

UAB

Universitat Autònoma de Barcelona



UNIVERSITAT AUTÒNOMA DE BARCELONA

ESCOLA D'ENGINYERIA

DEPARTAMENT D'ENGINYERIA QUÍMICA

Robust microbial construction and efficient processes for recombinant enzymes production in *Escherichia coli*

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

Martina Pasini

Setembre 2015

PAU FERRER ALEGRE, Professor agregat laboral del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona, CARLES DE MAS ROCABAYERA, Professor Titular del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona i GLORIA CAMINAL SAPERAS, Professora associada laboral del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona

CERTIFIQUEM:

Que la Biotecnòloga Industrial Martina Pasini ha dut a terme, sota la nostra direcció, el treball titulat "**Robust microbial construction and efficient processes for recombinant enzymes production in *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 2015

Dr. Pau Ferrer
Alegre

Dr. Carles de Mas
Rocabayera

Dra. Gloria Caminal
Saperas

“I think ninety-nine times and find nothing.

I stop thinking, swim in silence,

and the truth come to me”

-Albert Einstein

“Qualunque decisione tu abbia preso per il tuo futuro, sei autorizzato,

e direi incoraggiato a sottoporla ad un continuo esame,

pronto a cambiarla, se non risponde più ai tuoi desideri”

-Rita Levi Montalcini

Table of contents

Summary	I
Resum	III
Nomenclature	V
Abbreviations	VII
Strains abbreviations	IX
1 General introduction	1
1.1 Escherichia Coli	4
1.1.1 <i>E. coli</i> : platform for rpp	5
1.1.2 Bioprocess development for the production of recombinant proteins in <i>E.coli</i> : fed-batch operational mode.....	9
1.1.3 Stress factors in recombinant protein production	12
1.2 Methods of production	14
1.2.1 Sub-cellular localization.....	14
1.2.2 Conformational state.....	19
1.2.3 Expression system <i>E. coli</i> : <i>lac</i> operon	21
1.2.4 IPTG as inducer of recombinant protein	26
1.3 Gene expression and regulation	27
1.4 Synthetic biology	30
2 Context and objectives	33
2.1 Context of the thesis work	33
2.2 Objectives	34
3 Materials and Methods	37
3.1 Strains	37
3.1.1 <i>Escherichia coli</i> DH5 α	37
3.1.2 <i>Escherichia coli</i> DH5 α LacI ^{q1}	38
3.1.3 <i>Escherichia coli</i> M15	38
3.1.4 <i>Escherichia coli</i> M15 Δ glyA	38
3.1.5 <i>Escherichia coli</i> MG1655 LacI ^{q1}	39

3.1.6	<i>Escherichia coli</i> BL21 (DE3)*	39
3.2	Plasmids	40
3.2.1	pQE-40.....	40
3.2.2	pREP4.....	41
3.2.3	pUC-19.....	42
3.2.4	pSB1C3_GG plasmid	43
3.2.5	pASK-Pet.....	44
3.2.6	pET22b-PetAC.....	46
3.3	Molecular biology techniques	48
3.3.1	DNA isolation and manipulation.....	48
3.3.2	DNA transformation	51
3.3.3	DNA amplification by PCR.....	54
3.3.4	Golden Gate technique	61
3.4	Plasmid constructs	65
3.4.1	pQE-FucA	65
3.4.2	pQE $\alpha\beta$ FucA.....	66
3.4.3	pUC- <i>lacI</i>	67
3.4.4	pQE- <i>lacI</i> - $\alpha\beta$ FucA.....	67
3.4.5	BioBrick-based vectors	69
3.4.5.1	BioBricks constructs: pSB-J231XX-S_ <i>lacI</i>	69
3.4.5.2	BioBricks constructs: pSB-J231XX-S_ <i>glyA</i>	72
3.4.5.3	pQE-FucA-P3_ <i>lacI</i> -P4_ <i>glyA</i>	74
3.4.5.4	BioBricks constructs: pSB1C3-J231XX- <i>lacI-glyA</i>)	75
3.4.6	pQE-FucA_puzzle (J23110)	79
3.4.7	pQE-FucA_puzzle (J23110)_AmpR ⁻	81
3.4.8	pASK-FucA	83
3.4.9	pET22b-AC-FucA	84
3.5	Medium composition	85
3.5.1	Complex medium	85
3.5.2	Defined Medium.....	85
3.5.2.1	DM media for Shake flasks culture.....	87
3.5.2.2	MD media for fed-batch culture	87
3.5.3	Supplements.....	88
3.6	Cultivation conditions	89

3.6.1	Shake Flasks experiments setup.....	90
3.6.2	Bioreactors experiments setup	90
3.6.2.1	Biostat B bioreactor.....	91
3.6.2.2	Fermac 310/60 bioreactor	93
3.7	Analytical methods.....	94
3.7.1	Optical density.....	94
3.7.2	Colony Forming Unit (CFU), plasmid retention and plasmid curing	95
3.7.3	Substrate and by-products concentration.....	96
3.7.4	Flow cytometry.....	97
3.7.4.1	Component and function of a flow cytometer.....	97
3.7.4.2	Sample preparation and staining	99
3.7.4.3	Fluorescent Dyes	101
3.8	Protein Analysis.....	103
3.8.1	Enzyme activity for FucA	104
3.8.2	Total protein content	104
3.8.3	SDS-PAGE.....	105
4	Results I: Ab initio FucA overexpression in <i>E. coli</i>.....	109
4.1	Preliminary FucA expression studies with the M15[pREP4] and M15ΔglyA[pREP4] strains.....	112
4.1.1	FucA expression in the reference <i>E. Coli</i> M15[pREP4] strain	113
4.1.1.1	Flow cytometry analysis of the reference strain <i>E. coli</i> M15[pREP4].....	118
4.1.2	FucA expression in the <i>E. coli</i> M15ΔglyA[pREP4] strain	123
4.1.2.1	Flow cytometry analysis of M15ΔglyA[pREP4] strain.....	128
4.1.3	Comparison of the shake flasks cultures between M15[pREP4] and M15ΔglyA [pREP4] strains	132
4.2	Study of growth and FucA production in fed-batch cultures	134
4.2.1	M15[pREP4] strain.....	135
4.2.1.1	Off-gas analysis.....	138
4.2.1.2	Flow cytometry analysis of M15[pREP4] fed-batch cultures.....	141
4.2.2	M15ΔglyA[pREP4] strain	143
4.2.2.1	Off-gas analysis.....	146
4.2.2.2	Flow cytometry analysis of M15ΔglyA[pREP4] fed-batch cultures	148
4.2.3	Comparison of the fed-batch cultures between M15[pREP4] and M15ΔglyA [pREP4] strains	150

4.3 Conclusion.....	151
5 Results II: Design, construction and preliminary evaluation of an antibiotic-free single plasmid expression system	153
5.1 Curing the cells from the pREP4 plasmid.....	155
5.2 Cloning of <i>lacI</i> gene into the pQE α β FucA vector	157
5.3 Analysis of plasmid structural stability	158
5.3.1 Importance of the transcription levels of the <i>lacI</i> gene	159
5.4 Conclusion.....	163
6 Results III: Design and optimization of a single plasmid expression system	165
6.1 Tuning of <i>lacI</i> and <i>glyA</i> expression levels	166
6.1.1 BioBrick-based plasmid constructs for <i>lacI</i> expression level tuning	167
6.1.2 BioBrick-based constructs for <i>glyA</i> expression tuning	168
6.1.3 BioBricks-based construction of with the tuned <i>lacI</i> and <i>glyA</i> genes.....	170
6.2 Assembling the polycistronic <i>lacI-glyA</i> cassette	171
6.3 Engineered single-plasmid expression system, the Puzzle strain	176
6.3.1 Shake flasks cultures of the Puzzle strain	177
6.3.2 Flow cytometry analysis of the Puzzle strain.....	179
6.3.3 Fed-batch cultures of the Puzzle strain	184
6.3.3.1 Off-gas analysis.....	187
6.3.3.2 Flow cytometry analysis of the Puzzle strain	190
6.4 Conclusion.....	192
7 Results IV: Elimination of the antibiotic resistance encoding gene from the single-plasmid expression system	193
7.1 Deletion of the antibiotic resistance gene and Amp ^R strain characterization	194
7.1.1 Amp ^R - shake flask cultures.....	196
7.1.2 Flow cytometry analysis of Amp ^R strain.....	198
7.2 Fed-batch cultures of Amp ^R strain	203
7.2.1 Off-gas analysis.....	206
7.2.2 Flow cytometry analysis	208
7.3 Conclusion.....	210
8 Results V: Preliminary study of FucA secretion in the extracellular environment of <i>E. coli</i>.....	211

8.1	Preliminary FucA secretion studies	213
8.2	Conclusion.....	218
9	Comparison of <i>E. coli</i> M15 strains presented along this study	219
9.1	<i>E. coli</i> M15[pREP4] and M15 Δ glyA[pREP4] strains	219
9.2	Puzzle strain	223
9.3	AmpR ⁻ strain.....	226
10	General conclusions and future perspectives.....	229
11	References	233

Table of content

Summary

While antibiotic resistance marker genes are a powerful system for selection and maintenance of recombinant plasmids in hosts such as *E. coli*, its use has been considered unacceptable in many areas of biotechnology by regulatory authorities, particularly when the recombinant product will be used in the therapeutic field.

Previously, we have developed an expression system based on the pQE vector series (Qiagen) with an alternative plasmid selection marker based on glycine auxotrophy complementation (Vidal et al., 2008).

The pQE expression system is based on the IPTG-induced T5 promoter. Promoter leakiness in inducer absence may lead to structural instability of the expression vector, resulting in reduced expression levels, e.g. due to recombination events in the T5 promoter region. The lac repressor, encoded by the *lacI* gene, binds very tightly to the promoter, interfering with the transcription of the gene of interest. Therefore, *lacI* transcriptional level plays a key role in T5 promoter-based expression systems, influencing the basal transcriptional levels and the concentration of inducer.

Furthermore, although the overexpression of the plasmid-encoded protein of interest is a major factor in the metabolic burden, expression of other plasmid-genes may also contribute. Thus, in order to improve the system robustness, all the antibiotic resistance genes have been removed and the expression levels of the auxotrophic marker (*glyA*) and the lac repressor (*lacI*) genes have been fine-tuned, in order to minimize the metabolic burden related to plasmid-encoded genes.

In this study, an expression cassette has been constructed where the *lacI* and *glyA* genes have been placed under the control of selected synthetic constitutive promoters in order to obtain the sufficient *lacI* inhibitor ensuring lack of “promoter leakiness” and the minimal *glyA* transcriptional levels needed for plasmid maintenance and optimal cell growth in defined media. Moreover, in the expression vector the antibiotic resistance gene was replaced by the *lacI-glyA* cassette, thus yielding a completely antibiotic resistance gene free system.

Finally, in order to obtain high expression levels of the protein of interest, the capacity for recombinant *FucA* overexpression has been investigated at different scales (shake flasks and bioreactors) and operation modes, batch and fed-batch. Lastly, flow cytometry analyses were performed in order to analyze the bacterial populations at the single-cell level: changes in morphology (FSC and SSC) and physical and biochemical characteristics of individual cells within a bacterial population.

Resum

Encara que els gens de resistència a antibiòtics són una eina important per a la selecció i manteniment de plàsmids recombinants, les autoritats reguladores consideren el seu ús inacceptable en diverses àrees dins la indústria biotecnològica, sobretot quan el producte serà utilitzat amb finalitat terapèutica.

Un sistema d'expressió basat en el vector pQE (Quiagen) amb un marcador de selecció plasmídic alternatiu (basat en l'auxotròfia de glicina) va ser desenvolupat prèviament (Vidal et al., 2008).

El sistema d'expressió pQE es basa en el promotor T5, induïble per IPTG. La no-estabilitat del promotor en absència d'inductor pot provocar una inestabilitat estructural del vector, que portarà a nivells baixos d'expressió deguts, per exemple, a fenòmens de recombinació a la regió del promotor T5. El repressor *lac*, codificat pel gen *lacI*, s'uneix molt fortament al promotor, interferint en la transcripció del gen d'interès. Així doncs, els nivells transcripcionals del gen *lacI* tenen un paper crític en els sistemes d'expressió basats en el promotor T5, influenciant els nivells basals de transcripció i la concentració d'inductor.

Encara que la sobreexpressió de la proteïna d'interès és un factor important en quant a càrrega metabòlica, també cal tenir en compte la contribució deguda a l'expressió d'altres gens plasmídics. En aquest sentit, per tal de millorar la robustesa del sistema i minimitzar la càrrega metabòlica relacionada amb els gens codificats als plàsmids, s'han eliminat tots els gens de resistència a antibiòtics i s'han afinat els nivells

d'expressió del marcador auxotròfic (*glyA*) i del repressor *lac* (*lacI*).

En aquest treball s'han construït uns cassettes d'expressió on els gens *lacI* i *glyA* estan sota el control d'un seguit de promotors constituents sintètics, per tal d'obtenir suficient inhibidor *lacI* per evitar la "no-estabilitat del promotor" i assegurar el nivell mínim de transcripció de *glyA* necessaris tant per al manteniment del plàsmid com per al correcte creixement del cultiu en medi definit. A més, el gen de resistència a antibiòtic en el vector d'expressió va ser reemplaçat per la construcció *glyA-lacI*, obtenint un sistema totalment lliure de gens per a la resistència a antibiòtic.

Finalment, s'ha estudiat la capacitat per a la sobreexpressió de FucA recombinant a diferents escales (flascons agitats i bioreactor) i en diversos modes d'operació (discontínu i discontínu alimentat) per tal d'obtenir alts nivells d'expressió. Per últim, per tal d'analitzar la població bacteriana a nivell de cèl·lules individuals, els canvis morfològics (FSC i SSC) i les característiques físiques i bioquímiques de les diferents cèl·lules dins la població bacteriana, s'han realitzat anàlisis de citometria de flux.

Nomenclature

	Units	Description
F_s	$L \cdot h^{-1}$	feeding flow rate
m_{sx}	$g_{\text{glucose}} \cdot h^{-1} \cdot g^{-1} \text{DCW}$	maintenance coefficient
S_f	$g \cdot L^{-1}$	glucose concentration in the feeding solution
T	$^{\circ}C$	temperature
t	h	time
V	L	total volume
X	$g \text{DCW} \cdot L^{-1}$	biomass concentration
Y_{xs}	$g \text{DCW} \cdot g^{-1} \text{glucose}$	biomass-substrate yield
μ	h^{-1}	punctual specific growth rate
μ_{fix}	h^{-1}	fixed specific growth rate during the fed-batch
μ_{max}	h^{-1}	maximum specific growth rate

Abbreviations

AcCoA	Acetyl coenzyme A
AC	Autochaperon domain
AT	Autotransporter
ATP	Adenosine triphosphate
AU	Activity Unit
BioBricks	Standardized biological parts
BSA	Bovine serum albumin
cAMP	cyclic molecules of adenosine monophosphate
CAP	Catabolite Activator Protein
CDC	Proportion of carbon dioxide evolved
CFU	Colony forming unit
CR	Congo Red dye
DCW	Dry Cell Weight
ddH ₂ O	Distilled water
DHA	Dihydroxyacetone
DHAP	Dihydroxyacetone Phosphate
DM	Defined minimal Medium
DNA	Deoxyribonucleic acid
FCM	Flow Cytometry
FSC	Forward scatter
FucA	Fucose-1-Phosphate Aldolase
GTP	Guanosine triphosphate
HCDC	High cell-density cultures
I/X	inducer to biomass ratio
IBs	Inclusion bodies
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
Kb	Kilobase
kDa	Kilo Dalton

LB	Lysogeny Broth
MCS	Multiple cloning site
mRNA	Messenger RNA
NAD ⁺	oxidized form of Nicotinamide adenine dinucleotide
NADH	reduced form of Nicotinamide adenine dinucleotide
NRPs	Ribosomal peptides synthetases
nt	Nucleotide
OD ₆₀₀	Optical Density at $\lambda = 600\text{nm}$
OXC	Proportion of oxygen consumed
P.I.	Propidium iodide dye
PBS	Phosphate- buffered saline
PCR	Polymerase chain reaction
PI	Pre-induction
PKSs	Polyketide synthetases
pO ₂	Oxygen saturation
RBS	Ribosome binding site
RFP	Red fluorescence protein
RhuA	Rhamnulose-1-Phosphate Aldolase
RNA	Ribonucleic acid
RP	Recombinant protein
RPP	Recombinant protein production
RQ	Respiratory quotient
SD	Shine-Dalgarno sequence
SDS	Sodium dodecyl sulphate
SHMT	Serine hydroxymethyl transferase
SSC	Side scatter
TCA	Tricarboxylic acid cycle
TES	Trace elements saline
VBCN	Viable but not culturable phenotype
α -GDH	α -Glycerophosphate Dehydrogenase

Strains abbreviations

Abbreviation	Whole name
M15[pREP4]	<i>E. coli</i> M15[pREP4] pQE-FucA
M15ΔglyA[pREP4]	<i>E. coli</i> M15ΔglyA[pREP4] pQEαβFucA
M15ΔglyA[A]	<i>E. coli</i> M15ΔglyA pQEαβFucA
M15ΔglyA[B]	<i>E. coli</i> M15ΔglyA pQE-lacI-αβFucA
MG1655-FucA	<i>E. coli</i> MG1655 LacI ^{q1} pQEαβFucA
DH5α-FucA	<i>E. coli</i> DH5α LacI ^{q1} pQEαβFucA
J23100-S_lacI	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23100- S_lacI
J23111-S_lacI	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23111- S_lacI
J23110-S_lacI	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23110- S_lacI
J23117-S_lacI	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23117-S_lacI
J23100-S_glyA	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23100- S_glyA
J23111-S_glyA	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23111- S_glyA
J23110-S_glyA	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23110- S_glyA
J23117-S_glyA	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23117-S_glyA
M15ΔglyA[C00]	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23100-lacI-glyA
M15ΔglyA[C11]	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23111-lacI-glyA
M15ΔglyA[C10]	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23110-lacI-glyA
M15ΔglyA[C17]	<i>E. coli</i> M15ΔglyA pQE-FucA + pSB1C3-J23117-lacI-glyA
Puzzle	<i>E. coli</i> M15ΔglyA pQE-FucA_puzzle (J23110)
AmpR ⁻	<i>E. coli</i> M15ΔglyA pQE-FucA_puzzle (J23110)_AmpR ⁻

1 General introduction

Biotechnology is defined by the convention on Biological Diversity as “any technological application that uses biological system, living organisms or derivatives thereof, to make or modify products or processes for specific use” (United Nations 1992). Based on its application, different branches of biotechnology are distinguished: green (agricultural), red (medical), blue (marine) and white (industrial) biotechnology.

Biotechnology has been used for centuries by humankind to make beer, wine, bread, yoghurts and cheeses. However only with the beginning of the 20th first century, named the “Biotech century”, the nature of biotechnology has changed dramatically.

Since the 1970s, with the advent of recombinant DNA and the subsequent combination of traditional industrial microbiology and genetic engineering, a wide variety of expression systems for the production of protein or industrial enzymes of human purpose has become available (Schmidt 2004). For the first time, small medically relevant polypeptide hormones (somatostatin and insulin) were produced in *E. coli* as plasmid-encoded fusion to β -galactosidase. Since then the recombinant protein production (RPP) become a highly useful aspect of modern biotechnology.

Moreover, the initiation of the genomics era allowed for significant progress in heterologous (recombinant) protein production over the last two decades, finding applications in medicine, industry and research (Terpe 2006) (Li et al. 2014).

Various expression systems exist for RPP, from those based on bacteria, lower

eukaryotes (yeast and fungi), invertebrates (insect cells and larvae), vertebrates (cells and transgenic animals), plants (cells and transgenic plants) and even cell-free systems. The unique advantages and disadvantages of these expression organisms are briefly reviewed in Table 1.1 (Yin et al. 2007).

The wide range of expression systems available is necessitated by the huge amount of proteins targeted for production. Therefore, the selection of an appropriate expression system for expressing a recombinant protein is an important and probably the most challenging part of the process design. The host organism should ideally meet all the requirements of the destined application, which in turn strongly depends on the characteristics of the recombinant protein. Relevant criteria are (1) mass of the polypeptide and if it is a multi or single domain, (2) the number of disulfide bonds, (3) the desired post-translational modifications and (4) the final destination (intracellular expression, incorporation into the membrane or secretion into the extracellular broth). Proteins lacking proper post-translational modifications (e.g. a wrong glycosylation pattern), misfolded proteins or a low product yield typically indicate an expression system far away from optimal (Baneyx 1999a).

It is obvious that in an ideal situation, one universal system would enable of all possible recombinant genes in a fast, inexpensive, and proper manner with respect to yield, folding, and biological activity.

However, because of the limitations of existing systems, this situation is far from realistic (Sørensen 2010). It is necessary to carefully evaluate the properties of every new target and to experimentally screen the most promising expression hosts, for the

production of any kind of protein some day.

Table 1.1 Advantages and disadvantages of host cells for recombinant protein production

	Advantages	Disadvantages
Bacteria	<p>Large knowledge base and much experience available</p> <p>Gene expression rapidly and easily controlled</p> <p>Easy to grow with high yields (product can form up to 50% of total cell protein)</p>	<p>No post-translational modifications</p> <p>High endotoxin content in Gram negative bacteria</p> <p>Biological activity and immunogenicity may differ from natural protein</p>
Mammalian cells	<p>Same biological activity as native proteins</p> <p>Mammalian expression vectors available</p> <p>Can be grown in large scale cultures</p>	<p>Cells can be difficult and expensive to grow</p> <p>Manipulated cells can be genetically unstable</p> <p>Cells grow slowly, low productivity</p>
Yeast	<p>Lacks detectable endotoxins</p> <p>Facilitates glycosylation and formation of disulphide bonds</p> <p>Fermentation relatively inexpensive</p> <p>Well established large scale production and downstream processing</p> <p>Only 0.5 % native proteins are secreted. Isolation of secreted product is simplified</p>	<p>Gene expression less easily controlled</p> <p>Glycosylation not identical to mammalian systems</p> <p>High yield and stable production strains</p>
Fungi	<p>Well established systems for fermentation of filamentous fungi</p> <p>Growth inexpensive</p> <p>Can secrete large quantities of product into growth media, source of many industrial enzyme</p>	<p>High level of expression not yet achieved</p> <p>Generics not well characterized</p>

1.1 *Escherichia Coli*

Escherichia coli is a Gram-negative, facultative anaerobe and non-sporulating bacterium belonging to the Eubacteria kingdom. First isolated by Theodor Escherich, a German pediatrician, in 1885 and later named after him. It is commonly found in the lower intestine of warm-blooded organism and plays an important role in vertebrate anabolism and catabolism.

The *E. coli* cells are in general rod-shaped, measuring approximately 2 micrometers long and 0.5 micrometers in diameter, (Kubitshek 1990). This γ -proteobacterium divides through binary fission (about 1 division every 30 min) in optimal conditions and can be easily and inexpensively grown and cultured outside the intestine in a laboratory setting where it can live on a wide variety of substrates. It uses mixed-acid fermentation in aerobic conditions, producing lactate, succinate, ethanol, acetate and carbon dioxide.



Figure 1.1 *E. coli* cells electron micrograph (10000 X zoom). Image takes from Wikipedia.

E. coli is the most widely studied model organism, and an important species in the fields of biotechnology and microbiology, where it has served as the host organism for the majority of work with recombinant DNA.

The first complete DNA sequence of an *E. coli* genome (from the K-12 strain derivative MG1655) was published in 1997 (Blattner et al. 1997). It was found to be a circular DNA molecule of 4.6 million base pairs, containing 4288 annotated protein-coding genes. The coding density was found to be very high, with a mean distance between genes of only 118 base pairs.

Like all lifeforms, new strains of *E. coli* evolve through the natural biological processes of mutation, gene duplication and horizontal gene transfer (Lawrence and Ochman 1998); some strains can even develop traits that can be harmful to a host animal. *E. coli* K-12 and *E. coli* B strains are the most frequently used varieties for laboratory purposes.

1.1.1 *E. coli*: platform for rpp

The bacteria *E. coli* was the first host used to express recombinant proteins and it continues to dominate the bacterial expression systems; i.e. the first FDA-approved RP pharmaceutical (Humulin, Eli Lilly), recombinant human insulin was produced in *E. coli*, which even nowadays is still the principal producer (Johnson 1983).

One of the main reasons seems to be that this organism is very well known and established in each laboratory. So it is no surprise that *E. coli* systems are also most commonly used for industrial and pharmaceutical protein production.

Besides, the growth media for *E. coli* are quite inexpensive and there are relatively straightforward methods to scale-up bioproduction to extremely high cell densities. Moreover, in *E. coli* recombinant protein can be accumulated in very high percentage, up to 50 % of soluble protein (Sevastyanovich et al. 2009).

Furthermore, in contrast to eukaryotic cells, the bacterial chromosome is not enclosed inside of a membrane-bound nucleus but instead resides inside the bacterial cytoplasm. In there, the genome exists as a compact, usually circular, double-stranded piece of DNA, of which multiple copies may exist at any time. Thanks to its structure and its localization the bacterial genome can be easily modify. Main modifications have as purpose the optimization of the global production yields, focusing on the minimization of site product formation and/or on the improve of protein folding.

Additionally, along with chromosomal DNA, bacteria may also contain plasmids, which are small extra-chromosomal DNAs that can replicate independently and that can be easily gained, lost or transferred by cells. Plasmids range in size from a few thousand base pairs to more than 100 kilobases (kb). Besides, while the chromosomes contain all the essential information for living, plasmids, or expression vectors, usually contain genes that are advantageous but not essential to their host. For example, some bacterial plasmids encode enzymes that inactivate antibiotics.

These properties make the plasmids important tools for genetics and biotechnology purposes, where they are commonly used to clone and amplify or express particular genes. A wide variety of plasmids are commercially available for such uses; one possible way to classify them is according to the promoter they own (Table 1.2).

Generally a plasmid contains a number of features: i) replicon, origin of replication to allow the bacterial cells to replicate the plasmid DNA; ii) resistance marker, gene that confer resistance to particular antibiotics; iii) controlled promoter, that control the expression of the gene of interest; iv) multiple cloning site (MCS), suitable site for cloning the gene of interest. Mostly these plasmids have been engineered to optimize their use as vectors in DNA cloning. For instance, their length is reduced; many plasmid vectors are only 2 Kb in length, which is much shorter than in naturally occurring *E. coli* plasmids. Researchers can use the plasmids in order to introduce target genes into the microbes. Genetic systems have also been developed which allow the production of recombinant protein. One of the first useful applications of recombinant DNA technology, in 1978, was the manipulation of *E. coli* to produce human insulin (Johnson 1983).

Table 1.2 Some *E. coli* promoter systems that are in use for heterologous protein production and their characteristics

Expression system	Source	Bases of regulation	Reference
<i>lac</i> promoter (T5)	Natural or modified version of <i>E. coli lac</i> promoter	Expression is repressed by the <i>lac</i> repressor, LacI. This repression is lifted by addition of a ligand of LacI, which is naturally lactose or the non-metabolizable analogue, IPTG	(Gronenborn 1976)
T7 RNA Polymerase	Engineered <i>E. coli lac</i> promoter, T7 RNA polymerase gene	T7 promoter cannot be activated by the native <i>E. coli</i> RNA polymerase enzyme but with the T7 RNA polymerase (RNAP). Expression of RNAP is regulated by the <i>lacUV5</i> promoter, which is repressed as the <i>lac</i> system	(Studier 1991)
<i>tac/trc</i>	Hybrid of <i>lacUV5</i> and <i>trp</i> promoter	As <i>lac</i> system	(Brosius, Erfle, and Storella 1985)
<i>tetA</i> promoter/operator	Tetracycline operon of <i>E. coli</i>	Anhydrotetracycline	(Skerra 1994)
phage promoter p_L	Promoter and repressor cl from λ phage	p_L promoter is repressed by cl repressor protein. A temperature-sensitive version of cl (cl_{I857}) is stable at 30°C but unstable and, thus, lifts repression at 42°C. Induction requires growth at 42°C, which might not be beneficial for correct protein folding	(Elvin et al. 1990)
<i>araBAD</i> promoter (pBAD)	Arabinose operon of <i>E. coli</i>	<i>araBAD</i> promoter is repressed by the arabinose repressor AraC. Repression is lifted upon addition of arabinose	(Guzman et al. 1995)
<i>rhaBAD</i> promoter	Rhamnose operon of <i>E. coli</i>	<i>rhaBAD</i> is regulated by two activators, RhaS and RhaR, and the corresponding genes belong to one transcription unit which is located in opposite direction of <i>rhaBAD</i> . The induction requires L-rhamnose	(Haldimann, Daniels, and Wanner 1998)

1.1.2 Bioprocess development for the production of recombinant proteins in *E.coli*: fed-batch operational mode

When using *E. coli* as expression system for the production of proteins, high cell-density cultures (HCDC) are usually performed in order to increase volumetric protein yields. HCDC for culturing *E. coli* have been developed to improve productivity, and also to provide advantages such as reduced culture volume, enhanced downstream processing, reduce wastewater, lower production costs and reduced investment in equipment (Lee 1996) (Overton 2014).

HCDC are carried out in bioreactors, where several parameters influencing bacterial growth can be controlled, such as pH; concentration of dissolved oxygen (pO_2), temperature (T), foaming, etc. Three classical operational modes can be considered for the production of recombinant proteins: i) continuous cultivation, ii) discontinuous cultivation (batch) and iii) semi-continuous cultivation (fed-batch). The latter process has most often been used with *E. coli*, because it allows obtaining HCDC and because the growth rate can be modulated by controlled addition of substrate (Riesenber 1991).

Nevertheless, HCDC of this microbe has been difficult due to inhibitory by-product formation. In fact, under aerobic conditions when the glucose concentration in the medium is non-limiting, *E. coli* displays a high rate of substrate consumption and a corresponding high specific growth rate, leading to the secretion of the organic acid acetate (Fuentes et al. 2013).

The physiological background for this acid production under aerobic conditions is the

result of a metabolic imbalance, also known as overflow metabolism, where the rate of acetyl-coenzyme A (AcCoA) synthesis surpasses the capacity of the tricarboxylic acid (TCA) cycle to completely consume this metabolite. Therefore, the amount of AcCoA not consumed by the TCA cycle is diverted into the acetogenesis pathway where acetate is synthesized (Figure 1.2) (Eiteman and Altman 2006).

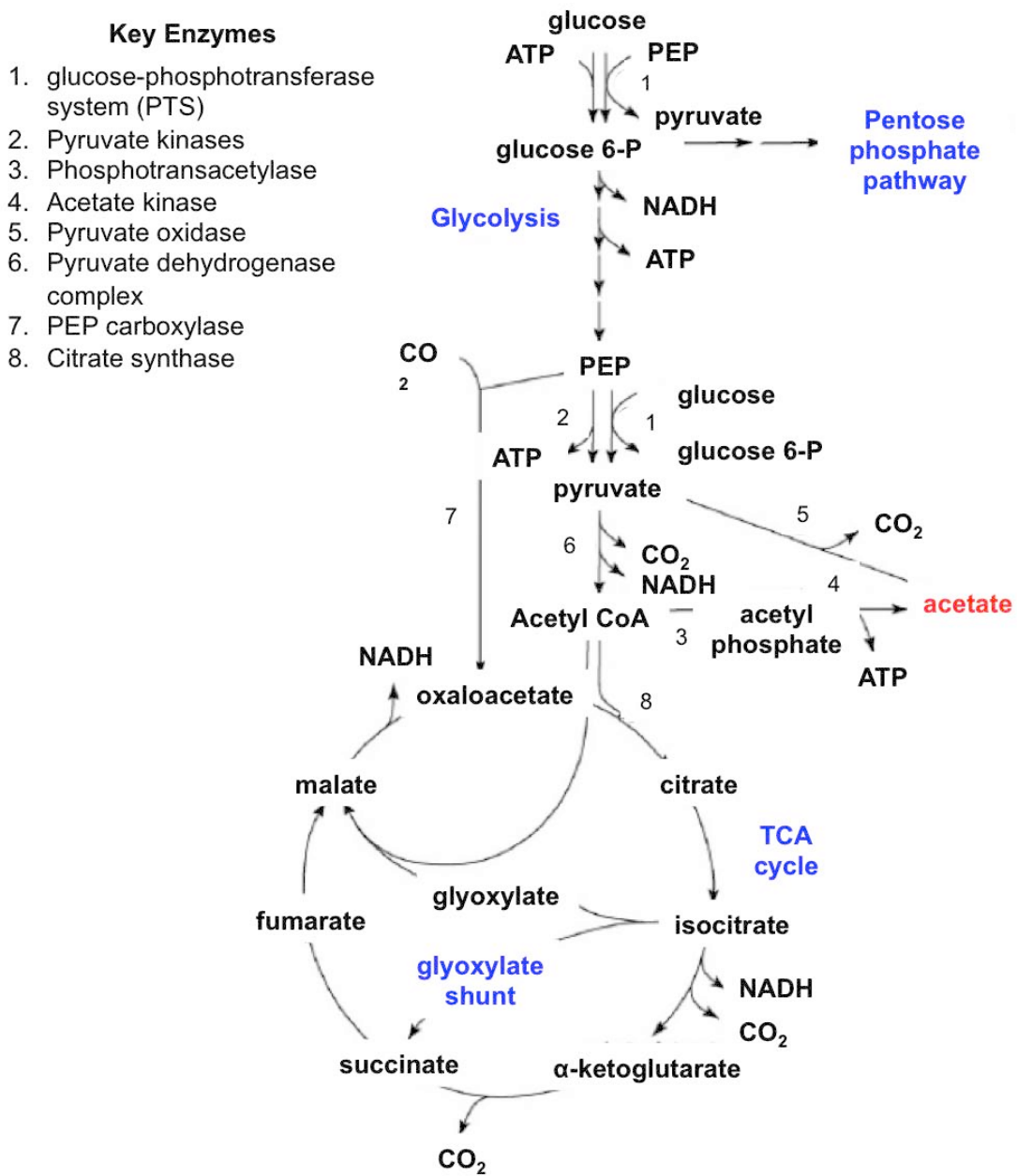


Figure 1.2 Acetate metabolism in *E. coli* (Eiteman and Altman 2006)

Thus, a culture of *E. coli* will generate acetate depending on the rate at which cells grow or when the cells surpass a threshold specific rate of glucose consumption (q_s) (Jensen and Carlsen 1990) (Xu, Jahic, and Enfors 1999). It also occurs aerobically when unlimited growth on excess of glucose inhibits respiration, a phenomenon called the Crabtree effect (Crabtree 1928). As a consequence, *E. coli* excretes 10 to 30 % of carbon flux from glucose to acetate in glucose-containing media (Holms 1996).

Acetate accumulation is a problem since this represents the loss of precursor molecule AcCoA, meaning a diversion of carbon that might otherwise have generated biomass. Besides, acetate and other short-chain fatty acids are toxic, whose accumulation at concentrations above 40 mM ($2.4 \text{ g}\cdot\text{L}^{-1}$) affects, not only the growth, but also the recombinant protein production (Phue et al. 2005). To solve this problem, many methods have been developed from genetic manipulation to the control of the glucose concentration in the fed-batch fermentation process (Kang et al. 2009).

Nevertheless, acetate is only partially oxidized, and so it is a potential source of both carbon and energy which can be assimilated through the glyoxylate shunt (Kleman and Strohl 1994). This "acetate switch" occurs when bacterial cells deplete their environment of acetate-producing carbon source, like glucose, and begin to rely on their ability to scavenge for environmental acetate. Therefore, the ability of *E. coli* to remove this potential toxin from its environment by consuming it permits to solve this problem under specific conditions. As explained before, genetic manipulation is a powerful technique to overcome acetate accumulation, however this can also be avoided in bioreactors by means of culture medium design and methods for carbon source feeding (Eiteman and Altman 2006).

1.1.3 Stress factors in recombinant protein production

The production of recombinant protein is one of the most energy and raw material consuming process (Rahmen et al. 2015). It is widely recognized that high level expression of heterologous proteins has a direct impact on cell's core metabolism and activates stress responses. Metabolic load is defined as the draining of host cell's resources, either in the form of energy such as ATP or GTP, or raw materials such as amino acids and nucleotides, that are required to express and synthesize the protein of interest (Glick 1995). Typically, cells overcome these effects by triggering stress-response mechanism that adapt and readjust the metabolism in order to restore functionality and viability. However, these cellular responses often affect negatively growth parameters such as growth rate, biomass yield, productivity and cellular viability (Carneiro, Ferreira, and Rocha 2012).

Certain amount of cellular energy is required to maintain the presence of the newly introduced plasmid DNA in the host cell. Besides, since the protein synthesis is the most energy consuming process in the cell, the extent of the stress responses is determined by the rates of transcription and translation, which may become limiting at in recombinant protein overproducing cells (Hoffman and Rinas 2004). Moreover, high synthesis rates can bring to misfolded products and to inclusion body formation. Thus, the higher the plasmid or foreign gene copy number, the greater the amount of energy required to maintain that DNA within the host cell (Glick 1995).

Additionally, one of the most prevalently observed changes in host-cell physiology after recombinant protein induction, is a decrease in the growth rate which can be

observed through lowered biomass concentration and increase oxygen consumption and carbon dioxide production after induction (Weber, Hoffmann, and Rinas 2002) (Neubauer, Lin, and Mathiszik 2003).

Furthermore, during growth of HCDC, overproduction of recombinant proteins often results in cell filamentation and/or encounter of bacteria with a viable but non-culturable phenotype, which means the lost of the ability to form colonies on agar plates, and/or cell death and lysis (Wyre and Overton 2014a) (Jeong and Lee 2003). In these cases, flow cytometry has been extensively used as a rapid process analytical technique to determine not only the fraction of dead cells but also the percentage of productive bacteria in a culture (Hewitt et al. 1999) (Wyre and Overton 2014b).

Several studies have reported the main challenges in recombinant protein production and potential factors influencing the metabolic burden, focusing on the different possible strategies to overcome these limitations and optimize the recombinant protein production (Makrides 1996) (Jeong and Lee 2003) (Neubauer et al. 2003).

The metabolic load imposed by the recombinant protein production and its main consequences in the whole metabolism of *E. coli* are quite relevant to consider when optimizing recombinant bioprocesses. Efforts in developing strategies to maximize the productivity of recombinant proteins in *E. coli* are well documented in the literature (Correa and Oppezzo 2011) (Li et al. 2012). Extensive research has been performed over the past years in order to improve recombinant protein production in this cell factory, including the optimization of process parameters such as growth temperature, media composition, induction conditions, as well as engineering novel expression

systems (Jana and Deb 2005) (Tomohiro, Skretas, and Georgiou 2011). However, further studies are needed. For example, the elimination of the antibiotic resistance gene which synthesis represents one of the major causes of a metabolic load (Vandermeulen et al. 2011).

1.2 Methods of production

Relating to the subcellular localization of the protein and the state in which it is produced, different strategies for RPP can be used in *E. coli*.

1.2.1 Sub-cellular localization

In *E. coli*, recombinant proteins are normally either directed to the cytoplasm or to the periplasm and, to a lesser extent, secreted into the extracellular medium (Figure 1.3).

Proteins directed to the cytoplasm are the most common and efficiently expressed. Cytoplasmatic production and storage requires the simplest expression systems and generally allows higher yields of protein (Jana and Deb 2005). However, the high expression of recombinant protein can often lead to the accumulation of aggregated, insoluble protein that forms inclusions bodies. Inclusion bodies can be a significant hindrance in obtaining soluble, active protein in some situations. Though, in some cases, inclusion bodies are advantageous because they are resistant to proteolysis, easy to concentrate by centrifugation, minimally contaminated with other proteins,

and, with some effort, able to be refolded in active and soluble proteins (Terpe 2006). Furthermore, formation of inclusion bodies can be prevented by: i) the co-overexpression of certain chaperones that aid in protein folding, such as DnaK-DnaJ and GroEL-GroES; ii) process temperature control, since the aggregation seems to be less likely to happen at lower temperature; iii) cell and process design (Ari et al. 2007) (Natl et al. 1998) (Nishihara et al. 1998).

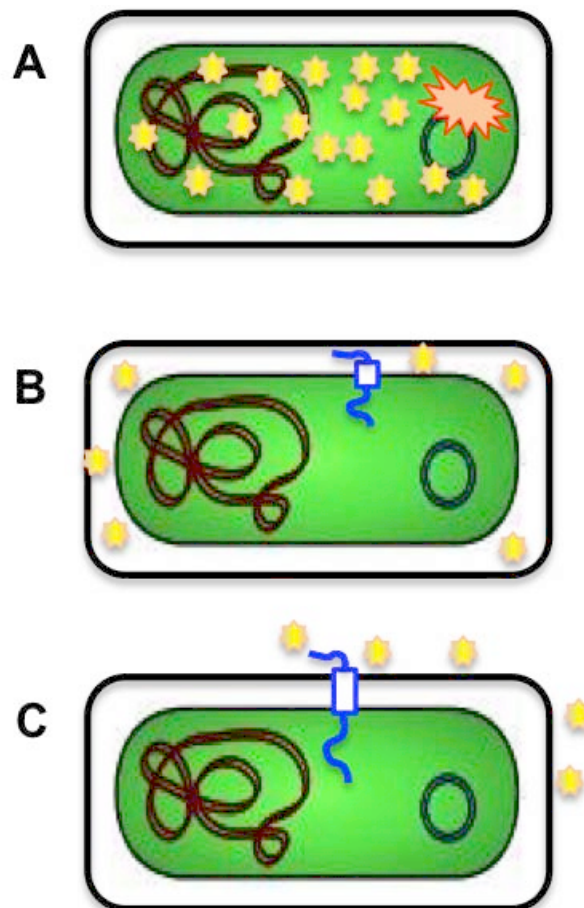


Figure 1.3 Production methods for RPP in *E. Coli*. A) cytoplasmic; B) periplasmic; C) secreted. The yellow stars represent the soluble RP, the orange cloud represents the inclusion bodies and the blue pipe represents the protein export machinery

The main advantage of periplasmic localization is that the periplasmic environment is less reducing than the cytoplasm and therefore favors disulphide bond formation.

Moreover, protease activity is lower in the periplasm allowing higher protein stability. Accumulation of RP in the periplasm requires the nascent polypeptide to be translocated across the cytoplasmic membrane. The principal mechanisms of translocation for RPP is through the SecYEG complex that exports unfolded proteins; and through the Tat systems, that thanks to the twin-arginine translocation can export folded proteins (Mergulhao, Summers, and Monteiro 2005). Most common problems for heterologous protein production in the periplasmic space are incomplete translocation across the inner membrane (Baneyx 1999a), proteolytic degradation, and insufficient capacity to be exported (Mergulhao et al. 2005).

Finally, secretion of RP into the extracellular environment is an attractive option, which, together with the positive effects of the periplasmic localization, is characterized by the simplicity of the downstream process. In fact extracellular production does not require membrane disruption and therefore it avoids intracellular proteolysis by periplasmic proteases, reduces the contaminants in the final product and allows continuous production of recombinant proteins (Lee 2004); besides, offers a better chance of correct protein folding (Cornelis 2000).

Due to these advantages secretion is a common strategy in host such as CHO cells, but it has been less effective when applied to *E. coli*, because wild type *E. coli* normally does not secrete proteins extracellularly except for a few classes of proteins such as toxins.

Several non-specific strategies for extracellular accumulation of recombinant proteins have been developed for *E. coli* including genetically or chemically altering strains to promote protein leakage from the periplasmic space to the culture medium (Lee 2004) (Jose and Meyer 2007). Besides, numerous highly specialized protein secretion systems exist, most being large protein complexes (Leo, Grin, and Linke 2012). There are at least six specialized secretion systems in Gram-negative bacteria: Type I, II, III, IV, V and VI (Lee 2004). The bacterial type I, II, III and chaperone-usher systems have been manipulated to secrete foreign proteins from *E. coli* and other Gram⁻ negative bacteria (Henderson, Navarro-garcia, and Nataro 1998)(Puertas et al. 2010). However, their use for RPP is hampered by the debatable nature of the secretion signals, their molecular complexity (which results in species and/or substrate specificity) and the limited accumulation of the target protein. In contrast, the Type V, or Autotransporter (AT), system has been utilized widely to successfully secrete a variety of heterologous target molecules to the bacteria (Leyton et al. 2010). Here only the type V mechanism will be discussed since it is the one used in this study.

Superficially the AT exhibits a tripartite domain organization: the N-terminal signal peptide mediates inner membrane translocation, the central passenger domain is the secreted functional moiety and the C-terminus forms a porin-like structure, termed β -barrel, within the outer membrane, which is essential for the translocation of the passenger domain to the bacterial cell surface (Leo et al. 2012) (Henderson et al. 2004). A hypothetical model is presented in Figure 1.4 (Jose and Meyer 2007).

The precursor protein contains a cleavable N-terminal signal sequence, which mediates Sec-dependent protein export into the periplasm, a passenger domain which

is required to achieve full surface exposure and functionality of the passenger protein, and a C-terminal domain which allows the rest of the peptide to reach the outside of the cell (Henderson et al. 2004) (Tajima et al. 2010). The effector portion of the molecule displays functional and structural heterogeneity and can be substituted with heterologous proteins (Jose and Meyer 2007) (Henderson et al. 2004).

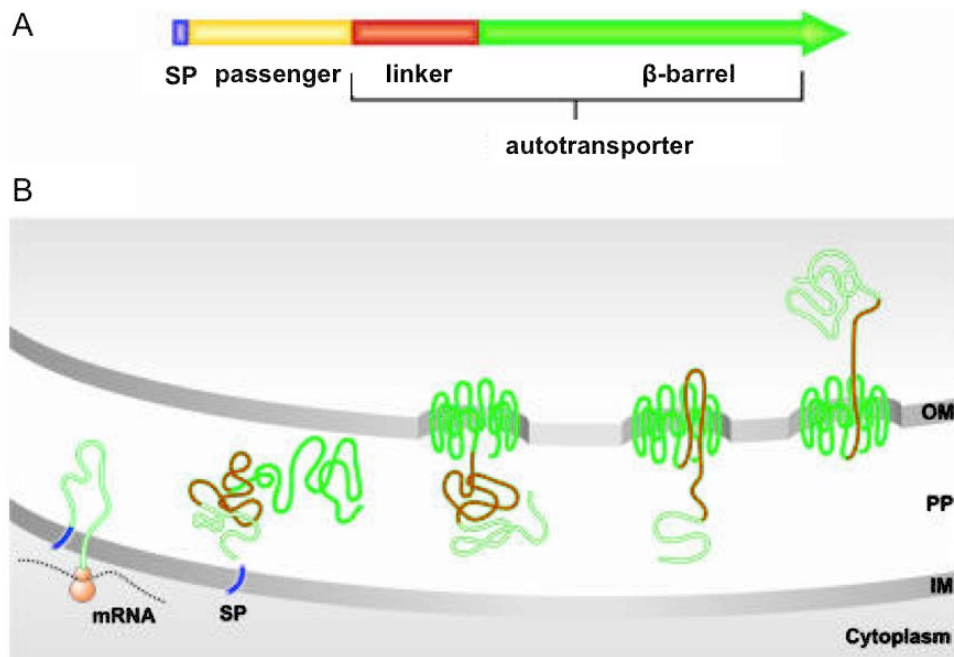


Figure 1.4 Secretion mechanism of the autotransporter proteins. A) Structure of the polyprotein precursor. B) Transport of the recombinant passenger. By the use of a typical signal peptide, a precursor protein is transported across the inner membrane. After arrival at the periplasm, the C-terminal part of the precursor folds a β -barrel structure within the outer membrane, and the passenger is transmitted to the cell surface. SP, signal peptide; IM, inner membrane; PP, periplasm; OM, outer membrane.

Previous studies have demonstrated the utility of ATs for Autodisplay of heterologous proteins on the bacterial cell surface (Jose and Meyer 2007). Furthermore, Sevastyanovich et al., have demonstrate that the AT module can be utilized not only

for cell surface display but also for the accumulation of heterologous proteins in the culture medium without the addition of exogenous protease (Sevastyanovich et al. 2012). In the expression vector design by Sevastyanovich, cleavage of the passenger domain of the serine protease ATs from their cognate β -barrel is affected by nucleophilic attack of β -barrel residues on a single residue in the α -helix, and the passenger domain can be completely replaced with a target protein (Sevastyanovich et al. 2012). Thus, the passenger domain can be replaced with a target protein of interest, generating a platform for secretion of heterologous protein in the extracellular environment in a soluble folded form.

1.2.2 Conformational state

Proteins can be made in two conformational states, either in their soluble folded form or as insoluble aggregates named inclusion bodies (IBs) (Terpe 2006). Protein solubility and propensity to aggregate has been central to biotechnology and biosciences through the era of recombinant protein expression. It is also becoming increasingly important in the area of formulation and preparation of biologics (therapeutic proteins), and in consideration of disorders arising from misfolding (Chan, Curtis, and Warwicker 2013).

Under normal growth condition *E. coli* produces native globular proteins that can fold in their stable and soluble form. Expression in soluble fraction is paramount for the expressed protein to be biologically active (Berrow et al. 2006). Besides, solubility is essential for efficient, large-scale purification. Despite its importance and utility RP not

always are produced in a soluble and homogeneous state. Though, high-level expression of heterologous protein is prone to cause soluble protein to accumulate as IBs. IBs are dense electron-refractile particles of aggregated protein, found in both cytoplasmic and periplasmic spaces. In the case of proteins having disulfide bonds, formation of protein aggregates is more probable, since the reducing environment of bacterial cytosol inhibits the formation of disulfide bonds. For many years IBs have been considered a waste product, essentially formed by misfolded and inactive proteins, and several approaches have been developed in order to overcome the problem associated with insolubility. However, in the last decade, with the advent of high-throughput screening methods, the scene has completely changed and IBs have gained considerable interest. Recent reviews, in fact, have presented properly folded proteins as natural components of inclusion bodies (Martínez-Alonso et al. 2009) (Sans et al. 2012). As IBs are relatively pure and proteolysis-resistant and contain intact recombinant proteins, several approaches have been reported to recover these aggregated forms as biologically active proteins (Lilie, Schwarz, and Rudolph 1998). In a typical procedure IBs are first denatured with a high concentration of denaturant such as urea or guanidinium chloride, and reduced with a reducing agent such as β -mercaptoethanol. Then, the refolding from denatured proteins to active proteins occurs by the removal of denaturant. Although these methods work well for many IBs, in many cases there is a significant amount of protein precipitation, resulting in a low recovery yield.

Therefore, it can be argued that production of correctly folded, soluble protein would be the preferable strategy.

1.2.3 Expression system *E. coli*: *lac* operon

Many expression systems have been developed for RPP in *E. coli*, comprising of different mechanism for induction, strains and additional features to aid production.

In particular, many promoter systems of *E. coli* are described as tools for protein expression, but only a few of them are commonly used (Table 1.2). A useful promoter must be strong enough, has a low basal expression level (i.e., it is tightly regulated), must be easily transferable to other *E. coli* strains, and the induction must be simple and cost-effective, and should be independent on the commonly used ingredients of culturing media. Therefore, to produce high levels of protein, it is often useful to clone the gene of interest downstream a well-characterized, regulated promoter.

The *E. coli lac* promoter, without any doubt, is the most extensively studied. This sequence is naturally present in the *lac* operon, which was the first genetic regulatory mechanism to be understood clearly and to become a foremost example of prokaryotic gene regulation. The *lac* operon in wild type *E. coli* comprises three structural genes: i) *lacZ*, encodes β -galactosidase (LacZ), an intracellular enzyme that cleaves the disaccharide lactose into glucose and galactose and is responsible for the production of allolactose, the natural inducer of *lac* operon; ii) *lacY*, encodes lactose permease (LacY), a transmembrane symporter that pumps β -galactosides into the cell using a proton gradient in the same direction; iii) *lacA*, encodes galactoside O-acetyltransferase (LacA), an enzyme that transfers an acetyl group from acetyl-CoA to β -galactosides (Figure 1.5). The *lac* genes are organized into an operon and oriented in the same direction.

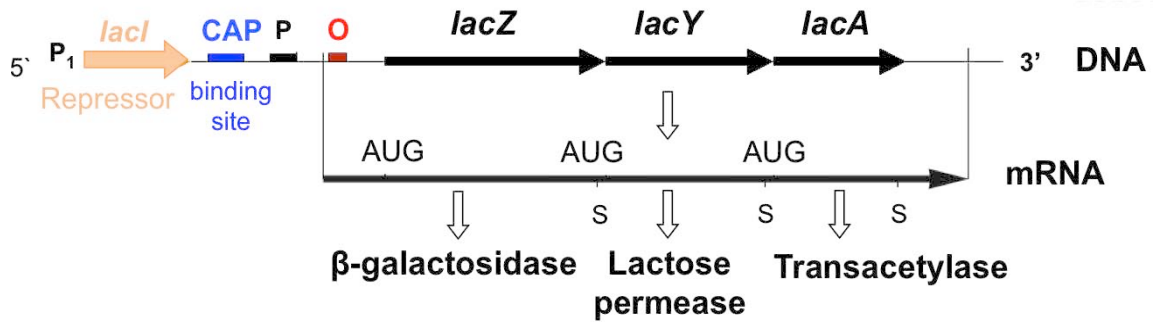


Figure 1.5 Schematic representation of the *lac* operon. The genes *lacZ*, *lacY* and *lacA* are transcribed from a single promoter (P) in a single mRNA from which three proteins are translated, LacZ, LacY and LacA respectively. The *lac* operator (O) is regulated by the repressor protein codify by the *lacI* gene, regulated by its own promoter (P_1). CAP is the catabolite activator Adapted from Wikipedia.

The first mechanism of control is regulated by the lac repressor (LacI), which is the product of the *lacI* gene (Guzman et al. 1995) (Dunaways et al. 1980).

In absence of lactose the LacI binds very tightly the operator region, downstream the promoter, impeding the RNA polymerase access to the promoter and starting the transcription (Figure 1.6). However, in presence of lactose, on the lactose isomer allolactose, LacI undergoes a conformational change that prevents DNA binding. With no repressor bound to the operator region, the RNA polymerase can access to the promoter and hence initiating the transcription.

Furthermore, since a chemical equilibrium exists between bound and unbound repressor molecules, the operator region is not continuously occupied by the LacI and thus, there is generally a low basal level of transcription of the *lac* genes.

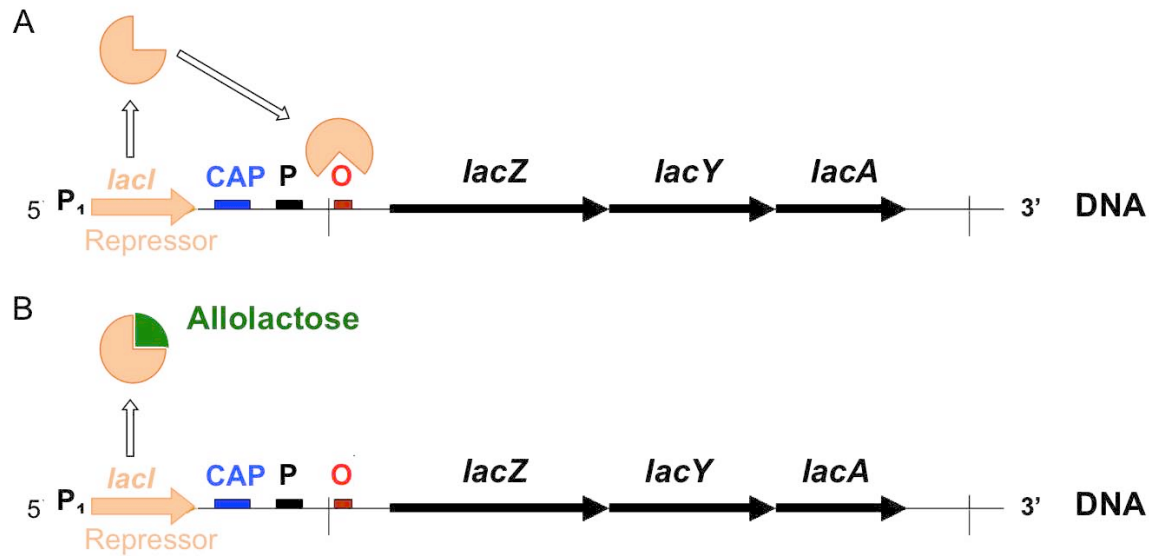


Figure 1.6 Model for the regulation of the *lac* operon. A) The Lac repressor or LacI binds to the *lac* operator (O) and prevents transcription of *lacZ*, *lacY*, and *lacA* genes. B) If lactose is present inside the cell, it is converted to allolactose. Allolactose binds to LacI and prevents its interaction with *lac* operon (O). This allows the expression of the *lacZ*, *lacY*, and *lacA* genes. Adapted from Wikipedia.

The second control mechanism depends on the binding of the catabolite activator protein (CAP) to the promoter region (Beckwith 1987). In absence of glucose in the culture medium cyclic adenosine monophosphate (cAMP) is produced. The cAMP can combine with the CAP protein, creating a complex that binds to the promoter. This increases the affinity of the promoter for RNA polymerase, thereby increasing transcriptional levels of the *lac* genes (Figure 1.7).

Catabolite repression allows *E. coli* to metabolize glucose preferentially prior to lactose (and other sugars) when a mixture of sugars is presented in the growth medium. As a contrary, the production of cAMP is reduced in high glucose concentration. Thus, even in the presence of inducer, a high glucose concentration leads to a low level of transcription from the *lac* promoter (Pinsach, de Mas, and López-Santín 2008).

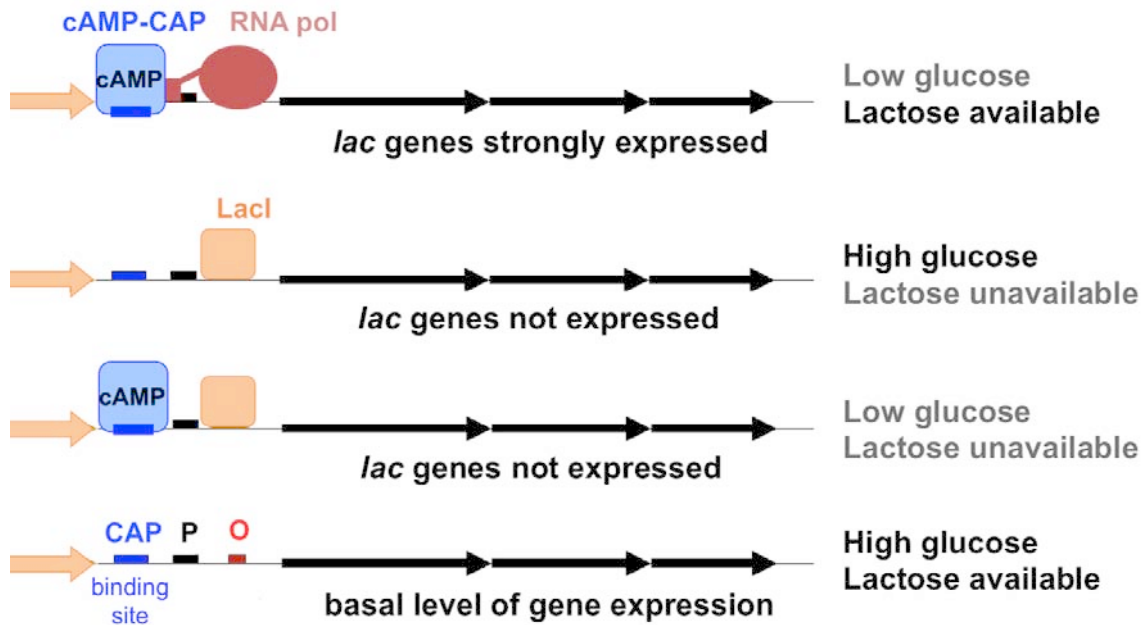


Figure 1.7 Effect of glucose and lactose concentration in the growth medium on the transcriptional levels of the *lac* promoter of *E. Coli*. Lacl, repressor protein, product of the *lacl* gene (orange); CAP, catabolite activator protein (blue); P, promoter of the *lacZ*, *lacY* and *lacA* genes; O, *lac* operator (red); cAMP, cyclic adenosine monophosphate; RNA pol, RNA polymerase; AUG, start codon. Adapted from Wikipedia

The *E. coli* lactose utilization is one of the well-known regulation mechanisms. Many promoters have been constructed from *lac*-derived regulatory element (Polisky, Bishop, and Gelfandt 1976). The *lac* and its closer relative *lacUV5* are rather weak promoters, but some synthetic-hybrid promoters derived from them, like *tac* and *trc* are much stronger (Table 1.2).

Other commonly used promoter systems for the regulation of recombinant protein production are presented in Table 1.2. In the last years the pET family vectors (Invitrogen) with its strong T7 promoter, have gained increasing popularity. The T7 sequence, isolated from the homonymous bacteriophage T7, is not recognized by *E.*

coli RNA polymerase. Gene expression with this system requires a host strain that carries a chromosomal copy of the gene encoding the T7 RNA polymerase. This RNA polymerase gene is usually under control of an inducible *lac* promoter (Baneyx 1999a).

Furthermore, the *araBAD* (or P_{BAD}) promoter, isolated from the arabinose operon of *Salmonella typhimurium* is another sugar-inducible promoter used in the pARA and pBAD families of expression vectors. The dimeric protein AraC is responsible for both negative and positive regulation of the *araBAD* promoter (Brautaset, Lale, and Valla 2009). In general, genes cloned under the control of the PBAD promoter can be efficiently expressed, and this system also allows tight regulation and inexpensive induction, where the gene expression can be turned on and off rapidly by changing the sugar concentration in the medium. In the presence of arabinose, in fact, the PBAD promoter is rapidly turned on, while in the absence of L-arabinose and in the presence of glucose, very low or undetectable levels of transcription occurs.

The same transcription regulation system can be found in the *rhaBAD* promoter, where L-rhamnose acts as an inducer, even if the *rhaBAD* promoter has been stated to be even more tightly regulated than *araBAD* (Brautaset et al. 2009) (Rosano and Ceccarelli 2014).

Finally, an alternative of promoters that need the addition of inducer molecule in order to obtain recombinant protein production could be the thermo-regulated expression systems. In temperature inducible promoters, such as P_{λ} , gene expression is inhibited at culture temperatures below 37°C (normally in the range of 28-32°C) whereas transcription by the host RNA polymerase ensues upon inactivation

of the mutant repressor by increasing the temperature above 37°C (Valdez-Cruz et al. 2010). Although these are strong and finely regulated promoter and toxic or expensive chemical inducers is avoided, their use in industrial applications is limited due to heat transfer issues in large bioreactors.

1.2.4 IPTG as inducer of recombinant protein

Induction of all *lac* promoters could be achieved by adding non-metabolizable lactose analog isopropyl- β -D-1- thiolgalactopyranoside (IPTG) (Figure 1.8). This molecule is analogue of allolactose, the natural inducer of the *lac* operon, but, unlike allolactose, IPTG is non-hydrolyzable by β -galactosidase. Thus, this inducer is entirely available for induction purpose maintaining the level of induction constant in an experiment. As it can be seen in Figure 1.8, IPTG is a 9 carbon sugar of MW = 238.3 g·mol⁻¹. Once IPTG enters into the cell, it binds and triggers the release of the LacI repressor protein from the *lac* operon, allowing the RNA polymerase to start the transcription of the gene cloned downstream the promoter.

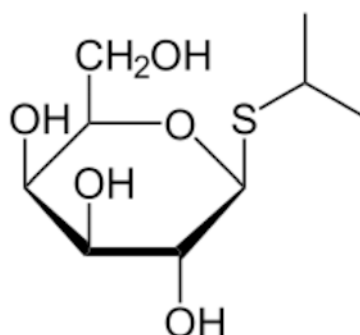


Figure 1.8 IPTG molecular structure. Image takes from Wikipedia

Concentration used depends on the strength of induction required, as well as the genotype of cells and/or plasmid used (Sørensen 2010). The induction point as well as the inducer concentration are the most critical factors which influence product yield (Donovan, Robinson, and Glick 1996). Low inducer concentration may result in an inefficient induction (low recombinant protein yields), whereas expensive inducers, like in the case of IPTG, added in excess can result in an important economic loss or in toxic effects, including reduced cell growth and/or recombinant protein production. It is common to set an in excess concentration of IPTG to ensure the total induction of the system in small scale experiments. A wide range of IPTG concentrations have been reported, although overexpression is commonly induced at IPTG concentrations of 500-1000 μM (Donovan et al. 1996). Nevertheless under these conditions, growth inhibition occurs as a result of the metabolic burden. When working with strong promoters, much lower concentrations (in the range between 50-100 μM) have been shown to be enough in order to obtain the maximum production, and higher concentration did not lead to an increase in the specific activity and/or yield (Vidal et al. 2005)(Ruiz et al. 2009).

1.3 Gene expression and regulation

Gene expression in bacteria, as in all cells, involves two steps: i) transcription of DNA into the messenger RNA (mRNA) by RNA polymerase, and ii) translation of the message into protein by the protein synthesis machinery.

