Development of innovative screening procedures and fermentation processes for the production of recombinant proteins in *E. coli*

By

Tania Selas Castiñeiras

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the University of Birmingham

for the degree of DOCTOR OF PHILOSOPHY

School of Chemical Engineering

College of Engineering and Physical Sciences

University of Birmingham

February 2017

Abstract

From the plethora of possible microbial hosts, *Escherichia coli* remains the most widely used "microbial factory" for recombinant protein production (RPP). However, despite the numerous advances, RPP in *E. coli* is still a significant challenge. Strong promoters to achieve high expression and protein production levels are often used, however, the underlying effects on the host physiology are often unseen.

This thesis reports the development of "stress-minimisation" approaches for the production of recombinant proteins, either targeting the cytoplasm or the periplasmic space. First, the fermentation conditions for the production of TNFα in the cytoplasm were optimised. The expression vector, culture medium, temperature, inducer concentration and induction point were optimised, yielding 5.35 $g \cdot L^{-1}$ of rhTNFα, 70% being accumulated as a soluble product.

Second, different approaches for the production of recombinant proteins targeting the periplasm were evaluated, using an antibody fragment, scFv163R4, as a model protein. The effect of different growth conditions and signal peptides on the production of the scFv163R4 were evaluated. However, the selection of the optimal signal peptide was proven to be challenging, establishing the requirement of a high-throughput screening assay. This resulted in the development of a screening assay using β-lactamase as a reporter protein, for the evaluation of mutant signal peptide libraries with improved translocation activity. Initial evaluations resulted in the selection of two mutant signal peptides with enhanced translocation of scFv163R4-A, yielding almost 1 $g \cdot L^{-1}$ of periplasmic scFv163R4-A. The generation of mutant signal peptide libraries in combination with the β-lactamase screening assay represents an important advance for the production of disulphide bonded proteins for the biotechnology industry.

Dedication

To Carlos, Rafa and Helena,

Acknowledgments

First of all, I would like to thank Professor Jeff Cole for his trust in me through all these years, for pushing me when required but also motivating me when I needed; without his support, none of this work would have been possible. I would also like to thank my supervisor, Dr Tim Overton, for his guidance during the project, his continuous support and encouragement, especially during the writing of this manuscript. Dr Steven Williams, for being my mentor, his continuous guidance, not only scientifically but also as a friend, answering all my questions (even if they involved maths!) and guiding me whenever I felt lost during the project. I could never thank him enough for all the time and effort spent on this project. Dr Daniel Smith and Tony Hitchcock, for always being enthusiastic about the project and his continuous support, help and advice during these 4 years. Dr Paola Salerno, for her useful scientific advices but more important than that, for being a friend. Dr Annelise, Parisa and Daniela, for all the support and patience during the last years and all the "Costa coffee moments".

A very special thanks to my mum, for her blind trust in me in a way that only a mum can do, and my dad for always being there. I am also grateful to my friends, specially, Carlos, Rafa and Helena, for being so understandable, providing me with useful advices and being there for me. To Carlos, because he has always encouraged me to continue, comforting me and reminding me that I could do it.

I would like to thank the funding bodies, which have made this project possible, InnovateUK, formerly the Technology Strategy Board (TSB), the Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC) and Cobra Biologics. Special thanks to Cobra Biologics, for providing me with a place to carry out my research, funding the project, and providing all the equipment required, and more important, thank you to all the lovely people who work there.

Table of contents

List of figures

List of tables

List of abbreviations

Amp	Ampicillin
Bp	Base pairs
CFU	Colony forming units
cGMP	Current good manufacturing practices
C-terminal	Carboxyl terminal
CV	Column volume
DO	Dissolved oxygen
DCW	Dry cell weight
DNA	Deoxyribonucleic acid
DMSO	Dimethylsulfoxide
EDTA	Ethylene diamine tetra acetic acid
ELISA	Enzyme-linked immunosorbent assay
EMA	European medicines agency
Fab	Fragment antigen-binding
FDA	Food and drug administration
HCDC	High-cell-density culture
HRP	Horseradish peroxidase
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
IBs	Inclusion bodies
Kan	Kanamycin
LB	Luria broth (Lennox)
mAb/Mab	Monoclonal antibody

M-H	Mueller-Hinton
MIC	Minimum inhibitory concentration
N-terminal	Amino terminal
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline – Tween 20
PCR	Polymerase chain reaction
RCB	Research cell bank
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPP	Recombinant protein production
scFv	Single-chain variable fragment
SDS	Sodium dodecyl sulphate
SRP	Signal recognition particle
TAT	Twin-arginine translocase
TB	Terrific broth
TCP	Total cell protein
TMB	3,3′,5,5′-tetramethylbenzidine
WCW	Wet cell weight

Chapter 1:

Introduction

1.1 Industrial recombinant protein production: an overview

1.1.1 The rise of the biotechnology industry

From ancient times, living organisms have been widely exploited for the development or production of different bioproducts. Early biotechnology dates back to 6,000 BC, as evidence has been found that beer was already being produced by the Sumerians. In the brewing process, malting grains converts starch into sugars, which are broken down into alcohols by yeast to produce beer. In 3,500 BC, other fermentation processes were developed in China, such as wine and soya fermentations. Soon after, the development of lactic fermentation allowed the production of other forms of food, such as cheese and yogurt (Buchholz and Collins, 2013).

In the 1800s, mainly beer, but also wine, acetic (e.g. vinegar) and lactic acid fermentations significantly contributed to the economy of the United Kingdom and other European countries. However, the nature of the fermentation was attributed to "secret living forces" emerging from spontaneous generation and the concept of the fermentation process is often found associated to mystic concepts. In 1830s, Schwann and Cogniard-Latour were the first to describe little globular bodies, the yeast of the beer, as responsible of alcoholic fermentation. Two decades later, Louis Pasteur developed the initial theories about fermentation, settling the scientific debate on the nature of fermentation in favour of the role of living organisms. Pasteur's work not only included the development of the first fermentation theories, but also a sterilisation process, pasteurisation, which is still used today by the dairy industry (Buchholz and Collins, 2013).

In 1928, Alexander Fleming discovered the mould *Penicillum*, and his work led to the purification of the antibiotic compound that we today known as penicillin. The large demand for this antibiotic during the Second World War facilitated the development of the penicillin manufacturing process. Penicillin initiated the era of antibiotics, being among the first medications to be effective against bacterial infections caused by Staphylococci and Streptococci (Buchholz and Collins, 2013).

The turning point in genetics arose from the establishment of a model for the molecular structure of deoxyribonucleic acid (DNA) by Watson and Crick, supported by the work carried out by Rosalind Franklin at Wilkins' laboratory (Buchholz and Collins, 2013). Only 20 years later, Boyer and Cohen developed the recombinant DNA technology, showing that genetically engineered DNA molecules can be cloned into foreign cells, laying the foundations of gene cloning (Buchholz, 2007; Buchholz and Collins, 2013). In 1975, Milstein and Köhler developed the hybridoma technology enabling the production of monoclonal antibodies, which led to a new era of therapeutic biologics (Weiner *et al.*, 2010). In 1976, Boyer in collaboration with Robert Swanson founded Genentech, which only a few years later developed the production of the recombinant human insulin. In 1982, the Food and Drugs Agency (FDA) approved the commercialisation of Humulin®, the recombinant human insulin produced in *Escherichia coli*, for the treatment of type I diabetes, being the first recombinant protein drug to reach the market. The precedent of its approval favoured the production of other recombinant proteins, such as human growth hormone, β-interferon and the hepatitis B vaccine. Nowadays, recombinant proteins, including hormones, growth factors, cytokines and monoclonal antibodies, represent a significant part of total revenue obtained by the pharmaceutical industry (Buchholz and Collins, 2013).

In 1997, an important announcement revolutionised public opinion when Wilmut and Campbell proclaimed the birth of the sheep "Dolly", the first mammal cloned from an adult somatic cell by nuclear transfer. A long public debate concerning the possible outcomes of animal cloning and its potential application to humans arose from this announcement.

Genetic engineering had not only transformed the pharmaceutical industry, but also had a great impact in other types of industries, such as agriculture, environmental technology, and small chemicals production. Plant biotechnology led to a new "green revolution" with the development of genetically modified (GM) crops, which can provide higher yields, minimising the use of herbicides and pesticides. The first transgenic product, Flavr Savr® tomato, was approved in 1994,

and since then, GM crops have been massively cultivated in the United States, Canada, Brazil and Argentina (Buchholz and Collins, 2013). Nonetheless, controversy still exist with respect to ethical and biosafety implications of GM products, with a still growing negative public opinion in Europe.

A detailed timeline of the history of biotechnology can be found in **Figure 1.1**.

1.1.2 Economic impact of the biotechnology industry

Since the advent of recombinant DNA technology in the 70s, the biotechnology market has grown to reach an average compound annual growth rate (CAGR) of 11.6% from 2012 to 2017, and it is expected to reach a value worth $414.5 billion by the end of 2017. Industrial biotechnology has global annual sales of over $250 billion with products varying from pharmaceuticals, chemicals, industrial enzymes, cosmetics, food or dietary supplements (Meyer and Schmidhalter, 2014). Just the market value of the biopharmaceuticals has reached a total cumulative sales value of $140 billion in 2013, which exceeds the reported gross domestic product (GDP) of three-quarters of the economies in the world (Walsh, 2014).

A list detailing the top 10 best-selling pharmaceutical products in 2015 can be found in **Table 1.1**, showing the penetration and economic impact of biotechnology, which accounted for 40% of the total sales of the pharmaceutical industry in 2013 (Spadiut *et al.*, 2014). The market is led by Humira®, the first full human monoclonal antibody drug approved by the FDA in 2002, and still today ranks first among the best-selling pharmaceutical products, accounting for $14 billion of sales in 2015. Other monoclonal antibody-derived products, included among the best-selling biopharmaceuticals, are Remicade, MabThera/Rituxan, Avastin and Herceptin. Humira® is closely followed in sales by Enbrel®, a chimeric protein generated by the fusion of the TNFα receptor to the constant end of the IgG1 antibody. Monoclonal antibodies, including Fc-fusions, generated sales of $75.7 billion in 2013 (Walsh, 2014).

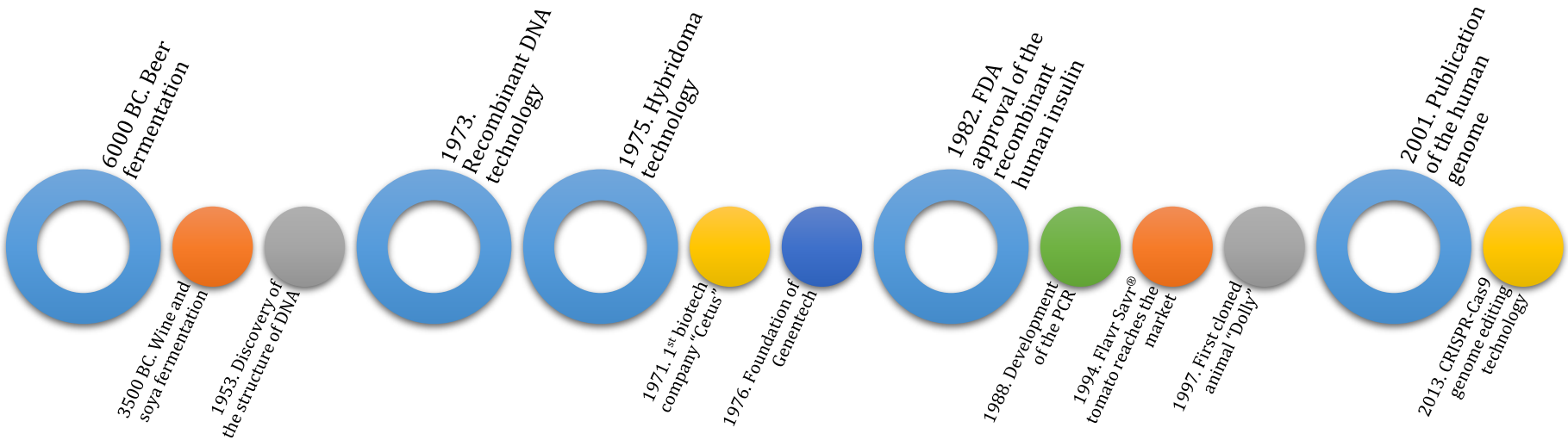


Figure 1.1. Timeline of the major breakthroughs in the history of the biotechnology industry (Adapted from Buchholz and Collins, 2013).

Table 1.1. The top worldwide 10 best-selling pharmaceuticals in 2015 (PharmaCompass, 2016).

Rank	Trading name	Generic name	Type of molecule	Company	Disease/Medical use	Total sales
1	Humira®	Adalimumab	Monoclonal antibody	AbbVie	Rheumatoid arthritis	$14 billion
2	Harvoni®	Ledipasvir/ Sofosbuvir	Small molecule	Gilead Sciences	Hepatitis C	$13.8 billion
3	Enbrel®	Etanercept	TNFα-receptor fusion protein	Amgen / Pfizer	Rheumatoid arthritis	$8.7 billion
4	Remicade®	Infliximab	Monoclonal antibody	Johnson & Johnson / Merck	Crohn's Disease Rheumatoid Arthritis	$8.4 billion
5	MabThera®/Rituxan®	Rituximab	Monoclonal antibody	Roche	Lymphoma, Leukaemia Autoimmune disorders	$7.1 billion
6	Lantus®	Insulin glargine	Recombinant protein	Sanofi-Aventis	Diabetes mellitus	$7 billion
7	Avastin®	Bevacizumab	Monoclonal antibody	Roche	Metastatic cancers	$6.8 billion
8	Herceptin®	Trastuzumab	Monoclonal antibody	Roche	Breast cancer	$5.8 billion
9	Revlimid®	Lenalidomide	Small molecule	Celgene Corporation	Multiple myeloma Myelodysplastic syndromes	$5.8 billion
10	Sovaldi®	Sofosbuvir	Small molecule	Gilead Sciences	Hepatitis C	$5.3 billion

Among non-antibody based products, insulins are the most lucrative product class, collectively generating sales of $21.5 billion in 2013. Lantus®, Levemir®, Novolog® and Humalog®, all insulin-derived products, are included between the 50 largest selling pharmaceutical products, constituting the bulk of products manufactured using microbial platforms (Walsh, 2014).

1.1.3 The next generation of biopharmaceuticals

Recombinant proteins, interferon and growth factors

Recombinant proteins constitute a significant portion of the biopharmaceutical market, including a broad range of products, such as growth factors, hormones, cytokines and anticoagulants. As depicted in **Figure 1.2**, among cytokines, interferon products (IFNα, IFNβ and IFNγ) represent 2% of the biologicals in preclinical or early phase clinical development. Growth factors, such as erythropoietin-stimulating agents (ESAs), colony-stimulating factors (CSFs), and anticoagulants, such as recombinant tissue-plasminogen activator (rt-PA), have been extensively commercialised by the pharmaceutical industry (Aggarwal, 2014).

Monoclonal antibodies, antibody fragments and derivatives

Today, over 200 recombinant proteins have reached the market, leading the number of new biological entities (NBE) approved during the last years (Walsh, 2006; Walsh, 2010). This exceptional market success is largely derived from the commercialisation of monoclonal antibodies, the fastest growing class of biopharmaceutical product (Carter, 2011; Spadiut *et al.*, 2014). Monoclonal antibody-based products accounted for 33% of the biological products in early phases of development in 2012 (**Figure 1.2**).

Since the approval of the first monoclonal antibody in 1986, known as muromonab-CD3 (OKT3), for acute allograft rejection, monoclonal antibodies have been engineered to optimise their

properties and reduce undesired effects. However, limitations are often associated with the development of monoclonal antibodies, such as cost-intensive manufacturing processes, low tissue penetration and rapid renal clearance (Carter, 2011). Small antibody fragments, such as antigen-binding (Fab) and single-chain variable (scFv) fragments and other engineered variants are emerging as credible product candidates, representing respectively the 6% and 4% of the monoclonal antibody-derived products in early phases of development (**Figure 1.2**) (Nelson and Reichert, 2009).

Vaccines

Recombinant vaccines represent 33% of the products in preclinical and clinical development **(Figure 1.2).** This sector has experienced a high growth mainly driven by the blockbuster vaccine Gardasil®, used to protect against certain strains of human papillomavirus (Aggarwal, 2014).

Gene-based therapies: antisense, RNA interference and gene therapy

Gene therapy can be defined as the therapeutic delivery of nucleic acids into a patient's cell with the aim of altering gene expression to treat a pathological process by the addition of DNA or RNA oligonucleotides or by the use of viral vectors (Kay, 2011).

Antisense oligonucleotides are single-stranded DNA molecules designed to be complementary to a specific gene's mRNA, inhibiting the expression of a target gene (Gleave and Monia, 2005). RNA interference (RNAi) are double-stranded RNA oligonucleotides used for sequence-specific gene silencing, using the cell's natural machinery to target specific mRNA within the cell, inducing its degradation and resulting in gene suppression (Kay, 2011). Both antisense and RNAi technologies are still facing numerous obstacles; however, an increasing number of these molecules are nowadays in clinical trials (Gleave and Monia, 2005; Walsh, 2005).

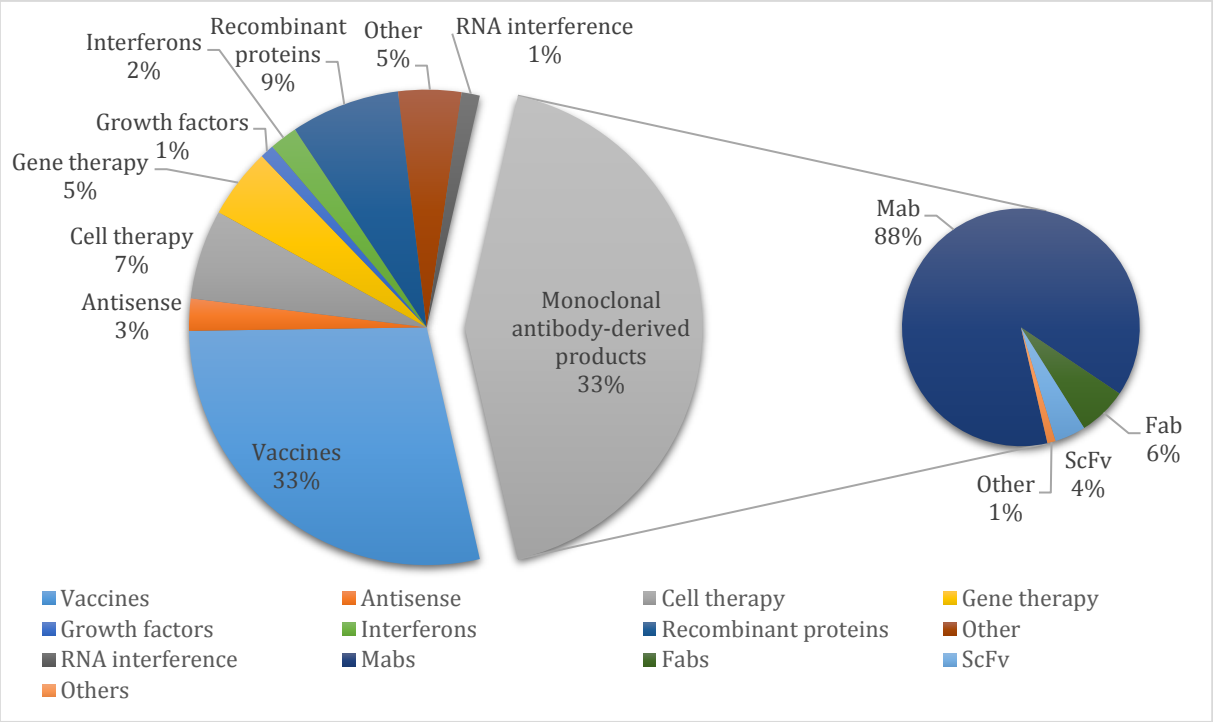


Figure 1.2. Biopharmaceuticals in preclinical or clinical development in 2012 (Adapted from Nelson and Reichert, 2009; PhRMA, 2013).

Viral platforms are the most credited DNA delivery method. So far, 70% of the gene therapy products in clinical trials have used virus as method of delivery, such as retrovirus, lentivirus, adenovirus or adeno-associated virus (Yin *et al.*, 2014). After the death of a patient in 1999 following a high dose of an adenoviral product, today gene therapy is facing a renaissance. In 2012, the lipase lipoprotein deficiency therapy Glybera® got approval by the European Medicine Agency (EMA), leading to the resurgence of gene therapy (Walsh, 2014).

Cell therapy

Cell therapy is based on the injection of cellular material, meaning intact living cells, into a patient. Several cell therapies have been recently approved, such as tissue-engineered skin, cartilage repair products and the first personalised cellular immunotherapy for cancer treatment. Cell therapy is still in early phases of development and facing numerous obstacles associated with the manufacturing process, such as the high cost of goods, the process variability and complexity of scale-up (Brandenberger *et al.,* 2011).

1.2 Recombinant protein production

1.2.1 Production hosts for recombinant protein production

E. coli as a host for recombinant protein production

Escherichia coli is a Gram-negative, rod-shaped bacterium, commonly found in the lower intestine of warm-blooded organisms. *E. coli* is the most widely studied prokaryotic model organism, playing a major role in bioengineering and industrial microbiology due to its fast growth, well-known genetics, physiology and biochemistry (Chou, 2007). *E. coli* is considered the main workhorse for the production of non-glycosylated recombinant proteins, being one third of the approved protein therapeutics produced using this host organism (Terpe, 2006).

Nevertheless, there are some important disadvantages of *E. coli* as an expression host for recombinant protein production (RPP), such as the high level of endotoxin production. Lipopolysaccharides, also known as endotoxin, the major component of the outer membrane of Gram-negative bacteria are pyrogenic for humans, so their removal from the final product is an essential requirement (Petsch and Anspach, 2000). In addition, *E. coli* lacks the machinery required to perform certain post-translational modifications, such as glycosylation (Terpe, 2006).

Escherichia coli K-12 and its derivatives are the most widely used strains in the biotechnology industry and classified in risk group 1 in biosafety guidelines. Commonly used K-12 strains include *E. coli* strains MG1655, DH1, DH5α and W3110 (Huang *et al.*, 2012).

Alternatively, *E. coli* B strains are also widely used for RPP due to their low acetate accumulation, even when high concentrations of glucose are used as carbon source (Huang *et al.*, 2012). Among others, *E. coli* BL21 is one of the most commonly used strains and has a proven track record for protein expression studies (Sørensen and Mortensen, 2005). *E. coli* BL21, and its parental strain B834, are deficient in *lon* and *ompT* proteases, associated to proteolytic degradation of recombinant proteins. Other strains, such as Origami (Novagen) are deficient in *trxB/gor* genes, which facilitate cytoplasmic disulphide bond formation; Tuner strains (Novagen) are *lacY* mutants which allows precise control of gene expression by fine tuning of the inducer concentration; or C41 and C43 strains, also known as Walker strains, for improved production of membrane proteins (Shadev *et al.*, 2008; Miroux and Walker, 1996). Further details regarding *E. coli* strains commonly used for RPP can be found in **Table 1.2**.

Other prokaryotic hosts for recombinant protein production

Pseudomonas spp., another Gram-negative host used for RPP, are widely known for their rapid growth and secretion ability, considered as a promising alternative to *E. coli* (Krzeslak *et al*., 2009).

Table 1.2. *E. coli* strains frequently used for recombinant protein production (Adapted from Terpe, 2006).

Strain	**Key features**
K-12 strains	
W3110	One of the oldest laboratory strains of *E. coli* K-12, cured of lambda and F
HMS174	*recA* mutant, Rif resistance
JM 83	Usable for secretion of recombinant proteins into the periplasm
AD494	*trxB* mutant, facilitates cytoplasmic disulphide bond formation
Origami	*trxB/gor* mutant, facilitates cytoplasmic disulphide bond formation
B strains	
B834	Parental strain of BL21, methionine auxotroph, ^{35}S-met labelling
BL21	Deficient in *lon* and *ompT* proteases
BLR	*recA* mutant, stabilises tandem repeats, deficient in *lon* and *ompT* proteases
C41 and C43	Mutants selected for expression of membrane proteins
Origami B	*trxB/gor* mutant, facilitates cytoplasmic disulphide bond formation, deficient in *lon* and *ompT* proteases
Rosetta	Enhances the expression of eukaryotic proteins that contain codons rarely used in *E. coli*: AUA, AGG, AGA, CGC, CUA, CCC and GGA. Deficient in *lon* and *ompT* proteases

Bacillus subtilis is a Gram-positive rod-shaped bacterium known for its great secretion ability for host enzymes or enzymes obtained from related strains. Nowadays, 60% of enzymes on the market are produced in *Bacillus spp.*, mainly alkaline proteases for detergents and amylases for baking and starch production (Demain and Vaishnav, 2009). However, secretion of heterologous proteins is often hampered in this host organism (Westers *et al*., 2004). Other Gram-positive hosts, as *Streptomyces* or *Lactoccocus* have been also used for RPP.

Eukaryotic hosts for recombinant protein production

Yeast has been used extensively by the biotechnology industry, *Saccharomyces cerevisiae* and *Pichia pastoris* being the preferential host choices for RPP. Yeast, as bacteria, can grow in non-expensive culture media to high cell densities, but can produce complex post-translational modifications (PTMs), such as glycosylation (Schmidt, 2004). However, the glycosylation pattern generated by yeast is not the same as those found in human proteins, as yeast produces N-linked oligosaccharides containing mainly mannose residues (hyperglycosylation). Hyperglycosylated proteins have a reduced half-life and can be immunogenic (Demain and Vaishnav, 2009).

Filamentous fungi, such as *Aspergillus niger*, are also used for the biotechnology industry mainly for the production of acids, antibiotics or enzymes, e.g. celullases and amylases (Schmidt, 2004).

Mammalian cells as a host for recombinant protein production

Mammalian cells have become the dominant option for RPP, and almost 70% of the recombinant proteins produced by the pharmaceutical industry are produced using mammalian cell lines (Demain and Vaishnav, 2009; Schmidt, 2004). Mammalian cultures are used mainly for the production of monoclonal antibodies due to their ability to correctly fold and assemble complex proteins and make post-translational modifications with a high similarity to human cells

(Schmidt, 2004). Numerous mammalian cell lines are available, widely used examples being Chinese Hamster Ovary (CHO), Baby Hamster Kidney (BHK), mouse myeloma derived-cells (NS0) and Human Embryo Kidney (HEK-293) cell lines (Wurm, 2004).

The optimisation of culture conditions for mammalian fermentation is the results of years of research and development of vectors, cell lines and medium composition. Mammalian cell cultures can grow adhered to the walls of culture containers, usually roller bottles filled with 10 - 30% of culture media, or in suspension in stirred tank reactors (STRs) or wave bioreactors, the last ones being the preferential choice for RPP (Wurm, 2004).

1.2.2 Expression systems for RPP in E. coli

1.2.2.1 Design of the expression vector

Gene expression efficiency and recombinant protein production is highly dependent on the expression vector design, including genetic elements such as the origin of replication, antibiotic resistance markers, promoters and translation initiation regions (**Figure 1.3**) (Sørensen and Mortensen, 2005). The gene dosage, promoter strength, mRNA stability and the efficiency of translation often have a great impact on the final rate of protein synthesis, and therefore, in the recombinant protein yield (Huang *et al.,* 2012).

Gene dosage

The origin of replication (*ori*) is the sequence where replication is initiated in the vector, which determines the copy number and consequently, the gene dosage. Both high copy (e.g. pUC vectors, 500 to 700 copies) and medium copy (e.g. pBR322, 15 to 20 copies) number vectors have been used for RPP, however, medium copy vectors are recommended to avoid segregational instability associated to high copy number vectors (Huang *et al.,* 2012).

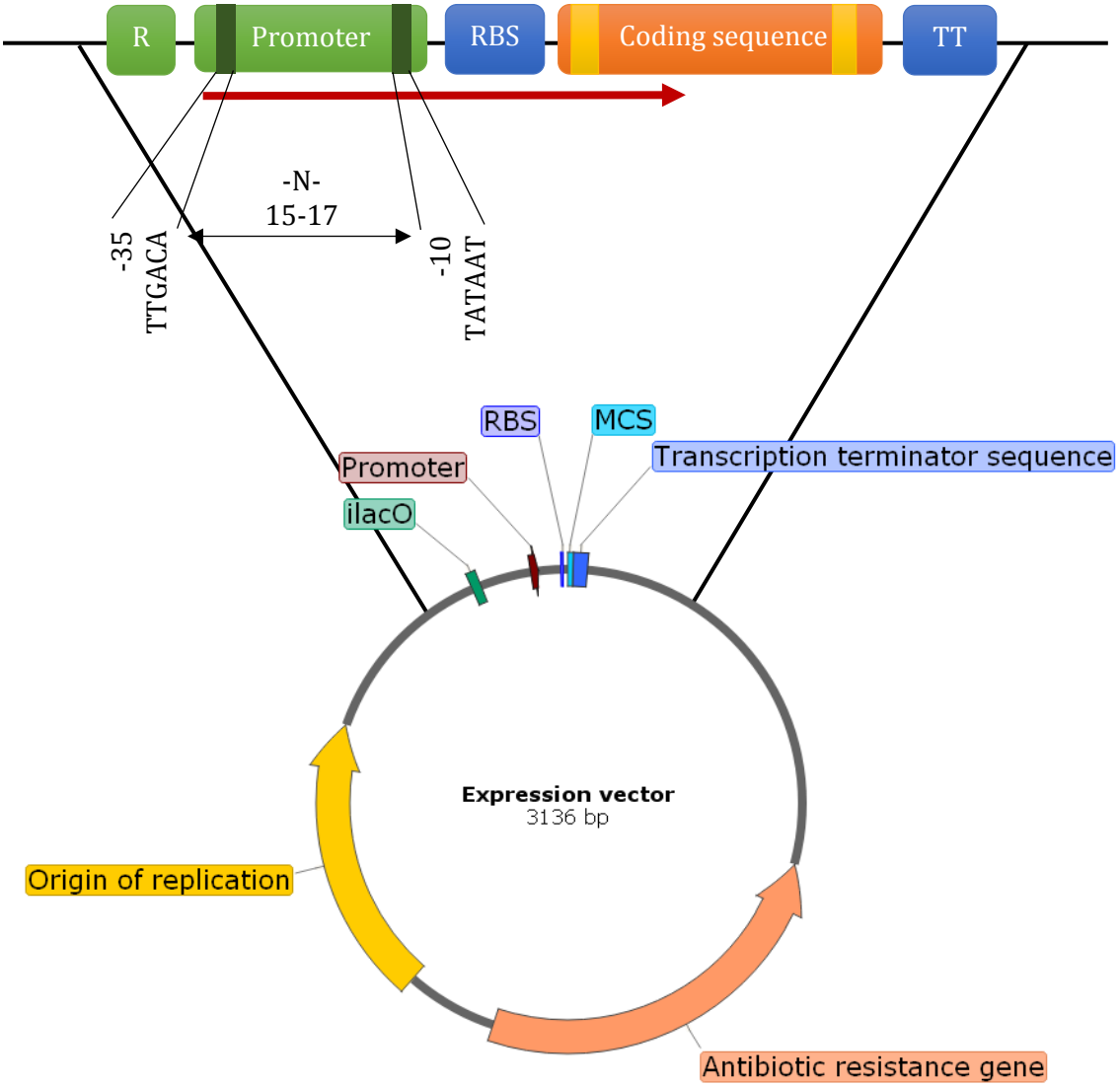


Figure 1.3. Diagram of a model expression vector. Artificially constructed vectors contain essential components including: the origin of replication (ori); the promoter upstream of the ribosome-binding site (RBS); the multiple cloning site (MCS) where the gene of interest will be inserted; transcriptional terminator sequences; the antibiotic resistance gene; and some vectors may contain a reporter protein or a protein purification tag (Adapted from Makrides, 1996).

Resistance markers

Resistance markers conferring resistance to antibiotics are regularly present in the plasmid backbone to ensure plasmid maintenance. Commonly used antibiotic resistance genes are ampicillin, which inhibits the formation of peptidoglycan cross-linking in the bacterial cell wall; and kanamycin, chloramphenicol or tetracycline, which bind to ribosomes interfering with protein synthesis (Sørensen and Mortensen, 2005).

Transcriptional regulation

Promoters are DNA sequences, localised 10 to 100 bp upstream of the ribosome-binding site. The RNA polymerase recognises two major sequence elements localised in the promoter sequence, -35 and -10 (Pribnow box) bases upstream of the transcriptional initiation (**Figure 1.3**). Initiation of transcription by the RNA polymerase is regulated by the sequence of the promoter, and interaction with transcription factors (TFs), which respond to external stimuli such as changes in the culture media or the addition of chemical compounds, such as sugars. A more detailed description of commonly used promoters for RPP can be found in **Section 1.2.2.2**.

The ribosome-binding site (RBS) is a sequence of nucleotides upstream of the start codon, which interacts with the 16S subunit, responsible for the recruitment of the ribosome during translation initiation. The RBS is located between 5 to 13 bp upstream of the start codon and contains a 6-base consensus sequence, AGGAGG, known as the Shine-Dalgarno sequence (**Figure 1.3**). Transcription terminators, inverted repeats prone to form a stem-loop structure, are often located downstream of the gene, stabilising the mRNA (Makrides, 1996).

Translational regulation

Translation efficiency is determined by the sequence and structure of the translation initiation region (TIR), which comprises the Shine-Dalgarno (SD) sequence, the start codon (AUG), the spacing between the SD and the start codon and translation enhancers. The modulation of TIR sequences is known to have a great impact on recombinant protein expression, as poor transcription efficiency often results in low recombinant protein yields (Huang *et al.,* 2012).

There are three start codons, AUG, GUG and UGG, the most commonly used is AUG, which is present in 91% of *E. coli's* genes. There are three stop codons, UAA, UGA and UAG but translational termination is preferably mediated by UAA. If desired, more than one stop codon can be included at the end of the coding sequence to avoid ribosome skipping (Huang *et al.,* 2012).

Codon usage

The codon usage in *E. coli* and other organisms vary affecting protein translation. Expression of genes containing rare codons can lead to premature translation termination, growth arrest or amino acid missincorporation. Codon optimisation may help to overcome this problem, substituting rare codons for those preferentially used by *E. coli* (Sahdev *et al.*, 2008). Another approach is the overexpression of tRNAs coding for rare codons from a vector, as *E. coli* Rosetta (DE3) which carries the pRARE plasmid (Sørensen and Mortensen, 2005; Huang *et al.,* 2012).

1.2.2.2 Commonly used E. coli expression systems

Lactose expression system and derivatives

The lactose operon contains the sequence coding for the genes required for the transport and catabolism of lactose in *E. coli*. In the absence of lactose, the synthesis of mRNA is repressed to

avoid the waste of energy used in the synthesis of proteins when not required. If lactose is not present in the media, the Lac repressor (LacI), binds to the operator sequence, blocking the transcription of the *lacZYA* genes located downstream of the promoter (**Figure 1.4 A**). In contrast, when lactose is present in the medium, the lactose molecules bind to the Lac repressor, inducing a conformational change, which triggers its dissociation from the operator sequence (**Figure 1.4 B**). The transcription of the *lacZYA* genes is higher when glucose is not present, as the cAMP levels increase binding to the catabolite activator protein (CAP). The cAMP-CAP complex interacts with the RNA polymerase stimulating the synthesis of mRNA (**Figure 1.4 C**) (Lodish *et al.,* 2011).

The lactose operon is the most studied regulatory mechanism in *E. coli*, and several promoters have been constructed from *lac*-derived regulatory elements (Terpe, 2006). The *lac* expression system is based on the *lac* operon, where the genes coding for the enzymes required for lactose metabolism have been substituted for a gene of interest. In addition, the inducer lactose is commonly substituted for isopropyl-β-D-thiogalactopyranoside (IPTG), a sugar analogue, whose uptake is not dependent on permeases. IPTG diffuses freely through the cytoplasmic membrane, so the level of induction depends on the concentration of the inducer present in the medium. In addition, IPTG cannot be cleaved by the β-galactosidase, and therefore is not metabolised (Schumann and Ferreira, 2004).

Several mutations of *lac* promoter have been identified, such as the *lacUV5*, which presents two point mutations within the -10 region of the promoter and one in the -66 region of the CAP binding site. This mutated version showed a 2.5-fold activity increase over the wild type promoter and a lower sensitivity to catabolite repression, allowing maximal expression in rich culture media (Terpe, 2006). As the level of the wild-type LacI repressor is not sufficient to completely repress the expression of the genes located downstream of the *lac* promoter, mutant variants have been isolated, such as $LacI^{q}$, allowing a lower basal expression when compared with the wild type (Sørensen and Mortensen, 2005).

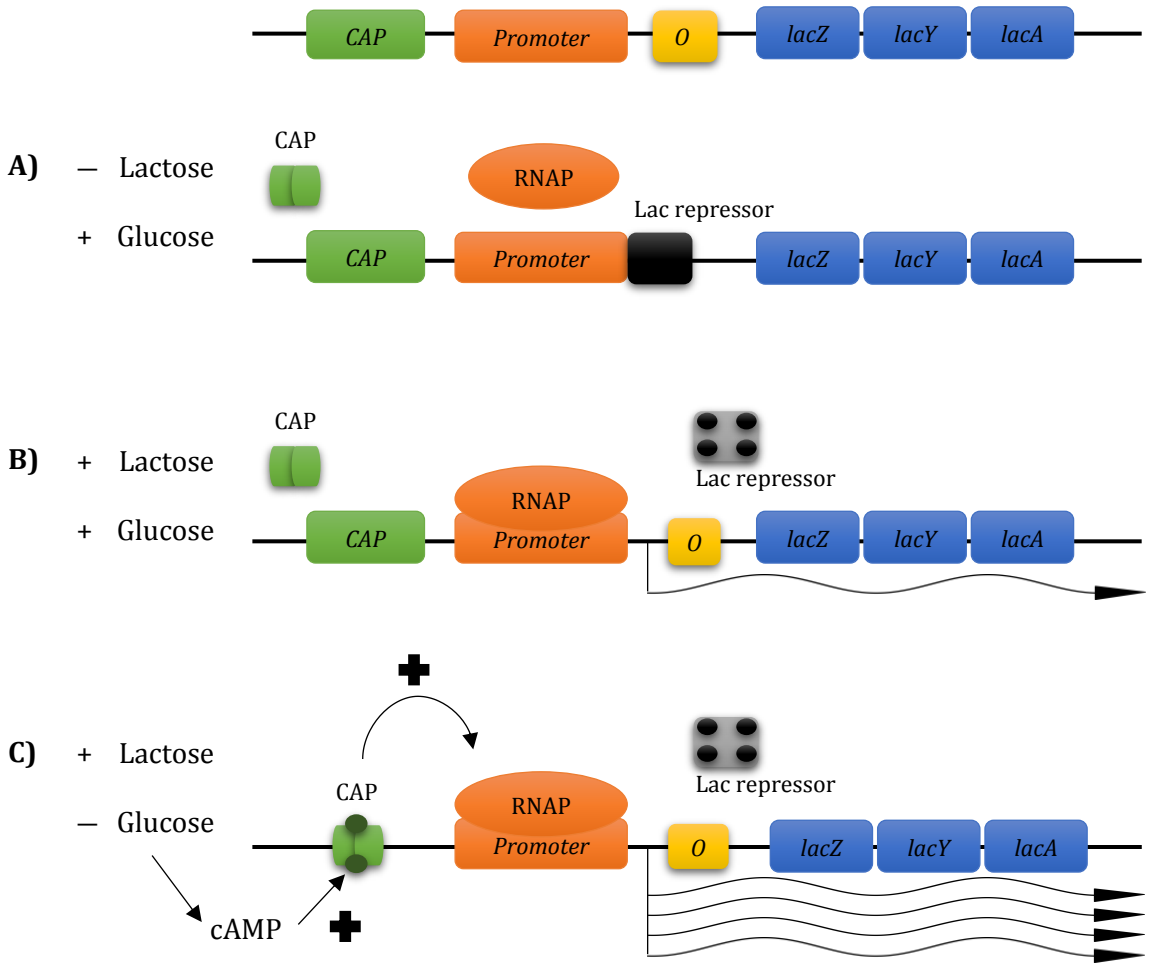


Figure 1.4. Regulation of the transcription of the *lac* operon. The transcription-control region of the *lac* operon includes three protein-binding regions: the CAP binding site, the *lac* promoter and the *lac* operator (O). When lactose is not present in the culture media, low amounts of mRNA are produced since the *lac* repressor is bound to the operator sequence, inhibiting the initiation of transcription by the RNA Polymerase (RNAP) (A). When lactose and glucose are present in the culture medium, the *lac* repressor binds lactose, dissociating from the operator sequence, allowing transcription at low rates (B). When only lactose is present in the culture medium, the level of cAMP increases as a consequence of the low intracellular levels of glucose, constituting the CAP-cAMP complex, which binds to the CAP binding site, interacting with the RNA polymerase, allowing gene transcription at high rates (C) (Adapted from Lodish *et al.,* 2013)

Synthetic promoters were created with the desire to further improve the strength of the *lac* promoter by constructing a *trp*/*lac* hybrid promoter combining the -35 and -10 regions of the tryptophan and lactose promoters, respectively. The increase in the spacing between the −35 and −10 consensus sequences 1 bp or 2 bp resulted in an improvement of the promoter efficiency, leading to the development of the *tac* and *trc* promoters (Terpe, 2006).

Nonetheless, even when hybrid promoters, such as *tac* and *trc*, allowed the production of proteins up to 15 to 30% of the total cell protein, its high basal expression under non-induced conditions represents an important drawback (Terpe, 2006). Stronger and more tightly regulated expression systems have been developed, as leaky expression of certain recombinant proteins may result toxic for the host organism.

T7 expression system

Podoviral polymerases are widely used by the biotechnology industry, especially the RNA polymerase (RNAP) derived from the bacteriophage T7. The T7 RNA polymerase is known for its high specificity for its cognate promoter, which cannot be efficiently recognised by the *E. coli* RNA polymerase. In addition, it is five times more active than *E. coli's* RNA polymerase (Fernandez and Hoeffler, 1999). The characterisation of the T7 RNA polymerase led to the development of a series of expression systems, known as plasmids for expression by T7 RNA polymerase or pET vectors, developed by Studier and collaborators (Studier, 1990; Rosenberg *et al.,* 1987; Studier and Moffart, 1986).

The most common strategy for the T7 polymerase delivery in *E. coli* makes use of BL21 (λDE3), whose chromosome contains a prophage (λDE3), encoding the T7 RNA polymerase (bacteriophage T7 gene 1) under the control of the *lacUV5* promoter (Sørensen and Mortensen, 2005). Since the T7 promoter is not completely repressed by LacI, some molecules of T7 RNA

polymerase are continuously expressed, generating considerable amounts of target mRNA even in the absence of inducer. Mutant versions of LacI repressor, such as $LacI^q$, or the co-expression of the T7 lysozyme, the natural inhibitor of the T7 polymerase, can be used to achieve a lower basal expression (Fernandez and Hoeffler, 1999).

The expression of the gene of interest is directed by the T7 RNA polymerase, with the gene coding for the T7 polymerase expressed from *E. coli's* chromosome under the control of the *Lac* promoter. When the inducer, either lactose or IPTG, is present in the media, the T7 RNA polymerase is expressed and recognises the T7 promoter present in the plasmid, allowing the transcription of the gene of interest (Fernandez and Hoeffler, 1999). Another possible alternative is to place the gene coding for the T7 polymerase under the control of the *araBAD* promoter, which allows a tighter regulation of the T7 expression system (**Figure 1.5**) (Wycuff and Matthews, 2000).

Arabinose expression system

E. coli can use other carbon sources apart from glucose or glycerol, such as arabinose. Expression of the enzymes required to metabolise arabinose into a suitable sugar form, which can be used in the pentose phosphate metabolic pathway, is positively regulated by the AraC protein. AraC is a transcription factor, which acts inducing the synthesis of enzymes required for the arabinose uptake (*araE*, *araF* and *araG*) and catabolism (*araB*, *araA* and *araD*). When arabinose is not present in the system, AraC binds simultaneously to two DNA half sites, O_1 and O_2, located 210 bp apart on the *araBAD* promoter (**Figure 1.6**). This binding is formed through a DNA loop which restricts the access of the RNA polymerase to the P_c and P_{BAD} promoters. When arabinose is present in the medium, arabinose binds AraC, inducing a conformational change which allows AraC to bind to the adjacent half-sites O_1 and I. Since AraC no longer constitutes a loop between the two DNA sequences, the RNA polymerase is able to transcribe the genes located downstream the promoter (Schleif, 2002; Sørensen and Mortensen, 2005).

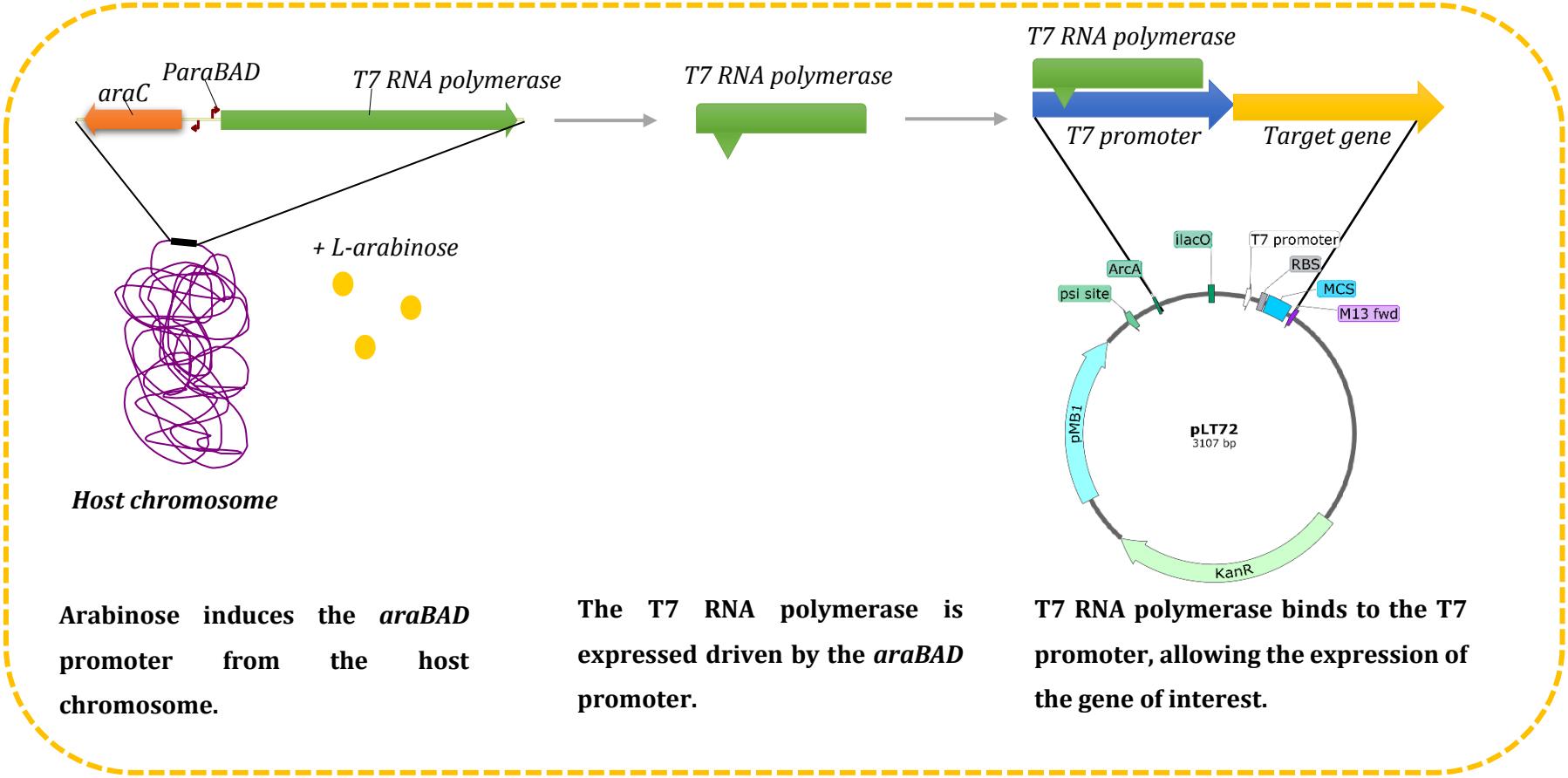


Figure 1.5. The mechanism of action of arabinose-inducible T7 expression system. In the T7 expression system, the addition of arabinose allows the expression of the T7 RNA polymerase from the host chromosome. The T7 RNA polymerase recognises the T7 promoter present on the expression vector, driving the expression of the gene of interest (Adapted from Sørensen and Mortensen, 2005).

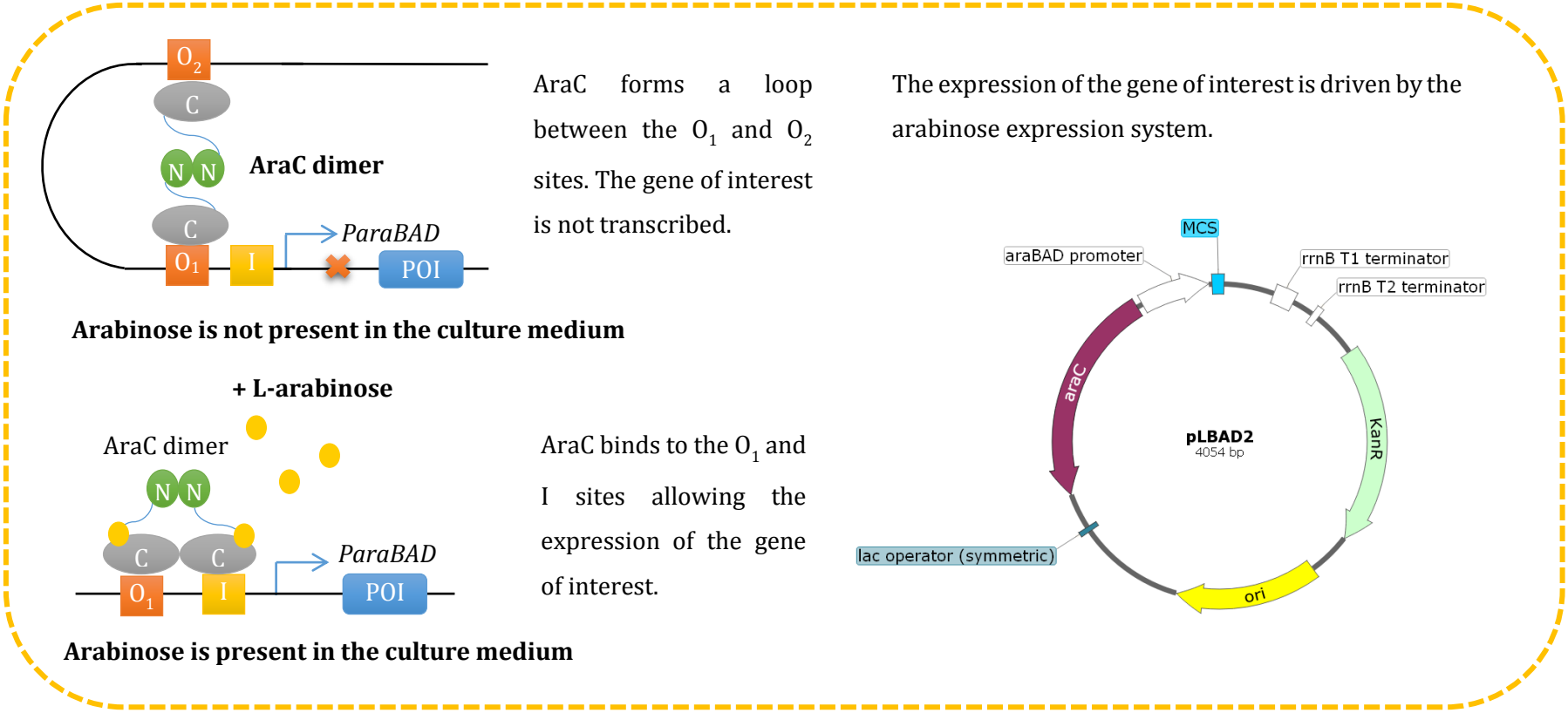


Figure 1.6. The mechanism of action of arabinose expression system. The arabinose operon is regulated positively by AraC. When arabinose is not present in the culture medium, AraC forms a loop between the O_1 and O_2 sites, restricting the access of the host RNA polymerase. When arabinose is added to the culture medium, AraC binds to the O_1 and I sites, allowing the transcription of the gene of interest (Adapted from Sørensen and Mortensen, 2005).

AraC also regulates the transcription of the arabinose transporters (*araE* and *araFGH*) which results in an all-or-nothing response under induction conditions. A more homogenous cell induction can be achieved in strains deficient in arabinose transport genes, where arabinose uptake is catalysed by arabinose independent transporters supplied from a constitutive promoter or an arabinose-independent inducible promoter (Sørensen and Mortensen, 2005; Khlebnikov *et al.,* 2000).

Guzman *et al.* (1995) extensively described the arabinose expression system, well known for its tight regulation. Low basal expression can be achieved with the *araBAD* promoter, highly desirable when "toxic proteins" need to be produced (Terpe, 2006).

Other expression systems

lac-derived expression systems are the most widely used by the biotechnology industry, however, a broad range of different expression systems are nowadays available for RPP (Terpe, 2006). The selection of the expression system will depend on the characteristics of the recombinant protein, expression level, intracellular and extracellular protein production level and cell growth characteristics. A detailed list of commonly used promoters for RPP can be found in **Table 1.3**.

λ pR/pL is a thermoregulated expression system derived from bacteriophage lambda (λ), commonly used for the production of protein therapeutics. The gene of interest is commonly inserted between the major leftward (pL) and/or the rightward (pR) promoters, which in nature controls the early lytic genes of the λ phage. The pR/pL promoters are induced shifting the culture temperature from 30 °C to 42 °C, as a result of the presence of a temperature-sensitive repressor cI857, a mutant cI repressor which is highly unstable at high temperatures (Valdez-Cruz *et al.*, 2010). Other expression systems such as the tetracycline (*tet*) system allows a high level of

Table 1.3. Commonly used expression systems for protein production in *E. coli* (Adapted from Terpe, 2006).

Expression system	**Inducer and concentration**	**Level of expression**	**Key features**	**Original reference**
Lac promoter	IPTG (0.05-2 mM)	Low to medium	Weak promoter Titrable High basal level expression	Gronenborn, 1976
Trc and tac promoters	IPTG (0.05-2 mM)	Moderate to high	Titrable High basal level expression	Brosius *et al*., 1985
T7 promoter	IPTG (0.05-2 mM)	Very high	Utilizes T7 RNA polymerase Very high level expression High basal level expression	Studier and Moffatt, 1986
ParaBAD promoter	L-arabinose (0.001-1.0%)	Variable from low to high	Tight regulation Low basal level expression Inexpensive inducer	Guzman *et al.,* 1995
PrhaBAD promoter	L-rhamnose (0.001-1.0%)	Variable from low to high	Tight regulation Low basal level expression Relative expensive inducer	Haldimann *et al*., 1998
TetA promoter	Anhydrotetracycline (200 μg · L^{-1})	Variable from medium to high	Tight regulation Independent of metabolic state	Skerra, 1994
PhoA promoter	Phosphate depletion	Moderately high	Media limitation Not titrable	Miyake *et al.*, 1985
$P_R P_L$ *promoter*	Temperature shift (30 °C → 42 °C)	Moderately high	High basal level by temperatures below 30 °C	Elvin *et al.,* 1990

expression when anhydrotetracycline (aTc) is added to the culture medium (Terpe, 2006). No specific strains are required (pASK75 vector), and the *tet* promoter is tightly regulated and not dependent of the metabolic state or the *E. coli* strain used (Samuelson, 2011). As with the arabinose expression system, the rhamnose (PrhaBAD) system is also known to be a tightly regulated system (Egan and Schleif, 1993).

Other promoters, such as the alkaline phosphatase promoter (PhoA), are active when a nutrient (phosphate) is depleted from the culture medium (Huang et *al.*, 2012). Also, as the T7 expression system has shown a high popularity as the first choice for RPP, other phage-derived expression systems, like the T5 (Qiagen), have been developed. In contrast with T7, the T5 promoter is recognised by the host RNA polymerase (Samuelson, 2011).

1.3 Routes of production of recombinant proteins in *E. coli*

E. coli cells, like other Gram-negative bacteria, are constituted of an inner and outer membrane, which divides the organism in three compartments: the cytoplasm, the periplasm and the extracellular space. Recombinant proteins can be produced targeting any of these compartments. Advantages and disadvantages of targeting the production of recombinant proteins to each cellular compartment can be found listed in **Table 1.4.**

1.3.1 Cytoplasmic recombinant protein production

Accumulation of recombinant protein in the soluble fraction

The production of recombinant proteins in the cytoplasm is often the most common choice, as successful production often leads to high titres of correctly folded (and likely to be active) protein products. Nevertheless, the production of proteins in a soluble form will depends on the host cell's capability and the total metabolic load imposed on the cells during RPP (Shadev *et al.,* 2008).

Table 1.4. Advantages and disadvantages of the production of recombinant proteins in the cytoplasm, periplasm and extracellular environment (Adapted from Choi and Lee, 2004).

Route of production	***Advantages***	***Disadvantages***
Cytoplasm (soluble)	High protein yields Soluble proteins likely to be active Solubilisation agents not required	High host protein concentration Proteolysis Limited disulphide bond formation
Cytoplasm (insoluble)	Low host protein concentration High purity of IBs Easy recovery and purification	Insoluble proteins less likely to be active Complex refolding Contamination with truncated proteins
Periplasm	High purity Disulphide bond formation True N-terminal (absence of methionine)	Protein accumulation in periplasmic IBs Lower yields Overload translocation pathways
Extracellular	Easier purification Disulphide bond formation Cell breakage not required	Leaky strains may be required Overload of export pathways Very dilute protein samples

High-level protein production often leads to the accumulation of misfolded or unfolded proteins in inclusion bodies (IBs). Several expression strategies have been developed to avoid this outcome, as the co-expression of chaperones, the reduction of the protein synthesis rate by using low cultivation temperatures and/or low inducer concentration, or the use of a highly soluble fusion partner (Fahner *et al.,* 2004). However, there is no general rule for the selection of an optimal strategy and different conditions need to be evaluated in a trial-and-error approach. In addition, the success of the different strategies seems to be protein-specific (Choi *et al*., 2006). A more detailed description of factors affecting protein folding during the production of recombinant proteins in the cytoplasm can be found in **Section 1.5.2**.

In addition, purification of recombinant proteins from the cytoplasm is a relatively difficult process, since this compartment contains the majority of cellular proteins (Baneyx and Mujacic, 2004). Downstream processes often require extensive development and optimisation, being often time-consuming and costly.

Accumulation of recombinant proteins in the insoluble fraction

As transcription and translation are fast and coupled process in *E. coli*, some eukaryotic proteins requiring longer time and/or folding chaperones assistance often fail to reach their native conformation, accumulating as misfolded proteins in the cytoplasm, where they tend to aggregate, forming inclusion bodies (Huang *et al.,* 2012). Inclusion bodies (IBs) are intracellular deposits of insoluble aggregates of unfolded or partially misfolded proteins (Alfasi, 2011; Gasser *et al.,* 2008), which can be founded in the cytoplasm or the periplasm depending on the presence of a secretory signal peptide (Baneyx and Mujacic, 2004).

Despite the fact that unfolded proteins accumulated in IBs can be biologically inactive, several commercial products, including interferon, interleukins and Fc-fusion proteins have been

produced and purified from IBs (Huang *et al.,* 2012). Under recombinant protein production conditions, inclusion bodies are mainly constituted by misfolded protein, accounting for 80% - 95% of the inclusion body content. Chaperones, as DnaK and GroEL, and co-chaperones, as DnaJ-GrpE and GroES, are also accumulated with the unfolded protein (Valdez-Cruz *et al.*, 2010).

Recombinant proteins accumulated in the inclusion bodies can be easy to recover and purify, since IBs can be easily isolated after cell disruption due to their high density. In addition, as a result of the host proteins concentration being lower when compared to the cytoplasmic soluble fraction, they are highly pure. Refolding of IBs to active proteins is often complex, requiring considerable optimisation for each target protein. In addition, chaotropic agents used during protein refolding can affect the integrity of refolded proteins (Huang *et al.,* 2012). In general, protein refolding can be unsuccessful and associated with high yield losses during processing. However, when successfully optimised, recombinant protein production in IBs and protein refolding has been proven a suitable production route for protein therapeutics.

Since most recombinant proteins for pharmaceutical purposes are human originated, they often lack the N-terminal methionine required for cytoplasmic production in *E. coli*, which can be a problem for recombinant proteins intended for human therapeutics (Choi *et al.,* 2006). When this is required, periplasmic protein production is the optimal choice, since upon translocation to the periplasm, the cleavage of the signal peptide generates an authentic N-terminus.

