8 Off-line measured of biomass (g-WCW/L) and scFv (mg/L) concentration values obtained in 10 experiments of validation performed in shake flasks for 140 hours.

	Biomass (g-WCW/L)					**scFv (mg/L)**
	t = 49 h	t = 73 h	t = 97 h	t = 121 h	t = 140 h	t = 140 h
Exp. A	26.5	27.7	29.2	29.9	29.1	5.2
Exp. B	27.2	27.2	28.8	29.2	29.6	4.5
Exp. C	29.7	31.3	32.7	32.0	31.8	4.4
Exp. D	30.2	31.0	31.3	30.8	31.3	3.1
Exp. E	28.0	31.0	30.0	29.9	31.0	6.3
Exp. F	30.5	31.1	31.5	29.7	31.1	7.2
Exp. G	29.2	30.7	29.8	30.4	29.3	6.9
Exp. H	29.8	30.7	31.5	31.3	31.2	3.7
Exp. I	28.4	30.8	29.9	30.3	30.0	4.6
Exp. J	19.8	22.0	23.5	23.8	22.4	6.8

The values of biomass and scFv concentration are the average of three replicas of each experiment. Figure 4.6 shows the result of the measured specific scFv productivity compared to the target design value used in Eq. 4.7.

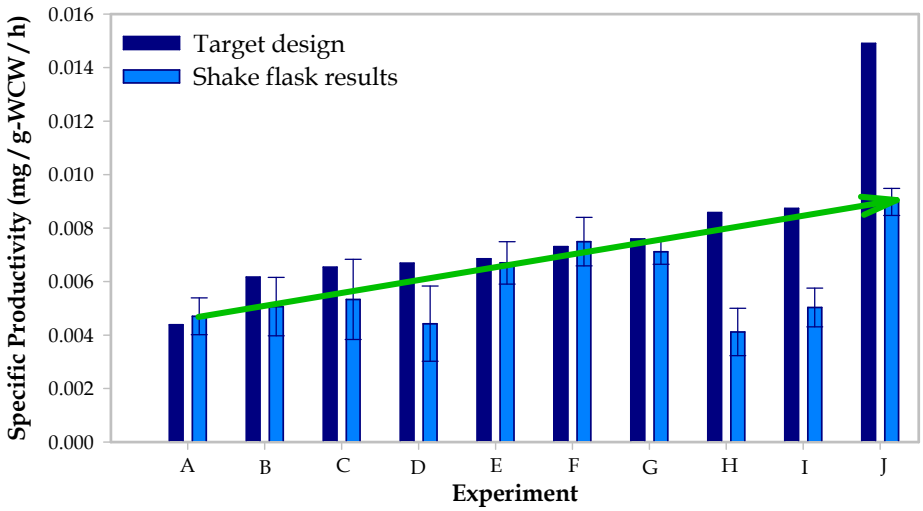


Figure 4.6 Specific productivity (predicted and obtained) for optimized media formulations using a recombinant *P. pastoris* X33 strain.

The results presented above clearly demonstrate that optimized CM formulations can be obtained with much higher product specific productivity than that of the baseline CM formulation (exp. A) using the information contained in the FEM. With exceptions of experiments H and I the measured productivity follows the tendency line of the target design value. Furthermore, the specific productivity of exp. J is 1.9 fold higher than the exp. A. The formulation tested in exp. J represents the final optimum obtained by the functional enviromics method.

Final optimized formulation of *Pichia* Trace Metals 1 (opt. PTM1): $CuSO_4.5H_2O$, 12.0 g/L, NaI, 0.16 g/L, $MnSO_4.H_2O$, 6.00 g/L, $Na_2MoO_4.2H_2O$, 0.40 g/L, H_3BO_3, 0.001 g/L, $CoCl_2.6H_2O$, 0.25 g/L, $ZnCl_2$, 40.0 g/L, $FeSO_4.7H_2O$, 3.25 g/L, Biotin, 0.4 g/L, H_2SO_4, 10.0 mL/L, BSM diluition, 0.25:1 (v/v).

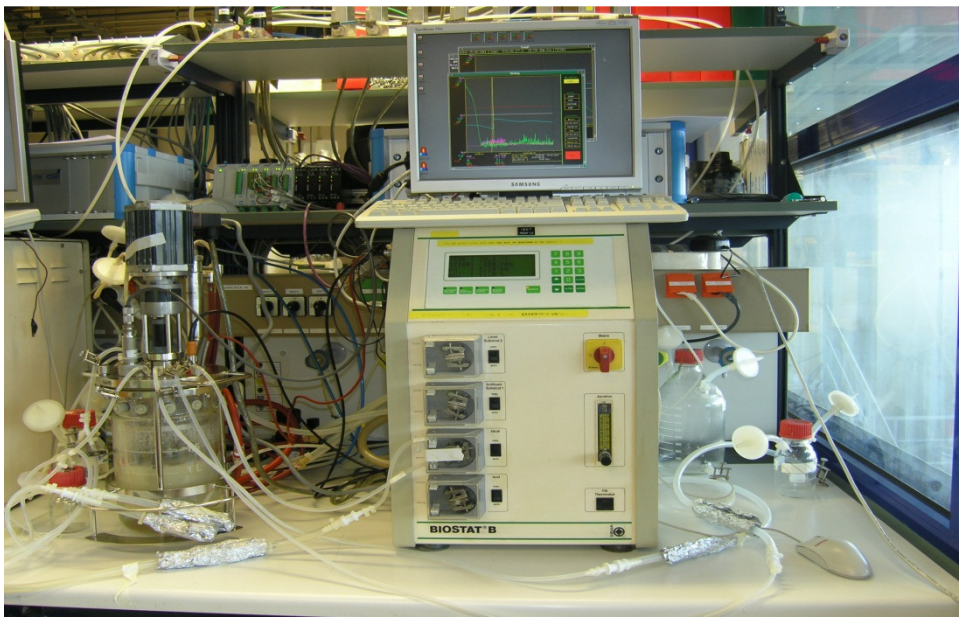
4.4.4 Validation in 2 L bioreactor experiments

The results obtained from 2 L bioreactor experiments are relevant for validation of the used of this new methodology. Shake flask experiments are easy to handle and cost-effective compared to experiments in bioreactors, but the lack of knowledge of the conditions and the lower controllability of volume, oxygen transfer, pH and substrate addition makes the shake flask data less informative for cultivation scale-up (Gupta and Rao 2003). Loss of productivity is often observed when cell cultures are transferred from shake flasks to bioreactors (Scragg *et al.* 1987). The decrease in production was attributed to the different physical conditions (such as degree of mixing, shear stress, and gas phase compositions) encountered between shake flasks and bioreactors.

In order to further validate the best formulation obtained in the previous section (exp. J), it was chosen for scale up to a 2 L bioreactor to confirm the shake flask experiments results. Two runs were performed to access the productivity improvement, the first using the baseline CM formulation (same as experiment A),

and the second using the optimized CM formulation (same as experiment J). Figure 4.7 shows the 2 L bioreactor used for these runs.

a)



b)

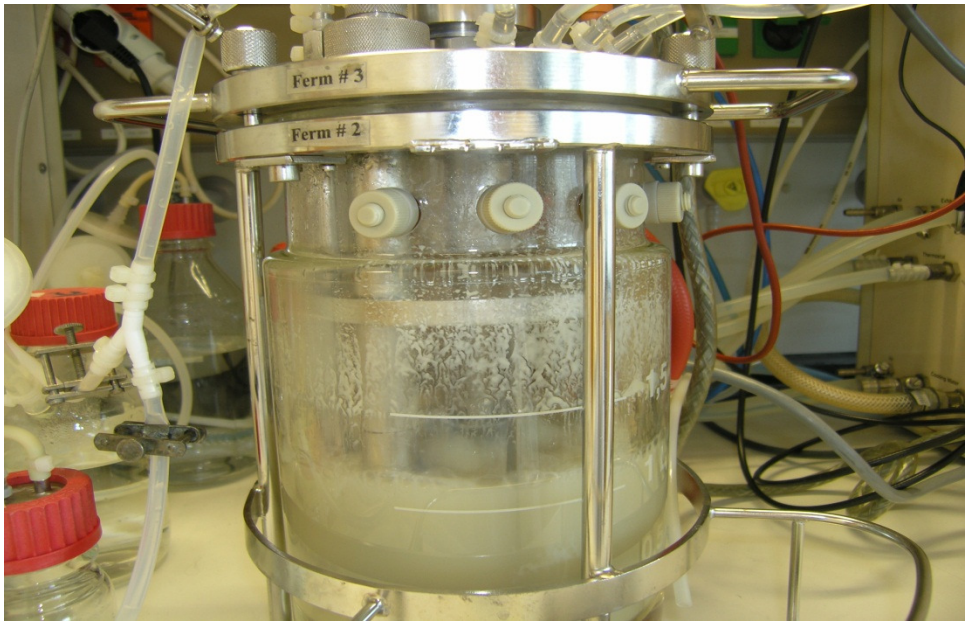


Figure 4.7 2 L bioreactor used for cultivation of the *P. pastoris* X33 strain. a) System used including 2 L culture vessel, the control cabinet and the Supervisory Control and Data Acquisition System. b) Detail of the 2 L culture vessel after 100 hours cultivation.

Figure 4.8 shows the product concentration evolution over time for the two experiments, using the baseline CM composition and the optimized CM composition.

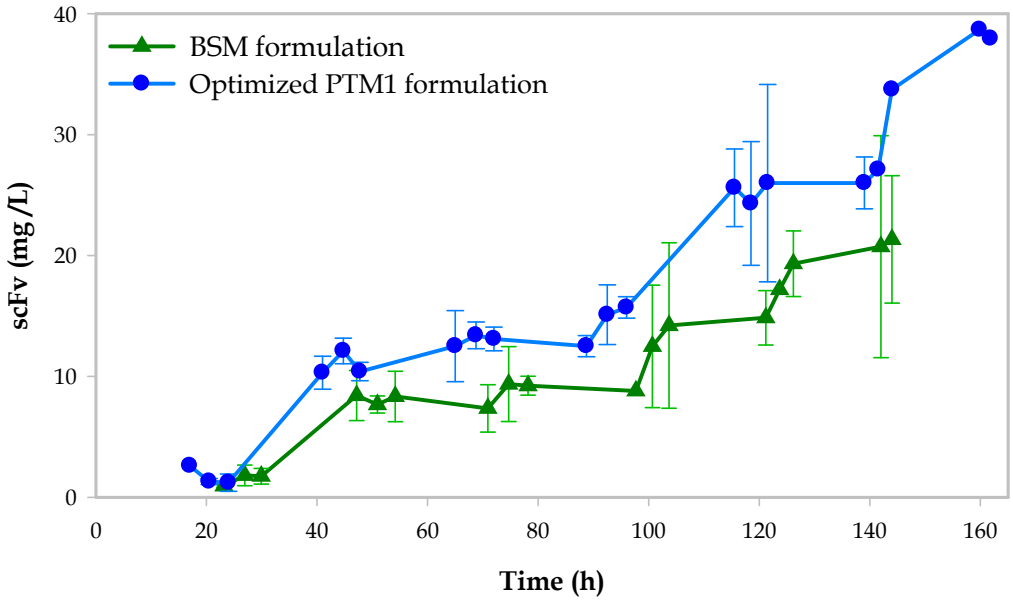


Figure 4.8 scFv concentration evolutions in 2 L bioreactor with *P. pastoris* X33 strain. Squares: baseline medium composition; Circles: optimized PTM1 composition.

These experiments performed at 2 L scale bioreactor confirmed the results obtained in shake flask cultures, being the productivity of the optimized CM composition 1.7 fold higher when compared to the baseline CM composition.

4.5 Conclusion

In this work, a novel cell functional enviromics method was developed to determine the optimal composition of CM. The method comprises the construction of a FEM through the execution of cell culture experiments and preferably high throughput analytical methods of the exometabolome, followed by medium factors values adjustment on the basis of said function of medium factors.

A FEM was built for a *P. pastoris* X33 strain expressing scFv. Ten optimized CM compositions were computed and tested in shake flask experiments to verify the increase of scFv productivity. The best optimized formula yielded 1.9 fold specific scFv productivity increases in relation to the baseline BSM formulation. Further validation experiments performed at 2 L bioreactor scale confirmed the shake flask results showing a 1.7 fold increase in specific scFv productivity in relation to the baseline BSM formulation.

The developed Functional Enviromics method presents several benefits in relation to the currently used methods for CM development. It enables engineering the cells metabolism by the adjustment of several target metabolic objectives simultaneously. It is less empirical and thus can be more easily generalized or extrapolated to different cell species sharing a given group of genes. As the function of genes tends to be conserved among different species also the function of environmental variables tends to be conserved among different species. The CM design is thus based on knowledge, which can be augmented over time without the need of repeating previously tested conditions. Once a sufficient base knowledge is built, CM design requires a much smaller number of experiments. In theory, if the FEM is sufficiently accurate, a single design step is required to optimize the metabolism of the target cells.

Bibliography

Cos O, Ramon R, Montesinos JL, Francisco Valero F. 2006. A simple model-based control for *Pichia pastoris* allows a more efficient heterologous protein production bioprocess. Biotechnology and Bioengineering 95(1): 145-154.

Dai XJ, Liu MQ, Jin HX, Jing MY. 2011. Optimisation of Solid-state Fermentation of Aspergillus niger JL-15 for Xylanase Production and Xylooligosaccharides Preparation. Czech Journal of food sciences 29(5): 557-567.

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

Forster J, Gombert AK, and Nielsen J. 2002. A functional genomics approach using metabolomics and in silico pathway analysis. Biotechnology and Bioengineering, 79(7): 703-712.

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

Gupta A and Rao G. 2003. A study of oxygen transfer in shake flasks using a non-invasive oxygen sensor. Biotechnol Bioeng 84: 351–358.

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

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. 2005. Metabolic footprinting and systems biology: The medium is the message. Nature Reviews Microbiology, 3(7): 557-565.

Kim HO, Lim JM, Joo JH, Kim SW, Hwang HJ, Choi JW, Yun JW. 2005. Optimization of submerged culture condition for the production of mycelial biomass and exopolysaccharides by Agrocybe cylindracea. Bioresour. Technol. 96: 1175-1182.

Luan YT, Wang W, Nolan R, Drapeau D. 2008. Rationally designed media for cell culture. Patent US20080108553.

Mandenius CF and Brundin A. 2008. Bioprocess Optimization Using Design-of-Experiments Methodology. Biotechnology Progress 24(6): 1191-1203.

Mantha D, Basha ZA, Panda T. 1998. Optimization of medium composition by response surface methodology for the production of tartaric acid by Gluconobacter suboxydans. Bioprocess Engineering 19(4): 285-288.

Ooijkaas LP, Wilkinson EC, Tramper J, Buitelaar RM. 1999. Medium optimization for spore production of Coniothyrium minitans using statistically-based experimental designs. Biotechnology and bioengineering 64(1): 92-100.

Pichia fermentation process guidelines. 2000. Invitrogen Corporation.

Preetha R, Jayaprakash NS, Philip R, Singh ISB. 2007. Optimization of medium for the production of a novel aquaculture probiotic, Micrococcus MCCB 104 using central composite design. Biotechnology and bioprocess engineering 12(5): 548-555.

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

Scragg AH, Morris P, Allan EJ, Bound P, Fowler MW. 1987. Effect of scale-up on serpentine formation by Catharanthus roseus suspension cultures. Enzyme Microb. Technol. 9: 619-624.

Spaargaren DH. 1996. The design of culture media based on the elemental composition of biological material. Journal of biotechnology 45(2): 97-102.

Trinh CT, Carlson R, Wlaschin A, Srienc F. 2006. Design, construction and performance of the most efficient biomass producing E-coli bacterium. Metabolic Engineering, 8(6): 628-638.

WO2004101808 – Genomic and proteomic approaches for the development of cell culture medium.

Zhang J and Greasham R. 1999. Chemically defined media for commercial fermentations. Applied microbiology and biotechnology 51(4): 407-421.

CHAPTER 5

Optimization of *Pichia* Main Salts (PMS) composition

Adapted from:

Ferreira AR, Marques A, Dias JML, Clemente J, Cunha AE, Oliveira R. Dynamic modeling and optimization of main elements consumption in *Pichia Pastoris* cultures. *(Submitted)*.

Abstract

This study investigates the dynamics of the main elements consumption in cultures of the methylotrophic *P. pastoris* GS115 expressing a scFv. Nine experiments were performed under different operational conditions in a pilot 50 L bioreactor. A mathematical model was developed describing the concentration profiles of biomass and of the five main elements (Mg, K, Ca, P and S), which were measured by ICP-AES. The biomass growth model was derived from the Monod equation while the elements consumption rates were defined by the respective element/biomass yields. The model calibration parameters were the maximum specific growth rate and the element/biomass yield coefficients. These parameters were estimated from the data of six cultivation experiments, of which four were used for calibration and two for validation. The model is able to describe accurately the measured concentration profiles of both calibration and validation experiments, while the parameters confidence bounds are generally very low denoting high statistical confidence. The resulting yield coefficients were used to design a new CM formulation of the main salts contained in BSM for a target biomass concentration. This new CM formulation was incorporated in the cell CM developed in Chapter 4 (opt. PTM1) resulting in the new medium called *Pichia* Functional Enviromics Medium 1 (PFEM1).

5.1 Introduction

Yeasts are capable to synthetize most of the building blocks required for biomass growth and recombinant proteins production. Therefore, they are able to grow in relatively simple CM containing a carbon source, a nitrogen source and a few essential inorganic elements, such as Mg, K and P (Zhang and Greasham). It is well-known that these elements play an important role in cell metabolism primarily as structural components of metalloenzymes. Despite its role is not totally well understood, they are essential, serving as structural components for proteins, as cofactors and as enzyme active sites in cellular systems (Plantz *et al.* 2007). Magnesium, potassium, strontium, calcium, copper, iron, manganese, chloride and zinc are reported to be essential elements for yeast (Spencer and Spencer 1997). Potassium is a fundamental cation for *S. cerevisiae* since it stimulates the culture metabolism and respiration (Martínez-Muñoz and Peña 2005). Both sodium and potassium are the main element involved in generation of electrochemical gradients across the plasma membrane (Ariño *et al*. 2010). On the other hand, magnesium and calcium are used as a cofactors of several enzymes, such ATPases on the yeast plasma membrane (Willsky 1979) and glycolytic and alcohologenic enzymes (Walker and Maynard 1997), Ca^{2+}-dependent-ATPases (Okorokov and Ludwig 1998) and aspartase (Depue and Moat 1961). Concerning to recombinant proteins production, Seo and Rhee 2004 have shown that the addition of element metals such Zn, Mn and Co influence the activity of the recombinant phospholipase C (PLC) produced by *P. pastoris*.

The most commonly used CM for the high cell density culture of *P. pastoris* is the BSM developed by Invitrogen, wherein Mg, K, Ca, P and S are the main elements of the medium. One of the major problems reported using BSM is the precipitation of some salts (such as phosphate and magnesium) in the medium, especially at pH higher than 5.0 (Cos *et al.* 2006, Ghosalkar *et al.* 2008), due to their high concentration in the medium (Brady *et al.* 2001, Cereghino *et al.* 2002, Damasceno *et al.* 2004). The saturation of these salts may also cause inhibition of cell metabolism (Plantz *et al.* 2007) which can be prevented by reducing the BSM salts

concentration. This reduction has shown no adverse effect on cell growth rates or biomass yields during growth on either glycerol or methanol (Brady *et al.* 2001). However for extended fed-batch culture time (more than 80 hours) a salt deficiency may occur. To prevent this situation a concentrated salt feed can be pumped into the culture. It was also suggested that the high salt concentration resulted in increased cell death, possibly caused by high osmotic pressure in the medium. Brady *et al.* 2001 have chosen a salt-reduced medium containing one-quarter of the BSM concentration. A study by Surribas *et al.* 2007 demonstrated that decreasing the BSM salts concentration improve the cell viability of GS115 strain and reduces cell lysis. Zhao *et al.* 2008 also confirmed that the cell dead rate before induction decreased from 12 % to 2 % with reduction of one-quarter of the BSM salt concentration. In another study, Jahic *et al.* 2006 reported that the reduction of the BSM salts concentration resulted in lower cell death and consequently lower proteases and other contaminating proteins concentration in the culture broth, which ultimately resulted in higher product concentrations. Jahic proposed a salts monitoring and control system based on on-line conductivity measurement. They were able to control culture solution conductivity at 8 mS/cm which lead to an increase in recombinant protein titer by 3.6 fold in comparison to the standard BSM.

In this work, we studied the individual elements consumption dynamics. ICP-AES was used for measuring selected elements concentration over time and a simple mathematical model was derived from the data obtained. Finally, based on the estimated element/biomass yields, the CM composition was redesigned in order to prevent an excess of certain elements in relation to other elements. Overall, this study enabled to estimate the elements requirements for biomass synthesis and to design a new optimized CM formulation which avoids the precipitation problems commonly reported for the BSM.

5.2 Materials and methods

5.2.1 Yeast strain, culture medium and inoculum

A stable *P. pastoris* GS115 (*Mut*$^+$) strain was used for this work. The culture was grown in a solution containing BSM given in Chapter 2.1. The protocol of the inoculum growth was the same as described in Chapter 2.1.

5.2.2 Bioreactor operation

Cultivations were carried out in a 42 L working volume bioreactor with standard instrumentation to measure on-line pH, temperature, pressure, agitation rate, airflow and DO. The standard operational conditions of the bioreactor were described in Chapter 2.2.

5.2.3 Analytical methods

Biomass concentration was quantified by OD_{600nm} and WCW. The recombinant protein produced (scFv) was analyzed by SDS-PAGE and Western blot and quantified by ELISA. Glycerol and methanol levels were measured in off-line samples by HPLC. To determine the concentration of some elements (Mg, K, Ca, P and S) present in the chemical medium ICP-AES is used. These analytical methods are described in Chapter 2.3.

5.2.4 Mathematical model

According to literature revised in Introduction, biomass growth is highly dependent on salts concentration in the CM. In order to describe their dynamics, a mathematical model for biomass growth on methanol and elements consumption was derived from the Monod equation applied to microbial growth limited by more than one substrate in a multiplicative form (Bailey and Ollis, 1986). As such, the specific growth rate was formulated as follows:

$$\mu = \mu_{max} \cdot \prod_{i=1}^{n} \frac{S_i}{S_i + K_{S,i}} \qquad (5.1)$$

with μ the specific growth rate, μ_{max} the maximum specific growth rate, S_i the concentration of compound i (carbon source or salt), $K_{S,i}$ the respective half-saturation constant and n the number of limiting substrates of biomass growth. The specific uptake rate equations for the salts were defined by the respective biomass yields:

$$q_{S,i} = Y_{S,i/X} \cdot \mu \tag{5.2}$$

with $q_{S,i}$ the specific uptake rate of element i and $Y_{S,i/X}$ the yield of element i on biomass. The material balance equations of biomass and salts in a fed-batch bioreactor without salts feeding are as follows.

$$\frac{dX}{dt} = (\mu - D) \cdot X \tag{5.3}$$

$$\frac{dS_i}{dt} = -q_{S,i} \cdot X - D \cdot S_i \tag{5.4}$$

with X the biomass concentration and D the dilution rate.

Parameters estimation

A program developed in MATLAB was used for simultaneous estimation of the model parameters, as developed by Teixeira (Teixeira *et al*. 2011). The program minimizes the mean of squared errors using the Levenberg-Marquardt algorithm. The material balance equations for fed-batch operation along with the postulated kinetic equations were numerically integrated using a 4th/5th order Runge-Kutta solver. Calculated biomass and salts concentrations were compared with the correspondent off-line measurements to calculate residuals. The final residuals and Jacobian matrix were used to calculate an approximation to the Hessian matrix, thereby assuming that the final solution is a local optimum. The Hessian matrix enabled the calculation of the parameters covariance matrix within 95 % confidence intervals.

5.3 Results and discussion

5.3.1 Cultivation experiments

In order to analyze the effect of the process operational conditions on scFv productivity, nine 50 L fed-batch cultivations were performed. Figure 5.1 shows the 50 L pilot bioreactor used for these experiments.

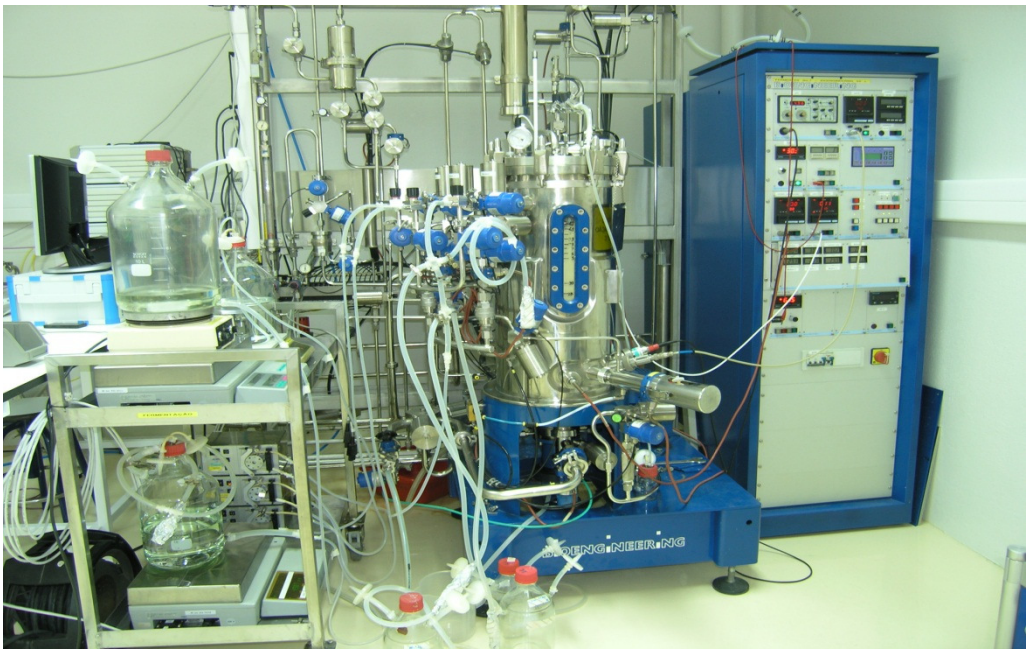


Figure 5.1 50 L pilot bioreactor used for the experimental work.

A fed-batch cultivation of *P. pastoris* GS115 comprehends three distinct phases: glycerol batch (GB) phase, glycerol fed-batch (GFB) phase and methanol fed-batch (MFB) phase. Figure 5.2 shows the main bioreactor operation parameters for a typical 50 L cultivation segmented in the three distinct phases.

In the GB phase the inoculum is added to the reactor and the culture is allowed to grow using the glycerol available in the medium (40 g/L). In this phase glycerol is used as the sole carbon source for cell growth. The DO starts at 100 % of air saturation concentration and progressively decreases mirroring the biomass growth profile. The duration of this phase may vary with the inoculum concentration but it takes on average approximately 30 hours for completion.

The following phase, the GFB phase, consists in the addition of glycerol using a predefined feeding strategy, normally an exponential glycerol feeding, for increasing biomass concentration up to a desired value (WCW= 230 - 450 g-WCW/L) in order to reach the required cell concentration for induction, being also ensured a residual glycerol concentration at induction time.

The last phase is the MFB phase where the glycerol feeding is replaced by a methanol addition that works as both inducer of the recombinant protein expression and as main carbon source. A transition phase between glycerol and methanol feeding is needed, as the cells require some time to express the AOX for metabolizing the methanol. Consequently, the methanol feeding rate starts at a low constant feeding rate, between 1.2 g/(L.h) and 6.0 g/(L.h) for 5 hours. Then, different feeding strategies for methanol feeding were experimentally investigated to maximize scFv expression by *P. pastoris*. Special care must be taken to keep the methanol concentration below toxic levels throughout the all MFB phase.