In prokaryotes, transcription and translation occur in the same compartment (cytosol),

as a result both processes are closely linked, and translational issues are more critical for gene expression than they are in the case of eukaryotes (Ferrer-Miralles et al. 2009). Transcription starts when RNA polymerase binds to the promoter region and proceeds to copy the DNA sequence into mRNA (Figure 1.9). RNA polymerase catalyzes the formation of phosphodiester bonds between ribonucleotides using DNA as a template. Transcription continues until core RNA polymerase encounters a transcription termination signal or terminator, where both the RNA polymerase and the newly synthesized RNA dissociate from the DNA. The transcription terminator is located downstream of the coding sequence and serve both as a signal to terminate transcription and as a protection element composed of stem loop structure, protecting the mRNA from exonucleolytic degradation and extending the mRNA half-life (Jana and Deb 2005).

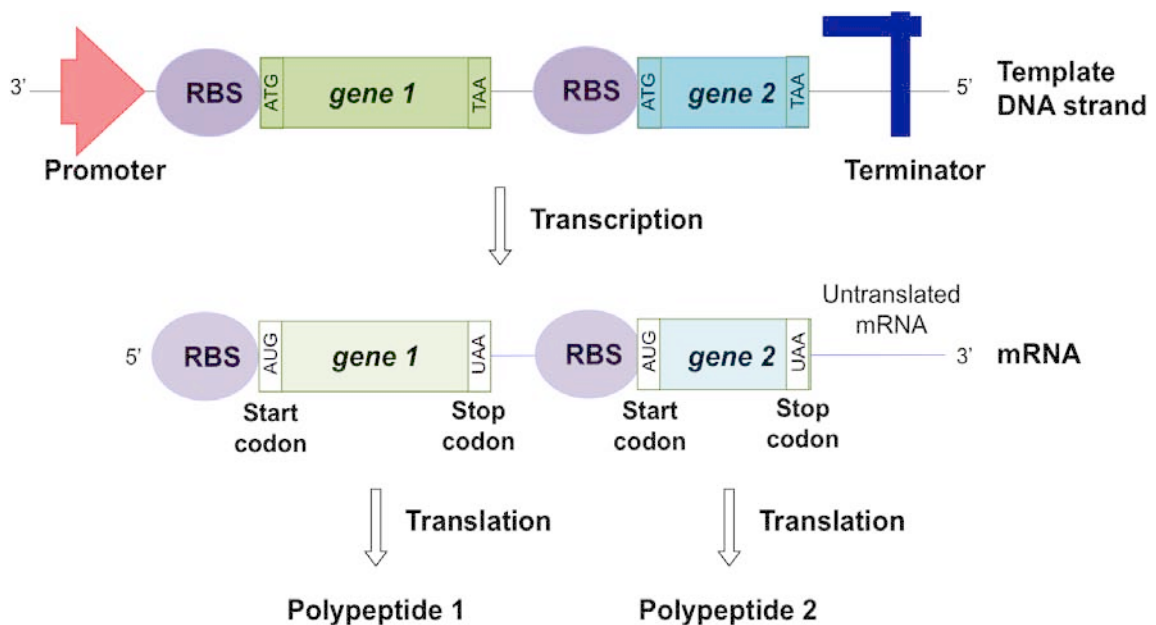


Figure 1.9 Schematic representation of the transcription and translation mechanism. Adapted from (Cooper 2000).

Once complete the mRNA synthesis, ribosomes immediately attach to it and start the translation of the messenger into the polypeptides.

The translational initiation region of most sequenced *E. coli* genes (91 %) contains the codon AUG. GUG is used by about 8 % of the genes and UUG is rarely used.

Proper binding of transcribed mRNA to the ribosomal machinery requires the presence of a ribosome binding site (RBS) comprising the Shine-Dalgarno (SD) sequence, after its discovery, that helps recruit the ribosome to the mRNA to initiate the protein synthesis (Mott and Berger 2007). The spacing between the SD site and the initiating AUG codon can vary from 5 to 13 nucleotides, and it influences the efficiency of translational initiation.

Finally, the presence of a stop signal in the mRNA is an indispensable component to the translation termination process. In addition to the three termination codons UAA/UGA/UAG in *E. coli*, this complex event involves specific interactions among the ribosome, mRNA and several release factors at the site of termination (Jana and Deb 2005).

Recent advances in the synthetic biology, allowed the development of new methods and tools to speed up and standardize strain engineering. Compared with conventional DNA cloning protocols, these advanced DNA assembly strategies offer an efficient approach to construct multi-gene pathways in a one-step, scar-less, and sequence-independent manner. In particular, the Parts Registry is a collection of standardized biological parts (BioBricks) that allow the fast assembly of new functions (Voigt 2006) (Heinz and Neumann-Staubitz 2010) (Vick et al. 2011). Individual parts or combinations of parts that encode defined functions can be independently tested and characterized

in order to improve the expression system (Yokobayashi, Weiss, and Arnold 2002). DNA construction based on the BioBrick theory has become a key part of most metabolic engineering projects and genetic circuits design. The BioBrick concept exploits the advantage that the same promoters, ribosome binding sites, expression tags, antibiotic resistances and origins of replication are frequently reused, with only the genes of interest being varied (Heinz and Neumann-Staubitz 2010) (Vick et al. 2011).

1.4 Synthetic biology

Recent advances in the synthetic biology, allowed the development of new methods and tools to speed up and standardize strain engineering. Compared with conventional DNA cloning protocols, these advanced DNA assembly strategies offer an efficient approach to construct multi-gene pathways in a one-step, scar-less, and sequence-independent manner. In particular, the Parts Registry is a collection of standardized biological parts (BioBricks) that allow the fast assembly of new functions (Voigt 2006) (Heinz and Neumann-Staubitz 2010) (Vick et al. 2011). Individual parts or combinations of parts that encode defined functions can be independently tested and characterized in order to improve the expression system (Yokobayashi et al. 2002). DNA construction based on the BioBrick theory has become a key part of most metabolic engineering projects and genetic circuits design. The BioBrick concept exploits the advantage that the same promoters, ribosome binding sites, expression tags, antibiotic resistances and origins of replication are frequently reused, with only the genes of interest being

varied (Heinz and Neumann-Staubitz 2010) (Vick et al. 2011). Thus, with a series of standardized biological parts, researchers should be able to quickly assemble biological parts (bricks) into different biological functions, pathways and mechanisms.

The recent advances on synthetic biology have allowed to produce novel compounds (Heinz and Neumann-Staubitz 2010), including high-performance enzymes with applications in the biotechnological sector such as the non ribosomal peptides synthetases (NRPs) and the polyketide synthetases (PKSs) (Li and Neubauer 2014) (Li et al. 2014).

2 Context and objectives

2.1 Context of the thesis work

This thesis has been developed into the “Bioprocesses Engineering and Applied Biocatalysis” research group. During the last years, the main activities of the group have been focused to process development in *E. coli* protein production of biocatalyst (aldolases) and their utilization in biocatalysis.

Previous works within the group related to recombinant protein production consisted in: i) media formulation for aldolase production (Durany et al. 2004) (Vidal et al. 2008); ii) implementation of glucose-limited batch operation with a predefined exponential profile in order to obtain recombinant aldolases in HCDC (Pinsach, de Mas, and Lopez-Santin 2006, 2008); iii) development of an alternative expression system based on glycine auxotrophy to ensure plasmid stability avoiding antibiotic supplementation was employed for the production of rhamnulose-1-phosphate aldolase (RhuA) (Vidal et al. 2008); iv) study of the mechanisms of transport involved in IPTG uptake by *E. coli* and optimization of RhuA production in fed-batch HCDC (Fernández-Castañé et al. 2010, 2012); v) development of a model of the whole recombinant production process, in order to obtain all the different equations needed to describe and predict HCDC of *E. coli* (Calleja et al. 2014).

The overall objective of this thesis is to develop a robust and manageable process for the production the fuculose 1-phosphate aldolase (FucA) in *Escherichia coli*. It is well

know that the over-expression of proteins in recombinant host cells often requires a significant amount of resources causing an increase in the metabolic load for the host. This results in a variety of physiological responses leading to altered growth parameters, including growth inhibition or activation of secondary metabolism pathways. Moreover, the expression of other plasmid-encoded genes such as antibiotic resistance genes or repressor proteins may also alter growth kinetics.

2.2 Objectives

The main objective of this work is the development of a second-generation system suitable for *Escherichia coli* expression with an antibiotic-free plasmid maintenance mechanism in order to completely avoid the presence of the antibiotic resistance gene and to achieve maximum protein yields.

The principal objective was then translated into several secondary objectives:

- Clone and expression of the *fucA* gene into the auxotrophic M15 Δ *glyA* strain (Vidal et al. 2008). Test of the auxotrophic expression system in shake flasks and fed-batch cultures in defined media and comparison with its predecessor
- Study of the possibility to create a single-based plasmid expression system in order to eliminate the presence of the antibiotic resistance genes.
- Minimization of the metabolic burden related to plasmid maintenance and heterologous protein expression.

- Further engineer the expression system in order to be completely devoid of antibiotic resistance marker genes.
- Test the new strains behavior and the FucA overproduction in *E. coli* HCDC

Moreover, study for the construction of an expression system able to secrete the recombinant protein in the extracellular medium.

3 Materials and Methods

3.1 Strains

Along this work, different *E. coli* strains have been employed both for the clonation and the overexpression of the recombinant protein Fuculose-1-Phosphate Aldolase. The entire list of the microorganisms used is presented below.

3.1.1 *Escherichia coli* DH5 α

Genotype: F⁻ Φ 80dlacZ Δ M15 endA1 recA1 hsdR17 (r_k⁻, m_k⁺), supE44 thi⁻¹ gyrA96 relA1 Δ (lacZYA-argF)U169 λ ⁻.

The *E. coli* DH5 α is a commercial strain (Invitrogen) used for vector propagation, cloning and gene expression. Like many cloning strains, DH5 α has several features that make it useful for recombinant DNA methods:

- endA1 mutation inactivates an intracellular endonuclease that degrades plasmid DNA during miniprep methods
- Δ (lacZ)M15 is the alpha acceptor allele needed for blue-white screening with many lacZ base vectors
- recA eliminates homologous recombination

3.1.2 *Escherichia coli* DH5 α LacI^{q1}

The *E. coli* DH5 α LacI^{q1} derives from the commercial DH5 α where the *lacI* gene has been introduced by conjugation into the genome under the control of the q1 promoter. The main characteristic of the q1 promoter is the deletion of 15 bp in the -35 region, upstream of *lacI*. This deletion created a consensus -35 hexamer, responsible for the increased *lacI* transcription (Glascock and Weickert 1998a).

3.1.3 *Escherichia coli* M15

Genotype: nal^s, str^s, rif^s, thi⁻, lac⁻, ara⁺, gal⁺, mtl⁻, F⁻, recA⁺, uvr⁺, lon⁺.

The *E. coli* M15 is a commercial strain (Qiagen) K12 derived (Villarejo and Zabin 1974; Zamenhof and Villarejo 1972). This strain has been used for cloning and expression of the fuculose-1-phosphate aldolase. The key features of this strain are:

- M15 deletion in the *lacZ* gene (Beckwith, 1964)

3.1.4 *Escherichia coli* M15 Δ glyA

E. coli M15 derived with the deletion of the *glyA* locus of the chromosome (Vidal et al. 2008). The *E. coli glyA* gene encodes for the enzyme serine hydroxymethyl transferase (SHMT). This 46 kDa enzyme has two activities: i) threonine aldolase activity, which catalyzes the reversible interconversion between L-threonine and glycine plus acetaldehyde, and ii) serine hydroxymethyl transferase activity, which catalyzes the reversible interconversion between serine and glycine.

3.1.5 *Escherichia coli* MG1655 LacI^{q1}

Phenotype: F⁻, [proAB lacIqZΔM15 Tn10 (TetR)], λ⁻, ilvG⁻, rfb⁻⁵⁰, rph⁻¹.

The *E. coli* MG1655 LacI^{q1} is a K12 MG1655 derived strain, where, like the DH5α LacI^{q1} strain, the lacI gene under the control of the q1 promoter was introduced by conjugation into the *E. coli* MG1655 genome, resulting in constitutive overproduction of the repressor protein.

3.1.6 *Escherichia coli* BL21 (DE3)*

BL21(DE3)*, from now on BL21*, is a derivative of *E. coli* BL21 that contains the rne131 mutation. rne is an essential gene encoding the enzyme RNase E, which produces a truncated enzyme that results in increased RNA stability and can increased yields of recombinant protein. The DE3 designation indicates the strains contain the DE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter.

3.2 Plasmids

The expression systems used during the course of this thesis project are presented below. All the expression vector-engineering purposes of the different plasmids are explained in detail in the corresponding result Chapters.

3.2.1 pQE-40

The commercial vector pQE-40 (Qiagen) was used as reference expression vector for the protein of interest, FucA. This expression vector is based on the IPTG-induced T5 promoter, derived from the T5 phage. This promoter is recognized by *E. coli* RNA polymerase, and has a double *lac* operator (*lac O*) repression module in series to provide tightly regulated and high-level expression of recombinant proteins (Figure 3.1).

Besides, noteworthy among its features:

- Ribosomal binding site RBSII with a high transcription level
- 6X his tag (histidine affinity tag coding sequence) at the 5' extreme of the multiple cloning site (MCS). This 6X his tag will be incorporated in the final polypeptide for downstream processes
- Two transcriptional termination regions: t_0 from phage lambda and T1 from the operon *rrnB*
- ColE1 replication origin (low copy number)

- β -lactamase coding sequence (*bla* gene), conferring ampicillin antibiotic resistance

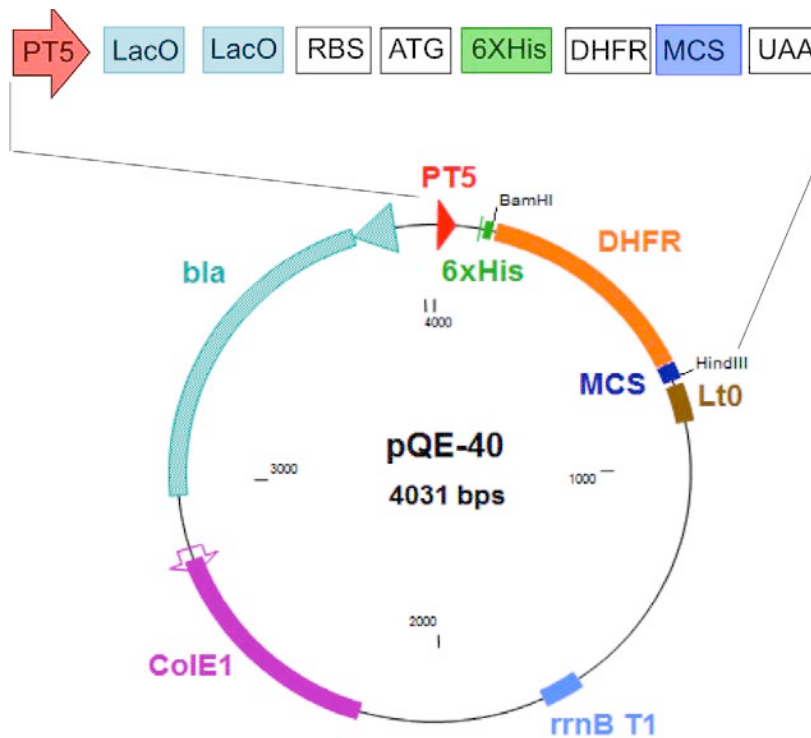


Figure 3.1 pQE-40 expression vector. PT5, promoter T5; LacO, 2 Lac operator regions; RBS, ribosome binding site; ATG, start codon; 6xHis, histidine affinity tag coding sequence; DHFR, dihydrofolate reductase gene; MCS, multi cloning site; UAA, stop codon; Lt0, lambda t0 transcriptional termination region; rrnBT1, transcriptional termination region; ColE1, replication origin; *bla*, ampicillin resistance gene.

3.2.2 pREP4

The pREP4 plasmid (Figure 3.2) is derived from pACYC and contains the p15A replicon. Multiple copies of pREP4 are present in the host cells that ensure the production of high levels of the lac repressor protein that binds to the operator sequences and tightly regulates recombinant protein expression. In fact, the extremely high

transcription rate initiated at the T5 promoter can only be efficiently regulated and repressed by the presence of high levels of the lac repressor protein.

The pREP4 main features are:

- *lacI^q* gene, constitutively expresses the lac repressor protein
- Medium-copy p15A replication origin that gives like 15-20 copy per cell
- *kan* gene which confers kanamycin resistance

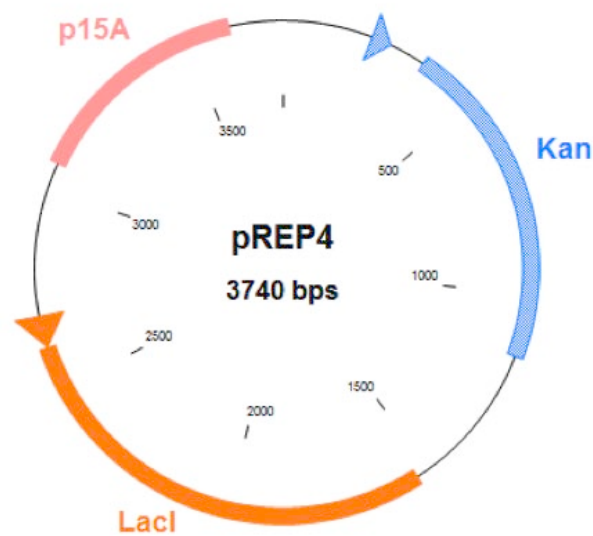


Figure 3.2 pREP4 plasmid. P15A, replication origin; *kan*, kanamycin resistance gene; *lacI* gene codify for the repressor protein.

3.2.3 pUC-19

pUC-19 is one of a series of plasmid cloning vectors created by Messing and co-workers in the University of California (the letter p stands for plasmid and UC represents the University in which it was created). pUC-19 is one of the most widely used vector molecules as the recombinants, or the cells into which foreign DNA has

been introduced, can be easily distinguished from the non-recombinants based on color differences of colonies on growth media (Figure 3.3).

The main characteristics of the pUC-19 are:

- *LacZ* gene, N-terminal fragment of the β -galactosidase
- MCS, multiple cloning site is split into the *lacZ* gene
- pMB1 replication origin, pUC-18 derived (copy number of 100-300 per cell)
- *bla* gene which confers ampicillin resistance

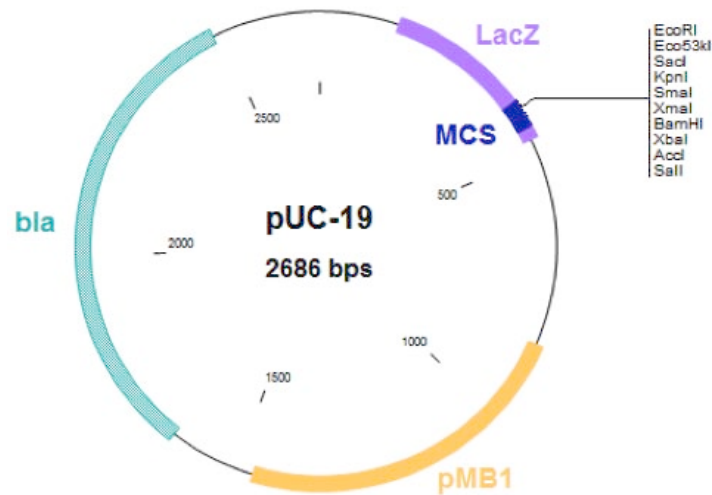


Figure 3.3 pUC-19 plasmid. *LacZ*, β -galactosidase gene; MCS, multi cloning site; pMB1, replication origin; *bla*, ampicillin resistance gene.

3.2.4 pSB1C3_GG plasmid

The pSB1C3_GG is a high copy BioBrick assembly plasmid that derived from the pSB1C3 vector. The original backbone has been modified with a PCR reaction introducing two recognition sequences for the type IIs restriction enzyme *BsaI* (Figure 3.3). The *BsaI* sites flank the *mCherry* reporter gene coding for red fluorescence

protein (RFP) and are oriented to eliminate themselves and the *mCherry* during digestion and ligation of the Golden Gate method (3.3.4). Both the replication origin and the antibiotic resistance marker were maintained as standard parts.

The pSB1C3_GG's keys features are:

- pMB1 replication origin, pUC-18 derived (copy number of 100-300 per cell)
- *mCherry* reporter gene, it codifies for a red fluorecence protein (RFP)
- *Bsal* restriction sites with two different overhangs (pink and green squares)
- *Cam*, chloramphenicol resistance gene

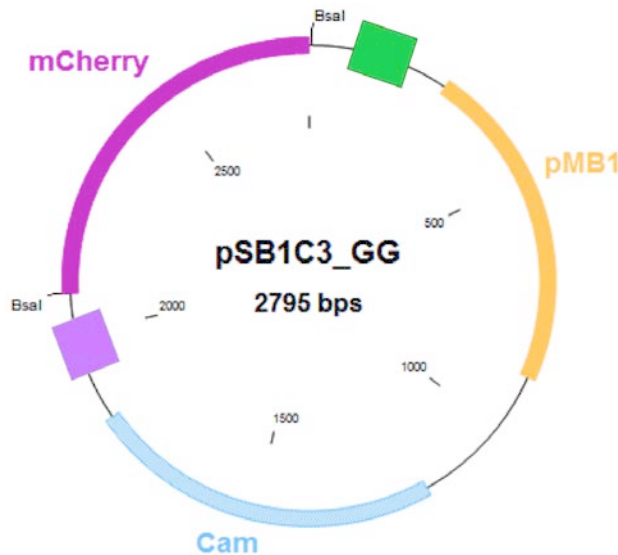


Figure 3.4 pSB1C3_GG plasmid. pMB1, replication origin; *Cam*, kanamycin resistance gene, *mCherry*, report gene.

3.2.5 pASK-Pet

The pASK-Pet derives from the pASK-IBA33plus (IBA BioTAGnology). The entire sequence is not available. The original backbone has been modified by cloning the *pet*

gene PCR-amplified from pBAD-Pet (Sevastyanovich et al. 2012). The *pet* gene codifies for the *E. coli* SPATE protein, which possesses high amino acid sequence identity and structural similarity: the passenger domain consists of a central β -helical stem decorated with several discursive subdomains and is connected to the characteristic β -barrel by a short α -helical peptide (Figure 3.5, 3.6).

The pASK-Pet's keys features are:

- ColE1 replication origin (Medium copy number, 15-60 copy per cell)
- *tet* promoter/operator (anhydrotetracycline inducible promoter)
- β -lactamase coding sequence (*bla* gene), conferring ampicillin antibiotic resistance

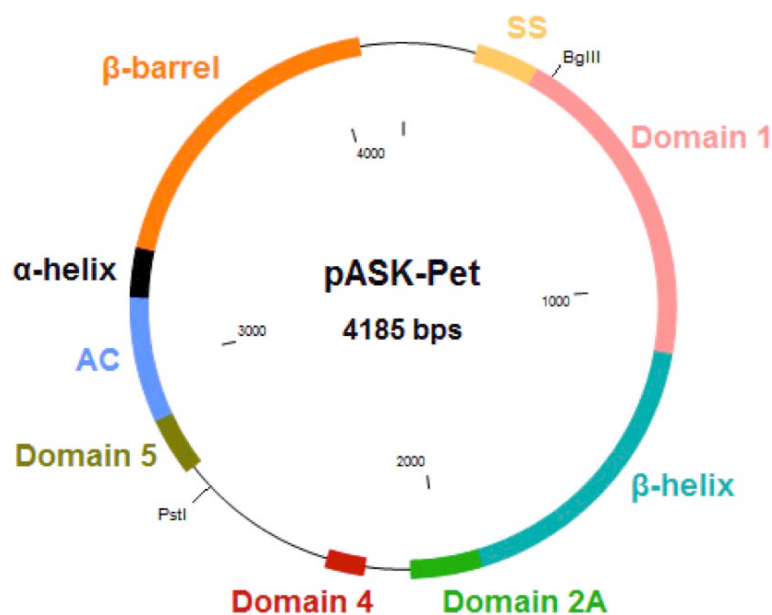


Figure 3.5 pASK-Pet expression vector. SS, signal sequence; Domain 1, serine protease domain; β -helix; Domain 2A; Domain 4; Domain 5; AC, auto-chaperone domain; α -helix region connects the AC to the β -barrel and contains the cleavage site

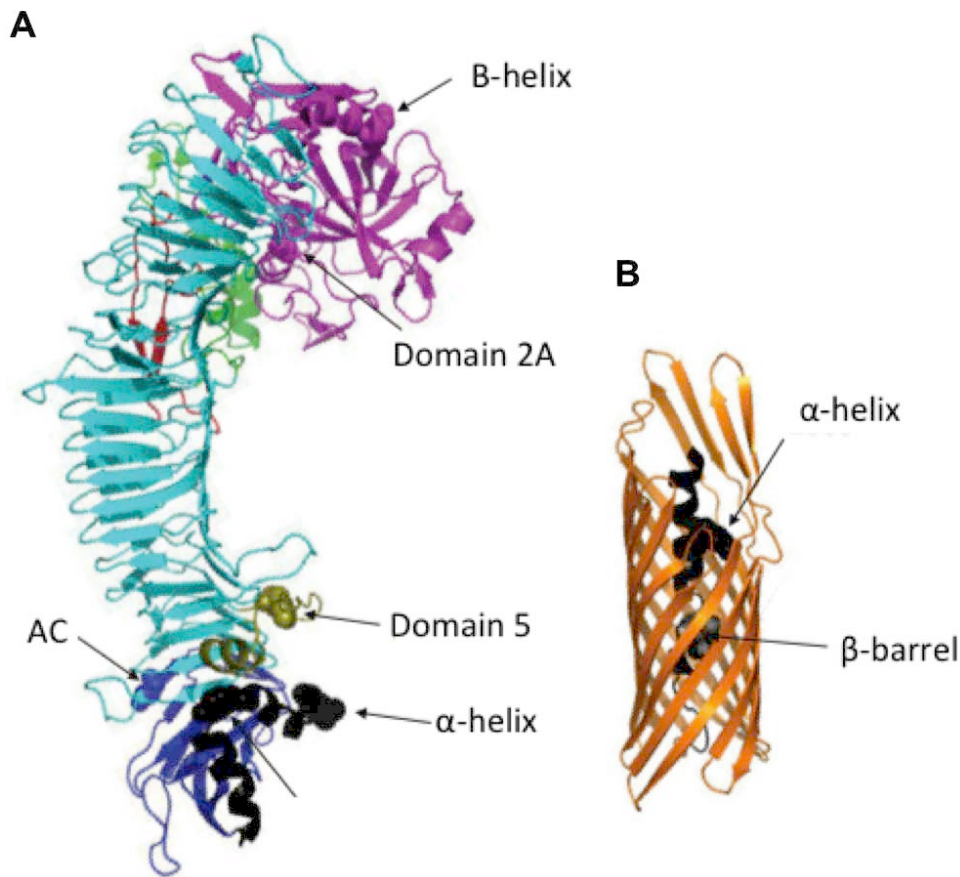


Figure 3.6 pASK-Pet expression vector. A) A three-dimensional model of the Pet passenger domain. Domain 1, encompassing the serine protease domain, colored in magenta; β -helix is in cyan; domain 2A is in green; domain 4 is in red; domain 5 is in olive; AC, autochaperon is in blue; α -helical region connecting the passenger domain to the β -barrel is in black. B) Cartoon representation of the Pet β -domain. The β -barrel is in orange and the α -elix is in black (Sevastyanovich et al. 2012).

3.2.6 pET22b-PetAC

The pET33b-PetAC derives from the pET22b plus backbone (Novagen). The entire sequence is not available. The original backbone has been modified by cloning the *pet* gene PCR-amplified from pBAD-Pet (Sevastyanovich et al. 2012). The *pet* gene has been modified by Sevastyanovich et al. 2012, in order to obtain the autotransporter sequence with only the auto-chaperone (AC) and the β -barrel domains (Figure 3.7).

The pET22b-PetAC's keys features are:

- T7 promoter, IPTG-induced
- 6X his tag (histidine affinity tag coding sequence)
- *lacI* gene, constitutively expresses the lac repressor protein
- β -lactamase coding sequence (*bla* gene), conferring ampicillin antibiotic resistance

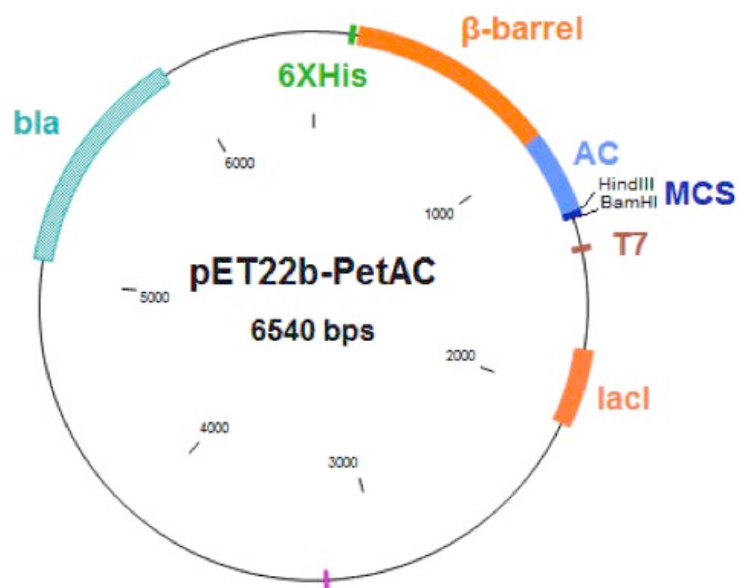


Figure 3.7 pET22b-PetAC expression vector. 6xHis, histidine affinity tag coding sequence; β -barrel; with the C-terminus which inserts into the outer membrane; AC, auto-chaperone domain; MCS, multi cloning site; T7, inducible promoter; *lacI*, *lacI* gene; *bla*, ampicillin resistance gene (Sevastyanovich et al. 2012)

3.3 Molecular biology techniques

Basic molecular biology techniques related to recombinant DNA technology have been used in this work. General protocols are described in laboratory manuals and in the manufacturer's instructions (Sambrook et al., 1989). Below, different methods and proceedings are exposed.

3.3.1 DNA isolation and manipulation

DNA extraction. Plasmid DNA isolation is an essential technique used in the process of molecular cloning to construct new clones and to analyze the ones obtained. For extraction of plasmid DNA from recombinant strains, commercial Promega PreYield™ plasmid miniprep and midiprep System kits were used. The DNA extraction is based on the alkaline lysis method invented by the researchers Birnboim and Doly in 1979. In fact, under alkaline conditions (pH 12.0-12.5) both chromosomal DNA and protein are denatured, while the plasmid DNA is bind to a resin, which is immobilized within a filter plate and finally eluted using a low-ionic-strength buffer.

Enzymatic modifications. DNA restriction procedures were carried out using restriction enzymes from Thermo Fisher Scientific (Waltham, MA, USA). Enzyme concentrations, DNA, buffer, ionic strength, temperature and reaction time in each case depend on the enzyme and the specific application. In general, a restriction enzyme unit is defined as the amount of enzyme required to completely digest 1 µg DNA in 60 min under optimum conditions. The endonucleases used in this work are listed in the Table 3.1.

Ligation of DNA fragments is a technique used to covalently link two ends of DNA, in particular, to join together the DNA fragment (insert) with the destination vector. The enzyme used for this reaction is T4 DNA ligase (Roche), which catalyzes the bonding of cohesive or blunt ended double-stranded DNA.

Table 3.1 Restriction enzymes used along this thesis work

Enzyme	Cleavage site
<i>Bam</i> HI	5'...G↓GATCC...3' 3'...CCTAG↑G...5'
<i>H</i> IndIII	5'...A↓AGCTT...3' 3'...TTCGA↑A...5'
<i>S</i> maI	5'...CCC↓GGG...3' 3'...GGG↑CCC...5'
<i>T</i> th111I	5'...GACN↓NNGTC...3' 3'...CTGNN↑NCAG...5'
<i>S</i> apI	5'...GCTCTTC(N) ¹ ↓...3' 3'...CGAGAAG(N) ⁴ ↑...5'
<i>M</i> scl	5'...TGG↓CCA...3' 3'...ACC↑GGT...5'
<i>X</i> baI	5'...T↓CTAGA...3' 3'...AGATC↑T...5'
<i>B</i> pu10I	5'...CC↓TNAGC...3' 3'...GGANT↑CG...5'
<i>K</i> pn2I	5'...T↓CCGGA...3' 3'...AGGCC↑T...5'
<i>B</i> saI	5'...GGTCTC(N) ¹ ↓...3' 3'...CCAGAG(N) ⁵ ↑...5'
<i>E</i> coO109I	5'...RG↓GNCCY...3' 3'...YCCNG↑GR...5'
<i>E</i> am1105I	5'...GACNNN↓NNGTC...3' 3'...CTGNN↑NNNCAG...5'

DNA electrophoresis. The agarose gel electrophoresis is based on the separation of DNA fragments subjected to an electric field in the presence of a mesh (agarose polymer) that impedes the migration towards the anode. DNA separation depends on several parameters such as the length of DNA, structural conformation, the pore size of the gel and the intensity of the applied current. The net negative charge of the DNA allows its migration in an electric field from the cathode toward the anode with a speed inversely proportional to the logarithm of its molecular weight. Thus, the smaller is a DNA molecule will migrate more easily through the gel. This technique not only allows to identify DNA fragments but also to isolate them for further manipulation steps. In this thesis work, DNA fragments were separated over agarose gels 1% or 2% (A9539_SIGMA) (depending on the size of the destination fragment), prepared in TAE containing Syber[®] safe DNA gel stain (*Life technology*). Samples were prepared with the commercial 6X DNA loading dye (Thermo scientific) and then loaded in the agarose gel. Gels were run at 90 V in TAE as running buffer. The simultaneous running of a known molecular weight marker allows establishing, by comparing the banding pattern, the molecular weight of the samples of interest. In order to extract DNA fragments after electrophoretic separation, the desired band was cut off the gel with a clean cutter under UV illumination and, finally, purified using the Promega Wizard[®] SV gel and PCR clean-up system.

3.3.2 DNA transformation

Transformation of heterologous DNA into *E.coli* was achieved by two different methods: chemical transformation and electroporation.

Chemical transformation

Chemical competence was conferred to *E.coli* by re-suspension in CaCl_2 solution at 0°C . Under these conditions, the Ca_2^+ ion was thought to create pores in the membrane, assist binding of the DNA to the cell membrane and mask the negative charge on the DNA, easing its passage through the hydrophobic cell membrane. The DNA was forced into the cells by applying a short 42°C heat shock, which resulted in a thermal current that sweeps the DNA into the cells.

Chemical competent cells were prepared as follows:

- Incubate the cells 2×10 mL in LB medium, with the corresponding selection conditions, at 37°C in agitation overnight (o/n)
- Inoculate 100 mL fresh LB with 5 mL of the o/n culture and incubate at 37°C up to an $\text{OD}_{600\text{nm}}$ of 0.7-0.8
- Place the cells in ice for 20-30 min (split the culture in 2×50 mL falcon tubes)
- Cells were collected by centrifugation at 4°C , 3000 g for 15 min
- Decant the supernatant and resuspend gently the pellet in about 100 mL of ice-cold 0.1 M CaCl_2
- Keep the cells on ice for 30 min, then harvest the cells by centrifugation at 4°C , 2000 g for 15 min

- Remove the supernatant and resuspend the pellet in 50 mL of ice cold 0.1 M CaCl₂ plus 15 % Glycerol
- Harvest the cells by centrifugation at 4°C, 1000 g for 15 min
- Decant the supernatant and resuspend the cells in 2 mL of ice cold 0.1 M CaCl₂ plus 15 % Glycerol
- Aliquot the cells into pre-chilled eppendorf tubes (50 µL each) and freeze immediately to -80°C

For the transformation the following steps were carried out:

- Add 100 µL of Electrocompetent cells to 50-100 ng of DNA and mix gently together by swirling with the pipette tip
- Incubate the mix on ice for 1 hour
- Heat-shock at 42°C for 1min without agitation and then put again the cells on ice for 1-2 min
- Immediately, add 1 mL of LB medium incubate the cells at 37°C with agitation for 90min, in order to allow the expression of antibiotic resistance
- Spin down for 2 min, then resuspend the pellet with 1 mL of fresh LB
- 50-100 µL of the cell culture were spread on LB and on selective plates with the proper antibiotic, and then incubated overnight at 37°C

Electroporation

Electroporation is a molecular biology technique in which a high voltage pulse of electrical current is applied to cells in order to increase the permeability of the cell membrane, creating transient pores, allowing chemicals or DNA to be introduced into the cell. Electroporation is carried out at 0°C to minimize heat damage to the cells.

Electrocompetent cells were prepared as detail next:

- Incubate the cells 2x10 mL in LB medium, with the corresponding selection conditions, at 37°C in agitation overnight
- Inoculate 100 mL fresh LB with 5 mL of the o/n culture and incubate at 37°C up to an OD_{600nm} of 0.7-0.8
- Place the cells in ice for 15-20 min (split the culture in 2x50 mL falcon tubes)
- Cells were collected by centrifugation at 4°C, 4000 g for 20min
- The pellet was resuspended in 45 mL of ice-cold 10 % glycerol; cells were recovered by centrifugation at 4°C, 4000 g for 20min
- Discard the supernatant and resuspend the pellet with 25 mL cold glycerol 10 %; cells were separated by centrifugation at 4°C, 4000 g for 20min
- The pellet was resuspended in 12.5 mL of ice-cold 10 % glycerol; cells were collected by centrifugation at 4°C, 4000 g for 20 min
- Discard the supernatant and resuspend the pellet with 1 mL cold glycerol 10 % (join it in one Eppendorf); centrifuge the cells for the last time at 4°C, 4000 g for 20 min
- The final pellet was resuspended in 100 ice-cold 10 % glycerol and aliquoted into pre-chilled eppendorf tubes (50 µL each) and freeze immediately to -80°C

For the transformation the following steps were carried out:

- Add 50 µL of Electrocompetent cells to 50-100 ng of DNA and mixed gently together by swirling with the pipette tip

- Transfer the mix to a cold 2.0 mm electroporation cuvette (Biorad), being careful of not letting air bubbles form and proceed to electroporation
- Electroshock the cells in a Gene Pulser II electroporator from Biorad with a pulse ($V = 2500$ v; $C = 25$ μ F; $R = 200$ Ω)
- Immediately, add 1 mL of LB medium and transfer the volume to a 15 mL falcon tube and incubate at 37°C with agitation for one hour, in order to allow the expression of antibiotic resistance
- Spread 50-100 μ L of the cell culture on LB plates with and without the proper antibiotic, and then incubate overnight at 37°C

3.3.3 DNA amplification by PCR

The polymerase chain reaction (PCR) is a molecular biology technique used for DNA amplification. PCR has been used to amplify a specific region of a DNA stand (DNA target) between 0.6 and 4 kilo base pairs (Kbp).

A basic PCR set up requires several components and reagents. These components include:

- DNA template that contains the DNA target to be amplified
- Two primers that are complementary to the 3' ends of each of the sense and anti-sense strand of the DNA target
- DNA polymerase
- Deoxynucleoside triphosphates (dNTPs), the building-blocks from which the DNA polymerase synthesizes a new DNA strand

- Buffer solution, providing a suitable chemical environment for optimum activity and stability of the DNA polymerase
- Bivalent cations, magnesium (Mg_2^+) or manganese (Mn_2^+) ions

PCR consists of a series of 20-40 repeated temperature changes, the cycles, where the three main cycles for DNA amplification consist in:

Denaturation of the double strand DNA by melting it with high temperature

Annealing of the primers to the single-stranded DNA template

Elongation of the new DNA by the DNA polymerase

The PCRs were carried out in a reaction volume between 25-50 μ L in a thermal cycler, following the guidelines provided by the manufacturer of each DNA polymerase. In particular, two different DNA polymerases have been used in this thesis work for DNA amplification: KOD Polymerase Novagen from Merck Biosciences, for fragments till 2.0 Kbp, and Phusion high-fidelity DNA polymerase (Thermo Scientific), for longer DNA fragments. The verification of cloning reactions and transformations colony PCR were performed using the GoTaq[®] master mix (Promega) and following routine protocols as described in Green and Sambrook (Green and Sambrook 2012).

All the primers designed to amplify the genes *LacI*, *fucA* and *glyA* (with its own promoter and terminator) are listed in Table 3.2. The underlined sequences correspond to the target specific enzyme restriction sites.

Table 3.2 Primers designed to amplify the genes *lacI*, *fucA* and *glyA*. T_m, melting temperature in °C.

Name	Sequence 5'-3'	T _m (°C)	Size (pb)	Enzyme
FucA FW	CGGTACGTGGATCCATGGAACGAAATAAACTTGCTCG	61	37	<i>Bam</i> HI
FucA REW	GCTAGTCCAAGCTTACTCTTCAATTCGTAACCCATAGG	61	37	<i>Hind</i> III
<i>LacI</i> FW	CGCATCTGAATGACCCAGTCTCACTGCCCGCTTTCC	61	36	<i>Th</i> 1111
<i>LacI</i> REW	CATGCGTGAAGAGCGTGGTGAATGTGAAACCAGTAAC	61	37	<i>Sap</i> I
αβT FW	GCTTGATCACCGGTGTGAAGACGAAAGGGC	61	31	<i>M</i> scl
αβT REW	CGTCCATGTCTAGAGGGCGGATTTGTCCTACTC	63	33	<i>X</i> baI

All the primers used for the Golden Gate techniques are listed in Table 3.3. The red bases indicate the RBS BBa_B0034 for the *lacI* gene and the green bases specify the RBS J61100 for the *glyA* gene. The recognition site of the *Bsa*I enzyme is underline and the different overhangs are represented in bold with a number.

The four promoters used (J23100, J23111, J23110 and J23117) in this study were selected from a small combinatorial library of constitutive promoters, the J23119 family, where the J23119 is the wild type (wt) promoter sequence and the strongest member of the family (Figure 3.8) (Registry for Standard Biological Parts, <http://parts.igem.com>).

Table 3.3 Primers used for the Golden Gate technique. T_m, melting temperature in °C

Name	Sequence 5'-3'	T _m (°C)	Size (pb)	Enzyme
GG_ <i>lacl</i> FW	TTGGTCTCTTAGC ² TCTAGAGAAAGAGGAGAAATA CTAGGTGAAACCAGTAACGTTATACG	59	60	<i>BsaI</i>
GG_ <i>lacl</i> REW	TTGGTCTCTTCCA ³ TTCAAGATCTTCACTGCCCGCTT TCC	61	39	<i>BsaI</i>
GG_ <i>lacl</i> -S REW	GGTCTCTCGCT ⁴ GCGAAAAACCCCGCCGAAGCGG GGTTTTTTCGCTCACTGCCCGCTTCC	61	61	<i>BsaI</i>
GG_ <i>glyA</i> FW	TTGGTCTCTTGGA ³ GAAAGAGGGGACAACTAGT ATGTTAAAGCGTGAAATGAA	58	53	<i>BsaI</i>
GG_ <i>glyA</i> -S FW	TTGGTCTCTTAGC ² GAAAGAGGGGACAACTAGTA TGTTAAAGCGTGAAATGAA	58	53	<i>BsaI</i>
GG_ <i>glyA</i> REW	TTGGTCTCTCGCT ⁴ GATAACGTAGAAAGGCTTCC	58	33	<i>BsaI</i>
PJ1 FW	GCTAAGGATGATTTCTGGAATTC	59	23	<i>Bpu10I</i>
PJ2 FW	GCTAAGGCGGACTGCAGGAAGAC	58	23	<i>Bpu10I</i>
PJ1-S FW	TCCGGAGATTTCTGGAATTCGAAGACG	58	27	<i>Bpu10I</i>
PJ REW	TGGCCA CGATAACGTAGAAAGGCTTCC	60	27	<i>Kpn2I</i>
PJ-S REW	TGGCCAACGCTGCGAAAAAACCC	60	23	<i>Kpn2I</i>

LegendRBS BBa_B0034 (http://parts.igem.org/Part:BBa_B0034)RBS BBa_J61100 (http://parts.igem.org/Part:BBa_J61100)Terminator: RBS BBa_B1002 (http://parts.igem.org/Part:BBa_B1002)***BsaI*** enzyme^{1,2,3,4} Four different bp overhang for the *BsaI* enzyme

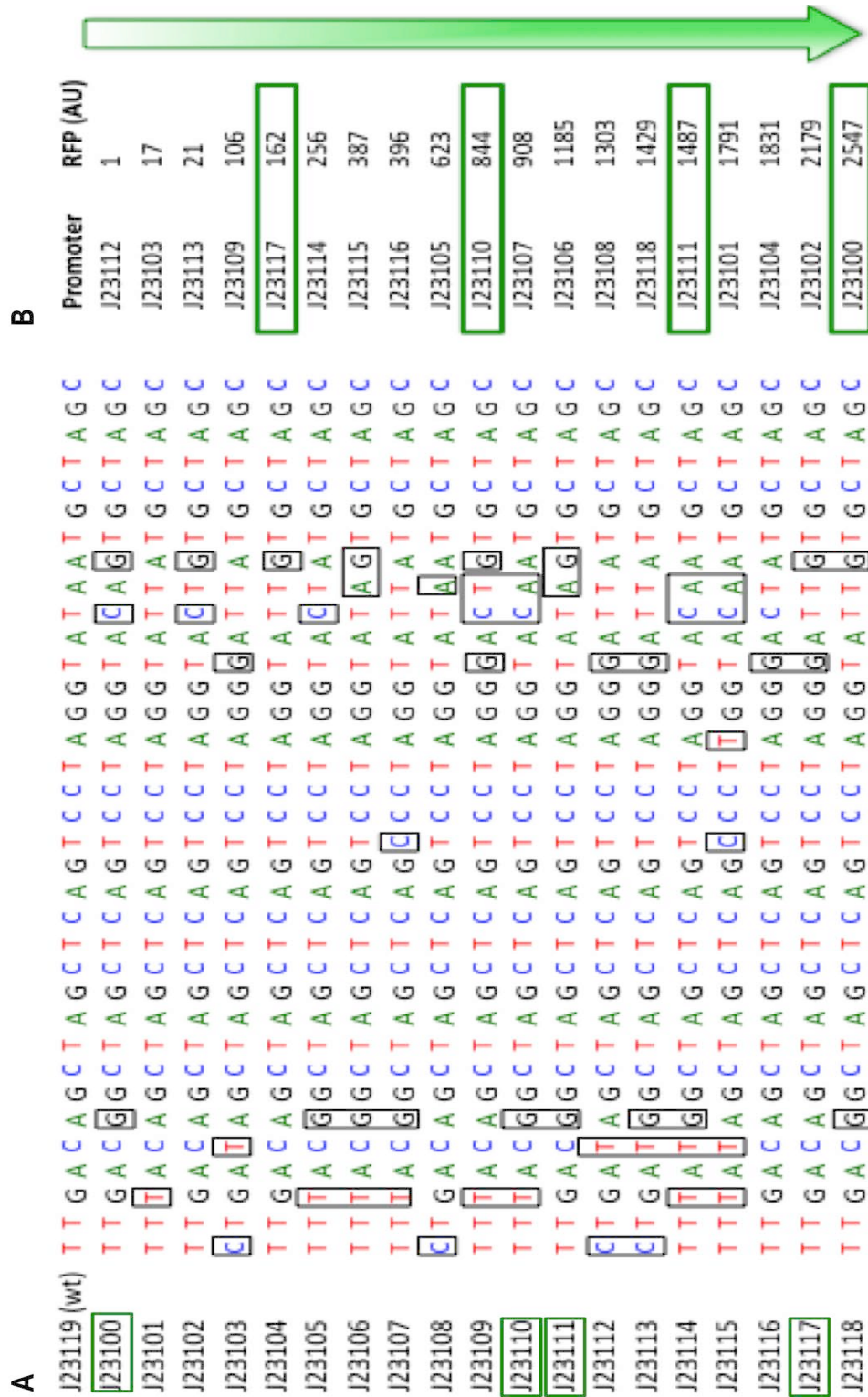


Figure 3.8 The J23119 series family. A) Sequences of the different constitutive promoters. B) Reported activities of the promoters are given as the red fluorescence protein (RFP) relative fluorescence (in arbitrary units, AU), The strength of the promoters goes from the weakest to the strongest, as the orientation of the green arrow on right hand side of the figure. The four promoters used in this study (J23100, J23110, J23111 and J23117) are marked with green boxes.

The strength of the promoters is calculated as the reported activities of red fluorescence protein (RFP), being the J23117 the reference promoter, that is, with a given relative transcription efficiency of 1 (in arbitrary units). The promoters J23110, J23111 and J23100 are 5.2, 9.2 and 15.7 fold strongest than J23117, respectively. Each promoter was synthesized by oligonucleotide hybridization including two *Bsa*I sites with 2 different overhangs at both, 5' and 3' terminus.

Table 3.4 Primers used for the synthesis of the promoters J23100, J23110, J23111 and J23117

Name	Sequence 5'-3'
J23100_FW	TTGGTCTCT <u>CGCT</u> ¹ TTGACGGCTAGCTCAGTCCTAGGTACAGTGCT <u>TAGC</u> ² TGAGAC CTT
J23100_REW	AAGGTCTCAGCTA ² GCACTGTACCTAGGACTGAGCTAGCCGTCAA <u>AGCG</u> ¹ AGAGA CCAA
J23110_FW	TTGGTCTCT <u>CGCT</u> ¹ TTTACGGCTAGCTCAGTCCTAGGTACAATGCT <u>TAGC</u> ² TGAGACC TT
J23110_REW	AAGGTCTCAGCTA ² GCATTGTACCTAGGACTGAGCTAGCCGTAAA <u>AGCG</u> ¹ AGAGA CCAA
J23111_FW	TTGGTCTCT <u>CGCT</u> ¹ TTGACGGCTAGCTCAGTCCTAGGTATAGTGCT <u>TAGC</u> ² TGAGAC CTT
J23111_REW	AAGGTCTCAGCTA ² GCACTATACCTAGGACTGAGCTAGCCGTCAA <u>AGCG</u> ¹ AGAGA CCAA
J23117_FW	TTGGTCTCT <u>CGCT</u> ¹ TTGACAGCTAGCTCAGTCCTAGGGATTGTGCT <u>TAGC</u> ² TGAGAC CTT
J23117_REW	AAGGTCTCAGCTA ² GCACAATCCCTAGGACTGAGCTAGCTGTCA <u>AGAG</u> ¹ CAGAGA CCAA

Legend

***Bsa*I** enzyme

^{1,2} Two different bp overhang for the *Bsa*I enzyme

Finally, all the genes obtained by PCR and all the different clonations along this thesis work were sequenced through DNA Sequencing Facility of the Genomic and Bioinformatics Service (SGB) of the UAB, in order to be sure that the sequences were correct. The Sanger method was applied for sequencing, using the pairs of primers presented in Table 3.5. The “Seq” primers were used to sequence all the different constructions based on the pQE-40 vector. Moreover, the number in the primer name refers to the starting position recognized by the primer on the pQE-40 plasmid. The VF2 and VR primers were used to sequence the golden gate constructions to ensure the correct cloning of the interest genes.

Table 3.5 Primers designed for the sequencing. T_m, melting temperature in °C, % GC, guanine and cytosine content

Name	Sequence 5'-3'	T _m (°C)	% GC	Size (pb)
Seq_336	GGTCGATTCTTCTCAGGAATG	60	47	21
Seq_771	GAGCTTGGACTCCTGTTGAT	61	50	20
Seq_1352	CAGCCAATCCCTGGGTGAGT	60	66	20
Seq_2075	GCGAGCGGTATCAGCTCACT	60	66	20
Seq_2762	GCAGCAGATTACGCGCAGAA	65	55	20
Seq_3540	GAATAGTGTATGCGGCGACC	62	55	20
Seq_4012	CGAGGCCCTTTCGTCTTCAC	60	65	20
Seq_rew 1011	GCTGAACGGTCTGGTTATAG	59	50	20
Seq_rew 1098	CATTCTTGCCCGCCTGATG	63	57	19
VF2	TGCCACCTGACGTCTAAGAA	62	50	20
VR	ATTACCGCCTTTGAGTGAGC	62	50	20

3.3.4 Golden Gate technique

The Golden Gate method offers standardized, quasi-scarless, multi-part DNA assembly, and is an excellent choice for combinatorial library construction. In conventional cloning, restriction enzymes bind to and cut at the same exact spot. Consequently, one conventional restriction enzyme only produces one type of sticky ends. That is the reason why in conventional cloning, only two DNA parts can be assembled in one step.

The Golden Gate method, instead, relies upon the use of type IIs endonucleases, whose recognition sites are outside of their recognition sequence. Thus, these enzymes are capable of producing multiple sticky ends at different DNA fragments in one reaction. There are several different type IIs endonucleases to choose from but in the Golden Gate method only a single type IIs endonuclease is used at the same time.

In this study the *BsaI* enzyme has been used (Table 3.1). The *BsaI* recognition sequence, GGTCTC, is separated from its four bp overhang by a single bp, and the activity of the enzyme is independent of this five base pairs. Importantly, binding sites of type IIs restriction enzymes are not palindromic and therefore are oriented towards the cutting site.

The principle of the cloning strategy is based on the ability of, in this case, the *BsaI* to cut distal from its recognition site. Two DNA ends can be designed to be flanked by a type IIs restriction, so the digestion of the fragments removes the enzyme recognition sites and generate ends with complementary 4 nt overhangs (Figure 3.9).

Such ends can be ligated seamlessly; creating a junction lacks the original site. This property allows cloning to be performed using one-step restriction-ligation.

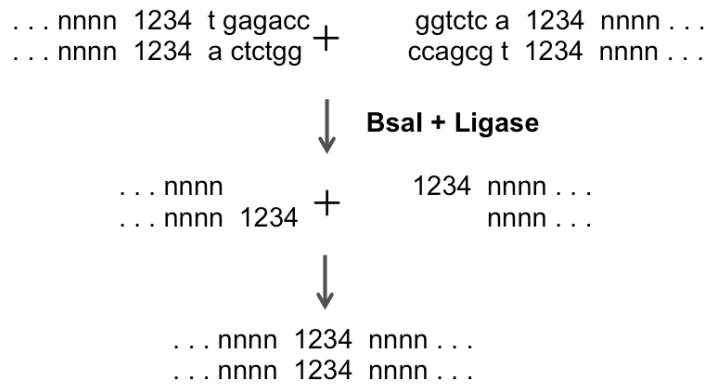


Figure 3.9 Principle cloning strategy of the Golden Gate

This method has multiple advantages:

- The recognition site of the enzyme is independent from the gene's sequence of interest and will be eliminated after subcloning
- The *Bsal* sites can be designed to have different cleavage site sequence in order to have directional cloning and preventing the relegation of empty vector. 256 different overhangs are possible for *Bsal* enzyme
- The overhang can be designed to be non-palindromic making the cloning very efficient since each DNA fragment become unable to ligate to another copy of the same molecule
- Restriction and ligation reactions can be performed together at the same time

Golden Gate Cloning is typically performed as an all-in-one-pot reaction, as said before. This means that all DNA parts, the type II restriction enzyme and a ligase are mixed in a PCR tube and put into a thermocycler. By cycling back and forth 25 to 50 times between 37°C and 16°C, the DNA parts get digested and ligated over and over again. A typical Golden Gate reaction is presented in Table 3.6.

Table 3.6 Golden Gate reaction

Assembly pieces	Quantity
pSB1C3_GG (Backbone)	60 fmol
Assembly pieces	60 fmol
Ligase Buffer	2.0 μ L
Bsal (10 U/L)	0.5 μ L
T4 ligase (5 U/L)	0.5 μ L
ddH ₂ O	Up to 15 μ L

The destination vector of the Golden Brick constructs is the pSB1C3_GG plasmid, designed with two *Bsal* sites with two different overhangs (Section 3.2.4).

The other assembly pieces of the Golden Gate reaction, with the correspondent *Bsal* restriction sites, are the *lacI* and *glyA* genes and one of the four constitutive promoters (Figure 3.5). For convenience, the starting BioBrick vector pSB1C3_GG, the promoters and the PCR products were diluted to a final concentration of 60 fmol· μ L⁻¹.

The assembly reaction was performed in a thermocycler as follows:

Table 3.7 Thermocycler reaction of the Golden Gate

Time (min)	Temperature (°C)	
3	37	25-50 cycles
4	16	
5	50	
5	80	

Bsal cuts optimally at 50°C, but as this temperature inactivate the T4 ligase, it was compromise at 37°C for 3 min to allow the restriction enzyme to cut. Then the

temperature is reduced to 16°C in order to let the T4 ligase to stick the compatible ends, produce by the enzyme, together. The pSB1C3 vector does not contain any other Bsal restriction sites; so that it is possible to repeat the cycle to ensure that any assembly pieces that have not been cut the first time has other chances. This makes the assembly reaction very efficient. The number of cycles depends on the number of different fragments to put together. Final digestion at 50°C linearized any remain plasmid and the inactivation step at 80°C prevents re-ligation.

Lastly, 2 µL of the mixture was transformed into *E. coli* DH5α competent cells and plated on LB medium containing 30 mg·L⁻¹ of chloramphenicol. The plates were incubated at 37°C overnight, and then screened for desired mutants (Figure 3.10). Only the white colonies present the acceptor vector with cloned inside the BioBrick construct. The pink colonies represent the plasmids self-ligated without the golden gate construct cloned where the *mCherry* gene can be expressed (Figure 3.10).

To demonstrate the correct assembly of BioBrick parts into the new expression vectors, colony PCR was performed for at least 6 white transformants for each construction. Finally sequencing was performed for positive clones.

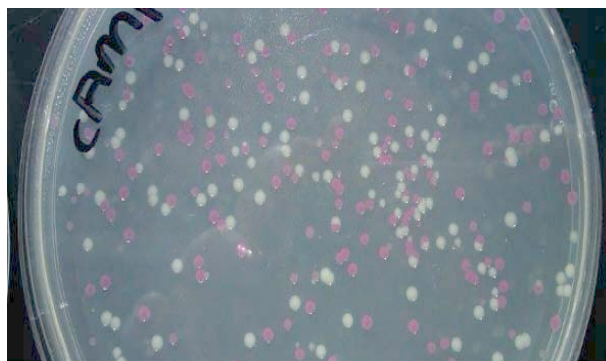


Figure 3.10 Petri dish with the *E. Coli* DH5α transformed with the Golden Gate reaction. White colonies represent the positive clones, while the pink colonies represent the plasmids self-ligated without the construct cloned inside.

3.4 Plasmid constructs

3.4.1 pQE-FucA

The commercial vector pQE-40 (Qiagen) was used as reference expression vector for the protein of interest, FucA. The *fucA* gene was amplified from the previously available pTrc*fuc* vector (Garcia-Junceda et al. 1995) using the FucA FW and REW primers (Table 3.2) (Figure 3.11). Thereafter, the 0.65 Kb PCR fragment was digested with the restriction enzymes *Bam*HI and *Hind*III (Table 3.1) and subsequently cloned into the linearized pQE-40, yielding pQE-FucA (Figure 3.12). The final product was then transformed into *E. coli* M15 [pREP4] cells, obtaining the *E. coli* M15[pREP4] pQE-FucA strain.

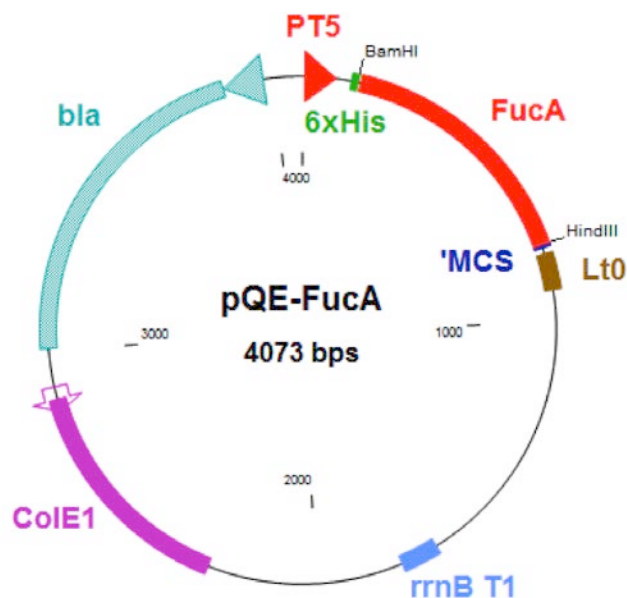


Figure 3.11 Scheme of pQE-FucA plasmid obtained by the ligation of pQE-40 vector and *fucA* PCR double digested with *Bam*HI and *Hind*III. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; rrnBT1, transcriptional termination region; ColE1, replication origin; *bla*, ampicillin resistance gene.

3.4.2 pQE $\alpha\beta$ FucA

The DNA fragment comprising the *glyA* gene was amplified from pQE $\alpha\beta$ Rham (designated as $\alpha\beta$ Terminator) (Vidal et al. 2008), including the promoter and the 3' termination region, using the primers $\alpha\beta$ T FW and $\alpha\beta$ T REW (Table 3.2). The resulting 1,784 kbp fragment was then digested with *Bsp*EI and *Xba*I (Table 3.1) and subsequently ligated into pQE-FucA to obtain pQE $\alpha\beta$ FucA (Figure 3.12). The ligation reaction was transformed into *E. coli* M15 Δ *glyA*[pREP4] yielding M15 Δ *glyA*[pREP4 pQE $\alpha\beta$ Rham].

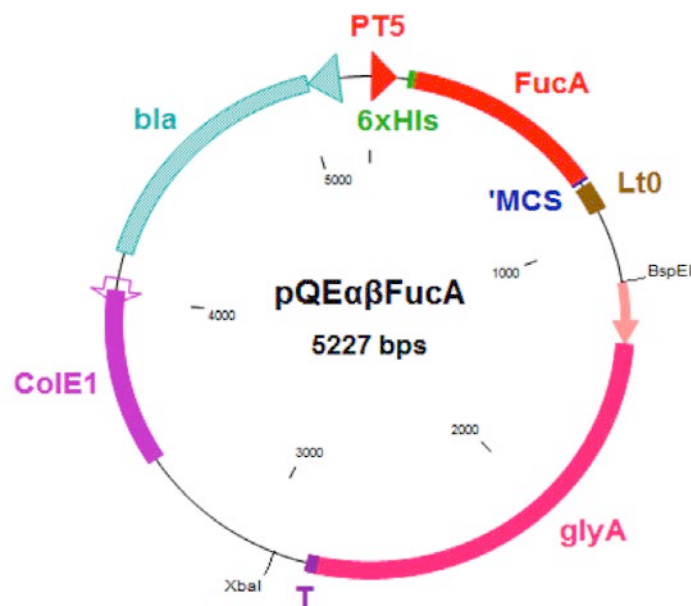


Figure 3.12 Scheme of pQE $\alpha\beta$ FucA plasmid, obtained by the ligation of the pQE-FucA plasmid and the $\alpha\beta$ T PCR-fragment, double digested with the *Bsp*EI and *Xba*I enzymes. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; *glyA*, *glyA* gene; T, terminator sequence of the *glyA* gene; ColE1, replication origin; *bla*, ampicillin resistance gene.

3.4.3 pUC-*lacI*

The *lacI* gene from the pREP4 was amplified and cloned into the pQE $\alpha\beta$ FucA plasmid. In order to do that the *lacI* fragment was amplified using the *LacI* FW and REW primers (Table 3.2) digested with *SmaI* restriction enzyme (Table 3.1) and cloned into the pUC-19 plasmid obtaining pUC-*lacI* (Figure 3.13).

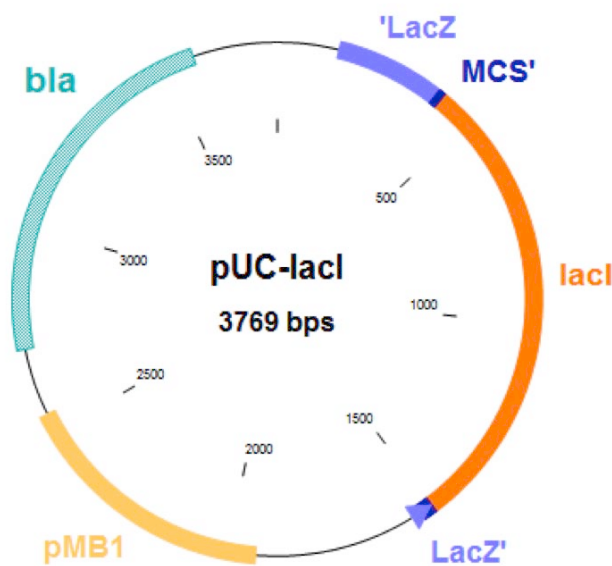


Figure 3.13 pUC-*lacI* vector, obtained by the ligation of the *lacI* gene, amplified from pREP4, together with the pUC-19 plasmid linearized with *SmaI* restriction enzyme. LacZ, β -galactosidase gene; MCS, multi cloning site; *lacI*, *lacI* gene; pMB1, replication origin; *bla*, ampicillin resistance gene

3.4.4 pQE-*lacI*- $\alpha\beta$ FucA

pUC-*lacI* vector is then transformed into DH5 α competent cells and selected in LB-agar plates with ampicillin. Besides, because this cloning vector contain the multiple cloning site at the *lacZ'* region, the positive recombinant plasmids can be verified via blue/white colony screening using LB agar plates containing X-gal and IPTG.