1.3.2 Periplasmic recombinant protein production

1.3.2.1 Structure of the signal peptide

Periplasmic and outer membrane proteins are initially synthesised in the cytoplasm as larger precursor proteins containing an amino-terminal signal peptide consisting of 20 to 40 amino acids, which following translocation to the periplasmic space is cleaved. Signal peptides are highly

conserved sequences containing three well-differentiated regions: an amino-terminus, a hydrophobic domain and the cleavage site (Choi and Lee, 2004; Oliver, 1985). SEC, SRP and TAT signal peptides share the same basic structure which can be found depicted in **Figure 1.7**.

The basic amino-terminus (N-region)

The basic amino-terminus consists of a stretch of two to ten amino acids, positively charged, which enables the precursor protein to insert itself into the inner membrane. A basic amino acid residue is often found after the start codon, frequently a lysine, coded by the codon AAA, which seems to favour the translocation of the precursor protein (Low *et al.,* 2013).

The hydrophobic core (H-region)

The hydrophobic core region has a variable length, being composed of 10 to 20 neutral hydrophobic amino acids residues, preferably leucine, alanine and valine, but phenylalanine, isoleucine, methionine and tyrosine can also be found at this region. Although it seems there is no strong preference for a specific amino acid to be located at a particular position, its distribution does not seem to be completely random (Choi and Lee, 2004; Oliver, 1985).

The processing site (C-region)

The processing site is a consensus sequence which contains the cleavage site recognised by the signal peptidase I. Alanine is the most frequent amino acid at positions -1 and -3, forming the so-called Ala-X-Ala box; however, other amino acid residues can be found at these positions (Choi and Lee, 2004). Amino acid residues located at -1 must have a small neutral side-chain, such as alanine, serine or glycine; and at position -3, includes any of the previous amino acids as well as

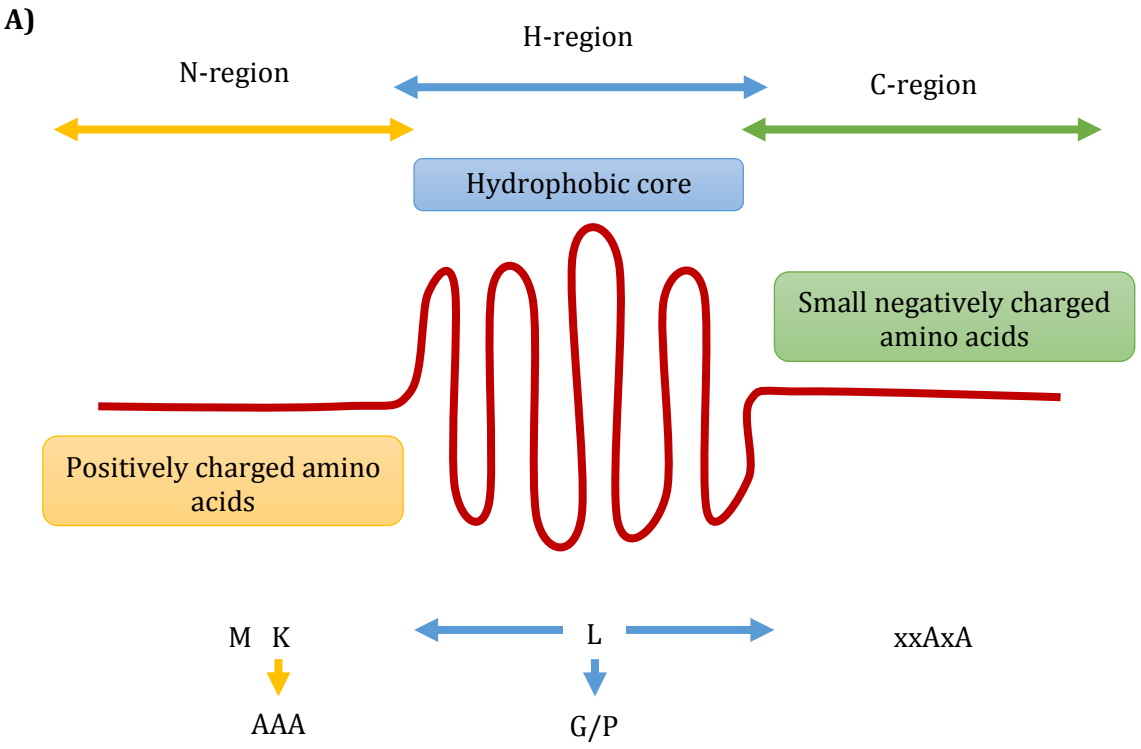


Figure 1.7. Structure of the signal peptide. Signal peptide sequences are constituted by three differentiated regions: a basic amino-terminus or N-region; the hydrophobic core or H-region; and the cleavage site or C-region. The N-region is constituted by positively charged amino acid residues. A lysine residue, coded by codon AAA, is commonly found at the second position of the amino acid sequence. The H-region is constituted by hydrophobic amino acids, preferably leucine, alanine and valine. A helix-breaker residue, such as glycine or proline, can often be found in this region. The C-region is constituted by small negatively charged amino acid residues. Alanine is the most frequent residue at the cleavage site, forming the so-called Ala-X-Ala box (A). Comparison of the structure of signal peptides targeting the SEC, SRP and TAT translocation pathways (B) (Adapted from Low *et al.,* 2013; Palmer and Berks, 2012).

leucine, isoleucine and valine. Any amino acid residue seems to be tolerated at position -2 and -4 (Oliver, 1985). The secondary structure at the processing site of precursor protein also plays a key role determining the exact cleavage site (Choi and Lee, 2004).

In addition, the export initiation domain, also known as the pro-region, is a stretch of approximately up to 30 amino acids long located downstream the processing site, which may influence the ejection of the precursor protein through the inner membrane. Mutations in this region may hamper or reduce protein export, altering the cleavage site, since a net negative charge in this region is required for efficient protein translocation (Low *et al.,* 2013).

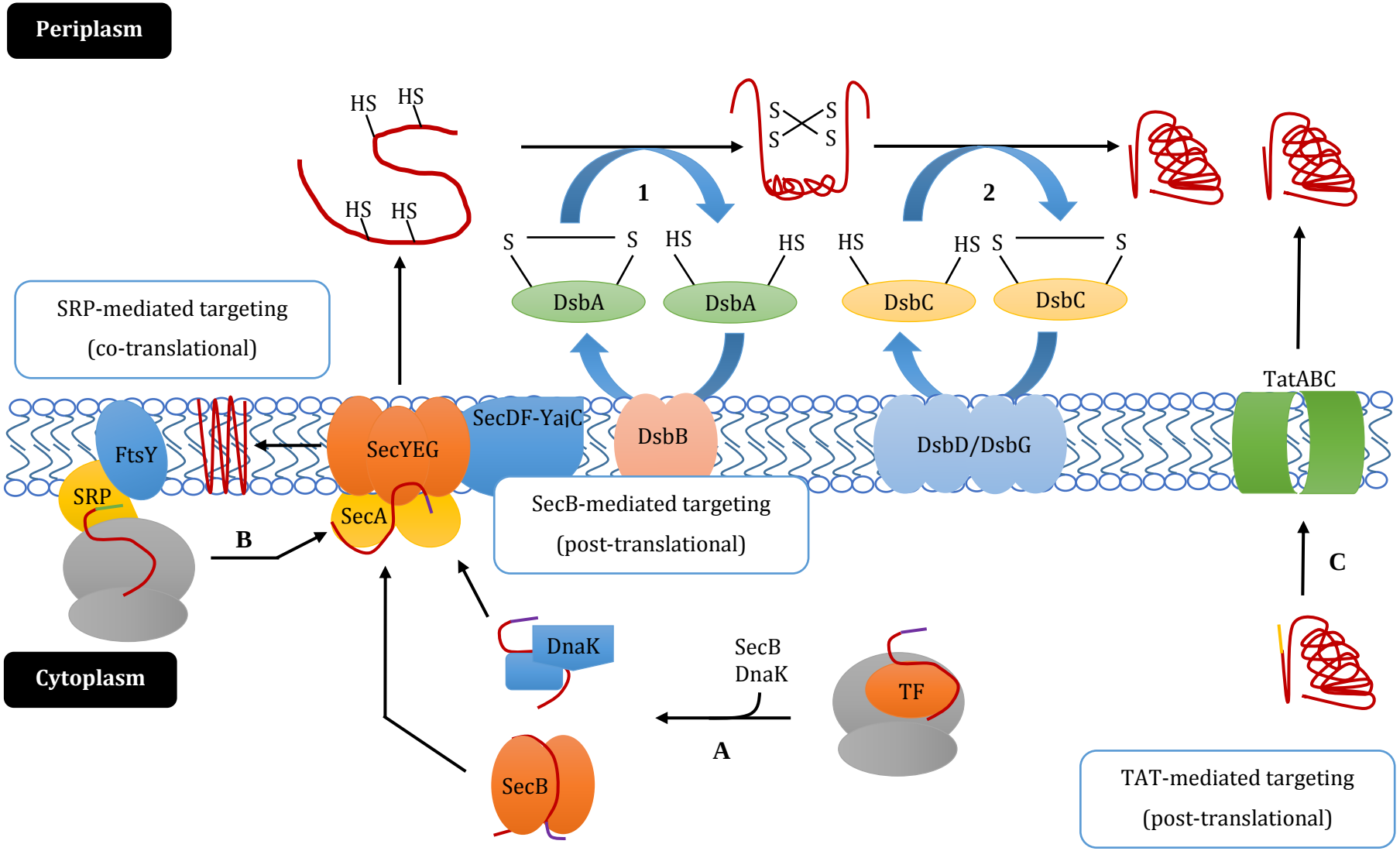
1.3.2.2 Periplasmic translocation pathways: SEC, SRP and TAT

Periplasmic translocation is a two-stage process required for the transport of proteins to the periplasmic space, the outer membrane and the extracellular environment. There are three main pathways used for protein translocation in Gram-negative bacteria: the SecB-dependent, the SRP-mediated and the twin-arginine translocation pathways, which can be found depicted in **Figure 1.8** (Mergulhao *et al.*, 2005).

SecB-dependent translocation pathway

In *E. coli*, 50% of the cellular proteins are exported to the periplasmic space, the outer membrane or to the extracellular environment, with more than 90% being translocated through the SEC translocation pathway (Low *et al.*, 2013). The translocation of a precursor proteins involves a cascade of interactions initiated from translation of the mRNA by the ribosome to the translocation of the protein to the periplasm by the SEC machinery (Mergulhao *et al.*, 2005; Low *et al.*, 2013).

Figure 1.8. Periplasmic protein translocation pathways. SecB-dependent translocation pathway (A). The majority of the protein intended to be translocated to the periplasm is directed to the SecB-dependent pathway. The trigger factor (TF) binds to nascent chains as they emerge from the ribosome. Subsequently, polypeptides are transferred to DnaK or SecB, which prevents its premature folding. SecB delivers the precursor protein to SecA, which initiates its translocation through the SecYEG translocon complex. SRP-mediated pathway (B). The majority of the proteins with highly hydrophobic signal peptides (green) or transmembrane segments are recognised by SRP, which delivers the nascent chains to FtsY and to the SecYEG complex. TAT translocation pathway (C). Fully folded or partially folded precursor proteins with signal peptides bearing the twin-arginine motif (yellow) are translocated by the TAT pathway, through the TatABC complex. DSB oxidative pathway (1). Cysteine residues present in disulphide-bonded proteins are introduced by DsbA, while DsbB helps to main DsbA in an oxidised form. DSB isomerization pathway (2). DsbC catalyses the isomerisation of incorrectly formed disulphide bonds. DsbC is maintained in a reduced form by DsbD (Adapted from Baneyx and Mujacic, 2004).


Periplasm
HS HS
HS HS
S S
S S
1
2
S S
S S
HS HS
HS HS
DsbA
DsbA
DsbC
DsbC
SRP-mediated targeting
(co-translational)
TatABC
FtsY
SRP
SecYEG
SecDF-YajC
DsbB
DsbD/DsbG
SecA
SecB-mediated targeting
(post-translational)
B
C
DnaK
SecB
DnaK
TF
A
SecB
Cytoplasm
TAT-mediated targeting
(post-translational)

During protein translocation, SecB binds the precursor protein or preprotein, preventing its premature folding (**Figure 1.8 A**). Unlike other chaperones, SecB binding does not require ATP hydrolysis, maintaining the precursor protein in a translocation-competent state (Danese and Silhavy, 1998). Once SecB is bound to a precursor protein, it will bind SecA, releasing it from SecB and transferring the preprotein to SecA. SecA targets the precursor protein to the membrane-bound translocation apparatus, the SecYEG complex (Sec61αγβ complex in eukaryotes) (Papanikou, 2007). SecA binds the SecYEG complex, and acting as an ATPase, causes the release of SecB from the membrane. SecE and SecY form a conducting channel, where SecG stimulates the translocation activity. Binding of the precursor protein to the membrane-bound SecA, results in the translocation of approximately 20 amino acids. Subsequently after the binding of ATP to SecA, another 15 to 20 amino acids are translocated. After the hydrolysis of ATP, the precursor protein is transferred from SecA into the translocation channel; SecA is then released from the membrane and exchanged for the cytosolic SecA. After several rounds of insertion and release of the SecA from the precursor protein, it is then translocated through the channel, process that can be facilitated by the proton motive force (Mergulhao *et al.*, 2005). Once the signal peptidase I cleaves the signal peptide, the now mature protein is released from the membrane by SecD (Low *et al.*, 2013).

Although the SEC translocation pathway has been extensively used for RPP, it is not able to translocate fully folded proteins. Therefore, proteins that fold rapidly in the cytoplasm, are often produced targeting the SRP or the TAT translocation pathways (Mergulhao *et al.*, 2005).

SRP-mediated translocation pathway

The signal recognition pathway targets primarily membrane proteins which directs, co-translationally, to the SecYEG translocon (Low *et al.*, 2013). This translocation pathway is also present in eukaryotic cells, where the signal recognition particle (SRP), a complex of six proteins

and 7S RNA mediates protein translocation across the endoplasmic reticulum (ER). In eukaryotes, SRP docks the precursor protein to the ER membrane by interaction with SRα, a subunit of the membrane-associated receptor for SRP, which favours the co-translational translocation of proteins (Danese and Silhavy, 1998).

In prokaryotes, the signal recognition particle (SRP), composed by Ffh protein and the 4.5S RNA subunit, homologous to two eukaryotic SRP constituents, the SRP54 and 7S RNA, binds to the nascent preprotein (Danese and Silhavy, 1998). Depending on the hydrophobicity of the signal peptide and the characteristics of the precursor protein, such as the presence of a transmembrane segments, SRP will bind to the nascent chain with high affinity, which interacts with the FtsY receptor, homologous to the eukaryotic SRα, displacing the trigger factor (TF) and initiating the translocation processes (**Figure 1.8 B**) (Luirinka and Sinning, 2004; Mergulhao *et al.,* 2005). The signal recognition particle shares a common attachment site and the binding of one or the other will determine whether the protein will be translocated co-translationally through the SRP-mediated pathway, or post-translationally by the SecB-dependent pathway. FtsY interacts with the protein-SRP complex and, in combination with Ffh, release the nascent chain to the SecYEG translocon (Baneyx and Mujacic, 2004).

For RPP, the SRP translocation pathway is targeted particularly for proteins that fold too quickly in the cytoplasm, as nascent protein chains are bound to SRP as they exit the ribosome tunnel, avoiding premature folding or protein aggregation (Lee and Jeong, 2013). The use of the SRP-mediated translocation pathway for phage display has been reported (Steiner *et al.,* 2006).

TAT pathway

The twin-arginine translocation, also known as TAT pathway, receives its name due to the presence of two consecutive and highly conserved arginine residues (RR) present in signal

peptides targeting this pathway. In contrast with SEC and SRP-dependent pathways, the TAT pathway mediates the translocation of fully folded or partially folded proteins by using the proton motive force (PTM). The average size of TAT signal peptides is 37 amino acids, containing an extended polar region with the consensus motif S/T-R-R-X-F-L-K and generally are less hydrophobic than SEC signal peptides (**Figure 1.7 B**) (Patel *et al.,* 2014; Palmer and Berks, 2012).

The main components of TAT pathway are the proteins TatA, TatB, TatC, TatD and TatE, however, their specific functions have not yet been completely elucidated (**Figure 1.8 C**). It has been proposed the involvement of TatA constituting a channel and TatB and TatC providing the initial binding site for the precursor protein, constituting the TatABC complex. Briefly, the twin-arginine motif of the signal peptide is recognised by TatC, the folded protein is then translocated through TatA. Once the precursor protein reaches the other side of the membrane, the signal peptide is cleaved, and TatA dissociates from the TatBC complex (Palmer and Berks, 2012; Mergulhao *et al.*, 2005).

The TAT pathway has been considered less efficient and slower than the SecB-dependent pathway, as it can be easily saturated. In consequence, the co-expression of the TatABC operon is required when this pathway is being targeted for RPP (Mergulhao *et al.*, 2005).

1.3.2.3 Disulphide bond formation

Disulphide bond formation is limited in the reductive environment of the cytoplasm, and disulphide bonded proteins are often targeted to the periplasm, which allows the formation of disulphide bonds due to its oxidising nature and the presence of the disulphide bond (DSB) generation machinery (Choi and Lee, 2004). The presence of disulphide bonds is a common post-translational modification which requires the participation of two cysteine residues, often crucial for the stability and activity of numerous proteins, and its absence may induce degradation and

proteolysis (Baneyx and Mujacic, 2004). The DSB machinery is involved in the formation of disulphide bonds by either introducing disulphide bonds in newly synthesised proteins, also known as the DSB oxidative pathway, or by rearranging the existing disulphide bonds, known as the DSB isomerization pathway (Heras *et al.,* 2009).

DSB oxidative pathway

The DSB oxidative pathway is responsible for the introduction of disulphide bonds into newly synthesised proteins that have been recently translocated to the periplasm. It involves two proteins: DsbA, a highly oxidising protein, and DsbB, which helps to maintain DsbA in an oxidised form. DsbA comprises a thioredoxin (TRX) domain which contains a pair of redox-active cysteines as part of the CXXC motif (**Figure 1.8 1**). During the formation of disulphide bonds, the two cysteines presents in the CXXC motif of DsbA become reduced, catalysing the oxidation of the cysteines present in the substrate protein. Subsequently, DsbA is regenerated to its original oxidised form by DsbB, ready to catalyse the formation of disulphide bonds in another protein (Heras *et al.,* 2009).

DSB isomerization pathway

The DSB oxidative pathway introduces disulphides bonds indiscriminately in any protein substrate containing more than two cysteine residues, therefore, a proofreading system that can correct the formation on non-native disulphide bonds is needed (**Figure 1.8 2**). The proofreading and shuffling activity is carried out by DsbC, a disulphide isomerase that catalyses the isomerization of disulphide bonds when incorrectly introduced by DsbA. As DsbA, DsbC contains a TRX domain with a CXXC motif, however this domain is linked to DsbG by amino-terminal dimerization, creating a hydrophobic cavity. This hydrophobic cavity allows the binding of

substrate proteins, and prevents the interaction with DsbB, separating the oxidative and isomerization pathways. DsbC and DsbG are maintained in its reduced form by DsbD, an integral membrane protein (Heras *et al.,* 2009).

1.3.2.4 Periplasmic recombinant protein production

For the production of recombinant proteins containing disulphide bonds, targeting the periplasmic space is often the preferred option (Choi and Lee, 2004). The three translocation pathways can be used for RPP and the selection among them is generally driven by the characteristics of target protein. Whereas the SEC translocation pathway only allows the translocation of unfolded proteins, TAT allows the transport of fully folded protein products. As the SRP pathway allows co-translational protein translocation, is generally being targeted for proteins that fold too quickly, hampering its translocation through SEC (Mergulhao *et al.*, 2005).

An advantage of periplasmic protein production is the lower concentration of host proteins in the periplasm when compared to the cytoplasm, which often allows an easier purification of the protein product. However, the limited space of the periplasm does not allow the accumulation of high amounts of recombinant protein, and yields are usually lower when compared to cytoplasmic protein production (Jonasson *et al.*, 2002). Leakage to the extracellular medium can be observed when the protein is accumulated to high titres, caused by the osmotic pressure build-up (Mergulhao *et al.*, 2005).

Nevertheless, several bottlenecks are often faced when the periplasmic space is targeted for RPP: the secretion efficiency is protein-dependent which can lead to low amounts of recombinant protein being secreted to the periplasm; misfolded proteins can be accumulated in IBs within the periplasm; and translocation of the protein to the periplasm does not always guarantee the correct formation of disulphide bonds. The selection of the optimal signal peptide and the overexpression

of chaperones may help to address some of the problems; however, different methods need to be evaluated (Choi and Lee, 2004).

1.3.3 Extracellular recombinant protein production

The secretion of proteins to the culture medium has been highlighted as an alternative route to reduce the cell stress induced during RPP, as recombinant proteins are no longer accumulated inside the cell. Six secretion mechanisms are known to mediate protein secretion to the extracellular environment in Gram-negative bacteria, recombinant proteins being generally secreted by the type I or by a second step of the type II secretion system (Ni and Chen, 2009).

The type I secretion system is the most frequently used for RPP, being capable of transporting proteins up to 800 kDa across the outer membrane (e.g. haemolysis (Hly) system). In contrast with the type I secretion system, the type II is a two-step process, which initially involves the translocation of the protein to the periplasm by SEC, SRP or TAT pathways, where it folds. In the second step, the protein is transported across the outer membrane. The use of other secretion systems, such as type III and V, have been reported but have not been widely used for RPP (Ni and Chen, 2009).

Recombinant proteins can also be released to the extracellular environment by non-specific periplasmic leakage. Periplasmic leakage can be induced from the osmotic pressure build-up, as a result of the accumulation of the recombinant protein, which is the driving force for transport across the outer membrane. RPP can induce alterations in the outer membrane leading to an increase the permeability, which may result in the release of the periplasmic content (Mergulhao *et al.*, 2005).

An increase in the permeability of the cell outer membrane can also be achieved by biochemical methods, such as the addition of detergents as Triton-X, or physical methods like osmotic shock

or sonication. Addition of some amino acids, such as glycine, to the culture medium also seem to favour the leakage of proteins to the extracellular environment (Chou and Lee, 2004).

Several disadvantages are often associated with the extracellular secretion of recombinant proteins, as protein leakage to the culture medium is often a consequence of a reduction of the integrity of the outer membrane. Export system can be easily overloaded by the high protein production, mainly when strong promoters are used. In addition, the secretion of proteins to the culture medium results in highly diluted protein samples (Chou and Lee, 2004).

1.4 High-cell-density culture techniques for RPP in *E. coli*

Recombinant DNA technology made possible the large-scale production of recombinant proteins, with increased effectiveness, at lower cost by the development of high-cell-density culture (HCDC) techniques (Choi *et al.,* 2006). Once the conditions for the production of a target protein had been optimised, the process can be transferred to a high-cell-density culture aiming to increase cell density and maximise product yields. Fermentation techniques provide control over chemical, physical and biological parameters that can affect cell growth and recombinant protein production, including temperature, pH, dissolved oxygen (DO) and culture medium composition (Huang *et al.,* 2012).

Culture medium

The culture medium used to grow *E. coli* to high cell densities is constituted by essential components, as a carbon source, nitrogen source, essential salts and trace elements. Three types of media are available: chemically defined medium (CDM), semi-defined medium (SDM) and complex medium. CDM has a defined composition and concentration of chemicals, whereas SDM and complex medium compositions contain non-defined components, such as protein

hydrolysates or yeast extract. SDM is the most widely used in industry since non-defined components provide essential trace nutrients, allowing the achievement of high cell densities. However, CDM is gaining more acceptance due to inconsistent performance and lot-to-lot variation associated to undefined medium components (Huang *et al.,* 2012).

Type of fermentation

Batch fermentation, in which cells are grown to a maximum specific growth rate with the nutrients supplied in the fermenter, is an easy technique to grow cells to moderately high cell densities in a short period of time. However, batch fermentation requires a high concentration of sugar to be present in the bioreactor to support cell growth, resulting in the formation of side metabolites, such as acetate, and in less extent, lactate and formate as a consequence of the overflow metabolism, which affects negatively cell growth and the host metabolism. Fed-batch fermentation is more widely used for industrial manufacturing, allowing the culture of *E. coli* up to a concentration greater than 100 grams of dry cell weight (DCW) per liter by continuous supply of nutrients, usually the carbon source, to ensure control over cell growth (Lee, 1996).

A highly concentrated nutrient solution, also known as feed solution, is often provided after a sharp DO and pH increase, which indicates that the nutrients in the batch medium have been exhausted. Several feeding strategies have been developed: pH-stat, where feeding is activated when the pH of the medium increases; DO-stat, when the feed solution will be added after a sharp DO increase caused by substrate depletion; and exponential feeding profiles based on calculating for prediction of use of a specific nutrient, such as the carbon source (Huang *et al.*, 2012).

The selection of the feeding strategy is critical since it may affect metabolic pathway fluxes, limiting the maximum attainable cell concentration, recombinant protein production or the formation of by-products. To date, there is no general rule for the selection of a feeding strategy

but different feeding approaches, such as constant feeding rate or exponential feeding profiles, have proven to be successful for RPP (Choi *et al.,* 2006).

Fermentation scale-up

The development of a fermentation process usually starts in a laboratory scale bioreactor, e.g. 1 – 10 L, to optimise the growth and protein production conditions. Once optimal conditions have been achieved, the process is then transferred to pilot scale, e.g. 50 – 200 L, aiming to determine the optimal operating parameters, before transfer to manufacturing scale, above 500 L. The scale-up process aims to achieve a high productivity and process consistency, nevertheless, factors affecting growth and protein expression may change during scale-up (Huang *et al.,* 2012). At large scale, poor mixing efficiency, oxygen limitation or heat generation, may lead to imbalanced nutrient distribution (Choi *et al.,* 2006; Enfors *et al.*, 2001).

Poor mixing efficiency can lead to limited dissolved oxygen and zonal nutrient distribution, as cells close to the feeding port are exposed to a high nutrient concentration whereas at other locations, cells may be starved of substrate, affecting cell growth and leading to amino acid misincorporation during RPP. Oxygen limitation can lead to the production of organic acids resulted from anaerobic reactions and the overflow metabolism, which decreases the pH and induces growth arrest, resulting in a low specific growth rate (μ) and low biomass and product yields (Overton, 2014; Enfors, 2001; Lee, 1996). In addition, cultures where there is a limitation of dissolved oxygen may result in low plasmid retention, even under selective pressure, resulting in poor protein yields. At large scale, when mixing efficiency is not optimal, the use of oxygen-enriched air or pure oxygen may help to provide sufficient oxygenation to the culture, otherwise, the overgrowth of plasmid-free cells may eventually dominate the culture (Glick, 1995). It is important to understand the limitations of large-scale fermentations, which should be kept in mind during the development of a fermentation process at laboratory scale (Choi *et al.,* 2006).

1.5 Common problems associated with recombinant protein production

Fermentation scientists often have to face one if not more of the general problems associated with RPP: misfolded or truncated protein; no protein production; growth arrest; or cell death. The accumulation of misfolded or truncated proteins and growth arrest are often the result of high level of recombinant protein synthesis, which can be controlled at induction level or by reducing the cultivation temperature to 18 - 25 °C to slow down the protein synthesis rate (Fahner *et al.,* 2004). Common problems encountered during RPP can be found described in **Table 1.5.**

1.5.1 Plasmid maintenance

Optimal gene transcription depends largely on the copy number, the structural and segregational plasmid stability. A plasmid is considered segregationally stable when all daughter cells carry at least one copy of the plasmid. Structural plasmid stability is if all the plasmids generated have the correct base sequence. Structural and segregational plasmid stability depends largely on the copy number, which is determined by the origin of replication, but also can be severely affected by cell stress during RPP (Huang *et al.,* 2012).

The introduction of foreign DNA in a host organism may impair the organism's normal metabolic functioning, as a result of the metabolic burden. The metabolic burden increases with plasmid size and copy number, the energy cost required for plasmid maintenance will be higher, being exacerbated when high copy number plasmids are used (Glick, 1995, Sørensen and Mortensen, 2005). A simple strategy is the use of low rather than high copy number vectors or the integration of the foreign DNA directly into the chromosome of the host strain. By chromosomal integration plasmid instability is no longer a problem and avoids the waste of cellular resources synthesising resistance marker proteins (Glick, 1995).

Table 1.5. Common problems and possible solutions to RPP (Adapted from Terpe, 2006).

Symptom	***Possible problem***	***Possible solutions***
Cell death or no colonies on agar plates	Too much recombinant protein expression Toxic protein High basal expression Cellular resources diverted from biomass to RPP	Weaker promoter Higher control over basal expression Tightly controlled promoter Lowering temperature Lowering inducer concentration
Insoluble disulphide-bonded protein (inclusion bodies)	Reduction of disulphide bonds	Minimise reduction in the cytoplasm Periplasmic protein production
Insoluble proteins (inclusion bodies)	Too much expression leading to poor recombinant protein folding	Attenuate expression by: weaker promoter, lowering inducer concentration, lowering temperature, decrease plasmid copy number, fusion of a hydrophilic affinity tag
Recombinant protein has no activity	Misfolded protein Affinity tag can decrease activity	Minimise reduction in the cytoplasm Periplasmic protein production Attenuate expression Change affinity tag
No protein *Truncated protein*	*E. coli* codon usage (codon bias)	Supply rare tRNAs Codon optimisation Lower temperature Tightly controlled promoter

1.5.2 Protein folding

The production of misfolded proteins is one of the most common situations during RPP. Chaperones generally assist newly synthesised proteins to achieve the correct folding conformation. However, the use of strong promoters in combination with high inducer concentrations means that the number of nascent polypeptide chains easily exceeds the chaperone number, leading to the accumulation of misfolded or unfolded proteins in inclusion bodies (IBs). IB formation is often the result of unbalanced equilibrium between protein aggregation and solubilisation, which may result in the induction of the heat shock response (Baneyx and Mujacic, 2004).

The cytoplasmic space is an extremely unfavourable protein-folding environment where transcription and translation processes are coupled and a protein chain is released from the ribosome every 35 seconds. In the macromolecular crowding of the cytoplasm, where protein concentration can reach 200 - 300 mg · mL^{-1}, small and single domain protein can easily fold, but large or multidomain proteins often require the assistance of molecular chaperones (Baneyx and Mujacic, 2004; Sørensen and Mortensen, 2005).

Molecular chaperones are proteins whose main activity is to assist protein folding. The trigger factor (TF), is the first chaperone to be encountered by the nascent chain as it emerges from the ribosome, assisting protein folding (**Figure 1.9**). Other folding chaperones, such as DnaK or GroEL, mediate refolding and unfolding by ATP-dependent conformational change; thus, preventing aggregation and IB formation. Folding chaperones often work in association with holding chaperones, such as IbpA and IbpB, helping to stabilise partially folded proteins and protecting heat-denatured proteins from aggregation. The solubilisation of aggregated proteins is carried out by disaggregating chaperones, such as ClpB or HtpG (Shadev *et al.*, 2008; Huang *et al.,* 2012).

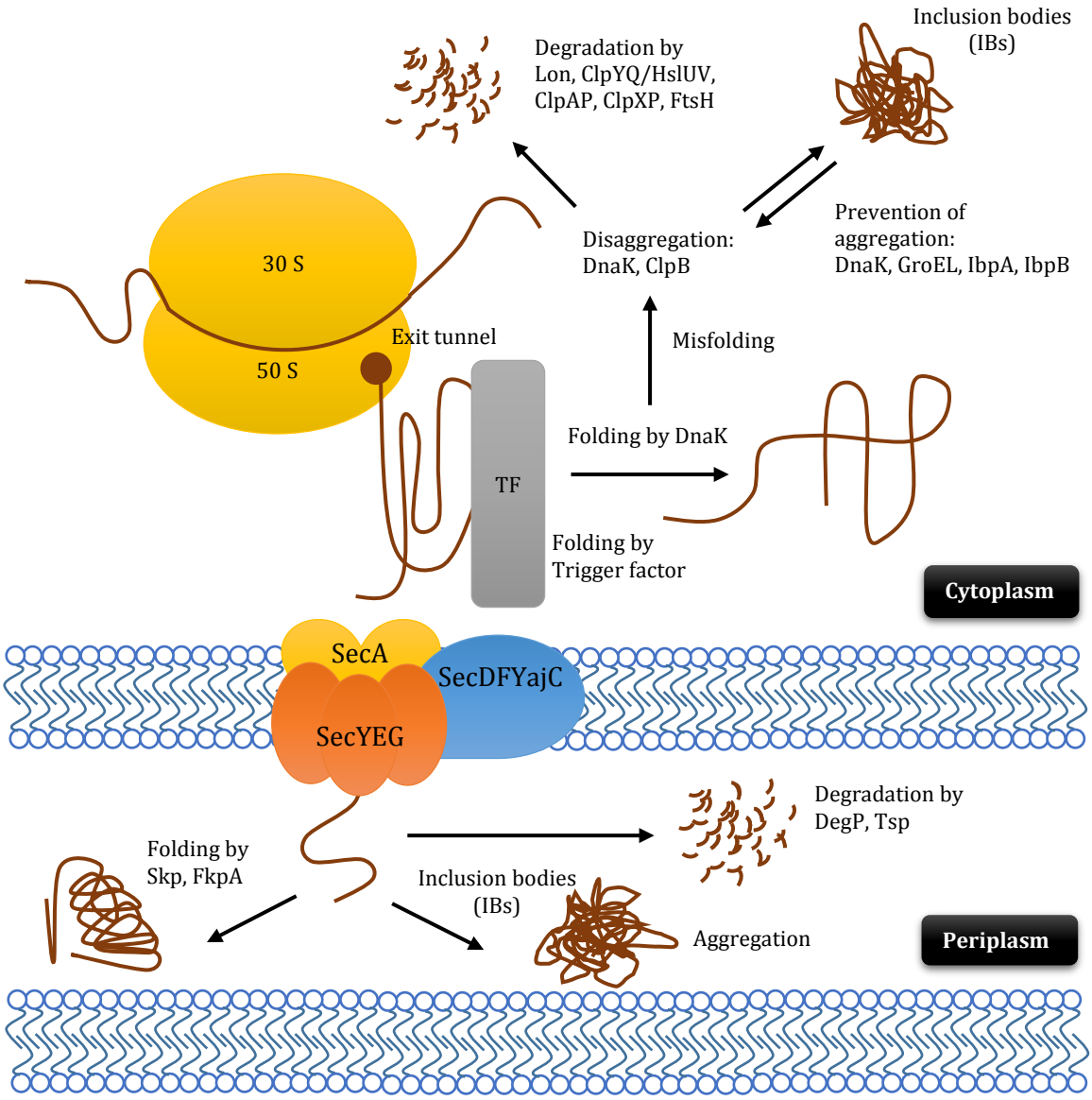


Figure 1.9. Protein folding and protease activity in *E. coli*. In the cytoplasm, proteins often associate to the trigger factor (TF) as soon as the newly synthesised protein exits the ribosomal tunnel, protecting the exposed hydrophobic patches from un-intended interactions. Incorrectly folded proteins are assisted by DnaK and GroEL, which reduce protein aggregation and promotes proteolysis of misfolded proteins. DnaK and ClpB also mediates the solubilisation or disaggregation of proteins. Disaggregation and IB formation is also prevented by GroEL, which operates protein transit between soluble and insoluble fractions. IbpA and IbpB are small heat shock proteins which protect heat-denatured proteins from irreversible aggregation. In the periplasm, partially folded proteins may aggregate, undergo proteolysis by DegP and Tsp or reach their native conformation assisted by Skp and FkpA (Adapted from Baneyx and Mujacic, 2004; Sørensen and Mortensen, 2005).

Despite all the progress made during the last years, the production of recombinant proteins frequently results in protein misfolding, low protein yields and laborious and time-consuming downstream process (Sørensen and Mortensen, 2005). Traditional approaches to improve folding include reducing the protein synthesis rate by using weaker promoters or decreasing the inducer concentration (Baneyx and Mujacic, 2004). The use of low cultivation temperatures is an alternative strategy, which has the advantages of slowing the protein synthesis and reducing the strength of hydrophobic interactions that contribute to protein misfolding (Baneyx and Mujacic, 2004). Increased chaperone expression has also been associated during growth at low temperatures. Also, proteases induced during RPP are less active at cultivation temperatures of 15 °C to 25 °C, leading to a significant reduction in proteolysis and degradation of expressed proteins (Shadev *et al.,* 2008). The co-expression of chaperones, such as GroEL/GroES, have proven to be effective facilitating folding of some recombinant proteins (Huang *et al.,* 2012).

1.5.3 Proteolysis

Proteolysis is a cellular mechanism used to minimise the accumulation of misfolded polypeptides within the cell, helping amino acid recycling (Baneyx and Mujacic, 2004). Misfolded, unfolded or truncated proteins are frequently unstable and prone to be targeted by proteolysis. Cell stress induced by high cultivation temperatures, protein overproduction or accumulation of misfolded proteins can increase proteolytic activity (Gottesman, 1996).

In the cytoplasm, five heat shock proteases (Lon, ClpYQ/HslUV, ClpAP, ClpXP and FtsH) initiate proteolytic degradation, degrading unfolded protein substrates via energy-dependent proteolysis. The process is completed by the action of peptidases that hydrolyse peptides of 2–5 residues in length. Lon protease plays a primary role in the degradation of many misfolded or unstable proteins. As the *lon* gene is under the control of a heat shock promoter, its synthesis can

be activated by the accumulation of unfolded proteins. Lon along with ClpYQ are considered to be primarily responsible for degradation of misfolded proteins (Baneyx and Mujacic, 2004).

The accumulation of misfolded proteins in the periplasm is often the result of high cultivation temperatures, oxidative stress or incorrect disulphide bond formation, DegP being the primary housekeeping protease in the periplasmic space **(Figure 1.9)**. DegP is regulated by the sigma factor E (σ^E) and its activity switches from protease to chaperone at low temperatures. A second protease, Tsp also exerts its activity in the periplasm (Baneyx and Mujacic, 2004).

Other common proteases located in the cell envelope are DegQ, DegS, Protease III and OmpT (Baneyx and Mujacic, 2004). OmpT is a protease found in the outer membrane, which often remains active after cell disruption and under denaturing conditions (Gottesman, 1996).

1.6 Reporter protein technology

Reporter proteins are defined as proteins with a specific activity that can be identified over a background of endogenous proteins (Naylor, 1999). Reporter proteins have been widely used as markers for cellular localisation, and are especially valuable in the identification of protein regions that promote membrane translocation. Novel approaches rely on the use of protein fusions as marker proteins whose enzymatic activity depends on their cellular localisation (Manoil *et al.,* 1990). Commonly used reporter proteins are detailed in **Table 1.7,** including β-galactosidase, as a cytoplasmic protein marker; and alkaline phosphatase and β-lactamase, as periplasmic markers.

β-galactosidase (LacZ)

β-galactosidase catalyses the cleavage of the disaccharide lactose to glucose and galactose, and was the first protein to be used as a reporter (Casadaban and Cohen, 1980). Its activity can be

Table 1.6. Commonly used reporter proteins (Adapted from Naylor, 1999).

Reporter protein	**Advantages**	**Disadvantages**
Chloramphenicol acetyltransferase	No endogenous activity in *E. coli*	Narrow linear range
β-galactosidase	Well characterised Colorimetric assays available	Endogenous activity *E. coli* Large molecular weight
Luciferase	High specific activity No endogenous activity *E. coli* Convenient assays	Requires substrate (luciferin), O_2 and ATP
Alkaline phosphatase	Periplasmic protein Colorimetric assays available	Endogenous activity *E. coli*
β-lactamase	No endogenous activity *E. coli* Periplasmic protein Colorimetric assays available	High molecular weight and poor cell permeability of fluorescent substrates
Fluorescent proteins	Autofluorescent No endogenous activity *E. coli* Mutants with altered spectral qualities available	Low sensitivity (no signal amplification)

quantified by using the colorimetric substrate O-nitrophenyl-galactopyranoside (OPNG) producing yellow coloured O-nitrophenol when cleaved by LacZ. β-galactosidase has been successfully expressed and used as a marker protein in both prokaryotes and eukaryotes (Vizcaino Caston, 2012).

β-galactosidase has been widely used for molecular biology, especially in gene cloning as part of blue/white screening. It has been also involved in numerous localisations studies as a fusion partner for the identification of components of the translocation machinery, as it is enzymatically active only when accumulated in the cytoplasm (Feilmeier *et al.,* 1998).

Alkaline phosphatase (PhoA)

In contrast to β-galactosidase, alkaline phosphatase is inactive in the cytoplasm. The lack of activity of the cytoplasmic alkaline phosphatase is the result of the lack of disulphide bonds; PhoA is active only when successfully translocated to periplasm. These characteristics made alkaline phosphatase the most common marker protein used to determine enzymatic or phenotypic activities of proteins depending on their subcellular location (Manoil *et al.,* 1990).

Fluorescent proteins (FPs)

During the last decade, cell biology imaging has been revolutionised by the discovery of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. GFP is generally expressed alone or fused to other proteins, resulting in visible fluorescence with the only requirement being O_2. Nonetheless, GFP is just one member of a large family of fluorescent proteins and mutagenesis of fluorescent proteins has expanded the possible choices, with diverse FP's in terms of spectra, brightness and folding (Giepmans *et al.*, 2006).

Nowadays, FPs are widely used as reporters for protein synthesis, as high-throughput screening tools or as sensors of physiochemical conditions. FPs are small and simple proteins which can be expressed in a large variety of host organisms, being easily detected using non-invasive techniques, such as microscopy or flow cytometry (Vizcaino-Caston *et al.,* 2012). GFP monitoring using flow cytometry has been used for the development of stress-minimising protocols for RPP, improving the yield of correctly folded recombinant proteins while avoiding IB formation (Sevastsyanovich *et al.,* 2009). However, solubility, correct folding and protein fluorescence are not always directly related, and it has been observed that inclusion bodies containing GFP fused to aggregation-prone proteins can still retain some fluorescence (Garcia-Fruitos *et al.,* 2005).

TEM-1 β-lactamase

The TEM-1 β-lactamase (Bla) is periplasmic enzyme with a molecular weight of 29 kDa, the product of the ampicillin resistance gene, which provides resistance to β-lactam antibiotics by hydrolysis of the β-lactam ring (Xing *et al.,* 2005).

β-lactamase has been widely used as a marker protein for the study of protein translocation. As alkaline phosphatase, β-lactamase displays a measurable and distinctive phenotype depending on subcellular localisation. When β-lactamase is localised in the periplasm, it acts hydrolysing β-lactam-derived antibiotics before they can inactivate penicillin-binding proteins (transpeptidases), which are essential for bacterial cell wall biosynthesis. In contrast, the cytoplasmic form of β-lactamase is unable to hydrolyse the antibiotic in the periplasm from its cytoplasmic localisation, and therefore, the host is sensitive to the antibiotic, resulting in cell lysis (Broome-Smith *et al.*, 1990).

β-lactamase has been widely used as marker protein in protein localisation studies, as fusion proteins provide a distinct phenotype, as colonies can be selected based on their resistance to

ampicillin (Broome-Smith *et al.,* 1990). The activity of β-lactamase can also be quantified by using colorimetric substrates, such as nitrocefin, CENTA or PADAC, the last one being no longer commercially available (Bebrone *et al.,* 2001). Fluorogenic substrates for imaging β-lactamase gene expression, such as CC1, CC2 and CR2 (CR2/AM) have also been developed (Gao *et al.,* 2003).

1.7 Scope of this project

This project is part of a 3-years research collaboration between the University of Birmingham and Cobra Biologics Ltd. The research project was funded by the Technology Strategy Board (TSB), now known as InnovateUK, the Biotechnology and Biological Sciences Research Council (BBSRC) and the Engineering and Physical Sciences Research Council (EPSRC) as part of the Knowledge Transfer Partnership (KTP) scheme. The aim of this project is the development of high productivity expression and fermentation processes for the production of recombinant proteins in *E. coli*, aiming to develop a knowledge-based approach, which can be applied to GMP production of recombinant proteins.

Initially, the work was focused in the optimisation of the conditions for the production of soluble recombinant proteins in the cytoplasm, using tumour necrosis factor α (TNFα) as a model protein. The aim was to optimise the fermentation conditions, including vector design, cultivation temperature, inducer concentration and induction point, to improve the production of soluble and active TNFα.

Subsequently, the work was expanded to the production of recombinant proteins targeting the periplasm of *E. coli*, using a single-chain antibody fragment, scFv163R4, as model protein. However, translocation of proteins into the periplasm is often inefficient due to the saturation of the translocation pathways, which can be a major bottleneck in periplasmic protein production.

Different cultivation conditions and signal peptides were evaluated with the aim to optimise the production of scFv163R4 in the periplasm.

The efficiency of protein translocation is highly dependent of the signal peptide used to direct the target protein to the periplasmic space. However, the selection of the optimal signal peptide is often challenging, and several signal peptides need to be evaluated in a trial and error approach. Aiming to expand the knowledge regarding periplasmic protein production, signal peptide libraries were generated by random mutagenesis and a high-throughput screening assay which allows the rapid evaluation of signal peptides was developed. The scFv163R4-A was used as a model protein fused to β-lactamase as a reporter protein. The resistance to β-lactam antibiotics was used as a selectable phenotype by directly screening for antibiotic resistance levels of *E. coli* colonies, and by the quantification of the β-lactamase activity of each individual clone. Signal peptides showing enhanced β-lactamase activity were evaluated by fed-batch fermentation to identify mutagenized signal peptides with improved translocation activity.

Chapter 2:

Material and methods

2.1 Materials

2.1.1 Suppliers

Chemical, reagents and media components were obtained from Sigma-Aldrich, Merck Millipore, Thermo Fisher Scientific, or Becton, Dickinson and Company (BD), unless stated otherwise.

2.1.2 Liquid growth media

Luria broth (LB) was prepared by dissolving 10 g · L^{-1} phytone peptone, 5 g · L^{-1} yeast extract and 5 g · L^{-1} NaCl in deionised water and autoclaving at 121 °C and 1.2 atmospheres (atm) for 20 minutes.

Terrific broth (TB) (Thermo Fisher Scientific) was prepared by dissolving 47 g · L^{-1} of premade terrific broth powder (11.8 g · L^{-1} SELECT peptone 140, 23.6 g · L^{-1} yeast extract, 9.4 g · L^{-1} K_2HPO_4 and 2.2 g · L^{-1} KH_2PO_4) and 4 mL · L^{-1} of glycerol in deionised water and sterilised by autoclaving.

2.1.3 Solid growth media

Luria Bertani agar was prepared by dissolving 10 g · L^{-1} phytone peptone, 5 g · L^{-1} yeast extract, 5 g · L^{-1} NaCl and 15 g · L^{-1} of extra pure agar in deionised water.

Mueller-Hinton agar (Sigma-Aldrich) was prepared by dissolving 38 g · L^{-1} of premade Mueller-Hinton agar powder (17 g · L^{-1} of agar, 2 g · L^{-1} of beef infusion solids, 17.5 g · L^{-1} of casein hydrolysate and 1.5 g · L^{-1} of starch) in deionised water.

All solid media was sterilised by autoclaving. Media were cooled to at least 50 °C before addition of antibiotics or other supplements. Agar plates were poured under sterile conditions, approximately 20 mL of agar per petri dish, and stored at 4 °C.

2.1.4 Antibiotics and other supplement solutions

Kanamycin stock solution (50 mg · mL^{-1} of kanamycin in 0.9% NaCl) from *Streptomyces kanamyceticus* was purchased from Sigma-Aldrich and stored at 4 °C. Ampicillin stock solution was prepared by dissolving 100 mg · mL^{-1} of ampicillin sodium salt (Sigma-Aldrich) in deionised water.

Media supplement stock solutions, such as glucose, was prepared by dissolving 200 g · L^{-1} of glucose in deionised water to generate a 20% glucose stock solution. 20% casamino acid stock solution (Sigma-Aldrich) was prepared by dissolving 200 g · L^{-1} of casamino acids in deionised water. Arabinose, used as inducer for arabinose-inducible expression systems, was prepared by diluting 200 g · L^{-1} of arabinose in deionised water to generate a 20% arabinose stock solution. Isopropyl β-D-1-thiogalactopyranoside (IPTG), used as inducer for lactose-inducible expression systems, was prepared by diluting IPTG in water (0.5 M IPTG stock solution). Lysozyme, used for periplasmic protein extraction, was prepared at final concentration of 2 mg · mL^{-1} in deionised water and stored at -20 °C.

All the solutions were sterilised by filtration through a sterile 0.2 μm filter.

Table 2.1 summarises the antibiotics and stock solutions used in the present work.

2.2 Buffers and solutions

2.2.1 Phosphate buffered saline (PBS)

Dulbecco's phosphate buffered saline (PBS) was prepared by dissolving 8 g · L^{-1} of NaCl, 0.2 g · L^{-1} of KCl, 2.16 g · L^{-1} of $Na_2HPO_4 \cdot 7\ H_2O$ and 0.2 g · L^{-1} of KH_2PO_4 in deionised water. The pH was adjusted to 7.2 – 7.4 by the addition of 1 M HCl and sterilised by autoclaving. PBS was used for serially dilute culture samples.

Table 2.1. Concentrations, solvents and storage conditions of antibiotics and stock solutions.

Compound	*Stock concentration*	*Working concentration*	*Solvent*	*Storage*
Ampicillin	100 mg · mL^{-1}	100 µg · mL^{-1}	Water	-20 °C
Kanamycin	50 mg · mL^{-1}	50 µg · mL^{-1}	0.9% NaCl	4 °C
Casamino acids	20% (w/v)	0.2%	Water	4 °C
Glucose	20% (w/v)	0.25% - 0.5%	Water	4 °C
Arabinose	20% (w/v)	0.002% - 0.2%	Water	4 °C
IPTG	0.5 M	1 mM	Water	-20 °C
Lysozyme	2 mg · mL^{-1}	-	Water	-20 °C

2.2.2 Buffers for DNA electrophoresis

The 50x TAE buffer stock was prepared by dissolving 242 g · L^{-1} of Tris Acetate, 18.61 g · L^{-1} of Na_2EDTA and 57.1 mL · L^{-1} of glacial acetic acid in deionised water. The pH should be approximately 8.2 - 8.4.

The electrophoresis running buffer was prepared by diluting the 50x TAE buffer stock in deionised water to a final concentration of 1x (v/v). The 10x sample buffer was prepared containing 0.025% (w/v) bromophenol blue, 10% (v/v) glycerol in 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

2.2.3 Buffers and solutions for SDS-PAGE

For 4-12% Bis-Tris NuPAGE SDS PAGE gels (Life Technologies), 1x (v/v) electrophoresis running buffer was prepared by diluting 20x NuPAGE MES SDS running buffer (Life Technologies) in deionised water. For reducing protein electrophoresis, 0.5 mL of NuPAGE antioxidant (Life Technologies) was added to 200 mL of running buffer and used to fill the inner chamber of the electrophoresis tank.

For 18% Tris-Glycine SDS-PAGE gels (Life Technologies), 1x (v/v) electrophoresis running buffer was prepared by diluting 10x Tris-Glycine SDS running buffer (Life Technologies) in deionised water. 0.5 mL of NuPAGE antioxidant was added to 200 mL of running buffer and used to fill the inner chamber of the electrophoresis tank.

Fixing solution was prepared by mixing 40% (v/v) methanol, 10% (v/v) glacial acetic acid with deionised water. The Colloidal Blue Staining kit (Life Technologies) was used for staining of SDS-PAGE. Staining solution was prepared by mixing 20% (v/v) methanol, 20% (v/v) staining solution A and 5% (v/v) staining solution B.

2.3 Bacterial strains and plasmids

Strains of *E. coli* and their genotypes used in the present work are listed in **Table 2.2**. ElectroSHOX competent cells (Bioline) were used for highly efficient transformation and for plasmid DNA production, whereas *E. coli* BL21 strains were used for protein expression studies.

Plasmids used during the present study are described in **Table 2.3**.

2.3.1 Preparing bacteria for long term storage

5 mL of LB was inoculated with a single colony of the strain to be stored and grown overnight at 37 °C. An aliquot of the culture was transferred to a cryovial and mixed with sterile glycerol to a final concentration of 20% (v/v). The culture was snap-frozen in dry ice and stored at – 80 °C.

2.3.2 Generation of research cell banks (RCB)

A starter culture of 5 mL of LB was inoculated with a single colony of the strain to be stored. The starter culture was used to inoculate a 250 mL shake-flask containing 50 mL of LB and grown at 37 °C. Once cultures have reached a target $OD_{600} \leq 2.0$, sterile glycerol was added to a final concentration of 20% (v/v). The culture was aliquoted in 2 mL cryovials, snap-frozen in dry ice and stored at – 80 °C.

2.3.3 Generation of master cell plates

A starter culture grown in a 96-well plate containing 200 µL of LB was inoculated with a single colonies of the strains to be stored and grown overnight at 37 °C. 96-deep well plates were inoculated with 10 µL of culture obtained from the 96-well inoculum plate and grown overnight at 37 °C. Sterile glycerol was added to a final concentration of 20% (v/v) and stored at –80 °C.

Table 2.2. *E. coli* strains genotype used in the present work.

***E. coli* strain**	***Genotype***	***Source***
E. shox	*F- mcrA Δ(mrr- hsdRMS- mcrBC) Φ80lacZ ΔM15 ΔlacX74 recA1 endA1 ara Δ139Δ(ara, leu)7697 galU galK λ- rpsL (StrR) nupG*	Bioline
BL21-A	*F- ompT gal dcm lon hsdSB (rB-mB-) [malB+]$_{K-12}$(λS) Δ(araBAD)*	Cobra Biologics
BL21-T7	*F- ompT gal dcm lon hsdSB (rB- mB-) [malB+]$_{K-12}$(λS) araBAD::T7RNAP*	Cobra Biologics

Table 2.3. Plasmid vectors used in the present work.

Plasmid name	*Description*	*Source*
Expression of recombinant human tumour necrosis factor α		
pLT72	Medium copy number (pMB1), T7 expression vector used for NdeI-SalI cloning (Kan^R)	Cobra Biologics
pLT72-TNFα	Derivative of pLT72, T7 expression vector used for the production of TNFα (Kan^R)	Cobra Biologics
pLT72-T7t-TNFα	Derivative of pLT72-TNFα, containing the T7 terminator sequence (Kan^R)	Cobra Biologics
pLT72-T7tKan-TNFα	Derivative of pLT72-TNFα, containing the T7 terminator and the kanamycin gene in reverse orientation (Kan^R)	Cobra Biologics
pLT72-T7tKanT2t-TNFα	Derivative of pLT72-TNFα, containing the T7 and T2 terminator sequences flanking the kanamycin gene in reverse orientation (Kan^R)	Cobra Biologics
Expression of recombinant scFv163R4 and scFv163R4-A		
pLBAD2	Medium copy number (pMB1), pBAD expression vector used for NdeI-SalI cloning (Kan^R)	Cobra Biologics
ptrc2-STII-scFv163R4	ptrc2 vector used as source of the sequence coding for STII-scFv163R4 (Kan^R).	Cobra Biologics
pLBAD2-STII-scFv163R4	Derivative of pLBAD2, pBAD expression vector used for the production of the STII-scFv163R4 (Kan^R)	This work
pLBAD2-DsbA-scFv163R4	pLBAD2-STII-scFv163R4 derivative in which the STII signal peptide has been replaced with the DsbA signal peptide using NdeI and PvuII restriction sites (Kan^R).	This work
pLBAD2-PelB-scFv163R4	pLBAD2-STII-scFv163R4 derivative in which the STII signal peptide has been replaced with the PelB signal peptide using NdeI and PvuII restriction sites (Kan^R).	This work

pLBAD2-STII-scFv163R4-A	pLBAD2-STII-scFv163R4 derivative in which the extra alanine located between the STII signal peptide and the scFv163R4 has been removed by overlapping-PCR (Kan^R).	This work
pLBAD2-DsbA-scFv163R4-A	pLBAD2-DsbA-scFv163R4 derivative in which the extra alanine located between DsbA signal peptide and scFv163R4 has been removed by overlapping-PCR (Kan^R).	This work
pLBAD2-PelB-scFv163R4-A	pLBAD2-PelB-scFv163R4 derivative in which the extra alanine located between PelB signal peptide and scFv163R4 has been removed by overlapping-PCR (Kan^R).	This work
pLBAD2-Bla -scFv163R4-A	pLBAD2-STII-scFv163R4 derivative in which the STII signal peptide has been replaced with Bla signal peptide using NdeI and PvuII restriction sites (Kan^R).	This work
pET163R4	pET-23d(+) vector used for cytoplasmic expression of the scFv163R4 (Amp^R).	Martineau *et al.*, 1998

Expression of β-lactamase

pUC18	High copy number vector used as source of the sequence coding for Bla-β-lactamase (Amp^R).	Norrander *et al.*, 1983
pLBAD2-Bla-β-lactamase	Derivative of pLBAD2, pBAD expression vector used for NdeI-SalI cloning of the Bla-β-lactamase (Kan^R)	This work
pLBAD2-β-lactamase (M)	Derivative of pLBAD2, pBAD expression vector used for NdeI-SalI cloning of the cytoplasmic β-lactamase (Kan^R).	This work

Expression of scFv163R4-A::β-lactamase

pMK-T-scFv163R4-A-linker-Bla	GeneArt vector used as source of the sequence coding for a fragment of the scFv163R4, the linker and the β-lactamase (Kan^R).	GeneArt

pLBAD2-STII-scFv163R4-A::Bla	pLBAD2-STII-scFv163R4-A derivative in which the sequence coding for the linker and β-lactamase from pMK-T-scFv163R4-A-linker-bla vector has been fused to the scFv163R4-A using NotI and SalI (Kan^R).	This work
pLBAD2-DsbA-scFv163R4-A::Bla	pLBAD2-DsbA-scFv163R4-A derivative in which the sequence coding for the linker and β-lactamase from pMK-T-scFv163R4-A-linker-bla vector has been fused to the scFv163R4-A using NotI and SalI (Kan^R).	This work
pLBAD2-PelB-scFv163R4-A::Bla	pLBAD2-PelB-scFv163R4-A derivative in which the sequence coding for the linker and β-lactamase from pMK-T-scFv163R4-A-linker-bla vector has been fused to the scFv163R4-A using NotI and SalI (Kan^R).	This work
pLBAD2-Bla-scFv163R4-A::Bla	pLBAD2-Bla-scFv163R4-A derivative in which the sequence coding for the linker and β-lactamase from pMK-T-scFv163R4-A-linker-bla vector has been fused to the scFv163R4-A using NotI and SalI (Kan^R).	This work
pLBAD2-scFv163R4-A::Bla (M)	pLBAD2-STII-scFv163R4-A::Bla derivative in which the sequence an NdeI site was introduced to generate the cytoplasmic version of the scFv163R4-A::Bla (Kan^R).	This work
Vectors used for the development of the β-lactamase screening		
pMA-RQ-SapIPelBSapI	GeneArt vector containing the sequence coding for PelB signal peptide flanked by the SapI restriction sequences, used as a template for the error-prone PCR (Amp^R).	GeneArt
pLBAD2ΔSapI-STII-scFv163R4-A::Bla	pLBAD2-STII-scFv163R4-A::Bla derivative in which the SapI restriction site within the plasmid backbone has been removed (Kan^R).	This work
pLBAD2-SapIXbaISapI-scFv163R4-A::Bla	pLBAD2ΔSapI-STII-scFv163R4-A::Bla derivative in which the STII signal peptide has been replaced by a sequence containing SapIXbaISapI using NdeI and PvuII (Kan^R).	This work

2.4 Recombinant DNA techniques

2.4.1 Primer design

In general, primers were designed with a length of 18-30 nucleotides, aiming for a melting temperature (Tm) between 55 °C and 65 °C, and within 5 °C of each other. Primers intended to be used to introduce mutations were designed containing the mismatched bases located in the middle of the primers. Snapgene software (GSL Biotech LLC) was used for primer design and visualisation of plasmid vectors.

Primers used during the present study are listed in **Table 2.4.**

2.4.2 Plasmid DNA extraction

Plasmid DNA was isolated at small scale using the QIAprep Spin Miniprep kit (Qiagen) following the protocol described by the manufacturer. Large-scale DNA purification was carried out using either the Qiagen Plasmid maxi kit (Qiagen) or the PureYield Plasmid Maxiprep System (Promega), according to the manufacturer's instructions. Plasmid DNA was generally extracted, when possible, from *E. shox* cultures.

2.4.3 Annealing and phosphorylation of oligonucleotides

Lyophilised oligonucleotides were obtained from Eurofins Genomics and resuspended in annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA). Oligonucleotides were annealed by heating at 95 °C for 5 min and allowed to cool at room temperature. Annealed oligonucleotides used for the introduction of signal peptides and other sequences can be found listed in **Table 2.5.**

Table 2.4. Primers used in the present study.