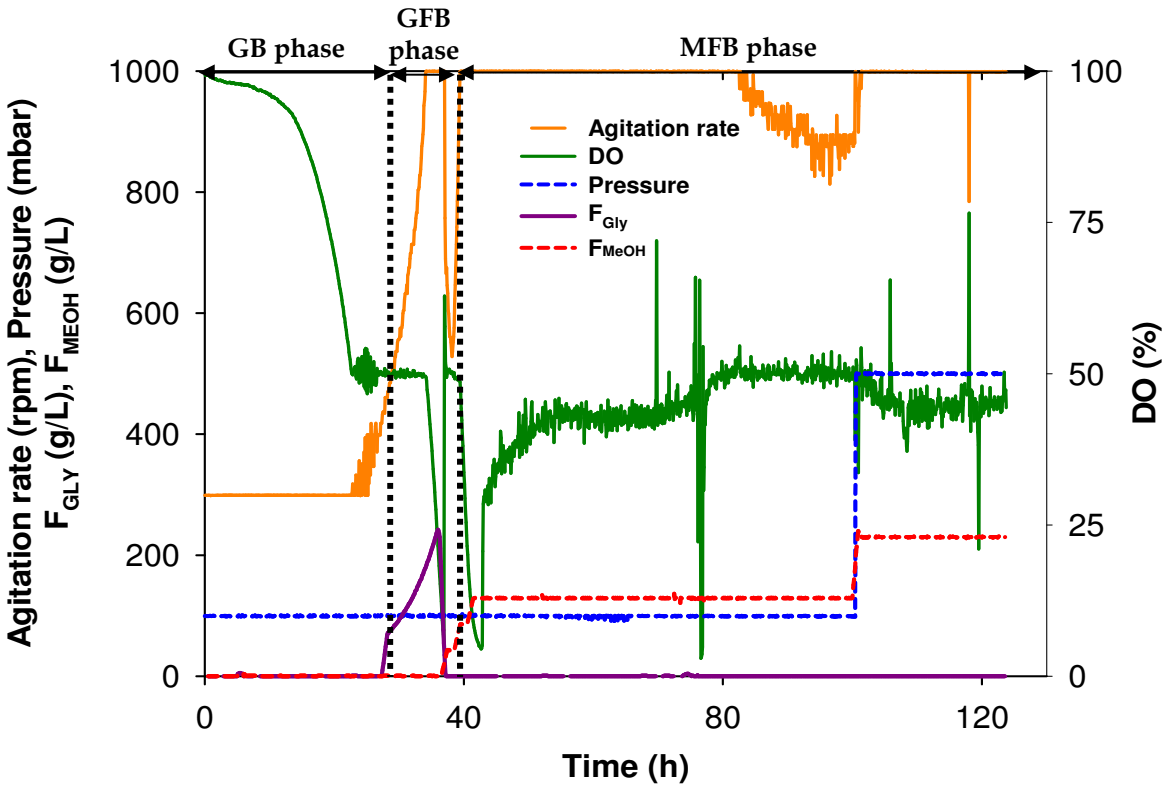


Figure 5.2 Typical *P. pastoris* GS115 cultivation parameters profile in a 50 L bioreactor over the three culture phase, GB, GFB and MFB. (Agitation rate, Pressure, DO – Dissolved oxygen, F_{Gly} – Glycerol feeding rate. F_{MeOH} – Methanol feeding rate).

The bioreactor operational parameters known to impact on the final protein yield in *P. pastoris Mut*$^+$ are the specific methanol feeding rate, pH, temperature and initial biomass concentration in the MFB phase. As such, several experiments were designed in order to perturb the specific methanol feeding rate, pH, temperature and initial biomass concentration in the MFB phase in such a way to generate an information rich data set. This step is essential to maximize the information content of the collected data permitting to develop a consistent mathematical model that captures the impact of these operational parameters on biomass and product concentrations while filtering random fluctuations observed in the process.

Table 5.1 shows the operational parameters for the MFB phase tested in the nine experiments performed. The details of the remaining operational conditions are presented in Appendix D.

Table 5.1 50 L fed-batch cultivation experiments performed and respective MFB phase operational conditions.

Exp.	Time (h)	Operational parameters		Measured variables		
		Temp. (°C)	pH	$X_{MFB,initial}$ (g-WCW/L)	$X_{MFB,final}$ (g-WCW/L)	scFv (mg/L)
D.A	101.8	30.0	5.0	316.9	457.3	5.9
D.B	127.2	30.0	5.0	446.7	585.0	15.6
D.C	145.3	23.0	7.0	295.4	587.7	16.1
D.D	145.8	23.7	7.0	268.1	573.1	14.3
D.E	124.0	30.0	5.0	301.5	434.2	11.9
D.F	162.2	23.7	4.0	139.2	632.3	30.8
D.G	150.0	23.7	4.0	274.2	479.6	30.7
D.H	153.0	30.0	6.5	259.6	475.4	52.5
D.I	149.0	30.0	7.0	244.2	428.1	8.4

Two temperature levels were investigated, 30 °C and 23 °C. Due to limitations in the bioreactor cooling system it was not possible to test temperatures below 23.0 °C. The pH set points used in these experiments were 4.0, 5.0, 6.5 and 7.0.

Two experiments were performed at baseline conditions (T=30 ^{o}C and pH 5.0 defined in the *Pichia* fermentation process guidelines, 2000), namely D.A and D.E (replicate of D.A). Experiment D.E was a replicate of experiment D.A because the latter was prematurely aborted due to air filter clogging.

Experiment D.C and D.D (replicate of D.C) clearly demonstrate high reproducibility concerning biomass and product concentration when all operational parameters are kept constant.

The values shown in Table 5.1 for biomass concentration at induction were achieved through modifications in the glycerol feeding strategy in the proceeding GFB phase. In experiments D.A, D.B, D.C, D.D and D.E glycerol feeding consisted in an exponential feeding profile according to the following equation:

$$F_{Gly} = F_{Gly,0} \times e^{\mu \times t} \tag{5.5}$$

With F_{Gly} being the glycerol feeding rate in g/h, $F_{Gly,0}$ the initial feeding rate (65 g/h) and μ the desired specific growth rate, set at 0.16 h^{-1}. In experiment D.B the GFB phase was extended from 9 to 16 h for increasing the biomass content at the induction point to evaluate the effect of the biomass concentration increase at MFB phase starting.

Experiment D.F, D.G, D.H and D.I used two consecutive exponential profiles:

$$F_{Gly} = \begin{cases} 30 \times e^{0.16 \times t} & t \le 8.5h \\ 116.9 \times e^{0.025 \times (t-8.5)} & 8.5 < t \le 12.5h \end{cases} \tag{5.6}$$

In the MFB phase, the feeding profile of methanol was always different for each experiment (data are shown in Appendix D).

5.3.2 Elements measurements by ICP-AES

Four to six samples per day were taken in all cultivations performed of off-line analysis. Elements concentration, S_i, in the supernatant of each of the samples taken was measured by ICP-AES. Figure 5.3 shows the percentage of variation of concentrations, χ_i, of Ca, K, Mg, P and S over time calculated from ICP-AES measurements using the following formula:

$$\chi_i = \frac{S_i(t) - S_i(0)}{S_i(0)} \times 100 \qquad (5.7)$$

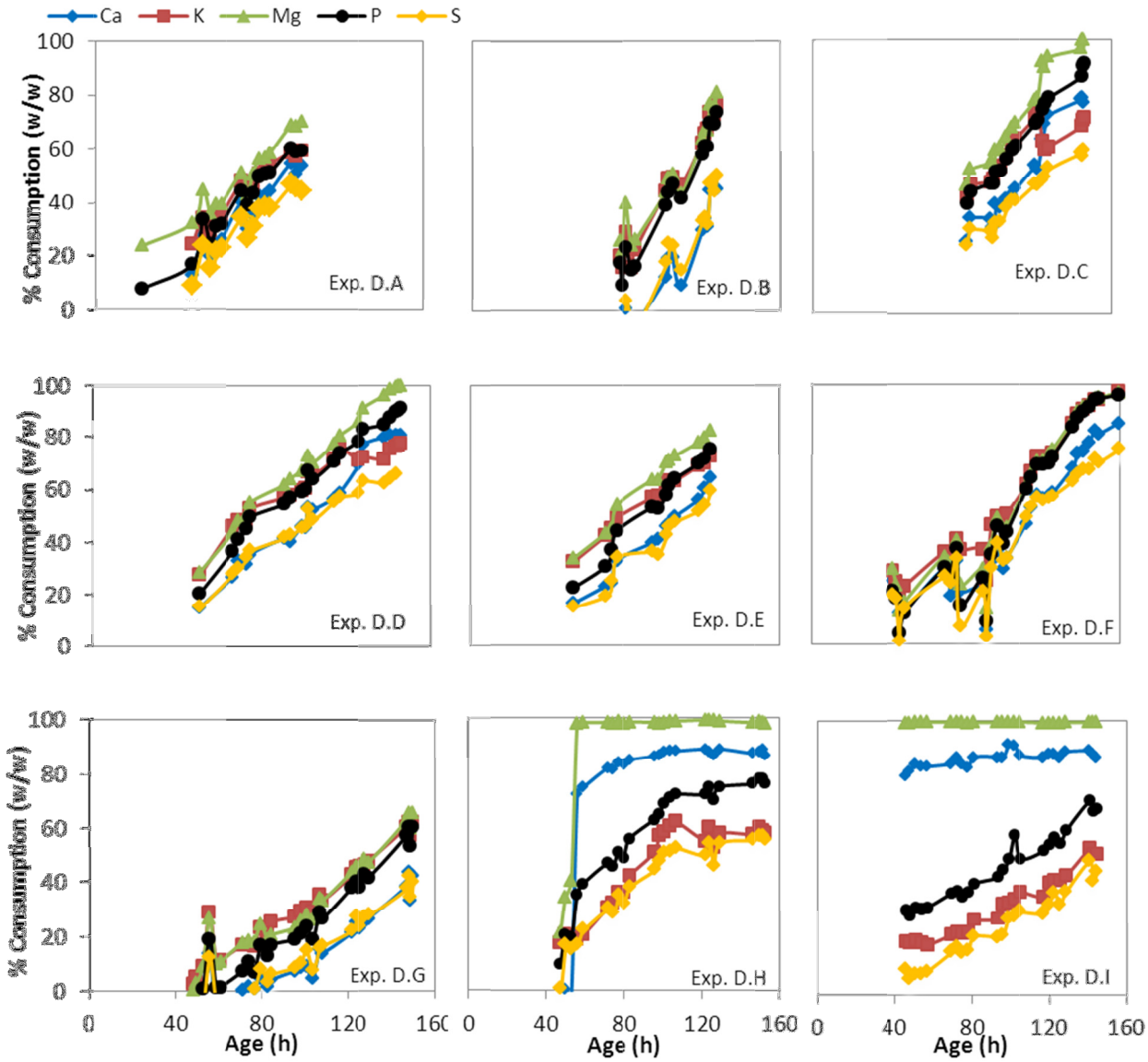


Figure 5.3 % consumption of Ca, K, Mg, P and S measured by ICP-AES over culture time for the nine runs performed.

Although shape of the consumption profile curves is very similar for each of the elements analyzed, the data shows that some elements are in excess in relation to others. More specifically, Mg, K and P tend to deplete sooner than Ca and S. This observation holds for all experiments with exception for experiments D.H and D.I, which were run at pH 6.5 and 7.0 respectively. These data suggest thus the occurrence of precipitation at higher pH values, more severe for Mg and Ca. The concentration of the remaining elements was also significantly affected by pH but precipitation was less severe, permitting to keep reasonable concentration in solution. Salts precipitation at pH higher than 5.0 during BSM preparation has been widely reported (Cos *et al.* 2006, Ghosalkar *et al.* 2008). In the particular case of experiment D.H, the data clearly shows that precipitation can also occur during the cultivation and even in the last MFB phase.

The lowest biomass concentration was reached in experiment D.I (428.1 g-WCW/L or 119.9 g-DCW/L final biomass concentration), which suffered from severe salts precipitation. The highest biomass concentration (632.3 g-WCW/L or 177.4 g-DCW/L final biomass concentration) was observed in experiment D.F, leading to a complete depletion of Mg, K and P.

Complete Mg depletion was observed in experiments D.C, D.D, D.H and D.I. In fact, Mg is always the first element to be depleted from the CM.

In experiments D.C and D.D, a pH shift from pH 5.0 to pH 7.0 at 120 h was tested. As response for these sudden pH increases some degree of precipitation may also be observed in these runs after pH shift.

All in all, these data shows that the composition of each element in BSM should be readjusted in order to mitigate precipitation problems and or limitation of some elements, namely Mg, K and P.

5.3.3 Kinetic modeling of elements consumption

The mathematical model previously described was calibrated with the ICP-AES data represented in Figure 5.3. The parameters were calibrated using data of four experiments (exp. D.A, exp. D.B, exp. D.D and exp. D.G). Model validation was performed with two independent experiments, exp. D.C and D.E. The three experiments with pH variations and/or severe precipitation of salts (exp. D.F, exp. D.H and exp. D.I) were excluded. The estimated maximum specific growth rate and element/biomass yields are given in Table 5.2 with the corresponding 95 % confidence intervals (CI). The confidence bounds for all parameters are generally quite low denoting the high sensitivity of model residuals to parameters and high statistical confidence of the estimated parameter values.

Table 5.2 Estimated parameters for the kinetic model.

Estimated parameters		CI
μ_max (h^{-1})	0.0504	0.0011
Y_Ca/X (g/g-DCW)	0.0009	0.0002
Y_K/X (g/g-DCW)	0.0630	0.0083
Y_Mg/X (g/g-DCW)	0.0129	0.0013
Y_P/X (g/g-DCW)	0.0871	0.0096
Y_S/X (g/g-DCW)	0.0151	0.0051

Figure 5.4 compares model predictions and experimental data for biomass and main elements (Mg, K, Ca, P and S) concentration for the six runs used for model calibration and validation. All modeling residuals of both calibration and validation experiments are within the 95 % confidence intervals for prediction. Moreover, modeling residuals of the validation data set are not higher than those of the calibration data set. This suggests that a simple stoichiometric description of salts consumption relatively to biomass growth is statistically consistent. *P. pastoris* biomass may thus be viewed as being composed by fairly constant inorganic salts contents. The two inorganic elements with higher amounts in the biomass are P and K.

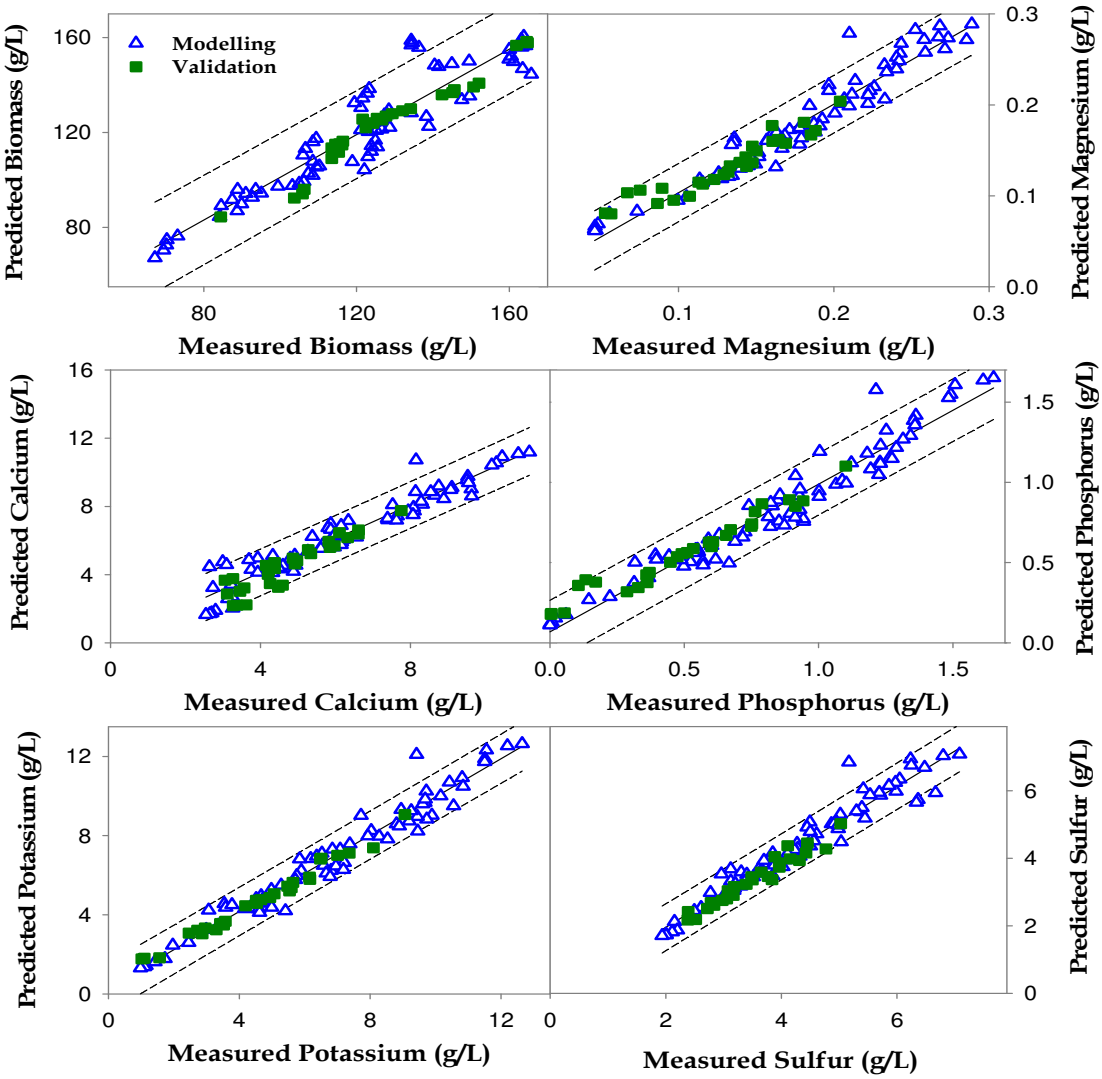


Figure 5.4 Modeling results for the biomass and salts concentration for six bioreaction experiments. Open symbols represent results from calibration experiments and full symbols the model validation experiments.

5.3.4 Optimization of medium formulation

The PMS medium formulation contains all the above elements (Mg, K, Ca P and S) through the addition of several salts. Table 5.3 shows the salts concentrations of several medium formulations published in the literature. Table 5.4 shows the corresponding elements concentration assuming complete salts dissociation and no precipitation. Table 5.5 shows the theoretical maximum biomass yielded by each element calculated on the basis of the yields presented in Table 5.2.

Table 5.3 Salts concentrations of the several medium formulations published in the literature.

Medium	Salts composition								
	$MgSO_4.7H_2O$ (g/L)	$CaSO_4$ (g/L)	$CaSO_4.2H_2O$ (g/L)	$CaCl_2.2H_2O$ (g/L)	H_2SO_4 (g/L)	KOH (g/L)	H_3PO_4 (mL/L)	$(NH_4)_2SO_4$ (g/L)	KH_2PO_4 (g/L)
PMS (Invitrogen)	14.90	0.93	--------	--------	18.20	4.13	26.70	--------	--------
FM22 (Laroche *et al.* 1994)	11.70	--------	1.00	--------	14.30	--------	--------	5.00	42.90
FM (d'anjou and Daugulis 2000)	4.70	--------	--------	0.36	--------	--------	--------	20.00	12.00
BSM/4 (Brady *et al.* 2001)	3.73	0.23	--------	--------	4.55	1.03	6.68	--------	--------
Opt. PMS	3.82	--------	0.22	--------	2.78	1.24	6.87	--------	--------

We have redesigned the composition of the main salts solution in order to decrease the standard deviation of the biomass that results from each salt. The final composition is presented in the last line of Table 5.3 (opt. PMS). This new formulation was named as "optimized *Pichia* main salts (opt. PMS)" solution.

Table 5.4 Elements concentration assuming complete salts dissociation and no precipitation.

Medium	Elements				
	P (g/L)	K (g/L)	S (g/L)	Mg (g/L)	Ca (g/L)
PMS (Invitrogen)	12.27	10.61	5.51	1.47	0.27
FM22 (Laroche *et al.* 1994)	19.53	31.06	5.55	1.15	0.23
FM (d'anjou and Daugulis 2000)	5.46	6.89	5.46	0.46	0.10
BSM/4 (Brady *et al.* 2001)	3.07	2.65	1.38	0.37	0.07
Opt. PMS	**3.16**	**1.98**	**1.05**	**0.38**	**0.05**

It may be seen large discrepancies in the calculated maximum biomass values (Table 5.5). For example, in the case of PMS the Mg concentration results in a maximum biomass concentration of 113.7 g/L while S is large excess resulting in a maximum biomass value of 364.4 g/L. As shown in Table 5.5 the biomass standard deviation in the opt. PMS formulation decreased to 38.9 % which is much lower than all other formulations presented.

Table 5.5 Theoretical maximum biomass yielded by each element

Medium	Biomass contribution by each element (g-DCW)					Biomass Standard deviation (%)
	P	K	S	Mg	Ca	
PMS (Invitrogen)	140.80	168.52	364.38	113.73	290.77	49.72
FM22 (Laroche *et al.* 1994)	224.11	493.32	367.03	88.97	247.69	53.83
FM (d'anjou and Daugulis 2000)	62.66	109.43	361.08	35.59	107.69	96.11
BSM/4 (Brady *et al.* 2001)	35.23	42.09	91.26	28.63	75.38	50.08
Opt. PMS	36.26	31.45	69.44	29.40	53.85	38.88

Summarizing, from the element yield values identified in this study, the composition of main salts was redefined as follows:

Optimized *Pichia* Main Salts (opt. PMS): H_3PO_4 85 % 6.87 mL/L, $CaSO_4.2H_2O$ 0.22 g/L, K_2SO_4 2.78 g/L, $MgSO_4.7H_2O$ 3.82 g/L and KOH 1.24 g/L.

This optimized PMS solution was added to the optimized (opt. PTM1) developed in the Chapter 4 to yield a new *P. pastoris* cell CM called "*Pichia* Functional Enviromics Medium 1 (PFEM1)". The composition of opt. PTM1 and PFEM1 are defined as follow:

Optimized *Pichia* Trace Metals solution 1 (opt. PTM1): $CuSO_4.5H_2O$ 12.00 g/L, NaI 0.16 g/L, $MnSO_4.H2O$ 6.00 g/L, $Na_2MoO_4.2H_2O$ 0.40 g/L, H_3BO_3 0.001 g/L, $CoCl_2.6H_2O$ 0.25 g/L, $ZnCl_2$ 40.00 g/L, $FeSO_4.7H_2O$ 3.25 g/L, H_2SO_4 10.00 mL/L and biotin 0.40 g/L.

Pichia Functional Enviromics Medium 1 (PFEM1): optimized *Pichia* Main Salts (opt. PMS) solution 995.65 mL/L, optimized *Pichia* Trace Metal 1 (opt. PTM1) solution 4.35 mL/L and glycerol 40.00 g/L.

PFEM1 is the medium that will be further tested in Chapter 6.

5.4 Conclusion

In this study we have investigated the consumption of the main elements in cultures of *P. pastoris* GS115 expressing a scFv through ICP-AES measurements and mathematical modeling. The experimental data revealed excess of Ca and S over Mg, P and K. In some cultures using the BSM medium, Mg, P and K deplete completely thus becoming limiting of biomass growth. Precipitation was detected for high pH between 6.5 and 7.0, which affected more Mg and Ca. In one culture, precipitation begins to occur during the MFB phase. We have concluded that a simple stoichiometric relationship is able to describe accurately the dynamics of elements concentrations over time, which allowed estimating with high statistical confidence the element/biomass yields. Based on the estimated yield values, an optimized CM formulation of the main elements was calculated for a given desired final biomass concentration. This new formulation was incorporated in the cell CM developed in Chapter 4 resulting in the new medium called *Pichia* Functional Enviromics Medium 1 (PFEM1).

Bibliography

Ariño J, Ramos J, Sychrová H. 2010. Alkali metal cation transport and homeostasis in yeasts. Microbiology and Molecular Biology Reviews 74(1): 95-120.

Bailey JE and Ollis DF. 1986. Biochemical Engineering Fundamentals, 2nd ed.; McGraw-Hill: New York, pp 500-507.

Brady CP, Shimp RL, Miles AP, Whitmore M, Stowers AW. 2001. High-level production and purification of P30P2MSP1 19, an important vaccine antigen for malaria, expressed in the methylotrophic yeast *Pichia pastoris*. Protein Expr Purif, 23: 468–475.

Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM. 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Current Opinion in Biotechnology 13(4): 329-332.

Cos O, Ramón R, Montesinos JL, Valero F. 2006. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review. Microbial Cell Factories 5:17.

d'Anjou MC and Daugulis AJ. 2000. Mixed-feed exponential feeding for fed-batch culture of recombinant methylotrophic yeast. Biotechnology Letters 22: 341-346.

Damasceno LM Pla I, Chang HJ, Cohen L, Ritter G, Old LJ, Batt CA. 2004. An optimized fermentation process for high level production of a single chain Fv antibody fragment in *pichia pastoris*. Protein Expr Purif. 37(1): 18-26.

Depue RH and Moat AG. 1961. Factors affecting aspartase activity. Journal of Bacteriology 82:383-6.

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

Jahic M, Knoblechner J, Charoenrat T, Enfors SO, Veide A. 2006. Interfacing *Pichia pastoris* cultivation with expanded bed adsorption. Biotechnol Bioeng. 93(6):1040-9.

Laroche Y, Storme V, Meutter J, Messens J and Lauwereys M. 1994. High–Level Secretion and Very Efficient Isotopic Labeling of Tick Anticoagulant Peptide (TAP) Expressed in the Methylotrophic Yeast, *Pichia pastoris*. Bio-technology (11): 1119-1124.

Martínez-Muñoz GA and Peña A. 2005. In situ study of K+ transport into the vacuole of *Saccharomyces cerevisia*. Yeast 22: 689-704.

Okorokov Lev A and Ludwig Lehle. 1998. Ca2+-ATPases of *Saccharomyces cerevisiae*: diversity and possible role in protein sorting. FEMS Microbiology Letters 162(1): 83–91.

Pichia fermentation process guidelines. 2000. Invitrogen Corporation.

Plantz BA, Nickerson K, Kachman SD, Schlegel VL. 2007. Evaluation of metals in a defined medium for *Pichia pastoris* expressing recombinant beta-galactosidase. Biotechnology Progress 23(3): 687-692.

Seo KH and Rhee JI. 2004. High-level expression of recombinant phospholipase C from Bacillus cereus in *Pichia pastoris* and its characterization. Biotechnology Letters 26(19): 1475-1479.

Spencer JFT and Spencer DM. 1997. Yeasts in Natural and Artificial Habitats. Springer-Verlag, Berlin- Heidelberg, p. 381.

Surribas A, Stahn R, Montesinos JL, Enfors SO, Valero F, Jahic M. 2007. Production of a Rhizopus oryzae lipase from Pichia pastoris using alternative operational strategies. Journal of Biotechnology 130(3): 291-299.

Teixeira AP, Dias JML, Carinhas N, Sousa M, Clemente JJ, Cunha AE, Stosch M von, Alves PM, Carrondo MJT, Oliveira R. 2011. Cell functional enviromics: Unravelling the function of environmental factors. Bmc Systems Biology 5.

Walker GM and Maynard AI. 1997. Accumulation of magnesium ions during fermentative metabolism in *Saccharomyces cerevisiae*. J Ind Microbiol Biotechnol 18(1): 1-3.

Willsky Gail R. 1979. Characterization of the Plasma Membrane Mg2+-ATPase from the Yeast, *Saccharomyces cerevisiae*. The Journal of Biological Chemistry 254(9):3326-3332.

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

Zhao HL, Xue C, Wang Y, Yao XQ, Liu ZM. 2008. Increasing the cell viability and heterologous protein expression of *Pichia pastoris* mutant deficient in PMR1 gene by culture condition optimization. Appl Microbiol Biotechnol 81(2): 235-41.

CHAPTER 6

Application of adaptive DO-stat feeding controller to *Pichia pastoris* cultures

Adapted from:

Ferreira AR, Ataíde F, Dias JML, Clemente J, Cunha AE, Oliveira R. 2012. Application of adaptive DO-stat feeding control to Pichia pastoris X33 cultures expressing a single chain antibody fragment (scFv). Bioprocess and Biosystems Engineering.

Abstract

In this chapter, fed-batch cultures of a *P. pastoris* X33 strain constitutively expressing the scFv under the control of the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter were performed in a pilot 50 L bioreactor. Due to the very high cell density achieved within the first 75 h, typically between 140 and 160 g-DCW/L of culture, most of the scFv is produced under high oxygen transfer limitation. To improve scFv productivity, direct adaptive dissolved oxygen (DO)-stat feeding controller that maximizes glycerol feeding under the constraint of available oxygen transfer capacity was developed and applied to this process. The developed adaptive controller enabled to maximize glycerol feeding through the regulation of DO concentration between 3 and 5 % of saturation, thereby improving process productivity. Set-point convergence dynamics are characterized by a fast response upon large perturbations to DO, follow by a slower but very robust convergence in the vicinity of the boundary with almost imperceptible overshoot. Such controller performance enabled operating closer to the 0 % boundary for longer periods of time when compared to a traditional proportional–integral–derivative algorithm, which tends to destabilize with increasing cell density. To finish, we study the application of this DO-stat feeding controller in two experiments of *P. pastoris* X33 expressing the scFv in a 50 L bioreactor with different medium, one experiment with the PFEM1 medium and other with the BSM medium. The results were very positive, 37.0 % increase of the average productivity (mass per unit time) with the PFEM1 medium in comparison to the DO-stat feeding controller with the BSM medium.

