UNIVERSITAT AUTÒNOMA DE BARCELONA
ESCOLA D'ENGINYERIA
DEPARTAMENT D'ENGINYERIA QUÍMICA

Modeling bioreactors for the production of recombinant proteins in high-cell density cultures of *Escherichia coli*

Memòria per a optar al grau de Doctor
per la Universitat Autònoma de Barcelona

Daniel Calleja Martínez

Setembre 2014

JOSEP LÓPEZ SANTÍN, Catedràtic del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona i CARLES DE MAS ROCABAYERA, Professor Titular del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona,

CERTIFIQUEM:

Que l'Enginyer Químic Daniel Calleja Martínez ha dut a terme, sota la nostra direcció, el treball titulat **"Modeling bioreactors for the production of recombinant proteins in high-cell density cultures of *Escherichia coli*"** que es presenta en aquesta memòria i que constitueix la seva Tesi per optar al grau de Doctor en Biotecnologia per la Universitat Autònoma de Barcelona.

I perquè es prengui coneixement i consti als efectes oportuns, signem la present a l'Escola d'Enginyera de la Universitat Autònoma de Barcelona.

Bellaterra, Setembre del 2014

Dr. Josep López Santín

Dr. Carles de Mas Rocabayera

“There must be a beginning of any great matter,
but the continuing unto the end until it be
thoroughly finished yields the true glory”
-Sir Francis Drake

“If everything seems to be under control,
you are just not going fast enough”
-Mario Andretti

Table of contents

Summary

The present work is focused on the study of recombinant protein production in high-cell density fed-batch cultures of *E.coli*. In particular, the development of a model capable to predict Rhamnulose-1-Phosphate Aldolase (RhuA) production is the objective.

Firstly, a qualitative and quantitative study about the variables involved in protein production has been made. This study has permitted the evaluation of the impact of the main experimental variables (I/X and biomass concentration at induction and the specific growth rate) on protein production (in mass and activity units). Using a Response Surface Methodology (RSM), that is a statistical methodology, it can be determined the optimal experimental conditions that conduce to a maximum in protein production, and set the operating working conditions.

Secondly, because a deeper study about the importance of IPTG in inducible *E.coli* systems is needed, a model describing inducer uptake has been developed, calibrated and validated. IPTG uptake model has been developed in two steps: a) using a *lac*Y deficient strain, non-specific transport mechanisms have been modeled; b) in addition to non-specific transport mechanisms, lactose permeases (specific transporting proteins for lactose –and IPTG) contribution has been added. It has been demonstrated that the model is capable to predict IPTG depletion from culture medium, not only for the model strain, but also for three different strains.

Thirdly, a coupled model, composed by three different ones (biomass growth, IPTG uptake and protein production) has been proposed. In this case, a new protein production model has been presented, using as inputs the time evolution of the variables involved in the other two models. Protein production rate (expressed in mass) can be related to the amount of inducer bound to the repressor. The binding equilibrium depends on the intracellular concentration of IPTG along time, which is an output of the IPTG uptake model. Otherwise, biomass growth model is able to predict

biomass concentration and the total volume into the bioreactor from the beginning of the batch phase.

Finally, the protein production model, coupled with the IPTG uptake model, has been extended to the production of different proteins (Fructose-6-Phosphate Aldolase and ω-Transaminase) using different expression systems. In this case, the expression system's dependent parameters have been identified, and the model has demonstrated that, estimating those parameters is also capable to predict, properly, the protein production along time.

To sum up, this work presents a new model, which contributes to the prediction of protein production in different inducible *E.coli* expression systems.

Resum

Aquest treball es centra en l'estudi de la producció de proteïnes recombinants en cultius semi-continus d'alta densitat cel·lular utilitzant *E.coli*. Particularment, l'objectiu és el desenvolupament d'un model que sigui capaç de predir l'evolució amb el temps de la producció de l'Aldolasa Ramnulosa-1-Fosfat (RhuA).

Primer, s'ha dut a terme un estudi a nivell qualitatiu i quantitatiu de les variables que juguen un paper fonamental en la producció de proteïna. Aquest estudi ha permès l'avaluació de l'impacte en la producció de proteïna (tant en unitats de massa com activitat) de les principals variables experimentals (I/X i biomassa en el moment d'inducció i la velocitat específica de creixement). La Metodologia de Superfície de Resposta (MSR) permet la determinació de les condicions òptimes de cultiu per tal de maximitzar la producció, a la vegada que permet la determinació de les condicions de treball més adequades.

Segon, s'ha desenvolupat, calibrat i validat un model de transport d'IPTG des del medi de cultiu a l'interior cel·lular. Aquest model aportarà un estudi més profund del sistema d'inducció. Està dividit en dues etapes: a) amb la utilització d'una soca deficient en lactosa permeases (les proteïnes responsables del transport específic de lactosa -i IPTG) s'han pogut determinar els mecanismes de transport no específics; b) la contribució al transport de les proteïnes específiques s'ha afegit als mecanismes no específics. S'ha demostrat que el model és capaç de predir, no només l'evolució de la concentració d'inductor per a la soca model, sinó que ho és també per a tres soques diferents.

Tercer, s'ha presentat un model acoblat (partint de tres diferents models: creixement cel·lular, transport d'IPTG i producció de proteïna). El model de producció de proteïna usa com a variables d'entrada les evolucions temporals de les variables dels altres dos models. La velocitat de producció de proteïna es pot relacionar amb la quantitat d'inductor unit al repressor. L'equilibri d'unió depèn de la concentració intracel·lular

d'inductor, que és calculada mitjançant el model de transport d'IPTG. Per altra banda, el model de creixement cel·lular és capaç de predir l'evolució amb el temps de la concentració de biomassa i de volum total en el bioreactor, des del principi de l'etapa discontínua.

Per últim, el model de producció de proteïna, juntament amb el de transport d'inductor, es pot estendre a la producció d'altres proteïnes recombinants (Fructosa-6-Fosfat Aldolasa i ω-Transaminasa) usant diferents sistemes d'expressió. Per a fer-ho, és necessari identificar els paràmetres que són dependents del sistema d'expressió i estimar-ne els valors. S'ha comprovat que el model és capaç de predir, acceptablement, l'evolució amb el temps de la producció de les noves proteïnes.

En resum, aquest treball aporta un nou model que contribueix a la predicció de la producció de proteïnes recombinants usant diferents sistemes d'expressió en *E.coli*.

Nomenclature

	units	description
F_B	$L \cdot h^{-1}$	alkali flow rate
F_s	$L \cdot h^{-1}$	feeding flow rate
g_s	--	function describing metabolic burden
$[IPTG]_e$	$\mu mol\ IPTG \cdot L^{-1}_{medium}$	extracellular IPTG concentration
$[IPTG]_{e,0}$	$\mu mol\ IPTG \cdot L^{-1}_{medium}$	extracellular IPTG concentration at induction
$[IPTG]_i$	$\mu mol\ IPTG \cdot L^{-1}_{cell}$	intracellular IPTG concentration
$K_{R\text{-}IPTG}$	mM	equilibrium constant for the repressor-IPTG bond
M	$molecules \cdot cell^{-1}$	mRNA molecules of lacI
m_{SX}	$g_{glucose} \cdot h^{-1} \cdot g^{-1}DCW$	maintenance coefficient
N	$plasmid \cdot cell^{-1}$	plasmid copy number
n	--	cooperativity of the binding lacI-IPTG
P	$mgRhuA \cdot g^{-1}DCW$	specific protein in mass
P_{ind}	$mgRhuA \cdot g^{-1}DCW$	basal specific protein in mass units
q_p	$mgRhuA \cdot g^{-1}DCW \cdot h^{-1}$	specific protein production rate
q_{p0}	$mgRhuA \cdot g^{-1}DCW \cdot h^{-1}$	initial specific protein production rate
r	$\mu mol\ IPTG \cdot h^{-1} \cdot L^{-1}_{medium}$	net transport rate
R_{free}	$molecule \cdot cell^{-1}$	free repressor
R_{IPTG}	$molecule \cdot cell^{-1}$	repressor bond to IPTG
S	$g \cdot L^{-1}$	glucose concentration
S_0	$g \cdot L^{-1}$	glucose concentration at batch beginning
S_f	$g \cdot L^{-1}$	glucose concentration in the feeding solution
t	h	time
t_{ind}	h	time since induction
U	$AU \cdot g^{-1}DCW$	specific protein in activity
U_{ind}	$AU \cdot mg^{-1}RhuA$	basal specific protein in activity units
V	L	total volume

$\mathbf{V_0}$	L	total volume at batch beginning
$\mathbf{V_{cel}}$	L_{cell}	volume of cells
$\mathbf{V_m}$	L_{medium}	culture medium volume
$\mathbf{X}$	$gDCW \cdot L^{-1}$	biomass concentration
$\mathbf{X_0}$	$gDCW \cdot L^{-1}$	biomass concentration at batch beginning
$\mathbf{X_{ind}}$	$gDCW \cdot L^{-1}$	biomass concentration at induction
$\mathbf{Y_{XS}}$	$gDCW \cdot g^{-1}$ glucose	biomass-substrate yield
$\mathbf{(Y_{XS})_{ap}}$	$gDCW \cdot g^{-1}$ glucose	apparent biomass-substrate yield
$\boldsymbol{\alpha_{mRNA}}$	h^{-1}	*lac*I transcription rate
$\boldsymbol{\alpha_R}$	h^{-1}	*lac*I transcription rate
$\boldsymbol{\lambda_{mRNA}}$	h^{-1}	mRNA degradation rate
$\boldsymbol{\lambda_R}$	h^{-1}	lacI degradation rate
$\boldsymbol{\mu}$	h^{-1}	punctual specific growth rate
$\boldsymbol{\mu_{fix}}$	h^{-1}	fixed specific growth rate during the fed-batch
$\boldsymbol{\mu_{max}}$	h^{-1}	maximum specific growth rate for M15 strain
$\boldsymbol{\tau_{R\text{-}O}}$	h^{-1}	time constant for the repressor-IPTG bond

Abbreviations

ATA	ω-Transaminase
AU	Activity Unit
cAMP	cyclic molecules of adenosine monophosphate
CAP	Catabolite Activator Protein
DCW	Dry Cell Weight
DHA	Dihydroxyacetone
DHAP	Dihydroxyacetone Phosphate
DM	Defined minimal Medium
DNA	Deoxyribonucleic acid
FruA	Fructose-1,6-Biphosphate Aldolase
FSA	Fructose-1-Phosphate Aldolase
FucA	Fuculose-1-Phosphate Aldolase
I/X	inducer to biomass ratio
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
lac permease	lactose permease
LB	Lysogeny Broth
NAD^+	oxidized form of Nicotinamide adenine dinucleotide
NADH	reduced form of Nicotinamide adenine dinucleotide
OD_{600}	Optical Density at λ= 600nm
pO_2	Oxygen saturation
RhuA	Rhamnulose-1-Phosphate Aldolase
RNA	Ribonucleic acid
RSM	Response Surface Methodology
TagA	Tagatose-1,6-Biphosphate Aldolase
α-GDH	α-Glycerophosphate Dehydrogenase

1 General introduction

1.1 Biotechnology

Biotechnology is defined by the OECD (Organization for Economic Co-operation and Development) as the application of science and technology to living organisms as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services.

Based on its application, it can be divided in different branches: red (for medical applications), green (for agricultural applications), blue (for marine applications) and white (for industrial applications) biotechnology.

- *White biotechnology* deals with finding the equilibrium between market demand satisfaction and the environmental impact. Competitive economic industrial processes can be reached using enzymes and microorganisms, with the possibility of water, energy and raw material savings.

- *Green biotechnology* deals with the development and using of modified cultures, allowing the reduction of soil working. This fact can conduct to fuel savings, pesticide use reductions and give the cultures more resistance to adverse conditions.

- *Red biotechnology* deals with the research and development of new medicines, vaccines and diagnosis.

- *Blue biotechnology* deals with the exploitation of the marine sources in industrial, health and alimentation processes.

Focusing on the white biotechnology, some major challenges are the production of therapeutically interesting proteins and the production of biocatalysts. It is a key technology wanting the transformation of the scientific knowledge into sustainable products.

1.2 Biocatalysis

Biocatalysis can be defined as the process that, using a biologic molecule, is able to perform a determined chemical reaction with high specificity and selectivity for the substrate and the final products, respectively. Biocatalysts are increasingly being introduced in industrial processes to assist in synthetic production of molecules (Pollard & Woodley, 2007).

Biocatalysis has some advantages and disadvantages if compared to traditional organic chemical processes (Aehle, 2007; Bommarius & Riebel, 2004; Buchholz, Kasche, & Bornscheuer, 2005). The main advantages are:

- using enzymes for a reaction allow the enhancement of the selectivity and/or the regioselectivity, and the differentiation between enantiomerically different substrates;
- usually, minor generation of subproducts;
- if enzymes are immobilized on a solid support, they can be used repeatedly because their easy recovery through mechanical methods;
- enzymes can be biologically degraded, and then their management as waste is easy
- and enzymes usually present activity at room temperature and pH near the neutrality. Moreover, they catalyze reactions in aqueous medium, and allow the development, in environmental friendly conditions, of industrial processes, far away from the extreme reaction conditions that are usual in chemical synthesis.

Otherwise, enzymatic processes have some disadvantages in front of the classical chemical processes:

- in presence of organic solvents, extreme pH and/or extreme temperature, enzymes present low stability;
- some metals (even at low concentrations) may conduct to enzyme inhibition;

- presence of other enzymes may conduct to side reactions, reducing the global process yield
- and economically, development needs, usually, a long time to be implemented in industrial processes. This long time conducts to the dismissal of the enzyme process in front of the classical chemical production.

1.2.1 Asymmetric synthesis. Aldolases

Aldolases (*EC 4.1.2.* ; *EC 4.1.3.*) are enzymes of liases group that catalyze the aldol addition, the formation of C-C bound. This C-C formation is made through the nucleophilic attack of a ketone (or donor compound) to an aldehyde (or acceptor compound) with high enantioselectivities (Sugiyama et al., 2007). Although the main of the aldolases present specificity for the donor, they are able to use other acceptor (chains between 1 and 6 carbons) (Samland & Sprenger, 2006). As a consequence, the new products have two (or more) new chiral centers (Palomo, Oiarbide, & García, 2002), as shown in Figure 1.

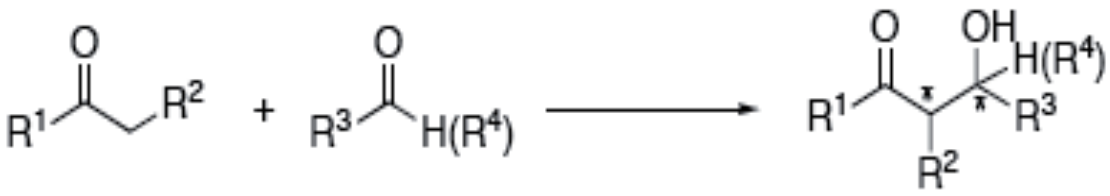


Figure 1.1. Typical aldol addition reaction. * are the chiral centers (Palomo, Oiarbide, & García, 2002)

In organic chemistry, the formation and degradation of C-C bounds is one of the key reactions, as well as one of the main processes in biological organisms. Long time and resources are needed for the chemical synthesis of products with determined stereochemistry, due to the different processes of protection and deprotection of the functional groups, in order to obtain enantiomerically pure compounds.

One of the main advantages of the aldolases as catalysts is, precisely, their ability in the synthesis of products with a specific stereochemistry. It is interesting their use in

the obtention of non-conventional sugars as statins (for the cholesterol treatment), epothilones (for the inhibition of the division of cancer cells), sialic acids (cellular signal) and amino polyols, precursors of iminocyclitols. Iminocyclitols have an important role in the inhibition of enzymes as glycosidases and glycolsiltransferases (Andersen et al., 2000; Liang et al., 2006; Saotome et al., 2000; Wrightet al., 2013). Its inhibition can represent interesting effects on the carbon hydrates catabolism and on the recognition cell-cell and cell-virus, and could be applied in therapeutic treatments as antitumor or anti-infective agents (Ichikawa et al., 2004).

Up to more than 30 aldolases are known (Machajewski & Wong, 2000).There are two different classifications for the aldolases: in function of their catalytic mechanism and in function of the donor compound (Fessner, 1998; Samland & Sprenger, 2006; Takayama, McGarvey, & Wong, 1997; Wymer & Toone, 2000).

Depending on the donor compound, they can use DHAP (dihydroxyacetone phosphate), pyruvate, phosphoenolpyruvate, acetaldehyde and glycine, as shown in Figure 1.2.

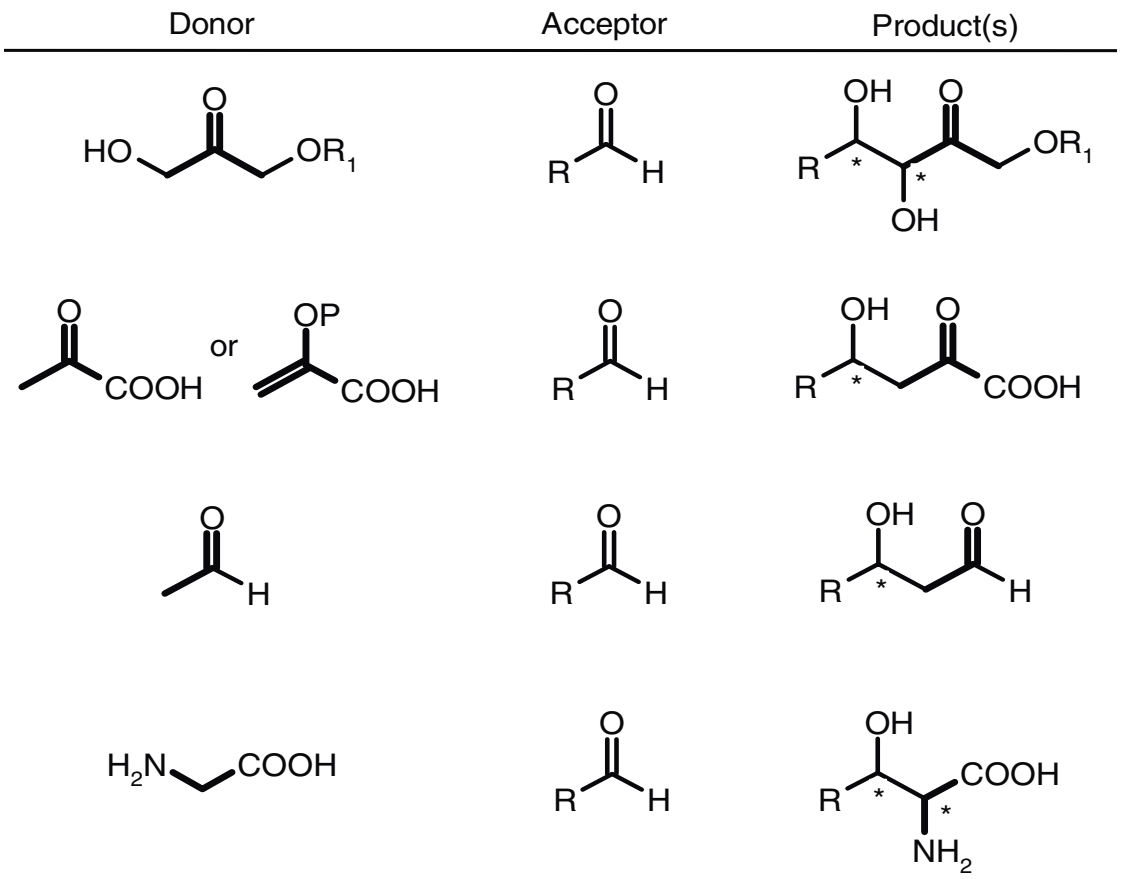


Figure 1.2. overview of the classification of aldolases by their specific donor substrate. (Samland & Sprenger, 2006)

Taking into account the reaction mechanism, the can be classified into 2 different classes:

- *class I*: present in plants and superior animals. Allow the deprotonation of the substrate forming a Schiff base between the substrate and one Lysine residue, located in the active center (Lorentzen et al., 2005; Takayama et al., 1997). Figure 1.3 shows the Schiff base formation mechanism.

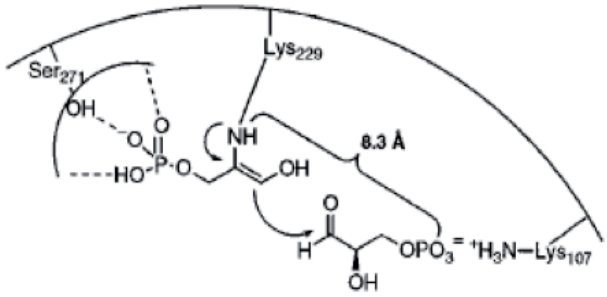


Figure 1.3. Schiff base formation mechanism for class I aldolases

- *class II*: present in bacteria and fungi. They require as cofactor a divalent cation (Zn^{2+}) to act as Lewis acid for the deprotonation of the substrate (Takayama et al., 1997). Figure 1.4 shows the activation mechanism for the donor substrate.

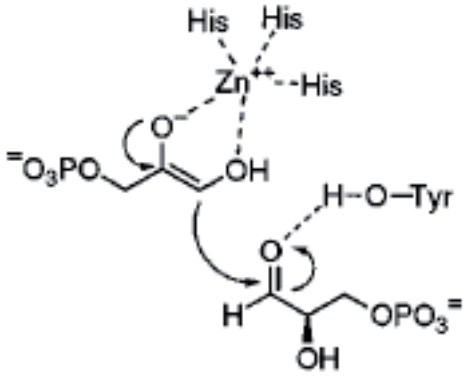


Figure 1.4. Donor substrate activation mechanism for class II aldolases

1.2.2 DHAP dependent aldolases

There are four different enzymes of the family of the DHAP dependent aldolases. They catalyze in vivo –with stereospecificity- the four possible diastereoisomers of the two chiral centers generated in the reaction (Fessner et al., 1991). They present specificity for DHAP as donor substrate, with a generation of complementary stereochemistry in C3 and C4 carbons, as shown in Figure 1.5.

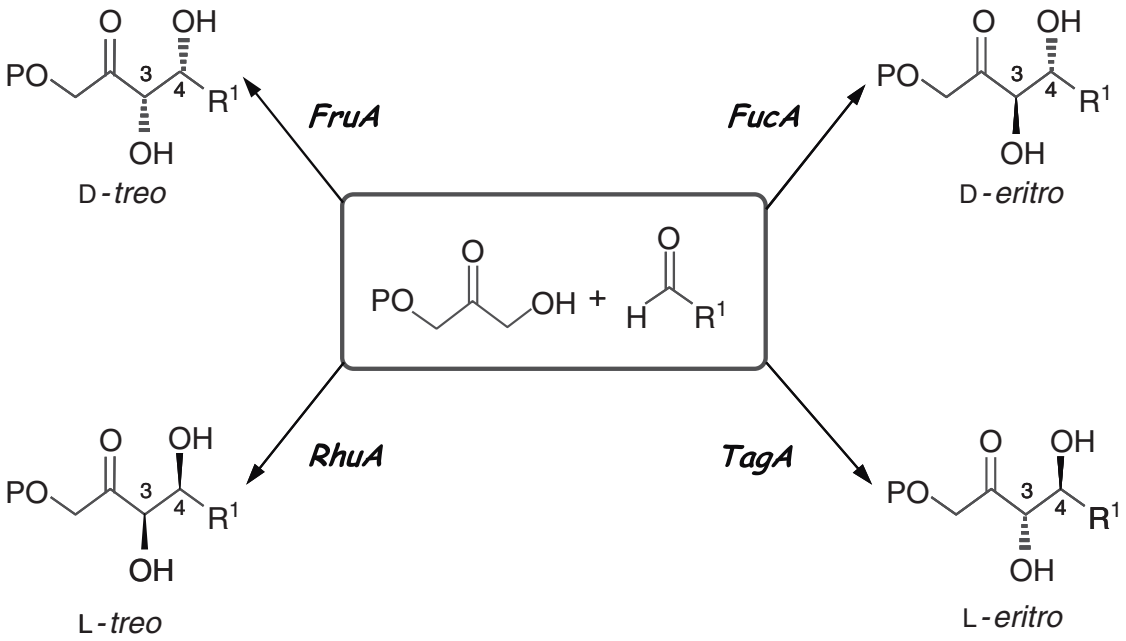


Figure 1.5. DHAP dependent aldolases, with the corresponding catalysis products.

These aldolases (Fructose-1,6-biphosphate aldolase (FruA); Tagatose-1,6-diphosphate aldolase (TagA); Fuculose-1-phosphate aldolase (FucA) and Rhamnulose-1-phosphate aldolase (RhuA)) have high affinity for the glyceraldehyde-3-phosphate and L-lactaldehyde, but they can use a great amount of other acceptor substrates (as aliphatic aldehydes) and this fact make the aldolases important tools in the synthesis of monosaccharaides and derivates. There are only two Fructose-6-Phosphate aldolase isoenzymes reported to be able to use dihydroxyacetone (DHA) as donor substrate (Schurmann & Sprenger, 2001).

TagA and FruA (from rabbit muscle) belong to class I. The free amino group from a specific Lysine in the active center of the enzyme reacts with DHAP, forming a Schiff base. This Schiff base is deprotonated and then reacts with the acceptor substrate.

FucA, RhuA and FruA (from microbe) belong to class II aldolases. In this case, the cofactor Zn^{2+} polarizes the carbonyl group of DHAP, and then reacts with de donor substrate of the reaction (Schurmann & Sprenger, 2001).

There are, also, aldolases that use DHA instead of DHAP as donor substrate. DHA dependent aldolases only catalyze the reaction were the obtained product takes D-*treo* configurations.

The target main protein in this work is Rhamnulose-1-phosphate Aldolase (RhuA). As said above, RhuA is a DHAP dependent class II aldolase, catalyzing the reaction between DHAP and L-lactaldehyde.

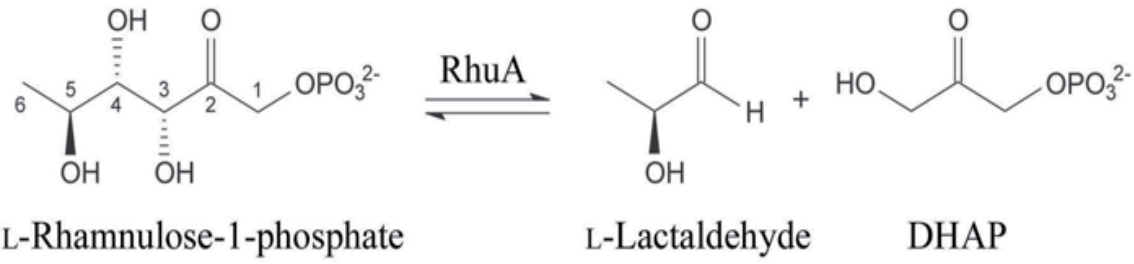


Figure 1.6. Reversible aldol addition between DHAP and L-lactaldehyde, catalyzed by RhuA.

RhuA has homotetrameric structure (Figure 1.7) with 274 amino acids every subunit and 30.15 kDa in total (Kroemer, Merkel, & Schulz, 2003; Kroemer & Schulz, 2002).

There are four active centers (one for subunit), constructed surrounding the cofactor from some residues acting as ligands, 3 hystidine residues and one glutamic acid from the same subunit, and one threonine residue from the near subunit (Kroemer & Schulz, 2002).

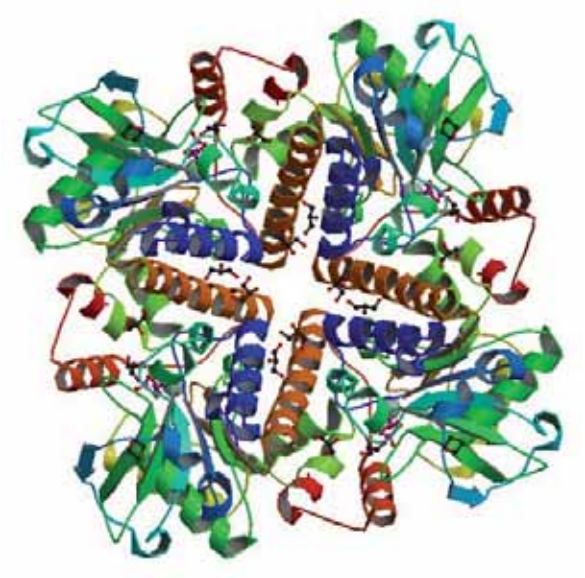


Figure 1.7. Tridimensional RhuA structure.

1.3 *Escherichia coli* as host for recombinant protein production

E.coli is a microbe living in the large intestine and in the colon of warm-blooded animals. It is a Gram-negative bacteria, mobile and facultative anaerobe. This kind of bacteria presents, usually, easy nutritional requirements, allowing the metabolization of a wide variety of carbon sources. Its duplication rate is around 1 division every half an hour (under optimal growth conditions). Generally, *E.coli* strains are not pathogenic, but some of them may conduct to diarrhea because the synthesis of the antigen K, that permit their colonization in the large intestine (Nataro & Kaper, 1998).

Genetically, *E.coli* presents, as a prokaryote, a circular DNA organization in its cytoplasm, without any separation membrane between the cytoplasm and the chromosomes. The genome was observed to contain significant number of transposable genetic elements, repeat elements, cryptic prophages and bacteriophage remnants (Blattner et al., 1997). For its use in research works and industrial applications, it is usual to employ different genotypic varieties with certain differences at genetic level in front of parent strains. The usual varieties in microbiological applications are derived from *E.coli* K12 or *E.coli* B strains. These changes with regard to parent strains are looking for the optimization of the global production yields. Some

of the genetic modifications are faced to the minimization of side products formation, which can affect the growth. Others are directed to the shut down of the action of proteases (that degrade the synthesis products).

Bacteria can use genetic elements able to replicate independent of the chromosomal DNA: plasmids. Plasmids are usually formed by a circular DNA chain and may have multiple copies into the cell.

Plasmids have genetic information that differences the cells with and without them. For example, one of the plasmid groups more extensively studied are the ones containing antibiotic resistance genes, allowing the host cell to grow in a culture medium supplemented with antibiotic. Plasmids are used as useful tools in biotechnological processes because, if they are correctly designed, they may contain the needed information for product synthesis that can represent up to 50% of the dry cell weight (Graumann & Premstaller, 2006).

The classical genetic structure of plasmids is:

- *Replicon*: is the replication origin. It controls the number of copies of the plasmid into cells (Sørensen & Mortensen, 2005)

- *Resistance marker*: it contains the necessary information to give cells antibiotic resistance (usually ampicillin, kanamycin and chloramphenicol) (Sørensen & Mortensen, 2005). However, due to the possibility of contamination of the product or biomass by antibiotics, which may be unacceptable from a medical or regulatory standpoint (Baneyx, 1999), it has been proposed to use genes as complementation of a deletion in chromosomal information (Vidal et al., 2008)

- *Regulable promoter*: needed for the control of the expression genes. One of the most commonly used systems are the *lac* operon derived, that can be induced by Isopropyl-β-D-1-thiogalactopyranoside (IPTG) (Donovan, Robinson, & Glick, 1996; Fernández-Castané, Caminal, & López-Santín, 2012; Santillán & Mackey, 2004)

- *Sequences*: transduction starting and terminators

- *Genetic information* of the product of interest

E.coli generally produces acetate as side product during its growth (even in aerobic conditions) when using glucose as carbon source. This acetic acid production inhibits the cellular growth (Eiteman & Altman, 2006; Kwon, Kim, & Kim, 1996; Sakamoto et al., 1994). Moreover, logically, its production derives the consumed carbon source from biomass and recombinant protein production (Eiteman & Altman, 2006). Acetic acid production takes place when cells are growing at high rate, when there is not oxygen enough to metabolize the totality of the carbon source (Wolfe, 2005).

Acetate production can be minimized by limiting the glucose-consumption rate, by controlling the feed rate of the cultivation, when working in in fed-batch mode. This can be achieved using a fed-batch growth strategy, where the glucose is added maintaining a fixed specific growth rate (when exponential feeding profile is used) (Pinsach, de Mas, & Lopez-Santin, 2006).

On the other hand, it can be also minimized using different carbon sources (as glycerol or fructose) with a tightly regulated assimilation and utilization (Aristidou, San, & Bennett, 1999; Zhang et al. , 1999)

Genetic engineering can be also used in acetate production reduction. The use of *E.coli* B genotypic variations can represent a reduction in acetate production if compared with *E.coli* K12 derived strains, even in glucose excess (Choi, Keum, & Lee, 2006; Noronhaet al., 2000).

E.coli is one of the most commonly used hosts for the expression of heterologous recombinant proteins, in laboratory work and also in industrial production (Cornelis, 2000). Industry about therapeutic proteins and synthesis of enzymes for the conduction of biotransformations is acquiring more importance along the years (Balzeret al., 2013). Genetic and proteomic systems have been developed allowing the use of *E.coli* for the production of recombinant proteins (using plasmids). In fact, one of the pioneer applications of recombinant DNA technology was the manipulation of *E.coli* strains for the production of human insulin. However, *E.coli* has some limitations in the production of heterologous proteins. It cannot be used in the production of

larger and more complex proteins containing multiple disulfide bonds and unpaired thiols, or in general, proteins that require post-translational modifications (Lee, 1996).

There are some advantages of using *E.coli* as host for recombinant protein production:

- rapid expression,
- high production yields,
- genetic modifications are easy,
- inexpensive culture medium preparation,
- it can be used for mass production, and it is cost effective
- and its genome is widely known.

Otherwise, there are some inconvenients in its use:

- proteins containing disulfide bonds are difficult to express,
- only for unglycosylated proteins,
- endotoxins may be produced with the protein,
- cell toxicity because acetate production
- if the protein is produced as inclusion bodies, they are inactive and need refolding.
- and, finally, *E.coli* usually does not secrete the protein produced, needing disruption processes in the downstream.

Processes producing recombinant proteins are highly dependents on the heterologous DNA stability into the host cells (Yee & Blanch, 1992). In order to assure the presence of the plasmids into the host cell, it is needed working with selective markers (e.g. antibiotic resistance). However, this maintenance of the foreign DNA conduce to cellular stress, especially when the target protein has high expression rate (Sørensen & Mortensen, 2005). The metabolic overcharge because the maintenance of plasmids is tightly related, obviously, to the number of copies present (Bailey, 1993)

Usually, during the overexpression of the target protein, inhibition in growth is observed, because the cellular metabolism is redirected to the production. Otherwise, high heterologous protein production rates can be associated to the inhibition of the

production of essential components for the cell growth, leading, also, to growth inhibitions (Hoffmann & Rinas, 2004).

These stress situations could lead in a decrease on the quantity and quality of the recombinant protein produced. Because this, a deeper study (modeling and optimization, for example) is needed in order to achieve important improvements in the global production yields.

1.4 IPTG inducible expression systems

As said before, IPTG is commonly used as inducer in systems based on the *lac* promoter.

IPTG is a non-metabolizable analogue of allolactose. As can be seen in Figure 1.8, IPTG is a 9 carbon sugar of MW = 238.3 $g \cdot mol^{-1}$.

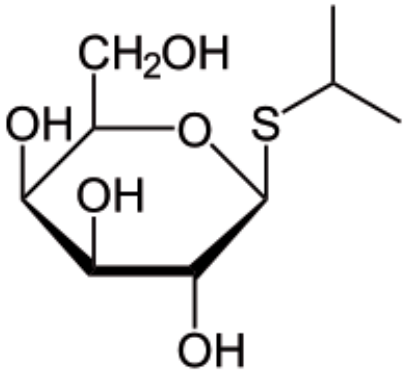


Figure 1.8. IPTG molecular structure.

There is a wide range of possible expression systems for different application. Lactose operon is one of the most commonly used and one of the most well known. *lac* operon in wild type *E.coli* consists of three genes (*lac*Z, *lac*Y and *lac*A) as depicted in Figure 1.9.

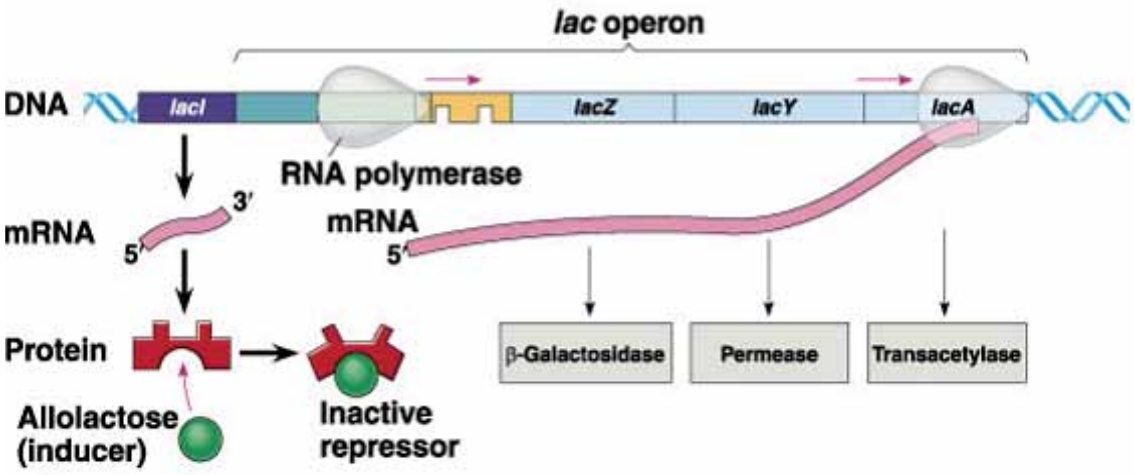


Figure 1.9. Molecular description of the *lac* operon.

The products resulting from the transcription of the three genes are:

- *lac*Z: β-galactosidase. Converts lactose to glucose and galactose.

- *lac*Y: lactose permeases. Proteins responsible for the active transport of lactose (or its analogues) from the culture medium to the cytoplasm.

- *lac*A: transacetylase or tiogalactoside transferase. May function as detoxifying agent for lactose analogs that are harmful for the cell.

Transcription from the promoter is regulated by the lac repressor (LacI, Protein in Figure 1.9), produced by *lac*I gene through its transcription (mRNA) (Horton, Lewis, & Lu, 1997; Lopezet al., 1998). Inhibition takes place in absence of lactose (or non metabolizable analogues as IPTG) because LacI does not allow the transcription from the *lac* promoter by binding the operator site. RNA polymerase is not able to bind properly the promoter, and transcription does not take place. Basal expression levels can be found because the binding between the repressor and the operator sites is a chemical equilibrium, thus it is a dynamic equilibrium. As the operator site is not continuously occupied, RNA polymerase is able to some gene transcription.

Lactose inside cells is converted to allolactose by the action of β-galactosidase. Allolactose is able to bind the repressor molecules, avoiding their linkage to the

operator sites. With no repressor bound to the operation region, a high transcription of the *lac* genes is observed. In the case of using IPTG, it is directly capable to bind the repressor, because, as said, it is a non-metabolizable analogue of allolactose.

In absence of glucose in the culture medium, cyclic molecules of adenosine monophosphate (cAMP) are produced. These molecules can produce a complex with CAP (Catabolite Activator Protein) –that increases the affinity of the promoter to RNA polymerase- increasing the transcription (Abeles, Frey, & Jencks, 1992).
Catabolite repression allows *E.coli* to metabolize glucose prior to lactose when the culture medium has a mixture of sugars. As the contrary to said above, high glucose concentrations reduce the production of cAMP meaning that, even in absence of inducer, high glucose concentration leads to low transcription levels. It has been published that glucose concentrations below 60 $g \cdot L^{-1}$ allow to achieve transcription levels high enough to have protein overexpression (Pinsach et al., 2008).

Once IPTG enters into the cell, it binds the LacI repressor molecules, and the transcription of the genes from the promoter of the *lac* operon is initiated. As IPTG is an expensive product, it is important to use the minimal inducer amount necessary to obtain the higher transcription level by optimizing the process (Hansen, Knudsen, & Sørensen, 1998).

1.5 Mathematical modeling

Mathematical modeling in chemical and biotechnological industry is used as a basic tool for the prediction and monitoring processes. Models are able to give detailed information, but with great saving in time and cost (Bailey, 1998; Brass, Hoeks, & Rohner, 1997; Thilakavathi, Basak, & Panda, 2006). Moreover, the exploitation of an accurate mathematical model can be derived in the optimization of the process and its control.
A model is a mathematical representation of a process, making a relationship between the inputs and the outputs of the study system, which objective is its description.

1.5.1 General model types

One of the firsts steps in modeling is the identification of the different variables of the process, and how they could be related (using fundamental/basic principles). Construction of the model depends on its application.

Depending the nature of the model, a distinction can be made:

- *Mechanistic model* (or theoretical or fundamental): based on the fundamental phenomena involved in the process. This kind of models is developed by applying mass, energy and momentum balances. There are some advantages using fundamental models: because the application of physico-chemical principles, it is possible to describe non-linear and dynamic processes; the model can be extrapolated, due to the application of engineering principles; the parameters of the model can be obtained bibliographically or by experimentation. Otherwise, description of these models needs a high effort, because of their complexity.

- *Empirical model*: sometimes, the process' complexity does not allow the understanding of the fundamental or basic phenomena involved. Then, the knowledge of the system is through experimental work. Empirical models are useful when, basically, a relationship between inputs and outputs of the system are going to be defined, without fundamental description of the processes involved. Examples of empirical model are: polynomial models (Lei, Ding, & Ding, 2014; Parker & Doyle, 2001); artificial neural network (Aoyama, Doyle, & Venkatasubramanian, 1995; Pendashteh et al., 2011). However, this kind of model has limited application (it can be only used into the range of determination).

- *Mixed models*: are the combinations of mechanistic and empirical models.

There are 6 different basic steps in modeling processes:

1. Problem analysis: it is needed to study the situation sufficiently to identify the problem precisely and understand its fundamental questions clearly. This stage determines the objective

2. Model formulation: abstraction of the system wanted to be modeled. The tasks in this stage are: gathering the relevant data to gain information about the system; making the assumptions reasonably; variables and units determination; establishment of relationships between variables, equations and functions determination

3. Solve the model: is the implementation of the model. It is usual to employ algebra, calculus, graphs, and computer software to solve the model. If the model is too complex and it cannot be solved, it is needed to return to step 2 and reformulate it, usually simplifying the model.

4. Model verification, interpretation and validation: once a solution has been found, it is needed examining the results, making sure their sense (verification and interpretation) and the solution must solve original new data (validation), satisfying the problem's requirements.

5. Report on the model: it is important for its utility. The report contains some steps, which parallel the steps of the modeling process.

6. Model maintenance: As the model's solution is used, it may be necessary or desirable to make corrections, improvements, or enhancements.

1.5.2 Mathematical models applied to bioprocesses

By definition, bioprocesses deal with biochemical systems, often employing growing cells. Even the simplest living cell is a system of such complexity that any mathematical description is a modest approximation (Bailey, 1998).

The use of mathematical tools to provide a rigorous, systematic and quantitative linkage between microscopic (understood as intracellular processes) phenomena and

macroscopic process performance was installed into research principles of Chemical Engineering late 60s, early 70s, mainly with the appearance of the textbook *Biochemical Engineering* by Aiba, Humphrey and Millis (Aiba, Humphrey, & Millis, 1973). Chemical Engineering principles were used in the biological sphere. This book marked a change on the tendency of empirical approaches to a new types of model, where rigorous mathematical reasoning is applied in the understanding of the processes (Bailey, 1998).

Systems involving living cells were systematically classified in order to give biochemical engineers a conception of the mathematical description and the experimental characterization, as shown in Figure 1.10 (Fredrickson, Megee, & Tsuchiya, 1970).

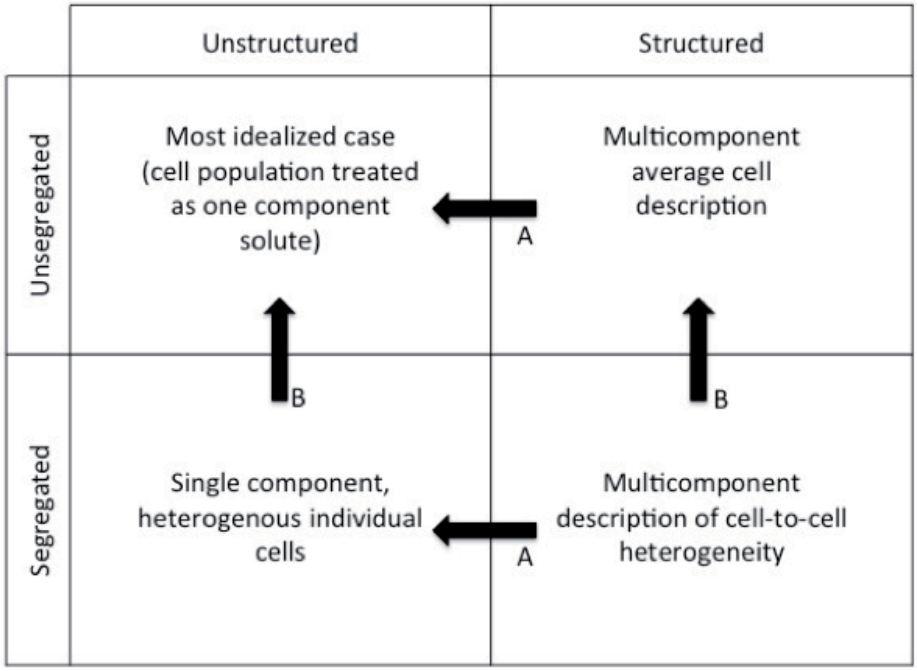


Figure 1.10. Classifications introduced by Fredrickson for mathematical (and other) representation of cell population. A: balanced growth approximation; B: "average cell" approximation.

As it can be seen in Figure 1.10, models concerning cell processes can be classified depending on the basic knowledge of the system (structured or unstructured) and depending on the assumption or not of differences between cell populations (segregated or unsegregated).

The four different types of models are going to be explained and illustrated with one example (each), applied or susceptible to be applied, to recombinant protein production.

As said, there are two different kinds of model depending on the basic knowledge of the system concerning cell processes:

- *Unstructured models* are the simplest for the modeling of a biological process. They consider the biomass as a unique variable. Growth descriptions are based on the uptake of substrates and the formation of products. Monod model is one of the basic unstructured model more widely used (Monod, 1949). Unstructured models are based on variables easily measured with standard laboratory equipment (spectrophotometers, kits for analysis, HPLC, …)

An example of unstructured model of protein production is the work made by Nadri and collaborators (Nadri et al., 2006). In this case, an *E.coli* strain was used, containing a plasmid with bioluminescence and β-galactosidase production. The model Equations for a fed-batch process are presented below (Equations (1.1) to (1.5)):

$$\frac{dX}{dt} = \mu X - \Phi_{death} X - DX \tag{1.1}$$

$$\frac{dS}{dt} = -\varphi_S \mu X - k_m X - D(S - S_{in}) \tag{1.2}$$

$$\frac{dP}{dt} = \varphi_P \Phi_P \mu X - \Phi_{death} P - DP \tag{1.3}$$

$$\frac{dI}{dt} = -DI \tag{1.4}$$

$$L = \varphi_L \frac{S}{(k_S + S)} \Phi_P X P \tag{1.5}$$

where, basically:

- X: biomass concentration
- S: substrate concentration
- P: protein concentration
- I: inducer concentration
- L: light intensity
- μ is the specific growth rate, Φ_{death} is the death kinetics, D is the dilution rate, φ_S is the substrate yield coefficient, k_m is the maintenance coefficient associate to the substrate, φ_P is the protein yield coefficient, Φ_P is the inducer effect, φ_L is the luminescence yield coefficient and k_S is the saturation constant associated to the oxygen

Qualitatively, it can be seen that biomass is considered as a unique solute component. The model is based on net biomass growth expressed through Equation (1.1), were terms on the right correspond to growth, death and dilution, respectively. Substrate and inducer uptake and products production (both protein and luminescence) are presented in Equations (1.2) to (1.5).
It is a good example of unstructured model, based on substrate uptake and product formation, using easily determining variables (glucose, protein and luminescence).

- *Structured models* include high detailed knowledge about the biochemical reactions taking place into the cells. Biomass is considered as multiple interrelated compartments (Gerlach et al., 2014; Tang, Chen, & Zhang, 2007). It is usual to divide the cell as: a) active compartment (ribosomes, mRNA, tRNA,…); b) plasmid DNA; c) codified product in the plasmid DNA; d) structural and genetic cell material. A deep knowledge about metabolic fluxes is required.

A clear example of structured modeling of protein production is the work carried out by Toksoy and collaborators (Toksoy Öner et al. , 2003). In this case, a four-compartment model is presented. It is shown in Figure 1.11, where the four compartments can be seen, as well as the outline of the biophase structure.

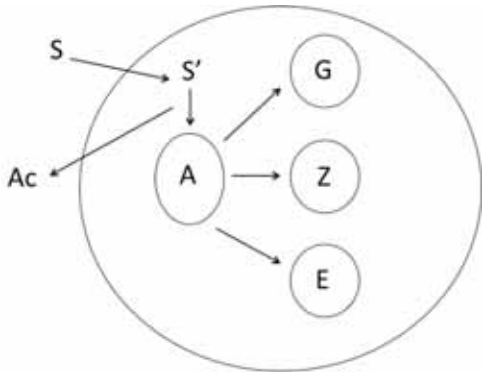


Figure 1.11. Schematic representation of the four-compartment model

In this case, A is the compartment for the active cell components (ribosomes, mRNA, tRNA); E contains only the recombinant enzyme protein; G compartment comprises the recombinant plasmids; and Z contains the structure forming cell material and chromosomal DNA.

Equations (1.6) to (1.9) show the synthesis rates of the particular cell components. Equations (1.10) to (1.16) show the mass balances.

$$R_A = k_{XA} X_A \left(\frac{S}{S+K_A}\right) \tag{1.6}$$

$$R_Z = k_{XZ} X_A \left(\frac{S}{S+K_Z}\right) \tag{1.7}$$

$$R_G = k_{XG} X_A X_G \left(\frac{S}{S+K_Z}\right) \tag{1.8}$$

$$R_E = k_{XE} X_A \left(\frac{S}{S+K_Z}\right)\left(\frac{X_G}{X_G+K_E}\right)\left(\frac{C_1}{C_1+K_1}\right) \tag{1.9}$$

$$\frac{dX_A}{dt} = \gamma_A R_A - R_Z - R_G - R_E - \mu X_A \tag{1.10}$$

$$\frac{dX_Z}{dt} = \gamma_Z R_Z - \mu X_Z \tag{1.11}$$

$$\frac{dX_G}{dt} = \gamma_Z R_G - \mu X_G \tag{1.12}$$

$$\frac{dX_E}{dt} = \gamma_Z R_E - \mu X_E \quad (1.13)$$

$$\frac{dx}{dt} = (1 - \phi)\mu x \quad (1.14)$$

$$\frac{dS'}{dt} = -R_A x \quad (1.15)$$

$$\frac{dAc}{dt} = R_A \eta \quad (1.16)$$

Where X_i is the concentration of the components in every compartment, S is the substrate concentration, S' is the substrate concentration inside the cell, and Ac is the concentration of the byproduct (acetic acid), μ is the specific growth rate, k_{Xi} are kinetic parameters, K_i are also kinetic parameters, ϕ is the segregation coefficient of the production plasmid (calculated using a probability of a plasmid-free and a plasmid-carrying daughter cell from a plasmid-carrying cell), γ_i are the yield coefficients for macromolecular synthesis and η is the acetic acid production yield.
Qualitatively, it can be seen that a high knowledge about the system and the relationships between the compartments are needed to propose a structured model describing protein production.