From the miniprep of the positive white transformants the *lacI* fragment is then double digested with *Th111I* and *SapI* and cloned into the pQE $\alpha\beta$ FucA vector yielding pQE-*lacI*- $\alpha\beta$ FucA (Figure 3.14).

E. coli M15 Δ *glyA* competent cells were transformed with the ligation product, and the clones were selected by growing on LB plates containing ampicillin. Positive clones were validated as described previously.

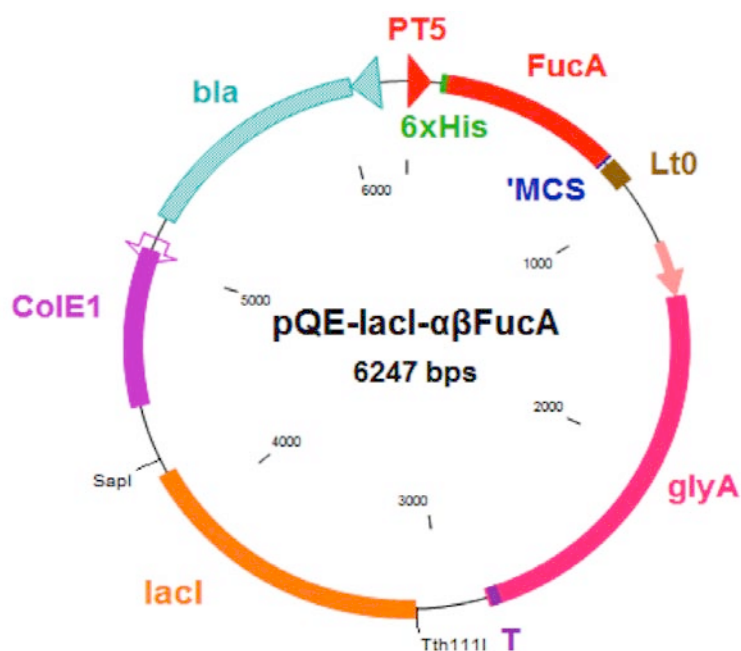


Figure 3.14 pQE-*lacI*- $\alpha\beta$ FucA plasmid, obtained by the cloning of the *lacI* gene from pUC-*lacI* into pQE $\alpha\beta$ FucA, both double digested with *Tth111I* and *SapI* restriction enzymes. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; *glyA*, *glyA* gene; T, terminator sequence; *lacI*, *lacI* gene; CoIE1, replication origin; *bla*, ampicillin resistance gene

3.4.5 BioBrick-based vectors

The BioBrick vectors were assembled using the golden gate technique (Engler, Kandzia, and Marillonnet 2008) (Section 3.3.4). In this study *BsaI* restriction enzyme and four nucleotide overhangs is used (Table 3.1).

3.4.5.1 BioBricks constructs: pSB-J231XX-S_ *lacI*

The *lacI* gene was amplified from the pREP4 plasmid with the GG_ *lacI* FW and GG_ *lacI*-S REW primers and a *de novo* strong RBS sequence (BBa_B0034, http://parts.igem.org/Part:BBa_B0034) and a terminator sequence (BBa_B1002; http://parts.igem.org/Part:BBa_B1002) were introduced (Indicated in blue in Table 3.3). The four promoters used in this study are J23100, J23111, J23110 and J23117, as said previously (Figure 3.8; Table 3.4). The promoter, *lacI* and *glyA* modules are represented in Figure 3.16. *BsaI* sites with two overhangs were introduced at the 5' and 3' prime in all assembled fragments to provide directional cloning and to prevent the religation of empty vector. Four BioBrick constructs were assembled, each one with a different constitutive promoter to tune the expression levels of *lacI* (Figure 3.15A).

All reaction DNA fragments were prepared equimolar to a concentration of 69 fmol· μL^{-1} . To each reaction, 0.5 μL of *BsaI* and T4 ligase were added. Final reactions were incubated in a thermocycler as follows: 25-30 cycles (37 °C, 3 min; 16 °C, 4 min) and final step 50 °C, 5 min and 80 °C, 5 min. Thus, reactions were performed in one-step restriction-ligation.

Thus, the golden gate reactions were performed and 2-5 μL of assembly BioBrick

constructs were transformed into 50 μ L of *E. coli* DH5 α competent cells and selected on LB plates containing chloramphenicol. The four vectors were named pSB1C3-J231XX-S_*lacI*, where the double X represents the last two digits of the promoter name (Figure 3.15B). Once obtained the four expression vectors, they were transformed together with the pQE-FucA plasmid into the M15 Δ *glyA*. The resulting final strains were named M15 Δ *glyA* pQE-FucA + pSB-J231XX-S_*lacI* (shortly J23100-S_*lacI*, J23111-S_*lacI*, J23110-S_*lacI* and J23117-S_*lacI* strains).

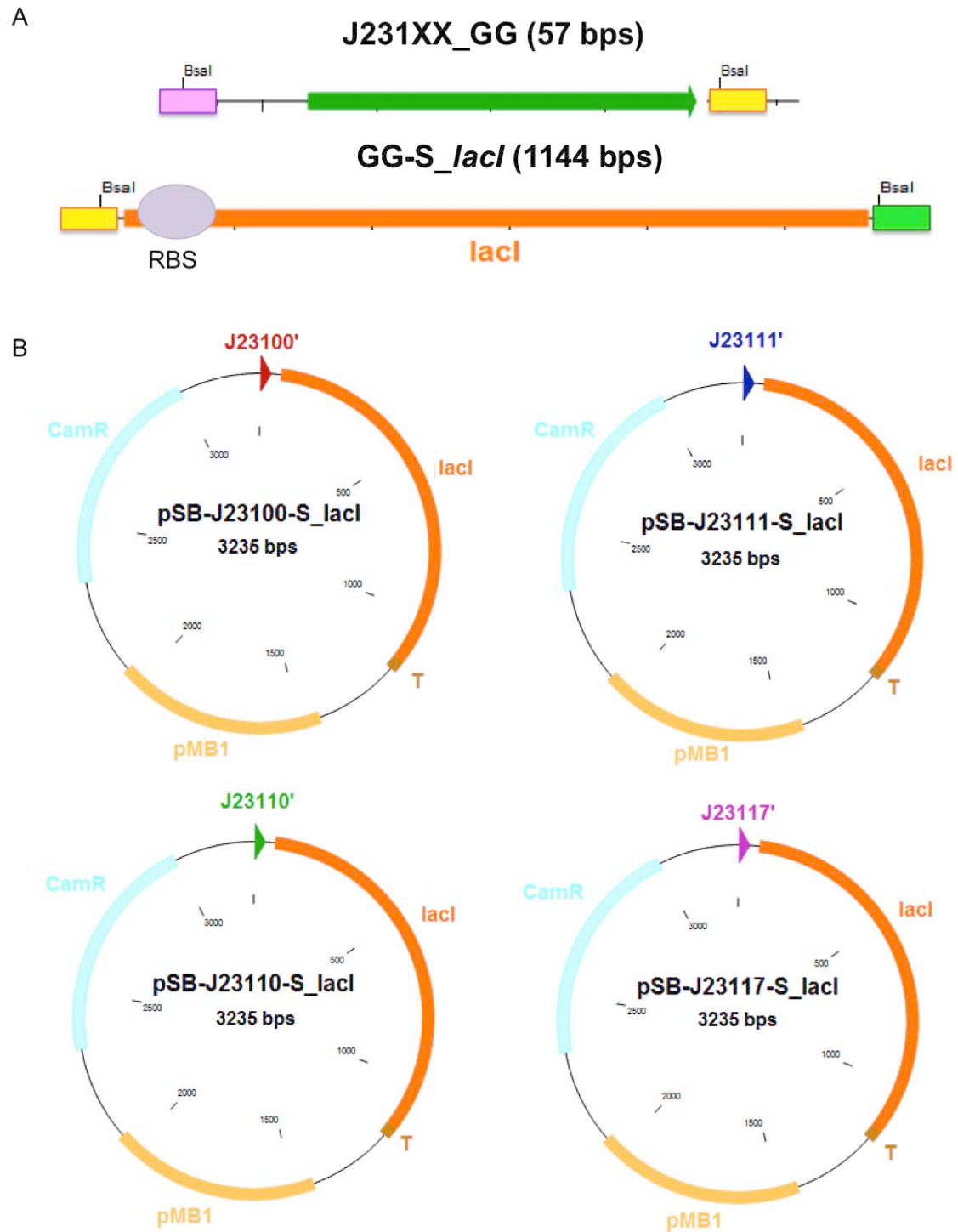


Figure 3.15 Golden gate cloning for the *lacI* tuning. **A**) PCR products of the promoters (J231XX_GG) and *lacI* gene (GG-S_ *lacI*). The different overhangs of the *Bsal* at the 5' and 3' terminus are represented as boxes with different colors, showing the complementary part for the directional cloning. **B**) Representation of the four pSB-J231XX vectors for the *lacI* gene, each one with one of the four constitutive promoters: J23100, J23111, J23110 and J23117. *lacI*, *lacI* gene; T, terminator sequence; pMB1, replication origin and *CamR*, kanamycin resistance gene.

3.4.5.2 BioBricks constructs: pSB-J231XX-S_ glyA

The DNA part containing the *glyA* gene was obtained from pQE α Rham using GG_ glyA-S FW and GG_ glyA REW primers comprised with a strong RBS sequence (BBa_J61100, http://parts.igem.org/Part:BBa_J61100), (Indicated in green in Table 3.3). The terminator sequence was maintained from the *glyA* native region, which was PCR-amplified from chromosomal *E. coli* K-12 (Vidal et al. 2008). As for the *lacI* tuning J23100, J23111, J23110 and J23117 promoters were used to create the pSB-J231XX-S_ glyA library (Figure 3.16).

Thus, the golden gate reactions were performed and again 2-5 μ L of assembly BioBrick constructs were transformed into 50 μ L of *E. coli* DH5 α competent cells and selected on LB plates containing chloramphenicol, obtaining the pSB-J231XX-S_ glyA constructs, where the double X represents the last two digits of the promoter name (Figure 3.16B).

Finally, the four expression vectors were transformed together with the pQE-FucA plasmid into the M15 Δ glyA competent cells obtaining the 4 final strains M15 Δ glyA pQE-FucA + pSB-J231XX-S_ glyA (shortly J23100-S_ glyA, J23111-S_ glyA, J23110-S_ glyA and J23117-S_ glyA strains).

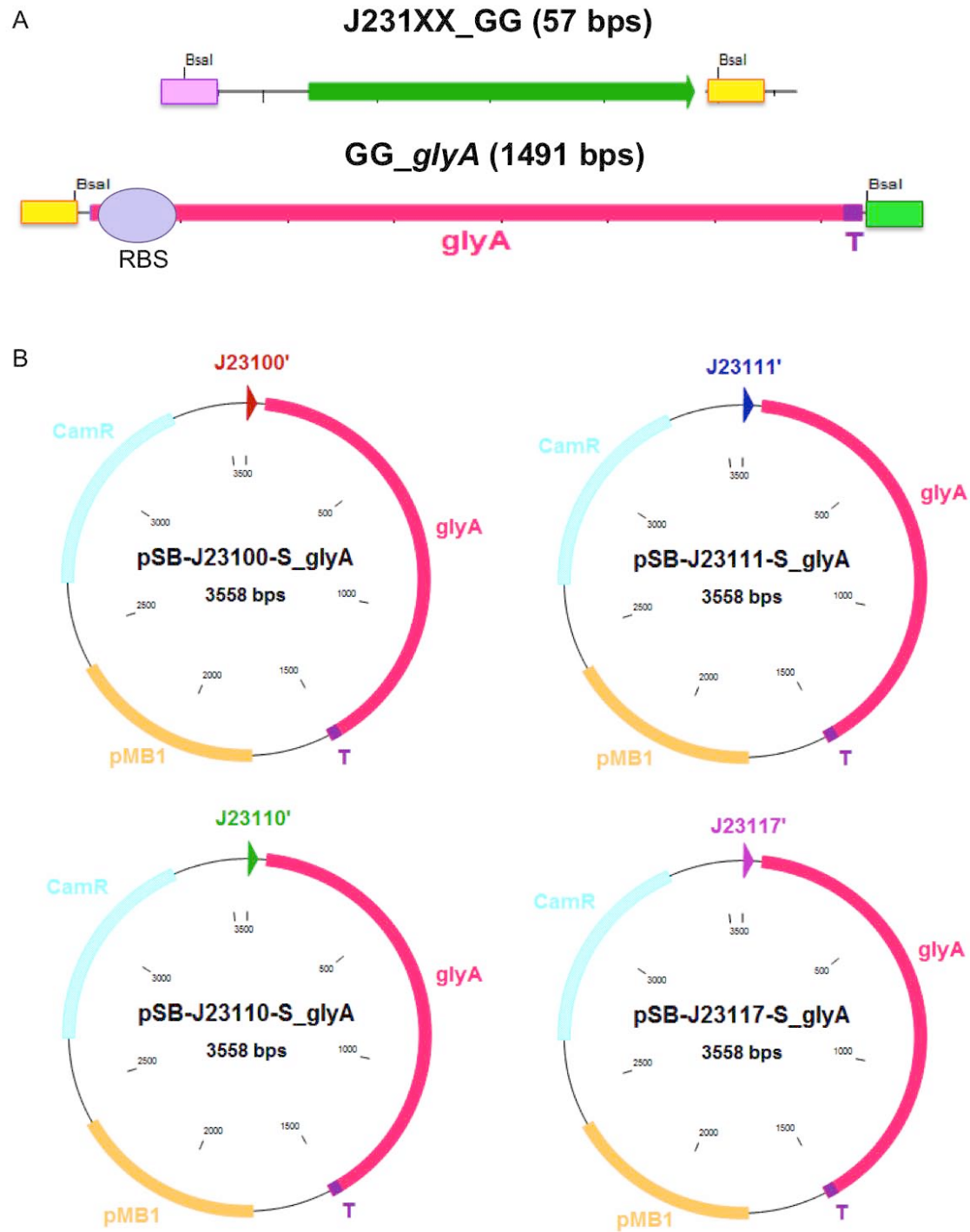


Figure 3.16 Golden gate cloning for the *glyA* tuning. **A**) PCR products of the promoters (J231XX_GG) and *glyA* gene (GG-S_*glyA*). The different overhangs of the *BsaI* at the 5' and 3' terminus are represented as boxes with different colors, showing the complementary part for the directional cloning. **B**) Representation of the four pSB-J231XX vectors for the *glyA* gene, each one with one of the four constitutive promoters: J23100, J23111, J23110 and J23117. *glyA*, *glyA* gene; T, terminator sequence; pMB1, replication origin and *CamR*, kanamycin resistance gene.

3.4.5.3 pQE-FucA-P3_lacI-P4_glyA

The construction of the single expression vector presents the *lacI* and *glyA* genes cloned downstream the J23110 (P3) and J23117 (P4) constitutive promoters, respectively.

The *lacI* gene was amplified from J23110-S_ *lacI* plasmid with the PJ1 FW and PJS REW primers (Table 3.3) and then double digest with the *Bp10I* and *Kpn2I* restriction enzymes (Table 3.1).

The *glyA* gene was amplified from the J23117-S_ *glyA* plasmid with the PJS FW and PJ REW primers (Table 3.3) and double digest with *Kpn2I* and *MscI* restriction enzymes (Table 3.1).

The destination vector pQE-FucA was digested with restriction enzymes *Bpu10I* and *MscI*. After agarose gel running and purification the two fragments together with the destination vector were ligated together obtaining the pQE-FucA-P3_ *lacI*-P4_ *glyA* vector (Figure 3.17). Finally, the ligation product was transformed into the M15 Δ *glyA* competent cells obtaining the M15 Δ *glyA* pQE-FucA_S-puzzle.

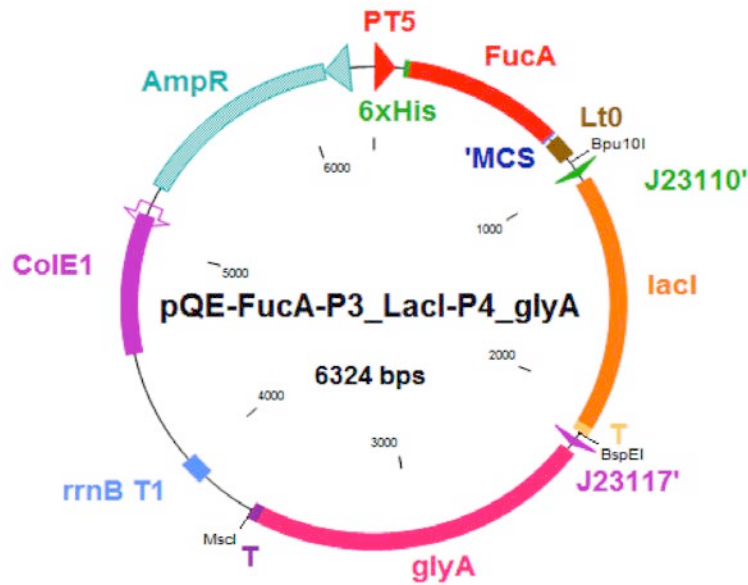


Figure 3.17 Final ligation step of the double digested PCR-amplified fragment into the pQE-FucA, obtaining the pQE-FucA-P3_ *lacI*-P4_ *glyA* vector. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; J23110, constitutive promoter; *lacI*, *lacI* gene; J23117, constitutive promoter; *glyA*, *glyA* gene *rrnBT1*, transcriptional termination region; ColE1, replication origin; *bla*, ampicillin resistance gene.

3.4.5.4 BioBricks constructs: pSB1C3-J231XX-*lacI*-*glyA*)

Both the *lacI* and *glyA* genes were PCR-amplified and were made compatible for the construction of the BioBrick vectors. The *lacI* gene was amplified from the pREP4 plasmid with the GG_ *lacI*_FW and GG_ *lacI*_REW primers and a *de novo* strong RBS sequence (BBa_B0034, http://parts.igem.org/Part:BBa_B0034) was introduced (Indicated in red in Table 3.3). The DNA part containing the *glyA* gene was obtained from pQEαβRham using GG_ *glyA*_FW and GG_ *glyA*_REW primers comprised with a strong RBS sequence (BBa_J61100, http://parts.igem.org/Part:BBa_J61100), being this RBS sequence different to the *lacI* gene in order to avoid homologous recombination (Indicated in green in Table 3.3). The terminator sequence was maintained from the

glyA native region, which was PCR-amplified from chromosomal *E. coli* K-12 (Vidal et al. 2008). The four promoters used in this study are J23100, J23111, J23110 and J23117, as said previously (Figure 3.5; Table 3.4). The promoter, *lacI* and *glyA* modules are represented in Figure 3.18. *Bsal* sites with two overhangs were introduced at the 5' and 3' prime in all assembled fragments to provide directional cloning and to prevent the religation of empty vector.

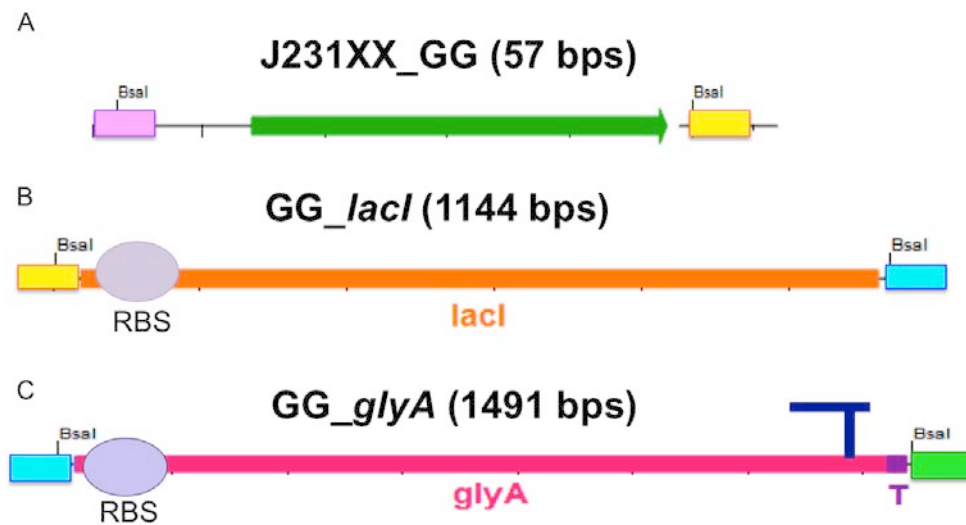


Figure 3.18 PCR products for the golden gate cloning. A) Promoters (J231XX); B) *lacI* gene and C) *glyA* gene. The different overhangs of the *Bsal* at the 5' and 3' terminus are represented as boxes with different colors, showing the complementary part for the directional cloning. *lacI*, *lacI* gene; *glyA*, *glyA* gene, RBS, ribosome binding site; T, terminator sequence.

All reaction DNA fragments were prepared equimolar to a concentration of $69 \text{ fmol} \cdot \mu\text{L}^{-1}$. To each reaction, $0.5 \mu\text{L}$ of *Bsal* and T4 ligase were added. Final reactions were incubated in a thermocycler as follows: 25-30 cycles ($37 \text{ }^\circ\text{C}$, 3 min; $16 \text{ }^\circ\text{C}$, 4 min) and final step $50 \text{ }^\circ\text{C}$, 5 min and $80 \text{ }^\circ\text{C}$, 5 min. Thus, reactions were performed in one-step restriction-ligation (Figure 3.19).

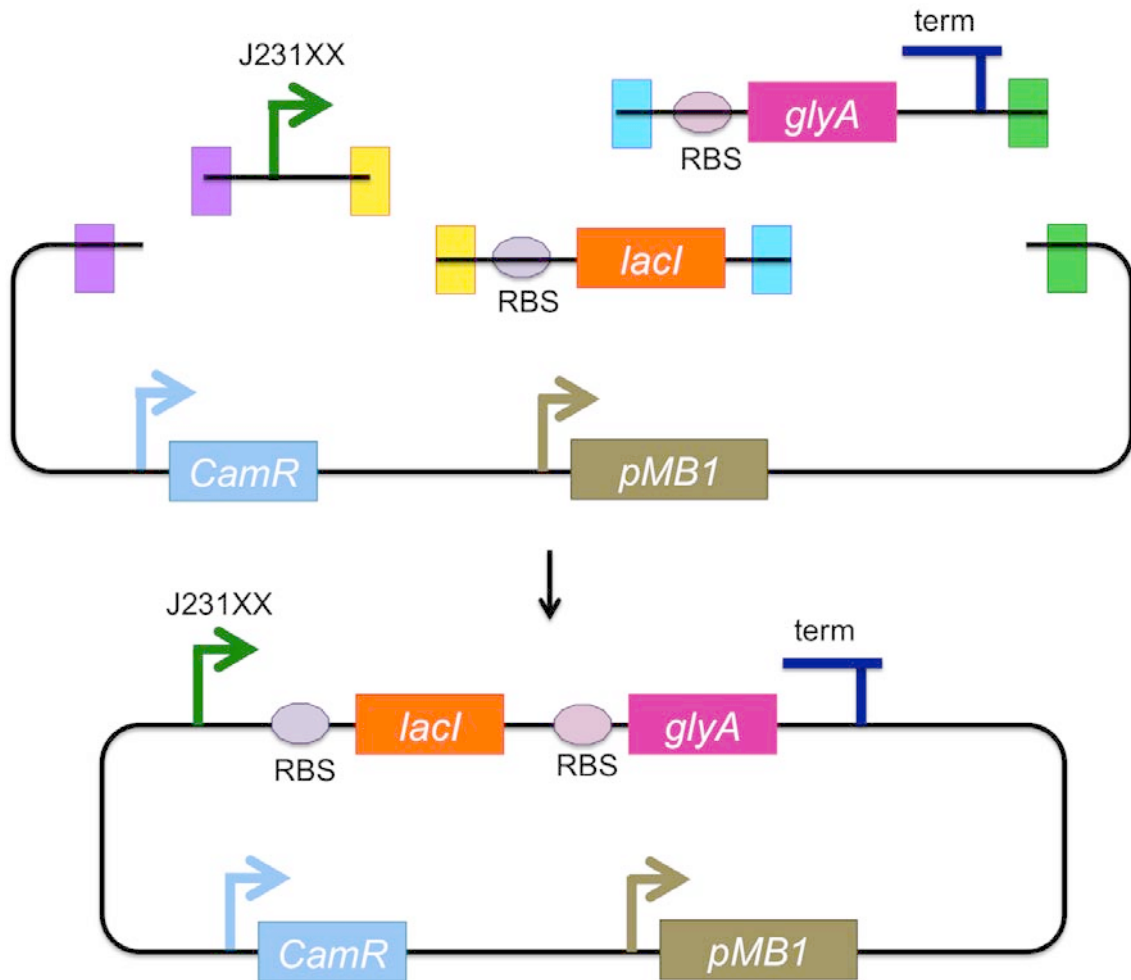


Figure 3.19 Schematic diagram of Golden Gate assembly method of the construction of the new BioBrick vectors. J231XX, constitutive promoter (where the “x” depends on the name of the promoter); *lacI*, *lacI* gene; *glyA*, *glyA* gene; RBS, ribosome binding site; *CamR*, kanamycin resistance gene; pMB1, replication origin. The different overhangs of the *BsaI* at the 5’ and 3’ terminus are represented as boxes with different colors, showing the complementary part for the directional cloning.

Four BioBrick constructs were assembled, each one with a different constitutive promoter to tune the expression levels of *lacI* and *glyA* genes. The four vectors were named pSB1C3-J231XX, where the double X represents the last two digits of the promoter name (Figure 3.20).

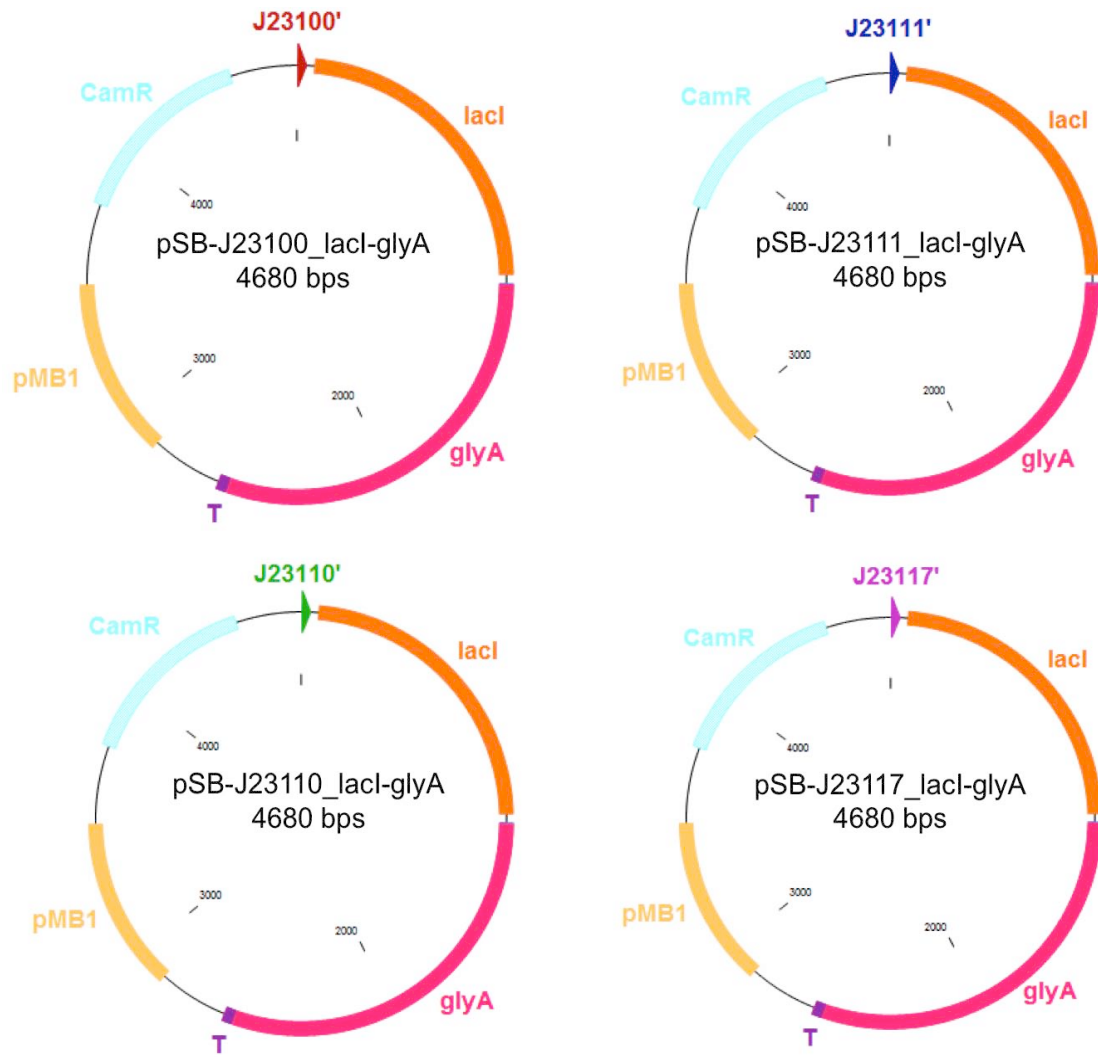


Figure 3.20 Representation of the four pSB-J231XX vectors, each one with one of the four constitutive promoters: J23100, J23111, J23110 and J23117. *lacI*, *lacI* gene; *glyA*, *glyA* gene; *T*, terminator sequence; *pMB1*, replication origin and *CamR*, kanamycin resistance gene.

Finally, 2-5 μL of assembly BioBrick constructs were transformed into 50 μL of *E. coli* DH5 α . Plasmids were selected by growing on LB plates containing chloramphenicol and isolated from pink/white colonies selection and validated as described previously.

Once obtained the four expression vectors, they were transformed together with the pQE-FucA plasmid into the M15 Δ *glyA* competent cells. The resulting final strains were named M15 Δ *glyA* pQE-FucA + pSB-J231XX (shortly M15 Δ *glyA*[C00], M15 Δ *glyA*[C11],

M15 Δ *glyA*[C10] and M15 Δ *glyA*[C17] strains).

3.4.6 pQE-FucA_puzzle (J23110)

The construction of vector derived from the pQE-FucA and the BioBrick vectors required a two-step assembly (Figure 3.21). The expression cassette J23110-*lacI-glyA* was amplified from pSB1C3-J23110-*lacI-glyA* using PJ/2_FW and PJ_REW primers (Table 3.3). Both, PCR product and destination vector pQE-FucA were digested with *Bpu10I* and *MscI* and subsequently extracted from agarose gel. The expression cassette was cloned into the pQE-FucA (double digested) obtaining the pQE-FucA_puzzle (J23110) (Figure 3.21C). Finally, the ligation product was transformed into the M15 Δ *glyA* generating M15 Δ *glyA* pQE-FucA_puzzle (J23110), from now on Puzzle strain. Transformants were isolated and correct plasmid sizes were verified.

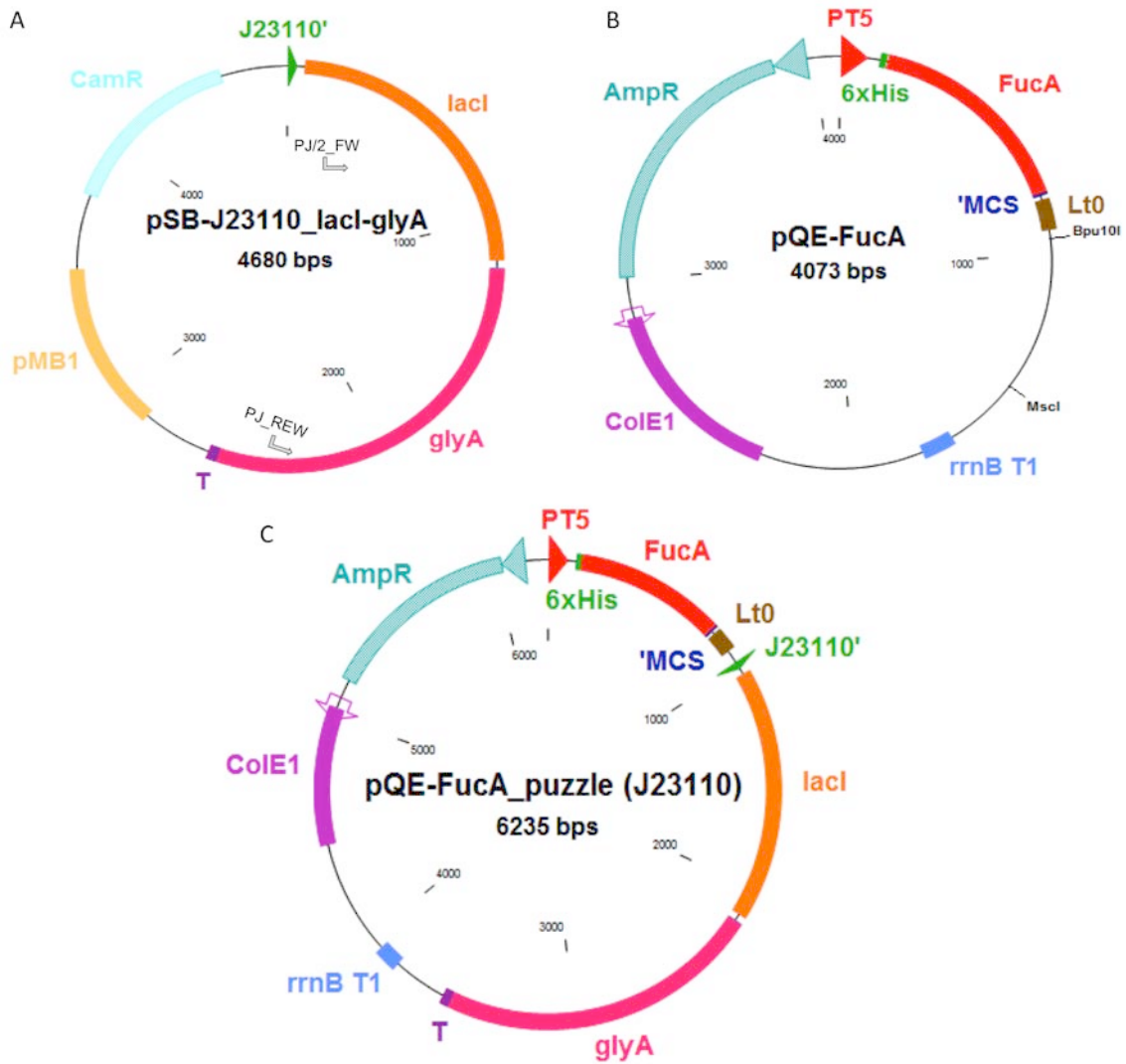


Figure 3.21 Strategy for pQE-FucA_puzzle (J23110) construction. **A)** PCR amplification *lacI-glyA* cassette using plasmid SB1C3-J23110 as template and the primers PJ/2_FW and PJ-REW. **B)** Double digestion of pQE-FucA with *Bpu10I* and *MscI*. **C)** Final ligation step of the double digested PCR-amplified fragment (in step A) into the pQE-FucA_puzzle (J23110) expression vector. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; J23110, constitutive promoter; *lacI*, *lacI* gene; *glyA*, *glyA* gene *rrnB*T1, transcriptional termination region; *ColE1*, replication origin; *bla*, ampicillin resistance gene; *pMB1*, replication origin and *CamR*, kanamycin resistance gene.

3.4.7 pQE-FucA_puzzle (J23110)_AmpR⁻

The ampicillin resistance gene (*bla*) was eliminated from the pQE-FucA_puzzle (J23110) plasmid using the Klenow fragment-based blunting DNA technique.

Precisely, the vector pQE-FucA_puzzle (J23110) was double digested with *Eco0109I* and *AhdI* in order to eliminate the *bla* gene (Figure 3.22A). The 5'-3' polymerase and the 3'-5' exonuclease activities of the DNA Polymerase I (Large) Klenow Fragment was used in order to end-removal and fill-in terminal unpaired nucleotides.

The blunting DNA reaction, composed of digested vector 0.5 mg, 1 µL of dNTPS 25 mM (Bioline) and 1 µL of DNA polymerase I (5U/µL) (NEB), was incubated at room temperature (RT) 20 min, followed by a heat inactivation step at 75 °C 10 min. The ligation reaction of the blunt ended DNA fragments and their respective plasmid backbones were carried out at 16 °C overnight using T4 ligase (Figure 3.22B).

The resulting ligation (Figure 3.22C) vector was transformed into *E. coli* M15Δ*glyA* and plated on defined medium (DM) plates, generating M15Δ*glyA* pQE-FucA_puzzle (J23110)_AmpR⁻, from now on AmpR⁻ strain.

Transformants were isolated and tested both, in DM and LB supplemented with ampicillin plates, as a positive and negative control, respectively. Selected transformants were able to grow in defined media but not in LB plates supplemented with ampicillin. Positive clones were validated as described previously.

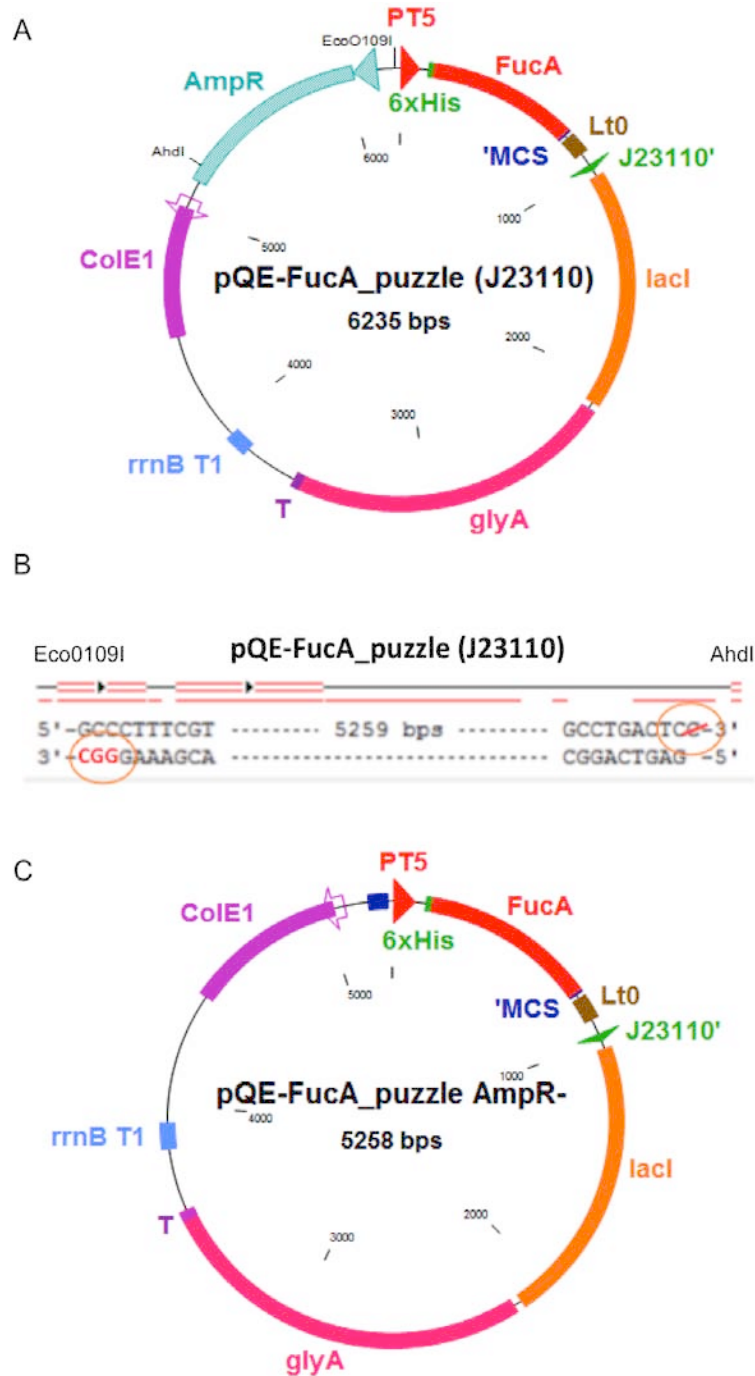


Figure 3.22 Strategy for pQE-FucA_{puzzle} AmpR^r construction. **A**) Double digestion of the pQE-FucA_{puzzle} (J23110) plasmid by *Eco0109I* and *AhdI* restriction enzymes; **B**) End removal and fill-in of terminal unpaired nucleotides by DNA polymerase I (large) Klenow fragments (NEB); **C**) Final ligation step obtaining pQE-FucA_{puzzle} AmpR^r expression vector. PT5, promoter T5; 6xHis, histidine affinity tag coding sequence; MCS, multi cloning site; Lt0, lambda t0 transcriptional termination region; J23110, constitutive promoter; *lacI*, *lacI* gene; *glyA*, *glyA* gene; *rrnBT1*, transcriptional termination region; *ColE1*, replication origin; *bla*, ampicillin resistance gene.

3.4.8 pASK-FucA

The *fucA* gene was amplified from the pQE-FucA vector using the FucA-sec FW and REW primers (Table 3.2). Thereafter, the 0.65 Kb PCR fragment was digested with the restriction enzymes *Bgl*II and *Sap*I (Table 3.1) and subsequently cloned into the linearized pASK-Pet, yielding pASK-FucA (Figure 3.23). The final product was then transformed into *E. coli* BL21* competent cells, obtaining the *E. coli* BL21* pASK-FucA strain.

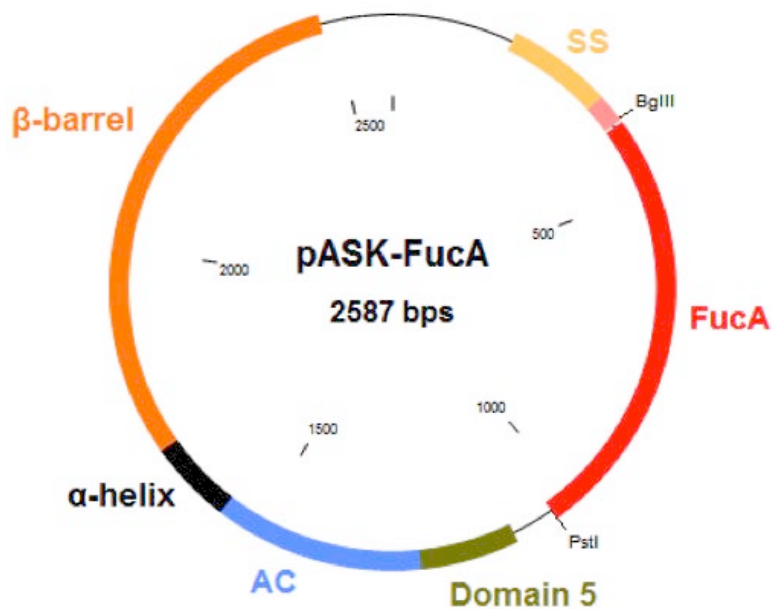


Figure 3.23 Scheme of pASK-FucA plasmid obtained by the ligation of pASK-Pet vector and *fucA* PCR double digested with *Bgl*II and *Pst*I. SS, signal sequence; FucA, *fucA* gene; AC, auto-chaperone domain; α -helix region connects the AC to the β -barrel and contains the cleavage site.

3.4.9 pET22b-AC-FucA

The *fucA* gene was amplified from the previously available pTrc*fuc* vector (Garcia-Junceda et al. 1995) using the FucA FW and REW primers (Table 3.2). Thereafter, the 0.65 Kb PCR fragment was digested with the restriction enzymes *Bam*HI and *Hind*III (Table 3.1) and subsequently cloned into the linearized pET22b-PetAC plasmid, yielding pET22b-AC-FucA (Figure 3.24). The final product was then transformed into *E. coli* BL21* competent cells, obtaining the *E. coli* BL21* pEt22b-AC-FucA strain.

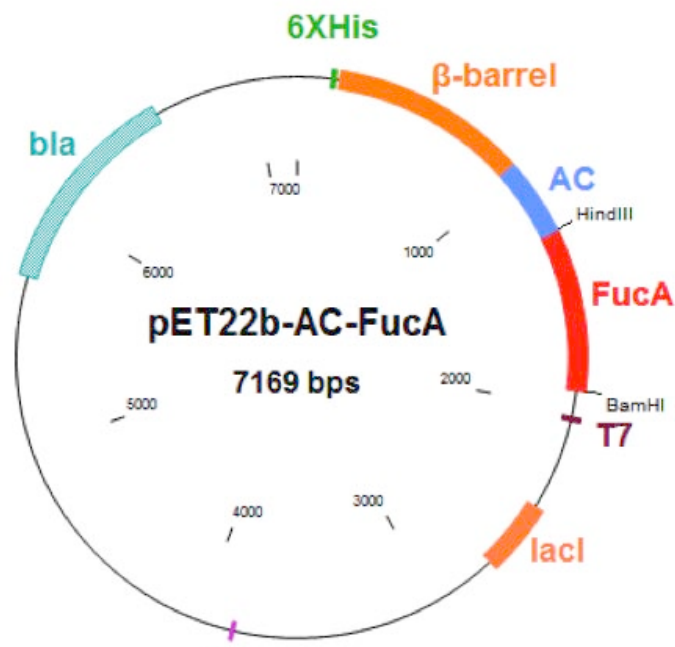


Figure 3.24 Scheme of pET22b-AC-FucA plasmid obtained by the ligation of pET22b-PetAC vector and *fucA* PCR double digested with *Bam*HI and *Hind*III. β -barrel domain; AC, auto-chaperone domain; FucA, *fucA* gene; T7 promoter; *lacI*, *lacI* gene; *bla*, ampicillin resistance gene.

3.5 Medium composition

3.5.1 Complex medium

Luria Bertani (LB) medium, containing $10 \text{ g}\cdot\text{L}^{-1}$ peptone, $5 \text{ g}\cdot\text{L}^{-1}$ yeast extract and $10 \text{ g}\cdot\text{L}^{-1}$ NaCl, was used for all pre-cultures preparations. For plate preparation, bacteriological grade agar was added to a final concentration of $15 \text{ g}\cdot\text{L}^{-1}$. Both LB and LB-agar were sterilized by autoclaving (120°C for 30 min).

3.5.2 Defined Medium

The Defined Medium (DM) is a glucose mineral medium used for growing *E. coli* (Vidal et al. 2005). All the different parts of the medium were prepared as stock solution as described below:

- Glucose $600 \text{ g}\cdot\text{L}^{-1}$
- Macro-elements (pH 7.3) (Table 3.8)

Table 3.8 Macro-elements composition

Component	Concentration ($\text{g}\cdot\text{L}^{-1}$)
K_2HPO_4	17.91
KH_2PO_4	3.59
$(\text{NH}_4)_2\text{SO}_4$	4.52
NaCl	2.76

- Trace elements solution (TES) (Table 3.9)

Table 3.9 Trace elements solution (TES)

Component	Concentration (g·L ⁻¹)
AlCl ₃ ·6H ₂ O	0.04
ZnSO ₄ ·7H ₂ O	1.74
CoCl ₂ ·6H ₂ O	0.16
CuSO ₄	1.55
H ₃ BO ₃	0.01
MnCl ₂ ·4H ₂ O	1.42
NiCl ₂ ·6H ₂ O	0.01
Na ₂ MoO ₄ ·H ₂ O	0.02

- MgSO₄·7H₂O 500 g·L⁻¹
- FeCl₃ 5 g·L⁻¹
- Thiamine 100 g·L⁻¹
- CaCl₂ 2H₂O 100 g·L⁻¹
- Phosphates (P) were not included in the feeding solution for fed-batch cultures in order to avoid co-precipitation with magnesium salts. Instead, a concentrated phosphate solution containing 500 g·L⁻¹ K₂HPO₄ and 100 g·L⁻¹ KH₂PO₄ was pulsed during the fed-batch phase to avoid their depletion when necessary (calculated yield $Y_{X/P} = 18 \text{ g DCW gP}^{-1}$)

Glucose and macro-elements were sterilized by autoclaving at 121°C for 30 min. TES, MgSO₄·4H₂O, FeCl₃, vitamin and CaCl₂·2H₂O were sterilized separately by filtration (0.2 mm syringe filter made from a blend of cellulose esters, Sartorius).

3.5.2.1 DM media for Shake flasks culture

Defined minimal medium composition, used for studies in Shake Flasks and for the preparation of inoculum for the fermenter, was prepared from sterile concentrated stock solutions in the proportion indicated below (Table 3.10):

Table 3.10 DM media for Shake Flasks culture per 100 mL

Component	Concentration (g·L ⁻¹)
Glucose	5.00
K ₂ HPO ₄	2.98
KH ₂ PO ₄	0.60
(NH ₄) ₂ SO ₄	0.75
NaCl	0.46
MgSO ₄ ·7H ₂ O	0.112
FeCl ₃	0.006
Thyamine	0.025
CaCl ₂ 2H ₂ O	1.42
TES	0.8 mL·L ⁻¹

3.5.2.2 MD media for fed-batch culture

Defined minimal medium for the bioreactor batch growth was prepared from stock solution as described in Table 3.11. After batch phase feeding medium was added to achieve high cell densities. The composition of the feeding is described in Table 3.12.

Table 3.11 DM composition per liter, for bioreactor batch growth

Component	Concentration (g·L ⁻¹)
Glucose	20.00
K ₂ HPO ₄	11.9
KH ₂ PO ₄	2.4
(NH ₄) ₂ SO ₄	3.00
NaCl	1.80
MgSO ₄ ·7H ₂ O	0.45
FeCl ₃	0.02
Thyamine	0.01
CaCl ₂ 2H ₂ O	1.44
TES	2.86 mL·L ⁻¹

Table 3.12 Feeding solution composition

Component	Concentration (g·L ⁻¹)
Glucose	490.00
MgSO ₄ ·7H ₂ O	9.56
FeCl ₃	0.49
Thyamine	0.33
CaCl ₂ 2H ₂ O	0.10
TES	63.0 mL·L ⁻¹

3.5.3 Supplements

The different antibiotics and IPTG solutions were prepared separately from the rest of the media. All the solutions were sterilized by filtration and stored in aliquots at -20°C.

Below, are presented the antibiotics and the induced used with their stock concentration:

- Ampicillin $100 \text{ mg}\cdot\text{mL}^{-1}$ (Sigma)
- Kanamycin $25 \text{ mg}\cdot\text{mL}^{-1}$
- Chloramphenicol $100 \text{ mg}\cdot\text{mL}^{-1}$
- IPTG 100 mM

Antifoam (Antifoam 204, Sigma), sterilized by autoclaving at 121°C for 30 min, was added manually to the bioreactor when foaming was observed.

3.6 Cultivation conditions

Pre-inoculum. From cryo-stock stored at -80°C , strains were grown in 50 mL Falcon tubes with 10 mL of LB medium, and supplemented with the corresponding antibiotic (ampicillin and kanamycin final concentration $100 \text{ mg}\cdot\text{L}^{-1}$, chloramphenicol $30 \text{ mg}\cdot\text{L}^{-1}$). Growth was performed overnight at 37°C with agitation, and after around 16 h of incubation, cultures reached 2.0-2.5 units of optical density at 600 nm (OD_{600}).

Inoculum Shake Flasks. 3 mL of the overnight pre-inoculum were transferred to a 500 mL-volume baffled shake flasks containing 100 mL of DM, with the same growing conditions as pre-inoculum cultures.

Inoculum Bioreactor. Once the inoculum in the shake flasks reached an OD_{600} of 1.1-1.2 a quantity of culture was aseptically added into the fermentation jar with the DM batch medium. This amount depends on the bioreactor setup (see section 3.6.2).

3.6.1 Shake Flasks experiments setup

For Shake Flasks RPP, the cultivations were performed at 37°C and 150 rpm, in 100 mL of DM medium. To induce *fucA* expression, an IPTG pulse with a concentration of 1 mM was carried out. Cells were induced when reached an OD_{600} of 1.5 and the induction was maintained for 4 hours, sampling before induction (PI) and after 1, 2 and 4 hours of induction. The PI sample is needed for negative control of basal recombinant protein production.

3.6.2 Bioreactors experiments setup

Bioreactor cultivation experiments were carried out using two types of reactors: the Biostat B and the Fermac 310/60. In both cases the temperature was maintained at 37°C and the pH was kept at 7.00 ± 0.05 by adding 15 % (v/v) NH_4OH solution (NH_4^+ from base addition served also as the sole nitrogen source). Dissolved oxygen levels were kept at 60 % of the saturation concentration. Once glucose was depleted, the feeding part of the fermentation was started with the addition of feed.

When culture reached the desired biomass concentration a single-pulse of 70 μM of IPTG was added in order to induce the overexpression of the recombinant protein.

Samples were removed from the bioreactor every 10-30 min during the induction period till the end of the culture. One sample was removed prior to induction for negative control of basal recombinant protein expression.

3.6.2.1 Biostat B bioreactor

For all the experiments performed at the Universitat Autònoma de Barcelona 80 mL of the inoculum (with an OD_{600} of 1.2) were added to 720 mL of DM with the composition described in Table 3.7. The fermentations were carried out using a Biostat B bioreactor (Braun Biotech Int.) with a 2 L vessel equipped with pH, dissolved oxygen and temperature probes (Mettler Toledo, Columbus, OH, USA). The digital control unit (DCU) controlled the pH, stirring, temperature and dissolved oxygen of bioreactor. Set point was fixed at 60 % saturation and pO_2 was controlled by adapting the stirring speed between 350 and 1120 rpm and by supplying $1.5 \text{ L}\cdot\text{h}^{-1}$ of air (enriched with pure oxygen when necessary). A microburette (MB, Crison Instruments MICRO BU 2030) with a 2.5 mL syringe (Hamilton) was used in fed-batch cultures for discrete feed addition. Figure 3.25 schematizes this experimental set up, and Figure 3.26 presents a picture of the bioreactor and its components.

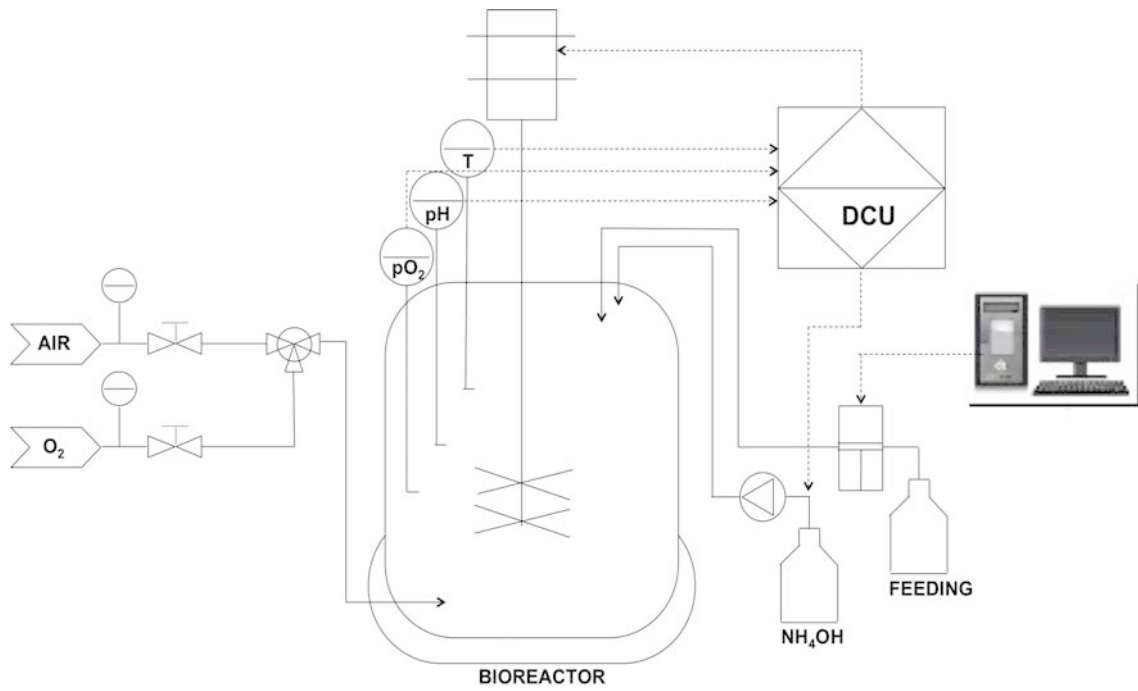


Figure 3.25 Schematic representation of the bioreactors experimental setup

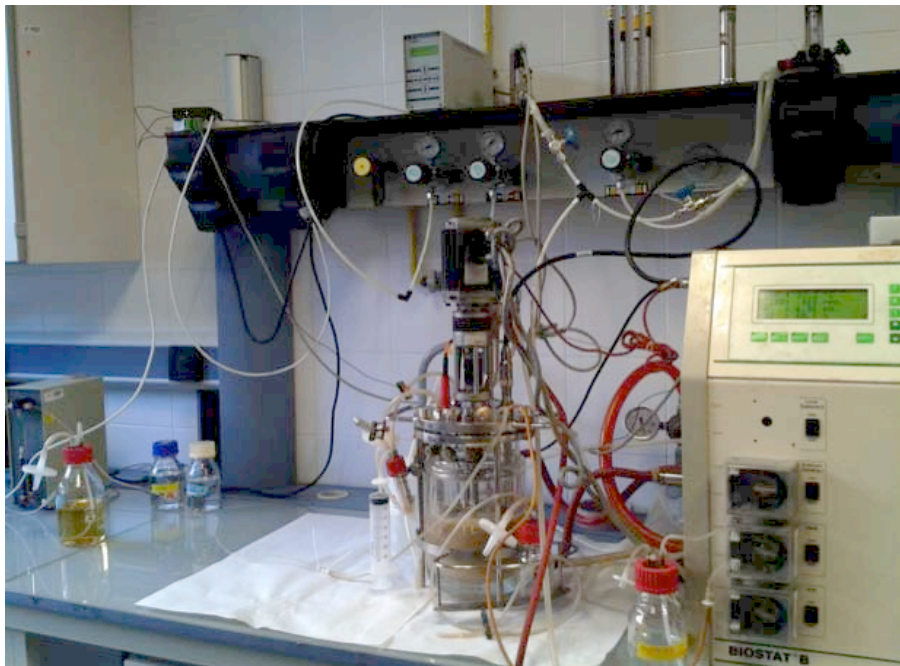


Figure 3.26 Picture of the Biostat B reactor setup

Predefined exponential feeding profile was performed in order to maintain the specific growth rate (μ_{fix}) at a fixed value (Pinsach et al. 2006). As reported in literature, the volume of feedstock at a time point can be estimated in order to maintain the μ_{fix} fixed, as shown in the equation 3.1.

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

where:

- S_f is the glucose concentration ($g \cdot L^{-1}$) in the feedstock
- m_{SX} is the cellular maintenance coefficient ($gS \cdot gDCW^{-1} \cdot h^{-1}$)
- μ_{Fix} is the fixed specific growth rate (h^{-1})
- Y_{XS} is the substrate yield ($gDCW \cdot gS^{-1}$)
- t (h) is the moment when the feeding starts
- Δt is the time interval for the next addition

3.6.2.2 Fermac 310/60 bioreactor

For the fed-batch experiments, performed at the School of Engineering at the University of Birmingham (Birmingham, UK), the bioreactor Fermac 310/60 (Electrolab) was used. The volume of the inoculum used was 200 mL and the DM volume was 1300 mL, for a total of 1500 mL. The 5 L cylindrical vessel was equipped with 4 baffles and an agitator with 2 six-bladed Rushton turbines. Aeration was achieved by sparging air with a rate of $3 L \cdot h^{-1}$ (2 vvm). Temperature and pH sensing were via an external PT100 sensor (Electrolab) and via a F-365 Fermprobe (Broadley James), respectively. pO_2 was measured with a D150 Oxyprobe (Broadley James), and it was maintained above a set

point of 60 % by increasing agitation to a maximum of 1000 rpm from a minimum of 200 rpm.

Off gas was automatically collected from compositional analysis using a PrimaDB gas-chromatograph mass spectrometer (GC-MS) (Thermo) and compared to atmospheric air in order to calculate the proportion of O₂ consumed (OXC), CO₂ evolved (CDC) and from these the respiratory quotient (RQ). This data was logged automatically by GasWorks v1.0 (Thermo). This data give information not only about the cells metabolism, but also allow monitoring the changes in the respiratory activity of the culture during the fed-batch.

3.7 Analytical methods

3.7.1 Optical density

Cell concentration was determined by optical density (OD₆₀₀) measurements at 600 nm using a spectrophotometer (Uvicon 941 Plus, Kontrol). The samples were diluted with distilled water (ddH₂O) when necessary to obtain a reading between 0.1 and 0.8, zeroing the machine with ddH₂O. OD₆₀₀ values were converted to biomass concentration expressed as Dry Cell Weight (DCW), with 1 OD₆₀₀ equivalent to 0.3 gDCW·L⁻¹ (Vidal 2006).

3.7.2 Colony Forming Unit (CFU), plasmid retention and plasmid curing

The purpose of plate counting is to determine cultivability under specific conditions and plasmid retention. The plate count is linear for *E. coli* over the range of 30-300 CFU on a standard size Petri dish. Therefore, cultures were serially diluted in sterile NaCl 0.09% solution, then plated onto non-selective plates and incubated at 37°C overnight. The plasmid retention was determined by growing the diluted culture on the non-selective and on the appropriate antibiotic selection media and determining the ratio of total cells plated that survives.

The elimination of a plasmid (curing) from a bacterial culture relies on the fact that some recombinant plasmids are stable and can be maintained through successive generations during cell division; however, differing growth and storage conditions can induce changes in the transformed population and the loss of the plasmid during replication (Trevors 1986). The efficiency of curing depends on the plasmid and on the bacterial host carrying it.

In order to cure the *E. coli* M15[pREP4] cells from the pREP4 plasmid, serial dilution have been performed in non-selective medium. As explain in materials and methods section 3.2.2, the pREP4 harboring the *kan* gene, which confers kanamycin resistance. The cure methods consists of: i) grow the cells in LB media with non-selective condition for the plasmid to cure (no kanamycin in the culture media) (Figure 3.27A); ii) dilute the cell culture twice a day; iii) after 6-8 serial dilution, plate some cells out into agar plates containing ampicillin (Figure 3.27B); iv) after growing at 37°C overnight, test positive colonies by streaking onto selective and non-selective plate for

the lost of plasmid (Figure 3.27C). The colony that can grow only in LB agar plus ampicillin but not plus kanamycin, means it has lost the pREP4 plasmid.

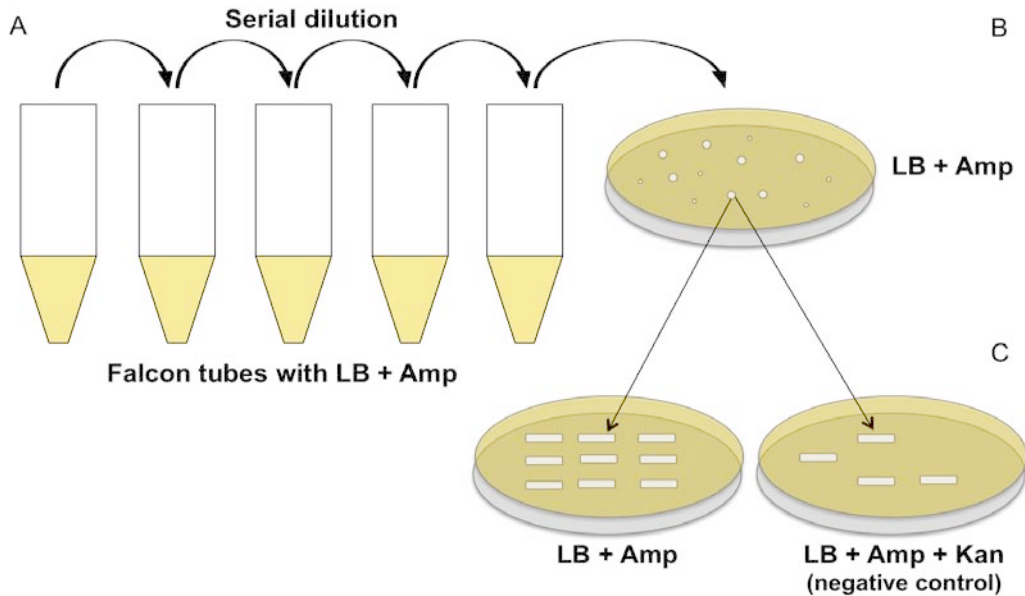


Figure 3.27 Schematic representation of the cure methods. A) Serial dilution of the *E. coli* M15[pREP4] cells in non-selective medium (LB + Amp); B) After 6-8 dilution plate the cells into non-selective agar plate (LB + Amp); C) positive and negative control in LB + Amp and LB + Amp + Kan plates, respectively.