6.1 Introduction

A basic concern in *P. pastoris* cultivations is controlling the cell growth rate and the corresponding biomass profile over time. From a theoretical point of view, the final product titer increases with the biomass time integral when the product is cell growth dissociated as illustrated later in this study. High cell density above 130 g–DCW/L has been frequently reported (Cereghino and Cregg 2000) however the maximum biomass concentration and duration in *P. pastoris* cultures are ultimately constrained by the bioreactor maximum oxygen transfer capacity (Berdichevsky *et al.* 2011, Cunha *et al.* 2004). Air supplementation with pure oxygen is possible at small scale but cost prohibitive at large scale production (Lee *et al.* 2003). Maximization of the driving force for oxygen transfer and therewith the biomass concentration can naturally be achieved when the process is run at a minimum DO concentration. Operating at low DO concentrations is in principal not critical for non–fermentative substrates such as glycerol or methanol. Moreover, very low DO concentrations were reported to improve protein yield in some *P. pastoris* processes (Baumann *et al.* 2008, Charoenrat *et al.* 2005, Trentmann *et al.* 2004).

Under oxygen transfer limitation, maximizing the cell growth rate can be achieved by applying a DO–stat substrate feeding controller that maximizes substrate feeding while keeping DO slightly above the threshold limitation value (Lee *et al.* 2003, Lim *et al.* 2003, Oliveira *et al.* 2005). This work studies how such a DO–stat substrate feeding controller affects the expression of scFv by *P. pastoris* X33 under GAP control. Accurate control of DO concentration at very low levels <5 % is, however, not straightforward with a standard proportional–integral–derivative (PID) controller. Due to large variations in cell density, the DO dynamics are time varying and tendentially faster at late process stages, which complicates the implementation of a stable and robust PID with constant convergence dynamics over the whole process operating domain. Chung 2000, studied the application of a proportional–integral (PI) algorithm to DO-stat *P. pastoris* controller and concluded that when the rate of oxygen transfer approaches in magnitude the rate

of oxygen consumption, the potential for PI destabilization is greatest, and that periodic PI retuning might be necessary to ensure stability.

To overcome such problems, in this work we have applied an adaptive controller algorithm, which is exponentially stable with quasi time-invariant convergence dynamics in the whole process operation domain permitting a stable operation at low DO concentration and therewith maximizing the overall process productivity.

6.2 Material and methods

6.2.1 Yeast strain, culture medium and inoculum

A stable *P. pastoris* X33 strain constitutively expressing a scFv antibody under the control of GAP promoter was used for this work. The composition of BSM used for the cultures was the same as described in Chapter 2.1 as well as the protocol of the inoculum growth. The optimized CM formulation (PFEM1) was the one described in Chapter 5.3.4.

6.2.2 Bioreactor operation

The *P. pastoris* X33 fed-batch cultivation process was divided in three phases:

Glycerol batch (GB) phase – The bioreactor was operated for 30 hours in batch mode starting with a glycerol concentration of 40 g/L. DO drops very slowly and remained close to saturation levels.

Glycerol fed–batch (GFB) phase – An exponential feeding program was initiated after 30 hours of operation,

$$F_{Gly} = F_{Gly,0} \times e^{\mu \times t} \qquad (6.1)$$

With F_{Gly} being the glycerol feeding rate in g/h, $F_{Gly,0}$ the initial feeding rate (65 g/h) and μ the desired specific growth rate, set at 0.16 h^{-1}. It was in this phase that cell density significantly increased and DO decreased more rapidly. Once the DO reached the threshold level of 50 %, it was kept at that level by automatic closed loop control, manipulating the agitation rate between 300 and 1000 rpm.

Oxygen transfer limitation (OTL) phase – Once the agitation rate reached the maximum level of 1000 rpm, DO decreased very rapidly and the exponential glycerol feeding program was then terminated. From this point on, the DO was kept constant at a low level (e.g. 3–5 %) by closed–loop manipulation of the glycerol feeding rate (DO–stat feeding controller strategy) using the adaptive controller algorithm developed in this work.

6.2.3 Analytical methods

Cell density was measured by OD_{600nm} and WCW and the product quantification by ELISA as described in Chapter 2.3.

6.2.4 Design of an adaptive DO-stat controller

DO dynamics in the OTL phase can be approximated by the following simplified material balance equation, which assumes glycerol limitation and negligible accumulation in the liquid phase:

$$\frac{dC_0}{dt} = -\frac{F}{V} \cdot S_0 \cdot Y_{OS} + k_L a \cdot \left(C_0^* - C_0\right) \tag{6.2}$$

With C_0 the dissolved oxygen concentration, F the glycerol solution feeding rate, V the broth volume, S_0 the glycerol concentration in the feed solution, Y_{OS} the observed oxygen/glycerol yield, $k_L a$ the oxygen mass transfer coefficient, C_O^* the saturation dissolved oxygen concentration at working conditions.

Equation 6.2 can be rearranged into the following form:

$$\frac{dx}{dt} = -a_p(t) \cdot x + k_p(t) \cdot F \tag{6.3}$$

With,

$$x = 100 - DO \tag{6.4}$$

$$k_p = \frac{S_0 \cdot Y_{OS}(t) \cdot 100}{V(t) \cdot C_O^*(t)} \tag{6.5}$$

$$a_p = k_L a(t) \tag{6.6}$$

Due to the time–varying nature of Y_{OS}, $k_L a$, V and C_O^*, Eq. 6.3 represents a linear time–varying dynamical system. Nevertheless, as shown later the dynamics of k_p and a_p are much slower than x (and DO), which renders Eq. 6.3 to be a linear quasi time–invariant dynamical system (Narendra and Annaswamy 1989).

Assuming the general form of DO dynamics given by Eq. 6.3, a control law was derived for controlling x to a desired set–point x^* by manipulating the glycerol feeding rate, F. A first order linear reference model was defined:

$$\tau_c \cdot \frac{dx}{dt} = -x + x^* \tag{6.7}$$

With τ_c being a first order time constant, a design parameter that can be used to set the controller speed of convergence. Combining Eq. 6.3 and 6.7, results in the following controller equation:

$$F = \theta(t) \cdot x + K(t) \cdot x^* \tag{6.8}$$

which can be interpreted as the controller that forces the process (Eq. 6.2) to mimic the reference model (Eq. 6.7). The θ and K parameters are related to process parameters according to the following equations:

$$\theta(t) = \frac{a_p(t) - \frac{1}{\tau_c}}{k_p(t)} \tag{6.9}$$

$$K(t) = \frac{1}{\tau_c \cdot k_p(t)} \tag{6.10}$$

Since a_p and k_p are in general unknown and time–varying, it is a necessary condition for the implementation of Eq. 6.8 to adapt θ and K on–line in order to ensure stability. We propose the following adaption scheme, which may be shown to be uniformly stable in a Lyapunov sense:

$$F = \hat{\theta}(t) \cdot x + \hat{K}(t) \cdot x^* \tag{6.11}$$

$$\tau_c \cdot \frac{d\hat{x}}{dt} + \hat{x} = x^* \tag{6.12}$$

$$\frac{d\hat{\theta}}{dt} = -\gamma \cdot x \cdot \left(x^* - \hat{x}\right) \tag{6.13}$$

$$\frac{d\hat{K}}{dt} = -\gamma \cdot x^* \cdot \left(x^* - \hat{x}\right) \tag{6.14}$$

Equations 6.11 and 6.12 are derived from Eqs. 6.8 and 6.7 by substituting θ, K and x by their estimates $\hat{\theta}$, $\hat{K}$ and $\hat{x}$, respectively. The choice of Eqs. 6.13 and 6.14 to estimate $\hat{\theta}$ and $\hat{K}$ arises from stability requirements. A dynamical system is said to be asymptotically stable in the Lyapunov sense if a Lyapunov candidate function exists that (1) is positive definite and (2) its first derivative is negative definite (Narendra and Annaswamy 1989). The form of Eqs. 6.13 and 6.14 ensures that the first derivative of the candidate Lyapunov function is negative definite. Note that the parameter γ is an additional degree of freedom that can be used to tune the speed of convergence of $\hat{\theta}$ and $\hat{K}$ to their true values.

6.3 Results and discussion

6.3.1 Adaptive DO-stat feeding controller

Figure 6.1 illustrates the application of the previously described glycerol feeding and DO control strategy to a pilot *P. pastoris* process. The GB, GFB and OTL phases lasted 26, 21 and 53 h, respectively. The adaptive controller worked between 47 and 100 h, starting at a cell density of 103.0 g-DCW/L and finalizing at 132.1 g-DCW/L. In this first batch, the process was subject to a set of perturbations in the DO set–point (between 3–50 % of oxygen saturation), agitation rate (between 900 and 1000 rpm) and pressure (left uncontrolled up to 800 mbar overpressure). These perturbations were intentional in order to assess control stability, tracking characteristics and robustness in relation to the tuning parameters τc and γ.

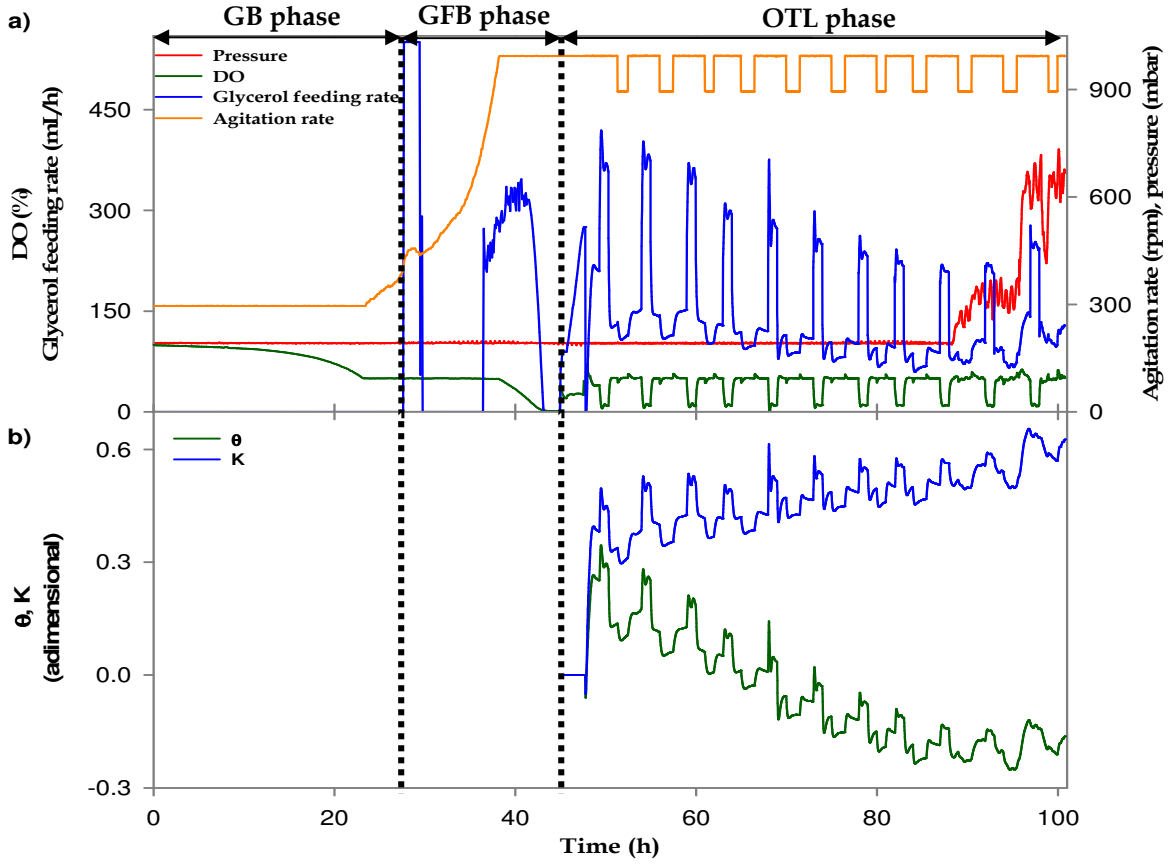


Figure 6.1 Overall glycerol feeding and dissolved oxygen adaptive controller results in a pilot 50 L *P. pastoris* process. a) DO, glycerol feeding rate, agitation rate and overhead pressure over time; b) adaptive parameters K and θ over time.

It is worthwhile to note the impact of DO set–point changes in the glycerol feeding rate (Fig. 6.1). As expected, decreasing the DO set–point resulted in a large increase of the glycerol feeding rate. For instance, decreasing the DO set–point from 50 % to 3 % in the beginning of the OTL phase resulted in more than threefold increase in the glycerol feeding rate, from 120 to 380 mL/h. This translates into an average glycerol feeding increase of 7.3 g–glycerol per 1 % DO per hour. A central aspect in this control strategy is that keeping the DO set–point constant, which implies constant oxygen transfer rate, does not maintain the feeding of glycerol at constant levels. Indeed, a continuous decrease of the glycerol feeding rate is observed until converging to a plateau. For a DO set–point of 3 %, the feeding rate starts at 380 mL/h converging to 200 mL/h just before the pressure perturbation at t = 88 hours. The same trend is observed for the DO–set point of 50 %. These results are consistent with a continuous decrease in the cell growth rate and consequent increase in the oxygen/glycerol yield. As pointed out by Jahic *et al*. 2002, the oxygen flux, q_O (g/L.h), has two terms, one for the anabolism and the other for the catabolism:

$$q_O = q_{S,an} \cdot Y_{OS,an} + q_{S,en} \cdot Y_{OS,en} \tag{6.15}$$

When the substrate is glycerol, glycerol flux for anabolism $q_{S,an}$ and glycerol flux for catabolism $q_{S,en}$, the theoretical yields are $Y_{OS,an}$ = 0 and $Y_{OS,en}$ = 1.217 (w/w) for anabolism and catabolism respectively. As such, the continuous cell growth rate decrease implies that an increasingly higher percentage of glycerol is catabolized for energy production with characteristic $Y_{OS,en}$ = 1.217 (w/w). Also, when the DO set-point is changed between 3 and 50 %, strong perturbations in the observed yield Y_{OS} and indirectly in θ and K are introduced. Sharp variations in the glycerol feeding rate shift the metabolism toward a higher or lower relative catabolic activity, which translates into abrupt variations in the observed yield Y_{OS}. Since θ and K are inversely proportional to $k_p(t)$ (Eqs. 6.9 and 6.10) and since $k_p(t)$ is directly proportional to Y_{OS} (Eq. 6.5), step changes in the DO set-point result into perturbations in θ and K in the opposite direction as shown in Fig. 6.1b.

In parallel to the cell density, cell growth rate and Y_{OS} yield, also the culture volume changes considerably over time. The volume starts at 15 L and ends up at 30 L in the end of the OTL phase. These variations, incorporated in the process model Eq. 6.3 to 6.6 in the form of time–varying parameters, illustrate the time–varying nature of DO dynamics. The proposed adaptive controller algorithm presents the advantage of automatically adjusting to such intrinsic time–varying dynamics, which is patent in the evolution of adaptive parameters θ and K in Fig. 6.1b. Besides the short–term adjustments caused by process perturbations, these parameters dynamics also reflect a long–term adaptation to the intrinsic time–varying DO dynamics.

6.3.2 Effect of tuning parameters τ_c and γ

Figure 6.2 shows the controller response to step changes in the DO set–point for different values of τ_c (60, 150 and 300 s) and constant γ = 1500. The full lines represent the reference model output for different τ_c values, while the lines with symbols represent the actually measured control response.

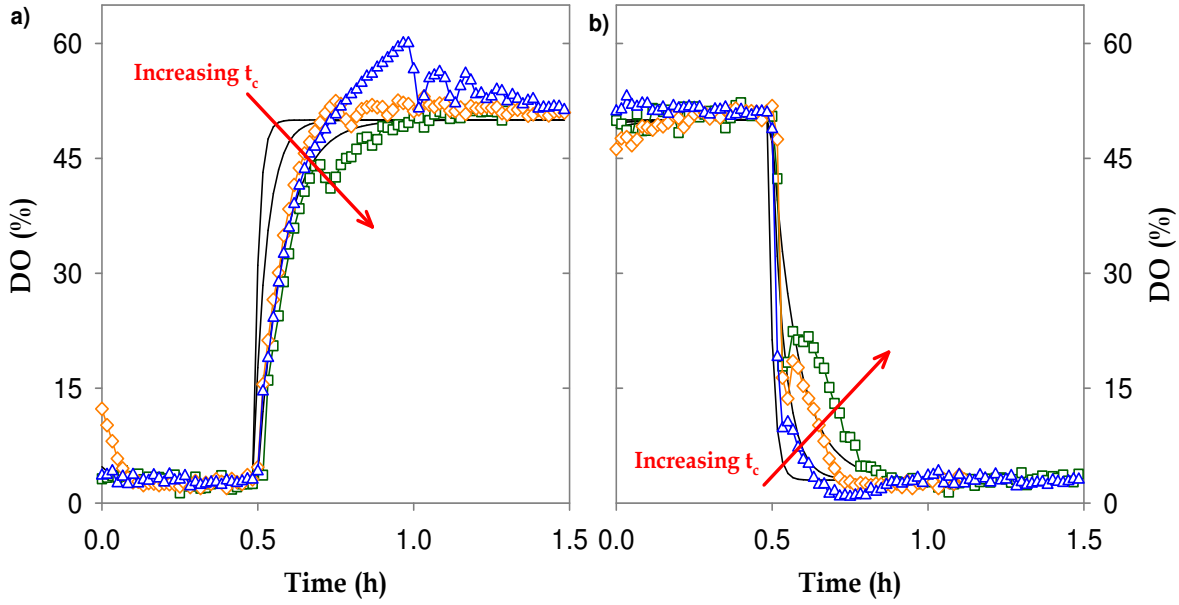


Figure 6.2 Adaptive DO–stat controller response to step changes in DO set–point: a) 3 → 50 % and b) 50 → 3 % for varying τ_c, (◊) – τ_c = 60; ( ) – τ_c = 150; (Δ) – τ_c = 300; with constant γ = 1500. Full lines represent the reference DO concentration (in % of saturation).

As general trend it can be observed that decreasing τ_c results in a faster response. The response to DO set–point down perturbations closely matches that of a first order dynamical response, which was a design condition, but the same cannot be said for up perturbations. Even for down perturbations, there is a slight overshoot when τ_c is 60 or 150 s, which is, however, practically eliminated when τ_c increases to 300 s. There are several factors that could explain the mismatch in relation to the reference model. A DO probe has a response time of typically 15-40 s, which alone can cause oscillations for low τ_c values. Simulations have shown that with a probe delay of 40 s and τ_c = 90 s, the oscillations are highly attenuated and almost imperceptible (results not shown). The markedly slower and more oscillatory response to up perturbations is, however, caused by glycerol accumulation dynamics. The difference in the responses for set-point up and down perturbations is mostly due to the fact that (1) the physical limit of the actuator is reached for up perturbations, i.e. the glycerol feeding rate cannot be lower than 0 (L/h); and (2) for up perturbations, the accumulated glycerol takes some time to be taken up. Assuming that the maximum glycerol uptake rate is 0.37 g/g.h (Jahic *et al.* 2002) and that the biomass varies between 125 g-DCW/L and 175 g-DCW/L in the oxygen transfer limitation phase, the time required to consume 1 g/L of glycerol varies between 55 and 78 s. Since increasing DO implies a decrease in glycerol concentration, the controller time constant cannot be set faster than the characteristic glycerol uptake rate by the cells. This explains why the initial response is practically insensitive to the τ_c value. The three curves diverge subsequently because of the accumulated error in the adaptation of θ and K, which is more severe for lower τ_c values. In fact, for τ_c = 60 s high frequency oscillations appeared suggesting an unstable operational state.

The other important factor to consider is the value of γ, which interferes in the adaptation dynamics of θ and K, namely in the adaptation speed. Figure 6.3 shows the effect of changing the γ value. The important point is that the

adaptation speed needs to be balanced with the controller speed. It can be seen in Fig. 6.3 that for τ_c = 60 s, decreasing the γ value brings about a faster convergence response with increasing overshoot, particularly for up DO set-point perturbations. As rule of thumb, the convergence to the set-point should not be set faster than the adaptation law convergence dynamics. Otherwise, the control will always exhibit oscillations and limit unstable behavior. The last perturbation in Fig. 6.3 shows the controller response with τ_c = 90 s and γ = 3000, which corresponds to a relatively fast controller tuning combined with a relatively slow adaptation law. It can be observed that the DO response is almost perfectly pictured by a first order convergence without overshoot or oscillations. The elimination of the overshoot was possible for both up and down perturbations and is particularly impressive in the convergence to the 3 % level.

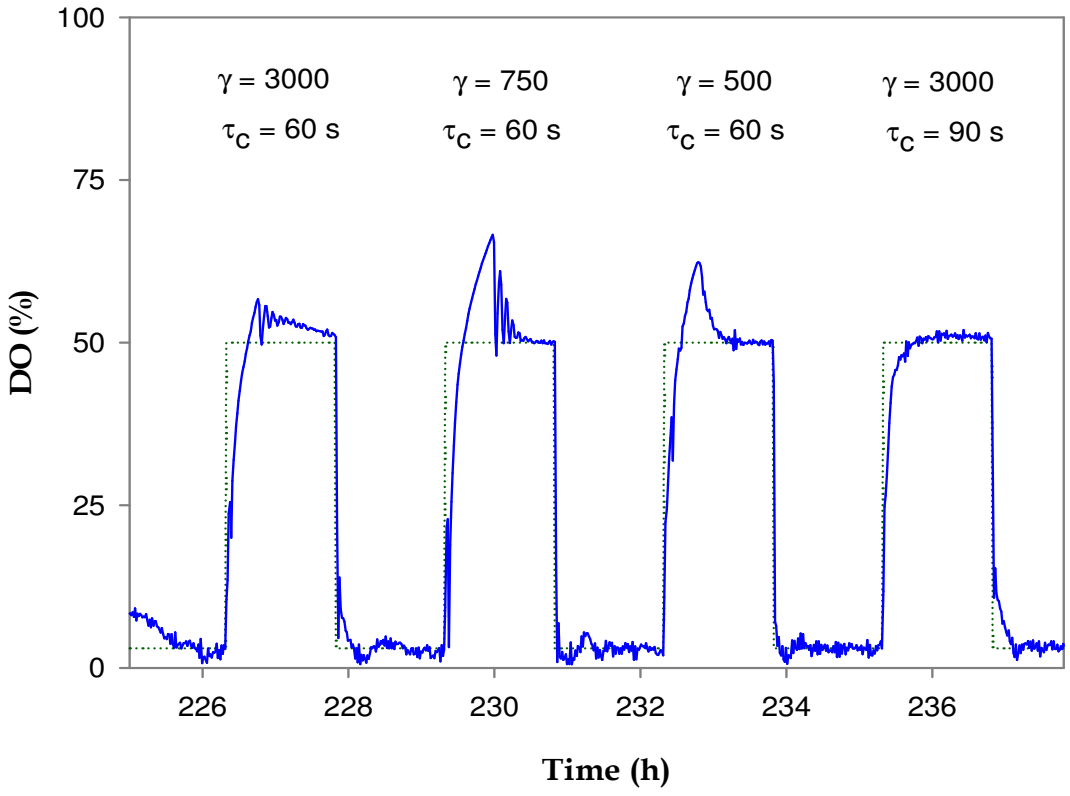


Figure 6.3 Response to step changes in DO set–point: 50 → 3 → 50 % for different values of τ_c (60 s and 90 s) and γ (500, 750 and 3000).

Summing up, we propose the following rules to tune the controller parameters:

1. The controller time constant, τ_c, should be set higher than the DO probe time delay (between 15 and 40 s) and also higher than the glycerol uptake time constant, estimated to vary between 55 and 80 s in the OTL phase. The value τ_c = 90 s seems to be a good choice.

2. The value of c sets the speed of adaptation of θ and K to their ''true'' values. As a rule of thumb, the speed of adaptation should not be set slower than the controller convergence time constant. Once τ_c is fixed, γ should be incremented from 500 up to the point where oscillations are sufficiently attenuated. In the present case, for τ_c = 90 s the value of γ = 3000 produced almost an oscillation free response. This particular tuning (τ_c = 90 s and γ = 3000) seems to be adequate for difficult boundary control problems and, therefore, it has been used in subsequent productivity assessment experiments as explained below.

6.3.3 Response to process charges

Figure 6.4 shows the controller response to changes in agitation rate from 1000 to 900 and then back to 1000 rpm for different values of τ_c and γ. Step changes in the agitation rate impact on the parameter a_p (=k_La) and in θ (see Eq. 6.9).

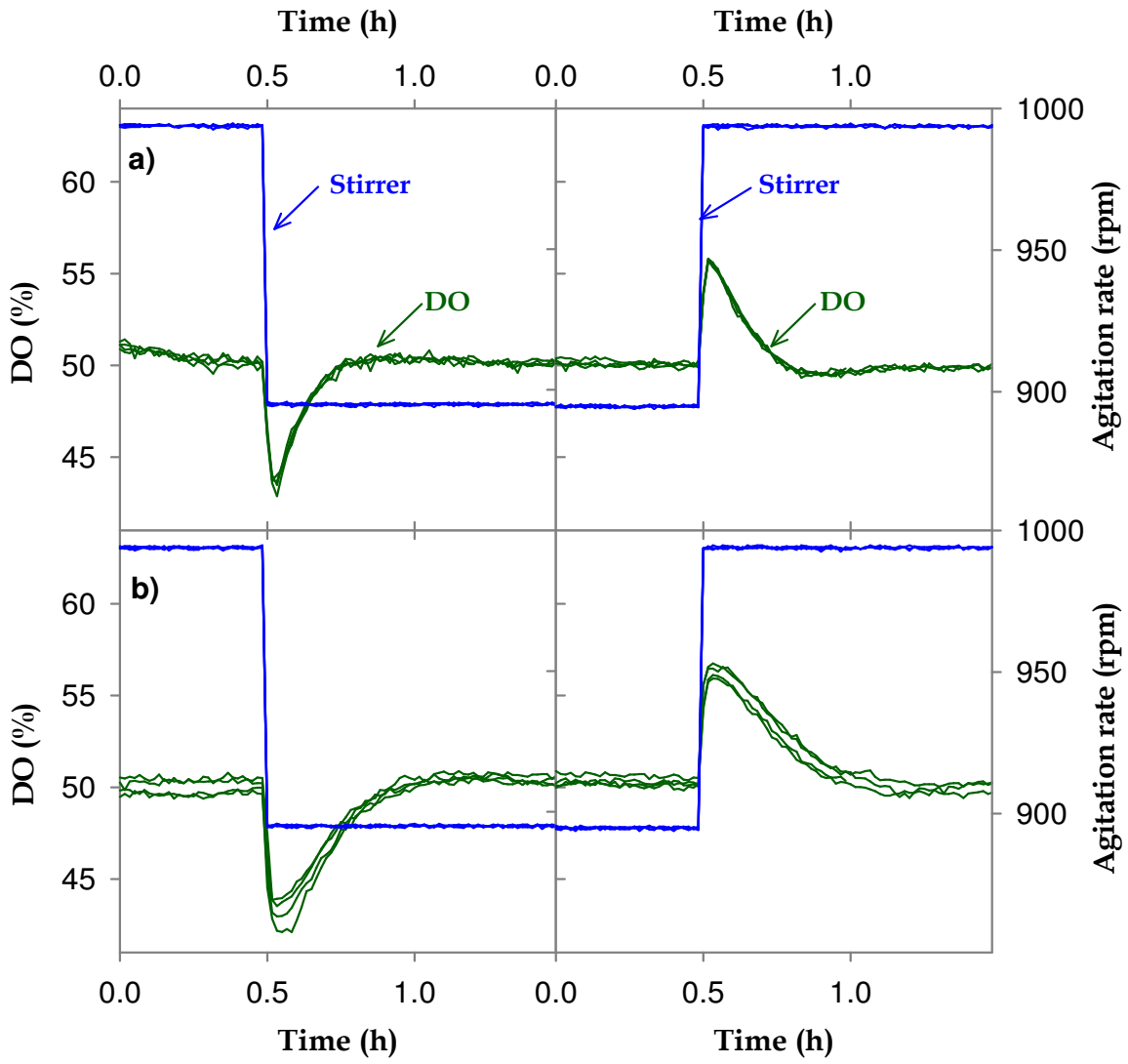


Figure 6.4 Response to step changes in agitation rate set–point 1000 → 900 → 1000 rpm for varying τ_c, (60; 120; 180; 240) and constant γ: a) γ = 750 and b) γ = 1500.