Primer name	*Sequence 5' to 3'*	*Description*
LBAD F	ttagcggatcctacctgacg	Anneal outside the MCS in pLBAD2 vectors
LBADtrc R	atcagaccgcttctgcgttc	
M13 F	Gtaaaacgacggccagtg	Anneal outside the MCS of pCR®-Blunt vector
M13 R	Cacaggaaacagctatgacc	
scFv F	aaacagcatatgaaaaagaatatcgcat	Introduction of NdeI and SalI to STII-scFv163R4
scFv R	Cttctctcatccgccaaaac	
STII mut F	tacaaatgcctatgcagaggtgcagctggtgg	Removal of extra alanine in scFv163R4 by overlapping-PCR
STII mut R	ccaccagctgcacctctgcataggcatttgta	
DsbA mut F	agcgcatcggcggaggtgcagctg	Removal of extra alanine in scFv163R4 by overlapping-PCR
DsbA mut R	cagctgcacctccgccgatgcgct	
PelB mut F	ccggcgatggccgaggtgcagctg	Removal of extra alanine in scFv163R4 by overlapping-PCR
PelB mut R	cagctgcacctcggccatcgccgg	
Bla SalI F	gtaaacttggtcgtcgacttaccaatgcttaatcagtgaggcacc	Introduction of NdeI and SalI to Bla-β-lactamase gene
Bla NdeI R	tgaaaaaggaagcatatgagtattcaacatttccgtgtcgcccttt	
Cyt Bla F	tgccttcctgttcatatgcacccagaaacgctg	Introduction of NdeI site to β-lactamase (M) gene
NdeI F	gctacaaatgcccatatggaggtgcagctggtg	Introduction of NdeI site to scFv163R4-A::Bla (M) gene
Seq Bla R	Gcacccaactgatcttcagc	Sequencing primer for β-lactamase fusions
SapI F	cgtcttttactggctctactcgctaaccaaaccgg	Removal of SapI in pLBAD2 backbone
LBADilac F	Aagggagaaaggcggacagg	

Table 2.5. Oligonucleotides for annealing and introduction of sequences.

Primer name	***Sequence 5' to 3'***	***Description***
DsbA Fa	Tatgaaaaagatttggctggcgctggct	Oligonucleotides coding for DsbA signal peptide
DsbA Fb	Ggtttagttttagcgtttagcgcatcggcggccgaggtgcag	
DsbA Ra	Ctgcacctcggccgccgatgcgctaaac	
DsbA Rb	gctaaaactaaaccagccagcgccagccaaatctttttca	
PelB F	tatgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgcccagccggcgatggccgccgaggtgcag	Oligonucleotides coding for PelB signal peptide
PelB R	ctgcacctcggcggccatcgccggctgggcagcgaggagcagcagaccagcagcagcggtcggcagcaggtatttca	
Bla F	tatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctgaggtgcag	Oligonucleotides coding for Bla signal peptide
Bla R	ctgcacctcagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactca	
SapIXbaISapI F	tatgtgaagagctctagagctcttctgccgaggtgcag	Introduction of SapI cloning sites into pLBAD2ΔSapI-STII-scFv163R4-A::Bla vector
SapIXbaISapI R	ctgcacctcggcagaagagctctagagctcttcaca	

The previously described protocol was used for the annealing of PelB F and PelB R, Bla F and Bla R, SapIXbaISapI F and SapIXbaISapI R. Annealed oligonucleotides were phosphorylated using the T4 Polynucleotide kinase (PNK), incubated at 37 °C for 30 min. T4 PNK was inactivated by incubation at 70 °C for 15 min.

For the annealing of the DsbA signal peptide, DsbA Fa and DsbA Rb oligonucleotides were phosphorylated using T4 PNK. The phosphorylated DsbA Fa and Dsb Rb oligonucleotides were mixed with an equal molar amount of DsbA Fb and Dsb Ra. Once annealed, oligonucleotides were phosphorylated as described previously.

2.4.4 Polymerase chain reaction (PCR)

Colony PCR

Colony PCR was used to amplify the flanking regions of expression vectors for the screening of sequence insertions using MyTaq™ HS Red Mix (Bioline). PCR reaction mixes were made up to 10 µL in deionised water, the relevant buffer supplied by the manufacturer and oligonucleotide primers to a final concentration of 1 µM. PCR reaction mixes were inoculated with one single fresh colony. DNA amplification was performed in a Hypaid or a MJ Research PCT thermal cycle and cycling conditions are described in **Table 2.6 (A).**

Phusion PCR

PCR was carried out using Finnzymes Phusion™ High-Fidelity DNA polymerase (Bioline) and reaction conditions were adjusted according to manufacturer's instructions. Appropriate oligonucleotide primers were used at a final concentration of 1 µM. PCR reaction mixes were made up to 50 µL in deionised water and the relevant buffer supplied by the manufacturer. Cycling conditions are described in **Table 2.6 (B).**

Table 2.6. PCR protocols used in the present study.

A. Colony PCR

Denaturation	Annealing	Extension	No. of cycles	Hold
95 °C, 2 min			1	
95 °C, 10 sec	60 °C, 10 sec	72 °C, 10 sec	25	
		72 °C, 10 min	1	
				4°C

B. Phusion PCR

Denaturation	Annealing	Extension	No. of cycles	Hold
98 °C, 30 sec			1	
98 °C, 10 sec	x °C[1], 20 sec	72 °C, 30 sec	25	
		72 °C, 10 min	1	
				4 °C

C. Touch-down PCR

Denaturation	Annealing	Extension	No. of cycles	Hold
98 °C, 30 sec			1	
98 °C, 10 sec	68 °C, 20 sec	72 °C, 30 sec	10 →	1 °C decrease per cycle
98 °C, 10 sec	58 °C, 20 sec	72 °C, 30 sec	20	
		72 °C, 5 min	1	
				4 °C

D. One-step overlap extension PCR

Denaturation	Annealing	Extension	No. of cycles	Hold
98 °C, 30 sec			1	Generation of chimeric product
98 °C, 10 sec	67 °C, 20 sec	72 °C, 30 sec	5	
		72 °C, 5 min	1	
				RT
98 °C, 30 sec			1	Amplification of chimeric product
98 °C, 10 sec	67 °C, 20 sec	72 °C, 30 sec	35	
		72 °C, 5 min	1	
				4 °C

[1] Variable annealing temperatures used for the Phusion PCR protocol. The annealing temperature was determined by the melting temperature of the primers used for DNA amplification.

Touch-down PCR (TD-PCR)

Touch-down PCR was used to increase the specificity of PCR reactions when there was a high difference in the annealing temperature between primers. Initially, the annealing temperature was set several degrees above the estimated melting temperature (Tm) of the primers, and gradually decreased during the first cycles (1 °C per cycle). Once reached the optimal Tm, the PCR was then continued at this temperature. Cycling conditions are described in **Table 2.6 (C).**

One-step overlap extension PCR (OE-PCR)

One-step overlap extension PCR, as described by Wei *et al.* (2012), was used to remove the extra amino acid residue, coding for alanine, located between the signal peptide and the scFv163R4 from the following vectors: pLBAD2-STII-scFv163R4, pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4. A summary of the protocol, using as example the generation of the STII-scFv163R4-A, can be found described in **Figure 2.1**. The protocol consisted of two rounds of PCR reactions. In the first PCR, the DNA fragment was amplified using primers LBAD F and STII mut R/DsbA mut R/PelB mut R primers. In the second PCR, the DNA fragment was amplified using the STII mut F/DsbA mut F/PelB mut F and LBADtrc R primers. The two PCR fragments generated shared a homology of approximately 30 bp. Thus, the sequence at the 3' end of the first PCR fragment was identical to the sequence at the 5' end of the second fragment. PCR reactions were conducted using Phusion DNA polymerase and the protocol detailed in **Table 2.6 (B)**.

The two PCR fragments generated were mixed, in which the overlapping sequences of the PCR fragments generated served as primers to allow the extension of the chimeric product. After 5 cycles, the two flanking primers, LBAD F and LBADtrc R, were used to amplify the full-length chimeric product. Cycling conditions can be found described in **Table 2.6 (D).**

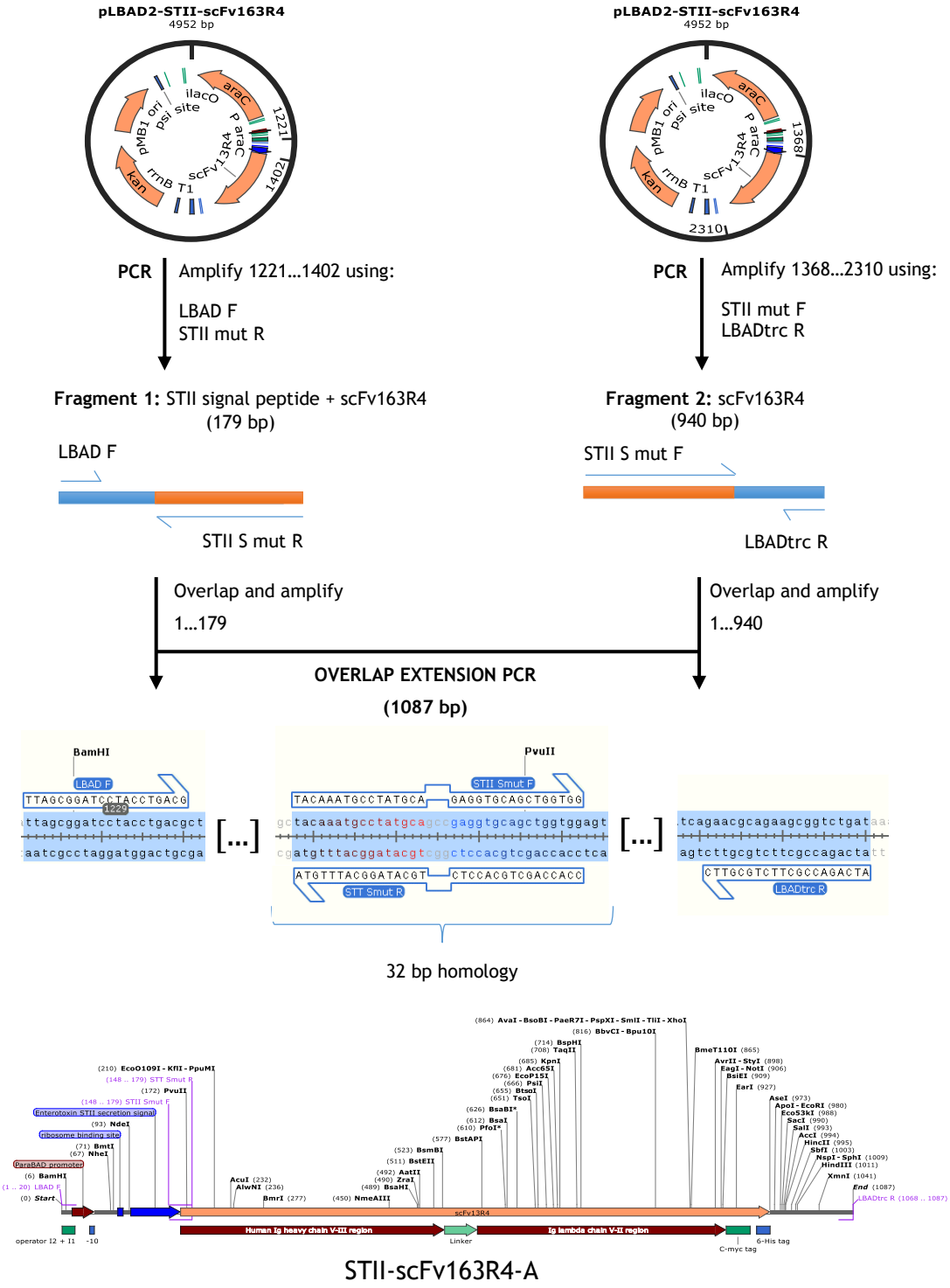


Figure 2.1. One-step overlap extension PCR. Diagram showing the one-step overlap extension PCR method used for the removal of the extra amino acid residue, coding for alanine, located between the signal peptide and the sequence of the scFv163R4, used for the generation of the STII-scFv163R4-A.

2.4.5 Restriction digestion

Restriction digests were carried out using between 0.1 to 5 µg of plasmid DNA, restriction digestion buffer, 1 to 5 U of restriction endonuclease, and nuclease-free water up to the desired volume. Restriction digests were incubated at the temperature and for the duration specified by the manufacturer. Restriction endonucleases were purchased from New England BioLabs (NEB).

2.4.6 Vector dephosphorylation

When required, plasmid vectors were dephosphorylated using calf intestinal alkaline phosphatase (CIP) or shrimp alkaline phosphatase (rSAP), both purchased from NEB. The desphophorylation reaction was carried out according to the manufacturer's recommendations.

2.4.7 Agarose DNA electrophoresis

DNA fragments were analysed using agarose gel electrophoresis. 2% to 0.8% (w/v) agarose solutions in 1x TAE buffer were dissolved by boiling in a microwave. Once cooled to 50 °C, 0.5 µg · mL^{-1} (v/v) of ethidium bromide were added before being poured into an electrophoresis tray. A DNA ladder (1 kb, Agilent Technologies) was used to facilitate DNA size estimation. DNA samples were mixed with 10x (v/v) loading dye before loading. Agarose gels were run at 100 V for 1 hour.

2.2.1 Agarose gel extraction

DNA fragments were excised from agarose gel using a scalpel and extracted using the QIAquick Gel Extraction Kit (Qiagen). 1 mM (w/v) guanosine was added to agarose solutions and running buffer of gels used for gel extraction, to protect the DNA from UV damage (Gründemann and Schömig, 1996). The QIAquick gel extraction kit was also used to purify PCR or digested products.

2.4.8 Ligation

T4 DNA ligase (Promega) was used to ligate DNA inserts and plasmid vectors. Volumes used for ligation reactions were calculated using a 3:1 insert-to-vector ratio. Ligation reactions were incubated for 3 h at room temperature or at 16 °C overnight, as described by the manufacturer.

2.4.9 DNA and protein sequencing

Vector modifications were confirmed by sequencing. DNA sequencing was carried out by Genewiz, formerly Beckman Coulter Genomics, with the exception of the sequencing of signal peptide libraries, which it was carried out by Eurofins Genomics. N-terminal sequencing, obtained from SDS-PAGE gel slices of protein samples was carried out by Alta Bioscience.

2.4.10 Alignment of DNA and protein sequences

The alignment of DNA and protein sequences was carried out using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), a multiple alignment sequence tool (EMBL-EBI) using the default settings. The calculation of the grand average of hydropathicity (GRAVY) was carried out using ProtParam tool (http://web.expasy.org/protparam/) using the default settings.

2.5 Bacterial transformation

2.5.1 Preparation of electrocompetent cells

A single colony was used to inoculate 10 mL of LB and grown overnight at 37 °C. The starter culture was used to inoculate a 1 L shake-flask containing 200 mL of LB and cultures were grown at 37 °C to an OD_{600} of 0.4 - 0.6. Bacteria were collected by centrifugation at 4,000 *g* at 4 °C for 15 min and resuspended in 100 mL of ice-cold 10% (v/v) glycerol. After 1-hour incubation on ice, cells were harvested by centrifugation and resuspended in 50 mL of ice-cold 10% glycerol. The

centrifugation step was repeated and bacteria was resuspended in 25 mL of ice-cold 10% glycerol. Cells were collected by centrifugation, resuspended in 0.5 mL of ice-cold 10% glycerol, aliquoted in cryovials and snap-frozen in dry ice. Competent cells were at stored at –80 °C.

2.5.2 Transformation of E. coli with plasmid DNA

Plasmid DNA (10 to 100 ng) was mixed with electrocompetent cells (30 to 50 μL) and transferred to a chilled electrocuvette. Bacteria were transformed by electroporation using MicroPulser electroporation apparatus (Bio-Rad) with the default settings for *E. coli* (1.8 kV). Immediately after electroporation, bacteria were resuspended in 1 mL of LB and incubated for at least 1 hour at 37 °C, to allow bacteria to recover and express the antibiotic resistance gene. A serially diluted aliquot of the cell suspension was plated onto selective LB agar and incubated overnight at 37 °C.

2.6 Construction of vectors

2.6.1 Vectors for the expression of scFv163R4

Construction of pLBAD2-STII-scFv163R4

The ptrc2-STII-scFv163R4 vector was used as source of the gene coding for the STII-scFv163R4. The STII-scFv163R4 gene was amplified by PCR using the scFv F primer, which introduced an NdeI restriction site, and the scFv R primer using the Phusion™ High-Fidelity DNA polymerase. The PCR product was cloned into pCR®-Blunt vector using the PCR Zero Blunt® PCR Cloning Kit (Life Technologies).

The pLBAD2 and the pCR®-Blunt-STII-scFv163R4 vectors were digested using NdeI and SalI-HF. Digested products were loaded in 0.8% agarose gel and run at 100 V for 1 hour. Subsequently, the digested pLBAD2 vector and the STII-scFv163R4 insert were extracted from the gel using the Qiagen gel extraction kit, ligated together using the T4 DNA ligase and transformed by

electroporation into *E. shox* competent cells. The pLBAD2-STII-scFv163R4 was extracted, purified and transformed by electroporation into *E. coli* BL21-A for expression studies.

Construction of pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4

A diagram depicting of the protocol used for the generation of the pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4 vectors can be found in **Figure 2.2**. The pLBAD2-STII-scFv163R4 vector was digested using NdeI and PvuII and dephosphorylated using CIP. The dephosphorylated pLBAD2-STII-scFv163R4 and the phosphorylated annealed oligonucleotides (annealing protocol described in **Section 2.4.3**) coding for DsbA or PelB signal peptides were ligated using the T4 DNA ligase. The pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4 vectors were transformed into *E. coli* BL21-A for expression studies.

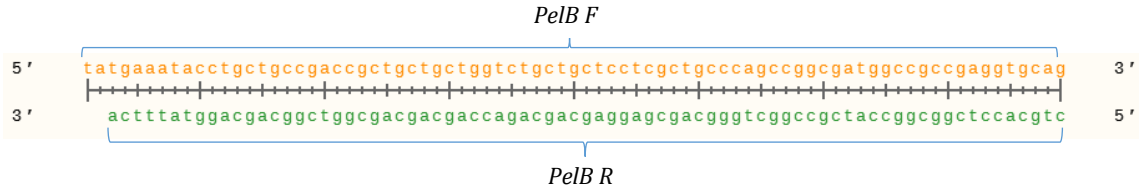
Construction of pLBAD2-STII-scFv163R4-A, pLBAD2-DsbA-scFv163R4-A and pLBAD2-PelB-scFv163R4-A

The removal of the extra alanine at position +1 from the scFv163R4 was carried out by generating a PCR product without the 3 bp coding for the alanine by one-step extension PCR, as described in **Section 2.4.4**.

The pLBAD2 vector and the PCR products coding for STII-scFv163R4-A, DsbA-scFv163R4-A and PelB-scFv163R4-A were digested using the NdeI and SalI-HF, purified and ligated using the T4 DNA ligase. The pLBAD2-STII-scFv163R4-A, pLBAD2-DsbA-scFv163R4-A and pLBAD2-PelB-scFv163R4-A vectors were transformed into *E. coli* BL21-A for expression studies.

A. *DsbA signal (annealed oligonucleotides):*

B. *PelB signal peptide (annealed oligonucleotides):*



C. *pBAD2-STII-scFv163R4*

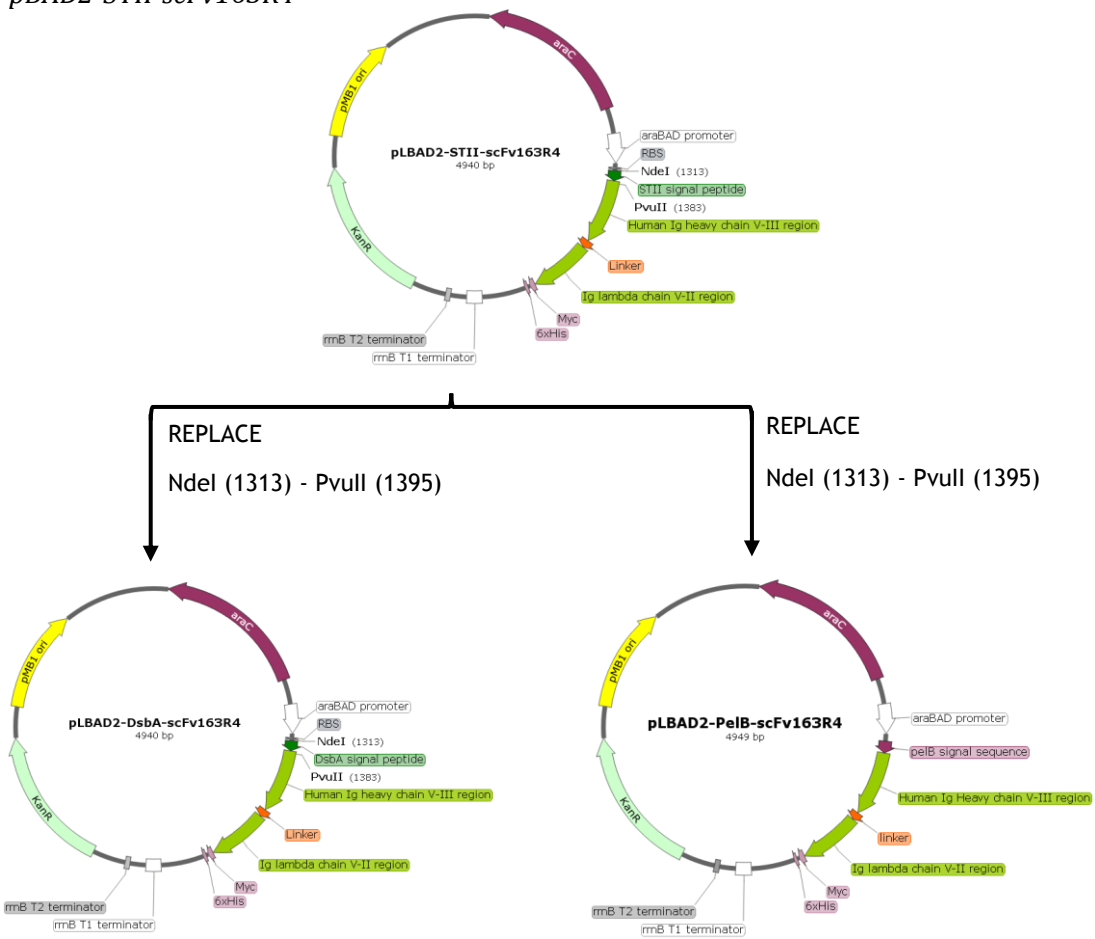


Figure 2.2. Construction of pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4. Diagram showing the annealing of oligonucleotides coding for the signal peptide DsbA (A) and PelB (B), the plasmid backbone used for cloning, pLBAD2-STII-scFv163R4 and the two final plasmid vectors, pLBAD2-DsbA-scFv163R4 and pLBAD2-PelB-scFv163R4 (C).

2.6.2 Vectors for the expression of β-lactamase and scFv163R4-A::β-lactamase

Construction of pLBAD2-Bla-β-lactamase and pLBAD2-β-lactamase (M)

For the generation of the pLBAD2-Bla-β-lactamase vector, the Bla-β-lactamase gene was amplified by Phusion PCR using the Bla SalI and Bla NdeI primers from the pUC18 vector, using the touch-down PCR protocol described in **Table 2.6 (C).** The pLBAD2 and the Bla-β-lactamase gene were digested using NdeI and SalI, purified and ligated using T4 DNA ligase.

For the generation of the pLBAD2-β-lactamase vector (M), the pLBAD2-Bla-β-lactamase vector was used as a template for the amplification of the β-lactamase gene, lacking the signal peptide. The β-lactamase gene was amplified using the Cyt Bla F primer, which introduces an NdeI site and the initiation codon (methionine), and the LBADtrc R primer, by Phusion PCR. The β-lactamase PCR product was cloned in the pCR®-Blunt vector using the PCR Zero Blunt® PCR Cloning Kit according to the manufacturer's instructions. The pLBAD2 and the pCR®-Blunt-β-lactamase were digested using NdeI and SalI, purified and ligated using T4 DNA ligase.

The pLBAD2-Bla-β-lactamase and the pLBAD2-β-lactamase (M) vectors were purified and transformed by electroporation into *E. coli* BL21-A for expression studies.

Construction of pLBAD2-STII-scFv163R4-A::β-lactamase, pLBAD2-DsbA-scFv163R4-A::β-lactamase, pLBAD2-PelB-scFv163R4-A::β-lactamase, pLBAD2-Bla-scFv163R4-A::β-lactamase and pLBAD2-scFv163R4-A::β-lactamase (M)

For the generation of the pLBAD2-STII-scFv163R4-A::Bla vector, the pLBAD2-STII-scFv163R4-A and the pMK-T-scFv163R4-A-linker-Bla were digested using NotI and SalI and ligated using T4 DNA ligase. The pLBAD2-DsbA-scFv163R4-A::Bla and pLBAD2-PelB-scFv163R4-A::Bla vectors were generated using the same protocol.

For the generation of the pLBAD2-Bla-scFv163R4-A::Bla vector, the pLBAD2-STII-scFv163R4-A::Bla was digested using NdeI and PvuII and dephosphorylated using CIP. The dephosphorylated pLBAD2-STII-scFv163R4-A::Bla and the phosphorylated annealed oligonucleotides (annealing protocol described in **Section 2.4.3**) coding for Bla signal peptide were ligated using the T4 DNA ligase.

The pLBAD2-scFv163R4-A::Bla (M) vector was generated using the pLBAD2-STII-scFv163R4-A::Bla vector as a template for the amplification of the scFv163R4-A::Bla gene. The scFv163R4-A::β-lactamase gene, lacking the signal peptide, was amplified using the NdeI F primer, which introduces an NdeI site and the initiation codon (methionine), and the Seq Bla R primer by Phusion-PCR. The pLBAD2-STII-scFv163R4-A::Bla and the scFv163R4-A::Bla PCR product were digested using NdeI and NotI.

The pLBAD2-STII-scFv163R4-A::Bla, pLBAD2-DsbA-scFv163R4-A::Bla, pLBAD2-PelB-scFv163R4-A::Bla, pLBAD2-Bla-scFv163R4-A::Bla and pLBAD2-scFv163R4-A::Bla (M) vectors were purified and transformed by electroporation into *E. coli* BL21-A for expression studies.

2.6.3 Vectors for the β-lactamase screening assay

Construction of pLBAD2-SapIXbaISapI-scFv163R4-A::β-lactamase

Initially, the removal of SapI/BspQI restriction site present in the vector backbone of the pLBAD2-STII-scFv163R4-A::Bla was required for the generation of the pLBAD2-SapIXbaISapI-scFv163R4-A::Bla, intended for the cloning of signal peptide libraries using the type IIs restriction enzyme SapI/BspQI.

For the generation of the pLBAD2ΔSapI-STII-scFv163R4-A::Bla, the SapI restriction site was removed by amplifying the sequence upstream the STII-scFv163R4-A::β-lactamase by Phusion-PCR using the SapI F and the STII mut R primers. The SapI F primer introduced a 1 bp change

which removed the restriction site. A second Phusion-PCR, was carried out using the pLBAD2-STII-scFv163R4-A::Bla as a DNA template, using LBADilac F and the product of the previous PCR as primers, to amplify a larger fragment containing the BspEI restriction site, required for cloning purposes. The product of the second PCR, which includes the sequence of the ori, the AraC and the STII signal peptide, was then amplified using the LBADilac F and the STII mut R primer. A diagram illustrating the steps leading to the generation of the PCR product used for the removal of the SapI/BspQI restriction site can be found in **Figure 2.3**.

The pLBAD2-STII-scFv163R4-A::Bla vector and the final PCR product was digested using BspEI and NdeI. The sequence of the pLBAD2ΔSapI-STII-scFv163R4-A::Bla vector was confirmed by DNA sequencing.

For the generation of the pLBAD2-SapIXbaISapI-scFv163R4-A::Bla vector, the pLBAD2ΔSapI-STII-scFv163R4-A::Bla was digested using NdeI and PvuII and dephosphorylated using CIP. The dephosphorylated pLBAD2ΔSapI-STII-scFv163R4-A::Bla and the phosphorylated annealed oligonucleotides (annealing protocol described in **Section 2.4.3**) coding for SapIXbaISapI were ligated using the T4 DNA ligase.

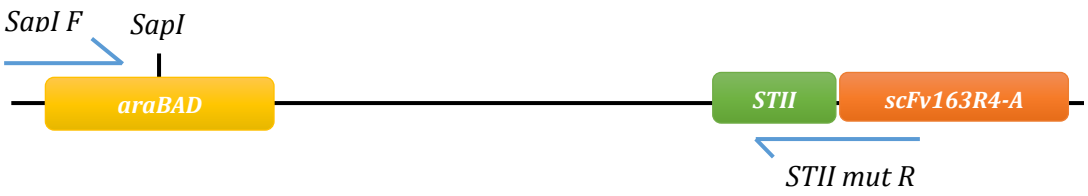
The pLBAD2-SapIXbaISapI-scFv163R4-A::Bla vector was transformed by electroporation into *E. shox* used as expression vector for the introduction of the signal peptide libraries for the β-lactamase screening assay.

2.7 Generation and cloning of signal peptide libraries

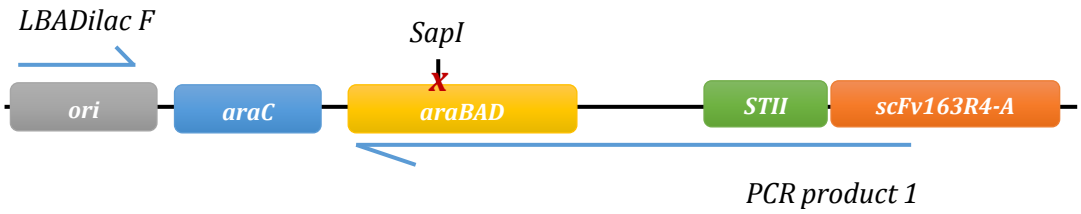
2.7.1 Error-prone PCR signal peptide library

Error-prone PCR was used for the introduction of random mutations in the PelB signal peptide sequence. The sequence coding for PelB signal peptide flanked by SapI/BspQI restriction sites was generated by GeneArt and provided in the pMA-RQ-SapIPelBSapI vector.

A. *1st Phusion-PCR: Removal of SapI restriction site using Sap F*



B. *2nd Phusion-PCR: Amplification of the PCR product using the product of the 1st Phusion-PCR as a primer*



C. *3rd Phusion-PCR: Amplification of the PCR product before restriction digestion*

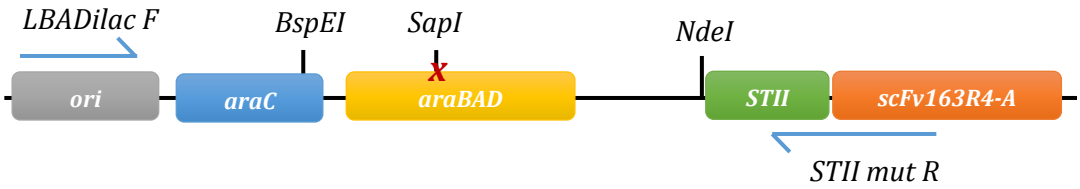


Figure 2.3. Removal of SapI/BspQI restriction site from plasmid backbone. Diagram illustrating the steps leading to the generation of the PCR product used for the removal of the SapI/BspQI restriction site located in the *araBAD* promoter.

Initially, 1 µg of the pMA-RQ-SapIPelBSapI vector was digested with SfiI with the aim to linearize the plasmid. 0.1 ng of the linearized plasmid was used as a template for the error-prone PCR, using the GeneMorph II Random Mutagenesis kit (Agilent). The PCR product obtained, named Ep-PCR 1, was used as template for a second error-prone PCR. Four sequential error-prone PCRs were carried out, obtaining the Ep-PCR 1, 2, 3 and 4 products, to achieve between 1 and 8 mutations per oligonucleotide. A pool containing equivalent amounts of each Ep-PCR product was also generated (Ep-PCR 5 pool), containing a mixture of signal peptide sequences with high and low number of mutations. The primers used and the cycling conditions are shown in **Table 2.7.**

2.7.2 Twist signal peptide library

A second signal peptide library also using PelB signal peptide sequence as a template was chemically synthesised by Twist Biosciences. As part of a Beta program, Twist Biosciences generated an oligonucleotide library that consisted on 10,000 oligonucleotides containing between 2 to 4 mutations per oligonucleotide. Once generated, the oligonucleotide pool was made double-stranded by PCR and amplified to a final concentration of 10 pM.

2.7.3 Restriction digestion and ligation

The expression vector, pLBAD2-SapIXbaISapI-scFv163R4-A::Bla, was digested using BspQI for at least 4 hours. The digested pLBAD2-SapIXbaISapI-scFv163R4-A::Bla and the signal peptide libraries, the EP-PCR products 1, 2, 3, 4 and 5 or the Twist oligonucleotides, were mixed and digested using BspQI for 1 hour, generating the cohesive ends required for cloning. Subsequently, the T4 DNA ligase and ATP to a final concentration of 1 mM was added, and further incubated for 3 hours at room temperature.

Table 2.7. Primers (A) and cycling conditions (B) used for error-prone PCR.

A. Primers

Primer name	***Sequence 5' to 3'***	***Description***
Mut F	gcatcaacgagctcgctcttctatg	Primers for error-prone PCR
Mut R	gcacccaactgatcttcagc	

B. Error-prone PCR

Denaturation	**Annealing**	**Extension**	**No. of cycles**	**Hold**
95 °C, 2 min	55 °C, 20 sec	72 °C, 60 sec	1	
95 °C, 30 sec	55 °C, 20 sec	72 °C, 60 sec	25	
95 °C, 30 sec	55 °C, 20 sec	72 °C, 10 min	1	
				4 °C

2.7.4 Bacterial transformation and plasmid DNA extraction

Ligation products were transformed into *E. shox* competent cells by electroporation. Immediately after, cells were resuspended in 950 µL of LB and incubated for at least 1 hour before being transferred to a 1 L shake-flask containing 200 mL of LB broth supplemented with 50 $\mu g \cdot mL^{-1}$ of kanamycin. After 12-18 hours, cells were harvested, lysed, and the plasmid DNA extracted using PureYield® Plasmid Maxiprep System.

Plasmid DNA stocks were transformed into *E. coli* BL21-A, and after 1-hour incubation, the cell suspension was transferred into a 250 mL shake-flask, containing 50 mL of LB supplemented with 50 $\mu g \cdot mL^{-1}$ of kanamycin and incubated overnight at 37 °C.

For the generation of research cell banks (RCBs), glycerol to a final concentration of 20% (v/v) was added to the culture, aliquoted in cryovials, snap-frozen in dry ice and stored at –80 °C.

2.7.5 Selection of signal peptide mutants based on ampicillin resistance

After incubation, 0.5 mL aliquots of each culture were transferred to a universal bottle containing 10 mL of LB supplemented with 50 $\mu g \cdot mL^{-1}$ of kanamycin, and grown at 37 °C. Once reached an OD_{600} of 1 - 2, cultures were serially diluted to obtain a concentration of 10^4 $CFU \cdot mL^{-1}$ and plated onto M-H agar containing 0.2% (v/v) arabinose, 50 $\mu g \cdot mL^{-1}$ of kanamycin and a concentration of ampicillin ranging from 3 to 1600 $\mu g \cdot mL^{-1}$. After 12 - 18 hours of incubation at 37 °C, colonies grown on ampicillin-supplemented plates were selected to be evaluated by β-lactamase assay.

2.8 β-lactamase enzymatic assay

The evaluation of β-lactamase activity in whole cells can be found described in Angus *et al.* (1982) and the β-lactamase enzymatic assay was carried out as described in O'Callaghan (1972).

2.8.1 Preparation of nitrocefin stock solution

For the preparation of the nitrocefin stock solution, 10 mg of nitrocefin (Carbosynth) was initially dissolved in 200 µL DMSO, and then mixed with 9.8 mL of 10 mM Na-HEPES pH 7.0. The nitrocefin stock solution (1 mg · mL^{-1}) was stored in the dark at -20 °C.

The nitrocefin working solution (0.1 mg · mL^{-1}) was prepared by the addition of 9 mL of Na-HEPES pH 7.0 to 1 mL of nitrocefin stock solution.

2.8.2 Growth of bacterial cells

For the preparation of inoculum plates, 96-well plates containing 200 µL of Terrific broth were inoculated with *E. coli* BL21-A clones selected for the β-lactamase enzymatic assay. The plates were incubated at 25 °C overnight. After incubation, 10 µL of the culture was transferred to a new 96-well plate containing 200 µL of Terrific broth and incubated at 25 °C. After 2 hours of incubation, 0.02% (v/v) arabinose was added to the plates and β-lactamase activity was evaluated at different intervals post-inoculation.

2.8.3 β-lactamase enzymatic assay

For the evaluation of β-lactamase activity, 96-well plates were centrifuged at 1,218 *g* for 15 minutes. After centrifugation, the supernatant was discarded or stored for further analysis. Cell pellets were resuspended in 10 mM Na-HEPES pH 7.0 to the appropriate dilution and 100 µL of nitrocefin working solution was added to the plates. Immediately after the addition of nitrocefin, the OD_{495} was measured continuously for 30 minutes.

2.8.4 Quantification of β-lactamase activity

The initial velocity (slope) was calculated by linear regression and adapted to obtain an $R \geq 0.9$ using GraphPad® software. The initial velocity of the hydrolysis of nitrocefin by β-lactamase was compared between different cell samples. Data was normalised by OD_{600}. The 5% uncertainty of the prediction of the slopes was also calculated.

2.9 Bacterial methods

2.9.1 Minimum inhibitory concentration experiments

Mueller-Hinton (M-H) agar was used for all minimum inhibitory concentration (MIC) experiments. A single colony was used to inoculate a starter culture of 10 mL of LB supplemented with 50 μg · mL^{-1} of kanamycin and grown at 37 °C. Once cultures reached an OD_{600} between 1 and 2, cultures were serially diluted in PBS to achieve a concentration of cells between 10^5 and 10^8 CFU · mL^{-1}. Serially diluted cultures were plated onto M-H agar containing 50 μg · mL^{-1} of kanamycin, 0.2% (v/v) arabinose and different concentrations of ampicillin ranging from 3 to 1600 μg · mL^{-1}. Control plates contained only kanamycin or arabinose and kanamycin. Unless stated otherwise, plates were incubated at 37 °C for 12-18 hours.

2.9.2 Recombinant protein production in shake-flasks

Starter cultures were grown overnight from a single colony of bacteria in 10 mL of LB medium supplemented with 50 μg · mL^{-1} of kanamycin and incubated at 30 °C. Inoculum was added to achieve a starting OD_{600} of 0.1. Cultures were grown in 50 mL of TB dispensed in 250 mL baffled shake-flasks and kanamycin was added to a final concentration of 50 μg · mL^{-1}. Unless stated otherwise, bacteria were aerobically grown in TB medium for all shake-flask experiments. Cultivation temperature varied with experiments.

2.9.3 Recombinant protein production in fed-batch fermentation

Starter cultures were grown in 10 mL of LB medium supplemented with 50 μg · mL^{-1} of kanamycin at 25 °C. Once reached an OD_{600} of 2, starter cultures were used to inoculate a 1 L baffled shake-flask containing 200 mL of LB supplemented with 50 mg · L^{-1} of kanamycin. Cultures grew at 25 °C until they reached an OD_{600} between 4 and 6.

Fed-batch fermentations for the production of rhTNFα were carried out in 7 L STR bench-top fermenters (Applikon), 5 L working volume, equipped with 3 rushton impellers and 4 baffles. The aeration rate was constant at 1 vvm and the dissolved oxygen (DO) was maintained above 20% being controlled by the stirrer speed. All fermentations started with an initial volume of 3 L of culture medium, sterilised within the bench-top fermenter by autoclaving for 20 minutes at 121 °C. Once cooled, the batch salts were supplemented with the post-autoclave additions and trace element solutions. A volume of 2 L of feed solution was prepared and sterilised by filtration. The composition of the batch medium, post-autoclave additions, and feed solutions are described in **Table 2.8**. The pH was maintained at 6.8 by the addition of 5 M NaOH, as base, and 5 M HCl, as acid, for the Cobra Biologics recipe, and at 7.0 by the addition of 5 M HCl and 20% NH_4OH for the Want *et al.* (2009) recipe. Polypropylene glycol (PPG) 2000, used as antifoam, was added when required to prevent blockage of the outlet filter. Fed-batch fermentations were monitored using BioXpert® software package (Applikon).

Fed-batch fermentations for the production of the scFv163R4-A::β-lactamase or scFv163R4-A were carried out using the Ambr250 modular (TAP Biosystems, Sartorius Stedim). The Ambr250 modular uses single-use 250 mL bioreactors equipped with 2 rushton impellers and 4 baffles. Fed-batch fermentations were started with an initial volume of 150 mL of batch medium and 100 mL of feed solution. All the culture medium components were filter-sterilised and added to the vessel prior to inoculation. The composition of the batch medium, post-autoclave additions, and feed solutions were obtained from Want *et al.* (2009) and can be found described in **Table 2.8**.

Table 2.8. Composition of the culture medium used during fed-batch fermentation.

	Cobra Biologics recipe	**Want *et al.* (2009) recipe**
Batch salts	13.3 g · L^{-1} K_2HPO_4 4 g · L^{-1} $(NH_4)_2SO_4$ 1.7 g · L^{-1} Citric acid 10 g · L^{-1} Yeast extract 0.16 mL · L^{-1} PPG 2000	14 g · L^{-1} $(NH_4)_2SO_4$ 35 g · L^{-1} Glycerol 20 g · L^{-1} Yeast extract 2 g · L^{-1} KH_2PO_4 16.5 g · L^{-1} K_2HPO_4 7.5 g · L^{-1} Citric acid 1.5 mL · L^{-1} Conc. H_3PO_4 0.66 mL · L^{-1} PPG 2000
Post-autoclave additions	1 mL · L^{-1} Trace elements solution A 10 mL · L^{-1} Trace elements solution B 10 g · L^{-1} Glycerol 1.2 g · L^{-1} $MgSO_4 \cdot 7H_2O$ 1 mL·L^{-1} 50 mg · mL^{-1} Kanamycin stock	34 mL · L^{-1} Trace elements solution 10 mL · L^{-1} 1 M $MgSO_4 \cdot 7H_2O$ 2 mL · L^{-1} 1 M $CaCl_2 \cdot 2H_2O$ 1 mL · L^{-1} 50 mg · mL^{-1} Kanamycin stock
Feed solution	600 g · L^{-1} Glycerol 5 g · L^{-1} $MgSO_4 \cdot 7H_2O$ 50 g · L^{-1} Yeast extract 10 g · L^{-1} KH_2PO_4 2.1 g · L^{-1} K_2HPO_4 1 mL· L^{-1} 50 mg · mL^{-1} kanamycin stock	714 g · L^{-1} Glycerol 30 mL· L^{-1} 1 M $MgSO_4 \cdot 7H_2O$ 1 mL · L^{-1} 50 mg · mL^{-1} kanamycin stock
Trace elements solution A	5 g · L^{-1} Citric acid 2 g · L^{-1} $CoCl_2 \cdot 6H_2O$ 1.2 g · L^{-1} $CuCl_2 \cdot 2H_2O$ 2.5 g · L^{-1} H_3BO_3 2 g · L^{-1} $Na_2MoO_4 \cdot 2H_2O$ 1.2 g · L^{-1} $MnCl_2 \cdot 4H_2O$	3.36 g · L^{-1} $FeSO_4 \cdot 7H_2O$ 0.84 g · L^{-1} $ZnSO_4 \cdot 7H_2O$ 0.15 g · L^{-1} $MnSO_4 \cdot H_2O$ 0.25 g · L^{-1} $Na_2MoO_4 \cdot 2H_2O$ 0.12 g · L^{-1} $CuSO_4 \cdot 5H_2O$ 0.36 g · L^{-1} H_3BO_3 48 mL · L_{-1} Conc. H_3PO_4
Trace elements solution B	6 g · L^{-1} $FeSO_4 \cdot 7H_2O$ 0.84 g · L^{-1} EDTA · $2H_2O$ 0.8 g · L^{-1} $ZnCl_2$	

The pH was kept at 7.0 by the addition of 10% NH_4OH, as base, and 1 M HCl, as acid. PPG 2000 was used as antifoam. The dissolved oxygen (DO) was maintained above 20% being controlled by the stirrer speed. When the increase in the stirrer speed was not sufficient, pure oxygen was added to the air flow. Fed-batch fermentations were monitored using Ambr250 runtime software package (TAP Biosystems, Sartorius Stedim).

For all the fed-batch fermentations, the inoculum was added to achieve a final OD_{600} of 0.1. The feed solution was started at different time points depending of the fermentation experiment, to achieve a specific growth rate of 0.1 h^{-1}. The feed profile was calculated using the following equation:

$$F = \left(\frac{1}{S}\right) \times \left(\frac{\mu}{Y_{XS}} + m\right) \times X_0 \times V \times e^{\mu t}$$

Where F is the feed rate in $L \cdot h^{-1}$, S is the substrate concentration in the feed in $g \cdot L^{-1}$, μ is the required specific growth rate, Y_{XS} is the yield coefficient in g biomass per g carbon source, m is the maintenance coefficient, X_0 is the biomass in g, V is the volume of the culture medium in L, μ is the specific growth rate and t is time. The yield coefficient (Y_{XS}) and the maintenance coefficient (m) used during the production of rhTNFα were obtained from the literature, corresponding to 0.22 (Passarinha *et al.*, 2009) and 0.025 (Babaeipour *et al.*, 2007). The Y_{XS} and m used during the production of scFv163R4-A::β-lactamase or scFv163R4-A were 0.4 and 0.04 (Wyre, 2015).

When cell paste was required for down-stream purification process, the culture was centrifuged at 3,500 *g* at 4 °C for 30 min. Pellets were resuspended in PBS and homogenised using a Dounce homogeniser. The homogenised cell paste was centrifuged at 7,500 *g* at 4 °C for 30 min. The cell paste was stored at -20 °C.

2.9.4 Measuring growth, cell viability and plasmid retention

The optical density of cultures at 600 nm (OD_{600}) was measured using an Amersham Pharmacia Ultrospec 1100 Pro UV/Visible Spectrophotometer using 10 mm path length plastic cuvettes. Culture samples were serially diluted in PBS and plated onto LB agar for cell culturability and plasmid retention evaluation. LB agar plates were incubated at 37 °C overnight. Colonies were transferred by replica plating to LB agar and LB agar supplemented with 50 mg · L^{-1} of kanamycin and incubated overnight at 37 °C. Colonies replica plated and grown on LB agar and LB agar supplemented with kanamycin were enumerated to assess plasmid retention.

2.10 Protein production analysis

2.10.1 Cell lysis by freeze-thaw

A volume of bacterial culture equivalent to a 1 mL at an OD_{600} of 1 was centrifuged at 12,000 *g* in a microcentrifuge for 10 min. Pellets were resuspended in 250 µL of 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$ supplemented with 1 µL of benzonase (Merck Millipore) and incubated on ice, corresponding to the whole cell lysate protein sample. Subsequently, samples were incubated on ice for 30 min after the addition of 0.12 mg · mL^{-1} of lysozyme. A minimum of 3 cycles of freeze (ethanol/dry ice bath) and thaw (37 °C) were carried out on each sample. Samples were centrifuged at 12,000 *g* for 30 min. The supernatant obtained after centrifugation constituted the soluble protein fraction. Pellets were resuspended in 250 µL of 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, constituting the insoluble protein fraction.

2.10.2 Cell lysis by sonication

As before, a volume of bacterial culture equivalent to a 1 mL at an OD_{600} of 1 was centrifuged at 12,000 *g* in a microcentrifuge for 10 min. Pellets were resupended in 250 µL of 50 mM Tris-HCl

pH 8, 10 mM $MgCl_2$ supplemented with 1 µL of benzonase and incubated on ice, corresponding to the whole cell lysate protein sample. Subsequently, cell samples were sonicated (Model 505 Sonic Dismembrator, Fisher Scientific) with 80% amplitude, for eight cycles (30 seconds on, 30 seconds off). After sonication, the cell lysate was centrifuged at 12,000 *g* for 30 minutes. The supernatant obtained after centrifugation constituted the soluble protein fraction and the insoluble pellet was resuspended in 250 µL of 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$.

The efficiency of cell lysis protocols was evaluated by quantifying the activity of the host protein β-galactosidase in the soluble and insoluble protein fractions, as efficient cell lysis should result in the complete release of β-galactosidase into the soluble fraction.

2.10.3 Periplasmic extraction by cold osmotic shock

A volume of bacterial culture equivalent to a 1 mL at an OD_{600} of 1 was centrifuged at 12,000 *g* in a microcentrifuge for 10 min. Pellets were resupended in 150 µL of ice-cold spheroplast buffer (100 mM Tris-HCl pH 8, 500 mM sucrose, 0.5 mM EDTA) and stored on ice for 5 min. 50 µL were transferred to a new centrifuge tube, constituting the whole lysate protein sample. The cell suspension was centrifuged at 12,000 *g* for 1 minute. The supernatant was discarded and the pellet was resuspended in 100 µL of ice-cold deionised water. Ice-cold $MgCl_2$ was added to a final concentration of 20 mM, the cell suspension was centrifuged at 12,000 *g* for 2 minutes. The supernatant was extracted, constituting the periplasmic protein fraction, and the pellet was resuspended in 100 µL of 10 mM Tris-HCl pH 8, containing the spheroplast protein fraction. If required, the soluble and insoluble protein fractions were separated by sonication or freeze and thaw. The cold osmotic shock protocol was extracted from Humphreys (2007).

2.10.4 Periplasmic extraction by modified cold osmotic shock

As before, a volume of bacterial culture equivalent to a 1 mL at an OD_{600} of 1 was centrifuged at 12,000 *g* in a microcentrifuge for 10 min. The cell pellet was resuspended in 50 µL of ice-cold spheroplast buffer (100 mM Tris-HCl pH 8.2, 500 mM sucrose, 5 mM EDTA). Subsequently, 0.8 mg · mL^{-1} of lysozyme were added, followed by the addition of 50 µL of ice-cold deionised water. The cell suspension was incubated for 5 minutes. $MgSO_4$ to a final concentration of 20 mM was added before centrifugation at 12,000 *g* for 2 minutes. The supernatant constituted the periplasmic protein fraction, and the pellet was resuspended in 100 µL of 100 mM Tris-HCl pH 8, constituting the spheroplast protein fraction. This periplasmic extraction protocol was obtained from Albiniak *et al.* (2013).

The efficiency of the periplasmic extraction protocols was evaluated by measuring the activity of the host protein alkaline phosphatase and quantifying the DNA content in the periplasmic and spheroplast protein fractions; as alkaline phosphatase activity should be only detected in the periplasmic fraction and DNA should only be present in the spheroplast fraction.

2.10.5 Analysis of recombinant proteins by SDS-PAGE

4-12% Bis-Tris NuPAGE SDS-PAGE gels were generally used to evaluate the production of recombinant proteins from samples obtained from shake-flask and fermentation experiments. 7 µL of protein sample were mixed with 2 µL 4x (v/v) NuPAGE sample buffer (Life Technologies) and 1 µL 10x (v/v) NuPAGE reducing agent (Life Technologies), and heated for 10 min at 70 °C. SDS-PAGE gels run for at least 45 minutes at 200 V.

For 18% Tris-Glycine SDS-PAGE gels, samples were prepared by mixing 4 µL of protein sample with 5 µL 2x (v/v) Tris-Glycine sample buffer (Life Technologies) and 1 µL 10x (v/v) NuPAGE reducing agent. Samples were heated for 2 min at 85 °C. SDS-PAGE gels run for 3 hours at 125 V.

SDS-PAGE gels were stained using the Colloidal Blue Staining kit (Life Technologies) according to the manufacturer's instructions. SDS-PAGE gels were stained for a minimum of 3 hours and destained with deionised water for a minimum of 12 hours.

2.10.6 Analysis of recombinant proteins by western blot

Transfer buffer was prepared by the addition of 20x (v/v) NuPAGE transfer buffer (Life Technologies), 10% (v/v) methanol and 1 mL · L^{-1} (v/v) NuPAGE antioxidant. Blotting pads, filter papers and nitrocellulose membranes were submerged in transfer buffer before assembly of the transfer chamber.

For western blots, SDS-PAGE gels were transferred to a 0.2 μm or 0.45 μm nitrocellulose membrane (Life Technologies) using the Xcell II blot module (Life Technologies). SDS-PAGE gels were transferred at 25 V for 1 h. Membranes were incubated in blocking buffer, 5% (w/v) skimmed milk (Sigma-Aldrich) in PBS for at least 1 hour.

Primary antibodies used for western blot included the following: an anti-TNFα antibody (ab9635, Abcam) using 1:2500 dilution was used for the detection of rhTNFα; an anti-myc antibody (R950-25, Life Technologies) using a 1:5000 dilution for the detection of the scFv163R4, the scFv163R4-A and the scFv163R4-A::β-lactamase; and an anti-β-lactamase antibody (sc-66062, Santa Cruz Biotechnology) using a 1:500 dilution for the detection of β-lactamase. All the blots were incubated with the appropriate dilution of primary antibody in blocking buffer for at least 1 h.

After washing with PBST buffer (0.05% (v/v) Tween-20 in PBS), blots were incubated with an anti-mouse IgG antibody conjugated with horseradish peroxidase (HRP) (A4416, Sigma-Aldrich) as a secondary antibody, using 1:5000 dilution in blocking buffer. After 1 hour incubation, blots were washed three times with PBST for 10 mins. Western blots were developed using 3,3′,5,5′-tetramethylbenzidine (TMB) liquid substrate for horseradish peroxidase (HRP) (Sigma-Aldrich).

2.10.7 Quantification of recombinant proteins from SDS-PAGE

AlphaEase® software (Alpha Innotech) was used to calculate the percentage of recombinant protein from total cell protein (TCP). Images were subjected to peak-to-peak background subtraction using the default settings. The percentage of soluble and insoluble recombinant protein was calculated by the software package. Densitometry was used to quantify the concentration of recombinant proteins and the ratio of protein in the soluble/insoluble or periplasmic/spheroplast protein fractions, being these results considered just an estimation of the protein concentration.

rhTNFα reference material (10602-HNAE, Life Technologies) was used to quantify the concentration of rhTNFα by densitometry from culture samples. A standard curve was generated by loading different concentrations of rhTNFα reference material in a SDS-PAGE gel. The concentration of rhTNFα from culture samples was quantified using a standard curve with the AlphaEase® software. The rhTNFα yields were calculated based on the final OD_{600} of the culture.

2.11 Protein purification methods

2.11.1 Purification of rhTNFα

The purification of rhTNFα obtained from fermentation studies was carried out by Dr Nicola Barison at Cobra Biologics. A proprietary purification protocol was used for the purification of rhTNFα, involving the disruption of cell paste obtained by fermentation by the use of a high-pressure cell disruption system (Constant systems). rhTNFα was purified from the clarified soluble protein fraction by a process comprising an ammonium sulphate precipitation, followed by an anion exchange (AEX) and heparin-affinity chromatography.

2.11.2 Purification of scFv163R4-A

The scFv163R4-A was extracted from culture medium samples obtained from fermentation experiments. Culture medium samples were clarified by using a 0.2 μm filter to remove cell debris. Imidazole and NaCl to a final concentration of 40 mM and 500 mM respectively, were added to the clarified culture medium sample, as a sample pre-treatment.

A 1 mL prepacked Nickel Sepharose 6 Fast Flow column (HisTrap FF, GE Healthcare), was used for the purification of the scFv163R4-A. Initially, the HisTrap FF column was washed with 5 column volumes (CV) of deionised water to remove any traces of the ethanol used for the storage of the column. The HisTrap column was equilibrated with 5 CV of binding buffer (PBS pH 7.2, 40 mM imidazole, 500 mM NaCl). 10 mL of the pre-treated culture medium sample, containing 40 mM imidazole and 500 mM NaCl, were loaded onto the column. Subsequently, the column was washed with 15 CV of binding buffer. The scFv163R4-A was eluted with 5 CV elution buffer A (PBS pH 7.2, 100 mM imidazole, 500 mM NaCl). A second elution step was carried out with 5 CV elution buffer B (PBS pH 7.2, 500 mM imidazole, 500 mM NaCl).

2.12 TNFα cytotoxic assay

2.12.1 Cell culture

The C3H mouse fibrosarcoma cell line L929, a cell line sensitive to the activity of TNFα, was used to evaluate the activity of rhTNFα produced by fed-batch fermentation. L929 cells were grown in T225 flasks with Eagle's minimum essential medium (EMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM glutamine and 0.1% $NaHCO_3$ (Gibco, Thermo Fisher Scientific). The cell culture was incubated at 37 °C and 5% CO_2 for three days. Once they had reached confluency, cells were passaged to a new T225 flask, by removing the culture medium and washing the cells with PBS. Trypsin was used to facilitate the detachment of the L929 cells in combination with gentle

shaking or rocking of the flask. Once detached from the surface of the flask, cells were resuspended in fresh EMEM medium and distributed to a new T225 flask.

2.12.2 Cytotoxic assay

The cytotoxic bioassay was carried out following the protocol described by McGee and Clemens (1994). For the cytotoxic bioassay, $6 \cdot 10^4$ cells per well were added to 96-well plates and incubated at 37 °C and 5% CO_2 for 18 hours. Once the cells had grown to reach confluency, the media was exchanged for fresh EMEM media containing actinomycin D, a cell growth inhibitor. The pre-treatment with actinomycin D prevents cell proliferation, sensitising the L929 cells to the activity of rhTNFα, which leads to apoptosis and subsequent cell death (McGee and Clemens, 1994). Different concentrations of rhTNFα reference material, purified rhTNFα or buffer were added to the plates and incubated for 12 - 18 hours at 37 °C and 5% CO_2. The activity of rhTNFα was quantified by the ability of the cells to uptake a staining solution containing crystal violet. The staining solution was subsequently released upon treatment with SDS, a detergent used to lyse the cells. Cells affected by the activity of rhTNFα were unable to uptake the crystal violet solution, which is removed during the washing steps.

After incubation, the culture medium was discarded and 200 µL staining solution (0.5% (v/v) crystal violet, 20 % (v/v) methanol) was added to each well. After 10 minutes, the staining solution was discarded by inversion of the plate, and the excess of staining was removed by the addition of deionised water. L929 cells were solubilised by the addition of 100 µL of 1% (w/v) sodium dodecyl sulphate (SDS) solution. The plate was incubated for 1 hour on a rotary shaker at 180 rpm. The uptake of crystal violet was quantified by measuring the OD_{580}.

2.12.3 Interpretation of results

Initially, the mean absorbance for each triplicate set of standards or samples assayed was calculated; using the mean absorbance data, the percentage of cytotoxicity was calculated using the following formula:

$$\%\ cytotoxicity = \left[1 - \frac{(absorbance\ of\ sample\ or\ standard)}{(absorbance\ of\ 0\ U \cdot mL^{-1}\ TNF\alpha)}\right] x\ 100$$

A TNFα standard curve was constructed by plotting the percentage cytotoxicity values for the TNFα standards using the GraphPad Prism® software. The standard curve was used to calculate the concentration of TNFα which generates a cytotoxicity value of 50% (LD_{50}). The specific activity was calculated assuming that 1 U · mg^{-1} of TNFα specific activity is defined by the quantity of TNFα that produces 50% of the lysis of L929 cells.

2.13 Enzyme-linked immunosorbent assay (ELISA) for the evaluation of the activity of scFv163R4-A

MaxiSorp flat-bottom 96-well plates (Thermo Fisher Scientific) were coated with β-galactosidase (Sigma-Aldrich) to a final concentration of 10 μg · mL^{-1} in PBS. The plate was incubated at 4 °C overnight.

After incubation, the coating solution was discarded and 200 μL of blocking buffer (5% (w/v) milk in PBS) was added to the plates, to block the remaining protein-binding sites. After 2 hours of incubation, plates were washed three times with PBS.

Cell extracts, obtained by sonication using the protocol described in **Section 2.10.2**, including total cell lysates, periplasmic and spheroplast protein fractions and culture medium samples were

added to the plate. The purified scFv163R4-A (protocol described in **Section 2.11.2**), used as standard reference material, was also added to the plates. Samples were diluted across the plate with blocking buffer and incubated for 1 hour at room temperature.

After 1 hour, plates were washed with PBS three times. The primary antibody, an anti-myc HRP-conjugated antibody (R951-25, Life Technologies), was diluted to 1:5000 in blocking buffer and added to the plates. Plates were incubated for 1 hour at room temperature. After incubation, 96-well plates were washed three times with PBS.

100 µL TMB liquid substrate for ELISA (Sigma-Aldrich) was added to each well of the plate. After 10 minutes, 100 µL of 1 M H_2SO_4 was added to stop the reaction. The OD_{450} was measured to quantify the amount of scFv163R4-A present in each sample.

The concentration of the scFv163R4-A was quantified by interpolating the result from the unknown samples using a sigmoidal standard curve generated with data obtained from the purified scFv163R4-A, used as standard. GraphPad® was used to calculate the concentration of active scFv163R4-A from results obtained by ELISA.