Otherwise, models can be classified depending if the effect of the overexpression on the cellular population is assumed or not:

- *Unsegregated models* do not distinguish between different types of cells in the culture. In protein production these models do not take into account differences in population because the production of heterologous protein (Kavanagh & Barton, 2008; Lee & Ramirez, 1992a; Mahadevan & Doyle, 2003).

One of the examples above correspond to an unstructured unsegregated (Nadri et al., 2006) model.

- *Segregated models* for protein production consider that the metabolic burden due to the production of the recombinant protein makes the cell population to be in different growth steps. It can make, also, a difference between plasmid carrying and non-carrying cells. (Andersson, Strandberg, & Enfors, 1996; Fan, Tuncay, & Ortoleva, 2007; Lopez-Vernaza & Leach, 2013; Zheng, Yao, & Lin, 2005).

In this case, a structured unsegregated (Toksoy Öner et al., 2003) model has been explained above.

Segregated models provide highly detailed information about cellular physiological state, but they require a large amount of data. This data is not easily available, experimentally nor bibliographically.

There is another type of models, based on statistical tools: Response Surface Methodology (RSM). They explore the relationship between several explanatory variables and one or more response variables. The main idea is to use a sequence of designed experiments to obtain an optimal response. As all the statistical models, this model is an approximation to reality.

Because the goal of the modeling, at the level that is going to be performed in this work, is the description, prediction and the possible optimization of the process, using its macroscopic properties for control and automation applications, an unstructured and unsegregated modeling seems to be the best option.

2 Objectives

This work has been developed into the Bioprocess Engineering and Applied Biocatalysis group, which main objective is researching on recombinant protein production, and its application to biocatalytic processes. Specifically, this work is focused on the modeling of the whole process, in order to obtain all the differential equations needed to describe and predict high cell density cultures of *E.coli* producing recombinant proteins.

The model will be calibrated using an M15-derived strain producing an aldolase (RhuA). This overall model is going to be found through the achievement of different sub-objectives:

- Qualitative and quantitative studies about how the different variables affect to protein production and its activity

- Quantitative IPTG uptake model development. This intermediate model will be capable to predict the inducer uptake from the culture medium to the intercellular space, and will predict, also, the inner concentration of IPTG for every time during the induction phase

- Development of the specific protein production (in mass and activity units) model, able to predict the RhuA production along time. This model needs the prediction obtained through the IPTG uptake model

- Coupling of three different models: biomass growth, IPTG uptake and protein production models.

- Extension of the specific protein production model to different strains and proteins, showing how the model is capable to predict and simulate other expression systems, and showing which model' parameters are needed to change to fit the model.

3 Materials and Methods

This chapter describes general procedures used in this work. Specific methodologies will be described in detail in its own chapter.

3.1 Strains and expression systems

Some different *E.coli* strains and expression systems have been employed in the overexpression of different recombinant proteins (Rhamnulose-1-Phosphate Aldolase, Fuculose-1-Phosphate Aldolase, Fructose-6-Phosphate Aldolase and ω-Transaminase). Details about Fructose-6-Phosphate aldolase and ω-Transaminase will be explained in Chapter 7.

Rhamnulose-1-Phosphate Aldolase (RhuA) was expressed as a fusion protein to a 6xhistidine tag at its N-terminal end, using the K-12 *E.coli* derived strain M15 Δ*gly*A[pQEαβrham][pREP4], under the control of the strong promoter T5 (Vidal et al., 2003; Vidal et al., 2008). The expression system was derived from the commercial *E.coli* M15[pREP4] pQE40 (Qiagen).

This strain presents a chromosomal deletion in *gly*A gene, responsible of the synthesis of Serine Hydroxymethyltransferase (SHMT) that has a key role in glycine biosynthetic route. *gly*A gene, for auxotrophy complementation, was inserted in the pQErham plasmid under the control of constitutive promoter P3, obtaining the plasmid pQEαβrham, as shown in Figure 3.1.

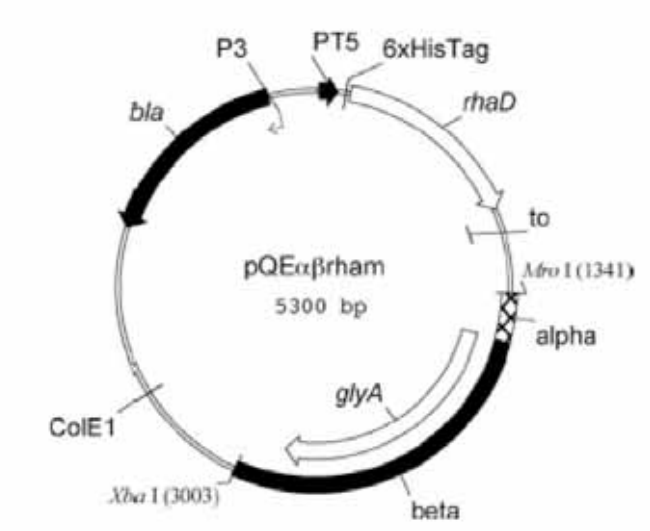


Figure 3.1. pQEαβrham plasmid

SHMT expression is codified in *gly*A gene. RhuA is codified by *rhaD* gene. Moreover, the sequence that codifies for the 6xhistidine tags (6xHisTag) that merged with N-terminus of RhuA allows its separation from all other cytoplasm proteins using Immobilized Metal Affinity Chromatography (IMAC), is also present in the plasmid.

bla gene codifies for β-lactamase, giving the strain antibiotic resistance (ampicillin). This resistance to antibiotics is useful in the first phase of growing (in complex medium, that may contain glycine), allowing keeping the plasmid.

Moreover, pREP4 plasmid is also present into the cells (Figure 3.2). This plasmid provides kanamycin (*neo* gene) resistance and expresses the repressor protein (LacI repressor), encoded by the *lacI* gene. There are multiple copies (12) (Pinsach, 2009) of pREP4 plasmid into the cells, in order to produce enough quantity of repressor molecules to regulate recombinant protein production.

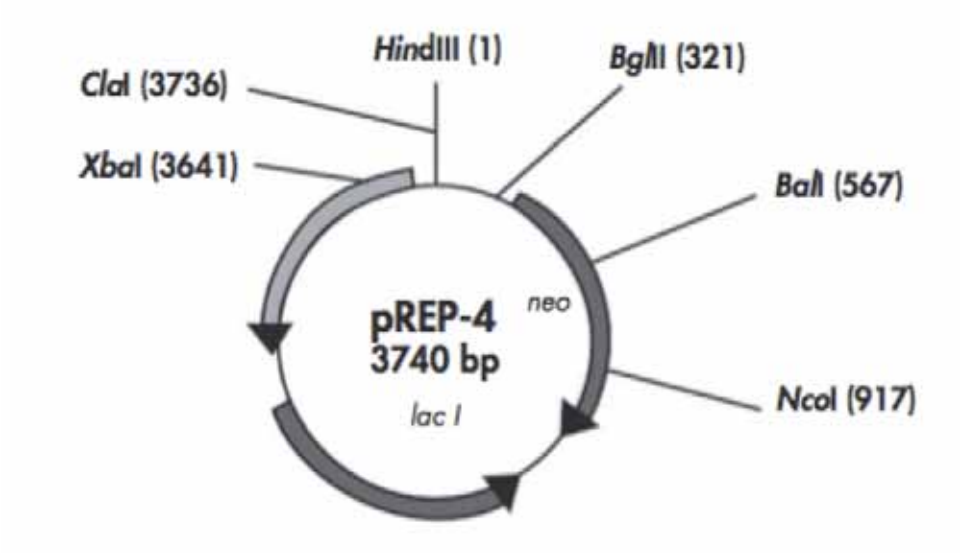


Figure 3.2. pREP4 vector

A new strain (*E.coli* M15 Δ*gly*AΔ*lac*Y [pQEαβrham][pREP4]) was constructed as described in literature (Fernández-Castané et al., 2012). This strain has a deletion in *lac*Y gene, which codifies for the lactose specific active transport proteins (lac permeases).

Fuculose-1-Phosphate Aldolase (FucA) producing strains were: *E.coli* M15 Δ*gly*A[pQEαβfucA][pREP4] (which has the same construction as *E.coli* M15 Δ*gly*A[pQEαβrham][pREP4], but containing the gene for FucA production) and *E.coli* M15 [pQE-fucA][pREP4] (strain without deletion in chromosomic *glyA* gene and without the αβ termination sequence).

Fructose-6-Phosphate Aldolase (FSA) is produced by *E.coli* BL21(DE3) strain (from now BL21 (DE3) FSA), as described in literature (Sánchez-Moreno et al., 2012). *mipB* gene, encoding FSA was inserted into a modified expression vector pET22b(+) (Kreimeyer et al., 2007). This strain was supplied by Prof. Dr. Pere Clapés from IQAC-CSIC, Barcelona.

ω-Transaminase (ATA) is produced using *E.coli* BL21 (DE3) strain (from now BL21 (DE3) ATA), harboring pLE1-A10-AcATA-D4 plasmid (supplied by c-LEcta) (Casablancas et al., 2013).

Strains for the production of FSA and ATA are described in detail in Chapter 7.

3.2 Culture medium and fermentation conditions

The working sequence for all the experiments was always the same. A detailed explanation is going to be presented in sections 3.2.1 to 3.2.3, but a general picture of it is described below.

Firstly, a cryobille from the stock stored at -80°C was made grow in rich medium. Once the strain has grown, it was placed in a shake flask containing defined medium, until it reached the exponential phase of the growth. At that point, the bioreactor was inoculated. Once the glucose of the batch is completely consumed, the fed-batch phase begins, through a predefined addition of nutrients. When the culture reached the desired biomass concentration, a pulse of inducer (IPTG) was added to the culture, making possible the overexpression of recombinant protein.

3.2.1 Pre-inoculum

Lysogeny Broth (LB) -containing 10g·L^{-1} of peptone, 5g·L^{-1} of yeast extract and 10g·L^{-1} of NaCl- was used for all pre-inoculum growth. It was sterilized by autoclaving (120°C, 30 min). When necessary, it was supplemented with antibiotic, as shown in Table 3.1.

Table 3.1. Antibiotic supplementation in pre-inoculum cultures

	M15 Δ*gly*A [pQEαβrham] [pREP4]	**M15 Δ*gly*AΔ*lac*Y [pQEαβrham] [pREP4]**	**M15 Δ*gly*A [pQEαβfucA] [pREP4]**	**M15 [pQE-fucA] [pREP4]**	**BL21 (DE3) FSA**	**BL21 (DE3) ATA**
Ampicillin (g·L^{-1})	0.1	0.1	0.1	0.1	0.1	--
Kanamycin (g·L^{-1})	0.1	0.1	0.1	0.1	--	0.05

From cryostock (-80 °C) in commercial Cryobilles (AES Chemunex), strains were grown in Falcon tubes containing 10 mL of LB medium and the supplementary antibiotic. Growths were performed at 37 °C and 150 rpm, reaching 1.8 to 2.1 units of optical density (OD_{600}) after around 16 hours of incubation.

3.2.2 Inoculum

A Defined minimum Medium (DM) using glucose as a sole carbon source (Durany et al., 2004) was employed for inoculum growth. Table 3.2 shows the DM composition used for the different strains. It can be seen that the composition is mainly the same, with little differences. All the different components in Table 3.2 were prepared separately, save K_2HPO_4, KH_2PO_4, NaCl and $(NH_4)_2SO_4$, that were prepared as a unique solution, called macroelements. Sterilization of glucose and macroelements were made by autoclaving (120°C, 30 min), while all the other components were sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim). Concentrations of the solutions prepared were: 600 gL^{-1} glucose, 17.91 gL^{-1} K_2HPO_4, 3.59 gL^{-1} KH_2PO_4, 4.52 gL^{-1} $(NH_4)_2SO_4$, 2.76 gL^{-1} NaCl, 500 gL^{-1} $MgSO_4 \cdot 7H_2O$, 5 gL^{-1} $FeCl_3$, 100 gL^{-1} $CaCl2 \cdot H_2O$. Trace elements shown in Table 3.3 were prepared as a unique solution, and sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim). The necessary quantity of medium is prepared by mixing the corresponding volume from the stock solutions.

Table 3.2. DM composition for inoculum growth per liter of medium. A: *E.coli* M15 strains; B: *E.coli* BL21 (DE3) FSA strain; C: *E.coli* BL21 (DE3) ATA strain.

Component ($g \cdot L^{-1}$)	A	B	C
Glucose	5	5	20
K_2HPO_4	2.973	2.973	11.9
KH_2PO_4	0.596	0.596	2.4
NaCl	0.458	0.458	1.8
$(NH_4)_2SO_4$	0.75	0.75	3
$MgSO_4 \cdot 7H_2O$	0.112	0.025	0.1
$FeCl_3$	0.006	0.0025	0.01
$CaCl_2 \cdot 2H_2O$	0.001	--	--
Trace elements solution (mL/L)	0.8	0.44	0.8
Thiamine	0.025	--	--
Ampicillin	--	0.0125	--
Kanamycin	--	--	0.05

Table 3.3. Trace elements solution concentration by liter of trace element.

Component	Concentration ($g \cdot L^{-1}$)
$AlCl_3 \cdot 6H_2O$	0.04
$ZnSO_4 \cdot 7H_2O$	1.74
$CoCl_2 \cdot 6H_2O$	0.16
$CuSO4 \cdot H_2O$	1.55
H_3BO_3	0.01
$MnCl_2 \cdot 4H_2O$	1.42
$NiCl_2 \cdot 6H_2O$	0.01
Na_2MoO_4	0.02

5 mL of pre-inoculum were transferred to shake flasks, containing 100 mL of DM described in Tables 3.2 and 3.3, with same cultivation conditions as pre-inoculum cultures, until reaching 1.2 OD_{600}, that ensures the culture to be in the exponential growth phase.

3.2.3 Bioreactor. Fed-batch operation

Defined minimal Medium was used in bioreactor growth. After batch phase, feeding medium was necessary to achieve high cell densities. DM composition for batch phase corresponds to Table 3.4. Feeding composition is shown in Table 3.5. All the components were prepared as explained in Section 3.2.2.

In order to avoid co-precipitation with magnesium salts, 5 mL of phosphates solution (500 gL^{-1} K_2HPO_4 and 100 gL^{-1} KH_2PO_4) were added manually to the culture every 30 OD_{600} increment. The solution was autoclaved (120°C, 30 min) for its sterilization.

Stock solution of kanamycin was prepared with a concentration of 100 $mg \cdot mL^{-1}$ and stored at -20°C. Ampicillin (100 $mg \cdot mL^{-1}$) 50% (v/v) ethanol-water stock was prepared and stored at -20°C. IPTG stock was prepared at 100 mM, and stored at -20°C. Antifoam (Antifoam 204, Sigma) was added to bioreactor whenever foam formation was observed. Kanamycin, ampicillin and IPTG stocks were sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim). Antifoam was autoclaved (120°C, 30 min).

Table 3.4. DM composition for bioreactor batch growth per liter of medium. A: *E.coli* M15 strains; B: *E.coli* BL21 (DE3) FSA strain; C: *E.coli* BL21 (DE3) ATA strain.

Component ($g \cdot L^{-1}$)	A	B	C
Glucose	20	20	20
K_2HPO_4	11.9	11.9	11.9
KH_2PO_4	2.4	2.4	2.4
NaCl	1.8	1.8	1.8
$(NH_4)_2SO_4$	3	3	3
$MgSO_4 \cdot 7H_2O$	0.45	0.1	0.1
$FeCl_3$	0.02	0.01	0.01
$CaCl_2 \cdot 2H_2O$	1.44	--	--
Trace elements solution (mL/L)	2.88	0.90	0.8
Thiamine	0.1	--	--
Ampicillin	--	0.05	--
Kanamycin	--	--	0.05

Table 3.5. Feeding solution composition per liter of feeding medium. A: *E.coli* M15 strains; B: *E.coli* BL21 (DE3) FSA strain; C: *E.coli* BL21 (DE3) ATA strain.

Component ($g \cdot L^{-1}$)	A	B	C
Glucose	478	478	478
$MgSO_4 \cdot 7H_2O$	9.56	9.56	9.56
$FeCl_3$	0.49	0.49	0.49
$CaCl_2 \cdot 2H_2O$	0.1	0.1	0.1
Trace element solution (mL/L)	63	63	63
Thiamine	0.33	--	--
Ampicillin	--	0.05	--
Kanamycin	--	--	0.005

For the batch phase, 80 mL of inoculum were added to 720 mL of DM with composition shown in Table 3.4. Bacterial growth was carried out in a Biostat® B bioreactor (Braun Biotech Int.) with 2L jar for *E.coli* M15 strains (BL21 (DE3) strain fermentations were performed in a different bioreactor, as explained specifically in

Chapter 7). The bioreactor is equipped with stirring, temperature, dissolved oxygen and pH controllers. Temperature was maintained at 37 °C and pH was kept at 7.00±0.05 by adding 15% (v/v) NH_4OH solution. Oxygen saturation (pO_2) value was maintained at 60% saturation by adapting the stirring speed between 350 and 1120 rpm and supplying 1.5 $L \cdot h^{-1}$ of air (enriched with pure oxygen when necessary). A reduction in oxygen consumption and an increase in pH can be used as identification for the batch phase ending. For the glucose limited feeding, a microburette (CRISON, microBU 2030; 2.5 mL syringe, Hamilton) was used (Ruiz et al., 2009; Vidal et al., 2005). Feeding was performed in order to maintain constant the specific growth rate (μ_{fix}) at a fixed value. Predefined exponential feeding profile based on mass balances and substrate uptake was used for it (Pinsach et al., 2006). When culture reached the desired biomass concentration, a pulse of IPTG was added, in order to have the desired inducer concentration into the reactor

As reported in literature (Pinsach et al., 2006), specific growth rate can be modulated (predefined exponential addition -open loop-) using a mathematical model based on mass balances. The volume of feeding solution with S_f concentration of glucose added to the culture medium in order to maintain the μ_{fix} in a time interval can be calculated using the Equation (3.1)

$$V_{ad}(t) = \frac{1}{S_f}\left(\frac{m_{SX}}{\mu_{fix}} + \frac{1}{Y_{XS}}\right) \cdot X(t) \cdot V(t) \cdot \left(exp\left(\mu_{fix} \cdot \Delta t\right) - 1\right) \quad (3.1)$$

where:

- S_f is the glucose concentration ($g \cdot L^{-1}$) in the feeding medium
- m_{SX} is the cellular maintenance coefficient
- μ_{fix} is the fixed specific growth rate (h^{-1}) wanted to maintain during Δt
- Y_{XS} is the substrate yield
- t the moment the feeding is started (h)
- Δt is time interval

The scheme of the experimental setup is shown in Figure 3.3.

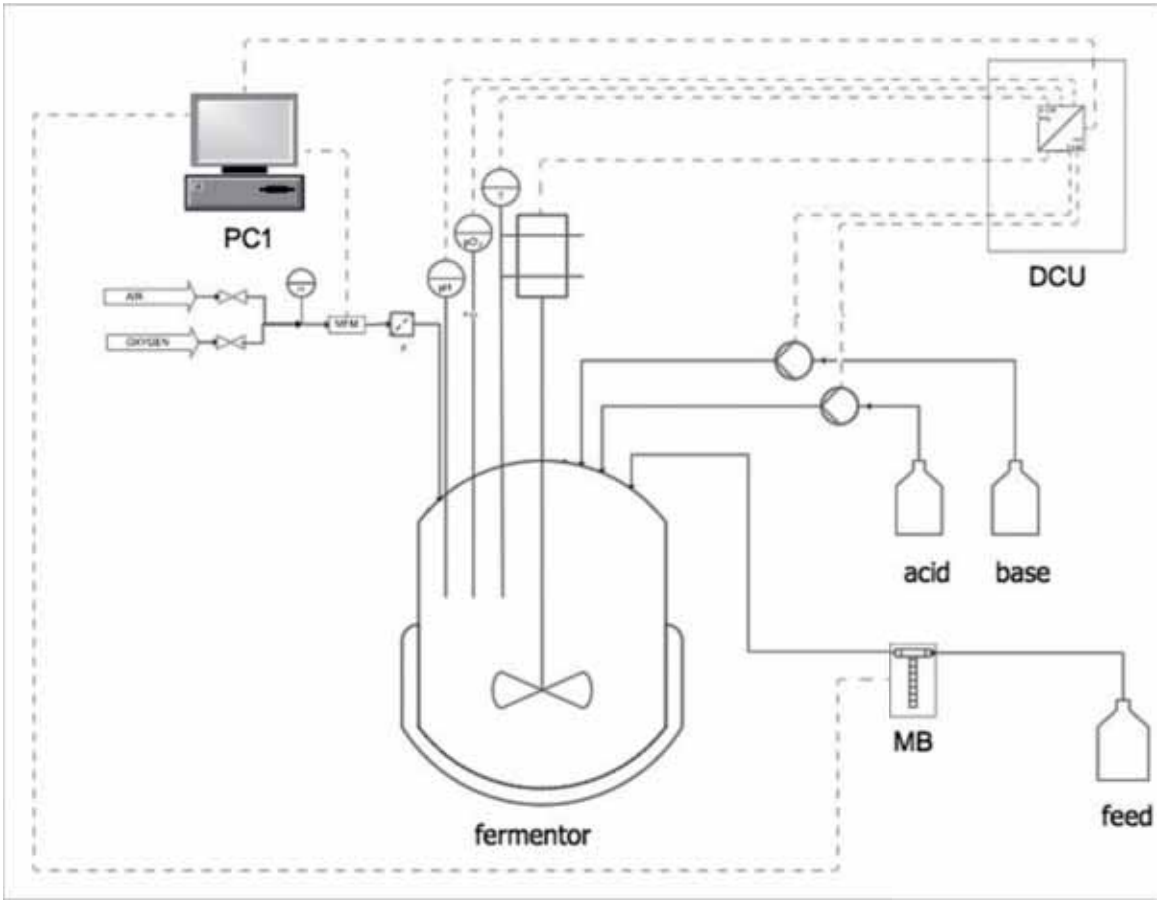


Figure 3.3. Scheme of the experimental bioreactor setup

3.3 Analytical methods

3.3.1 Biomass

Cellular concentration was determined by optical density measurements at 600 nm of wavelength (OD_{600}), using a spectrophotometer (Uvicon 941 Plus, Kontrol). OD_{600} values were converted to biomass concentration expressed as Dry Cell Weigh (DCW), with 1 OD_{600} equivalent to 0.3 $gDCW \cdot L^{-1}$ (Pinsach et al., 2008). This factor has been assumed to be the same for all the strains, or, in any case, the difference assumed to be lower than the error associated to measurements.

In order to use the Lambert-Beer Law, that makes possible a relationship between the absorbance and the quantity of mass, it is necessary to be in the linear region of

absorbance. This linear region is between 0.1 and 0.8 units of absorbance. Any sample above 0.8 cannot be related to its real biomass concentration, and it is necessary to make a dilution. Dilution was made with water, when necessary, to have a value of absorbance in the linear range.

3.3.2 Glucose

Glucose concentration in culture supernatant was determined enzymatically using an YSI 2700 (Yellow Spring Instruments) after separation from biomass. Supernatant of 1.5 mL of sample was separated by centrifugation (13400 rpm, 5 min) and filtration (0.45 µm).

3.3.3 Inducer

IPTG analysis was performed by Liquid Chromatography (Shimadzu) with an UV/Vis detector, that was operating at 210 nm of wavelength, coupled to a Mass Spectrometer equipped with ESI (Electro Spray Ionization) (Shimadzu) interface and a single quadrupole. The delivery system was a LC-10AD.

Medium samples (1.5 mL) from the bioreactor were centrifuged (13400 rpm, 5 min) and the supernatant was filtered (0.45 µm), and diluted (with milliQ water) if necessary in order to have a concentration within the measurement range (between 1 µM and 10 µM).

Mass spectrometry and liquid chromatography conditions are described in literature (Fernández et al., 2010). The mobile phase was prepared from milliQ water and formic acid at pH=3, operating in isocratic mode. It was necessary one hour for the equilibration of the column (300 x 7.8 mm ICSep ICE-COREGEL 87H3, Transgenomic®) before the first injection. Flow rate was 0.6 $mL \cdot min^{-1}$ and 10 µL of sample (or standard) were directly injected on the column.

The flux was diverted using a valve (FCV-20H2, Shimadzu) in order to allow only inducer peak to reach the spectrometer. A Shimadzu SIL-10AD automatic injector was used, and data analysis was processed using Lab Solutions 3.04 software.

Both column and autoinjector were kept at room temperature. Retention time for IPTG was 12.4 min, then the flow was directed to mass spectrometer 11 min after injection, and diverted after 14.5 min, with 30 min as a total run time.

ESI source and positive/negative ionization mode were used. Nebulizing and drying were made with nitrogen. The conditions for the MS data acquisition with the ESI mode were: 1,6 kV probe voltage; 1.5 $L \cdot min^{-1}$ nebulizing gas; 250 °C in the CDL (Curved Desolvation Line); 200 °C in the heat block; and the acquisition mode was by SCAN, between 100 and 400 *m/z*.

3.3.4 Protein

3.3.4.1 Sample preparation

Samples from bioreactor fermentation were diluted with milliQ water to an OD_{600} of 4 (for comparison purposes) (with a final volume of 1mL). Pellet resulting after centrifugation (13400 rpm, 5 min) was resuspended with 1mL of lysis buffer (100 mM Tris-HCl; pH=7.5) –maintaining an OD_{600} of 4. Cell suspension in lysis buffer was kept in ice, and sonicated in order to disrupt the cells. Sonication consisted in 4 cycles of 15 s pulses at 50 W (with 2 min between pulses), using a VC50 (Vibracell®, Sonics & Material) with microtip probe. Cellular debris was separated from supernatant by centrifugation (14000 rpm, 10 min, 4°C). Supernatant contained the intracellular protein produced in the induction phase of the culture growth.

3.3.4.2 Total protein

Total protein concentration was determined by Bradford Method. This method uses Coomassie® Protein Assay Reagent Kit (Thermo Scientific, US) and Bovine Serum Albumin (BSA) as standard. To determine the absorbance of the different samples, Microplate system (Microtiter Plate Flat, SUDELAB 900011) was used. 200 µL of

Coomassie reagent were placed in the Microplate. 7 µL of sample or standard were added to the 200 µL of Coomassie reagent. After 15 minutes of reaction, the absorbance was read with ThermoScientific® Multiskan FC equipment.

3.3.4.3 Specific protein

Once total protein is quantified, it is needed to specify the amount that corresponds to the target protein. The determination of the percentage of target protein was determined using NuPAGE® 12% Bis-Tris gels using MES-SDS as running buffer, following the manufacturer's manual (Invitrogen, US). Samples were prepared with 5 µL of sample buffer 4X; 3 µL milliQ water; 2 µL reducing agent and 10 µL of the protein sample. 10 minutes of incubation in the dry bath were necessary to denaturalize the sample (70°C, 300 rpm). One of the wells of the gels was charged with 5 µL of protein standard marker, and the others with 15 µL of sample.

After 40 minutes of running at 200 V, gels were rinsed with distilled water, and covered with gel fixation solution (40% methanol, 10% acetic acid in water) for 1 hour. After fixing gels, they were rinsed with distilled water and covered with Bio-Safe™ Coomassie (Bio-Rad®) in order to dye the protein bands. After 1 hour dying, gels were rinsed again with distilled water, and covered with water for 1 or 2 hours.

Percentage of the protein was calculated by densitometry using the specific software Image Lab© (Bio-Rad Laboratories).

3.3.4.4 Specific activity

- RhuA

RhuA activity can be determined through the initial NADH disappearance rate in the aldolase reaction, which is coupled to a second reaction catalyzed by α-Glycerophosphate Dehydrogenase (α-GDH) (Vidal et al., 2003), as can be seen in Reactions (a) and (b).

$$Substrate \xleftrightarrow{aldolase} DHAP + L - lactaldehyde \quad \text{(a)}$$

$$DHAP + NADH \xrightarrow{\alpha-GDH} \alpha - glycerophosphate + NAD^{+} \quad \text{(b)}$$

Reaction (a) is the natural aldolase reaction, where the substrate (rhamnulose-1-phosphate) is transformed to L-lactaldehyde and DHAP. Rhamnulose-1-phosphate is prepared in the laboratory (Vidal et al., 2003). In presence of NADH, DHAP is rapidly reduced to α-glycerophosphate through α-GDH enzyme, while NADH is oxidized to NAD^+ as shown in Reaction (b).

The evolution of NADH disappearance was followed by spectrophotometry at 340 nm wavelength and 25 °C, using a spectrometer UV-VIS Cary (VARIAN). Activity of the aldolase is proportional to NADH disappearance rate. One unit of activity is defined as the amount of aldolase capable to convert 1 µmol of substrate in DHAP and L-lactaldehyde per minute.

Table 3.6 shows the concentration of every reagent, as well as of the activity buffer, for Reactions (a) and (b) depending on the aldolase used.

Table 3.6. Reagents concentration for the activity assays

Reagent	Concentration
NADH	0.15 mM
Substrate	2 mM
α-GDH	2.5 $U \cdot mL^{-1}$
TRIS-KCl	50-100 mM

All reagents were added into a 1mL quartz cuvette, to a final volume of 980 µL. Sample volume, in this case, was 20 µL.

The calculation of the volumetric activity is described by Equation (3.2).

$$AU \cdot mL^{-1} = \frac{\Delta Abs_{340}}{\varepsilon} \frac{V_t}{V_e} D_f L_p \quad (3.2)$$

where:

- ΔAbs_{340} is the absorbance variation per minute at 340 nm wavelength (min^{-1})
- ε is the NADH molar extinction coefficient (6.22 $mM^{-1} \cdot cm^{-1}$)
- V_t is the total assay volume (mL)
- V_e is the volume of the aldolase sample (mL)

- D_f is the dilution factor of the sample (if necessary, when the absorbance read is over the limits of detection of the hardware)
- L_p is the optical path of the cuvette (cm)

Activity assay protocols for FSA and ATA are described in their specific chapter (Chapter 7).

3.4 Modeling and parameter estimation

In order to simulate and estimate the parameters of the model, as well as to perform the numerical integration of the different differential equations, commercial specific softwares were employed. Specifically, MathWorks™Matlab® and PSE© gPROMS®.

4 Preliminary studies. Fed-batch fermentations for recombinant protein production

Fed-batch fermentations for recombinant protein production using *E.coli* as host are commonly used for its simplicity in manipulation and good results. Fed-batch growth allows the culture to reach high cell densities before induction, which is important in terms of production.

Previous work in the research group dealt with recombinant aldolase production in fed-batch cultures growing at constant specific growth rate, and three variables have shown to be key in protein production: initial inducer concentration ($[IPTG]_{e,0}$, µM); fixed specific growth rate during fed-batch (μ_{fix}, h^{-1}) and biomass concentration at induction (X_{ind}, $gDCW \cdot L^{-1}$).

Inducer dosage has been one of the most studied factors. Two different induction strategies have been proposed: pulse inducer addition and continuous induction. Continuous induction consists of the modulation of the transcription rate by continuous inducer feeding in a constant ratio to biomass. This strategy allows extending the production phase and achieving optimal yields in some systems. On the other hand, induction by pulse consists of the addition of the necessary dose of IPTG, able to induce the culture, in one single pulse.

As reported in literature (Pinsach et al., 2008; Ruiz et al., 2009)), continuous induction applied to aldolase expression system leads to protein activities similar to those obtained using an optimal pulse induction, but with lower IPTG requirements. However, because of its simplicity, pulse addition (when the culture reaches the desired biomass concentration) has been chosen as the induction strategy for this work. One of the constraints of pulse-induced processes is the metabolic burden imposed on the host cell by the production of the recombinant protein. In consequence, there is a decrease of specific growth during the induction phase of the cultivation related to the amount of IPTG at induction (Ruizet al., 2011).

Specific fixed growth rate depends on the nutrient feeding during the fed-batch phase of the experiments. Nutrients for culture growth are added through the predefined

exponential feeding profile presented in Chapter 3 (Materials and Methods, Equation (3.1)), leading to a constant specific growth rate. This feeding takes into account the biomass-substrate yield (Y_{XS}) as well as the desired growth rate (μ_{fix}) (Pinsach et al., 2006). However, once the induction begins, it can be seen that specific growth rate decreases due to overexpression of the recombinant protein, and makes necessary to introduce either manual or automatic control in order to avoid glucose accumulation in the culture medium. Glucose accumulation (as a result of specific growth rate decrease) leads to lower protein activities, as reported in literature (Donovan et al., 1996; Pinsach et al., 2008).

Fed-batch cultivations make possible working at higher cell concentrations than in batch or continuous reactors (Maresová, Stepánek, & Kyslík, 2001; Shiloach et al., 1996; Turner, Gregory, & Turner, 1994). However, the high oxygen requirements in high cell population growth (oxygen limitation can occur), as well as the increase of viscosity and the changes in surface tension of the growing culture (low mass transfer rates, hindering nutrients and oxygen transfer), determine the maximum biomass concentration at induction time in order to allow the culture to grow and produce recombinant protein.

This chapter will present a qualitative and quantitative study of the importance of these three variables (inducer concentration, biomass concentration at induction and fixed specific growth rate) using the *E.coli* M15 Δ*gly*A [pQEαβrham][pREP4] strain. It will contribute to the familiarization with the experimental methodology as well as to give a global knowledge about protein production in recombinant *E.coli* cultures.

In previous works of the research group, the interactive effects of the three parameters mentioned before was studied using a Response Surface Methodology (RSM) (Fernández, 2012). With RSM it was possible to obtain equations (one for the specific protein in mass units (mg RhuA·g^{-1}DCW) and another one for specific protein in activity units (AU·g^{-1}DCW)), expliciting the crossing effect of the variables and obtaining a maximum for specific protein production in mass and in activity.

4.1 Previous work

Previous knowledge was used to delimitate the ranges of variation for the three variables, which are presented in Table 4.1.

Table 4.1. experimental ranges for the studied variables at induction time.

Variable	Units	Range
Inducer concentration ($[IPTG]_{e,0}$)	µM	24-96
Biomass concentration at induction time (X_{ind})	$gDCW \cdot L^{-1}$	20-40
Specific growth rate (μ_{fix})	h^{-1}	0.06-0.22

IPTG concentrations around a maximum of 96 µM have been identified to be sufficient for reaching the maximum protein production (Fernández-Castané et al., 2012; Pinsach et al., 2006; Sevastsyanovich et al., 2009). The lower value for induction in this case is 24 µM, following the RSM methodology that fixes it.

E.coli, in fed-batch cultures, can grow up to biomass concentration higher than 100 $gDCW \cdot L^{-1}$ (Lee, 1996). However, as said in the introduction of this chapter, oxygen limitations may compromise protein production. For this reason, the experimental range of biomass concentration at induction time was set between 20 $gDCW \cdot L^{-1}$ and 40 $gDCW \cdot L^{-1}$. The upper limit of the range has been selected for a final expected biomass concentration close to the maximum compatible with the oxygen transfer capacity of the system. Induction when biomass concentration is higher than 40 $gDCW \cdot L^{-1}$ leads on oxygen limitations that will not allow the culture to grow. The lower limit was set to accomplish the RSM range.

The maximum specific growth rate for this strain is $\mu_{max} = 0.55\ h^{-1}$ (in DM medium), but in order to avoid acetate accumulation -which clearly inhibits culture growth at specific growth rates over 0.25 h^{-1} (Ruiz et al., 2009), the effect of growth rate has been studied in a range between 0.06 h^{-1} and 0.22 h^{-1}. Below specific growth rates of 0.06 h^{-1}, the process is extremely slow, leading to bad productivities and experiments hard to follow because their duration.

Subsequently, the experimental space was defined according to the knowledge of the system, aiming the study of the combined effect of those three variables in protein production and its activity.

Experiments performed changing these three different experimental conditions are presented in Table 4.2. It can be seen that maximum specific protein in mass is reached when inducer is maximum, biomass minimum and specific growth rate maximum. These conditions correspond to an extreme of the curve. Specific protein (in activity) maximum is reached when induction is made at low biomass concentration, low inducer concentration and specific growth rate maximum, which represent an extreme of the experimental space.

On the other hand, in terms of maximum total protein production in mass units (mgRhuA·L^{-1}), it is reached at maximum biomass concentration at induction (data not shown). In the same way, average high values of activity (AU·L^{-1}) are found when the biomass concentration at induction point was 40 gDCW·L^{-1} (data not shown). This reason makes necessary reaching the maximum possible biomass concentration before induction if maximum total protein production and activity is wanted.

Table 4.2. Experiments performed at different induction conditions.

Experiment	Inducer (μM)	Biomass (g·L^{-1})	Growth rate (h^{-1})	Specific protein in mass (mg RhuA·g^{-1} DCW)	Specific protein in activity (AU·g^{-1} DCW)
1	24	20	0.1	93.0	489.0
2	96	20	0.1	102.6	569.4
3	24	40	0.1	73.6	375.2
4	96	40	0.1	56.8	301.4
5	24	20	0.22	90.9	756.9
6	96	20	0.22	122.0	596.2
7	24	40	0.22	59.2	288.0
8	96	40	0.22	90.8	388.5
9	60	30	0.06	83.9	194.3
10	60	30	0.16	89.3	522.6
11	60	30	0.16	83.6	509.1
12	60	30	0.16	86.6	500.2
13	60	30	0.16	85.1	506.2
14	60	30	0.16	85.2	498.4

Equation (4.1) showing the effect of the three variables on the specific protein in mass units was found using RSM (Fernández, 2012).

$$Specific\ RhuA\ in\ mass\ (mg \cdot g^{-1}DCW) = 147.27 + 0.32 \cdot I - 2.41 \cdot X - 324.79 \cdot \mu - 9.09 \cdot 10^{-3} \cdot I \cdot X + \\ +4.4 \cdot I \cdot \mu + 0.51 \cdot X \cdot \mu - 4.28 \cdot 10^{-3} \cdot I^2 + 0.03 \cdot X^2 + 430.86 \cdot \mu^2 \quad (4.1)$$

Equation (4.2) shows the surface response of specific activity of the protein:

$$Specific\ RhuA\ in\ activity\ (AU \cdot g^{-1}DCW) = \\ -288.35 + 3.03 \cdot I + 8.40 \cdot X + 7576.32 \cdot \mu + 0.12 \cdot X \cdot I - 18.11 \cdot I \cdot \mu - \\ -112.05 \cdot X \cdot \mu - 0.02 \cdot I^2 - 0.15 \cdot X^2 - 6338.25 \cdot \mu^2 \quad (4.2)$$

Unfortunately, using I, X and μ as variables, the optimal point is found in a corner of the experimental space.

4.2 Variable change

Even though inducer concentration, biomass concentration at induction and specific growth rate can be studied individually, there are published evidences that a key parameter in protein production optimization is the relationship between inducer and biomass concentrations at induction time (I/X, μmol IPTG·g^{-1}DCW) (Pinsach, de Mas, & López-Santín, 2008; Ruiz et al., 2011). This parameter was not included in Response Surface Methodology (RSM) previous work, but the resultant equation can be rewritten allowing the study of I/X effect on protein production and activity. In this way, a maximum in specific protein in mass as well as in activity is wanted, but using I/X as study variable instead of inducer concentration.

Equation (4.1) can be rewritten to take into account the effect of the inducer to biomass ratio (I/X) as shown in Equation (4.3):

$$Specific\ RhuA\ in\ mass\ (mg \cdot g^{-1}DCW) = 147.27 + 0.32 \cdot X \cdot {}^{I}/_{X} - 2.41 \cdot X - 324.79 \cdot \mu - 9.09 \cdot 10^{-3} \cdot X \cdot {}^{I}/_{X} \cdot X + 4.4 \cdot X \cdot {}^{I}/_{X} \cdot \mu + 0.51 \cdot X \cdot \mu - 4.28 \cdot 10^{-3} \cdot \left(X \cdot {}^{I}/_{X}\right)^{2} + 0.03 \cdot (X)^{2} + 430.86 \cdot \mu^{2} \quad (4.3)$$

In order to compare the total production depending on biomass concentration at the end (in mg·L^{-1} of cultivation), two different scenarios will be simulated: induction at 20 gDCW·L^{-1} and at 40 gDCW·L^{-1}.

Equation (4.3) has been used to obtain surface and contour plots for specific protein in mass units. Moreover, profiles corresponding to the specific growth rate where the maximum is reached were also calculated. The results are presented in Figures 4.1 to 4.6.

On the other hand, as it has been done with specific protein in mass, previous equation from Alfred Fernández Doctoral Thesis (Fernández, 2012) (Equation (4.2)) has been changed to take into account the effect of I/X, resulting in Equation (4.4):

$$Specific\ RhuA\ in\ activity\ (AU \cdot g^{-1}DCW) = -288.35 + 3.03 \cdot X \cdot {}^{I}/_{X} + 8.40 \cdot X + 7576.32 \cdot \mu + 0.12 \cdot X^{2} \cdot {}^{I}/_{X} - 18.11 \cdot X \cdot {}^{I}/_{X} \cdot \mu - 112.05 \cdot X \cdot \mu - 0.02 \cdot \left(X \cdot {}^{I}/_{X}\right)^{2} - 0.15 \cdot X^{2} - 6338.25 \cdot \mu^{2} \quad (4.4)$$

Equation (4.4) has been used to obtain surface and contour plots for specific protein activity. Moreover, profiles corresponding to the specific growth rate where the maximum is reached are also presented. Figures 4.7 to 4.12 show the results.

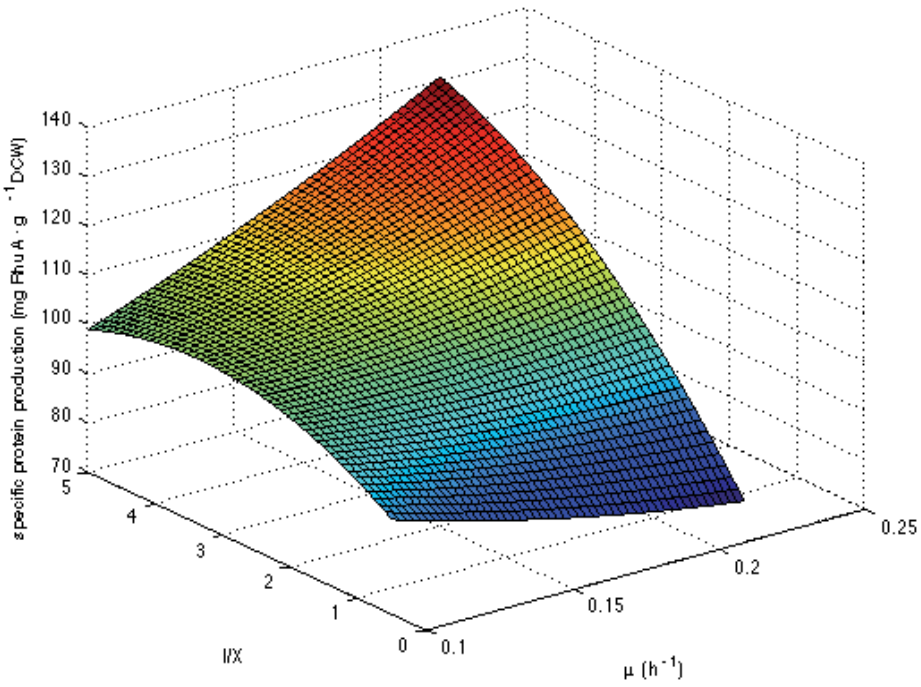


Figure 4.1. Surface plot of specific protein in mass induced at 20 gDCW·L^{-1}

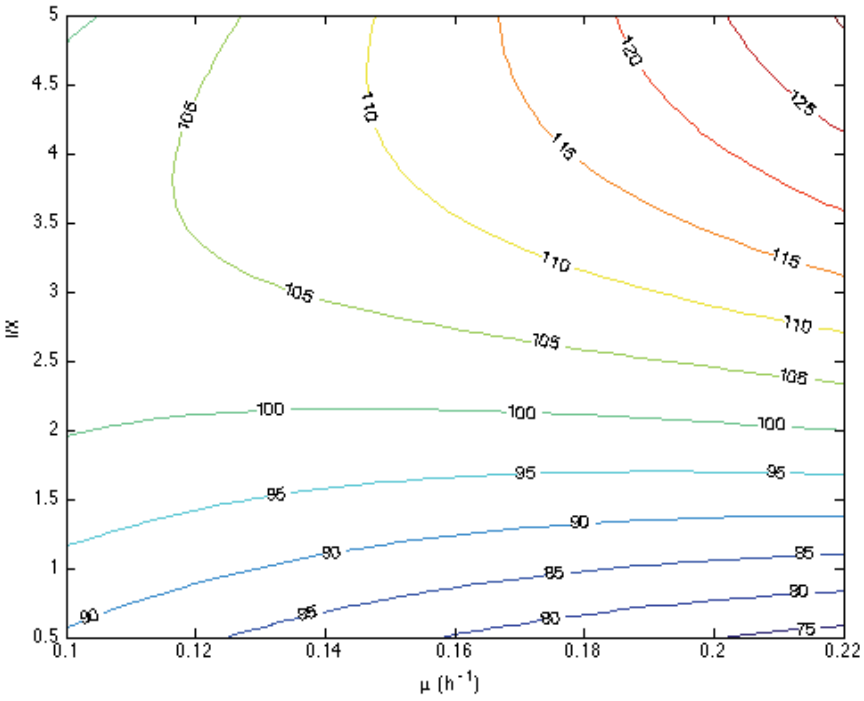


Figure 4.2. Contour plot of specific protein in mass induced at 20 gDCW·L^{-1}

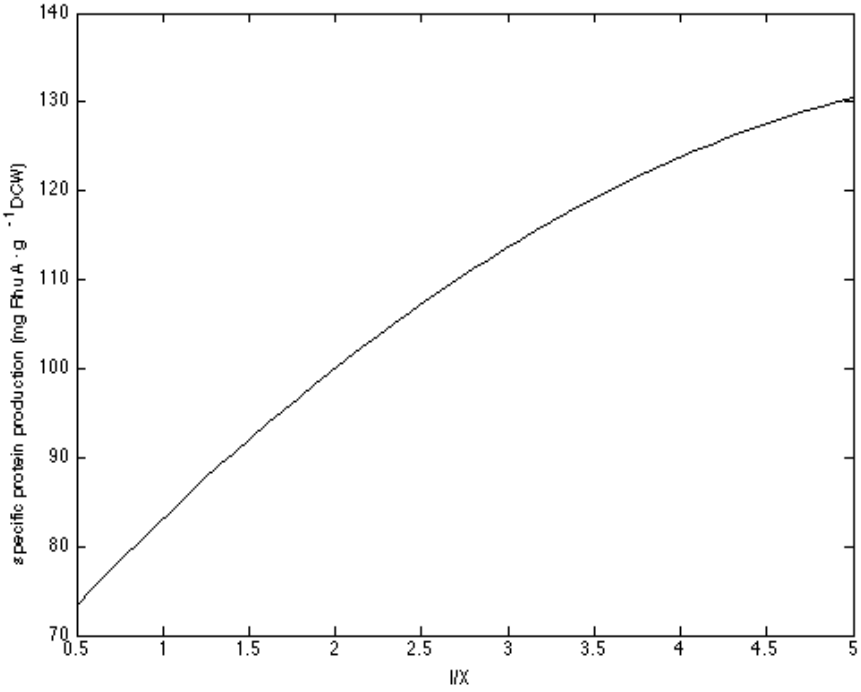


Figure 4.3. Specific protein in mass profile for μ= 0.22 h^{-1} induced at 20 $gDCW \cdot L^{-1}$

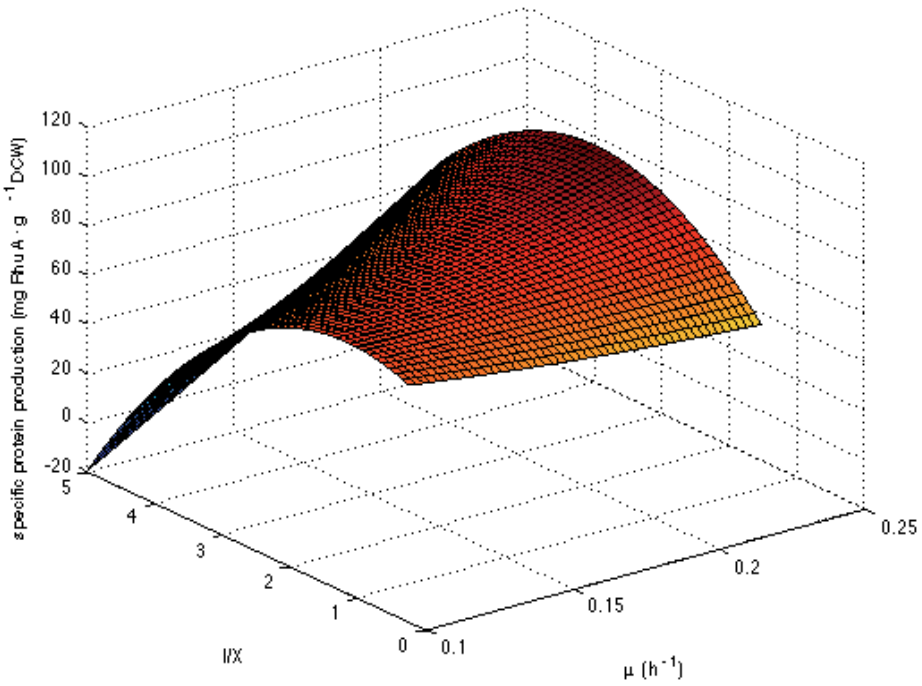


Figure 4.4. Surface plot of specific protein in mass induced at 40 $gDCW \cdot L^{-1}$