3.7.3 Substrate and by-products concentration

For analysis of substrate and by-products concentration along the cultures, one milliliter of culture medium was separated from biomass by centrifugation at 14000 rpm for 6 min and filtered (0.45 mm membrane filter of cellulose esters, Millipore) prior to analysis. Glucose concentration was determined enzymatically on an YSI 2070 system (Yellow Spring System). Acetic and lactic acid were analyzed by HPLC (Hewlett Paackard 1050) equipped with an ICsep COREGEL 87H3 ICE-99-9861 (Transgenomic)

column and an IR detector (HP 1047), using 6 mM H₂SO₄ (pH 2.0) as mobile phase, a flow rate of 0.3 mL·min⁻¹, at 40 °C.

3.7.4 Flow cytometry

Bacteria were analysed using a BD-Accury C6 flow cytometer (BD, Oxford, UK).

The flow cytometry (FCM) is flow-based method that enables the simultaneous measurements of multiple physical and chemical/biological characteristics of single particles, which typically are cells. Originally developed by Gucker et al. (1947), modern flow cytometers are employed in cell counting, cell sorting and cell classifications. As the name suggests, the FCM system transports particles in a fluid stream, one cell at time at high speed, to a laser beam for the interrogation of particles properties. One of the main advantages of FCM is its ability to analyze individual cells. Some flow cytometers also include additional features like the cell sorter device (FACS) in order to separate two or more cell populations for further analysis.

3.7.4.1 Component and function of a flow cytometer

Briefly, a light source is directed onto a hydrodynamically focused stream of fluid to create a zone for interrogation (Figure 3.28). Once cells reach the flow cell they pass through a laser beam and the interaction is measured. One or multiple laser beams illuminate the cells in order to measure light scattering properties and to excite fluorescent molecule (Tracy, Gaida, and Papoutsakis 2010).

These fluorescent molecules can be endogenous (auto-fluorescence) or fluorescent molecules purposefully employed to tag or probe biological structures. Multiple detectors allow gathering fluorescence emitted at different wavelengths, to obtain multiparametric information from any single cell. Light signals captured by these detectors are converted into electronic signals by the cytometer's electronic system, which then filters and processes these signals using computer-based algorithms.

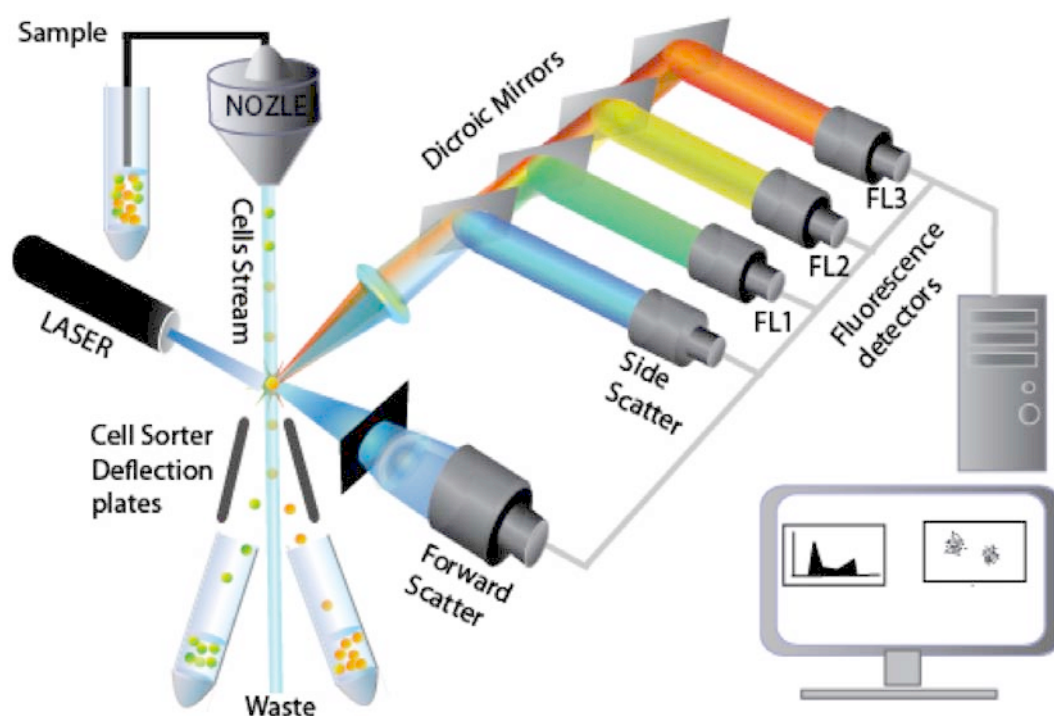


Figure 3.28 Operation of a typical flow cytometer. Sample leave nozzle into stream of sheath fluid, achieving laminar flow and aligning cells in single file before passing to the flow cell. Cells interact with a laser, Forward Scatter is detected at $< 10^\circ$ from incidental light, light at 90° passes through a series of filters and mirrors allowing detection of Side Scatter and fluorescence.

As mentioned before the light scatter is collected at two angles:

- Forward Scatter (FSC). The primary data output is the amount of light scattered by a cell. The FSC is detected at angles normally $< 10^\circ$ from that of the incidental laser light. This interaction of light can reveal important information about the cell size and shape of the cells;
- Side Scatter (SSC). The SCC is detected at 90° to the incidental light and is primarily indicative of internal granularity and intercellular complexity of the cells.

3.7.4.2 Sample preparation and staining

The samples, taken directly from the cultures, were diluted with $0.22 \mu\text{M}$ filtered phosphate-buffered saline (PBS) to a final concentration of 10^5 - 10^6 mL^{-1} and analysed using a BD Accuri C6 flow cytometer at an appropriate data rate (Table 3.13).

Table 3.13 Operating parameters for the BD-Accuracy C6 FCM data acquisition

BD-Accuracy C6 flow cytometer	
Optical settings	Fixed, manufacturers setting. No artificial voltage manipulation or color compensation applied
Fluidics Settings	Default slow fluidics setting: $14 \mu\text{L}\cdot\text{mL}^{-1}$ flow rate $10 \mu\text{m}$ core size
Data rate	1000 - $2500 \text{ events}\cdot\text{s}^{-1}$
Events collected	20000 (in gate P1)

Samples were stained with different fluorescent dyes:

- Bis-oxonol (BOX, Sigma) for determining transmembrane potential;
- Propidium Iodide (P.I., Sigma) for determining membrane integrity;
- Congo Red (CR, BDH Chemicals) to determine amyloid content and presence of inclusion bodies.

Sample were excited using a 488 nm solid-state laser and fluorescence was detected using a 533/30 BP filter for BOX (FL1-A) and 670 nm LP filter for P.I. and CR (FL3-A). Particulate noise was eliminated using a FSC-H threshold. 20,000 data points were collected at a maximum rate of 2,500 events·s⁻¹. Data was analysed using the BD Accury™ C6 software. For routine analysis the pre-set fluidics setting “Slow” was used to maximize the passage time of cells through the laser due to the relatively small size of *E. coli*.

The samples were co-stained with BOX and P.I. to simultaneously assessed cell viability and vitality. Combining measures of cell membrane integrity (viability) and plasma membrane potential (vitality) in a single sample allows differentiation between dormant, depolarized cells (VBNC, viable but not culturable phenotype) that retain membrane integrity and may recover if culture conditions are optimized, and dead cells with depolarized membranes and compromised membrane integrity.

Figure 3.29 shows a control sample, with live and ethanol-killed *E. coli* M15 cells, on a P.I. (FL3-A) vs BOX (FL1-A) plot. Bacterial cells were harvested from a stationary-phase LB shake flasks culture, washed and diluted in ethanol. After 5 min, the sample was co-stained with both BOX/P.I.. Populations of live (BOX⁻P.I.⁻) and dead (BOX⁺P.I.⁺) cells are clearly visible. A third population of damaged cells (BOX⁺P.I.⁻) can be seen in the upper-

left quadrant, representing the cells with their cellular membrane intact but with the plasma membrane depolarized. Depending on culture conditions, these cells may remain dormant and damaged ($\text{BOX}^+\text{P.I.}^-$), progress to death ($\text{BOX}^+\text{P.I.}^+$), or regenerate their membrane potential and move into the live quadrant ($\text{BOX}^-\text{P.I.}^-$).

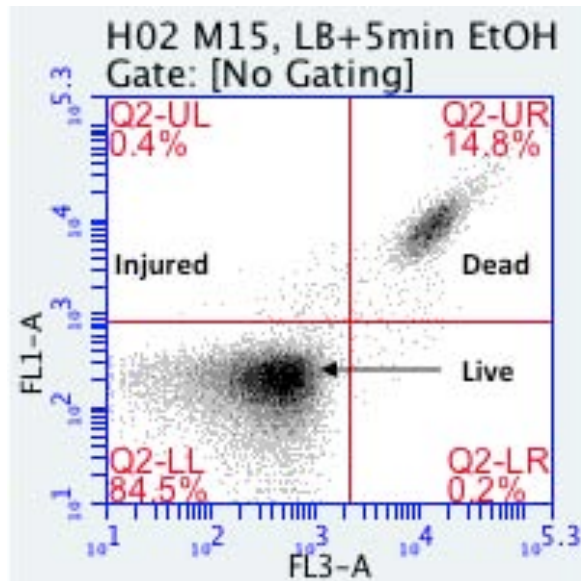


Figure 3.29 Viability and vitality of *E. coli* M15 cells, cultured for 18 hours. The sample was co-stained with BOX/P.I. 20,000 *E. coli* events/sample (gated on SSC-A vs FSC-A scatter) were acquired on a BD Accuri C6 with FSC-H threshold = 12,000. Results showed P.I. (FL3-A) vs BOX (FL1-A). Live cells are $\text{BOX}^-\text{P.I.}^-$, Damaged are $\text{BOX}^+\text{P.I.}^-$ and dead are $\text{BOX}^+\text{P.I.}^+$.

3.7.4.3 Fluorescent Dyes

Bis (1,3 dibutylbarbituric acid) thrimetine oxonol. BOX is the preferred membrane potential assay for *E. coli* and Gram-negative bacteria. Box is lipophilic, anionic and accumulates within cells with depolarized membrane but it is excluded from healthy cell with an active membrane potential. Thence membrane potential is used as test of viability, since BOX will only cause a cell to become fluorescent when either the cell is

membrane compromised (dead) or membrane depolarized (damaged cells). It is excited at 488 nm and its emission green fluorescence can be measured in the FL1 channel using the standard 533/30 nm emission filter. A 5 $\mu\text{g BOX mL}^{-1}$ stock solution was made up in ddH₂O and added to samples at 0.05 $\mu\text{g BOX mL}^{-1}$ (10 $\mu\text{L dye per 1 mL sample}$).

Propidium Iodide. P.I. is a red-fluorescent intercalating agent that binds to DNA, by little or no sequence preference, and with a stoichiometry of one dye per 4-5 base pairs of DNA. When bound to DNA P.I.'s fluorescence is enhanced 20- to 30-fold, its excitation maximum shifts 30-40 nm to the red and its emission maximum shifts 15 nm to the blue. It is excited at 535 nm and its fluorescence emission can be measured in the FL3 channel using 617 nm emission filter. As the cell membrane is impermeable to P.I., it will only be able to enter a cell and bind the DNA when the cell membrane is compromised. Therefore, P.I. is used to identifying dead cells in a population.

A 200 $\mu\text{g P.I. mL}^{-1}$ stock solution was made up in ddH₂O and added to samples at 4 $\mu\text{g P.I. mL}^{-1}$ (20 $\mu\text{L dye per 1 mL sample}$).

Congo Red. CR is a benzidine derivative, which can react with structural polysaccharides, like amyloids. Bacterial inclusion bodies have often been regarded as disordered precipitates of non-specifically coagulated polypeptide chains. Nevertheless, Carrió M. et al. (Carrió et al. 2005) have shown that bacterial IBs formation is a very selective reaction mechanistically similar to amyloid formation. Therefore, CR seems to be a good choice of dye for detecting IBs due to its ability to recognize amyloid deposits. Besides, CR displays a maximum absorbance at 497 nm

that shifts to red, bright fluorescence emission at 614 nm, once it binds amyloid material.

Stock solution of CR (BDH Chemicals) $2 \text{ mg}\cdot\text{mL}^{-1}$ solution was made by dissolving 2 mg of dye in 1 mL of dimethylsulphoxide (DMSO)(Fisher) and stored at 4°C ; as DMSO freezes at 19°C the solution was allowed to melt at room temperature before use. CR from stock was added to samples at a final concentration of $40 \mu\text{g CR mL}^{-1}$.

3.8 Protein Analysis

Samples from culture broths were withdrawn, adjusted to a final OD_{600} of 4 or 2, for Shake Flasks and Fed-batch cultures respectively, centrifuged and then processed as previously described (Vidal et al. 2003). Briefly, after rejecting the supernatant, the pellets were resuspended in 100 mM Tris-HCl (pH 7.5) keeping the same OD_{600} previously adjusted. Cell suspensions were placed in ice and sonicated with four 15 s pulses at 50 W with 2 min intervals in ice between each pulse, using a Vibracell™ model VC50 (Sonics & Materials). Cellular debris was then removed by centrifugation and the clear supernatant was collected for product analysis.

Two different methods were used to quantify the product in each sample. One was used to calculate the FucA activity while the second one was used to quantify the amount of FucA amongst the rest of intracellular soluble proteins .

3.8.1 Enzyme activity for FucA

Determination of FucA activity was carried out with a coupled assay (Figure 3.30). In the first step, fucose 1-phosphate is cleaved to L-lactaldehyde and DHAP. In the second step, DHAP is reduced using rabbit muscle α -GDH and NADH. In the assay, the sample to be titrated was mixed with a solution containing 0.14 mM NADH, 2.0 mM fucose 1-phosphate, 1.7 AU·mL⁻¹ GDH, 50 mM KCl and 100 mM Tris-HCl at pH 7.5. The reaction was incubated at 25°C and the absorbance at 340 nm monitored in a UV-VIS Cary (Varian) spectrophotometer ($\epsilon = 6.22 \text{ mM}^{-1}\text{cm}^{-1}$).

One unit of FucA activity is defined as the amount of enzyme required to convert 1 μmol of fucose-1-phosphate in DHAP and L-lactaldehyde for minute at 25°C and pH 7.5.

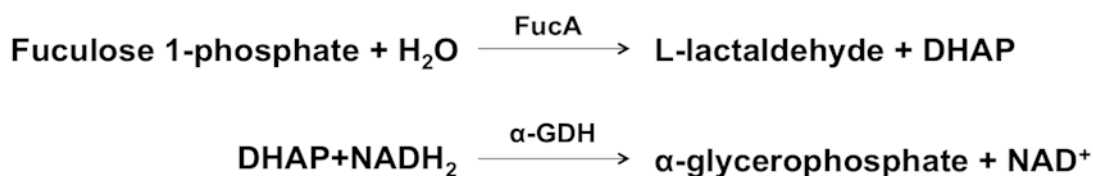


Figure 3.30 Schematic representation of the Fucose 1-phosphate assay

3.8.2 Total protein content

The Bradford protein assay is an analytical procedure used to measure the concentration of total protein in a sample. This method uses Comassie[®] Protein Assay Reagent Kit (Thermo Scientific, US) and Bovine Serum Albumin (BSA) as standard. To

determine the absorbance of the different samples, Microplate systems (Microtiter Plate Flat, SUDELAB 900011) were used. The absorbance was read with ThermoScientific® Multiskan FC equipment.

3.8.3 SDS-PAGE

Electrophoresis has been used to determine the percentage of FucA among the rest of intracellular soluble proteins. SDS-PAGE gels were performed in two different ways.

For all the experiments performed at the Universitat Autònoma de Barcelona, NuPAGE® 12 % Bis Tris gels with MES-SDS as running buffer have been used, following the manufacturer's manual (Invitrogen, US). Samples were prepared with 5 µL of sample buffer 4X, 3 µL of milliQ water, 2 µL of reducing agent and 10 µL of the sample. After 10 min incubation at 70°C and 300 rpm in a dry bath, 15 µL of each sample were loaded into the gel. One of the wells of the gels was loaded with 5 µL of protein standard marker (Bio-Rad®). After 40min running at 200 V, gels were rinsed with distilled water and then covered with gel fixation solution (40 % methanol, 10 % acetic acid in water) for 1 hour. Finally, gels were covered with Bio-Safe™ Comassie (Bio-Rad®) for 1-2 hours, till the protein bands appear. Percentage of protein was quantified by densitometry using the specific software Image Lab® (Bio-Rad Laboratories).

For the SDS-PAGE performed both at the Institute of Systems and Synthetic Biology at the University of Evry (Genopole-CNRS, Paris, FR), and at the School of Engineering at

the University of Birmingham (Birmingham, UK) protein gels were prepared as explain in Table 3.14.

A commercially produced stock solution of 30 % (w/v) acrylamide, 0.8% (w/v) bisacrylamide (National Diagnostics) was used for both resolving and stacking gels. The Tris-HCl buffers were prepared dissolving the required amount of Tris in ddH₂O and adjusting the pH with HCl. A commercially produced 20 % (w/v) sodium dodecyl sulphate (SDS) (Bio-Rad) was used. An 80 mg·mL⁻¹ of ammonium persulphate (APS) was prepared in ddH₂O and stored at -20°C in 1 mL aliquots.

Table 3.14 Casting gels preparation

	Resolving Gel (15 %)	Stacking Gel (6 %)
Acrylamide solution	10 mL	3 mL
Tris-HCl 0.75 M pH 8.3	10 mL	-
Tris-HCl 1.25 M pH 6.8	-	1.5 mL
SDS	100 µL	75 µL
APS	200 µL	150 µL
TEMED	10 µL	7.5 µL
ddH ₂ O	-	10.5 mL

Sample buffer was prepared mixing together 2 g SDS, 20 mL glycerol and 5 mg bromophenol blue (sulphone form) in 92 mL of 10X diluted Tris-HCl 0.75 M pH 8.3 buffer, and finally divided into 1 mL aliquots. Immediately prior to use 87 µL of β-mercaptoethanol were added to the tube. 10 µL of each sample were resuspended in 60 µL of sample buffer (with β-mercaptoethanol) and boiled for 10 min and 15 µL were

loaded into the gel. 1 L of running buffer was prepared mixing 5 mL of 20 % SDS with 100 mL of electrode buffer stock solution (150 g glycine, 30 g Tris in 1 L) and ddH₂O.

The gels were run at 100 V for half an hour and then at 200 V for about an hour. After running the gels, these were rinsed with distilled water and then stained with SimplyBlue SafeStain solution (Invitrogen) for 1 hour. Relative amount of each protein was quantified by densitometry using the specific software Image Lab[®] (Bio-Rad Laboratories).

Finally dried SDS-PAGE were scanned with a Canoscan 9000 F (Canon) and quantified by densitometry with the software package ImageJ (Bio-Rad[®]).

4 Results I: *Ab initio* FucA overexpression in *E. coli*

FucA over expression studies in our laboratory were initiated when the gene *fucA* was isolated from *E. coli* K12 and cloned into a plasmid vector under the control of the *Trc* promoter (pTrcfucA) and overexpressed as a recombinant his-tag fusion protein by *E. coli* XL1 Blue MRF' (Garcia-Junceda et al. 1995). Subsequent studies were focused on the effect of the medium composition and the amount of inducer on FucA production (Durany et al. 2004). Besides, the behavior and control of the fermentation process as well as the effect of the time of induction on the cell growth and recombinant protein expression were further optimized (Durany, de Mas, and López-Santín 2005).

After these initial studies, that showed poor over-expression levels, our Group switched to another commercially available expression system (Qiagen), consisting of the *E. coli* M15 strain and two-plasmid system including the pQE-40 expression vector, which uses the stronger T5 promoter, and the pREP4 vector for the constitutive expression of the LacI repressor protein (Vidal et al. 2003). This system was used to overproduce a wide range of aldolases, including a DHAP-dependent aldolase, the rhamnulose-1-phosphate aldolase (RhuA) (Vidal et al. 2005). Besides, preliminary studies of FucA production with the Qiagen expression system in shake flasks cultures were presented by Vidal in his thesis work (Vidal 2006).

Furthermore, this reference expression system was further evolved in our laboratory by introducing a novel auxotrophic marker-based plasmid maintenance system based

on amino acid auxotrophy complementation. It is well known that antibiotics and antibiotic resistance genes have been traditionally used for the selection and maintenance of recombinant plasmids in hosts such as *Escherichia coli*. Although a powerful selection tool, the choice to explore an antibiotic-free alternative plasmid selection system, arises from the fact that the use of antibiotics has been considered unacceptable in many areas of biotechnology by regulatory authorities (Glenting and Wessels 2005). The use of selection markers that confer resistance to antibiotics may cause potential biosafety and clinical hazards, such the risk of transforming the patient's microflora, the possibility of horizontal spread of resistance genes or the rapid emergence of multidrug-resistant pathogens. Moreover, in recombinant protein production for therapeutics, the antibiotic must be eliminated from the final product. Another problem arises from the potential loss of selective pressure as a result of antibiotic degradation; as an example ampicillin can be degraded by β -lactamases in less than 30 min in high cell density cultures (Jung et al. 1988).

This evolved expression system used here was obtained by (i) knocking out the *glyA* gene (encoding for the enzyme serine hydroxymethyl transferase, SHMT) from the M15 prototrophic strain using a PCR-based method (Datsenko and Wanner 2000), thereby obtaining the glycine-auxotrophic M15 Δ *glyA* strain, and (ii) by constructing a pQE-40 derived plasmid containing the auxotrophic marker gene (*glyA*) under the control of a weak constitutive promoter (Vidal et al. 2008). The SHMT is a 46 kDa enzyme, member of the α -class of pyridoxal phosphate enzymes (EC 2.1.2.1), acting as the first enzyme in the assimilation of C1 compound and it is the main enzyme responsible for the synthesis of intracellular glycine when glucose is the main carbon

source (Plamann and Stauffer 1983). In particular, SHMT has two activities: i) threonine aldolase activity, which catalyzes the reversible interconversion between L-threonine and glycine plus acetaldehyde, and ii) serine hydroxymethyl transferase activity, which catalyzes the reversible interconversion between serine and glycine (Plamann and Stauffer 1983). This autotrophic plasmid selection system was initially tested for heterologous protein expression using RhuA as model protein (Vidal et al. 2005).

The present work is focused on the production of the DHAP-dependent aldolase FucA, with the objective of maximizing specific activity and productivity of the recombinant enzyme.

The first aim of our study was to clone and express the *fucA* gene into the auxotrophic M15 Δ *glyA* strains. Besides, in order to obtain high intracellular expression levels of the protein of interest, the auxotrophic expression system together with its predecessor system have been tested and compared under protein production conditions in shake flasks and fed-batch cultures in defined media.

4.1 Preliminary FucA expression studies with the M15[pREP4] and M15 Δ glyA[pREP4] strains

As described in Materials and Methods the *fucA* gene has been cloned into the commercial vector pQE-40 and the auxotrophic expression vector, yielding the strains M15[pREP4] pQE-FucA, from now on M15[pREP4] (Section 3.4.1) and M15 Δ glyA[pREP4] pQE $\alpha\beta$ FucA, from now on M15 Δ glyA[pREP4] (Section 3.4.2), respectively.

Moreover, in previous studies of our group (Durany et al. 2005), FucA production was obtained in high cell density cultures in bioreactors, using a defined mineral medium (DM). This medium, formulated according to specific yields of main nutrients to biomass for *E. coli* and reported general nutritional requirements for bacteria (Durany et al. 2005) (Yee and Blanch 1993), uses glucose as the sole carbon source (Materials and Methods section 3.5.2).

In order to use the same operational strategy, preliminary studies of the reference expression system *E. coli* M15[pREP4] and the auxotrophic *E. coli* M15 Δ glyA[pREP4] were assessed in shake flasks with both LB and DM media. First of all, strains were compared in terms of maximum specific growth rate in both LB and DM.

Subsequently, heterologous expression was induced in mid-exponential phase cultures using 1mM IPTG. In order to ensure maximum FucA production, the induction phase was maintained for 4 hours. Biomass, FucA and acetate production, as well as glucose consumption profiles along time have been analysed.

Lastly, flow cytometry analyses were performed for both the samples pre-induction and post-induction, in order to analyze the bacterial populations at the single-cell level: changes in morphology forward scatter and side scatter measurement (FSC and SSC) and physical and biochemical characteristics of individual cells within a bacterial population. Samples were co-stained with BOX/PI to simultaneously assess cell viability and vitality.

4.1.1 FucA expression in the reference *E. Coli* M15[pREP4] strain

Initially the performance of the reference FucA expression system *E. coli* M15[pREP4] was assessed in shake-flask cultures in both LB and defined media (DM). The maximum specific growth rate (μ_{\max}) in both LB and DM media was calculated being 0.49 ± 0.01 h⁻¹, although the lag phase in the defined media was a little bit longer (Figure 4.1). This is probably due to the necessity of the cells to adapt to the new environmental conditions.

Besides, under aerobic conditions when the glucose concentration in the medium is non-limiting, *E. coli* displays a high rate of substrate consumption and a corresponding high specific growth rate (μ_{\max}). Under these conditions, *E. coli* secretes acetic acid. This is the result of a metabolic imbalance, also known as overflow metabolism, where the rate of acetyl-coenzyme A (AcCoA) synthesis surpasses the capacity of the tricarboxylic acid (TCA) cycle to completely consume this metabolite.

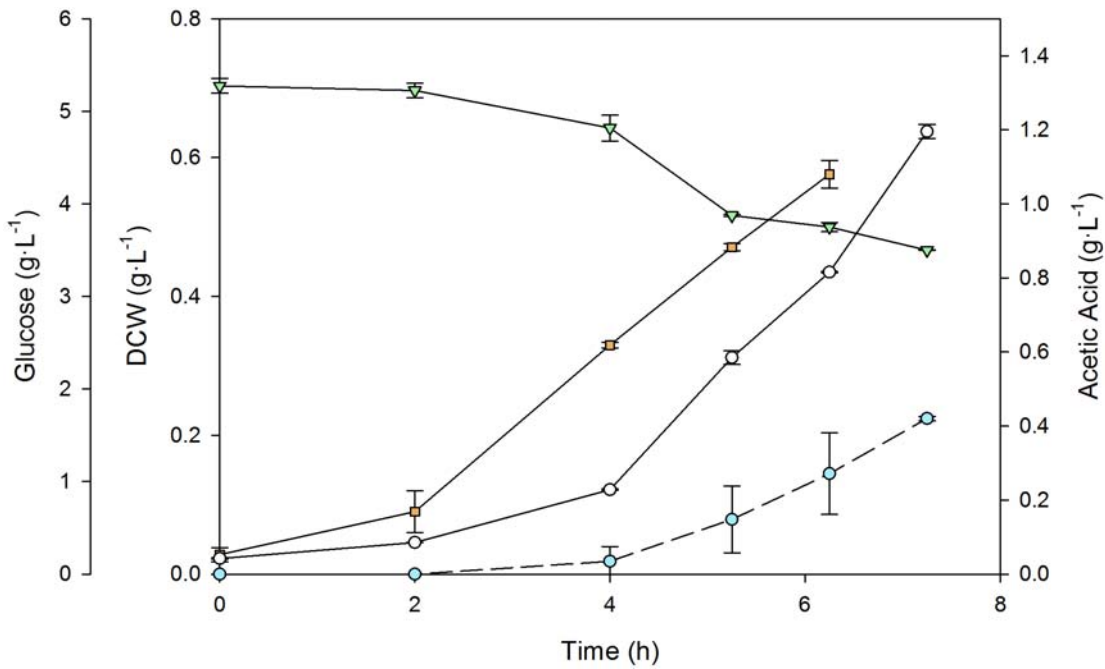


Figure 4.1 Shake flasks cultures of *E. coli* M15[pREP4] without induction. Profiles along time of (□) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$) in LB media and (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (▽) Glucose concentration ($\text{g}\cdot\text{L}^{-1}$) and (●) Acetic Acid ($\text{g}\cdot\text{L}^{-1}$) in DM medium.

Therefore, the amount of AcCoA not consumed by the TCA cycle is diverted into the phosphotransacetylase (Pta)-acetate kinase (Ack) pathway where acetate is synthesized (Fuentes et al. 2013) (Kang et al. 2009). Moreover, acetate production represents a diversion of carbon that might otherwise have generated biomass or the protein product (Phue and Å 2005). Furthermore, acetate production should be avoided, because its accumulation in culture media reduces cell growth and productivity.

The calculation of the substrate uptake rate ($q_s \text{ exp}$) during the exponential phase for the reference strain M15[pREP4] was $1.41 \pm 0.05 \text{ gGlc}\cdot\text{g}^{-1}\text{DCW}\cdot\text{h}^{-1}$, while the final amount of acetate was $0.43 \pm 0.06 \text{ g}\cdot\text{L}^{-1}$.

Furthermore, FucA overexpression experiments were performed in triplicates using DM medium, this time with induction using IPTG. Results are presented in Figure 4.2. Reference M15[pREP4] strain presents, again, μ_{\max} of $0.49 \pm 0.01 \text{ h}^{-1}$ till the pre-induction point, but the cell growth starts to slow down after IPTG induction, indicated by an arrow in Figure 4.2 A. The substrate uptake rates along the induction phase (q_s ind) is $0.37 \pm 0.04 \text{ gGlc}\cdot\text{g}^{-1}\text{DCW}\cdot\text{h}^{-1}$, is reduced by 77 % in relation to the q_s of the exponential phase (q_s exp). This concomitantly reduces the final concentration of acetate in the fermentation broth to $0.41 \pm 0.02 \text{ g}\cdot\text{L}^{-1}$, 10 % lower compared with the non-induced strain (Figure 4.2B).

Besides, in terms of FucA production, M15[pREP4] strain reached a final amount of $181 \pm 5 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$, with an activity of $721 \pm 82 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ (Figure 4.2A). The Figure 4.3 shows clearly the difference between FucA levels during the pre-induction and post-induction phases, where the amount of the recombinant protein reaches around the 30 % of total cell protein. Moreover, in terms of *fucA* expression, the Figure 4.2A shows how after only 1 hour of induction, the FucA activity reaches already the 70 % of the maximum. Finally, in Table 4.1 the SHMT concentration values ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) of the M15[pREP4] strain are represented. As it can be seen, the higher value is represented by the pre-induction, being $27 \pm 14 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$.

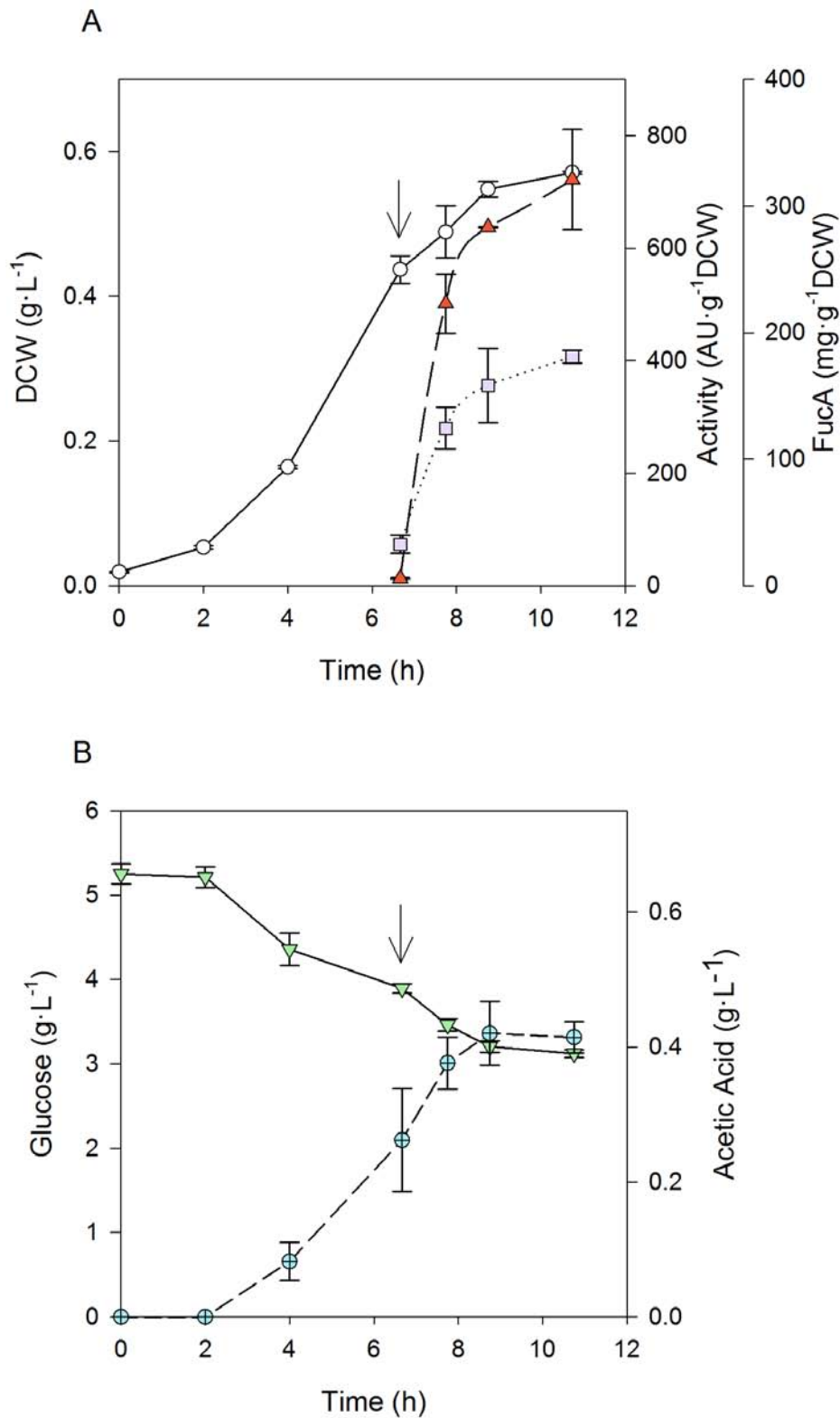


Figure 4.2 Shake flask cultures of *E. coli* M15[pREP4] strain in DM medium. **A**) (●) Biomass DCW (g·L⁻¹), (▲) enzyme activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹); **B**) (▼) Glucose concentration (g·L⁻¹) and (●) Acetic Acid (g·L⁻¹). 1mM IPTG pulse is indicated by an arrow.

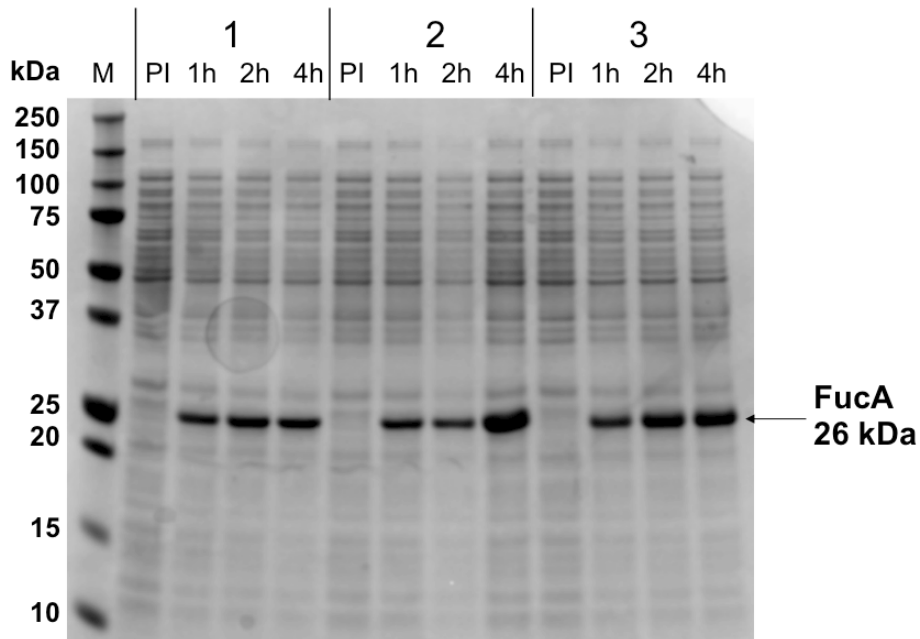


Figure 4.3 SDS-PAGE of shake flask cultures of *E. coli* M15[pREP4] strain in DM medium. Lane M: molecular weight marker; 1, 2, 3 are the biological replicates; PI, pre-induction; 1 h, 2 h and 4 h correspond to the time after induction. The band corresponding to the 26 kDa FucA is indicated in the figure.

Table 4.1 SHMT production ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) along the induction phase for the M15[pREP4] strain. PI (pre-induction) and 1 h, 2h and 4h correspond to the time after induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	27 ± 14
1h	14 ± 1
2h	12 ± 4
4h	13 ± 1

4.1.1.1 Flow cytometry analysis of the reference strain *E. coli* M15[pREP4]

For the flow cytometry analysis three experimental conditions were tested, being: A) 1mM IPTG plus ampicillin, B) 70 μ M IPTG plus ampicillin and C) 70 μ M IPTG without antibiotic.

As an example, profiles of the M15[pREP4] strain growing on shake flask cultures in DM plus ampicillin are shown in Figure 4.4. In these cultures, FucA expression was induced with a 1mM IPTG pulse. After induction, indicated with an arrow in the figure, the μ estimated on the basis of OD variation over time decreased from the pre-induction to the post-induction phase (Figure 4.4).

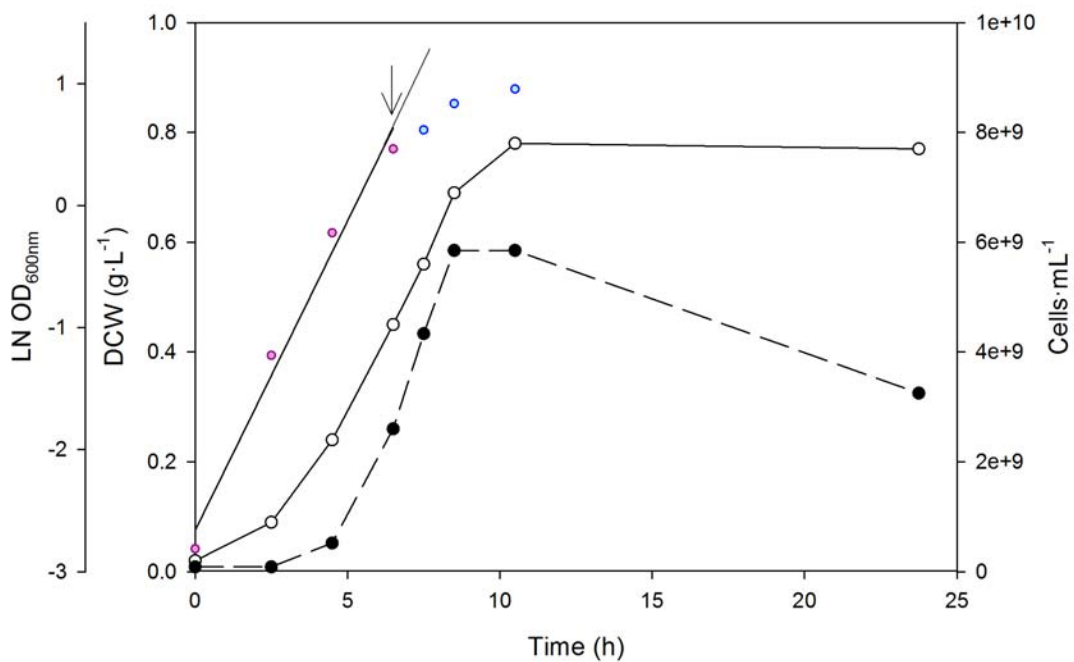


Figure 4.4 Shake flasks culture of the *E. coli* M15[pREP4] strain in defined media plus antibiotic. (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (c) concentration determined using flow ($\text{cell}\cdot\text{mL}^{-1}$), (●) LN OD₆₀₀ of the pre-induction phase and (●) LN OD₆₀₀ of the post-induction phase. 1mM IPTG induction is indicated with an arrow and the curve that fits the pre-induction samples is represented.

This reduction is further confirmed by the measurements of cell concentration ($\text{cells}\cdot\text{mL}^{-1}$) monitored by flow cytometry, which does not show any further increase after 2 h of induction. Moreover, whereas the value of the DCW ($\text{g}\cdot\text{L}^{-1}$) remains invariant from the 4th to the 18th hour of induction, cell concentration decreases around 50 % over this period, as shown in Figure 4.4.

As it can be seen in Figure 4.5A, the curve representing the FSC of the 18 h post-induction sample shifts to the right compared to the curve of the pre-induction sample showing higher forward scatter. This change does not occur in the negative control cultivation without induction, where the FSC curves corresponding to pre-induction and non-induced sample after 18h overlap (Figure 4.5B).

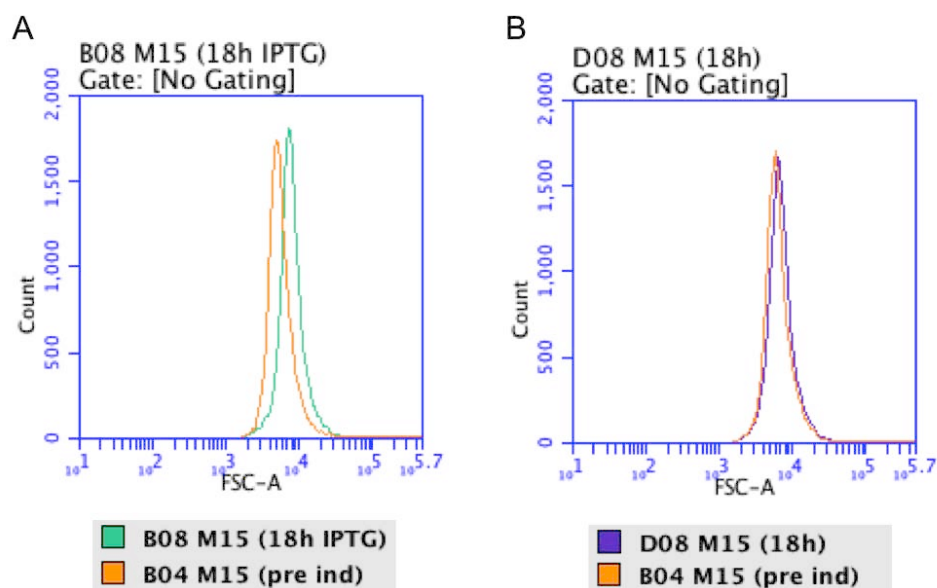


Figure 4.5 Flow cytometry analysis of the forward-scatter light (FSC-A) versus count, of the M15[pREP4] strain's samples in DM medium. A) pre-induction (pre ind, in orange) and of 18 hours after induction (18 h IPTG, green); B) negative control not induced where the pre-induction sample is in orange (pre ind) and the non-induced sample after 18h is in purple (18 h).

Thus, one possible explication is that the cells after being induced undergo a morphological change of their shape, becoming more elongated, showing higher forward scatter. Conversely, this can bring to an overestimation of optical density values (OD_{600}). In fact, as definition, OD or absorbance, is a measure of the amount of light that is absorbed or scattered by a bacteria suspension. The bacteria absorb or scatter light depending on their concentration, size and shape.

Unfortunately, a limitation of OD measurements is the inability to distinguish between dead and living cells and to perceive changes in cell morphology, which leads to an overestimation of the cell concentration (Hudson and Mott 1994).

Analysis of samples by flow cytometry using as fluorescent dyes BOX and P.I., revealed that the population of the pre-induction (corresponding to cells harvested at 6.5 h of growth in shake flasks, i.e. exponential growth phase) was relatively homogenous and 99 % healthy (Figure 4.6, quadrant Q1 of the pre-induction sample plot).

As the induction time progresses from 1 h to 4 h, the percentage of healthy population decreases slowly. As a consequence the amount of unhealthy and dead cells increases. This viability change reaches its peak at 18 h of induction, where more than 34 % of the population is permeable to BOX, indicating loss of transmembrane potential. Moreover, about of 7 % of the cells are also permeable to the P.I. dye meaning that they are dead. The higher amount of damaged cells in Q2 quadrant is probably due to both the presence of ampicillin in the media that forces the cells to maintain the expression vector, and the higher induction pulse of IPTG.

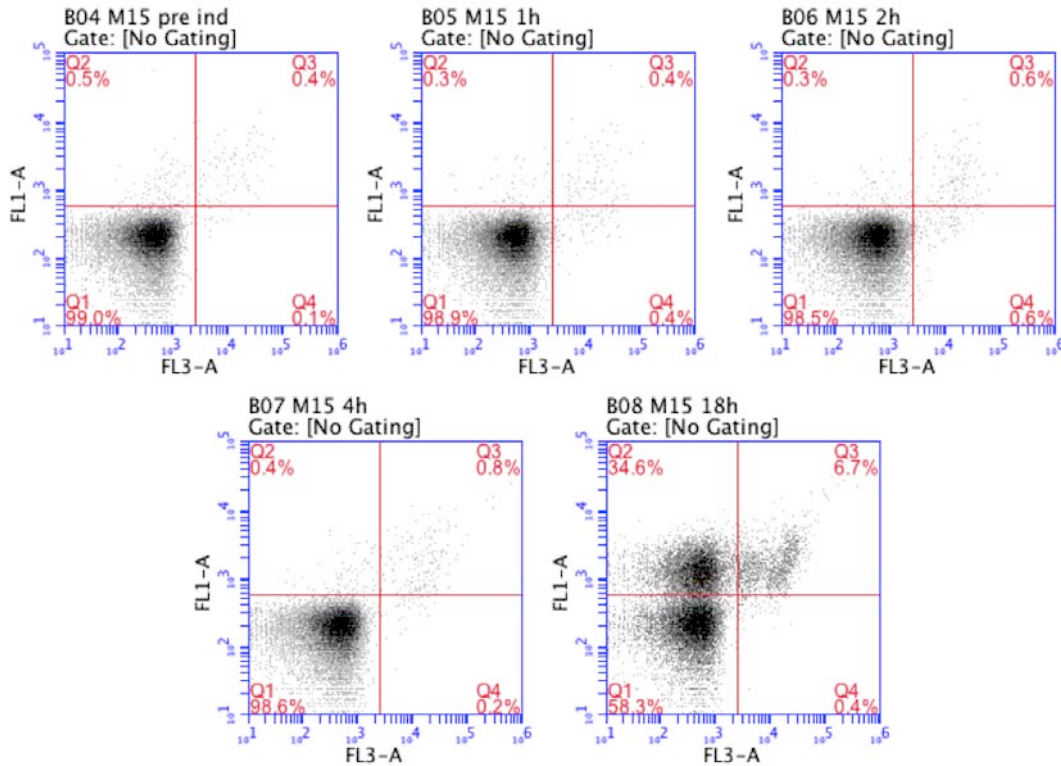


Figure 4.6 FL1-A (BOX fluorescence) versus FL3-A (P.I. fluorescence) plots of cells of *E. coli* M15[pREP4] shake flask cultures in DM medium plus antibiotic with a 1 mM pulse of IPTG. Plots corresponding to cells harvested during pre-induction (pre ind) and post-induction at 1 h, 2 h, 4 h and 18 h are represented. Samples were co-stained with both BOX/P.I. (See Materials and Methods section 3.7.4.2). The numbers in each quadrant show percentage of bacteria population in each subgroup.

A comparison of the three experimental conditions tested: A) 1mM IPTG plus ampicillin, B) 70 μ M IPTG plus ampicillin and C) 70 μ M IPTG without antibiotic, is represented in Figure 4.7. After 4 hours of induction, in all the three conditions, the cells are healthy and live with less than 1 % of damaged or dead cells, except for the 4 h of induction in the condition of 70 μ M IPTG without antibiotic, where the amount of damaged cells increased till 2.5 % (Figure 4.7C). Differently, at 18 h of induction, when the growth rate of the cultures declines and the cells enter the stationary phase, the % of damaged or dead cells increases significantly.

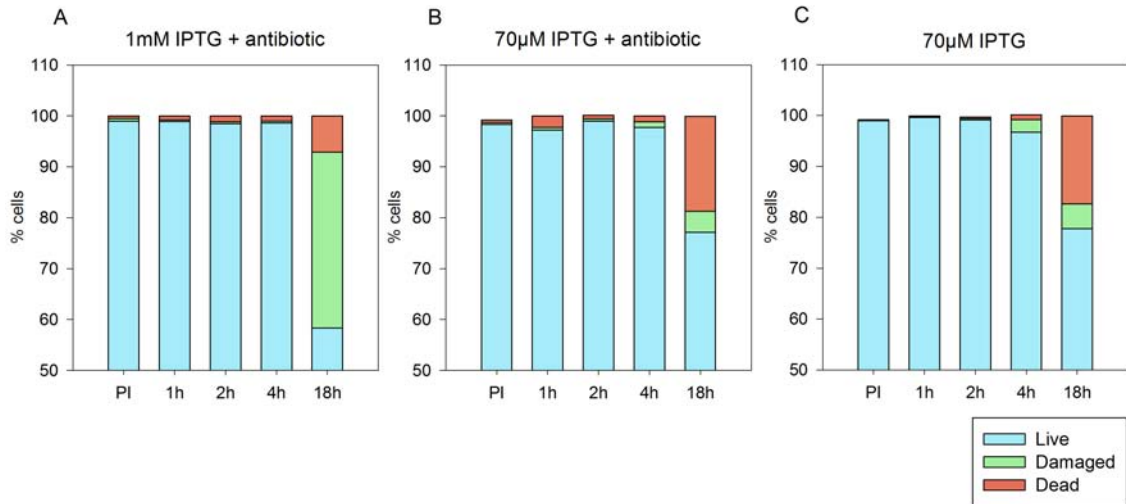


Figure 4.7 Comparison of the three experimental condition in shake flask cultures for the *E. coli* M15[pREP4] strain. PI, pre-induction samples; 1 h, 2 h, 4 h and 18 h represent the time after IPTG induction. A) DM + antibiotic and 1 mM IPTG; B) DM + antibiotic and 70 μ M IPTG; C) DM without antibiotic and 70 μ M IPTG. Live cells are represented in light blue, damaged cells in green and dead cells in orange.

In Figure 4.7A, the experiment with 1mM induction brings to the higher amount of damaged cells, being 34.6 % plus 7.1 % of dead cells (Figure 4.7A).

The reduction of the inducer concentration leads to an improvement of the physiological state of the cells, reaching only a 4 % of damaged cells (Figure 4.7B, C). It must be pointed out that the reduction of the IPTG pulse causes a 2.4-fold increase of the number of the dead cells.

Unexpectedly, the elimination of the antibiotic in the culture medium does not seem to affect the cell viability/vitality after induction with a 70 μ M pulse of IPTG (Figure 4.7C). In fact, fast growing bacteria with a metabolic burden due to recombinant protein expression tend to lose the plasmid in the absence of antibiotic selection pressure (Hägg et al. 2004).

However, the presence of plasmid-free segregants is also not totally prevented in presence of the ampicillin, since the concentration of the antibiotic used for the plasmid selection often decreases during long-term cultivation as a result of dilution and/or enzymatic degradation.

4.1.2 FucA expression in the *E. coli* M15 Δ *glyA*[pREP4] strain

Once obtained the recombinant *E. coli* M15 Δ *glyA*[pREP4] strain, the growth and the FucA overproduction were assessed in shake flask cultures. Firstly, the growth rate of this strain in both LB and DM media were compared (Figure 4.8).

As expected, the auxotrophic strain M15 Δ *glyA*[pREP4] is able to grow in defined media, without glycine supplementation, proving that the *glyA* gene under control of the constitutive promoter P3 is correctly expressed, providing the auxotrophic strain the sufficient amount of SHMT (Figure 4.8). Thanks to the presence of this enzyme, the cells can degrade threonine and obtain the endogenous amount of glycine. It is important to underline that, while the growth in LB media is characterized by a μ_{max} of 0.48 h^{-1} , this value slightly decreases in DM media, reaching a value of 0.44 h^{-1} .

The calculation of the substrate uptake rate during the exponential phase ($q_s \text{ exp}$) for the auxotrophic strain is $1.14 \pm 0.01 \text{ gGlc} \cdot \text{gDCW}^{-1} \cdot \text{h}^{-1}$, with the final amount of acetate being $0.28 \pm 0.01 \text{ g} \cdot \text{L}^{-1}$.

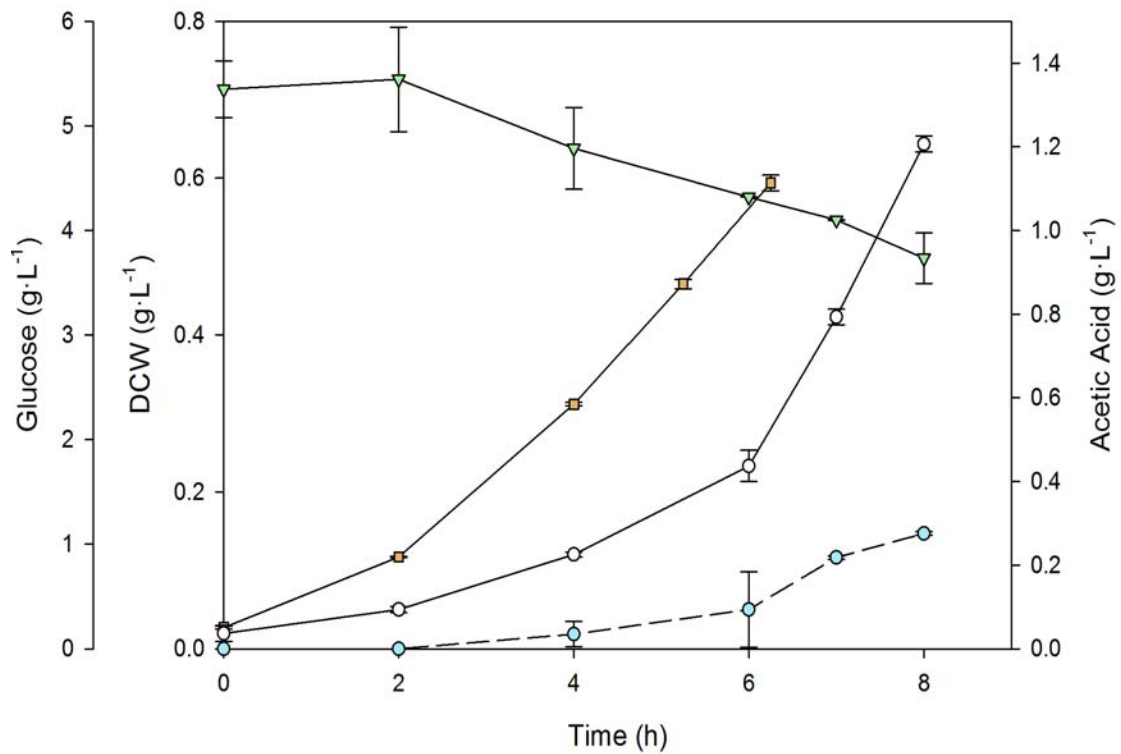






Figure 4.8 Shake flask cultures of *E. coli* M15ΔglyA[pREP4] without induction. Profiles along time of () Biomass DCW (g·L⁻¹) in LB media and () Biomass DCW (g·L⁻¹), () Glucose concentration (g·L⁻¹) and () Acetic Acid (g·L⁻¹) in DM medium.

As for the reference M15[pREP4] strain, shake flask cultures were performed, where heterologous expression was induced at mid-exponential phase with 1mM IPTG in order to compare biomass, FucA and acetate production, as well as glucose consumption profiles along time (Figure 4.9).

The calculated substrate uptake rate along induction phase (q_s ind) for the M15ΔglyA[pREP4] strain is 0.50 ± 0.13 gGlc·gDCW⁻¹·h⁻¹ a 44 % the value of the q_s exp. Besides, the final concentration of acetate is 0.54 ± 0.03 g·L⁻¹, as it can be observed in Figure 4.9 B. Moreover, whereas the M15 strain reached a final production of 181 ± 5 mgFucA·g⁻¹DCW, with an activity of 721 ± 82 AU·g⁻¹DCW, these values were reduced to

$67 \pm 37 \text{ mgFucA} \cdot \text{g}^{-1}\text{DCW}$ and $291 \pm 24 \text{ AU} \cdot \text{g}^{-1}\text{DCW}$, respectively, in the M15 Δ *glyA* strain (Figure 4.9A).

Overall, the results presented above suggested a physiological impact of the plasmid-encoded SHMT overexpression observed in the M15 Δ *glyA*[pREP4] strain. In fact, significant amounts of SHMT, accumulated in the cytoplasm of M15 Δ *glyA*[pREP4] cells, can be clearly observed in SDS-PAGE analysis (Figure 4.10). Besides, by comparing the Table 4.2 with Table 4.1 it can be seen that the SHMT production ($\text{mgSHMT} \cdot \text{g}^{-1}\text{DCW}$) increased more than 4.5-fold when moving from the M15[pREP4] to the M15 Δ *glyA*[pREP4] strain. This observation suggested that *glyA* overexpression from the plasmid imposed a significant burden to the metabolism of the host cell, thereby affecting negatively FucA expression levels and the μ_{max} .

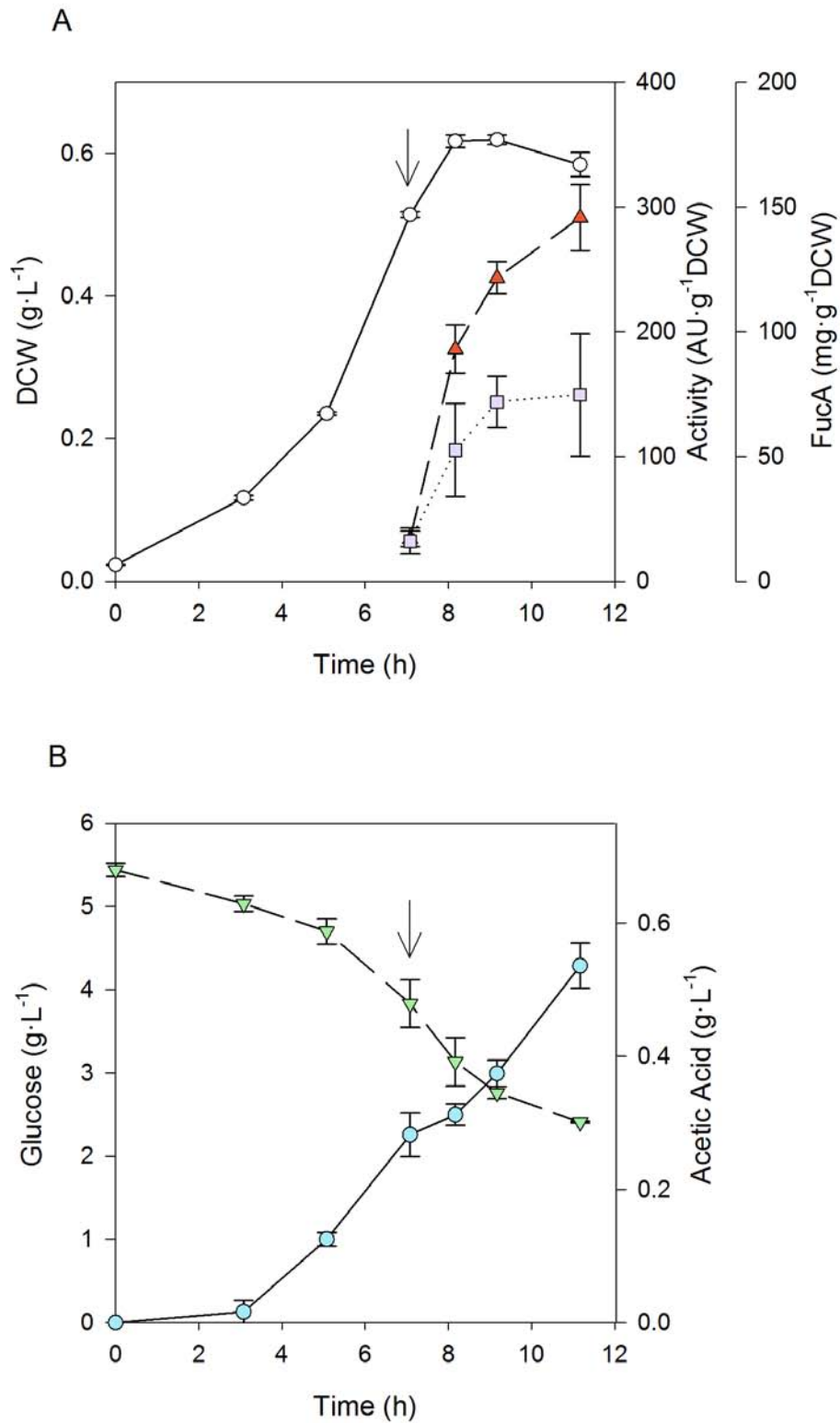


Figure 4.9 Shake flask cultures of *E. coli* M15 Δ glyA[pREP4] strain in DM medium. **A**) (●) Biomass DCW(g·L⁻¹), (▲) enzyme activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹); **B**) (▼) Glucose concentration (g·L⁻¹) and (●) Acetic Acid (g·L⁻¹). 1mM IPTG pulse is indicated with an arrow.

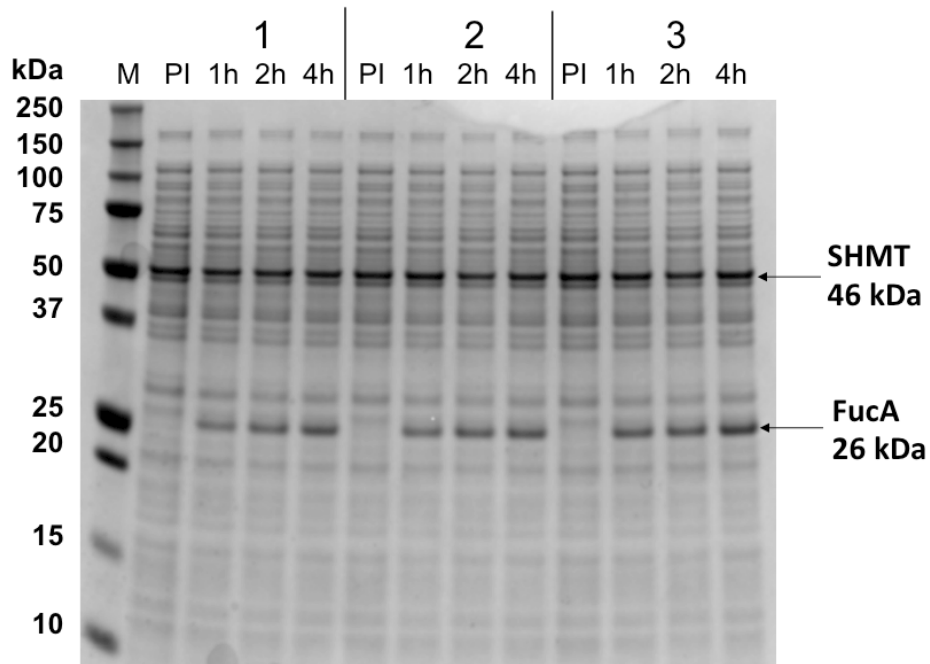


Figure 4.10 SDS-PAGE of shake flasks cultures of *E. coli* M15 Δ glyA[pREP4] strain in DM medium. Lane M: molecular weight marker. 1, 2, 3 correspond to each of the biological replicates. PI, pre-induction; 1 h, 2 h and 4 h correspond to the time after induction. The bands corresponding to the 26 kDa FucA and the 46 kDa SHMT are indicated.

Table 4.2 SHMT production ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) along the induction phase for the M15 Δ glyA[pREP4] strain. PI (pre-induction) and 1 h, 2h and 4h correspond to the time after induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	92 ± 14
1h	95 ± 7
2h	91 ± 11
4h	90 ± 14

4.1.2.1 Flow cytometry analysis of M15ΔglyA[pREP4] strain

As for the reference M15 strain, three experimental conditions were performed for the flow cytometer analysis: A) 1mM IPTG plus ampicillin, B) 70 μM IPTG plus ampicillin and C) 70 μM IPTG without antibiotic.

Cultures of *E. coli* M15ΔglyA[pREP4] growing in DM plus ampicillin and induced with a 1 mM pulse of IPTG are represented as an example in Figure 4.11. The μ decreases upon IPTG induction; however, unlike the reference strain, the cells concentration (cells·mL⁻¹) increased steadily until the 4th hour of induction; thereafter cell concentration starts decreasing too (Figure 4.11).