Chapter 3:

Production of recombinant human tumour necrosis factor α

3.1 Introduction

The development of novel fermentation strategies that allow the production of soluble and high quality recombinant proteins is still a major area of interest for the biotechnology industry. Today, a large number of human proteins have been produced using *E. coli* as a host organism, such as insulins, interleukins or growth factors, among others. With the aim to exploit different approaches to favour the production of soluble recombinant proteins by high-cell-density fermentation, a pharmaceutically relevant protein, human tumour necrosis factor α (hTNFα), was selected as model protein for these studies.

Tumour necrosis factor α is a cell signalling protein involved in systemic inflammation and its primary role is the regulation of immune cells. It is produced mainly by macrophages, but it can also be produced by other cell types, such as CD4+ lymphocytes, neutrophils, eosinophils, natural killer cells, mast cells, or neurons. TNFα is able to induce fever, apoptotic cell death, inflammation, cachexia and to inhibit viral replication and tumourigenesis.

In human cells, TNFα is primary produced in two biologically active forms, as a 233 amino acid transmembrane protein arranged in homotrimers, which is released from the membrane by proteolytic cleavage; and a soluble 51 kDa trimeric form. The soluble form of TNFα is an homotrimer comprising 17 kDa protomers forming a "jelly roll" structure. rhTNFα is currently on the market, being approved by the EMA in 1999, for the treatment of soft-tissue sarcoma and commercialised by Boehringer Ingelheim, under the trade name of Beromun® (tasonermin).

The aim of the present work was the optimisation of the fermentation conditions for the production of recombinant human tumour necrosis factor α (rhTNFα) using *E. coli* BL21-T7, aiming to achieve a production of 5 $g \cdot L^{-1}$ of rhTNFα for a target OD_{600} of 100. The expression of rhTNFα was driven by the T7 expression system, comprising the T7 RNA polymerase under the control of the arabinose promoter. Different conditions for the production of rhTNFα were evaluated by shake-flask studies, with the aim to favour the accumulation of the recombinant

protein in a soluble and functional form, minimising its accumulation as aggregates, in inclusion bodies (IBs). Factors affecting bacterial growth, such as medium supplements, temperature, induction point or inducer concentration were investigated in an attempt to improve the yield of soluble and active protein. Optimal conditions selected during shake-flask experiments were used for the production of rhTNFα by fed-batch fermentation at 5 L scale. rhTNFα produced during fermentation was recovered from the soluble protein fraction, purified and tested for activity.

3.2 Results

3.2.1 Selection of culture medium for the optimisation of the production of rhTNF α

The expression of rhTNFα was driven by the T7 arabinose-inducible expression system, using *E. coli* BL21-T7 as host organism. A detailed description of this expression system can be found in **Chapter 1, section 1.2.2.2**. Initially, a shake-flask experiment was carried out comparing two culture media formulations, Luria broth (LB) and Terrific broth (TB), with the aim to evaluate the growth and the production of rhTNFα using the T7 expression system.

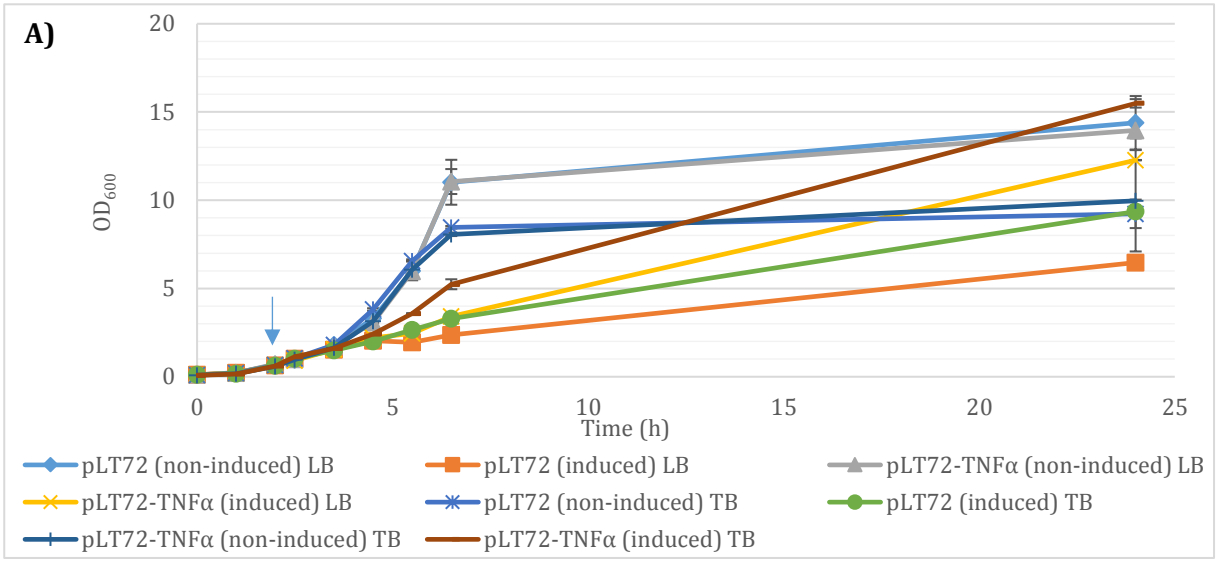
Shake-flask experiments were carried out with *E. coli* BL21-T7 carrying the empty vector (pLT72) or the vector containing the sequence coding for rhTNFα (pLT72-TNFα) under induced and non-induced conditions in both culture media. Cultures were grown at 30 °C and induced with 0.2% arabinose at an OD_{600} of 1. Upon induction, casamino acids were added to the culture medium to a final concentration of 2%, as it has been reported than the addition of casamino acids could have a beneficial effect on the production of rhTNFα (Oshima *et al.*, 1989). Cell culturability, as an indication of cell viability, was evaluated by serially diluted culture samples being plated onto non-selective LB agar and incubated at 37 °C. After 12-18 hours of incubation, colonies were replica-plated onto non-selective and selective LB agar to evaluate plasmid retention. All shake-flask experiments were carried out in duplicate.

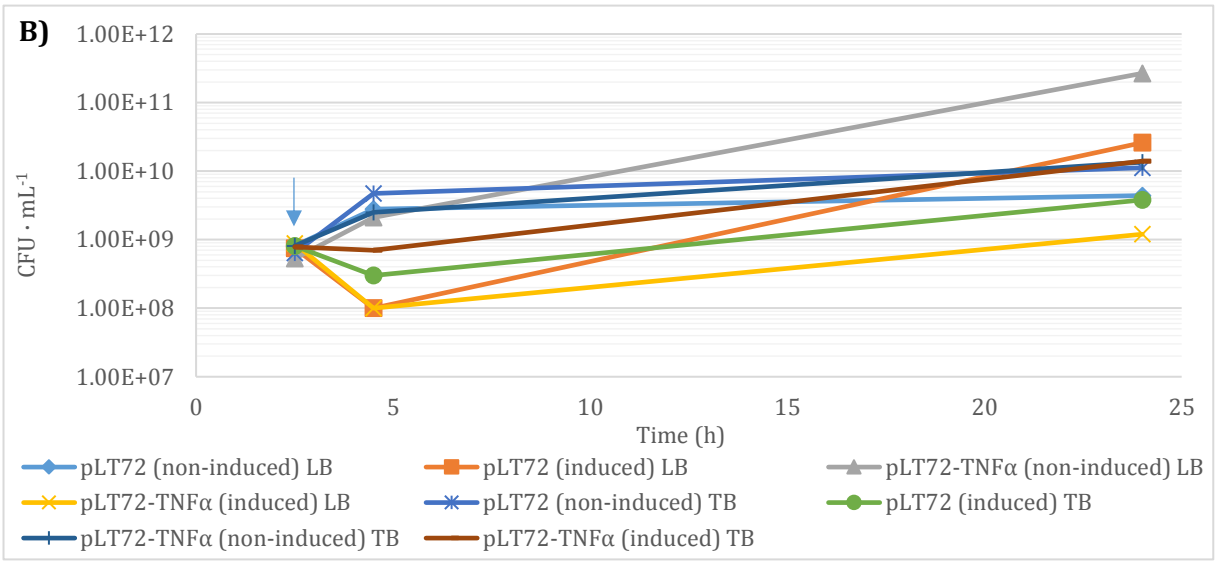
In general terms, cultures grown in LB and TB showed some differences in terms of growth, cell viability and plasmid retention. Cultures growing in LB reached higher cell densities than those growing in TB, reaching a final OD_{600} of 14 under non-induced conditions (**Figure 3.1 A**). In contrast, non-induced cultures growing in TB only reached an OD_{600} of 10, after 24 hours of growth. Induced cultures, either carrying the empty vector (pLT72) or the vector containing the sequence coding for rhTNFα (pLT72-TNFα), generally reached a lower cell density and grew slower than non-induced cultures. The addition of arabinose had a negative effect on growth, not only in cultures producing rhTNFα, but also in cultures carrying the empty vector. Where the expression of rhTNFα was induced by the addition of arabinose, cultures growing either in LB or TB, reached an OD_{600} of 12 and 15, after 24 hours of growth.

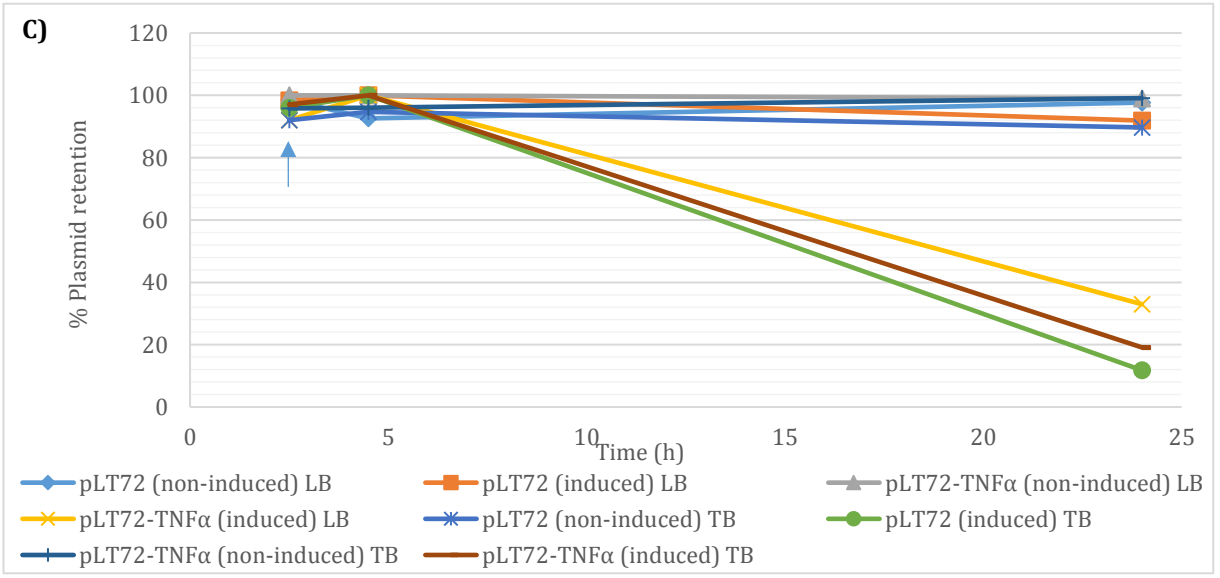
The slower growth of cultures expressing rhTNFα can be explained as the result of the high metabolic burden imposed on the cells during RPP, by the diversion of cell resources to the overproduction of rhTNFα. The negative effect on cell growth observed upon addition of arabinose was not only the result of protein production since this effect was also observed on cultures carrying the empty vector (**Figure 3.1 A**). This effect could not be caused by the addition of arabinose itself, as arabinose has not been reported to have a toxic effect, in contrast with other commonly used inducers, such as IPTG, where toxicity has been reported (Choi *et al.,* 2000; Kosinski *et al.,* 1992). In addition, arabinose cannot be utilised as a carbon source by *E. coli* BL21-T7, as the arabinose utilisation genes, *araB*, *araA* and *araD,* have been completely or partially removed from bacterial chromosome, being replaced by the gene coding for the T7 RNA polymerase.

A decrease in the number of colony forming units (CFU) after 2 hours of induction was also observed, being higher in cultures growing in LB than in TB (**Figure 3.1 B**). The CFU recovered 24 hours after inoculation reaching 10^9 to 10^{10} CFU · mL^{-1}. No loss of cell culturability was observed in non-induced cultures.

Figure 3.1. Selection of culture medium for the optimisation of the production of rhTNFα. *E. coli* BL21-T7 cultures carrying the empty vector (pLT72) or the vector coding for rhTNFα (pLT72-TNFα) were grown at 30 °C; half of cultures were induced with 0.2% arabinose at an $OD_{600} \approx 1$ (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 2.5 h, 4.5 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Colonies were replica-plated onto kanamycin-supplemented LB agar to reveal plasmid retention (C). Data shown are single values for CFU and plasmid retention and mean values from replica flasks for OD_{600}, error bars are ±1 standard deviation.


A)
20
15
10
5
0
OD600
0
5
10
15
20
25
Time (h)
pLT72 (non-induced) LB
pLT72 (induced) LB
pLT72-TNFα (non-induced) LB
pLT72-TNFα (induced) LB
pLT72 (non-induced) TB
pLT72 (induced) TB
pLT72-TNFα (non-induced) TB
pLT72-TNFα (induced) TB


B)
1.00E+12
1.00E+11
1.00E+10
1.00E+09
1.00E+08
1.00E+07
CFU · mL-1
0
5
10
15
20
25
Time (h)
pLT72 (non-induced) LB
pLT72 (induced) LB
pLT72-TNFα (non-induced) LB
pLT72-TNFα (induced) LB
pLT72 (non-induced) TB
pLT72 (induced) TB
pLT72-TNFα (non-induced) TB
pLT72-TNFα (induced) TB


C)
120
100
80
60
40
20
0
% Plasmid retention
0
5
10
15
20
25
Time (h)
pLT72 (non-induced) LB
pLT72 (induced) LB
pLT72-TNFα (non-induced) LB
pLT72-TNFα (induced) LB
pLT72 (non-induced) TB
pLT72 (induced) TB
pLT72-TNFα (non-induced) TB
pLT72-TNFα (induced) TB

Plasmid retention remained around 100% for all cultures 2 hours after induction, but it decreased drastically for arabinose-induced cultures 24 hours post-inoculation. Induced cultures producing rhTNFα (pLT72-TNFα) showed a plasmid retention of only 10 to 35% after 24 hours of growth (**Figure 3.1 C**). The overgrowth of plasmid-free cells could explain the unexpectedly high cell density after 24 hours of growth observed in rhTNFα producing cultures for both culture media (**Figure 3.1 A**), as a result of the low plasmid retention.

The production of rhTNFα was evaluated by SDS-PAGE in samples collected at time points pre- and post-induction (**Figure 3.2**). Bacterial pellets obtained after 24 hours of growth were fractionated using the freeze-thaw method to obtain the soluble and insoluble protein fractions. Lysozyme was added to facilitate cell lysis, and it can be observed in the soluble and insoluble protein fractions, running as a 14.3 kDa protein. The quantity of rhTNFα increased over the time after induction, reaching a maximum of 20% of the total cell protein (TCP) after 24 hours of growth for both culture media evaluated. After 24 hours of growth, a high quantity of rhTNFα was accumulated in a soluble form, accounting for 55% of the total rhTNFα.

An unexpected protein band was also detected in samples obtained from induced cultures, both cultures carrying the empty vector (pLT72) and the vector containing the sequence coding for rhTNFα (pLT72-TNFα), which accounted for a 10% of TCP in samples obtained from cultures carrying the empty vector (pLT72). It was hypothesised that this unknown protein could be the product of the kanamycin resistance gene, aminoglycoside 3' phosphotransferase (APH). APH catalyses the addition of phosphate from ATP to the 3'-hydroxyl group of a 4,6-disubstituted aminoglycoside, disrupting the mechanism of action of the antibiotic kanamycin. APH has a molecular weight of 31 kDa, similar to the unknown protein present in samples obtained from induced cultures (**Figure 3.2**). The lack of terminator sequences downstream the multiple cloning site of the backbone of the vector pLT72 may have allowed the read-through by the T7 RNA polymerase, resulting in the overproduction of APH.

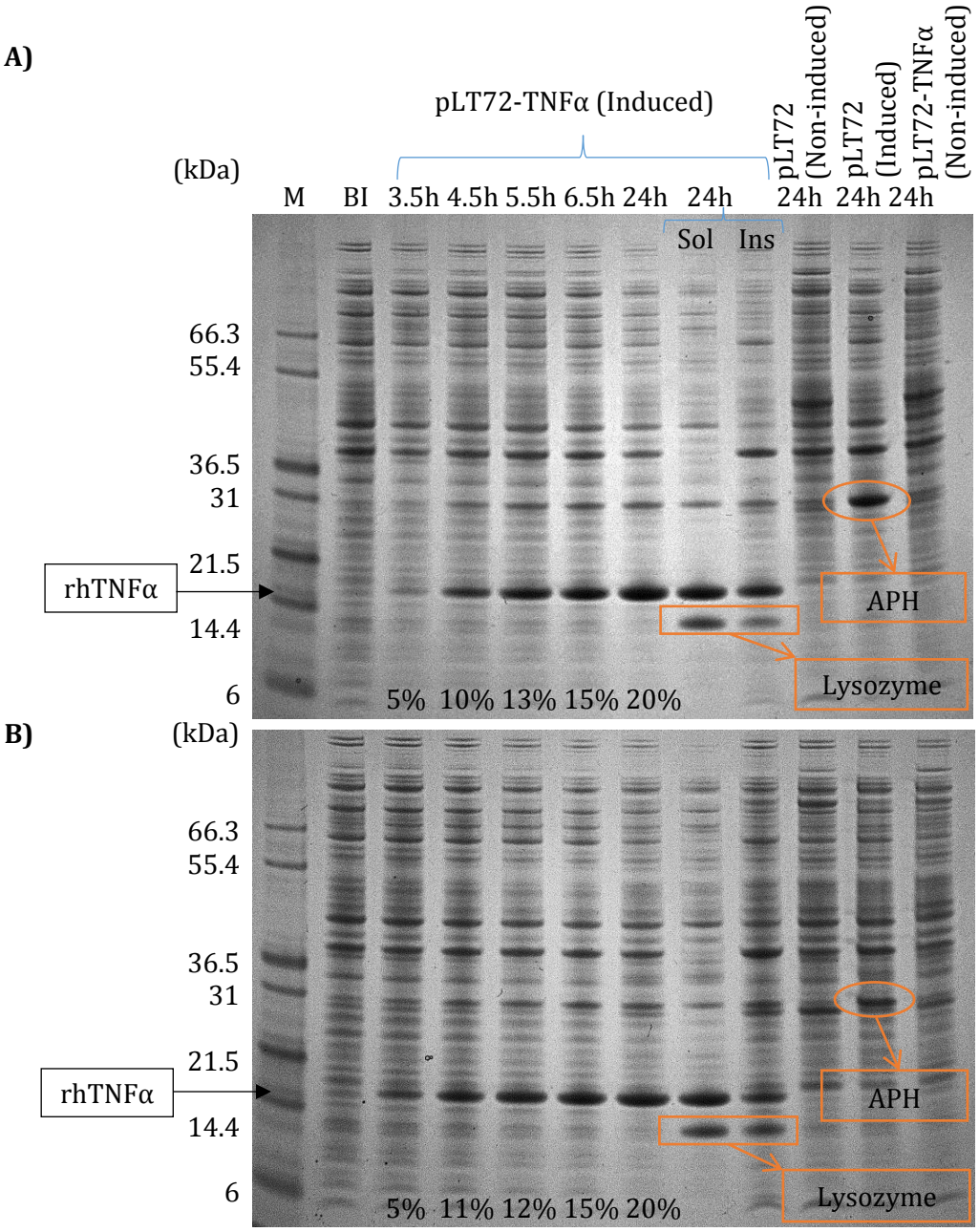


Figure 3.2. Accumulation of rhTNFα in cultures grown in LB and TB culture media. SDS-PAGE gels showing the accumulation of rhTNFα from whole cell lysates from cultures grown in LB (A) and TB (B) media. Samples were obtained from induced and non-induced cultures carrying the empty vector (pLT72) or the vector encoding rhTNFα (pLT72-TNFα). The quantity of rhTNFα is expressed as a percentage of whole cell protein at the bottom of the gel. Samples collected at 24 hours were also fractionated to obtain soluble (sol) and insoluble (ins) protein fractions facilitated by the addition of lysozyme (orange square). The overproduction of the presumptive APH (orange circle) can be observed in all the samples obtained from cultures induced with arabinose.

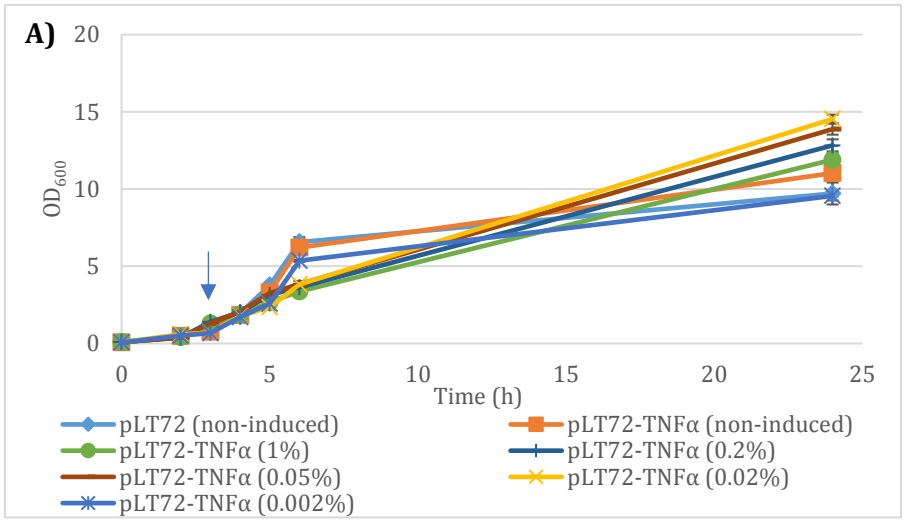
In general terms, both culture media showed similar results in terms of growth, cell viability, plasmid retention and protein production. Both media formulations were found adequate for growth experiments. As a result of the improved buffering capacity of terrific broth, this medium formulation was selected as standard for all the shake-flask studies carried in the present work.

3.2.2 Effect of inducer concentration on the production of rhTNFα

As observed in the previous experiment; growth arrest, low cell viability and plasmid loss occurred upon recombinant protein induction. The previous experiment was carried out using a concentration of arabinose of 0.2%, the standard concentration recommended by the literature to induce the arabinose promoter (Guzman *et al*., 1995; Terpe, 2006). However, this concentration may be too high due to the amplification loop of the T7 RNA polymerase, resulting in a high metabolic burden, compelling the cells to lose the plasmid. Therefore, different concentrations of arabinose, ranging from 1% to 0.002%, were evaluated to determine whether the use of lower concentrations of arabinose could have a beneficial effect on growth and rhTNFα production.

Cultures were grown at 30 °C in terrific broth supplemented with casamino acids to a final concentration of 2%. The expression of rhTNFα was induced with a concentration of arabinose ranging from 1% to 0.002% at an OD_{600} of 1. Cells carrying the empty vector (pLT72) or the vector coding for rhTNFα (pLT72-TNFα) under non-induced conditions were used as controls.

The use of different concentrations of arabinose had a great effect on bacterial growth, and overall, all the induced cultures grew slower upon induction than non-induced cultures (**Figure 3.3 A**). As expected, the use of high concentrations of arabinose (1 to 0.02%) induced growth arrest but reaching a higher cell density than the control cultures, probably as a result of the overgrowth of plasmid-free cells. Growth arrest was not observed for cultures induced with the lowest concentration of arabinose (0.002%) evaluated in this experiment.



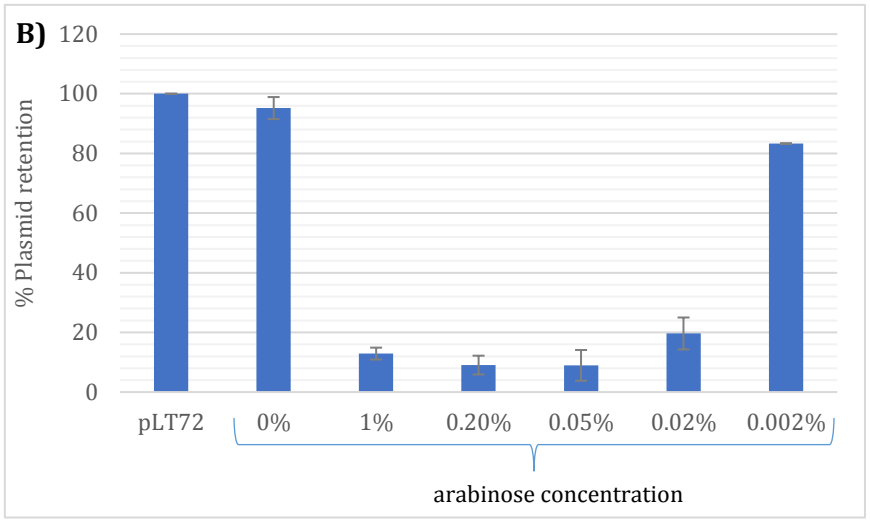
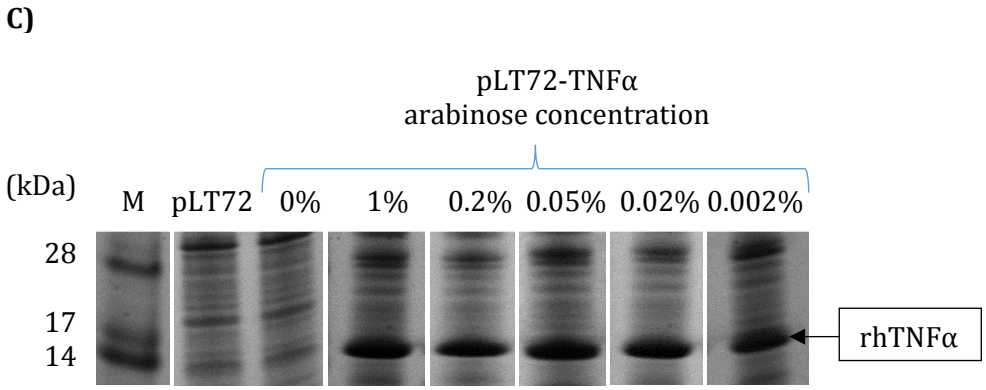


Figure 3.3. Effect of inducer concentration on the production of rhTNFα. Cultures grown at 30 °C were induced with a concentration of arabinose ranging from 1% to 0.002%, at OD_{600} ≈ 1 (blue arrow). The OD_{600} was measured at intervals post-inoculation (A). Plasmid retention, after 24 hours of growth, was obtained by replica-plating colonies in selective and non-selective LB agar (B). SDS-PAGE showing the accumulation of rhTNFα from whole cell lysates obtained 24 h post-inoculation (C). Cultures carrying the empty vector (pLT72) and the vector coding for rhTNFα (pLT72-TNFα) under non-induced conditions were used as control. Data shown are mean values from two replica flasks for OD_{600} and plasmid retention, error bars are ±1 standard deviation.



In terms of plasmid retention, the use of high concentrations of arabinose led to a drastic plasmid loss (**Figure 3.3 B**). Cultures induced with 1%, 0.2%, 0.05% and 0.02% showed a plasmid retention lower than 20%, after 24 hours of growth. In contrast, cultures induced 0.002% arabinose showed a great improvement on plasmid retention, with more than 80% of the cells being plasmid-positive at the end of the experiment.

The production of rhTNFα was shown to be similar independently of the concentration of arabinose used to induce protein production, being the lowest concentration of arabinose evaluated, 0.002%, enough to fully induce the T7 expression system (**Figure 3.3 C**). In addition, the majority of the rhTNFα was found to be accumulated in the soluble protein fraction for all the induced cultures.

Overall, a concentration of arabinose 100 times lower than the standard recommended concentration seems to be sufficient to completely induce protein production without entailing a reduction in the production of rhTNFα, but minimising the growth arrest and plasmid loss.

3.2.3 Effect of induction point on the production of rhTNFα

The effect of different induction points was evaluated by shake-flask. However, growth experiments carried out in shake-flasks only support growth up to a maximum OD_{600} of 10 to 20, limiting the evaluation of the effect of the induction point to low cell densities. Mid (induction at OD_{600} of 20-40) and late (induction at OD_{600} of 60-80) induction need to be evaluated by fed-batch fermentation.

Cultures were grown in terrific broth supplemented with 2% casamino acids at 30 °C and induced with 0.2% arabinose. Three different induction points were evaluated at OD_{600} of 0.5, 2 and 3. The addition of inducer in previous studies was carried out at an OD_{600} of 1, and therefore, this induction point was not included in this experiment.

In general terms, there were no drastic differences between cultures induced at different cell densities (**Figure 3.4 A**). Cultures induced at an OD_{600} of 0.5 grew slower upon protein induction but reached a higher cell density than cultures induced at an OD_{600} of 2 or 3. No significant differences in cell culturability were observed (data not shown).

In terms of production of rhTNFα, there were no significant differences between the three induction points evaluated in this experiment, showing similar levels of production of rhTNFα after 2 and 4 hours of induction (**Figure 3.4 B**).

Overall, an early induction of the production of rhTNFα did not have a negative effect on growth, however, further evaluations of the effect of different induction points at high cell densities need to be evaluated by fed-batch fermentation. Since no major differences after protein induction were observed, induction at an OD_{600} of 0.5 was selected as optimal for the following experiments.

3.2.4 Effect of temperature on the production of rhTNFα

The effect of different cultivation temperatures on RPP was also evaluated, since it has been reported than low cultivation temperatures favour the accumulation of some recombinant proteins in a soluble form (Alfasi, 2011; Sevastsyanovich *et al.*, 2010). Previous shake-flask experiments were carried out at 30 °C and therefore, the effect of an increase of the cultivation temperature to 37°C and a decrease to 25 °C were evaluated during this experiment. The effect of the temperature was not only evaluated in terms of rhTNFα yield but more specifically, in terms of soluble and insoluble protein product being accumulated.

Cultures were grown in terrific broth supplemented with 2% casamino acids and grown at a temperature of 37 °C, 30 °C or 25 °C. The expression of rhTNFα was induced by the addition of 0.2% arabinose at an OD_{600} of 1.

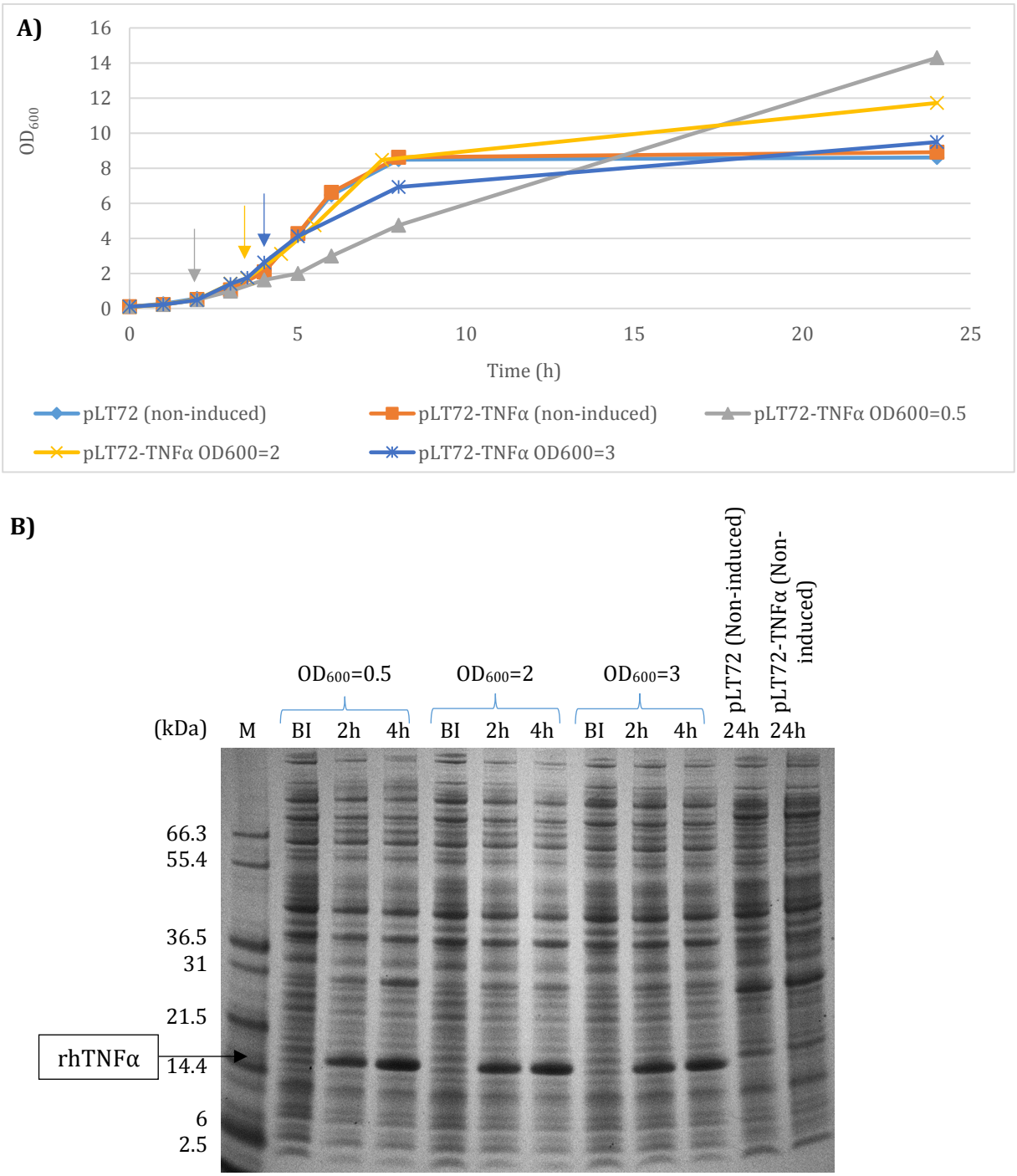


Figure 3.4. The effect of induction point on the production of rhTNFα. *E. coli* BL21-T7 carrying the empty vector (pLT72) or the vector coding for rhTNFα (pLT72-TNFα) were grown at 30 °C under induced and non-induced conditions. Cultures were induced with 0.2% arabinose at an OD_{600} of 0.5, 2 or 3 (arrows). The OD_{600} was measured at intervals post-inoculation (A). SDS-PAGE gel showing the accumulation of rhTNFα from whole cell lysates before induction (BI), 2 hours and 4 hours after induction (B). Data shown are mean values from two replica flasks for OD_{600}, error bars are ±1 standard deviation.

The cultivation temperature had a significant effect on the proportion of rhTNFα accumulated in the soluble and insoluble protein fractions (**Figure 3.5**). Cultures grown at the highest cultivation temperature evaluated, 37 °C, showed that only 54% of the total rhTNFα produced was being accumulated in a soluble form, the remaining 46% was accumulated in the insoluble fraction. At 30 °C, 73% of total rhTNFα was found accumulated in the soluble fraction. A further decrease to 25 °C showed an increase of the production of soluble rhTNFα up to 90% (**Figure 3.5 A**). The differences in the accumulation of rhTNFα in the soluble and insoluble fractions for the three temperatures was evaluated by SDS-PAGE and quantified by densitometry (**Figure 3.5 B**).

In general terms, the correct folding of rhTNFα was favoured by lower cultivation temperatures, helping to slow down the protein synthesis rate, allowing the correct folding of the recombinant protein and avoiding the formation of IBs (Sahdev *et al.,* 2008; Sevastsyanovich *et al.,* 2010).

3.2.5 Effect of temperature and inducer concentration on the production of rhTNFα

The effect of different inducer concentrations in combination with lower cultivation temperatures was further evaluated with the aim to understand whether the plasmid loss observed in cultures producing rhTNFα could be minimised. Cultures were grown in terrific broth supplemented with 2% casamino acids at 25 °C. The expression of rhTNFα was induced by the addition of different concentrations of arabinose, ranging from 0.2% to 0.001%, at an OD_{600} of 0.5.

Cultures carrying the empty vector (pLT72) and the vector containing the sequence coding for rhTNFα (pLT72-TNFα) under non-induced conditions were used as control, achieving plasmid retention of >95% after 24 hours of growth (**Figure 3.6 A**). Cultures induced with 0.2% arabinose, showed a plasmid retention lower than 10%. In contrast, plasmid loss was minimal, accounting for less than 5%, in cultures induced with low concentrations of arabinose, ranging from 0.02%

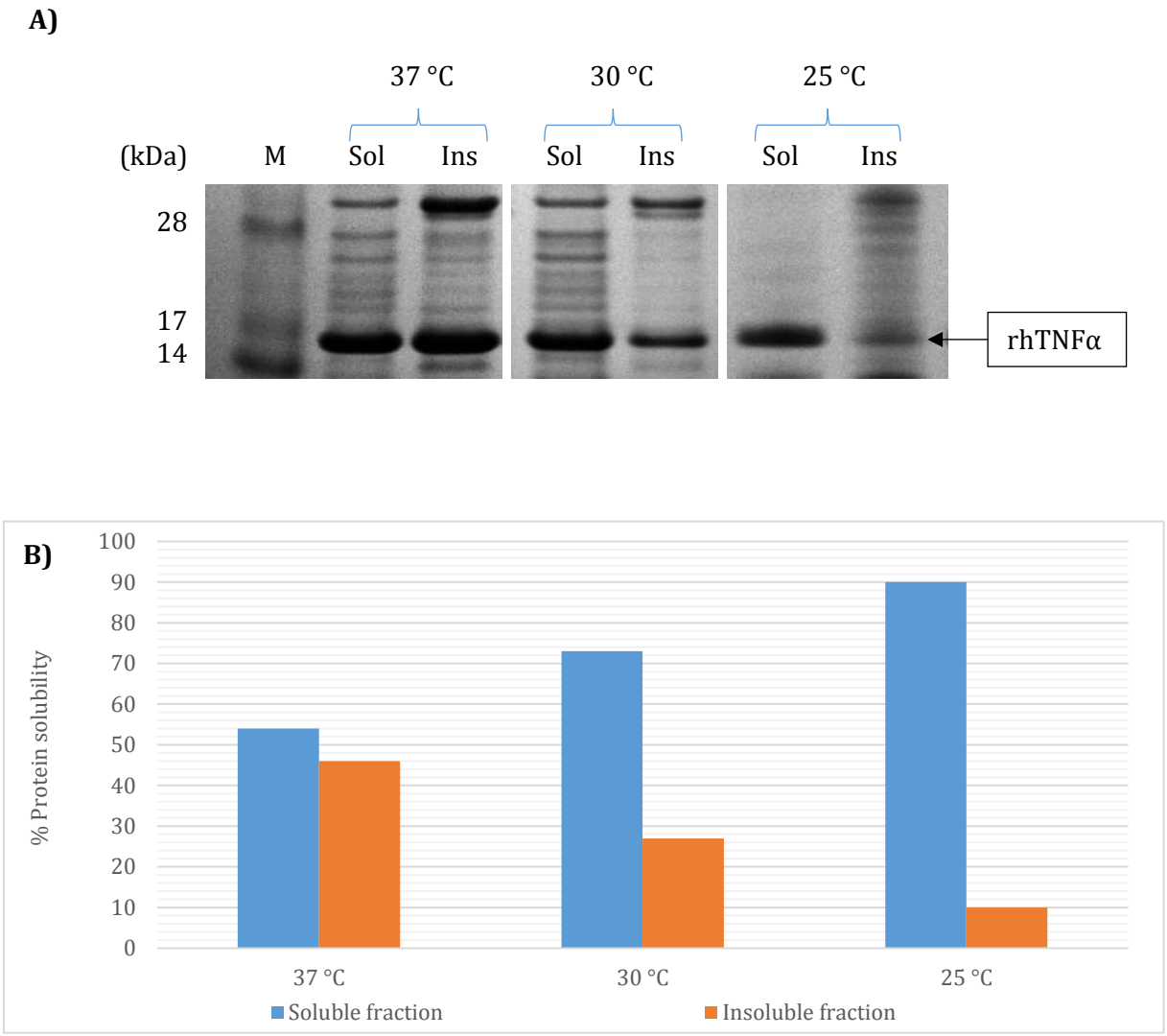
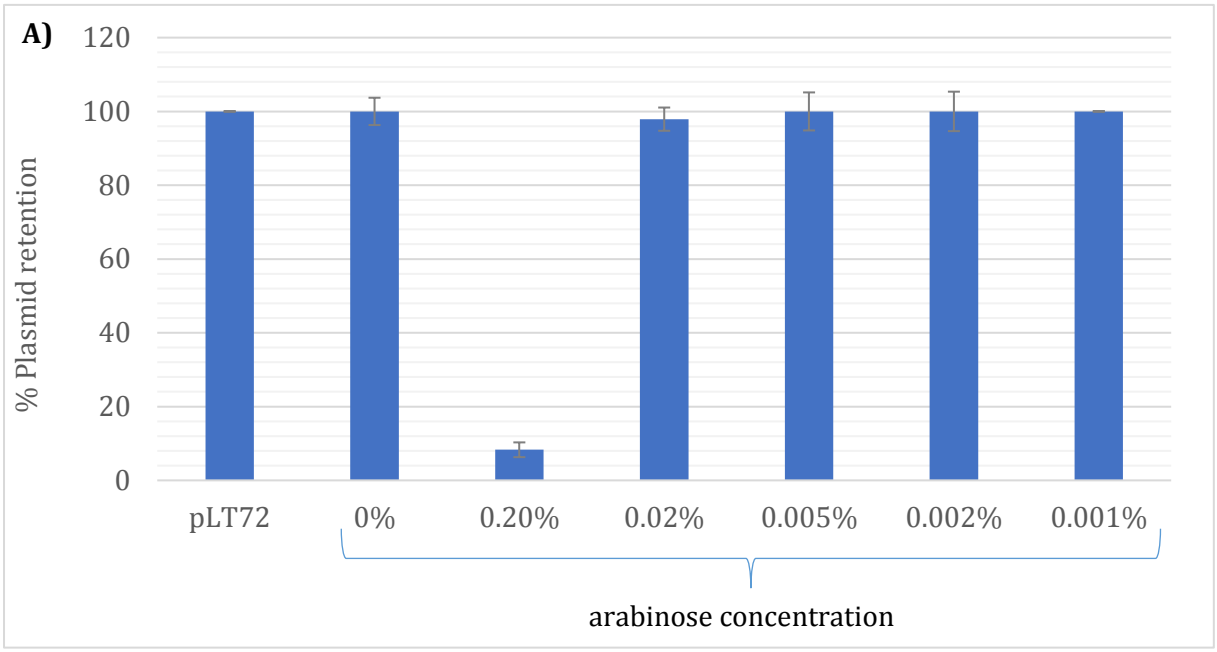


Figure 3.5. Effect of temperature on the production of rhTNFα. SDS-PAGE gel showing the accumulation of the rhTNFα in the soluble and insoluble fraction 4 hours after induction. Cultures were induced with 0.2% arabinose and grown at 37 °C, 30 °C or 25 °C. Samples were fractionated to obtain the soluble (sol) and insoluble (ins) cell fractions (A). The accumulation of rhTNFα in the soluble and insoluble fraction was quantified by densitometry from SDS-PAGE gels by comparing the accumulation of rhTNFα in the soluble and insoluble protein fractions. Data shown here are single values.



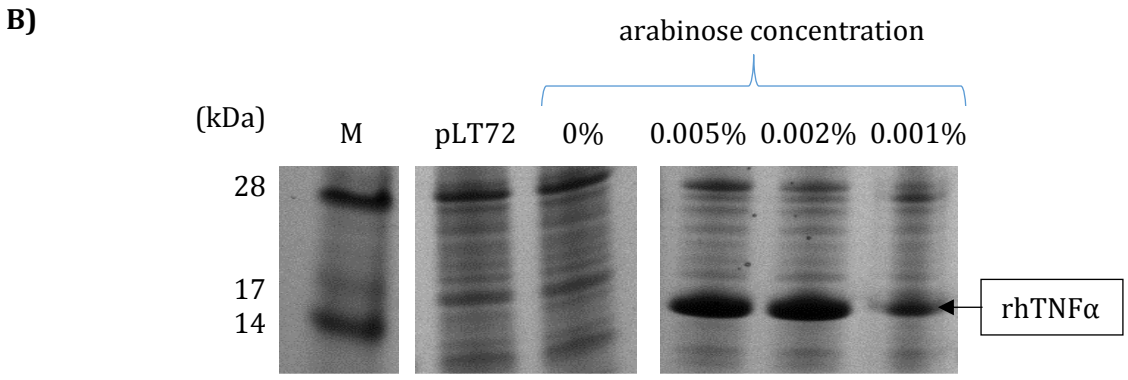


Figure 3.6. The effect of temperature and inducer concentration on the production of rhTNFα. Plasmid retention of cultures carrying the empty vector (pLT72) or the vector coding for rhTNFα (pLT72-TNFα) grown at 25 °C under non-induced conditions or induced at an OD_{600} ≈ 0.5 with different concentrations of arabinose, ranging from 0.2% to 0.001%. Serial dilutions of culture samples obtained after 24 hours of growth were plated onto non-selective LB agar. Colonies were replica plated onto kanamycin-supplemented LB agar to reveal the plasmid retention. Data shown are mean values from two replica flasks, error bars are ±1 standard deviation (A). SDS-PAGE gel showing the accumulation of rhTNFα from whole cell lysates after 24 hours of growth (B).

to 0.001%. A decrease in the temperature to 25 °C showed a great improvement when compared with previous results obtained at 30 °C, when plasmid retention was lower than 85% and 20%, for cultures induced with 0.002% and 0.02% arabinose, respectively (**Figure 3.3 C**). The use of a concentration of arabinose of 0.001% showed a high plasmid retention; however, the accumulation of rhTNFα was also lower than when higher concentrations of arabinose were used, not being enough to fully induce the T7 promoter (**Figure 3.6 B**).

Lower cultivation temperatures in combination with a lower concentration of arabinose minimised the physiological stress imposed on the cells, favouring the plasmid retention and enhancing the production of soluble rhTNFα. As cell stress can often be found associated with plasmid loss, the reduction of the inducer concentration also had a positive effect on plasmid maintenance. In fed-batch fermentations, the low plasmid retention can severely affect the final product yield, since only a small percentage of the cell population is still carrying the plasmid, and therefore producing the recombinant protein of interest (Palomares *et al.*, 2004). Overall, a temperature of 25 °C and a concentration of arabinose of 0.005% were found to be optimal for the production of rhTNFα.

3.2.6 Selection of the optimal vector design to minimise the overproduction of APH

Initial experiments carried out to evalute the production of rhTNFα in *E. coli* showed the overproduction of an unknown protein of approximately 31 kDa, for cultures grown under induced conditions (**Figure 3.2**). It was considered that this unknown protein could be the protein that confers resistance to kanamycin, the aminoglycoside-3'-phosphotransferase (APH).

To confirm whether this unknown protein was APH and to reduce its hypothetical overproduction, several vector variations were generated. The main objective was to minimise the read-through of the T7 RNA polymerase by the addition of terminator sequences downstream the sequence coding for the rhTNFα (pLT72-T7t-TNFα) or by changing the orientation of the

kanamycin gene (pLT72-T7tKan-TNFα). The addition of two different terminator sequences flanking the kanamycin resistance gene was also evaluated (pLT72-T7tKanT2t-TNFα) (**Figure 3.7**). The new plasmid vectors were kindly generated by Dr Steven Williams at Cobra Biologics.

Cultures were grown in terrific broth supplemented with 2% casamino acids at 30 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. These growth conditions are known to show high plasmid loss and were selected to determine whether the new vector designs could show an improvement in plasmid retention. Non-induced cultures carrying the empty vector (pLT72) and the vector coding for rhTNFα (pLT72-TNFα) were used as controls.

The control cultures, carrying the empty vector (pLT72) and the vector coding for rhTNFα (pLT72-TNFα) under non-induced conditions showed a plasmid retention of 100% after 24 hours of growth (**Figure 3.8 A**). As observed in previous experiments, the original vector design (pLT72-TNFα) showed a plasmid retention lower than 25%. In contrast, cultures carrying the vector containing the T7 terminator showed a slight improvement in plasmid retention, with more than 45% of the cells being plasmid-positive. Plasmid retention was further improved in cultures carrying the vector with the T7 terminator and the sequence coding for the kanamycin resistance gene in the opposite orientation to the TNFα gene (pLT72-T7tKan-TNFα), with a plasmid retention of 80%. Plasmid retention was lower for the plasmid containing both the T7 and the T2 terminators sequences (pLT72-T7tKanT2t-TNFα), with a plasmid retention of 45.7%.

There were no significant differences between the four plasmid vector backbones in terms of production of rhTNFα (**Figure 3.8 B**). The accumulation of APH can be observed in cultures carrying the original vector design. A slight decrease in the accumulation of this protein can be observed in cultures carrying the vector containing the T7 terminator sequence, and even further reduced for vectors with the sequence coding for kanamycin resistance gene in reverse orientation, confirming that the unknown protein was APH.

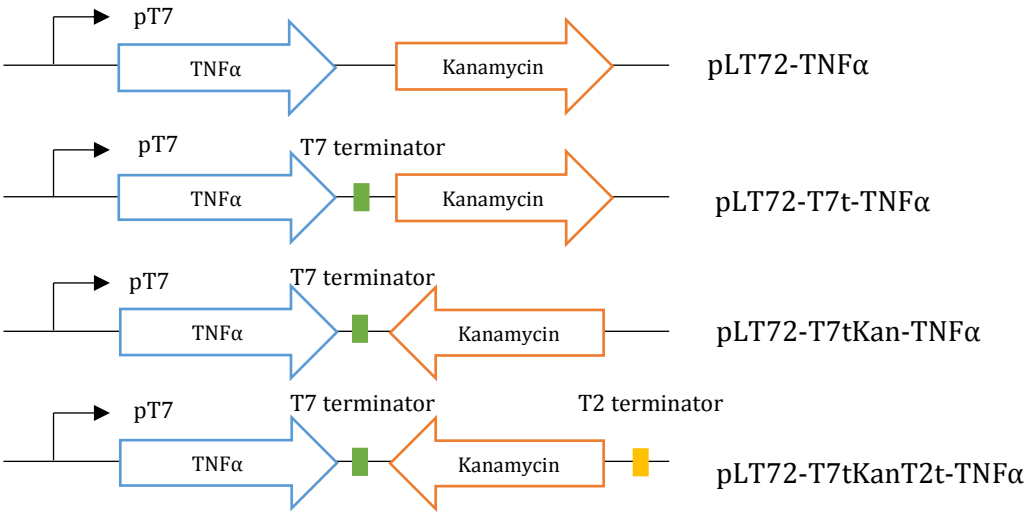


Figure 3.7. Diagram of the new vector designs to minimise the overproduction of APH. The original vector design (pLT72-TNFα) was modified in three ways: with the T7 terminator sequence downstream the sequence coding for the TNFα gene (pLT72-T7t-TNFα); with the T7 terminator sequence and the sequence coding for the kanamycin resistance gene in the opposite orientation to the TNFα gene (pLT72-T7tKan-TNFα); and with the T7 terminator sequence and the sequence coding for the kanamycin resistance gene in reverse orientation and the T2 terminator sequence (pLT72-T7tKanT2t-TNFα).

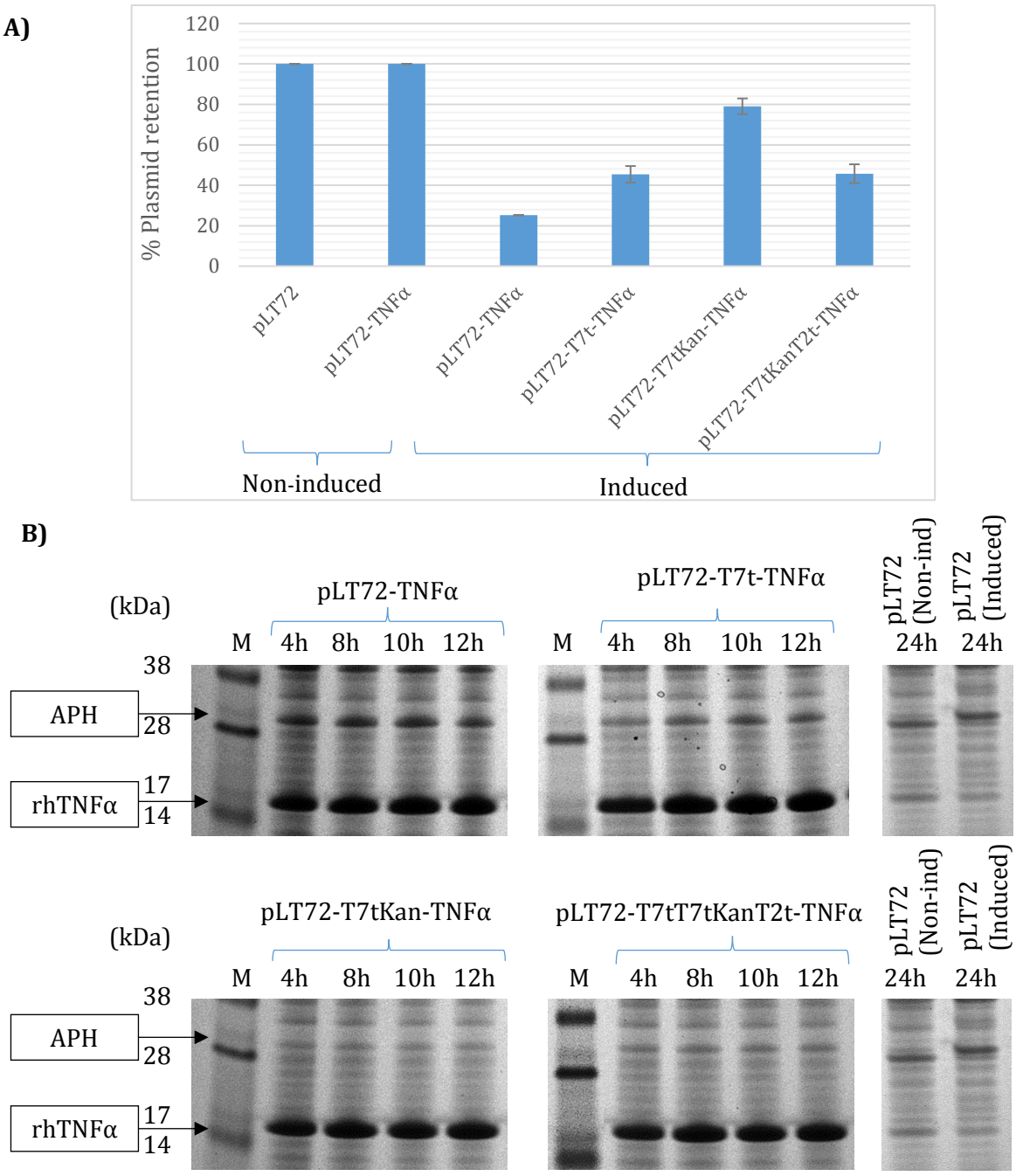


Figure 3.8. Selection of optimal vector design to minimise the overproduction APH. Cultures carrying the empty vector (pLT72) and the different vectors coding for rhTNFα (pLT72-TNFα, pLT2-T7t-TNFα, pLT72-T7tKan-TNFα and pLT72-T7tKanT2-TNFα), were grown at 30 °C under non-induced and induced conditions. Arabinose to a final concentration of 0.02% was added at an OD_{600} ≈ 0.5. Plasmid retention, after 24 hours of growth, was evaluated by replica plating onto non-selective LB agar. Data shown are mean values from two replica flasks, error bars are ±1 standard deviation (A). SDS-PAGE gel showing the accumulation of rhTNFα and APH from whole cell lysates at 4 h, 8 h, 10 h, 12 h and 24 h post-inoculation (B).

Overall, the vector design carrying the T7 terminator sequence and the sequence coding for the kanamycin resistance gene in reverse orientation (pLT72-T7tKan-TNFα) showed a lower accumulation of APH and a higher plasmid retention than other vector designs. This result confirmed that the high plasmid loss observed in previous experiments was caused by the overproduction of both proteins, rhTNFα and APH.

3.2.7 Production of rhTNFα by fed-batch fermentation

The optimal conditions for the production of rhTNFα defined during shake-flask studies were transferred to 5 L scale to produce rhTNFα using a bench-top fermenter. All the fed-batch fermentations were carried out at a temperature of 25 °C and cultures were induced with 0.005% arabinose at an OD_{600} of 0.5, since these conditions were found to be optimal to reduce plasmid loss and favour the accumulation of soluble rhTNFα. The dissolved oxygen was maintained at 20% by continuous increments of the stirrer speed and unless stated otherwise, the pH was kept at 6.8 by the addition 5 M NaOH, as a base, and 5 M HCl, as acid.

Initial fermentations were carried out using a semi-defined culture medium composition obtained from Cobra Biologics supplemented with 2% casamino acids and using the original design of the vector (pLT72- TNFα). Subsequent fermentations were carried out using the improved design of the vector (pLT72-T7tKan-TNFα) with and without casamino acid supplementation. A second semi-defined culture medium composition published in Want *et al.* (2009) was also evaluated. Glycerol was used as a carbon source for both culture media recipes. The addition of the feed solution was started 10 hours after inoculation, upon glycerol consumption, as determined by a previous batch fermentation. The specific growth rate (μ) was maintained at 0.1 h^{-1} being controlled by the exponential feeding regime.

a) Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-TNFα

The production of rhTNFα was evaluated by fed-batch fermentation at 5 L scale. The culture was grown in a semi-defined culture medium obtained from Cobra Biologics supplemented with 2% casamino acids, at a temperature of 25 °C and induced with 0.005% arabinose at an OD_{600} of 0.5.

The culture grew constantly after inoculation reaching a final OD_{600} of 90.8 after 48 hours of growth (**Figure 3.9 A**). Initially, the culture grew exponentially reaching a maximum specific growth rate of 0.5 during the first hours after inoculation. The growth rate was reduced after 6 hours of growth, ranging from 0.05 to 0.1 h^{-1} between 24 and 48 hours post-inoculation. The CFU increased during the first hours of growth, followed by a decrease after induction, from $8 \cdot 10^8$ CFU · mL^{-1} to 10^8 CFU · mL^{-1} (**Figure 3.9 B**). The CFU recovered after 24 hours, showing a continuous increase between 24 and 48 hours post-inoculation.

Plasmid retention was almost 100% during the first hours of growth, followed by a decrease between 11 and 24 hours, from 97% of the cells being plasmid-positive to only 42% (**Figure 3.9 B**). After 24 hours, the culture was overtaken by plasmid-free bacteria and plasmid retention was further decreased. Only 2% of the bacteria still carried the plasmid 48 hours post-inoculation.

The concentration of rhTNFα increased over the time after induction reaching a maximum of 25% of the total cell protein (TCP) between 24 and 32 hours of growth, however, it decreased to 20% at the end of the fermentation (**Figure 3.10 A**). The biomass increase observed between 32 and 48 hours of growth did not correspond with an increase in the concentration of rhTNFα, as a result of the low plasmid retention observed at the end of the fermentation. The low plasmid retention and lower accumulation of rhTNFα, caused a loss of productivity which affected the final protein yield. The production of APH can be noticed after 9 hours of growth since the original plasmid vector (pLT72-TNFα) was used for this fermentation.