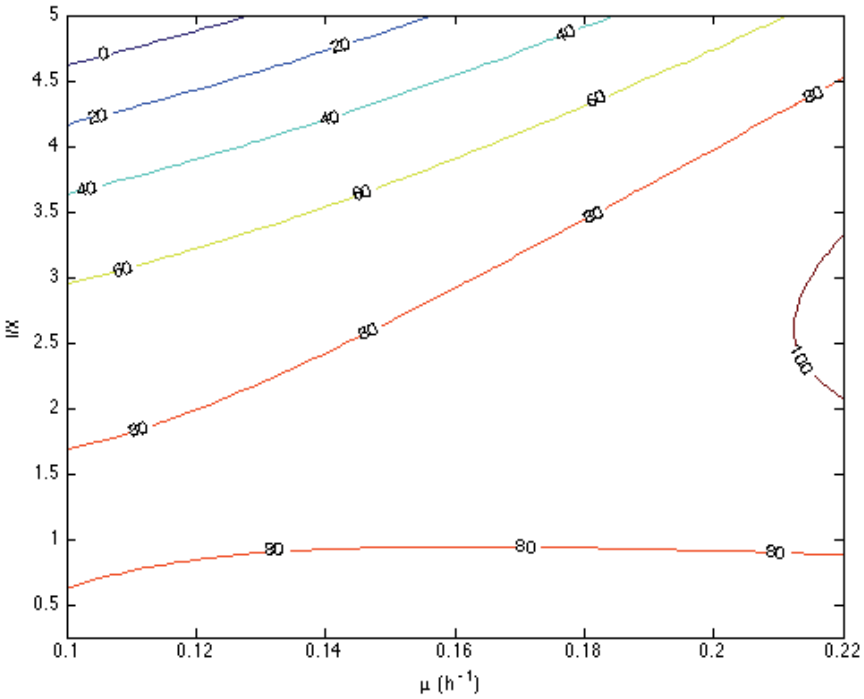


Figure 4.5. Contour plot of specific protein in mass induced at 40 gDCW·L^{-1}

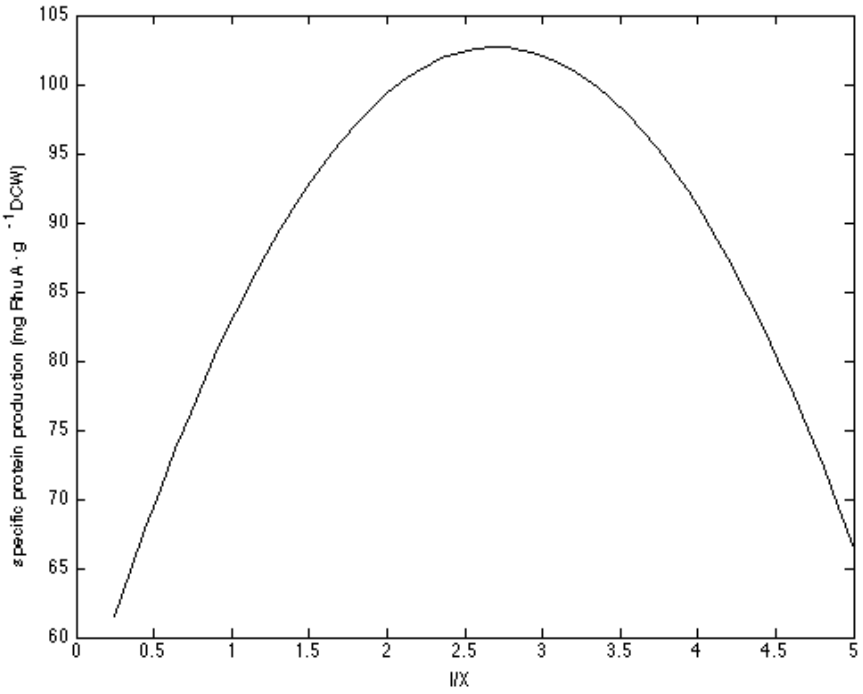


Figure 4.6. Specific protein in mass profile for μ= 0.22 h^{-1} induced at 40 gDCW·L^{-1}

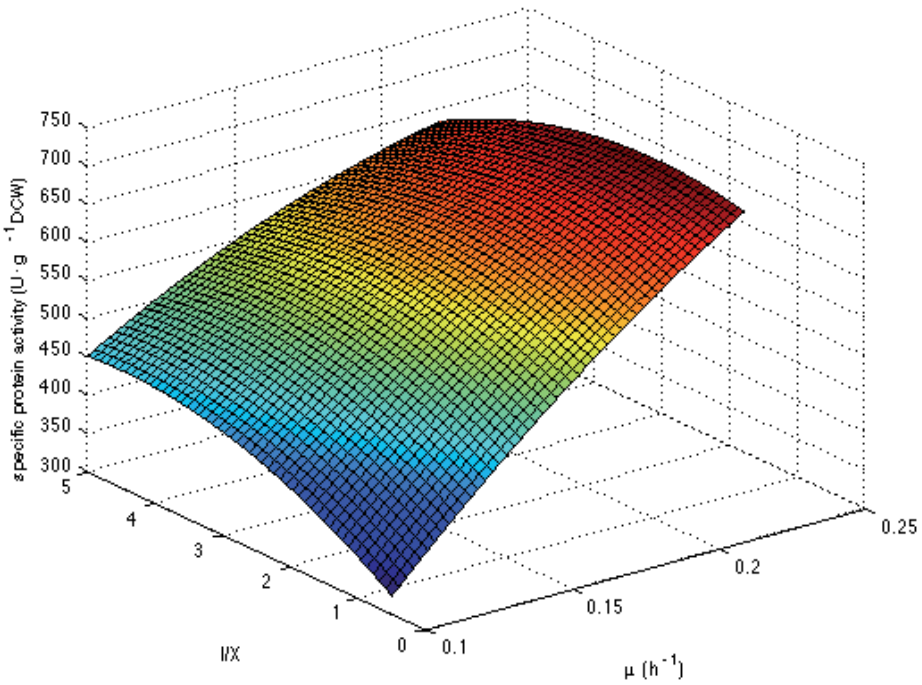


Figure 4.7. Surface plot of specific protein in activity induced at 20 gDCW·L^{-1}

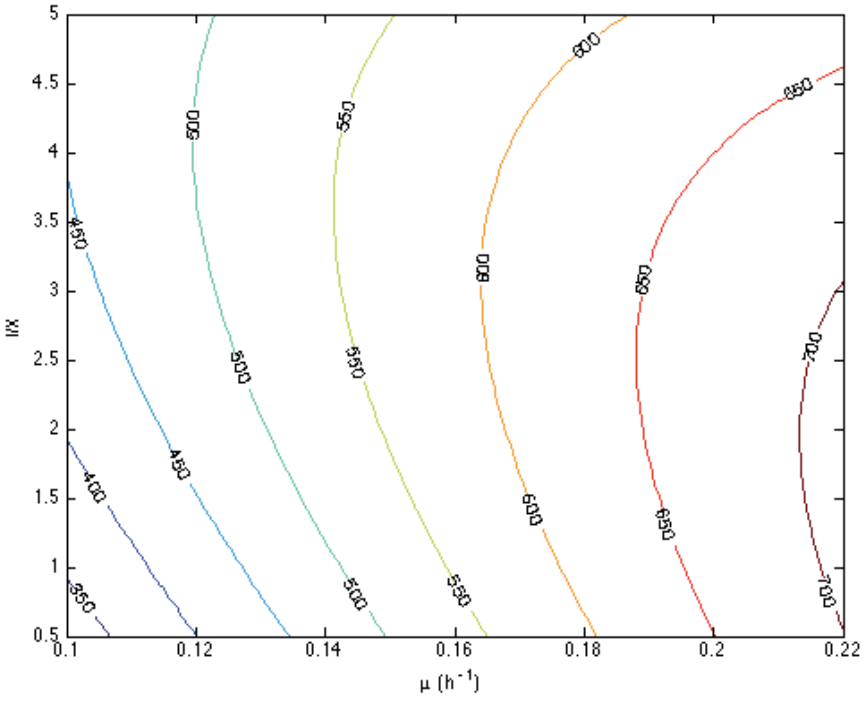


Figure 4.8. Contour plot of specific protein in activity induced at 20 gDCW·L^{-1}

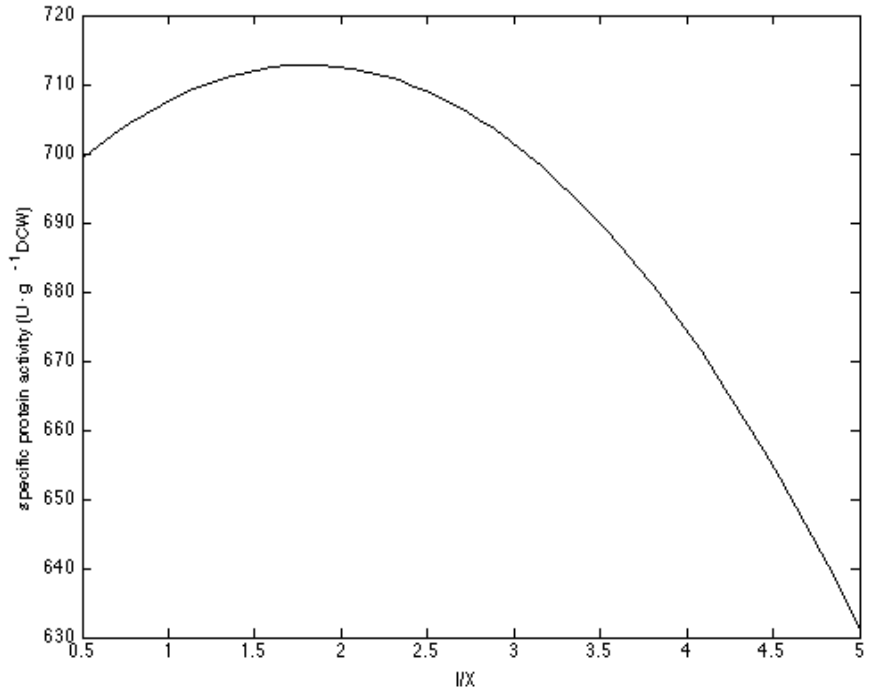


Figure 4.9. Specific protein in activity profile for μ= 0.22 h^{-1} induced at 20 gDCW·L^{-1}

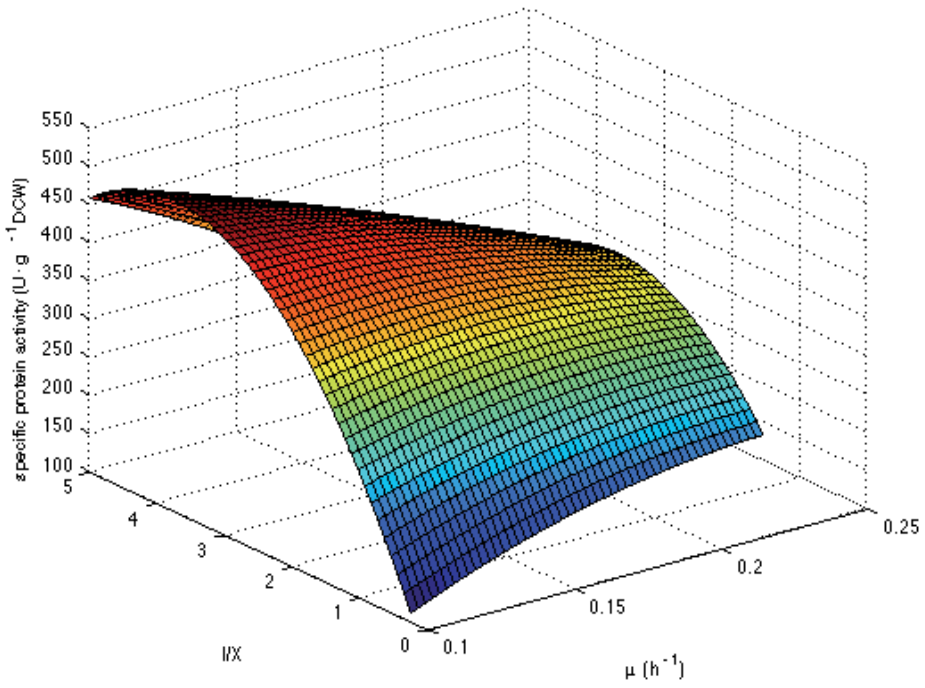


Figure 4.10. Surface plot of specific protein in activity induced at 40 gDCW·L^{-1}

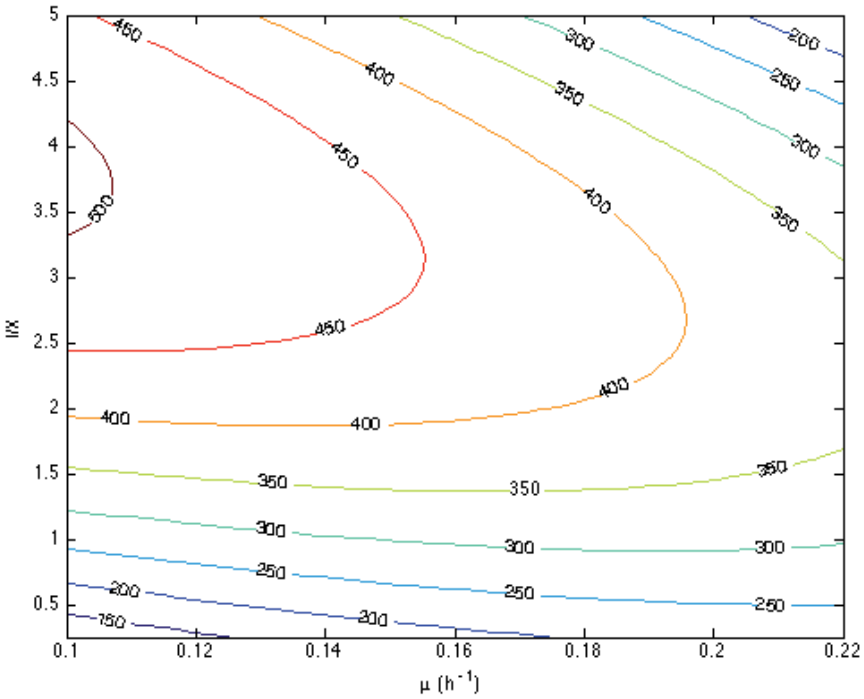


Figure 4.11. Contour plot of specific protein in activity induced at 40 gDCW·L^{-1}

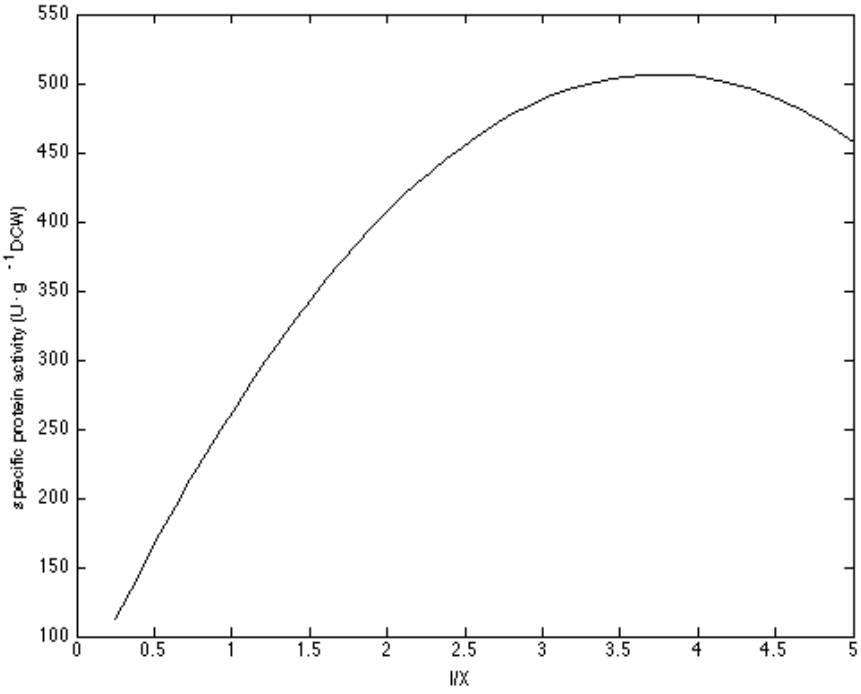


Figure 4.12. Specific protein in activity profile for μ= 0.1 h^{-1} induced at 40 gDCW·L^{-1}

Figure 4.12 shows, in this case, the profile at μ = 0.1 h^{-1}, when the activity is maximum.

Comparing the results obtained in Figures 4.1 to 4.6, it can be seen that specific protein in mass units is higher when induction is performed at biomass concentration of 20 gDCW·L^{-1} than the obtained when biomass has 40 gDCW·L^{-1} of concentration. In terms of total protein in mass production, as biomass is twice higher, induction at 40 gDCW·L^{-1} is going to reach much more total protein (in fact, 260 mgRhuA·L^{-1} for induction at 20 gDCW·L^{-1}; and 400 mgRhuA·L^{-1} if the induction is at 40 gDCW·L^{-1}). This fact makes interesting to work with the maximum biomass concentration possible at the end of the batch, taking into account the process limitations.

Moreover, it seems clear that specific production in mass rises with the growth rate. This fact could be explained due to the up-regulation of the genetic machinery at higher growth rates, and therefore allowing a higher production of the heterologous protein (RhuA). Accordingly, the standard specific growth rate can be set to μ_{fix} = 0.22 h^{-1}. Figures 4.1 to 4.6 show that in terms of specific protein in mass, it is important to work at maximum affordable growth rate. In that case, two different situations can be found. On one hand, if the induction has been made at 20 gDCW·L^{-1}, there is not a maximum in the range studied. This means that it is possible that a higher concentration of IPTG would be needed to achieve the maximum specific protein in mass units. On the other hand, when the culture is induced at a biomass concentration of 40 gDCW·L^{-1}, a maximum is reached at I/X ratio of 2.5. Then, in order to optimize the total amount of protein, it is necessary to work at maximum specific growth rate possible (avoiding growth inhibitions), inducing at higher possible biomass concentration (avoiding oxygen and mass transfer limitations) and no more than 2.5 µmol IPTG·g^{-1}DCW is needed.

Otherwise, Figures 4.7 to 4.12 show the results in terms of specific protein in activity, where two different situations can also be found. On one hand, induction at low biomass concentration needs to work at maximum specific growth rate, and the maximum is located at I/X ratio of 2 µmol.$gDCW^{-1}$. On the other hand, if high biomass concentration is present in the bioreactor at induction time, the model reaches its maximum when specific growth rate is as low as possible, leading to bad productivities and slow processes. Specific activity when biomass concentration at induction is high takes lower values (increase of 20% when induction takes place at 20 gDCW·L^{-1}).

Moreover, I/X ratio takes a value of around 4 μmol.gDCW^{-1}. This fact could seem to be in contradiction with the maximum found in specific protein in mass, but it can be attributed to down regulation of the genetic machinery at low growth rates which allow a better folding of the proteins (Gasser et al., 2008).

Specific growth rate for maximum specific protein in mass is different than for the maximum specific protein in activity. In this case, it will be necessary to decide a compromise between total quantity of protein into the fermentor and its activity.

Inducing at high biomass concentration allows obtaining higher amount of protein, which can be maximized working at high specific growth rate. However, working at low μ will lead to achieve higher protein activity. Then, a compromise in the value of fixed growth rate is needed to achieve a proper quantity of aldolase with good activity. As it is more interesting to achieve high activity values instead of the production of inactive protein, it can be seen that setting the standard experimental conditions to 20 gDCW·L^{-1} and 0.22 μ_{fix} will lead to higher specific activity of the protein, and better activity productivity. The volumetric specific activity productivity when induced at 20 gDCW·L^{-1} is 550 AU·L^{-1}·h^{-1}, while it is 500 AU·L^{-1}·h^{-1} when induced at 40 gDCW·L^{-1}.

4.3 Conclusions

It has been possible to investigate the process variables effect taking into account the I/X ratio and achieving a significant empirical mathematical model for maximum specific RhuA production in mass and in activity units.

Not extremely high concentration of inducer is needed to achieve a maximum in specific protein production (both in activity and mass), which not only depends on inducer concentration, but also on the specific growth rate and inducer to biomass concentration ratio. Only 2-4 μmol IPTG·g^{-1}DCW are needed.

Otherwise, since IPTG concentration at induction is an important parameter, further study about its role in inductive expression systems will be needed.

5 Modeling IPTG transport phenomena in fed-batch high-cell density cultures of *E.coli*

Although a maximum in activity production has been found in Chapter 4, it is necessary to pursue with IPTG transport studies. RSM models use statistics for the determination of optimal points, but they are not time-dependent. RSM gives an idea about the optimal experimental conditions, but does not describe the process. Fundamental-based models want to describe the evolution of the different variables involved in the process, making a description of it.

As seen in the previous chapter, IPTG dose added to the culture in order to induce the overexpression of the recombinant protein is a determinant factor in aldolase production (Donovan et al., 1996; Durany et al., 2004). Correlation between aldolase production and inducer concentration is complex (Ruiz et al., 2011) and some different hypothesis can be found in literature. Some of them suggest that the probability of binding between IPTG and repressor depends on the intracellular concentration of inducer (Vilar, Guet, & Leibler, 2003), while some others support that induction system is stochastic (Rao, Wolf, & Arkin, 2002), using every variable independently.

lac operon derived systems (described in General Introduction section), as productive systems, with additional copies of its constituents (i.e. *lac*I repressor gene and operator sites) present in plasmids are usually not described in literature (Vilar & Saiz, 2013) for the global process of protein production, and if described, *lac* operon dynamics is modeled with scarce and indirect experimental data. Electrical membrane potential, as well as interaction between membrane proteins and metabolites have been proposed in literature to explain the IPTG transport mechanism between the culture medium and the intracellular space (Cuppoletti & Segel 1975; Garcia et al. 1982). Nevertheless, more recent publications in the field formulate some different hypothesis about inducer transport: in addition to semipermeable membrane that allows IPTG diffusion through itself, active transport must be taken into account because of lactose permeases (which expression is also promoted by IPTG) (Jensen & Hammer, 1998; Jensen, Westerhoff, & Michelsen, 1993). IPTG could go out from

intracellular space by the same mechanism as foe entering. However, other authors consider that effects of active transport from intracellular to extracellular space are negligible, as well as the effect of diffusion in IPTG transport from culture medium to the intracellular space (because its minor contribution) (Noel, Pilyugin, & Narang, 2009).

In addition to the controversy about transport mechanism, a bistability behavior of the operon, due to population diversity in function of the induction, has been postulated (segregated models). Depending on the amount (or concentration) of inducer, fully or non-fully induced cultures can be found, with different behavior both in IPTG transport and protein production (Laurent, Charvin, & Guespin-Michel, 2005; Noel et al., 2009). Moreover, other variables, such as the union between repressor and inducer, and glucose and lactose permease concentration, can influence in inducer transport mechanism (Ozbudak et al., 2004; Santillán & Mackey, 2004; Santillán, 2008). Only indirect evidences (like β-galactosidase activity or fusion of fluorescence proteins with the *lac* operon) have been used to contrast all these hypotheses, leading to the proposition of theoretical models, but the validation of the system (and its equations) needs direct measurements of the elements involved.

An inducer quantification method has been recently developed, allowing the obtainment of direct measurements of extracellular IPTG (Fernández et al., 2010). This method, based on HPLC-MS (High Performance Liquid Chromatography linked to Mass Spectrometry) allows the determination of extracellular inducer concentration, and hence the estimation of the intracellular concentration, making possible modeling the experimental behavior and validate the different hypotheses.

In this chapter, a validated mathematical model of IPTG uptake, using real experimental data from fed-batch cultures is reported by the first time. The experimental system was the production of Rhamnulose-1-Phosphate Aldolase (RhuA). The mathematical model, describing extracellular IPTG profiles as well as intracellular inducer accumulation along the time, will be built and calibrated. The model that is going to be developed will incorporate the effects on IPTG uptake of both active and passive transport phenomena, related to extracellular and intracellular IPTG concentrations along time. Moreover, the effect of biomass concentration and the

specific growth rate at induction point will be studied. The model will be validated using experimental data of three different strains, two of them producing Fuculose-1-Phosphate Aldolase (FucA), and one producing Fructose-6-Phophate Aldolase (FSA).

5.1 Model balances

An unsegregated, unsteady, unstructured and based on first principles model has been proposed to describe inducer transport in IPTG induced fed-batch cultures of *E.coli* M15Δ*gly*A [pQEαβrham][pREP4] producing RhuA. The selection of the type of model has been made following the explanation of the General Introduction Chapter. Cells are considered as single compartment perfectly stirred, meaning that inner concentration is assumed uniform.

Total volume of cells in the reactor can be calculated using experimental data of the total volume and experimental biomass concentration along the time:

$$\frac{dV_{cel}}{dt} = 0.0023\frac{d(XV)}{dt} \qquad (L \cdot h^{-1}) \tag{5.1}$$

constant 0.0023 is the specific cell volume in $L_{cell} \cdot g^{-1}DCW$, according to literature (Bennett et al., 2008).

Total biomass in the reactor (XV, gDCW), can be calculated from experimental biomass concentration (X) and experimental volume (V) data. Resulting values of (XV) are adjusted to a time-dependent function using splines that allows the evaluation of the derivative. Total volume in the bioreactor is also adjusted to a time-dependent spline function.

Culture medium volume can be calculated by difference between V and V_{cel}

$$\frac{dV_m}{dt} = \frac{dV}{dt} - \frac{dV_{cel}}{dt} \qquad (L \cdot h^{-1}) \tag{5.2}$$

IPTG evolution can be written following mass balances as:

$$\frac{d[IPTG]_e}{dt} = \frac{-r \cdot V_m - [IPTG]_e \frac{dV_m}{dt}}{V_m} \qquad (\mu M \cdot h^{-1}) \qquad (5.3)$$

$$\frac{d[IPTG]_i}{dt} = \frac{r \cdot V_m - [IPTG]_i \frac{dV_{cel}}{dt}}{V_{cel}} \qquad (\mu M \cdot h^{-1}) \qquad (5.4)$$

Equations (5.3) and (5.4) describe the variation with time of extracellular and intracellular IPTG concentration ($[IPTG]_e$, $[IPTG]_i$, respectively) as a function of the net transport rate (r) and a dilution term. Net transport rate (r, $\mu M \cdot h^{-1}$) is the target of the modeling, because it is necessary to calculate the IPTG profiles.

In order to estimate the parameters of the model, only measured values of $[IPTG]_e$ have been used because experimental values of $[IPTG]_i$ were not reliable enough due to the experimental errors associated to sample processing (Fernández-Castané et al., 2012).

5.2 IPTG transport rate modeling

Previous studies comparing IPTG transport in a *lac*Y deficient strain (from now *lac*Y mutant strain) –gene codifying for lactose permease formation- and its parent strain showed qualitative differences. As it is shown in Figure 5.1 (where t_{ind} is time after induction) -as an example-, parent strain exhibits faster initial transport rate than *lac*Y mutant strain when all experimental conditions are the same (inducer concentration, biomass concentration at induction and specific growth rate), indicating significant differences in transport mechanisms that should be studied separately (Fernández-Castané et al., 2012; Fernández-Castané et al., 2012). Lactose permeases mediate the specific transport mechanism of IPTG uptake. lacY mutant strain (lactose permeases deficient) does not have specific active transport mechanism, and the fact that exhibits a lower initial transport rate could mean that active transport is the main mechanism of IPTG uptake for the parent strain. Then, the strategy for modeling will be, firstly, to study the effect of the non-specific transport. This is necessary because specific active transport is masking all the other transport mechanisms.

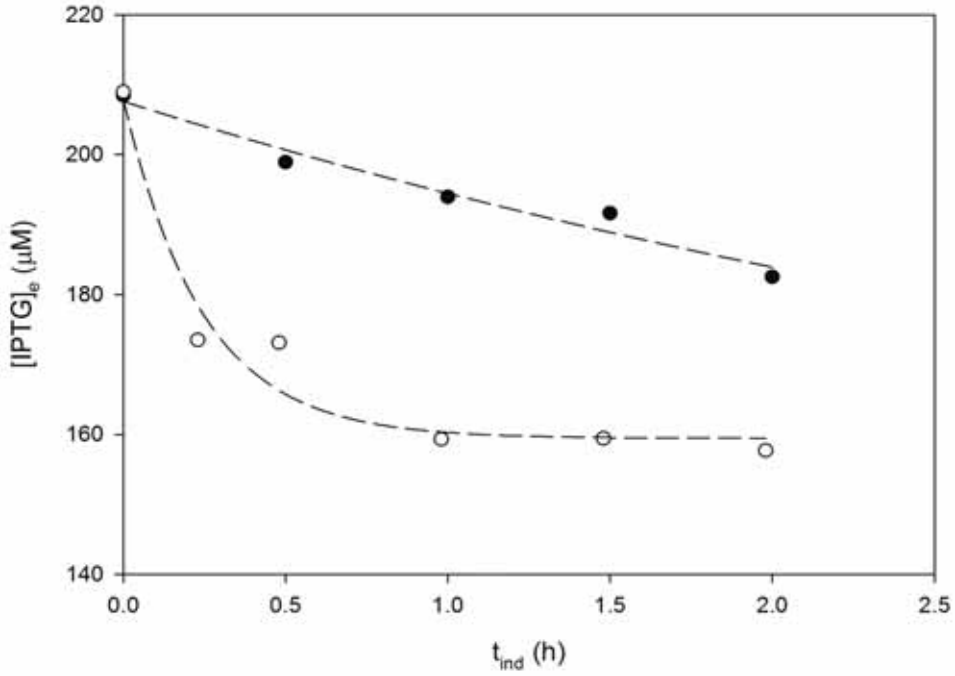


Figure 5.1. $[IPTG]_e$ depletion comparison for parent strain and *lac*Y mutant strain. (o) parent strain; (•) *lac*Y mutant strain; (--) trend curves.

5.2.1 IPTG transport rate model for *lac*Y mutant strain

E.coli M15 Δ*gly*AΔ*lac*Y [pQEαrham][pREP4] strain was used to determine the transport mechanism in absence of lactose permeases.

Four different experiments, using the *lac*Y mutant strain, induced at 10 µM, 20 µM, 54 µM and 200 µM, were conducted. All of them were induced at the standard conditions explained in Chapter 4: when reached a biomass concentration of 20 $gDCW \cdot L^{-1}$ and fed-batch cultivations were grown at specific growth rate of $\mu=0.22\ h^{-1}$.

The first approximation of modeling was assuming that only diffusion through cellular membrane takes place in absence of lactose permeases (the specific transport proteins). However, as it can be seen in Figure 5.2, the calculation of intracellular concentration (from extracellular experimental data) of IPTG (as difference between total amount added and the remaining in culture medium) for every experimental point shows that $[IPTG]_i$ can be higher than $[IPTG]_e$, indicating that, even in absence of the specific transport proteins, diffusion can not be the only transport mechanism.

In fact, in order to allow intracellular IPTG concentration to be higher than extracellular and to match the experimental extracellular measures, IPTG transport rate has to take into account some active transport contribution (not related to lactose permeases) as well as diffusive mechanism.

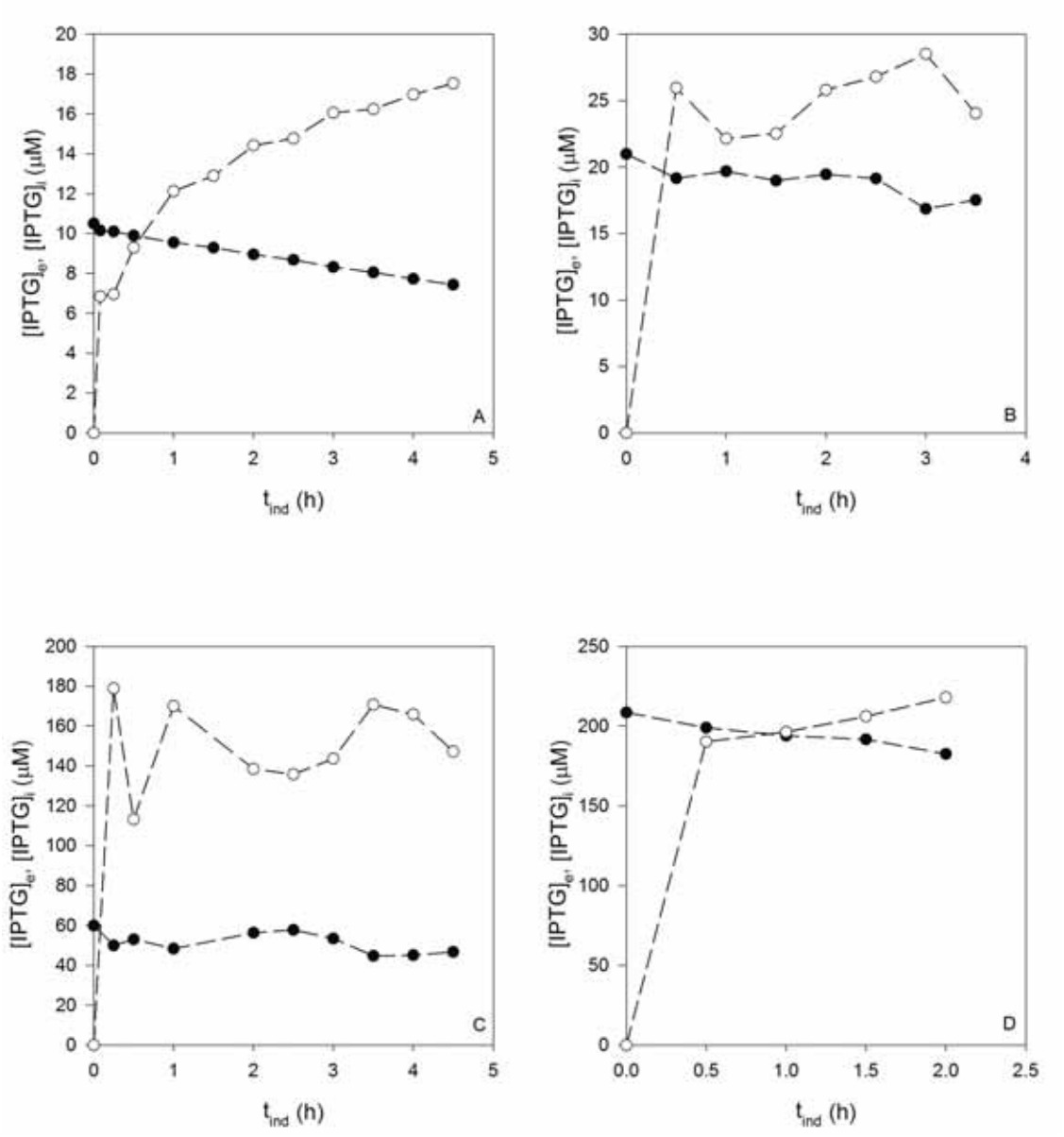


Figure 5.2. (•) experimental $[IPTG]_e$ and (o) point by point calculated $[IPTG]_i$ for *lac*Y mutant strain fed-batch cultures. A: $[IPTG]_{e,0}$=10 μM; B: $[IPTG]_{e,0}$=20 μM; C: $[IPTG]_{e,0}$=54 μM; D: $[IPTG]_{e,0}$=200 μM.

Equation 5.5 presents both contributions (diffusion and active transport). Diffusion rate depends on the difference between extracellular and intracellular inducer concentration, and also on a specific mass transfer coefficient k_ca (h^{-1}).

$$r = k_c a([IPTG]_e - [IPTG]_i) + \frac{k\,[IPTG]_e}{K_M + [IPTG]_e} \quad (5.5)$$

For nonspecific active transport, a Michaelis-like expression has been chosen because the similarity between transport proteins and enzymes can be extended to the kinetics of their action (Alberts et al., 2002). In this specific case, assuming that the transport is nonspecific for lactose (neither IPTG), K_M value is expected to be much higher than $[IPTG]_e$, and the expression for IPTG transport rate can be rewritten as shown in Equation (5.6).

$$r = k_c a([IPTG]_e - [IPTG]_i) + K' [IPTG]_e \quad (5.6)$$

where $K' = k/K_M$.

The model corresponding to Equations (5.1) to (5.4), using Equation (5.6) as IPTG net transport rate was fitted to experimental data obtained from the experiments presented above. Every experiment has 10 to 15 extracellular IPTG concentration measurements, and the model has 2 different parameters to be estimated.

The values found for the two parameters of Equation (5.6) are presented in Table 5.1. As a result of model fitting and parameter estimation, Figure 5.3 shows the experimental extracellular measurements, and the predicted evolution of IPTG both extracellular and intracellular.

Table 5.1. Estimated values for IPTG transport rate in *lac*Y mutant strain

Parameter	Value	Units
$k_c a$	$0.213 \pm 2 \cdot 10^{-3}$	h^{-1}
K'	$0.0893 \pm 4 \cdot 10^{-4}$	h^{-1}

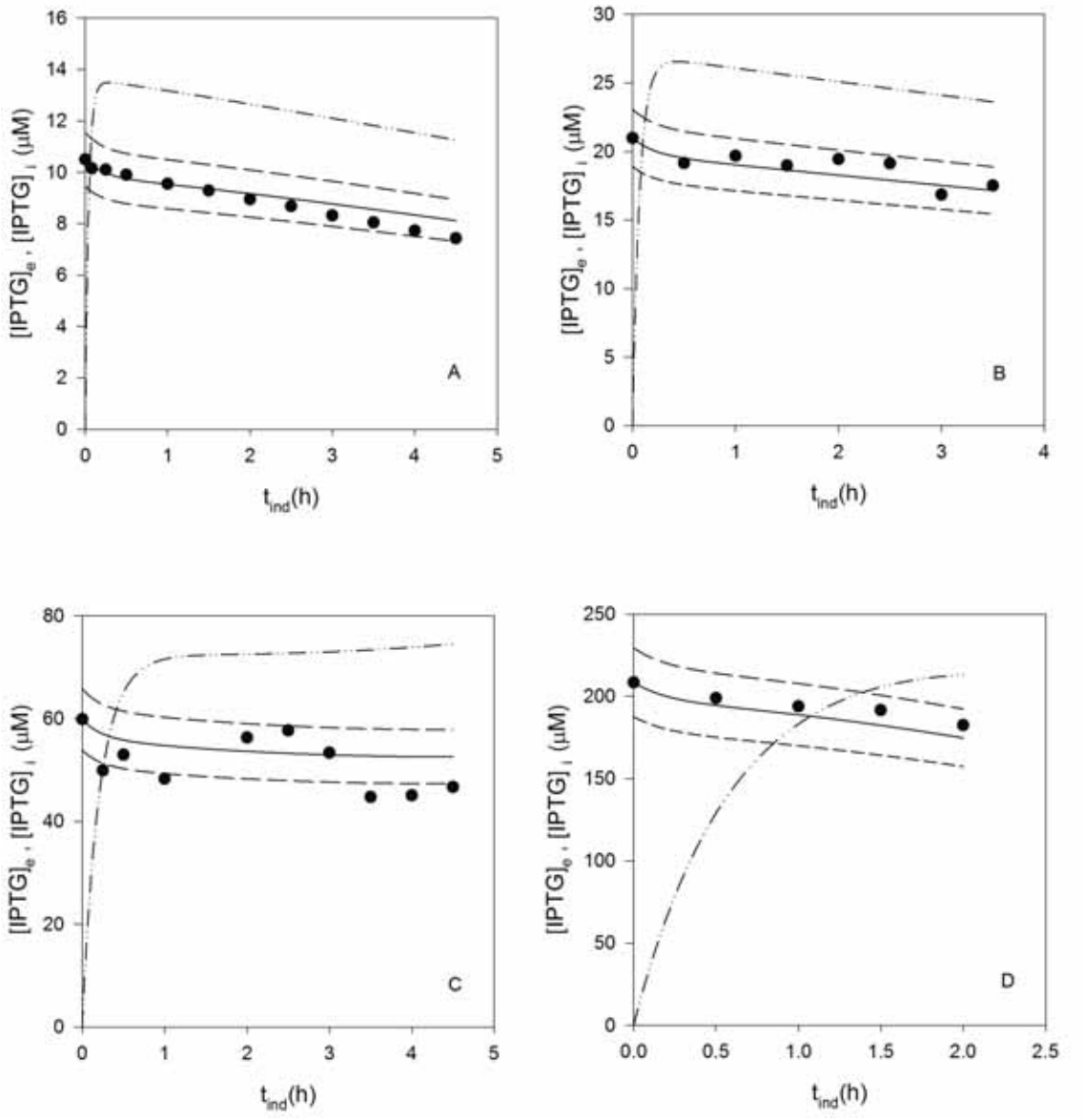


Figure 5.3. Model fitting for *lac*Y mutant strain. (•) experimental data, (-)$[IPTG]_e$ model prediction, (- -) 10% error interval, (-···-)$[IPTG]_i$ model prediction. (A) $[IPTG]_{e,0}$=10μM; (B) $[IPTG]_{e,0}$=20μM; (C) $[IPTG]_{e,0}$=54μM; (D) $[IPTG]_{e,0}$=200μM.

As it can be seen in Figure 5.3, the model is able to predict properly the IPTG depletion from medium culture to the intracellular space for *lac*Y mutant strain, within a 10% error. The error associated to analytical measurements was already estimated to be in the 10% range (Fernández et al., 2010), without accounting for other possible errors associated, for example, to the measure of OD_{600}.

In summary, IPTG transport rate for *lac*Y mutant strain can be explained as diffusion plus nonspecific active transport.

5.2.2 IPTG transport rate model for parent strain

The starting point for parent strain IPTG uptake modeling is *lac*Y mutant strain fitted equation (Equation 5.6). As a first approach, a new term will be added to Equation (5.6) describing the contribution of lactose permeases to IPTG transport.

Inducer transport rate depends on extracellular and intracellular IPTG concentration, and literature describes the important role of lactose permeases in IPTG uptake (Laurent, Charvin, & Guespin-Michel, 2005; Noel et al., 2009; Ozbudak et al., 2004; Santillán & Mackey, 2004; Santillán, 2008). Some of these models include equations trying to describe how IPTG induce permeases formation, which, in turn, catalyzes IPTG transport. In this work, as no experimental data of lactose permeases activity are available, their concentration was not explicitly included in the model and a simplified equation based on extracellular and intracellular IPTG concentrations is proposed:

$$r = k_c a([IPTG]_e - [IPTG]_i) + K' [IPTG]_e + \frac{k'[IPTG]_e}{K_M + [IPTG]_i^2} \qquad (\mu M \cdot h^{-1}) \qquad (5.7)$$

Equation (5.7) shows that transport rate depends on the same mechanisms as for *lac*Y mutant strain (diffusion and nonspecific active transport) and specific active transport due to lactose permeases. This specific active transport depends directly on the extracellular concentration of inducer, and it is inhibited by the intracellular IPTG concentration. The square term for intracellular concentration matches with Michaelis like model with substrate inhibition.

Moreover, because some different biomass concentration and specific growth rate at induction time can be used, it is necessary to study their effect in transport rate equation, giving the model wider prediction range. Figure 5.4 shows the experimental dependence of initial normalized IPTG transport rate with biomass concentration at induction (X_{ind}) and with the specific growth rate of the fed-batch phase (μ_{fix}). It can be seen that initial transport rate depends linearly with biomass concentration at induction, while the dependence is exponential with specific growth rate.

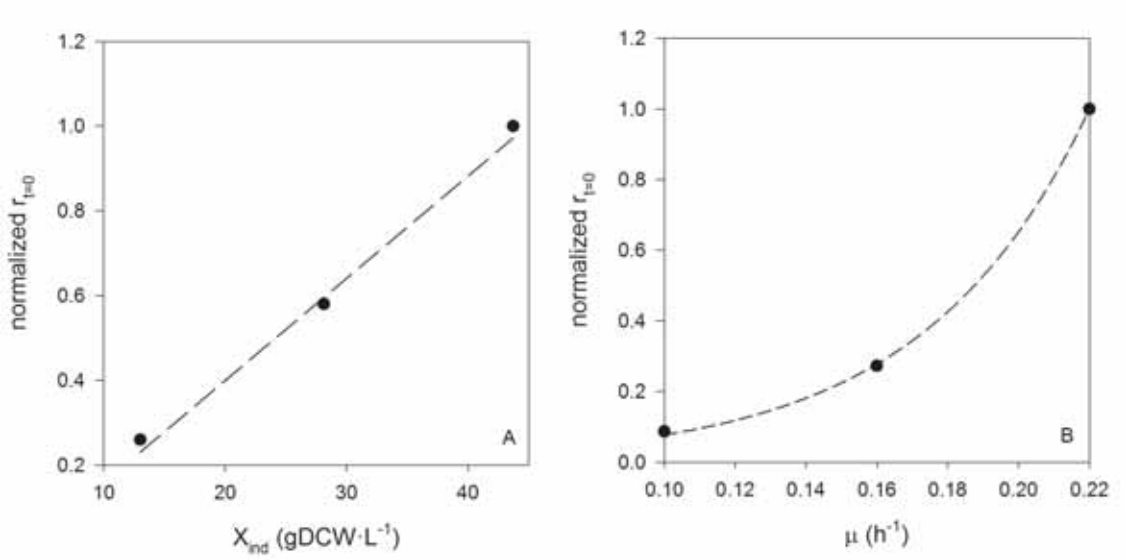


Figure 5.4. Initial transport rate dependence with (A) biomass concentration at induction and (B) fixed specific growth rate.

With the information in Figure 5.4, Equation (5.7) can be modified to take into account these new contributions. Linear dependence with biomass concentration can be attributed to the fact that, the more cells in the culture, the more IPTG can be transported with direct proportionality. On the other hand, specific growth rate affects, exponentially, gene transcription rate (*lac*Y gene in this case) and its expression. Moreover, these contributions will only affect the active transport terms, because are the ones depending on the cellular concentration and its growth. Diffusion takes into account biomass concentration evolution with time on an implicit way (through $[IPTG]_e$ and $[IPTG]_i$ evolution with time, Equations (5.3) and (5.4).

Equation (5.8) shows the final expression for parent strain IPTG transport rate.

$$r = k_c a([IPTG]_e - [IPTG]_i) + K_1 X_{ind}\left(K'[IPTG]_e + \frac{k'[IPTG]_e}{K_M + [IPTG]_i^2}\right)\exp\left(K_2 \mu_{fix}\right) \qquad (5.8)$$

with X_{ind} as biomass concentration at induction and μ_{fix} as the specific growth rate at induction point.

Nevertheless, parameter K' from equation (5.6) already includes the effect of X_{ind} and μ_{fix}, because all the mutant strain experiments were performed at same biomass concentration at induction (20 g.L^{-1}) and at the same fixed specific growth rate (0.22 h^{-1}). Then, in fact, K' must be rewritten as shown in equation (5.9)

$$K' = 0.0893\ h^{-1} = K'' \cdot K_1 X_{ind} \cdot exp(K_2 \mu) = K'' \cdot 20K_1 \cdot exp(0.22K_2) \quad (5.9)$$

were all the terms are constant. Because of that, all of them were included in a sole parameter in equation (5.6) describing mutant strain transport. This fact changes equation (5.8), yielding the expression shown in equation (5.10), where the parameter of the non-specific active transport changes according to equation (5.9)

$$r = k_c a([IPTG]_e - [IPTG]_i) + K_1 X_{ind} \left(K''[IPTG]_e + \frac{k'[IPTG]_e}{K_M + [IPTG]_i^2}\right) \exp\left(K_2 \mu_{fix}\right) \quad (5.10)$$

with $$K'' = \frac{0.0893}{20K_1 \cdot \exp(0.22K_2)} \quad (5.11)$$

5.2.2.1 Induction zones

Parent strain exhibited two different IPTG transport rate behaviors. As already published (Fernández-Castané et al., 2012), a gradual extracellular inducer concentration decrease with time is found when the concentration of IPTG at induction point is below 30 µM, because of low initial transport rates (see Figure 5.5). On the other hand, when IPTG concentration at induction point is above 60 µM, faster initial transport rates are observed (one to two orders of magnitude higher) and IPTG presents an initial rapid depletion from medium culture (as shown in Figure 5.5 as an example). Some works in literature hypothesize that partial induction leads to a lower role of *lac* permeases in transport than at high IPTG concentration because the *lac* operon is induced by the same IPTG (Fernández-Castané et al., 2012; Hansen et al., 1998; Jensen & Hammer, 1998; Jensen et al., 1993).

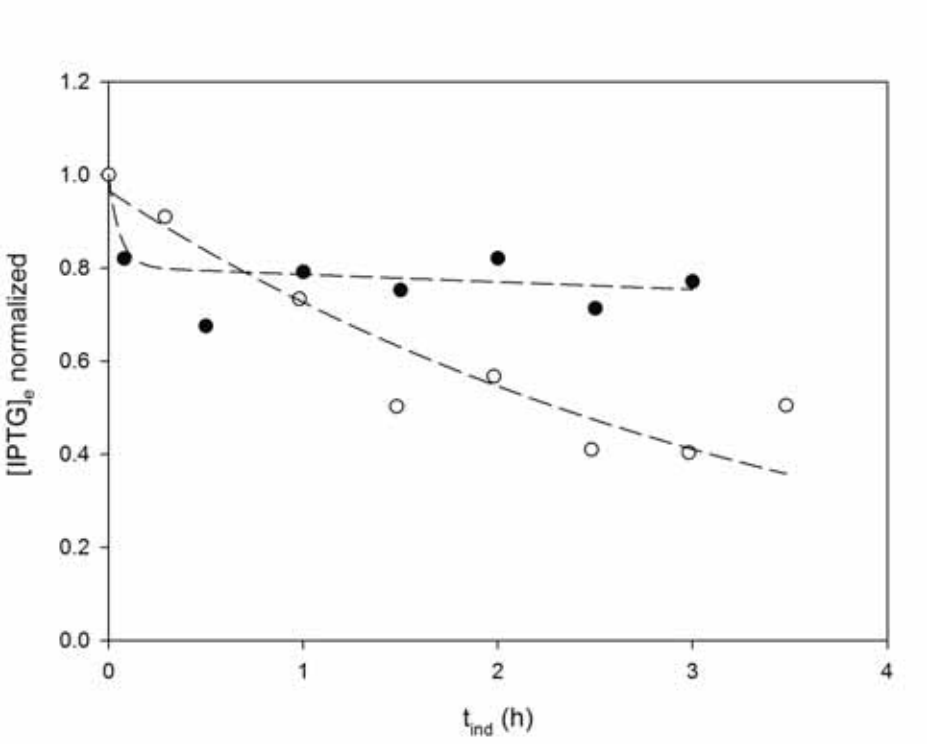


Figure 5.5. Qualitative comparison between high and low induction zones. (•) high induction zone normalized $[IPTG]_e$ data; (o) low induction zone normalized $[IPTG]_e$ data; (--) trend curves.

Experiments have been conducted under a wide range of conditions across the experimental spaces. The high induction zone is within the following experimental space (10 different experiments):

- Inducer concentration: $60\mu M \leq [IPTG]_{e,0} \leq 1000\mu M$
- Biomass concentration at induction time: $13 gDCWL^{-1} \leq X_{ind} \leq 47 gDCWL^{-1}$
- Specific growth rate: $0.06h^{-1} \leq \mu_{fix} \leq 0.22h^{-1}$

The low induction zone (10 experiments performed) has the following boundaries:

- Inducer concentration: $8\mu M \leq [IPTG]_{e,0} \leq 27\mu M$
- Biomass concentration at induction time: $20 gDCWL^{-1} \leq X_{ind} \leq 40 gDCWL^{-1}$
- Specific growth rate: $0.1h^{-1} \leq \mu_{fix} \leq 0.22h^{-1}$

The boundary limits for the variables have been selected to be the same as in Chapter 4, but expanding the upper limit of inducer concentration to 1000 μM and the lower

limit to 8 µM in order to obtain additional experimental information on IPTG transport at extremely high and low inducer concentrations.

Figures 5.6 and 5.7 show the distribution of the experiments into these defined experimental spaces. It is not any kind of Experimental Design Distribution, but experiments have been distributed into the experimental space to get representative results of the whole range of experimental conditions. As it can be seen, there are three different axes, showing the experimental boundaries for biomass concentration at induction, specific growth rate and IPTG concentration at induction point.