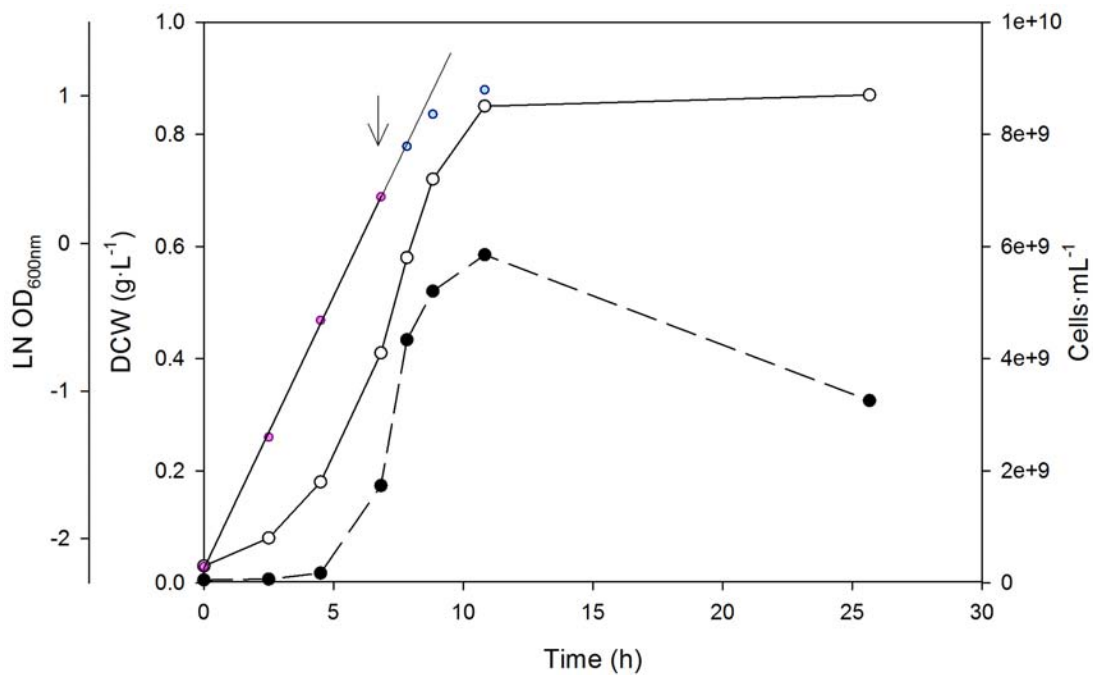


Figure 4.11 Shake flasks culture of the *E. coli* M15ΔglyA[pREP4] strain in defined media plus antibiotic. (●) Biomass DCW (g·L⁻¹), (●) concentration determined using FCM (cell·mL⁻¹), (●) LN OD₆₀₀ of the pre-induction phase and (○) LN OD₆₀₀ of the post-induction phase. 1mM IPTG induction is indicate with an arrow.

As for the reference strain, explanation of this fact can be found in the FSC analyses growing cells (Figure 4.12), where the curve representing the FSC of the 18 h IPTG-induced cells shifts to the right compared to the reference (pre-induced) cells (Figure 4.12A). Again, this morphological change is not observed in the negative control without induction (Figure 4.12B), where the two FSC curves corresponding to pre-induction and 18 h of IPTG-induction, are superimposed. As explained before, the cells may undergo a morphological change after being induced, becoming more elongated.

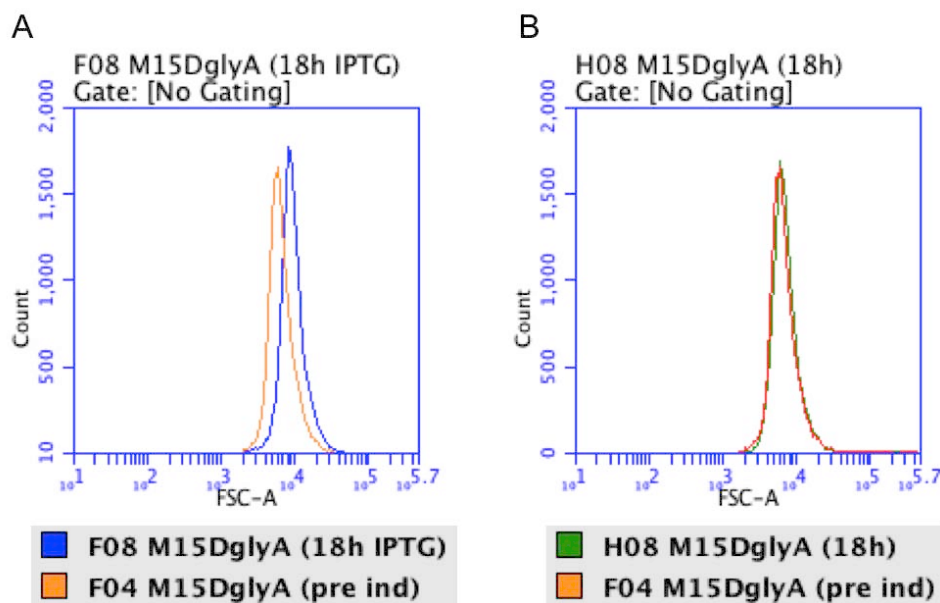


Figure 4.12 Flow cytometry analysis of the forward-scatter (FSC) of the samples of the pre-induction (pre ind, in orange) and of 18 hours after induction of both samples induced (18 h IPTG, green) and negative control not induced (18 h, purple). Samples of shake flasks cultures of *E. coli* M15 Δ glyA[pREP4] strain in DM medium.

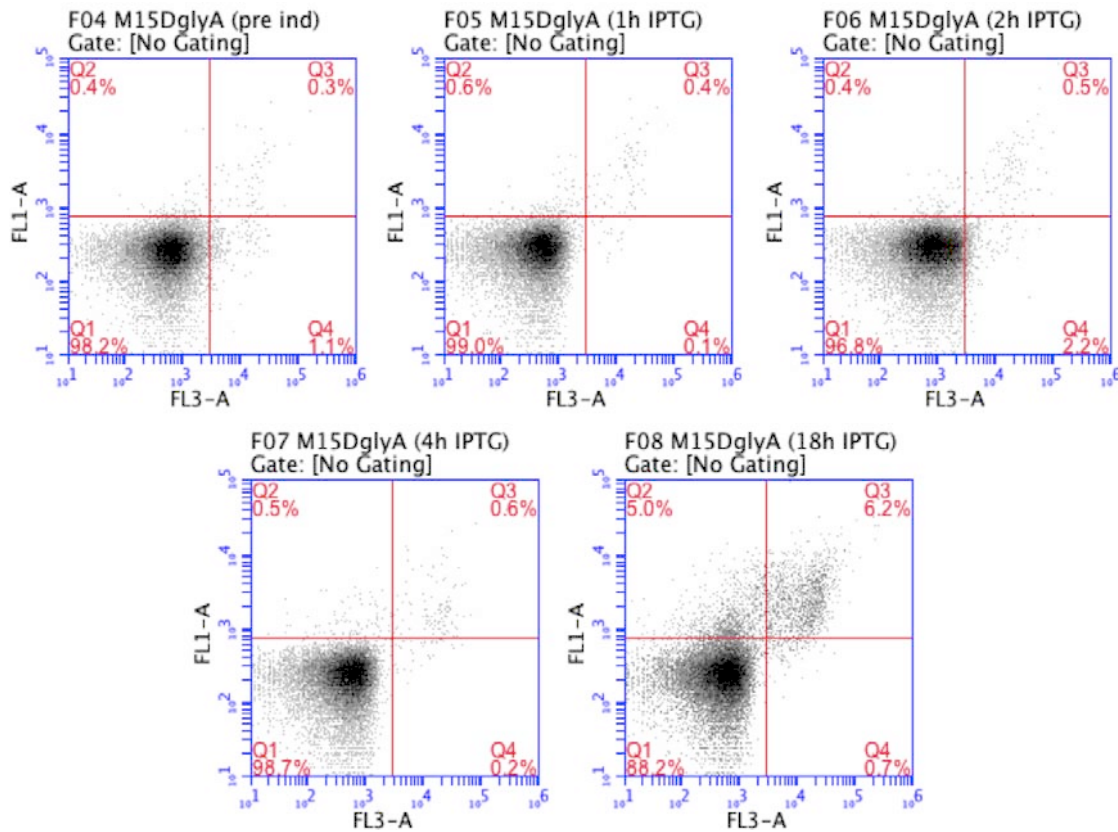


Figure 4.13 FL1-A (BOX fluorescence) versus FL3-A (P.I. fluorescence) plots of cells of *E. coli* M15ΔglyA[pREP4] shake flask cultures in DM medium plus antibiotic with a 1 mM pulse of IPTG. Pre-induction (pre ind), and 1 h, 2 h, 4 h and 18 h post-induction plots are represented. Samples were co-stained with both BOX/P.I. (See Materials and Methods section 3.7.4.2). The numbers in each quadrant show percentage of bacteria population in each subgroup.

Finally, the M15ΔglyA[pREP4] strain was cultivated in two additional conditions, that is 70 μM IPTG induction and 70 mM IPTG induction and absence of antibiotics in the growth medium. Flow cytometry analyses of these cultures were compared with the corresponding analyses performed under the reference growth condition (1mM IPTG). The results are shown in Figure 4.14. Notably, induced cultures in the presence of antibiotic selection pressure present the same amount of live, damaged, and dead cells, both at the 4 and 18 hours after induction, irrespectively of the amount of IPTG

used (Figure 4.14A, B). Strikingly, cell populations of the cultures without antibiotic addition present the same distribution at 4 h of induction but a 1.65-fold increase of dead cells after 18 h of induction (Figure 4.14C). This might reflect increased plasmid loss levels (i.e. leading to reduced cell viability counts in minimal medium) when the plasmid maintenance system relies exclusively on the auxotrophic selection marker.

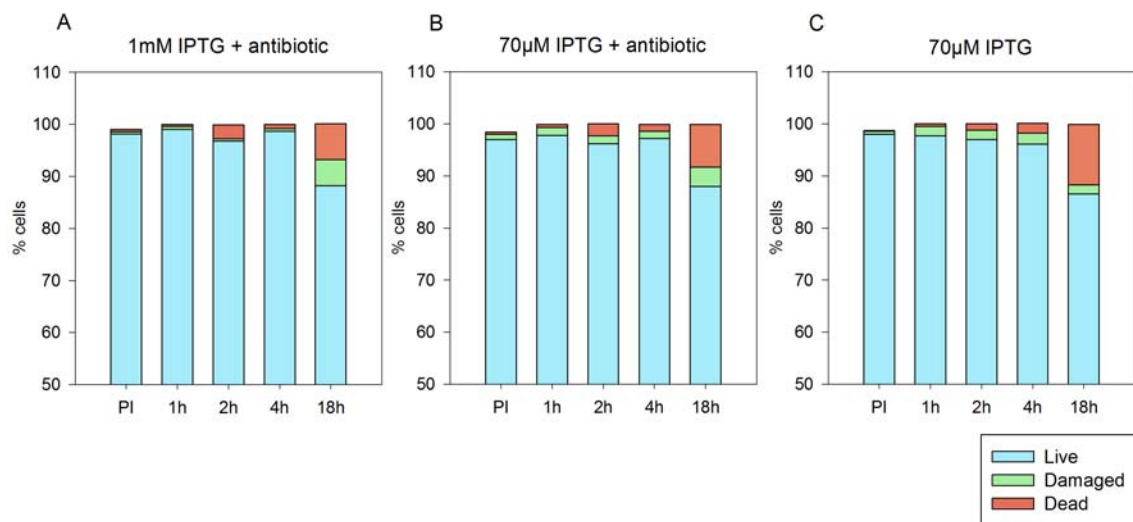


Figure 4.14 Comparison of the three experimental conditions in shake flask cultures for *E. coli* M15Δ*glyA*[pREP4] strain. PI, pre-induction samples; 1 h, 2 h, 4 h and 18 h represent the time after induction. A) DM + antibiotic and 1mM IPTG; B) DM + antibiotic and 70 µM IPTG; C) DM without antibiotic and 70 µM IPTG. Live cells are represented in light blue, damaged cells in green and dead cells in orange.

4.1.3 Comparison of the shake flasks cultures between M15[pREP4] and M15ΔglyA [pREP4] strains

The comparison between the reference M15[pREP4] and the M15ΔglyA[pREP4] *E. coli* strains demonstrated that the later presents slightly lower specific growth rate, decreasing from $0.49 \pm 0.01 \text{ h}^{-1}$, of the reference strain, to $0.44 \pm 0.01 \text{ h}^{-1}$ of the M15ΔglyA[pREP4] strain (Table 4.3). Besides, the reduction of the μ_{max} brings to a decrease of the $q_{\text{s exp}}$, from 1.41 ± 0.05 to $1.14 \pm 0.01 \text{ gGlc}\cdot\text{g}^{-1}\text{DCW}\cdot\text{h}^{-1}$. This effect may be caused by the increase in the metabolic burden due to the maintenance of the expression vector in the M15ΔglyA strain. The presence of the *glyA* gene in the vector results in a higher load of this gene due to the multiple copies of the plasmid. Furthermore, by comparing the $q_{\text{s ind}}$ values of both strains, it can be clearly seen that the M15ΔglyA[pREP4] strain showed an increase in the specific glucose uptake rate from 0.37 ± 0.04 to $0.50 \pm 0.13 \text{ gGlc}\cdot\text{g}^{-1}\text{DCW}\cdot\text{h}^{-1}$. As a consequence, the M15ΔglyA[pREP4] strain accumulated higher amounts of acetate throughout, reaching a final concentration of $0.54 \pm 0.03 \text{ g}\cdot\text{L}^{-1}$ as it can be observed in Figure 4.9 B. This is coherent with previous studies on acetate under aerobic conditions, pointing at the unbalance between glycolysis rates and the TCA-cycle limited capacity of *E. coli* (Jensen and Carlsen 1990). Furthermore, it has been reported that the recombinant protein production is significantly reduced by acetate accumulation (Lee 1996). Such effect can be observed in this study, were both FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$) and FucA mass ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) decrease more than 50 % when comparing the M15ΔglyA[pREP4] strain to the reference M15[pREP4] (Table 4.3). Noteworthy, the metabolic burden is caused not only due to the overexpression of the protein of interest but also to the expression of

other plasmid-encoded genes, that is, the *glyA* overexpression may also contribute (Glick 1995). In fact, the SHMT production ($\text{mgSHMT}\cdot\text{g}^{-1}\text{DCW}$) increased more than 4.5-fold when moving from the M15[pREP4] to the M15 Δ *glyA*[pREP4] strain (Table 4.1 and 4.2).

Furthermore, thanks to the flow cytometry analysis, we demonstrated that the cells after being induced undergo a changing in their morphology, maybe becoming more elongated. This change could bring to overestimate the optical density values ($\text{OD}_{600\text{nm}}$). Moreover, comparing two different inducer concentrations in the shake flasks culture, we proved that the reduction of the IPTG pulse leads to an improvement of the physiological state of the M15[pREP4] cells (Figure 4.7), although for the auxotrophic strain this doesn't affect so much (Figure 4.14).

Table 4.3 μ_{max} (h^{-1}), maximum FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$), maximum FucA mass ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$), substrate uptake rate ($\text{gGlc}\cdot\text{gDCW}^{-1}\cdot\text{h}^{-1}$) along the induction phase for M15[pREP4] and M15 Δ *glyA*[pREP4] strains in shake flasks cultures.

<i>E. coli</i> strains	μ_{max} (h^{-1})	FucA activity ($\text{AU}\cdot\text{gDCW}^{-1}$)	FucA mass ($\text{mg}\cdot\text{gDCW}^{-1}$)	qS exp ($\text{gGlc}\cdot\text{gDCW}^{-1}\cdot\text{h}^{-1}$)	qS ind ($\text{gGlc}\cdot\text{gDCW}^{-1}\cdot\text{h}^{-1}$)
M15[pREP4]	0.49 ± 0.02	721 ± 82	181 ± 5	1.41 ± 0.05	0.37 ± 0.04
M15 Δ <i>glyA</i> [pREP4]	0.44 ± 0.01	291 ± 24	67 ± 37	1.14 ± 0.01	0.50 ± 0.13

4.2 Study of growth and FucA production in fed-batch cultures

In previous studies, our research group designed a minimal balanced media and successfully was used in *E. coli* fed-batch cultures with high cell density. An exponential feed profile was used to control the specific growth rate by limiting the carbon source reducing acetate production while maintaining appropriate levels of the other nutrients.

Thus, in order to improve the understanding of the behavior of the cells during recombinant protein production, fed-batch cultures have been carried out for both the M15[pREP4] and M15 Δ glyA[pREP4] strains, in defined medium with glucose as sole carbon source.

The batch phase lasted until glucose was totally consumed. Substrate feeding was performed following a predetermined exponential function calculated assuming a constant apparent biomass-substrate yield with a constant specific growth rate of 0.22 h⁻¹.

Glucose was maintained to value close to zero along the whole fed-batch and induction periods by off-line measuring glucose concentration and by stopping the feed supply manually. Glucose accumulation has to be eluded in order to avoid proteolysis, as reported in literature (Ruiz et al. 2011), and also to avoid metabolism overflow (i.e. acetate accumulation).

Once reached the desired biomass (20 gDCW·L⁻¹) a single pulse of IPTG at a final concentration of 70 μ M was added to the bioreactor in order to induce the overexpression of the recombinant protein. During the induction phase, plasmid

stability was monitored in order to evaluate the impact of FucA overexpression on plasmid segregation. Indeed, during production with recombinant microorganisms, one of the most important problems is plasmid instability, a tendency of the transformed cells to lose their engineered vectors. This loss can result in a significant negative impact on the production of the recombinant protein of interest.

4.2.1 M15[pREP4] strain

The cultivation profile of the reference strain is summarized in Figure 4.15. As it can be seen, the concentration of the biomass progressively grew till 2.5 h of induction and then stopped around 4 h after induction, because the cells began to be unhealthy. The final biomass was experimentally determined as $40.1 \text{ gDCW}\cdot\text{L}^{-1}$. Glucose and acetate began to accumulate in the medium up to 5.0 and $0.8 \text{ g}\cdot\text{L}^{-1}$, respectively, after 2.5 h of induction, without arriving at inhibitory levels (Figure 4.15).

It is well known that a high-energy demand for the synthesis of the recombinant protein can decrease the observed biomass yield coefficient. Besides, a higher energy demand increases the catabolic fluxes for energy generation and reduces the material available for biosynthesis and cell growth, i.e. negatively affecting the growth rate (Hoffmann, Weber, and Rinas 2002)(Weber et al. 2002). Additionally, metabolic adaptation to the increased energy demands could be the cause of the increase in acetate formation. Moreover, since the set values of μ 0.22 h^{-1} and $Y_{X/S}$ were not modified during the induction phase, the addition program assumed these parameters constant even though they changed. Through this phase, more substrate than required

was consequently added and glucose accumulated. In order to avoid this accumulation the addition of glucose was stopped manually.

The evolution profiles of the FucA production are shown in Figure 4.15. FucA levels reached a maximum value of $240 \pm 16 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$, with an activity of $1214 \pm 47 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$, that correspond of a biomass of $41.4 \pm 0.3 \text{ gDCW}\cdot\text{L}^{-1}$. Moreover, the evolution of enzyme concentration can be followed in Figure 4.16 and in Table 4.4. The FucA production, in terms of both specific activity levels ($\text{AU}\cdot\text{gDCW}^{-1}$) and yield ($\text{mg}\cdot\text{gDCW}^{-1}$), decreased when glucose is accumulated in the media. This event can be attributed to the presence of proteases (Boada 2009).

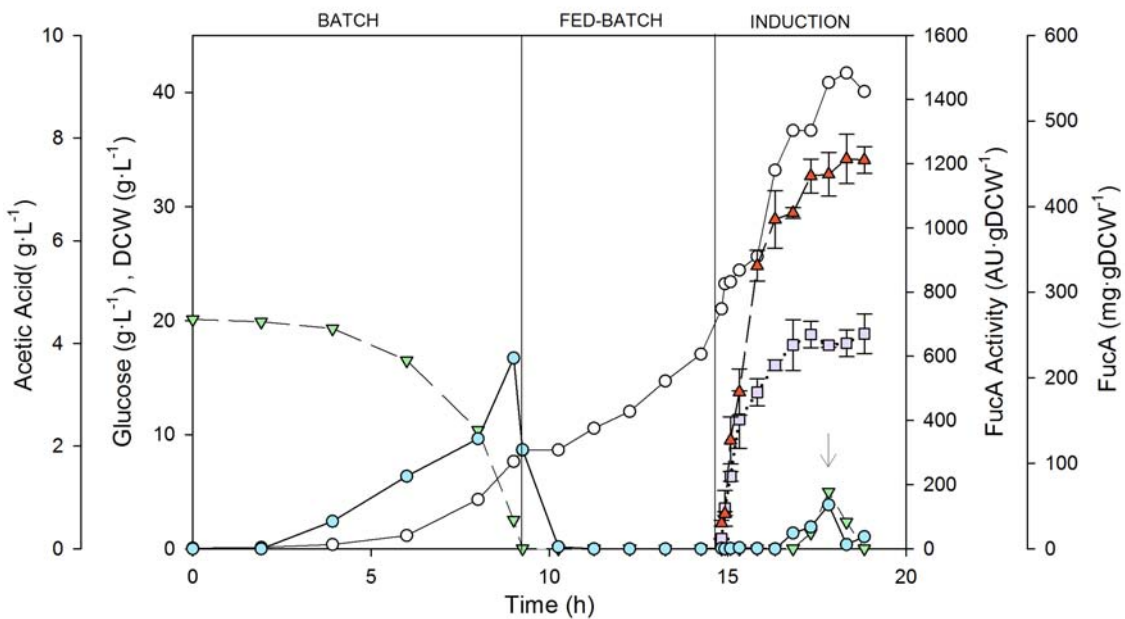


Figure 4.15 M15[pREP4] fed-batch culture: [IPTG], $70 \mu\text{M}$; X_{ind} , $20 \text{ g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . Schematic representation of (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (▼) Glucose concentration ($\text{g}\cdot\text{L}^{-1}$), (▲) FucA activity ($\text{AU}\cdot\text{gDCW}^{-1}$), (■) specific mass ($\text{mgFucA}\cdot\text{gDCW}^{-1}$) and (○) Acetic Acid ($\text{g}\cdot\text{L}^{-1}$) along time. Batch, fed-batch and induction phases are indicated. The arrow indicates the stop of the feeding.

Briefly, proteolysis is one of the mechanisms used by cells to get rid of misfolded, incorrectly synthesized or unwanted proteins. *E. coli* contains several proteases situated in various intracellular compartments, including the cytoplasm, periplasm, inner membrane, and outer membrane. These enzymes are the majors responsible for the inactivation of the protein. The mechanism of proteolysis in *E. coli* consists of a first step comprising partial cleavage of the protein by energy-dependent proteases followed by a second step cleavage of the protein where partially cleaved proteins are degraded into amino acids by energy independent proteases (Boada 2009). Besides, since this decrease in the specific activity levels does not correlate with the profiles of soluble product, as analyzed by SDS-PAGE (Figure 4.16), our results suggested that the proteolysis is just partial.

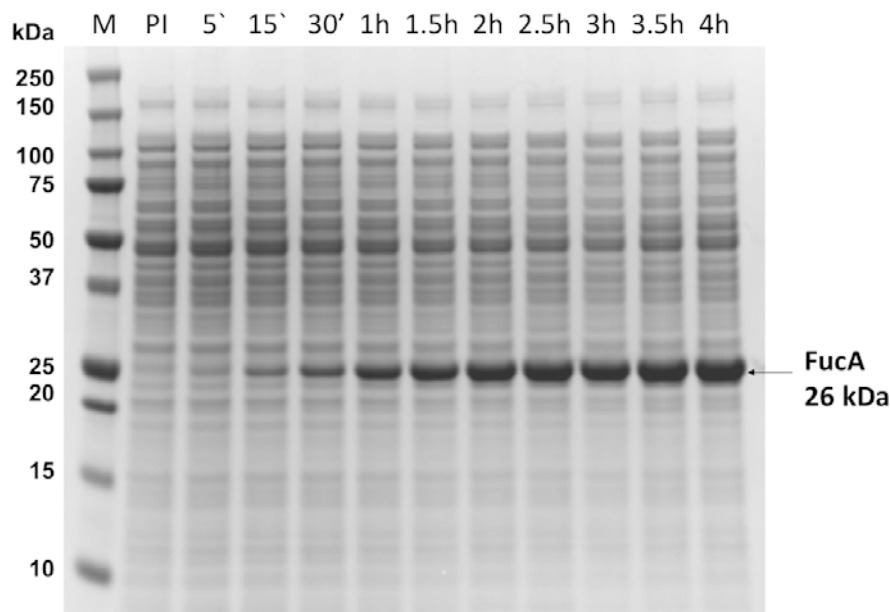


Figure 4.16 SDS-PAGE of fed-batch cultures samples of the M15[pREP4] strain: [IPTG], 70 μM ; X_{ind} , 20 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . Lane M: molecular weight marker; PI, pre-induction; from 5' to 4.5 h correspond to the time after induction. The bands corresponding to the 26 kDa FucA and the 46 kDa SHMT are indicated.

Table 4.4 Values of SHMT and FucA protein ($\text{mg}\cdot\text{gDCW}^{-1}$) along the induction phase of fed-batch culture's samples of the M15[pREP4] strain. PI, represents the pre-induction, while from 5' to 4 h are the samples along the induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)	FucA mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	39 ± 3	12 ± 1
5'	32 ± 2	47 ± 20
15'	29 ± 3	85 ± 6
30'	27 ± 3	151 ± 33
1 h	28 ± 2	183 ± 16
1.5 h	29 ± 2	215 ± 6
2 h	25 ± 1	238 ± 30
2.5 h	24 ± 2	250 ± 15
3 h	24 ± 3	238 ± 2
3.5 h	23 ± 4	240 ± 16
4 h	23 ± 6	251 ± 23

4.2.1.1 Off-gas analysis

To perform off-gas analysis, new fed-batch fermentations, with similar operation conditions, were performed for duplicate in 5 L fermenters with an initial volume of 1.5 L (See Materials and Methods section 3.6.2.2). Cultivation data of M15[pREP4] strain are shown in Figure 4.17. An initial glucose concentration of only $10 \text{ g}\cdot\text{L}^{-1}$ secured that the oxygen transfer rate was sufficient to support fully aerobic growth during the entire experiments.

The partial pressure of oxygen ($p\text{O}_2$) decreases along the batch phase exponentially up to the set point of 50 %, where an agitation controller is started in order to maintain the level.

The increase of both pO_2 and pH values, indicate the complete consumption of the $10 \text{ g}\cdot\text{L}^{-1}$ of initial glucose and the end of the batch phase (Figure 4.17), (Soosan, Chang, and Pan 2009). Once the glucose in the medium is finished, the bacteria shifts the so called “acetate switch”, where the acetate is assimilated because it is the only carbon source in the medium. Thus the pH of the culture increased as a result of the depletion of the acetate from the medium (Figure 4.17). The end of the batch can be also followed in Figure 4.18 where both the oxygen (O_2) and carbon dioxide (CO_2) curves, after exponential profiles, suddenly increase and decrease, respectively.

After 12 hours, when the initial glucose was exhausted and the bacteria consumed all the acetate in the media, continuous feeding of a highly concentrated glucose solution was started, indicated in the Figure 4.17 as fed-batch phase. The fixed growth rate of 0.22 h^{-1} was maintained until the end of the fermentation; when the cells reached and $6 \text{ g}\cdot\text{DCWL}^{-1}$ IPTG was added to the culture in order to induce the recombinant protein production. The induction phase last about 4.5 h where the final biomass was estimated as $10 \text{ gDCW}\cdot\text{L}^{-1}$. Due to the lack of a direct glucose measurement, in order to know when the glucose is accumulated into the medium, the feeding is turned off along the induction phase. Thus, if the pH increases, it indicates that cells are using the acetate as carbon source, i.e. meaning no glucose is accumulated. The points where the feeding is stopped due to glucose accumulation are indicated in Figure 4.17. These correspond to 2.5 h of induction, followed by a little increase of the pH due to acetate switch. The change of metabolism from glucose to the acetate consumption can be

observed in Figure 4.18 where, like at the end of the batch, both the O_2 and the CO_2 curves undergo a perturbation.

The respiratory quotient (RQ) curve in Figure 4.18 represents the molar ratio of CO_2 and O_2 during fermentation. The RQ value along the aerobic metabolism becomes 1.0 according to the chemical equation:

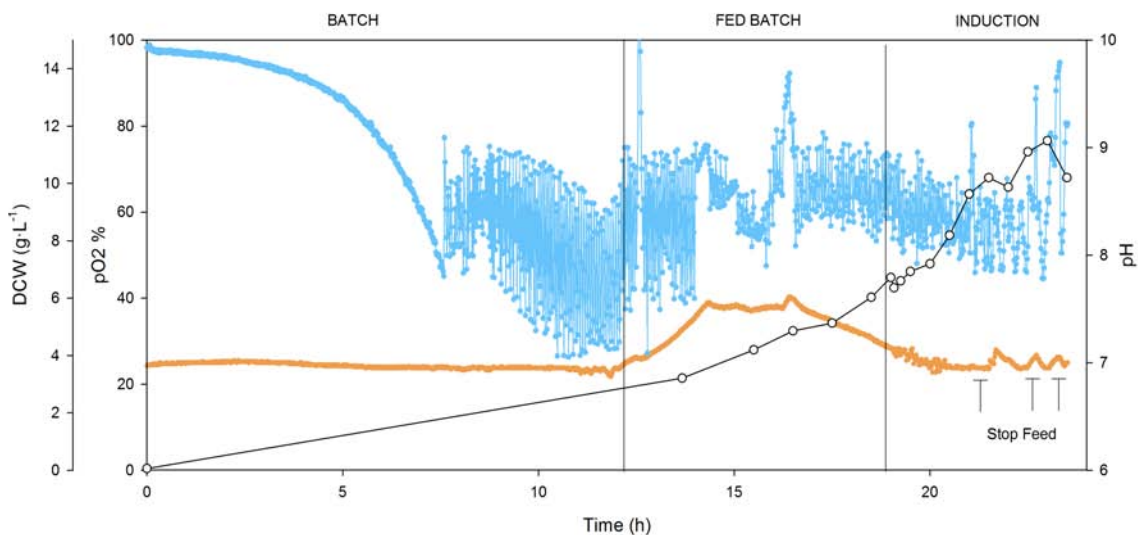


Figure 4.17 *E. coli* M15[pREP4] fed-batch culture: [IPTG], 70 μM ; X_{ind} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . pO_2 , in light blue, was set at 50 % and controlled by a gas flow rate of 3 $\text{L}\cdot\text{min}^{-1}$ of air and agitation. The pH, in orange, was set at 7 and maintained by adding 15 % (v/v) NH_4OH solution. (c) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$) along time. Batch, fed-batch and induction phases are indicated. The arrows indicate the stop of the feeding.

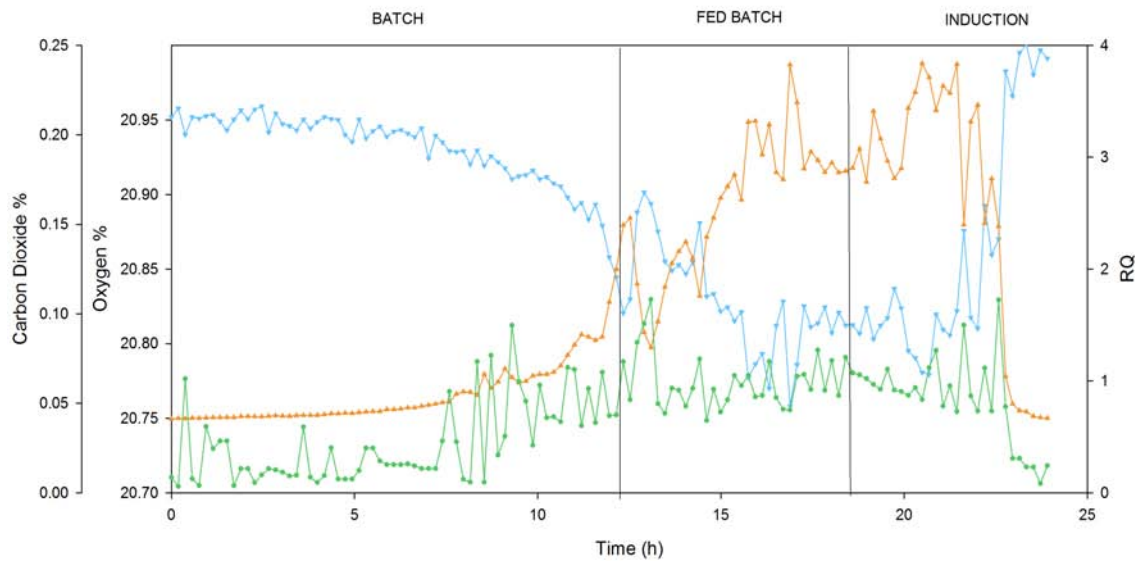


Figure 4.18 Gas-MS data of M15[pREP4] fed-batch culture: [IPTG], 70 μM ; X_{ind} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . % of oxygen in light blue, % of carbon dioxide in orange and the respiratory quotient (RQ) in green. Batch, fed-batch and induction phases are indicated.

4.2.1.2 Flow cytometry analysis of M15[pREP4] fed-batch cultures

Cell samples were taken along the fed-batch culture and were analysed by flow cytometry (Material and Methods section 3.7.5). Figure 4.19 shows the results of live, damaged and dead cells determined by BOX/P.I. co-staining.

As it can be seen, from the first batch sample to the end of the batch the proportion of live bacteria decreased by 14 %, while, as a consequence, the amount of damaged and dead cells increase. This decrease is probably due to unhealthy state of the cells caused by FucA overexpression.

Along the fed-batch the % of the damaged cells returns to the initial values, while more than 10 % of cells result dead.

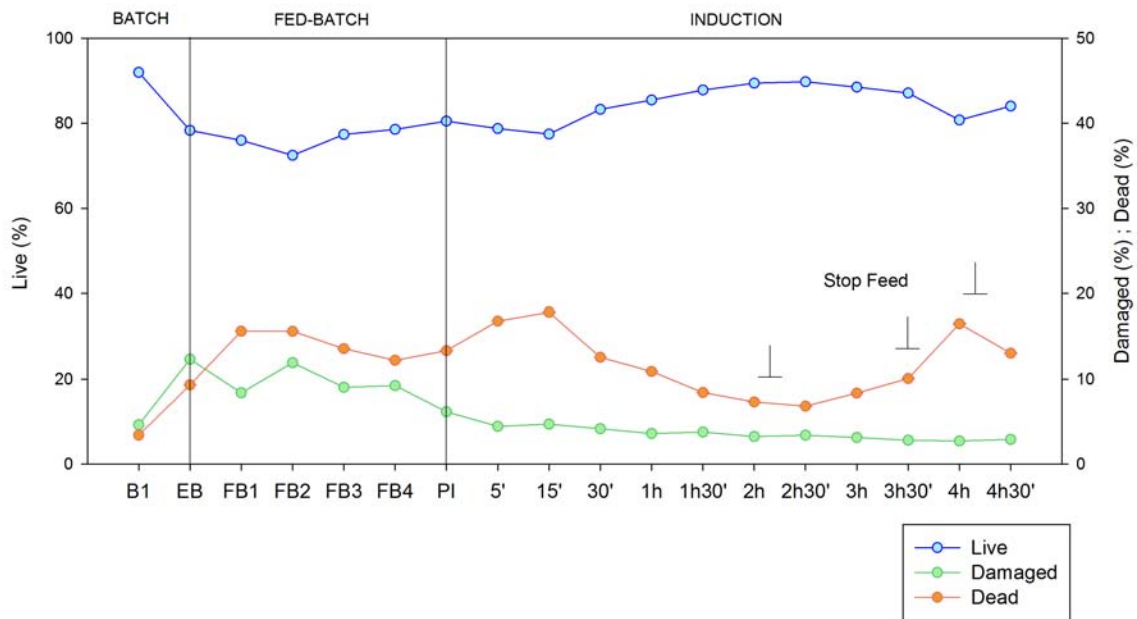


Figure 4.19 FCM analysis of *E. coli* M15[pREP4] fed-batch: [IPTG], 70 μM ; X_{indr} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . FCM data represents live (blue), damaged (green) or dead (orange) cells determined by BOX/P.I. co-staining (See Materials and Methods section 3.7.4.2). B1 and EB, first and the last samples of the batch, respectively; FB1, FB2, FB3 and FB4, four samples of the fed-batch phase; PI, pre-induction and from 5' to 4h30' different induction samples. Batch, Fed-batch and induction phases are indicated. The arrows indicate the stop of the feeding.

The curve representing the dead cells reaches the maximum value with the sample number 15' that corresponding to 15 minutes after the pulse of IPTG (Figure 4.19). This reflects the stress caused to the cells due to the induction. Besides, the slightly increase of the % of CR^+ population, representing the % of inclusion bodies present inside the cells, indicates that the induction of the recombinant protein expression system affects the cells, too (Figure 4.20).

It is well known that the production of a foreign protein causes an additional stress for the host strain and leads to a decrease in the overall cell fitness. At the end of the induction phase, the dead population corresponds to about 16 %. These results are

concordant with those presented previously, where the end of the fermentation process was characterized by a decrease in the DCW ($\text{g}\cdot\text{L}^{-1}$) and the accumulation of the glucose.

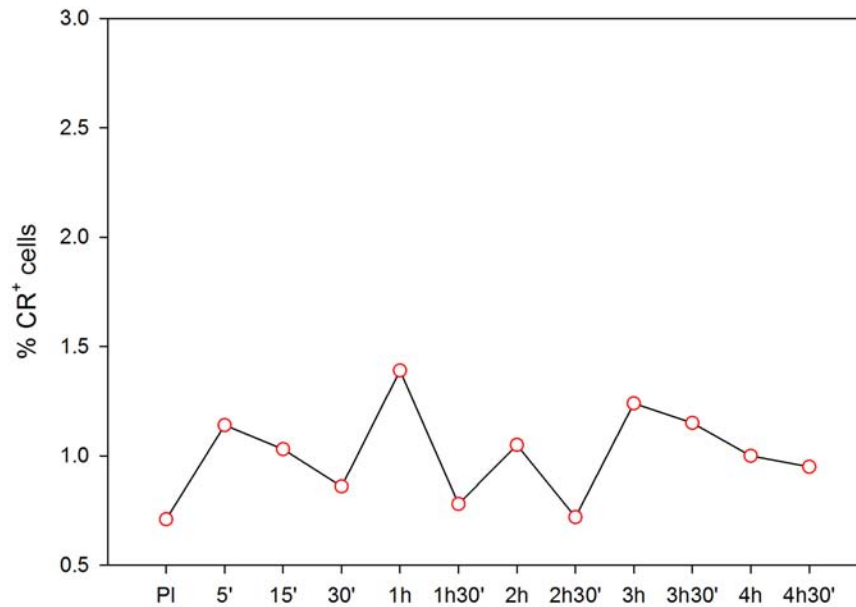


Figure 4.20 FCM analysis of *E. Coli* M15[pREP4] fed-batch culture: [IPTG], 70 μM ; X_{ind} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . % CR⁺ cells from FCM analysis. Samples stained with CR for 1 hour (See Materials and Methods section 3.7.4.3). PI, pre-induction; 5' to 4h30', induction samples.

4.2.2 M15 Δ glyA[pREP4] strain

The same μ -limiting cultivation conditions were used for the auxotrophic strain *E. coli* M15 Δ glyA[pREP4] in fed-batch cultures. The results are shown in Figure 4.21. Again, once the 20 $\text{g}\cdot\text{L}^{-1}$ of glucose substrate was consumed, the fed-batch started supplying the feed solution in order to allow a constant specific growth rate of 0.22 h^{-1} ; and again, the overexpression of FucA was induced at 20 $\text{gDCW}\cdot\text{L}^{-1}$ with a 70 μM pulse of

IPTG. After the induction the biomass concentration progressively slowed down, ending with a final value of $39 \text{ gDCW}\cdot\text{L}^{-1}$.

As observed previously, glucose and acetate began to accumulate in the medium after 2 h of induction, reaching as maximum value $2.5 \text{ g}\cdot\text{L}^{-1}$ and $0.5 \text{ g}\cdot\text{L}^{-1}$ respectively (Figure 4.21).

FucA production profile are shown in Figure 4.21, where the maximum values are $165 \pm 13 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$, with an activity of $806 \pm 12 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$, that correspond of a biomass of $39.6 \pm 0.6 \text{ gDCW}\cdot\text{L}^{-1}$. Moreover, the evolution of FucA concentration together with the production of SHMT protein can be followed in Figure 4.22. As it can be clearly seen, there is a constitutive overexpression of the SHMT protein both in the pre induction and along the induction phase (Figure 4.22 and Table 4.4). During the induction phase the expression system is focused on the recombinant protein production, meaning that most part of the energy-demand of the cells is used for *fucA* expression. This can be found in the decrease of the $\text{mgSHMT}\cdot\text{gDCW}^{-1}$ and the corresponding increase of the amount of the FucA protein (Table 4.4). As explained before, this metabolic burden forced to the cells could cause the glucose and acetate accumulation.

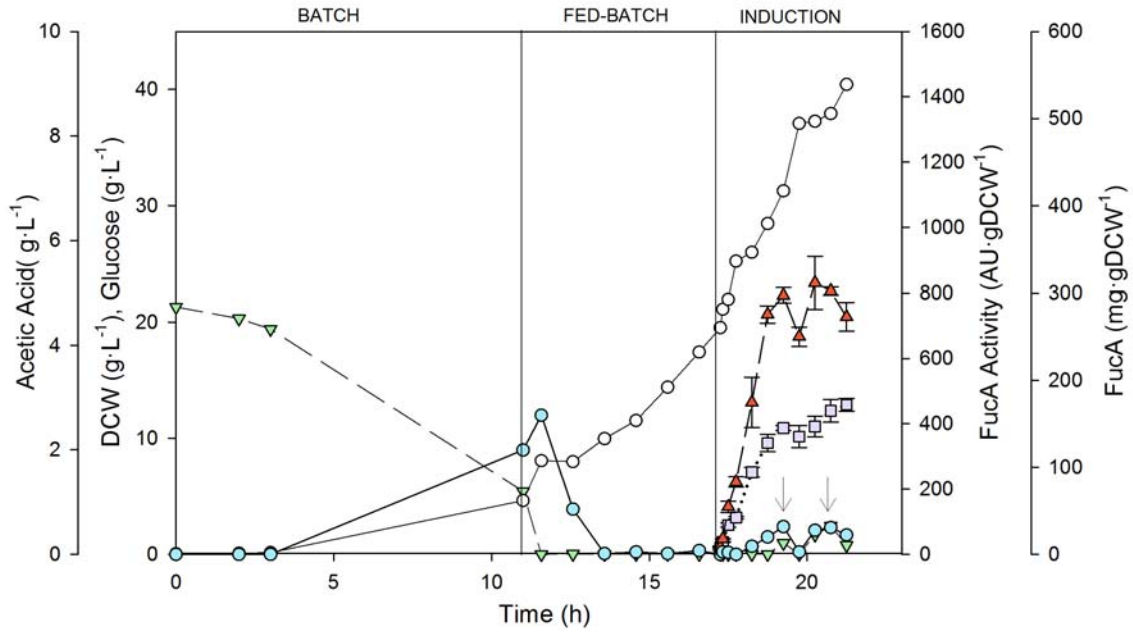


Figure 4.21 M15Δ*glyA*[pREP4] fed-batch culture: [IPTG], 70 μM; X_{ind} , 20 g·L⁻¹; μ 0.22 h⁻¹. Schematic representation of (●) Biomass DCW (g·L⁻¹), (▽) Glucose concentration (g·L⁻¹), (▲) FucA activity (AU·gDCW⁻¹), (◻) specific mass (mgFucA·gDCW⁻¹) and (◐) Acetic Acid (g·L⁻¹) along time. Batch, fed-batch and induction phases are indicated. The arrows indicate the stop of the feeding.

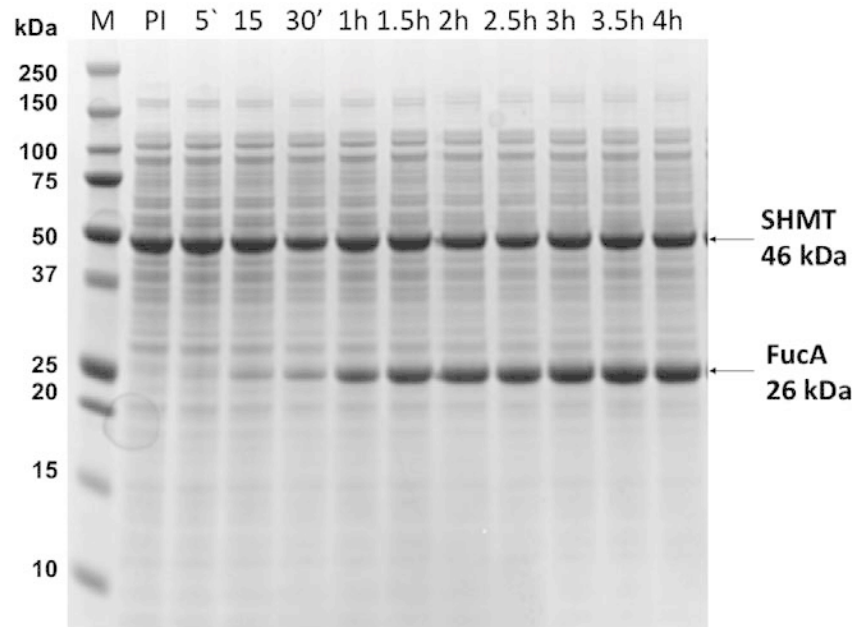


Figure 4.22 SDS-PAGE of fed-batch culture's samples of the M15Δ*glyA*[pREP4] strain: [IPTG], 70 μM; X_{ind} , 20 g·L⁻¹; μ 0.22 h⁻¹. Lane M: molecular weight marker; PI, pre-induction; from 5' to 4.5h correspond to the time after induction. The 26 kDa FucA and the 46 kDa SHMT are indicated in the figure.

Table 4.5 Values of SHMT and FucA protein ($\text{mg}\cdot\text{gDCW}^{-1}$) along the induction phase of fed-batch culture's samples of the M15 Δ glyA[pREP4] strain. PI, represents the pre-induction, while from 5' to 4 h are the samples along the induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)	FucA mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	145 \pm 2	7 \pm 1
5'	163 \pm 2	14 \pm 1
15'	150 \pm 2	34 \pm 2
30'	106 \pm 2	42 \pm 1
1 h	129 \pm 1	94 \pm 6
1.5 h	141 \pm 3	128 \pm 10
2 h	131 \pm 7	149 \pm 3
2.5 h	103 \pm 5	135 \pm 12
3 h	113 \pm 7	147 \pm 12
3.5 h	124 \pm 9	165 \pm 13
4 h	132 \pm 4	172 \pm 7

4.2.2.1 Off-gas analysis

Like for the reference strain M15[pREP4], fed-batch fermentation in 5 L fermenters (Electrolab) of M15 Δ glyA[pREP4] strain were performed for duplicate with similar conditions of that used with the Biostat B reactor. The results are shown in Figure 4.23. Again, initial glucose concentration of 10 g L^{-1} was completely consumed by the bacteria along the batch phase. The partial pressure of oxygen (pO_2) decreased along the batch phase exponentially until reaching the fixed set point at 50 %. After about 13 h the end batch phase can be identified by the increase of both pO_2 and pH values. Like for the reference strain, along the batch phase the cells produce acetate in order to balance glycolysis overflow. After an hour of acetate switch, where the pH reaches a value of 7.5, the feed started to enter the reactor. Compare to the reference strain

the accumulation of glucose appears more or less at the same time, as indicated in Figure 4.23. Besides, after the second stop of the feeding it can be seen how the pH value does not change, suggesting that glucose is still present in the media. This glucose accumulation indicates a cellular stress due to the metabolic burden, also indicated by the decrease of the DCW ($\text{g}\cdot\text{L}^{-1}$) at the end of the fermentation. Figure 4.24 represents the results of the off-gas MS analysis. The values of both oxygen and carbon dioxide along the batch phase present exponential profiles. The end of initial glucose along the batch phase is indicated by a slight perturbation of both the O_2 and CO_2 values. As for the reference strain, RQ value, of the M15 Δ g/yA[pREP4] strain, increases along the batch phase and then is maintained at 1. This means that when the sugar-feeding rate is controlled, there is no production of intermediate products.

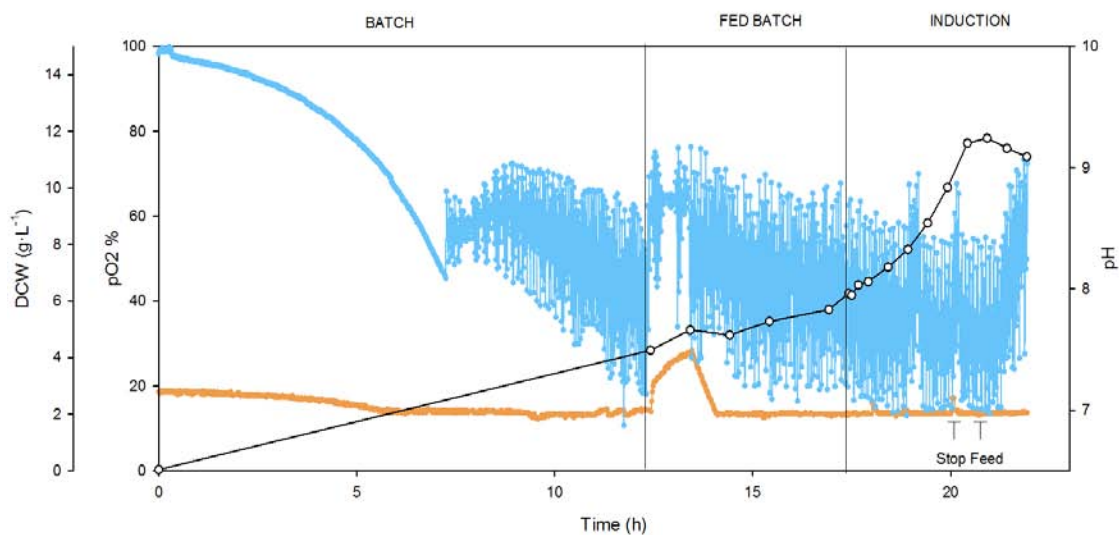


Figure 4.23 M15 Δ g/yA[pREP4] fed-batch culture: [IPTG], 70 μM ; X_{ind} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . pO_2 , in light blue, was set at 50 % and controlled by a gas flow rate of 3 $\text{L}\cdot\text{min}^{-1}$ of air and agitation. The pH, in orange, was set at 7 and maintained by adding 15 % (v/v) NH_4OH solution. (c) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$) along time. Batch, fed-batch and induction phases are indicated. The arrow indicates the stop of the feeding.

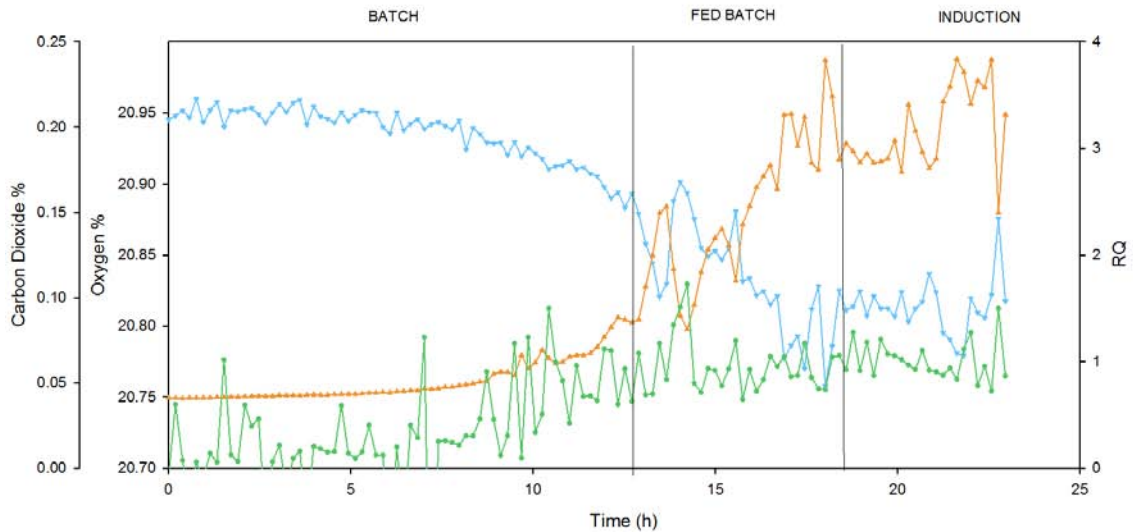


Figure 4.24 Gas-MS data of *M15ΔglyA[pREP4]* fed-batch culture: [IPTG], 70 μM ; X_{ind} , 6 $\text{g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . % of oxygen in light blue, % of carbon dioxide in orange and the respiratory quotient (RQ) in green. Batch, fed-batch and induction phases are indicated.

4.2.2.2 Flow cytometry analysis of *M15ΔglyA[pREP4]* fed-batch cultures

AS for the reference strain, samples along the fed-batch culture were analysed using a flow cytometer with BOX/P.I. co-staining. Contrary to the reference strain, the *M15ΔglyA[pREP4]* cells remained homogeneous and healthy along the induction phase. However, the population at the end of the batch presented 17 % of dead cells (Figure 4.25). As it can be seen in Figure 4.25, this increase corresponds to the point of acetate accumulation as indicated with arrows in the figure. Besides, in order to test the possible formation of inclusion bodies FCM analysis using CR as dyes were performed for the pre induction and induction samples (Figure 4.26). The end of the fed-batch culture is marked by a clearly increase of the % of inclusion bodies, reaching 3.0 %, suggesting an increase in the cellular stress.

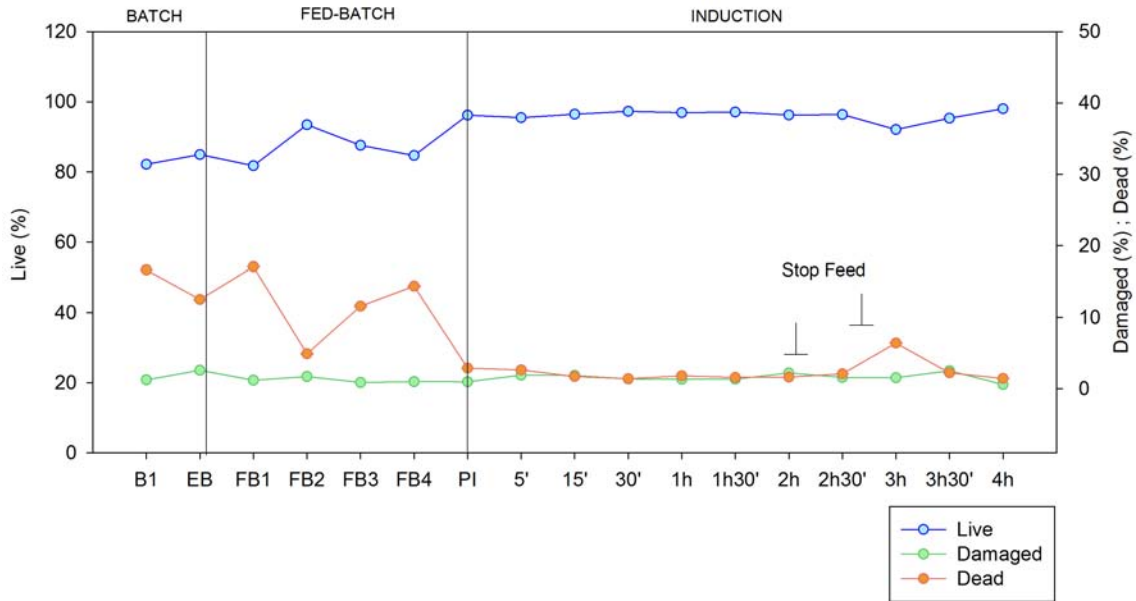


Figure 4.25 FCM analysis of *E. Coli* M15ΔglyA[pREP4] fed-batch culture: [IPTG], 70 μM; X_{ind} , 6 g·L⁻¹; μ 0.22 h⁻¹. FCM data represents live (blue), damaged (green) or dead (orange) cells determined by BOX/P.I. co-staining (See Materials and Methods section 3.7.4.2). B1 and EB, first and the last samples of the batch, respectively; FB1, FB2, FB3 and F4, four samples of the fed-batch phase; PI, pre-induction and 5' to 4h, different induction samples. Batch, fed-batch and induction phases are indicated. The arrow indicates the stop of the feeding.

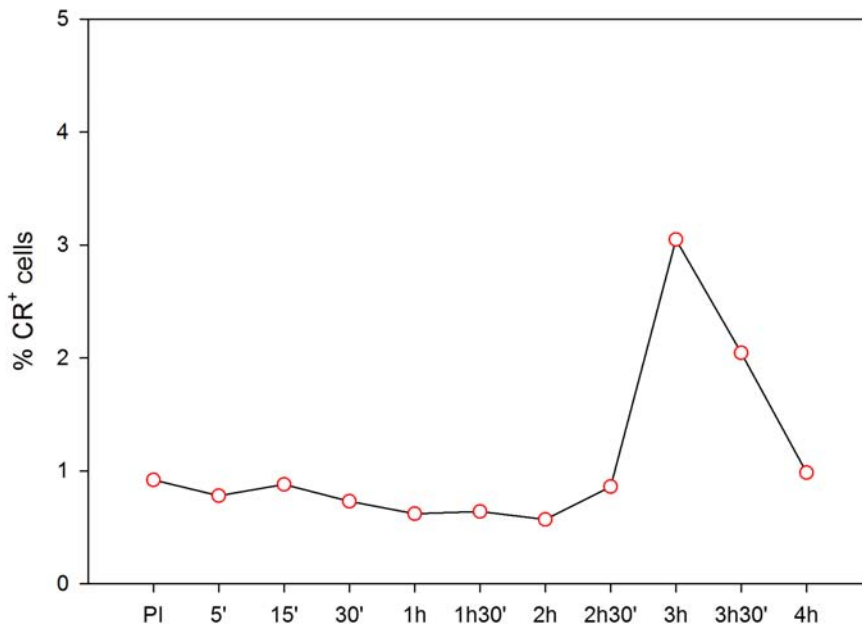


Figure 4.26 FCM analysis of *E. Coli* M15ΔglyA[pREP4] fed-batch culture. % CR⁺ cells from FCM analysis. Samples stained with CR for 1 hour (See Materials and Methods section 3.7.4.3). PI, pre-induction; 5' to 4h, induction samples.

4.2.3 Comparison of the fed-batch cultures between M15[pREP4] and M15ΔglyA [pREP4] strains

The comparison between the reference M15[pREP4] and the M15ΔglyA[pREP4] in fed-batch cultures demonstrated, again, that the maximum production of 240 ± 16 mgFucA·g⁻¹DCW, with an activity of 1214 ± 47 AU·g⁻¹DCW, in the reference strain decrease to 165 ± 13 mgFucA·g⁻¹DCW and 806 ± 12 AU·g⁻¹DCW respectively, in the M15ΔglyA strain (Table 4.6). As explained before, these results could be related to the concomitant SHMT overexpression observed after induction in the M15ΔglyA strain. In fact, significant amounts of SHMT, accumulated as soluble protein in the cytoplasm, increased more than 4.5-fold comparing the M15ΔglyA[pREP4] with the M15[pREP4] strains (Table 4.4, 4.5).

Table 4.6 Maximum FucA activity (AU·g⁻¹DCW), maximum FucA mass (mg·g⁻¹DCW) and % of inclusion bodies, representing the % of CR⁺ cells, for M15[pREP4] and M15ΔglyA[pREP4] fed-batch cultures.

E. coli strains	FucA activity (AU·gDCW ⁻¹)	FucA mass (mg·gDCW ⁻¹)	Inclusion bodies (%)
M15[pREP4]	1214 ± 47	240 ± 16	1.4
M15ΔglyA[pREP4]	806 ± 12	165 ± 13	3.0

Moreover, in both the reference M15[pREP4] and M15ΔglyA[pREP4] strains, along the batch phase the cells produce acetate in order to balance glycolysis overflow. Besides, the accumulation of glucose that indicates a cellular stress due to metabolic burden appears more or less at the same time in both strains.

Lastly, compared with the M15[pREP4] strain, M15 Δ *glyA*[pREP4] appears less affected by the recombinant protein expression presenting a more homogeneous and healthy physiological profile, although the percentage of inclusion bodies at the end of the induction in fed-batch cultures is higher (Table 4.6).

4.3 Conclusion

In this Chapter, FucA was expressed in both the reference *E. coli* M15[pREP4] strain and the auxotrophic *E. coli* M15 Δ *glyA*[pREP4] strain. The results presented in this chapter suggested that *glyA* overexpression imposed a significant burden to the metabolism of the host cell, thereby affecting negatively FucA expression levels. Thus, the regulation of the *glyA* expression levels is an important parameter to be taken into account for further improvement of the expression system and optimization of protein yields.

5 Results II: Design, construction and preliminary evaluation of an antibiotic-free single plasmid expression system

In Chapter 4, we have shown that the *glyA*-based antibiotic-free system of plasmid maintenance in *E. coli* is a promising alternative approach for FucA expression, as previously demonstrated for RhuA by Vidal et al. (Vidal et al. 2008)

The presence of a second plasmid (pREP4) with an additional antibiotic resistance marker, carrying the *Lacl* repressor encoding gene, impedes the achievement of an expression system completely devoid from the use of antibiotics.

Moreover, previous results have demonstrated that increased expression levels of plasmid-encoded proteins, in particular the expression levels of the *glyA* gene, might burden the cell's metabolism, leading to a decrease in activity and specific productivity profiles compared to the original system. According to Glick (Glick 1995), protein biosynthesis is an energy intensive process and it is therefore not surprising that the overproduction of one or more foreign proteins encoded within the plasmid DNA may cause plasmid-containing cells to lose a portion of the plasmid DNA; where it is often all or part of the foreign gene of interest that is deleted from the plasmid. For instance, the constitutive expression of plasmid-encoded proteins including the antibiotic resistance gene is one of the major causes of metabolic burden exerted to the host

cells (Bentley et al. 1990)(Lenski, Simpson, and Nguyen 1994). Usually, the marker protein represents up to 20 % of total cellular protein, thereby exceeding the levels that are needed for proper selection and maintenance (Rozkov et al. 2004). In this context, results obtained in the previous Chapter suggest that SHMT expression levels driven from the plasmid-encoded *glyA* gene may cause some metabolic burden to the cell, i.e. having a negative impact on the μ_{\max} and FucA production. In fact, the *glyA* gene encoded in the high-copy plasmid leads to substantially higher amounts of its product (SHMT) accumulated as soluble protein in the cytoplasm, compared to the reference strain containing a single copy of *glyA* in the genome.

In this Chapter, in order to decrease the extent of the metabolic load and to implement our pre-existing strategy for antibiotic resistance-free selection, we focused on the elimination of the pREP4 plasmid. The objectives were to i) obtain an expression system based on a single plasmid and ii) clone the *lacI* gene from the pREP4 plasmid to the pQE- expression vector, thereby obtaining a single-plasmid expression system.

In addition, it is envisaged that further studies to fine-tune the *glyA* and *lacI* expression levels will be needed in order to overcome the limitations mentioned above.

5.1 Curing the cells from the pREP4 plasmid

In this study, serial culture dilutions have been performed in order to cure the *E. coli* M15Δg/yA[pREP4] strain from the pREP4 plasmid. As explained in the Materials and Methods section 3.7.2, the colony that can grow only in LB agar plus ampicillin but not plus kanamycin, means it has lost the pREP4 plasmid. From the 28 single colonies plate (Figure 5.1), all of them could grow in LB agar plates plus ampicillin (Figure 5.1 A) but only the 32 % of them cannot grow in LB agar plates plus kanamycin (Figure 5.1 B). Two clones (No. 1 and 7) were selected and, in order to ensure the complete cure of the pREP4 plasmid were grown in LB liquid culture plus kanamycin. The no-capability of growing in presence of kanamycin confirmed the elimination of the pREP4.

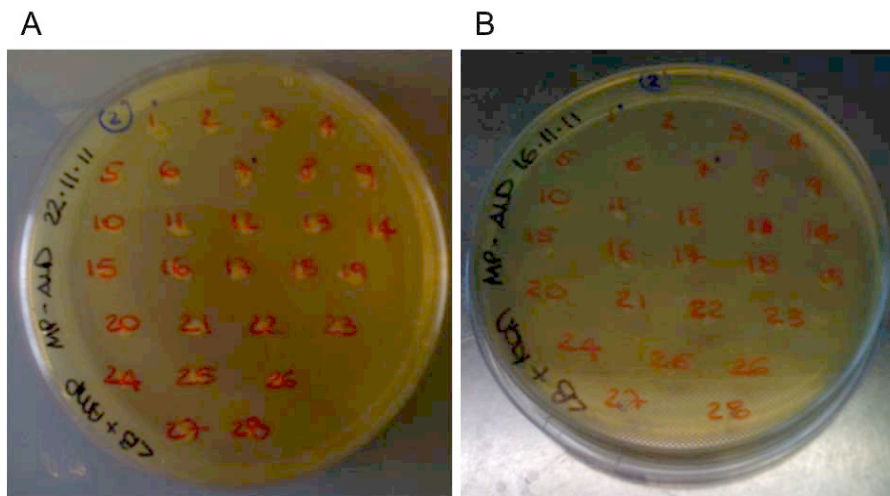


Figure 5.1 Pictures of two LB-agar plates representing 28 single colonies after 8 serial dilutions plated in: A) LB agar plus ampicillin; B) LB agar plus kanamycin. The two selected clones (No 1 and 7) are indicated with blue points.