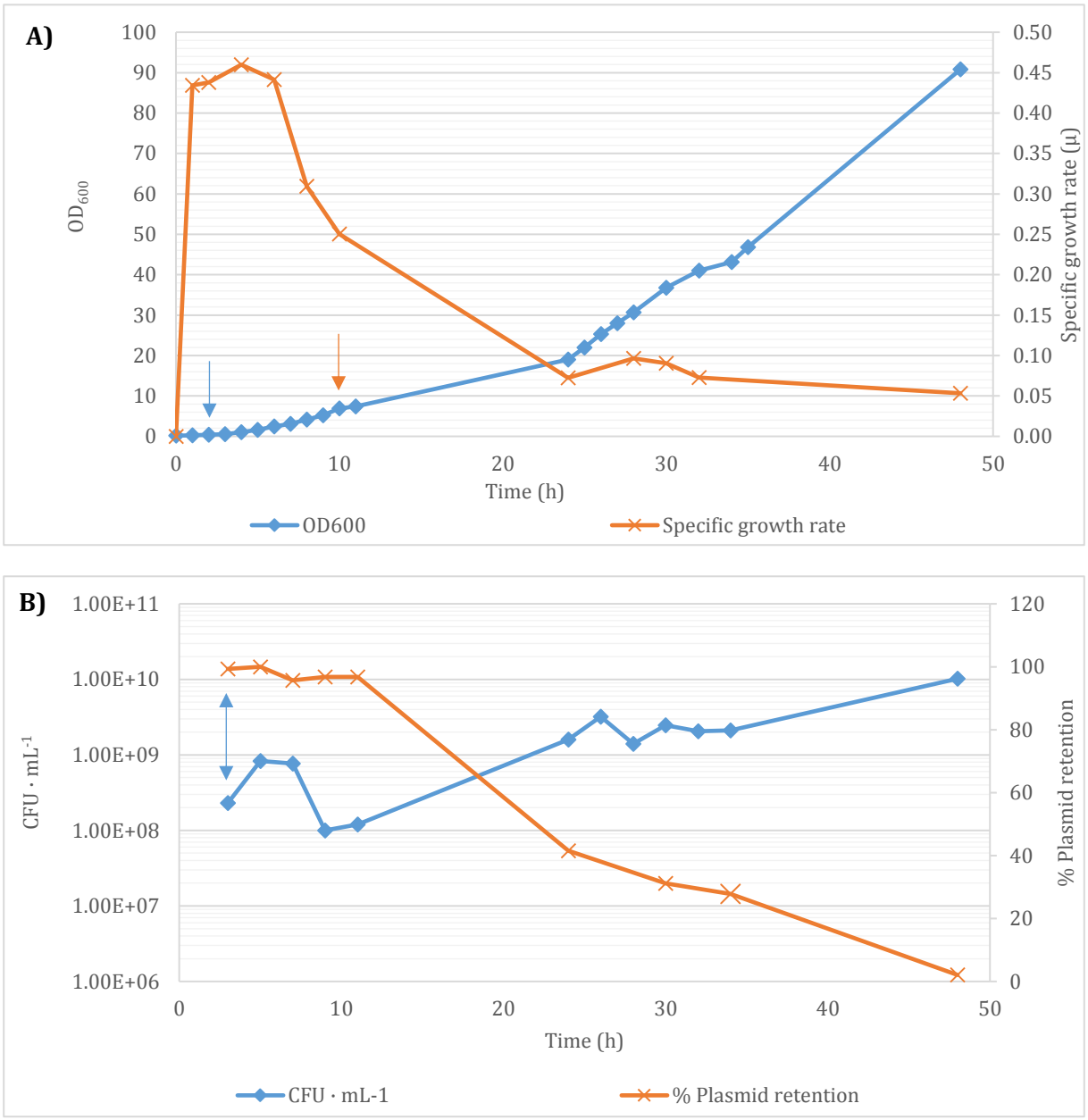


Figure 3.9. Fed-batch fermentation for the production of rhTNFα using BL21-T7 pLT72-TNFα. *E. coli* BL21-T7 carrying the vector coding for rhTNFα (pLT72-TNFα) was grown at 25 °C and induced with 0.005% arabinose at an OD_{600} ≈ 0.5 (blue arrows). The OD_{600} was measured at intervals post-inoculation and the specific growth rate was calculated. The addition of feed solution was started 10 hours post-inoculation (orange arrow) (A). Serial dilutions of culture samples post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction. Colonies were replica plated onto kanamycin-supplemented LB agar to reveal the plasmid retention (B).

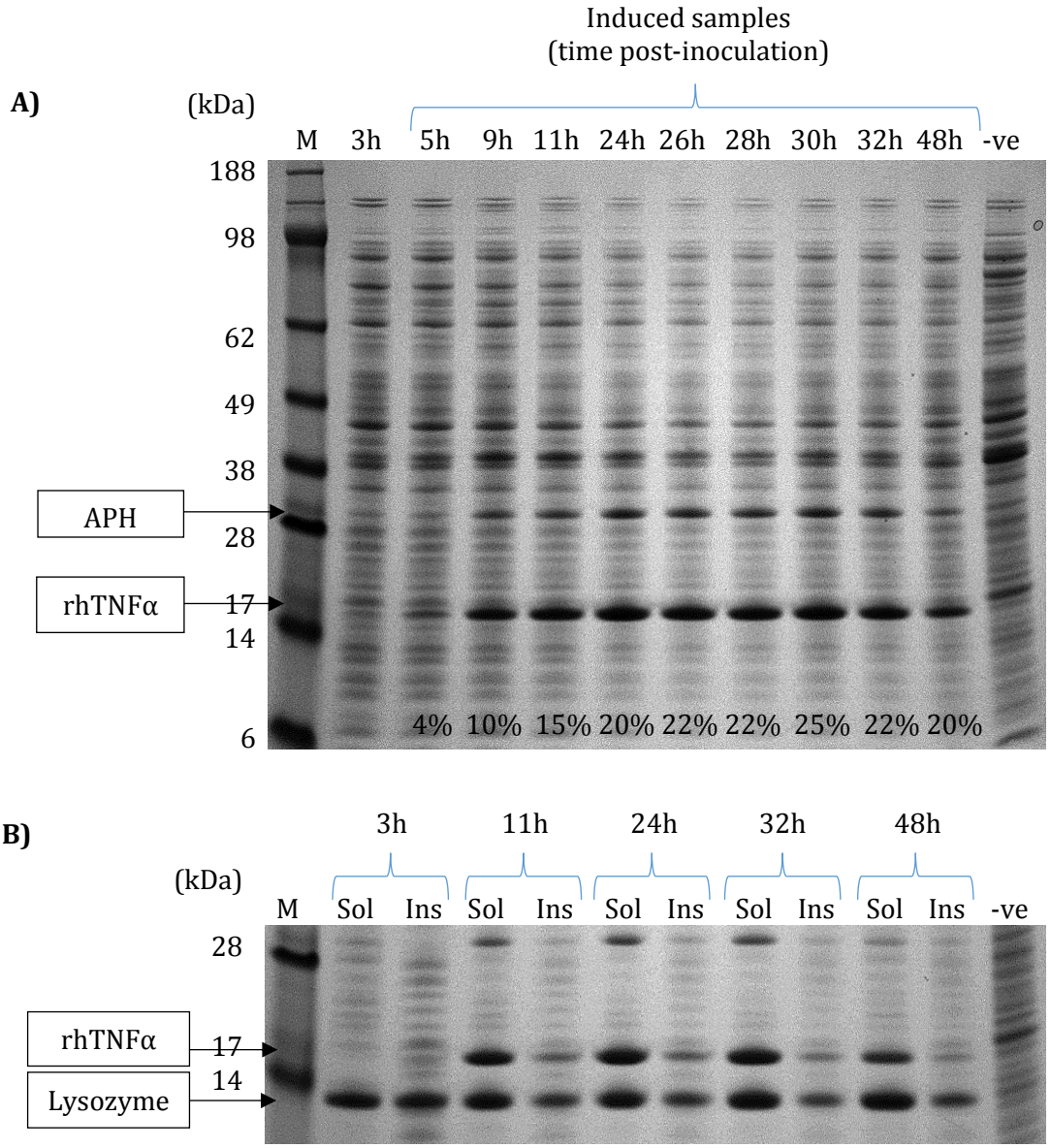


Figure 3.10. Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-TNFα. The production of rhTNFα from whole cell lysates obtained before induction (3h) and after induction (5h, 9h, 11h, 24h, 26h, 28h, 30h, 32h and 48h post-inoculation). The quantity of rhTNFα is expressed as a percentage of whole cell protein at the bottom of the gel (A). Samples obtained from whole cell lysates were disrupted to obtain the soluble (sol) and insoluble (ins) cell fractions facilitated by the addition of lysozyme, to reveal the proportion of soluble and insoluble rhTNFα (B). The accumulation of aminoglycoside-3'-phosphotransferase (APH) can also be observed in samples taken post-induction. A sample obtained from a culture carrying the empty vector (pLT72) after 24 hours of growth was used as a negative control.

rhTNFα was mainly accumulated in the soluble fraction, being only 8% of TCP accumulated in the insoluble fraction, concluding that the optimisation of the fermentation conditions successfully yielded a soluble product (**Figure 3.10 B**). However, further studies need to be carried out to improve plasmid retention during fed-batch fermentations. The rhTNFα in the soluble fraction was purified and used to evaluate the activity of the rhTNFα (**Sections 3.2.9 and 3.2.10**).

b) Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7tKan-TNFα

The previous fermentations showed a high accumulation of soluble rhTNFα, however, the final yield of rhTNFα was affected by the severe plasmid loss observed at the end of the fermentation. The improved vector design containing the T7 terminator and the gene coding for the kanamycin resistance gene in reverse orientation (pLT72-T7tKan-TNFα) was used with the aim to improve plasmid retention and minimise the overproduction of APH. Cultivation conditions were the same as described for the previous fermentation.

The culture grew steadily reaching a final OD_{600} of 68 after 48 hours of growth (**Figure 3.11 A**). The CFU also increased over the time, with small fluctuations over the course of the fermentation, reaching $1.8 \cdot 10^{10}$ CFU $\cdot$ mL^{-1} after 48 hours of growth. Plasmid retention was almost 100% (**Figure 3.11 B**), in contrast with the previous fermentation carried out with the original vector design (pLT7-TNFα), where only 2% of the cells were plasmid-positive at the end of the fermentation (**Figure 3.9 B**).

The concentration of rhTNFα increased after induction, reaching an accumulation of 20% of total cell protein after 48 hours of growth. The protein was accumulated mainly in the soluble fraction, and less than 20% was accumulated in an insoluble form (**Figure 3.12**). The use of the improved vector design helped to minimise plasmid loss during fermentation, avoiding the overgrowth of plasmid-free cells, and therefore, non-productive, at the last stages of the fermentation process.

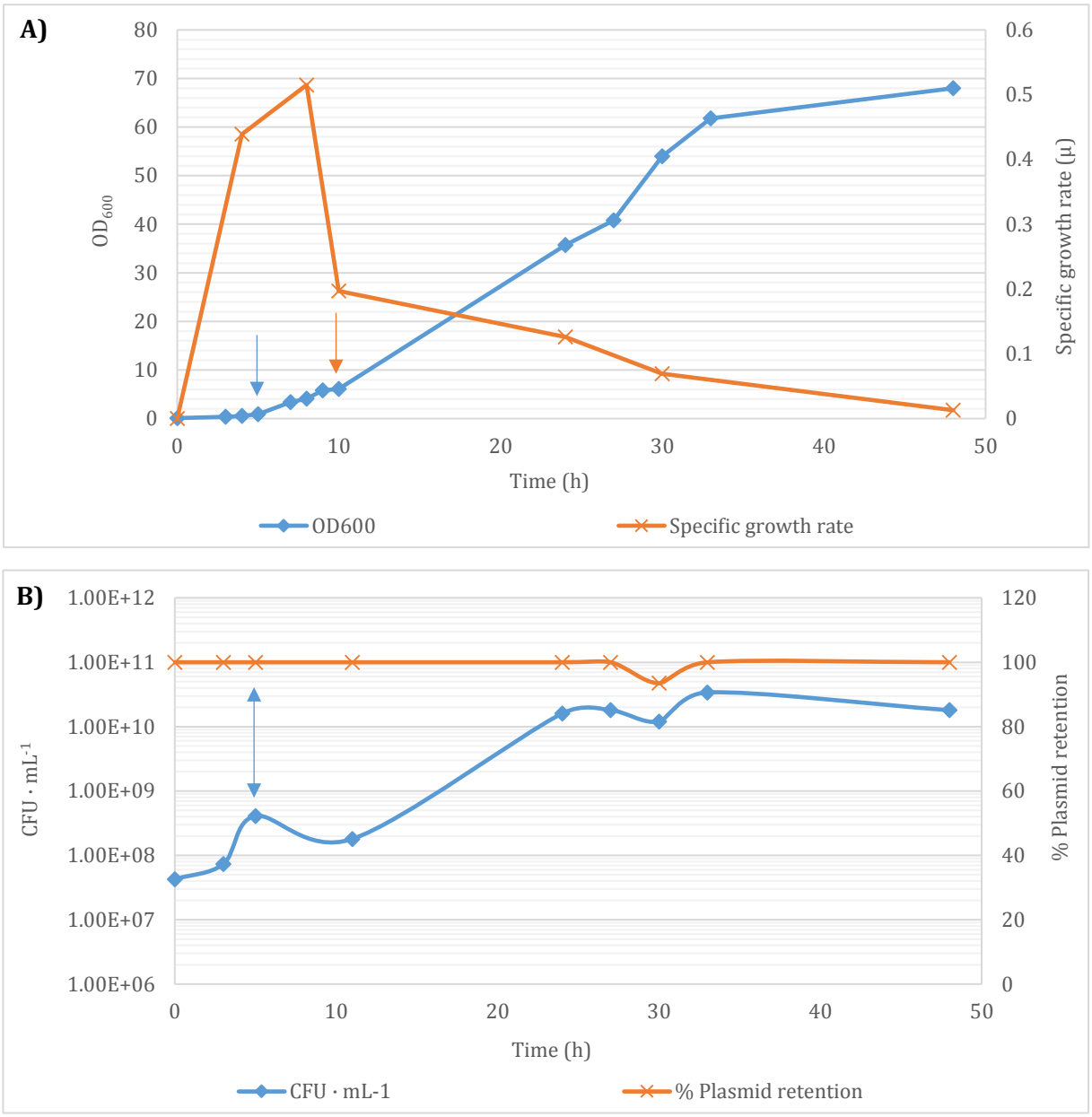


Figure 3.11. Fed-batch fermentation for the production of rhTNFα using BL21-T7 pLT72-T7tKan-TNFα. *E. coli* BL21-T7 carrying the vector coding for rhTNFα (pLT72-T7tKan-TNFα) was grown at 25 °C and induced with 0.005% arabinose at an OD_{600} ≈ 0.5 (blue arrows). The OD_{600} was measured at intervals post-inoculation and the specific growth rate was calculated. The addition of feed solution was started 10 hours post-inoculation (orange arrow) (A). Serial dilutions of culture samples post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction. Colonies were replica plated onto kanamycin-supplemented LB agar to reveal the plasmid retention (B).

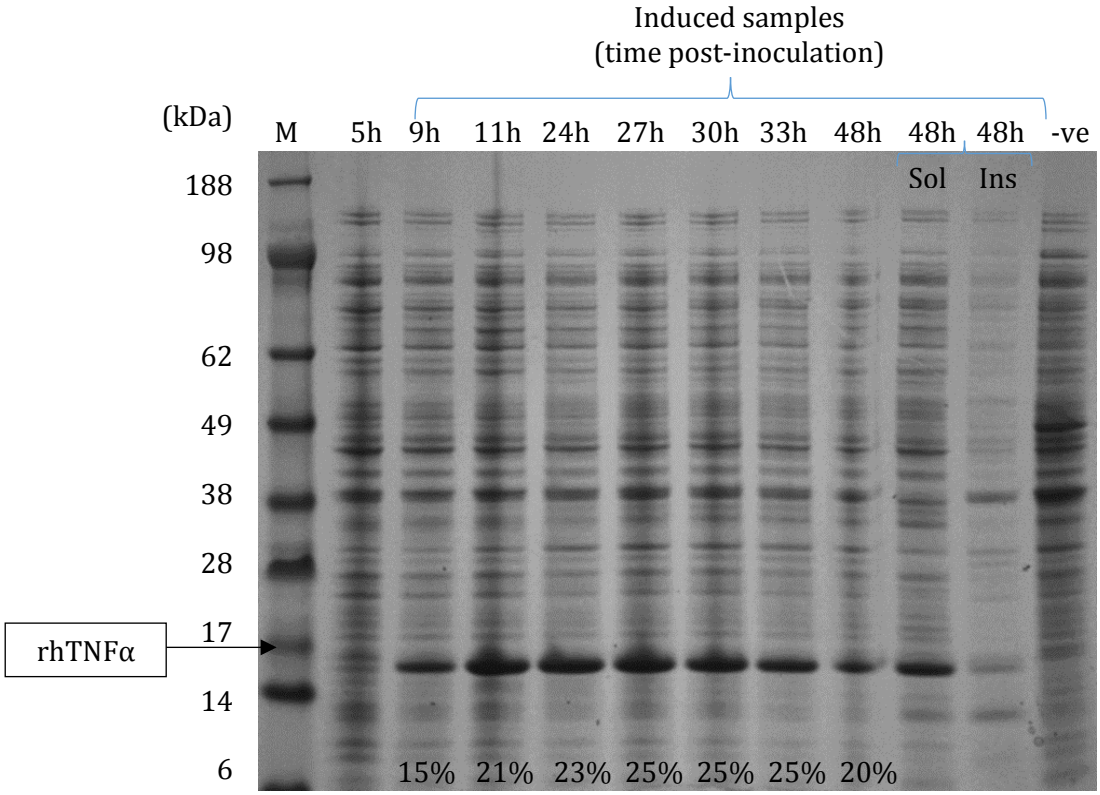


Figure 3.12. Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7tKan-TNFα. The production of rhTNFα from whole cell lysates obtained before induction (5h) and after induction (9h, 11h, 24h, 27h, 30h, 33h and 48h). Samples obtained from whole cell lysates were disrupted by sonication to obtain the soluble (sol) and insoluble (ins) protein fractions, to reveal the proportion of soluble to insoluble rhTNFα. A sample obtained from a culture carrying the empty vector (pLT72) after 24 hours of growth was used as a negative control. The quantity of rhTNFα is expressed as a percentage of whole cell protein at the bottom of the gel.

c) Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7Kan-TNFα without the addition of casamino acids

In order to comply with the current Good Manufacturing Practices (cGMP), which provides the minimum requirements that a manufacturer must meet to ensure that the product is of high quality and do not represent any health risk to the consumer, any animal-derived products need to be excluded from the culture medium. Casamino acids are a mixture of amino acids and small peptides obtained from acid hydrolysis of casein and therefore, they should be removed from the medium final formulation.

With the objective of developing a fermentation process fully compliant with the cGMP guidelines, casamino acids were removed from the semi-defined culture medium obtained from Cobra Biologics. The medium was supplemented with 14 $g \cdot L^{-1}$ of ammonium sulphate and 0.3 $g \cdot L^{-1}$ of calcium chloride, since the concentration of ammonium and calcium could be limiting due to the absence of casamino acids. Cultivation conditions were the same as described in **Section 3.2.7 B**.

The biomass concentration increased steadily over the time reaching a final OD_{600} of 72. However, no great increase in the cell biomass was observed between 32 and 48 hours and the growth rate decreased from 0.1 h^{-1} to 0.01 h^{-1} between 24 and 48 hours post-inoculation (**Figure 3.13 A**). The CFU also increased over time reaching $5.1 \cdot 10^{10}$ $CFU \cdot mL^{-1}$. The plasmid retention was almost 100% during the first hours post-inoculation, however, it decreased to 35% after 24 hours, and less than 10% of the cells still carried the plasmid at the end of the fermentation (**Figure 3.13 B**).

The concentration of rhTNFα increased over the time after induction, reaching a maximum of 26% of TCP. A decrease in the concentration of rhTNFα was observed between 32 and 48 hours of growth, from 25% to 18%, which could be the result of proteolysis or the high plasmid loss observed at the end of the fermentation. The concentration of rhTNFα in the soluble fraction was lower than in previous fermentations, and only 50% of the rhTNFα produced was found in the soluble fraction (**Figure 3.14**).

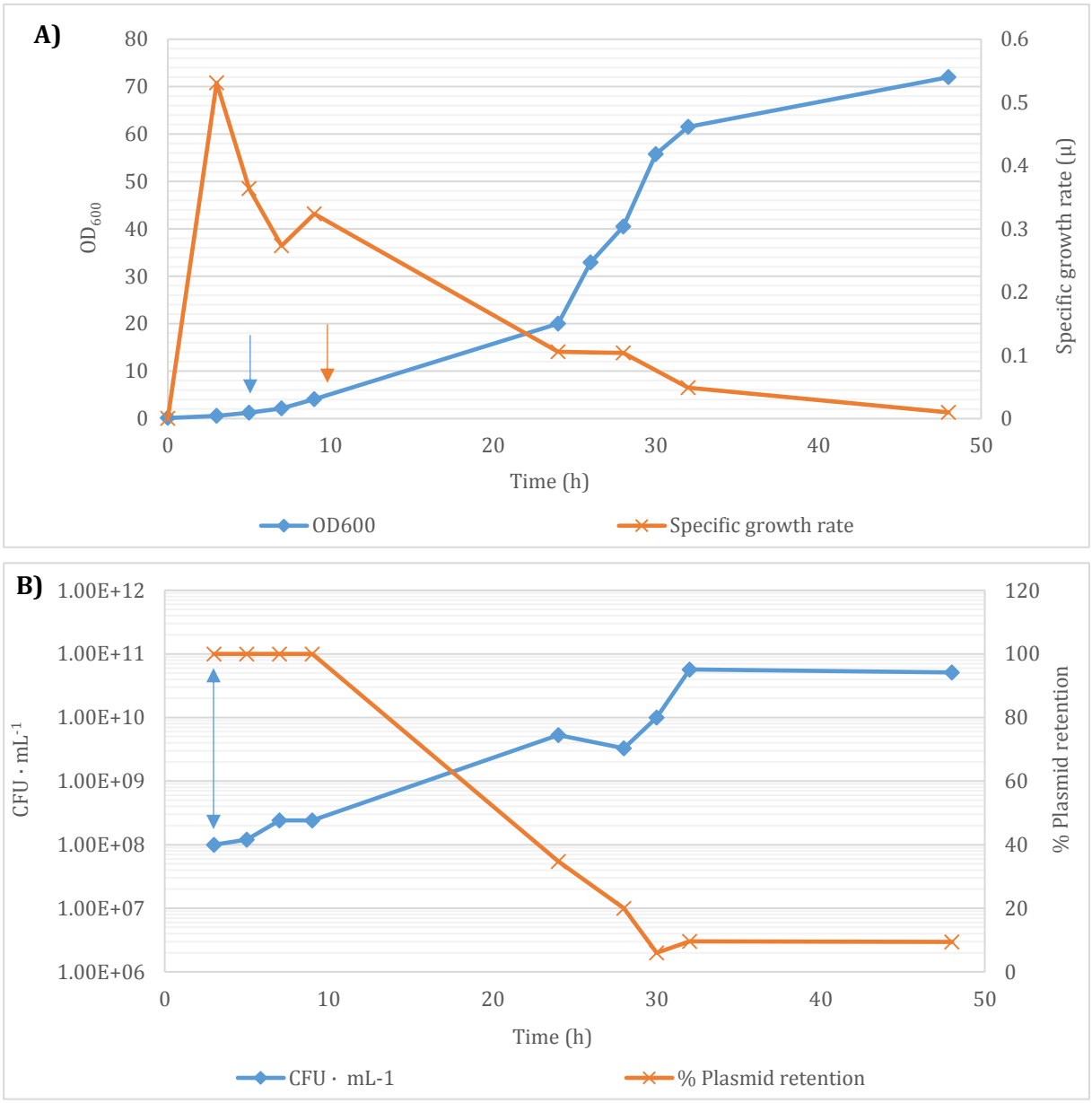


Figure 3.13. Fed-batch fermentation for the production of rhTNFα using BL21-T7 pLT72-T7tKan-TNFα without casamino acids. *E. coli* BL21-T7 carrying the vector coding for rhTNFα (pLT72-T7tKan-TNFα) was grown at 25 °C and induced with 0.005% arabinose at an OD_{600} ≈ 0.5 (blue arrows). The OD_{600} was measured at intervals post-inoculation and the specific growth rate was calculated. The addition of feed solution was started 10 hours post-inoculation (orange arrow) (A). Serial dilutions of culture samples post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction. Colonies were replica plated onto kanamycin-supplemented LB agar to reveal the plasmid retention (B).

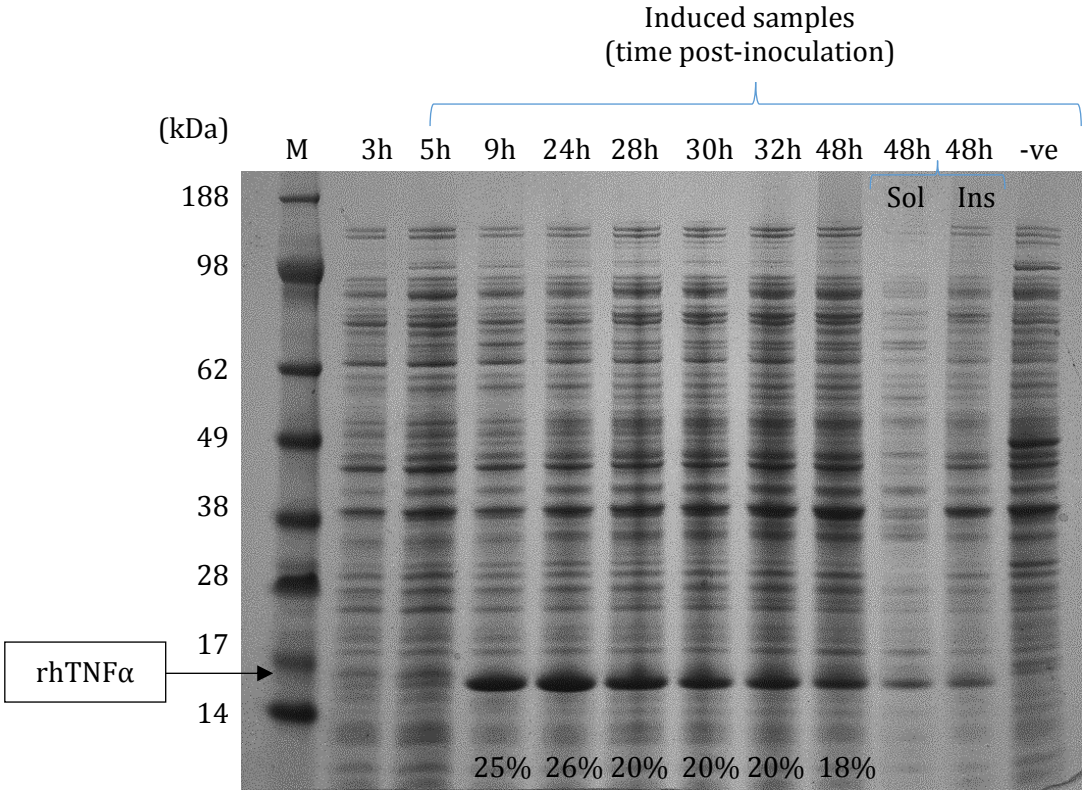


Figure 3.14. Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7tKan-TNFα without casamino acids. The production of rhTNFα from whole cell lysates obtained before induction (3h) and after induction (5h, 9h, 24h, 28h, 30h, 32h and 48h). Samples obtained from whole cell lysates were disrupted by sonication to obtain the soluble (sol) and insoluble (ins) protein fractions, to reveal the proportion of soluble to insoluble rhTNFα. A sample obtained from a culture carrying the empty vector (pLT72) after 24 hours of growth was used as a negative control. The quantity of rhTNFα is expressed as a percentage of whole cell protein at the bottom of the gel.

d) Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7tKan-TNFα with an alternative culture medium formulation

Previous fermentations showed an improvement of plasmid retention by using the improved vector design, however, a dramatic plasmid loss was still observed when casamino acids were removed from the culture medium. As casamino acids are not compliant with the cGMP guidelines, a different semi-defined culture medium composition reported to be used by the other biopharmaceutical companies (Hodgson *et al.*, 2011), was evaluated with the aim to develop a fully compliant GMP fermentation process for the production rhTNFα.

The culture was grown in a semi-defined culture medium obtained from Want *et al.* (2009), using glycerol as carbon source. The culture was grown at 25 °C and induced with 0.005% arabinose at an OD_{600} of 0.5. The pH was maintained at 7.0 by the addition of 20% NH_4OH, as a base, and 5 M HCl, as acid. The addition of the feed solution was started after 10 hours of inoculation and the specific growth rate was maintained at 0.1 h^{-1} by the use of an exponential feeding profile.

The culture grew constantly after induction reaching a final OD_{600} of 111 after 48 hours of growth (**Figure 3.15 A**). The growth rate increased during the first hours of growth reaching a maximum of 0.8 h^{-1} and decreasing drastically to 0.1 h^{-1}, after 24 hours of growth. Cell culturability was high through the fermentation, and the CFU increased steadily with the increase cell biomass. The plasmid retention was almost 100% through the fermentation (**Figure 3.15 B**).

The concentration of rhTNFα increased over time after induction, reaching a maximum of 30% of the total cell protein between 24 and 30 hours of growth. At the end of the fermentation, rhTNFα constituted 25% of the TCP, being primarily accumulated in the soluble fraction (**Figure 3.16**).

Overall, this fermentation process successfully generated a greater yield of rhTNFα than previous fermentations, being the majority accumulated in a soluble form, minimising the plasmid loss by using a culture medium compliant with cGMP guidelines.

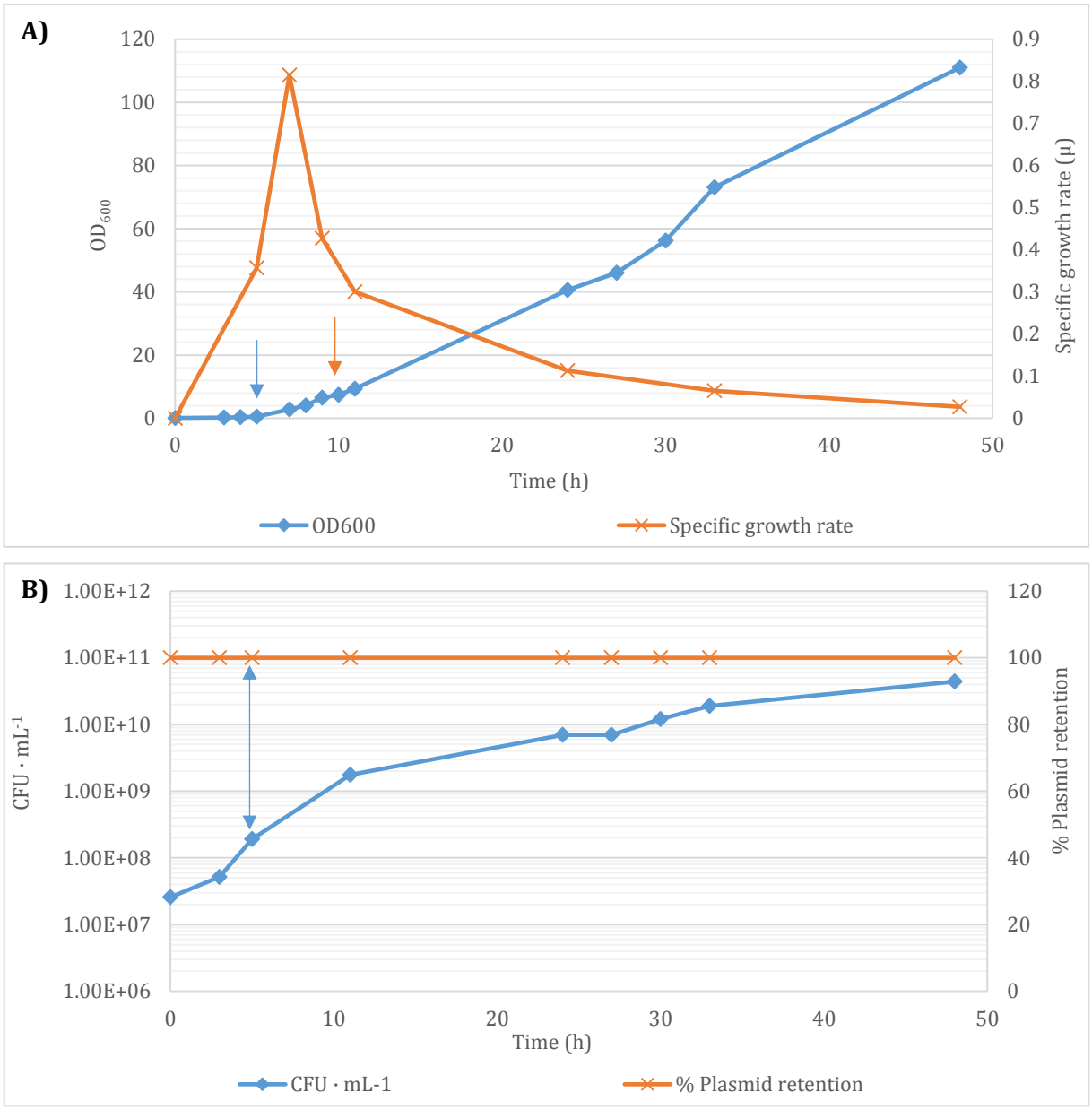


Figure 3.15. Fed-batch fermentation for the production of rhTNFα using BL21-T7 pLT72-T7tKan-TNFα with the alternative culture medium formulation. *E. coli* BL21-T7 carrying the vector coding for rhTNFα (pLT72-T7tKan-TNFα) was grown at 25 °C and induced with 0.005% arabinose at an OD_{600} ≈ 0.5 (blue arrows). The OD_{600} was measured at intervals post-inoculation and the specific growth rate was calculated. The addition of feed solution was started 10 hours post-inoculation (orange arrow) (A). Serial dilutions of culture samples post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction. Colonies were replica plated onto kanamycin-supplemented LB agar to reveal the plasmid retention (B).

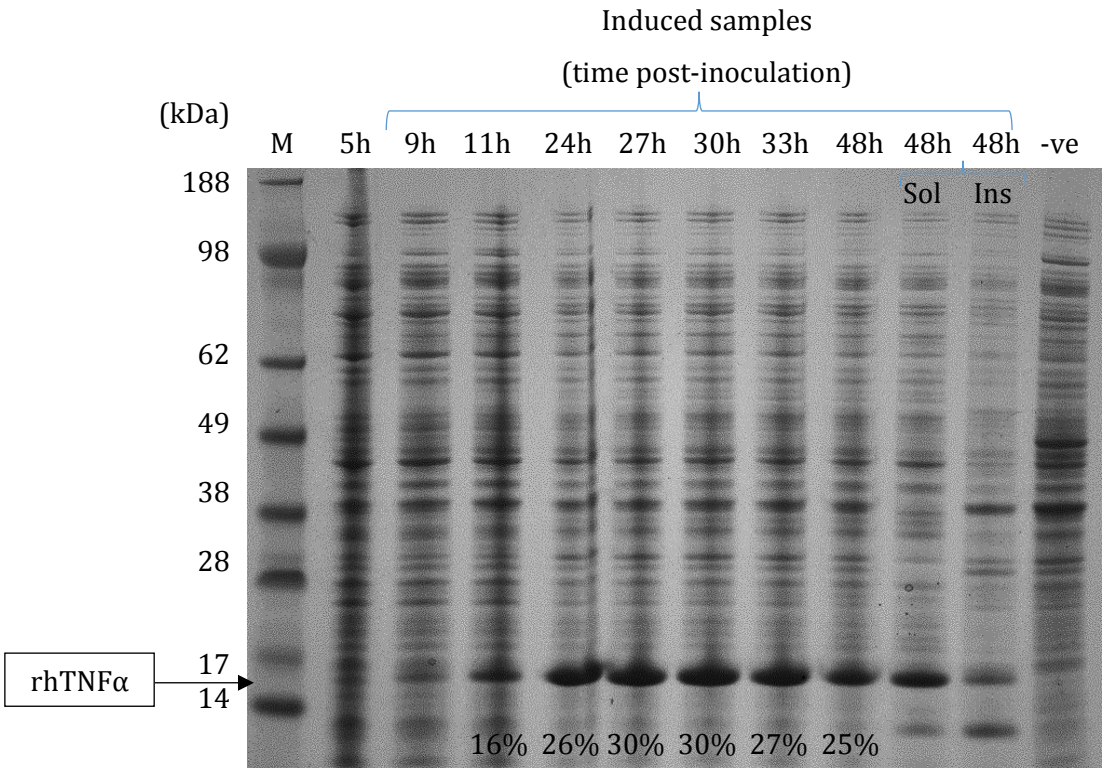


Figure 3.16. Production of rhTNFα by fed-batch fermentation using BL21-T7 pLT72-T7tKan-TNFα with the alternative culture medium formulation. The accumulation of rhTNFα from whole cell lysates obtained before induction (5h) and after induction (9h, 11h, 24h, 27h, 30h, 33h and 48h). Samples obtained from whole cell lysates were disrupted to obtain the soluble (sol) and insoluble (ins) cell fractions facilitated by the addition of lysozyme, to reveal the proportion of soluble to insoluble rhTNFα. A sample obtained from a culture carrying the empty vector (pLT72) after 24 hours of growth was used as a negative control. The quantity of rhTNFα is expressed as a percentage of whole cell protein at the bottom of the gel.

3.2.8 Evaluation of results obtained from fermentation studies

A summary of the results obtained from fed-batch fermentations experiments can be found in **Table 3.1**. Results shown in this table are the result of single fermentation experiments. Data corresponding to OD_{600} and wet cell weight (WCW) was measured at the end of each fermentation. For fermentation 1, the dry cell weight (DCW) was measured and compared to the WCW to obtain the ratio of DCW:WCW, which it was found to be approximately 1:4. This ratio was used to calculate the DCW for fermentations 2, 3 and 4. The yield of rhTNFα was determined using densitometry from SDS-PAGE gels of whole cell lysates and calculated to obtain the final rhTNFα yield for total biomass generated at harvest.

Initial fermentations showed a high biomass production, with a final OD_{600} of 90.8, equivalent to a wet cell weight of 85 $g \cdot L^{-1}$, and a high accumulation of rhTNFα, accounting for 20% of the total cell protein content. The accumulation of rhTNFα, corresponded to a final yield of 3.82 $g \cdot L^{-1}$ of rhTNFα, the majority being accumulated in a soluble form. However, low plasmid retention was observed with only 2% of the cells being plasmid positive after 48 hours of growth, affecting the final rhTNFα yield. It was considered that the suboptimal plasmid design allowed the read-through of the T7 RNA polymerase, which resulted in the overexpression of the kanamycin resistance gene, could be a contributing factor to the high plasmid loss observed during fed-batch fermentations (**Table 3.1, fermentation 1**). This problem was partially overcome by the use of an improved plasmid design containing the T7 terminator sequence and by changing the orientation of the kanamycin gene. By using the improved vector design, the plasmid retention was maintained at 100% through the whole fermentation process achieving a similar accumulation of rhTNFα to previous fermentations. In contrast with the previous fermentation, the biomass production was slightly lower, reaching a final OD_{600} of 68, equivalent to 106 $g \cdot L^{-1}$ of WCW (**Table 3.1, fermentation 2**). The large differences obtained between the OD_{600} and the

Table 3.1. Summary of results obtained at harvest from fermentation studies carried out for the production of rhTNFα.

	Fermentation 1	**Fermentation 2**	**Fermentation 3**	**Fermentation 4**
E. coli strain	BL21-T7	BL21-T7	BL21-T7	BL21-T7
Plasmid name	pLT72-TNFα	pLT72- T7tKan-TNFα	pLT72-T7tKan-TNFα	pLT72-T7tKan-TNFα
Culture medium	Semi-defined medium (Cobra Biologics)	Semi-defined medium (Cobra Biologics)	Semi-defined medium (Cobra Biologics)	Semi-defined medium (Want *et al.*, 2009)
Casamino acids	Yes	Yes	No	No
Temperature	25 °C	25 °C	25 °C	25 °C
Arabinose concentration	0.005%	0.005%	0.005%	0.005%
Induction point	$OD_{600} \approx 0.5$	$OD_{600} \approx 0.5$	$OD_{600} \approx 0.5$	$OD_{600} \approx 0.5$
Plasmid retention	2%	100%	9.4%	100%
Final OD_{600}	90.8	68	72	111
Wet cell weight (WCW)	85 g · L^{-1} [1]	106 g · L^{-1}	96 g · L^{-1}	133 g · L^{-1}
Dry cell weight (DCW)	20.4 g · L^{-1}	26.5 g · L^{-1} [2]	29.4 g · L^{-1} [2]	33.25 g · L^{-1} [2]
% of rhTNFα of TCP[3]	20 %	20%	18%	25%
% of soluble rhTNFα	92%	80%	50%	70%
rhTNFα yield	3.82 g · L^{-1}	1.92 · g L^{-1}	1.84 · g L^{-1}	5.35 g · L^{-1}

[1] WCW could be lower than expected due to cell paste loses due to resuspension and washing steps of the bacterial pellets.
[2] Dry cell weight calculated from wet cell weight using a 1:4 ratio (1 g · L^{-1} of wet cell weight equivalent to 0.25 g · L^{-1} of dry cell weight).
[3] Percentage of total cell protein (TCP) corresponding to rhTNFα.

WCW could be the result of cell filamentation, evaluated by microscopy, which may have resulted in the underestimation of biomass based on the OD_{600} measurement (Vizcaino Caston, 2012). The lower WCW obtained for fermentation 1 was probably caused by the loss of cell paste during the resuspension and washing of the cell pellet; step that was not carried out for fermentation 2, 3 and 4.

As previous fermentations were carried out using a culture medium supplemented with casamino acids, a fed-batch fermentation in absence of those was also carried out to provide Cobra Biologics with a cGMP compliant fermentation process. Similar to previous fermentation, a final OD_{600} of 72 was obtained, equivalent to 96 $g \cdot L^{-1}$ of WCW. Unfortunately, a lower yield of rhTNFα was obtained, accounting for 18% of the total cell protein, which corresponded to only 1.84 $g \cdot L^{-1}$ rhTNFα. Also a high plasmid loss was observed and less than 10% of the cells still carried the plasmid after 48 hours of growth (**Table 3.1, fermentation 3**).

As it became obvious than the addition of casamino acids provided an essential component to the medium, a new medium formulation extracted from Want *et al*. (2009) was evaluated for the production of rhTNFα. The evaluation of a new culture medium was found to be easier than attempting the supplementation of Cobra's culture medium recipe with different amino acid, vitamin or trace element solutions, an extremely time-consuming approach. With the new medium formulation, a higher biomass was obtained, with a final OD_{600} of 111. High plasmid retention was also achieved with 100% of the cells being plasmid-positive after 48 hours of growth. A higher accumulation of rhTNFα was also obtained, accounting for 25% of the total cell protein being rhTNFα (**Table 3.1, fermentation 4**).

In general terms, all the fed-batch fermentations carried out showed a high concentration of rhTNFα in a soluble form. The new medium recipe evaluated showed the most promising results, with a final yield of rhTNFα of 5.35 $g \cdot L^{-1}$, being majorly accumulated in a soluble form. The production of rhTNFα was similar to protein titres obtained by other manufacturing companies,

such as Fujifilm Diosynth, that for a target OD_{600} of 100 reported a final yield of rhTNFα ranging between 1.5 and 6.5 $g \cdot L^{-1}$, depending on the expression system used (Hodgson *et al.*, 2011).

3.2.9 Purification of rhTNFα produced by fed-batch fermentation

The purification of rhTNFα was carried out by Dr Nicola Barison at Cobra Biologics. A proprietary purification protocol was used for the purification of rhTNFα. Briefly, the cell paste obtained from fermentation studies was resuspended and cells were disrupted by the use of a high-pressure cell disruption system. rhTNFα was purified from the soluble fraction. The final product presented a purity greater than 95%, as quantified by densitometry from SDS-PAGE gels (**Figure 3.17**).

3.2.10 Evaluation of the activity of rhTNFα produced by fed-batch fermentation

The final goal of any fermentation process is the production of high quantities of a soluble and active product, and protein solubility is often a good indication of activity. However, this statement is not always true, and recombinant proteins can be accumulated in a soluble form but with an incorrect conformation, which limits or abolishes its activity. The C3H mouse fibrosarcoma cell line L929, a cell line sensitive to the activity of TNFα, was used to evaluate the activity of rhTNFα produced by fed-batch fermentation.

rhTNFα produced by fed-batch fermentation, corresponding to the bacterial paste purified by Dr Nicola Barison at Cobra Biologics, was evaluated for activity (**Table 3.1, fermentation 1**). The TNFα cytotoxicity assay is based on the ability of L929 cells to uptake a staining solution containing crystal violet. Cells affected by the activity of rhTNFα are unable to uptake the staining solution, which is removed during the washing steps. In contrast, cells not exposed to rhTNFα are able to uptake the staining solution, which subsequently released upon treatment with SDS. The uptake of the staining solution was quantified by measuring the absorbance at 580 nm.

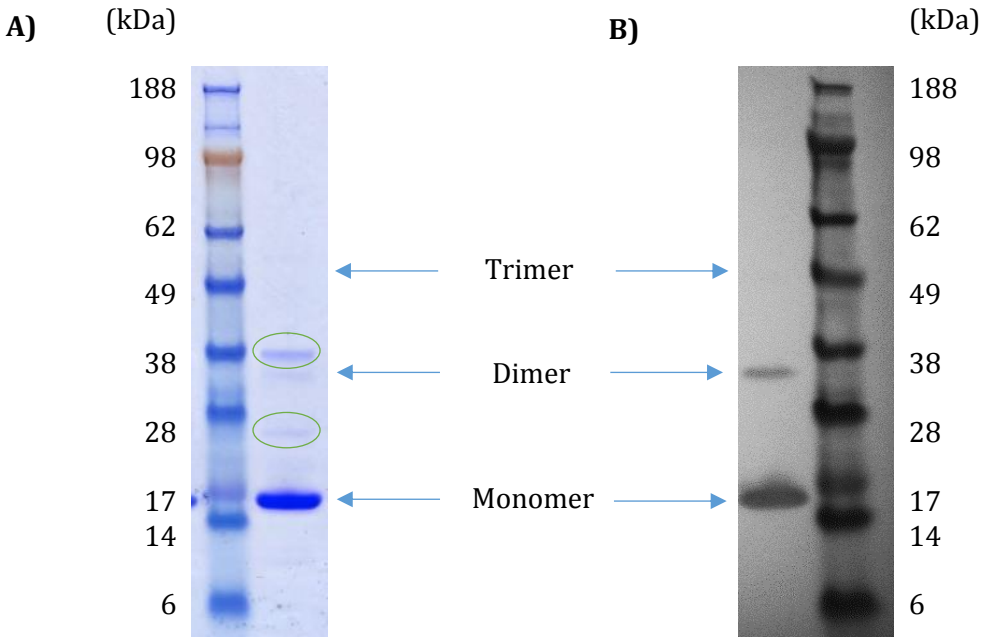


Figure 3.17. Final material obtained after the purification of rhTNFα. The rhTNFα was accumulated mainly in the monomeric form. Traces of the rhTNFα dimer and trimer can also be detected by SDS-PAGE (A) and western blot (B). Traces of other contaminant proteins can also be observed (green circles). The final product presented a purity greater than 95%. The western blot was developed using an anti-TNFα antibody, which confirmed the identity of the monomer, dimer and trimer forms of the rhTNFα produced by fed-batch fermentation.

The activity of rhTNFα was calculated as the percentage of cytotoxicity, which is calculated by comparing the absorbance of the cells exposed to the activity of rhTNFα and the absorbance of the cells exposed to the buffer, which is used as blank for the cytotoxic assay (0% cytotoxicity). In repeated assays following the protocol described by McGee and Clemens (1994), rhTNFα samples generated a 50% of cytotoxicity for L929 cells at a concentration of 0.35 ng · mL^{-1}, being in the concentration range of 0.05 to 20 ng · mL^{-1} where TNFα exerts its biological activity for most *in vitro* applications. The specific activity was found to be $2.8 \cdot 10^6$ U · mg^{-1}, being one unit of specific activity defined by the quantity of TNFα that produces 50% of the lysis of L929 cells. It has been published that the specific activity of TNFα is in the range of 10^6 U · mg^{-1} to 10^8 U · mg^{-1} (Meager *et al.,* 1989), suggesting that the rhTNFα produced by fed-batch fermentation fulfilled the expected specific activity for this protein. Overall, the result of the cytotoxicity assay showed that the optimisation of the fermentation conditions has led to the successful production of soluble and active rhTNFα.

3.3 Discussion

3.3.1 Metabolic stress induced by recombinant protein production in E. coli

The optimisation of the conditions for recombinant proteins production (RPP) is usually a challenging process, since nature has not optimised *E. coli* as a protein factory for the pharmaceutical industry (Hoffmann and Rinas, 2004). The "Holy Grail" of RPP is still the availability of generic approaches that could be applied to all the different types of recombinant proteins (Sevastsyanovich *et al.*, 2010). However, after more than 30 years of industrial production of biopharmaceuticals, there is still no single approach that can be applied to all protein products. Different variables are routinely evaluated, including the host strain, the expression vector, medium composition, induction protocol, cultivation temperature or chaperones co-expression, among others (Sevastsyanovich *et al.*, 2010).

Plasmid replication and gene expression often interfere with the physiological balance of host cells, which often results in a high metabolic burden imposed on the host cells. Metabolic burden can be defined as the portion of resources that it is redirected from the host's metabolism – either in the form of energy or metabolic resources - for maintenance and expression of the foreign DNA, RNA or proteins (Glick, 1995; Hoffmann and Rinas, 2004). The extent of the metabolic burden will depend on the size and copy number of the plasmid vector, the levels of production of the recombinant protein of interest, the composition of the growth medium and the metabolic state of the cells (Glick, 1995). In consequence, different conditions need to be evaluated in a trial-and-error approach in order to reduce the metabolic burden and cell stress with the aim to favour the successful production of recombinant proteins.

3.3.2 Plasmid maintenance and overexpression of APH

The metabolic burden is not only the result of high levels of protein synthesis, but it can also be induced by DNA replication. However, it is considered that the metabolic burden associated with plasmid maintenance is often negligible for an average plasmid size (5-10 kbp), and alterations associated to plasmid-positive cells have been mainly described when high copy number plasmids are used. However, cultures carrying the empty vector under induced conditions did show signs of growth inhibition when compared to non-induced cultures (**Figure 3.1**). This effect could not be explained as part of the physiological effects attributed to the overexpression of rhTNFα, since this vector does not contain the sequence coding for the recombinant protein and instead, it could be the result of the constitutive expression of the antibiotic resistance gene (Hoffmann and Rinas, 2004).

The growth arrest, loss of cell culturability and plasmid loss observed for cultures carrying the empty vector was caused by the read-through of the T7 RNA polymerase due to the lack of terminator sequences downstream the rhTNFα gene (**Figures 3.1**). The overexpression of the

kanamycin resistance gene resulted in a high production of the aminoglycoside-3'-phosphotransferase (APH), which accounted for 10% of the total cell protein (**Figures 3.2**). The addition of terminator sequences and the change in the orientation of the kanamycin resistance gene showed a four-fold improvement in plasmid retention and a considerable reduction in the production of APH (**Figure 3.8**). This highlights the importance of the optimal plasmid design in order to reduce unnecessary extra metabolic load to the host cell, which may have a debilitating effect on the host organism, hampering recombinant protein production (Glick, 1995).

3.3.3 Inhibition of growth, loss of cell viability and metabolic burden

The inhibition of growth, and the less documented loss of cell culturability, is often the result of the high cell stress upon RPP. There are two main causes of the cell stress imposed upon the host organism, the first is caused by the diversion of metabolic resources from biosynthesis and cell division; and the second is caused by the accumulation of unfolded or incorrectly folded proteins, due to the inability of the cellular apparatus to keep pace with the protein synthesis rate (Sevastsyanovich *et al.,* 2010).

Initial experiments carried out to optimise the conditions for the production of rhTNFα showed signs of high cell stress in induced cultures. The use of strong promoters, such as the T7 expression system, causes metabolic stress by draining cell resources to maintain the high level of protein production demanded by the T7 RNA polymerase (Hoffmann and Rinas, 2004). Growth arrest, low cell viability and plasmid loss was observed in all cultures induced with high concentrations of arabinose (**Figure 3.1**). Hofmann and Rinas (2004) reported that the growth rate and the level of accumulation of recombinant proteins is inversely correlated as a function of the inducer concentration. Cultures producing rhTNFα showed low cell viability and plasmid loss, which was minimised by a decrease of the arabinose concentration. A 100 times lower concentration of arabinose than the standard recommended by the literature showed an eight-fold increase in

plasmid retention (**Figure 3.3 B**). In addition, plasmid-positive cells often grow slower than plasmid-free cells, outcompeting plasmid containing cells, which may lead to a concomitant decrease in the amount of protein being produced (Glick, 1995). This effect was even more obvious during fed-batch fermentation, where plasmid-free cells outgrew the productive culture, entailing a reduction of the final rhTNFα yield (**Figures 3.9 and 3.10**).

High overproduction of recombinant proteins may also affect cell division, showing a decline in the ability to form colonies on agar plates, being "viable but not culturable", since cells still exhibit metabolic activity but are unable to divide when grown on agar plates. No increase in the CFU associated with the increase in biomass was often observed during the production of rhTNFα in the late stages of the fermentation (**Figures 3.11 B and 3.13 B**).

3.3.4 Effect of temperature on the production of correctly folded rhTNFα

Protein production is the most energy consuming process in the cell, thus, energy generation becomes critical for the host organism during RPP (Hoffmann and Rinas, 2004). The level of induction, the gene codon bias and the mRNA stability determine the expression rates and correct folding of recombinant proteins. Slowing down the rate of product synthesis is often the most obvious approach to prevent the accumulation of unfolded products which can be achieved by reducing the cultivation temperature and inducer concentration, favouring the accumulation of a soluble and correctly folded product (Sevastsyanovich *et al.*, 2010; Fahner *et al.*, 2004).

Three different cultivation temperatures were evaluated for the production of rhTNFα, achieving the highest production of soluble rhTNFα at the lowest temperature evaluated. The accumulation of soluble rhTNFα increased with a reduction of the cultivation temperature, reaching a total of 90%, when cultures were grown at 25 °C. As expected, when higher cultivation temperatures were used, almost half of the total protein produced was accumulated in an insoluble form (**Figure 3.5**) (Fahner *et al.,* 2004). When a low cultivation temperature (25 °C) and a low concentration of

inducer (0.002% arabinose) were used, the plasmid retention reached almost 100%, but still maintained the high level of protein production (**Figure 3.6**), observed under harsher conditions, as a temperature of 37 °C and a strong induction with 0.2% arabinose (**Figure 3.5**). Once these conditions were translated to fed-batch fermentation, high yields of rhTNFα were found in the soluble protein fraction (**Figure 3.10**).

3.3.5 Production, purification and evaluation of activity of rhTNFα

Once the conditions for the production of rhTNFα had been optimised, high-cell-density cultures were carried out to increase rhTNFα titres. Two different semi-defined culture media were evaluated by fed-batch fermentation during the production of rhTNFα, in combination with the addition of casamino acids as a supplement (**Table 3.1**). The composition of both semi-defined culture media was similar: glycerol was used as a carbon source, ammonium sulphate was used as a nitrogen source and both formulations were supplemented with yeast extract. However, the casamino acid supplementation shown to have a great impact on plasmid retention when Cobra's growth medium was used (**Table 3.1**). In contrast, the culture medium composition extracted from Want *et al.* (2009) showed superior growth without the addition of casamino acids, exceeding a final OD_{600} of 100, the target cell density set by Cobra Biologics. A high plasmid retention and greater production of rhTNFα was also achieved when compared with fermentations using Cobra's semi-defined medium recipe (**Figures 3.9 – 3.16, Table 3.1**). In general, a higher cell density and production of rhTNFα was obtained with Want's medium formulation, with a final OD_{600} of 111 in contrast with 68 and 72, and accumulation of rhTNFα of 25%, comparing with 18 to 20%, of the total cell protein (**Table 3.1**). These results confirm the importance of an optimal culture composition for the production of recombinant proteins, and how it can help to improve growth and obtain a higher yield of the final product.

The nutrient feed strategy also plays a key role allowing the high productivity of recombinant proteins. During the production of rhTNFα, an exponential feeding strategy was used during fermentation, allowing the culture to grow at a constant specific growth rate, by using the carbon source, glycerol, as growth-limiting nutrient. A specific growth rate of 0.1 h^{-1} was selected as optimal with the aim to slow down growth and protein synthesis rates favouring the production of a correctly folded product (Choi *et al.*, 2006).

The optimal conditions selected during the shake-flask studies, such as a cultivation temperature of 25 °C or a low concentration of inducer (0.002% arabinose), minimised plasmid loss and led to a high production of rhTNFα favouring the accumulation of soluble protein. Nonetheless, solubility is not necessarily a synonym of activity, even though the production of a recombinant protein in a soluble form is often a good indication of activity. The activity of rhTNFα evaluated by TNFα cytotoxic bioassay, confirmed that rhTNFα produced by fed-batch fermentation was not only soluble but also an active protein product.

3.4 Conclusions and future work

Recombinant protein production still remains a challenge and different approaches need to be evaluated in a trial-and-error approach. The strategy used for the production of rhTNFα was based on the use of low concentrations of inducer to minimise cell stress during RPP; low cultivation temperature to decrease the protein synthesis rate and reduce protein misfolding; and a slow specific growth rate regulated by the feed regime which allowed a controlled growth during the production phase. In general terms, the optimisation of the conditions for production of rhTNFα was successful; a maximum yield of 5.35 $g \cdot L^{-1}$ of rhTNFα was obtained by fed-batch fermentation.

However, during the production of recombinant proteins by high-cell-density cultures, highest productivities have been reported when the growth and production phases are separated, by delaying the induction time until the cell density reaches the desired value (Wyre and Overton, 2014; Choi *et al.*, 2006). This approach needs to be explored since it has been reported that late induction may allow the production up to 12 g L^{-1} of rhTNFα (Hodgson *et al.*, 2011).

Nevertheless, the production of recombinant proteins targeting the cytoplasm is not an adequate approach for the production of protein therapeutics containing disulphide bonds. Disulphide-bonded recombinant proteins, are often targeted to the periplasm, which allows the formation of disulphide bonds due to its oxidising nature and the presence of the disulphide bond (DSB) generation machinery (Choi and Lee, 2004). Periplasmic protein production is increasing in relevance for the pharmaceutical industry, since antibody fragments are often produced using this route (Huang *et al.,* 2012). Therefore, the development of new approaches and screening tools is essential to ensure the successful production of recombinant protein targeting the periplasmic space.

Chapter 4:

Production of the single-chain antibody fragment 163R4

4.1 Introduction

The advent of the hybridoma technology in 1975 facilitated the production of monoclonal antibodies with defined antigen specificities and, nowadays, a large number of monoclonal antibodies have been approved for the treatment of cancer, inflammatory diseases or viral infections (Weiner *et al.*, 2010). A major breakthrough in the field was the production of different formats of antibody fragments, such as antigen-binding fragments (Fab) or single-chain antibody variable fragments (scFv) (**Figure 4.1**), which are currently on the market or in late phases of clinical trials, emerging as credible alternatives to monoclonal antibodies. Examples of approved antigen-binding antibody fragments are Lucentis® (ranibizumab), ReoPro® (abciximab) and Cimzia® (certolizumab pegol) (Nelson and Reichert, 2009). In contrast to full-length monoclonal antibodies, antibody fragments are minimally immunogenic due to smaller size, but retain the binding specificity of the parental monoclonal antibody. Antibody fragments can be produced at large scale with a lower cost using microbial hosts, and their smaller size allows the construction of fusion molecules, such as fusion proteins, nanoparticles or immunosensors, without a loss of binding function, guiding the way to a broad range of new therapeutic applications (Hagemeyer *et al.,* 2009).

ScFvs are commonly produced in *E. coli,* containing a leader sequence fused to the N-terminus that allows their translocation to the periplasmic space. The oxidising nature of the periplasm and the presence of the disulphide bond generation machinery allows the formation of the disulphide bonds that links the two β-sheets present in each antibody domain (**Figure 4.1**). As the disulphide bonds make a greater contribution to the stability of the scFv, lower yields of soluble and correctly folded scFv are expected when the protein is not translocated to the periplasm. With the main objective of exploring different approaches for periplasmic protein production, a single-chain antibody variable fragment, the scFv163R4, was selected as model protein.