These perturbations caused a deviation to the setpoint below 10 % of oxygen saturation concentration. The convergence to the set-point follows exactly the expected profile, i.e. a first-order convergence without oscillations in agreement with the reference model. The response in the case of τ_c = 60 s has the greatest amplitude, which is consistent with the theory, but no high frequency oscillations are observed in any of the cases. One important point here is that the DO is relatively high and varies only very little (50 ± 7 %). At lower DO concentrations, the glycerol uptake kinetics is likely to be more nonlinear and to result in a more oscillatory response.

Figure 6.5 shows the controller behavior when the overhead pressure is left uncontrolled between 200 and 800 mbar.

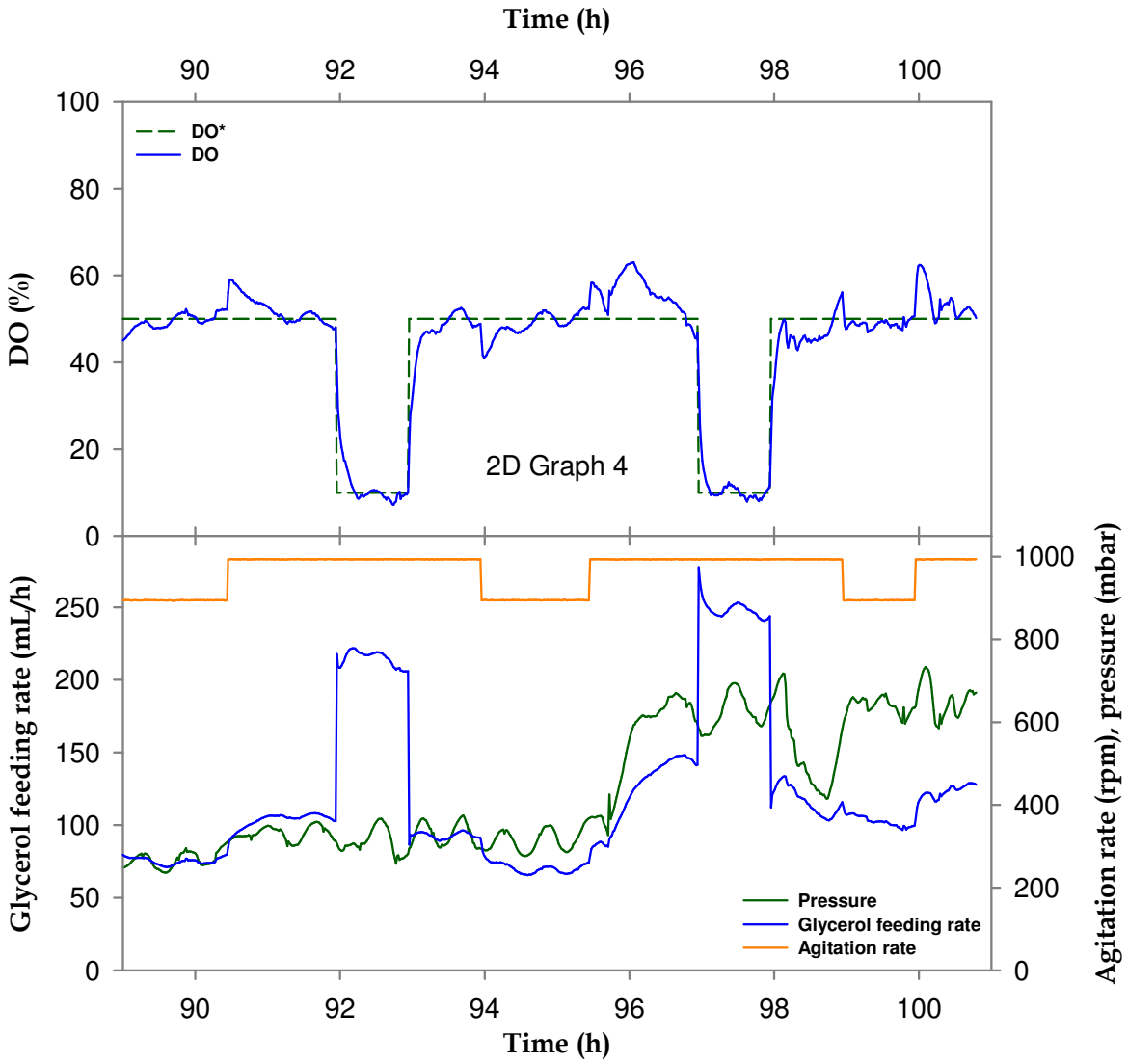


Figure 6.5 DO responses to perturbations in overhead pressure left uncontrolled between 200 and 800 mbar.

According to the Henry's law, 800 mbar overpressure corresponds to a 1.8 times higher oxygen solubility than at 1 bar. This corresponds to an almost twofold increase in the oxygen transfer rate, which in principle would permit increasing the glycerol feeding rate. Indeed, an increase in glycerol feeding rate from 200 mL/h up to 267 mL/h is observed (33.5 % increase), attenuated by the fact that most of the glycerol is being consumed catabolically with characteristic Y_{OS} = 1.217 w/w (Jahic *et al.* 2003). Although the quality of the control degrades considerably due to erratic pressure variations, its behavior is markedly stable.

Fluctuations in the overhead pressure are transmitted to the DO control, but they are effectively dampened. Moreover, the controller response to large step changes in the set–point between 50 – 3 % is effectively handled even under such fast changing pressure variations. It is remarkable that given the high perturbations in pressure, the controller still manages to control DO to levels as low as 3 % of saturation.

6.3.4 Analysis of process performance

Product expression kinetics

In this section it is shown, that the utilization of the adaptive control strategy has a direct impact on the productivity, in that it allows prolonging the process at very high biomass concentration which in turn results in higher final product titer and productivity. Figure 6.6 shows in the on–line measured cumulative glycerol feeding profile, off–line measured biomass concentration over time and product titer over time in arbitrary units, obtained in four independent pilot *P. pastoris* X33 cultures. The four batches were executed at a constant temperature of 30 ^{o}C and constant DO set–point of 5 % during the OTL phase. The variations in the times of the GB phases shown in Figure 6.6a) are due to differences in the initial concentrations of the inoculum which was being optimized at this stage. The time intervals of the GFB phases are A (57.0 – 67.9 h), B (76.3 – 88.2 h), C (31.6 – 45.4) and D (58.3 – 73.1 h). Batches A and B suffered pH perturbations at some point of operation, specifically from pH 5.0 to 4.0 at time 88.9 hours in batch A and from pH 5.0 to 4.0 and to 3.0 at times 96.2 and 123.1 hours respectively in batch B. The batches C and D pH are constant at 5.0 over time.

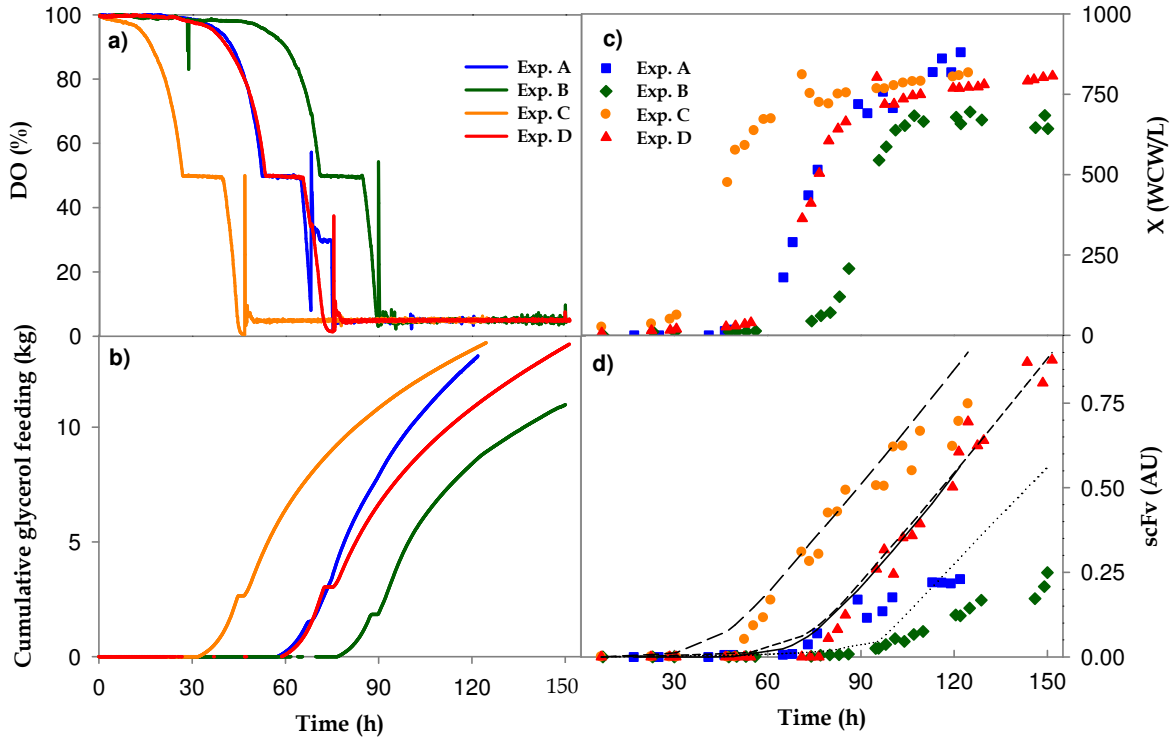


Figure 6.6 DO, Glycerol feeding, biomass and product profiles obtained in 4 independently operated 50 L bioreactor experiments (exp. A, exp. B, exp. C and exp. D) using the adaptive DO–set feeding controller with τ_c = 90 and γ = 3000. a) on–line measured of DO (%); b) on–line measured of cumulative glycerol feeding (Kg); c) off–line measured of biomass concentration (g–WCW/L) over time; d) scFv concentration over time in arbitrary units (AU).

Figure 6.6d) plots product titer along cultivation time with symbols representing measurements and lines the simulation of a cell growth dissociated kinetic model:

$$\frac{d(P \cdot V)}{dt} = \beta \cdot (X \cdot V) \tag{6.16}$$

With constant specific protein synthesis rate [β = 1.44×10^{-4} AU/(g–WCW·h)] from the beginning to the end of the process. Remarkably, this very simple model is able to reproduce very faithfully the product build up in batches A and B and also in C and D up to the time point where pH perturbed. These results support a constant specific protein synthesis rate of 1.44×10^{-4} AU/(g–WCW·h) at a temperature of 30 °C and pH 5.0, which is however pH dependent and possibly temperature dependent.

For cell growth dissociated product kinetics with constant specific productivity, the product titer can be calculated by integrating Eq. 6.16, taking the following simple form:

$$P(t) = \frac{\beta}{V(t)} \cdot \int_0^t X(\tau) \cdot V(\tau) \cdot d\tau \tag{6.17}$$

This equation shows that both product titer and productivity strongly depend on the time integral of biomass. Batches C and D, which have higher biomass concentration (Fig. 6.6c) are also the ones with higher productivity (Fig. 6.6d). Batch A is at first sight an exception to this scenario since its biomass profile is almost coincident to that of batch B but this exception can be explained by the pH drop from 5.0 to 4.0 at time 88.9 hours, which seems to have not the same detrimental effect on cell growth as it has on product synthesis.

Effect of glycerol feeding

The integral of biomass concentration can be controlled through the glycerol feeding rate. Fig. 6.6b) shows the cumulative glycerol feeding for the four batches. Overall, these results evidence a consistent relationship between glycerol feeding, cell growth and product synthesis. Higher glycerol feeding rates bring about higher biomass concentrations, higher biomass time integrals and eventually higher product titer as defined by Eq. 6.17. Again batch A arises as an apparent exception. The cell biomass growth profile in batches A and D are similar but more glycerol is consumed in batch D due to higher process duration. To strengthen this hypothesis, the cumulative glycerol feeding in both experiments is practically coincident up to the time point when pH is perturbed in batch A.

Effect of adaptive DO-stat glycerol feeding controller

Figure 6.7b) shows the adaptive glycerol feeding controller results for batch D, which is the one with highest final product titer. The DO control remained remarkably accurate during the whole OTL phase. The glycerol feeding rate decreases smoothly over time concomitantly with a continuous decrease of specific growth rate. The $\theta(t)$ and $K(t)$ parameters also change very smoothly over time to compensate for the time–varying process dynamics corroborating with the previously raised hypothesis of linear quasi time–invariant DO dynamics. The convergence dynamics to the set–point remain fairly constant and the quality of the control does not degrade with increasing cell density as observed in PID control, Fig. 6.7c). The PID (standard form) is, as usually, tuned for the initial phase of the cultivation. In the presented case the well-known Ziegler-Nichols method that utilizes ultimate gain and frequency found application for the tuning of the PID parameters.

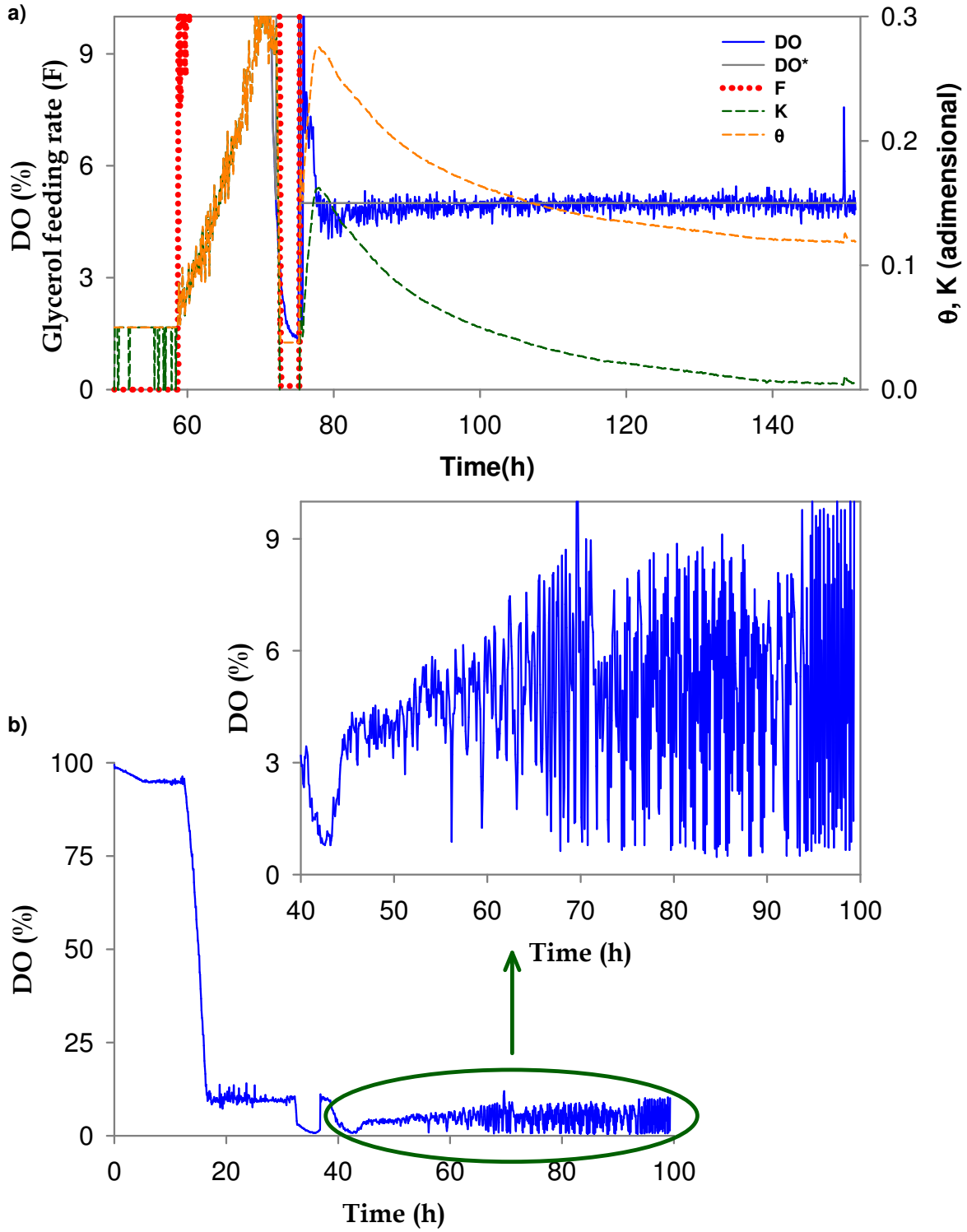


Figure 6.7 Comparison between adaptive DO–stat feeding and PID DO–stat feeding. a) adaptive DO–stat feeding results for batch B with τ_c = 90 and γ = 3000; plotted variables over time are measured DO, DO* set–point, glycerol feeding rate, controller adaptive parameters, $K(t)$ and $\theta(t)$; b) DO–PID control results with DO set–point 5 %; plotted variable is measured DO over time.

These parameters were thereupon fine-tuned in order to decrease overshooting responses to disturbances, resulting, for a given sampling period of 2 s, in the following parameter values K_P = 0.2, τ_I = 2500 s, τ_D = 1320 s. However, when this PID is applied, the DO concentration becomes noisier with increasing frequency and amplitude of oscillations over time. The degradation of control quality at late but still productive process stages was, in case of applying a PID, repeatedly observed in many other batches, even for higher DO set–point than 5 % (results not shown). This performance decrease is due to the increase in cell density and the corresponding decrease of the oxygen time constant, i.e. the ultimate gain and frequency vary along the cultivation time. As shown by Chung 2000 for a *P. pastoris* process, a PI controller may diverge at some point of operation unless a gain scheduling strategy is implemented.

The quality of the DO–stat feeding controller is highly relevant for the productivity of *P. pastoris* processes. When exponential stability and quasi time invariant response are ensured, the DO can be safely controlled closer to the 0 % boundary than when using a PID algorithm, as clearly evidenced in Fig. 6.7b). The results achieved of adaptive controller with DO in the range of 3–5 % were very satisfactory, which are not possible to obtain with a PID algorithm. By operating closer to the 0 % boundary, the glycerol feeding can be maximized, which translates into higher biomass concentration and ultimately in higher product productivity as shown before in Fig. 6.6. The ability to maximize the glycerol feeding is particularly relevant in the end of the process since it enables to prolong the process at almost constant biomass concentration. This becomes possible because enough carbon is supplied to sustain maintenance requirements thus maintaining a healthy culture for longer periods of time. The ability to prolong the process at very high biomass concentration has a large impact on the integral of biomass over time resulting in higher final product titer and productivity.

6.3.5 Implementation of adaptive DO–stat controller with optimized culture medium PFEM1

It was demonstrated in section 6.3.4 substantial improvements of scFv productivity due to the adaptive DO-stat controller with four experiments using the BSM medium (baseline formulation). Here, we study the application of this DO-stat feeding controller together with the optimized CM developed in Chapters 4 and 5 (PFEM1). More specifically, the *P. pastoris* X33 strain expressing the scFv was cultured under DO-stat feeding controller in the optimized CM PFEM1 (experiment E) in very similar conditions to the ones verified in the experiment D, which was the culture that yielded more product with the BSM medium.

Due to the low initial salts concentration in PFEM1, it was necessary to feed a concentrated salts solution (concentrated PMS) along time during the GFB phase. The composition of the concentrated feeding solution was the following:

Concentrated PMS: H_3PO_4 85 % 68.70 mL/L, $CaSO_4.2H_2O$ 2.20 g/L, K_2SO_4 27.80 g/L, $MgSO_4.7H_2O$ 38.20 g/L, KOH 12.40 g/L.

Since salts consumption correlates with biomass growth, a PMS solution feeding program was implemented based on the ammonia feeding rate, F_{base}. To note that the ammonia feeding control loop is based on the addition of a 25 % NH_4OH solution to regulate pH at a constant level. In this way it is ensured that the ratio salts/ammonia feeding is held constant throughout the cultivation. It was however found through simulations (results not shown) two phases of PMS feeding which maintain approximately constant the concentration of salts in the bioreactor. The final feeding program was as follows:

$$\begin{cases} F_{salts} = 1.95 \times F_{base} & \text{in the GB phase} \\ F_{salts} = 2.89 \times F_{base} & \text{in the GFB and OTL phases} \end{cases} \quad (6.18)$$

The experiment D was performed at a constant temperature of 30 ^{o}C, constant pH 5.0 and constant DO set–point of 5 % during the OTL phase, using the BSM

medium. In experiment E, all the above parameters were the same, except for the CM. Instead of the BSM, the PFEM1 formulation was used together witch the concentrated PMS feeding program specified by Eq. 6.18.

A comparison between the two experiments can be seen in Figures 6.8 and 6.9. The differences in the duration of the GB phases shown in Figure 6.8 are due to differences in the initial concentrations of the inoculum, which was optimized in experiment E resulting in a much shorter lag phase. In experiment D the GB phase took 59 hours, decreasing by approximately 30 hours in experiment E.

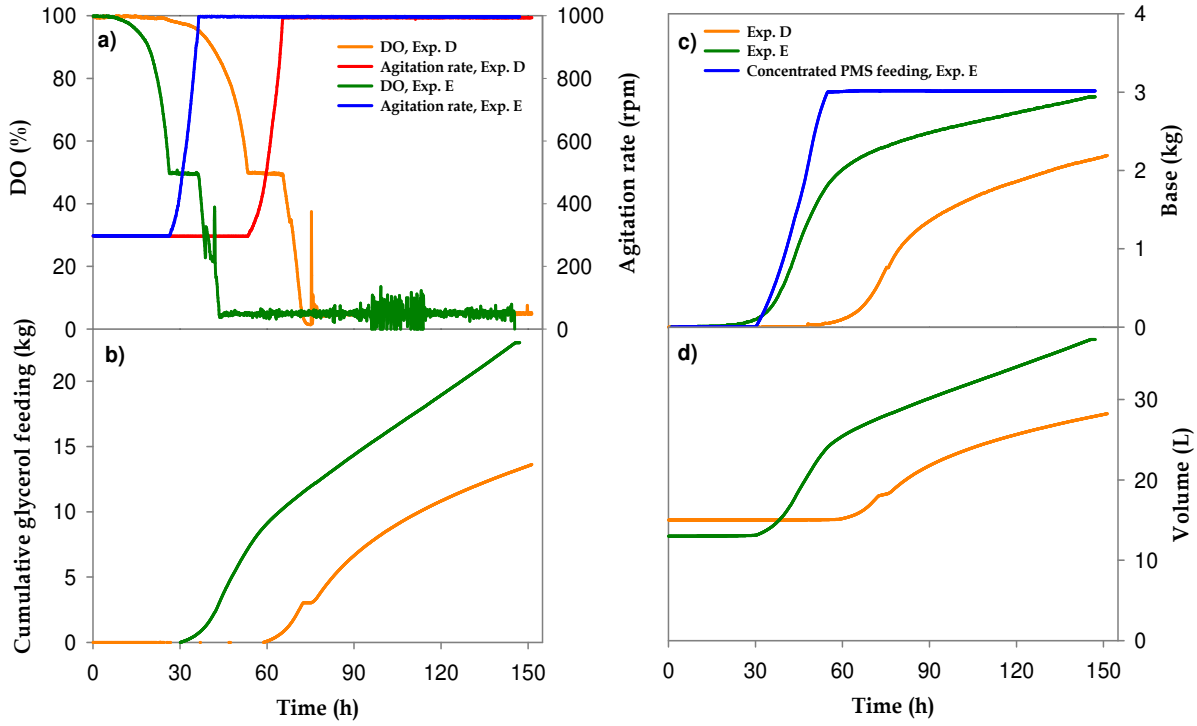


Figure 6.8 DO and agitation rate, Glycerol feeding, Base feeding and total Volume profiles obtained in 2 independently operated 50 L bioreactor experiments (exp. D and exp. E). a) on–line measured of DO (%) and agitation rate (rpm); b) on–line measured cumulative glycerol feeding (Kg); c) on–line measured cumulative Base feeding (Kg) and cumulative PMS feeding (Kg); d) calculated culture Volume (L).

To maintain constant DO set-point of 5 % the feeding of glycerol was substantially higher in the PFEM1 experiment (exp. E) than in the BSM experiment (exp. D) as

seen in Figure 6.8b. Also the base addition was considerably higher in the PFEM1 experiment than in the BSM experiment (Figure 6.8c). The direct consequence of this was a much higher culture volume in the PFEM1 experiment than in the BSM experiment (37.5 L and 28.2 L respectively). To note that the total volume of salts solution fed was 3.0 L, which is 8 % of the total culture volume only.

From several experiments performed in the past, it could be observed that the duration of the lag phase does not correlate with productivity variations as long as the duration of the lag is discounted from the productivity calculation. As such, in the productivity analysis shown below in Figures 6.9 we have synchronized both experiments to the time point where the glycerol feeding starts in Figures 6.9c) (Biomass) and 6.9d) (Product).

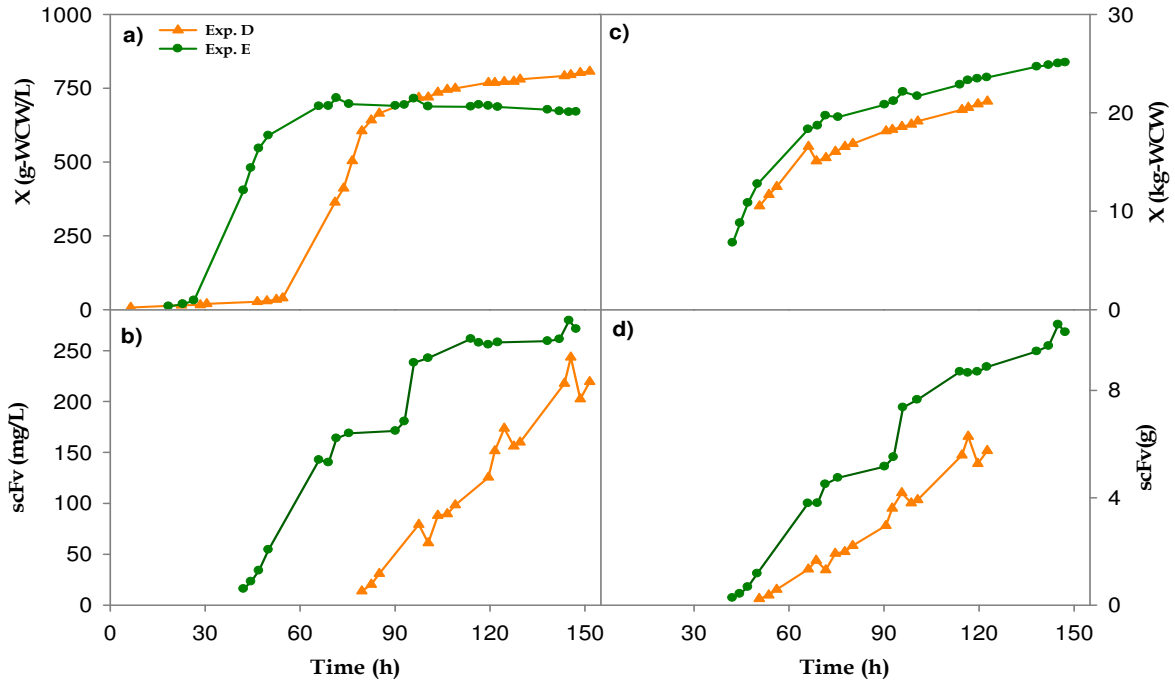


Figure 6.9 Biomass and scFv concentration profiles obtained in 2 independently operated 50 L bioreactor experiments (exp. D and exp. E). a) off-line measured biomass concentration (g-WCW/L) over time; b) off-line measured scFv concentration (mg/L) over time; c) total biomass (Kg-WCW) over time discounted by the lag phase in exp. E; d) total scFv (g) over time discounted by the lag phase in exp. E.

The final biomass concentration was lower in the PFEM1 experiment than in the BSM experiment (Figure 6.9a). However, given the much higher volume in the

PFEM1 experiment, the total biomass generated on a weight basis is 18.6 % higher in the PFEM1 than in the BSM experiment. This agrees with the differences in glycerol and ammonia feeding shown in Figure 6.8. The fact that PFEM1 generates lower biomass concentrations is beneficial for the production as it enabled to extend the duration of the cultivation by 24.6 hours (117.3 h and 92.7 h in the PFEM1 and BSM experiments respectively).