The two clones of the new strain *E. coli* M15 Δ *glyA* pQE α β FucA, shortly M15 Δ *glyA*[A], were tested for FucA expression in shake-flask cultures in LB medium supplemented just with ampicillin.

As in previous experiments, a 1 mM IPTG pulse was added to the culture in the mid-exponential phase and samples at 1 h and 2 h of induction were harvested for further analysis. An increase in the basal FucA expression was expected, due to the removal of the repressor protein encoded by the *lacI* gene present on the pREP4 plasmid (Gruber et al. 2008). Strikingly, no FucA production was detected in these cultures. Figure 5.2 clearly shows the expression levels of the recombinant protein were below the detection limit, since no band of 26 kDa was present.

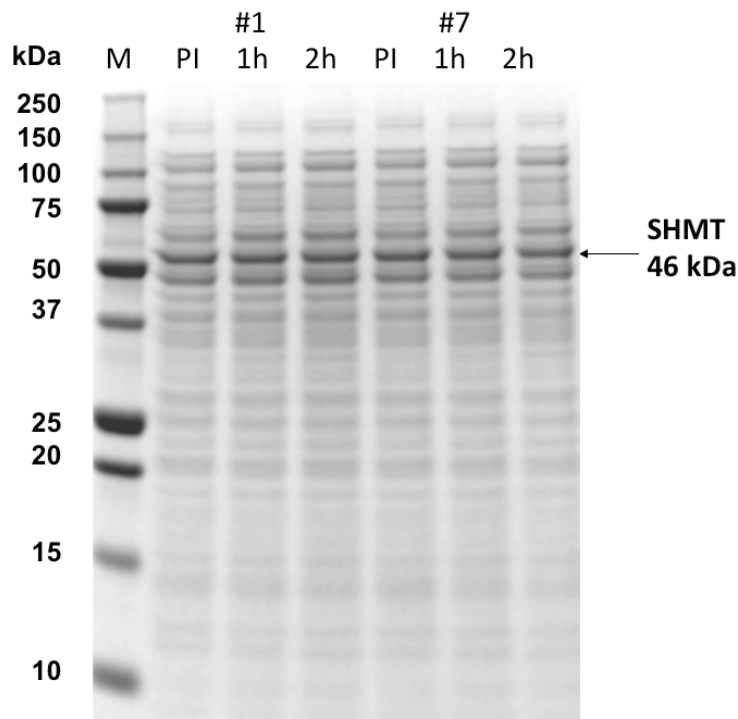


Figure 5.2 SDS-PAGE of shake flasks culture's samples of *E. coli* M15 Δ *glyA*[A] strain in DM medium. Lane M: molecular weight marker; PI, pre-induction; 1 h and 2h correspond to the time after induction. The band corresponding to the 46 kDa SHMT is indicated in the figure. No band of FucA (26 kDa) can be identified.

5.2 Cloning of *lacI* gene into the pQE $\alpha\beta$ FucA vector

To further understand the lack of recombinant FucA expression in the *E. coli* M15 Δ *glyA*[A] strain, the *lacI* gene was amplified from pREP4 and cloned into the pQE $\alpha\beta$ FucA plasmid (Materials and Methods section 3.4.3 and 3.4.4), yielding the M15 Δ *glyA* pQE-*lacI*- $\alpha\beta$ FucA strain, from now on M15 Δ *glyA*[B].

Thus, shake-flask cultures of three transformants were performed and induced with a 1 mM IPTG pulse in order to test FucA production. However, as it can be seen in figure 5.3, FucA expression could not be detected.

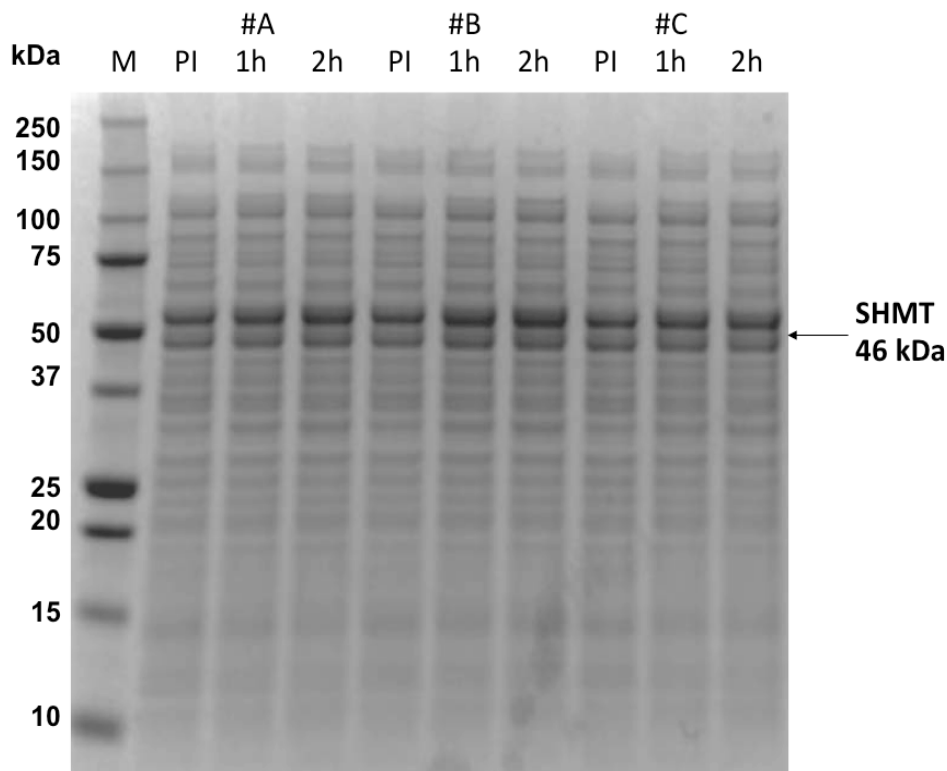


Figure 5.3 SDS-PAGE of shake flasks culture's samples of *E. coli* M15 Δ *glyA*[B] strain. Lane M: molecular weight marker; PI, pre-induction; 1 h and 2h correspond to the time after induction for the three clones A, B and C. The band corresponding to the 46 kDa SHMT is indicated in the figure. No band of FucA (26 kDa) can be identified.

In order to ensure there was no loss of the expression vector from the cells, a plasmid segregational stability assay was carried out at different cultivation times before and after induction by growing M15 Δ *glyA*[B] cells on both LB and LB-Amp agar plates. The experiment confirmed that the cells maintained the expression vector and, consequently, the lack of FucA expression was not the result of plasmid loss (segregational instability), thus pointing at a potential plasmid structural instability.

5.3 Analysis of plasmid structural stability

As mentioned in the Materials and Methods section 3.2.1, the T5 promoter has a double *lacO* region in order to guarantee a strong repression under non-induction conditions. The Lac repressor, encoded by the *lacI* gene, binds very tightly to the promoter and ensures efficient repression of the strong T5 promoter interfering with the transcription of the gene of interest. In order to further investigate whether the promoter leakiness in repressor absence may lead to structural instability, DNA isolated from several non-producing M15 Δ *glyA*[B] constructs was sequenced (Figure 5.7A). Interestingly, sequence analysis revealed that 32 bp fragment of the expression vector including the -10 region and one of the homologous operator regions of the T5 promoter had been deleted (Figure 5.7B). Besides, the remaining vector backbone was sequenced, showing a correct sequence without any mutational variability. Thus, this deletion in the operator/promoter region of our expression vector, which is occurring when the strain does not encode *lacI*, is the reason for the complete lack of protein expression under these conditions.

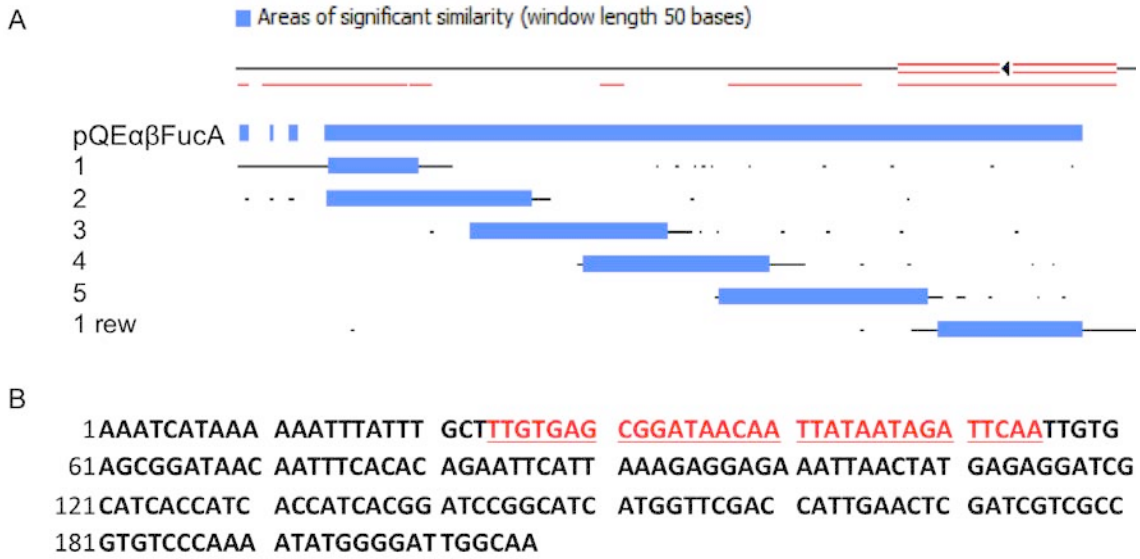


Figure 5.4 A) Sequencing of the entire pQEαβFucA plasmid, where the reference molecule is pQEαβFucA; 1, 2, 3, 4, 5 are the forward primers; 1 rew is the reverse primer used to sequencing. B) PT5 sequence with in red the 32 bp deletion region.

5.3.1 Importance of the transcription levels of the *lacI* gene

To further confirm the relevance of *lacI* transcriptional levels as key parameter in T5 promoter-based expression systems, influencing both the basal and induced expression levels, the pQEαβFucA construct was transformed into the *E.coli* MG1655 *LacI*^{q1} and DH5α *LacI*^{q1} strains, yielding *E.coli* MG1655 *LacI*^{q1} pQEαβFucA and DH5α *LacI*^{q1} pQEαβFucA, from now on MG1655-FucA and DH5α-FucA. As explained in the Materials and Methods section 3.1.2 and 3.1.5, these strains have in their chromosome the *lacI* gene under the control of the strong and constitutive q1 promoter. Once obtained the transformants, duplicated shake flasks cultures with 1mM IPTG pulse were performed in order to test recombinant protein expression (Figure 5.5).

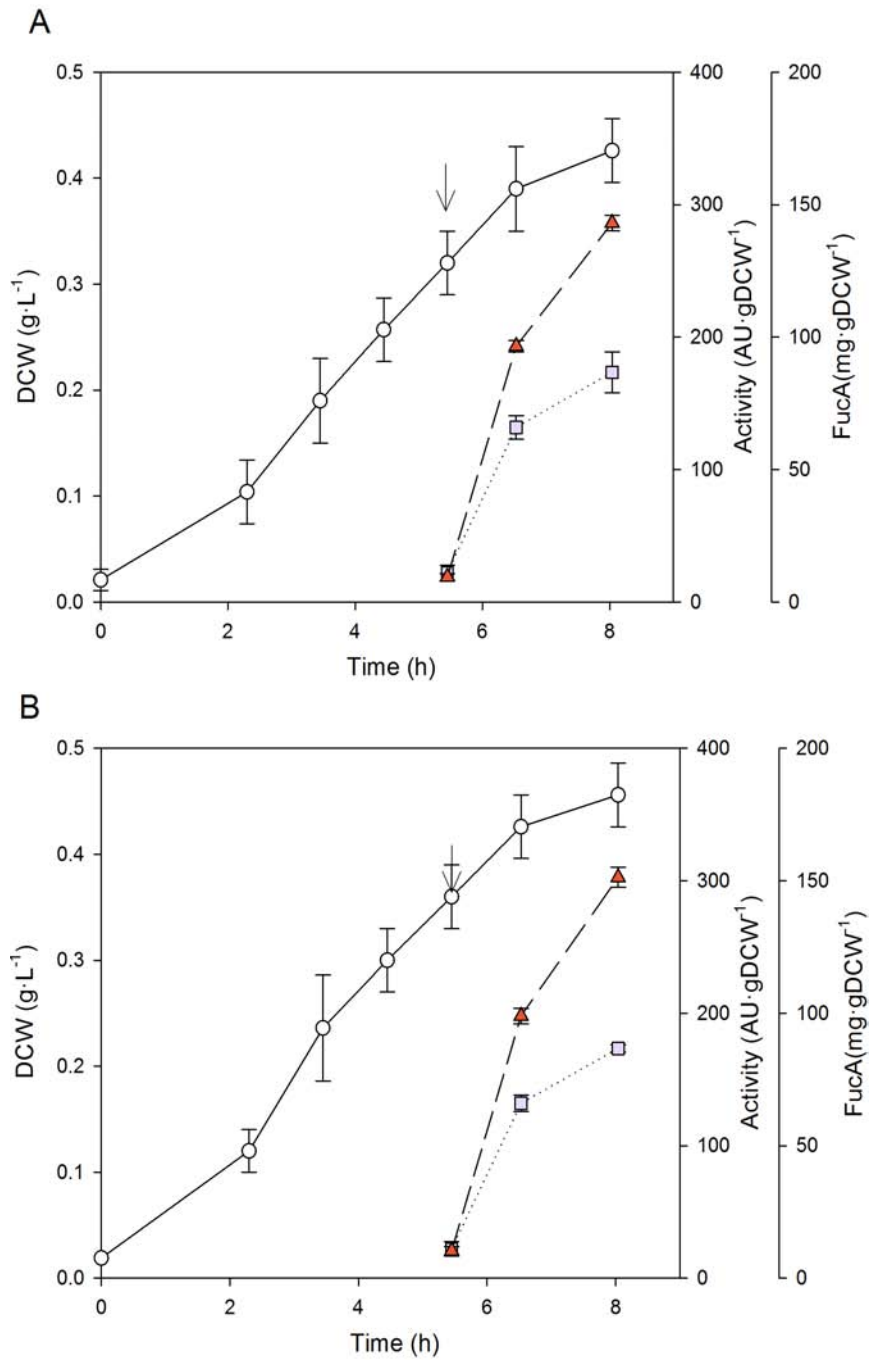


Figure 5.5 Shake flask cultures of A) *E. coli* MG1655-FucA strain and B) *E. coli* DH5-FucA strain in LB medium supplemented with ampicillin. (●) Biomass DCW (g·L⁻¹), (▲) FucA activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹).

This time, FucA protein was correctly expressed, as it can be seen in Figure 5.6, where a 26 kDa band appears 1 h and 2 h after induction.

Moreover, to ensure the FucA was functional folded, the activity test was performed, leading to a positive result for both strains. In particular, $87 \pm 8 \text{ mgFucA} \cdot \text{gDCW}^{-1}$ with an activity of $286 \pm 6 \text{ AU} \cdot \text{gDCW}^{-1}$ and $96 \pm 1 \text{ mgFucA} \cdot \text{gDCW}^{-1}$ with an activity of $303 \pm 8 \text{ AU} \cdot \text{gDCW}^{-1}$ were reached after 2 h of induction of *E. coli* MG1655-FucA and *E. coli* DH5 α -FucA strains, respectively (Figure 5.5).

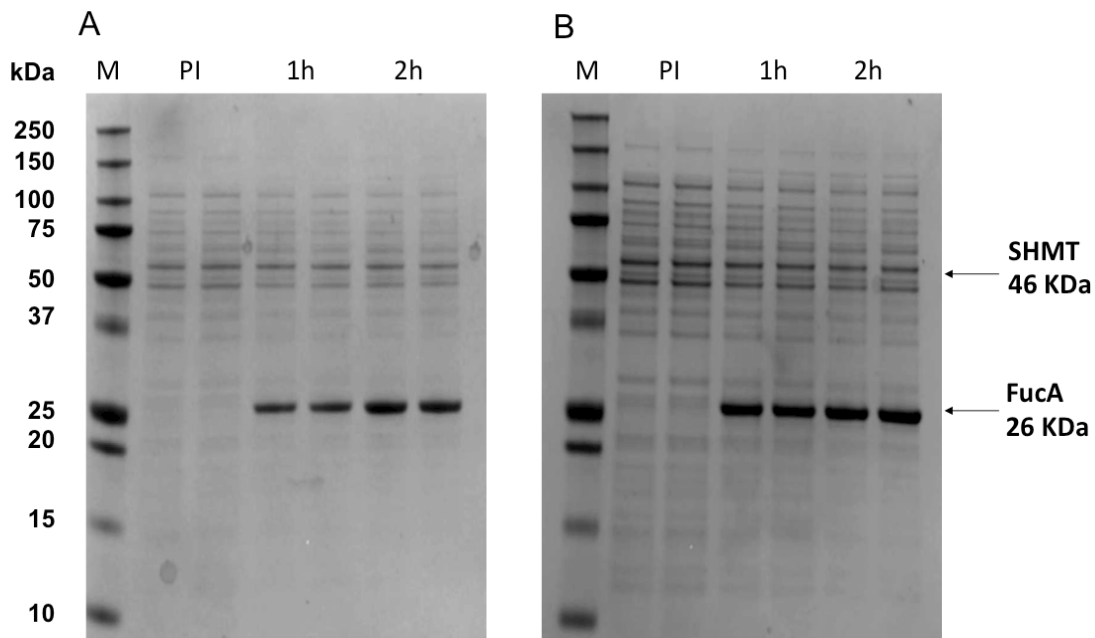


Figure 5.6 SDS-PAGE of shake flasks of A) *E. coli* MG1655-FucA strain and B) *E. coli* DH5 α -FucA strain cultures growing in LB medium plus ampicillin. Lane M: molecular weight marker; PI, pre-induction; 1 h and 2h correspond to the time after induction. Duplicate for each samples. The band corresponding to the 26 kDa FucA and the 46 kDa SHMT are indicated in the figure.

Interestingly, in this work we report the lack of FucA expression in the system with no *lacI* gene. This effect may be related to T5 promoter leakiness in absence of LacI

repressor protein, leading to plasmid structural instability due to recombination events of the homology regions, as suggested in the sequencing data. According to that, Kawe et al. have demonstrated that, by comparing different expression system with or without the *lacI* gene together in strains expressing different levels of LacI, the absence of a net overexpression of *lacI* led to an efficient promoter deletion in the plasmid (Kawe, Horn, and Plückthun 2009).

However, the exact underlying mechanism for deletion of part of the promoter region still needs further investigations.

Alternatively, a possible explanation for the lack of FucA expression in the single plasmid system may be that the copy number of the *lacI* gene increases too much when cloned into the pQE- vector, resulting in significantly higher levels of intracellular LacI. In fact, the pQE- vector is based on the plasmid replication origin ColE1 that presents 15-60 copies per cell, while the pREP4 presents the P15A replicon with about 10-12 copies per cell (Baneyx 1999b). Furthermore, the expression of recombinant proteins using the T5 promoter and its tight regulation in order to regulate the gene expression has been widely studied (Xu and S. Matthews 2009)(Gruber et al. 2008)(Glascock and Weickert 1998b). High level of Lac repressor can suppress basal expression of recombinant protein (i.e. improve "leakiness"), but can simultaneously decrease maximal level of production (Gruber et al. 2008). Accordingly, the cloning of the *lacI* gene in the pQE-vector, which presents a copy number 2-fold higher compared with the pREP4, resulted in low detection levels of FucA.

5.4 Conclusion

Since the pQE- expression system used along this study does not carry the *lacI* gene, a net overexpression of Lac repressor must be supplied, either from the bacterial chromosome or from the expression plasmid, in order to avoid possible homology recombination events of the two *lacO* regions.

The change of *lacI* copy number, moving from the pREP4 plasmid to the pQE- vector, is a key parameter to take into account for recombinant protein production.

In conclusion, these results confirmed that the regulation of *lacI* transcriptional levels is needed, in order to optimize the regulation and induction of the foreign gene expression.

6 Results III: Design and optimization of a single plasmid expression system

Results presented in Chapter 4 pointed at the regulation of the *glyA* expression levels as an important parameter to be taken into account for further improvement of the expression system. In fact, the first generation of the *glyA*-based auxotrophic system maintenance system contained the *glyA* gene under the control of the P3 constitutive promoter, resulting in relatively high amounts of its product. The SHMT, accumulated as soluble protein in the cytoplasm of the cells, could impose a slight metabolic burden to the cell, i.e. having a negative impact on the μ_{\max} and FucA production. Besides, in Chapter 5, we demonstrated the relevance of *lacI* transcriptional levels as key parameter in T5 promoter-based expression systems, influencing both the basal and induced transcriptional levels.

In this Chapter, an expression cassette was settled where the *lacI* and *glyA* genes were cloned under the control of the four constitutive promoters. The BioBrick vector pSB1C3_GG is used as backbone for the construction of a series of four vectors, each co-expressing the *lacI* and *glyA* genes under the selected constitutive promoters. That is, each expression vector differs from the other just for the constitutive promoters. The aim was to find the suitable promoter with strength enough to synthesize the minimum amount of LacI inhibitor preventing “promoter leakiness” (i.e. basal expression levels under non-inducing conditions), as well as the minimal *glyA*

transcriptional levels required to maintain plasmid-bearing cells and optimal cell growth in defined media.

Recent advances in the synthetic biology, allowed the development of new methods and tools to speed up and standardize strain engineering. Despite the large diversity of existing DNA shuffling protocols for collecting variant gene libraries (Zuo and Rabie 2010) (Gibson 2011), the golden gate method based on the type II restriction enzymes has been used for assembling the BioBrick vectors. Compared with conventional DNA cloning protocols, these advanced DNA assembly strategies offer an efficient approach to construct multi-gene pathways in a one-step, scar-less, and sequence-independent manner.

6.1 Tuning of *lacI* and *glyA* expression levels

Four BioBrick constructs were assembled, each one with a different constitutive promoter to tune the expression levels of *lacI* and *glyA* genes, separately

The BioBrick vectors were assembled using the Golden Gate technique (Materials and Methods section 3.3.4).

6.1.1 BioBrick-based plasmid constructs for *lacI* expression level tuning

As explained in Materials and Methods section 3.4.5.1 four BioBrick constructs were assembled, each one with a different constitutive promoter to modulate the expression levels of *lacI* (Figure 3.15), and then transformed into *E. coli* competent cells yielding M15 Δ *glyA* pQE-FucA + pSB-J231XX-S_*lacI* (shortly J23100-S_*lacI*, J23111-S_*lacI*, J23110-S_*lacI* and J23117-S_*lacI* strains). After colony PCR and sequencing, one correct assembled clone for each strain was selected for further testing in shake flasks cultures.

The over-expression of FucA for the 4 resulting selected transformants J231XX-S_*lacI* was tested in shake-flasks cultures per triplicate. Cultures were grown in LB media and the expression of the recombinant protein was induced with 1mM IPTG at an OD of 1.5. Clearly, the defined media can not be used because this strain is devoid of the auxotrophic gene.

The FucA production both in term of mass and activity for the 4 constructions is summarized in figure 6.2. For each strain a single induction value, after one hour of IPTG addition, is represented. As it can be clearly seen, the best expression vector is the J23110-*lacI* strains (P3 in figure 6.1), with the highest values of both FucA mass and activity, being $59 \pm 6 \text{ mg}\cdot\text{g}^{-1}\text{DCW}$ and $228 \pm 15 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ respectively. The J23100 promoter is the worst, leading to low detection values of FucA (P1 in figure 6.2).

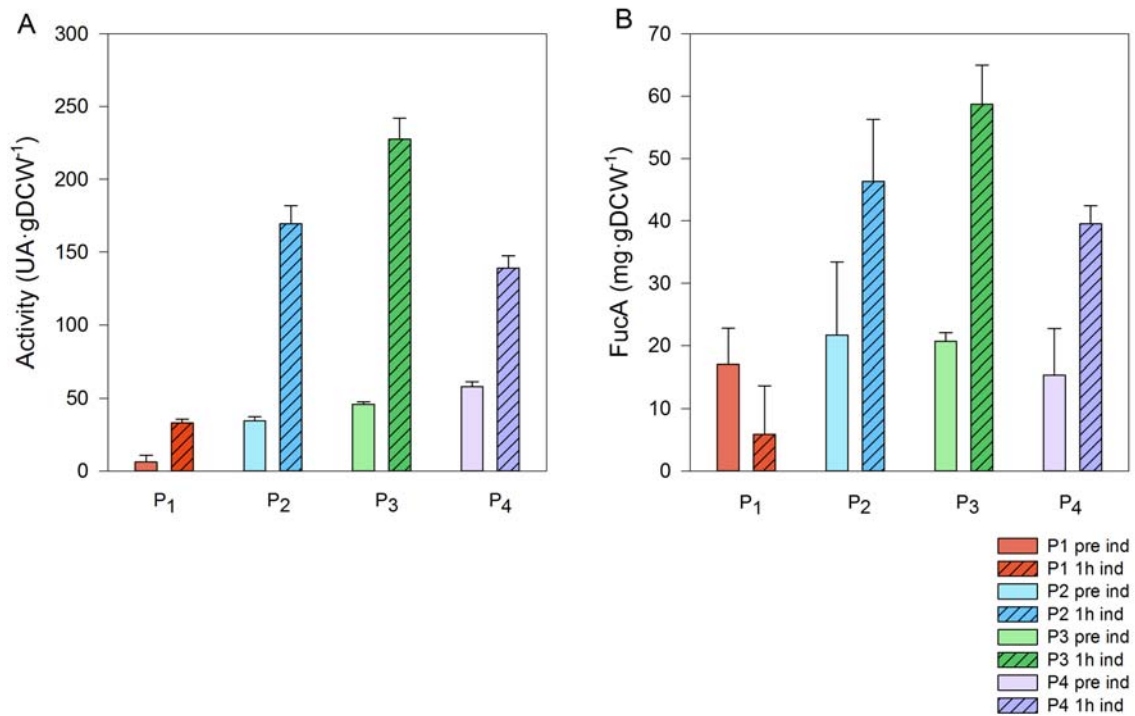


Figure 6.1 Shake flask cultures of M15Δ*glyA* pQE-FucA pSB1XX-S_ *lacI* strains series. A) FucA activity (AU·g⁻¹DCW) and B) specific mass production (mgFucA·g⁻¹DCW) profiles along time in LB medium for the 4 selected transformants with the 4 different constitutive promoters: P1) J23100-*lacI*, P2) J23111-*lacI*, P3) J23110-*lacI*, P4) J23117-*lacI*.

6.1.2 BioBrick-based constructs for *glyA* expression tuning

Golden gate reactions were performed for the construction of the four pSB-J231XX-S_ *glyA* BioBrick vector for the tuning of the *glyA* gene in order to modulate the expression levels of *glyA* (Materials and Methods section 3.4.5.2, Figure 3.16). Thus, the four expression vectors were transformed together with the pQE-FucA plasmid into the M15Δ*glyA* competent cells yielding the 4 final strains M15Δ*glyA* pQE-FucA + pSB-J231XX-S_ *glyA* (shortly J23100-S_ *glyA*, J23111-S_ *glyA*, J23110-S_ *glyA* and J23117-S_ *glyA* strains).

The growth for the 4 resulting selected transformants J231XX-S_ *glyA* was tested in shake-flasks cultures in defined medium, in order to find the minimal *glyA* transcriptional levels required to maintain plasmid-bearing cells and optimal cell growth in defined media.

The evolution of the DCW ($\text{g}\cdot\text{L}^{-1}$) along time is represented in Figure 6.2. As observed in Figure 6.2, the specific growth rate measured showed similar values for all the transformants except for those with the J23100 promoter, which presented the higher μ_{max} being $0.51 \pm 0.01 \text{ h}^{-1}$.

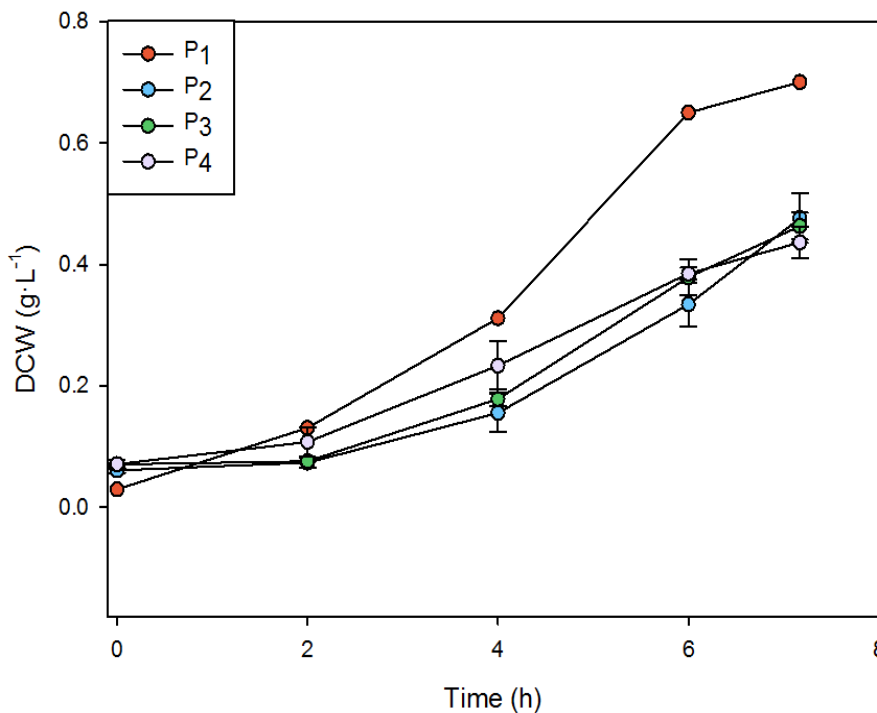


Figure 6.2 Shake flask cultures of M15Δ*glyA* pQE-FucA pSB1XX-S_ *glyA* strains series. DCW ($\text{g}\cdot\text{L}^{-1}$) along time in defined medium shake flasks cultures for the 4 selected transformants with the 4 different constitutive promoters: red) J23100-*glyA*, light blue) J23111-*glyA*, green) J23110-*glyA*, pink) J23117-*glyA*.

The J23117 promoter, with the minimal amount of SHMT produced, is the one chosen for the *glyA* expression together with the J23110 promoter used for the *lacI* expression.

6.1.3 BioBricks-based construction of with the tuned *lacI* and *glyA* genes

The construction of the single expression vector presents the *lacI* and *glyA* genes cloned downstream the J23110 (P3) and J23117 (P4) constitutive promoters, respectively (Materials and Methods section 3.4.5.3). The final product, named pQE-FucA-P3_ *lacI*-P4_ *glyA* (Figure 3.17) was finally transformed into M15 Δ *glyA* obtaining the M15 Δ *glyA* pQE-FucA-P3_ *lacI*-P4_ *glyA*. Transformants were isolated and correct plasmid sizes were verified by restriction analysis and sequencing. Unfortunately, the two promoters, probably due to a high sequence homology, undergo to a recombination event that brings to loose a region of the promoter, represented by a gap in Figure 6.3.

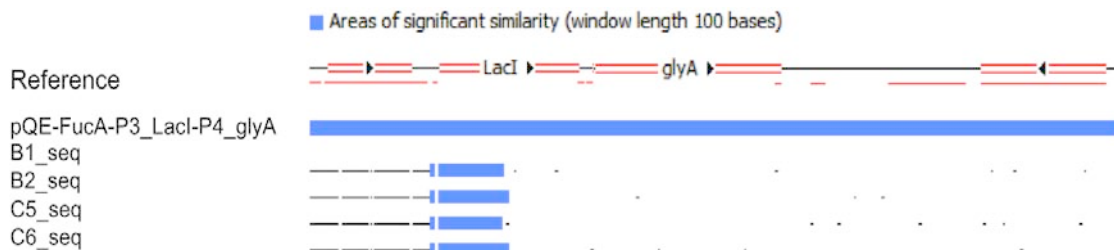


Figure 6.3 Sequencing of the pQE-FucA_S-puzzle transformants. pQE-FucA-P3_ *lacI*-P4_ *glyA* is the final vector; B1, B3, C5, C6 the sequenced sequence of 4 positive transformants.

6.2 Assembling the polycistronic *lacI-glyA* cassette

In order to avoid the promoters' recombination, four different expression cassettes were constructed where the *lacI* and *glyA* genes were placed under the control of a set of four constitutive promoters, thereby obtaining a polycistronic expression cassette, covering a wide range of transcriptional efficiencies. The resulting selected transformants obtained, each one with a *lacI-glyA* cassette with a different constitutive promoter, were named pSB-J231XX, where the double X represents the last two digits of the promoter name (Figure 3.20).

Thus, they were transformed together with the pQE-FucA plasmid into the M15 Δ *glyA* competent cells, resulting in the final strains named M15 Δ *glyA* pQE-FucA + pSB-J231XX (shortly M15 Δ *glyA*[C00], M15 Δ *glyA*[C11], M15 Δ *glyA*[C10], M15 Δ *glyA*[C17] strains).

The over-expression of FucA for the 4 resulting selected transformants was tested in shake-flasks cultures. Cultures were grown on 5 g·L⁻¹ of initial glucose concentration and when the OD reached 1.5, FucA expression was induced with 1 mM IPTG. Samples were collected before induction, as well as 1h, 2h and 4h post-induction for analysis. The evolution of cultures is summarized in figures 6.4 and 6.5.

The specific growth rate measured in the different cultures showed a similar behavior for all the transformants except for those with the M15 Δ *glyA*[C00] promoter. The strongest promoter is the one that presented a μ_{\max} of 0.62 ± 0.05 h⁻¹, 1.3 fold higher than the reference strain M15[pREP4] (Table 6.1).

The higher growth rate of the M15 Δ *glyA*[C00] strain could be explained as follows: higher constitutive *lacI* expression level may lead to a reduction of the *fucA* expression and subsequently decrease the metabolic burden.

As it can be seen in figure 6.4C and in table 6.1, the best expression vector is the one harboring the constitutive promoter M15 Δ *glyA*[C10], with a production of $93 \pm 3 \text{ mg} \cdot \text{g}^{-1} \text{DCW}$ and $641 \pm 54 \text{ AU} \cdot \text{g}^{-1} \text{DCW}$ for the mass and activity respectively. *FucA* expression can be seen also in figure 6.6 where the SDS-PAGE of shake flasks cultures samples of the M15 Δ *glyA*[C10] and M15 Δ *glyA*[C11] strains for triplicates are presented (Figures 6.4A and 6.4B, respectively). The intensity of the *FucA* band is more accentuated in the M15 Δ *glyA*[C10] strain. Moreover, it can be seen that the SHMT band, that represents the *glyA* product, is reduced passing from the strong J23111 to the weaker J23110 promoters.

Conversely, the use of a stronger constitutive promoter for the *lacI* and *glyA* expression such as the M15 Δ *glyA*[C00] resulted in low detection levels of *FucA*, both, in terms of mass and activity.

Furthermore, when looking at the glucose consumption profile (Figure 6.5) it can be observed that the 4 different constructs have a similar trend during the exponential phase, while comparing the q_s values during the induction phase it can be clearly seen that the M15 Δ *glyA*[C00] strain is the one with higher specific uptake rate of glucose, being $0.79 \pm 0.10 \text{ g} \cdot \text{g}^{-1} \text{DCW} \cdot \text{h}^{-1}$ (Table 6.1). Consistently, this strain is the one with the highest yield of acetate reaching $1.50 \pm 0.10 \text{ gAc} \cdot \text{g}^{-1} \text{DCW}$.

Furthermore, the *M15ΔglyA*[C10] strain presented the lowest q_s ($0.44 \pm 0.06 \text{ g} \cdot \text{g}^{-1} \cdot \text{DCW}^{-1} \cdot \text{h}^{-1}$) and acetate yields ($0.70 \pm 0.12 \text{ g} \cdot \text{g}^{-1} \cdot \text{DCW}$) compared to the other three constructs.

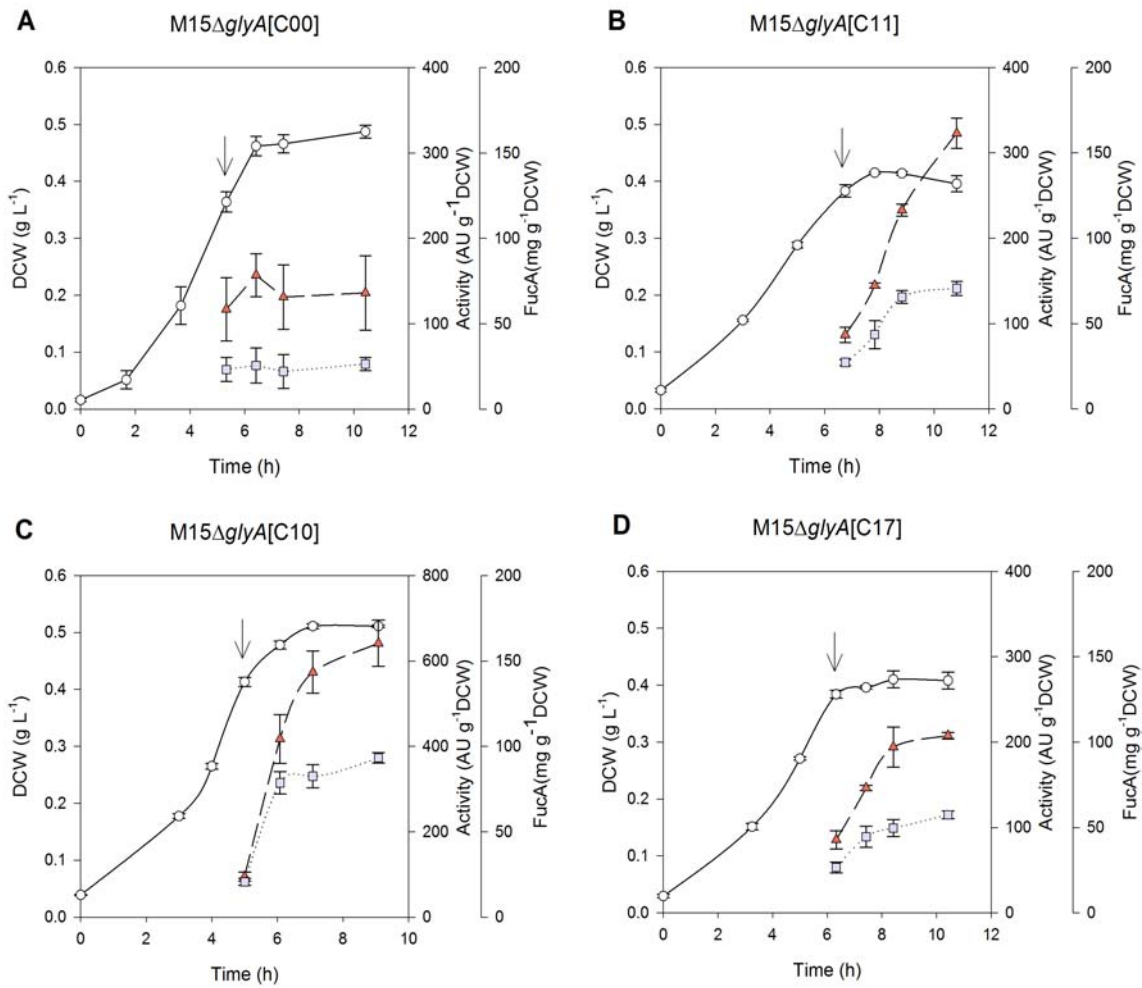


Figure 6.4 Shake flask cultures in DM medium of the 4 different selected transformants *M15ΔglyA* pQE-FucA pSB1C3 with the 4 different constitutive promoters: A) J23100 B) J23111 C) J23110 D) J23117. (●) Biomass DCW ($\text{g} \cdot \text{L}^{-1}$), (▲) FucA activity ($\text{AU} \cdot \text{gDCW}^{-1}$), (■) specific mass ($\text{mgFucA} \cdot \text{gDCW}^{-1}$). 1mM IPTG pulse is indicated by an arrow.

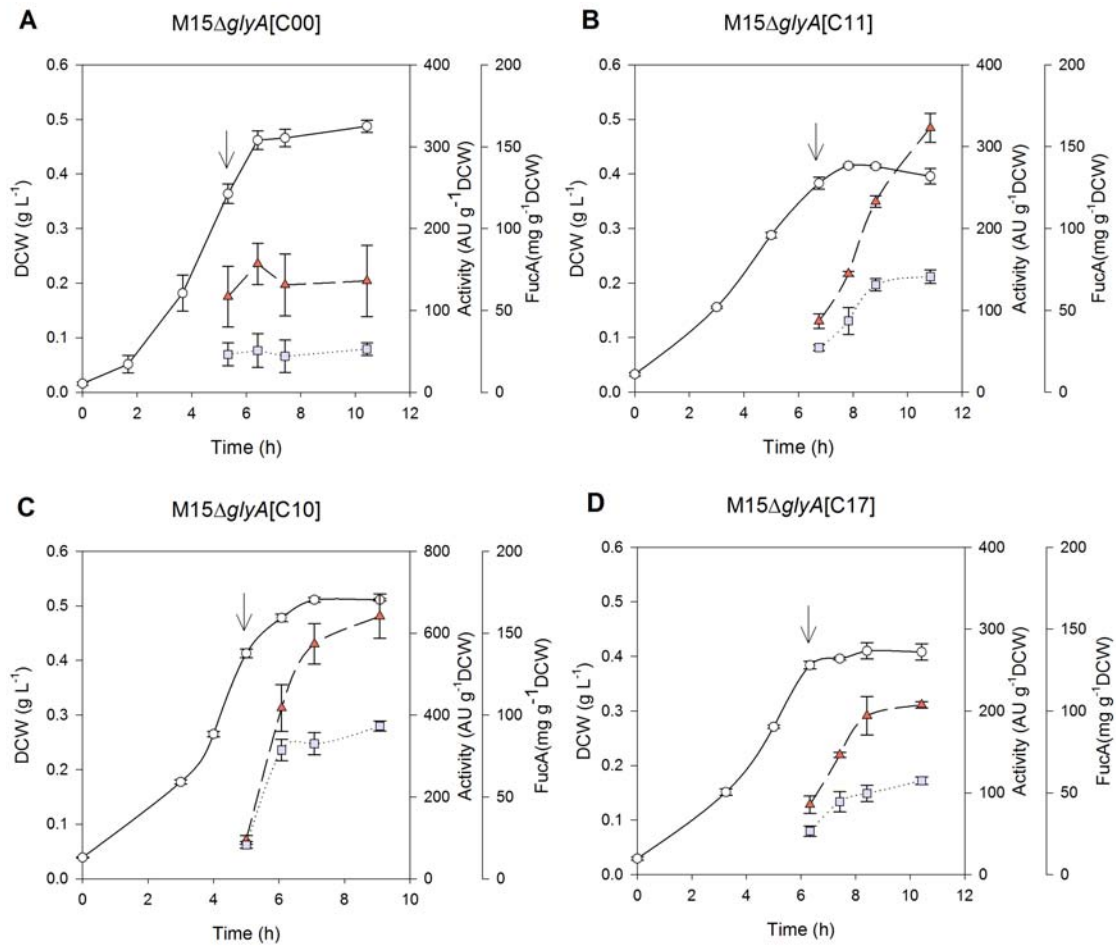


Figure 6.5 Shake flask cultures in DM medium of the 4 different selected transformants *M15ΔglyA* pQE-FucA pSB1C3 with the 4 different constitutive promoters: A) J23100 B) J23111 C) J23110 D) J23117. (●) Biomass DCW(g·L⁻¹), (▼) Glucose concentration (g·L⁻¹) and (▲) Acetic Acid (g·L⁻¹). 1mM IPTG pulse is indicated by an arrow.

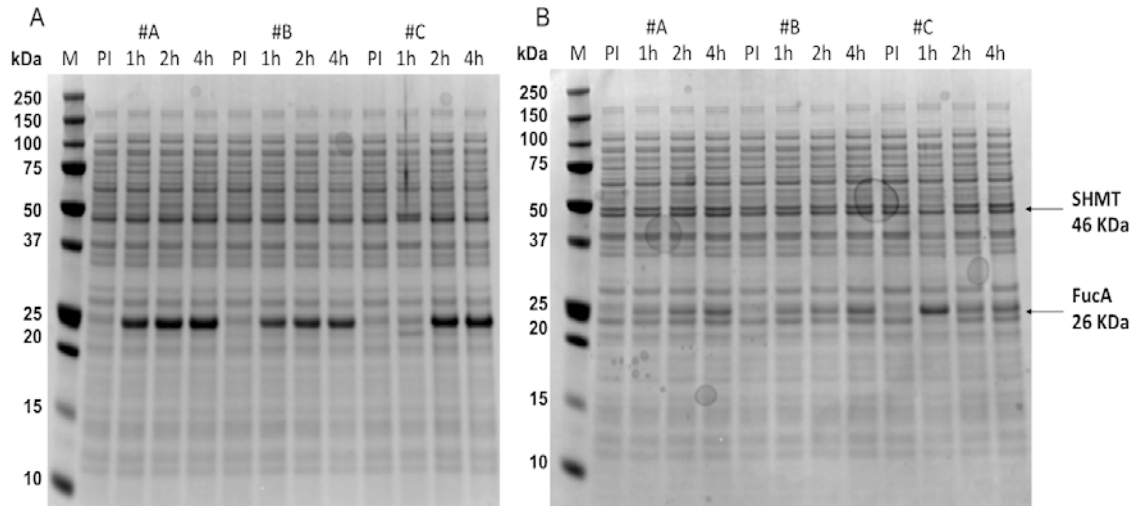


Figure 6.6 SDS-PAGE of shake flasks cultures of A) *E. coli* M15ΔglyApQE-FucA pSB1C3-J23110 and B) *E. coli* M15ΔglyApQE-FucA pSB1C3-J23111 strains in DM media. Lane M: molecular weight marker; 1, 2, 3 are the replicates; PI, pre-induction; 1 h, 2h and 4h correspond to the time after induction. The 26 kDa FucA and the 46 kDa SHMT bands are indicated in the figure.

Table 6.1 Maximum specific growth rate (μ_{\max} h⁻¹), FucA activity (AU·g⁻¹DCW), FucA mass (mg·g⁻¹DCW), q_s of the induction phase and the maximum acetate yield (g·g⁻¹DCW) for each of the four selected transformants M15ΔglyA pQE-FucA pSB1C3 with the 4 different constitutive promoters: J23100, J23111, J23110 and J23117, from the strongest to the weakest. The values represent the sample after 2 hour of induction.

Promoter	μ_{\max} (h ⁻¹)	FucA Activity (AU·g ⁻¹ DCW)	FucA mass (mg·g ⁻¹ DCW)	q_s (g·g ⁻¹ DCW·h ⁻¹)	Acetate yield (g·g ⁻¹ DCW)
J23100	0.62 ± 0.05	136 ± 43	26 ± 4	0.79 ± 0.10	1.5 ± 0.10
J23111	0.37 ± 0.01	323 ± 18	71 ± 4	0.42 ± 0.09	1.12 ± 0.09
J23110	0.48 ± 0.01	641 ± 54	93 ± 3	0.44 ± 0.06	0.70 ± 0.12
J23117	0.41 ± 0.01	207 ± 4	57 ± 2	0.58 ± 0.01	1.11 ± 0.05

6.3 Engineered single-plasmid expression system, the Puzzle strain

According to the results presented above, the selected expression cassette for *lacI* and *glyA*, was the one where both genes were placed under the control of the constitutive promoter J23110. This promoter, which is in the lower range of the tested *lacI* and *glyA* transcriptional levels, seems to down-regulate their transcriptional levels. This suggested that the reduced expression of *lacI* and *glyA* expression seems to have reduced the energy demand and the building blocks necessary for *glyA* synthesis. In addition, T5 promoter leakiness was minimized, resulting in an overall reduced metabolic burden. This result is reflected in the fact that the μ_{\max} of this strain was comparable to that of the M15[pREP4] reference strain ($0.48 \pm 0.01 \text{ h}^{-1}$) under pre-induction conditions.

The next goal was the construction of a single expression vector harboring both the *fucA* gene under control of the inducible T5 promoter and the *lacI-glyA* cassette cloned under the J23110 constitutive promoter, and the elimination of the PSB1C3-J231XX plasmid from the M15 Δ *glyA* cells.

The selected expression cassette was cloned into the pQE-FucA obtaining the pQE-FucA_puzzle (J23110) (Figure 3.21). Finally, the ligation product was transformed into the M15 Δ *glyA* generating M15 Δ *glyA* pQE-FucA_puzzle (J23110), from now on Puzzle strain (Materials and Methods section 3.4.6)

6.3.1 Shake flasks cultures of the Puzzle strain

Cultures of the new strain were carried out in triplicate at 37°C in defined medium (DM) and FucA production was induced with 1mM IPTG. The μ_{\max} of the Puzzle strain being $0.45 \pm 0.01 \text{ h}^{-1}$, was still comparable to those from the preceding 2-plasmid construct and original M15[pREP4] reference strain (being 0.44 ± 0.01 and $0.49 \pm 0.02 \text{ h}^{-1}$, respectively).

Interestingly, FucA production values increased even more in the strain Puzzle strain compared to the previous system with two plasmids, with a maximum FucA mass and FucA specific activity of $97 \pm 14 \text{ mg}^{-1} \text{ FucA} \cdot \text{g}^{-1} \text{DCW}$ and $984 \pm 32 \text{ AU} \cdot \text{g}^{-1} \text{DCW}$, respectively (Figure 6.7A). Besides, comparing these values with those obtained with the M15[pREP4] reference strain ($181 \pm 5 \text{ mg FucA} \cdot \text{g}^{-1} \text{DCW}$ and $721 \pm 82 \text{ AU} \cdot \text{g}^{-1} \text{DCW}$), it can be observed how, even though the amount of the recombinant protein was still lower, the specific activity increased 1.4-fold. Besides, the puzzle strain presented a reduction in the amount of acetate production, being $0.57 \pm 0.03 \text{ g} \cdot \text{g}^{-1} \text{DCW}$, compared to $0.73 \pm 0.04 \text{ g} \cdot \text{g}^{-1} \text{DCW}$ of the M15[pREP4]. Finally, SHMT values were calculated for each time point of induction in shake flasks cultures. The results are presented in Table 6.1, where the pre-induction sample with $66 \pm 17 \text{ mgSHMT} \cdot \text{g}^{-1} \text{DCW}$ represents the sample with the higher amount of SHMT. It can be seen that this values decreased 1.4-fold comparing with the M15 Δ glyA[pREP4] strain, where the pre-induction values was $92 \pm 14 \text{ mgSHMT} \cdot \text{g}^{-1} \text{DCW}$ (Table 4.2).

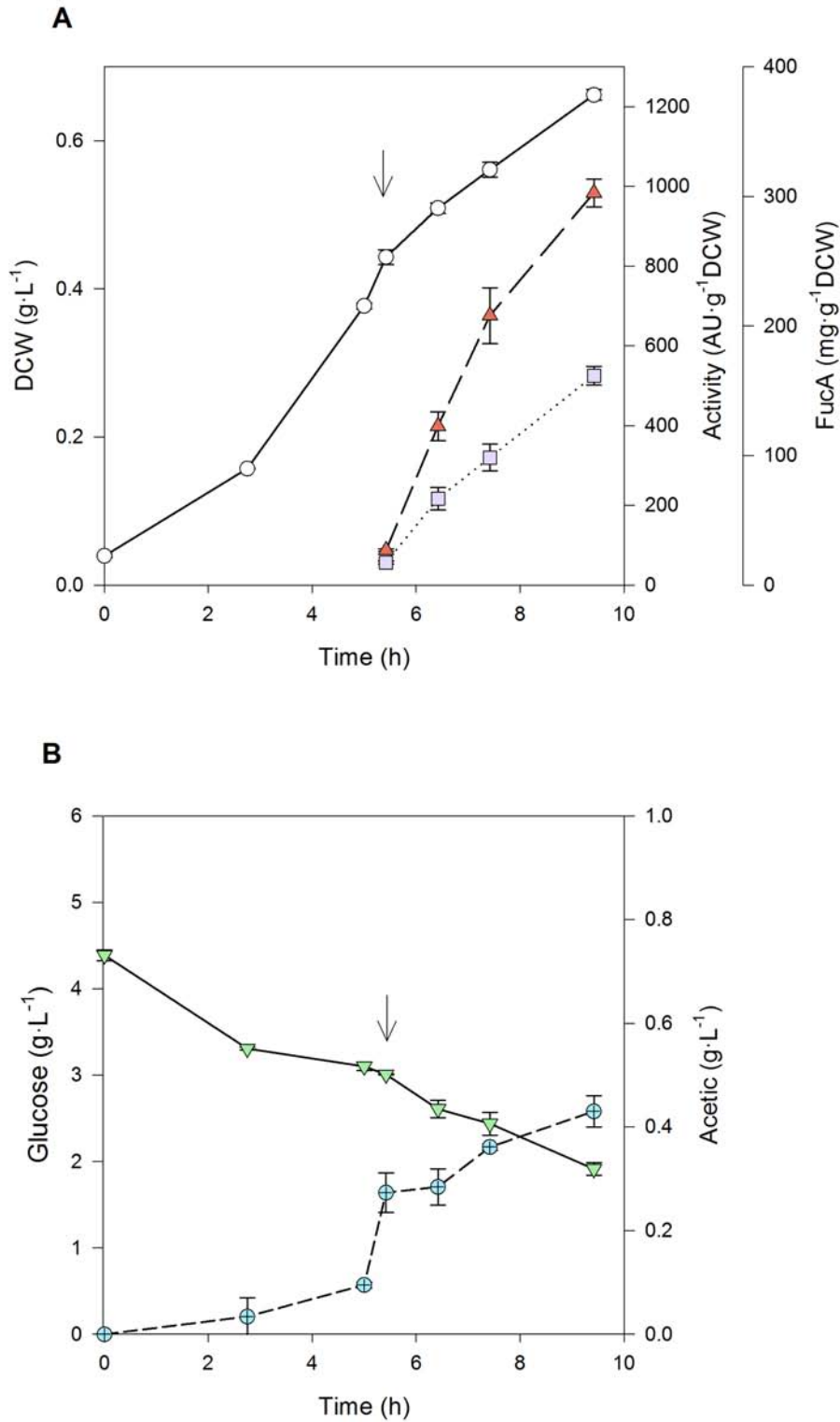


Figure 6.7 Shake flask cultures in DM medium of the Puzzle strain **A-1), (▲) FucA activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹). **B**) (▼) Glucose concentration (g·L⁻¹) and (⊕) Acetic Acid (g·L⁻¹). The arrow indicates the 1mM IPTG pulse for the induction.**

Table 6.1 SHMT production ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) along the induction phase for the Puzzle strain. PI (pre-induction) and 1 h, 2 h and 4 h correspond to the time after induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	66 ± 17
1 h	62 ± 10
2 h	54 ± 14
4 h	54 ± 7

6.3.2 Flow cytometry analysis of the Puzzle strain

Cell viability and vitality of the Puzzle strain growing in shake flasks was monitored by flow cytometry analysis, as performed with previous expression strains.

Three experimental conditions were tested, being: A) 1mM IPTG plus ampicillin, B) 70 μM IPTG plus ampicillin and C) 70 μM IPTG without antibiotic.

As an example, profiles of the Puzzle strain growing on shake flask cultures in DM plus ampicillin are shown in Figure 6.8. In these cultures, FucA expression was induced with a 1mM IPTG pulse.

As observed for previous strains in Chapter 4, growth rate started to slow down just after IPTG induction. The values of the μ , calculated from the pre induced to the 4 h post-induction samples, decreased from 0.45 to 0.13 h^{-1} during this period. This reduction is confirmed also by the cell concentration values ($\text{cells}\cdot\text{mL}^{-1}$) that cease to increase after only 2 h of induction. Moreover, whereas the value of the DCW ($\text{g}\cdot\text{L}^{-1}$)

remained invariant, cell concentration decreased 35 % from 4 to 18 h of induction, as observed in Figure 6.8.

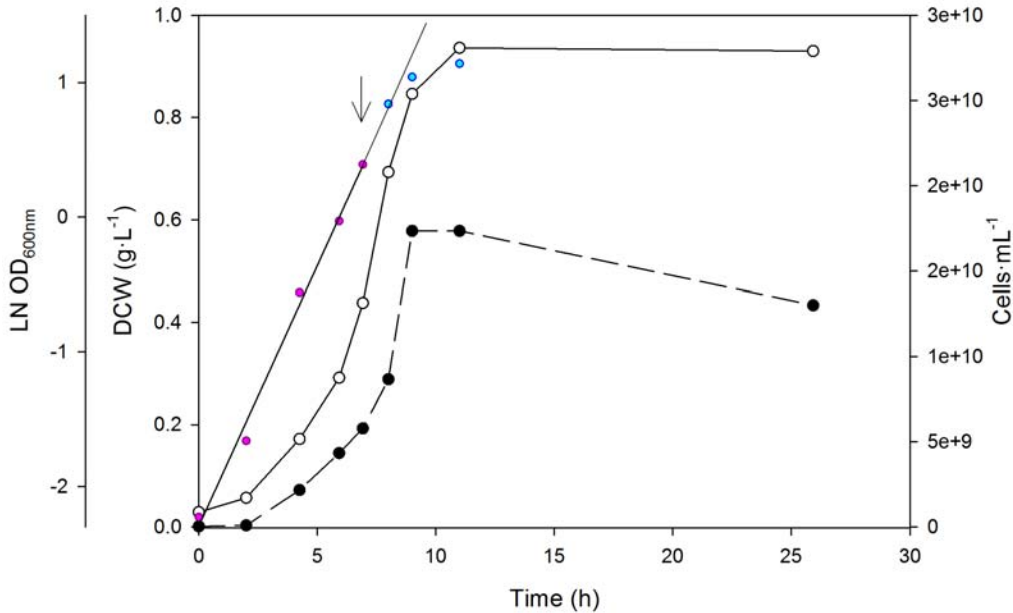


Figure 6.8 Shake flasks culture of *E. coli* Puzzle strain in DM media plus antibiotic. (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (◻) concentration determined using flow ($\text{cell}\cdot\text{mL}^{-1}$). (●) LN OD_{600} of the pre-induction phase and (◻) LN OD_{600} of the post-induction phase. 1mM IPTG induction is indicated with an arrow and the curve that fits the pre-induction samples is represented.

Furthermore, the curve representing the FSC of the 18 h post-induction sample shifted to the right compared to the curve of the reference sample of pre-induction. As explain in Chapter 4, this increment of the forward scatter may indicate that the cells after being induced undergo a morphological change of their shape, becoming more elongated (Figure 6.9A). This change does not occur to the negative control, without induction (Figure 6.9B).

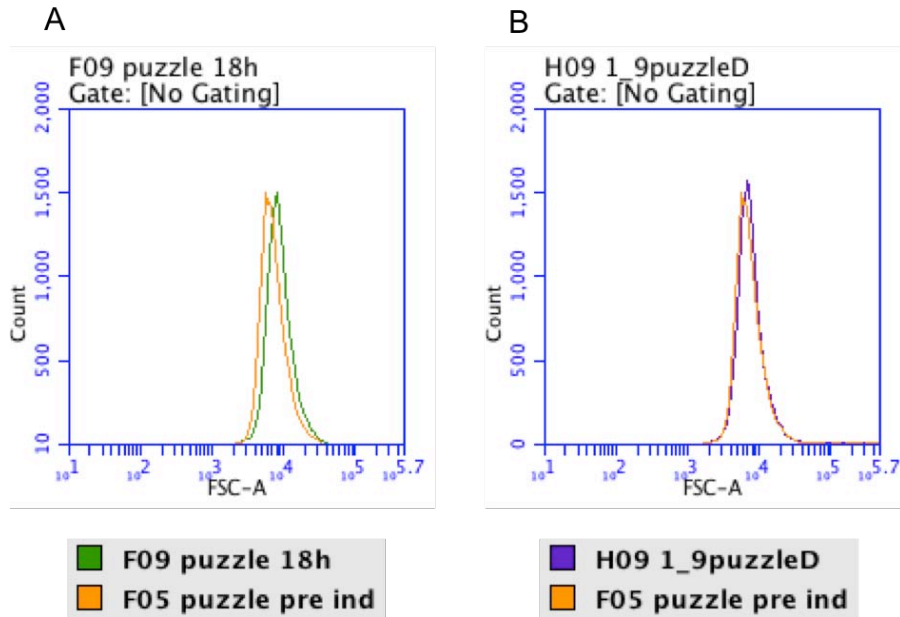


Figure 6.9 Flow cytometry analysis of the forward-scatter light (FSC-A) versus the count number of the cells. Samples of the pre-induction (pre ind, orange) and of 18 hours after induction of both samples induced (green) and negative control not induced (purple). Samples of shake flasks cultures of *E. coli* M15Δ*glyA* pQE-FucA puzzle strain in DM media plus antibiotic.

Analysis of samples by flow cytometry using as fluorescent dyes BOX and P.I., are represented in Figure 6.10. Cells analysed after 6.9 h of growth correspond to pre-induced cells in their exponential growth phase. Analyses suggest that the 95.5 % of the population at exponential growth is healthy (Q1 quadrant of the pre ind sample). During the first 2 h of induction the percentage of viable population stays higher than 90 %. However, after 4 h of induction, a change in the population is observed, where viable cells represent only 72.7 % of the population, and 16.1 % and 11.2 % of the population appear to be damaged and dead, respectively. This viability change reaches its peak when cells enter into the stationary phase (18 h of cultivation), where approximately 30 % of the population is permeable to both BOX and P.I. (quadrant Q3

of the 18 h-sample Figure. 6.10). The higher amount of damaged cells in Q3 quadrant is probably due the higher induction pulse of IPTG.

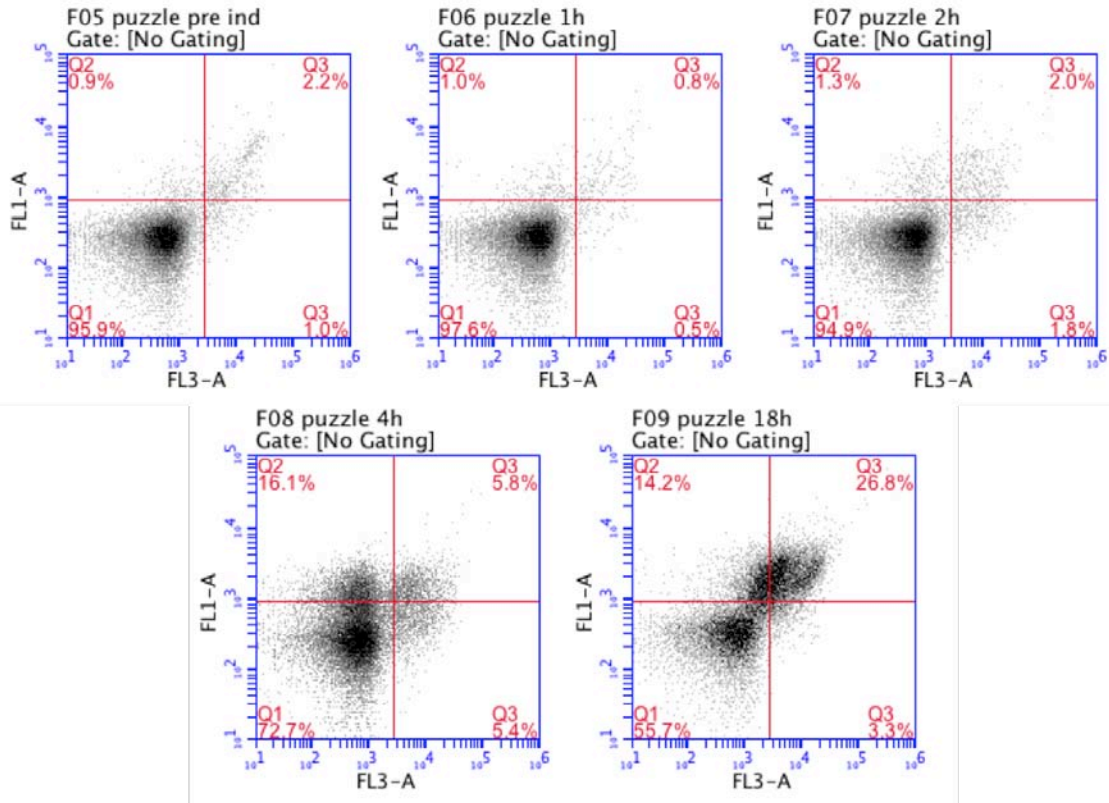


Figure 6.10 Flow cytometry analysis of *E. Coli* M15ΔglyA pQE-FucA puzzle. Shake flasks cultures performed in DM media plus antibiotic. Pre-induction (pre ind), after 1 h, 2 h, 4 h and 18 h induction are represented. Samples were co-stained with both BOX/P.I. (See Materials and Methods section 3.7.4.2). The numbers in each quadrant show percentage of bacteria population in each subgroup.