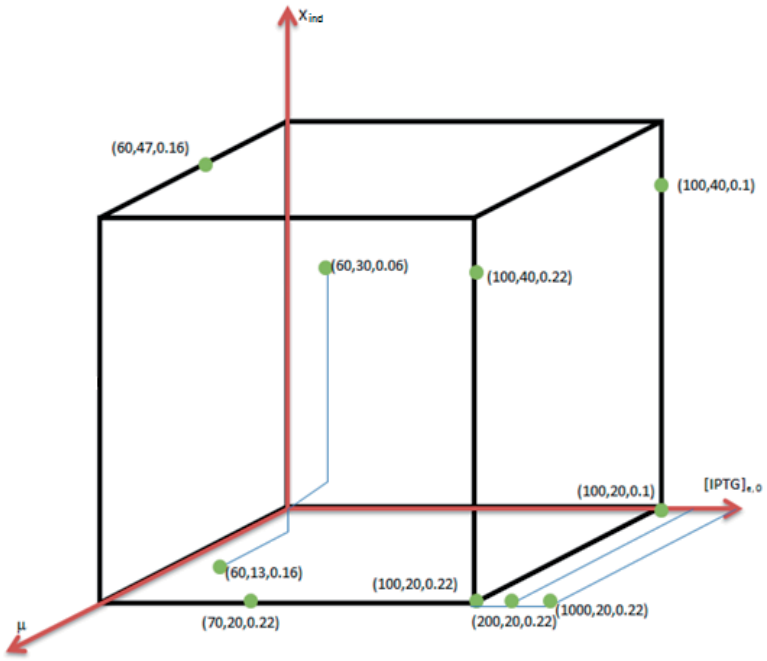


Figure 5.6. Experimental distribution for high induction zone. Nomenclature: $[IPTG]_{e,0}$, X_{ind} , μ_{fix}. (inducer concentration at induction time, biomass concentration at induction time, fed-batch specific growth rate at induction time)

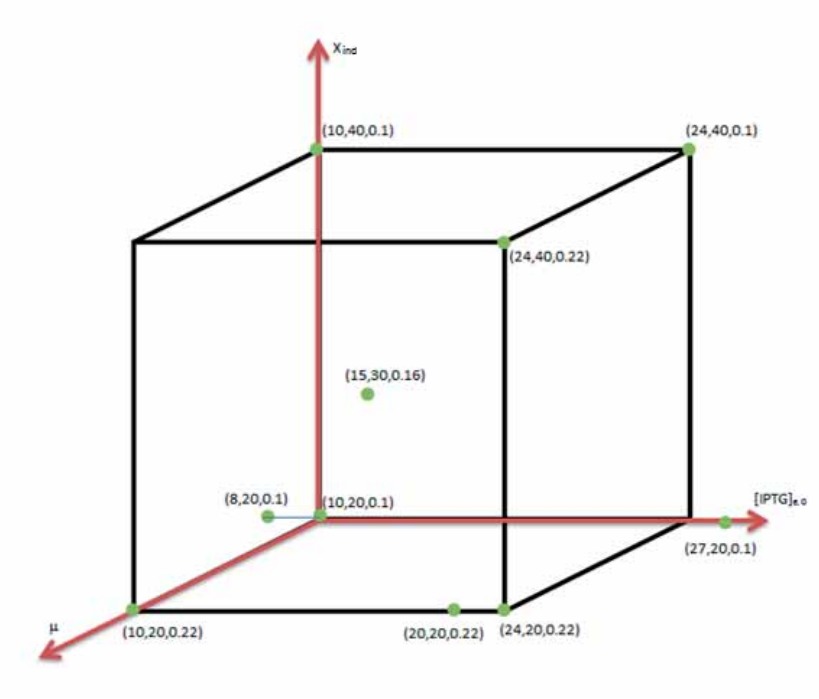


Figure 5.7. Experimental distribution for low induction zone. Nomenclature: $[IPTG]_{e,0}$, X_0 , μ_{fix}. (inducer concentration at induction time, biomass concentration at induction time, fed-batch specific growth rate at induction time)

Due to the different behavior, k' and K_M estimated values are considered to be different depending on the induction zone of the experiment. On the other hand, the effect of biomass concentration as well as the effect of the specific growth rate on transport rate will be assumed to be the same whatever the induction zone. For this reason, K_1 and K_2 estimated values are going to be the same for both zones. Moreover, the values of $k_c a$ found for *lac*Y mutant strain will be fixed according to Table 5.1, but the value for K' will be changed to K'' according to equations (5.10) and (5.11). It is important to say that this change will not increase the number of parameters in the estimation, while simply the value is being updated. All this assumptions can be used in a final expression of transport rate, shown in Equation (5.12)

$$r = 0.213([IPTG]_e - [IPTG]_i) + K_1 X_{ind}\left(K''[IPTG]_e + \frac{k'[IPTG]_e}{K_M + [IPTG]_i{}^2}\right)\exp\left(K_2\mu_{fix}\right) \qquad (5.12)$$

Values found for the parameters of both zones are shown in Table 5.2. Because there are common and specific parameters for the two zones, the estimation has been made at the same time with all the experiments. In function of the induction zone of every

experiment, the parameters for the specific active transport will be different, and they can be considered as different parameters. On the other hand, the parameters of biomass concentration contribution as well as specific growth rate will be the same in both zones, and they must be estimated at the same time for all the experiments.

Table 5.2. Estimated values for transport rate parameters in parent strain.

Parameter	Common Value	Value low induction zone	Value high induction zone	Units
k'	--	$0.800 \pm 1 \cdot 10^{-3}$	31±1	$\mu M^2 h^{-1}$
K_M	--	$4.9 \cdot 10^2 \pm 0.8 \cdot 10^2$	$2.4 \cdot 10^4 \pm 2 \cdot 10^3$	μM^2
K_1	$2.00 \pm 3 \cdot 10^{-2}$	--	--	$L \cdot g^{-1}$DCW
K_2	21.3±0.3	--	--	h

with the values of K_1 and K_2 from Table 5.2 is possible to calculate the value of K'',

using equation (5.11): $$K'' = \frac{0.0893}{20 \cdot 2.0 \cdot \exp(0.22 \cdot 21.3)} = 2.0 \cdot 10^{-5} \ \left(h^{-1}\right)$$

Figures 5.8 to 5.11 show the experimental data of $[IPTG]_e$, as well as the predicted values along time for both intracellular and extracellular IPTG concentrations.

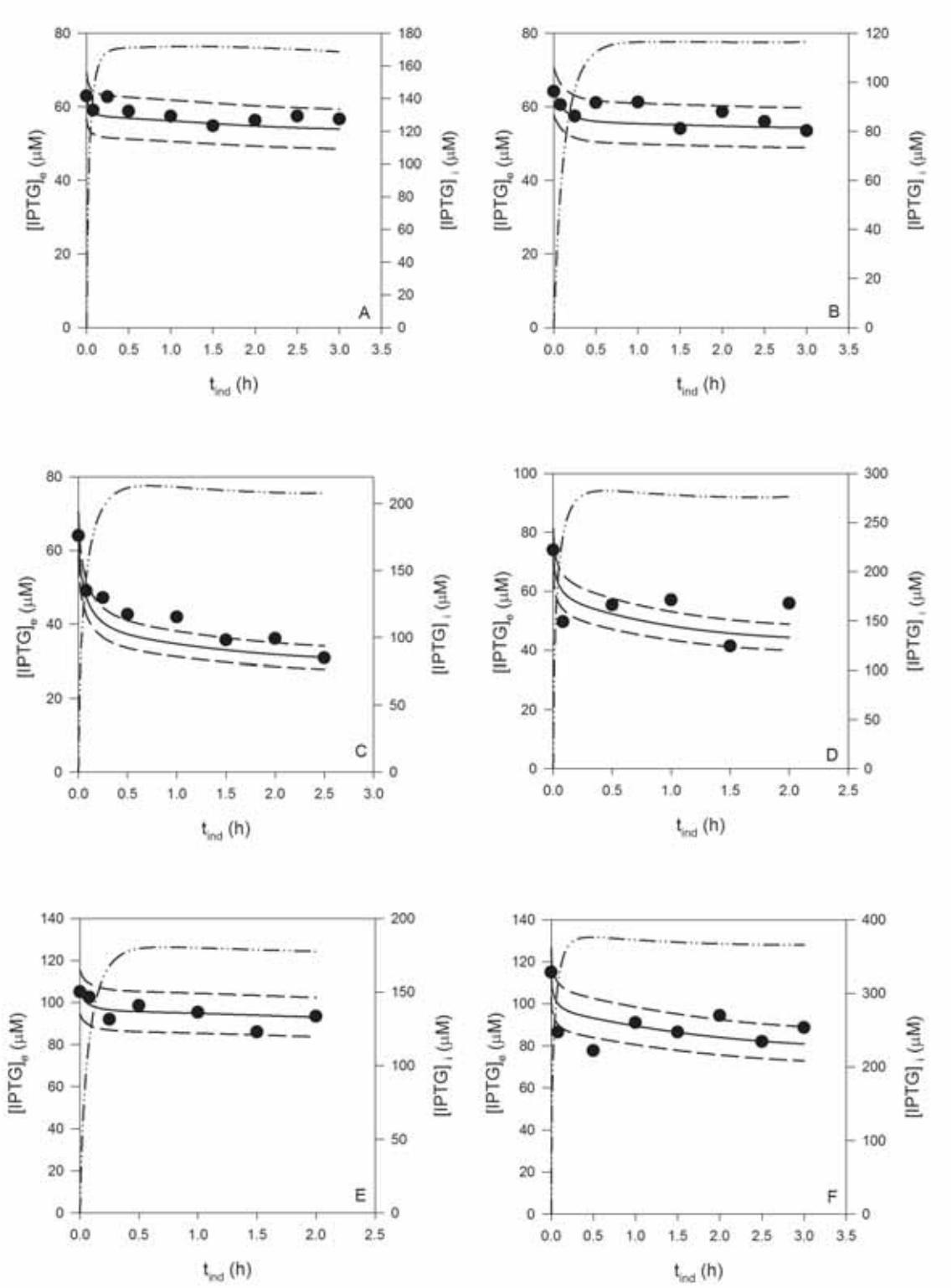


Figure 5.8. High induction model fitting for parent strain. (•) experimental data, (-)$[IPTG]_e$ model prediction, (- -) 10% error interval, (-···-)$[IPTG]_i$ model prediction. Experimental conditions: (A) 60μM, 13gDCWL^{-1}, 0.16h^{-1}; (B) 60μM, 30gDCWL^{-1}, 0.06h^{-1}; (C) 60μM, 47gDCWL^{-1}, 0.16h^{-1}; (D) 70μM, 20gDCWL^{-1}, 0.22h^{-1}; (E) 100μM, 20gDCWL^{-1}, 0.1h^{-1}; (F) 100μM, 20gDCWL^{-1}, 0.22h^{-1}. Nomenclature: $[IPTG]_{e,0}$, X_{ind} , μ_{fix}

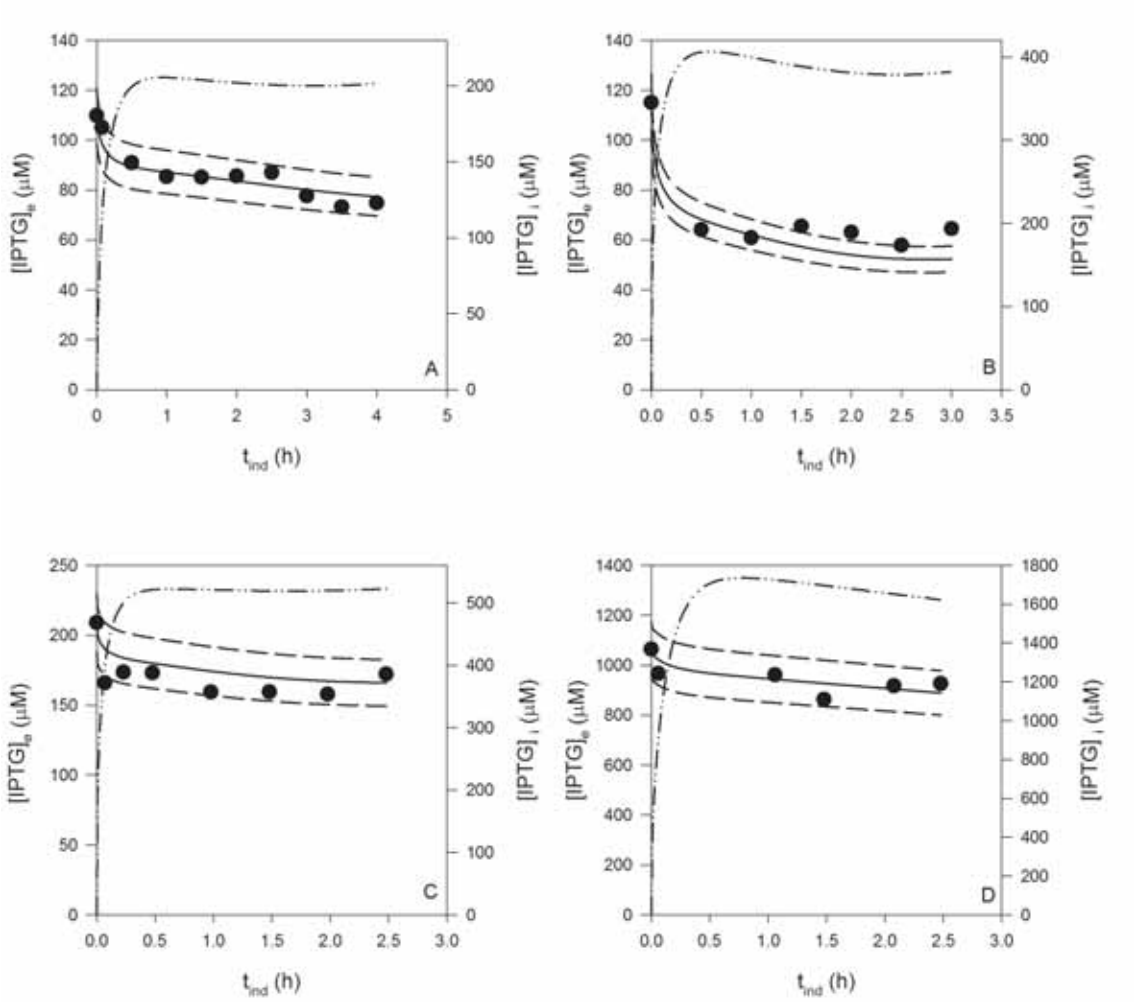


Figure 5.9. High induction model fitting for parent strain. (•) experimental data, (-)$[IPTG]_e$ model prediction, (- -) 10% error interval, (-···-)$[IPTG]_i$ model prediction. Experimental conditions: (A) 100μM, 40gDCWL^{-1}, 0.1h^{-1}; (B) 100μM, 40gDCWL^{-1}, 0.22h^{-1}; (C) 200μM, 20gDCWL^{-1}, 0.22h^{-1}; (D) 1000μM, 20gDCWL^{-1}, 0.22h^{-1}. Nomenclature: $[IPTG]_{e,0}$, X_{ind} , μ_{fix}

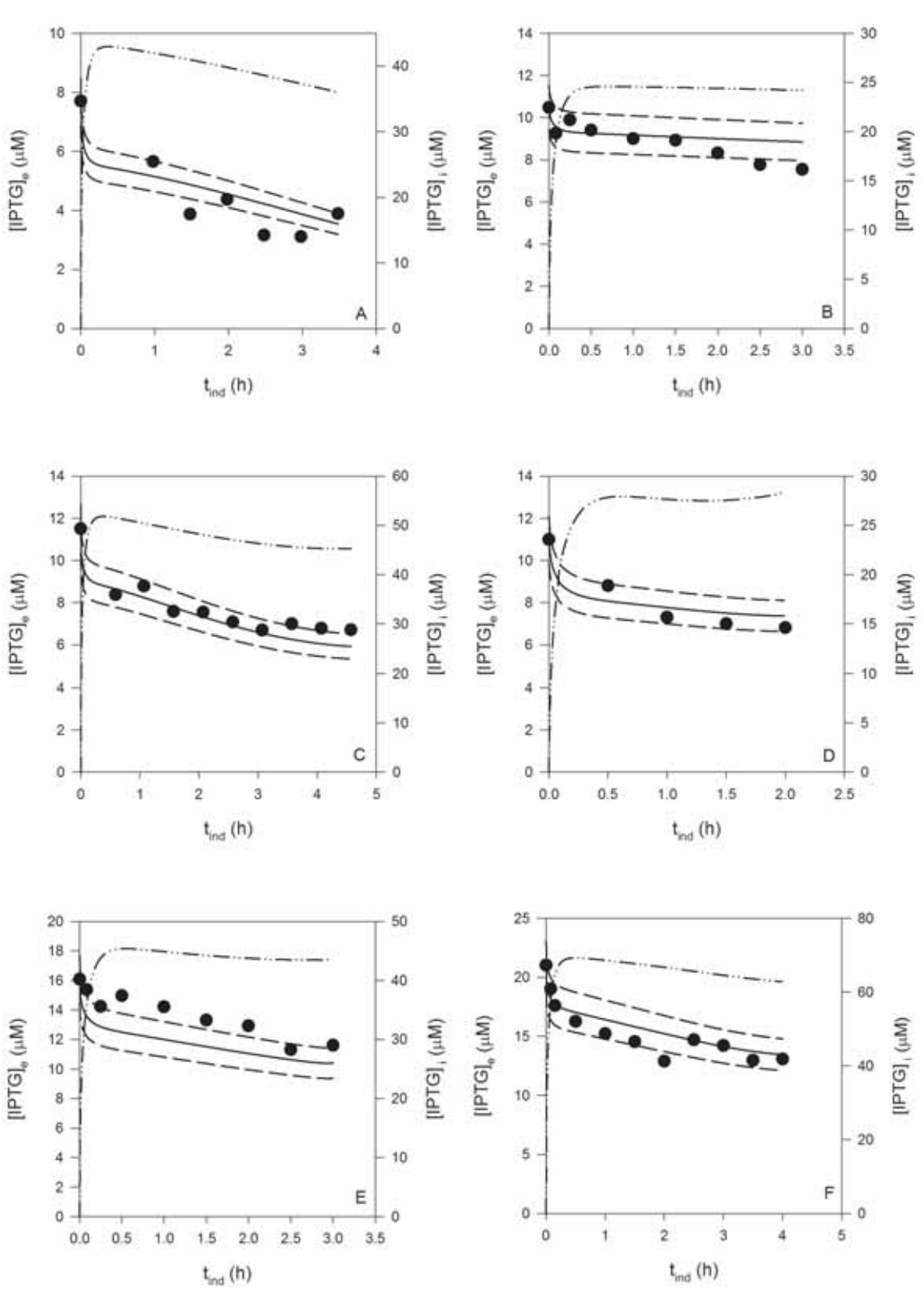


Figure 5.10. Low induction model fitting for parent strain. (•) experimental data, (-)[IPTG]$_e$ model prediction, (- -) 10% error interval, (···-)[IPTG]$_i$ model prediction. Experimental conditions: (A) 8μM, 20gDCWL^{-1}, 0.22h^{-1}; (B) 10μM, 20gDCWL^{-1}, 0.1h^{-1}; (C) 10μM, 20gDCWL^{-1}, 0.22h^{-1}; (D) 10μM, 40gDCWL^{-1}, 0.1h^{-1}; (E) 15μM, 30gDCWL^{-1}, 0.16h^{-1}; (A) 20μM, 20gDCWL^{-1}, 0.22h^{-1}. Nomenclature: [IPTG]$_{e,0}$, X_{ind} , μ_{fix}

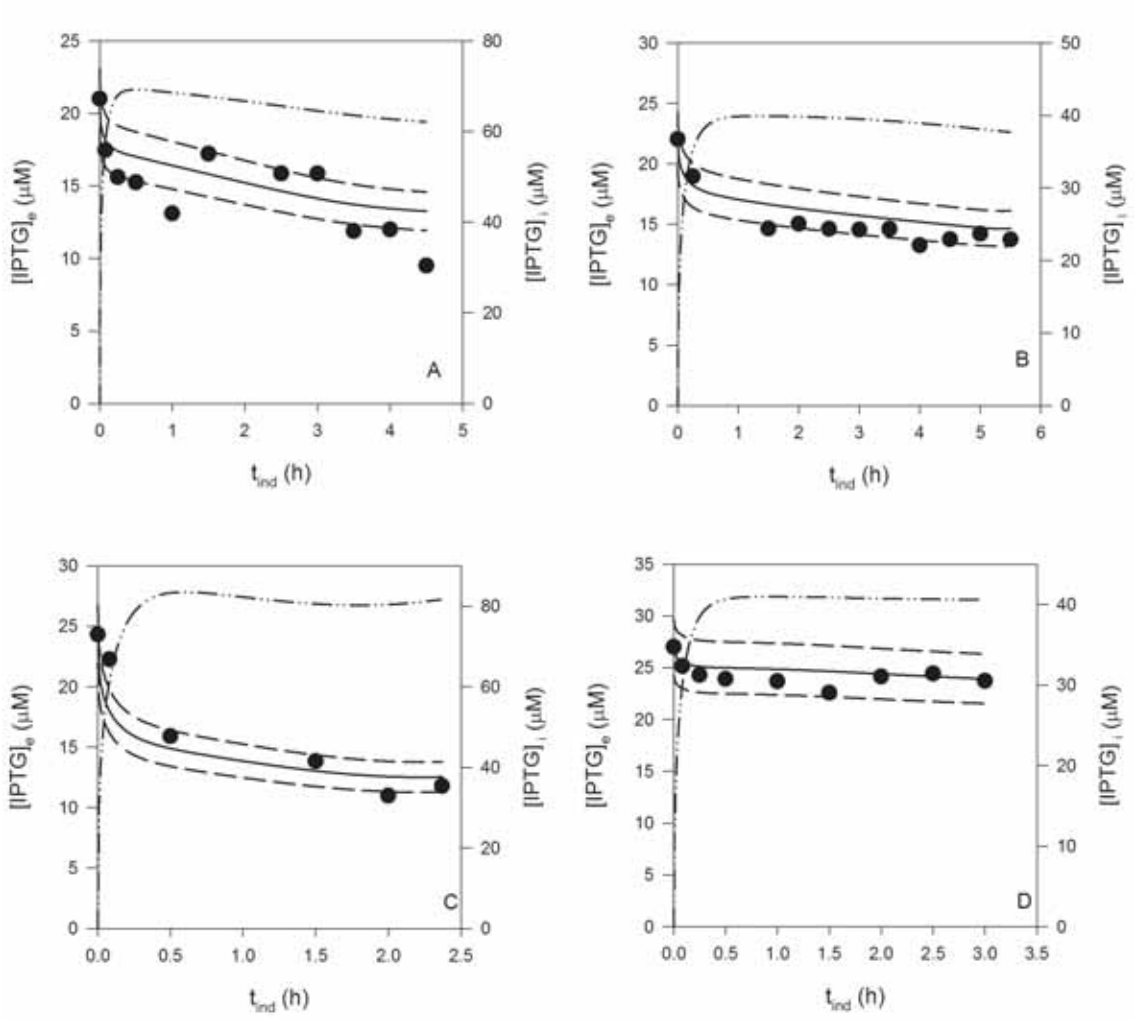


Figure 5.11. Low induction model fitting for parent strain. (•) experimental data, (-) $[IPTG]_e$ model prediction, (- -) 10% error interval, (-···-)$[IPTG]_i$ model prediction. Experimental conditions: (A) 24μM, 20gDCWL^{-1}, 0.22h^{-1}; (B) 24μM, 40gDCWL^{-1}, 0.1h^{-1}; (C) 24μM, 40gDCWL^{-1}, 0.22h^{-1}; (D) 27μM, 20gDCWL^{-1}, 0.1h^{-1}. Nomenclature: $[IPTG]_{e,0}$, X_{ind} , μ_{fix}

Figures 5.8 to 5.11 show that the model fitted, with its estimated parameters, is able to predict the extracellular IPTG concentration profile after induction within 10% of error for all the experiments. Figure 5.12 shows the parity plots for the high and low induction zones. This kind of representation allows taking an overview of the global fitting, comparing the experimental $[IPTG]_e$ with the values predicted by the model.

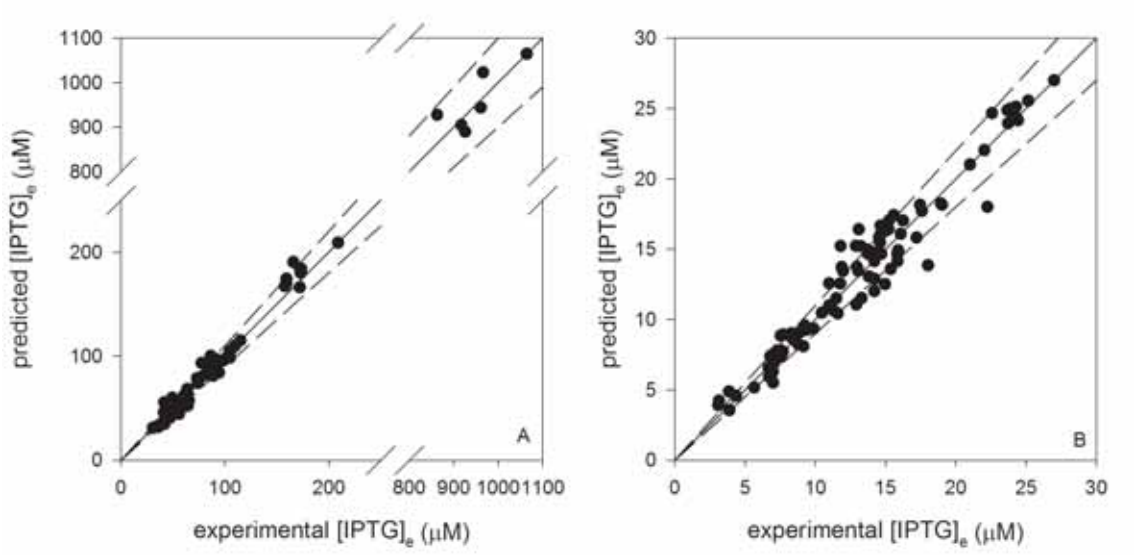


Figure 5.12. Parity plot with 10% error. (A) high induction zone. (B) low induction zone.

5.2.2.2 Overall model

Although a model divided into 2 different zones is able to predict the extracellular IPTG uptake for the experiments shown, a unique model –using a single set of parameters- would be significantly more useful. This overall model should be able to predict IPTG behavior in both zones.

In this way, as k' and K_M parameters have different values depending on, solely, the inducer concentration at induction, a relationship between their value and external IPTG concentration at induction point can be found. A Hill type sigmoidal equation (Goutelle et al., 2008) can be used to fit the value of the parameters depending on IPTG concentration at induction point ($[IPTG]_{e,0}$), allowing an overall prediction of IPTG concentration evolution with time, employing only the value of biomass concentration at induction, the specific growth rate and the inducer concentration at induction as input variables, as well as the experimental total volume.

Equation (5.13) corresponds to the general expression for a Hill's sigmoidal function, and its parameter values are shown in Table 5.3.

$$y = a + \frac{b \cdot \left([IPTG]_{e,0}\right)^c}{d^c + \left([IPTG]_{e,0}\right)^c} \tag{5.13}$$

where: y is the parameter K_1 or k'.

Table 5.3. Estimated values for the Hill function parameters as function of $[IPTG]_{e,0}$

Parameter	Value for k'	Value for K_M
a	$0.800\pm1\cdot10^{-3}$ (μM^2h^{-1})	$4.9\cdot10^2\pm0.8\cdot10^2$ (μM^2)
b	30 ± 1 (μM^2h^{-1})	$2.4\cdot10^4\pm2\cdot10^3$ (μM^2)
c	56.5±0.4 (--)	57.8±0.3 (--)
d	40.3±0.4 (μM)	40.3±0.4 (μM)

Figure 5.13 shows how Hill's sigmoidal is capable to give the right values for both K_M and k' parameters as function of extracellular IPTG concentration at induction time.

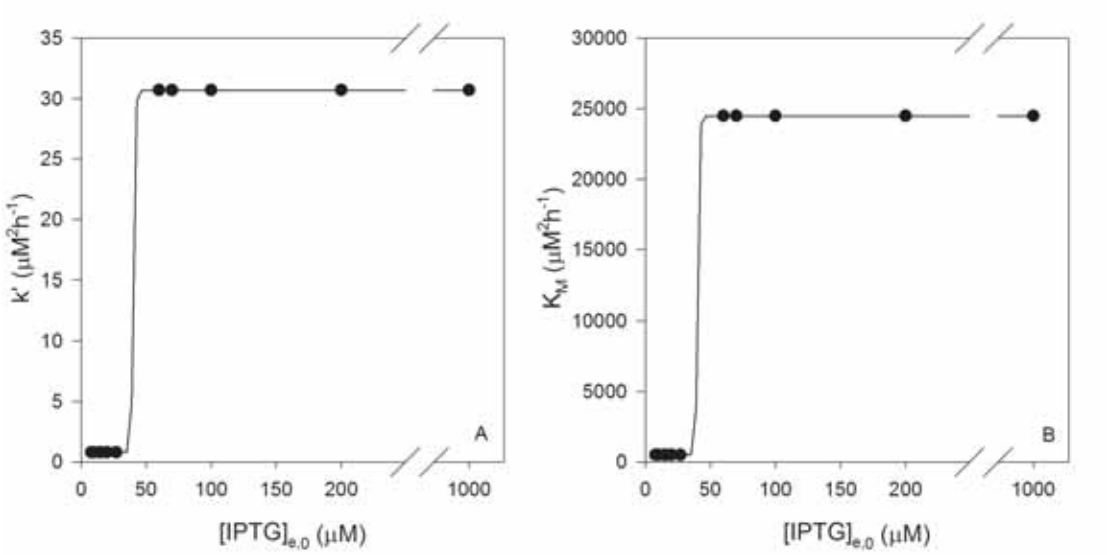


Figure 5.13. Hill's sigmoidal fitting for K_M and k' parameters as function of $[IPTG]_{e,0}$. (A) Adjust for k'; (B) Adjust for K_M.

Then, the final expression for the net transport rate is described in Equations (5.14) to (5.16)

$$r = 0.213([IPTG]_e - [IPTG]_i) + 2.0 \cdot X_{ind}\left(2.0 \cdot 10^{-5}[IPTG]_e + \frac{k'[IPTG]_e}{K_M + [IPTG]_i{}^2}\right)\exp\left(21.3\mu_{fix}\right) \quad (5.14)$$

$$k' = 0.8 + \frac{29.9 \cdot [IPTG]_{e,0}{}^{56.5}}{40.3^{56.5} + [IPTG]_{e,0}{}^{56.5}} \quad (5.15)$$

$$K_M = 489.6 + \frac{2.4 \cdot 10^4 \cdot [IPTG]_{e,0}{}^{57.8}}{40.3^{57.8} + [IPTG]_{e,0}{}^{57.8}} \quad (5.16)$$

According to the presented results, the model is capable to predict both fully induced (high induction) and partially induced (low induction) cultures. Parameter k' includes the lactose permeases transcription and production rates. Moreover, k' values from Table 5.2 suggest that fully induced cultures present a higher transcription rate than partially induced ones. This suggestion agrees with qualitative behavior of the *lac* operon described in literature (Fernández-Castané et al., 2012; Hansen et al., 1998; Jensen & Hammer, 1998; Jensen et al., 1993; Noel et al., 2009; Ozbudak et al., 2004).

5.3 Model validation

Once the model has been developed, fitted and calibrated, independent data must be used for its validation. In order to show that the model can be extrapolated, some different strains have been selected:

- M15 Δ*glyA* [pQEαβFucA][pREP4]. The same strain with expression of a different protein (Fuculose 1-Phosphate Aldolase, FucA).

- M15 [pQE-FucA][pREP4]. Parent strain without deletion of *glyA* gene and without the αβ termination sequence, expressing FucA.

- BL21 (DE3) FSA. Different strain producing a different aldolase (Fructose-6-Phosphate Aldolase, DHA dependent).

Induction concentration for FucA producing strains was 70 µM, while BL21 strain was induced with 100 µM of IPTG. All the experiments were induced the standard conditions described in Chapter 4 Chapter 4: when biomass concentration was 20 $gDCW \cdot L^{-1}$ and the specific growth rate during the fed-batch was $0.22h^{-1}$.

Using Equations (5.1) to (5.4) and Equations (5.14) to (5.16) -as IPTG transport rate-, both extracellular and intracellular inducer concentration along time were predicted, and the extracellular one was compared with experimental data.

Figure 5.14 shows the result of the simulation, and can be seen that the model is also able to predict reasonably well the behavior of different strains, producing different proteins. It is important to take into account that the value of the parameters is the same as for the M15 strain producing RhuA.

The last case has to be specially emphasized, because corresponds to a strain with different genetic background than the employed for calibration, and the model, employing the same parameters' values, was able to predict properly the experimental behavior.

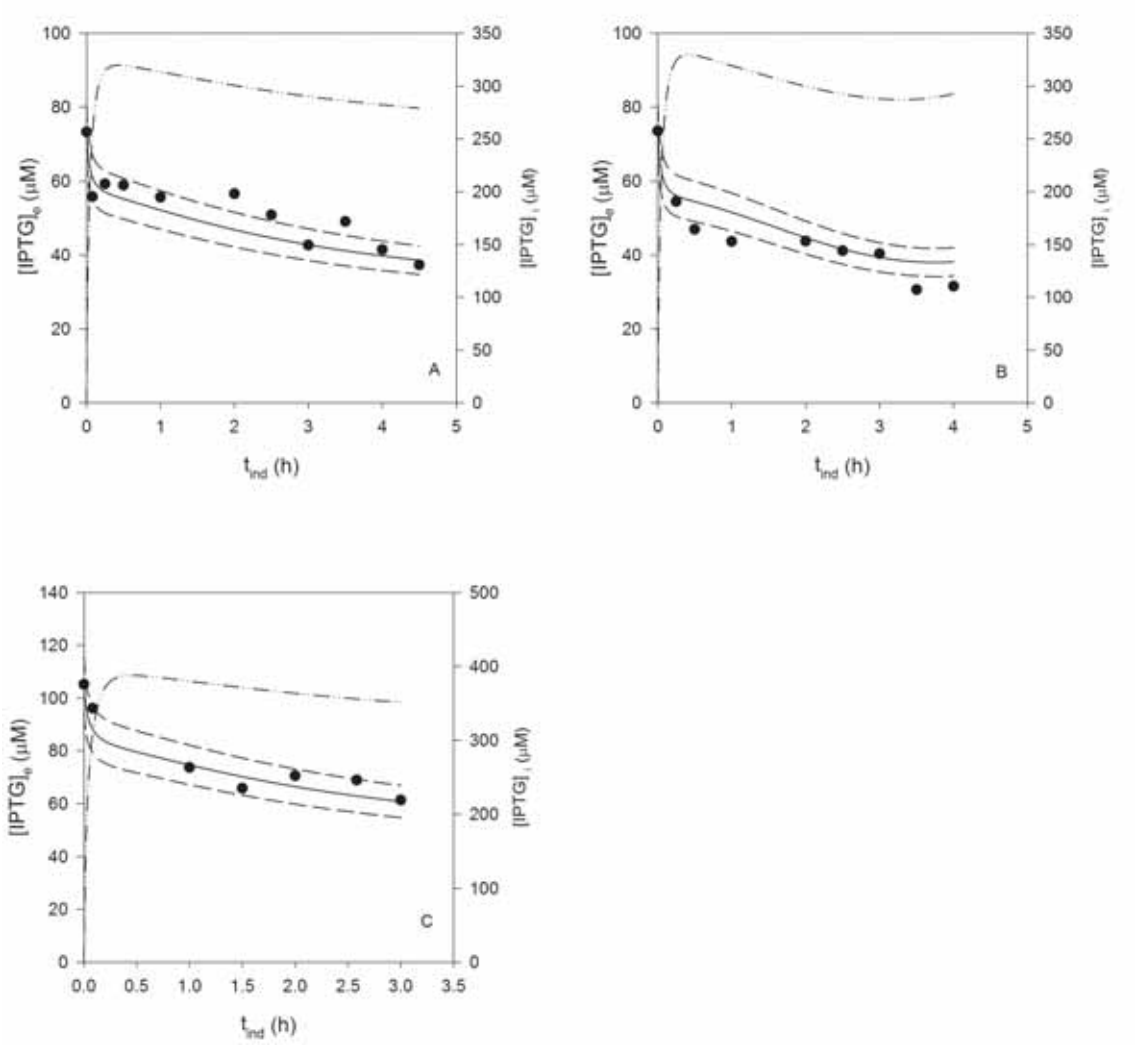


Figure 5.14. Model validation with other strains. (•) experimental data, (-) $[IPTG]_e$ model prediction, (- -) 10% error interval, (-···-)$[IPTG]_i$ model prediction. (A) M15 Δ*glyA* [pQEαβFucA][pREP4]; (B) M15 [pQE-FucA][pREP4]; (C) BL21 (DE3) FSA.

5.4 Conclusions

The unsegregated, unsteady, unstructured and based on first principles model proposed in this chapter has been calibrated with experimental data of extracellular IPTG concentration under real fed-batch fermentations, with a wide range of operational conditions of biomass, specific growth rate and IPTG concentration at induction time.

Firstly, the study of a *lac*Y mutant strain showed that, in absence of the specific transport protein for lactose, not only diffusion takes place in the transport, while

nonspecific active transport (mediated by other proteins) is needed to match the experimental data. Moreover, it has been experimentally observed that intracellular concentration can be higher than extracellular one. This fact cannot be explained if diffusion was the only transport mechanism.

On the other hand, parent strain showed faster initial IPTG transport rate, which can be explained adding another term to the net transport rate equation. This new term corresponds to the contribution of lactose permeases to the transport. Diffusion and nonspecific transport parameters have been fixed to be the same for both strains. Moreover, active transport terms are modified depending on the biomass concentration at induction and the specific growth rate during the fed-batch.

Parent strain shows two different behaviors depending on inducer concentration. The model is able to describe both behaviors using different values for lactose permeases' depending parameters.

The model has been validated using three different strains, producing different proteins with different expression systems. It has been demonstrated that the model can be extrapolated to other *E.coli* strains. Nevertheless, it is expected that other expression systems or strains may need additional calibration.

6 Modeling recombinant protein production. Overall model: biomass growth, IPTG uptake and protein production

As it has been seen in Chapter 4, IPTG dosage is a determinant factor in the overexpression of recombinant proteins (Donovan et al., 1996; Durany et al., 2004). Then, a model capable to predict its uptake has been developed in Chapter 5.

Although a lot of literature is available on heterologous protein production in *E.coli*, every expression system is different depending on the promoter, the interaction between the host and the expression vector, as well as the characteristics of the recombinant protein (Goyal, Sahni, & Sahoo, 2009).

Some mathematical models have been developed in an effort to describe systems employing promoters such as *lac* and the temperature inducible λ for the production of recombinant proteins in *E.coli* (Betenbaugh & Dhurjati, 1990; Nancib, Mosrati, & Boudrant, 1993; Tamerleret al., 2001; Toksoy Öner et al., 2003).

Cultivation conditions and media composition affect parameters such as plasmid copy number, the maintenance of the recombinant plasmid, the effect of protease activity on the recombinant protein of interest (Goyal et al., 2009; Lee, 1996; Yee & Blanch, 1992). Then, the production of recombinant protein in *E.coli* needs to take into account the study of those different cultivation conditions.

However, models found in literature only use the initial IPTG concentration added to the culture, because there was not any model of the IPTG uptake using experimental IPTG data for its calibration. Using only initial IPTG added to the culture medium, it is not possible to model properly how the repressor and the inducer are being bound.

Specifically, there is a model in literature (Ceroni et al., 2010) that describes the transcription of the *lac*I gene, producing repressor molecules. It is a perfect example of unsegregated and unstructured model: there is not any kind of cellular population difference, and it is supposed that all the reactions, production and degradation of the different components take place in the same compartment. Their approach includes equations for the gene transcription, as well as the expressions for the production of free repressor (not bound to IPTG) and repressor-IPTG molecules, as a function of the initial IPTG concentration added to the culture. This model allows the calculation of a

time evolution of the repressor-IPTG molecules that are going to be used in the protein production model. Moreover, it is described the production of a green fluorescence protein, making a relationship between the binding of IPTG and repressor with the production of recombinant proteins. However, model dependence with IPTG only is described by the initial dosage, instead of use the intracellular IPTG concentration along time.

The main aim in this chapter is to develop a model able to predict specific protein in mass and in activity units as function of time. This model should be able, coupled with biomass growth model (Ruiz et al., 2011) and IPTG uptake model (Chapter5), to describe specific protein production (in mass and in activity) in an exponential fed-batch culture, starting from defined initial conditions. The model should predict cell growth during non-induced batch and fed-batch growth (volume and concentration of biomass and substrate along time), as well as growth and protein production after induction.

6.1 Model development

The proposed model is an extension of the IPTG uptake model presented in Chapter 5. The new model will also be an unsteady, unsegregated, unstructured and based on first principles model composed by three compartments as depicted in the scheme shown in Figure 6.1. In the last compartment, the rate of specific protein production -expressed either in mass units (P) or activity units (U) per gram of cells- profile is calculated as a function of the intracellular inducer concentration ($[IPTG]_i$) as well as the repressor-inducer dynamics in the protein model compartment. The inducer uptake model compartment allows calculating the intracellular inducer profile. Otherwise, biomass concentration evolution -for non-induced and induced growth- is the output of the growth model compartment.

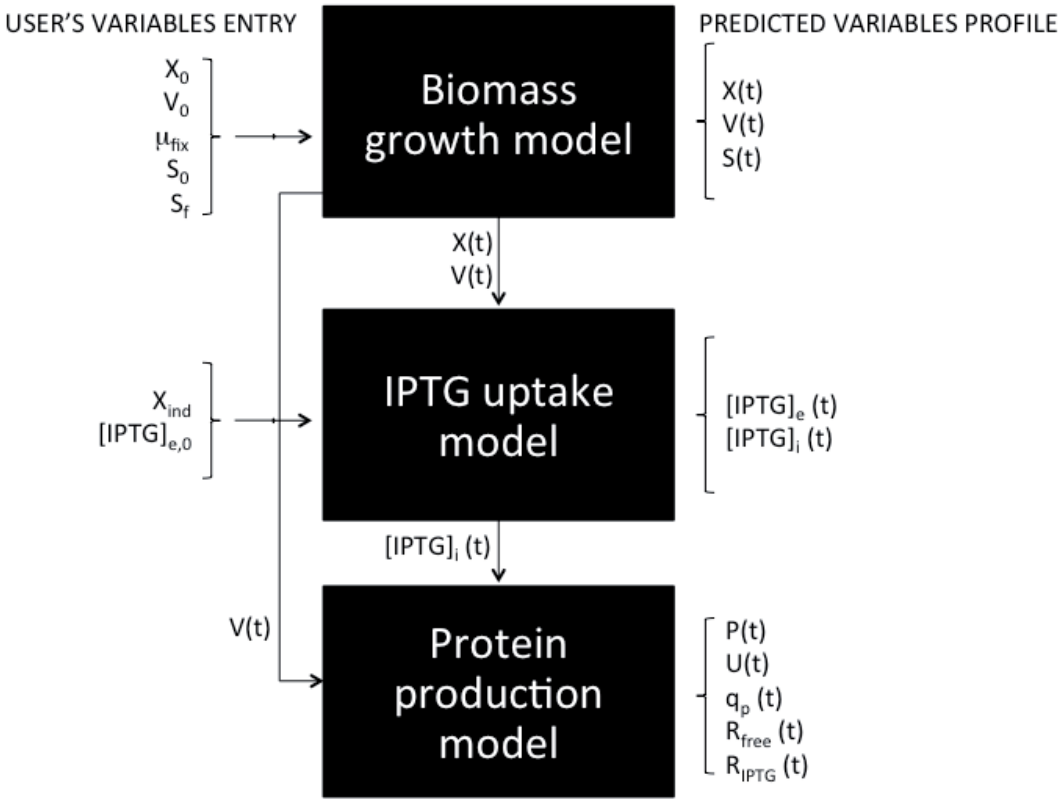


Figure 6.1. Modeling compartments diagram.

All the differential equations needed for describing the process, corresponding to each compartment, are presented in Tables 6.1 and 6.2, splitting the model into non-induced and induced phases. Moreover, the meaning of every variable is also explained in the text following the equations. The model describing biomass and substrate concentration evolution for both non-induced and induced growth was previously validated and reported (Ruiz et al., 2011). In summary, material balances for a fed-batch growth with predefined exponential feed addition are coupled with a prediction of the IPTG uptake rate and inducer extracellular and intracellular concentration along the time (Chapter 5). Once the intracellular IPTG concentration is predicted, it can be used for the calculation of the amount of repressor that can be bound to IPTG (R_{IPTG}), allowing the transcription of the target protein genes. The amount of repressor bound to IPTG will permit the estimation of the specific protein production rate in mass units (q_p), and is necessary in the mass balance for calculating

the specific protein production. Moreover, the specific activity of the recombinant protein will be predicted, using an average of the protein activity per mg of aldolase. Table 6.1 shows the non-induced biomass growth model.

Table 6.1. Model equations for non-induced phase of growth

BIOMASS GROWTH MODEL	
$\frac{dV}{dt} = F_s + F_B$	(6.1)
$F_s = \frac{1}{S_f}\left(\frac{\mu_{fix}}{Y_{XS}} + m_{SX}\right) \cdot X \cdot V$	(6.2)
$F_B = K_B \frac{dX}{dt}$	(6.3)
$\frac{dX}{dt} = \mu X - \frac{\frac{dV}{dt}X}{V}$	(6.4)
$\frac{dS}{dt} = \frac{\left(F_S \cdot S_f - S\frac{dV}{dt}\right)}{V} - \left(\frac{\mu}{Y_{XS}} + m_{sx}\right)X$	(6.5)
$\mu = \frac{\mu_{max} \cdot S}{K_s + S} exp\left(\frac{-S}{K_{iS}}\right)$	(6.6)

Non-induced Biomass Growth Model (Ruiz et al., 2011) needs as input the initial batch biomass concentration (X_0), the initial total volume into the bioreactor (V_0), the initial substrate concentration (S_0), the fixed specific growth rate for the fed-batch (μ_{fix}) and the glucose concentration in the feeding solution (S_f). Using Equations (6.1) to (6.6), a prediction of biomass growth as well as volume variation and substrate concentration along time is obtained for batch and fed-batch phases of the culture. Non-induced fed-batch phase starts when the initial glucose of the batch phase has been completely consumed. Induced phase starts when the culture reaches the desired biomass concentration (X_{ind}).

Equation (6.1) calculates the total volume (V, L) change as the sum of the feeding flow rate (F_s, $L \cdot h^{-1}$) and the added alkali flow rate (F_B, $L \cdot h^{-1}$)(pH control). In this case, in addition to the equation for volume variation described in literature, the alkali flow has been added.

Equation (6.2) represents the feeding flow rate (F_s, $L \cdot h^{-1}$), considering substrate-limiting fed-batch operation. It depends on glucose concentration in the feeding solution (S_f, $g \cdot L^{-1}$), the fixed specific growth rate (μ_{fix}, h^{-1}), the biomass-substrate yield (Y_{XS}, $gDCW \cdot g^{-1}$ glucose), the maintenance coefficient (m_{SX}, g glucose$\cdot g^{-1}DCW \cdot h^{-1}$), the biomass concentration (X, $gDCW \cdot L^{-1}$) and the total volume (V, L) into the bioreactor.

Equation (6.3) shows the time evolution of alkali flow rate (F_B, $L \cdot h^{-1}$) used in pH control, which depends on the biomass growth evolution, and it is a new equation added to literature's work.

Equation (6.4) indicates the biomass concentration (X, $gDCW \cdot L^{-1}$) variation along time as function of generation (growth) and dilution. It depends on the specific growth rate of the culture (μ, h^{-1}), biomass concentration, total volume and its rate of change.

Equation (6.5) is the substrate balance. It can be calculated as the amount added (depending on the feeding flow rate and its concentration) minus the amount that is being consumed by the culture and the dilution term.

Equation (6.6) calculates the specific growth rate of the culture with substrate inhibition represented by the exponential term modifying the Monod equation (Aiba equation). During the batch phase, the specific growth rate changes freely in function of the substrate concentration, but in the fed-batch phase it is maintained at μ_{fix} value by limiting the addition of substrate.

Table 6.2 shows the induced biomass growth model, the IPTG uptake model and the protein production model.

Table 6.2. Model equations for induced phase

BIOMASS GROWTH MODEL

$$\frac{dV}{dt} = F_S + F_B \quad (6.7)$$

$$F_S = \frac{1}{S_f}\left(\frac{\mu_{fix}}{(Y_{XS})_{ap}}\right) \cdot X \cdot V \quad (6.8)$$

$$F_B = K_B \frac{dX}{dt} \quad (6.9)$$

$$\frac{dX}{dt} = (\mu \cdot g_S)X - \frac{\frac{dV}{dt}X}{V} \quad (6.10)$$

$$\frac{dg_S}{dt} = -\frac{k_{s1}\left(\frac{[IPTG]_{e,0}}{X_{ind}} - 0.4\right)}{k_{s2} + \left(\frac{[IPTG]_{e,0}}{X_{ind}} - 0.4\right)} g_S, \quad g_S(0) = 1 \quad (6.11)$$

$$\frac{dS}{dt} = \frac{\left(F_S \cdot S_f - S\frac{dV}{dt}\right)}{V} - \left(\frac{\mu}{(Y_{XS})_{ap}}\right)X \quad (6.12)$$

$$\mu = \frac{\mu_{max} \cdot S}{K_s + S} exp\left(\frac{-S}{K_{iS}}\right) \quad (6.13)$$

IPTG UPTAKE MODEL

$$\frac{dV_{cel}}{dt} = 0.0023 \frac{d(XV)}{dt} \quad (6.14)$$

$$\frac{dV_m}{dt} = \frac{dV}{dt} - \frac{dV_{cel}}{dt} \quad (6.15)$$

$$\frac{d[IPTG]_e}{dt} = \frac{-r \cdot V_m - [IPTG]_e \frac{dV_m}{dt}}{V_m} \quad (6.16)$$

$$\frac{d[IPTG]_i}{dt} = \frac{r \cdot V_m - [IPTG]_i \frac{dV_{cel}}{dt}}{V_{cel}} \quad (6.17)$$

$$r = k_c a([IPTG]_e - [IPTG]_i) + K_1 X_{ind}\left(K''[IPTG]_e + \frac{k'\ [IPTG]_e}{K_M + [IPTG]_i^2}\right) exp(K_2 \mu_{fix}) \quad (6.18)$$

PROTEIN PRODUCTION MODEL

$$\frac{dM}{dt} = \alpha_{mRNA} N - \lambda_{mRNA} M \quad (6.20)$$

$$\frac{dR_{free}}{dt} = \frac{-1}{\tau_{R-O}}\left[R_{free}\left(\frac{[IPTG]_i/_{1000}}{K_{R-IPTG}}\right)^n - R_{IPTG}\right] - \lambda_R R_{free} + \alpha_R M \quad (6.21)$$

$$\frac{dR_{IPTG}}{dt} = \frac{1}{\tau_{R-O}}\left[R_{free}\left(\frac{[IPTG]_i/_{1000}}{K_{R-IPTG}}\right)^n - R_{IPTG}\right] - \lambda_R R_{IPTG} \quad (6.22)$$

$$\frac{dq_p}{dt} = -\frac{K_{P1}\left(\frac{dR_{IPTG}}{dt}\right)}{1 + K_{P2} P} q_p \quad (6.23)$$

$$q_{p0=}\left(33.80 + \frac{658.20 \cdot \frac{[IPTG]_{e,0}}{X_{ind}}}{41.48 + \frac{[IPTG]_{e,0}}{X_{ind}}}\right) \exp(3.73\mu_{fix}) \quad (6.24)$$

$$\frac{dP}{dt} = q_p - \frac{P\frac{dV}{dt}}{V} \quad (6.25)$$

$$\frac{dU}{dt} = K_q \frac{dP}{dt} \quad (6.26)$$

Induced Biomass Growth Model (Ruiz et al., 2011) takes as starting point the values obtained from the non-induced model, when biomass reaches the desired concentration (X_{ind}).