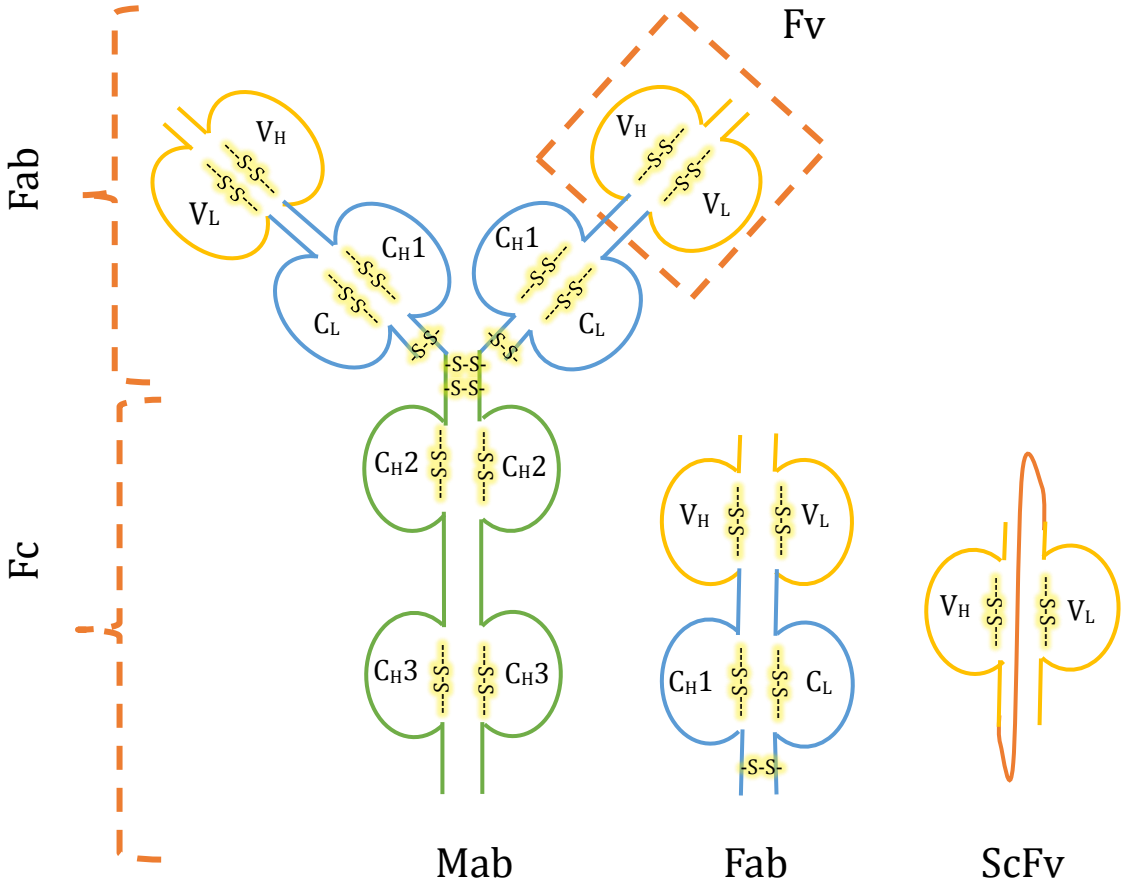


Figure 4.1. Schematic representation of a Mab, a Fab and a scFv. A full monoclonal antibody (Mab) is constituted by the fragment crystallisable (Fc) and the antigen-binding (Fab) regions. The Fc interacts with cell surface receptors, also called Fc receptors, and some proteins of the complement system and the Fab regions constitutes the region of the antibody that recognises and binds to the antigen. The Fab region is constituted by the constant and variable domains, with the variable domain containing the antigen-binding site, comprising the complementary determining regions (CDR). The antigen-binding site (Fab) is constituted by the constant and the variable regions of the heavy and light chains of an antibody. The Fv region the smallest antibody fragment that still contains the antigen-binding site. The single-chain antibody variable fragment (scFv) is a fusion of the variable regions of the heavy (VH) and light (VL) chains by a short peptide linker, which avoids the dissociation of the two antibody chains (Adapted from Plückthun, 1991).

Initially, a series of growth experiments were carried out for the optimisation of the conditions for the production of the single-chain antibody Fv fragment 163R4 (scFv163R4) using *E. coli* BL21. Martineau *et al.* (1998) described the generation of mutant variations of anti-β-galactosidase scFv, leading to the selection of the scFv163R4, due to its superior production in the cytoplasm. The scFv163R4 was used as a model protein for the present work and its expression was driven by the arabinose expression system, comprising the gene coding for the scFv163R4 under the control of the arabinose promoter. In addition, the scFv163R4 has C-terminal 6xHistidine and myc tag, to facilitate its purification and detection by western blot. Different conditions affecting the production of scFv163R4, such as temperature, inducer concentration or the effect of catabolite repression were evaluated by shake-flask experiments, with the aim to favour the production of the antibody fragment in the periplasm and to improve the yields of soluble and active protein.

In general terms, proteins are synthesised in the cytoplasm, as precursor proteins, requiring a short amino acid sequence, the signal peptide, to direct their translocation to the periplasmic space. Therefore, the selection of the optimal signal peptide is key for periplasmic protein production (Choi and Lee, 2004). The use of different signal peptides may also determine the translocation pathway, the SecB-dependent, SRP-mediated and TAT translocation pathways, depending on the amino acid composition of the signal peptide. The SEC and SRP translocation pathways are used for the translocation of unfolded polypeptides from the cytoplasm, post-translationally or co-translationally, respectively. In contrast, the TAT translocation pathway preferentially allows the secretion of folded or partially folded proteins to the periplasmic space (Holland, 2004). A more detailed description of the periplasmic translocation pathways can be found in **Chapter 1, section 1.3.2.2**. Among others, commonly used signal peptides for periplasmic protein production are: PelB, OmpA, PhoA or STII targeting the SEC pathway, DsbA, TolB or SfmC targeting the SRP pathway or TorA, which targets the TAT complex (Choi and Lee, 2004). Here, only signal peptides targeting either the SEC or the SRP translocation pathways were evaluated.

The use of different signal peptides has a great impact on the production and processing of recombinant proteins, and the efficiency of secretion may vary depending on the type of protein to be produced. At present, there is no general rule for the selection of the optimal signal peptide that can guarantee successful translocation to the periplasm; different signal peptides need to be evaluated in a trial-and-error approach (Choi and Lee, 2004). In order to accelerate the selection of the optimal signal peptide to facilitate the translocation of the scFv163R4, a fusion protein, comprising the scFv163R4 fused to the TEM-1 β-lactamase, was used to rapidly screen and evaluate the production of the recombinant protein in the periplasmic space. The TEM-1 β-lactamase, also known as penicillinase, is an enzyme that provides resistance to β-lactam antibiotics by hydrolysis of the β-lactam ring. Hence, the production of periplasmic β-lactamase confers protection to ampicillin, which can be evaluated by minimum inhibitory concentration (MIC) experiments and colony screening, and its activity can be quantified by enzymatic assay.

4.2 Results

4.2.1 Effect of the cultivation temperature on the production of scFv163R4

Preliminary experiments were carried out by shake-flask with the aim to evaluate the optimal conditions for the production of the single-chain antibody 163R4 (scFv163R4). The expression of the scFv163R4 was driven by the arabinose expression system, using *E. coli* BL21-A (*ΔaraBAD),* as host organism. A detailed description of this expression system can be found in **Chapter 1, section 1.2.2.2.** Initial screening included the evaluation of the optimal temperature for the production of the scFv163R4. The heat stable enterotoxin II (STII) signal peptide was used to direct the translocation of the scFv163R4 to the periplasm, mediated by the SecB-dependent translocation pathway.

Shake-flask experiments were carried out with *E. coli* BL21-A carrying a plasmid containing the sequence coding for the scFv163R4 under the control of the arabinose promoter (pLBAD2-STII-

scFv163R4). Cultures were grown in terrific broth, at 30 °C or 25 °C, induced with 0.02% arabinose at an OD_{600} of 0.5. Cultures grown under non-induced conditions were used as a control for this experiment. Cell culturability, used as indication of cell viability, was evaluated by serially diluted culture samples being plated onto non-selective LB agar. Plasmid retention was evaluated by replica plating. Shake-flask experiments were carried out in duplicate.

As expected, cultures grown at 30 °C showed a faster growth than cultures grown at 25 °C, although large differences were observed after induction of the expression of the scFv163R4 (**Figure 4.2 A**). Cultures grown at 30 °C under induced conditions, showed a decrease in the cell density, from an OD_{600} of 17.9 to 7.5 between 9 and 24 hours of growth, probably as a result of cell lysis. In contrast, cultures grown at 25 °C showed a normal growth, showing a similar growth profile than cultures grown under non-induced conditions. In terms of cell culturability, a drastic reduction in the colony forming units (CFU) can be observed between 8 and 24 hours post-inoculation (**Figure 4.2 B**). The loss of culturability was more severe for cultures grown at 30 °C after 6 hours of induction, but cultures grown at both temperatures reached a similar number of CFU after 24 hours of growth. Plasmid retention showed that 90 to 100% of the cells were plasmid-positive after 24 hours of growth for all the cultures (data not shown).

Culture samples were collected at time points pre- and post-induction and processed fresh to extract the periplasmic protein fraction, by cold osmotic shock (Humphreys, 2007). Once extracted, the periplasmic and the spheroplast fractions were stored at -20 °C for further analysis. When required, spheroplasts were resuspended and disrupted by sonication. Subsequently, the soluble and insoluble cytoplasmic fractions were obtained after centrifugation.

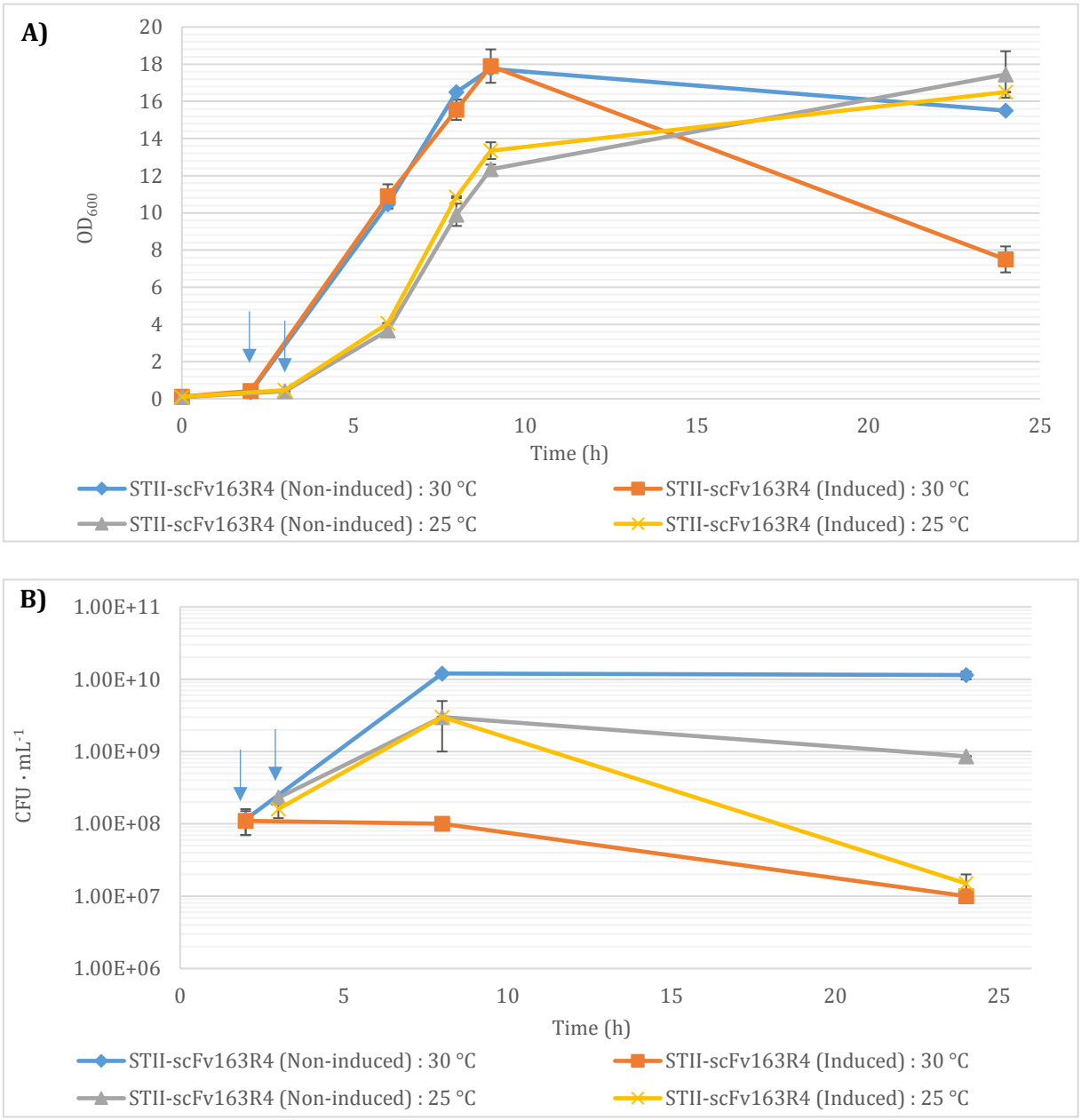


Figure 4.2. Selection of the cultivation temperature for the production of scFv163R4. *E. coli* BL21-A carrying the vector coding for the scFv163R4 were grown at 30°C or 25 °C; half of cultures were induced with 0.02% arabinose at an OD_{600} ≈ 0.5 (blue arrow). The translocation of the scFv163R4 to the periplasm was facilitated by the STII signal peptide. The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 2 h, 3 h, 8 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

In terms of protein production, the production of the scFv163R4 was slightly higher in cultures grown at 30 °C than at 25 °C, as it can be observed comparing the concentration of the scFv163R4 from whole cell lysates obtained 24 hours post-inoculation (**Figure 4.3**). Nevertheless, the high concentration of the scFv163R4 in the whole cell lysate did not correlate with a higher concentration of protein in the periplasm, being accumulated mainly in the insoluble protein fraction. Cultures grown at 25 °C also showed a high production of the scFv163R4, the majority of the protein being also accumulated in an insoluble form. Similar amounts of recombinant protein were accumulated in the periplasm, independently of the cultivation temperature, the ratio of total to periplasmic accumulation of recombinant protein being slightly higher for those cultures grown at 25 °C. This suggests than the use of lower cultivation temperatures may have helped to slow down the protein synthesis rate, resulting in a lower production of recombinant protein but a higher quantity being produced in a soluble form.

Nevertheless, results obtained from the partitioning could also lead to misleading conclusions, since the presumptive high concentration of the scFv163R4 in the cytoplasmic fraction could be the result of an incomplete extraction of the periplasmic content. Hence, it is important to highlight that no differences in the molecular weight of the scFv163R4 were observed between samples of different protein fractions (**Figure 4.3**). During periplasmic protein production, once the recombinant protein is successfully translocated to the periplasmic space, the signal peptide is cleaved off by the signal peptidase I, yielding the mature form of the protein, with a slightly lower molecular weight. However, no difference in the molecular weight of the precursor (30.9 kDa), accumulated in the spheroplast, and mature form (28.3 kDa), accumulated in the periplasm, was observed by SDS-PAGE, so it was not possible to determine whether the scFv163R4 was successfully translocated to the periplasmic space or if the signal peptide had been cleaved off. Further experiments were required to confirm if the scFv163R4 was translocated to the periplasm.

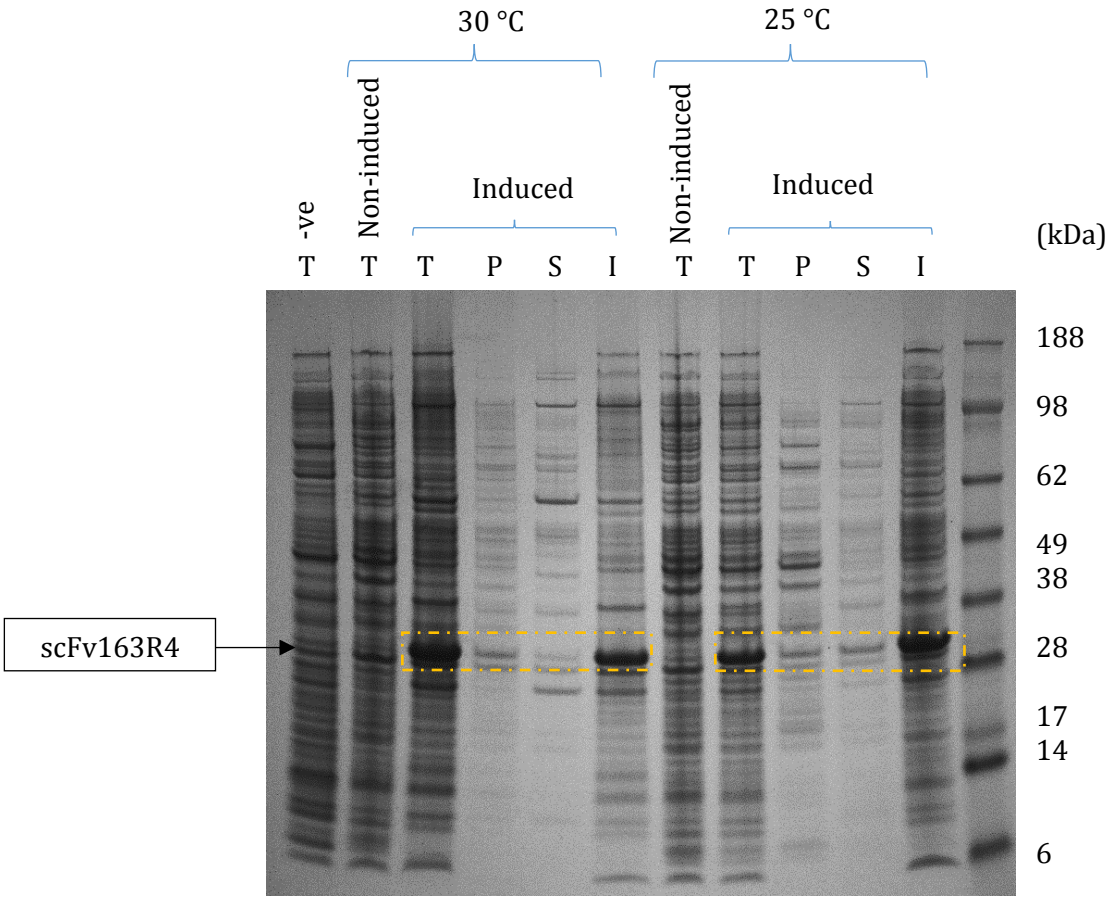


Figure 4.3. Effect of temperature on the production of scFv163R4. SDS-PAGE gels showing the accumulation of scFv163R4 from total (T), periplasmic (P), cytoplasmic soluble (S) and cytoplasmic insoluble (I) protein fractions. Samples were obtained from induced and non-induced cultures carrying the vector encoding scFv163R4 (yellow square) after 24 hours of growth. Cultures were grown at a temperature of 30 °C and 25 °C and induced with 0.02% arabinose. The translocation of the scFv163R4 to the periplasmic space was facilitated by the STII signal peptide. Samples obtained from whole cell lysates of cultures carrying the empty vector (pLBAD2) after 24 hours of growth, were used as a negative control.

4.2.2 Effect of catabolite repression (0.5% glucose) on the production of scFv163R4

Previous experiments failed to achieve a high production of the scFv163R4 in the periplasmic space, the majority being accumulated in an insoluble form. Since a lower cultivation temperature may have facilitated the production of a higher quantity of scFv163R4 in the periplasm, it was considered that a decrease of the activity of the arabinose promoter could further improve the production of mature scFv163R4. Wycuff and Matthews (2000) reported that the addition of glucose was able to repress the activity of the arabinose promoter, allowing the modulation of the production of target protein by the addition of different concentrations of arabinose. With the aim to understand whether catabolite repression could be used as a strategy to improve the protein production in the periplasm, combinations of different concentrations of arabinose and glucose were evaluated.

Shake-flask experiments were carried out with *E. coli* BL21-A carrying a plasmid containing the sequence coding for the scFv163R4 and grown at 25 °C. Cultures grown under non-induced conditions were used as a control. As in the previous experiment, the STII signal peptide was used to direct the translocation of the scFv163R4 to the periplasmic space. Protein expression was induced with different concentrations of arabinose, ranging from 0.2% to 0.0002%, at an OD_{600} of 0.5. In order to evaluate the effect of glucose repression on the production of the scFv163R4, upon induction, glucose was added to a final concentration of 0.5% to all the cultures. Cell culturability, as an indication of cell viability, was evaluated by serially diluted culture samples being plated onto non-selective LB.

In terms of growth, no significant differences were observed between cultures induced with different concentrations of arabinose and cultures grown under non-induced conditions (**Figure 4.4 A**). All the cultures reached a high cell density after 24 hours of growth, with a final OD_{600} ranging from 17 to 19.

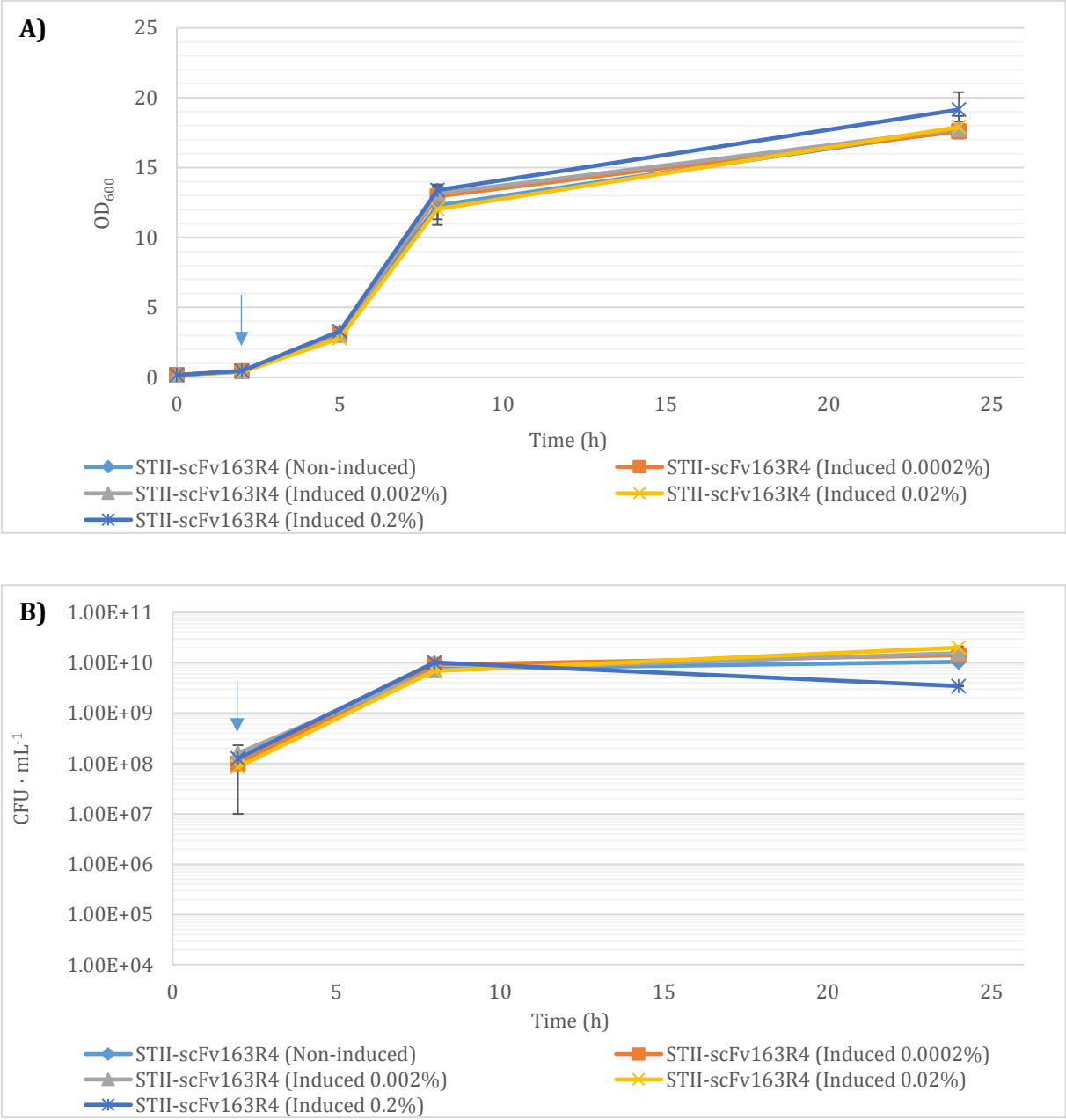


Figure 4.4. Effect of catabolite repression (0.5% glucose) on growth and cell viability. *E. coli* BL21-A carrying the vector coding for the scFv163R4 were grown at 25 °C; cultures were induced with different concentrations of arabinose, ranging from 0.0002% to 0.2% of arabinose, at an OD_{600} ≈ 0.5 (blue arrow). The translocation of the scFv163R4 was facilitated by the STII signal peptide. The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 2 h, 8 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

Minor differences were also observed in terms of cell culturability, and only slight decrease in the CFU was observed for cultures induced with the highest concentration of inducer evaluated, 0.2% arabinose, after 24 hours of growth (**Figure 4.4 B**). The loss of cell culturability was minimal, with a decrease from 10^{10} to $3.5 \cdot 10^{9}$ CFU $\cdot$ mL^{-1}, between 8 and 24 hours post-inoculation. In general terms, no significant differences were observed between cultures induced with different concentrations of arabinose.

As expected, the catabolite repression had a strong effect on the production of the scFv163R4 (**Figure 4.5 A**). When low concentrations of arabinose, from 0.0002% to 0.002%, were used to induce protein production, no protein was detected 24 post-inoculation. However, when the concentration of arabinose was increased to 0.02%, the production of the scFv163R4 could be detected by western blot after 24 hours of growth. The highest accumulation of scFv163R4 was observed in samples obtained from cultures induced with 0.2% arabinose, accounting for 9.4% of the total cell protein (TCP), and the titration of the production of scFv163R4, induced by the catabolite repression, can be observed at a concentration of arabinose ranging from 0.02% and 0.2% after 24 hours of growth.

In order to evaluate whether the glucose repression had an effect on the proportion of the scFv163R4 being translocated to the periplasm, the periplasmic fraction was extracted by cold osmotic shock. A high concentration of scFv163R4 was detected in the periplasm for cultures induced with 0.2% arabinose and repressed with 0.5% glucose, corresponding to 22% of the total scFv163R4 produced being accumulated in the periplasm (**Figure 4.5 B**). Nonetheless, a high concentration of the recombinant protein was still detected in the cytoplasmic soluble and insoluble protein fractions, corresponding to 20% and 58% of the total produced scFv163R4, respectively.

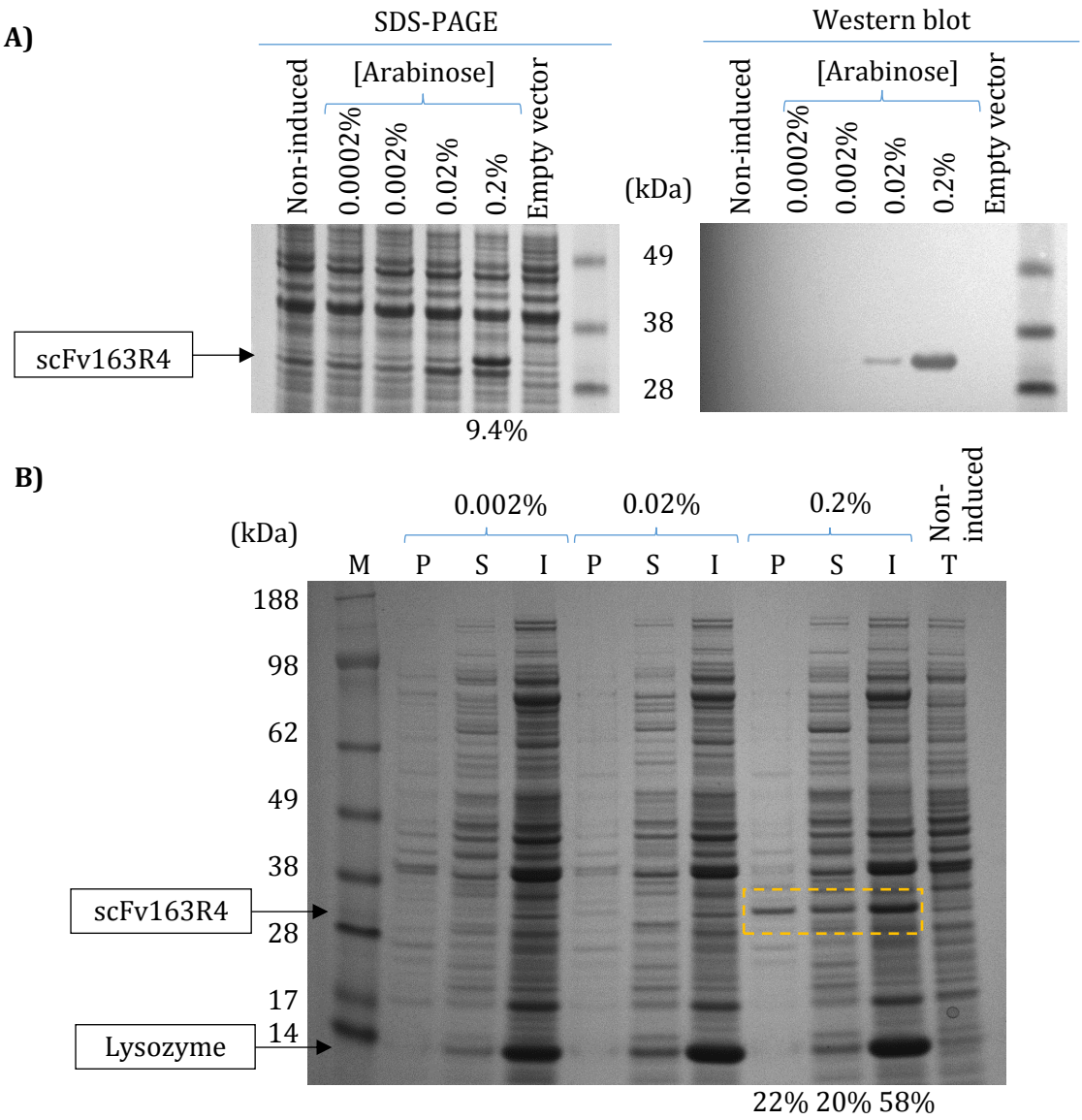


Figure 4.5. Effect of catabolite repression (0.5% glucose) on the production of scFv163R4. SDS- PAGE gel showing the accumulation of scFv163R4 from whole cell lysates obtained from cultures grown at 25 °C, supplemented with 0.5% glucose and induced with different concentrations of arabinose, ranging from 0.2% to 0.0002%, at an OD_{600} of 0.5. Samples were obtained after 24 hours of growth. A sample obtained from a culture carrying the empty vector after 24 hours of growth was used as a negative control. The production of the scFv163R4 was confirmed by western blot using an anti-myc antibody (A). The accumulation of scFv163R4 (yellow square) from total (T), periplasmic (P), cytoplasmic soluble (S) and cytoplasmic insoluble (I) protein fractions after 24 hours of growth (B). The quantity of the scFv163R4 is expressed as percentage of whole cell protein and the ratio of protein accumulated in the periplasm (A) and spheroplast protein fractions are expressed as percentage at the bottom of the gel (B).

The addition of glucose upon induction allowed the repression of the arabinose promoter even in the presence of high concentrations of arabinose. The concentration of glucose in the medium was reduced progressively during cultivation, since *E. coli* preferentially uses glucose as a carbon source, allowing the slow de-repression of the arabinose promoter, inducing the expression of the scFv163R4 (**Figure 4.5 B**). This slow de-repression seems to favour the translocation of the protein to the periplasm, and a higher quantity of scFv163R4 was found in the periplasm when compared with previous experiments (**Figure 4.3**).

4.2.3 Effect of catabolite repression (0.25% glucose) on the production of scFv163R4

As previous experiments showed a great improvement in the production of the scFv163R4 in the periplasm, the effect of glucose repression on the arabinose promoter was further evaluated. With the objective to find the optimal balance between the concentration of inducer and repressor, a concentration of glucose of 0.25% was evaluated in combination with a concentration of inducer ranging from 0.0002% to 0.2% arabinose.

Shake-flask experiments were carried out with *E. coli* BL21-A carrying a plasmid containing the sequence coding for the scFv163R4 fused to the STII signal peptide, under the control of the arabinose promoter. Cultures were grown at 25 °C and induced with a concentration of arabinose ranging from 0.2% to 0.0002%, at an OD_{600} of 0.5. In order to evaluate the effect of glucose repression on the production of the scFv163R4, upon induction, glucose was added to a final concentration of 0.25% to all the cultures. Cultures grown under non-induced conditions were used as a control for this experiment.

In terms of growth, no significant differences were observed between cultures induced with different concentrations of arabinose and cultures grown under non-induced conditions. All the cultures reached a high cell density after 24 hours of growth, with a final OD_{600} ranging from 17 to 20 (**Figure 4.6 A**). However, significant differences were detected in terms of cell culturability

(**Figure 4.6 B**). Cultures induced with high concentrations of arabinose, ranging from 0.2% to 0.02%, showed a decrease in CFU when compared to cultures induced with lower concentrations of arabinose. This decrease in the CFU was not observed in cultures grown under non-induced conditions, being probably the result of the cell stress induced by the high levels of production of the scFv163R4.

The production of the scFv163R4 was detected on a western blot even when low concentrations of arabinose, such as 0.0002% or 0.002%, were used to induce protein expression (**Figure 4.7 A**), in contrast with previous experiments, where high concentrations of arabinose were required to induce the production of scFv163R4 (**Figure 4.5 A**). The production of the scFv163R4 was detected even at the lowest concentration of arabinose, 0.0002%, despite the addition of glucose to the culture medium (**Figure 4.7 A**).

In terms of the proportion of protein accumulated in each protein fractions, the accumulation of the scFv163R4 was detected in the periplasmic fraction for cultures induced with the highest concentrations of arabinose, ranging from 0.02% to 0.2% (**Figure 4.7 B**). The concentration of the scFv163R4 in the periplasm was higher, accounting for 21% of the total scFv163R4 produced when 0.02% arabinose was used to induce protein production, whereas 6% and 73% was found accumulated in the soluble and insoluble cytoplasmic fractions, respectively. Only 15% of the total scFv163R4 was accumulated in the periplasm when the culture was induced with 0.2% arabinose, being 14% and 71% accumulated in the cytoplasmic soluble and insoluble fractions.

In general terms, the concentration of the scFv163R4 in the periplasmic fraction was higher when glucose was added to repress the activity of arabinose promoter for both concentrations evaluated, 0.5% and 0.25% (**Figures 4.5 and 4.7**), showing a great improvement from initial shake-flasks experiments (**Figure 4.3**). However, a high concentration of recombinant protein can

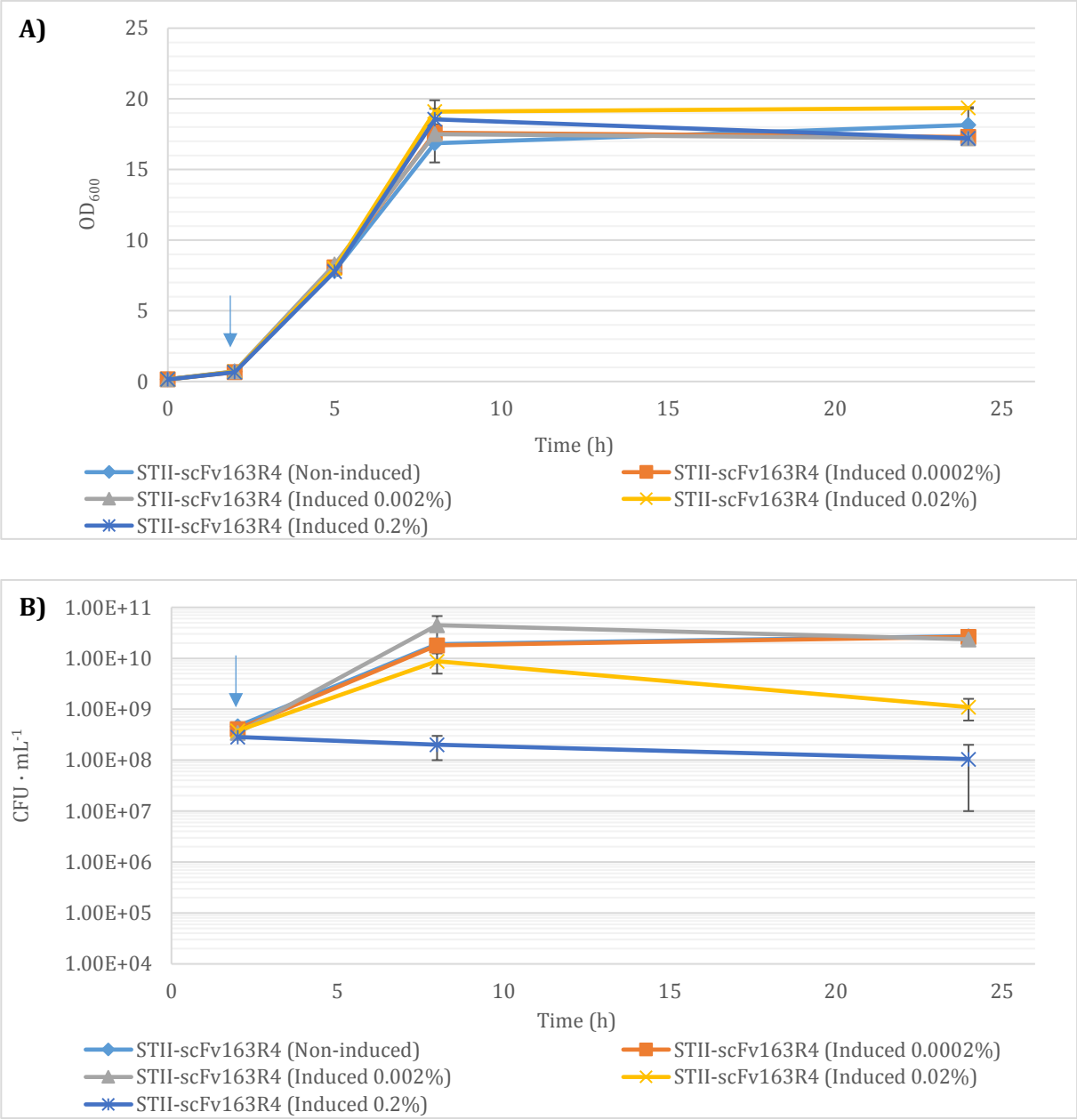


Figure 4.6. Effect of catabolite repression (0.25% glucose) on growth and cell viability. *E. coli* BL21-A carrying the vector coding for the scFv163R4 were grown at 25 °C; cultures were induced with different concentrations of arabinose, ranging from 0.0002% to 0.2% of arabinose, at an OD_{600} ≈ 0.5 (blue arrow). The translocation of the scFv163R4 was facilitated by the STII signal peptide. The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 2 h, 8 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

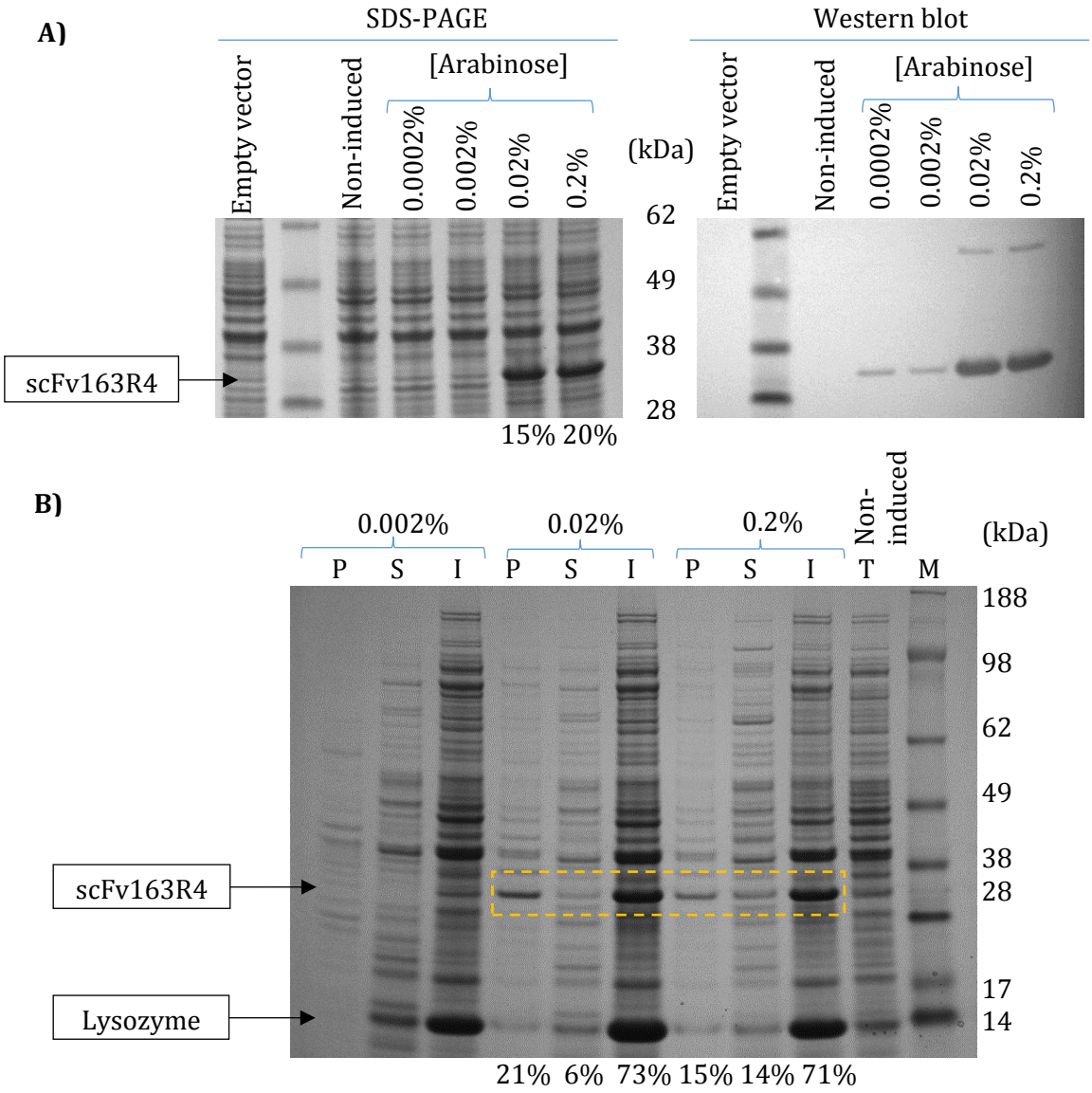


Figure 4.7. Effect of catabolite repression (0.25% glucose) on the production of scFv163R4. SDS-PAGE gel showing the accumulation of scFv163R4 from whole cell lysates obtained from cultures grown at 25 °C, supplemented with 0.25% glucose and induced with different concentrations of arabinose, ranging from 0.2% to 0.0002%, at an OD_{600} of 0.5. Samples were obtained after 24 hours of growth. A sample obtained from a culture carrying the empty vector after 24 hours of growth was used as a negative control. The production of the scFv163R4 was confirmed by western blot using an anti-myc antibody (A). The accumulation of scFv163R4 (yellow square) from total (T), periplasmic (P), cytoplasmic soluble (S) and cytoplasmic insoluble (I) protein fractions after 24 hours of growth (B). The quantity of the scFv163R4 is expressed as percentage of whole cell protein and the ratio of protein accumulated in the periplasm (A) and spheroplast protein fractions are expressed as percentage at the bottom of the gel (B).

be still found in the insoluble protein fraction. The production of the scFv163R4 was higher for cultures repressed with 0.25% glucose, accounting for 20% of the total cell protein (TCP) for cultures induced with 0.2% arabinose (**Figure 4.7**). In contrast, when a concentration of 0.5% glucose was used to repress the activity of the arabinose promoter, only 9.4% of TCP corresponded to the scFv163R4, with the same concentration of inducer (**Figure 4.5**).

Overall, the use of glucose repression seems to allow the modulation of the production of scFv163R4 by varying the concentration of arabinose used to induce the activity of the arabinose promoter. Fine tuning the concentration of arabinose used to the induce protein expression and the amount of glucose required to repress the activity of the arabinose promoter seems to provide better control over the level of protein expression (Wycuff and Matthews, 2000). However, this approach is not feasible at larger scale, since *E. coli* cells preferentially use glucose as a carbon source, being difficult to maintain the glucose concentration in the medium without losing the control of the growth rate during fed-batch fermentations. In addition, the slow de-repression of the arabinose promoter produced by the consumption of the glucose present in the culture medium would be an extremely complex process to scale-up and reproduce at high-cell-density fermentations.

4.2.4 Effect of different signal peptides on the production of scFv163R4

The efficiency of protein translocation varies extremely depending on the signal peptide and the type of protein to be translocated to the periplasm. In addition, protein translocation is strongly influenced by the first amino acids of the sequence of the target protein. At the moment, there is no universal signal peptide that guarantees successful translocation for a given recombinant protein and several signal peptides need to be screened in a trial-and-error approach (Choi and Lee, 2004). In previous experiments, the translocation of the scFv163R4 to the periplasmic space was facilitated by the STII signal peptide. However, this signal peptide may not be the optimal for

the production of the scFv163R4 and therefore, the use of other signal peptides should be evaluated.

For this purpose, two new signal peptides were selected: the disulphide oxidoreductase A (DsbA) and the pectate lyase B (PelB). The DsbA signal peptide targets the SRP translocation pathway, allowing the co-translational translocation of recombinant proteins to the periplasm (Bibi, 2007). On the other hand, the PelB signal peptide is a non-*E. coli* signal peptide obtained from *Pectobacterium carotovorum*, previously known as *Erwinia carotovora,* which targets the Sec translocation pathway, analogously to the STII signal peptide.

Shake-flask experiments were carried out with *E. coli* BL21-A carrying a plasmid containing the sequence coding for the scFv163R4, fused to the DsbA (pLBAD2-DsbA-scFv163R4) or the PelB (pLBAD2-PelB-scFv163R4) signal peptides, under the control of the arabinose promoter. Cultures were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. Catabolite repression was used in this experiment for screening purposes, with the aim to evaluate whether the use of different signal peptides could favour a higher production of the scFv163R4 in the periplasm under the similar growth conditions than in previous experiments. Upon induction, glucose to a final concentration of 0.5% or 0.25% was added to the culture medium. Non-induced cultures, used as a control, were supplemented with glucose to a final concentration of 0.5%.

Cultures carrying the plasmid vector containing the sequence coding for DsbA and PelB signal peptides fused to the scFv163R4 under non-induced conditions reached a final OD_{600} of 19 and 20, after 24 hours of growth (**Figure 4.8 A**). Cultures grown under induced conditions where the translocation of the scFv163R4 to the periplasm was facilitated by the PelB signal peptide also reached high cell density, with a final OD_{600} of 18 after 24 hours of growth, for both concentrations of glucose evaluated. In contrast, cultures carrying the DsbA signal peptide showed a severe growth arrest and reached a final OD_{600} of 8.7 for cultures supplemented with 0.5% glucose, and

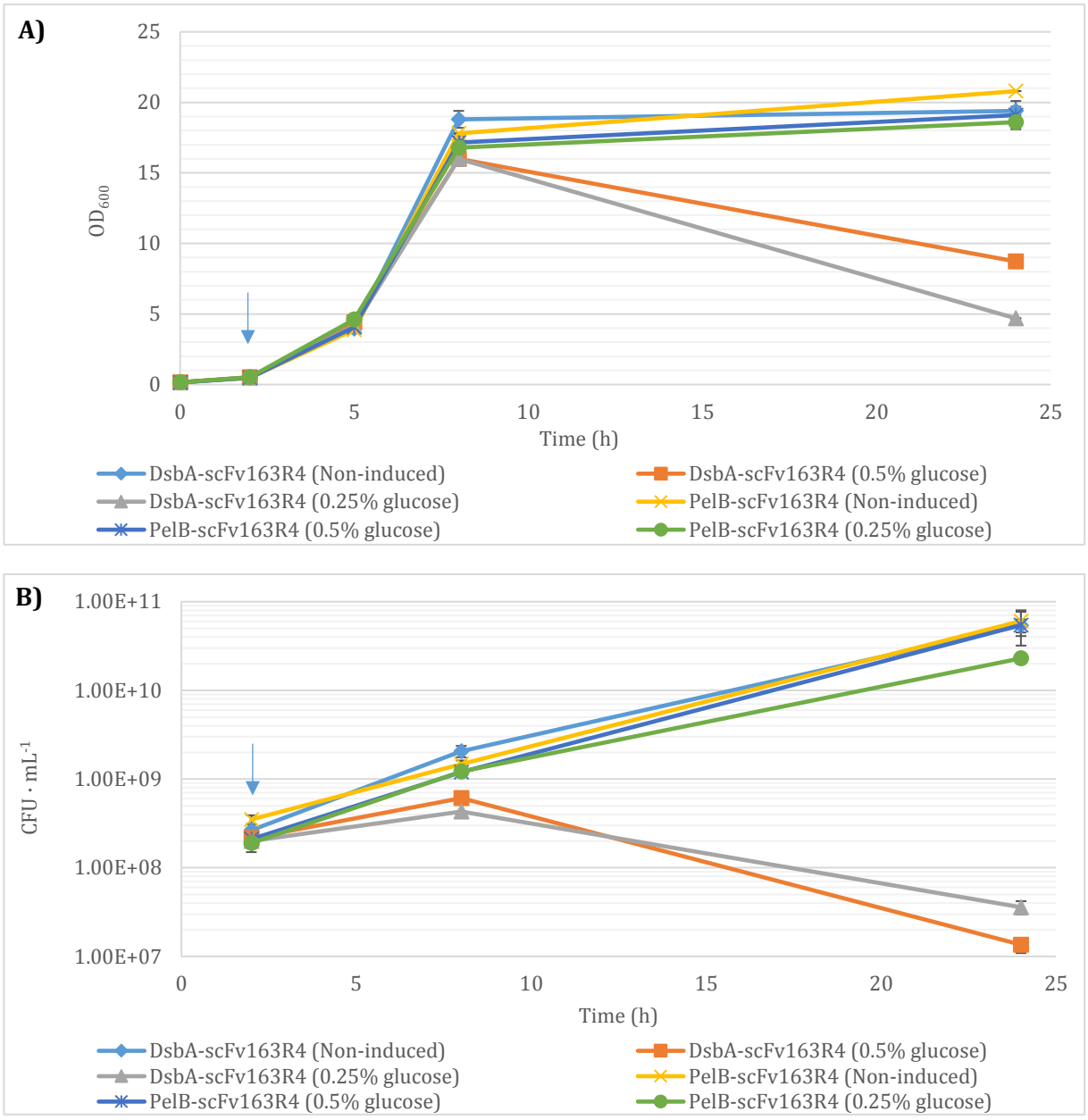


Figure 4.8. Effect of different signal peptides on growth and cell viability. *E. coli* BL21-A carrying the vector coding for the scFv163R4 fused to the DsbA or PelB signal peptides were grown at 25 °C. Cultures were induced with 0.02% of arabinose and repressed by the addition of 0.25% or 0.5% of glucose at an OD_{600} ≈ 0.5 (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 2 h, 8 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

4.7 for cultures supplemented with 0.25% glucose, after 24 hours of growth. The decrease in cell density observed between 8 and 24 hours post-inoculation was probably the result of cell lysis.

A similar pattern to the growth profile was observed in terms of cell culturability (**Figure 4.8 B**). Non-induced cultures carrying the vector coding for either the DsbA or the PelB signal peptides showed a high cell culturability, reaching approximately $5 \cdot 10^{10}$ CFU $\cdot$ mL^{-1} after 24 hours of growth. Cultures carrying the PelB signal peptide and supplemented with either 0.25% or 0.5% glucose also showed a high cell culturability, comparable to non-induced cultures. However, cultures carrying the vector coding for the DsbA signal peptide fused to the scFv163R4 showed a drastic decrease in cell viability, with approximately only 10^7 CFU $\cdot$ mL^{-1} after 24 hours of growth for cultures supplemented with either 0.5% or 0.25% of glucose. The loss of cell culturability was so severe that after 24 hours of growth, the number of CFU was lower than the CFU obtained before induction, 2 hours post-inoculation. The decrease in the CFU observed for cultures carrying the DsbA signal peptide after 24 hours of growth correlated with low final OD_{600}, which it was probably caused by cell lysis.

In terms of the production of the scFv163R4, there were significant differences depending on the signal peptide used to favour the translocation of the scFv163R4 (**Figure 4.9 A**). When DsbA was used as a signal peptide, a high production of recombinant protein was observed soon after induction, and high levels of protein production were detected after 24 hours of growth, accounting for 18% and 25% of the total cell protein, in cultures supplemented with 0.5% and 0.25% glucose, respectively. Unfortunately, this high level of protein production correlated with poor growth and low cell viability (**Figure 4.8**). In contrast, when PelB was used as a signal peptide, the production of the antibody fragment was lower, when compared with samples taken from cultures carrying the DsbA signal peptide (**Figure 4.9 A**).

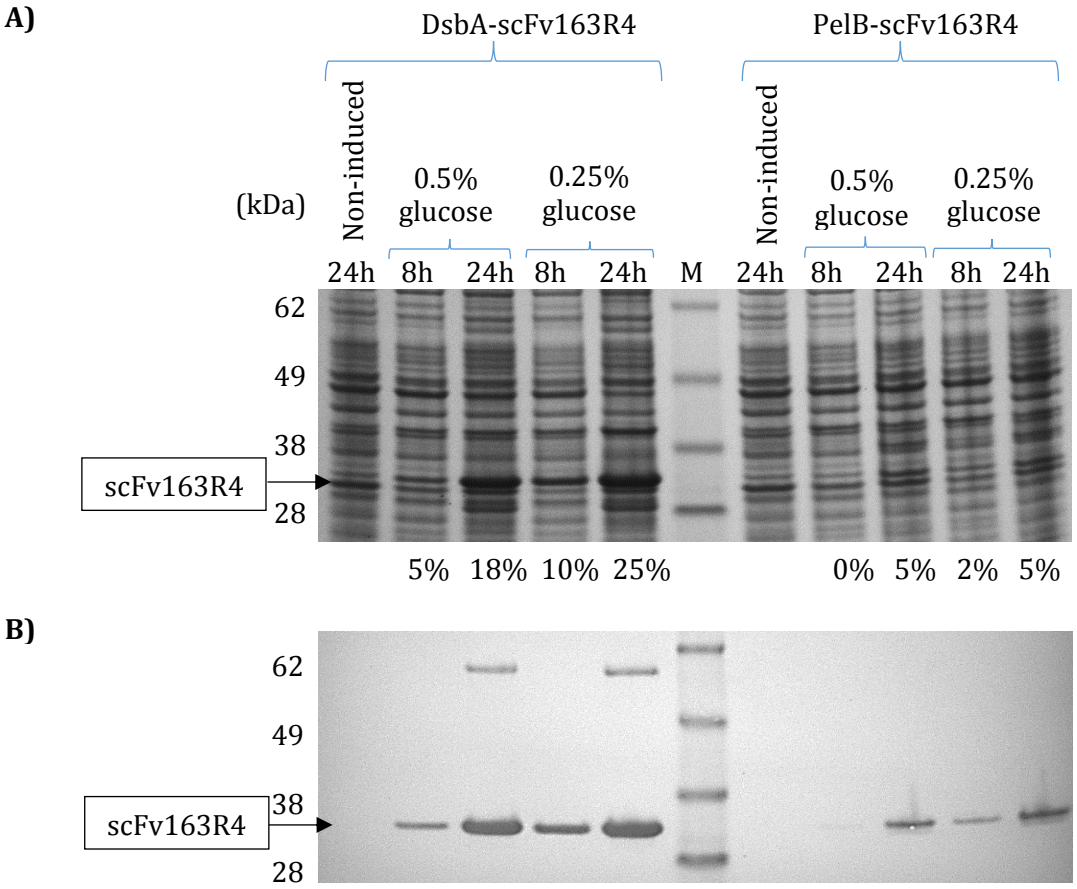


Figure 4.9. Effect of different signal peptides on the production of scFv163R4. SDS-PAGE gel showing the accumulation of scFv163R4 from whole cell lysates. Samples were obtained from induced and non-induced cultures carrying the vector encoding the scFv163R4 fused to the DsbA or PelB signal peptides after 8 and 24 hours of growth. Cultures were induced with 0.02% of arabinose and repressed by the addition of 0.5% or 0.25% glucose. The quantity of scFv163R4 is expressed as a percentage of whole cell protein at the bottom of the gel (A). The accumulation of the scFv163R4 was confirmed by western blot using an anti-myc antibody (B).

As in previous experiments, the repression of the arabinose promoter was higher 8 hours post-inoculation, and large differences were observed depending on the concentration of glucose used (**Figure 4.9 B**). As a result of the de-repression of the arabinose promoter caused by the consumption of glucose by the cells, no significant differences were observed 24 hours post-inoculation, independently of the concentration of glucose used, for cultures carrying the PelB signal peptide. A higher production of the scFv163R4 was observed for cultures carrying the DsbA signal peptide, for cultures and repressed with 0.25% glucose, accounting for 25% of the total cell protein.

The addition of glucose repressed the activity of the arabinose promoter, and lower scFv163R4 yields were obtained at a glucose concentration of 0.5% than with 0.25%. (**Figure 4.10**). Cultures carrying the STII signal peptide showed an increase in the production of the scFv163R4, from 9% to 12%, when the concentration of glucose added to the culture medium was reduced from 0.5% to 0.25%. Comparably, the production of the scFv163R4 for cultures carrying the DsbA signal peptide was increased from 18% to 25%, for a concentration of glucose of 0.5% and 0.25%, respectively. In contrast with results obtained with the STII and DsbA signal peptides, the addition of glucose had a limited effect on cultures where the production of the scFv163R4 was facilitated by the PelB signal peptide, accounting for a maximum of 5% of the total cell protein, independently of the concentration of glucose used.

4.2.5 Effect of different signal peptides on the production of scFv163R4-A

The vector pET163R4, from where the original sequence of the scFv was obtained, was designed for the cytoplasmic expression of the scFv163R4, and its sequence can be found described in Martineau *et al.* (1998). Since the original sequence was designed for its production targeting the cytoplasm, an extra alanine at the first position of the scFv163R4 was added after the initiation

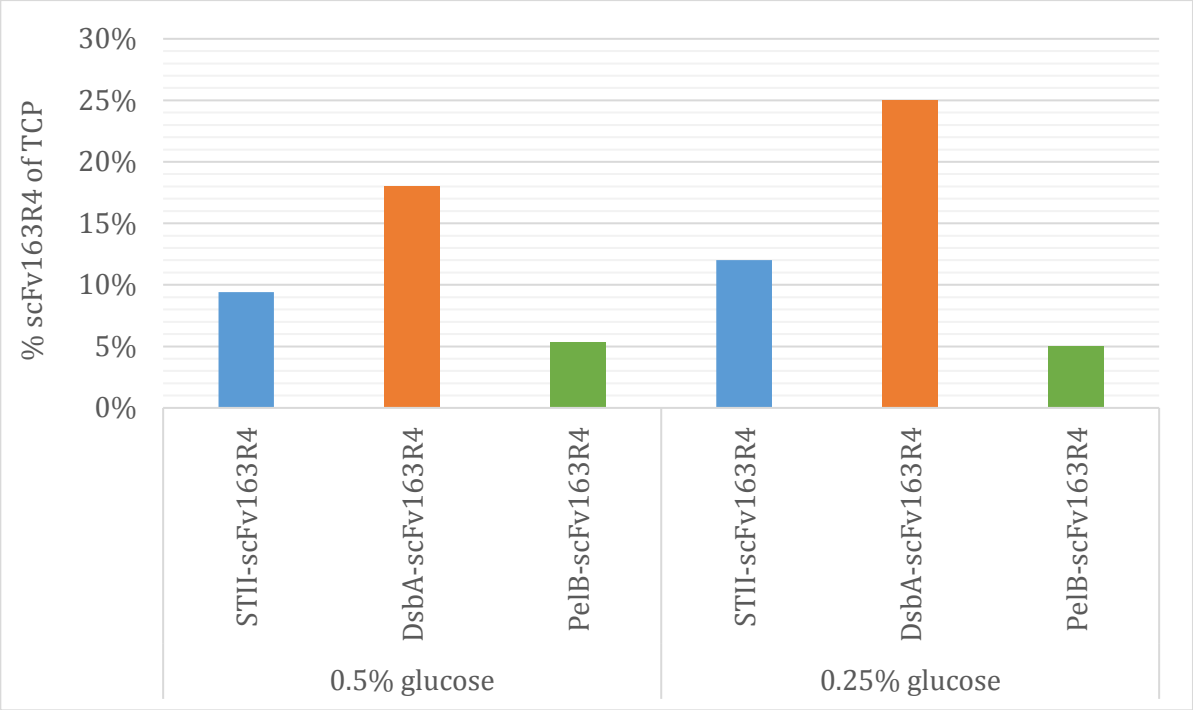


Figure 4.10. Effect of catabolite repression and the use of different signal peptides on the production of scFv163R4. Production of the scFv163R4 in cultures grown at 25 °C, induced with 0.02% arabinose and supplemented with 0.5% or 0.25% glucose at an OD_{600} of 0.5. The production of the scFv163R4 is represented as a percentage of the total cell protein (TCP) obtained after 24 hours of growth. Data shown here are single values obtained by quantifying the accumulation of the scFv163R4 in whole cell lysates by densitometry from SDS-PAGE gels.

codon. The introduction of the extra amino acid was probably resulted in an increase of the stability of scFv163R4 when produced as a soluble protein in the cytoplasm. This extra amino acid was no longer required for its production in the periplasm, since the stability of the protein is favoured by the formation of the two disulphide bonds within the light and heavy chains of the scFv, and it should not be present in order to produce the protein with its native N-terminus.

Shake-flask experiments were carried out with *E. coli* BL21-A carrying a plasmid containing the sequence coding for the scFv163R4, without the alanine at the first position of the mature amino acid sequence, renamed scFv163R4-A. With the aim to evaluate the impact of the removal of the extra amino acid on the production and translocation of the scFv163R4-A to the periplasm, the STII, DsbA or PelB signal peptides were fused to the sequence of the scFv163R4-A (pLBAD2-STII-scFv163R4-A, pLBAD2-DsbA-scFv163R4-A and pLBAD2-PelB-scFv163R4-A). Cultures were grown at 25 °C, induced with 0.02% arabinose at an OD_{600} of 0.5. Upon induction, glucose was added to a final concentration of 0.25%. Cell viability was evaluated by serially diluted culture samples being plated onto non-selective LB agar.