The comparison of the three growth conditions, being: A) 1mM IPTG plus ampicillin, B) 70 μM IPTG plus ampicillin and C) 70 μM IPTG without antibiotic, is represented in Figure 6.11. As it can be observed, the experiment with 1mM induction leads to the higher amount of damaged and dead cells both at 4 h of induction and even more at 18 h of induction.

The reduction of the inducer concentration leads to an improvement of the physiological state of the cells, reflected in very low levels of damaged and dead cells (3.6 % and 14.8 %, respectively) after 18 h of induction (Figure 6.11B). Moreover, the elimination of the addition of the antibiotic in the culture medium, reduce even more the damaged and dead cells to values of 1 % and 7.9 % respectively (Figure 6.11C). Plasmid segregational stability analyses of this series of cultivations were carried out confirming that the cells, both in presence or absence of the antibiotic, maintain the expression vector. Therefore, the reduction of the IPTG pulse, together with the elimination of the antibiotic in the medium, lead to an improvement of the physiological state of the population.

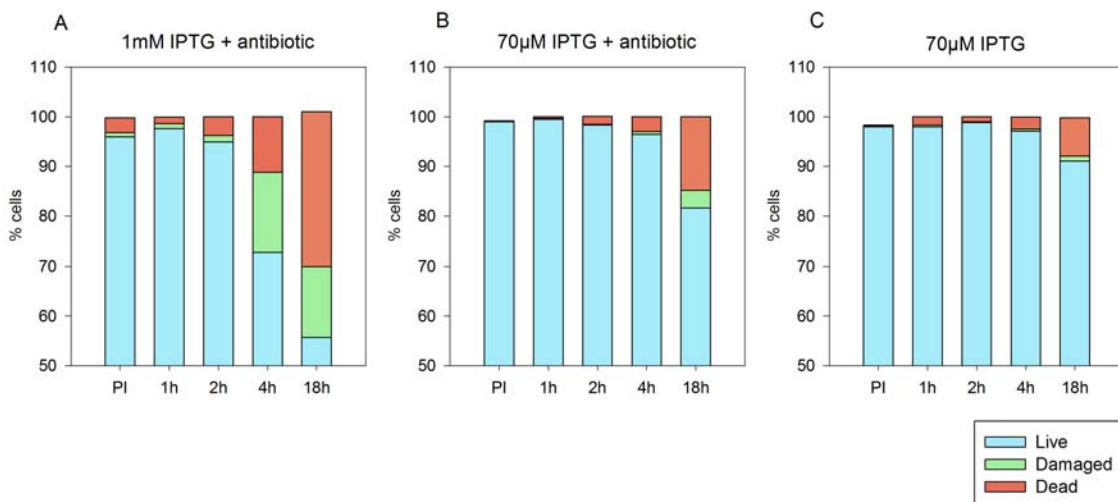


Figure 6.11 Comparison of the three experimental conditions in shake flask cultures for the *E. coli* M15ΔglyA pQE-FucA puzzle strain. PI, pre-induction samples; 1 h, 2 h, 4 h and 18 h represent the time after induction. A) DM + antibiotic and 1 mM IPTG; B) DM + antibiotic and 70 μM IPTG; C) DM without antibiotic and 70 μM IPTG. Live cells are represented in light blue, damaged cells in green and dead cells in orange.

6.3.3 Fed-batch cultures of the Puzzle strain

Fed-batch cultures have been performed in order to prove the robustness of the single expression system compared with the previous 2-plasmid construct. Like in Chapter 4, all the fed-batch experiments were carried out in defined medium with glucose as sole carbon source with a constant specific growth rate for the feeding of 0.22 h^{-1} . A pulse of IPTG $70 \mu\text{M}$ was added once reached $20 \text{ gDCW}\cdot\text{L}^{-1}$ of biomass, in order to induce the overexpression of the FucA protein. The time-profiles of the biomass, glucose consumption, FucA mass and specific activity and acetate production of $\text{M15}\Delta\text{glyA}$ pQE-FucA puzzle strain are represented in Figure 6.12.

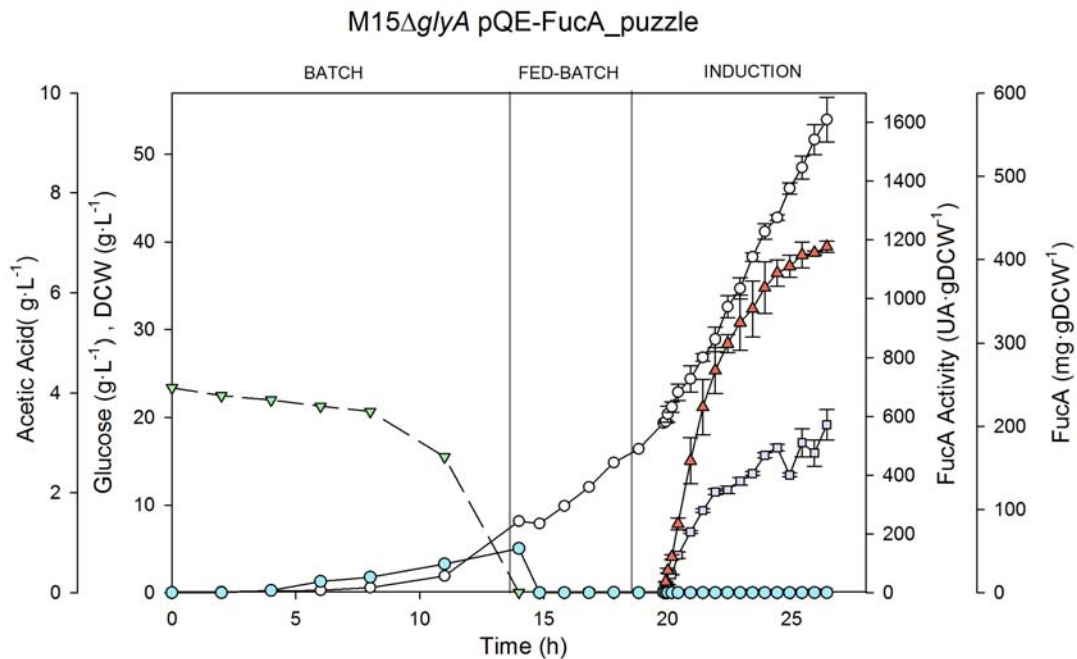


Figure 6.12 *E. coli* Puzzle Fed-Batch culture: [IPTG], $70\mu\text{M}$; X_{ind} , $20 \text{ g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . Schematic representation of (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (▼) Glucose concentration ($\text{g}\cdot\text{L}^{-1}$), (▲) FucA activity ($\text{AU}\cdot\text{gDCW}^{-1}$), (■) specific mass ($\text{mgFucA}\cdot\text{gDCW}^{-1}$) and (◐) Acetic Acid ($\text{g}\cdot\text{L}^{-1}$) along time. Batch, fed-batch and induction phases are indicated.

Along the batch phase it can be observe a constant increment of acetate in the culture media. It is proved that acetate excretion is physiologically due to a metabolic overflow mechanism, caused by an imbalance between the rapid uptake of glucose and its conversion into biomass and products, diverting AcCoA from the TCA-cycle toward acetate. At the end of the batch phase, around 15 h, when all the glucose is consumed, the secreted acetate began to be re-used by the cells (Figure 6.12).

Contrary to the fed-batch results presented in Chapter 4, neither glucose nor acetate accumulation occur along the fed-batch phase. Moreover the concentration of the biomass never stops growing, even after 6.5 h of induction. The fermentation was stopped at a final biomass of $54 \pm 3 \text{ gDCW}\cdot\text{L}^{-1}$ due to aeration limitations. Indeed, the fermentation ended with a final volume of 1.3 L with a final pO_2 value of 10 %.

The evolution profiles of the FucA production are shown in Figure 6.12. FucA levels reached a maximum value of $201 \pm 18 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$, with an activity of $1176 \pm 19 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$. Moreover, an example of the evolution of FucA concentration together with the production of SHMT protein can be seen in Figure 6.13. As it can be clearly seen, there is a constitutive overexpression of the SHMT protein both in the pre-induction and along the induction phase (Table 6.2 and Figure 6.13). However, by comparing the values of SHMT presented in table 6.2 with the ones of table 4.1, it can be observed a reduction of the expression levels of the *glyA* gene. As an example, after 4 h of induction SHMT mass are 85 ± 0.5 and 132 ± 4 for the Puzzle and M15[pREP4] strains, respectively. This 1.6-fold reduction of the accumulation of the SHMT in the cytoplasm, moving from the P3 to the J23110 promoter, leads to an improvement of

the metabolic burden imposed to the cells. This improvement is reflected in the physiological state of the cells that are less damaged.

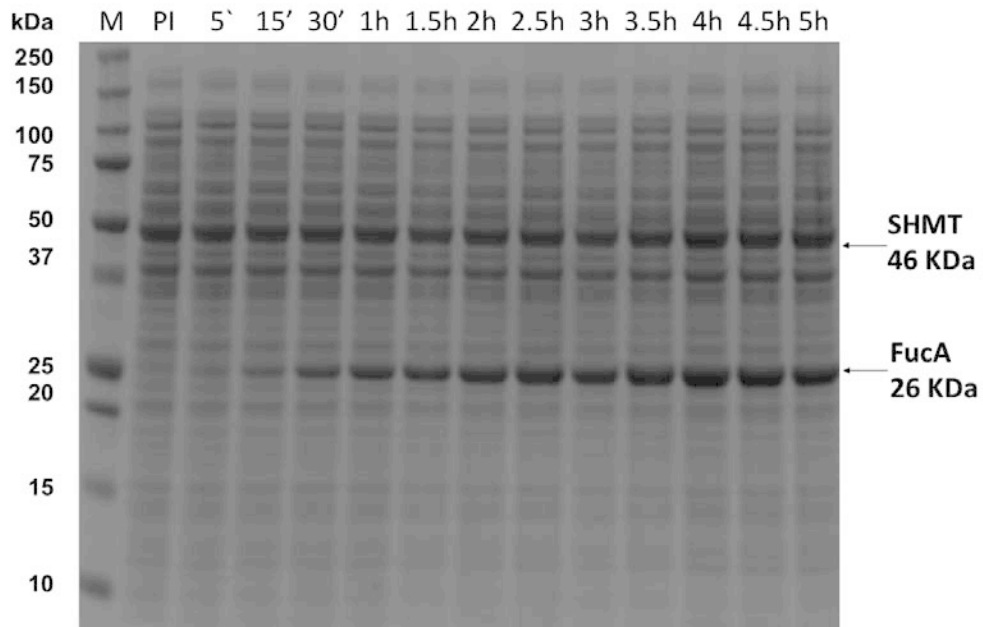


Figure 6.13 SDS-PAGE of Fed-Batch culture's samples of the M15 Δ *glyA* pQE-FucA puzzle strain: [IPTG], 70 μ M; X_{ind} , 20 g·L⁻¹; μ 0.22 h⁻¹. Lane M: molecular weight marker; PI, pre-induction; from 5' to 5h correspond to the time after induction. The 26 kDa FucA and the 46 kDa SHMT bands are indicated.

Table 6.2 Values SHMT and FucA protein ($\text{mg}\cdot\text{gDCW}^{-1}$) along the induction phase. PI, represents the pre-induction sample, while samples from 5` to 4 h are those taken along the induction phase.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)	FucA mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	94 ± 1	6 ± 0.4
5`	98 ± 1	9 ± 0.3
15'	93 ± 1	22 ± 0.3
30'	92 ± 7	45 ± 4
1h	84 ± 0.4	73 ± 2
1.5 h	87 ± 4	99 ± 2
2h	83 ± 4	121 ± 3
2.5h	79 ± 3	124 ± 5
3 h	82 ± 0.4	134 ± 5
4 h	85 ± 0.5	164 ± 1
4.5 h	83 ± 2	175 ± 5
5 h	80 ± 2	143 ± 5
5.5 h	82 ± 2	181 ± 18
6 h	81 ± 1	167 ± 16
6.5 h	81 ± 0.5	202 ± 18

6.3.3.1 Off-gas analysis

Puzzle cells grew under the same conditions described previously. The results of the off-gases of fed-batch fermentations are shown in Figures 6.14 and 6.15.

The partial pressure of oxygen (pO_2) decreases along the batch phase exponentially up to the set point of 50 %, where an agitation controller is started in order to maintain the level. The exponential drop of the pO_2 curve indicates the rapid increase of the respiration capacity of the bacteria. After about 11 h, the complete consumption of the $10 \text{ g}\cdot\text{L}^{-1}$ of initial glucose indicates the end of the batch phase and it is characterized by an increase of both pO_2 and pH values (Sooan et al. 2009). Once finish the glucose, the bacteria starts to assimilate the acetate produced during the

batch phase due to the glycolysis overflow. Along this acetate switch phase, the pH value increased, reaching value of 7.2, as a result of the consumption of the acid from the media (Figure 6.14). Besides the end of the batch can be followed in Figure 6.15 observing both the oxygen (O_2) and carbon dioxide (CO_2) curves that after exponential profiles, suddenly increase and decrease, respectively.

Along the fed-batch and induction phase continuous feeding of a high concentrate glucose solution was added to the media in order to maintain the fixed growth rate of 0.22 h^{-1} . Contrary to the reference strain M15[pREP4] and the auxotrophic M15 Δ glyA[pREP4], presented in Chapter 4, no glucose accumulation is detected all along the fed-batch and induction phase. In fact, in figure 6.14, it can be seen how the turn off of the feeding is closely correlate with the immediate increase of the pO₂ and the pH values. In fact, the increase of the pO₂ indicates that the cells stop to consume oxygen and to grow; while the increase of the pH signifies that the cells are metabolizing the acetate, meaning no glucose is accumulated.

When the cells reached and $6\text{ gDCW}\cdot\text{L}^{-1}$ IPTG was added to the culture in order to induce the recombinant protein production. It was possible to maintain the fermentation till the 7.5 h of induction. At the end of the experiments it is possible to see that the pO₂ levels decrease to reach 2 % due to the limitations of the aeration system and that the pH decreases reaching value of 6.7 (Figure 6.14).

RQ value increases along the batch phase and then is maintained at 1 (Figure 6.15). This means that when the sugar-feeding rate is controlled, there is no production of intermediate products.

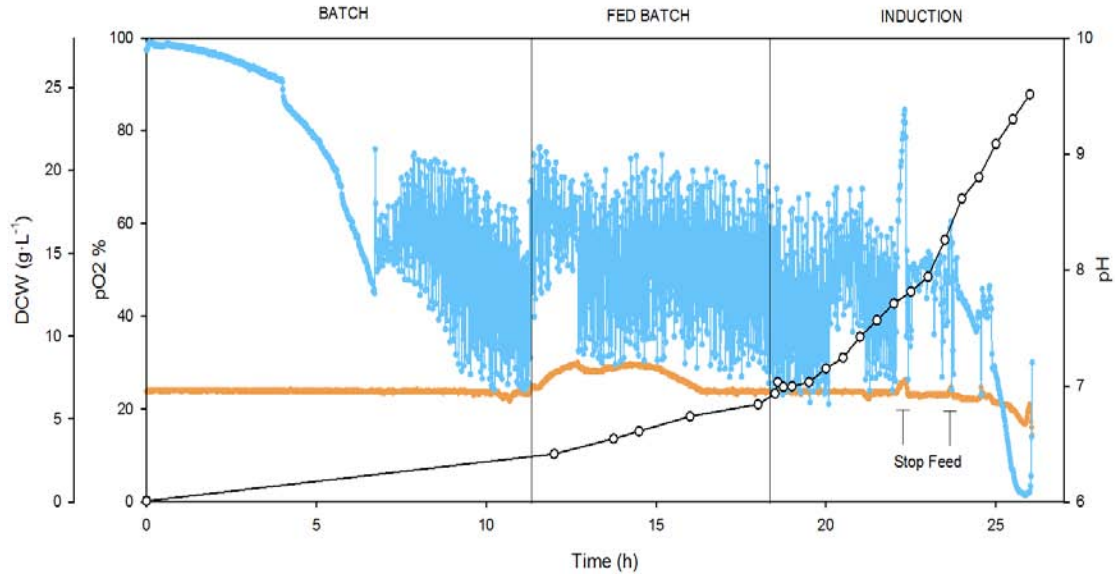


Figure 6.14 *E. Coli* M15Δ*glyA* pQE-FucA puzzle fed-batch culture: [IPTG], 70μM; X_{ind} , 6 g·L⁻¹; μ 0.22 h⁻¹. pO₂, in light blue, was set at 50% and controlled by a gas flow rate of 3 L·min⁻¹ of air and agitation. The pH, in orange, was set at 7 and maintained by adding 15% (v/v) NH₄OH solution. (c) Biomass DCW (g·L⁻¹) along time. Batch, Fed-Batch and induction phases are indicated. The arrows indicate the stop of the feeding.

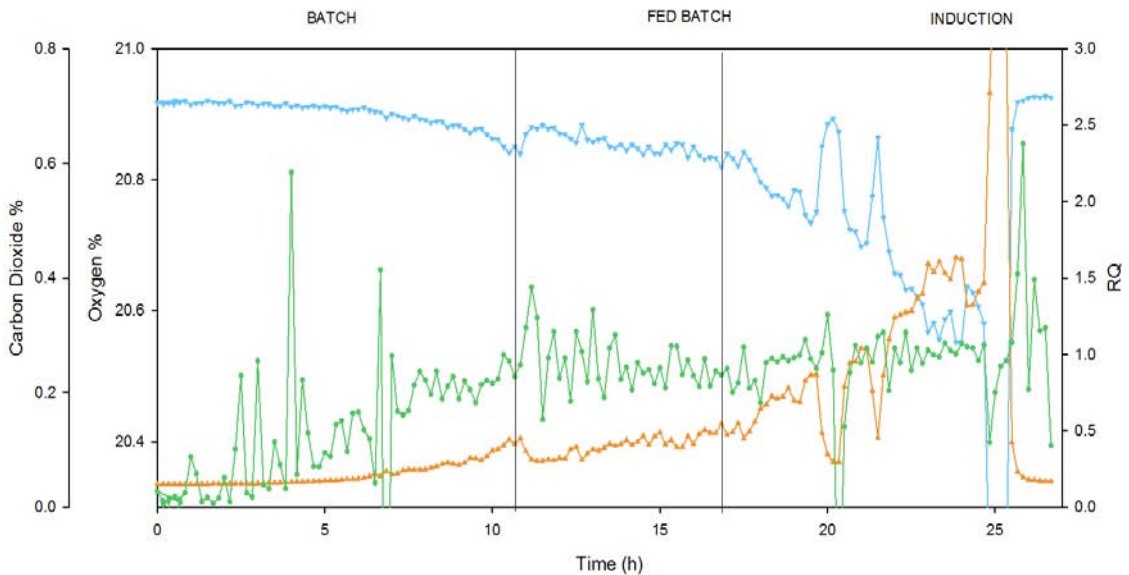


Figure 6.15 Gas-MS data of M15Δ*glyA* pQE-FucA puzzle Fed-Batch culture: [IPTG], 70μM; X_{ind} , 6 g·L⁻¹; μ 0.22 h⁻¹. % of oxygen in light blue, % of carbon dioxide in orange and the respiratory quotient (RQ) in green. Batch, Fed-Batch and induction phases are indicated.

6.3.3.2 Flow cytometry analysis of the Puzzle strain

Cell samples along the fed-batch culture were collected for flow cytometer analysis. In Figure 6.16 FCM data represented the % of live, damaged or dead cells determined by BOX/P.I. co-staining. A part from an initial perturbation, probably due to the stress of moving from the shake flask to the reactor and from the presence of the antifoam that could interfere with the analysis, the cells results are homogeneous and for the 95 % live and healthy almost till the end of the fermentation.

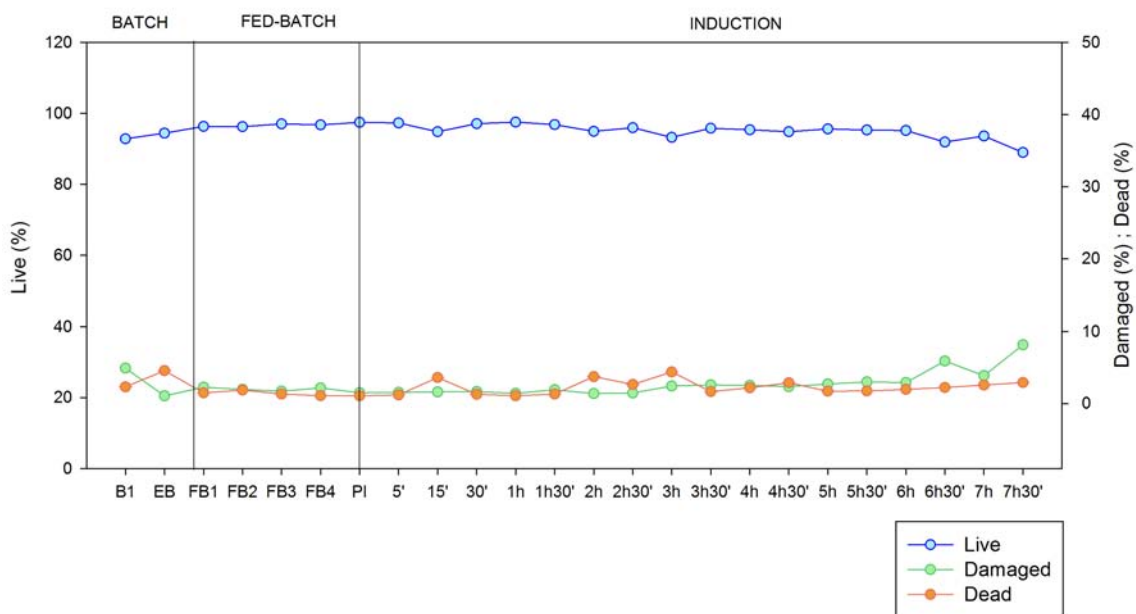


Figure 6.16 FCM analysis of *E. coli* Puzzle fed-batch per duplicate: [IPTG], $70\mu\text{M}$; X_{indr} , $6\text{ g}\cdot\text{L}^{-1}$; μ 0.22 h^{-1} . FCM data represents live (blue), damaged (green) or dead (orange) cells determined by BOX/P.I. co-staining (See Materials and Methods section 3.7.4.2). B1 and EB, first and the last samples of the batch, respectively; FB1, FB2, FB3 and FB4, four samples of the fed-batch phase; PI, pre-induction and 5 to 7h30', different induction samples. Batch, Fed-Batch and induction phases are indicated.

Only after 6.5 h of induction a slight increase of both damaged and dead cells appears, with values of 8 % and 3 %, respectively. This decrease is probably due to the non-

optimal conditions at the end of the fermentation, due to the low values of pO_2 and pH at the end of the fermentation.

This stress caused to the cells can be observed also in figure 6.17, where a slightly increase of the % of CR^+ population, representing the % of inclusion bodies present inside the cells, appears at the of the induction.

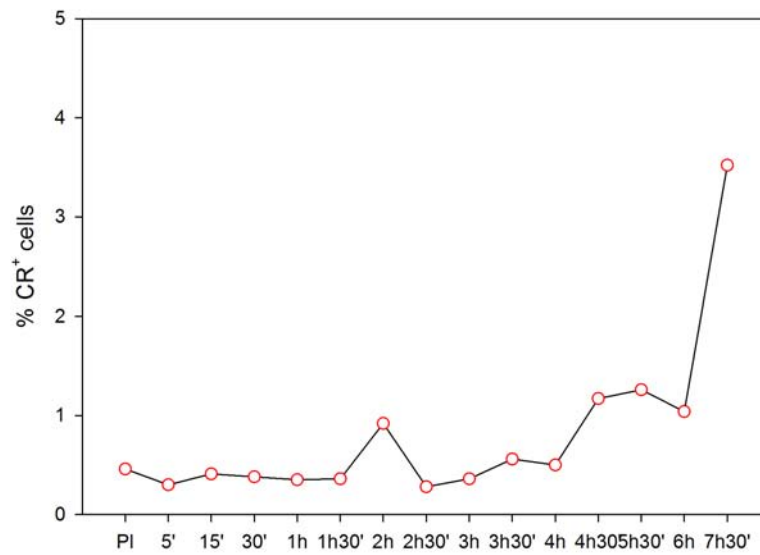


Figure 6.17 FCM analysis of *E. coli* M15 Δ glyA pQE-FucA puzzle fed-batch culture: [IPTG], 70 μ M; X_{ind} , 6 g \cdot L⁻¹; μ 0.22 h⁻¹. % CR⁺ cells from FCM analysis. Samples stained with CR for 1 hour (See Materials and Methods section 3.7.4.3). PI, pre-induction; 5 to 7h30, induction samples.

6.4 Conclusion

These results suggest that transcriptional tuning of *lacI* expression levels is a key factor in a better *fucA* expression regulation, leading to a higher FucA specific activity. Fine-tuning the co-expression of the two genes allowed to i) reduce the metabolic burden related to plasmid-encoded genes and, ii) optimize the regulation and induction of the foreign gene expression, when engineering parts of the reference two-plasmid system into a single plasmid.

Besides, the transcriptional tuning of *lacI* expression levels is a key factor to improve *fucA* expression regulation, leading to a higher FucA specific activity. Moreover, the tuning of *glyA* levels has a positive effect on the reduction of the metabolic load due to expression of plasmid-encoded genes (also reflected in the reduced acetate production). In fact, SHMT values were almost 50% reduced comparing the Puzzle with the M15 Δ *glyA*[pREP4] strains. These results are in accordance with the observation by Mairhofer et al. (Mairhofer et al. 2013), who demonstrated that the folding machinery is severely overstrained in the plasmid-based expression system compared with the plasmid-free cells due to the different transcriptional profiles.

7 Results IV: Elimination of the antibiotic resistance encoding gene from the single-plasmid expression system

An alternative selection strategy based on auxotrophic selection for recombinant FucA production was presented in Chapter 4. The principle of the M15 Δ *glyA* system is based on the chromosomal deletion of the *glyA* gene that is complemented by its cloning into expression vector.

Besides, in Chapter 6, an expression cassette was constructed where the *lacI* and *glyA* genes have been placed under the control of a set of synthetic constitutive promoters in order to obtain the sufficient *lacI* inhibitor protein to minimize “promoter leakiness” and the minimal *glyA* transcriptional levels needed for plasmid maintenance and optimal cell growth in defined media. The *lacI-glyA* cassette cloned under the J23110 constitutive promoter has been selected and cloned into the pQE-FucA plasmid yielding a single-vector expression system as described in Chapter 6.

In this Chapter, the β -lactamase gene (*bla*) has been removed from the pQE40 backbone with the aim to further improve the system robustness and in order to obtain an expression system completely devoid of antibiotic resistance genes;. Finally, the capacity for a recombinant FucA overexpression has been investigated in shake flask and fed-batch cultures in DM medium.

7.1 Deletion of the antibiotic resistance gene and Amp^R strain characterization

The ampicillin resistance gene (*bla*) was eliminated from the pQE-FucA_puzzle (J23110) plasmid using the Klenow fragment-based blunting DNA technique (See Materials and Methods Section 3.4.7).

After selection of transformants in DM plates, 24 clones were tested in order to select the best clones in terms of growth and FucA production after 1h of induction (induced with 1mM IPTG). Subsequently, characterization experiments were scaled-up to 500 mL shake flasks with 100 mL of defined media. The induction period was 4h the work was performed for four selected clones (Figure 7.2)

As shown in the figure 7.1A, the four clones present a similar growth and glucose consumption profiles. Besides, in the pre-induction samples all the four transformants show a relatively lower production ($\approx 20 \text{ mgFucA}\cdot\text{gDCW}^{-1}$) compared to previously constructed strains. Finally, FucA production after 4 h of induction presented values around $257 \pm 22 \text{ mg FucA}\cdot\text{gDCW}^{-1}$, like in the case of the clone #8 (Figure 7.1A).

7 Results IV: Elimination of the antibiotic resistance encoding gene from the single-plasmid expression system

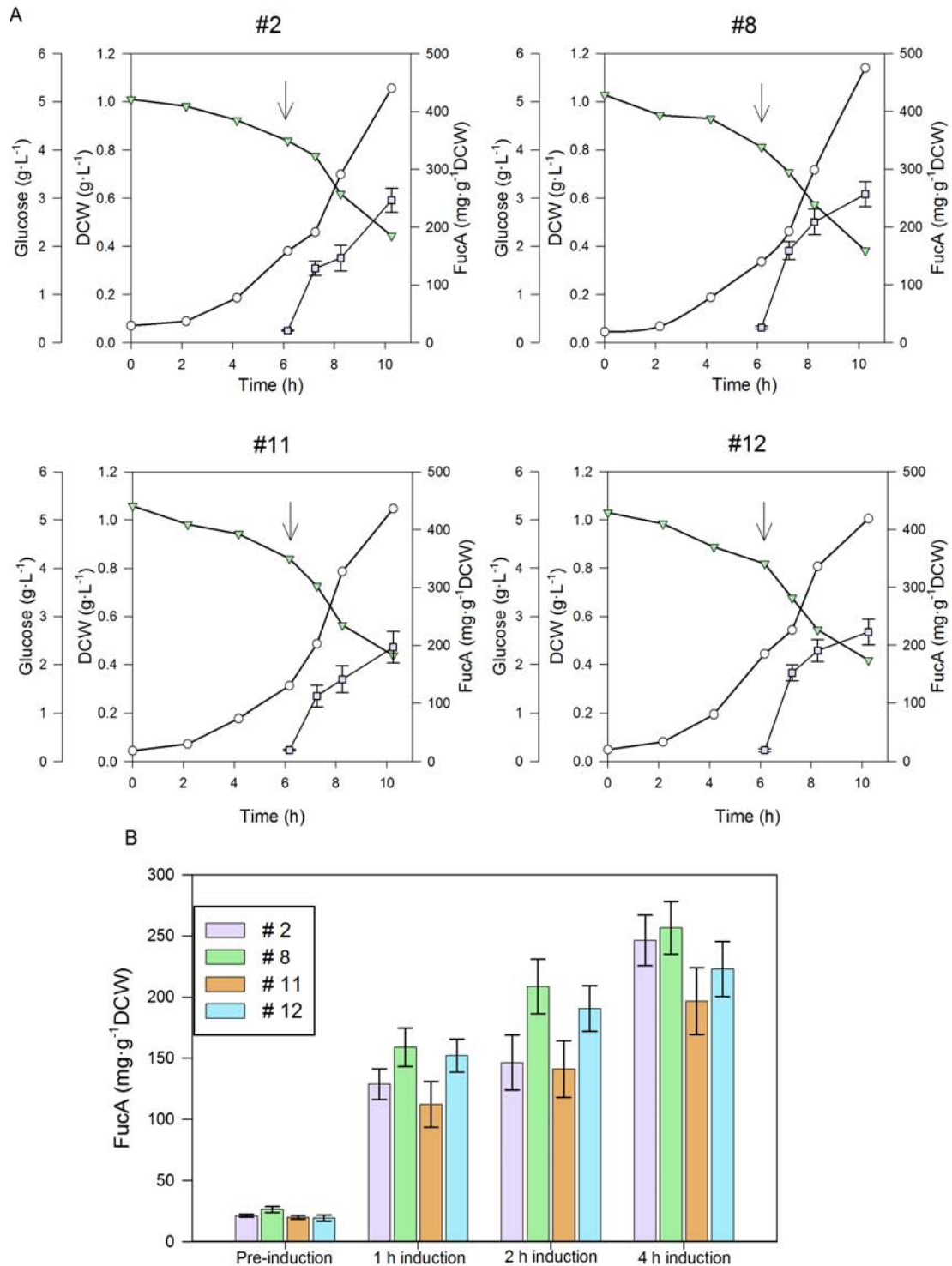


Figure 7.1 Shake flask cultures of 4 M15ΔglyA pQE-FucA_puzzle AmpR⁻ clones (#2, #8, #11, #12). **A**) (●) Biomass DCW (g·L⁻¹), (■) specific mass (mgFucA·g⁻¹DCW⁻¹); (▼) Glucose concentration (g·L⁻¹). 1mM IPTG pulse is indicated by an arrow. **B**) Comparison of specific mass production (mgFucA·g⁻¹DCW) for the 4 transformants for the samples pre induction, 1h 2h and 4h induction.

7.1.1 AmpR- shake flask cultures

E. coli AmpR⁻ was assessed in shake-flask cultures per triplicate in DM medium without antibiotic supplementation.

The time-profiles of biomass, glucose consumption, acetate and FucA (mass and specific activity) were analyzed. Results are presented in Figure 7.2.

A slightly decrease in the μ_{\max} $0.41 \pm 0.01 \text{ h}^{-1}$ was detected compared to the strains M15[pREP4] and Puzzle, which showed a μ_{\max} of $0.49 \pm 0.02 \text{ h}^{-1}$ and $0.45 \pm 0.01 \text{ h}^{-1}$, respectively (as shown in Chapter 4 and 6, respectively). In terms of FucA production, the point of maximum activity corresponds to $1309 \pm 42 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ with a production in mass of $219 \pm 5 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$ after 4 h of induction (Figure 7.2A).

Finally, SHMT values ($\text{mgSHMT}\cdot\text{g}^{-1}\text{DCW}$) were calculated for each time point of induction in shake flasks cultures. The results are presented in Table 7.1, where the pre induction sample with $53 \pm 1 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$ represents the sample with the higher amount of SHMT. It can be seen that this pre-induction values decreased more than 1.7-fold comparing with M15 Δ *glyA*[pREP4], being $92 \pm 14 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$ (Table 4.2); and 1.2-fold comparing with Puzzle strain, $66 \pm 17 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$ (Table 6.1).

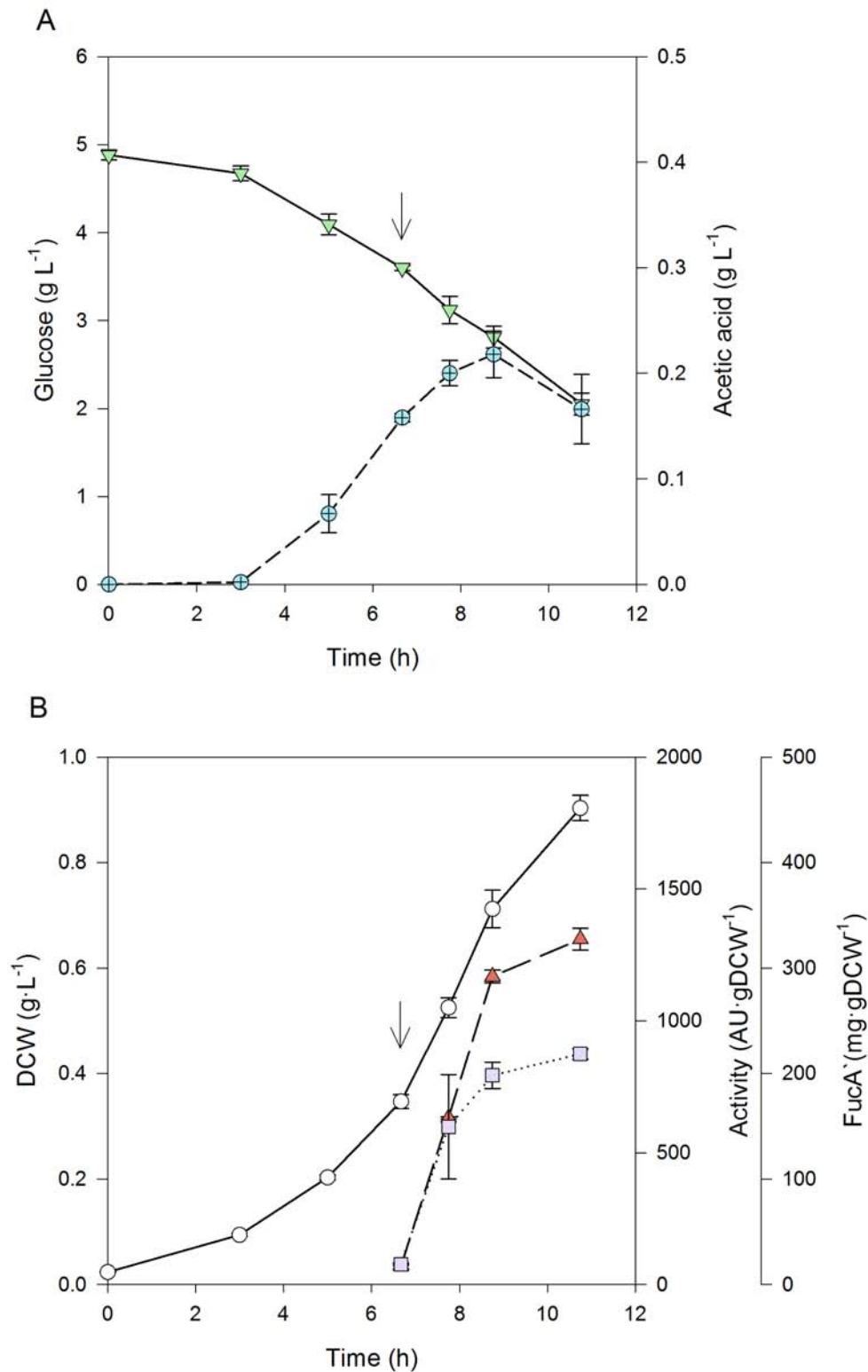


Figure 7.2 Shake flask cultures of *E. coli* AmpR⁻ strain in DM medium. **A**) (●) Biomass DCW (g·L⁻¹), (▲) FucA activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹); **B**) (▼) Glucose concentration (g·L⁻¹) and (●) Acetic Acid (g·L⁻¹). 1mM IPTG pulse is indicated by an arrow.

Table 7.1 SHMT production ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) along the induction phase for the AmpR^- strain. PI (pre-induction) and 1 h, 2 h and 4 h correspond to the time after induction.

	SHMT mass ($\text{mg}\cdot\text{gDCW}^{-1}$)
PI	53 ± 1
1h	44 ± 11
2h	48 ± 6
4h	50 ± 10

7.1.2 Flow cytometry analysis of AmpR^- strain

Cell viability of the AmpR^- strain growing in shake flasks was monitored by flow cytometry analysis, as performed with previous expression strains.

Only two experimental conditions varying inducer concentration were tested: A) 1mM IPTG and B) 70 μM IPTG, since this strain do not grow in presence of ampicillin antibiotic.

As an example, profiles of the AmpR^- strain growing in shake flask cultures in DM with an induction pulse of 1 mM IPTG, are presented in Figure 7.3.

As observed for previous strains in Chapters 4 and 6, growth rate started to slow down just after IPTG induction. The values μ , calculated from the pre-induced to the 4 h post-induction samples, decreased from 0.41 to 0.20 h^{-1} during this period. This reduction is confirmed also by cell concentration values ($\text{cells}\cdot\text{mL}^{-1}$) that falls after the 2 h of induction.

Moreover, at the end of the cultivation (stationary phase), whereas the value of DCW ($\text{g}\cdot\text{L}^{-1}$) remain constant, cell concentration decreased more than 10 % from 4 to 18 h of induction, as indicated in Figure 7.3. This has been observed previously for the M15[pREP4] and M15 ΔglyA [pREP4] (Chapter 4, section 4.1.1.1 and 4.1.2.1, respectively) and for the Puzzle strain (Chapter 6, section 6.3.2), where post-induction cell concentration ($\text{cells}\cdot\text{mL}^{-1}$) values decrease after 4 to 18 h of induction, whereas DCW ($\text{g}\cdot\text{L}^{-1}$) remains unchanged.

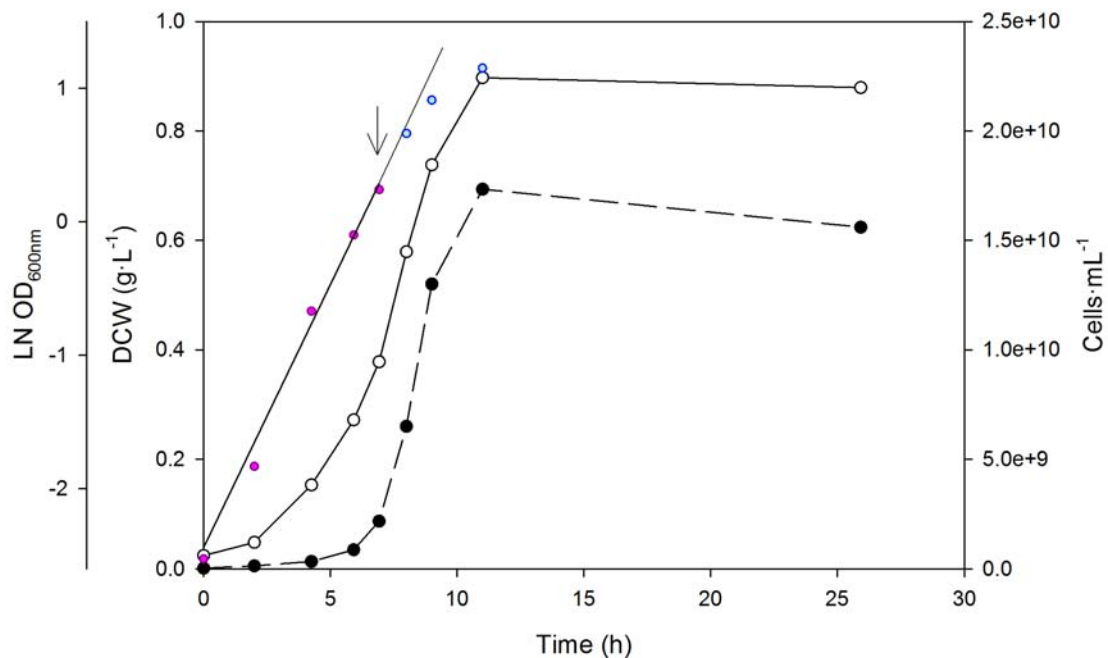


Figure 7.3 Shake flasks culture of the *E. coli* AmpR⁻ strain in defined media plus antibiotic. (●) Biomass DCW ($\text{g}\cdot\text{L}^{-1}$), (●) concentration ($\text{cell}\cdot\text{mL}^{-1}$), (●) LN OD₆₀₀ of the pre-induction phase and (●) LN OD₆₀₀ of the post-induction phase. 1mM IPTG induction is indicate with an arrow.

Furthermore, the forward Scatter (FSC) signal of the 18 h post-induction sample shifted to the right compared to the curve of the reference sample of pre-induction. As explained in Chapters 4 and 6, this increment of the forward scatter may indicate that

the cells undergo a morphological change of their shape after being induced and therefore, becoming more elongated (Figure 7.4A). This change does not occur in the negative control cultivation without induction, where the FSC curves corresponding to pre-induction and non-induced sample after 18h overlap (Figure 7.4B).

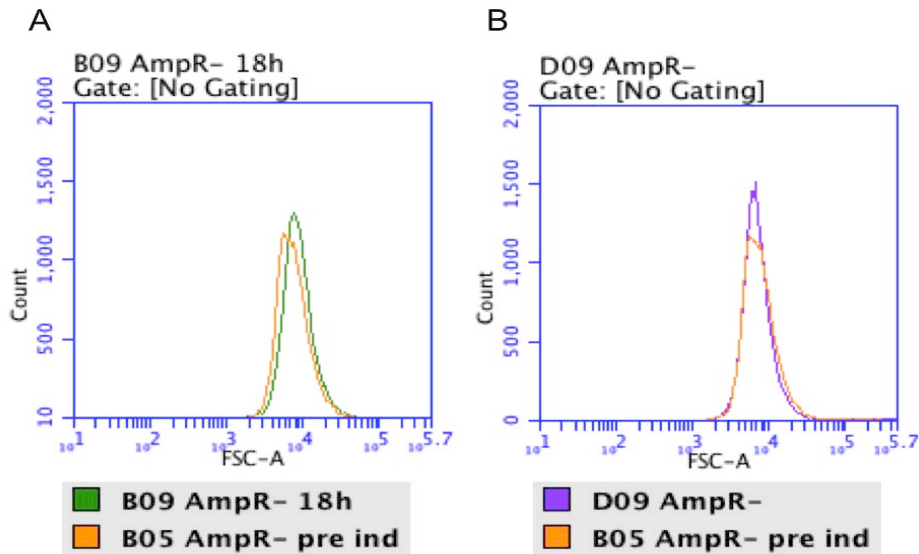


Figure 7.4 Flow cytometry analysis of the forward-scatter light (FSC) of the samples of the pre induction (pre ind, in orange) and of A) 18 hours after induction of both samples induced (18h IPTG, green) and B) negative control not induced (AmpR- purple). Samples of shake flasks cultures of *E. coli* M15 Δ glyA pQE-FucA puzzle strain in DM media plus antibiotic, 37°C with agitation.

Analysis of samples by flow cytometry using as fluorescent dyes BOX and P.I., are represented in Figure 7.5. The population of the pre-induction time point, representing the sample of 6.9 h of growth in shake flasks, correspond to the exponential phase, was relatively homogenous and 95.8 % healthy (Figure 7.5, quadrant Q1 of the pre-induction sample plot).

As the induction time progresses from 1 h to 4 h, the percentage of healthy population maintains values higher than 90 %. However, at the 4 h of induction it can be seen a change in the population, which presents 64.0 % viable, 5.7 % damaged and 23.3 % dead populations. The viability of the 18 h sample remained unchanged, where 6.0 % of the cells are damaged whereas 17.5 % of the population is permeable to both BOX and P.I. (quadrant Q3 of 18h sample).

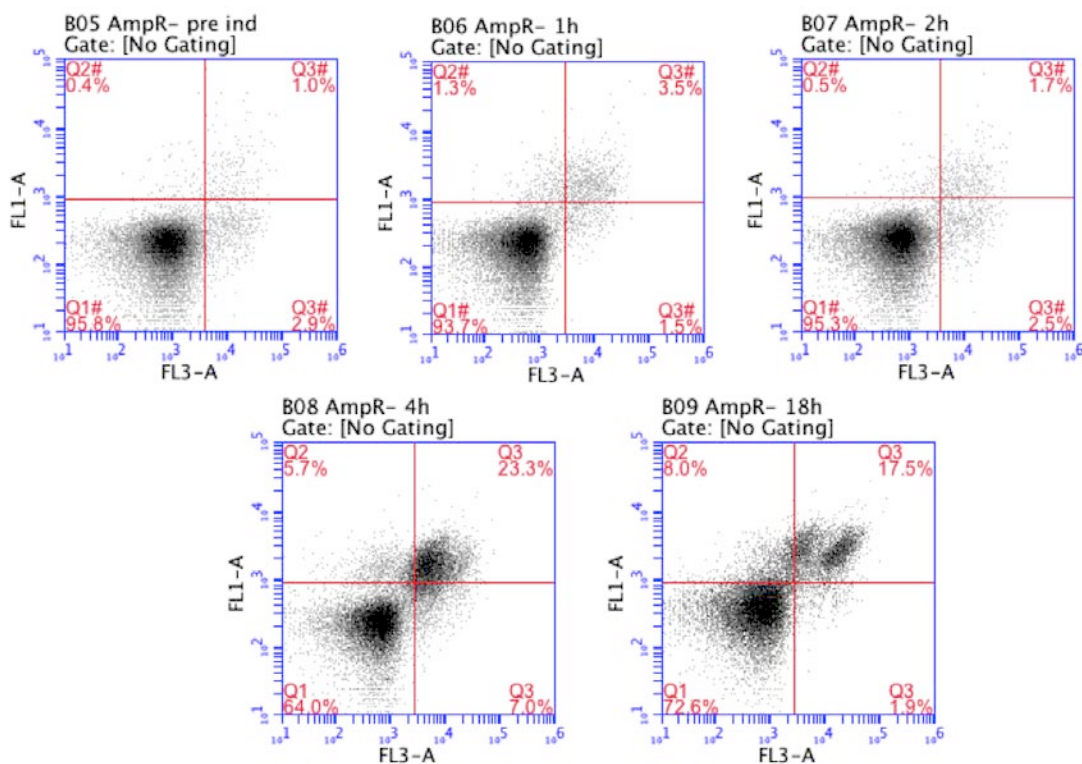


Figure 7.5 Flow cytometry analysis of *E. Coli* M15Δ*glyA* pQE-FucA puzzle Amp^R. Shake flask cultures performed in DM medium plus antibiotic. Pre induction (pre ind), after 1 h, 2 h, 4 h and 18 h induction are represented. Samples were co-stained with BOX/P.I. (See Materials and Methods section 3.7.4.2). The numbers in each quadrant show percentage of bacteria population in each subgroup.

Figure 7.6 summarizes the comparison of two induction conditions tested: A) 1 mM IPTG and B) 70 μM IPTG. As it can be observed, the experiment with 1 mM IPTG

induction leads to a higher proportion of damaged and dead cells. Co-staining analysis provides clear evidence that the reduction of the inducer concentration from 1 mM to 70 μ M has a positive effect on cell physiology, varying from 30 % and 6 % (Figure 7.6A) to 5 % and 1 % (Figure 7.6B) of dead and damaged cells after 4 h of induction, respectively.

Moreover, the examination of membrane potential of cells induced with a pulse of 70 μ M IPTG, shows how this value remains less than 1 % along the whole culturing time (Figure 7.6B). Only after 4 h of induction, around 5 % of the population was permeable to both P.I. and BOX indicating loss of viability. Furthermore, plating post-induction samples revealed that more than 90 % of the bacteria were able to form colonies on selective agar. These results mean a plasmid segregational stability along the induction phase.

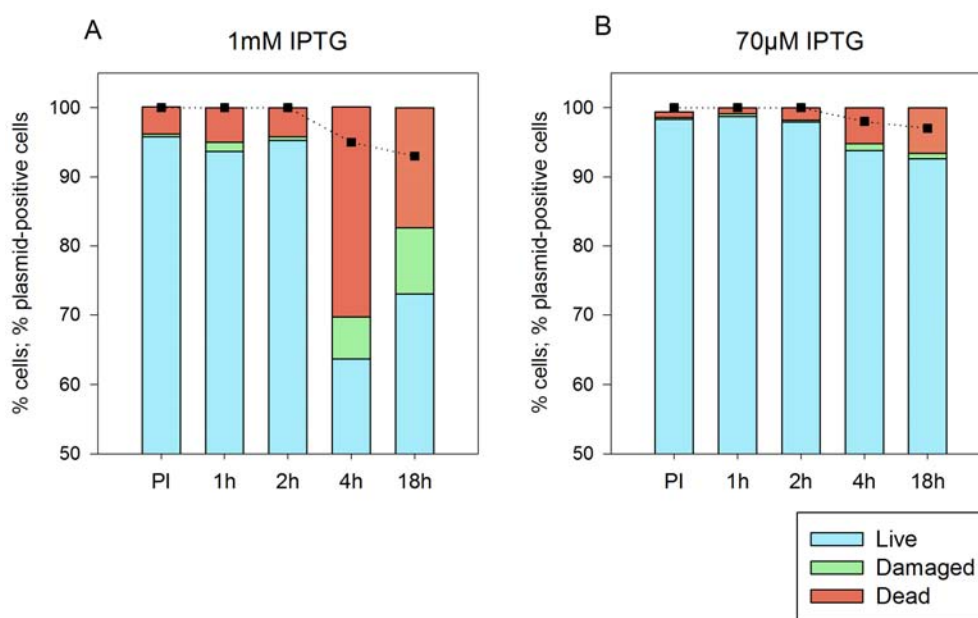


Figure 7.6 Comparison of the two experimental conditions in shake flasks cultures for the *E. coli* Amp^R strain. PI, pre-induction samples; 1 h, 2 h, 4 h and 18 h represent the time after induction. A) DM and 1 mM IPTG; B) DM and 70 μ M IPTG. Live cells in light blue, damaged cells in green and dead cells in orange; (n) plasmid retention (%).

7.2 Fed-batch cultures of Amp^R strain

To evaluate the FucA production potential in high-cell density cultures of *E. coli* M15Δ*glyA* pQE-FucA puzzle Amp^R under controlled conditions, fed-batch fermentations were performed. As described in previous Chapters 4 and 6, triplicate fed-batch cultures were performed in defined media with glucose as a sole carbon source (Figure 7.7). After inoculation, a growth rate of 0.41 h⁻¹ was observed. Along the batch phase, it can be observed a constant increment of acetate in the culture media yielding a final concentration of 0.9 g·L⁻¹. After 13 h of cultivation, when all the 20 g·L⁻¹ of glucose and the 0.9 g·L⁻¹ of secreted acetate was consumed, the feeding phase was started (indicated as the fed batch phase in figure 7.7). All along the fed-batch and induction phase a constant specific growth rate of 0.22 h⁻¹ was maintained with a pre-defined exponential feeding profile. As for the previous experiments, in Chapters 4 and 6, a pulse of 70 μM IPTG was added once reached 20 gDCW·L⁻¹ of biomass, in order to induce the overexpression of the FucA protein. The time-profiles of FucA mass and specific activity are represented in Figure 7.7. The cultivation was finished after 7 h of induction, for a total of 26.3 h of fermentation, when the concentration of dissolved oxygen dropped below 20 %. At the end of the process, biomass concentration reached 53.5 ± 0.7 gDCW·L⁻¹. This end point corresponds with the maximum FucA levels of production, being 245 ± 13 mgFucA·g⁻¹DCW, with an activity of 1322 ± 19 AU·g⁻¹DCW.

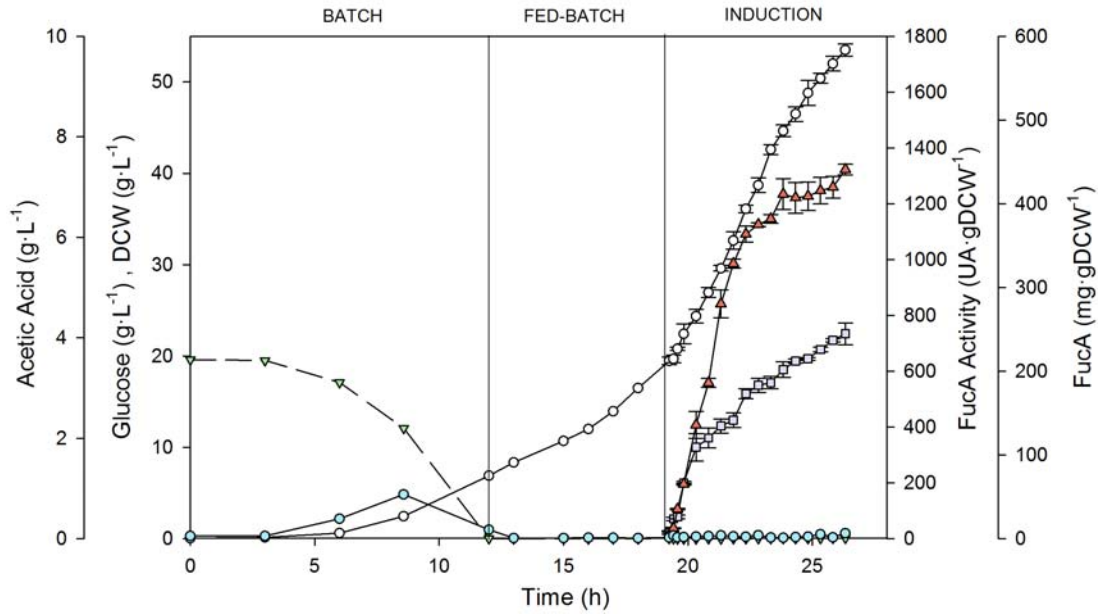


Figure 7.7 AmpR⁻ fed-batch culture: [IPTG], 70 μ M; X_{ind} , 20 $g \cdot L^{-1}$; μ 0.22 h^{-1} . Schematic representation of (●) Biomass DCW ($g \cdot L^{-1}$), (▼) Glucose concentration ($g \cdot L^{-1}$), (▲) FucA activity (AU·gDCW⁻¹), (□) specific mass (mgFucA·gDCW⁻¹) and (○) Acetic Acid ($g \cdot L^{-1}$) along time. Batch, fed-batch and induction phases are indicated.

Bands of soluble protein in cytoplasm are shown in SDS-PAGE analysis (Figure 7.8). Besides, the evolution of FucA concentration together with the amount of SHMT protein can be observed in table 7.2. As expected, there is a constitutive synthesis of SHMT protein, whereas the production of FucA is ascending. By comparing SHMT values between the AmpR⁻ and the Puzzle strains (Table 7.1 and Table 6.2 respectively) it can be seen that the expression levels of the *glyA* gene are slightly lower. Moreover, the J23110 promoter leads to a reduction of the *glyA* transcription levels that brings to an improvement of the metabolic load imposed to the cells. This improvement can be identified in the absence of both glucose and acetate accumulation along the fed-batch and induction phases.

7 Results IV: Elimination of the antibiotic resistance encoding gene
from the single-plasmid expression system

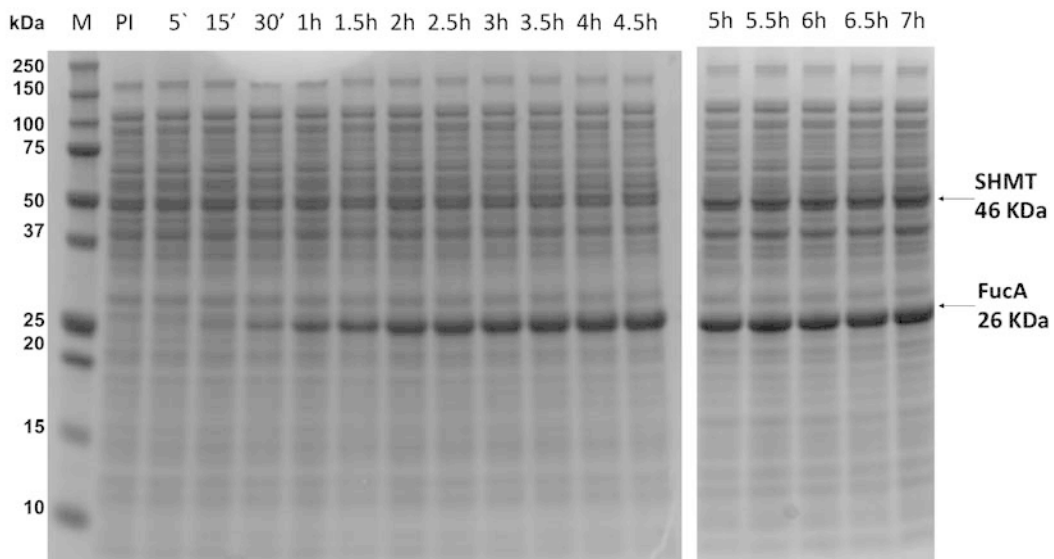


Figure 7.8 SDS-PAGE of Amp^R Fed-Batch culture: [IPTG], 70 μ M; X_{ind} , 20 g \cdot L⁻¹; μ 0.22 h⁻¹. Lane M: molecular weight marker; PI, pre-induction; from 5' to 7 h correspond to the time after induction. The 26 kDa FucA and the 46 kDa SHMT are indicated in the figure.

Table 7.2 SHMT and FucA protein (mg \cdot gDCW⁻¹) of Fed-Batch culture's samples of the Amp^R strain. PI, pre-induction, from 5' to 7 h are the samples along the induction phase.

	SHMT mass (mg \cdot gDCW ⁻¹)	FucA mass (mg \cdot gDCW ⁻¹)
PI	78 \pm 9	9 \pm 1
5'	76 \pm 8	24 \pm 2
15'	68 \pm 10	26 \pm 2
30'	75 \pm 8	67 \pm 2
1h	82 \pm 11	109 \pm 16
1.5 h	72 \pm 7	120 \pm 12
2h	82 \pm 8	134 \pm 8
2.5h	67 \pm 9	142 \pm 9
3 h	78 \pm 14	173 \pm 6
3.5 h	72 \pm 7	183 \pm 8
4 h	72 \pm 13	186 \pm 7
4.5 h	81 \pm 13	202 \pm 9
5 h	80 \pm 11	212 \pm 2
5.5 h	79 \pm 9	215 \pm 2
6 h	76 \pm 10	226 \pm 3
6.5 h	77 \pm 7	237 \pm 2
7 h	80 \pm 9	245 \pm 13

7.2.1 Off-gas analysis

Off-gas analysis was performed for the AmpR⁻ strain, per duplicate using the 5L Fermac 310/60 bioreactor. Results are presented in Figures 7.9 and 7.10. The batch phase is characterized by the exponential decrease of the pO₂ up to the set point of 50 %, where the agitation controller starts.

At the 11 hours of cultivation the concentration of dissolved oxygen and the pH started to increase. Those events were considered as markers of glucose exhaustion from the culture broth, and indicated the start of the feeding phase. The increment of pH to 7.2 indicates the acetate switch phase, where the bacteria, due to the absence of an alternative carbon source, consume the acetic acid.

Furthermore, the end of the batch can be identified in figure 7.10, where the values of the oxygen increase and the values of the carbon dioxide decrease, indicating a cellular respiratory change.

Along the fed-batch and induction phase continuous feeding of a high concentrate glucose solution was added to the media in order to maintain the fixed growth rate of 0.22 h⁻¹, by manually increasing the flow rate of the feeding pump. The induction is performed around the 17 hours of the culture, where the cells reach 6 g·DCWL⁻¹. No glucose accumulation is detected all along the fed-batch culture, indicating constant carbon source consumption. Decrease in the O₂ trend and increase in the CO₂ trend continued until the end of the cultivation. The end of the fermentation is characterized by an exhaustion of the pO₂ values, due to the high oxygen demands of the culture with a final DCW of 23 g·L⁻¹.

7 Results IV: Elimination of the antibiotic resistance encoding gene
from the single-plasmid expression system

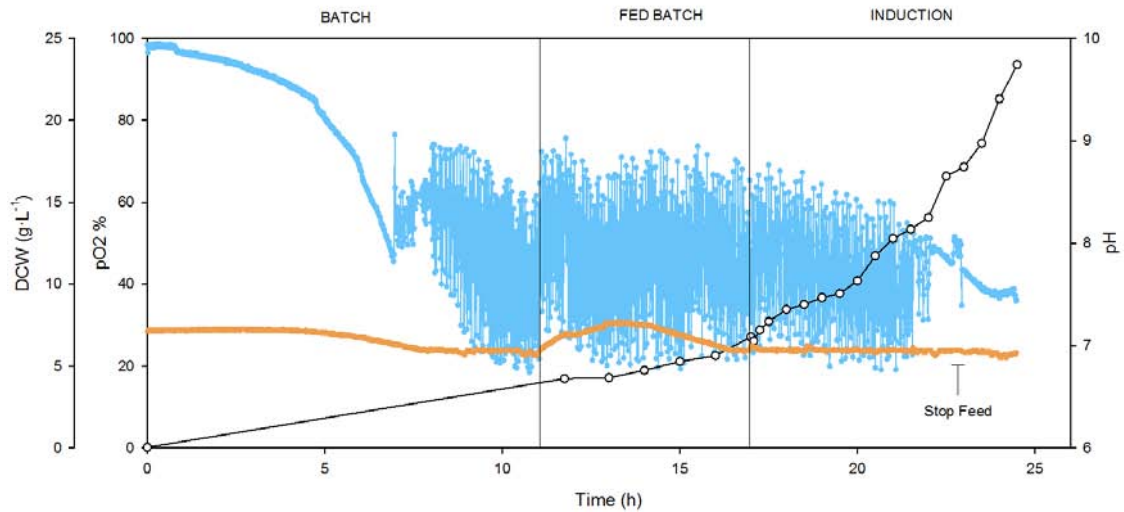


Figure 7.9 *E. coli* AmpR⁻ fed-batch culture: [IPTG], 70 μ M; X_{ind} , 6 g·L⁻¹; μ 0.22 h⁻¹. pO₂, in light blue, was set at 50 % and controlled by a gas flow rate of 3 L·min⁻¹ of air and agitation. The pH, in orange, was set at 7 and maintained by adding 15% (v/v) NH₄OH solution. (c) Biomass DCW (g·L⁻¹) along time. Batch, Fed-Batch and induction phases are indicated. The arrow indicates the stop of the feeding.