Equation (6.7) represents the total volume (V, L) variation, calculated on the same way as Equation (6.1). Again, the contribution of the alkali flow has been added.

Equation (6.8) differs from Equation (6.2) in the use of an apparent biomass-substrate yield ($(Y_{XS})_{ap}$, gDCW·g^{-1} glucose) which is different because it contains the maintenance coefficient.

Equation (6.9) has the same form as Equation (6.3), determining the time evolution of the alkali flow rate used in pH control.

Equation (6.10) represents the biomass concentration evolution with time, but, in this case, the specific growth rate decreases due to the metabolic burden caused by the overexpression of recombinant protein (Lee & Ramirez, 1992b). This effect could be described using a function g_s, through Equation (6.11).

Equation (6.11) describes the nonspecific effect of the induction on g_s and, consequently, on specific growth rate. It depends on the relationship between the amount of inducer ($[IPTG]_{e,0}$) added and the biomass concentration at induction point (X_{ind}), being 0.4 µmol. gDCW^{-1} the minimum necessary ratio to trigger protein overexpression (Ruiz et al., 2011).

Equation (6.12) is the mass balance for the substrate, and Equation (6.13) allows the calculation of the specific growth rate.

IPTG Uptake Model. IPTG uptake model from Chapter 5 requires as inputs the prediction of biomass concentration and volume variation with time as well as new user's variables: biomass and IPTG concentration at induction time.

Equation (6.14) can predict the total volume of cells (V_{cel}, L_{cell}) in the culture, using the specific cell volume (0.0023 $L_{cell} \cdot g^{-1}$DCW)(Bennett et al., 2008).

Equation (6.15) is able to predict the evolution of the medium volume (V_m, L_m) as the difference between the total volume in the bioreactor and the total volume of cells.

Equations (6.16) and (6.17) are the material balances for IPTG (e:extracellular; i:intracellular, µM) as a function of the IPTG transport rate (r, $\mu M \cdot h^{-1}$) and a dilution term.

Equation (6.18) represents the IPTG transport rate. As seen in Chapter 5, an overall model can be used to describe transport phenomena in IPTG uptake, only as a function of $[IPTG]_{e,0}$, X_{ind} and μ_{fix}. Then, IPTG transport rate can be described by a diffusion term, non-specific active transport and specific active transport, taking into account the effects on transport of the biomass concentration at induction and the fixed specific growth rate.

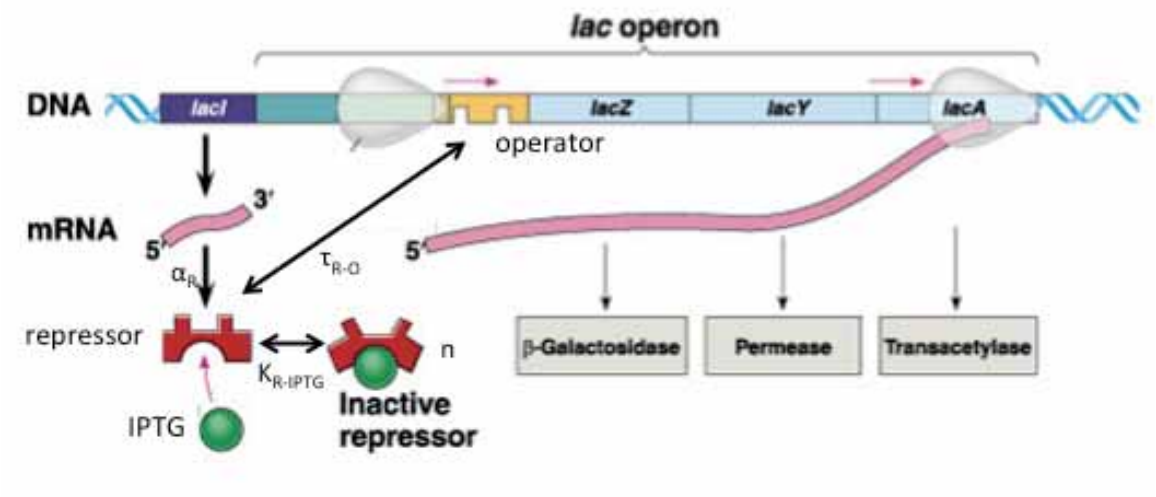
Using Equations (6.14) to (6.18) it is possible to calculate the intracellular IPTG concentration evolution along time, which is a key variable in protein production. Intracellular IPTG molecules can bind the repressor, making RNA polymerase able to transcribe the target gene and producing RhuA in this case.

Protein Production Model has to be able to predict specific protein both in mass and activity units. This model has two different parts. Firstly, using an adaptation of Ceroni's model (Ceroni et al., 2010) (as explained below) it is possible to predict the evolution with time of the concentration of repressor molecules free or bound to IPTG. Once the culture is induced, intracellular IPTG molecules could bind the repressor allowing the transcription of the gene of interest.

Equation (6.20) calculates the number of LacI mRNA molecules (M, molecules·cell^{-1}) into the cell along the time. For its calculation, the transcription rate (α_{mRNA}, h^{-1}), pREP4 plasmid (containing *lac*I gene) copy number (N, plasmid·cell^{-1}) –that in fact is the number of copies of the gene - and the degradation rate of LacI mRNA molecules (λ_{mRNA}, h^{-1}) are needed.

Equation (6.21) is used to calculate the amount of free repressor (R_{free}, molecule·cell^{-1}).

Equation (6.22) predicts the amount of repressor bounded to IPTG (R_{IPTG}, molecule·cell^{-1}) using the same variables than Equation (6.21).



Scheme 7.1. *lac* operon with the kinetic and equilibrium parameters.

For its calculation (see Scheme 6.1 for a general picture of the system), the values found in literature (Ceroni et al., 2010) for the IPTG-LacI equilibrium binding constant ($K_{R\text{-}IPTG}$, mM), the protein degradation rate (λ_R, h^{-1}), the protein transcription rate (α_R, h^{-1}) and the cooperativity of the binding LacI-IPTG (n) have been used because they are dependent of the interaction between IPTG and repressor. Otherwise, the value for the time constant of LacI binding to the operator ($\tau_{R\text{-}O}$, h^{-1}) will be a model-fitting parameter because it depends on the operator used in the particular expression system and has to be fitted to experimental data.

Moreover, as it is possible to calculate the profile of intracellular concentration of IPTG along time from Equation 6.17, repressor concentration (both free and IPTG bound) will be calculated using the actual intracellular IPTG concentration and Equations (6.21) and (6.22) will use $[IPTG]_i$ predicted values instead of the initial amount of inducer proposed in Ceroni's work.

Equation (6.23) describes the experimental specific protein production in mass units rate (q_p, $mgRhuA \cdot g^{-1}DCW \cdot h^{-1}$) exponential decay with time, which is observed in all the induced growths. It is illustrated for an experimental run in Figure 6.2. Moreover, specific protein production rate will be a function of the rate of change of repressor bound to inducer (R_{IPTG}) calculated by Equation (6.22).

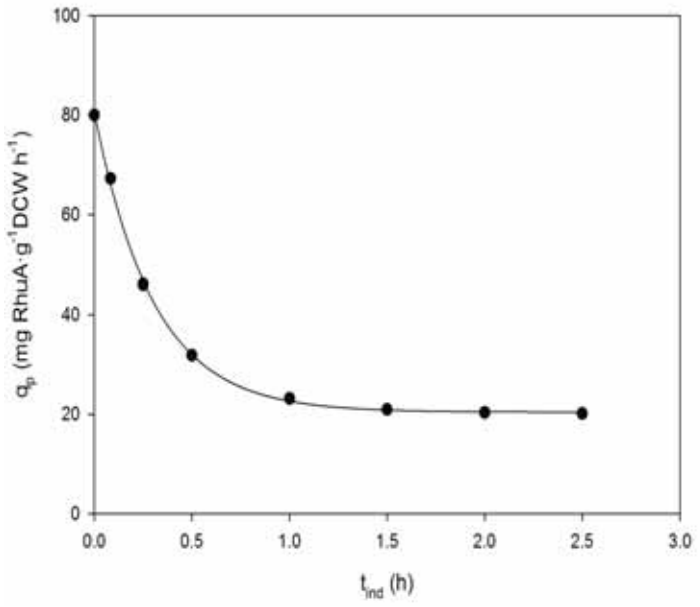


Figure 6.2. Example of experimental specific protein production rate exponential decay along time

Equation (6.24) predicts the initial specific production rate (q_{p0}, mgRhuA·g^{-1}DCW·h^{-1}) as a function of the experimental conditions ($[IPTG]_{e,0}$ to X_{ind} ratio and μ_{fix}) of each cultivation. On one hand, the dependence with growth rate presents an exponential trend. On the other hand, a dependence with inducer to biomass concentration ratio ($[IPTG]_{e,0}/X_{ind}$) at induction moment and specific growth rate has been experimentally found. In fact, $[IPTG]_{e,0}/X_{ind}$ has been previously discussed as key parameter in inducer transport and protein production in Chapter 4. The dependence on inducer to biomass concentration ratio at induction time can be explained as shown in Figure 6.3A. Using the fitting in Figure 6.3 (A) and the exponential dependence with specific growth rate, initial specific protein production rate can be described using Equation (6.24). Figure 6.3B shows the parity plot for initial specific protein production, in order to check the goodness of the Equation (6.24).
The value for the parameters is going to be presented in Table 6.4.

Equation (6.25) shows the mass balance for the specific protein in mass (P, mgRhuA·g^{-1}DCW), as a function of the specific protein production rate and the dilution term as a function of the total volume.

Equation (6.26) can predict the specific protein in activity (U, Activity units·g^{-1}DCW) using the specific activity per mass unit of aldolase as a measure of the protein quality (K_q, Activity Units·mg^{-1}RhuA). Protein quality varies along time, and data for all productive fed-batch are presented in Figure 6.4 (experiments presented in Section 6.2), as a function of the inducer to biomass ratio. As can be seen in Figure 6.4, the value for the protein specific quality varies between 5 and 7 AU·mg^{-1}RhuA for the range of I/X for all the experiments. For modeling purposes, it has been considered a constant parameter to be fitted.

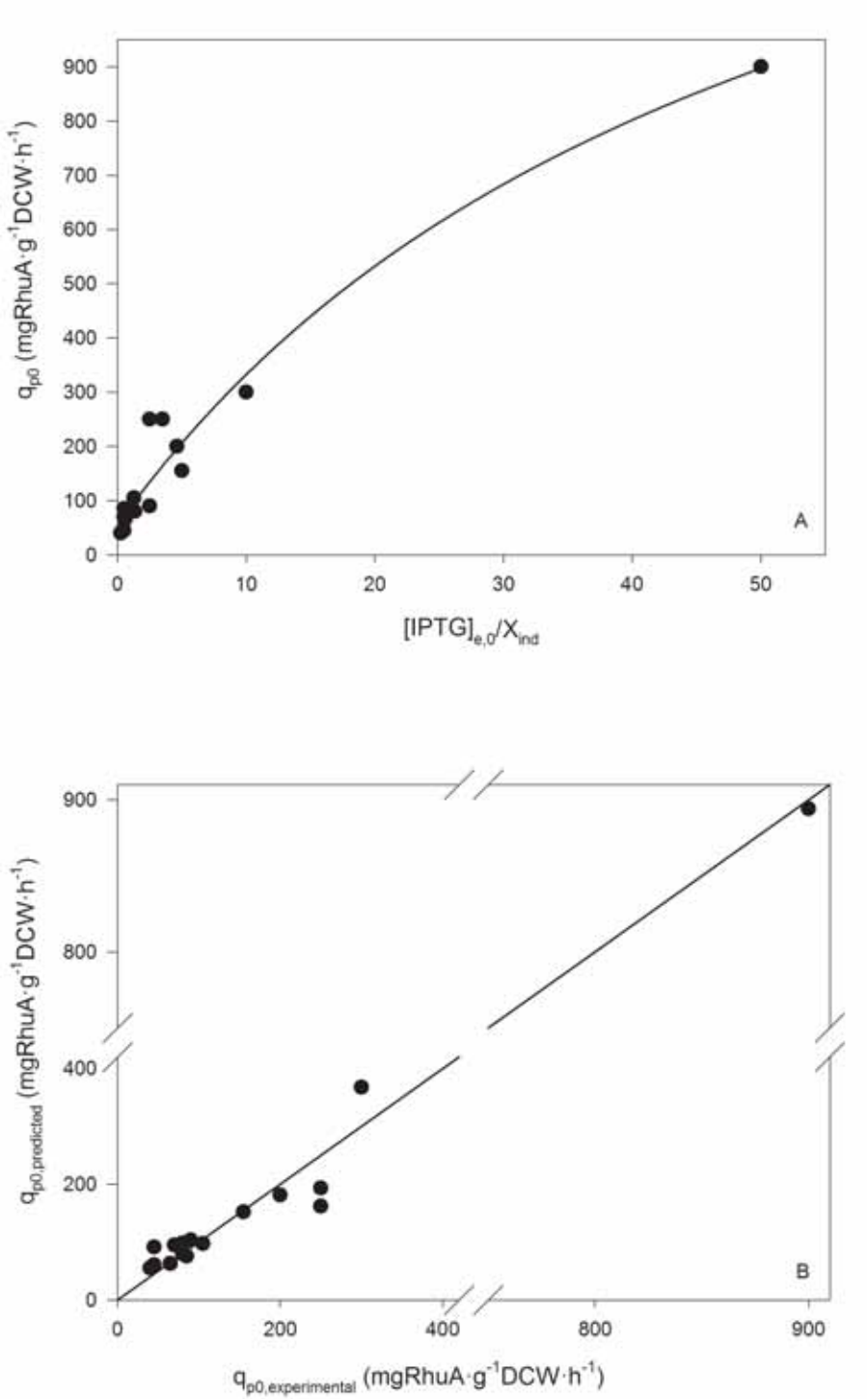


Figure 6.3. (A) initial specific protein production rate dependence with inducer to biomass ratio. (B) parity plot of initial specific protein production rate.

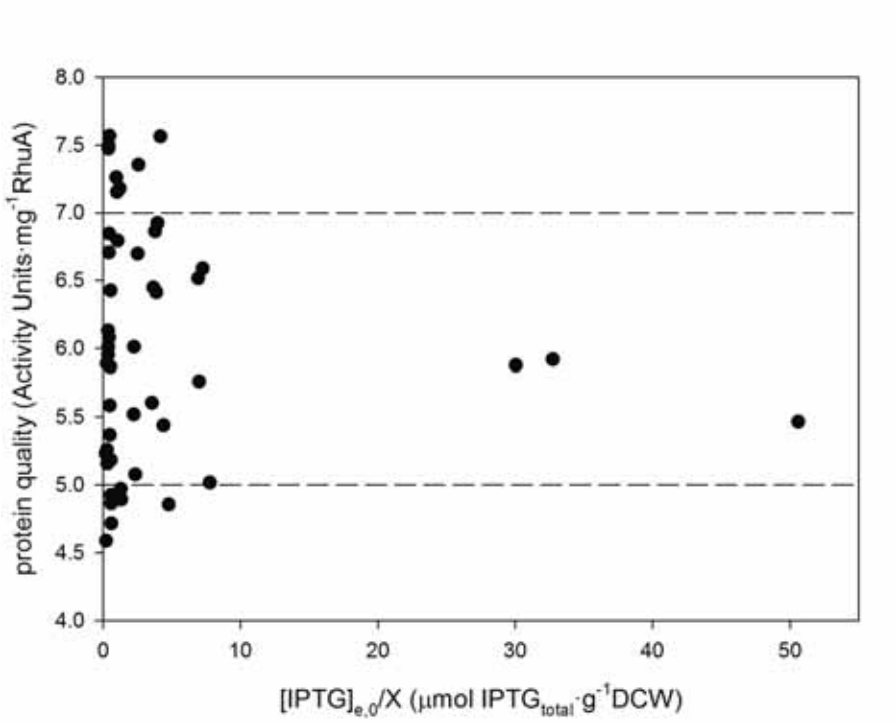


Figure 6.4. Experimental protein specific quality distribution in function of the inducer to biomass ratio.

6.2 Model fitting

A set of 15 *E.coli* fed-batch experiments has been employed to get experimental data for the recombinant Rhamnulose-1-Phosphate aldolase production process, monitoring their performance by determining specific protein in mass and in activity along time. These 15 experiments are a selection from the 20 different experiments used for the IPTG uptake model. This selection was performed taken into account the availability of reliable data about aldolase concentration and activity. To give the model the capacity to predict the evolution with time of specific protein in mass and in activity units, a wide range of experimental conditions must be used in the fitting process, expecting any experiment located into the defined space to follow the prediction of the model. As indicated in Chapter 5, the selected experimental variables –because they are key in protein production (Ruiz et al., 2011)- were inducer concentration and biomass concentration at induction time and the specific growth rate in the fed-batch phase of the culture, The range of operational conditions used in

this work is the one selected for the IPTG uptake model, and it is reminded in Table 6.3. As explained in Chapter 5, inducer range goes from 8μM (being 4μM the minimum concentration for RhuA overexpression) (Pinsach, 2009) to 1mM. Biomass concentration at induction time was up to 47 gDCW·L^{-1}, which is the maximum one allowing a final biomass concentration compatible with the aeration capacity of the fermentor. Finally, specific growth rate ranged from 0.1 to 0.22 h^{-1}, avoiding excessive acetic acid production (Ruiz et al., 2009).

Table 6.3. Operational conditions range

Operational variable	Range
$[IPTG]_{e,0}$ (μM)	$8 \leq [IPTG]_{e,0} \leq 1000$
X_{ind} (gDCW·L^{-1})	$13 \leq X_{ind} \leq 47$
μ_{fix} (h^{-1})	$0.1 \leq \mu_{fix} \leq 0.22$

The model proposed through equations (6.20) to (6.26) has only 4 unknown parameters that should be fitted. K_{P1} and K_{P2} are related to specific protein production in mass units, K_q is needed to describe the specific activity of the protein produced, and τ_{R-O}, the time constant for the linkage between the repressor and the operator, because it depends directly on the expression system, or, more exactly, on the operator-repressor bond. Then, depending on the operator used in the expression of the protein, a new time constant for the union between that operator and the repressor will be needed.

The values for all the other parameters of the models presented in Tables 6.1 and 6.2 -fixed parameters- are presented in Table 6.4.

Table 6.4. Values for the fixed parameters of the model. [*](Ruiz et al., 2011); [§] (Chapter 5); [^] (Pinsach, 2009); [⌘](Ceroni et al., 2010)

Parameter	Value	Units
Y_{XS}*	0.38	gDCW·g^{-1}glucose
m_{XS}*	<1·10^{-4}, taken as zero in the simulations	g glucose·g^{-1}DCW·h^{-1}
μ_{max}*	0.55	h^{-1}
K_S*	8.38·10^{-3}	g glucose·L^{-1}
K_{iS}*	46.9	g glucose·L^{-1}
$(Y_{XS})_{ap}$*	0.47	gDCW·g^{-1}glucose
k_{S1}*	0.50	h^{-1}
k_{S2}*	5.73·10^{-2}	µmol IPTG·g^{-1}DCW
K_B	BATCH: 4.3·10^{-3} FED-BATCH: 2·10^{-3} INDUCED PHASE: 4·10^{-6}	L·(gDCW·L^{-1})$^{-1}$
k_ca§	0.213±2·10^{-3}	h^{-1}
K''§	2·10^{-5}	h^{-1}
K_1§	2.00±3·10^{-2}	L·g^{-1}DCW
K_2§	21.3±0.3	h
k'§	$0.8 + \frac{29.9 \cdot [IPTG]_{e,0}{}^{56.5}}{40.3^{56.5} + [IPTG]_{e,0}{}^{56.5}}$	µM^2·h^{-1}
K_M§	$489.6 + \frac{2.4 \cdot 10^4 \cdot [IPTG]_{e,0}{}^{57.8}}{40.3^{57.8} + [IPTG]_{e,0}{}^{57.8}}$	µM^2
N^	12	plasmid·cell^{-1}
α_{mRNA}⌘	27	h^{-1}
λ_{mRNA}⌘	16.26	h^{-1}
α_R⌘	172.8	h^{-1}
λ_R⌘	1.284	h^{-1}
$K_{R\text{-}IPTG}$⌘	0.289	mM
n⌘	0.6	--

Some parameter values are reported without associated error, reproduced as indicated literature.

After fitting the model to the experimental data, the values found for the four parameters are those shown in Table 6.5.

Table 6.5. Parameter's estimated values for specific protein production and specific protein activity model.

Parameter	Value	Units
τ_{R-O}	6.30±0.07	h
K_{P1}	$8.01 \cdot 10^{-2} \pm 2 \cdot 10^{-4}$	cell·molecule^{-1}h^{-1}
K_{P2}	$1.311 \cdot 10^{-1} \pm 5 \cdot 10^{-4}$	gDCW·mg^{-1}RhuA
K_q	$6.1300 \pm 2 \cdot 10^{-4}$	AU·mg^{-1}RhuA

As can be seen in Table 6.5, the estimated specific activity per mg of aldolase, K_q value, is 6.13 AU.mg^{-1}, corresponding to the mean of the values of protein quality for all the experiments (Figure 6.4).

Figures 6.5 to 6.7 show the experimental data of specific protein production in mass units, as well as the predicted values along time using the model described through Equations (6.14) to (6.26). Growth model has no been used for the calibration of the protein production model. For this reason, total volume and biomass concentration are experimental values and they are adjusted to a time-dependent spline function.

Otherwise, Figures 6.8 to 6.10 show the experimental specific protein in activity units and its prediction using the model.

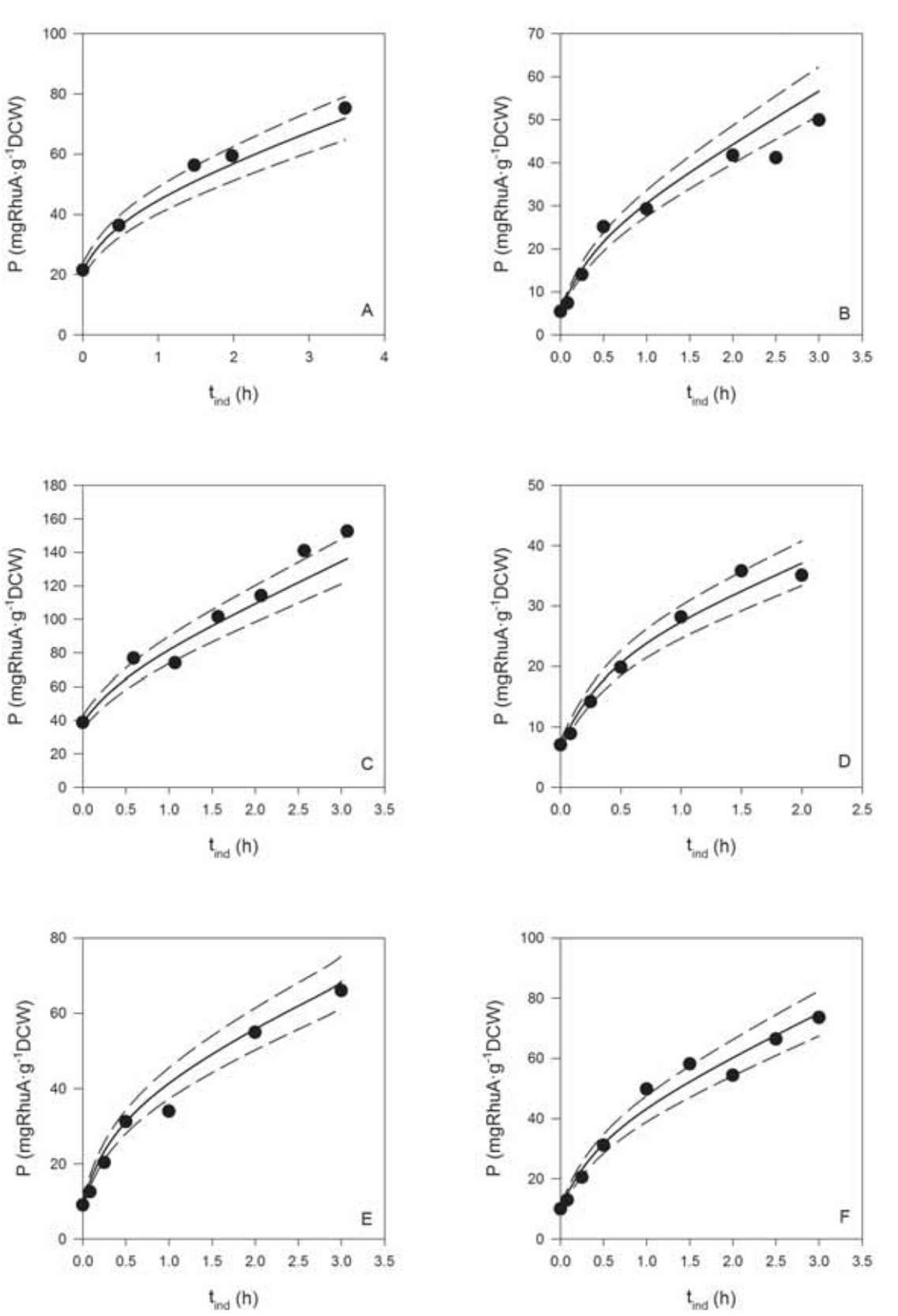


Figure 6.5. Specific protein in mass units model fitting. (•) experimental data, (-)model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively: (A) 8μM, 20gDCWL^{-1}, 0.22h^{-1}; (B) 10μM, 20gDCWL^{-1}, 0.1h^{-1}; (C) 10μM, 20gDCWL^{-1}, 0.22h^{-1}; (D) 10μM, 40gDCWL^{-1}, 0.1h^{-1}; (E) 15μM, 30gDCWL^{-1}, 0.16h^{-1}; (F) 24μM, 40gDCWL^{-1}, 0.1h^{-1}.

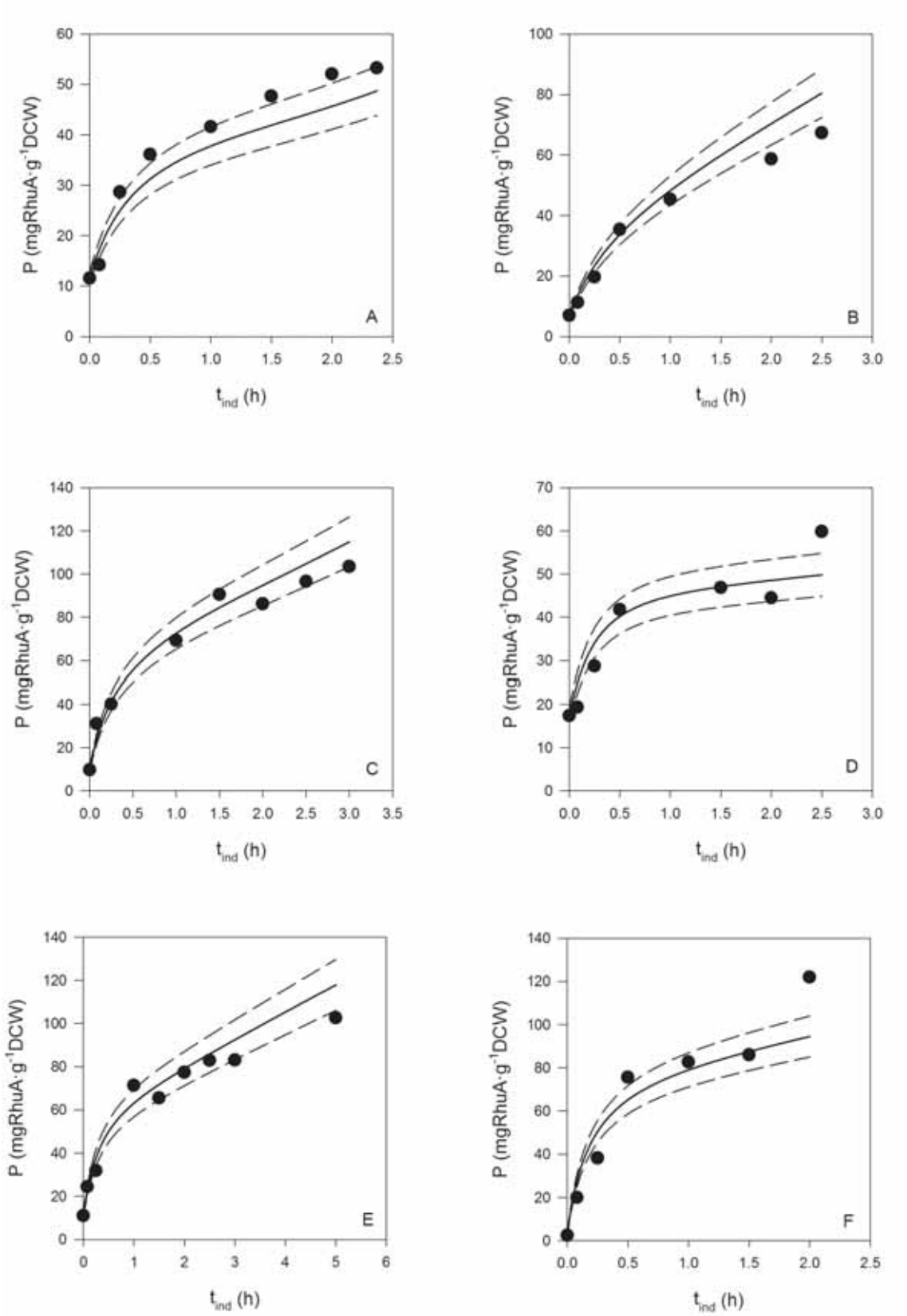


Figure 6.6. Specific protein in mass units model fitting. (•) experimental data, (-)model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively: (A) 24μM, 40gDCWL^{-1}, 0.22h^{-1}; (B) 27μM, 20gDCWL^{-1}, 0.1h^{-1}; (C) 60μM, 13gDCWL^{-1}, 0.16h^{-1}; (D) 60μM, 47gDCWL^{-1}, 0.16h^{-1}; (E) 100μM, 20gDCWL^{-1}, 0.1h^{-1}; (F) 100μM, 20gDCWL^{-1}, 0.22h^{-1}

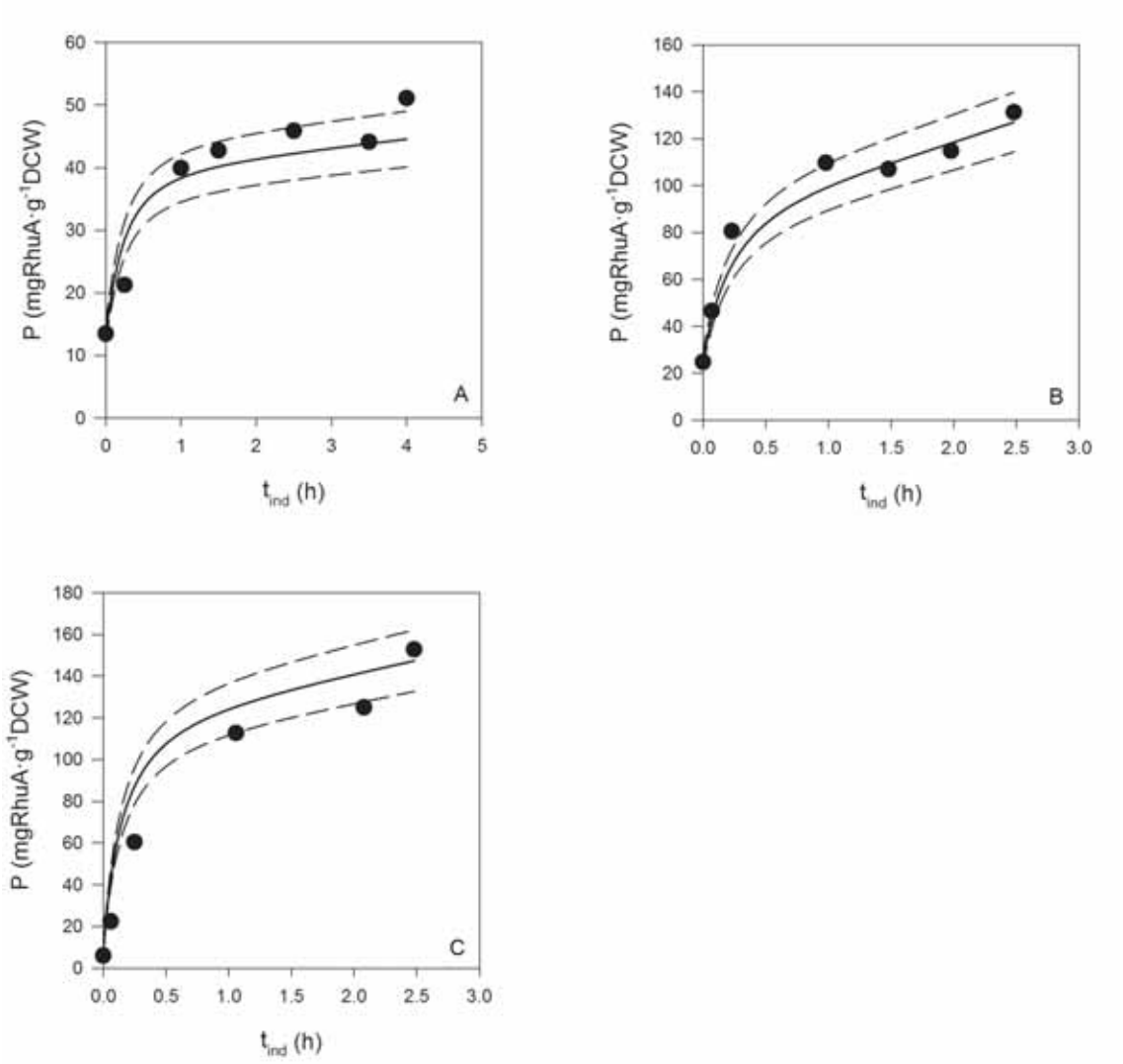


Figure 6.7. Specific protein production model fitting. (•) experimental data, (-)specific protein production model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively:; (A) 100μM, 40gDCWL^{-1}, 0.1h^{-1}; (B) 200μM, 20gDCWL^{-1}, 0.22h^{-1}; (C) 1000μM, 20gDCWL^{-1}, 0.22h^{-1}.

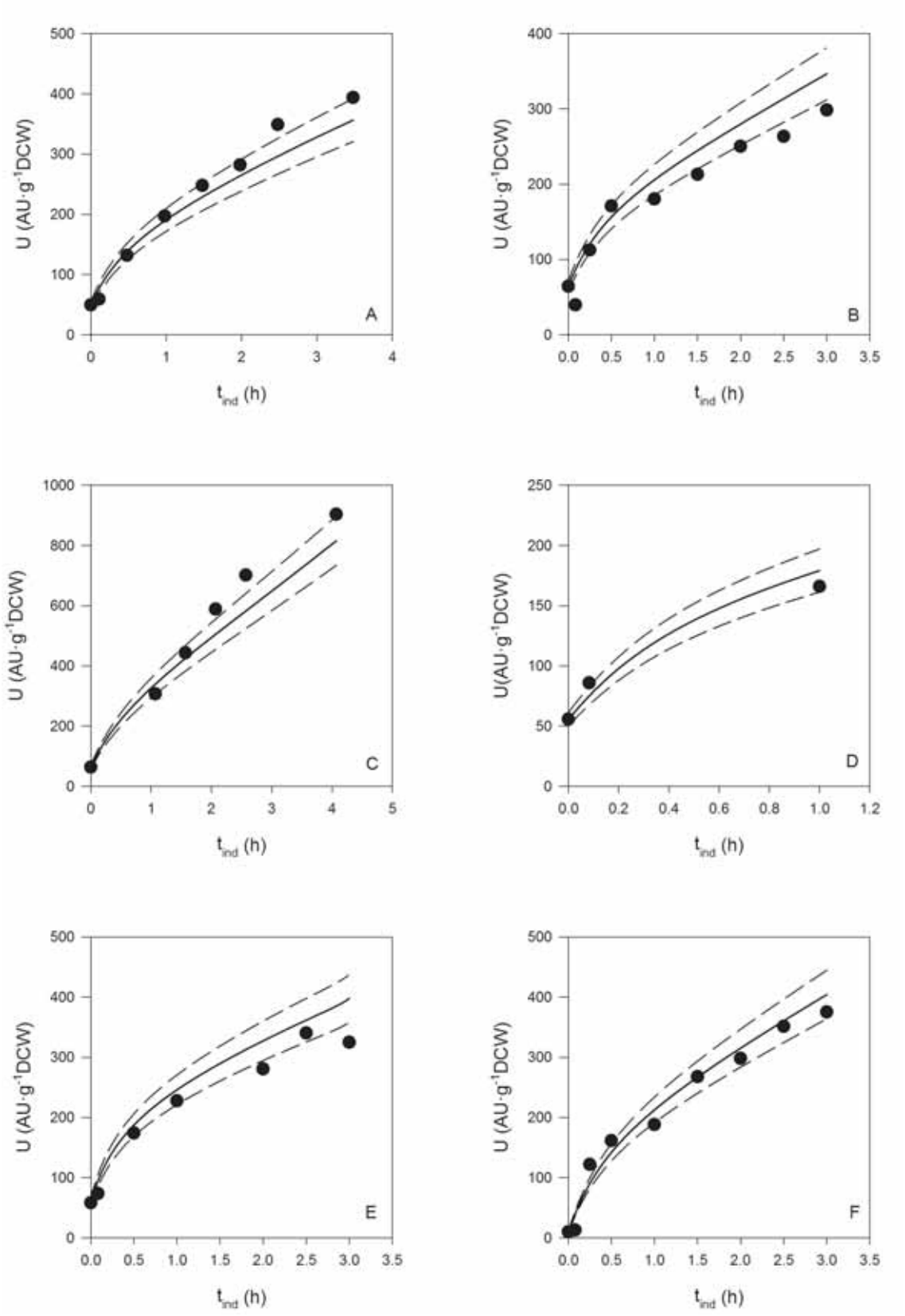


Figure 6.8. Specific protein in activity units model fitting. (•) experimental data, (-)model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively: (A) 8μM, 20gDCWL^{-1}, 0.22h^{-1}; (B) 10μM, 20gDCWL^{-1}, 0.1h^{-1}; (C) 10μM, 20gDCWL^{-1}, 0.22h^{-1}; (D) 10μM, 40gDCWL^{-1}, 0.1h^{-1}; (E) 15μM, 30gDCWL^{-1}, 0.16h^{-1}; (F) 24μM, 40gDCWL^{-1}, 0.1h^{-1}.

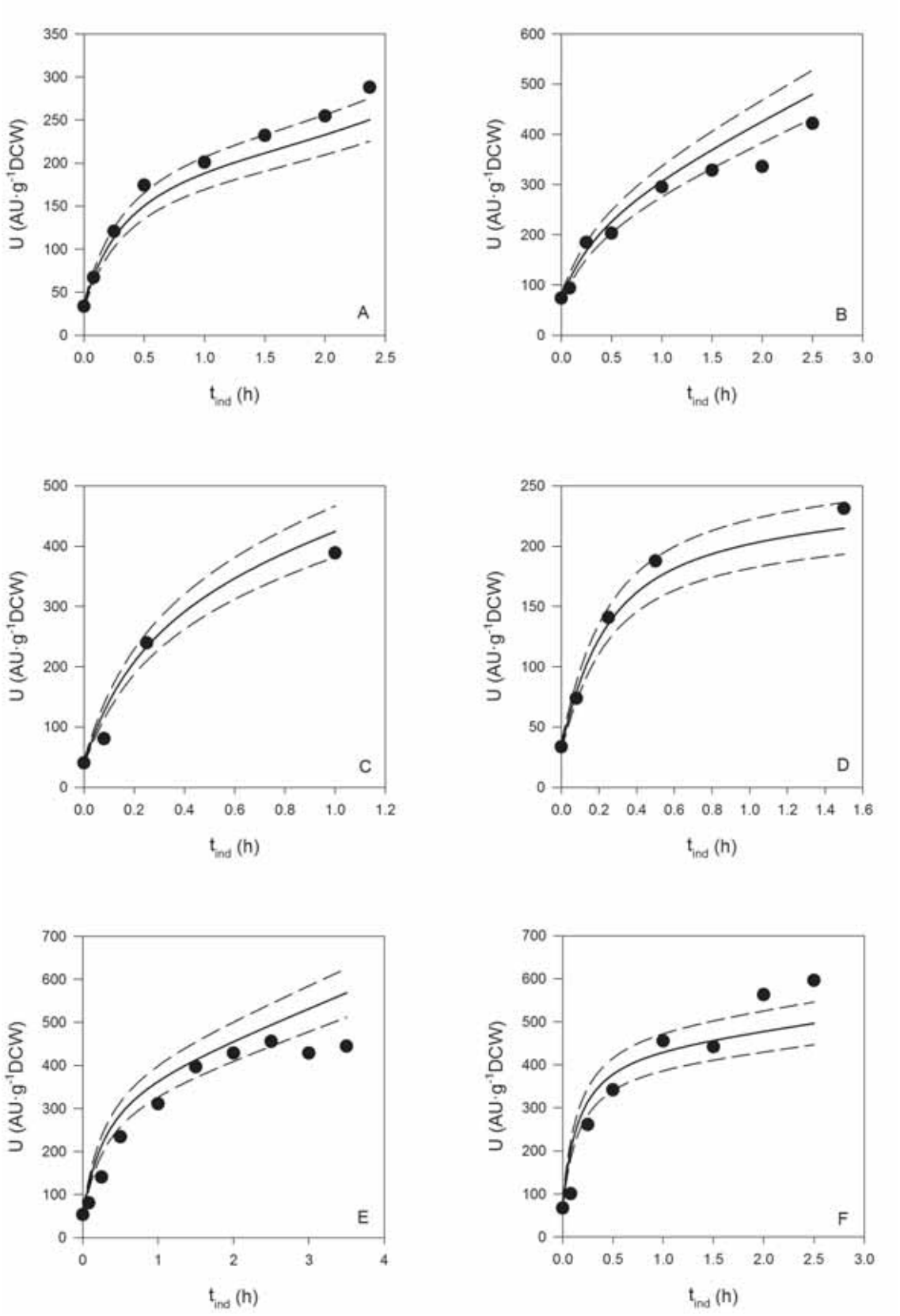


Figure 6.9. Specific protein in activity units model fitting. (•) experimental data, (-)model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively: (A) 24μM, 40gDCWL^{-1}, 0.22h^{-1}; (B) 27μM, 20gDCWL^{-1}, 0.1h^{-1}; (C) 60μM, 13gDCWL^{-1}, 0.16h^{-1}; (D) 60μM, 47gDCWL^{-1}, 0.16h^{-1}; (E) 100μM, 20gDCWL^{-1}, 0.1h^{-1}, (F) 100μM, 20gDCWL^{-1}, 0.22h^{-1}.

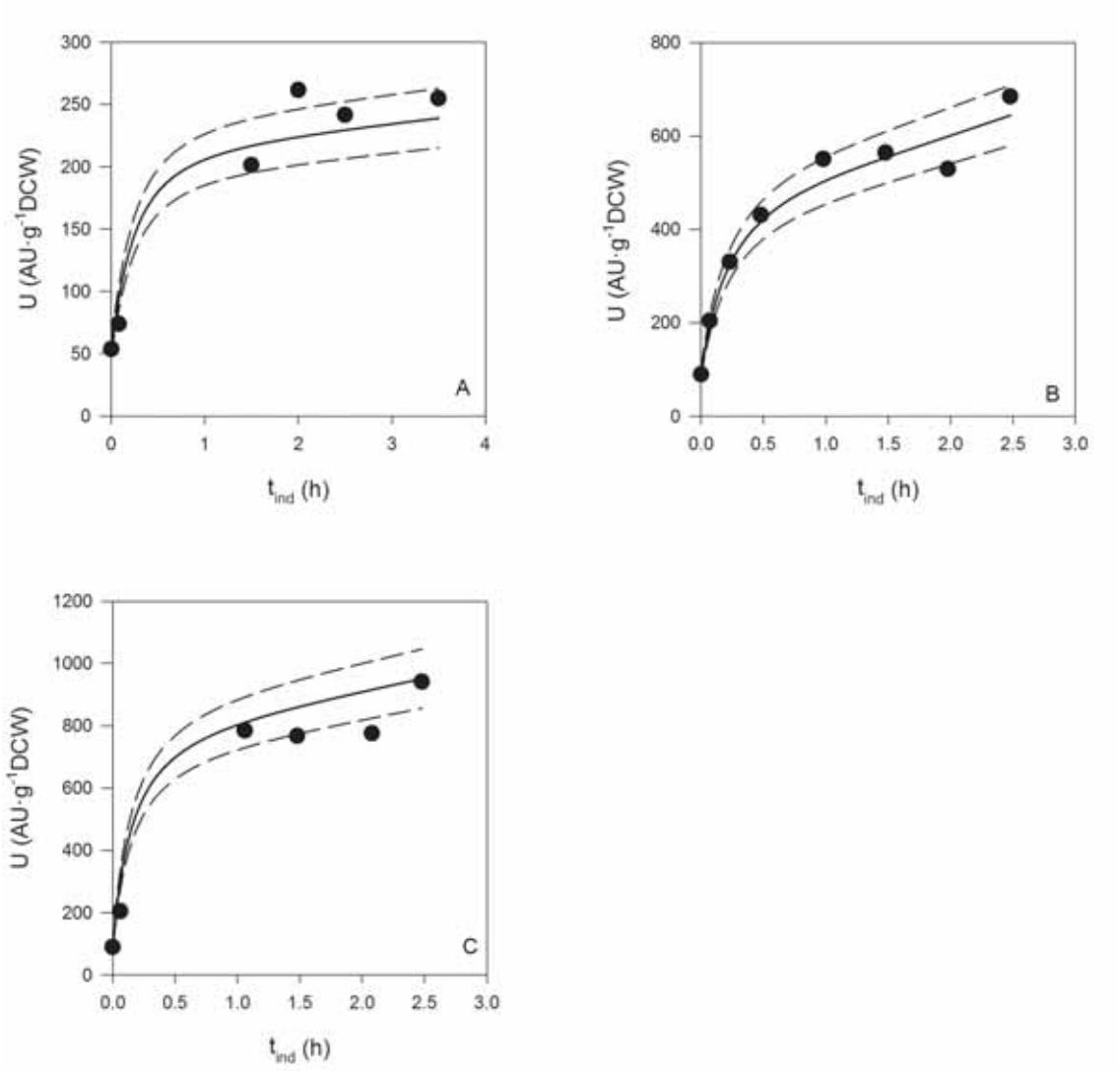


Figure 6.10. Specific protein in activity units model fitting. (•) experimental data, (-)model prediction, (- -) 10% error interval. Experimental conditions $[IPTG]_{e,0}$, X_{ind} , μ_{fix}, respectively:; (A) 100μM, 40gDCWL^{-1}, 0.1h^{-1}; (B) 200μM, 20gDCWL^{-1}, 0.22h^{-1}; (C) 1000μM, 20gDCWL^{-1}, 0.22h^{-1}.

Figures 6.5 to 6.10 show that the model fitted, with its estimated parameters, is able to reasonably predict both specific protein in mass and specific protein in activity Moreover, as for the IPTG uptake model, the error is showed as a ±10% area.

In order to give an overall idea of the goodness of the fitting, Figure 6.11 shows the parity plots for specific protein in mass and in activity units.

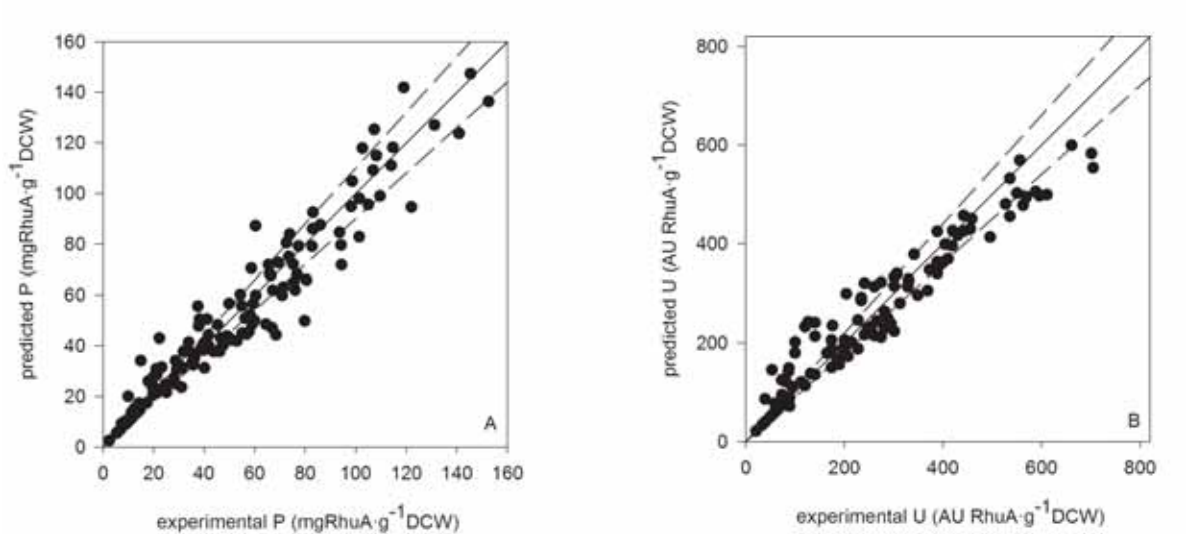


Figure 6.11. Parity plots with 10% error. A: specific protein in mass. B: specific protein in activity.

As it can be seen in Figure 6.11, the model is capable to predict, properly, the experimental specific protein both in activity and mass units for a whole range of different cultivation conditions, with quite small deviation.

6.3 Model validation

Once the model has been proposed and calibrated, a validation of the whole model has to be done. According to the compartments depicted in Figure 6.1, the protein production model has been coupled with the IPTG uptake and the biomass growth ones in order to have a simulation of the whole behavior of the protein production process. It is important to remark that this simulation is the result of three different models.

For validation purposes, a new experiment using different conditions at the beginning of induction phase than the ones used in the fitting was performed. The experimental conditions are selected to be within the experimental space used in the model fitting and they are depicted in Table 6.6. Table 6.6 shows, also, the initial values for the different variables, in order to make possible their integration.

Table 6.6. Initial values for the different variables.