As for the product, the final scFv concentration in the PFEM1 experiment is 21.3 % higher than in the BSM experiment (268.7 mg/L and 220.7 mg/L respectively, based on the average concentration of the last four points shown in Figure 6.9b). If however calculated on a mass basis, given the much larger volume in the PFEM1 experiment, the final product mass produced increases 76.7 % in the PFEM1 experiment in relation to the BSM experiment (Figure 6.9d). The average productivity calculated on the basis of final total mass of scFv produced and of the cultivation duration discounted by the lag phase duration is 37.0 % higher in the PFEM1 than in the BSM experiment (84.6 mg/h and 61.8 mg/h respectively)

6.4 Conclusion

Due to the very high cell densities achieved in *P. pastoris* cultivations, oxygen transfer limitation occurs relatively early in the cultivation, typically between 2 and 3 days of cultivation. In the oxygen transfer limitation phase, the concentration of biomass can be increased with a DO-stat substrate feeding controller. DO-stat feeding controller maximizes glycerol feeding rate by regulating DO to a constant set-point slightly above the metabolic limitation boundary. The performance of such boundary control problems is linked to the dynamics of convergence and robustness of the underlying control algorithm, since systematic deviations to the set-point will result in cumulatively production losses. In this work, it has been developed an adaptive controller algorithm with improved convergence dynamics for this particular boundary control problem. When exposed to large process perturbations it reacts very fast to bring DO close to the boundary, thus eliminating very rapidly most of the deviation to the set-point. Then, in the vicinity of the set-point, it converges slower to bring the process on track again while minimizing the risk of violating the boundary. The accuracy of this controller can have a significant impact on the process productivity. For the case of cell growth dissociated kinetics studied in this work, the final product titer increases with the biomass concentration time integral, which can be significantly augmented by applying an accurate DO-stat glycerol feeding controller at very low DO set-points in the range of 3 – 5 %. Finally, the application of the DO-stat feeding controller together with the optimized CM formula PFEM1 resulted in an overall increase of 76.7 % of the final scFv mass produced and a 37.0 % increase of the average productivity (mass per unit time) in comparison to the DO-stat feeding controller with the BSM medium formulation.

Bibliography

Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D. 2008. Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. Biotechnol Bioeng 100 (1): 177-183.

Berdichevsky M, d'Anjou M, Mallem MR, Shaikh SS, Potgieter TI. 2011. Improved production of monoclonal antibodies through oxygen-limited cultivation of glycoengineered yeast. J Biotechnol 155 (2): 217-224.

Cereghino JL and Cregg JM, 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, FEMS Microbiology Reviews 24, 45-66.

Charoenrat T, Ketudat-Cairns M, Stendahl-Andersen H, Jahic M, Enfors SO. 2005. Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes. Bioprocess Biosyst Eng 27 (6): 399-406.

Chung JD. 2000. Design of metabolic feed controllers: Application to high-density fermentations of *Pichia pastoris*, Biotechnol Bioeng, 68: 298-307

Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto R, Moosmayer D, Donner P, Carrondo MJT. 2004. Methanol Induction Optimization for scFv Antibody Fragment Production in *Pichia pastoris*. Biotechnology and Bioengineering 86: 458-467.

Jahic M, Gustavsson M, Jansen AK, Martinelle M, Enfors SO. 2003. Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes. J Biotechnol 102 (1): 45-53.

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 (6): 385-393.

Lee CY, Lee SJ, Jung KH, Katoh S, Lee EK. 2003. High dissolved oxygen tension enhances heterologous protein expression by recombinant *Pichia pastoris*. Process Biochem 38 (8): 1147-1154.

Lim HK, Choi SJ, Kim KY, Jung KH. 2003. Dissolved-oxygen-stat controlling two variables for methanol induction of rGuamerin in *Pichia pastoris* and its application to repeated fed-batch. Appl Microbiol Biotechnol, 62: 342-348

Narendra KS, Annaswamy AM. 1989. Stable Adaptive Systems. Prentice-Hall, New Jersey

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. Journal of Biotechnology 116: 35-50.

Trentmann O, Khatri NK, Hoffmann F. 2004. Reduced oxygen supply increases process stability and product yield with recombinant *Pichia pastoris*. Biotechnol Prog 20 (6): 1766-1775.

CHAPTER 7

Conclusions

Yeasts have relatively simple nutritional needs and most species can grow quite well under aerobic conditions when supplied with glucose, ammonium salts, inorganic ions, and a few basic elements for the maximum growth such as Mg, K, Cl, S, and P. Different concentration of these components can affect growth, but more importantly, the optimal CM composition for cell growth is not necessarily the same for heterologous protein expression. Indeed, certain medium components may be toxic at certain levels (for instance methanol in the case of *P. pastoris*), and many may be involved in complex interactions in the same or competing pathways within the cell. Systematic methodologies for the optimization of CM composition based on biological mechanisms are still lacking. The efforts developed in this PhD thesis, were firstly centered in the development of a new framework for systematic optimization of CM composition. The second main objective in this thesis was a sound proof-of-concept by the application of this framework for the optimization of CM composition and cultivation conditions of a *P. pastoris* strain expressing scFv.

The first important outcome of this thesis was a computational algorithm called Projection to Latent Pathways (PLP). PLP identifies a minimal set of metabolic pathways (more specifically the elementary flux modes, EFM, of the metabolic network of the target cells) with the highest degree of correlation with measured envirome variables. PLP may be viewed as a constrained version of PLS, attuned to the structure of the biological system under study. PLP as PLS maximizes correlation and minimizes redundancy (i.e. eliminates pathways we low correlation with the environment). Mathematically, while in PLS the loadings and score are abstract variables, in PLP loadings and scores refer to well defined metabolic structures, namely EFM and EFM weighting factors. Specifically, PLP explores EFM as "principle components" of a metabolic network.

This new computational method PLP was afterwards applied to develop new approach for CM design termed "Cell Functional Enviromics (CFE)". CFE consists in the systematic characterization of the effect of environmental variables (i.e.

medium factors) on cellular function. The CFE method was applied to optimize the composition of *Pichia* Trace Metals supplements with the aim of increasing the scFv titer. To achieve this goal, shake flasks experiments were performed using a *P. pastoris* X33 strain and samples of each experiment were analyzed for cell density, product titer and other analytical methods necessary to run the PLP algorithm, namely HPLC, MS and NMR. A *P. pastoris* functional enviromics map (FEM) was constructed and, with this FEM, 10 PTM1 supplements were optimized. The experimental validation of the 10 optimized PTM1 supplements clearly demonstrate that optimized CM formulations can be obtained with much higher product specific productivity than that of the baseline CM formulation. The best PTM1 formula increased the scFv productivity by twofold. Following, the best PTM1 formulation was validated in 2 L bioreactor *P. pastoris* X33 experiments, resulting in 1.7 fold increase in specific scFv productivity in relation to the baseline formulation proposed by Invitrogen. These successful results can be attributed to the rational knowledge-based design approach followed by CFE. The main advantages of CFE in relation to DoE methods are i) to decrease the number of experiments and ii) to increase the incremental productivity gain per experiment performed.

The BSM developed by Invitrogen is the most commonly medium used for *P. pastoris*, however, it has been reported that at pH greater than 5.0 precipitation of some salts occurs. This can be prevented by reducing the concentration of salts in the medium. For this purpose, we have measured the five main elements (Mg, K, Ca, P and S) concentration over time by ICP-AES and then used the data to derive a simple mathematical model that describe the concentration profiles of biomass and of these elements over time. Nine experiments were performed under different operational conditions in a pilot 50 L bioreactor with *P. pastoris* GS115 expressing the scFv. The experimental results revealed excess of Ca and S over Mg, P and K, and the data also suggested the occurrence of precipitation at pH 6.5 and pH 7.0, which seems to be more severe for Mg and Ca. Based on the estimated maximum specific growth rate and element/biomass yields provided by the

model, an optimized formulation of the elements of PMS (opt. PMS) was calculated for a given desired final biomass concentration. The biomass standard deviation for the PMS is much higher than the new formulation proposed in this study (107.2 g/L and 17.1 g/L for PMS and opt. PMS respectively). The opt. PMS was incorporated in the cell CM previously developed, opt. PTM1, resulting in a new formulation called *Pichia* Functional Enviromics Medium 1 (PFEM1).

In a final stage we have studied the optimization of the bioreactor operation. Direct adaptive dissolved oxygen (DO)-stat feeding controller that maximizes glycerol feeding under the constraint of available oxygen transfer capacity was developed and applied to the bioreactor. Four fed-batch cultures of a *P. pastoris* X33 strain constitutively expressing the scFv were performed in a pilot 50 L bioreactor. After 2 or 3 days of cultivation, the cell densities achieved are very high, and as a consequence, oxygen transfer becomes limiting. The developed adaptive controller enabled to maximize glycerol feeding through the regulation of DO concentration between 3 and 5 % of saturation. Such control performance enabled operating closer to the 0 % boundary for longer periods of time when compared to a traditional PID algorithm, which tends to destabilize with increasing cell density. With this accuracy the use of this controller can have a significant impact on the productivity. Finally, it was applied the DO-stat feeding controller together with the optimized CM formula, PFEM1. Two experiments of *P. pastoris* X33 expressing the scFv in a 50 L bioreactor with different medium, one experiment with the PFEM1 and other with the BSM, were performed. Summing up, the results show an increase of 76.7 % of the final scFv mass produced and 37 % increase of the average productivity (mass per unit time) for the PFEM1 in comparison to the DO-stat feeding controller with the BSM.

7.1 Future work

The work developed in this thesis resulted in a new chemically defined culture media formulation (PFEM1) for the yeast *P. pastoris.* Experiments performed in shake flasks, 2 L bioreactor and 50 L bioreactor demonstrated a clear increase of productivity of a *P. pastoris* X33 strain expressing scFv. In order to strengthen the value-added by this new formulation it is necessary to test it in the future with other strains and other recombinant proteins. It would be important to repeat the pilot experiments for the *P. pastoris* GS115 strain expressing the same scFv and also to test other strains expressing different proteins. Regarding the latter, it would be important to assess small proteins and complex proteins such as full length antibodies. Only after collecting this data, a definite final judgment can be made regarding the relative advantages/disadvantages of PFEM1 and BSM.

In this thesis, a FEM was for the first time constructed for the yeast *P. pastoris* using the PLP algorithm. The experimental data available was that of a relatively small number of metabolites obtained by HPLC and other biochemical assays. The resolution of this map (higher number of medium ingredients and higher number of EFMs) can be increased in the future by applying high-throughput metabolomics, namely NMR and/or GC-MS.

Another important aspect that was not addressed in this thesis is the quality of the expressed recombinant protein. It would be important to include quality attributes of the target protein in the *P. pastoris* FEM and in the CFE design method. Examples of such quality attributes are protein folding and glycosylation. There is today a tremendous lack of knowledge of how the medium affects the quality of proteins. The CFE method could shed light into this very relevant topic for the industry.

In relation to the CFE method, additional assessment should be performed to clearly demonstrate the advantage of a rational metabolic function oriented design method. We have addressed in this thesis only the EFM that corresponding to the

optimization of the protein. However, CFE can be used in theory to optimize any EFM or even all EFM simultaneously. Examples of future assessment would be to i) decrease the specific growth rate through down regulation of biomass generating EFM, ii) to decrease oxygen consumption and heat generation through the down regulation of EFM with high oxygen yield and iii) to decrease product expression (instead of increasing as pursued in this thesis) through the down regulation of EFM that produce product.

Finally, an important challenge for the CFE methodology developed in this thesis would be to extend it to other industry relevant expression systems, such as Chinese Hamster Ovary cells (*CHO*), stem cells and *E. coli*. FEM can be constructed for these expression systems in a similar way of that performed for *P. pastoris*. Then, an important challenge would be to reverse engineer the complex media used to grow mammalian cell lines into synthetic chemically defined media by setting all the components, such as, carbohydrates, amino acids, vitamins, minerals, lipids, buffers and growth factors guided by the respective FEMs and using the rational CFE design approach.

Appendix

Appendix A

BCA Protein Assay kit from Pierce was used to quantify total protein concentration of a purified sample of scFv to be used as standard for the all analytical methods requiring internal scFv standard (positive controls and calibration curves).

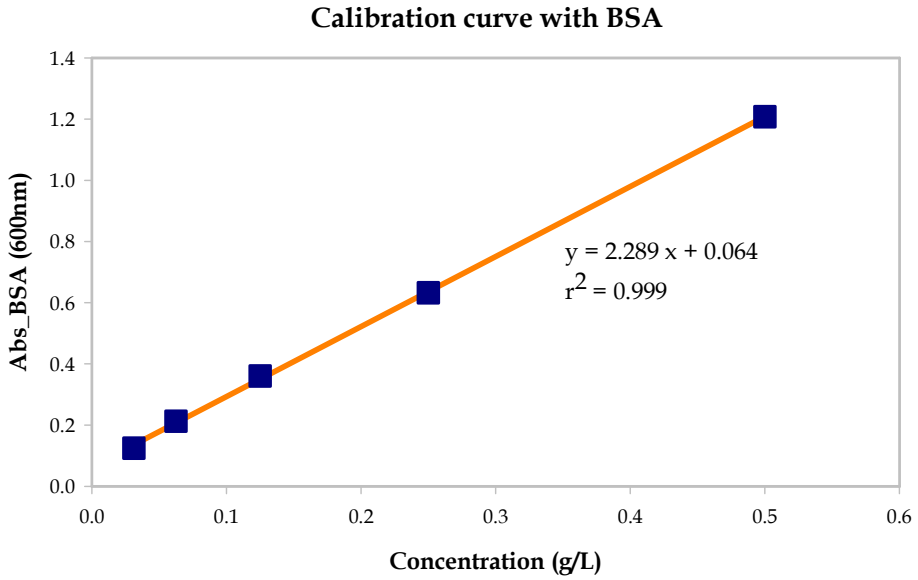


Figure A.1 Calibration curve for BCA Protein Assay Kit with BSA as standard protein.

Table A.1 Data for scFv standard by BCA Protein Assay Kit.

Dilution factor	Abs0	Abs1	Abs2	Abs average	Conc. scFv (g/L)
2	0.068	2.391	2.429	2.410	1.991
4	0.066	1.618	1.546	1.582	2.535
8	0.066	0.964	0.965	0.965	**2.913**
16	0.067	0.584	0.592	0.588	**3.194**
32	0.067	0.358	0.355	0.357	**3.151**
64	0.068	0.238	0.242	0.240	**3.045**
128	0.065	0.153	0.147	0.150	1.057

Abs0 average = 0.067; Abs_average=average(Abs1+Abs2);

Conc. scFv = ((Abs_average – Abs0 average – 0.064) / 2.289) X Dilution;

Average concentration of standard scFv = 3.08 g/L.

ELISA technique

ELISA method was used for scFv quantification in samples from the experiments performed. An example of ELISA titration results from a process sample is presented below.

Table A.2 Data from calibration curve for the ELISA plate.

Dilution factor	Conc. std scFv (ng/mL)	Abs PBS	Abs1 std scFv	Abs2 std scFv	Abs3 std scFv	Abs std scFv average
2000	1540.000	0.296	1.581	1.624	1.588	1.598
4000	**770.000**	**0.351**	**1.224**	**1.172**	**1.187**	**1.194**
8000	**385.000**	**0.35**	**0.869**	**0.817**	**0.801**	**0.829**
16000	**192.500**	**0.385**	**0.644**	**0.619**	**0.627**	**0.630**
32000	**96.250**	**0.357**	**0.493**	**0.484**	**0.472**	**0.483**
64000	**48.125**	**0.364**	**0.42**	**0.423**	**0.397**	**0.413**
128000	24.063	0.36	0.365	0.386	0.373	0.375
256000	12.031	0.322	0.358	0.338	0.338	0.345

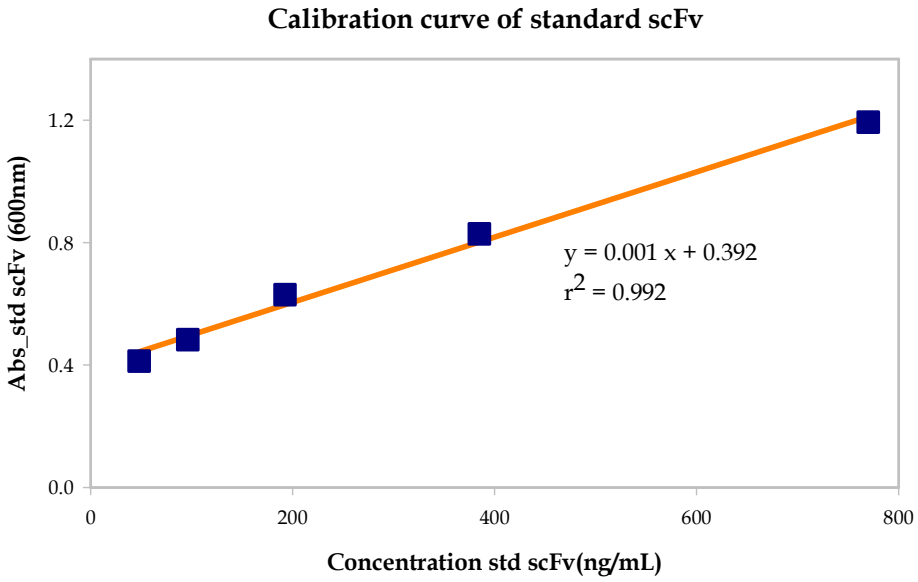


Figure A.2 Calibration linear curve of standard scFv.

Table A.3 scFv concentration determined for the sample.

Abs1	Abs2	Abs average	Product (mg/L)
1.533	1.551	1.542	26.972
1.256	1.246	1.251	40.295
0.906	0.933	0.920	**49.493**
0.654	0.643	0.649	**48.142**

scFv concentration of sample = 48.82 ± 0.96 mg/L.

Appendix B

Table B.1 List of metabolic reactions of a recombinant *P. pastoris* X33 and GS115 strain expressing a protein of empirical formula $CH_{1.965}N_{0.271}O_{0.535}S_{0.006}$.

	Uptake reactions
R1	1 ATP + 1 GlyOH → 1 ADP + 1 GAP + 1 $NADH_2$
R2	1 MeOH + 0.5 O_2 → 1 ForA + 1 H_2O
R3	1 ForA + 1 H_2O → 1 For + 1 $NADH_2$
R4	1 X5P + 1 ForA + 1 ATP → 1 ADP + 2 GAP
R5	1 H_3PO_4 + 2 ATP + 2 H_2O → 2 ADP + 3 Pi
R6	1 H_2SO_4 + 2 ATP + 4 $NADH_2$ → 2 ADP + 1 PPi + 3 H_2O + 1 H_2S
	Glycolysis/Glyconeogenesis
R7	1 G6P ↔ 1 F6P
R8	1 F6P + 1 ATP → 2 GAP + 1 ADP
R9	2 GAP + 1 H_2O → 1 F6P + 1 Pi
R10	1 GAP + 1 ADP + 1 Pi ↔ 1 PEP + 1 ATP + 1 H_2O + 1 $NADH_2$
R11	1 PEP + 1 ADP → 1 Pyr + 1 ATP
	Pentose Phosphate Pathway
R12	1 G6P + 1 H_2O → 1 R5P + 2 $NADH_2$ + 1 CO_2
R13	1 R5P ↔ 1 X5P
R14	1 R5P + 1 X5P ↔ 1 E4P + 1 F6P
R15	1 E4P + 1 X5P ↔ 1 GAP + 1 F6P
R16	1 ATP + 1 R5P → 1 AMP_en + 1 PRPP
	Anaplerotic reactions
R17	1 Pyr + 1 H_2O + 1 CO_2 + 1 ATP → 1 OA + 1 ADP + 1 Pi
R18	1 Mal → 1 Pyr + 1 CO_2 + 1 $NADH_2$
R19	1 OA + 1 ATP ↔ 1 PEP + 1 ADP + 1 CO_2
	Tricarboxylic acid (TCA) cycle
R20	1 OA + 1 AcCoA + 1 H_2O → 1 Cit
R21	1 Cit → 1 ICit
R22	1 ICit ↔ 1 aKG + 1 $NADH_2$ + 1 CO_2
R23	1 aKG + 1 ADP + 1 Pi → 1 Succ + 1 ATP + 1 $NADH_2$ + 1 CO_2
R24	1 Succ → 1 Fum + 1 $NADH_2$
R25	1 Fum + 1 H_2O ↔ 1 Mal
R26	1 Mal ↔ 1 OA + 1 $NADH_2$
	Glyoxylate Shunt
R27	1 ICit → 1 Glx + 1 Succ
R28	1 AcCoA + 1 Glx + 1 H_2O → 1 Mal

	C3/C4 Metabolism pathway
R29	1 Pyr → 1 AcCoA + 1 CO_2 + 1 $NADH_2$
	Proteins Biosynthesis
R30	1 Asp + 1 Ser + 1 ATP + 1 H_2O + 1 H_2S → 1 Cys + 1 ADP + 1 Pi + 1 Pyr + 1 NH_3 + 1 CO_2 + 1 $NADH_2$
R31	1 Glu + 1 Gln + 1 Asp + 4 ATP + 1 CO_2 + 2 H_2O + 1 $NADH_2$ → 1 Arg + 1 aKG + 3 ADP + 1 AMP_en + 3 Pi + 1 PPi + 1 Fum
R32	2 Glu + 1 AcCoA → 1 Lys + 1 aKG + 1 CO_2
R33	2 Pyr + 1 AcCoA + 1 Glu → 1 Leu + 2 CO_2 + 1 aKG
R34	1 PRPP + 1 ATP + 3 H_2O + 1 Gln → 1 His + 1 AICAR + 2 $NADH_2$ + 2 PPi + 1 Pi + 1 aKG
R35	1 Glu + 1 ATP + 1 NH_3 → 1 Gln + 1 ADP + 1 Pi
R36	1 aKG + 1 $NADH_2$ + 1 NH_3 → 1 Glu + 1 H_2O
R37	1 Glu + 1 OA → 1 Asp + 1 aKG
R38	1 Asp + 1 Gln + 1 ATP + 1 H_2O → 1 Asn + 1 Glu + 1 AMP_en + 1 PPi
R39	1 Glu + 1 Pyr → 1 Ala + 1 aKG
R40	1 E4P + 2 PEP + 1 $NADH_2$ + 1 ATP → 1 Chor + 1 ADP + 4 Pi
R41	1 Chor + 1 PRPP + 1 Gln + 1 Ser → 1 Trp + 1 GAP + 1 Glu + 1 CO_2 + 1 H_2O + 1 Pyr + 2 Pi
R42	1 Chor + 1 Glu → 1 Tyr + 1 $NADH_2$ + 1 aKG + 1 CO_2
R43	1 Chor + 1 Ala → 1 Tyr + 1 $NADH_2$ + 1 Pyr + 1 CO_2
R44	1 Chor + 1 Glu → 1 Phe + 1 aKG + 1 H_2O + 1 CO_2
R45	1 Chor + 1 Ala → 1 Phe + 1 Pyr + 1 H_2O + 1 CO_2
R46	1 Glu + 2 Pyr + 1 $NADH_2$ → 1 Val + 1 H_2O + 1 CO_2 + 1 aKG
R47	1 Glu + 1 GAP + 2 H_2O → 1 Ser + 1 aKG + 1 Pi + 2 $NADH_2$
R48	1 Gly + 1 MnTHF + 1 H_2O → 1 Ser + 1 THF
R49	1 Asp + 2 ATP + 2 $NADH_2$ + 1 H_2O → 1 Thr + 2 ADP + 2 Pi
R50	1 Asp + 1 ATP + 2 $NADH_2$ + 1 H_2S + 1 MnTHF + 3 $NADH_2$ → 1 Met + 1 ADP + 1 Pi + 1 THF + 1 H_2O
R51	1 Ser + 1 THF → 1 Gly + 1 MnTHF + 1 H_2O
R52	1 Glx + 1 Ala → 1 Gly + 1 Pyr
R53	1 Glu + 1 ATP + 2 $NADH_2$ → 1 Pro + 1 ADP + 1 Pi + 1 H_2O
R54	1 Glu + 1 Pyr + 1 Thr + 1 $NADH_2$ → 1 aKG + 1 ILeu + 1 H_2O + 1 CO_2 + 1 NH_3
R55	0.1245 Ala + 0.0437 Arg + 0.0277 Asn + 0.0808 Asp + 0.0019 Cys + 0.0285 Gln + 0.0819 Glu + 0.0787 Gly + 0.0179 His + 0.0524 ILeu + 0.0803 Leu + 0.0776 Lys + 0.0138 Met + 0.0364 Phe + 0.0448 Pro + 0.0502 Ser + 0.0518 Thr + 0.0076 Trp + 0.0277 Tyr + 0.0719 Val + 4.3 ATP + 4.3 H_2O => 1 ProteinsBiomass + 4.3 Pi + 4.3 ADP

	Lipids Biosynthesis
R56	9 AcCoA + 8 ATP + 1 GAP + 17 $NADH_2$ + 1 H_2O → 8 ADP + 9 Pi + 1 MAG
R57	18 AcCoA + 16 ATP + 1 GAP + 34 $NADH_2$ + 1 H_2O → 16 ADP + 17 Pi + 1 DAG
R58	27 AcCoA + 24 ATP + 1 GAP + 51 $NADH_2$ + 1 H_2O → 24 ADP + 25 Pi + 1 TAG
R59	18 AcCoA + 18 ATP + 23 $NADH_2$ + 10 O_2 → 18 ADP + 6 PPi + 6 Pi + 1 Zym + 7 H_2O + 9 CO_2
R60	1 Met + 1 Zym + 2 O_2 + 1 $NADH_2$ → 1 Asp + 1 Erg + 2 H_2O + 1 H_2S
R61	1 Ser + 18 AcCoA + 16 ATP + 1 GAP + 1 CTP + 37 $NADH_2$ => 16 ADP + 1 PPi + 16 Pi + 1 CMP_en + 1 PDS
R62	1 PDS → 1 PDE + 1 CO_2 + 3 $NADH_2$
R63	1 PDE + 3 Met + 6 H_2O → 3 Asp + 3 H_2S + 1 PDC + 6 $NADH_2$
R64	1 G6P + 18 AcCoA + 16 ATP + 1 GAP + 1 CTP + 34 $NADH_2$ + 1 H_2O → 16 ADP + 1 PPi + 17 Pi + 1 CMP_en + 1 PDMI
R65	0.0548 MAG + 0.0548 DAG + 0.0548 TAG + 0.1233 PDC + 0.0411 PDS + 0.0959 PDE + 0.1507 Erg + 0.3151 Zym + 0.1096 PDMI → 1 Lip
	Carbohydrates Biosynthesis
R66	1 F6P + 1 H_2O => 1 Pi + 1 Man
R67	1 G6P + 1 UTP + 1 H_2O => 1 Glc + 1 PPi + 1 UDP
R68	1 Glc => 1 Gal
R69	1 Gln + 1 AcCoA + 1 F6P + 1 UTP => 1 Glu + 1 PPi + 1 GlcNAc + 1 UDP
R70	1 PEP + 1 CTP + 1 GlcNAc + 3 H_2O + 1 ATP => 1 PPi + 2 Pi + 1 NeuAc + 1 CMP_en + 1 ADP
R71	0.1554 NeuAc + 0.1554 GlcNAc + 0.3250 Man + 0.2088 Glc + 0.1554 Gal → 1 Sug
	Nucleotides Biosynthesis
R72	1 Asp + 2 Gln + 1 Gly + 5 ATP + 1 PRPP + 1 FTHF + 1 CO_2 + 3 H_2O → 2 Glu + 1 THF + 5 ADP + 5 Pi + 1 PPi + 1 Fum + 1 AICAR
R73	1 AICAR + 1 FTHF → 1 THF + 1 IMP + 1 H_2O
R74	1 IMP + 1 Asp + 1 GTP → 1 AMP + 1 Fum + 1 Pi + 1 GDP
R75	1 IMP + 1 Gln + 1 ATP + 2 H_2O → 1 GMP + 1 Glu + 1 PPi + AMP_en + 1 $NADH_2$
R76	1 IMP + 1 NH_3 + 1 ATP + 1 H_2O → 1 GMP + 1 PPi + AMP_en
R77	1 PRPP + 1 Asp + 1 Gln + 2 ATP + 1 H_2O → 1 UMP + 1 Glu + 1 $NADH_2$ + 1 PPi + 2 Pi + 2 ADP
R78	1 NH_3 + 1 UMP → 1 CMP + 1 H_2O
R79	1 CMP + 1 $NADH_2$ → 1 dCMP + 1 H_2O
R80	1 GMP + 1 $NADH_2$ → 1 dGMP + 1 H_2O
R81	1 AMP + 1 $NADH_2$ → 1 dAMP + 1 H_2O