In terms of growth, large differences were observed between non-induced cultures, used as control, and cultures grown under induced conditions (**Figure 4.11 A**). These differences were even greater depending on the signal peptide used to direct the translocation of the scFv163R4-A to the periplasm. Growth arrest and possible cell lysis were observed in cultures where the translocation of the scFv163R4-A was facilitated by the STII and DsbA signal peptides after 24 hours of growth. A decrease in the cell density was observed between 9 and 24 hours of growth, from 18 to 12, for cultures carrying the STII signal peptide; and from 14 to 6.5, for cultures carrying the DsbA signal peptide. Growth arrest was not observed in cultures where the translocation of the recombinant protein was facilitated by the PelB signal peptide.

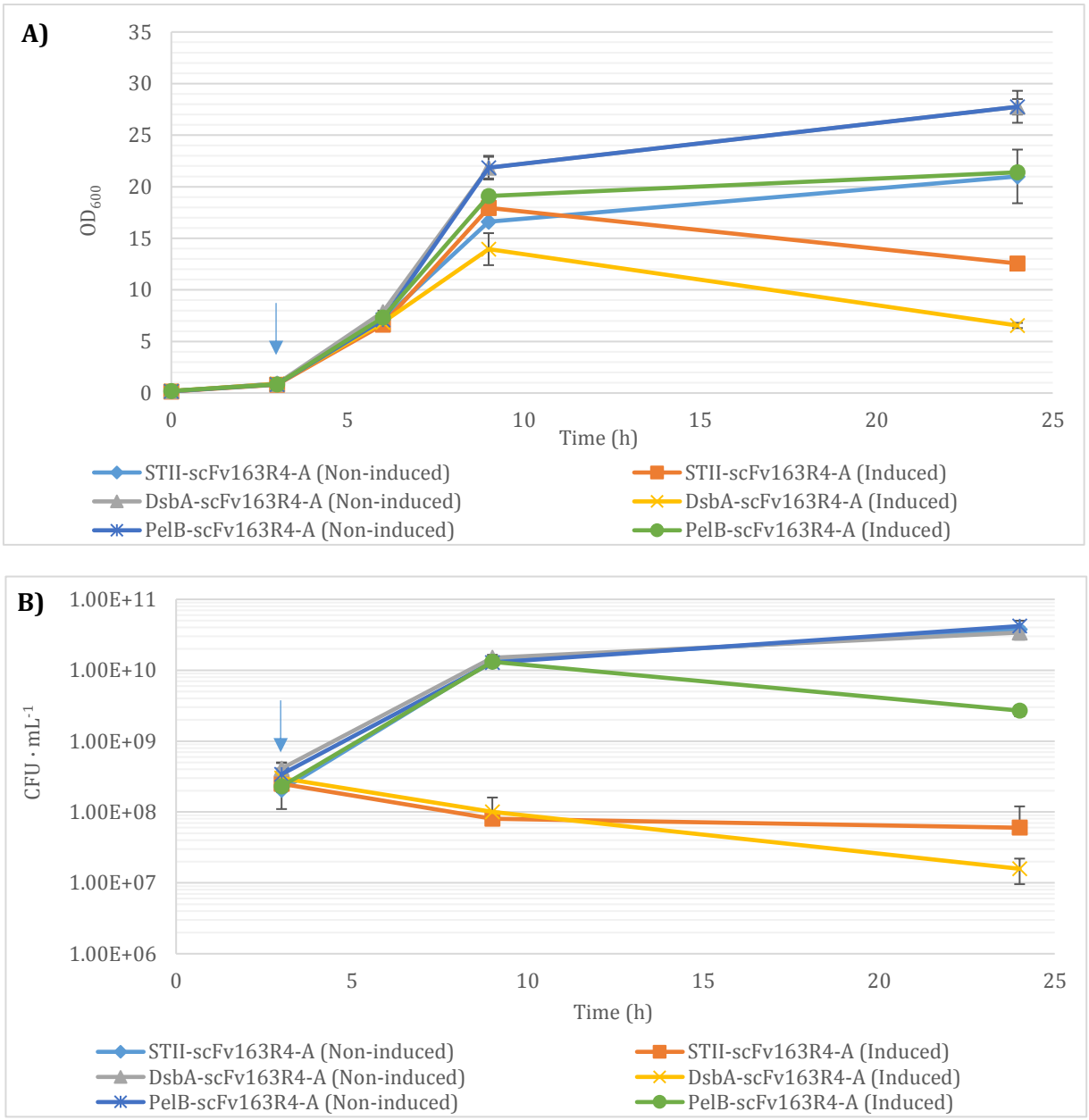


Figure 4.11. Effect of the removal of the extra alanine on growth and cell viability. *E. coli* BL21-A carrying the vector coding for the scFv163R4-A fused to the STII, DsbA and PelB signal peptides were grown at 25 °C. Cultures were induced with 0.02% of arabinose and repressed by the addition of 0.25% of glucose at an OD_{600} ≈ 0.5 (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 3 h, 9 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

In terms of cell culturability, a decrease in the number of colony forming units was observed for all the cultures grown under induced conditions. The decrease in the CFU was more severe for cultures carrying the STII and DsbA signal peptides (**Figure 4.11 B**). The loss of cell culturability can be already observed after 9 hours of growth, reaching $6 \cdot 10^7$ CFU · mL^{-1} and $1.6 \cdot 10^7$ CFU · mL^{-1}, 24 hours post-inoculation, for cultures carrying the STII and DsbA signal peptides, respectively. A less severe decrease in the CFU was observed for cultures carrying the PelB signal peptide, from $1.3 \cdot 10^{10}$ to $2.7 \cdot 10^9$ CFU · mL^{-1}, between 9 and 24 hours post-inoculation.

In terms of the production of the scFv163R4-A, the total accumulation of the antibody fragment was higher for cultures where the translocation of the scFv163R4 was directed by the STII and DsbA signal peptides (**Figure 4.12 A**). The production of the scFv163R4-A could also be detected in cultures where the translocation of the protein was facilitated by the PelB signal peptide, however, the production of the antibody fragment was lower when compared with cultures carrying the the STII and DsbA signal peptides. The accumulation of the scFv163R4 was detected in whole cell lysates, periplasmic and spheroplast protein fractions (**Figure 4.12 B**). Unfortunately, the concentration of scFv163R4-A detected in the periplasm was low, and the recombinant protein was detected mainly in the spheroplasts.

In general terms, the STII and DsbA signal peptides favoured a high production of the scFv163R4-A, but it was not correlated with a higher accumulation of the scFv163R4 in the periplasmic fraction. The high production of the scFv163R4-A was associated with growth arrest, cell lysis and low cell viability when compared with cultures grown under non-induced conditions. In contrast, cultures where protein translocation was facilitated by the PelB signal peptide showed a lower production of the scFv163R4-A when compared with the STII and DsbA signal peptides, nevertheless, a higher concentration of scFv163R4-A was detected in the periplasm. Better growth and cell viability was observed in cultures carrying the PelB signal peptide, and therefore, PelB seems to be a more appropiate signal peptide for the production of the scFv163R4-A.

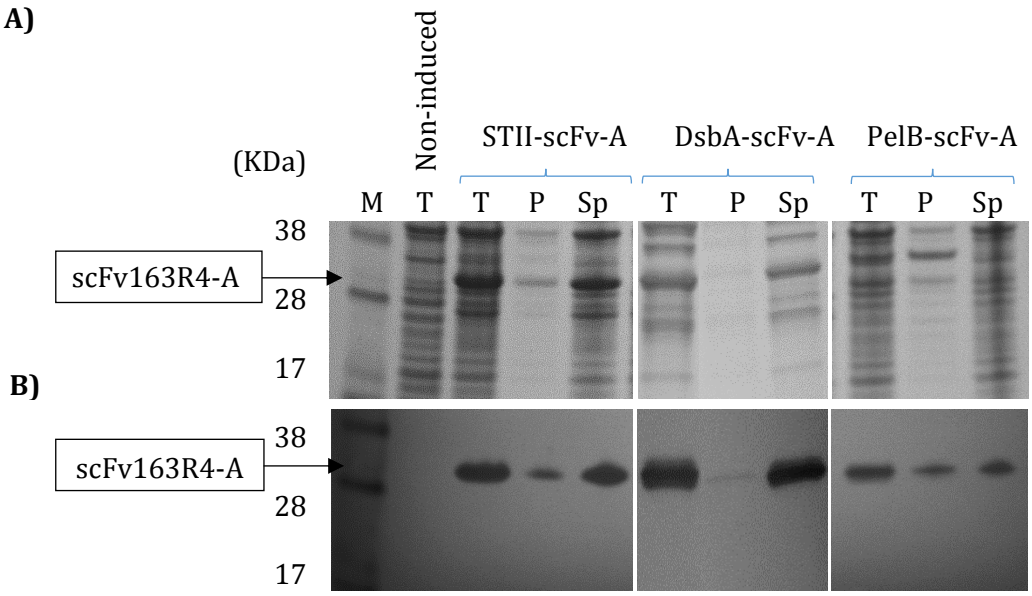


Figure 4.12. Effect of the removal of the extra alanine on the production of scFv163R4-A. SDS-PAGE gel showing the accumulation of scFv163R4-A from total (T), periplasm (P) and spheroplast (Sp) protein fractions. Samples were obtained from induced and non-induced cultures carrying the vector encoding the scFv163R4-A fused to the STII, DsbA and PelB signal peptides after 24 hours of growth. Cultures were induced with 0.02% of arabinose and repressed by the addition of 0.25% of glucose (A). The accumulation of the scFv163R4-A was detected by western blot using a mouse anti-myc antibody (B). Lanes shown here were obtained from different blots and rearranged for clarity.

In general terms, no great differences were observed compared to previous growth experiments, where the extra alanine at the first position of the amino acid sequence was still present. As the scFv163R4 was used as a model for periplasmic protein production, subsequent growth experiments were carried out by using the sequence coding for the scFv163R4-A, sometimes abbreviated to scFv-A, which it does not contain the extra amino acid.

4.2.6 Confirmation of the accumulation of scFv163R4-A in the periplasmic space

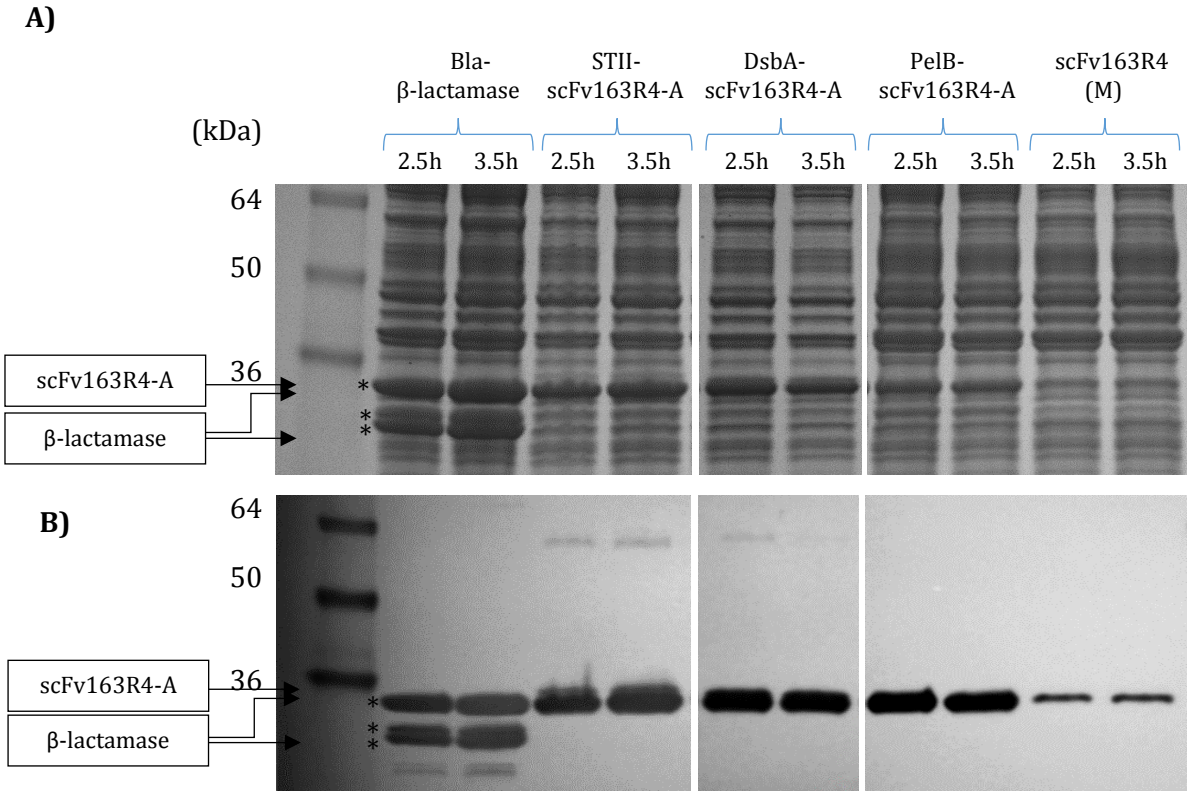
In general terms, the expression of the scFv163R4-A is initiated in the cytoplasmic space, where the protein is translated, and subsequently translocated to the periplasm, by the SEC or SRP pathways, depending on the signal peptide used. If the protein is translocated to the periplasmic space, the signal peptide is cleaved off, and the mature form of the protein is then accumulated in the periplasm. In contrast, if the scFv163R4 is not translocated, the antibody fragment is accumulated, as a precursor, in the cytoplasmic compartment.

The identification and quantification of the production of the scFv163R4-A was carried out by SDS-PAGE. SDS-PAGE gels and western blots were proven to be excellent tools to evaluate the production of the scFv163R4-A from whole cell lysates, nonetheless, it still remained difficult to determine whether the antibody fragment was being correctly translocated to the periplasm. Previous experiments have shown only one form of the protein being produced and no differences in molecular weight between the precursor (≈30 kDa) and the mature form (≈28 kDa) of the scFv163R4-A were observed by SDS-PAGE, either periplasmic or cytoplasmic.

In order to identify whether the scFv163R4-A is being correctly translocated to the periplasmic space, cultures were grown under "stressful" conditions as an attempt to favour the accumulation of the precursor protein, helping to identify whether the majority of scFv163R4 produced was accumulated as precursor or the mature form of the protein. Cultures carrying the vector coding for the signal peptides STII, DsbA or PelB fused to scFv163R4-A were grown in falcon tubes with

20 mL of terrific broth at 37 °C. Cultures carrying the vector coding for the cytoplasmic version of the scFv163R4 were used to facilitate the identification of the mature form of the antibody fragment. Cultures carrying a vector coding for β-lactamase were used as a used as model protein to help visualise the differences between the precursor and the mature form of the β-lactamase in a SDS-PAGE gel, since it has a similar molecular weight to scFv163R4. With the exception of the vector coding for the cytoplasmic version of scFv163R4, all the cultures were induced with 0.2% arabinose at an OD_{600} of 0.5. The production of the cytoplasmic scFv163R4 was induced with 1 mM IPTG (pET163R4 vector).

The production of the β-lactamase and scFv163R4-A in whole cell lysates was evaluated by SDS-PAGE and western blot. The precursor (31.6 kDa) and the mature (29.1 kDa) form of the β-lactamase can be easily identified in both colloidal blue stained SDS-PAGE and western blot (**Figure 4.13**). Two bands corresponding to the mature form of the β-lactamase were identified and they corresponded with the oxidised and reduced forms of the β-lactamase (Georgiou *et al.*, 1986). In contrast, the accumulation of the precursor form of the scFv163R4-A was not detected, and only one single protein band was identified in samples obtained from cultures overproducing scFv163R4-A. The scFv163R4-A, independently of the signal peptide used, showed a similar molecular weight to the cytoplasmic version of scFv163R4 (28.2 kDa), without a signal peptide, suggesting that it corresponded with the mature form of scFv163R4. No traces of the precursor form of the scFv163R4 were detected from samples obtained from cultures carrying the STII, DsbA or PelB signal peptides. Samples excised from SDS-PAGE gels were analysed by N-terminal sequencing to confirm whether the mature or the precursor form of the scFv163R4-A was being produced.



C)

	STII-scFv163R4-A	DsbA-scFv163R4-A	PelB-scFv163R4-A	scFv163R4 (M)
Major sequence	EVQLV	EVQLV	EVQLV	AEVQL
Minor sequence	MK–––			

Figure 4.13. Evaluation of the production of the precursor and mature scFv163R4-A. 10-20% Tris-Glycine SDS-PAGE gel showing the accumulation of scFv163R4-A from whole cell lysates. Samples were obtained from induced cultures carrying the vector encoding the STII, DsbA and PelB signal peptides fused to scFv163R4-A. Cultures expressing β-lactamase with its native signal peptide were used as control. Cultures were grown at 37 °C and induced with 0.2% of arabinose. The cytoplasmic form of the scFv163R4 was used as control and its production from the pET163R4 vector was induced by the addition of 1 mM IPTG. Samples were taken at 2.5 and 3.5 h after induction (A). The accumulation of the scFv163R4-A was detected by western blot using a mouse anti-myc antibody. The accumulation of the β-lactamase was confirmed by western blot using an anti-TEM-1 β-lactamase antibody. The precursor and the mature form of the β-lactamase can be observed and the two forms (reduced and oxidised) of the mature β-lactamase can be observed (asterisks) (B). Lanes shown here were obtained from different blots and rearranged for clarity. N-terminal sequencing results (C).

Whole cell lysates samples obtained from cultures where the translocation of the scFv163R4-A was directed by the STII, DsbA and PelB signal peptides, showed predominantly, the amino acid sequence corresponding to the mature scFv163R4-A, detected by N-terminal sequencing (**Figure 4.13 C**). No secondary amino acid sequences were detected for cultures where the translocation of the scFv163R4 was driven by the STII and PelB signal peptides. Samples obtained from cultures where the translocation of the scFv163R4-A was directed by DsbA showed traces of the secondary amino acid sequence corresponding to the precursor form of the protein, but the strongest signal detected corresponded to the mature form of the scFv163R4. The cytoplasmic version of the scFv163R4 was also sent for N-terminal sequencing, confirming that the amino acid sequence of the mature form of the scFv163R4 had been determined correctly.

N-terminal sequencing results confirmed that all the signal peptides evaluated, STII, DsbA and Bla, were able to direct the translocation of the signal peptide to the periplasmic space. Primarily the sequence coding for mature form of the protein was detected in samples obtained from whole cell lysates, confirming that the signal peptide was cleaved off. The scFv163R4 detected in the spheroplast from previous experiments (**Figures 4.3, 4.5, 4.7 and 4.12**) was probably the result of an inefficient extraction of the periplasmic protein content.

4.2.7 Effect of different signal peptides on the production of scFv163R4-A::β-lactamase

In order to further understand the effect of different signal peptides to direct the translocation of recombinant proteins to the periplasmic space, the sequence coding for the scFv163R4-A was fused by the C-terminus via a four amino acid linker to β-lactamase. The TEM-1 β-lactamase was selected as reporter protein, since it confers resistance to β-lactam antibiotics, such as ampicillin, and its activity can be quantified by measuring its enzymatic activity. In addition, it is a disulphide-

bonded protein, which is naturally translocated by *E. coli* to the periplasm, where it exerts its activity deactivating β-lactam antibiotics.

A growth experiment was carried out with the objective of evaluating the production of β-lactamase and the fusion protein scFv163R4-A::β-lactamase in the periplasmic space. Several plasmid constructs containing the sequence coding for β-lactamase were generated: wild type β-lactamase with its native signal peptide (Bla-β-lactamase); cytoplasmic versions of the β-lactamase (β-lactamase (M)) and scFv163R4-A fused to β-lactamase (scFv163R4::Bla (M)); scFv163R4-A::β-lactamase fused to the STII, DsbA, PelB and Bla signal peptides (STII-scFv163R4-A::Bla, DsbA-scFv163R4-A::Bla, PelB-scFv163R4-A::Bla and Bla-scFv163R4-A::Bla) (**Figure 4.14**). Cultures carrying the empty vector (pLBAD2) were used as a negative control. Cultures were grown in terrific broth at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. Cell culturability, as an indication of cell viability, was evaluated by serially diluted culture samples being plated onto non-selective LB agar.

In terms of growth, large differences were observed depending on the signal peptide used to direct the translocation of the scFv163R4-A::β-lactamase to the periplasm (**Figure 4.15 A**). Cultures carrying the DsbA and Bla signal peptides showed signs of growth arrest and cell lysis between 7 and 9 hours post-inoculation, reaching an OD_{600} of 2 and 4, after 24 hours of growth.

A healthier growth profile was observed for cultures where the translocation of the scFv163R4-A::β-lactamase was directed by the STII and PelB signal peptides, reaching an OD_{600} of 14, after 24 hours of growth (**Figure 4.15 A**). Growth arrest was not observed for the control cultures carrying the empty vector, the periplasmic β-lactamase and the cytoplasmic β-lactamase and scFv163R4-A::β-lactamase, reaching a final OD_{600} between 18 and 20, after 24 hours of growth.

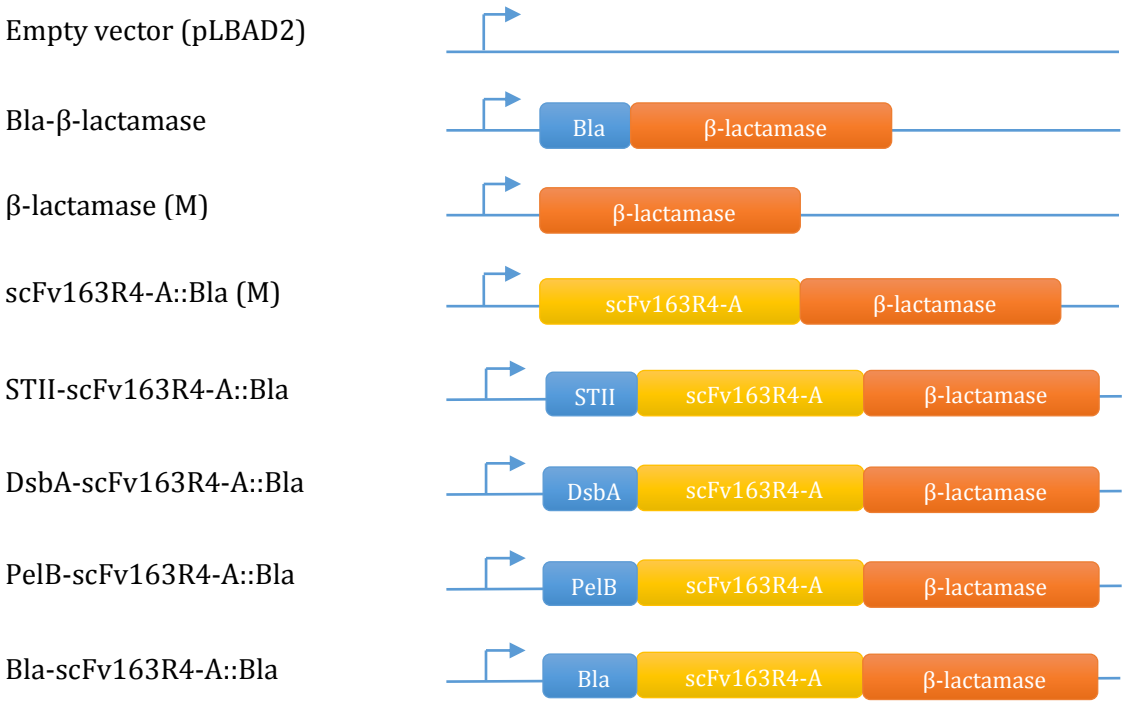


Figure 4.14. Representation of plasmid constructs generated for the β-lactamase screening. The signal peptide (blue), scFv163R4-A (yellow) and the reporter protein (orange). Cytoplasmic versions of β-lactamase and scFv163R4-A::β-lactamase are referred by the protein name followed by (M) as their sequence code for the mature form of the protein.

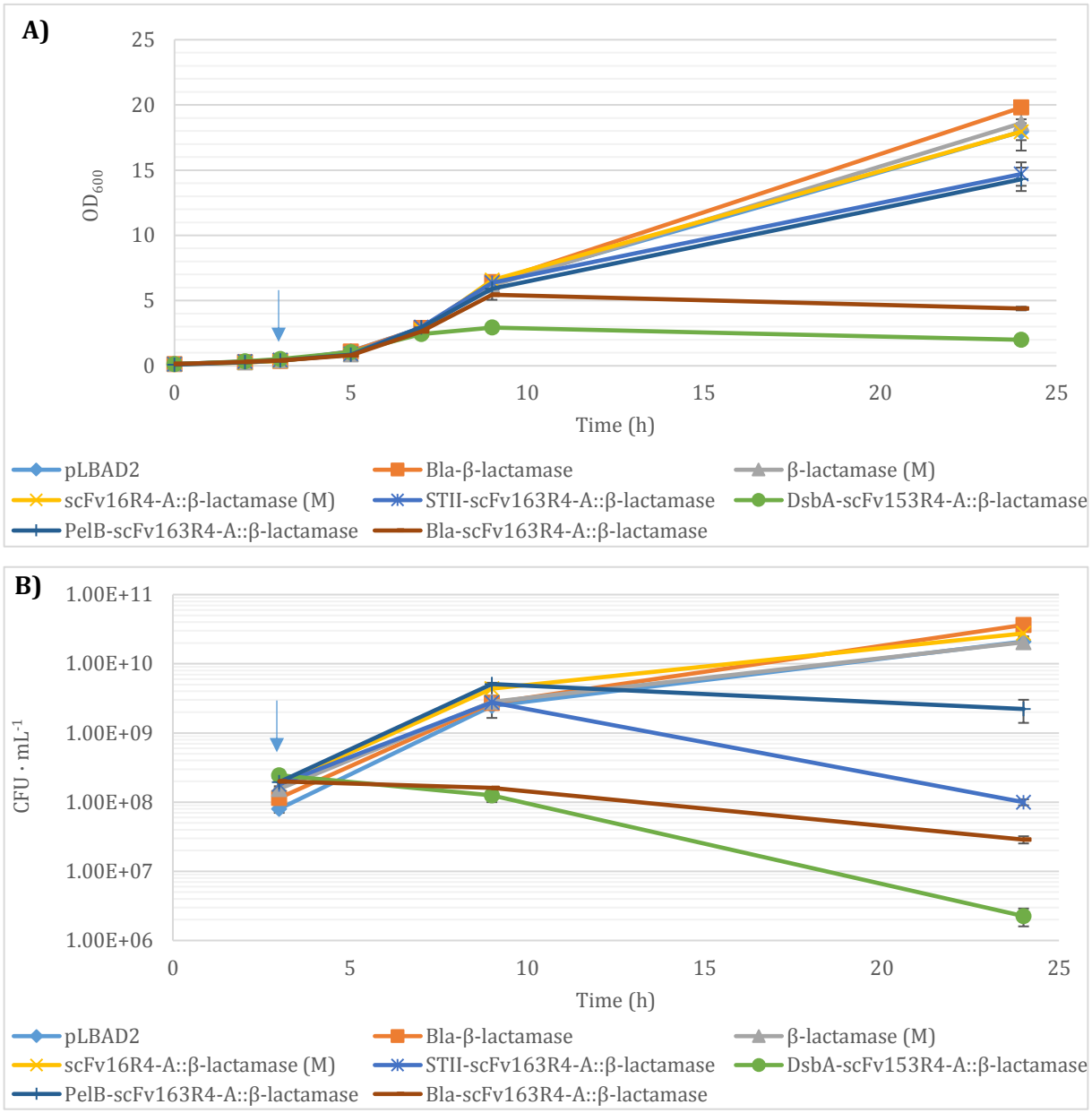


Figure 4.15. Evaluation of growth and cell viability of constructs generated for the β-lactamase screening. *E. coli* BL21-A carrying the empty vector (pLBAD2), the vector coding for β-lactamase, the cytoplasmic β-lactamase and scFv163R4-A::β-lactamase and the vector coding for the STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase were grown at 25 °C. Cultures were induced with 0.02% of arabinose at an OD_{600} ≈ 0.5 (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 3 h, 9 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data are shown as mean values from replica flasks, error bars are ±1 standard deviation.

Low cell culturability was observed for all the cultures producing the scFv163R4-A::β-lactamase in the periplasmic space (**Figure 4.15 B**). A decrease in the CFU was observed for cultures carrying the DsbA and the Bla signal peptides, reaching only $2.25 \cdot 10^6$ CFU $\cdot$ mL^{-1} and $2.9 \cdot 10^7$ CFU $\cdot$ mL^{-1} after 24 hours of growth. The decrease in the CFU was noticeable 6 hours post-induction, being less severe for cultures carrying the STII and PelB signal peptides. No loss of cell culturability was observed for any of the control cultures.

The production of β-lactamase in the periplasm and in the cytoplasm was also evaluated, as these two constructs will be used as control cultures to evaluate the activity of the fusion protein scFv163R4-A::β-lactamase by minimum inhibitory concentration experiments and by enzymatic assay. The production of β-lactamase was evaluated by SDS-PAGE gel and western blot. The periplasmic and spheroplast protein fractions were obtained by modified cold osmotic shock. A high concentration of β-lactamase was detected in samples corresponding to the total and periplasmic protein fractions (**Figure 4.16**). β-lactamase was also detected in the spheroplast fraction, possibly as a result of residual periplasmic protein content present in the sample. In contrast with previous experiments, the precursor form of the β-lactamase, containing the Bla signal peptide, was not detected (**Figure 4.13**). β-lactamase was also detected in culture medium samples (**Figure 4.16**). Samples obtained from cultures expressing the β-lactamase lacking its signal peptide (β-lactamase (M)) showed a high accumulation of the protein in the total and spheroplast protein fractions, and as expected, β-lactamase was not detected in the periplasmic fraction.

As a result of the high molecular weight of the fusion protein scFv163R4-A::β-lactamase (57.5 kDa) when compared to the scFv163R4-A (28.2 kDa), it was required to evaluate if the levels of production of the scFv163R4-A::β-lactamase would be similar to those obtained with the scFv163R4 (**Figure 4.10**) and scFv163R4-A (**Figure 4.12**). Large differences in the production of the scFv163R4-A::β-lactamase were observed depending on the signal peptide used to directed

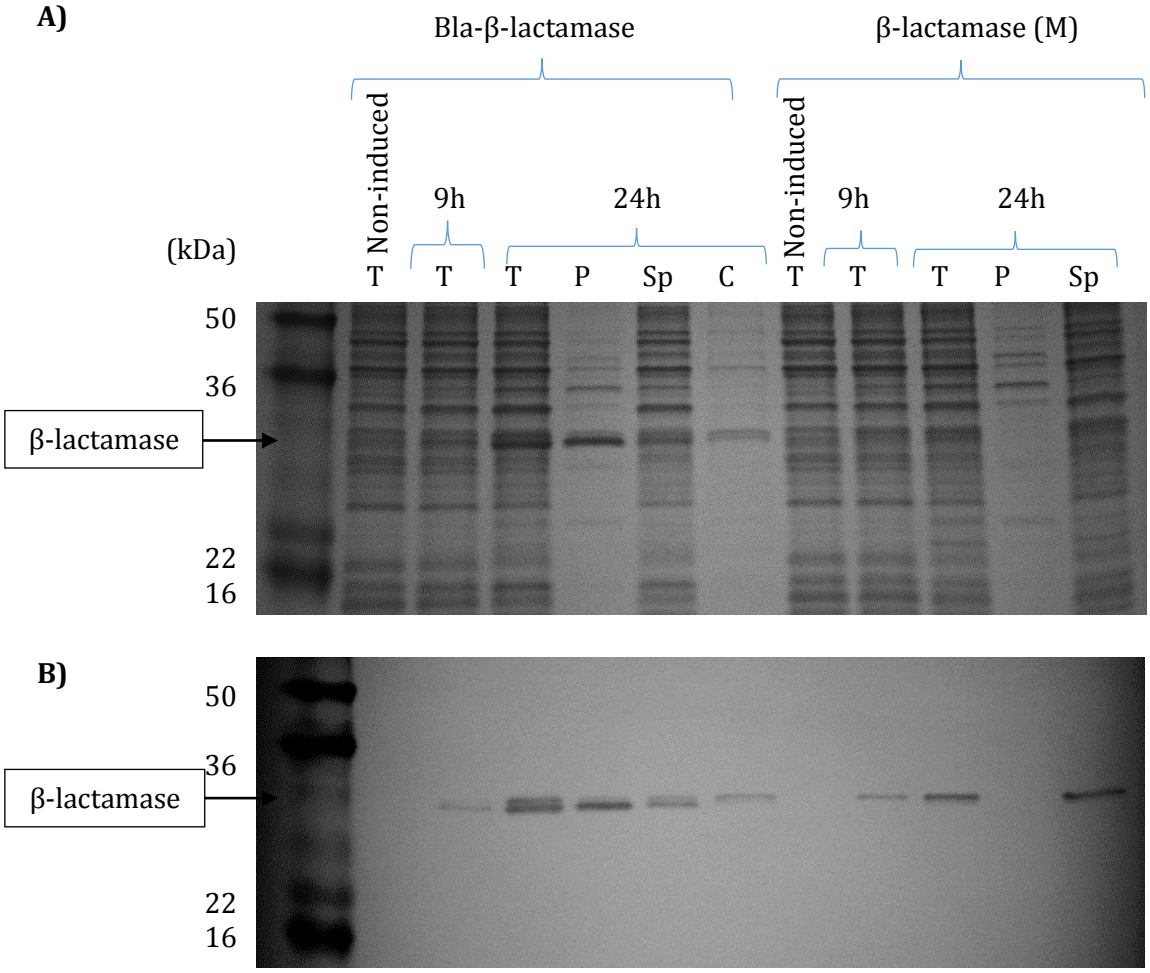


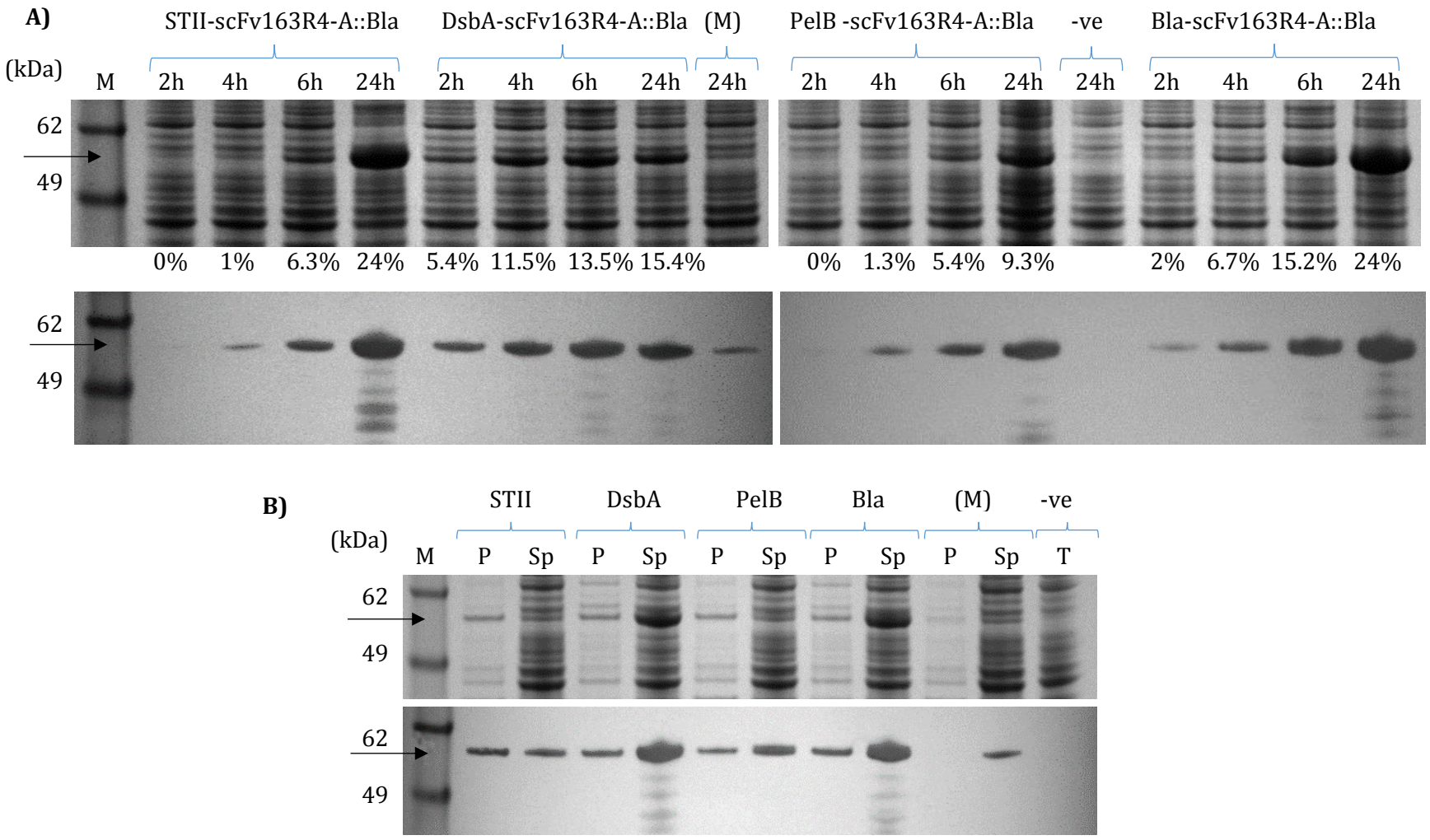
Figure 4.16. Evaluation of the production of β-lactamase. 18% Tris-Glycine SDS-PAGE gel showing the accumulation of β-lactamase from whole cell lysates (T), periplasm (P) and spheroplast (Sp) protein fractions or the culture medium (M). Samples were obtained from cultures containing the plasmid coding for the periplasmic (Bla-β-lactamase) and cytoplasmic β-lactamase (β-lactamase (M)), at 9 and 24 hours post-inoculation. The periplasm and spheroplast protein fractions were obtained by modified cold osmotic shock. Samples obtained from cultures grown under non-induced conditions after 24 hours of growth were used as a negative control (A). β-lactamase was detected by western blot using a mouse anti-β-lactamase antibody (B).

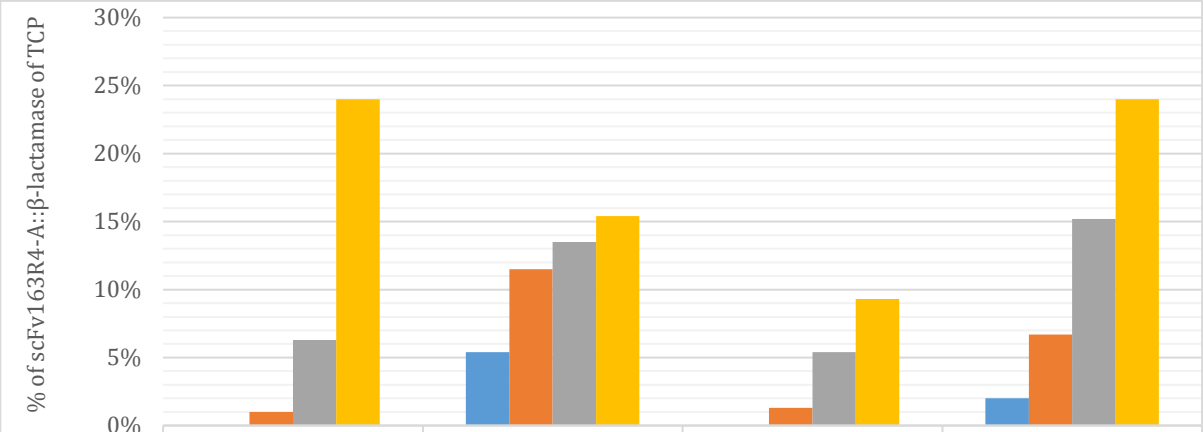
the translocation of the fusion protein to the periplasm (**Figure 4.17**). The differences observed are not only in terms of the amount of protein being produced but also with regards to the kinetics of its production. The use of the DsbA and Bla signal peptides seem to favour a rapid and high production of the fusion protein, which it can be already detected 2 hours post-inoculation, accounting for 5.4% and 2% of the total cell protein (TCP) (**Figures 4.17 and 4.18**). The production of the scFv163R4-A::β-lactamase reached 15.4% of TCP after 24 hours of growth for cultures carrying the DsbA signal peptide, lower than production levels achieved with STII and Bla signal peptides. This was probably the result of cell lysis and the leakage of the cell content to the culture medium as suggested by the low OD_{600} reached at the end of the experiment (**Figure 4.15**). Other signal peptides, as the STII leader peptide, seems to favour a slower production of the scFv163R4-A::β-lactamase, but obtaining a high protein production, reaching 24% of TCP, after 24 hours of growth. In contrast, the use of PelB as a signal peptide seems to favour a slower and lower production of the scFv163R4-A::β-lactamase, accounting for 9.3% of TCP after 24 hours of growth (**Figures 4.17 and 4.18**), which seems to correlate with a better growth and cell viability (**Figure 4.15**).

All the signal peptides evaluated in this experiment allowed the production of the fusion protein scFv163R4-A::β-lactamase in the periplasmic space. Unfortunately, a high concentration of the fusion protein was detected in the spheroplasts (**Figure 4.17 B**). The low concentration of scFv163R4-A::β-lactamase could be the result of the inefficient periplasmic extraction, as no differences in molecular weight, between the precursor (≈60 kDa) and the mature (57.5 kDa) protein were observed when compared with the cytoplasmic scFv163R4-A::β-lactamase.

In general terms, results obtained showed great differences in growth, cell viability, kinetics of protein production and translocation of the scFv163R4-A::β-lactamase to the periplasm. DsbA and Bla leader sequences favoured a rapid production of the fusion protein, but entailing a detrimental

Figure 4.17. Evaluation of the production of scFv163R4-A::β-lactamase using the STII, DsbA, PelB and Bla signal peptides. SDS-PAGE gel showing the accumulation of scFv163R4-A from total whole cell lysates obtained from induced cultures carrying the vector encoding the scFv163R4-A::Bla fused to the STII, DsbA and PelB signal peptide at 2, 4, 6 and 24 hours post-inoculation (A). The periplasmic (P) and spheroplast (Sp) protein fractions were obtained by modified cold osmotic shock from samples obtained 6 hours post-induction (B). The accumulation of the scFv163R4-A::Bla was detected by western blot using a mouse anti-myc antibody. Samples obtained from cultures carrying the empty vector taken after 24 hours of growth were used as a negative control. Samples obtained from cultures overproducing the scFv163R4-A::Bla in the cytoplasm (M) taken after 24 hours of growth were used to compare the migration of the mature form of the fusion protein on an SDS-PAGE gel. The quantity of scFv163R4 is expressed as a percentage of whole cell protein at the bottom of the gel.


A)
STII-scFv163R4-A::Bla
DsbA-scFv163R4-A::Bla
(M)
PelB -scFv163R4-A::Bla
-ve
Bla-scFv163R4-A::Bla
(kDa)
M
2h
4h
6h
24h
62
49
0%
1%
6.3%
24%
5.4%
11.5%
13.5%
15.4%
0%
1.3%
5.4%
9.3%
2%
6.7%
15.2%
24%
B)
STII
DsbA
PelB
Bla
(M)
-ve
P
Sp
T



	STII-scFv163R4-A::β-lactamase	DsbA-scFv163R4-A::β-lactamase	PelB-scFv163R4-A::β-lactamase	Bla-scFv163R4-A::β-lactamase
4 hours	0.0%	5.4%	0.0%	2.0%
6 hours	1.0%	11.5%	1.3%	6.7%
8 hours	6.3%	13.5%	5.4%	15.2%
24 hours	24.0%	15.4%	9.3%	24.0%

Figure 4.18. Quantification of the production of scFv163R4-A::β-lactamase using the STII, DsbA, PelB and Bla. Production of the scFv163R4-A::β-lactamase in cultures grown at 25 °C, induced with 0.02% arabinose at an OD_{600} of 0.5. The production of the scFv163R4 is represented as a percentage of the total cell protein (TCP) obtained after 24 hours of growth. Data shown here are single values obtained by quantifying the accumulation of the scFv163R4 in whole cell lysates by densitometry from SDS-PAGE gels.

effect on growth and cell culturability (**Figure 4.18**). In contrast, the use of PelB as a signal peptide yielded a lower concentration of the scFv163R4-A::β-lactamase but showed a better growth and cell viability, which is an essential requirement for its production by fed-batch fermentation. The accumulation of the scFv163R4-A::β-lactamase with regards to the total cell protein at each time point is depicted in **Figure 4.18**, quantified by densitometry from SDS-PAGE gels, as a representation of the kinetics of protein production of each signal peptide.

4.2.8 Minimum inhibitory concentration experiments

Minimum inhibitory concentration (MIC) experiments were carried out in order to further understand whether the scFv163R4-A::β-lactamase was being translocated to *E. coli's* periplasmic space and conferring protection against β-lactam antibiotics. In order to confer protection to β-lactam antibiotics, β-lactamase needs to be accumulated in the periplasm, to impede the binding of β-lactam antibiotics to the enzyme transpeptidase, which is required for the formation of peptidoglycan cross-links in the bacterial cell wall. MIC experiments are considered the "gold standard" to determine the sensitivity of a microbial organism to a chemical, the minimum inhibitory concentration being the lowest concentration of an antimicrobial that is able to prevent visible growth (Andrews, 2001).

E. coli BL21-A cultures carrying the empty vector (pLBAD2) or the vector containing the sequence coding for β-lactamase, the cytoplasmic β-lactamase or scFv163R4-A::β-lactamase, and the STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase were grown at 37 °C. At an OD_{600} of 1 to 2, cultures were serially diluted and plated onto Mueller-Hinton (M-H) agar containing 50 $\mu g \cdot mL^{-1}$ of kanamycin, 0.2% arabinose and different concentrations of ampicillin ranging from 3 to 1600 $\mu g \cdot mL^{-1}$. Control plates containing only kanamycin were used as a control. All the plates were supplemented with kanamycin to ensure plasmid maintenance during the experiment. Plates were incubated at 37 °C for 12 - 18 hours.

As expected, cultures used as a negative controls, the empty vector and the cytoplasmic β-lactamase and the scFv163R4-A::β-lactamase, were not able to grow on M-H agar plates containing ampicillin (**Figure 4.19**). Cultures overproducing β-lactamase with its native signal peptide, used as a positive control, showed normal visible growth up to a concentration of ampicillin of 200 $\mu g \cdot mL^{-1}$. Growth was observed up to a concentration of 1600 $\mu g \cdot mL^{-1}$, but the CFU decreased with increasing concentrations of ampicillin and colonies were smaller.

Great differences were observed from cultures expressing the fusion protein constructs depending on the signal peptide used to favour the translocation of the scFv163R4-A::β-lactamase to the periplasm (**Figure 4.19**). Visible growth was observed on plates seeded from cultures carrying the STII and PelB signal peptides, up to a concentration of ampicillin of 200 $\mu g \cdot mL^{-1}$ and 100 $\mu g \cdot mL^{-1}$, respectively. In contrast, agar plates seeded from cultures carrying the DsbA and Bla signal peptides showed no visible growth on any of the plates supplemented with ampicillin. Growth was only observed on the control plates containing kanamycin, although no colonies were visible on plates supplemented with 0.2% arabinose but without ampicillin, suggesting that the induction of recombinant protein production in cultures carrying the DsbA and Bla signal peptides had a detrimental effect in cell culturability. The low cell culturability correlated with the results obtained from previous experiments (**Figure 4.15 B**), the lack of colonies being the result of the effect of protein induction and not the deficiency in the translocation of the fusion protein to the periplasm.

With the objective to evaluate whether the lack of colonies was the result of the cell stress induced by high levels of production of the fusion protein, minimum inhibitory concentration experiments were repeated at lower cultivation temperature, 25 °C, and protein expression was induced with a lower concentration of arabinose, 0.02%, aiming to decrease the impact of protein induction on the cells.

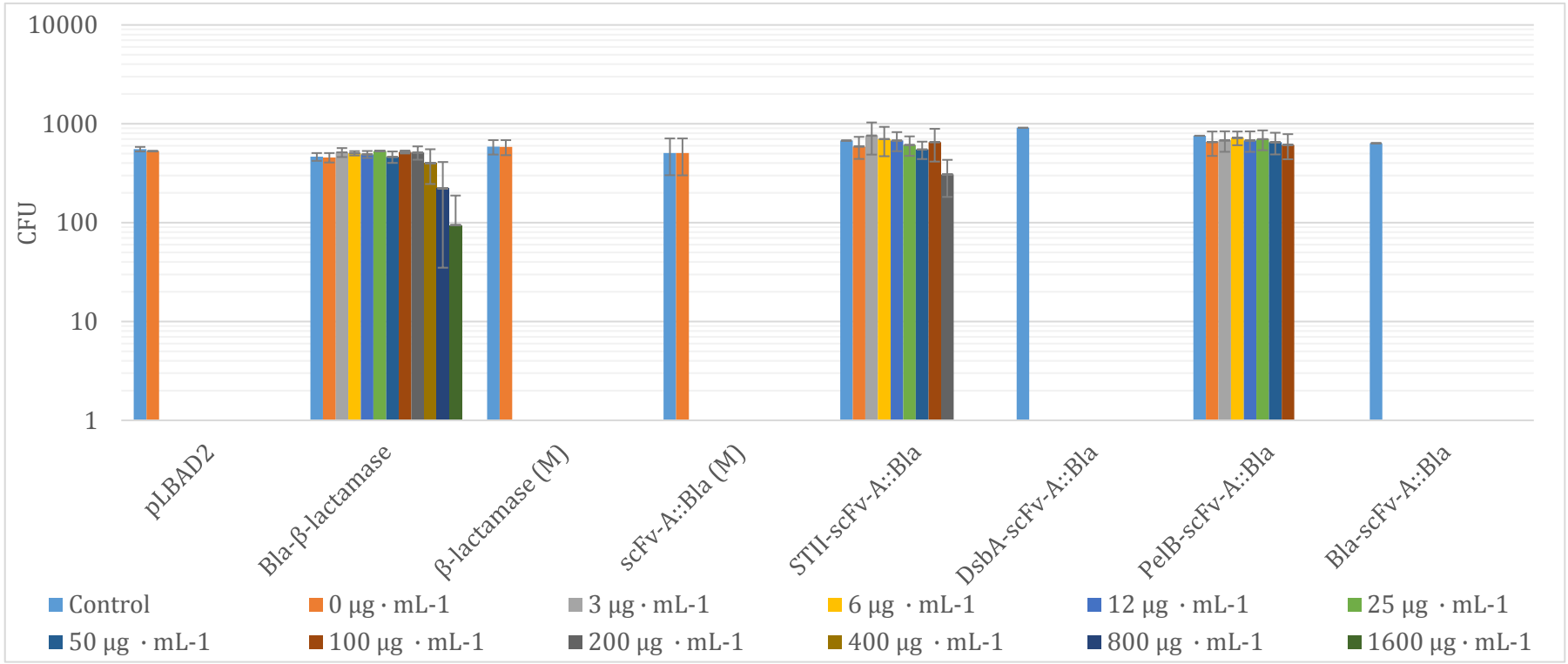


Figure 4.19. MIC evaluation of the production of β-lactamase and scFv163R4-A::β-lactamase at 37 °C with 0.2% arabinose. *E. coli* BL21-A carrying a plasmid containing the sequence coding for the empty vector (pLBAD2), the β-lactamase, the cytoplasmic β-lactamase and the scFv163R4-A::β-lactamase, and STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase were grown in TB at 37 °C. Samples serially diluted were plated onto Mueller-Hinton (M-H) agar containing kanamycin to a final concentration of 50 μg · mL^{-1}, 0.2% arabinose and different concentrations of ampicillin ranging from 3 to 1600 μg · mL^{-1}. An M-H agar plate supplemented only with kanamycin to a final concentration of 50 μg · mL^{-1} was used as control. M-H agar plates were incubated at 37 °C for 12 - 18 hours. Data are shown as mean values from two independent experiments, error bars are ±1 standard deviation.

When grown at 25 °C and induced with 0.02% arabinose, cultures overexpressing the wild type β-lactamase, showed visible growth up to a concentration of 100 μg · mL^{-1}; and less than 10 CFU were observed on plates containing 200 μg · mL^{-1} of ampicillin (**Figure 4.20**). No colonies were visible on plates supplemented with ampicillin for cultures carrying the empty vector, the cytoplasmic β-lactamase and the scFv163R4-A::β-lactamase, confirming that the translocation of the β-lactamase to the periplasm is essential to confer resistance to ampicillin. Under "non-stressful conditions", incubation at 25 °C and 0.02% arabinose, visible growth was observed on plates supplemented with different concentrations of ampicillin for all the cultures where the translocation of the scFv163R4-A::β-lactamase was favoured by a signal peptide (**Figure 4.20**). The minimum inhibitory concentration for cultures carrying the STII, DsbA, PelB and Bla signal peptides was 12, 50, 12 and 25 μg · mL^{-1}, respectively. Colonies were detected for all the cultures at higher concentrations of ampicillin, but the CFU was lower than the control plates and decreased with increasing concentrations of ampicillin.

In conclusion, MIC experiments showed that the scFv163R4::β-lactamase was being produced and translocated to the periplasmic space, demonstrated by the growth observed on M-H agar plates supplemented with ampicillin. Visible growth was observed for cultures carrying the DsbA and Bla signal peptides when grown under "non-stressful" conditions (**Figure 4.20**), in contrast with previous experiments where no growth was detected under induced conditions (**Figure 4.19**).

In addition, ampicillin selection can be used as a screening tool for selection of signal peptides which allows the translocation of the fusion protein scFv163R4-A::β-lactamase to the periplasmic space. Depending on the cultivation and induction conditions, MIC experiments can also be used to evaluate viability upon protein induction, and as an indication of which signal peptides allowed a high production of recombinant protein but entailing a detrimental effect on cell viability.

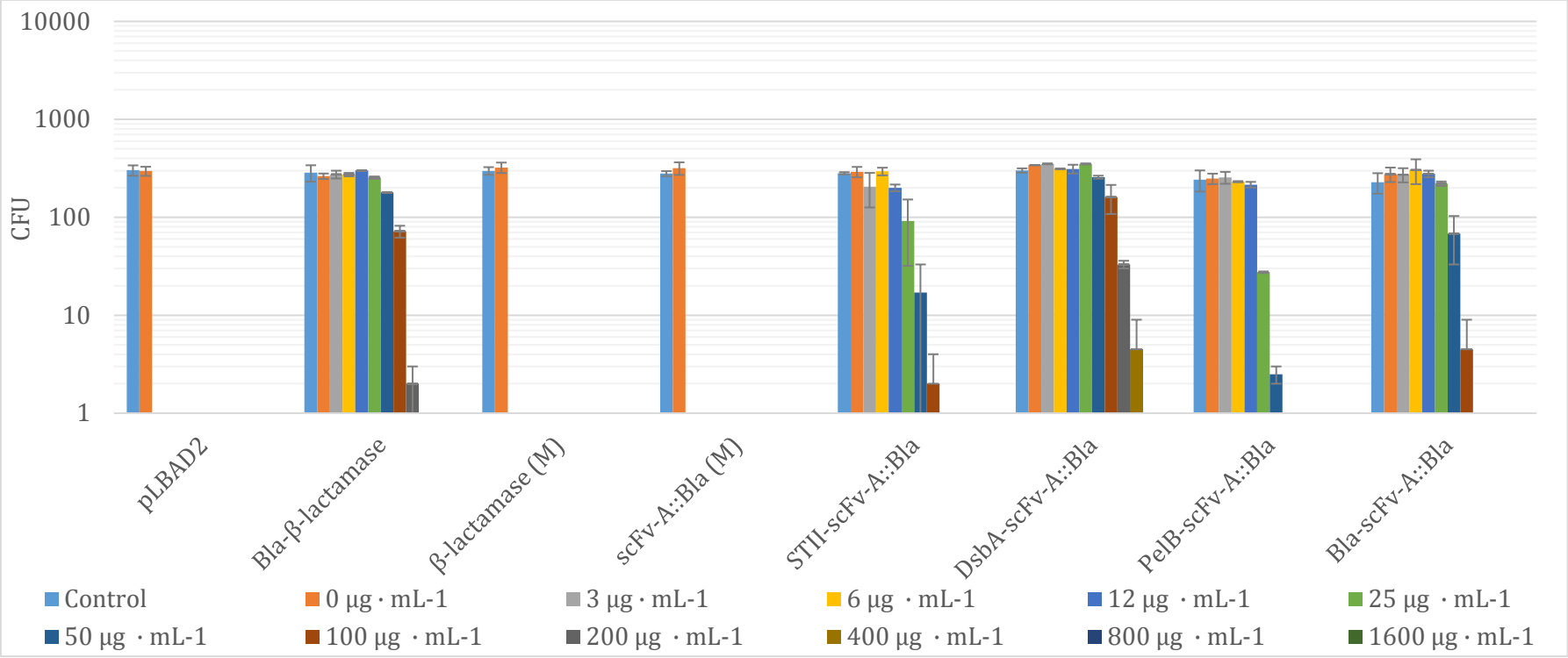


Figure 4.20. MIC evaluation of the production of β-lactamase and scFv163R4-A::β-lactamase at 25 °C with 0.02% arabinose. *E. coli* BL21-A carrying a plasmid containing the sequence coding for the empty vector (pLBAD2), the β-lactamase, the cytoplasmic β-lactamase and the scFv163R4-A::β-lactamase, and STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase were grown at 25 °C. Samples serially diluted were plated onto Mueller-Hinton (M-H) agar containing kanamycin to a final concentration of 50 μg · mL^{-1}, 0.02% arabinose and different concentrations of ampicillin ranging from 3 to 1600 μg · mL^{-1}. An M-H agar plate supplemented only with kanamycin to a final concentration of 50 μg · mL^{-1} was used as control. M-H agar plates were incubated at 25 °C for 48 hours. Data are shown as mean values from two independent experiments, error bars are ±1 standard deviation.

4.2.9 β-lactamase enzymatic assay for the production of scFv163R4-A::β-lactamase

Results observed during the MIC experiments showed that the four signal peptides evaluated allowed the translocation of the scFv163R4-A::β-lactamase to the periplasmic space, however, MIC experiments cannot be used to quantify the production of the fusion protein. With the aim to relatively quantify the production of the scFv163R4-A::β-lactamase, the activity of the β-lactamase was measured by enzymatic assay on whole live cells (Angus *et al.*, 1982). The activity of β-lactamase was evaluated by using nitrocefin, a chromogenic cephalosporin, as a substrate. β-lactamase catalyses the hydrolysis of the nitrocefin β-lactam ring, producing a shift of ultraviolet absorption from intact nitrocefin (yellow ~380 nm) to degraded nitrocefin (red ~495 nm) allowing the quantification of the activity of the β-lactamase. *E. coli* BL21-A carrying the empty vector or the vector containing the sequence coding for β-lactamase, the cytoplasmic β-lactamase, the cytoplasmic scFv163R4-A::β-lactamase and the STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase were grown at 25 °C and induced with 0.02% arabinose 2 hours post-inoculation. Upon nitrocefin addition, the relative activity of β-lactamase was determined by continuous measurement of the OD_{495} from cells obtained before induction (2h) and after induction (4, 6, 8 and 24 hours post-inoculation). The initial velocity (slope) was calculated by linear regression, adapted to obtain an $R \geq 0.9$, and compared between the different samples, as a quantification of the relative β-lactamase activity.