Variable	Value	Units
$\mathbf{[IPTG]_{e,0}}$	84.14	**μ**M
$\mathbf{X_{ind}}$	21.5	$gDCW \cdot L^{-1}$
$\mathbf{\mu_{fix}}$	0.20	h^{-1}
$\mathbf{V_0}$	0.800	L
$\mathbf{X_0}$	0.054	$gDCW \cdot L^{-1}$
$\mathbf{S_0}$	21	$g \cdot L^{-1}$
M	20	$molecules \cdot cell^{-1}$
$\mathbf{R_{free}}$	2700	$molecules \cdot cell^{-1}$
$\mathbf{R_{IPTG}}$	0	$molecules \cdot cell^{-1}$
$\mathbf{P_{ind}}$	8.5	$mgRhuA \cdot g^{-1}DCW$
$\mathbf{U_{ind}}$	59.3	$AU \cdot mg^{-1}RhuA$
S_f (feeding concentration)	485	$g \cdot L^{-1}$
$\mathbf{[IPTG]_i}$	0	μM

The initial values for M and R_{free} are calculated at equilibrium. As the culture is grown since the batch phase (without induction), LacI mRNA molecules as well as repressor molecules are being produced, until the equilibrium between the production and degradation rate are reached. For the calculation of M, Equation (6.20) is needed. When degradation and formation rates are equals, the variation with time of M is zero. On the same way, R_{free} evolution with time when degradation and production rates are equals (in addition to the third term that will be zero because there are not IPTG), will be zero. These values are the ones used as initial point for the induction phase.

P_{ind} and U_{ind} correspond, respectively, to the basal specific protein in mass units and in activity units. In this case, these are the experimental values that are close of the mean of basal protein for all the experiments.

Figure 6.12 shows the experimental data and model prediction values of biomass concentration, total volume and glucose along time, for the three different growth phases: batch, fed-batch (non-induced) and induction phase.

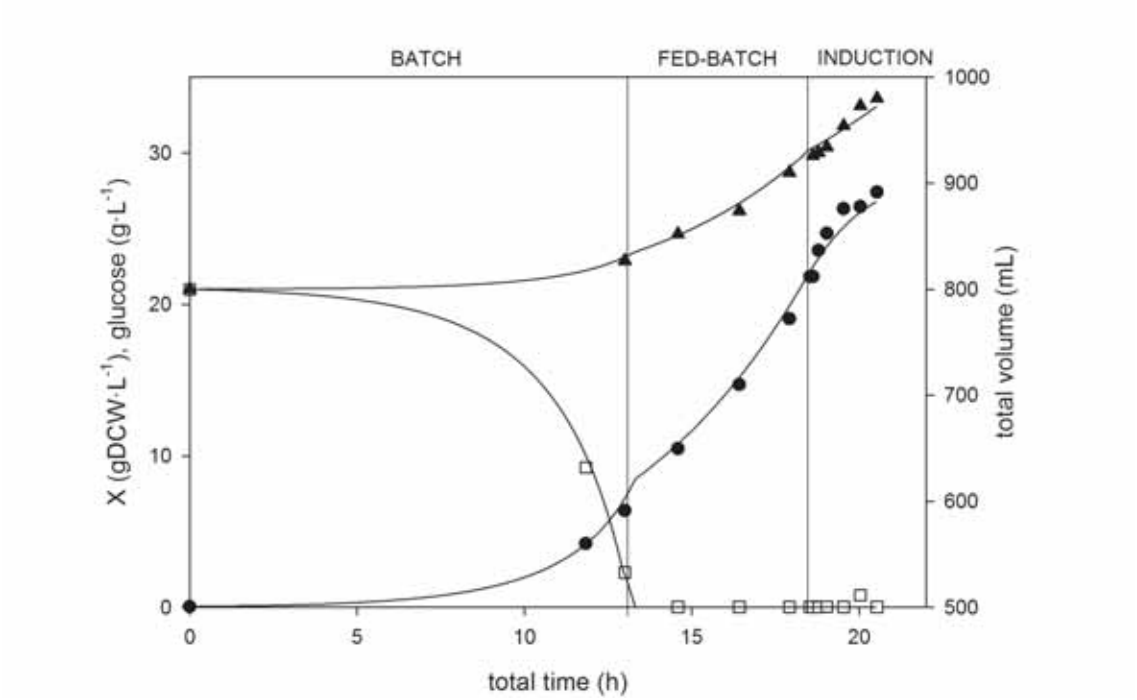


Figure 6.12. Validation experiment. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration; (-) model predicted curves.

As it can be seen, the model is capable to predict properly the evolution of glucose and biomass concentration, as well as the total volume, along all the experiment.

Figure 6.13 A shows the experimental evolution of the extracellular IPTG concentration, as well as its model prediction and calculated intracellular inducer concentration. It can be seen that the prediction of the experimental data of extracellular IPTG fits properly. Figure 6.13 B shows free and bound repressor predicted evolutions. Bound repressor evolution with time is used for the calculation of the specific protein in mass units evolution with time.

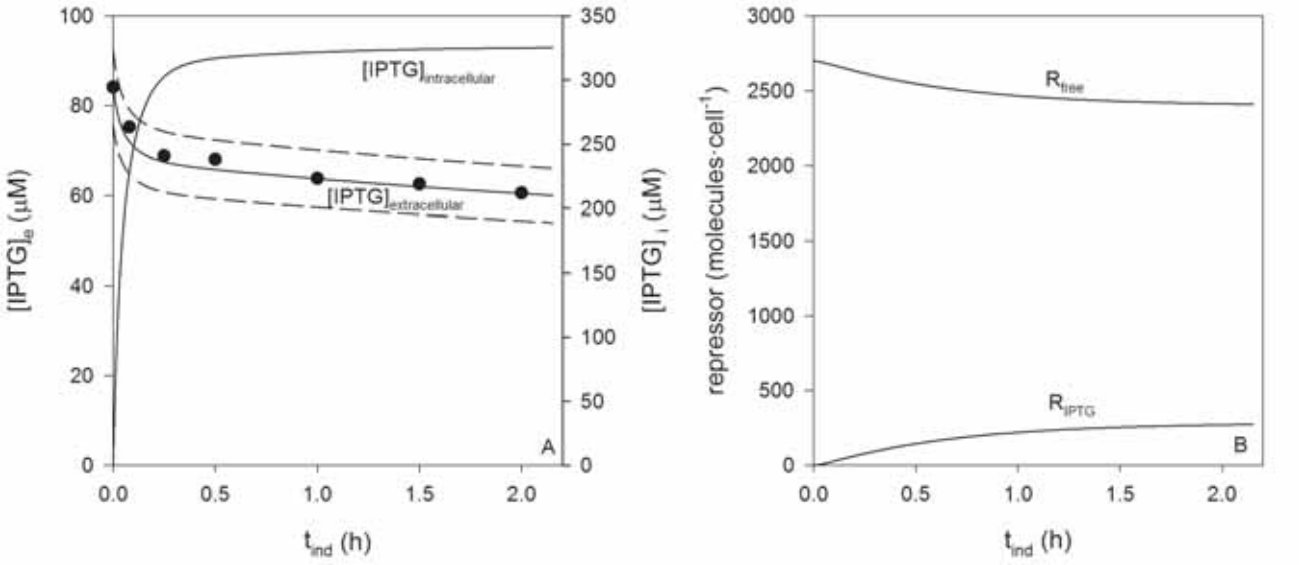


Figure 6.13. Validation experiment. A: (•) experimental extracellular IPTG concentration; (-) model predicted curves, (- -) 10% error. B: repressor evolution with time after induction.

Finally, Figure 6.14 shows both specific protein in mass and in activity units.

Figures 6.12 to 6.14 show that the overall coupled model is able to predict reasonably the specific protein in mass as well as in activity units. Not only the protein model, while the biomass growth model and the IPTG uptake models are predicting the experimental behavior of the whole experiment.

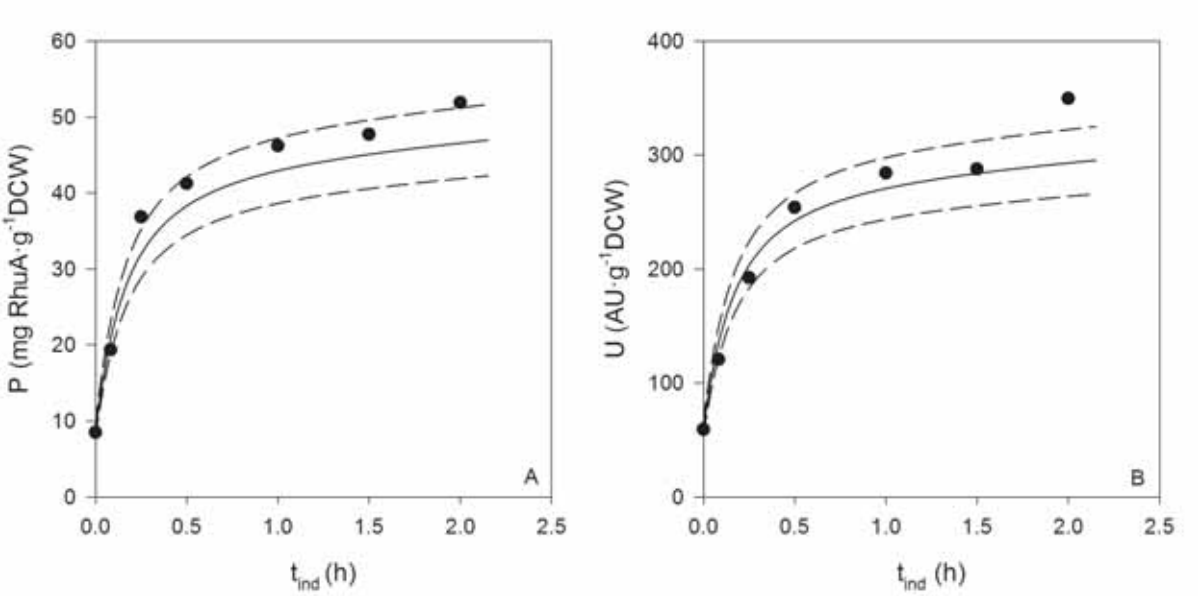


Figure 6.14. Validation experiment. Specific protein in mass (A) and in activity (B) evolution with time after induction. (•) experimental data; (-) predicted evolution, (- -) 10% error.

6.4 Conclusions

Experimental data of specific protein production as mass and activity under productive fed-batch fermentations have been used for the calibration of the model presented.

Firstly, using the calculated intracellular IPTG concentration from the IPTG uptake model (from Chapter 5) it has been possible to predict the evolution of repressor molecules bounded to the inducer, as well as the ones able to bind the operator sites. At this point, the specific protein in mass units (P) production can be related to the repressor-IPTG bounding equilibrium, and the model describes properly its evolution for all the experiments.

Secondly, aldolase's specific activity (U) can be related to the mass of protein using the quality of the protein (K_q, AU·mg^{-1}RhuA).

Secondly, once the model for specific protein in mass and activity units has been developed and calibrated, it is possible to couple it with biomass growth model, as well as with the IPTG uptake model.

The overall coupled model allows the simulation of the whole experiment needing, only, the initial values shown in Table 6.6, which can be chosen for every experiment.

7 Protein production model extension

Although a model of specific protein production (in mass and activity units) for a unique recombinant protein is important, it is significantly more useful the possibility of the model to be extended to other expression systems, producing other proteins. Nowadays, *E.coli* BL21 (DE3) strains are the most widely used for high-yield expression of recombinant proteins (Robichon et al. , 2011)

Two different proteins have been selected to be produced in engineered *E.coli* BL21 (DE3) strain: Fructose-6-phosphate aldolase (FSA) and ω-Transaminase (or Aminotransferase, ATA).

Firstly reported by Sprenger in 2001 (Schurmann & Sprenger, 2001), one of the important discoveries in the family of aldolases was Fructose-6-phosphate Aldolase. This aldolase is able to accept DHA (the unphosphorylated form of the dihydroxyacetone) as a donor. This fact represents an advantage over the expensive and unstable DHAP (which is used by the well-known DHAP dependents aldolases, for instance RhuA) (Sánchez-Moreno et al., 2012). Moreover, DHAP-dependent aldolases require a phosphatase to remove the phosphate group.

One additional advantage of using FSA is its ability to react with four different donors (DHA, 1-hydroxybutan-2-one, hydroxyacetone and glyceraldehyde) (Castillo et al., 2010; Schürmann, Schürmann, & Sprenger, 2002; Sugiyama et al., 2007).

FSA has been used in the synthesis of different polyhydroxylated compounds (such as iminocyclitols and carbohydrates) (Guérard-Hélaine et al., 2011; Sugiyama et al., 2007). FSA catalyzes the aldol reaction from D-fructose-6-phosphate to D-glyceraldehyde-3-phosphate and, as said, the non-phosphorylated DHA. Figure 7.1 shows the basic FSA catalyzed reaction.

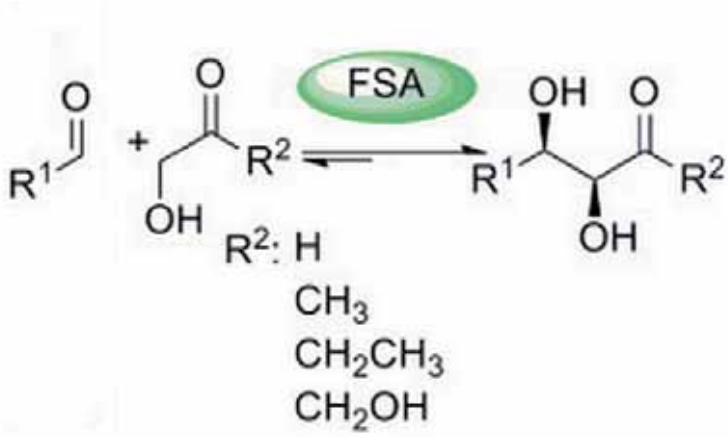


Figure 7.1. Schematic representation of a FSA catalyzed reaction

On the other hand, transaminases are a group of highly interesting enzymes that catalyze the transformation of ketones to chiral amines (Constable et al., 2007; Nugent & El-Shazly, 2010; Panke, Held, & Wubbolts, 2004). ω-Transaminase is a PLP (pyridoxal-5'-phosphate) dependent enzyme that makes possible the transfer of an amine group from amino-donors (as arylamine or amino acids) to a pro-chiral acceptor ketone. This reaction yields a second chiral amine and a ketone (Shin & Kim, 2002). Its use is gaining importance because of its potential for the resolution of racemic amines, and in the asymmetric synthesis of optically active amines (Casablancas et al., 2013; Koszelewski et al., 2010; Shin, Kim, & Shin, 2001).

Figure 7.2 shows the reaction catalyzed by ATA.

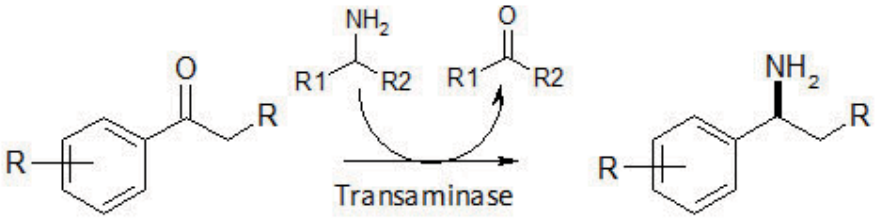


Figure 7.2. reaction catalyzed by ATA

As indicated, the main objective in this chapter is the extension of the model presented in Chapter 6 to other expression systems, producing two different proteins (FSA and ATA).

7.1 Specific material and methods

The work presented in this Chapter was undertaken at The University of Sydney, (Sydney, Australia) in the Fermentation and Wastewater group's Laboratories (School of Chemical and Biomolecular Engineering), under the supervision of Dr. John Kavanagh.

7.1.1 Strains and expression systems

The strains used (BL21 (DE3) FSA and BL21 (DE3) ATA) were already introduced in Chapter 3.

Fructose-6-Phosphate Aldolase (FSA) is produced by *E.coli* BL21 (DE3) strain, as described in literature (Sánchez-Moreno et al., 2012). The gene *mipB*, encoding FSA was inserted into a modified expression vector pET22b(+), presented in Figure 7.3 (Kreimeyer et al., 2007).

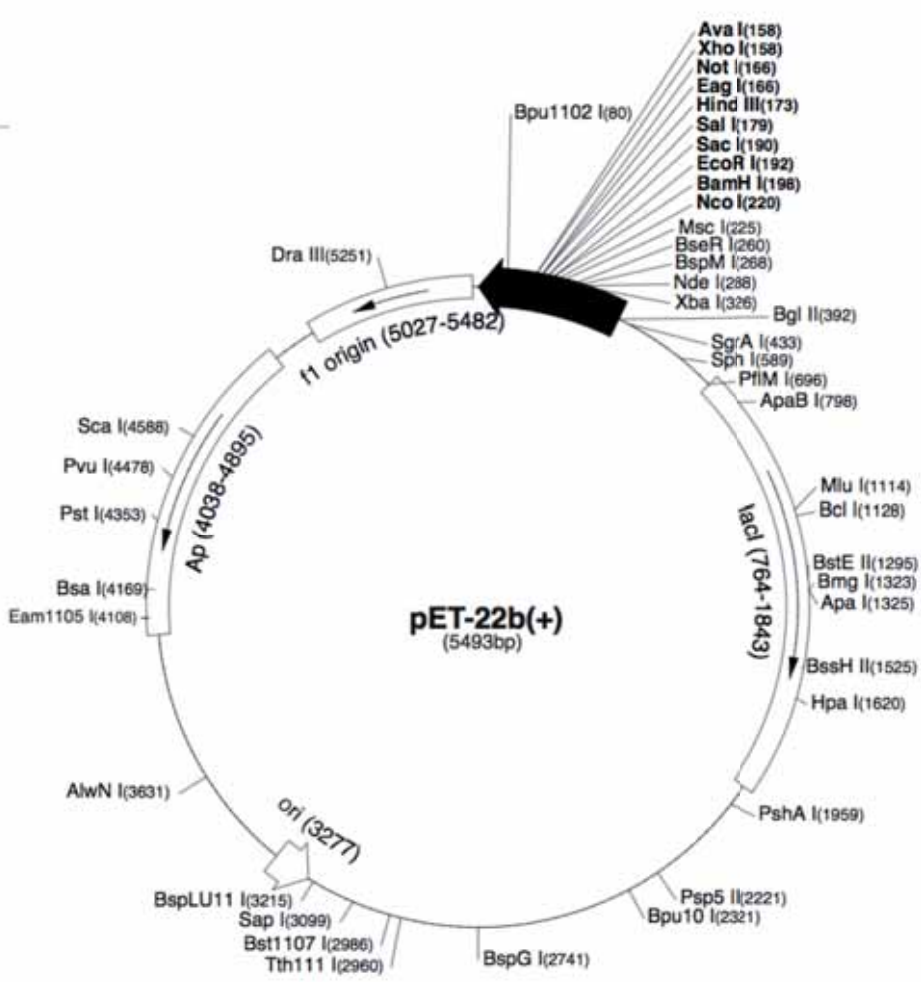


Figure 7.3. pET22b(+) plasmid

The main parts of the pET-22b(+) vector shown in Figure 7.3 are the *bla* gene (shown as Ap in Figure 7.3) that codifies for β-lactamase, giving the strain antibiotic resistance (ampicillin) and *lac*I gene, codifying for the LacI repressor protein. *mipB* gene is not shown, but it has been cloned just after the promoter (black arrow)

ω-Transaminase (ATA) is produced using *E.coli* BL21 (DE3) strain, harboring pLE1-A10-AcATA-D4 plasmid (supplied by c-LEcta) (Casablancas et al., 2013). Although the details of the vector cannot be depicted (because it is propriety of the company), it includes the *lac*I gene (codifying for the LacI repressor proteins) and the gene *neo*, codifying for kanamycin resistance.

7.1.2 Fed-batch fermentations

For the new fermentations, using the BL21 (DE3) strains, the methodology is the same as for the ones made in UAB until the inoculation of the fermentor. In this case, a New Brunswick BioFlo® Fermentor (Eppendorf) with 3L vessel was used. It was equipped with temperature, dissolved oxygen and pH control. As for the fermentations with the M15 strain, temperature was maintained at 37 °C and pH at 7.00±0.05 by addition of 15% (v/v) NH_4OH solution. pO_2 value vas maintained at 60% of the saturation, adapting the stirring speed between 300 and 1250 rpm, with 1.5 $L \cdot h^{-1}$ of air. For glucose limiting feeding, a peristaltic pump installed in the DCU was used. Induction phase was carried out as indicated in Chapter 3.

7.1.3 Analytical methods

7.1.3.1 Biomass and total protein quantification

Biomass and total protein quantification were carried out as described in Chapter 3.

7.1.3.2 Activity assay

Enzymatic activity tests for every protein were performed as described below:

- FSA

As for RhuA, FSA activity is determined through the NADH disappearance rate (30 °C at 340 nm wavelength). In this case, three coupled reactions are needed, as shown in Reactions (a), (b) and (c).

$$Fructose-6-Phosphate \xrightarrow{FSA} glyceraldehyde-3-phosphate + DHA \quad \text{(a)}$$

$$glyceraldehyde-3-phosphate \xrightarrow{TPI} DHAP \quad \text{(b)}$$

$$DHAP + NADH \xrightarrow{\alpha-GDH} \alpha-glycerophosphate + NAD^{+} \quad \text{(c)}$$

Reaction (a) is the natural aldolase reaction for FSA, a DHA-dependent enzyme. Reaction (b) –catalyzed by triosephosphate isomerase (TPI) enzyme- produces the DHAP necessary for reaction (c).

Table 7.1 shows the reagents concentration for the assay.

Table 7.1. Reagents concentration for FSA enzymatic assay.

Component	Concentration
NADH	0.1 mM
Fructose-6-Phosphate	5 mM
TPI/ α-GDH	3 $U \cdot mL^{-1}$
Imidazole	50 mM

In this case, all reagents are added into a 1mL quartz cuvette, to a final volume of 950 µL. Once the reagents are into the cuvette, as for RhuA activity test, it is read the absorbance until it shows a stable value. Then, 50 µL of sample is added, the cuvette is shake and the variation of absorbance is evaluated.

As for RhuA, the volumetric activity can be calculated using Equation (7.1)

$$AU \cdot mL^{-1} = \frac{\Delta Abs_{340}}{\varepsilon} \frac{V_t}{V_e} D_f L_p \quad (7.1)$$

where:

- ΔAbs_{340} is the absorbance variation per minute at wavelength 340 nm (min^{-1})
- ε is the NADH molar extinction coefficient (6.22 $mM^{-1} \cdot cm^{-1}$)
- V_t is the total assay volume (mL)
- V_e is the volume of the aldolase sample (mL)
- D_f is the dilution factor of the sample
- L_p is the optical path of the cuvette (cm)

- ATA

ATA activity was determined spectrophotometrically at 30 °C and 240 nm wavelength following the production of acetophenone. Reaction (d) shows the reaction for the activity assay.

$$1-phenylethylamine + pyruvate \xrightarrow{ATA} acetophenone + L-alanine \quad \text{(d)}$$

This reaction needs pyridoxal phosphate as cofactor.

Table 7.2 shows the reagents concentration for the assay.

Table 7.2. reagents concentration for ATA enzymatic assay.

Component	Concentration
Phosphate buffer	45 mM
Phenylethylamine	9.9 mM
Pyruvate	9.9 mM
Pyridoxal phosphate	0.09 mM
Dimethyl sulfoxide	1.125% (v/v)

In this case, all reagents are added into a 1mL quartz cuvette, to a final volume of 900 µL. Once the reagents are into the cuvette, as for RhuA activity test, it is read the absorbance until it shows a stable value. Then, 100 µL of sample is added, the cuvette is shake and the variation of absorbance is evaluated.

Volumetric activity can be calculated using Equation (7.1), but in this case, the increase of absorption caused by the production of acetophenone will be followed. Moreover, it is needed to use the acetophenone extinction coefficient ($0.28 mM^{-1} \cdot cm$).

7.2 Specific protein model

The equations describing the specific protein in mass and in activity have been developed and validated in Chapter 6, but for its extension, it is necessary to determine which parameters are going to be different, that means, which ones are expression system's dependent.

Equations are reproduced in Table 7.3

Table 7.3. Specific protein production and activity model equations.

PROTEIN PRODUCTION MODEL	
$\frac{dM}{dt} = \alpha_{mRNA}N - \lambda_{mRNA}M$	(7.2)
$\frac{dR_{free}}{dt} = \alpha_R M - \lambda_R R_{free} - \frac{1}{\tau_{R-O}}\left[R_{free}\left(\frac{[IPTG]_i/_{1000}}{K_{R-IPTG}}\right)^n - R_{IPTG}\right]$	(7.3)
$\frac{dR_{IPTG}}{dt} = \frac{1}{\tau_{R-O}}\left[R_{free}\left(\frac{[IPTG]_i/_{1000}}{K_{R-IPTG}}\right)^n - R_{IPTG}\right] - \lambda_R R_{IPTG}$	(7.4)
$\frac{dq_p}{dt} = -\frac{K_{P1}\left(\frac{dR_{IPTG}}{dt}\right)}{1+K_{P2}P}q_p$	(7.5)
$\frac{dP}{dt} = q_p - \frac{P\frac{dV}{dt}}{V}$	(7.6)
$\frac{dU}{dt} = K_q\frac{dP}{dt}$	(7.7)

E.coli BL21 (DE3) strain carries LacI repression system in its genome (Jeong et al., 2009; Lebedevaet al., 1994; Xuet al., 2012). Moreover, as explained above, plasmids inserted in cells carry, also, the gene responsible for the production of LacI molecules. As these other expression systems use the same repressor as the M15 strain producing RhuA, the value of the parameters describing the interaction between IPTG and repressor molecules ($K_{R\text{-}IPTG}$, n) and the ones describing the LacI production and degradation (α_{mRNA}, λ_{mRNA}, α_R, λ_R) will be the same for the three strains.

However, the number of copies of *lac*I gene for both BL21 (DE3) strains, as well as the number of plasmids into the cells, is unknown. For this reason, the parameter N (number of copies of *lacI* gene) will be a parameter to be fitted.

Moreover, $\tau_{R\text{-}O}$ (time constant for the linkage between the repressor and the operator) is dependent on the expression system, and it is one more parameter to be changed.

Finally, the parameters related to specific protein production (K_{P1} and K_{P2}) as well as the parameter that describes the specific activity of the protein produced (K_q) will be different depending on the protein produced.

Thus, there are 5 parameters to be estimated in order to extend the model validated for RhuA production to FSA and ATA, using *E.coli* BL21 (DE3) as host.

Otherwise, as Equations (7.3) and (7.4) show, it is needed the prediction of the intracellular IPTG concentration along time for the calculation of the free and the bound repressor molecules. This prediction can be made using the IPTG uptake model presented in Chapter 5 and used in Chapter 6. It is reproduced in Table 7.4

Table 7.4. IPTG uptake model equations

IPTG UPTAKE MODEL	
$\frac{dV_{cel}}{dt} = 0.0023\,\frac{d(XV)}{dt}$	(7.8)
$\frac{dV_m}{dt} = \frac{dV}{dt} - \frac{dV_{cel}}{dt}$	(7.9)
$\frac{d[IPTG]_e}{dt} = \frac{-r \cdot V_m - [IPTG]_e \frac{dV_m}{dt}}{V_m}$	(7.10)
$\frac{d[IPTG]_i}{dt} = \frac{r \cdot V_m - [IPTG]_i \frac{dV_{cel}}{dt}}{V_{cel}}$	(7.11)
$r = k_c a([IPTG]_e - [IPTG]_i) + K_1 X_{ind}\left(K''[IPTG]_e + \frac{k\prime[IPTG]_e}{K_M + [IPTG]_i^2}\right) exp(K_2 \mu_{fix})$	(7.12)

In this case, the BL21 (DE3) FSA strain was already been used for the validation of the IPTG uptake model (Chapter 5). It can be assumed that BL21 (DE3) ATA strain will conserve the same transport mechanisms as BL21 (DE3) FSA , thus the IPTG uptake model has been considered to be applicable to both strains.

For the fitting of the protein production sub-model to the new experiments, experimental data of total volume as well as biomass concentration have been used (as was made for the fitting of the IPTG uptake model in Chapter 5 and protein production and activity model in Chapter 6).

Three different experiments for every strain, with different cultivation conditions were conducted.

7.3 BL21 (DE3) FSA model fitting

Three different experiments with different cultivation conditions were conducted. The experimental conditions are shown in Table 7.5.

Table 7.5. Experimental conditions for BL21 (DE3) FSA strain

	FSA1	FSA2	FSA3
$[IPTG]_{e,0}$ (μM)	100	70	100
X_{ind} (gDCW·L^{-1})	10.5	12.5	21.6
μ_{fix} (h^{-1})	0.15	0.2	0.22

As it is depicted in Table 7.5, fed-batch production cultures have been performed under experimental conditions inside the experimental space employed for model calibration.

The result of the estimation of the different parameters is shown in Table 7.6.

Table 7.6. Parameter's estimated values for specific protein production and specific protein activity model.

Parameter	Value	Units
N	42±5	number of copies
τ_{R-O}	16±1	h
K_{P1}	0.0700±8·10^{-4}	cell·molecule^{-1}h^{-1}
K_{P2}	0.23±0.09	gDCW·mg^{-1}RhuA
K_q	5.720±1·10^{-3}	AU·mg^{-1}RhuA

As it is shown in Table 7.6, the value for the parameter N is much higher than the value used for the M15 strain, that means that a large number of copies of *lac*I gene are present into cells. It can be related to a large number of copies of the corresponding plasmid. τ_{R-O} is higher than for the M15 strain. This fact means that the dynamic equilibrium is slower. K_{P1} and K_{P2} take values similar to M15 strain. This means that protein production dynamics are similar.

Figure 7.4 presents the results of the fitting and it can be seen that the model is able to predict properly specific protein both in mass and activity for an *E.coli* BL21 (DE3) strain, producing Fructose-6-Phosphate Aldolase.

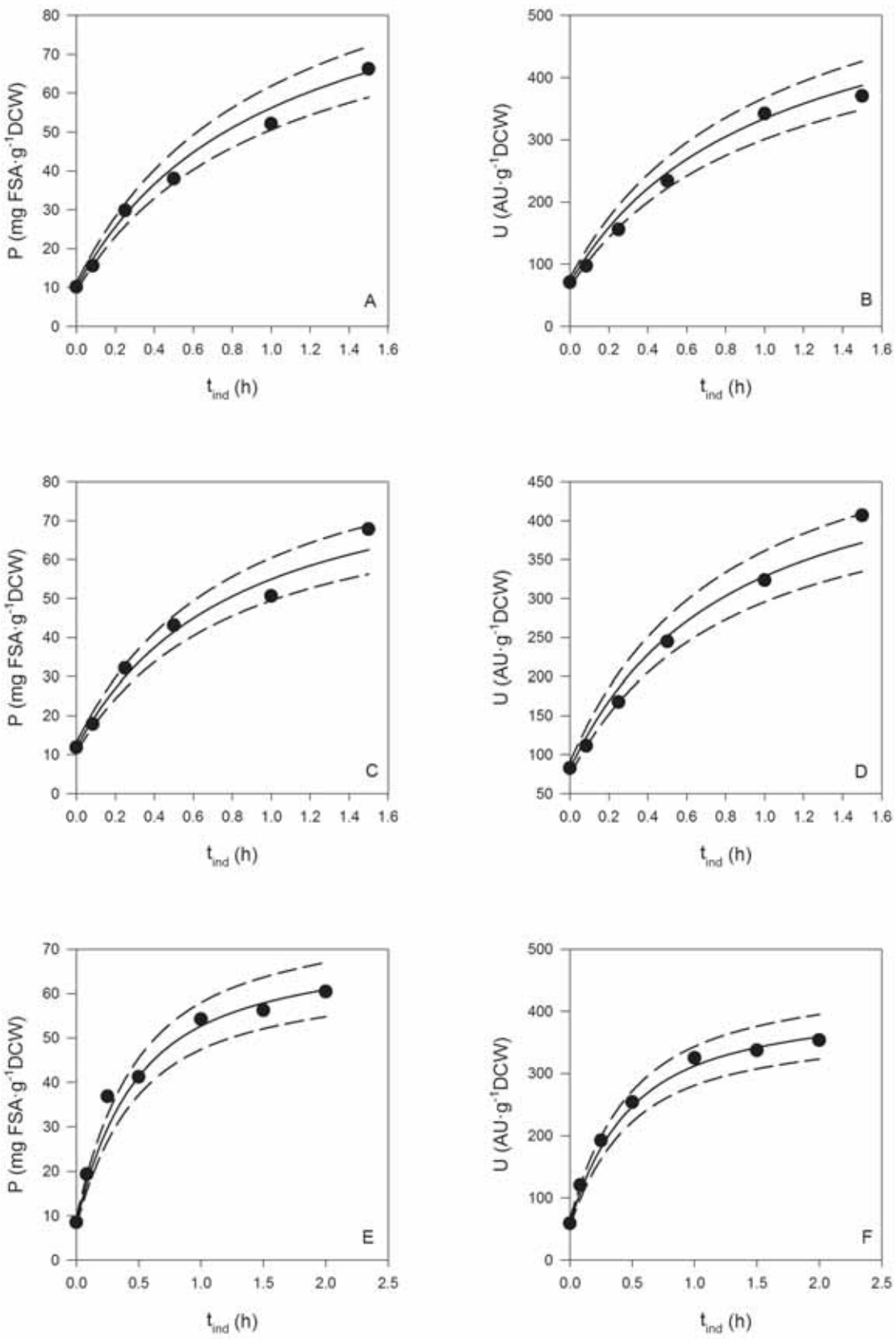


Figure 7.4. Specific protein in mass and activity model fitting. (•) experimental data, (-) model fitting, (- -) 10% error area. A: FSA1 specific protein in mass; B: FSA1 specific protein in activity; C: FSA2 specific protein in mass; D: FSA2 specific protein in activity; E: FSA3 specific protein in mass; F: FSA3 specific protein in activity.

7.4 BL21 (DE3) ATA model fitting

Three different experiments with different cultivation conditions were conducted. The different experimental conditions are shown in Table 7.7.

Table 7.7. Experimental conditions for BL21 (DE3) ATA strain

	ATA1	ATA2	ATA3
$[IPTG]_{e,0}$ (μM)	100	100	70
X_{ind} (gDCW·L^{-1})	10.8	15.3	20.8
μ_{fix} (h^{-1})	0.22	0.22	0.15

As it is depicted in Table 7.7, again, fed-batch experiments performed for data acquisition are into the experimental space used in the calibration of the model in Chapter 6.

The result of the estimation of the different parameters is shown in Table 7.8.

Table 7.8. Parameter's estimated values for specific protein production and specific protein activity model.

Parameter	Value	Units
N	46±4	number of copies
τ_{R-O}	173±9	h
K_{P1}	0.5±0.2	cell·molecule^{-1}h^{-1}
K_{P2}	0.4±0.2	gDCW·mg^{-1}RhuA
K_q	5.160±1·10^{-3}	AU·mg^{-1}RhuA

Table 7.8 shows, as Table 7.7 makes, that the number of copies of *lac*I gene (that means number of copies of plasmid) is much higher than the number of copies present in M15 strains used previously in this work. Nevertheless, the number is similar to the one estimated for BL21 (DE3) FSA strain. This fact, taking into account that the number of copies of the gene present in the genome is the same for both BL21 (DE3) strains, means that the number of copies of plasmid in both strains is similar. τ_{R-O} is much higher than for the M15 strain. This fact means that the dynamic equilibrium is much slower. K_{P1} and K_{P2} take values are higher than the ones for M15 strain. This means, on the one hand, that production is faster, and on the other, that there is a higher saturation.

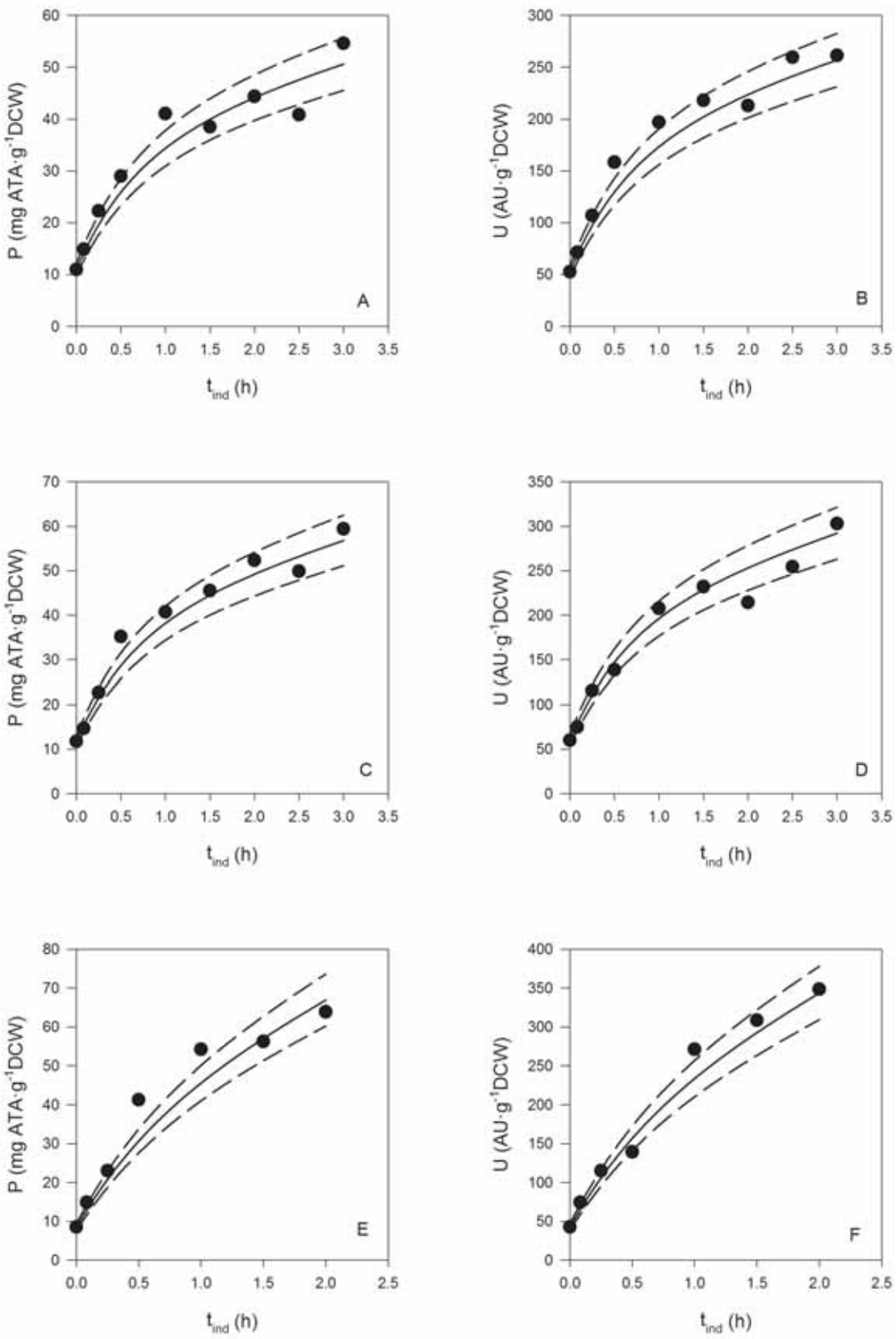


Figure 7.5. Specific protein in mass and activity model fitting. (•) experimental data, (-) model fitting, (- -) 10% error area. A: ATA1 specific protein in mass; B: ATA1 specific protein in activity; C: ATA2 specific protein in mass; D: ATA2 specific protein in activity; E: ATA3 specific protein in mass; F: ATA3 specific protein in activity.

Figure 7.5 shows that the model is able to predict properly specific protein both in mass and in activity units for a *E.coli* BL21 (DE3) strain, producing an ω-transaminase.

Although Figures 7.4 and 7.5 show the goodness of the fitting, as made in Chapters 5 and 6, a good way to see how good is the model is using parity plots. Figure 7.6 shows the parity plots for the specific protein in mass and specific activity for both strains, showing the goodness of the model fitted.

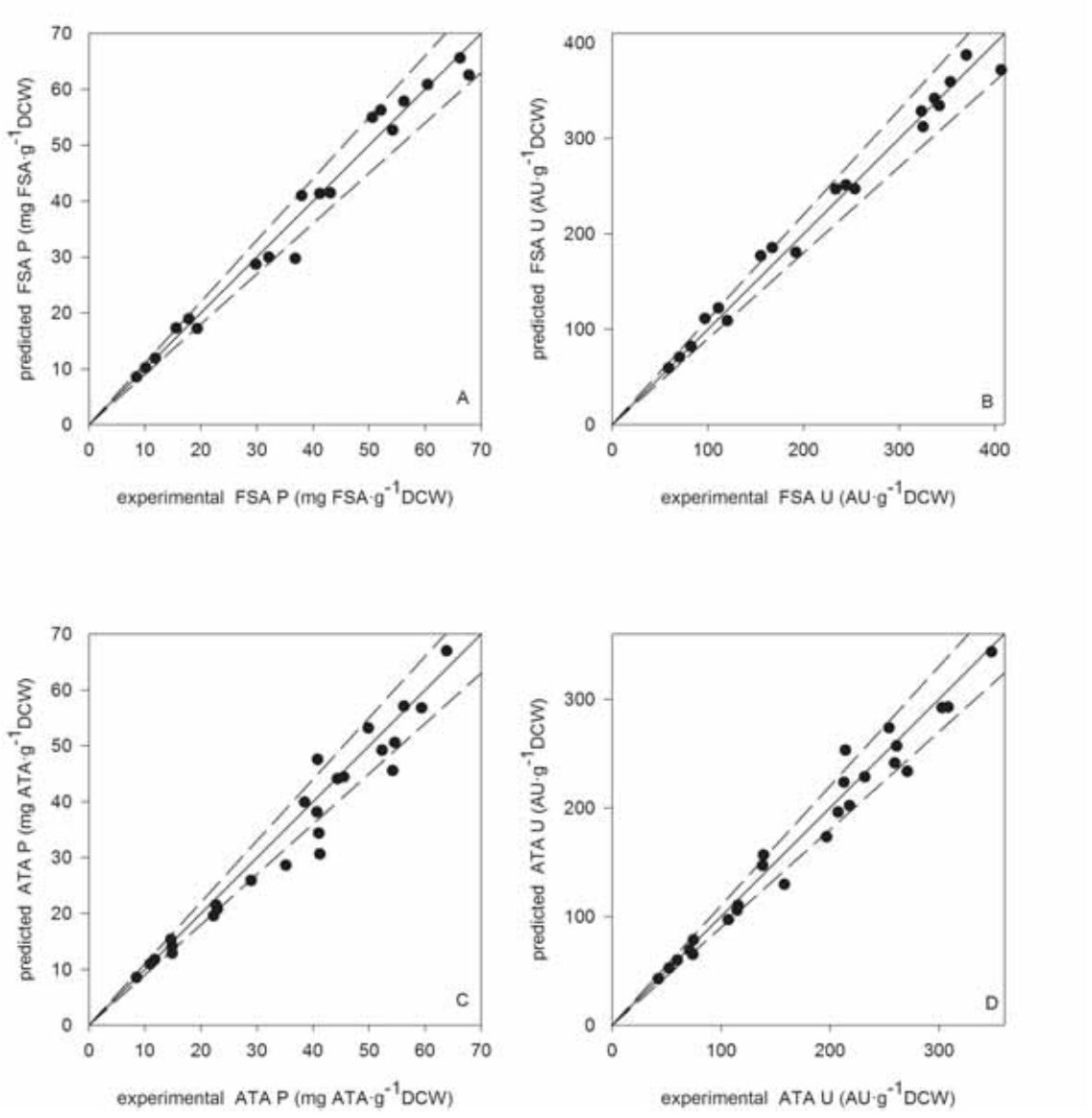


Figure 7.6. Parity plots with 10% error. A: specific FSA in mass; B: specific FSA in activity; C: specific ATA in mass; D: specific ATA in activity.

7.5 Conclusions

The model of specific protein production in mass units and in activity units developed, calibrated and validated for the production of Rhamnulose-1-phosphate Aldolase can be used for the production of different proteins -Fructose-6-phosphate Aldolase and one Aminotransferase- using a BL21 strain with different expression systems.

It is needed to identify the system-dependent' parameters. In this case, 5 of the parameters of the model are dependents on the expression system (depending on the strain: N, τ_{R-O}; depending on the protein produced: K_{P1}; K_{P2}; K_q).

The number of copies of *lac*I gene (N) (codifying for the lac repressor) is similar for both strains, meaning that they have a similar number of plasmid copies.

τ_{R-O} for both expression systems is higher than for M15 strain, meaning a slower binding equilibrium between repressor and operator.

Higher K_{P1} means a higher protein production rate, and higher K_{P2} is related with a higher saturation because protein production.

K_q value depends on the way the activity is defined.

As a final conclusion, it is possible to simulate and predict the production of several recombinant proteins in different *E.coli* strains and plasmids, demonstrating that the overall model developed can constitute a powerful tool for optimization and control of production processes.

8 General Conclusions and future perspectives

In this work, a coupled model capable to predict specific protein in mass and in activity units for fed-batch cultures of recombinant *E.coli* producing Rhamnulose-1-Phosphate aldolase has been developed, calibrated and validated.

The determination of the key experimental variables in heterologous protein production, and a preliminary study about their effect on specific protein in mass and activity units, has been carried on. Inducer, biomass concentration at induction and the specific growth rate are the control variables. It has been seen that working at low biomass concentration at induction allows obtaining high values of specific activity, and that low specific growth rates lead to low productivities. Concerning inducer concentration, it has been demonstrated that a low concentration is enough provided the inducer to biomass ratio is in the range 2-4 μmol IPTG·g^{-1} DCW. These three variables have been shown to be key in the model development.

The coupled model for protein production has three different sub-models calibrated individually:

- Biomass growth model: this model was proposed by Ruiz (Ruiz, González, de Mas, & López-Santín, 2011).
- Inducer uptake model: a quantitative study about inducer transport has been carried out. A model has been developed, calibrated and validated using experimental data. Net transport rate mechanism has been described as diffusion plus active transport. This active transport can be mediated by the specific transporting proteins (lactose permesases, specific active transport) and by other transport proteins (non-specific active transport).
- Protein production model: the part of the model corresponding to specific protein in mass and in activity units has been developed and calibrated. The model is based on the interaction between inducer and repressor.

The overall model is able to predict, properly, all the experimental data obtained in the fermentation: total volume, glucose concentration, biomass concentration, extracellular IPTG concentration, specific protein in mass and specific activity. The model is able, too, to predict intermediate variables that have not been measured: intracellular IPTG concentration, amount of free repressor and amount of repressor bound to IPTG.

The model has been extended to simulate and predict the production of different proteins expressed by different *E.coli* strains. For the extension of the model has been necessary to evaluate, the number of *lac*I gene copies (N) and the time constant for the linkage between the repressor and the operator (τ_{R-O}) that are strain dependent, and K_{P1}, K_{P2} and K_q that will change depending on the protein produced.

As final conclusion, it is possible to make a simulation and a prediction of different heterologous proteins produced in different *E.coli* strains, using different plasmids through an overall model. This kind of models represents an important tool for process optimization and control.

The obtained results of this work make possible further studies. The future perspectives include:

- Calibration of the biomass growth model to the BL21 strains
- Optimization of the volumetric productivity using the overall model
- Application of the model for control purposes.
- Coupling a metabolic model to the overall model, in order to give a more detailed explanation of the whole process

9 References

Abeles, R. H., Frey, P. A., & Jencks, W. P. (1992). *Biochemistry* (p. 838). Boston: Jones and Bartlett.

Aehle, W. (2007). *Enzymes in Industry* (3rd ed.). Wiley-VCH.

Aiba, S., Humphrey, A. E., & Millis, N. F. (1973). *Biochemical Engineering* (2nd ed.). Nwe York and London: Academic Press.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (4th ed.). New York: Garland Science.

Andersen, S. M., Ekhart, C., Lundt, I., & Stütz, A. E. (2000). Syntheses of sugar-related trihydroxyazepanes from simple carbohydrates and their activities as reversible glycosidase inhibitors. *Carbohydrate Research*, *326*(1), 22–33.

Andersson, L., Strandberg, L., & Enfors, S. O. (1996). Cell segregation and lysis have profound effects on the growth of *Escherichia coli* in high cell density fed batch cultures. *Biotechnology Progress*, *12*(2), 190–5.

Aoyama, A., Doyle, F. J., & Venkatasubramanian, V. (1995). Control-affine fuzzy neural network approach for nonlinear process control. *Journal of Process Control*, *5*(6), 375–386.

Aristidou, A. A., San, K. Y., & Bennett, G. N. (1999). Improvement of biomass yield and recombinant gene expression in *Escherichia coli* by using fructose as the primary carbon source. *Biotechnology Progress*, *15*(1), 140–5.

Bailey, J. E. (1993). Host-vector interactions in *Escherichia coli*. *Advances in Biochemical Engineering/biotechnology*, *48*, 29–52.

Bailey, J. E. (1998). Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology Progress*, *14*(1), 8–20.

Balzer, G. J., Thakker, C., Bennett, G. N., & San, K.-Y. (2013). Metabolic engineering of *Escherichia coli* to minimize byproduct formate and improving succinate productivity through increasing NADH availability by heterologous expression of NAD(+)-dependent formate dehydrogenase. *Metabolic Engineering*, *20*, 1–8.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, *10*(5), 411–21. Retrieved from

Bennett, B. D., Yuan, J., Kimball, E. H., & Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols*, *3*(8), 1299–311.

Betenbaugh, M. J., & Dhurjati, P. (1990). A comparison of mathematical model predictions to experimental measurements for growth and recombinant protein production in induced cultures of *Escherichia coli*. *Biotechnology and Bioengineering*, *36*(2), 124–34.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*, *277*(5331), 1453–62. Retrieved from

Bommarius, A. S., & Riebel, B. R. (2004). *Biocatalysis – Fundamentals and Applications*. Wiley-VCH.

Brass, J. ., Hoeks, F. W. J. M. ., & Rohner, M. (1997). Application of modelling techniques for the improvement of industrial bioprocesses. *Journal of Biotechnology*, *59*(1-2), 63–72.

Buchholz, K., Kasche, V., & Bornscheuer, U. T. (2005). *Biocatalysts and Enzyme Technology* (1st ed., p. 448). Wiley-VCH, Weinheim.

Casablancas, A., Cárdenas-Fernández, M., Álvaro, G., Benaiges, M. D., Caminal, G., De Mas, C., López-Santín, J. (2013). New ammonia lyases and amine transaminases: Standardization of production process and preparation of immobilized biocatalysts. *Electronic Journal of Biotechnology*, *16*(3).

Castillo, J. A., Guérard-Hélaine, C., Gutiérrez, M., Garrabou, X., Sancelme, M., Schürmann, M., Lemaire, M. (2010). A Mutant D-Fructose-6-Phosphate Aldolase (Ala129Ser) with Improved Affinity towards Dihydroxyacetone for the Synthesis of Polyhydroxylated Compounds. *Advanced Synthesis & Catalysis*, *352*(6), 1039–1046.

Ceroni, F., Furini, S., Giordano, E., & Cavalcanti, S. (2010). Rational design of modular circuits for gene transcription: A test of the bottom-up approach. *Journal of Biological Engineering*, *4*(1), 14.