R82	1 UMP + 1 MnTHF + 2 $NADH_2$ → 1 dTMP + 1 THF + 1 H_2O
R83	0.2329 AMP + 0.3059 UMP + 0.2329 GMP + 0.2283 CMP + 0.4 ATP + 0.4 H_2O → 1 RNA + 0.4 Pi + 0.4 ADP
R84	0.2000 dCMP + 0.3000 dTMP + 0.3000 dAMP + 0.2000 dGMP + 0.4 ATP + 0.4 H_2O → 1 DNA + 0.4 Pi + 0.4 ADP
R85	0.9471 RNA + 0.0529 DNA → 1 Nuc
	Biomass Synthesis (C_mmol/Cmmol DCW)
R86	0.5716 ProteinsBiomass + 0.0359 Nuc + 0.3855 Sug + 0.0073 Lip → 6.0872 X
	Product Biosynthesis
R87	0.0658 Ala + 0.0494 Arg + 0.0123 Asn + 0.0329 Asp + 0.0206 Cys + 0.0535 Gln + 0.0370 Glu + 0.1358 Gly + 0.0000 His + 0.0329 ILeu + 0.0782 Leu + 0.0288 Lys + 0.0082 Met + 0.0453 Phe + 0.0494 Pro + 0.1564 Ser + 0.0782 Thr + 0.0165 Trp + 0.0494 Tyr + 0.0494 Val + 4.3 ATP + 4.3 H_2O → 4.6502 Product + 4.3 Pi + 4.3 ADP
	Oxidative Phosphorylation (P/O ratio = 2)
R88	2 ADP + 2 Pi + 1 $NADH_2$ + 0.5 O_2 → 2 ATP + 3 H_2O
	Biosynthesis and Interconversion of one-carbon units
R89	1 MnTHF + 1 H_2O → 1 FTHF + 1 $NADH_2$
R90	1 FTHF + 1 AICAR + 1 OA + 1 NH_3 + 1 ATP + 1 GTP + 1 $NADH_2$ → 1 THF + 2 Pi + 1 GDP + 1 Fum + 1 AMP_en + 1 ADP + 1 H_2O
R91	1 MnTHF + 2 H_2O → 1 THF + 1 $NADH_2$ + 1 For
R92	1 THF + 1 Gly → 1 MnTHF + 1 CO_2 + 1 NH_3 + 1 $NADH_2$
R93	1 For → 1 $NADH_2$ + 1 CO_2
	Energy Interconversion
R94	1 PPi + 1 H_2O → 2 Pi
R95	1 ATP + 1 AMP_en → 2 ADP
R96	1 ATP + 1 CMP_en → 1 CDP + 1 ADP
R97	1 ATP + 1 CDP → 1 CTP + 1 ADP
R98	1 ATP + 1 GMP_en → 1 GDP + 1 ADP
R99	1 ATP + 1 GDP → 1 GTP + 1 ADP
R100	1 ATP + 1 UMP_en → 1 UDP + 1 ADP
R101	1 ATP + 1 UDP → 1 UTP + 1 ADP
R102	1 ATP + 1 H_2O → 1 ADP + 1 Pi

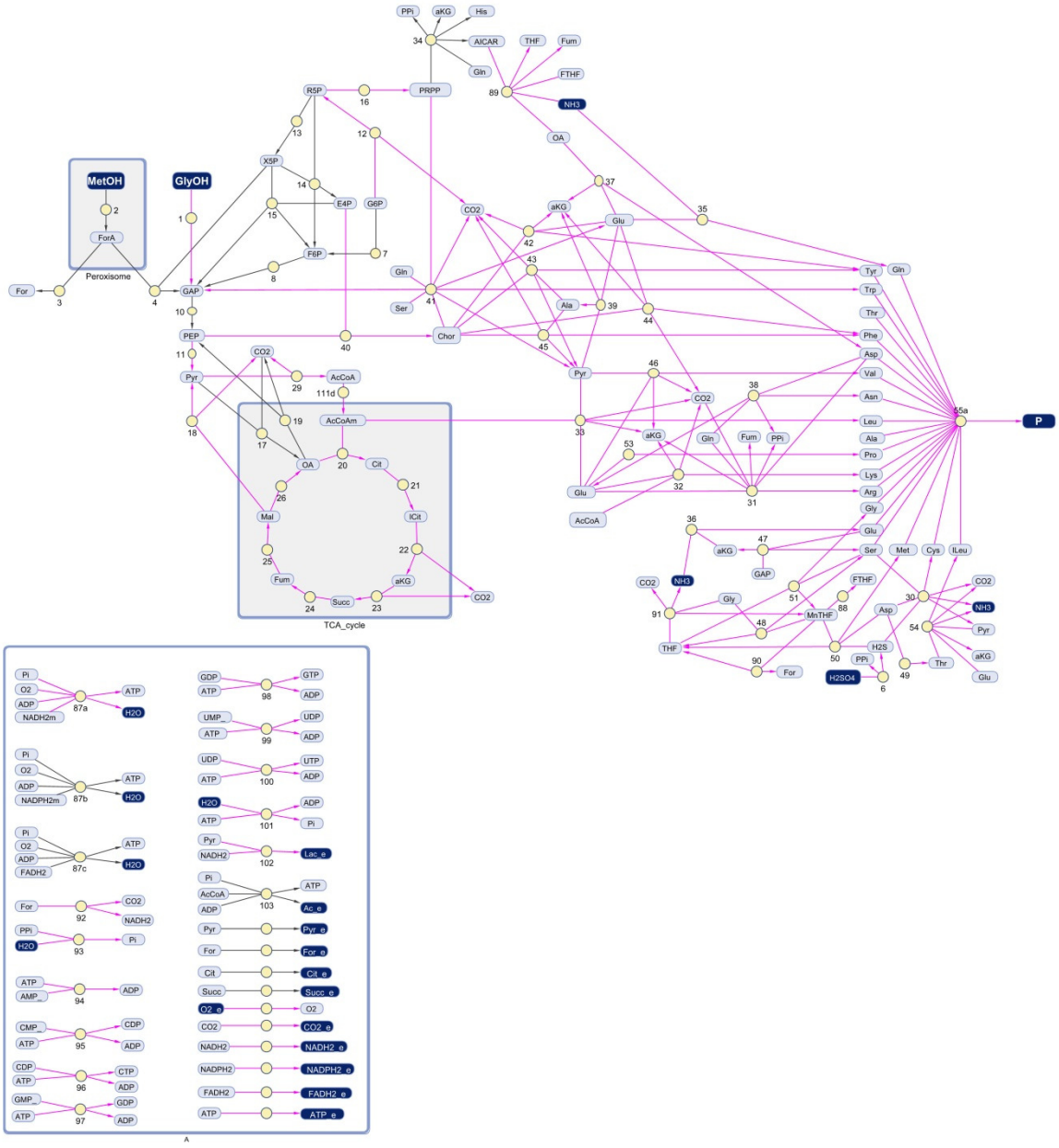


Figure B.1 Metabolic network of a *P. pastoris* X33 strain expressing a scFv.

Appendix C

Table C.1 Concentration of PTM1 components used in each experiment.

	$CuSO_4$ -$5H_2O$ (g/L)	NaI (g/L)	$MnSO_4$-H_2O (g/L)	Na_2MoO_4-$2H_2O$ (g/L)	H_3BO_3 (g/L)	$CoCl_2$ (g/L)	$ZnCl_2$ (g/L)	$FeSO_4$-$7H_2O$ (g/L)	Biotin (g/L)	H_2SO_4 (g/L)
Exp. 1	6.00	0.08	0.30	0.02	0.02	0.50	20.00	6.50	0.02	0.50
Exp. 2	6.00	0.008	0.30	0.20	0.002	0.50	2.00	65.00	0.02	0.50
Exp. 3	6.00	0.008	3.00	0.20	0.02	0.50	20.00	65.00	0.20	0.50
Exp. 4	6.00	0.08	0.30	0.02	0.02	5.00	2.00	65.00	0.20	5.00
Exp. 5	6.00	0.08	3.00	0.20	0.02	0.50	2.00	6.50	0.20	5.00
Exp. 6	0.60	0.08	0.30	0.20	0.02	5.00	20.00	6.50	0.20	0.50
Exp. 7	0.60	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 8	6.00	0.08	3.00	0.20	0.002	5.00	20.00	65.00	0.02	0.50
Exp. 9	6.00	0.008	0.30	0.20	0.002	0.50	20.00	65.00	0.20	5.00
Exp. 10	0.60	0.08	3.00	0.02	0.002	0.50	2.00	6.50	0.20	0.50
Exp. 11	0.60	0.08	0.30	0.20	0.002	5.00	2.00	65.00	0.02	0.50
Exp. 12	0.60	0.008	0.30	0.02	0.02	0.50	20.00	65.00	0.20	0.50
Exp. 13	0.60	0.008	3.00	0.20	0.02	5.00	2.00	6.50	0.02	0.50
Exp. 14	6.00	0.08	0.30	0.20	0.002	5.00	20.00	6.50	0.20	0.50
Exp. 15	6.00	0.008	0.30	0.20	0.02	5.00	2.00	6.50	0.02	5.00
Exp. 16	6.00	0.08	3.00	0.02	0.02	0.50	2.00	65.00	0.02	0.50
Exp. 17	6.00	0.008	3.00	0.02	0.002	5.00	2.00	6.50	0.20	0.50
Exp. 18	0.60	0.08	3.00	0.02	0.002	5.00	20.00	65.00	0.02	5.00
Exp. 19	0.60	0.008	0.30	0.02	0.02	5.00	20.00	6.50	0.02	0.50
Exp. 20	0.60	0.008	3.00	0.20	0.002	0.50	20.00	6.50	0.02	5.00
Exp. 21	0.60	0.08	0.30	0.20	0.02	0.50	2.00	65.00	0.02	5.00
Exp. 22	0.60	0.008	0.30	0.02	0.002	5.00	2.00	65.00	0.20	5.00
Exp. 23	6.00	0.08	0.30	0.02	0.002	0.50	20.00	6.50	0.02	5.00
Exp. 24	6.00	0.008	3.00	0.02	0.02	5.00	20.00	65.00	0.02	5.00
Exp. 25*	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 26	0.60	0.008	0.30	0.02	0.002	0.50	2.00	6.50	0.02	0.50
Exp. 27	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 28	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 29	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 30	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00
Exp. 31	6.00	0.08	3.00	0.20	0.02	5.00	20.00	65.00	0.20	5.00

Table C.2 Concentration of BSM components used in each experiment.

	H_3PO_4 85% (mL/L)	$CaSO_4$-$2H_2O$ (g/L)	K_2SO_4 (g/L)	$MgSO_4$ -$7H_2O$ (g/L)	KOH (g/L)	Glycerol (g/L)
Exp. 1	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 2	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 3	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 4	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 5	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 6	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 7	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 8	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 9	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 10	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 11	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 12	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 13	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 14	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 15	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 16	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 17	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 18	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 19	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 20	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 21	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 22	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 23	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 24	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 25*	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 26	13.35	0.47	9.10	7.45	2.07	20.00
Exp. 27	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 28	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 29	26.70	0.93	18.20	14.90	4.13	40.00
Exp. 31	26.70	0.93	18.20	14.90	4.13	40.00

The elementary flux modes were found to be the most important of the new method used shown in Table 4.5, Chapter 4, are specified below:

```
e10 = [ 7.22  0.00  0.00  0.00  0.13  0.03  -0.46  6.19  7.59  3.95  3.49
        0.00  -0.15 -0.01 -0.14 0.16  1.19  0.20   0.00  0.51  0.51  0.51
        0.00  0.00  0.20  0.00  0.00  0.00  1.62   0.00  0.08  0.14  0.15
        0.03  0.82  2.46  0.65  0.05  0.30  0.13   0.01  0.05  0.00  0.00
        0.07  0.13  0.32  0.00  0.19  0.04  0.20   0.00  0.08  0.10  1.85
        0.00  0.00  0.00  0.01  0.00  0.01  0.01   0.00  0.00  0.02  0.41
        0.45  0.19  0.39  0.19  1.25  0.05  0.05   0.03  0.03  0.00  0.06
        0.03  0.00  0.00  0.00  0.00  0.11  0.01   0.12  3.24  0.00  10.11
        0.14  0.03  0.02  0.00  0.02  1.48  0.36   0.20  0.20  0.00  0.06
        0.00  0.84  0.00];

e12 = [ 9.00  0.00  0.00  0.00  0.00  0.15  0.00   1.94  2.18  6.50  5.36
        0.00  -0.25 0.16  -0.41 0.08  2.07  0.00   0.00  1.12  1.12  1.12
        0.00  0.00  0.25  0.25  0.00  0.00  1.67   0.11  0.25  0.15  0.40
        0.00  0.67  6.05  1.20  0.06  0.34  0.57   0.08  0.25  0.00  0.23
        0.00  0.25  1.69  0.00  0.57  0.04  0.70   0.00  0.25  0.17  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  5.12  13.91
        0.00  0.00  0.65  0.00  0.65  0.46  0.40   0.00  0.00  0.00  0.00
        0.00  0.00  0.00];

e5 = [  1.08  0.00  0.00  0.00  0.00  0.00  -3.23  0.00  1.08  0.00  14.00
        3.23  2.15  1.08  1.08  0.00  14.00 0.00   14.00 0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  7.54
        0.00  0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00  0.00
        0.00  0.00  0.00];
```

```
e9 = [ 47.58  0.00  0.00  0.00  0.82  0.19  -4.48  0.00  9.66  25.76  22.76
       1.50   0.00  0.43  -0.43 1.07  7.78  1.28   0.00  3.34  3.34   3.34
       0.00   0.00  1.28  0.00  0.00  0.00  10.56  0.02  0.53  0.94   0.97
       0.22   5.16  16.07 4.23  0.33  2.28  0.87   0.09  0.00  0.33   0.00
       0.44   0.87  2.07  0.00  1.26  0.25  1.31   0.00  0.54  0.63   12.08
       0.01   0.01  0.01  0.07  0.02  0.04  0.03   0.02  0.02  0.15   2.65
       2.97   1.27  2.53  1.27  8.15  0.35  0.35   0.18  0.00  0.18   0.40
       0.17   0.01  0.01  0.01  0.01  0.72  0.04   0.76  21.14 0.00   69.30
       0.93   0.22  0.12  0.00  0.12  9.67  2.32   1.32  1.32  0.00   0.40
       0.00   5.50  46.69];

e3 = [ 37.03  0.00  0.00  0.00  0.00  0.60  -1.50  0.00  1.50  26.37  21.74
       1.50   0.00  1.16  -1.16 0.34  9.42  1.03   0.00  4.54  4.54   4.54
       0.00   0.00  1.03  0.00  0.00  0.00  6.76   0.43  1.03  0.60   1.63
       0.00   2.74  24.57 4.88  0.26  3.34  2.31   0.34  0.00  1.03   0.00
       0.94   1.03  6.85  0.00  2.31  0.17  2.83   0.00  1.03  0.68   0.00
       0.00   0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00   0.00
       0.00   0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  0.00   0.00
       0.00   0.00  0.00  0.00  0.00  0.00  0.00   0.00  0.00  20.80  59.99
       0.00   0.00  2.65  0.00  2.65  1.88  1.63   0.00  0.00  0.00   0.00
       0.00   0.00  13.34];
```

Experiment C.A

A 2 L bioreactor with 0.75 L of BSM and PTM1 medium was inoculated with 100 mL of inoculum. pH was controlled at 5.0 using a 25 % NH_4OH solution, temperature was kept constant at 30 °C. The bioreactor was operated in glycerol batch phase during approximately 100 hours. Then a second glycerol batch phase was initiated by adding 100 mL of a solution containing glycerol 99 % w/v and 12 mL/L of baseline formulation of PTM1, being the culture extended for additional 40 hours. The operational conditions are presented in Table C.3.

Table C.3 Bioreaction parameters, experiment C.A.

Variable	First GB phase	Second GB phase
Temperature (°C)	30.0	30.0
pH	5.0	5.0
Final DO (%)	93.7	100.0
Time phase ended (h)	97.6	144.0
Final WCW (g/L)	86.0	147.0
Final Product (scFv – mg/L)	8.9	21.3

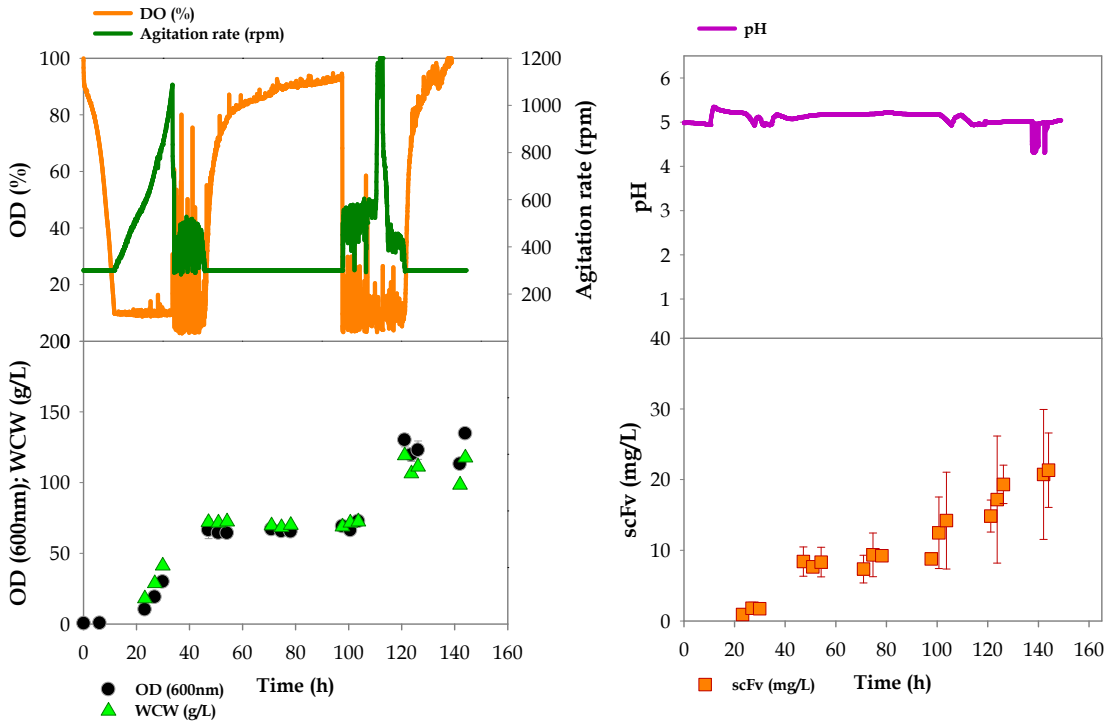


Figure C.1 Experiment C.A online and offline data.

Experiment C.B

A 2 L bioreactor with 0.75 L of BSM and optimized PTM1 medium was inoculated with 100 mL of inoculum. pH was controlled at 5.0 using a 25 % NH_4OH solution, temperature was kept constant at 30 ºC. The bioreactor was operated in glycerol batch phase during approximately 90 hours. Then a second glycerol batch phase was initiated by adding 100 mL of a solution containing glycerol 99 % w/v and 12 mL/L of optimized PTM1, being the culture extended for additional 50 hours. The operational conditions are presented in Table C.4.

Table C.4 Bioreaction parameters, experiment C.B.

Variable	First GB phase	Second GB phase
Temperature (ºC)	30.0	30.0
pH	5.0	5.0
Final DO (%)	91.8	84.9
Time phase ended (h)	88.7	160.0
Final WCW (g/L)	91.0	173.5
Final Product (scFv – mg/L)	12.5	33.7

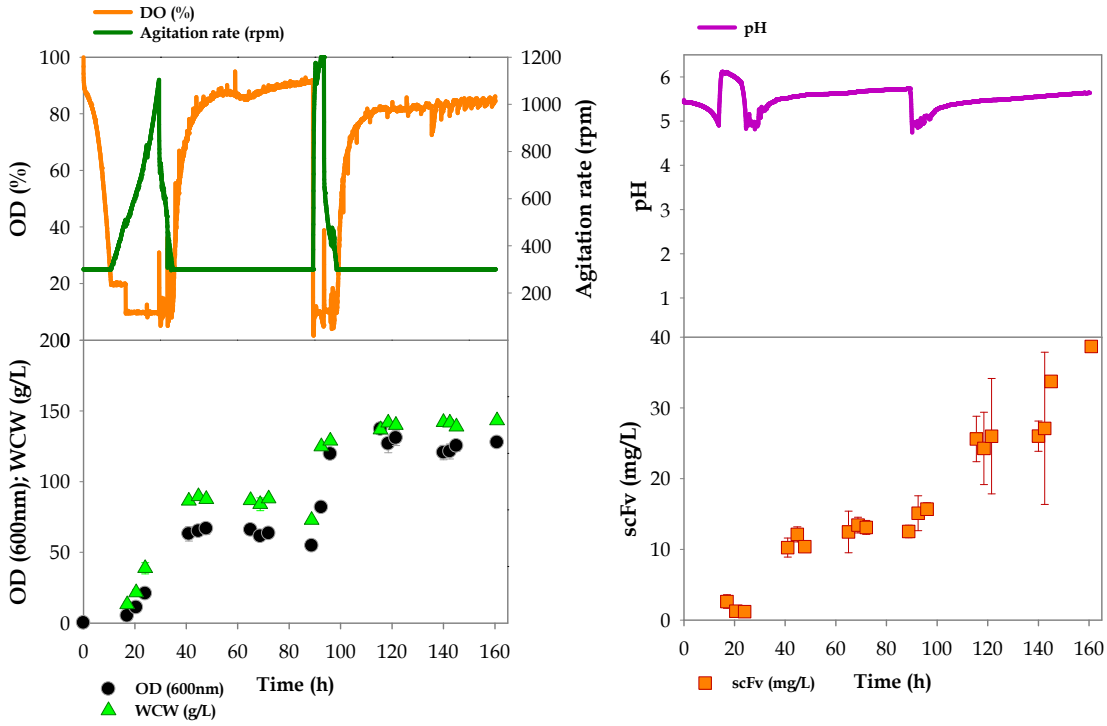


Figure C.2 Experiment C.B online and offline data.

Appendix D

Experiment D.A

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C and pH was controlled at 5.0 with 1.6 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.1 followed by methanol addition at constant feeding rate after the methanol adaptation phase (Table D.1). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.2.

Table D.1 Glycerol and methanol addition profiles, experiment D.A.

Glycerol fed batch phase	Methanol fed batch phase
$F = F_0 \times e^{(\mu \times t)}$;	$t_{MFB} > 40$ h → $F_{MeOH1} = 130.0$ g/h.
$F_0 = 65.0$ g/h; $\mu = 0.16$ h^{-1}; t=9.0 h;	

Table D.2 Bioreaction parameters, experiment D.A.

Variable	Glycerol phase	Methanol phase
Temperature (^{o}C)	30.0	30.0
pH	5.0	5.0
Final DO (%)	50.0	50.0
Glycerol added (Kg)	1.3	----
Methanol added (kg)	----	7.1
Base added (Kg)	0.8	0.8
Time after induction (h)	----	63.4
Total time (h)	----	101.8
Final WCW (g/L)	----	457.3
Final Product (scFv – mg/L)	----	5.9

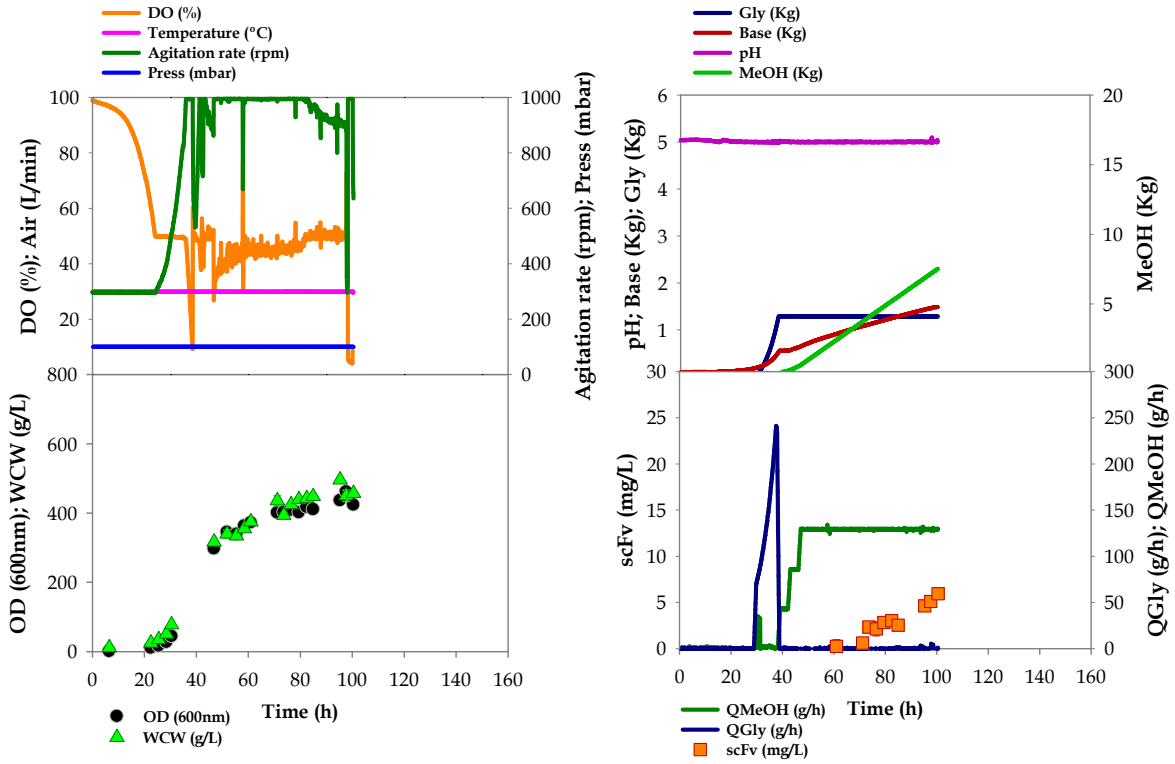


Figure D.1 Experiment D.A online and offline data.

Experiment D.B

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ºC in the first two phases and after, in the methanol fed-batch phase decrease at 20 ºC and the end of this phase increased again for 30 ºC. pH was controlled at 5.0 with 2.1 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.3 followed by methanol addition at constant feeding rate after the methanol adaptation phase (Table D.3). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.4.

Table D.3 Glycerol and methanol addition profiles, experiment D.B.

Glycerol fed batch phase	Methanol fed batch phase
$F= F_0 \times e^{(\mu \times t)}$;	$t_{MFB} \leq 40$ h → $F_{MeOH1}= 85.0$ g/h;
$F_0= 65.0$ g/h; $\mu= 0.16$ h^{-1}; t=16.0 h;	$t_{MFB} > 40$ h → $F_{MeOH2}= 215.0$ g/h.

Table D.4 Bioreaction parameters, experiment D.B.