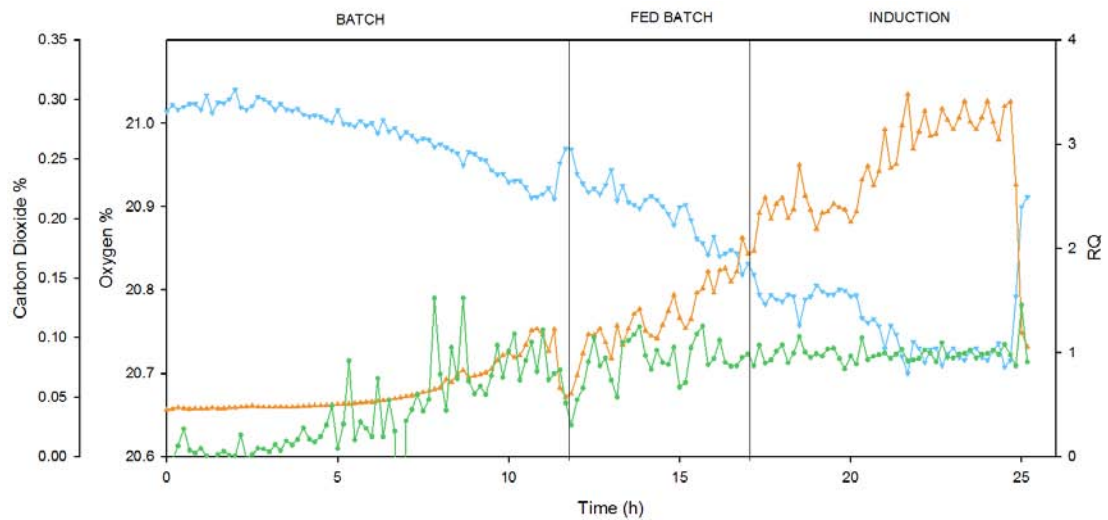


Figure 7.10 Gas-MS data of *E. coli* AmpR⁻ Fed-Batch culture: [IPTG], 70 μ M; X_{ind} , 6 g·L⁻¹; μ 0.22 h⁻¹. % of oxygen in light blue, % of carbon dioxide in orange and the respiratory quotient (RQ) in green. Batch, Fed-Batch and induction phases are indicated.

7.2.2 Flow cytometry analysis

Flow cytometry analysis of fed-batch samples demonstrated that no changes in *E. coli* physiological state appear with respect to BOX/P.I. co-staining all along the cultivation (Figure 7.11). Apart from an initial perturbation of the cells due to the presence of the antifoam in the media, from the samples taken during the batch and fed-batch phases co-stained with BOX/P.I. only one population of cells could be identified. No change in fluorescence was measured, indicating no cell taking the stain. Cells are healthy with intact and polarized cytoplasmic membranes. Indeed, in the Figure 7.11, more than 95 % of bacteria are live, while only the 3 % and the 1 % result damaged and dead, respectively. Subsequent samples analyzed during the period of induction showed that the population of healthy cells remains constant. However, after 13 hours of induction, two sub-populations can be identified. The new population corresponds to dead cells with permeabilised membrane taking both stain, BOX/P.I.. It can be seen that at the end of the cultivation the percentage of dead cells is up to 6.5 % of the total population (Figure 7.11). This increase is probably due to the non-optimal conditions presented at the end of the fermentation, i.e. low pO₂ level.

On the other hand, this stress condition can be also found in the flow cytometry analysis stained with the Congo Red (CR) (Figure 7.12). Thus, a clearly CR⁺ population is observed at the end of the cultivation, with a value of 3 %. It can be also seen that the percentage of CR⁺ cells is lower than 1 % along the fermentation. In conclusion, an increase of the stress condition could have generated a largest amount of aggregated protein.

7 Results IV: Elimination of the antibiotic resistance encoding gene from the single-plasmid expression system

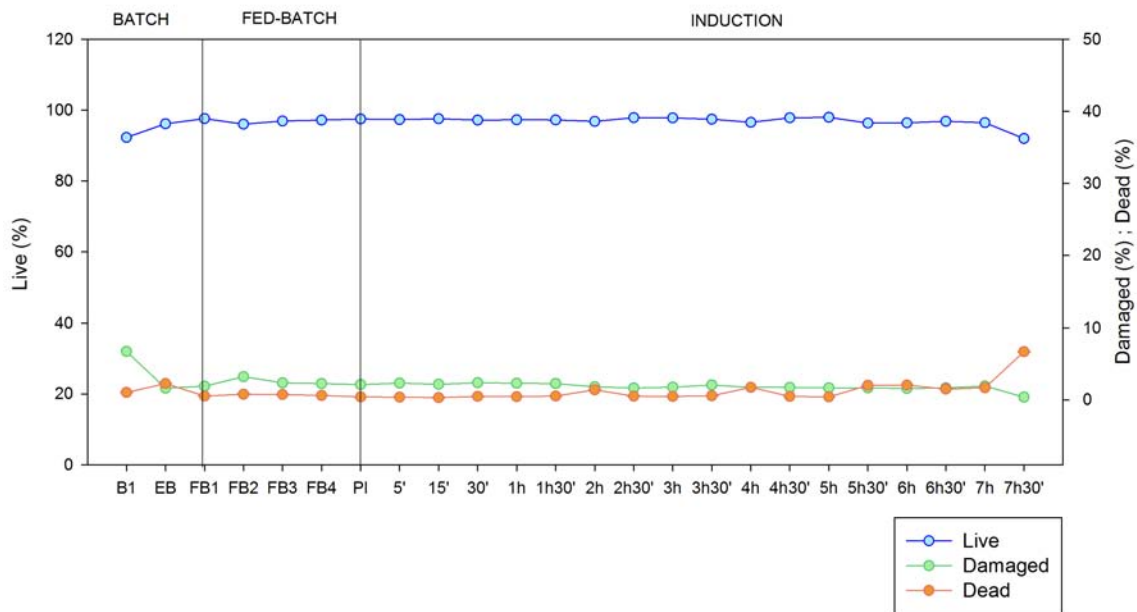


Figure 7.11 FCM analysis of *E. Coli* AmpR⁻ fed-batch: [IPTG], 70 μ M; X_{ind} , 6 g \cdot L⁻¹; μ 0.22 h⁻¹. FCM data represents live (blue), damaged (green) or dead (orange) cells determined by BOX/P.I. co-staining (See Materials and Methods section 3.7.4.2). B1 and EB, first and the last samples of the batch, respectively; FB1, FB2, FB3 and FB4, four samples of the fed-batch phase; PI, pre-induction and 5' to 7h30', different induction samples. Batch, Fed-Batch and induction phases are indicated.

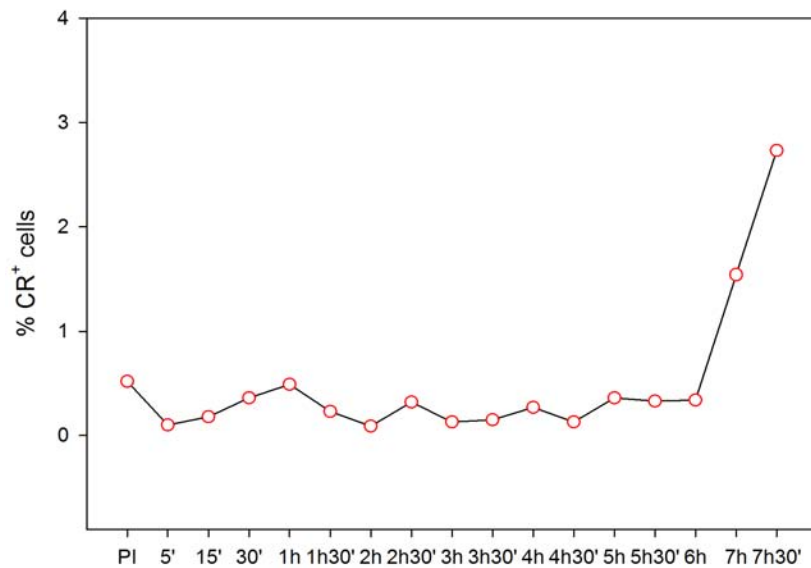


Figure 7.12 FCM analysis of *E. Coli* AmpR⁻ fed-batch culture: [IPTG], 70 μ M; X_{ind} , 6 g \cdot L⁻¹; μ 0.22 h⁻¹. % CR⁺ cells from FCM analysis. Samples stained with CR for 1 hour (See Materials and Methods section 3.7.4.3). PI, pre-induction; 5' to 7h30', induction samples.

7.3 Conclusion

The complete deletion of the antibiotic resistance gene (*bla*) has been achieved resulting the derived AmpR⁻ strain.

However AmpR⁻ strain showed a slightly decrease in the μ_{\max} , a value of $0.41 \pm 0.01 \text{ h}^{-1}$, it presented a significant increase both for the FucA specific mass and FucA specific activity. After 7 h of induction in fed-batch cultures it was found maximum activity that corresponded to $1322 \pm 42 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ and $245 \pm 13 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$.

The engineered antibiotic-free expression system have shown an increase in production up to 1.1-fold in terms of both FucA yield ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) and FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$) compared to previous expression systems, indicating that the aforementioned approaches are of paramount importance in order to increment the protein production in terms of mass and activity.

8 Results V: Preliminary study of FucA secretion in the extracellular environment of *E. coli*

As demonstrated in the previous Chapters, several cell engineering strategies have been implemented in order to optimize the recombinant FucA production in *E. coli*, yielding to maximum levels of $245 \pm 13 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$, with an activity of $1322 \pm 19 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ in the final expression strain Amp^R. Notably, intracellular overexpression of FucA has not resulted in the formation of inclusion bodies.

Nonetheless, intracellular expression still poses potential disadvantages over secretory expression from the downstream process point of view. Thus, being able to secrete the recombinant protein in the culture medium or periplasm of *E. coli* could be very advantageous over intracellular expression. Nowadays, there are different expression vectors that allow targeting the expression of the interested product into specific cellular compartments of *E. coli*, namely cytoplasm, periplasmic space and extracellular media. Secretory production provides several advantages including reduction of protease activity and of contaminants in the medium, enhanced biological activity and overcoming problems with degradation, aggregation and incorrect folding, avoidance of the cellular disruption process to recover the target protein (Mergulhao et al. 2005) (Pines and Inouye 1999). All these advantages lead to simplified downstream processes, allowing for enhanced protein activity and higher product stability and productivity.

Studies in the understanding of the secretion machinery and mechanism in *E. coli* have led the development of different approaches to reach this goal (Lee 2004). The type I, II, III, IV and V secretion pathways are widespread among the Gram-negative bacteria (Mergulhao et al. 2005). In particular, the type V, or Autotransporter (AT), system has been utilized widely to successfully secrete a variety of heterologous target molecules from the bacteria (Leyton et al. 2010). The AT involves the use of the *Sec* system for crossing the inner membrane. Proteins which use this pathway, like in this case the Pet protein, have the capability to form a porin like structure, termed the β -barrel, with their C-terminus which inserts into the outer membrane, allowing the rest of the peptide (passenger domain) to reach the outside of the cell (Jose and Meyer 2007). The passenger domain displays functional and structural heterogeneity and can be substituted with heterologous protein (Henderson et al. 2004). Sevastyanovich et al., have developed a versatile platform, based on an AT module, which can be used for secretion of heterologous fusion proteins to the culture medium in a soluble folded form (Sevastyanovich et al. 2012).

The aim of this Chapter was to clone and express the *fucA* into the Sevastyanovich AT system in order to secrete the protein of interest in the extracellular environment. The secretion systems were selected and initially tested for heterologous FucA production in shake flask cultures on LB medium, using as host *E. coli* BL21*, in agreement with Sevastyanovich et al., (Sevastyanovich et al. 2012).

8.1 Preliminary FucA secretion studies

The *fucA* gene has been cloned into the Sevastsyanovich vectors pASK-Pet and pET22b-PetAC, yielding the strains BL21* pASK-FucA (Materials and Methods section 3.4.8) and BL21* pET22b-AC-FucA (Materials and Methods, section 3.4.9), respectively. The main difference between the two systems is the reduction of the AT system, consisting only of the AC and the β -barrel domains, in the pET22b-PetAC expression system (Materials and Methods section 3.2.6). Thus, resulting in the reduction of the length of the functional translocation domain and as a consequence a reduction of the aminoacids upstream the β -barrel.

Another difference between the two-secretion vectors is the induction system based on the *tet* and the T7 promoters for the pASK-FucA and pET22b-AC-FucA, respectively. Therefore, FucA expression was induced in mid-exponential phase cultures by adding 200 $\mu\text{g}\cdot\text{L}^{-1}$ anhydrotetracycline and 0.5 mM IPTG for the pASK-FucA and pET22b-AC-FucA, respectively.

Biomass evolution profile together with FucA production, in terms of both specific activity levels ($\text{AU}\cdot\text{gDCW}^{-1}$) and yield ($\text{mg}\cdot\text{gDCW}^{-1}$), are represented in Figure 8.1.

Both strains presented a similar growing profile with a μ_{max} of $0.72 \pm 0.01 \text{ h}^{-1}$, calculated till the pre-induction point.

Besides, in terms of FucA production, some differences between the two strains can be observed in Figure 8.1. While BL21* pASK-FucA cells presented low detection levels of FucA (Figure 8.1A), the BL21*pET22b-AC-FucA strain reached a final amount of $33 \pm 11 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$ (Figure 8.2B).

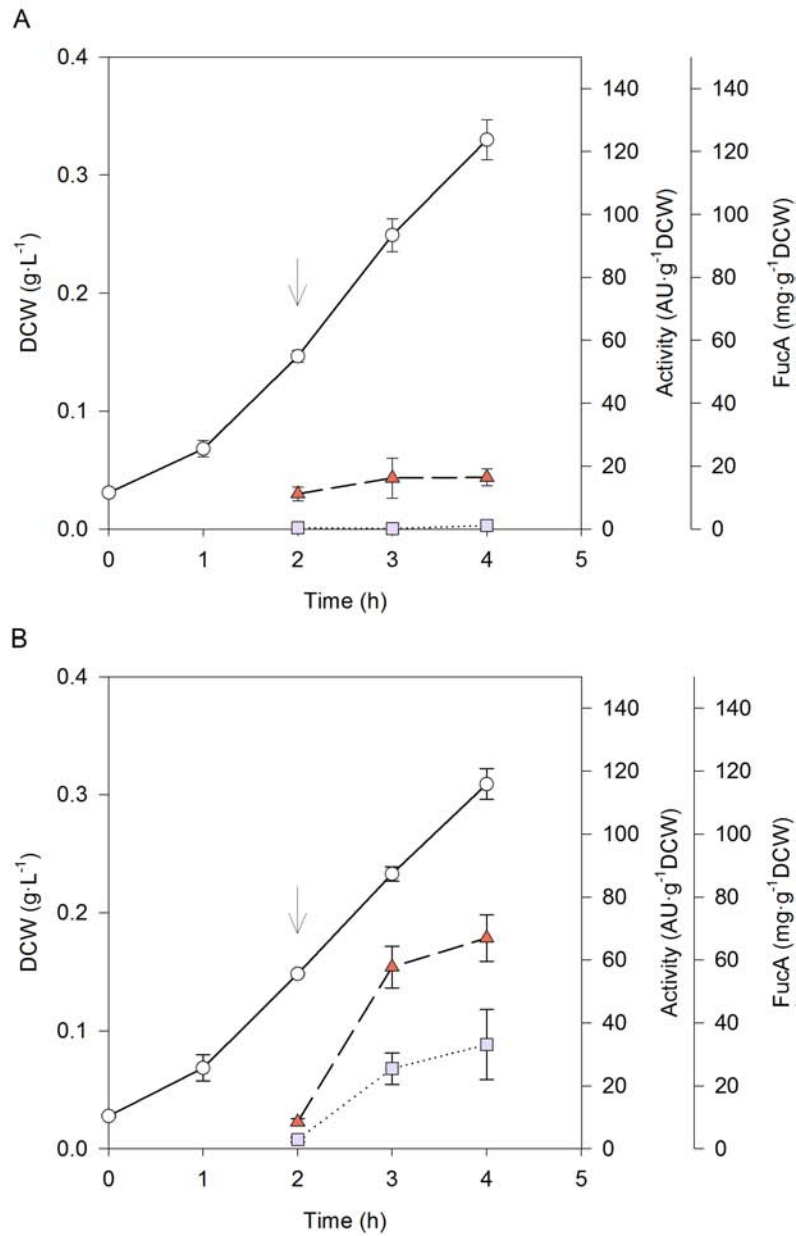


Figure 8.1 Shake flask cultures of A) BL21*pASK-FucA and B) BL21*pPET22b-AC-FucA strains in LB medium. (●) Biomass DCW (g·L⁻¹), (▲) FucA activity (AU·gDCW⁻¹), (■) specific mass (mgFucA·gDCW⁻¹). The induction point is indicated with an arrow.

Of course, to be useful as a method of secretion the AT system must be able to secrete folded and functional proteins in the culture medium. To test if FucA was natively

folded after secretion activity test was carried out, resulting in a value of the activity of $67 \pm 7 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ after 2 h of induction (Figure 8.1B).

Furthermore, the Figure 8.2 shows the difference between the two strains, in the pre-induction and post-induction phases. BL21*pPET22b-AC-FucA cells presents two bands in the 1 h and 2 h induction samples. The band present at the 60 kDa correspond to the 26 kDa of FucA protein plus the kDa residual of the autochaperon domain (Sevastyanovich et al. 2012). The second band could represent the outer membrane of the cleaved Pet β -barrel (Sevastyanovich et al. 2012).

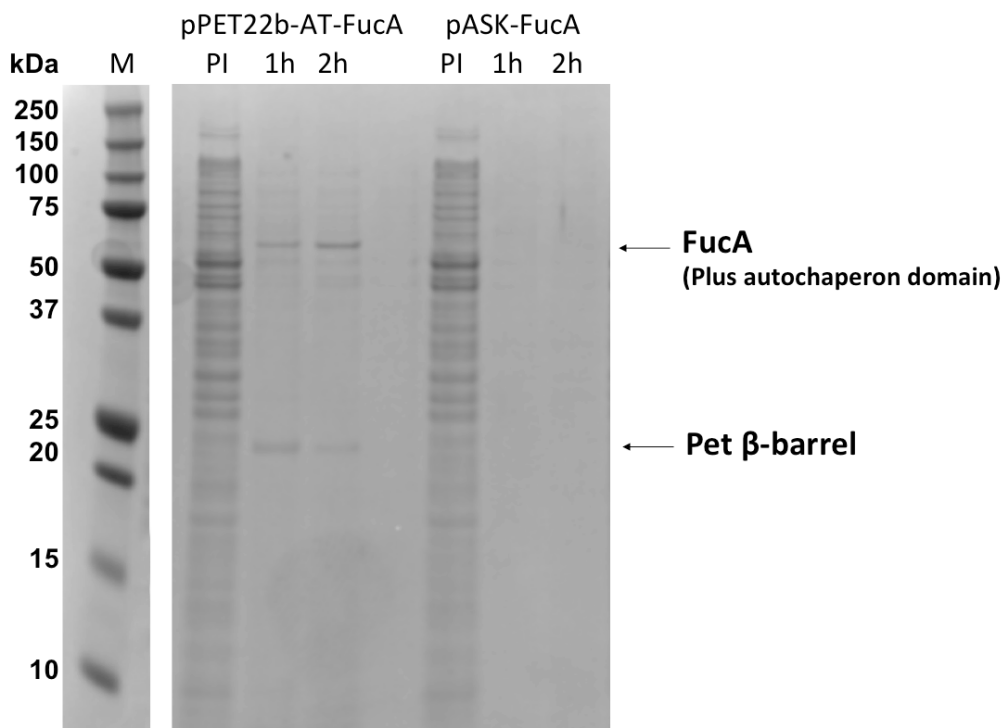


Figure 8.2 SDS-PAGE of shake flask cultures of pPET22b-AC-FucA and pASK-FucA, in LB medium. Lane M: molecular weight marker; PI, pre-induction; 1 h and 2 h correspond to sample post-induction.

Additionally, to demonstrate that the presence of heterologous proteins in the culture medium was due to secretion rather than cell lysis upon induction of expression, flow cytometer analyses were performed. Staining with both BOX/P.I. was used to assess cell viability and the integrity of the cell envelope of the bacteria (Figure 8.3). Importantly, for both strains, flow cytometry analyses revealed that the majority of cells, more than 95 %, remain healthy and live during protein secretion with only negligible increases in the number of P.I. + BOX-positive cells after induction of protein expression compared to uninduced cultures (Figure 8.3 A,B).

These data suggest that the presence of secreted protein in the BL21*pPET22b-AC-FucA strain in the culture media is not due to cell lysis but active secretion.

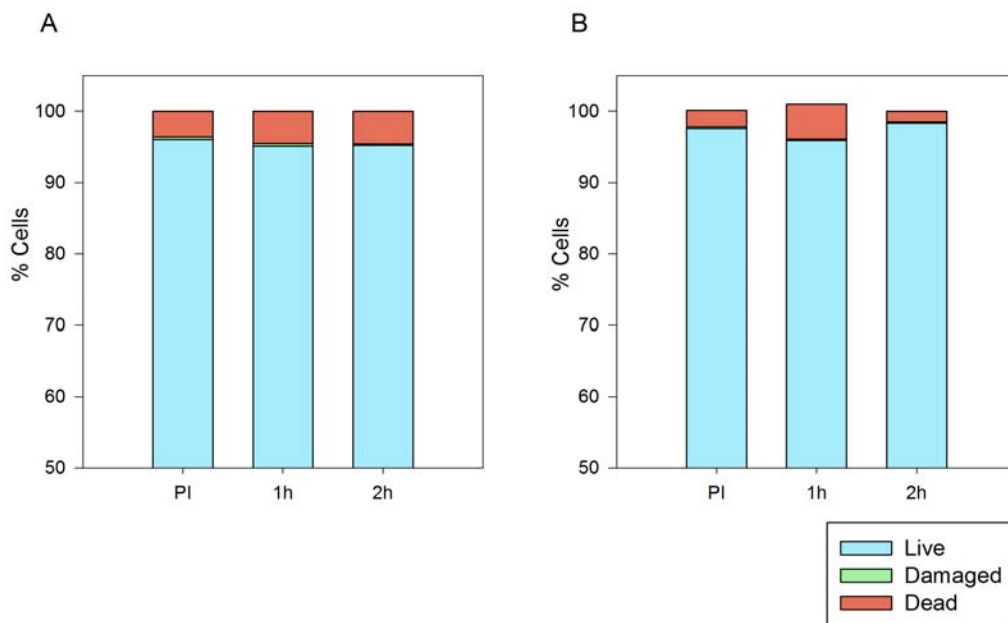


Figure 8.3 Flow cytometer analysis of the shake flask cultures of A) BL21*pASK-FucA and B) BL21*pPET22b-AC-FucA strains in LB medium. PI, pre-induction samples; 1 h and 2 h represent the samples post-induction. Live cells are represented in light blue, injured cells in green and dead cells in orange.

This Chapter demonstrates that the AT module of the pPET22b-AC-FucA expression system, with the modified autotransporter sequence with only the AC and the β -barrel domains, can be used for the targeted secretion of FucA protein into the culture media of *E. coli*. Importantly, this platform can be used for the specific accumulation of folded and functional proteins in the culture medium (Figure 8.1 B).

Conversely, the pASK-FucA vector may not lead to the secretion of the protein of interest due to the presence of the complete sequence of the AT module, that may impede the correct production and folding of FucA.

However, the ability to selectively accumulate target proteins in the culture medium, away from the majority of the process impurities associated with expression in other compartments, makes this system attractive for adoption in industrial recombinant protein production applications since extracellular expression of proteins in a folded active form reduces downstream processing.

Nevertheless, effective utilization of the AT module necessitates to increase the yields of FucA production, in terms of both specific activity levels ($\text{AU}\cdot\text{gDCW}^{-1}$) and yield ($\text{mg}\cdot\text{gDCW}^{-1}$). In fact, the comparison between the Amp^R and the BL21*pPET22b-AC-FucA *E. coli* strains after 2 h of induction in shake flask cultures, demonstrated that the later presents a 4.5-fold reduction in the yield, decreasing from $150 \pm 50 \text{ mg}\cdot\text{gDCW}^{-1}$ to $33 \pm 11 \text{ mg}\cdot\text{gDCW}^{-1}$, and a 9.4-fold decrease of the activity levels, decreasing from $632 \pm 5 \text{ AU}\cdot\text{gDCW}^{-1}$ to $67 \pm 7 \text{ AU}\cdot\text{gDCW}^{-1}$.

Importantly, these low yields values of FucA production in the secrete system were affected by several factors: i) the shake flask cultures performed in this Chapter were in LB medium and in non-optimized conditions; ii) the IPTG pulse, for induce *fucA* expression, was carried out at lower biomass concentration compare to the previous strain; iii) the induction was maintained for only 2 h.

Accordingly, it could be interesting to clone the AT module of the pET22b-PET-AC vector into the optimized auxotrophic expression system presented along this study, being pQE-FucA_puzzle (J23110)_Amp^R. In this way recombinant FucA secretion in the extracellular environment experiments could be performed in DM medium and in optimal operational conditions. Finally, could be interesting to performed high cell density fed-batch cultures in order to achieve high biomass concentration.

8.2 Conclusion

The AT module can be used for secretion of the FucA protein to the culture medium in a soluble folded form. Although the production of the protein of interest presents a significant reduction in term of production, the FucA protein secreted is correctly folded and active.

In conclusion, these promising results represent a good starting point to further investigate and optimize recombinant protein secretion in *E. coli*.

9 Comparison of *E. coli* M15 strains presented along this study

Along this work, a novel expression system based on an antibiotic-free plasmid maintenance mechanism has been developed. The stepwise design approach resulted in increased production levels, both in terms of FucA yield ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) and FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$), compared to the previous expression system.

In this Chapter a comparison between all the different strains used along this thesis is presented, focusing mainly on the fed-batch cultures results.

9.1 *E. coli* M15[pREP4] and M15 Δ glyA[pREP4] strains

The comparison between the reference M15[pREP4] and the M15 Δ glyA[pREP4] *E. coli* strains, presented in the Chapter 4, demonstrated that the later presents slightly lower specific growth rate, decreasing from $0.49 \pm 0.01 \text{ h}^{-1}$, of the reference strain, to $0.44 \pm 0.01 \text{ h}^{-1}$ of the M15 Δ glyA[pREP4] strain (Table 9.1).

Besides, in terms of production, both FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$) and FucA mass ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) decreased more than 35 % when comparing the M15 Δ glyA[pREP4] strain to the reference M15[pREP4] (Figure 9.1 and Table 9.1).

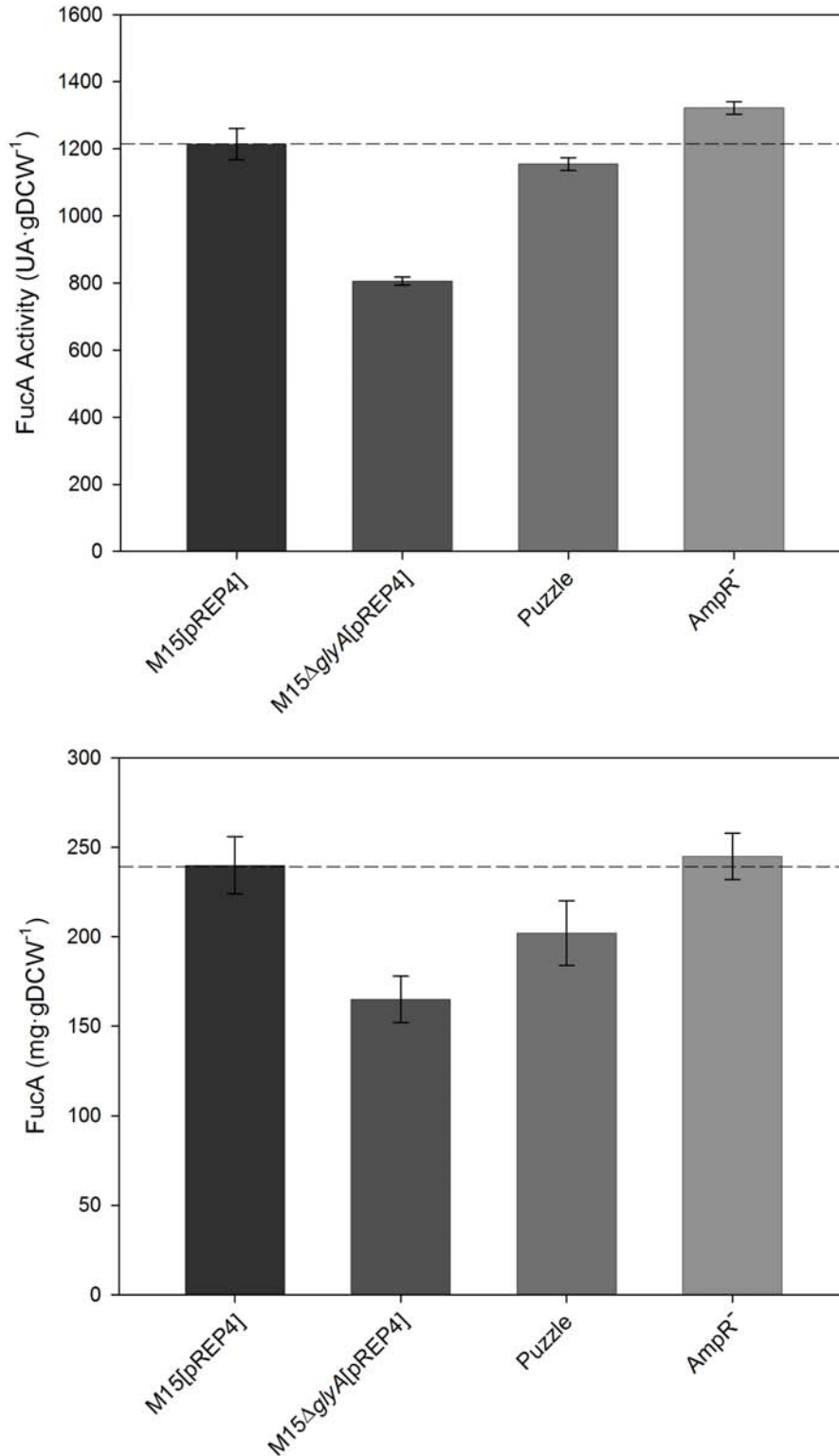


Figure 9.1 Maximum enzyme activity (AU·g⁻¹DCW) and **B**) Maximum specific mass (mgFucA·g⁻¹DCW) for the principal strains presented along this study along the induction phase of fed-batch cultures ([IPTG], 70 μM; X_{ind}, 20 g·L⁻¹; μ 0.22 h⁻¹). A dash-dot line indicates the value of the M15[pREP4] reference strain.

Table 9.1 Maximum FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$), maximum FucA mass ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$) and μ_{\max} along the induction phase in fed-batch culture in DM medium for the principal strains presented along this study.

<i>E. coli</i> strains	μ_{\max} (h^{-1})	FucA activity ($\text{AU}\cdot\text{g}^{-1}\text{DCW}$)	FucA mass ($\text{mg}\cdot\text{g}^{-1}\text{DCW}$)
M15[pREP4]	0.49 ± 0.02	1214 ± 47	240 ± 16
M15 Δ <i>glyA</i> [pREP4]	0.44 ± 0.01	806 ± 12	165 ± 13
Puzzle	0.45 ± 0.01	1176 ± 19	202 ± 18
AmpR ⁻	0.41 ± 0.01	1322 ± 19	245 ± 13

These effects may be caused by the increase in the metabolic burden due to the maintenance of the expression vector in the M15 Δ *glyA* strain. In fact, the *glyA* gene encoded in the high-copy plasmid leads to substantially higher amounts of its product (SHMT) accumulated in the cytoplasm, compared to the reference strain containing a single copy of *glyA* in the genome. The overexpression of the SHMT protein both in the pre-induction and during the induction phase in fed-batch cultures can be observed Figure 4.21. Moreover, the SHMT production of the M15 Δ *glyA*[pREP4] strain, being around $130 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$, increased more than 4.5-fold comparing with the M15[pREP4] strain, being around $20 \text{ mgSHMT}\cdot\text{g}^{-1}\text{DCW}$ (Table 9.2).

Furthermore, in both the reference and the auxotrophic strains the concentration of the biomass progressively grew till 2.5 h of induction and then stopped around 4 h after induction, reaching a final value of $40 \text{ gDCW}\cdot\text{L}^{-1}$. Thank to the flow cytometry analysis it can be seen that at the end of the fed-batch the cells began to be unhealthy, as the amount of damaged and dead cells increased.

Table 9.2 Values of SHMT and FucA protein ($\text{mg}\cdot\text{gDCW}^{-1}$) along the induction phase of fed-batch culture's samples of the principal strains presented along this study. PI, represents the pre-induction, while from 5' to 7 h are the samples along the induction.

	M15[pREP4]	M15 Δ <i>glyA</i> [pREP4]	Puzzle	AmpR ⁻
PI	39 ± 3	145 ± 2	94 ± 1	78 ± 9
5'	32 ± 2	163 ± 2	98 ± 1	76 ± 8
15'	29 ± 3	150 ± 2	93 ± 1	68 ± 10
30'	27 ± 3	106 ± 2	92 ± 7	75 ± 8
1 h	28 ± 2	129 ± 1	84 ± 0.4	82 ± 11
1.5 h	29 ± 2	141 ± 3	87 ± 4	72 ± 7
2 h	25 ± 1	131 ± 7	83 ± 4	82 ± 8
2.5 h	24 ± 2	103 ± 5	79 ± 3	67 ± 9
3 h	24 ± 3	113 ± 7	82 ± 0.4	78 ± 14
3.5 h	23 ± 4	124 ± 9	85 ± 0.5	72 ± 7
4 h	23 ± 6	132 ± 4	83 ± 2	72 ± 13
4.5 h			80 ± 2	81 ± 13
5 h			82 ± 2	80 ± 11
5.5 h			81 ± 1	79 ± 9
6 h			81 ± 0.5	76 ± 10
6.5 h				77 ± 7
7 h				80 ± 9

Besides both M15[pREP4] and M15 Δ *glyA*[pREP4] strains started to accumulate glucose and acetate in the medium after 2 h of induction, without arriving to inhibitory levels (Figure 4.15 and 4.21, respectively).

This accumulation is probably due to the high-energy demand imposed to the cell for the synthesis of the recombinant protein. Higher energy demand increases the

catabolic fluxes for energy generation and reduces the material available for biosynthesis and cell growth, i.e. negatively affecting the growth rate (Hoffmann et al. 2002)(Weber et al. 2002).

Additionally, this is coherent with previous studies of *E. coli* cultures growing under aerobic conditions in the presence of excess glucose, pointing at the unbalance between glycolysis rates and the TCA-cycle limited capacity of *E. coli* (Lee 1996).

The accumulation of glucose in the medium was due to the feed addition system used. In fact, since the set values of μ 0.22 h^{-1} and $Y_{x/S}$ were not modified during the induction phase, the addition program assumed these parameters constant even though they changed. Through this phase, more substrate than required was consequently added and glucose accumulated. Thus, in order to avoid this accumulation the addition of glucose was stopped manually.

9.2 Puzzle strain

Interestingly, in Chapter 5 the lack of FucA expression in the system with no *lacI* gene was reported. This effect may be related to T5 promoter leakiness in absence of LacI repressor protein, leading to plasmid structural instability due to recombination events, as supported by the sequencing data (Figure 5.4). Alternatively, a possible explanation for the lack of FucA expression in the single plasmid system may be that the copy number of the *lacI* gene increases when cloned into the pQE-40 vector, resulting in significantly higher levels of intracellular LacI. In fact, the pQE vector is

based on the plasmid replication origin ColE1, which presents a copy number 2-fold higher compared with the P15A replicon of pREP4 (Baneyx 1999b).

These results together with the necessity to regulated the *glyA* overexpression confirmed that *glyA* and *lacI* co-expression were required. For this reason, four different expression cassettes were constructed where the *lacI* and *glyA* genes were placed under the control of a set of four constitutive promoters, covering a wide range of transcriptional efficiencies (See Chapter 6 section 6.2). According to the results shown in Table 6.1, the selected expression cassette for *lacI* and *glyA*, was the one where both genes were placed under the control of the constitutive promoter J23110. This promoter, which is in the lower range of the tested *lacI* and *glyA* transcriptional levels, seems to down-regulate their transcriptional levels. Finally, to further optimize the expression system, a single vector harboring both the *fucA* gene under control of the inducible T5 promoter and the *lacI-glyA* cassette cloned under the J23110 constitutive promoter was constructed, getting the Puzzle strain.

Chapter 6 showed that FucA production values improved in the Puzzle strain compared to the previous system with two plasmids (See Chapter 6 section 6.3).

In particular, while the μ_{\max} , being $0.45 \pm 0.01 \text{ h}^{-1}$, was still comparable to that from the M15 Δ *glyA*[pREP4] strain, being $0.4 \pm 0.01 \text{ h}^{-1}$; the maximum FucA mass and FucA specific activity increased up to $202 \pm 18 \text{ mg}^{-1} \text{ FucA} \cdot \text{g}^{-1} \text{ DCW}$ and $1176 \pm 19 \text{ AU} \cdot \text{g}^{-1} \text{ DCW}$, respectively (Figure 9.1 and Table 9.1). Comparing these values with those obtained with the M15[pREP4] reference strain ($236 \pm 16 \text{ mg FucA} \cdot \text{g}^{-1} \text{ DCW}$ and $1214 \pm 47 \text{ AU} \cdot \text{g}^{-1}$

DCW), it can be observed how the specific activity was comparable even though the amount of the recombinant protein was relatively lower.

Remarkably, the Puzzle strain, unlike the previous two-plasmids expression system, did not present neither acetate production nor glucose accumulation along the induction phase of the fed-batch cultures (Figure 6.12). Besides, by observing the FCM data it can be seen that the cells were homogenous and for more than 95 % live and healthy all along the fermentation (Figure 6.16).

These results suggested that transcriptional tuning of *lacI* and *glyA* expression levels is a key factor to improve *fucA* expression regulation, leading to a higher FucA specific activity. In addition, T5 promoter leakiness was minimized, and *glyA* expression levels reduced, resulting in an overall reduced metabolic burden. In fact, as it can be seen in Table 9.2, SHMT values were almost 50 % reduced in comparison with the Puzzle with the M15 Δ *glyA*[pREP4] strains. Thus, the tuning of *glyA* levels had a positive effect on the reduction of the metabolic load due to expression of plasmid-encoded genes (also reflected in the reduced acetate production).

These results are in accordance with the observation by Mairhofer et al. (Mairhofer et al. 2013), who demonstrated that the folding machinery is severely overstrained in the plasmid-based expression system compared with the plasmid-free cells due to the different transcriptional profiles.

9.3 AmpR⁻ strain

Lastly, in Chapter 7, an expression system completely devoid of antibiotic resistance genes was constructed by removing the *bla* gene from the expression vector, resulting in the *E. coli* AmpR⁻ strain (See Chapter 7 section 7.1).

A slightly decrease in the μ_{\max} , a value of $0.41 \pm 0.01 \text{ h}^{-1}$ was observed compared to the M15[pREP4] and Puzzle strains, which showed a μ_{\max} of $0.49 \pm 0.02 \text{ h}^{-1}$ and $0.45 \pm 0.01 \text{ h}^{-1}$, respectively (Table 9.1). This strain presented a significant increase both for the FucA specific mass and FucA specific activity. After 7 h of induction in fed-batch cultures it was found maximum activity that corresponded to $1322 \pm 19 \text{ AU}\cdot\text{g}^{-1}\text{DCW}$ with a specific mass of $245 \pm 13 \text{ mgFucA}\cdot\text{g}^{-1}\text{DCW}$ (Figure 9.1). Moreover, by comparing the SHMT values between the AmpR⁻ and the Puzzle strains, Table 9.1, it can be seen, as expected, that the expression levels of the *glyA* gene are similar.

Comparing with all the previous constructs, FucA over-production using the antibiotic free-plasmid system is higher than any previous developed system studied. In particular, as it can be seen in Table 9.1 and in Figure 9.1, FucA yields in the AmpR⁻ strain is: i) 1.1-fold higher comparing with the M15[pREP4]; ii) more than 1.2-fold higher comparing with the Puzzle strain and iii) 1.6-fold higher referred to the M15 Δ *glyA*[pREP4] strain (Figure 9.1B). Noteworthy, the FucA activity, in terms of $\text{AU}\cdot\text{g}^{-1}\text{DCW}$, increased through the different stepwise improvements performed along this work. The best performing engineered strain reached 1.6-fold higher values compared to the first M15 Δ *glyA*[pREP4] strain (Figure 9.1A).

The AmpR⁻ strain, like its antecedent Puzzle strain, did not produce acetate nor accumulate glucose along the induction phase of the fed-batch cultures (Figure 7.7). Besides, as described previously for the Puzzle strain, AmpR⁻ cells resulted healthy with intact and polarized cytoplasmatic membrane all along the fermentation the (Figure 7.11).

10 General conclusions and future perspectives

Regarding the main goals proposed at the beginning of this work, the development of a second-generation system suitable for *Escherichia coli* expression with an antibiotic-free plasmid maintenance mechanism has been successfully studied. Besides, the study of an expression system able to secrete the recombinant protein in the extracellular medium has been positively achieved.

In this work we have applied rapid assembly strategies for the construction of improved expression systems that are useful for recombinant protein production. Using as a reference expression system commercially available, we have obtained an improved system that resulted in higher protein yields and devoid of antibiotic supply.

More in details, we identified the high gene dosage and consequently over-transcription of the *glyA* gene, which encodes for the auxotrophic selection marker protein, in the high-copy plasmid as one of the major factors responsible for the metabolic burden. Compared to the M15[pREP4] reference strain containing a single copy of *glyA* in the genome, the auxotrophic M15 Δ *glyA*[pREP4] strain presented an overexpression of the SHMT protein, thereby affecting negatively FucA expression levels.

Moreover, we demonstrated that the regulation of *lacI* transcriptional levels, which encodes for the Lac repressor, is a key parameter to take into account for recombinant protein production.

This case-study demonstrates that tuning the expression levels of *lacI* and *glyA* genes, results into a reduction of the metabolic burden leading to a better stability of expression system. This fact allows an improvement of the recombinant protein production due to the alleviation of the metabolic burden and a reduction of acetate secretion and glucose accumulation in fed-batch cultures.

Furthermore, we have developed an expression system completely devoid of antibiotic resistance genes that further improve and optimize the recombinant FucA production in these experimental conditions.

The main advantage of this engineered expression system devoid of antibiotic resistance markers is that it can be used as a platform for the production of a wide range of heterologous proteins where the use of antibiotics is restricted.

Our work allows versatile and tuneable levels of expressed proteins at will, and we envisage that it can be potentially used in a wide range of applications and biotechnological processes with a significant reduction of production time and upstream costs.

Furthermore, we demonstrated that the AT module of the pPET22b-AC-FucA expression system, can be used for the targeted secretion of FucA protein into the culture media of *E. coli*. Significantly, this platform can be used for the specific accumulation of folded and functional proteins in the culture medium

The obtained results from this thesis work make room for further studies. Future perspectives in this research field include:

- Extend the optimized expression system obtained along this study to other recombinant proteins. Besides, comparison with the FucA expression system would be helpful to further understand and optimize the process
- Optimize the secretion expression system in order to obtain higher yields of the FucA protein, i.e. clone the AT module into the optimized auxotrophic expression system presented along this study. In this way recombinant FucA secretion in the extracellular environment experiments could be performed in DM medium and in optimal operational conditions.
- Study of the secretion mechanism in high cell-density cultures in order to evaluate the recombinant protein secretion with higher cell concentration. It could be also interesting to perform an economic study at downstream level, making a comparison between intracellular and extracellular recombinant protein production.

11 References

- Arı, Anna, Antonio Villaverde, Andrea Vera, and Nuria Gonza. 2007. "The conformational quality of insoluble recombinant proteins is enhanced at Low growth temperatures." *Biotechnology and Bioengineering* 96(6):1101–6.
- Baneyx, François. 1999. "Recombinant protein expression in *Escherichia coli*." *Current Opinion in Biotechnology* 10:411–21.
- Beckwith, J. 1987. "The Lactose Operon In *Escherichia coli* and *Salmonella typhimurium*." *Cellular and molecular biology* (Neidhardt FC ed). American Society for Microbiology, pp. 1444–52 vol. 18.
- Bentley, William E., Noushin Mirjalili, Dana C. Andersen, Robert H. Davis, and Dhinakar S. Kompala. 1990. "Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant acteria." *Biotechnology and Bioengineering* 35:668–81.
- Berrow, Nick S. et al. 2006. "Recombinant protein expression and solubility screening in *Escherichia coli* : a comparative study." *Biological Crystallography* 1218–26.
- Blattner, Frederick R. et al. 1997. "The complete genome sequence of *Escherichia coli* K-12." 277:1453–62.
- Boada, Jaume Pinsach. 2009. "Development of recombinant aldolase production process in *Escherichia coli*." Doctoral thesis.
- Brautaset, Trygve, Rahmi Lale, and Svein Valla. 2009. "Positively regulated bacterial expression systems." *Microbial Biotechnology* 2(1):15–30.
- Brosius, J., M. Erfle, and J. Storella. 1985. "Spacing of the -10 and -35 regions in the *Tac* promoter. Effect on its in vivo activity." *Journal of Biol. Chem.* 260:2539–41.
- Calleja, Daniel, Alfred Fernández-Castañé, Martina Pasini, Carles de Mas, and Josep López-Santín. 2014. "Quantitative modeling of inducer transport in fed-batch cultures of *Escherichia coli*." *Biochemical Engineering Journal* 91:210–19.
- Carneiro, Sónia, Eugénio C. Ferreira, and Isabel Rocha. 2012. "Metabolic responses to recombinant bioprocesses in *Escherichia coli*." *Journal of Biotechnology* 1–13.

- Carrió, Mar, Nuria González-Montalbán, Andrea Vera, Antonio Villaverde, and Salvador Ventura. 2005. "Amyloid-like properties of bacterial inclusion bodies." *Journal of molecular biology* 347:1025–37.
- Chan, Pedro, Robin A. Curtis, and Jim Warwicker. 2013. "Soluble expression of proteins correlates with a lack of positively-charged surface." *Scientific reports* 3:3333–39.
- Cooper, Geoffrey M. 2000. "The Cell. A Molecular Approach". Sunderland (MA): Sinauer Associates.
- Cornelis, Pierre. 2000. "Expressing genes in different *Escherichia coli* compartments." *Current Opinion in Biotechnology* 450–54.
- Correa, Agustín, and Pablo Opezzo. 2011. "Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening." *Biotechnology journal* 6:715–30.
- Crabtree, Herbert Grace. 1928. "Clx. The carbohydrate metabolism of certain pathological overgrowths." *Biochem J* 22(5):1289–98.
- Datsenko, Kirill A., and Barry L. Wanner. 2000. "One-step inactivation of chromosomal Genes in *Escherichia coli* K-12 using PCR products." *Proceedings of the National Academy of Sciences of the United States of America* 97(12):6640–45.
- Donovan, R. S., C. W. Robinson, and B. R. Glick. 1996. "Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter." *Journal Of Industrial Microbiology* 145–54.
- Dunaways, Marietta et al. 1980. "Kinetic studies of inducer binding to *lac* repressor. Operator complex*." *The Journal of Biological Chemistry* 255(21):10115–19.
- Durany, Olga, Glòria Caminal, Carles de Mas, and Josep López-Santín. 2004. "Studies on the expression of recombinant fucose-1-phosphate aldolase in *E. coli*." *Process Biochemistry* 39(11):1677–84.
- Durany, Olga, Carles de Mas, and Josep López-Santín. 2005. "Fed-batch production of recombinant fucose-1-phosphate aldolase in *E. coli*." *Process Biochemistry* 40:707–16.
- Eiteman, Mark A., and Elliot Altman. 2006. "Overcoming acetate in *Escherichia Coli* recombinant protein fermentations." *Trends in biotechnology* 24(11):530-36.
- Elvin, CM et al. 1990. "Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*." *Gene* 87:123–26.

- Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. 2008. "A one pot, one step, precision cloning method with high throughput capability." *PLoS one* 3(11):1–9.
- Fernández-Castañé, Alfred, Jordi Ruiz, Gloria Caminal, and Josep Lopez-Santin. 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-Castañé, Alfred, Claire E. Vine, Glòria Caminal, and Josep López-Santín. 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):391–98.
- Ferrer-Miralles, Neus, Joan Domingo-Espín, José Luis Corchero, Esther Vázquez, and Antonio Villaverde. 2009. "Microbial factories for recombinant pharmaceuticals." *Microbial cell factories* 8:17.
- Fuentes, Laura G. et al. 2013. "Modification of glucose import capacity in *Escherichia coli*: physiologic consequences and utility for improving dna vaccine production." *Microbial cell factories* 12:42–52.
- Garcia-Junceda, Eduardo, Shen Gwo-Jenn, Sugai Takeshi, and Chi-Huey Wong. 1995. "A new strategy for the cloning, overexpression and one step purification of three dhp-dependent aldolases: Rhamnulose-1-phosphate aldolase, Fuculose-1-phosphate aldolase and Tagatose-1,6-diphosphate aldolase." *Biorganic & Medicinal Chemistry* 3(7):945–53.
- Gibson, Daniel G. 2011. "Enzymatic assembly of overlapping DNA fragments." *Methods in enzymology*, pp. 349-61 vol. 498.
- Gluscock, Christopher B., and Michael J. Weickert. 1998. "Using chromosomal *lacI*^{Q1} to control expression of genes on high-copy-number plasmids in *Escherichia coli*." *Gene* 223:221–31.
- Glenting, Jacob, and Stephen Wessels. 2005. "Ensuring safety of DNA vaccines." *Microbial cell factories* 4:26.
- Glick, Bernard R. 1995. "Metabolic load and heterologous gene expression." *Biotechnology Advances* 13(2):247–61.
- Green, Michael R., and Joseph Sambrook. 2012. *Molecular Cloning, a laboratory manual* (Cold spring Harbor laboratory press) vol 1.
- Gronenborn, Bruno. 1976. "Overproduction of phage lambda repressor under control of the lac promoter of *Escherichia coli*." *Mol. Gen. Genet.* 148:243–50.

- Gruber, David F., Vincent A. Pieribone, Barbara Porton, and Hung-Teh Kao. 2008. "Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system." *Protein Expression and Purification* 60(1):53–57.
- Guzman, LM, D. Belin, MJ Carson, and J. Beckwith. 1995. "Tight regulation, modulation, and high-level expression by vectors containing the arabinose *PBAD* promoter." *Journal of Bacteriology* 177:4121–30.
- Hägg, Peter, Johanna Wa De Pohl, Farhad Abdulkarim, and Leif A. Isaksson. 2004. "A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*." *Journal of Biotechnology* 111:17–30.
- Haldimann, A., L. Daniels, and B. Wanner. 1998. "Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon." *Journal of Bacteriology* 180:1277–86.
- Heinz, Neumann, and Petra Neumann-Staubitz. 2010. "Synthetic biology approaches in drug discovery and pharmaceutical biotechnology." *Appl Microbiol Biotechnol* 87:75–86.
- Henderson, Ian R., Fernando Navarro-Garcia, Mickaël Desvaux, Rachel C. Fernandez, and Dlawer Ala'Aldeen. 2004. "Type V protein secretion pathway: the autotransporter story." *Microbiology and molecular biology reviews* 68(4):692–744.
- Henderson, Ian R., Fernando Navarro-garcia, and James P. Nataro. 1998. "The great escape : structure and function of the autotransporter proteins." *Trends in Microbiology* 6:1615–21.
- Hewitt, Christopher J., Gerhard Nebe-von Caron, Alvin W. Nienow, and Caroline M. Mcfarlane. 1999. "The use of multi-parameter flow cytometry to compare the physiological response of *Escherichia coli* W3110 to glucose limitation during batch , fed-batch and continuous culture cultivations." *Journal of Biotechnology* 75:251–64.
- Hoffman, F., and U. Rinas. 2004. "Stress induced by recombinant protein production in *Escherichia coli*." *Advances in Biochemical Engineering/Biotechnology* 89:73–92.
- Hoffmann, Frank, Jan Weber, and Ursula Rinas. 2002. "Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: 1. readjustment of metabolic enzyme synthesis." *Biotechnology and bioengineering* 80(3):313–19.
- Holms, Harry. 1996. "Flux analysis and control of the central metabolic pathways in *Escherichia coli*." *FEMS Microbiology Reviews* 19:85–116.

- Hudson, J. Andrew, and Sandra J. Mott. 1994. "Comparison of Lag times obtained from optical density and viable count data for a strain of *Pseudomonas fragi*." *Journal of Food Safety* 14:329–39.
- Jana, S., and J. K. Deb. 2005. "Strategies for efficient production of heterologous proteins in *Escherichia coli*." *Applied Microbiology and Biotechnology* 67:289–98.
- Jensen, E. Bech, and S. Carlsen. 1990. "Production of recombinant human growth hormone in *Escherichia coli*: expression of different precursors and physiological effects of glucose, acetate, and salts." *Biotechnology and Bioengineering* 36:1–11.
- Jeong, K. J., and S. Y. Lee. 2003. "Enhanced production of recombinant proteins in *Escherichia coli* by filamentation suppression." *Applied and Environmental Microbiology* 69(2):1295–98.
- Johnson, Irving S. 1983. "Human insulin from recombinant DNA technology." *Science* 219:632–37.
- Jose, Joachim, and Thomas F. Meyer. 2007. "The autodisplay story, from discovery to biotechnical and biomedical applications." *Microbiology and molecular biology reviews : MMBR* 71(4):600–619.
- Jung, G., P. Denèfle, J. Becquart, and JF Mayaux. 1988. "High-cell density fermentation studies of recombinant *Escherichia coli* strains expressing human interleukin-1 beta." *Ann Inst. Pasteur Microbiol.* 139(1):129–46.
- Kang, Zhen, Yanping Geng, Yong Zhen Xia, Junhua Kang, and Qingsheng Qi. 2009. "Engineering *Escherichia coli* for an efficient aerobic fermentation platform." *Journal of biotechnology* 144(1):58–63.
- Kawe, Martin, Uwe Horn, and Andreas Plückthun. 2009. "Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors." *Microbial Cell Factories* 8:1–8.
- Kleman, Gary L., and William R. Strohl. 1994. "Metabolism by *Escherichia coli* in high-cell-density." *Applied and Environmental Microbiology* 60(11):3952–58.
- Kubitshek, H. E. 1990. "Cell volume increase in *Escherichia coli* after shifts to richer media." *journal of Bacteriology* 172(1):94–101.
- Lawrence, Jeffrey G., and Oward Ochman. 1998. "Molecular archaeology of the *Escherichia coli* genome." *PNAS* 95:9413–17.
- Lee, S. Y. and Choi J. .. 2004. "Secretory and extracellular production of recombinant proteins using *Escherichia coli*." *Biotechnology and Bioengineering* 625–35.

- Lee, Sang Yup. 1996. "High cell-density culture of *Escherichia coli*." *Trends in biotechnology*, pp.98-105 vol. 14.
- Lenski, Richard E., Suzanne C. Simpson, and Toai T. Nguyen. 1994. "Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness." *Journal of Bacteriology* 3140–47.
- Leo, Jack C., Iwan Grin, and Dirk Linke. 2012. "Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane." *Philosophical transactions of the Royal Society of London*. 367:1088–1101.
- Leyton, Denisse L. et al. 2010. "The Unusual extended signal peptide region is not required for secretion and function of an *Escherichia coli* autotransporter." *FEMS microbiology letters* 311(2):133–39.
- Li, Jian, Jennifer Jaitzig, Friederike Hillig, Roderich Süßmuth, and Peter Neubauer. 2014. "Enhanced production of the nonribosomal peptide antibiotic valinomycin in *Escherichia coli* through small-scale high cell density fed-batch cultivation." *Applied Microbiology Biotechnology* 98:591–601.
- Li, Jian, and Peter Neubauer. 2014. "*Escherichia coli* as a cell factory for heterologous production of nonribosomal peptides and polyketides." *New biotechnology* 31(6):579–85.
- Li, Mingji et al. 2012. "A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*." *Microbial Cell Factories* 11:19–29.
- Lilie, Hauke, Elisabeth Schwarz, and Rainer Rudolph. 1998. "Advances in refolding of proteins produced in *E. coli*." *Current Opinion in Biotechnology* 9:497–501.
- Mairhofer, Juergen, Theresa Scharl, Karoline Marisch, Monika Cserjan-Puschmann, and Gerald Striedner. 2013. "Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions." *Applied and Environmental Microbiology* 79(12):3802–12.
- Makrides, Savvas C. 1996. "Strategies for achieving high-level expression of genes in *Escherichia coli* †." *Microbiological Reviews* 60(3):512–38.
- Martínez-Alonso, Mónica, Nuria González-Montalbán, Elena García-Fruitós, and Antonio Villaverde. 2009. "Learning about protein solubility from bacterial inclusion bodies." *Microbial cell factories* 8:4–8.
- Mergulhao, F. J. M., D. K. Summers, and G. A. Monteiro. 2005. "Recombinant protein secretion in *Escherichia coli*." *Biotechnology Advances* 23:177–202.

- Mott, Melissa L., and James M. Berger. 2007. "DNA replication initiation: mechanisms and regulation in bacteria." *Nature reviews. Microbiology* 5(5):343–54.
- Nations, United. 1992. "Text of the Convention in Biological Diversity. Article 2: Use of Terms".
- Natl, N. Proc et al. 1998. "The ins and outs of a molecular chaperone machine." *Trends Biochem Sci* 138–43.
- Neubauer, P., H. Y. Lin, and B. Mathiszik. 2003. "Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*." *Biotechnology and bioengineering* 83(1):53–64.
- Nishihara, Kazuyo, Masaaki Kanemori, Masanari Kitagawa, and Takashi Yura. 1998. "Chaperone coexpression plasmids : differential and synergistic roles of dnaK-dnaJ-grpe and groEL-groES in assisting folding of an allergen of japanese cedar pollen, Cryj2, in *Escherichia coli*." *Applied and Environmental Microbiology* 64(5):1694–99.
- Overton, Tim W. 2014. "Recombinant protein production in bacterial hosts." *Drug discovery today* 19(5):590–601.
- Phue, Je-nie, and Joseph Shiloach. 2005. "Impact of dissolved oxygen concentration on acetate accumulation and physiology of *E. coli* BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions." *Metabolic Engineering* 7:353–63.
- Phue, Je-nie, Santosh B. Noronha, Ritabrata Hattacharyya, Alan J. Wolfe, and Joseph Shiloach. 2005. "Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and northern blot analyses." *Biotechnology*.
- Pines, Ophry, and Masayori Inouye. 1999. "Expression and secretion of proteins in *E. coli*." *Molecular Biotechnology* 12.
- Pinsach, Jaume, Carles de Mas, and Josep Lopez-Santin. 2006. "A simple feedback control of *Escherichia coli* growth for recombinant aldolase production in fed-batch mode." *Biochemical Engineering Journal* 29:235–42.
- Pinsach, Jaume, Carles de Mas, and Josep López-Santín. 2008. "Induction strategies in fed-batch cultures for recombinant protein production in *Escherichia coli*: application to Rhamnulose 1-phosphate aldolase." *Biochemical Engineering Journal* 41:181–87.

- Plamann, M. D., and G. V Stauffer. 1983. "Characterization of the *Escherichia coli* gene for serine hydroxymethyltransferase." *Gene* 22:9–18.
- Polisky, Barry, Robert J. Bishop, and David H. Gelfand. 1976. "A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria." *Biochemistry* 73(11):3900–3904.
- Puertas, Juan-miguel et al. 2010. "Influence of specific growth rate over the secretory expression of recombinant potato carboxypeptidase inhibitor in fed-batch cultures of *Escherichia coli*." *Process Biochemistry* 45:1334–41.
- Rahmen, Natalie et al. 2015. "Exchange of single amino acids at different positions of a recombinant protein affects metabolic burden in *Escherichia Coli*." *Microbial cell factories* 14:10.
- Riesenberg, Dieter. 1991. "High-cell-density cultivation of *Escherichia coli*." *Current Opinion in Biotechnology* 2:380–84.
- Rosano, Germán L., and Eduardo A. Ceccarelli. 2014. "Recombinant protein expression in *Escherichia coli*: advances and challenges." *Frontiers in microbiology* 5:172.
- Rozkov, A. et al. 2004. "Characterization of the metabolic burden on *Escherichia Coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence." *Biotechnology and Bioengineering* 88(7):909–15.
- Ruiz, J. et al. 2009. "Alternative production process strategies in *E. coli* improving protein quality and downstream yields." *Process Biochemistry* 44:1039–45.
- Ruiz, Jordi, Glòria González, Carles de Mas, and Josep López-Santín. 2011. "A semiempirical model to control the production of a recombinant aldolase in high cell density cultures of *Escherichia coli*."
- Sans, Cristina et al. 2012. "Inclusion bodies of Fuculose-1-phosphate aldolase as stable and reusable biocatalysts." *Biotechnology progress* 28(2):421–27.
- Schmidt, F. R. 2004. "Recombinant expression systems in the pharmaceutical industry." *Applied Microbiology Biotechnology* 65:363–72.
- Sevastyanovich, Yanina et al. 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.
- Sevastyanovich, Yanina R. et al. 2012. "A generalised module for the selective extracellular accumulation of recombinant proteins." *Microbial cell factories* 11(1):69.

- Skerra, A. 1994. "Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*." *Gene* 151:131–35.
- Sooan, Shin, Dong-eun Chang, and Jae Gu Pan. 2009. "Acetate consumption activity directly determines the level of acetate accumulation during *Escherichia coli* W3110 growth." *Journal of Microbiology and Biotechnology* 19:1127–34.
- Sørensen, Hans Peter. 2010. "Towards universal systems for recombinant gene expression." *Microbial Cell Factories* 9:27.
- Studier, FW. 1991. "Use of bacteriophage T7 Lysozyme to improve an inducible T7 expression system." *Molecular Biology* 219:37–44.
- Tajima, Nami, Fumihiko Kawai, Sam-Yong Park, and Jeremy R. H. Tame. 2010. "A novel intein-like autoproteolytic mechanism in autotransporter proteins." *Journal of molecular biology* 402(4):645–56.
- Terpe, Kay. 2006. "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Applied Microbiology Biotechnology* 211–22.
- Tomohiro, Makino, Georgios Skretas, and George Georgiou. 2011. "Strain engineering for improved expression of recombinant proteins in bacteria." *Microbial cell factories* 10:32–42.
- Tracy, Bryan P., Stefan M. Gaida, and Eleftherios T. Papoutsakis. 2010. "Flow cytometry for Bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes." *Current Opinion in Biotechnology* 21(1):85–99.
- Trevors, J. T. 1986. "Plasmid curing in bacteria." *FEMS Microbiology Reviews* 32(3-4):149–57.
- Valdez-Cruz, Norma a, Luis Caspeta, Néstor O. Pérez, Octavio T. Ramírez, and Mauricio a Trujillo-Roldán. 2010. "Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda *pL* and/or *pR* Promoters."
- Vandermeulen, Gaëlle, Corinne Marie, Daniel Scherman, and Véronique Prétat. 2011. "New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials." *Molecular therapy: the journal of the American Society of Gene Therapy* 19(11):1942–49.
- Vick, Jacob E. et al. 2011. "Optimized compatible set of BioBrick™ vectors for metabolic pathway engineering." *Appl Microbiol Biotechnol* 92(6):1275–86.

- Vidal, Luis. 2006. "Producción de aldolasas recombinantes : de la biología molecular al desarrollo de procesos." Doctoral thesis.
- Vidal, Luis, Olga Durany, T. Suau, Pau Ferrer, and Gloria Caminal. 2003. "High-level production of recombinant his-tagged Rhamnulose 1-phosphate aldolase in *Escherichia coli*." *Journal of Chemical Technology and Biotechnology* 78:1171–79.
- Vidal, Luis, Pau Ferrer, Gregorio Alvaro, M. Dolors Benaiges, and Gloria Caminal. 2005. "Influence of induction and operation mode on recombinant rhamnulose 1-phosphate aldolase production by *Escherichia coli* using the T5 promoter." *Journal of Biotechnology* 118:75–87.
- Vidal, Luis, Jaume Pinsach, Gerald Striedner, and Pau Ferrer. 2008. "Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in *Escherichia coli*." *journal of Bacteriology* 134:127–36.
- Voigt, Christopher A. 2006. "Genetic parts to program bacteria." *Current Opinion in Biotechnology* 17:548–57.
- Weber, Jan, Frank Hoffmann, and Ursula Rinas. 2002. "Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: redirection of metabolic fluxes." *Biotechnology and bioengineering* 80(3):320–30.
- Wyre, Chris, and Tim W. Overton. 2014. "Flow cytometric analysis of *E. coli* on agar plates: implications for recombinant protein production." *Biotechnology letters* 36(7):1485–94.
- Xu, Bo, Mehmedalija Jahic, and Sven-Olof Enfors. 1999. "Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*." *Biotechnology progress* 15:81–90.
- Xu, Jia, and Kathleen S. Matthews. 2009. "Flexibility in the inducer binding region is crucial for allostery in the *Escherichia coli* lactose repressor." 48(22):4988–98.
- Yee, L., and H. W. Blanch. 1993. "Recombinant trypsin production in high cell density fed-batch cultures in *Escherichia coli*." *Biotechnology and Bioengineering* 41:781–90.
- Yin, Jiechao, Guangxing Li, Xiaofeng Ren, and Georg Herrler. 2007. "Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes." *Journal of biotechnology* 127(3):335–47.
- Yokobayashi, Yohei, Ron Weiss, and Frances H. Arnold. 2002. "Directed evolution of a genetic circuit." *Applied Biological Sciences* 99(26):16587–91.

Zuo, Peijun, and Bakr M. Rabie. 2010. "One-step dna fragment assembly and circularization for gene cloning." *Current issues molecular biology* 12:11–16.