Choi, J. H., Keum, K. C., & Lee, S. Y. (2006). Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chemical Engineering Science*, *61*(3), 876–885.

Constable, D. J. C., Dunn, P. J., Hayler, J. D., Humphrey, G. R., Leazer, Jr., J. L., Linderman, R. J., … Zhang, T. Y. (2007). Key green chemistry research areas-a perspective from pharmaceutical manufacturers. *Green Chemistry*, *9*(5), 411.

Cornelis, P. (2000). Expressing genes in different *Escherichia coli* compartments. *Current Opinion in Biotechnology*, *11*(5), 450–4.

Cuppoletti, J., & Segel, I. H. (1975). Kinetic analysis of active membrane transport systems: equations for net velocity and isotope exchange. *Journal of Theoretical Biology*, *53*(1), 125–44.

Donovan, R. S., Robinson, C. W., & Glick, B. R. (1996). Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. *Journal of Industrial Microbiology*, *16*(3), 145–54.

Durany, O., Caminal, G., de Mas, C., & López-Santín, J. (2004). Studies on the expression of recombinant fuculose-1-phosphate aldolase in *E. coli*. *Process Biochemistry*, *39*(11), 1677–1684.

Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends in Biotechnology*, *24*(11), 530–6.

Fan, J., Tuncay, K., & Ortoleva, P. J. (2007). Chromosome segregation in *Escherichia coli* division: a free energy-driven string model. *Computational Biology and Chemistry*, *31*(4), 257–64.

Fernández, A., Ruiz, J., Caminal, G., & López-Santín, J. (2010). Development and validation of a liquid chromatography-mass spectrometry assay for the quantitation of IPTG in *E. coli* fed-batch cultures. *Analytical Chemistry*, *82*(13), 5728–34.

Fernández-Castané, A., Caminal, G., & López-Santín, J. (2012). Direct measurements of IPTG enable analysis of the induction behavior of *E. coli* in high cell density cultures. *Microbial Cell Factories*, *11*, 58.

Fernández-Castané, A., Vine, C. E., Caminal, G., & López-Santín, J. (2012). Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. *Journal of Biotechnology*, *157*(3), 391–8.

Fernández, A (2012). Study of transport mechanisms involved in IPTG uptake by *e.coli* in high cell density cultures. Doctoral Thesis.

Fessner, W. D. (1998). Enzyme mediated C-C bond formation. *Current Opinion in Chemical Biology*, *2*(1), 85–97.

Fessner, W.-D., Sinerius, G., Schneider, A., Dreyer, M., Schulz, G. E., Badia, J., & Aguilar, J. (1991). Diastereoselective Enzymatic Aldol Additions: L-Rhamnulose and L-Fuculose 1-Phosphate Aldolases from *E. coli*. *Angewandte Chemie International Edition in English*, *30*(5), 555–558.

Fredrickson, A. ., Megee III, R. ., & Tsuchiya, H. . (1970). Mathematical Models for Fermentation Processes. *Advances in Applied Microbiology*, *13*, 419–465.

Garcia, M. L., Patel, L., Padan, E., & Kaback, H. R. (1982). Mechanism of lactose transport in *Escherichia coli* membrane vesicles: evidence for the involvement of a histidine residue(s) in the response of the *lac* carrier to the proton electrochemical gradient. *Biochemistry*, *21*(23), 5800–5805.

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

Gerlach, I., Brüning, S., Gustavsson, R., Mandenius, C.-F., & Hass, V. C. (2014). Operator training in recombinant protein production using a structured simulator model. *Journal of Biotechnology*, *177*, 53–9.

Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M., & Maire, P. (2008). The Hill equation: a review of its capabilities in pharmacological modelling. *Fundamental & Clinical Pharmacology*, *22*(6), 633–48.

Goyal, D., Sahni, G., & Sahoo, D. K. (2009). Enhanced production of recombinant streptokinase in *Escherichia coli* using fed-batch culture. *Bioresource Technology*, *100*(19), 4468–74.

Graumann, K., & Premstaller, A. (2006). Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnology Journal*, *1*(2), 164–86.

Guérard-Hélaine, C., Légeret, B., Fernandes, C., Prévot, V., Forano, C., & Lemaire, M. (2011). Efficient immobilization of fructose-6-phosphate aldolase in layered double hydroxide: improved stereoselective synthesis of sugar analogues. *New Journal of Chemistry*, *35*(4), 776.

Hansen, L. H., Knudsen, S., & Sørensen, S. J. (1998). The effect of the *lac*Y gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Current Microbiology*, *36*(6), 341–7.

Hoffmann, F., & Rinas, U. (2004). Stress induced by recombinant protein production in *Escherichia coli*. *Advances in Biochemical Engineering/biotechnology*, *89*, 73–92.

Horton, N., Lewis, M., & Lu, P. (1997). *Escherichia coli lac* repressor-*lac* operator interaction and the influence of allosteric effectors. *Journal of Molecular Biology*, *265*(1), 1–7.

Ichikawa, Y., Shue, Y.-K., Orida, N. K., Lotz, M., Wong, C.-H., & Okumu, F. W. (2004). Treatment of degenerative cartilage conditions in a mammal with glycosidase inhibitors.

Jensen, P. R., & Hammer, K. (1998). Artificial promoters for metabolic optimization. *Biotechnology and Bioengineering*, *58*(2-3), 191–5.

Jensen, P. R., Westerhoff, H. V, & Michelsen, O. (1993). The use of *lac*-type promoters in control analysis. *European Journal of Biochemistry / FEBS*, *211*(1-2), 181–91.

Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S.-H., Kim, J. F. (2009). Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 644–52.

Kavanagh, J., & Barton, G. W. (2008). Productivity improvement of recombinant *Escherichia coli* fermentation via robust optimization. *Bioprocess and Biosystems Engineering*, *31*(2), 137–43.

Koszelewski, D., Tauber, K., Faber, K., & Kroutil, W. (2010). ω -Transaminases for the synthesis of non-racemic α -chiral primary amines. *Trends in Biotechnology*, *28*(6), 324–32.

Kreimeyer, A., Perret, A., Lechaplais, C., Vallenet, D., Médigue, C., Salanoubat, M., & Weissenbach, J. (2007). Identification of the last unknown genes in the fermentation pathway of lysine. *The Journal of Biological Chemistry*, *282*(10), 7191–7.

Kroemer, M., Merkel, I., & Schulz, G. E. (2003). Structure and catalytic mechanism of L-rhamnulose-1-phosphate aldolase. *Biochemistry*, *42*(36), 10560–8.

Kroemer, M., & Schulz, G. E. (2002). The structure of L-rhamnulose-1-phosphate aldolase (class II) solved by low-resolution SIR phasing and 20-fold NCS averaging. *Acta Crystallographica. Section D, Biological Crystallography*, *58*(Pt 5), 824–32.

Kwon, S., Kim, S., & Kim, E. (1996). Effects of glycerol of β -lactamase production during high cell density cultivation of recombinant *Escherichia coli*. *Biotechnology Progress*, *12*(2), 205–8.

Laurent, M. (2005). Bistability and hysteresis in epigenetic regulation of the lactose operon. Since Delbruck, a long series of ignored models. *Cellular and Molecular Biology*, *51*(7), 583–94.

Lebedeva, M. I., Rogozhkina, E. V., Tsyba, N. A., & Mashko, S. V. (1994). A new T7 RNA polymerase-driven expression system induced via thermoamplification of a recombinant plasmid carrying a T7 promoter-*Escherichia coli lac* operator. *Gene*, *142*(1), 61–66.

Lee, J., & Ramirez, W. F. (1992a). Mathematical modeling of induced foreign protein production by recombinant bacteria. *Biotechnology and Bioengineering*, *39*(6), 635–46.

Lee, J., & Ramirez, W. F. (1992b). Mathematical modeling of induced foreign protein production by recombinant bacteria. *Biotechnology and Bioengineering*, *39*(6), 635–46.

Lee, S. Y. (1996). High cell-density culture of *Escherichia coli*. *Trends in Biotechnology*, *14*(3), 98–105.

Lei, Y., Ding, L., & Ding, Y. (2014). Generalization ability of fractional polynomial models. *Neural Networks : The Official Journal of the International Neural Network Society*, *49*, 59–73.

Liang, P.-H., Cheng, W.-C., Lee, Y.-L., Yu, H.-P., Wu, Y.-T., Lin, Y.-L., & Wong, C.-H. (2006). Novel five-membered iminocyclitol derivatives as selective and potent glycosidase inhibitors: new structures for antivirals and osteoarthritis. *Chembiochem : A European Journal of Chemical Biology*, *7*(1), 165–73.

Lopez, P. J., Guillerez, J., Sousa, R., & Dreyfus, M. (1998). On the mechanism of inhibition of phage T7 RNA polymerase by *lac* repressor. *Journal of Molecular Biology*, *276*(5), 861–75.

Lopez-Vernaza, M. A., & Leach, D. R. F. (2013). Symmetries and asymmetries associated with non-random segregation of sister DNA strands in *Escherichia coli*. *Seminars in Cell & Developmental Biology*, *24*(8-9), 610–7.

Lorentzen, E., Siebers, B., Hensel, R., & Pohl, E. (2005). Mechanism of the Schiff base forming fructose-1,6-bisphosphate aldolase: structural analysis of reaction intermediates. *Biochemistry*, *44*(11), 4222–9.

Machajewski, T., & Wong, C. (2000). The Catalytic Asymmetric Aldol Reaction. *Angewandte Chemie (International Ed. in English)*, *39*(8), 1352–1375.

Mahadevan, R., & Doyle, F. J. (2003). On-line optimization of recombinant product in a fed-batch bioreactor. *Biotechnology Progress*, *19*(2), 639–46.

Maresová, H., Stepánek, V., & Kyslík, P. (2001). A chemostat culture as a tool for the improvement of a recombinant *E. coli* strain over-producing penicillin G acylase. *Biotechnology and Bioengineering*, *75*(1), 46–52.

Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Review of Microbiology*, *3*(1), 371–394.

Nadri, M., Trezzani, I., Hammouri, H., Dhurjati, P., Longin, R., & Lieto, J. (2006). Modeling and observer design for recombinant *Escherichia coli* strain. *Bioprocess and Biosystems Engineering*, *28*(4), 217–25.

Nancib, N., Mosrati, R., & Boudrant, J. (1993). Modelling of batch fermentation of a recombinant *Escherichia coli* producing glyceraldehyde-3-phosphate

dehydrogenase on a complex selective medium. *The Chemical Engineering Journal*, *52*(2), B35–B48.

Nataro, J. P., & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, *11*(1), 142–201.

Noel, J. T., Pilyugin, S. S., & Narang, A. (2009). The diffusive influx and carrier efflux have a strong effect on the bistability of the *lac* operon in *Escherichia coli*. *Journal of Theoretical Biology*, *256*(1), 14–28.

Noronha, S. B., Yeh, H. J., Spande, T. F., & Shiloach, J. (2000). Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using (13)C-NMR/MS. *Biotechnology and Bioengineering*, *68*(3), 316–27.

Nugent, T. C., & El-Shazly, M. (2010). Chiral Amine Synthesis - Recent Developments and Trends for Enamide Reduction, Reductive Amination, and Imine Reduction. *Advanced Synthesis & Catalysis*, *352*(5), 753–819.

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., & Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, *427*(6976), 737–40.

Palomo, C., Oiarbide, M., & García, J. M. (2002). The aldol addition reaction: an old transformation at constant rebirth. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, *8*(1), 36–44.

Panke, S., Held, M., & Wubbolts, M. (2004). Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Current Opinion in Biotechnology*, *15*(4), 272–9.

Parker, R. S., & Doyle, F. J. (2001). Optimal Control of a Continuous Bioreactor Using an Empirical Nonlinear Model. *Industrial & Engineering Chemistry Research*, *40*(8), 1939–1951.

Pendashteh, A. R., Fakhru'l-Razi, A., Chaibakhsh, N., Abdullah, L. C., Madaeni, S. S., & Abidin, Z. Z. (2011). Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network. *Journal of Hazardous Materials*, *192*(2), 568–75.

Pinsach, J. (2009) Development of recombinant aldolase production process in *Escherichia coli.* Doctoral Thesis

Pinsach, J., de Mas, C., & Lopez-Santin, J. (2006). A simple feedback control of *Escherichia coli* growth for recombinant aldolase production in fed-batch mode. *Biochemical Engineering Journal*, *29*, 235–242.

Pinsach, J., de Mas, C., & López-Santín, J. (2008). Induction strategies in fed-batch cultures for recombinant protein production in *Escherichia coli*: Application to rhamnulose 1-phosphate aldolase. *Biochemical Engineering Journal*, *41*(2), 181–187.

Pinsach, J., de Mas, C., López-Santín, J., Striedner, G., & Bayer, K. (2008). Influence of process temperature on recombinant enzyme activity in *Escherichia coli* fed-batch cultures. *Enzyme and Microbial Technology*, *43*(7), 507–512.

Pollard, D. J., & Woodley, J. M. (2007). Biocatalysis for pharmaceutical intermediates: the future is now. *Trends in Biotechnology*, *25*(2), 66–73.

Rao, C. V, Wolf, D. M., & Arkin, A. P. (2002). Control, exploitation and tolerance of intracellular noise. *Nature*, *420*(6912), 231–7.

Robichon, C., Luo, J., Causey, T. B., Benner, J. S., & Samuelson, J. C. (2011). Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Applied and Environmental Microbiology*, *77*(13), 4634–46.

Ruiz, J., González, G., de Mas, C., & López-Santín, J. (2011). A semiempirical model to control the production of a recombinant aldolase in high cell density cultures of Escherichia coli. *Biochemical Engineering Journal*, *55*(2), 82–91.

Ruiz, J., Pinsach, J., Álvaro, G., González, G., de Mas, C., Resina, D., & López-Santín, J. (2009). Alternative production process strategies in *E. coli* improving protein quality and downstream yields. *Process Biochemistry*, *44*(9), 1039–1045.

Sakamoto, S., Iijima, M., Matsuzawa, H., & Ohta, T. (1994). Production of thermophilic protease by glucose-controlled fed-batch culture of recombinant *Escherichia coli*. *Journal of Fermentation and Bioengineering*, *78*(4), 304–309.

Samland, A. K., & Sprenger, G. A. (2006). Microbial aldolases as C-C bonding enzymes--unknown treasures and new developments. *Applied Microbiology and Biotechnology*, *71*(3), 253–64.

Sánchez-Moreno, I., Nauton, L., Théry, V., Pinet, A., Petit, J.-L., de Berardinis, V., Lemaire, M. (2012). FSAB: A new fructose-6-phosphate aldolase from *Escherichia coli*. Cloning, over-expression and comparative kinetic characterization with FSAA. *Journal of Molecular Catalysis B: Enzymatic*, *84*, 9–14.

Santillán, M. (2008). Bistable behavior in a model of the *lac* operon in *Escherichia coli* with variable growth rate. *Biophysical Journal*, *94*(6), 2065–81.

Santillán, M., & Mackey, M. C. (2004). Influence of catabolite repression and inducer exclusion on the bistable behavior of the *lac* operon. *Biophysical Journal*, *86*(3), 1282–92.

Saotome, C., Kanie, Y., Kanie, O., & Wong, C. H. (2000). Synthesis and enzymatic evaluation of five-membered iminocyclitols and a pseudodisaccharide. *Bioorganic & Medicinal Chemistry*, *8*(9), 2249–61.

Schürmann, M., Schürmann, M., & Sprenger, G. A. (2002). Fructose 6-phosphate aldolase and 1-deoxy-d-xylulose 5-phosphate synthase from *Escherichia coli* as tools in enzymatic synthesis of 1-deoxysugars. *Journal of Molecular Catalysis B: Enzymatic*, *19-20*, 247–252.

Schurmann, M., & Sprenger, G. A. (2001). Fructose-6-phosphate aldolase is a novel class I aldolase from *Escherichia coli* and is related to a novel group of bacterial transaldolases. *The Journal of Biological Chemistry*, *276*(14), 11055–61.

Sevastsyanovich, Y., Alfasi, S., Overton, T., Hall, R., Jones, J., Hewitt, C., & Cole, J. (2009). Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. *FEMS Microbiology Letters*, *299*(1), 86–94.

Shiloach, J., Kaufman, J., Guillard, a S., & Fass, R. (1996). Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by *Escherichia coli* BL21 (λDE3) and *Escherichia coli* JM109. *Biotechnology and Bioengineering*, *49*(4), 421–8.

Shin, J.-S., & Kim, B.-G. (2002). Exploring the Active Site of Amine:Pyruvate Aminotransferase on the Basis of the Substrate Structure−Reactivity Relationship: How the Enzyme Controls Substrate Specificity and Stereoselectivity. *The Journal of Organic Chemistry*, *67*(9), 2848–2853.

Shin, J.-S., Kim, B.-G., & Shin, D.-H. (2001). Kinetic resolution of chiral amines using packed-bed reactor. *Enzyme and Microbial Technology*, *29*(4-5), 232–239.

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, *115*(2), 113–28.

Sugiyama, M., Hong, Z., Liang, P., Dean, S. M., Whalen, L. J., Greenberg, W. A., & Wong, C. (2007). D-Fructose-6-phosphate aldolase-catalyzed one-pot synthesis of iminocyclitols. *Journal of the American Chemical Society*, *129*(47), 14811–7.

Takayama, S., McGarvey, G. J., & Wong, C. H. (1997). Microbial aldolases and transketolases: new biocatalytic approaches to simple and complex sugars. *Annual Review of Microbiology*, *51*, 285–310.

Tamerler, C., Ulgen, K., Kirdar, B., & Ilsen Önsan, Z. (2001). A structured model for intracellular EcoRI endonuclease production by recombinant *E. coli* 294. *Process Biochemistry*, *36*(7), 621–627.

Tang, S., Chen, J., & Zhang, Z. (2007). Structured models for recombinant human interleukin-11 fermentation. *Biochemical Engineering Journal*, *35*(2), 210–217.

Thilakavathi, M., Basak, T., & Panda, T. (2006). Modeling of enzyme production kinetics. *Applied Microbiology and Biotechnology*, *73*(5), 991–1007.

Toksoy Öner, E., Kirdar, B., Önsan, Z., & Ülgen, K. . (2003). Modeling of the induced expression for high-level production of a foreign protein by recombinant *E. coli* under the control of the T7 phage promoter. *Process Biochemistry*, *39*(3), 315–323.

Turner, C., Gregory, M. E., & Turner, M. K. (1994). A study of the effect of specific growth rate and acetate on recombinant protein production of *Escherichia coli* JM107. *Biotechnology Letters*, *16*(9), 891–896.

Vidal, L., Durany, O., Suau, T., Ferrer, P., Benaiges, M., & Caminal, G. (2003). High-level production of recombinant His-tagged rhamnulose 1-phosphate aldolase in *Escherichia coli*. *Journal of Chemical Technology & Biotechnology*, *78*(11), 1171–1179.

Vidal, L., Ferrer, P., Alvaro, G., Benaiges, M. D., & Caminal, G. (2005). Influence of induction and operation mode on recombinant rhamnulose 1-phosphate aldolase production by *Escherichia coli* using the T5 promoter. *Journal of Biotechnology*, *118*(1), 75–87.

Vidal, L., Pinsach, J., Striedner, G., Caminal, G., & Ferrer, P. (2008). Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in *Escherichia coli* . *Journal of Biotechnology*, *134*(1-2), 127–36.

Vilar, J. M. G., Guet, C. C., & Leibler, S. (2003). Modeling network dynamics: the *lac* operon, a case study. *The Journal of Cell Biology*, *161*(3), 471–6.

Vilar, J. M. G., & Saiz, L. (2013). Reliable prediction of complex phenotypes from a modular design in free energy space: an extensive exploration of the lac operon. *ACS Synthetic Biology*, *2*(10), 576–86.

Wolfe, A. J. (2005). The acetate switch. *Microbiology and Molecular Biology Reviews : MMBR*, *69*(1), 12–50.

Wright, D. W., Moreno-Vargas, A. J., Carmona, A. T., Robina, I., & Davies, G. J. (2013). Three dimensional structure of a bacterial α-l-fucosidase with a 5-membered iminocyclitol inhibitor. *Bioorganic & Medicinal Chemistry*, *21*(16), 4751–4.

Wymer, N., & Toone, E. J. (2000). Enzyme-catalyzed synthesis of carbohydrates. *Current Opinion in Chemical Biology*, *4*(1), 110–9.

Xu, J., Banerjee, A., Pan, S.-H., & Li, Z. J. (2012). Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains. *Protein Expression and Purification*, *83*(1), 30–6.

Yee, L., & Blanch, H. W. (1992). Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Bio/technology (Nature Publishing Company)*, *10*(12), 1550–6.

Zhang, X.-W., Sun, T., Liu, X., Gu, D.-X., & Tang, Z.-Q. (1999). Production of granulocyte-macrophage colony stimulating factor (GM-CSF) by high cell density fermentation of secretory recombination *Escherichia coli*. *Process Biochemistry*, *34*(1), 55–58.

Zheng, Z.-Y., Yao, S.-J., & Lin, D.-Q. (2005). Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *Escherichia coli*. *Bioprocess and Biosystems Engineering*, *27*(3), 143–52.

10 Appendix

10.1 Growth profiles

The following figures show the time evolution of biomass and glucose concentration, as well as the total volume for all the experiments used in this work.

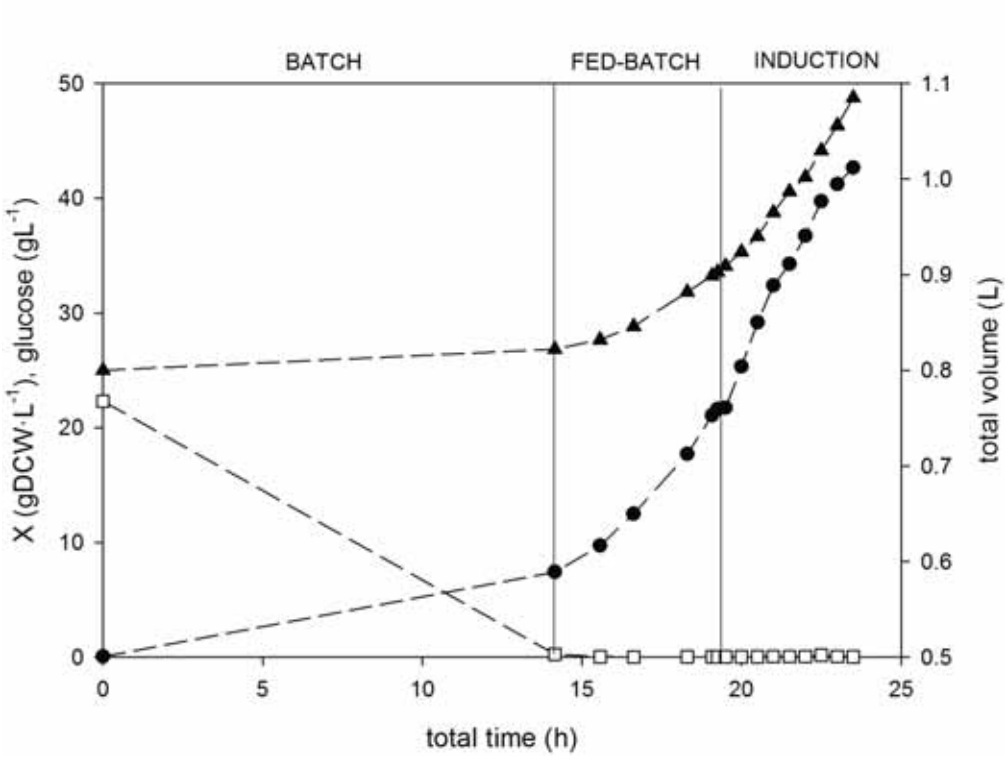


Figure 10.1. lac Y mutant strain. $[IPTG]_{e,0}$ = 10 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

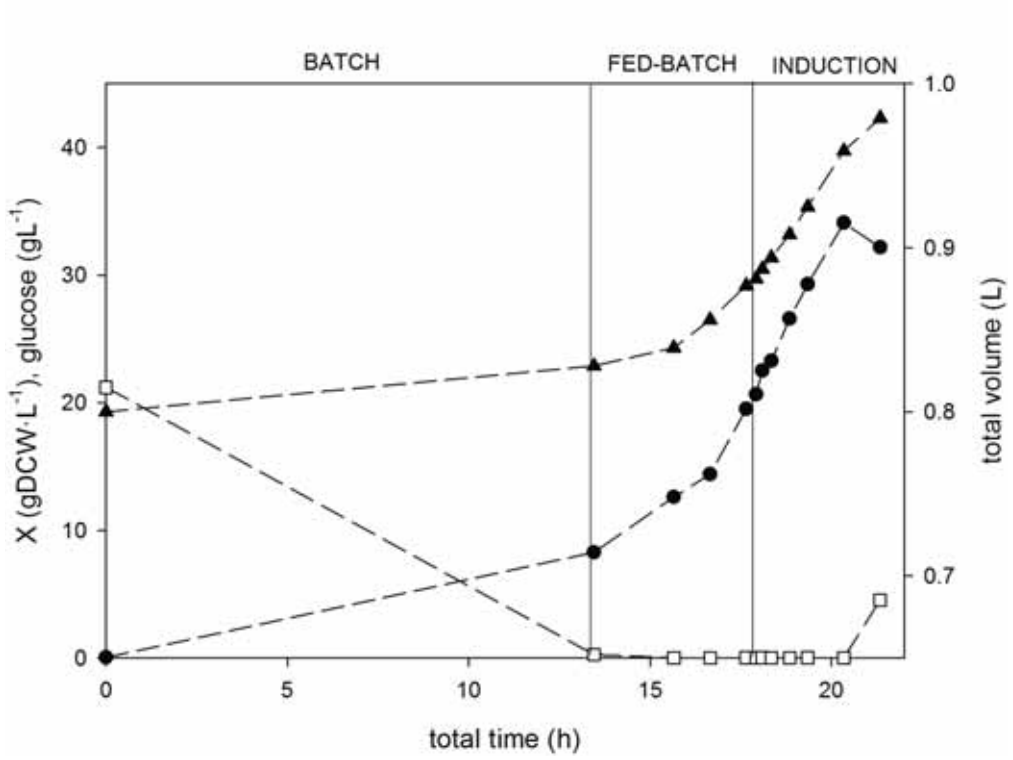


Figure 10.2. lac Y mutant strain. $[IPTG]_{e,0}$ = 20 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

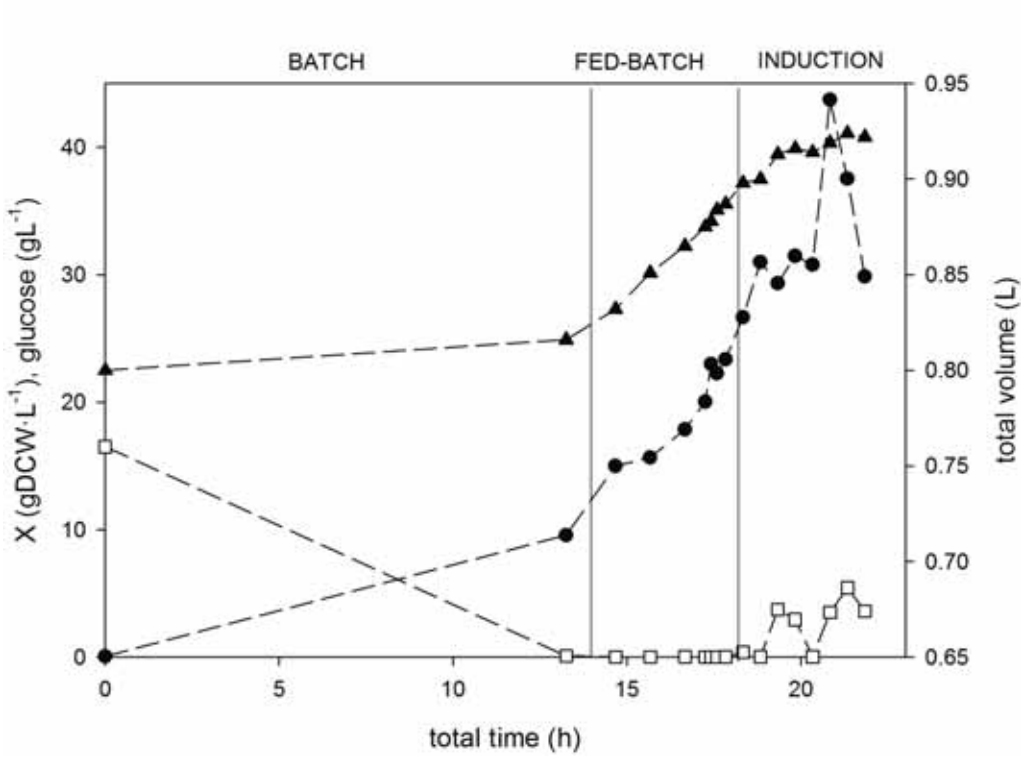


Figure 10.3. lac Y mutant strain. $[IPTG]_{e,0}$ = 54 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

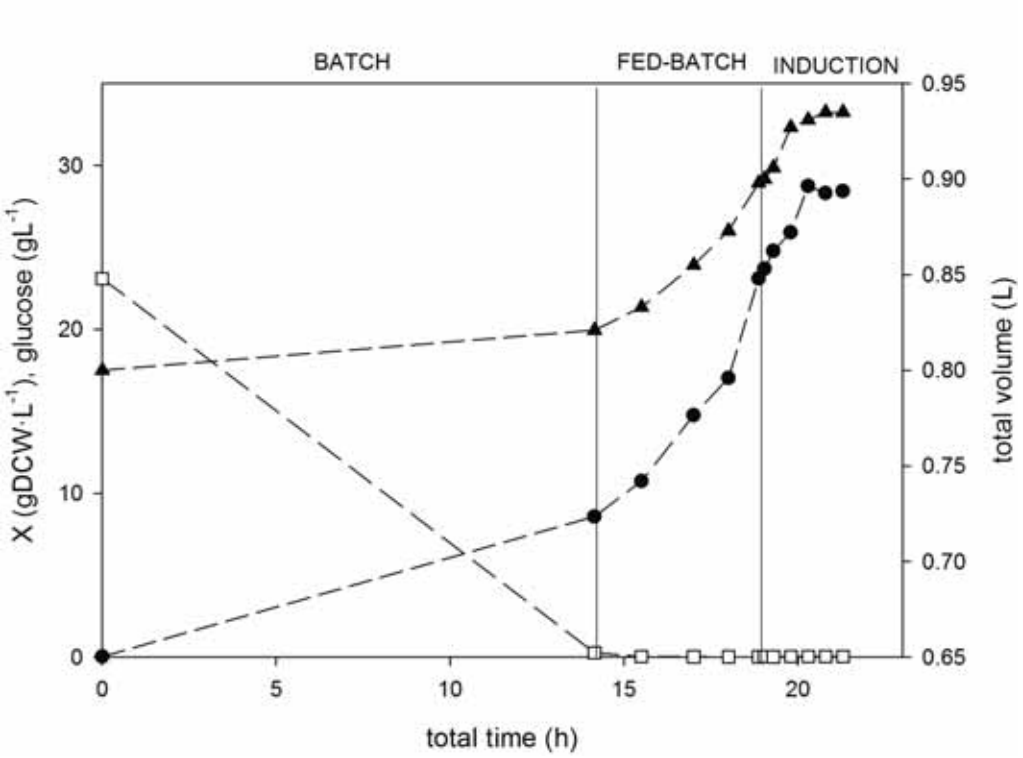


Figure 10.4. lac Y mutant strain. $[IPTG]_{e,0}$ = 200 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

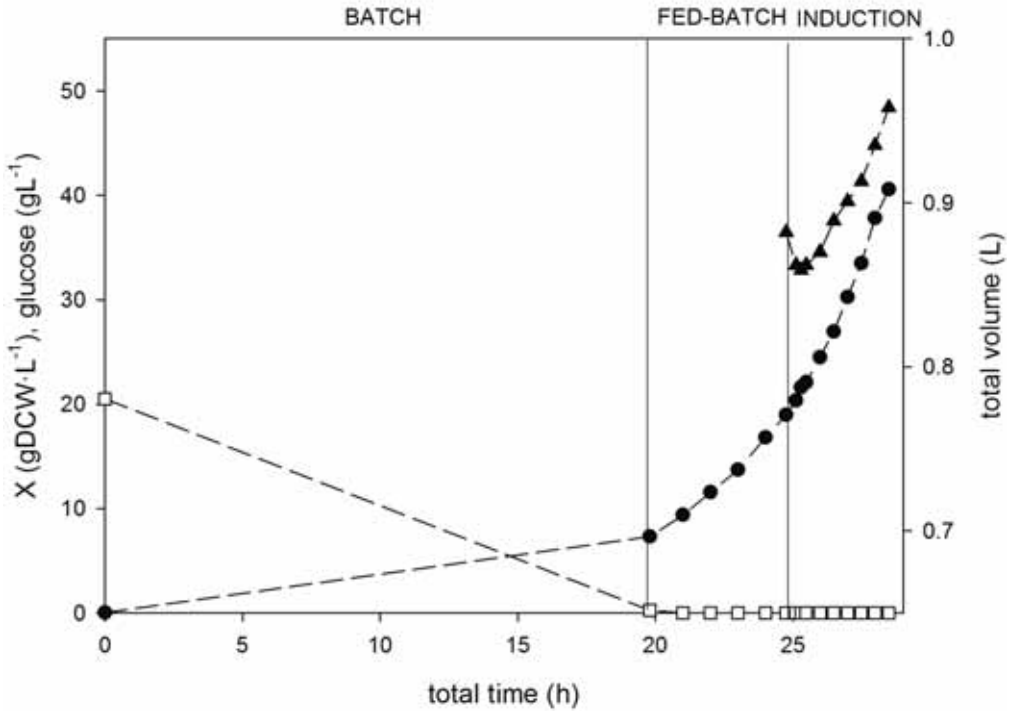


Figure 10.5. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 8 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

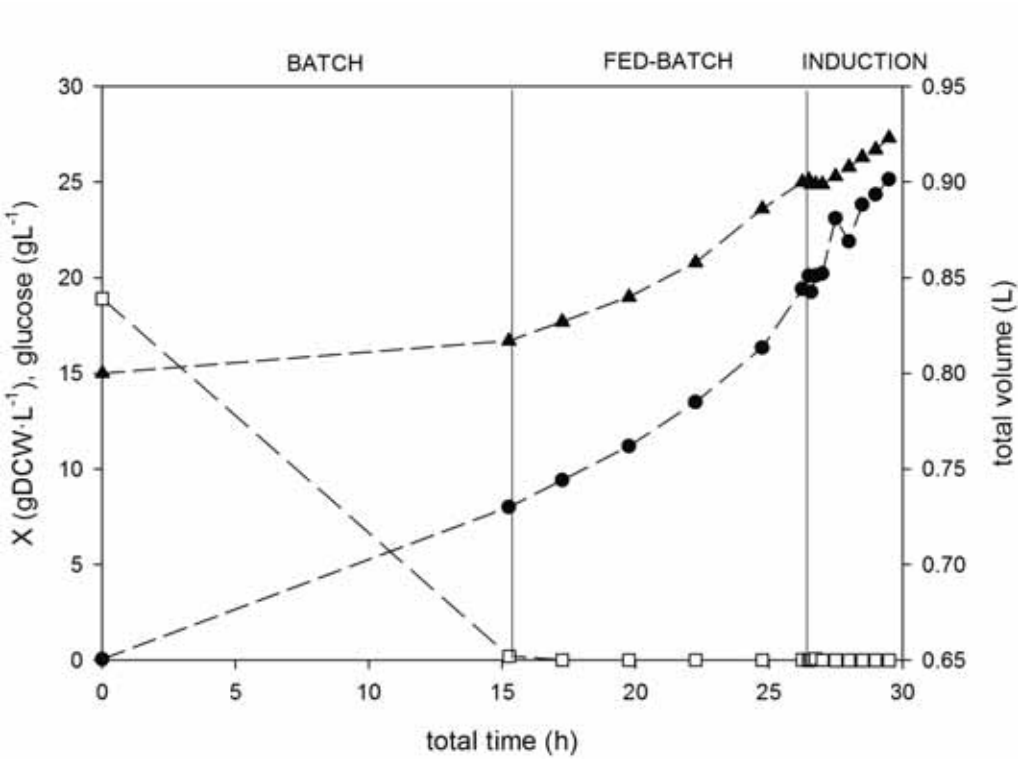


Figure 10.6. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 10 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

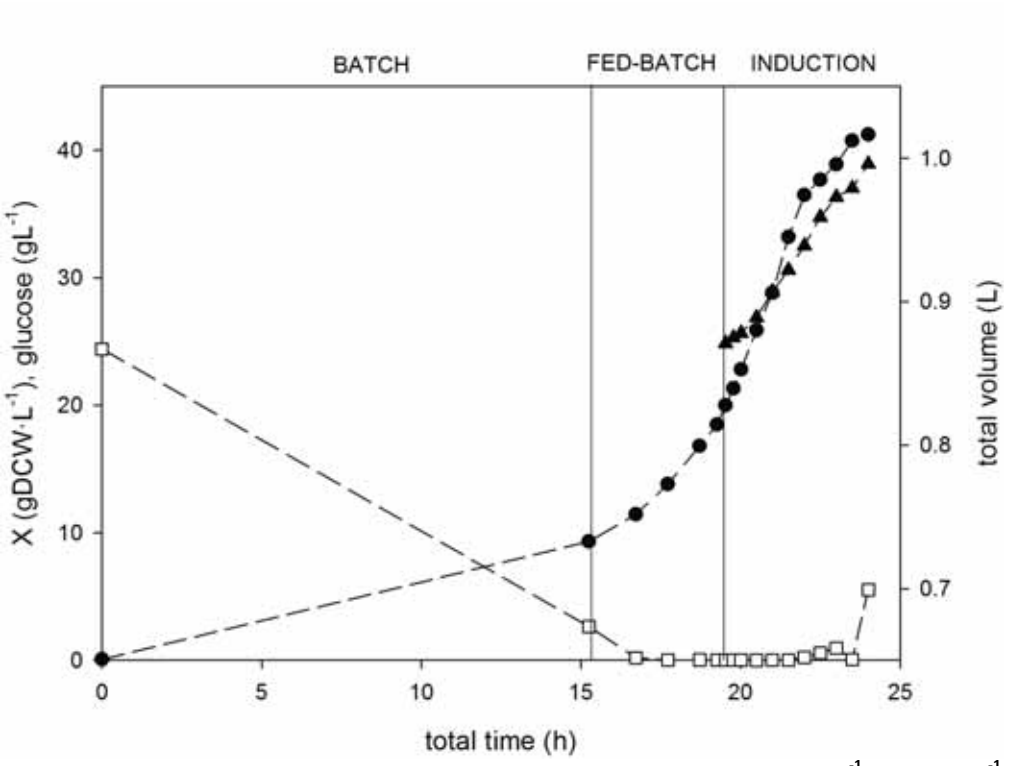


Figure 10.7. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 10 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

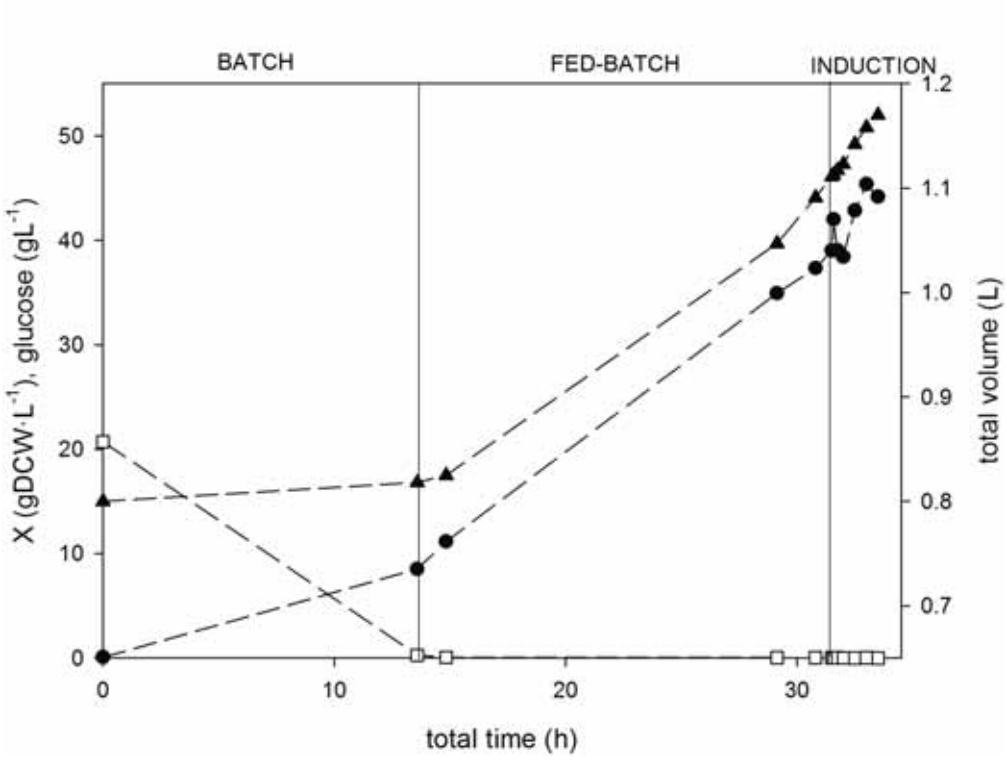


Figure 10.8. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 10 µM; X_{ind} = 40 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

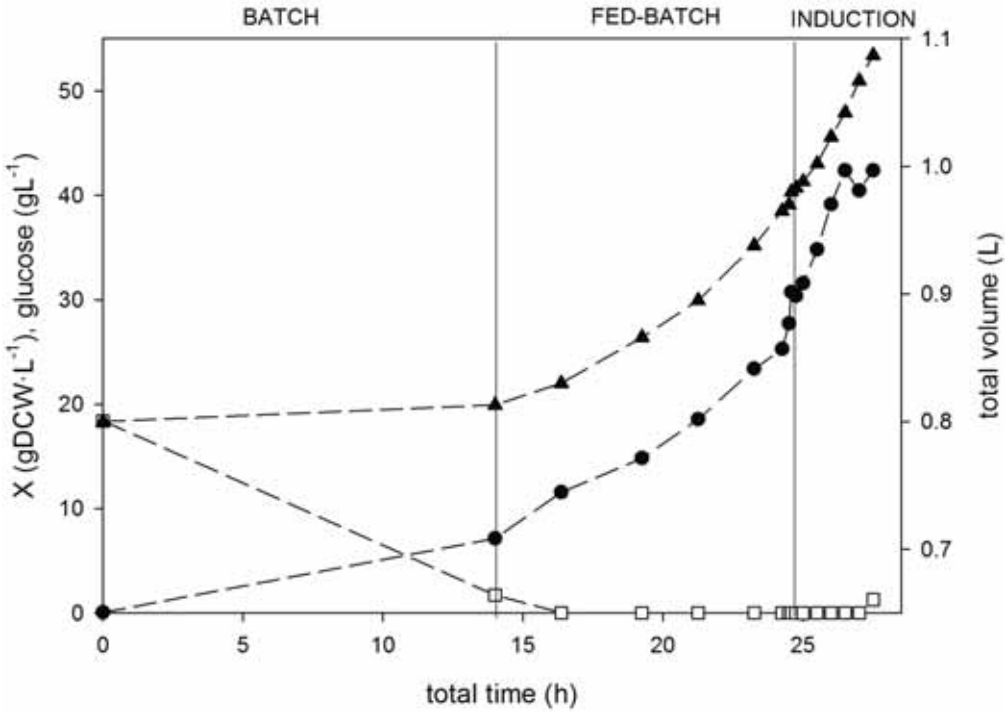


Figure 10.9. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 15 µM; X_{ind} = 30 gDCW·L^{-1}; μ_{fix} = 0.16 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

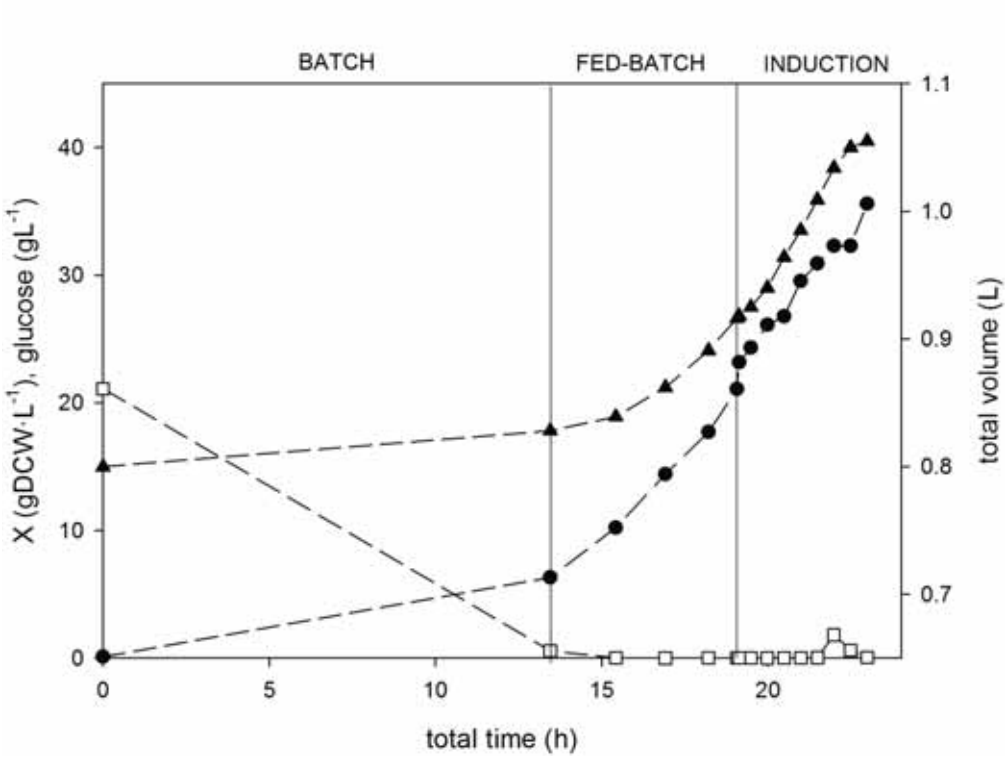


Figure 10.10. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 20 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

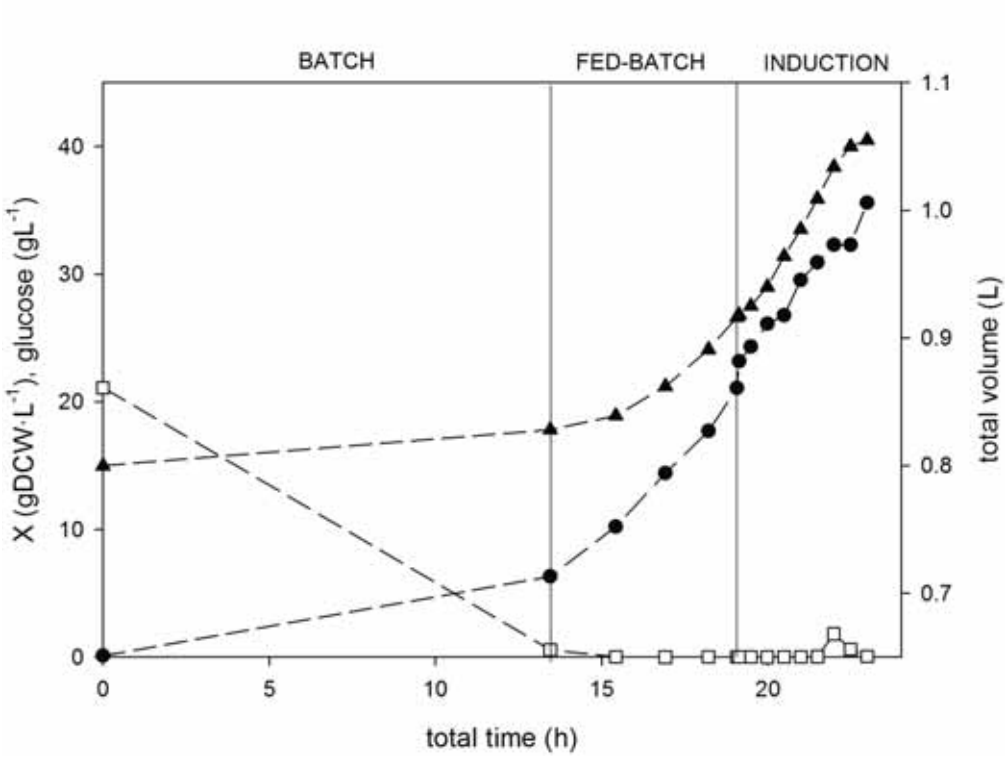


Figure 10.11. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 24 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

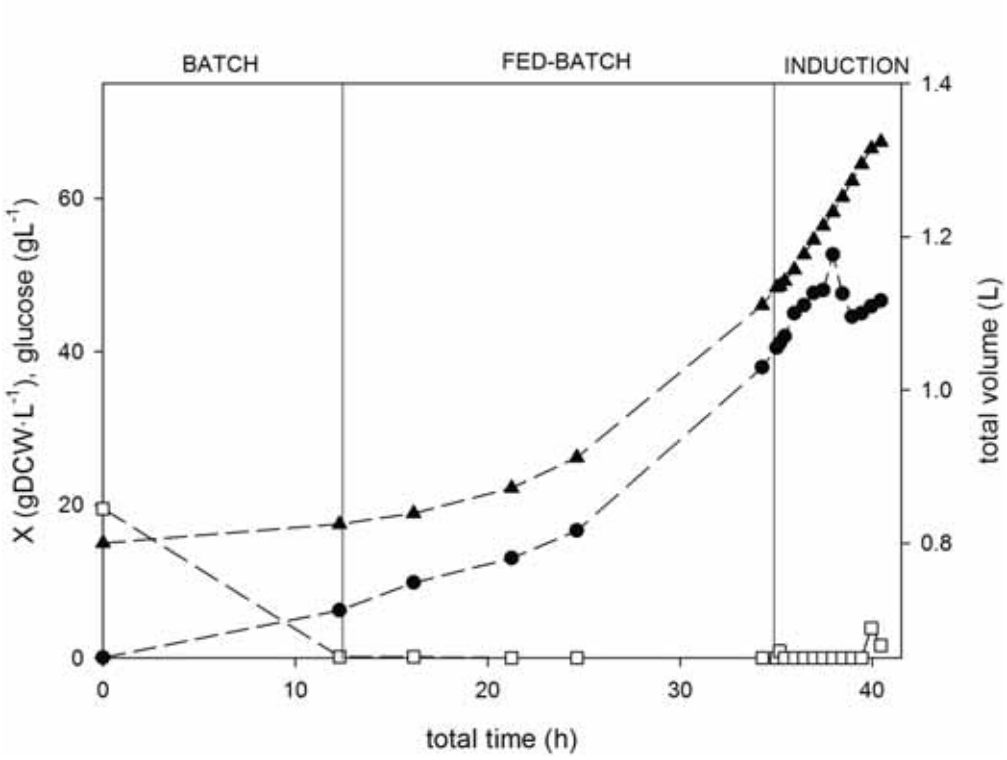


Figure 10.12. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 24 µM; X_{ind} = 40 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