Variable	Glycerol phase	Methanol phase
Final Temperature (ºC)	30.0	20.0 - 30.0
pH	5.0	5.0
Final DO (%)	50.0	50.0
Glycerol added (Kg)	2.8	----
Methanol added (kg)	----	9.6
Base added (Kg)	0.7	1.4
Time after induction (h)	----	71.2
Total time (h)	----	127.2
Final WCW (g/L)	----	585.0
Final Product (scFv – mg/L)	----	15.6

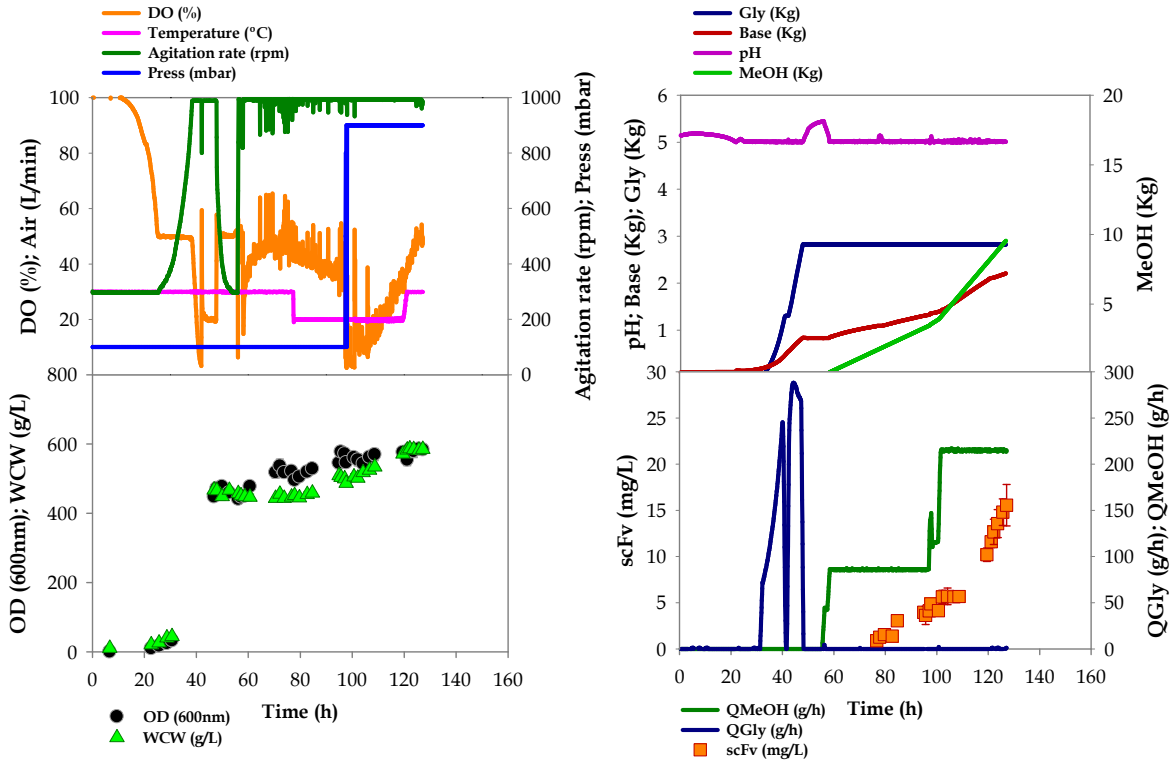


Figure D.2 Experiment D.B online and offline data.

Table D.5 Experiment D.B, SDS-PAGE and Western Blot results of selected samples.

PSM	SDS-PAGE										Western Blot									
Ap. MW	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
181.8																				
115.5																				
82.2																				
64.2																				
48.8																				
37.1																				
25.9																				
19.4																				
14.8																				
6.0																				

Legend: 1) PSM; 2) Sample time= 82.7 h; 3) Sample time= 84.7 h; 4) Sample time= 100.7 h; 5) Sample time= 104.2 h; 6) Sample time= 106.7 h; 7) Sample time= 119.5 h; 8) Sample time= 121.1 h; 9) Sample time= 123.7 h; 10) Sample time= 127.2h.

Experiment D.C

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in the glycerol batch and fed-batch phases, after decrease at 23 ^{o}C until the end. The pH was controlled first at 5, after 70 h of induction increase pH= 7 controlled with 3.0 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.6 followed by methanol addition at constant feeding rate after the methanol adaptation phase (Table D.6). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.7.

Table D.6 Glycerol and methanol addition profiles, experiment D.C.

Glycerol fed batch phase	Methanol fed batch phase
$F= F_0 \times e^{(\mu \times t)}$;	$t_{MFB} \leq 2$ h → F_{MeOH1}= 43.0 g/h;
F_0= 65.0 g/h;	$2 < t_{MFB} \leq 4$ h → F_{MeOH2}= 86.0 g/h;
μ= 0.16 h^{-1}; t= 9.0 h;	$4 < t_{MFB} \leq 62$ h → F_{MeOH3}= 130.0 g/h;
	$t_{MFB} > 62$ h → F_{MeOH4}= 230.0 g/h;

Table D.7 Bioreaction parameters, experiment D.C.

Variable	Glycerol phase	Methanol phase
Final Temperature (^{o}C)	30.0	23.0
pH	5.0	7.0
Final DO (%)	34.5	2.0
Glycerol added (Kg)	1.3	----
Methanol added (kg)	----	18.8
Base added (Kg)	0.4	2.6
Time after induction (h)	----	110.0
Total time (h)	----	145.3
Final WCW (g/L)	----	587.7
Final Product (scFv – mg/L)	----	16.1

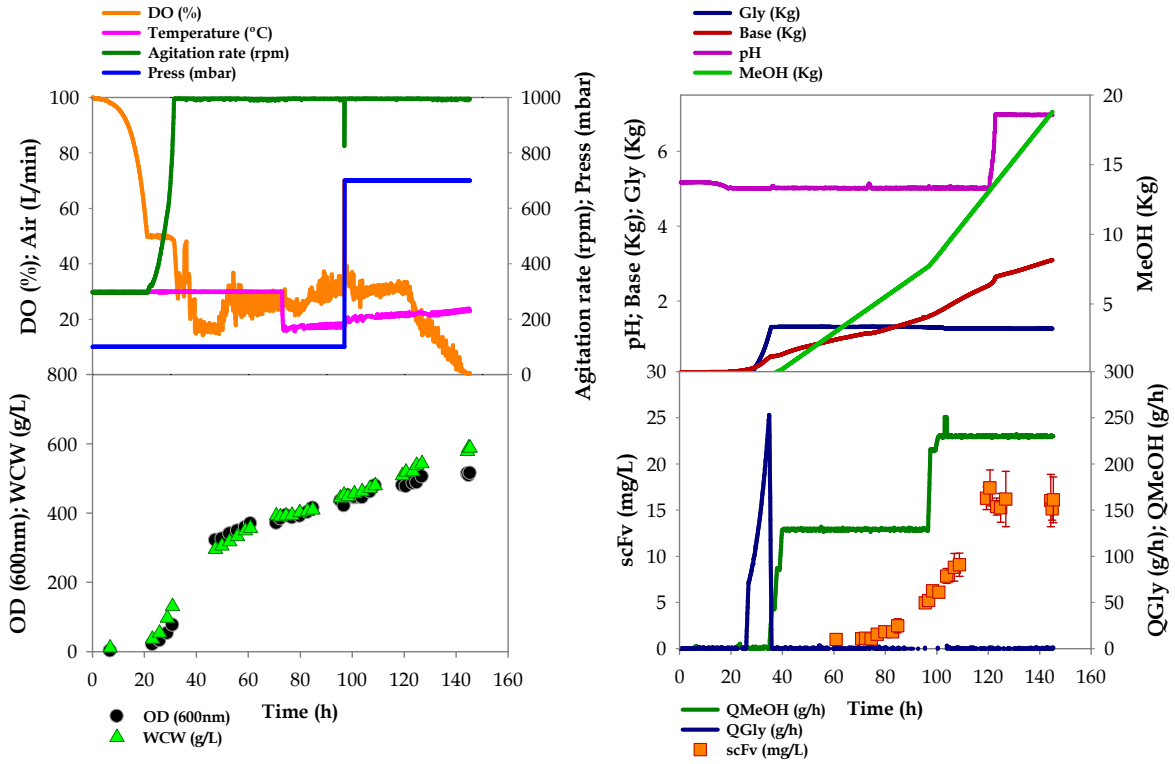


Figure D.3 Experiment D.C online and offline data.

Experiment D.D

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in the glycerol batch and fed-batch phases, after decrease at 23.7 ^{o}C until the end. The pH was controlled first at 5, after 70 h of induction increase pH= 7 controlled with 3.1 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.8 followed by methanol addition at constant feeding rate after the methanol adaptation phase (Table D.8). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.9.

Table D.8 Glycerol and methanol addition profiles, experiment D.D.

Glycerol fed batch phase	Methanol fed batch phase
$F= F_0 \times e^{(\mu \times t)}$;	$4 < t_{MFB} \leq 62$ h → F_{MeOH3}= 130.0 g/h;
F_0= 65.0 g/h; μ= 0.16 h^{-1}; t= 9.0 h;	$t_{MFB} > 62$ h → F_{MeOH4}= 230.0 g/h;

Table D.9 Bioreaction parameters, experiment D.D.

Variable	Glycerol phase	Methanol phase
Final Temperature (^{o}C)	30.0	23.7
pH	5.0	7.0
Final DO (%)	52.7	2.0
Glycerol added (Kg)	1.2	----
Methanol added (kg)	----	19.0
Base added (Kg)	0.5	2.6
Time after induction (h)	----	110.6
Total time (h)	----	145.8
Final WCW (g/L)	----	573.1
Final Product (scFv – mg/L)	----	14.3

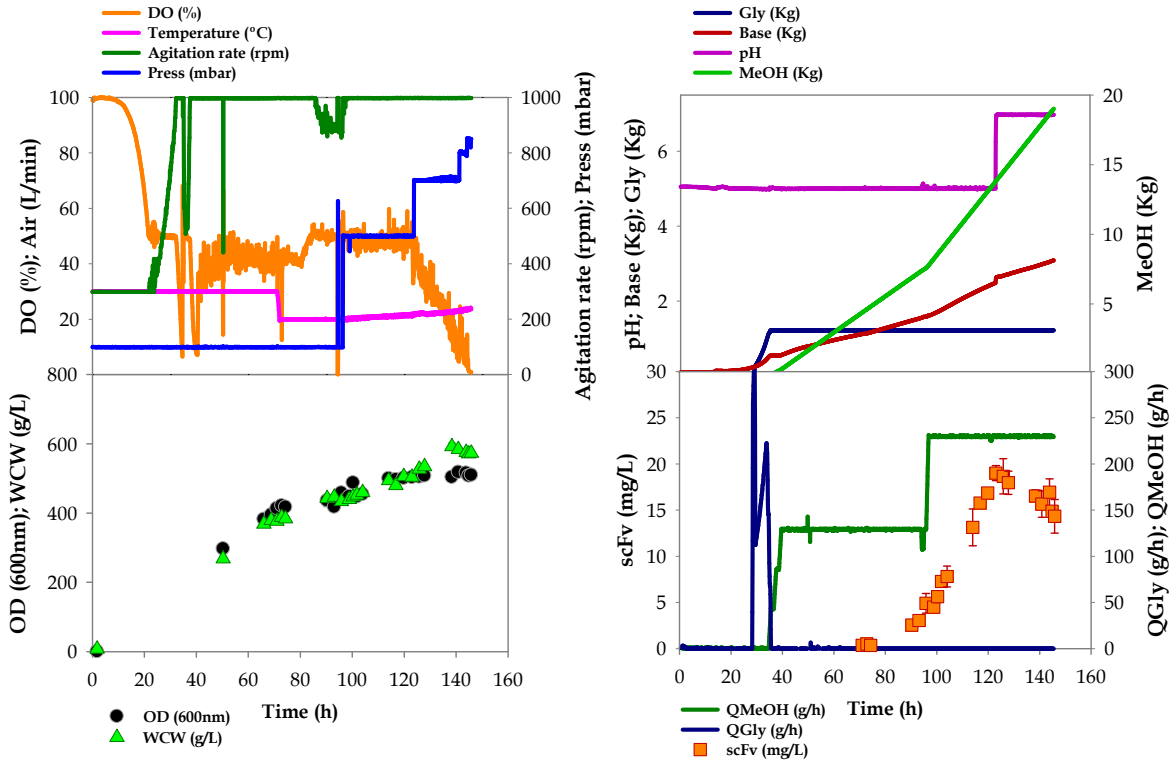


Figure D.4 Experiment D.D online and offline data.

Experiment D.E

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in all phases. The pH was controlled at 5 with 1.9 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.10 followed by methanol addition at constant feeding rate after the methanol adaptation phase (Table D.10). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.11.

Table D.10 Glycerol and methanol addition profiles, experiment D.E.

Glycerol fed batch phase	Methanol fed batch phase
$F = F_0 \times e^{(\mu \times t)}$;	$t \leq 2$ h → F_{MeOH1}= 43.0 g/h;
F_0= 65.0 g/h;	$2 < t \leq 4$ h → F_{MeOH2}= 86.0 g/h;
μ= 0.16 h^{-1}; t= 9.0 h;	$4 < t \leq 62$ h → F_{MeOH3}= 129.0 g/h;
	t > 62 h → F_{MEOH4}= 230.0 g/h;

Table D.11 Bioreaction parameters, experiment D.E.

Variable	Glycerol phase	Methanol phase
Final Temperature (^{o}C)	30.0	30.0
pH	5.0	5.0
Final DO (%)	10.5	44.4
Glycerol added (Kg)	1.3	----
Methanol added (kg)	----	13.4
Base added (Kg)	0.4	1.5
Time after induction (h)	----	87.2
Total time (h)	----	124.0
Final WCW (g/L)	----	434.2
Final Product (scFv – mg/L)	----	11.9

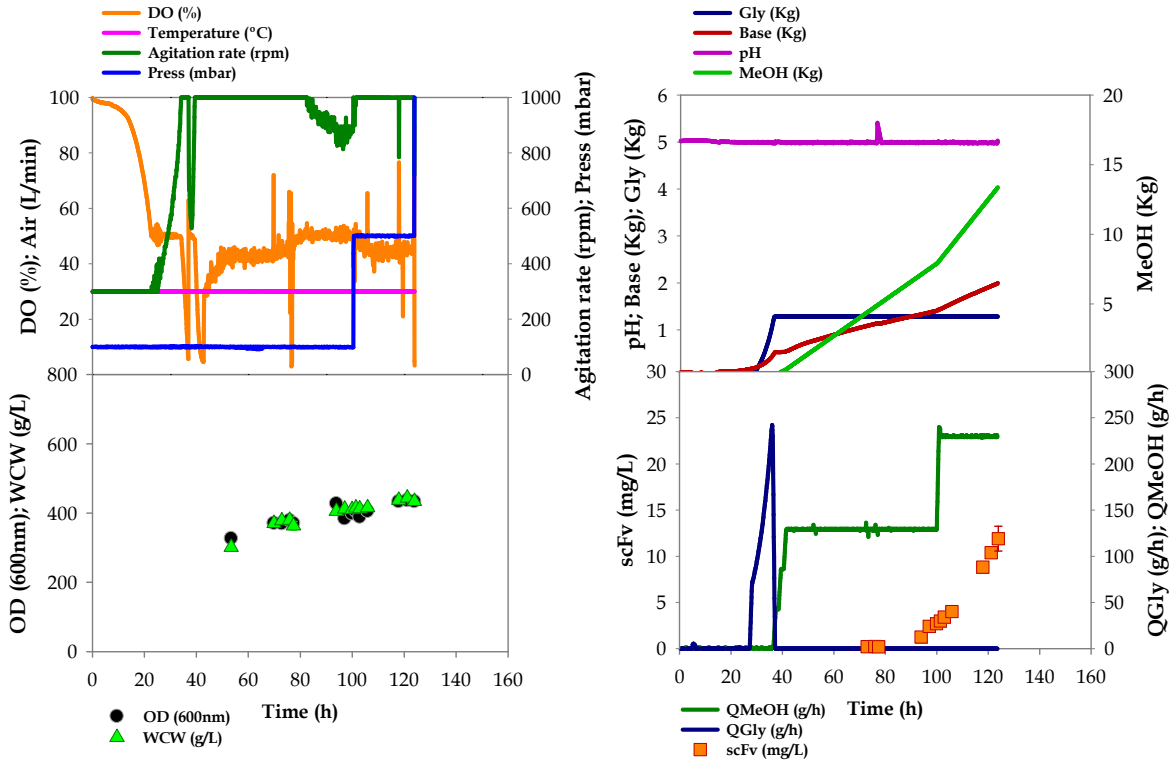


Figure D.5 Experiment D.E online and offline data.

Experiment D.F

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in the glycerol batch and fed-batch phases, after decrease at 23.7 ^{o}C until the end. The pH was controlled at 5, after glycerol fed-batch phase controlled at 4 with 3.0 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.12 followed by methanol addition controlled according to exponential profile after the methanol adaptation phase (Table D.12). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.13.

Table D.12 Glycerol and methanol addition profiles, experiment D.F.

Glycerol fed batch phase	Methanol fed batch phase
$F = F_0 \times e^{(\mu \times t)}$;	$t \leq 27$ h → F_{MeOH1}= 30.0 g/h;
F_0= 30.0 g/h; μ_1= 0.16 h^{-1}; t_1= 5.0 h;	27 < t ≤ 52 h → DO setpoint= 60%;
μ_2= 0.025 h^{-1}; t_2= 5.0 h;	52 < t ≤ 76 h → DO setpoint= 40%;
	76 < t ≤ 100 h → DO setpoint= 5%;
	t > 100 h → DO setpoint= 40%.

Table D.13 Bioreaction parameters, experiment D.F.

Variable	Glycerol phase	Methanol phase
Final Temperature (^{o}C)	30.0	23.7
pH	5.0	4.0
Final DO (%)	49.3	39.5
Glycerol added (Kg)	0.6	----
Methanol added (kg)	----	19.6
Base added (Kg)	0.3	2.8
Time after induction (h)	----	122.5
Total time (h)	----	162.2
Final WCW (g/L)	----	632.3
Final Product (scFv – mg/L)	----	30.8

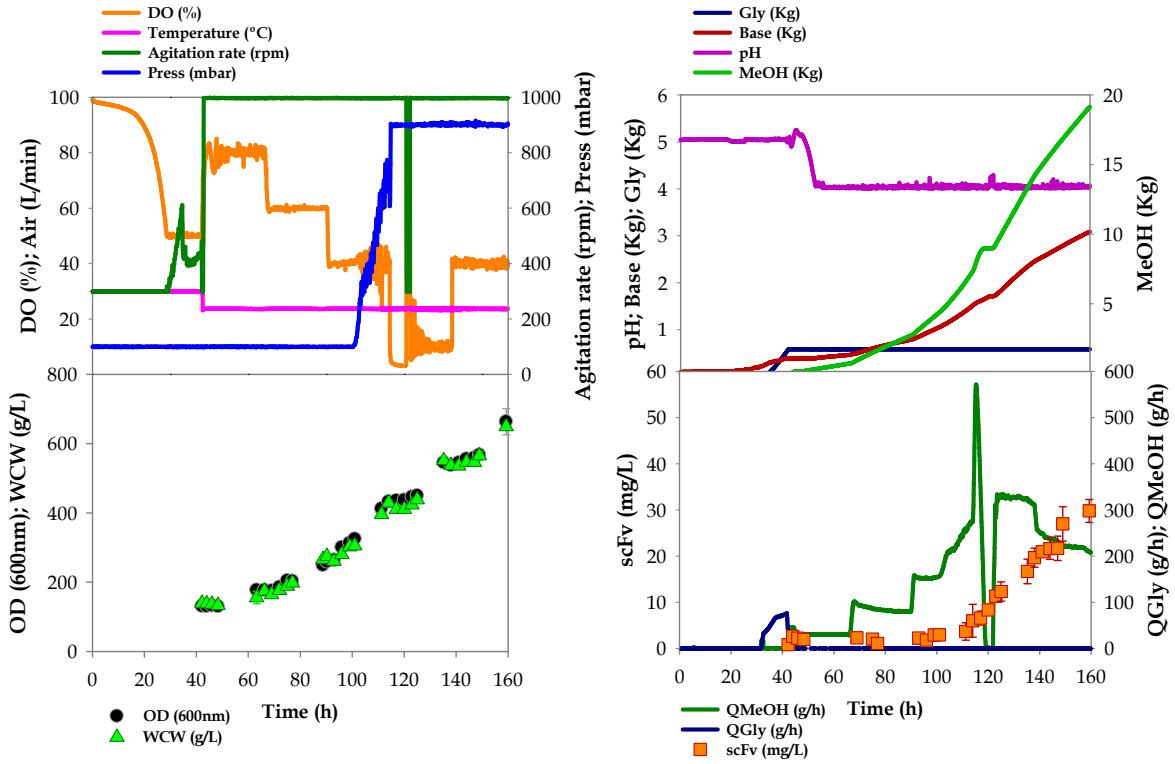


Figure D.6 Experiment D.F online and offline data.

Experiment D.G

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in the glycerol batch and fed-batch phases, after decrease at 23.7 ^{o}C until the end. The pH was controlled at 5, after glycerol fed-batch phase controlled at 4 with 1.8 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.14 followed by methanol addition controlled according to the profile after the methanol adaptation phase (Table D.14). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.15.

Table D.14 Glycerol and methanol addition profiles, experiment D.G.

Glycerol fed batch phase	Methanol fed batch phase
$F= F_0 \times e^{(\mu \times t)}$;	$F= F_0 \times e^{(\mu \times t)}$;
F_0= 30.0 g/h; μ_1= 0.16 h^{-1}; t_1= 8.5 h; μ_2= 0.025 h^{-1}; t_2= 4.0 h;	F_0= 38.5 g/h; μ_1= 0.013 h^{-1}; t_1> 100 h;

Table D.15 A Bioreaction parameters, experiment D.G.

Variable	Glycerol phase	Methanol phase
Final Temperature (^{o}C)	30.0	23.7
pH	5.0	4.0
Final DO (%)	50.7	39.9
Glycerol added (Kg)	1.0	----
Methanol added (kg)	----	9.8
Base added (Kg)	0.4	1.4
Time after induction (h)	----	111.1
Total time (h)	----	150.0
Final WCW (g/L)	----	479.6
Final Product (scFv – mg/L)	----	30.7

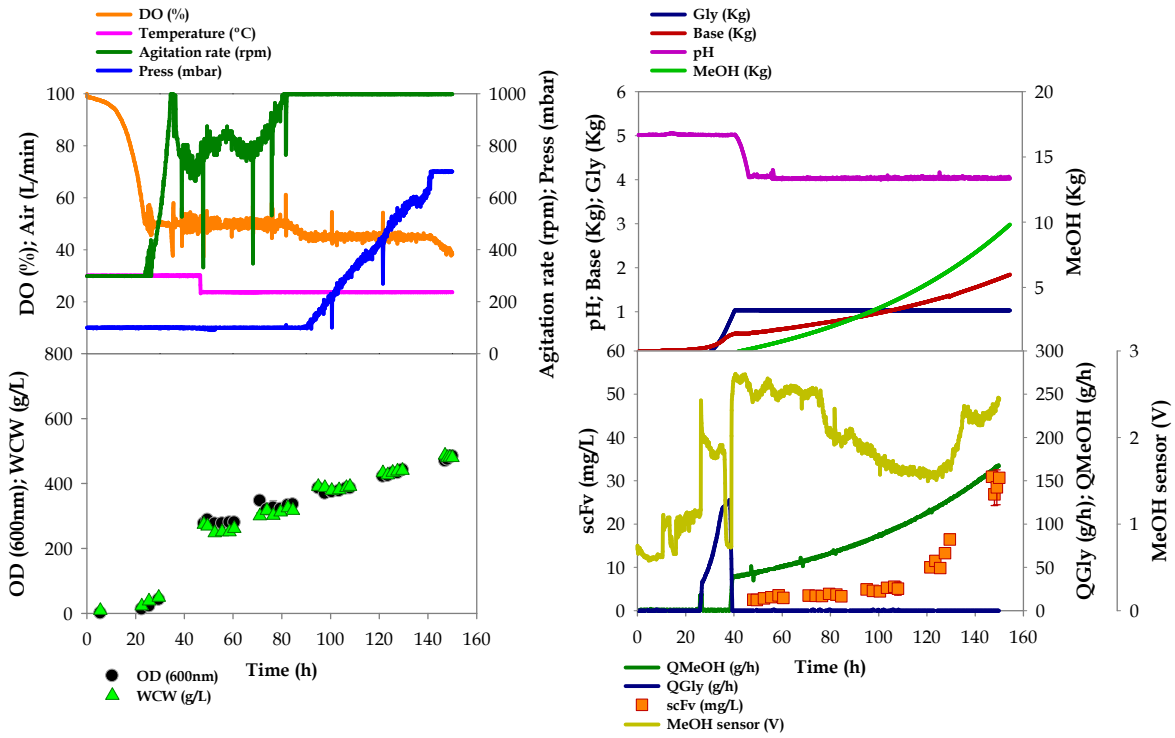


Figure D.7 Experiment D.G online and offline data.

Experiment D.H

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in all the batch phases. The pH was controlled at 5, after glycerol fed-batch phase controlled at 6.39 with 1.8 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.16 followed by methanol addition controlled according to the profile after the methanol adaptation phase (Table D.17). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.18.

Table D.16 Glycerol addition profile, experiment D.H.

Glycerol fed batch phase
$F= F_0 \times e^{(\mu \times t)}$;
F_0= 30.0 g/h; μ_1= 0.16 h^{-1}; t_1= 8.5 h;
μ_2= 0.025 h^{-1}; t_2= 4.0 h;

Table D.17 Methanol addition profile, experiment D.H.

Time (h)	MeOH (kg)	Time (h)	MeOH (kg)	Time (h)	MeOH (kg)	Time (h)	MeOH (kg)
0.00	0.00	26.00	3.33	52.00	9.36	78.00	11.14
2.00	0.20	28.00	3.66	54.00	9.69	80.00	11.23
4.00	0.41	30.00	4.02	56.00	9.95	82.00	11.32
6.00	0.63	32.00	4.41	58.00	10.16	84.00	11.40
8.00	0.87	34.00	4.83	60.00	10.30	86.00	11.48
10.00	1.12	36.00	5.28	62.00	10.40	88.00	11.57
12.00	1.37	38.00	5.78	64.00	10.50	90.00	11.65
14.00	1.64	40.00	6.33	66.00	10.59	92.00	11.73
16.00	1.91	42.00	6.91	68.00	10.69	94.00	11.82
18.00	2.18	44.00	7.48	70.00	10.78	96.00	11.90
20.00	2.45	46.00	8.02	72.00	10.87	98.00	11.98
22.00	2.72	48.00	8.52	74.00	10.96	100.00	12.07
24.00	3.02	50.00	8.97	76.00	11.05		

NOTE: This table was simulated by Matlab.

Table D.18 A Bioreaction parameters, experiment D.H.

Variable	Glycerol phase	Methanol phase
Final Temperature (ºC)	30.0	30.0
pH	5.0	6.3
Final DO (%)	55.7	50.9
Glycerol added (Kg)	1.0	----
Methanol added (kg)	----	12.3
Base added (Kg)	0.4	1.4
Time after induction (h)	----	108.5
Total time (h)	----	153.0
Final WCW (g/L)	----	475.4
Final Product (scFv – mg/L)	----	52.5

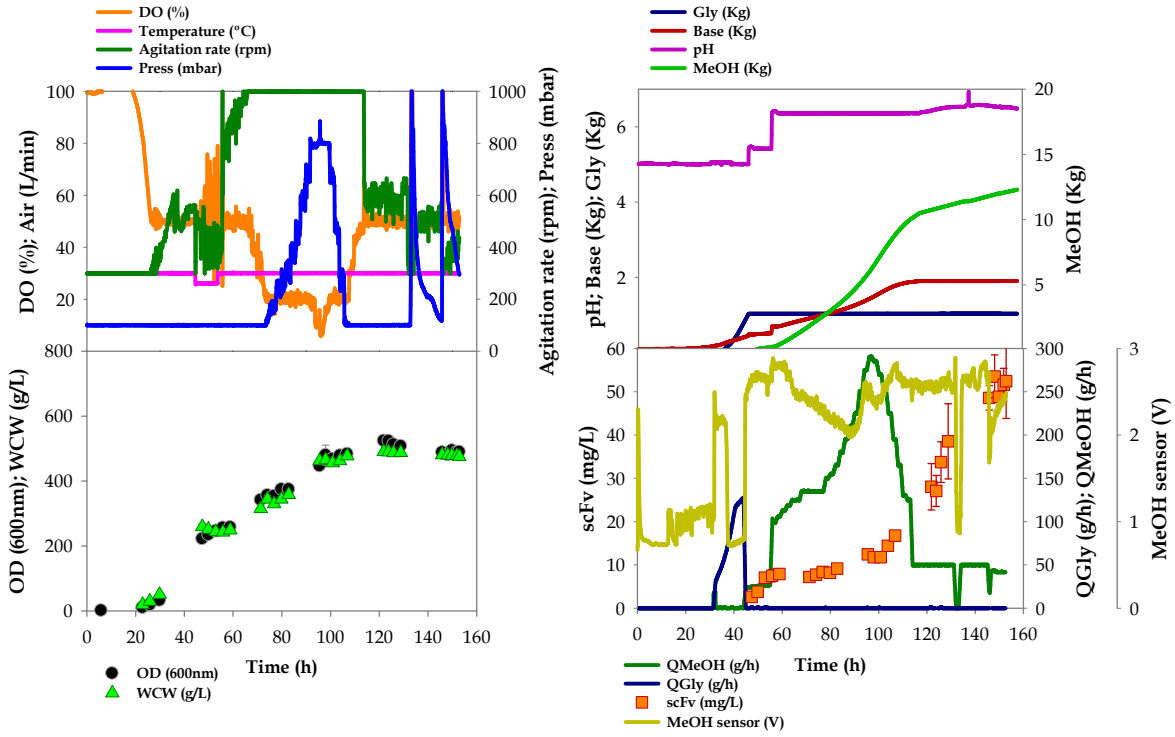


Figure D.8 Experiment D.H online and offline data.

Experiment D.I

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in all the batch phases. The pH was controlled at 5, after glycerol fed-batch phase controlled at 6.9 with 1.7 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to exponential profile present in Table D.19 followed by methanol addition was controlled according to the profile after the methanol adaptation phase (Table D.20). Methanol and glycerol additions were gravimetrically controlled. The operational conditions are present in Table D.21.

Table D.19 Glycerol addition profile, experiment D.I.

Glycerol fed batch phase
$F = F_0 \times e^{(\mu \times t)}$;
F_0= 30.0 g/h; μ_1= 0.16 h^{-1}; t_1= 8.5 h;
μ_2= 0.025 h^{-1}; t_2= 4.0 h;

Table D.20 Methanol addition profile, experiment D.I.

Methanol fed batch phase
$F = a_1 \times t^2 + a_2 \times t + a_3$;
a_1 = 0.00072;
a_2 = 0.01734;
a_3 = 0.17364; t = 100 h;

NOTE: This table was simulated by Matlab.

Table D.21 Bioreaction parameters, experiment D.I.

Variable	**Glycerol phase**	**Methanol phase**
Final Temperature (^{o}C)	30.0	30.0
pH	5.0	6.9
Final DO (%)	54.0	34.5
Glycerol added (Kg)	1.0	----
Methanol added (kg)	----	10.5
Base added (Kg)	0.4	1.3
Time after induction (h)	----	106.0
Total time (h)	----	149.0
Final WCW (g/L)	----	428.1
Final Product (scFv – mg/L)	----	8.4

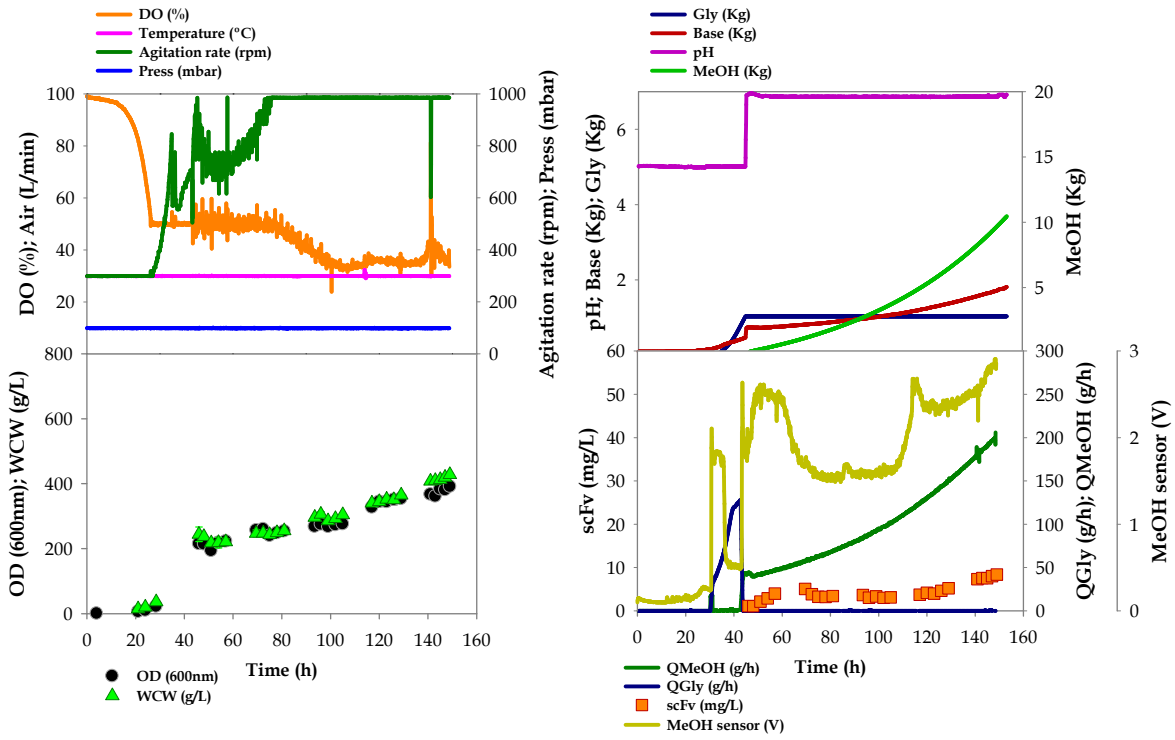


Figure D.9 Experiment D.I online and offline data.

Appendix E

Experiment E.A

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in all the batch phases. The pH was controlled at 5.0 in the beginning, in glycerol fed-batch phase pH decrease to 4.0, pH controlled with 2.1 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to DO set–point of constant 5% during the oxygen transfer limitation phase with a BSM (baseline formulation) medium. Glycerol addition was gravimetrically controlled. The operational conditions are present in Table E.1.

Table E.1 Bioreaction parameters, experiment E.A.

Variable	Glycerol phase
Temperature (^{o}C)	30.0
Final pH	4.0
Final DO (%)	5.0
Glycerol added (Kg)	13.0
Base added (Kg)	2.1
Time (h)	122.0
Final WCW (g/L)	880.8
Final Product (scFv – mg/L)	57.5

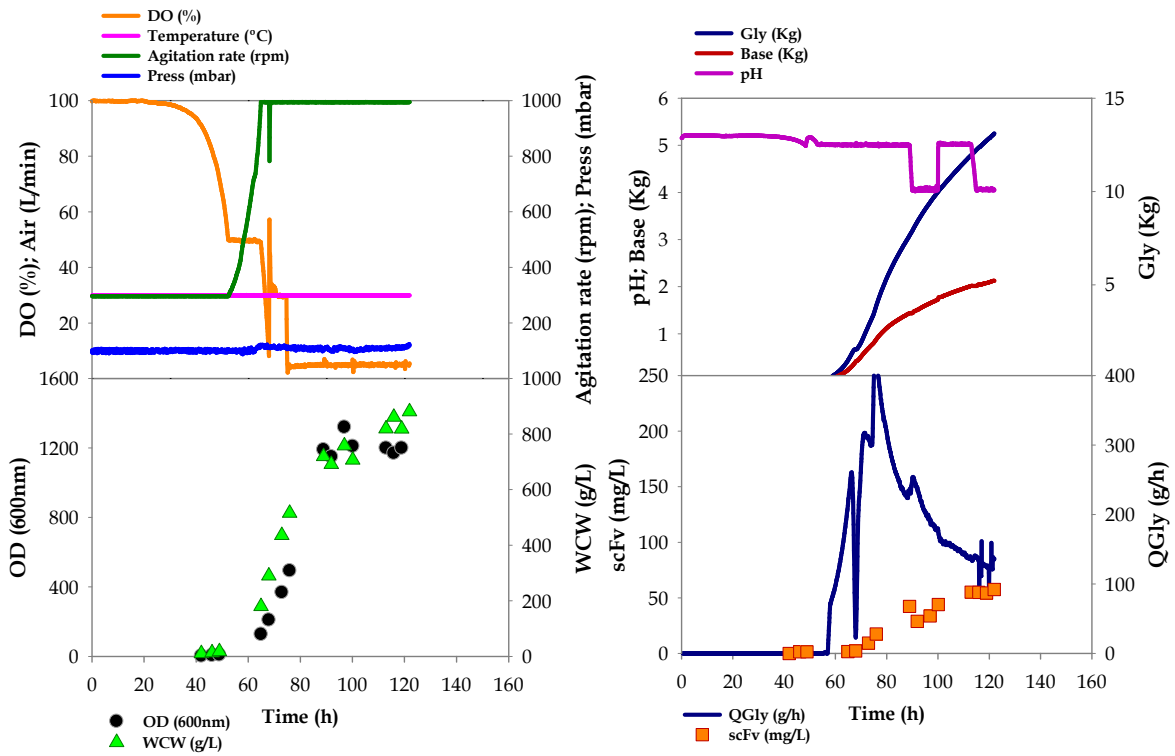
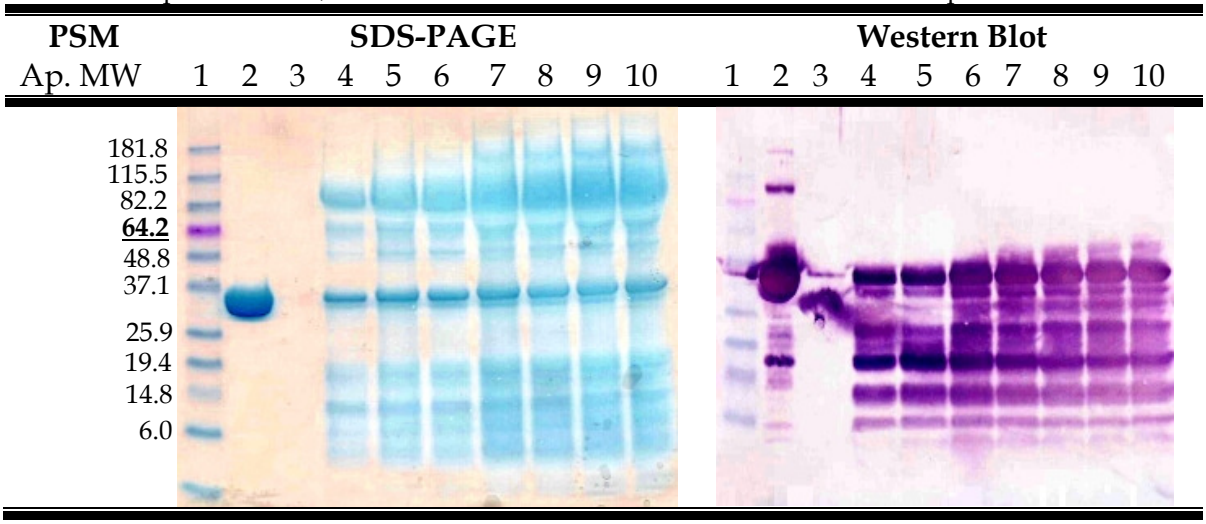


Figure E.1 Experiment E.A online and offline data.

Table E.2 Experiment E.A, SDS-PAGE and Western Blot results of selected samples.



Legend: 1) PSM; 2) Standard scFv 1; 3) Standard scFv 2; 4) Sample time= 89 h; 5) Sample time= 92 h; 6) Sample time= 97 h; 7) Sample time= 100 h; 8) Sample time= 113 h; 9) Sample time= 116 h; 10) Sample time= 119 h.

Experiment E.B

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ^{o}C in all the batch phases. The pH was controlled at 5.0 in the beginning, in glycerol fed-batch phase pH decrease to 4.0 and next decrease to 3.0, pH controlled with 2.1 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to DO set–point of constant 5% during the oxygen transfer limitation phase with a BSM (baseline formulation) medium. Glycerol addition was gravimetrically controlled. The operational conditions are present in Table E.3.

Table E.3 Bioreaction parameters, experiment E.B.

Variable	Glycerol phase
Temperature (^{o}C)	30.0
Final pH	3.0
Final DO (%)	5.0
Glycerol added (Kg)	11.0
Base added (Kg)	2.1
Time (h)	150.0
Final WCW (g/L)	642.7
Final Product (scFv – mg/L)	62.3

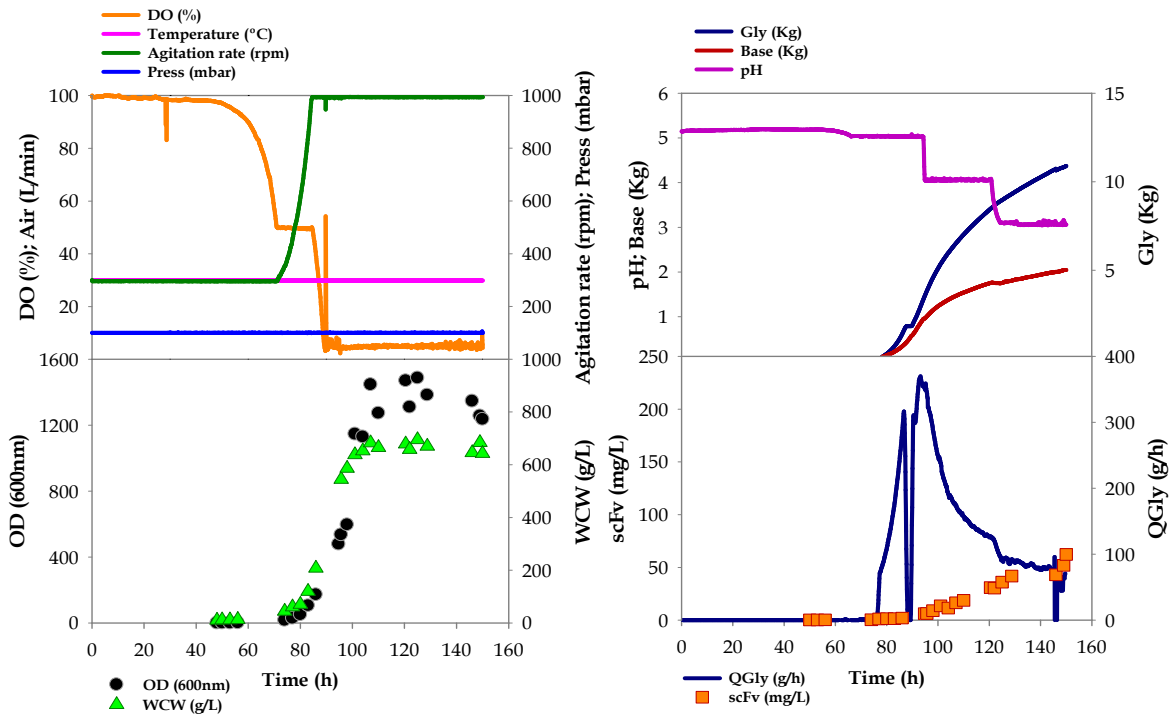
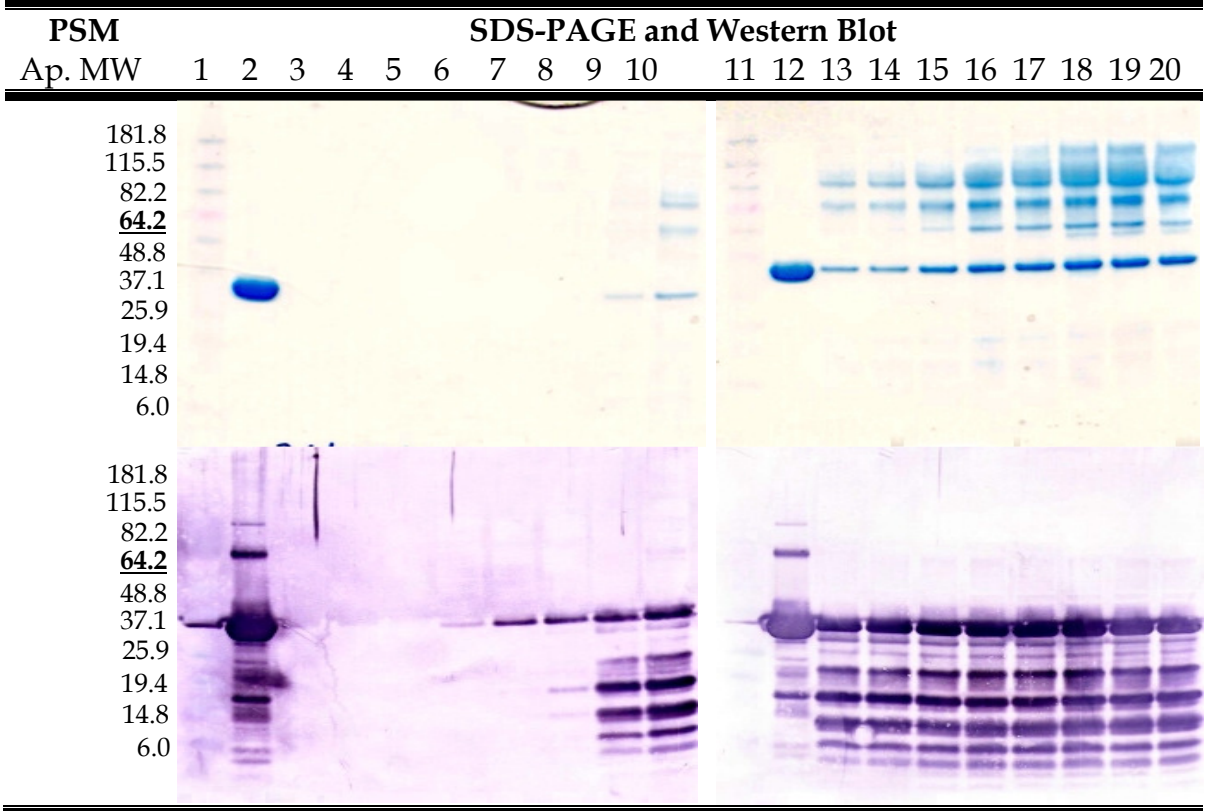


Figure E.2 Experiment E.B online and offline data.

Table E.4 Experiment E.B, SDS-PAGE and Western Blot results of selected samples.



Legend: 1) PSM; 2) Standard scFv; 3) Sample time= 23 h; 4) Sample time= 30 h; 5) Sample time= 53 h; 6) Sample time= 56 h; 7) Sample time= 77 h; 8) Sample time= 83 h; 9) Sample time= 96 h; 10) Sample time= 101 h; 11) PSM; 12) Standard scFv; 13) Sample time= 107 h; 14) Sample time= 110 h; 15) Sample time= 121 h; 16) Sample time= 122 h; 17) Sample time= 125 h; 18) Sample time= 129 h; 19) Sample time= 146 h; 20) Sample time= 150 h.

Experiment E.C

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 27 °C in all the batch phases. The pH was controlled at 5.0 with 2.2 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to DO set-point of constant 5 % during the oxygen transfer limitation phase with a BSM (baseline formulation) medium. Glycerol addition was gravimetrically controlled. The operational conditions are present in Table E.5.

Table E.5 Bioreaction parameters, experiment E.C.

Variable	Glycerol phase
Final Temperature (°C)	27.0
pH	5.0
Final DO (%)	5.0
Glycerol added (Kg)	13.7
Base added (Kg)	2.2
Time (h)	124.5
Final WCW (g/L)	816.5
Final Product (scFv – mg/L)	187.0

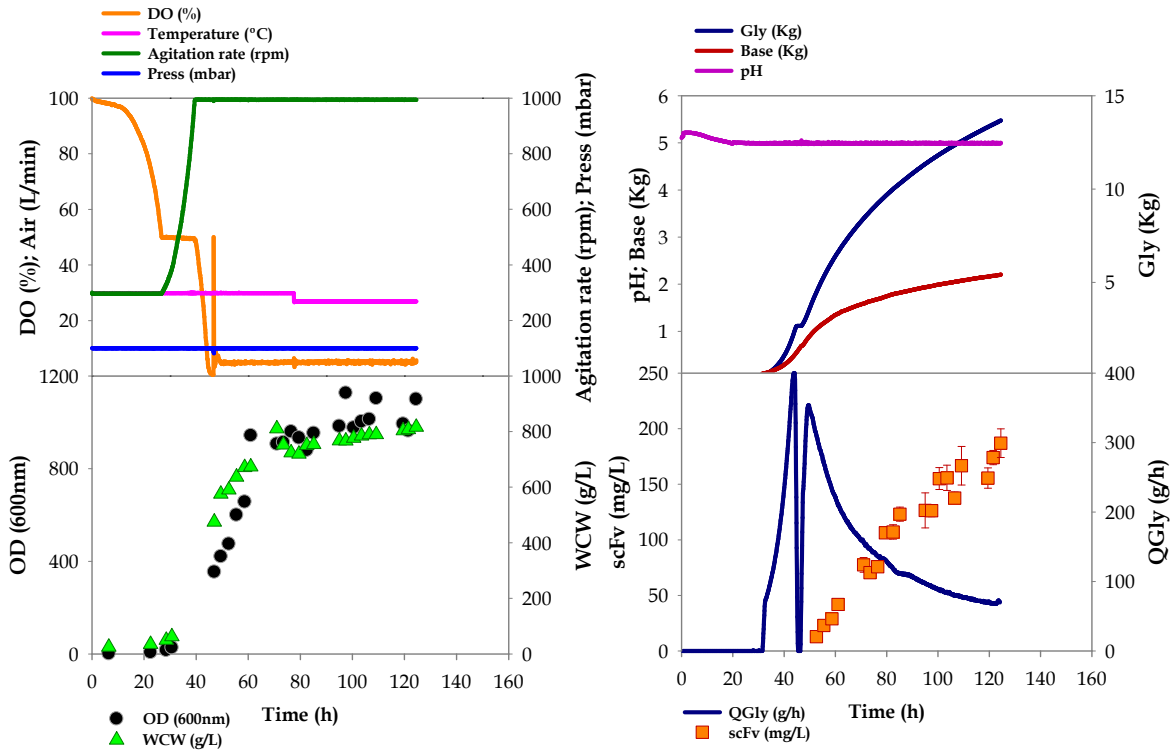


Figure E.3 Experiment E.C online and offline data.

Experiment E.D

A 50 L bioreactor was inoculated with 750 mL inoculum in BSM medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ºC in all the batch phases. The pH was controlled at 5.0 with 2.2 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to DO set-point of constant 5 % during the oxygen transfer limitation phase with a BSM (baseline formulation) medium. Glycerol addition was gravimetrically controlled. The operational conditions are present in Table E.6.

Table E.6 Bioreaction parameters, experiment E.D.

Variable	Glycerol phase
Temperature (ºC)	30.0
pH	5.0
Final DO (%)	5.0
Glycerol added (Kg)	13.6
Base added (Kg)	2.2
Time (h)	151.5
Final WCW (g/L)	806.5
Final Product (scFv – mg/L)	219.3

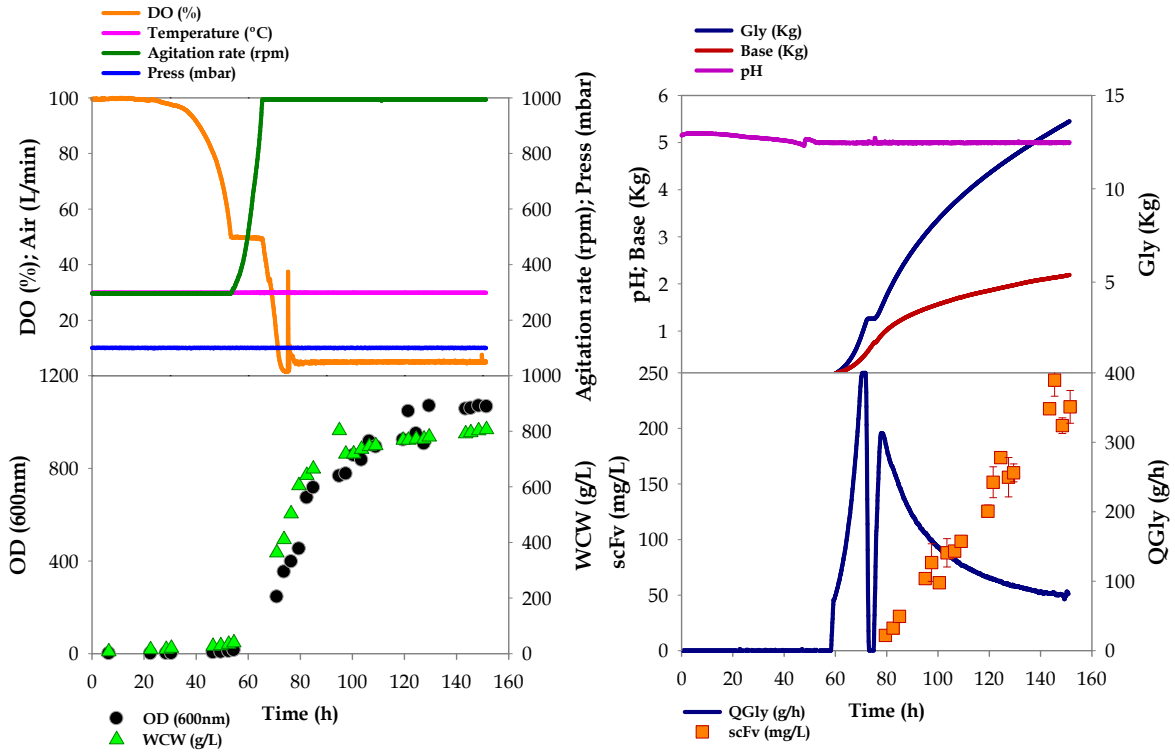


Figure E.4 Experiment E.D online and offline data.

Experiment E.E

A 50 L bioreactor was inoculated with 750 mL inoculum in PFEM1 medium to an initial volume of 15 L. In this experiment temperature was kept constant at 30 ºC in all the batch phases. The pH was controlled at 5.0 with 2.9 L of 25 % NH_4OH in all bioreaction phases. Addition of glycerol solution was controlled according to DO set-point of constant 5 % during the oxygen transfer limitation phase with a PFEM1 medium. Glycerol addition was gravimetrically controlled. Due to the low initial salts concentration in PFEM1, it was necessary to feed a concentrated salts solution (concentrated PMS, e. g. 10 times concentrated) along time during the GFB phase. The operational conditions are present in Table E.7.

Table E.7 Bioreaction parameters, experiment E.E.

Variable	Glycerol phase
Temperature (ºC)	30.0
pH	5.0
Final DO (%)	5.3
Glycerol added (Kg)	22.9
Base added (Kg)	2.9
Concentrated PMS added (Kg)	3.0
Time (h)	147.2
Final WCW (g/L)	669.6
Final Product (scFv – mg/L)	271.2

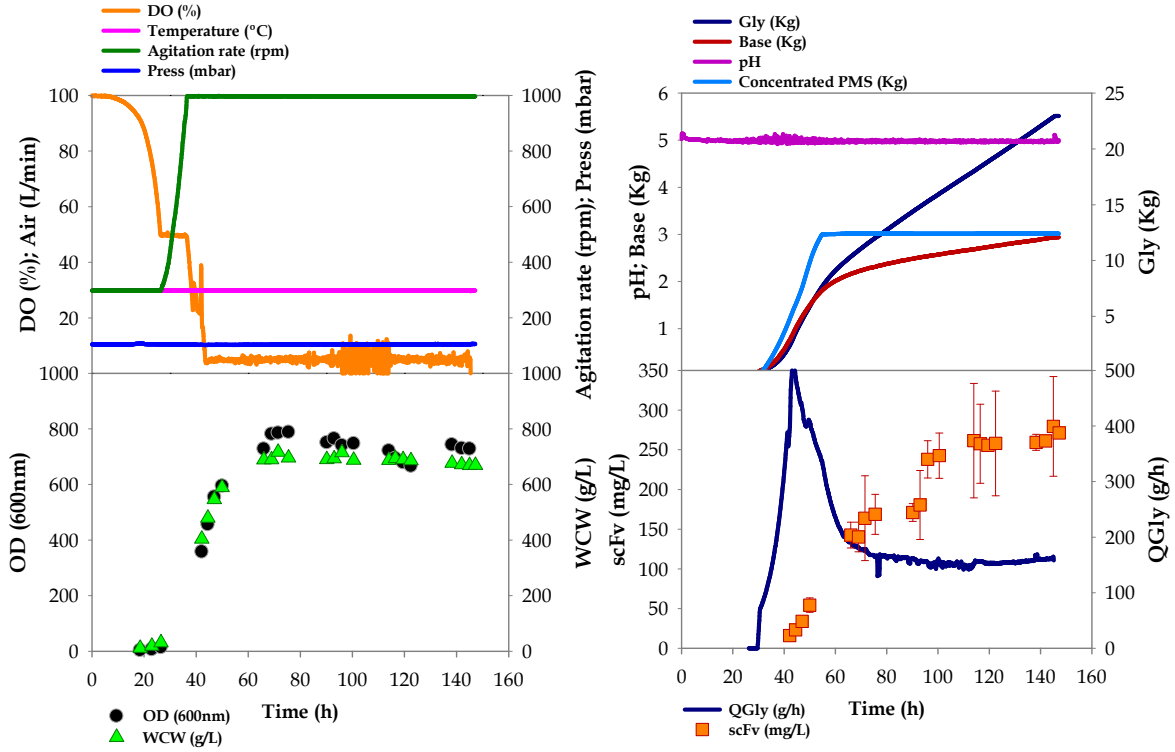


Figure E.5 Experiment E.E online and offline data.