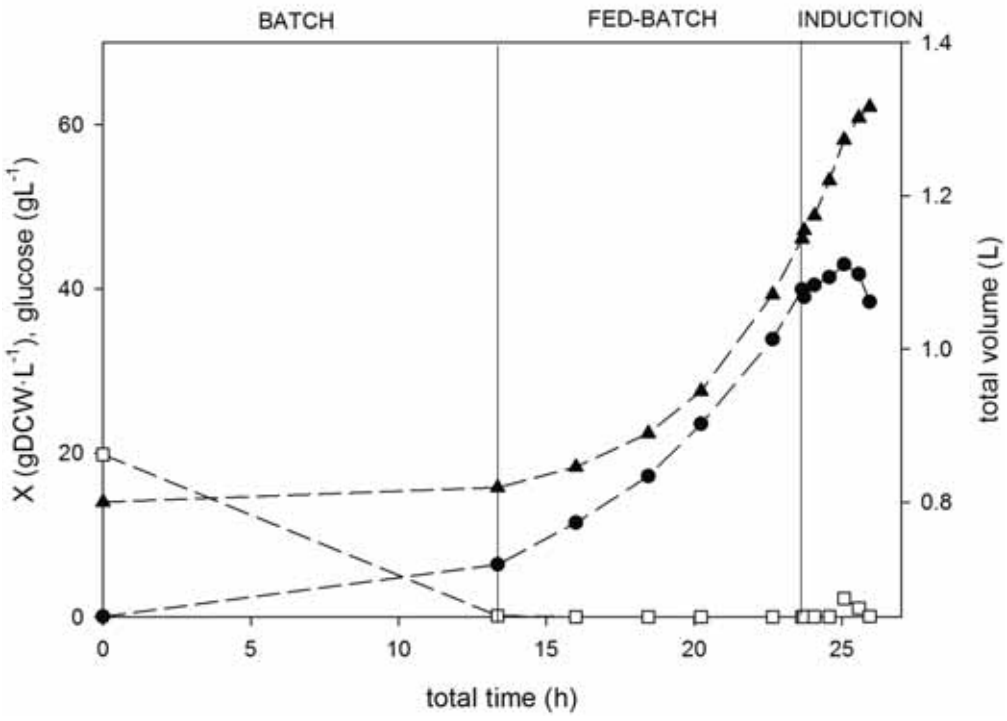


Figure 10.13. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 24 µM; X_{ind} = 40 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

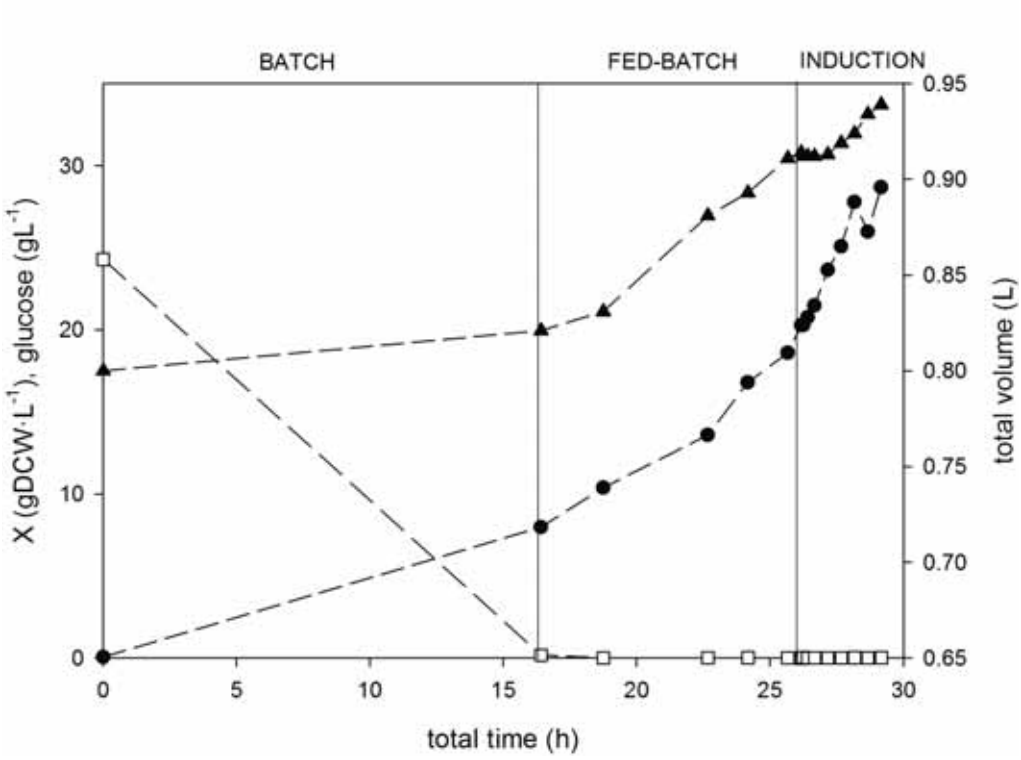


Figure 10.14. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 27 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

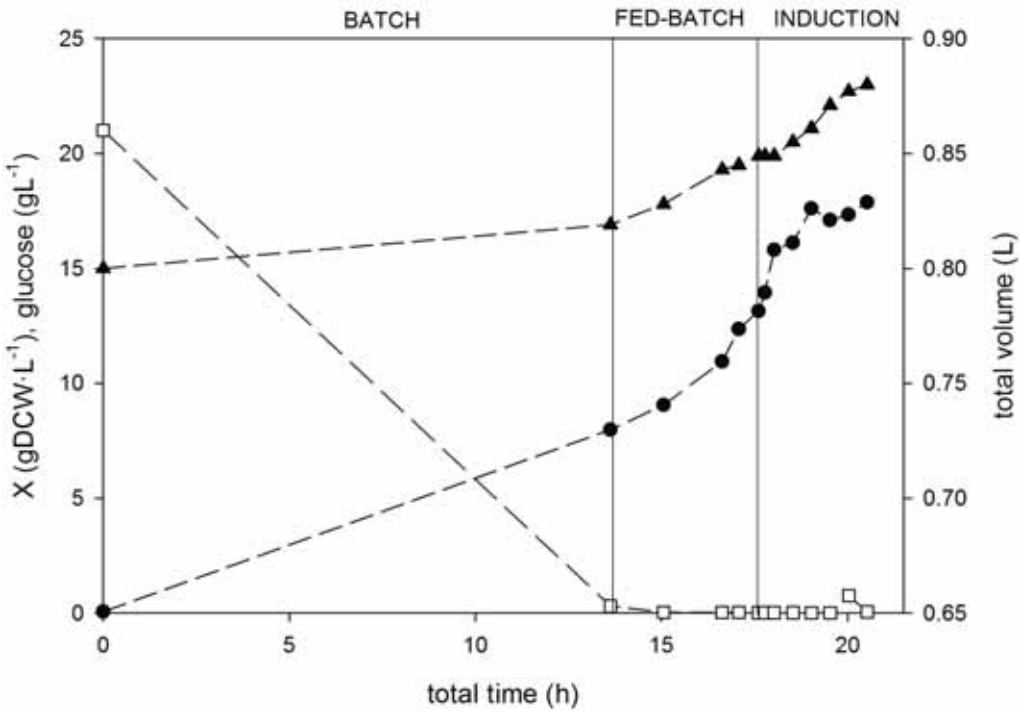


Figure 10.15. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 60 µM; X_{ind} = 13 gDCW·L^{-1}; μ_{fix} = 0.16 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

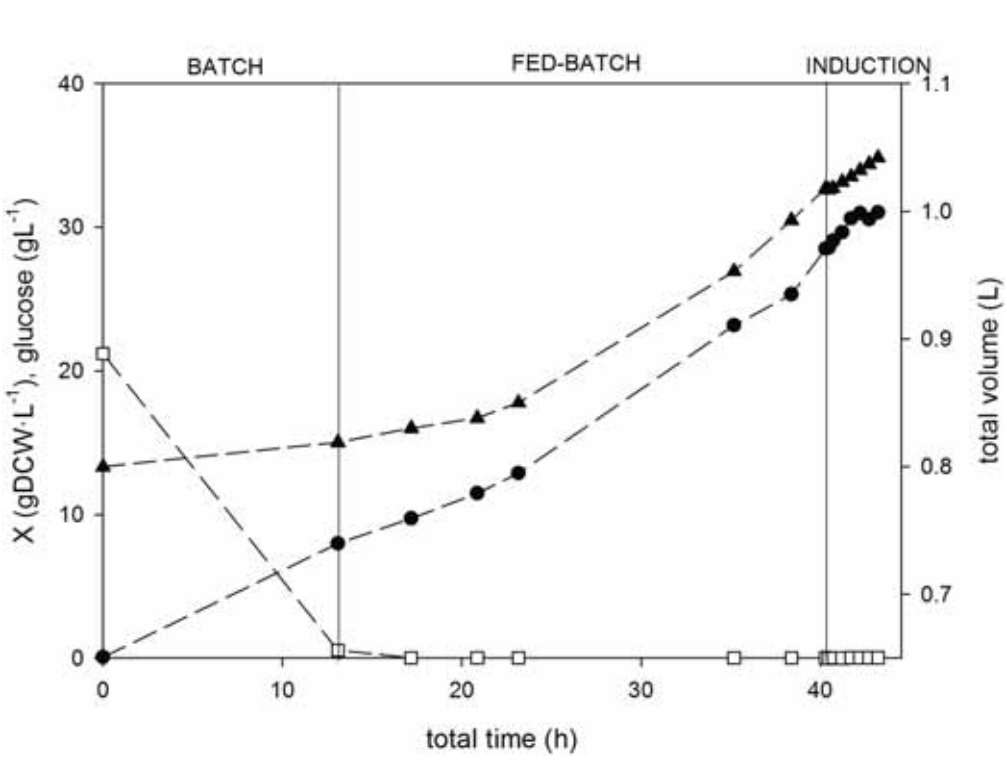


Figure 10.16. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 60 µM; X_{ind} = 30 gDCW·L^{-1}; μ_{fix} = 0.06 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

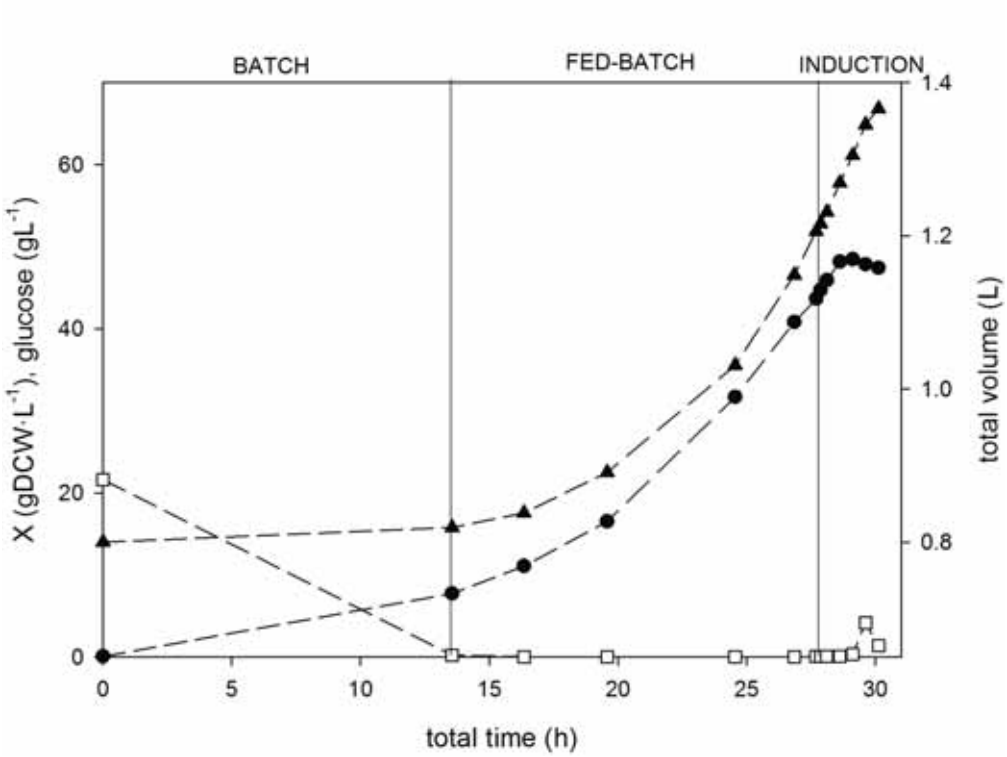


Figure 10.17. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 60 µM; X_{ind} = 47 gDCW·L^{-1}; μ_{fix} = 0.16 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

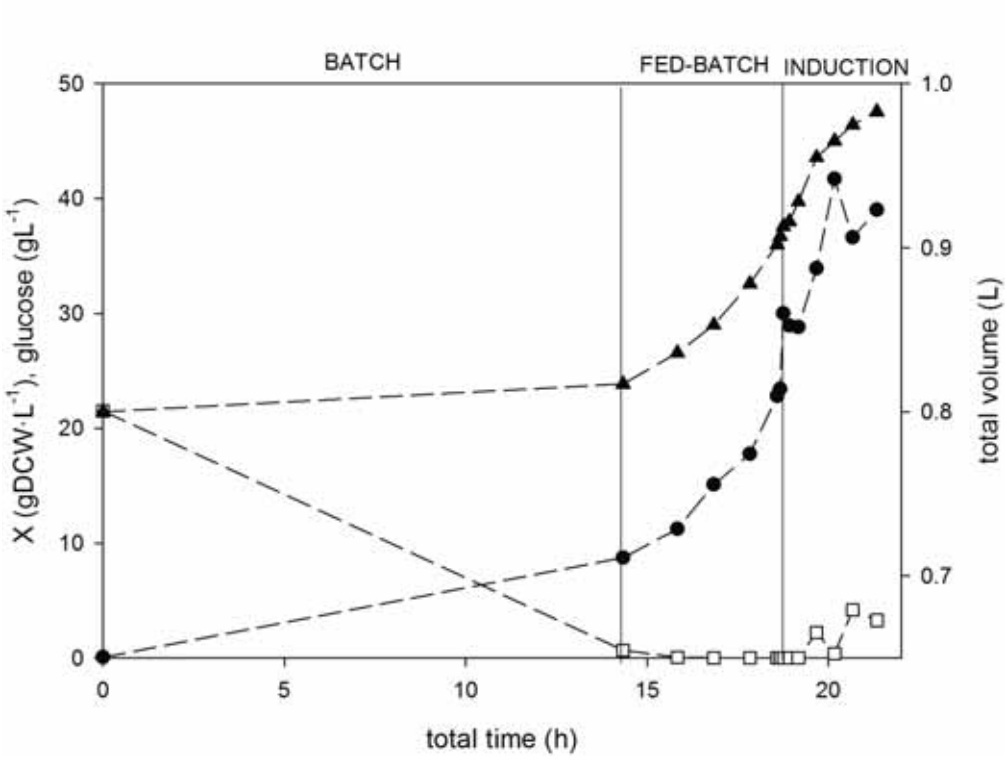


Figure 10.18. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 70 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

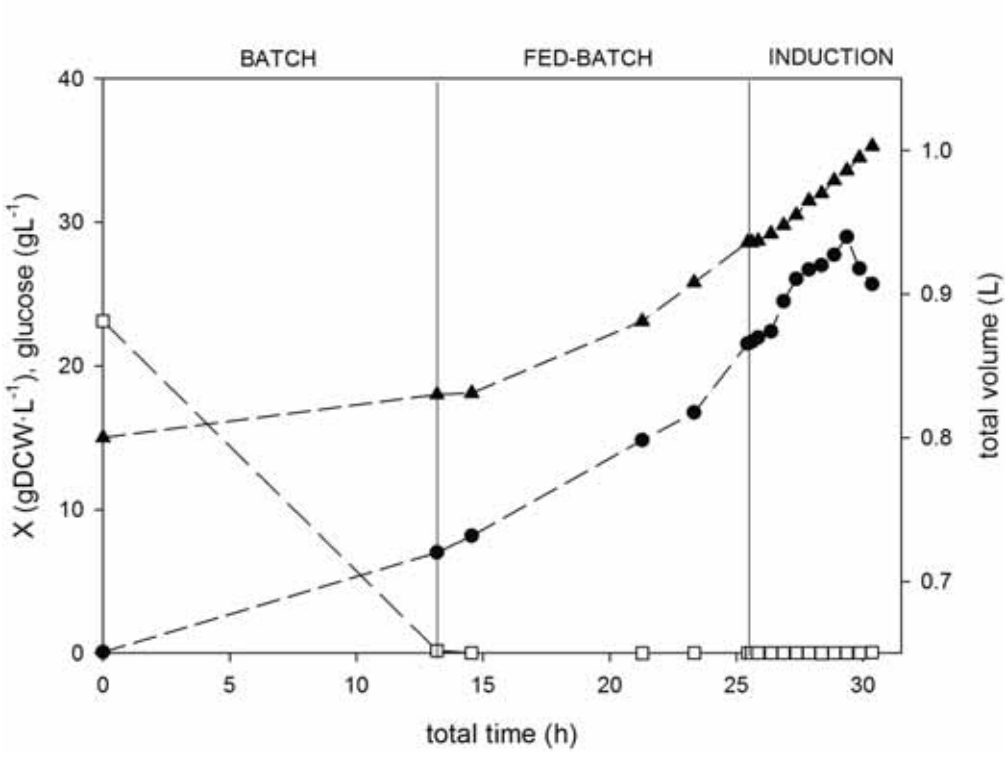


Figure 10.19. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 100 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

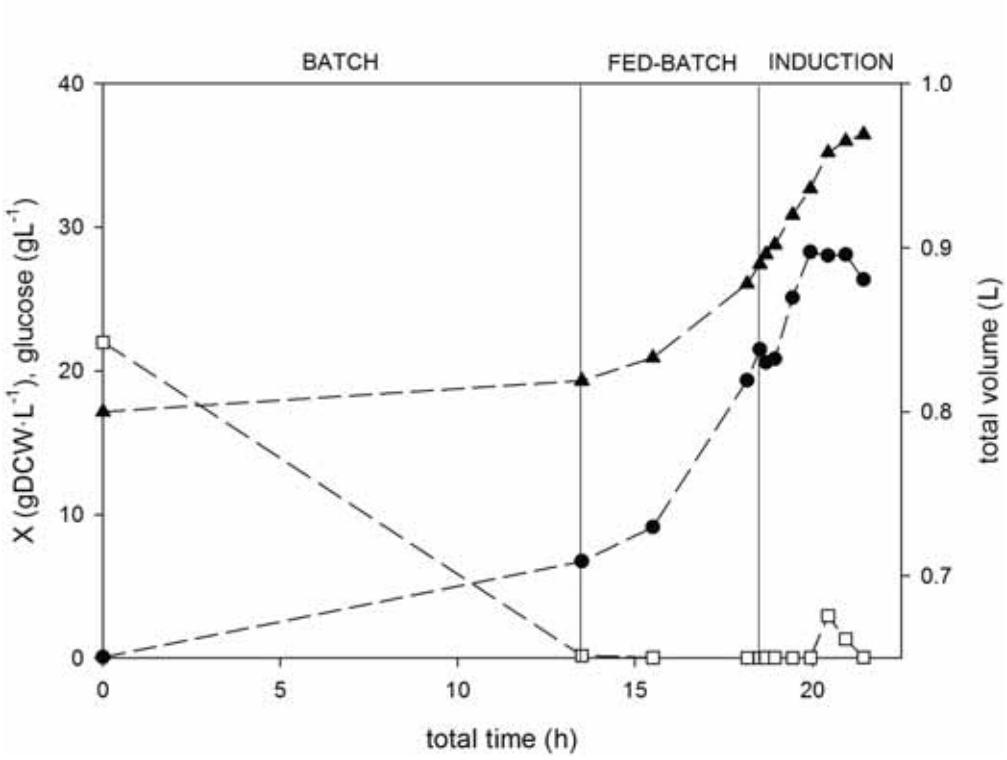


Figure 10.20. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 100 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

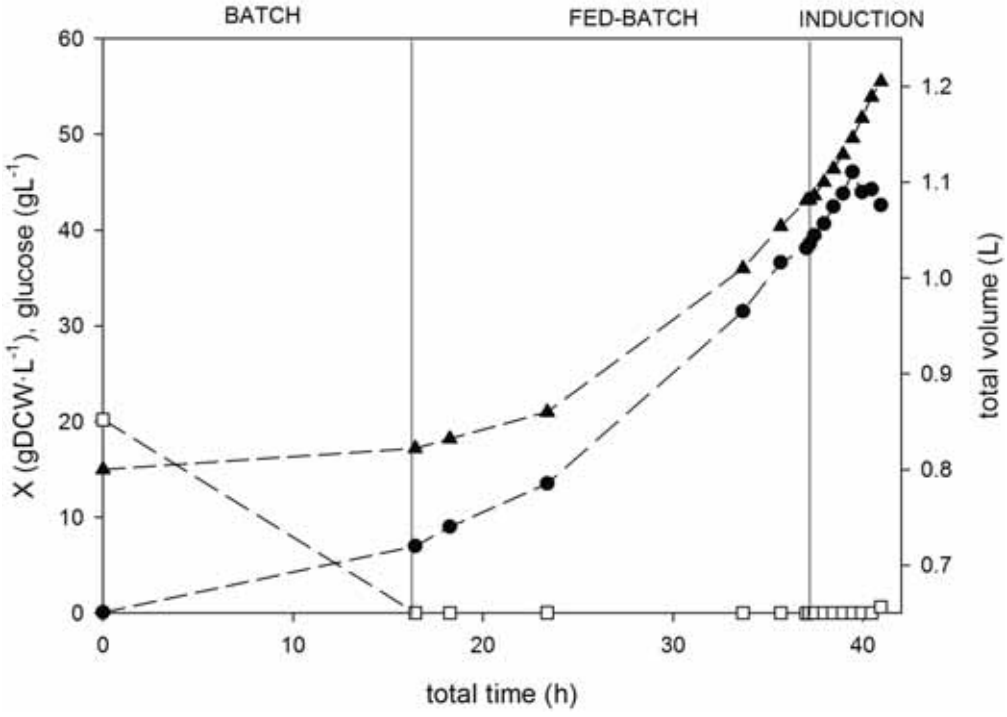


Figure 10.21. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 100 µM; X_{ind} = 40 gDCW·L^{-1}; μ_{fix} = 0.1 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

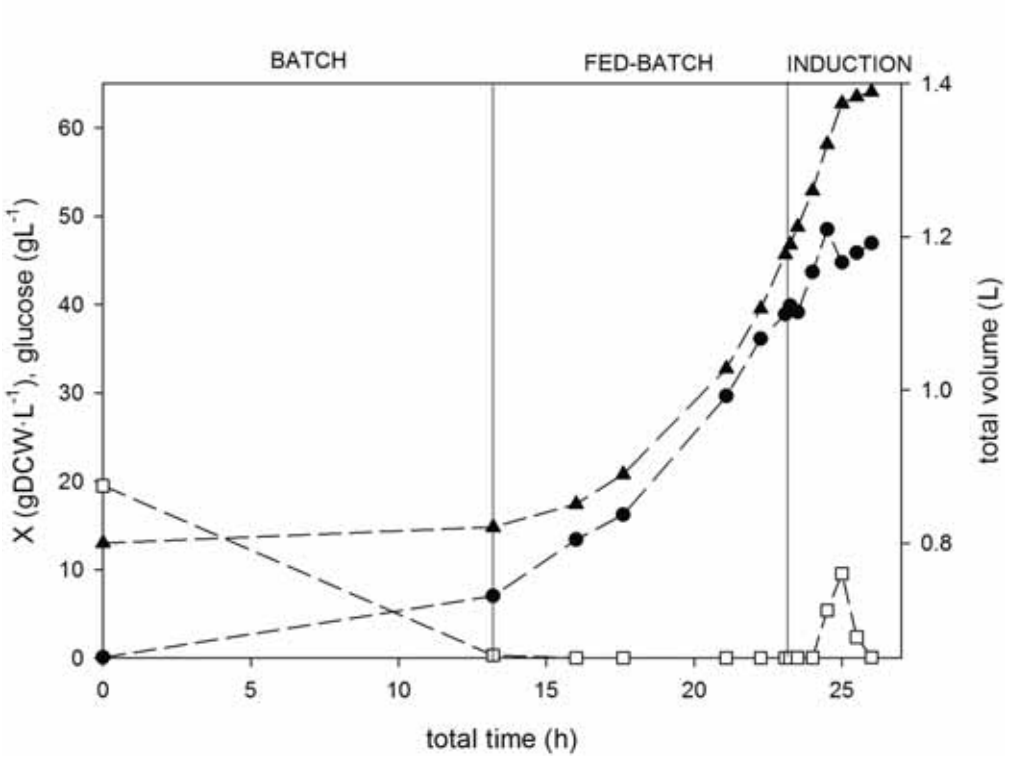


Figure 10.22. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 100 µM; X_{ind} = 40 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

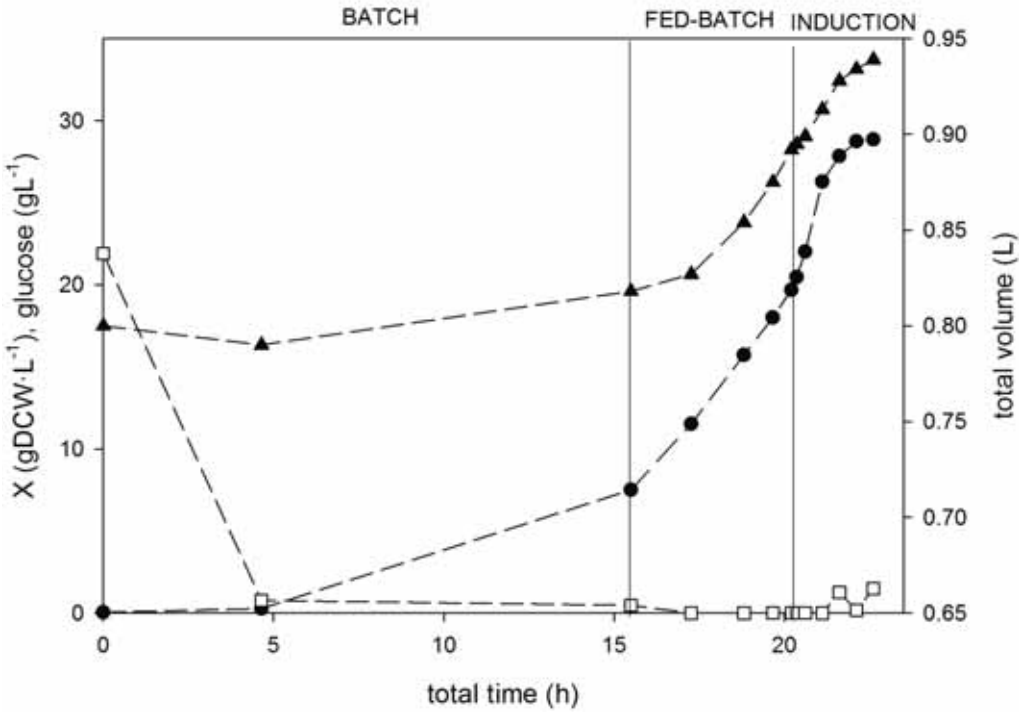


Figure 10.23. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 200 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

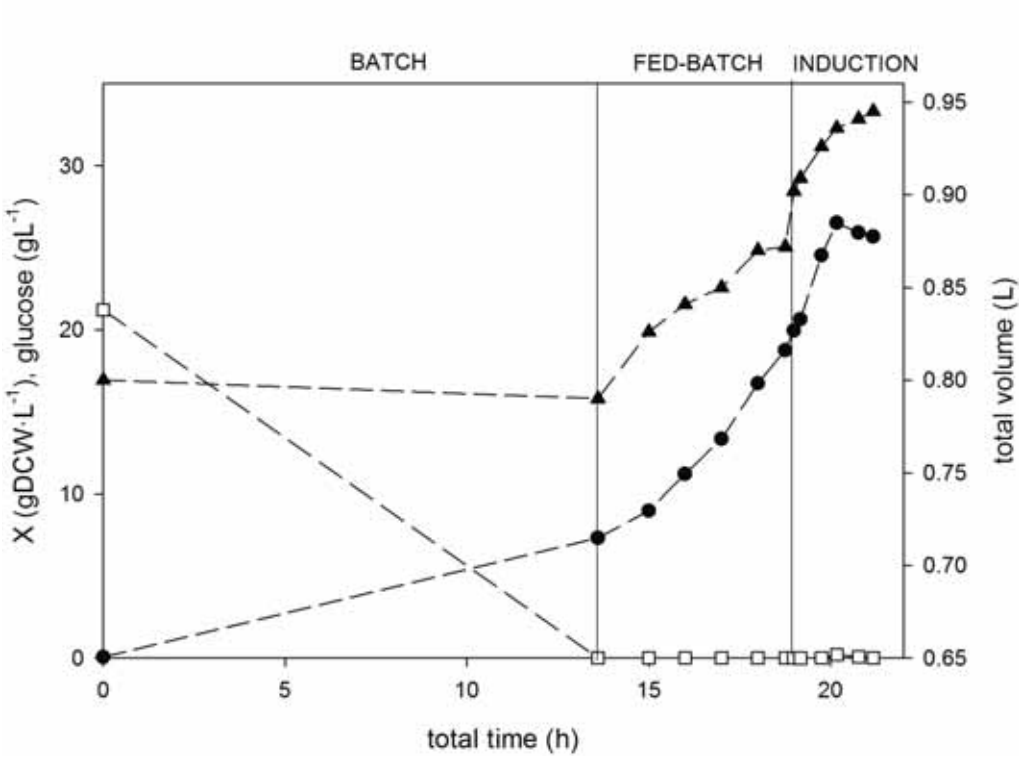


Figure 10.24. M15 Δ*gly*A[pQEαβrham][pREP4] strain. $[IPTG]_{e,0}$ = 1000 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

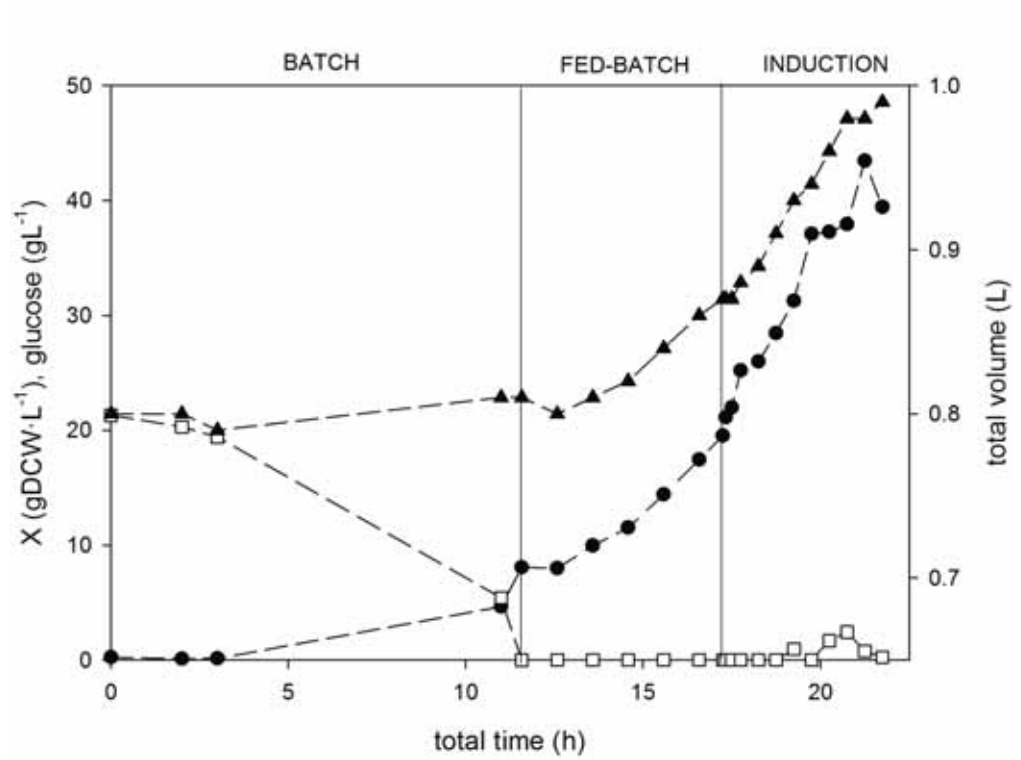


Figure 10.25. M15 Δ*gly*A[pQEαβfucA][pREP4] strain. $[IPTG]_{e,0}$ = 70 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}. (▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

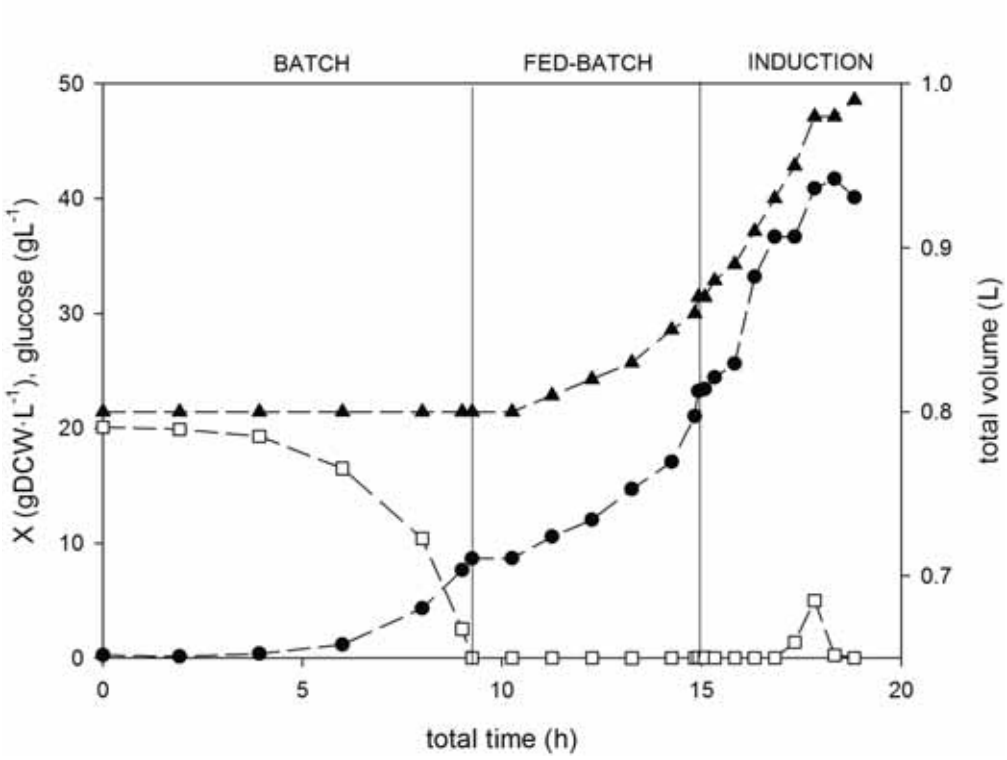


Figure 10.26. M15 [pQE-FucA][pREP4] strain. $[IPTG]_{e,0}$ = 70 µM; X_{ind} = 20 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

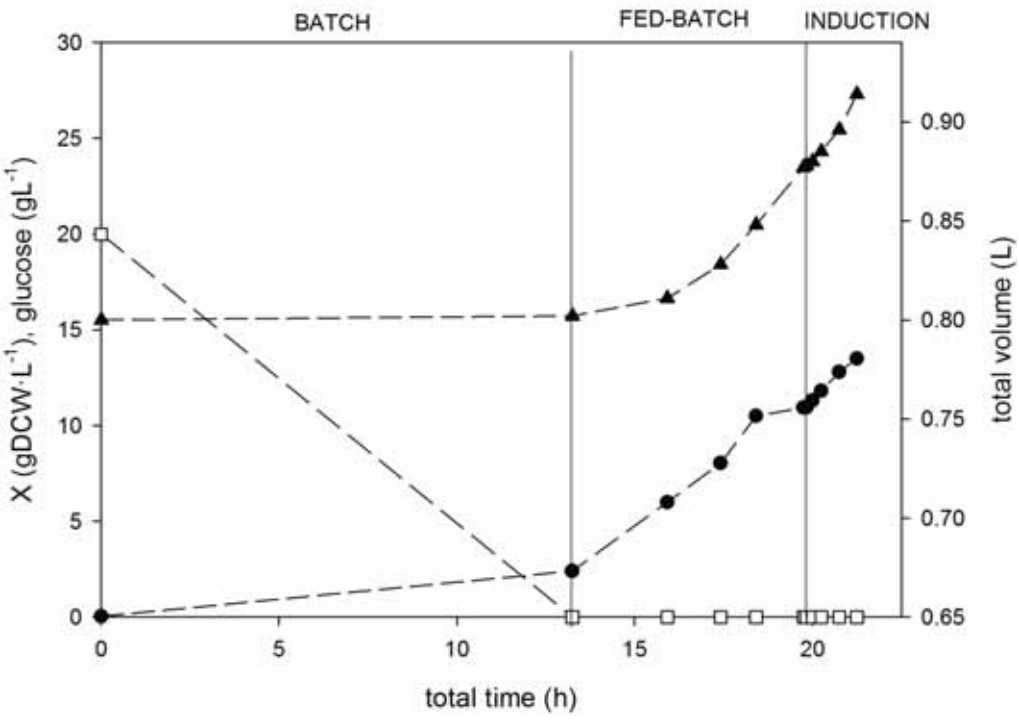


Figure 10.27. BL21 (DE3) FSA strain. $[IPTG]_{e,0}$ = 100µM; X_{ind} = 10.5 gDCW·L^{-1}; μ_{fix} = 0.15 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

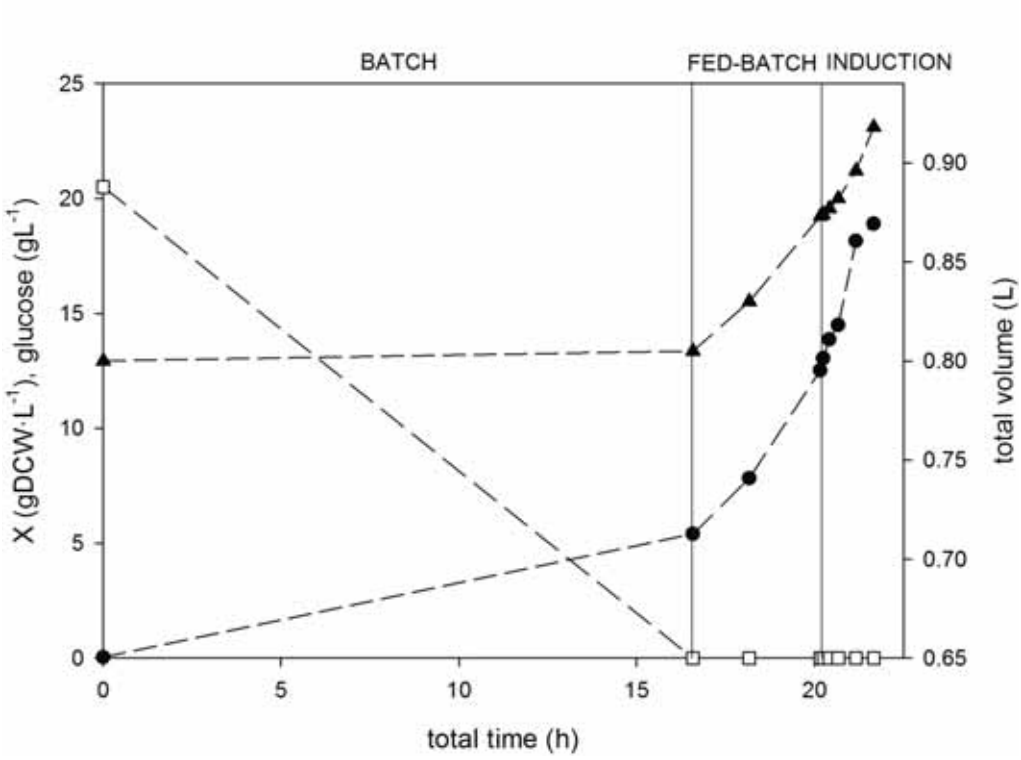


Figure 10.28. BL21 (DE3) FSA strain. $[IPTG]_{e,0}$ = 70µM; X_{ind} = 12.5 gDCW·L^{-1}; μ_{fix} = 0.2 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

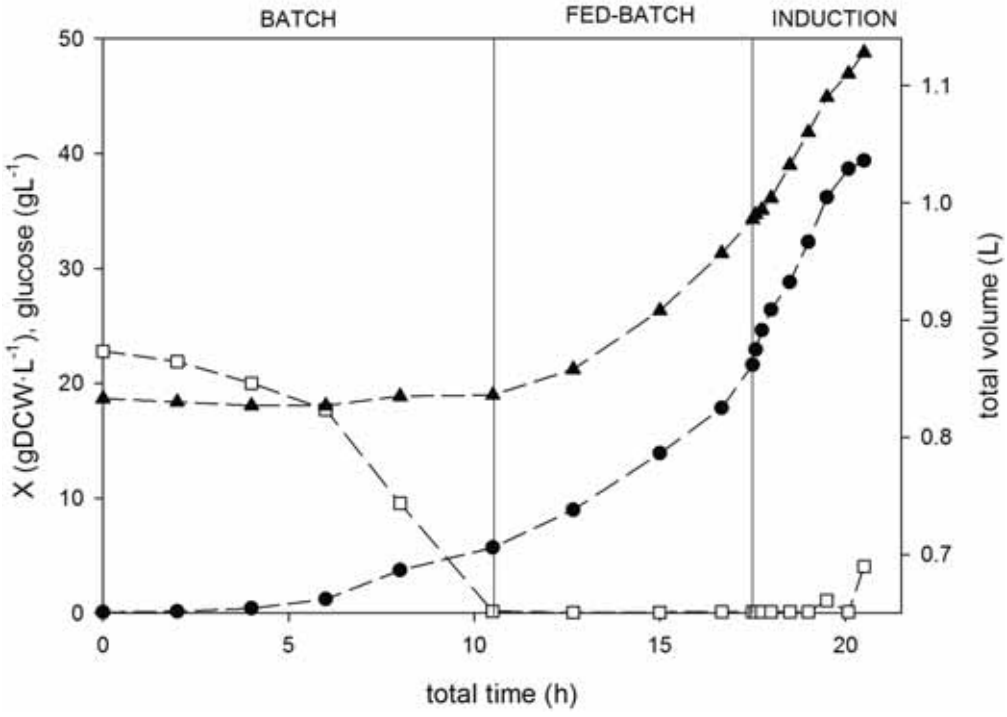


Figure 10.29. BL21 (DE3) FSA strain. $[IPTG]_{e,0}$ = 100µM; X_{ind} = 21.6 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

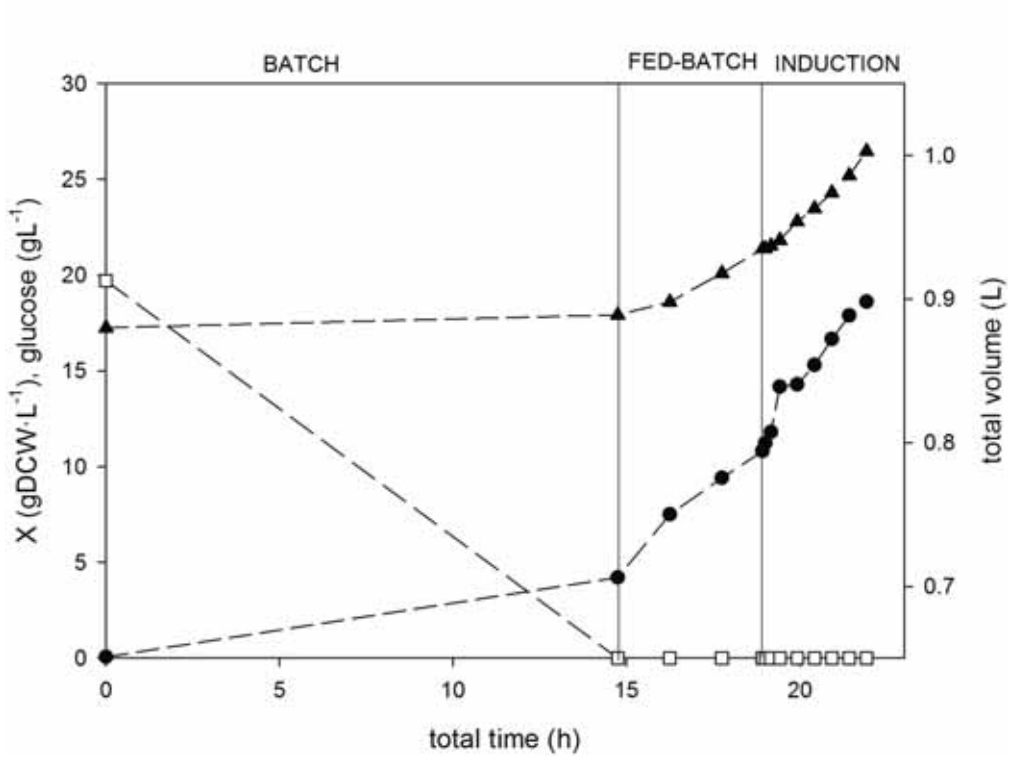


Figure 10.30. BL21 (DE3) ATA strain. $[IPTG]_{e,0}$ = 100µM; X_{ind} = 10.8 gDCW·L^{-1}; μ_{fix} = 0.22 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

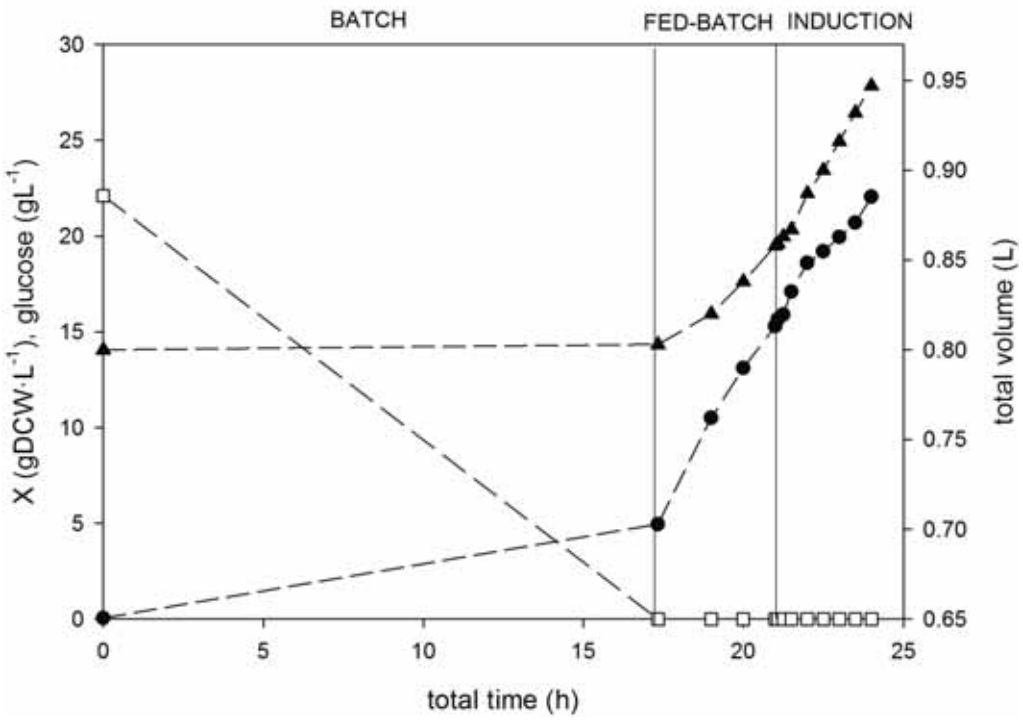


Figure 10.31. BL21 (DE3) ATA strain. $[IPTG]_{e,0}$ = 100 µM; X_{ind} = 15.3 gDCW·L^{-1}; μ_{fix} = 0.2 h^{-1}.

(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

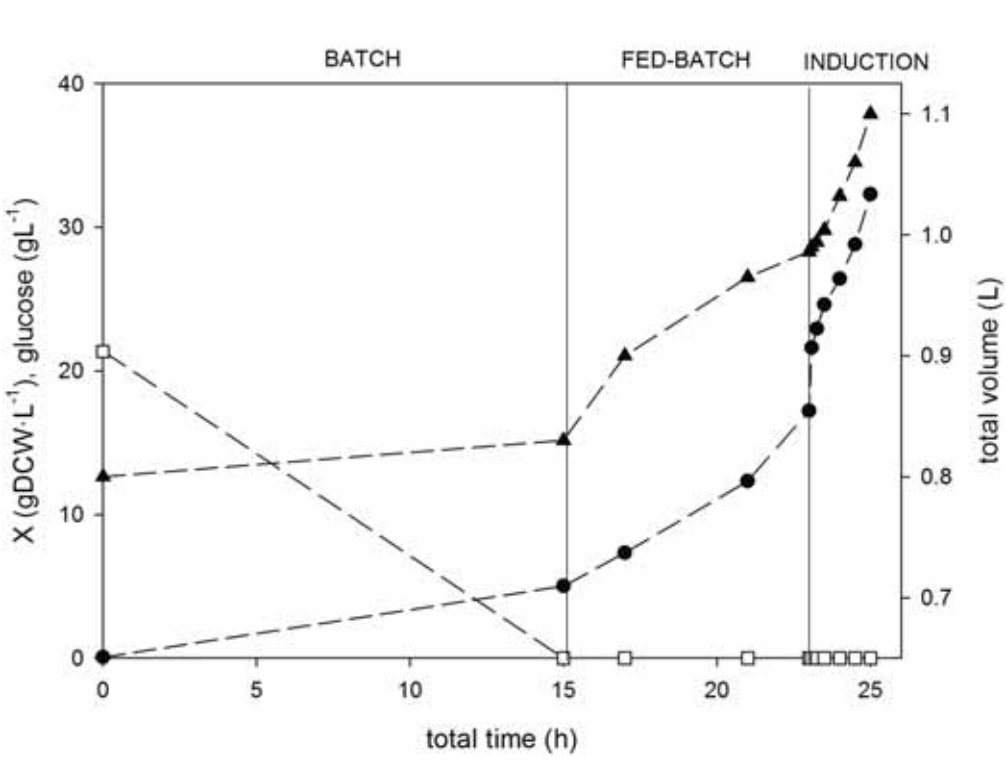


Figure 10.32. BL21 (DE3) ATA strain. $[IPTG]_{e,0}$ = 70μM; X_{ind} = 20.8 gDCW·L^{-1}; μ_{fix} = 0.15 h^{-1}.
(▲) experimental total volume; (•) experimental biomass concentration; (□) experimental glucose concentration.

10.2 List of publications

Calleja, D., Fernández-Castañé, A., Pasini, M., de Mas, C., López-Santín, J. (2014). Quantitative modeling if inducer transport in fed-batch cultures of *Escherichia coli. Biochemical Engineering Journal, 91,* 201-219.

Under preparation: (tentative titles)

Calleja, D., de Mas, C., López-Santín, J*.,* Overall modeling for recombinant protein production in high-cell density cultures o*f E.coli.*

Calleja, D., de Mas, C., López-Santín, J*.,* Protein production modeling using different *E.coli* expression systems.