As expected due to tight regulation of the arabinose promoter, no β-lactamase activity was detected for any of the cultures before induction, 2 hours post-inoculation (**Figure 4.21**). No significant β-lactamase activity was detected at any time point for cultures carrying the empty vector (pLBAD2), used as negative control. Cultures carrying the wild type β-lactamase, used as a positive control, showed high levels of β-lactamase activity. Unexpectedly, the cytoplasmic version of β-lactamase (β-lactamase (M)), did also show a high β-lactamase activity post-induction. It has been previously reported than β-lactamase produced in the cytoplasm could

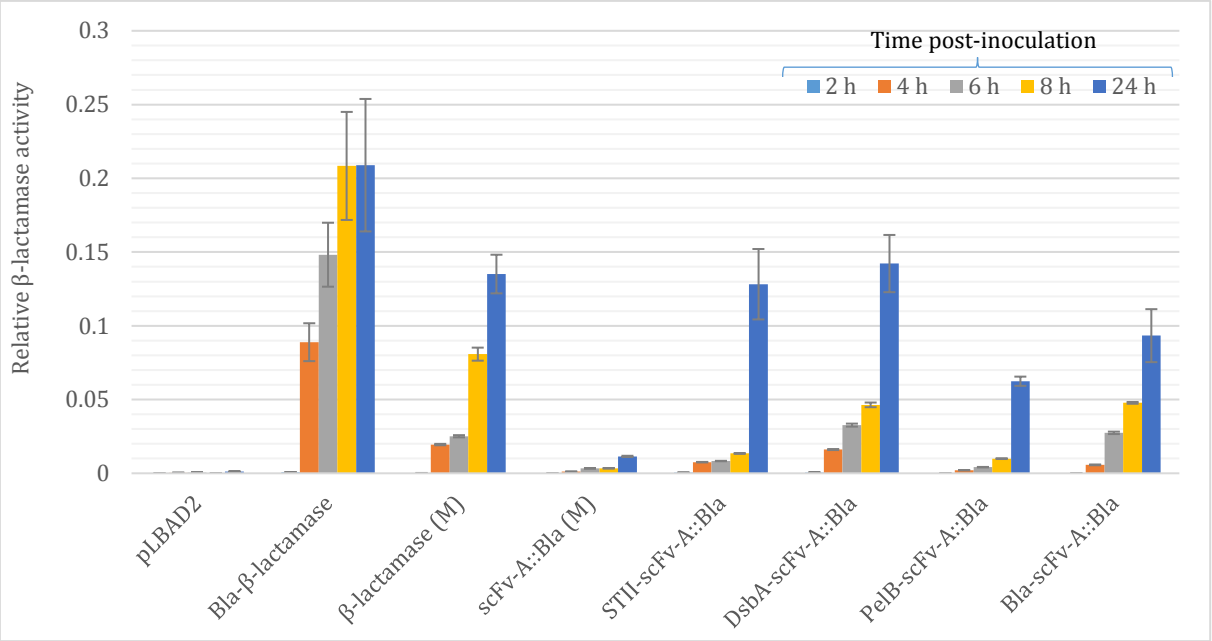


Figure 4.21. Evaluation of β-lactamase activity by enzymatic assay. *E. coli* BL21-A carrying the empty vector (pLBAD2), a plasmid containing the sequence coding for the β-lactamase, the cytoplasmic β-lactamase and the scFv163R4-A::β-lactamase, and STII, DsbA, PelB and Bla signal peptides fused to the scFv163R4-A::β-lactamase construct were grown at 25 °C and induced with 0.02% arabinose after 2 hours of growth. Whole cells samples were evaluated live for β-lactamase activity by quantification of the hydrolysis of the substrate nitrocefin by continuous measurement of the OD_{495} before induction (2h) and after induction (4, 6, 8 and 24 hours post-inoculation). The OD_{600} was also quantified at each time point and used to normalise the data obtained by the nitrocefin enzymatic assay. The initial velocity (slope) was calculated by linear regression, adapted to obtain an $R \geq 0.9$, and compared between the different samples. The error bars represent the 5% uncertainty in the calculation of the slopes.

retain some degree of activity (Kadonaga *et al.*, 1984) and therefore, being able to catalyse the hydrolysis of the nitrocefin. It is unclear if the substrate, nitrocefin, is able to diffuse through the cytoplasmic membrane, resulting in the detection of high levels of β-lactamase activity. In contrast, *E. coli* cells expressing the fusion protein (scFv163R4-A::β-lactamase (M)) in the cytoplasm, did not show high levels of β-lactamase activity, probably due to low production of the scFv163R4-A::β-lactamase or the proteolytic degradation thereof.

Cultures where the production of the scFv163R4-A::β-lactamase was directed to the periplasmic space by the STII, DsbA, PelB and Bla signal peptides, showed great differences in terms of β-lactamase activity (**Figure 4.21**). For cultures carrying the DsbA and Bla signal peptides, a high β-lactamase activity was detected 2 hours post-induction, which increased over time, showing rapid production of the protein early after induction. Cultures where the translocation of the scFv163R4-A::β-lactamase was directed by the STII signal peptide, the β-lactamase activity increased gradually during the first hours post-induction, but reaching a high level after 24 hours of growth. In contrast, for cultures carrying the PelB signal peptide, low β-lactamase activity was detected during the first hours post-induction, which reached a lower level of β-lactamase activity after 24 hours of growth than cultures carrying the STII, DsbA and Bla signal peptides.

Overall, the β-lactamase enzymatic assay has proven to be a useful tool for the quantification of the production of the scFv163R4-A::β-lactamase, not only as an indication of total protein production but also regarding how rapidly the protein is being produced. Unfortunately, β-lactamase also showed some degree of activity when produced and accumulated in the cytoplasm. However, low levels of activity were detected when the cytoplasmic compartment was targeted for the production of the fusion protein, the scFv163R4-A::β-lactamase (**Figure 4.21**).

In order to seek confirmation of whether the fusion protein scFv163R4-A::β-lactamase was being correctly translocated to the periplasmic space, samples obtained from whole cell lysates of cultures carrying the PelB signal peptide fused to the scFv163R4-A::β-lactamase were sent for N-

terminal sequencing. Results showed that the protein was being translocated to the periplasm, and only the mature sequence of the protein lacking the leader peptide was detected. No traces of minor sequences were detected in the sample (data not shown).

4.3 Discussion

The development of a successful periplasmic protein production process depends on the healthy balance between gene regulation, cell physiology and protein folding. In an attempt to maximise recombinant protein yields, a common strategy is the use of strong regulatory sequences to achieve a high transcription, translation and secretion efficiency (Fu *et al.,* 2005). Unfortunately, this approach often causes cell stress induced by the high level of protein production, resulting in the accumulation of incorrectly folded and aggregated protein products (Plückthun *et al*., 1996).

Stress minimisation conditions, such as low inducer concentrations in combination with low cultivation temperatures, were applied to the production of the scFv163R4, in order to improve not only the total protein yield but also the amount of target protein being translocated to the periplasmic space. In general terms, the use of lower cultivation temperatures favoured the production of the scFv163R4, as well as growth and cell viability, and a higher concentration of recombinant protein was found to be accumulated in the periplasm (**Figure 4.3**). This results correlated with previously published data that suggested that the optimal cultivation temperature for the production of antibody fragments is 25 °C (Plückthun *et al*., 1996).

4.3.1 Catabolite repression as a tool to modulate the activity of the arabinose promoter

The successful expression of high quantities of antibody fragments in *E. coli's* periplasm is a compromise between protein synthesis and protein folding, in order to maximise the amount of functional protein produced (Plückthun *et al*., 1996). Wycuff and Matthews (2000) showed that

the use of glucose, which repress the activity of the arabinose promoter, in combination with a broad range of arabinose concentrations could be used as strategy to adjust protein expression levels. The expression of recombinant proteins is therefore modulated by the addition of arabinose, used as inducer, and glucose, used as a repressor, helping to achieve a greater control over protein expression.

The strong repression of the arabinose promoter by the addition of 0.5% glucose was observed when combined with low arabinose concentrations, from 0.002% to 0.0002%. At these concentrations, no protein accumulation was detected in samples obtained from whole cell lysates (**Figure 4.5**). In contrast, at a higher concentration of arabinose, 0.2%, the slow de-repression of the arabinose promoter favoured the production of the scFv163R4 and the translocation of the protein to the periplasmic space (**Figure 4.5**). A concentration of glucose of 0.25%, in combination with a concentration of arabinose of 0.02% arabinose, showed a similar accumulation in the periplasm, without compromising growth and cell viability (**Figures 4.6 and 4.7**).

In general, glucose repression combined with different concentrations of arabinose provided a greater control over the level of protein expression (Wycuff and Matthews, 2000). However, even when this approach has been proven to be useful to reduce the accumulation of certain recombinant proteins in IBs and favour the translocation of proteins to the periplasm, as interferon α (Lee and Jung, 2007), this approach would not feasible at large scale. The production at fermentation scale without compromising the control over the growth rate would have been difficult, due to the preferential use of glucose as a carbon source (Aidelberg *et al.*, 2014).

4.3.2 Selection of signal peptides for periplasmic protein production

The optimal selection of the leader sequence is the first step for the efficient secretion of a recombinant protein and the level of production often has a great impact in growth and cell viability (Yoon *et al.,* 2010). There are three main translocation pathways, the Sec, the SRP and the

Tat, commonly used for periplasmic protein production in *E. coli*, the Sec pathway being the major gateway for protein secretion (Fu et *al.,* 2005).

The great differences observed between different signal peptides, in terms of protein production kinetics, had a strong effect on growth and viability. Fu *et al.* (2005) described growth arrest and cell death induced by the "saturated translocation" of a secretory exoglucanase when the conversion of pre-mature to mature exoglucanase was unable to match the production levels. Similarly, results obtained when DsbA was used as a leader sequence to favour the translocation of the scFv163R4, a high production of the recombinant protein was detected 6 h post-induction, even in the presence of glucose (**Figures 4.10**). The high level of protein production was also accompanied by growth arrest, loss of cell viability and cell lysis with and without glucose repression (**Figure 4.8 and 4.11**). The high levels of production of the precursor form of the scFv163R4-A could have saturated the translocation channel or could be the result of SRP depletion. Since SRP mediates the translocation of inner membrane proteins, whereas the majority of secretory and outer membrane proteins are translocated by the SecB-dependent pathway, its depletion may also have a strong negative effect on protein homeostasis in the entire bacterial envelope resulting in cell lysis and death (Zhang *et al.,* 2012).

In contrast, the PelB signal peptide favoured a lower production of the antibody fragment and a lower final yield (**Figure 4.10**), but showed a healthier growth profile, with minimum loss of cell viability upon induction with and without glucose repression (**Figure 4.8 and 4.11**). Nonetheless, PelB favoured the translocation of the protein to the periplasm, and only the sequence corresponding to the mature protein was detected in whole cell lysates obtained from cultures overexpressing either the scFv163R4-A and or the scFv163R4-A::β-lactamase when analysed by N-terminal sequencing (**Figure 4.13**).

4.3.3 β-lactamase as a reporter protein for periplasmic RPP

The effect of different signal peptides on cell growth and cell viability was simple to evaluate, but whether the protein was successfully being translocated to the periplasmic space was still hard to identify. N-terminal sequencing can provide with the definitive result by determining whether the leader sequence is still present, however, this is an expensive test that cannot be applied for screening of a large number of samples. The fusion of recombinant proteins by the C-terminus to β-lactamase can be used instead, to give a more detailed insight of the localisation of the recombinant protein as well as its possible use for high-throughput screening experiments.

With this tool in hand, it was easier to establish a link between growth, cell viability, protein production and the amount of protein being translocated to the periplasm, determining that some signal peptides allowed a high and rapid production of recombinant protein but entailed a deleterious effect on growth and cell viability, such as DsbA and Bla signal peptides (**Figures 4.15 to 4.18**). This effect was independent of the translocation pathway being targeted, SRP pathway for DsbA and the Sec pathway for Bla, suggesting that even when SRP depletion may have a strong effect on bacterial physiology, cell death was probably caused by the rapid production of the scFv163R4-A::β-lactamase. This result suggests a strong relationship between not only the amount of protein being produced but also the kinetics of protein production and how the saturation of the translocation pathways may affect cell physiology (Fu et *al.,* 2005).

The effect on cell physiology was so severe, that no colonies were obtained, independently of the concentration of ampicillin, when high concentrations of inducer were present in M-H agar plates (**Figure 4.19**) for cultures carrying the DsbA and Bla signal peptides. This was not the result of insufficient production of scFv163R4-A::β-lactamase to confer protection to ampicillin, but the low cell fitness and growth impairment caused by the production and translocation of the scFv163R4-A::β-lactamase. Colonies were only obtained after a 10-fold decrease of the arabinose concentration, and a reduction in the cultivation temperature from 30 °C to 25 °C (**Figure 4.20**),

used to slow down the protein synthesis rate. The quick and high levels of protein produced early after induction favoured by DsbA and Bla signal peptides was confirmed by the quantification of the relative β-lactamase activity (**Figure 4.21**), and correlates with the results observed by SDS-PAGE and western blot (**Figures 4.17 and 4.18**).

In contrast, when the translocation of the scFv163R4-A::β-lactamase was directed by the STII signal peptide, high accumulation of the protein can be observed after 24 hours of growth (**Figure 4.17**) but no growth arrest or cell lysis was observed (**Figure 4.15 A**). A decrease in the CFU was observed upon induction, but the loss of cell culturability was less severe than when DsbA and Bla signal peptides were used (**Figure 4.15 B**). A similar profile was observed when PelB was used as a signal peptide, showing a slightly lower protein production (**Figure 4.17**), but better cell culturability (**Figure 4.15 B**). Both signal peptides were able to produce enough protein in the periplasmic space to allow the growth of colonies in the presence of high concentrations of ampicillin at both cultivation temperatures and concentrations of arabinose evaluated (**Figures 4.19 and 4.20**).The production of functional scFv163R4-A::β-lactamase was also confirmed by enzymatic assay, where the relative β-lactamase activity increased over the time after induction, reaching the highest levels after 24 hours of growth, with the STII samples showing slightly higher β-lactamase activity than PelB (**Figure 4.21**).

4.4 Conclusions and future work

The selection of the optimal signal peptide for periplasmic protein production should not be evaluated solely on the basis of the levels of protein production but also considering growth and cell viability as key for the development of a successful production process. As high induction and rapid accumulation of recombinant proteins may lead to growth arrest, signal peptides that favour a gradual increase in protein production over time may result in a better growth and cell viability.

This is essential for the development of high-cell-density fermentation processes, since protein production can be maintained for a longer period of time.

There is no general rule selecting a signal peptide that guarantees successful translocation of the protein to the periplasm. Several signal peptides need to be evaluated in a trial-and-error approach, an often costly and time-consuming process (Choi *et al.*, 2004). The development of a high-throughput screening would not only help to evaluate the different signal peptides available from nature but also the rapid assessment of signal peptide libraries, containing millions of signal peptide variations. Signal peptide libraries, either generated by chemical synthesis or mutagenesis, may massively expand the catalogue of leader sequences or provide variations of the existing ones with improved properties. Therefore, the development of a screening assay for the evaluation of signal peptide libraries would be a desirable accomplishment, but to date no reporter assays have been developed (Mansell *et al.,* 2010). As β-lactamase has proven to be a great tool as a reporter protein to evaluate production and activity of proteins in the periplasmic space, its use expands the future possibilities for the screening and evaluation of the production of recombinant proteins in the periplasm.

Chapter 5:

β-lactamase screening assay for periplasmic protein production

5.1 Introduction

E. coli is quickly gaining status as a reliable host organism for periplasmic protein expression. However, despite the recent advances, periplasmic expression of recombinant proteins is a still a significant challenge. A better understanding of periplasmic translocation pathways has accelerated the development of new screening and applications in this field, specially, the understanding of signal peptides that facilitate the translocation of recombinant proteins to the periplasm (Low *et al.,* 2013). In addition, optimisation of the conditions for periplasmic protein production is often a costly, tedious and time-consuming process. For this reason, systematic studies are essential to determine the relationship between cell physiology, cultivation conditions and the cell production capability (Rodriguez-Carmona *et al.,* 2012).

Signal peptides direct the translocation of recombinant proteins to their site of function, such as the periplasm, the outer membrane or the culture medium; the selection of the optimal signal peptide is essential for the successful production of proteins in the periplasm. Several studies have expanded the knowledge regarding the structure of signal peptides, and it is well known that signal peptide sequences are often constituted by a short positively charged n-region, followed by a non-polar hydrophobic core and a neutral polar region, which is found immediately preceding the cleavage site (Heggeset *et al.,* 2013). A more detailed description of the structure of the signal peptide can be found in **Chapter 1, Section 1.3.2.1**. However, even when the structure of signal peptide has been described and the production of disulphide bonded-proteins in the periplasm has been established, no reporter assays have yet been developed to evaluate protein translocation (Mansell *et al.,* 2010).

In the last decades, directed evolution has emerged as a powerful method that mimics the process of natural selection for altering the properties of proteins (Heggeset *et al.,* 2013). In directed evolution experiments, mutations are introduced at random positions of a target DNA sequence, generating large and diverse libraries. Selection strategies or high-throughput screening are

required to select and identify those clones containing mutations that improve the desired property (Hanson-Manful and Patrick, 2013). As in directed evolution experiments for selection of proteins, signal peptide sequences can be evolved by random mutagenesis, which allows the creation of signal peptide libraries which can be evaluated by selection of clones with improved translocation properties and screening of individual variants which exert a higher translocation activity.

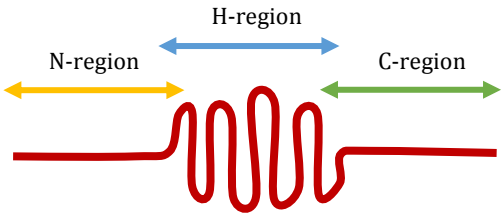
With the aim to evaluate the translocation of scFv163R4-A to the periplasmic space, a protein fusion was generated where a signal peptide was fused to the N-terminus of the protein of interest, scFv163R4-A, and the mature TEM-1 β-lactamase (Bla) sequence was fused to the C-terminus of the scFv163R4-A. The use of β-lactamase as a reporter protein allows selection using β-lactam antibiotics and the quantification of the production and translocation of the fusion protein to the periplasmic space by enzymatic assay.

The steps involved in the development of β-lactamase screening for periplasmic protein production are depicted in **Figure 5.1**. Initially, signal peptide libraries were generated by error-prone PCR or by chemical synthesis of variants of the PelB signal peptide sequence (1). Signal peptide libraries were introduced by seamless cloning into a plasmid vector containing the gene coding for the scFv163R4-A::β-lactamase, under the control of the arabinose promoter (2). Plasmid libraries were transformed in *E. coli* (3) and clones containing a variant of the signal peptide which allowed the translocation of the scFv163R4-A::β-lactamase to the periplasm were selected on ampicillin-supplemented agar plates (4). This step allowed the selection and isolation of only those clones containing signal peptide sequences, which allowed the production of the fusion protein in the periplasm. Subsequently, clones selected at high concentrations of ampicillin were individually screened using the β-lactamase enzymatic assay, allowing the relative quantification of β-lactamase activity of each individual clone (5).

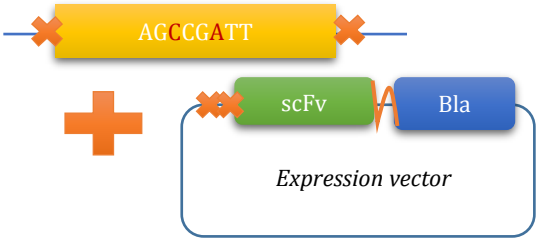
Figure 5.1. Diagram of the β-lactamase screening for periplasmic protein production. The development of the β-lactamase screening comprises seven main stages: (1) generation of signal peptide libraries, either by error-prone PCR or chemical synthesis of oligonucleotide libraries, both containing random mutations in the signal peptide sequence. (2) Seamless cloning of signal peptide libraries using type IIs restriction enzymes, containing the sequence coding for the protein of interest fused to the reporter protein β-lactamase, under the control of the arabinose promoter. (3) Introduction of the plasmid DNA into *E. coli*, amplification of the vector containing the signal peptide libraries and generation of research cell banks containing the signal peptide libraries. (4) Selection of clones containing signal peptides favouring the translocation of the protein of interest fused to β-lactamase. The production of the protein of interest fused to β-lactamase in the periplasmic space confers *E. coli* cells resistance to β-lactam antibiotics, such as ampicillin, allowing the selection of only those clones containing signal peptides which allowed the production and translocation of the fusion protein to the periplasm. (5) Screening of positive clones and relative quantification of the β-lactamase activity of host cells. Clones showing a high β-lactamase activity were selected to be further evaluated for periplasmic protein production by shake-flask growth experiments and fed-batch fermentation. (6) Evaluation of highly productive clones, containing signal peptide (SP) sequences with a higher production of the recombinant protein in the periplasm than the wild type signal peptide sequence. (7) Evaluation of selected clones by high cell density fermentation for the production of the protein of interest (POI) fused to the reporter protein β-lactamase in the periplasmic space. Subsequently, evaluation of selected clones by high cell density fermentation for the production of recombinant protein of interest in the periplasm, without the reporter protein, β-lactamase.

1. Generation of signal peptide libraries

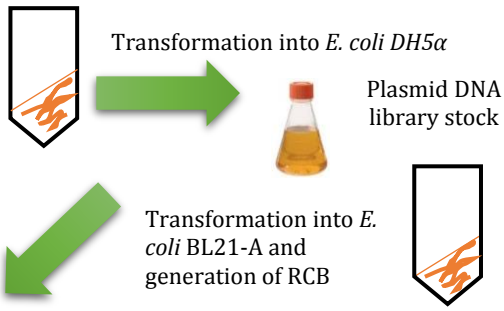
- *Random mutagenesis signal peptide library (Error-prone PCR library)*
- *Chemically synthesised signal peptide library (Twist library)*



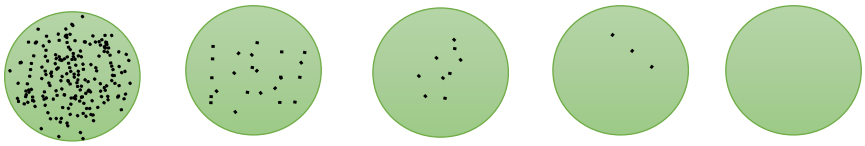
2. Cloning of the signal peptides libraries into the expression vector



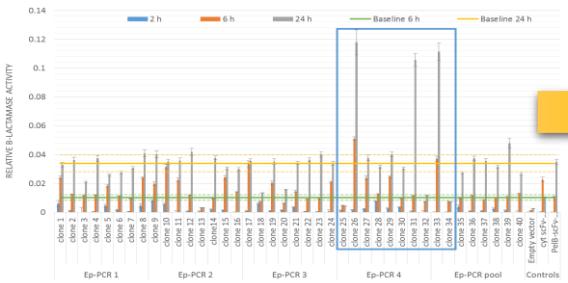
3. Transformation, generation of plasmid DNA stock and research cell banks (RCBs)



4. Transformation into E. coli BL21-A and selection of clones using ampicillin



5. Clone screening and selection by β-lactamase assay



6. Shake-flask growth evaluation

7. Fed-batch fermentation evaluation

Ambr250 modular: optimisation and development of fermentation processes

- Evaluation of the production of the SP-POI-β-lactamase
- Evaluation of the production of the SP-POI

After the initial selection of ampicillin-resistant clones, the β-lactamase screening allowed the selection of clones containing mutant signal peptides with a higher β-lactamase activity than the PelB wild type signal peptide, and theoretically, the selection of clones with a higher production of the fusion protein in the periplasmic space. The detection of β-lactamase activity was used not just as measurement of translocation of the protein to the periplasmic space, but as an indication of correct folding and disulphide-bond formation, based on the assumption that if the scFv163R4-A was correctly folded, β-lactamase should be correctly folded as well. Overall, the β-lactamase screening assay was developed as a screening tool, helping to find mutations, which favoured a higher translocation of the protein to the periplasm, without compromising cell physiology and viability. More important, this approach allows the rapid screening of a large number of signal peptide mutants for the production of therapeutic proteins containing disulphide bonds, facilitating the development of high productivity fermentation processes.

5.2 Results

5.2.1 Generation of signal peptide libraries

The most important step for the successful generation of a large and diverse DNA library is the selection of the mutagenesis strategy, which will depend on the type of DNA library to be generated and the desired degree of complexity. In addition, the efficiency of the subsequent cloning steps also limits the size and complexity of the library to be generated. At present, there are several methods that allow the introduction of random mutations in a DNA sequence, such as the use of XL1-red cells or degenerate oligonucleotides (Chusacultanachai and Yuthavong, 2004). However, error-prone PCR seems to be the optimal choice for the introduction of completely random mutations in a given sequence, allowing the generation of a highly diverse signal peptide library. A second mutant signal peptide library was also generated and produced by chemical synthesis.

Signal peptide libraries were introduced into the expression vector by seamless cloning using type IIS restriction endonuclease, which allows the cloning of two or more DNA fragments joined precisely, without the addition of unwanted nucleotides between the two DNA fragments. Type IIS restriction enzymes are a class of restriction endonucleases that allow the cleavage of DNA sequences outside of their non-palindromic asymmetric recognition sites, e.g. SapI, BsaI or FokI. Both the plasmid vector and the insert need to contain the restriction recognition sequences, which after restriction digestion would be removed, generating compatible cohesive ends allowing the seamless cloning of the DNA fragments (Lu, 2005).

Both different signal peptide libraries, generated by error-prone PCR and by chemical synthesis, used PelB signal peptide sequence as a template. This signal peptide was used as a template since results previously obtained showed that PelB allowed the production and translocation of the scFv163R4-A::β-lactamase to the periplasmic space, maintaining a good growth profile and cell viability (**Chapter 4, Figures 4.15 to 4.18**). In addition, the production of the mature form of the scFv163R4-A and the scFv163R4-A::β-lactamase in the periplasm had been confirmed by N-terminal sequencing (**Chapter 4, Sections 4.2.6 and 4.2.9**).

Error-prone PCR library

Error-prone PCR (Ep-PCR) is one of the most common methods used to introduce random mutations into a given DNA sequence, being based in the misincorporation of nucleotides by an error-prone DNA polymerase during the synthesis of DNA product (Hanson-Manful and Patrick, 2013). The GeneMorph II Random Mutagenesis kit, which contains a blend of two different DNA polymerases, Mutazyme I polymerase and a mutant Taq DNA polymerase, was used for the generation of the error-prone PCR signal peptide library. This enzyme blend allowed an equivalent rate of transitions versus transversions, which means a similar mutation rate of A's and T's versus G's and C's was achieved, obtaining a minimal mutation bias (Chusacultanachai and

Yuthavong, 2004). The error prone PCR was carried out by the amplification of the PelB signal peptide sequence from the plasmid vector pMA-RQ-SapIPelBSapI using the Mut F and Mut R primers (**Figure 5.2**). This plasmid vector contains the PelB signal peptide sequence flanked by the restriction recognition sequence of BspQI restriction enzyme, an isoschizomer of SapI. BspQI restriction endonuclease was used instead of SapI, due to its improved storage stability and digestion efficiency. The Mut F and Mut R primers aligned to the flanking DNA sequences of the PelB DNA template, containing the BspQI recognition sites, the first codon (ATG initiation codon) and the last codon (GCC codon) of the PelB signal peptide sequence. This enabled the misincorporation of nucleotides only in the core sequence of the signal peptide, avoiding mutations in the flanking DNA sequences, as required for the seamless cloning of the signal peptide into the expression vector.

In order to achieve a high mutation rate, between 9 to 16 mutations · kb^{-1}, a low amount of DNA, 0.1 ng, was used as a template. This mutation rate allowed the generation of 1 or 2 mutations in the DNA template per error-prone PCR. However, this mutation rate was considered not sufficient for the generation of a highly diverse signal peptide library. With the aim to achieve a higher number of mutations, sequential error-prone PCRs were carried out by using the product of the previous Ep-PCR as a template, increasing the number of mutations in the final DNA product. Four sequential Ep-PCR were carried out (Ep-PCR 1, Ep-PCR 2, Ep-PCR 3 and Ep-PCR 4), aiming to achieve between 1 and 8 mutations per copy of the signal peptide sequence. A pool containing equivalent amounts of each Ep-PCR product was also generated (Ep-PCR 5 pool), containing a mixture of signal peptide sequences with high and low number of mutations.

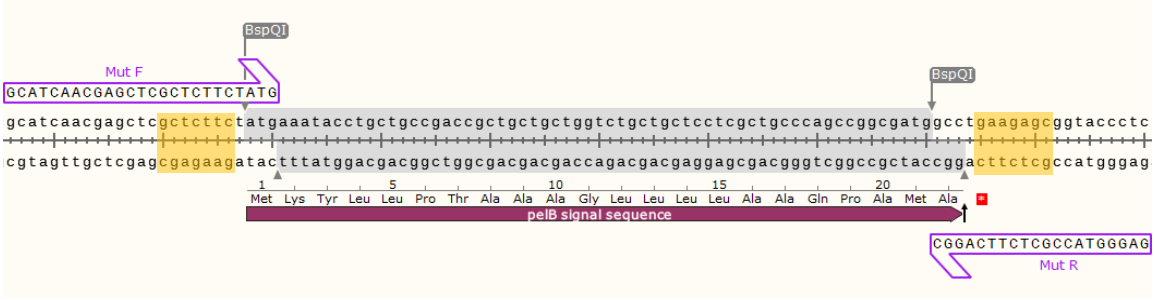


Figure 5.2. DNA template used for the generation of signal peptide libraries. The PelB signal peptide sequence was used as a template for the error-prone PCR. The sequence containing the BspQI restriction recognition sequence (yellow) is located flanking the PelB signal peptide sequence (grey), used for seamless cloning of the signal peptide mutants into the expression vector. The error-prone PCR was carried out by using the Mut forward and reverse primers to amplify the DNA template. The GeneMorph II Random Mutagenesis kit was used to amplify and introduce random mutations in the PelB signal peptide sequence. Sequential error-prone PCRs were carried out using the product from the previous PCR as template, with the aim of increase the number of mutations in the signal peptide sequence.

Twist Bioscience library

A second signal peptide library was also chemically synthesised by Twist Biosciences. Twist Biosciences owns an innovative silicon-based 9,600 well platform, enabling the parallel production of diverse oligonucleotide pools with high accuracy. As part of a Beta program, Twist Biosciences generated an oligonucleotide library, which consisted on the generation of 10,000 oligonucleotides containing between 2 to 4 mutations per oligonucleotide. The PelB signal peptide sequence was flanked by the BspQI restriction sites, the same DNA template used the generation of the error-prone PCR libraries (**Figure 5.2**). Once generated, the oligonucleotide pool was made double-stranded and amplified to obtain a final concentration of 10 pM.

5.2.2 Construction of the expression vector

Classic gene cloning techniques often involves the use of restriction digestion and DNA ligation, the commonly called cut-and-paste reaction, which was considered the gold standard for cloning and DNA modification during the last decades. Nevertheless, this type of cloning requires the addition of operational sequences containing the restriction recognition sequence, which introduces undesired sequences coding for extra amino acid residues at the DNA junction. This addition of extra amino acid residues may affect the activity of recombinant proteins and the generation of the recombinant protein containing non-native amino acid sequences (Lu, 2005).

In contrast, seamless cloning using type IIS restriction endonucleases was used for the generation of signal peptide-protein fusions, since it is essential to ensure that no extra sequences are added between the signal peptide and the protein of interest. The type IIS restriction enzyme BspQI, an isoschizomer of SapI, used for the cloning of the signal peptides into the expression vector, recognises a 7 bp non-palindromic recognition sequence, leaving a 3 bp 5' overhang after restriction digestion (**Figure 5.2**). BspQI recognition sequences were introduced into the expression vector (pLBAD2-SapIXbaISapI-scFv163R4-A::β-lactamase) upstream the sequence

coding for the scFv163R4-A::β-lactamase (**Figure 5.3**). Both signal peptides libraries, the error-prone and the twist libraries, were generated containing the signal peptide sequence flanked by the BspQI recognition sequences. This approach allowed the cloning of the signal peptide libraries into the expression vector seamlessly, in one single digestion and ligation reaction, allowing a high cloning efficiency. This reaction was carried out for the introduction of the Ep-PCR 1, Ep-PCR 2, Ep-PCR 3, Ep-PCR 4 and Ep-PCR 5 (pool) products. The same approach was used for the introduction of the Twist library into the expression vector.

5.2.3 Transformation, generation of research cell banks and plasmid DNA stock

Once the ligation reactions were completed, the ligated product was transformed into the commercially available ElectroSHOX competent cells by electroporation. Immediately after, cells were resuspendended in LB and incubated for at least 1 hour before being transferred to a 1 L shake-flask containing 200 mL of LB broth. After 12-18 hours, cells were harvested, lysed, and the plasmid DNA extracted, generating a plasmid DNA stock containing the signal peptide libraries cloned into the expression vector. ElectroSHOX competent cells were used due to their high transformation efficiency of ligated products and the generation of high quality plasmid DNA stocks of the expression vector containing the signal peptide libraries.

The expression vector library was transformed into *E. coli* BL21-A, used as expression strain. As described previously, after 1-hour incubation, the cell suspension was transferred into a 250 mL shake-flask and further incubated for 12-18 hours. At this stage, 0.5 mL aliquots of each culture were serially diluted for the selection of clones able to grow on ampicillin-supplemented agar plates. Only those clones able to produce and translocate the scFv163R4-A::β-lactamase to the periplasmic space were able to grow in the presence of ampicillin. The culture was also used for the generation of small research cell banks (RCB), for future evaluations of signal peptide libraries.

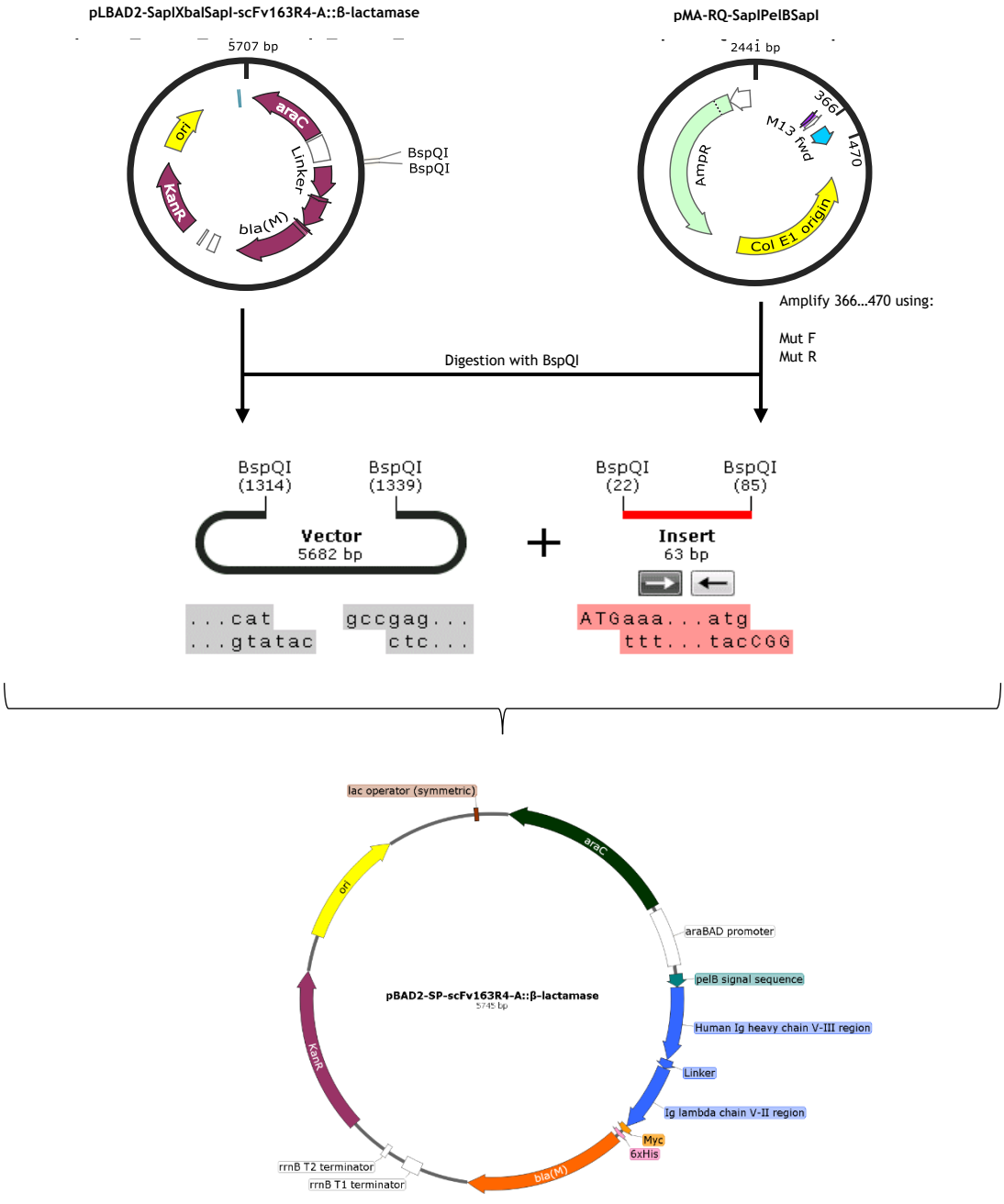


Figure 5.3. Generation of the Ep-PCR signal peptide library and introduction into the expression vector. The error-prone PCR library was generated by amplifying the PelB signal peptide from the pMA-RQ-SapIPelBSapI vector by ep-PCR using the Mut F and Mut R primers (top right). The DNA insert and the expression vector (pLBAD2-SapIXbaISapIscFv163R4-A::β-lactamase) were ligated by the compatible cohesive ends generated by BspQI restriction digestion (middle). The expression vector contained the signal peptide library, a mutant variant of the PelB signal peptide sequence, fused to the sequence coding for the antibody fragment, scFv163R4-A, fused to the reporter protein, β-lactamase (bottom).

Electroporated cell suspensions were transferred directly to shake-flasks, instead of being selected by plating on selective agar, to ensure that the library size was not being diminished with each transformation step. Kanamycin was added to all incubation steps to ensure plasmid maintenance, avoiding the overgrowth of plasmid-free cells. Plasmid retention was evaluated after each step of the cloning process, by plating a small aliquot of the cell suspension onto selective and non-selective LB agar and comparing the number of colonies. Overgrowth of plasmid-free cells was not observed at any step of the process and the addition of kanamycin seemed to be sufficient to impede the overgrowth of *E. coli* cells, which had not incorporated the expression vector.

5.2.4 Selection of signal peptide mutants based on ampicillin resistance

A small inoculum obtained from the transformed culture was used to inoculate a starter culture, grown at 37°C until an OD_{600} between 1 and 2. Subsequently, cultures were serially diluted to obtain a concentration of 10^4 CFU · mL^{-1} and plated onto M-H agar containing 0.2% arabinose, 50 μg · mL^{-1} of kanamycin, and ampicillin to a final concentration ranging from 3 to 1600 μg · mL^{-1}. After 12 - 18 hours incubation at 37 °C, colonies grown on ampicillin-supplemented plates were evaluated by β-lactamase enzymatic assay. As it has been previously observed in cultures carrying the DsbA and Bla signal peptides, non-healthy cells were unable to grow under "strong induction conditions" (**Chapter 4, Section 4.2.8**), and therefore, a high concentration of inducer (0.2% arabinose) and high cultivation temperature (37 °C) were used to minimise the growth of clones with low cell viability or severely affected by protein production.

Generally, visible growth was detected up to a concentration of ampicillin 200 μg · mL^{-1}, and no colonies grew at higher concentrations of ampicillin. Colonies grown under induced conditions, independently of the addition of ampicillin, showed large differences in terms of morphology when compared to colonies grown on the control plates, supplemented only with kanamycin

(**Figure 5.4**). Different colony morphologies were identified on plates supplemented with arabinose: larger colonies with either a flat or a raised morphology or small colonies. In addition, there seems to be a correlation between the colony morphology and the Ep-PCR product contained in the expression vector, as the number of large-size colonies increased with Ep-PCR products containing a higher number of mutations. The different colony morphologies were not detected on control plates, supplemented only with kanamycin, showing large-size colonies. A more homogenous colony morphology was observed on plates seeded with the control culture carrying the PelB wild type signal peptide and no different colony morphologies were observed within the same plate.

In general, the appearance of different colony morphologies seemed to be the result of the high induction, and not by the selective pressure generated by the addition of ampicillin. As had been observed before, high levels of production of the scFv163R4-A::β-lactamase could produce a detrimental effect on cell culturability observed in cultures where the translocation of the scFv163R4-A::β-lactamase was directed by the DsbA and Bla signal peptides (**Chapter 4, Section 4.2.8**). The same effect was observed during the selection of clones carrying the Twist library.

5.2.5 Selection of clones obtained from EP-PCR library by β-lactamase assay

Clones selected by their ability to grow on plates supplemented with ampicillin were evaluated for β-lactamase activity with the aim to screen for clones containing signal peptides which allowed a higher production of scFv163R4-A::β-lactamase compared with the PelB signal peptide. Colonies obtained from ampicillin-supplemented agar plates were grown in duplicate in 96-well plates. Cultures were grown in terrific broth at 25 °C and induced with 0.02% arabinose, 2 hours post-inoculation. The relative activity of β-lactamase was determined using nitrocefin as a

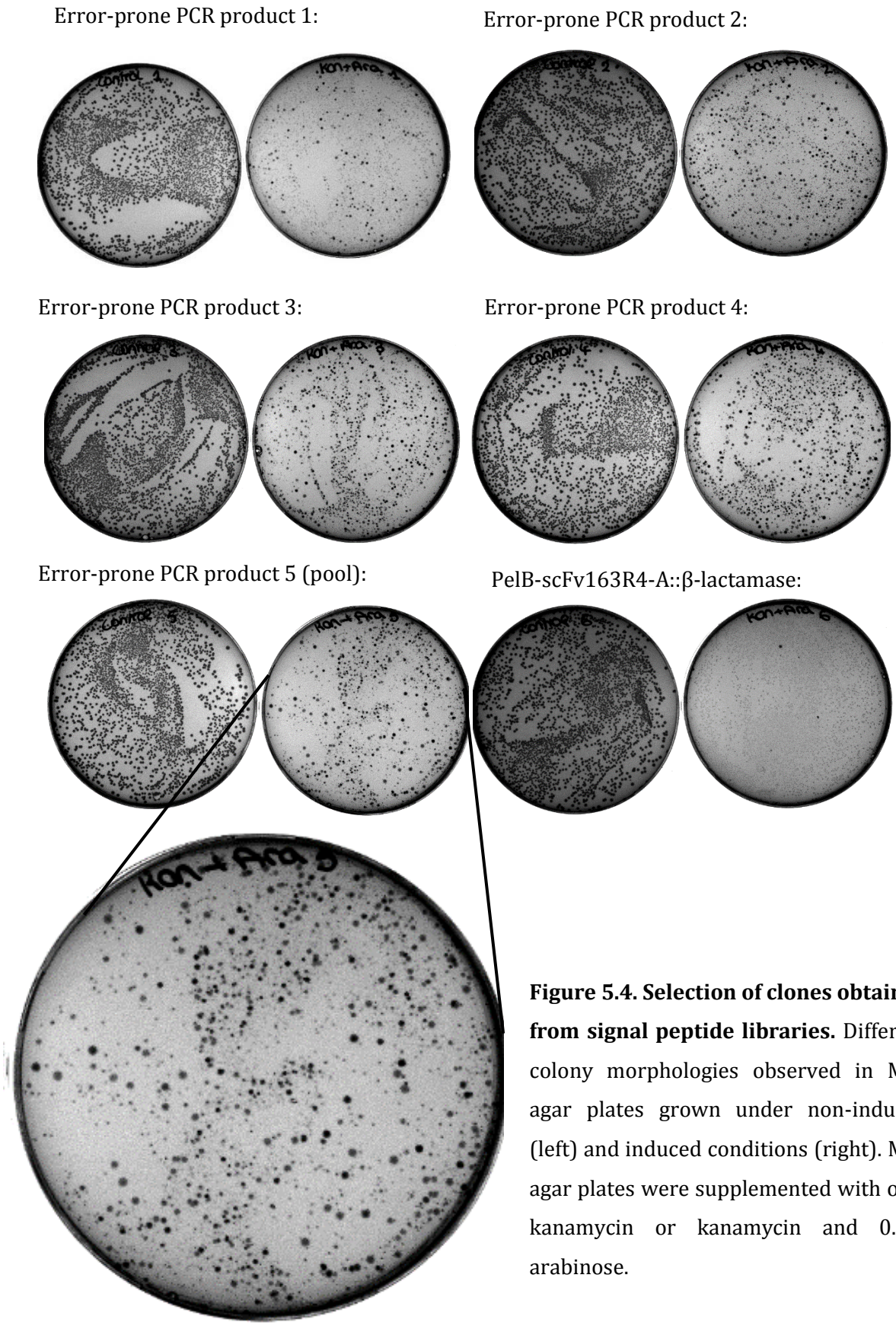


Figure 5.4. Selection of clones obtained from signal peptide libraries. Different colony morphologies observed in M-H agar plates grown under non-induced (left) and induced conditions (right). M-H agar plates were supplemented with only kanamycin or kanamycin and 0.2% arabinose.

substrate by continuous measurement of the OD_{495} before induction (2 h) and after induction (8 and 24 hours post-inoculation). The initial velocity (slope) was calculated by linear regression, as a quantification of the relative β-lactamase activity.

As part of the β-lactamase screening assay, 200 clones were evaluated for β-lactamase activity, divided between five 96-well plates (Libraries A, B, C, D and E) containing 40 clones each. For each library, 8 clones were selected corresponding to each Ep-PCR product (Ep-PCR 1, Ep-PCR 2, Ep-PCR 3, Ep-PCR 4 and Ep-PCR 5). Cultures carrying the empty vector (pLBAD2), the cytoplasmic scFv163R4-A::β-lactamase (M) and the wild type PelB signal peptide fused to the scFv163R4-A::β-lactamase were used as control cultures. The nitrocefin enzymatic assay was carried out on live cells.

As expected, the majority of the clones showed similar levels of β-lactamase activity to PelB signal peptide (**Figure 5.5**). A few clones, such as clones 25 and 34 obtained from Library C, showed reduced or no β-lactamase activity. Some clones showed a higher β-lactamase activity than PelB wild type, such as clones 26, 31 and 33 (Library C) and clones 27 and 29 (Library D), with a 3-fold increase in β-lactamase activity after 24 hours of growth. Some clones, such as 26 and 33 (Library C) showed a 4-fold increase in β-lactamase activity in samples obtained 8 hours post-inoculation. Unexpectedly, some clones showed higher β-lactamase activity at 8 hours than at 24 hours of growth, such as clones 7 and 19 (Library A) and clone 18 (Library B).

In general terms, there seems to be a correlation between the expected number of mutations in the signal peptide sequence and the level of β-lactamase activity observed between different clones. As the Ep-PCR 4 is the product of the fourth sequential error-prone PCR, contained a higher number of mutations, and overall, a higher number of clones with high β-lactamase activity were detected when compared with clones selected from Ep-PCR products 1 or 2 (**Figure 5.5**).

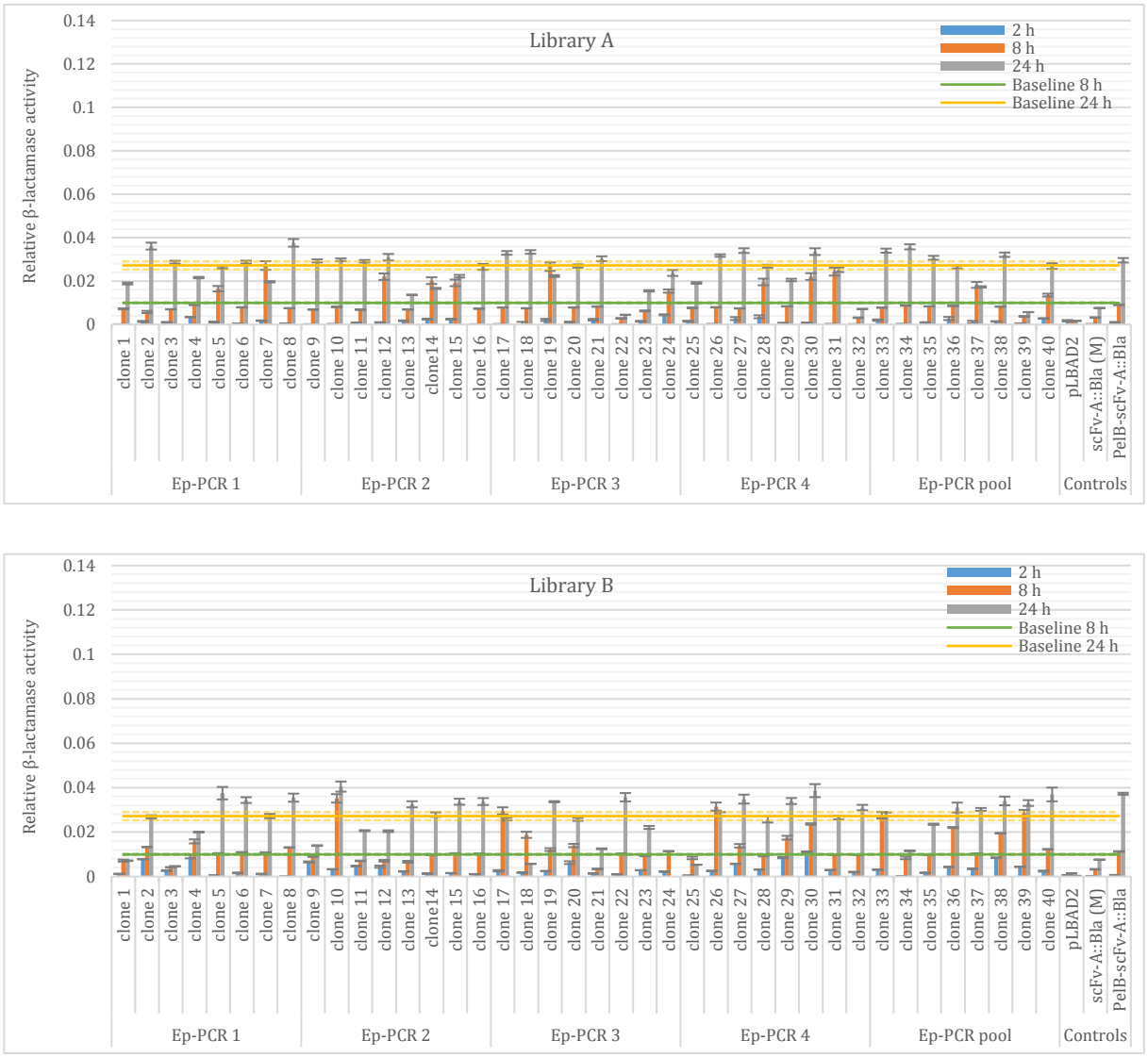


Figure 5.5. β-lactamase enzymatic assay for the screening of the EP-PCR library. *E. coli* BL21-A clones containing the vector coding for the error-prone PCR library signal peptides fused to the scFv163R4-A::β-lactamase were grown at 25 °C and induced with 0.02% arabinose, 2 hours post-inoculation. Cultures were evaluated for β-lactamase activity by quantification of the hydrolysis of the substrate nitrocefin by continuous measurement of the OD_{495} before (2 h) and after induction (8 h and 24 h post-inoculation). All the data was normalised by OD_{600}. The average relative β-lactamase activity after 8 hours (green line) and 24 hours (yellow line) of growth for the wild type PelB signal peptide fused to the scFv163R4-A::β-lactamase is represented. The dotted lines represent ±2 standard deviations. The initial velocity (slope) was calculated by linear regression, adapted to obtain an $R \geq 0.9$, and compared between the different samples. Error bars represent the 5% uncertainty in the calculation of the slopes.

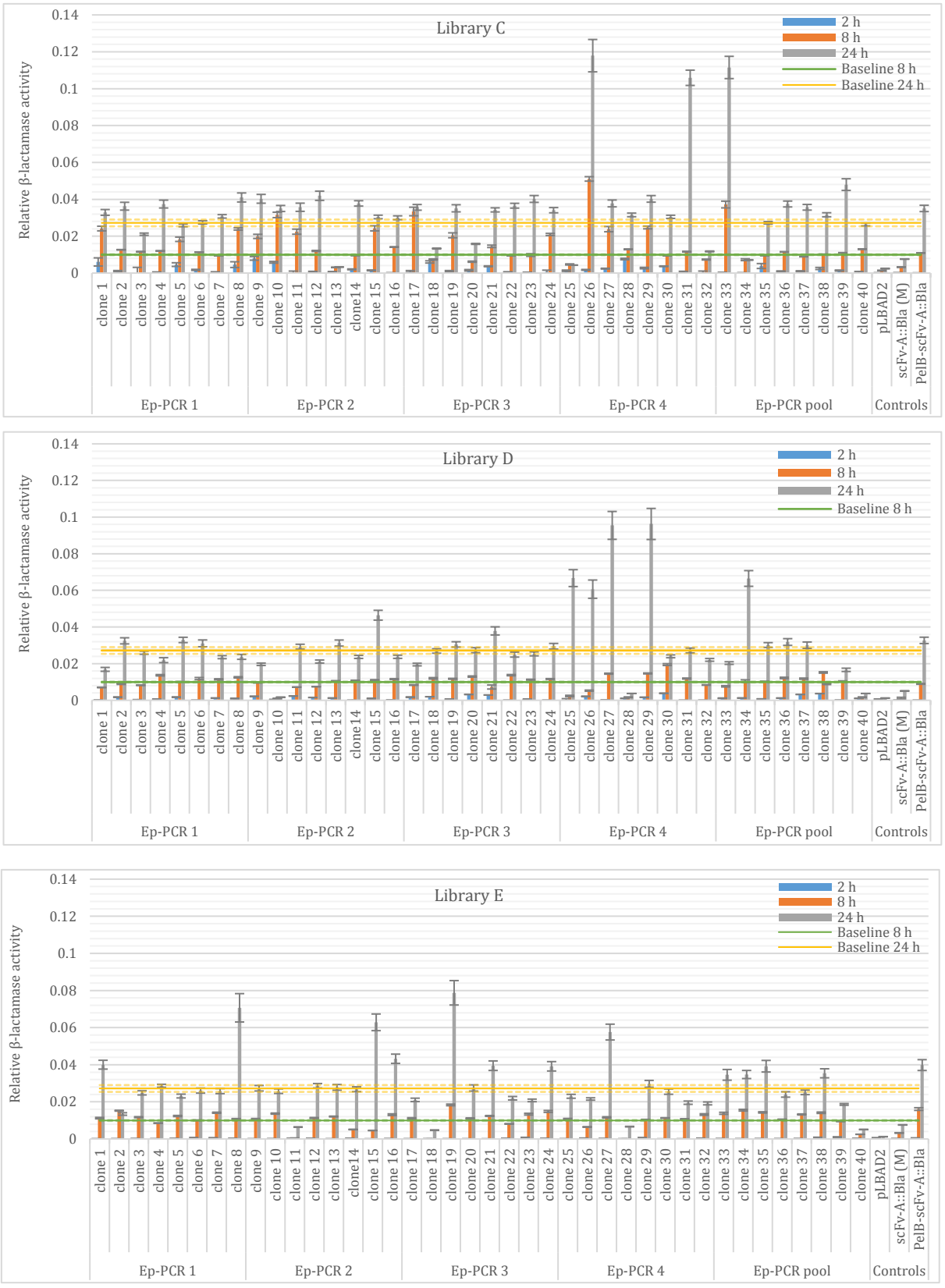


Figure 5.5. (Continuation)

5.2.6 Selection of clones obtained from Twist library by β-lactamase assay

With the aim to compare both signal peptide libraries, all the steps described in the previous sections were repeated using the signal peptide library generated by Twist Biosciences. Colonies selected from M-H agar plates were grown in duplicate in 96-well plates containing terrific broth at 25 °C, and induced with 0.02% arabinose, 2 hours post-inoculation. The relative activity of β-lactamase was determined using nitrocefin as a substrate by continuous measurement of the OD_{495} before induction (2h) and after induction (8 and 24 hours post-inoculation).

As with Ep-PCR libraries, 200 clones were evaluated for β-lactamase activity, divided between 5 different 96-well plates (Libraries A, B, C, D and E) containing 40 clones each. The empty vector (pLBAD2), the cytoplasmic scFv163R4-A::β-lactamase and the PelB signal peptide fused to the scFv163R4-A::β-lactamase, were used as controls.

In general terms, a higher number of clones with high β-lactamase activity were detected from clones obtained from the Twist library (**Figure 5.6**), when compared to the Ep-PCR library (**Figure 5.5**). This result was not unexpected, since the Twist library was designed to have a similar number of mutations as the Ep-PCR 4, where the highest number of high productive clones were detected. Also some clones from the Twist library had higher activity than most active clones from the Ep-PCR library (**Figure 5.6**). A high number of clones showed high levels of β-lactamase activity, such as clone 3 and 11 (Library A), clone 4 (Library D) and clones 18, 33, 35 and 39 (Library E), showing approximately a 5-fold increase over the wild type PelB, in samples obtained after 24 hours of growth. Also, a 2.5-fold increase in β-lactamase activity was detected in clones 15 and 19 (Library A) and clones 13, 24, 27 and 35 (Library D), 24 hours post-inoculation.

Overall, the β-lactamase activity assay allowed the rapid screening of a large number of clones, which by other methods could have been an extremely time-consuming process. A few clones with barely any β-lactamase activity were observed in the screening, showing that a minimal

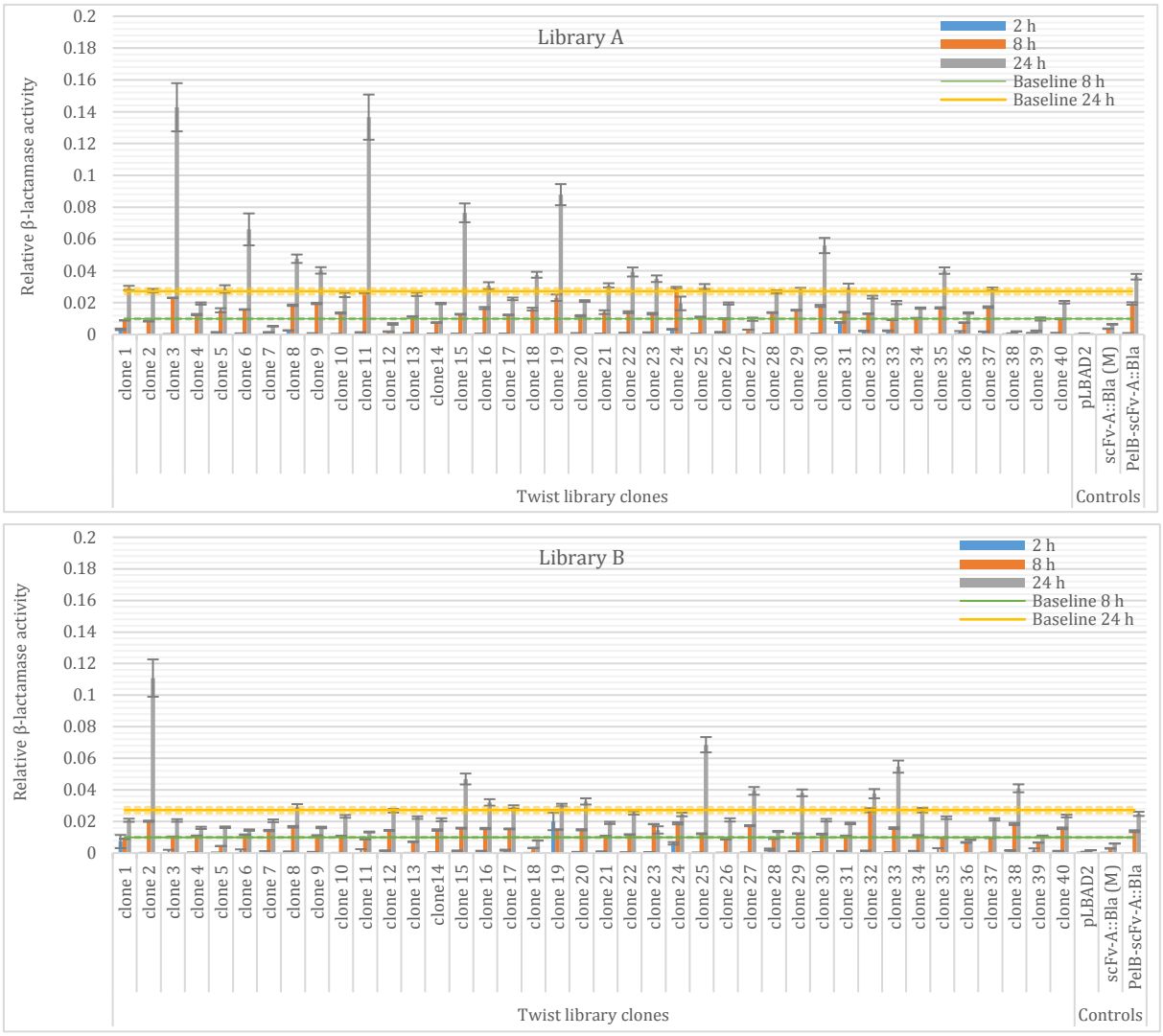


Figure 5.6. β-lactamase enzymatic assay for the screening of the Twist library. *E. coli* BL21-A clones containing the vector coding for the Twist library signal peptides fused to the scFv163R4-A::β-lactamase were grown at 25 °C and induced with 0.02% arabinose, 2 hours post-inoculation. Cultures were evaluated for β-lactamase activity by quantification of the hydrolysis of the substrate nitrocefin by continuous measurement of the OD_{495} before (2 h) and after induction (8 h and 24 h post-inoculation). All the data was normalised by OD_{600}. The average relative β-lactamase activity after 8 hours (green line) and 24 hours (yellow line) of growth for the wild type PelB signal peptide fused to the scFv163R4-A::β-lactamase is represented. The dotted lines represent ±2 standard deviations. The initial velocity (slope) was calculated by linear regression, adapted to obtain an $R \geq 0.9$, and compared between the different samples. Error bars represent the 5% uncertainty in the calculation of the slopes.

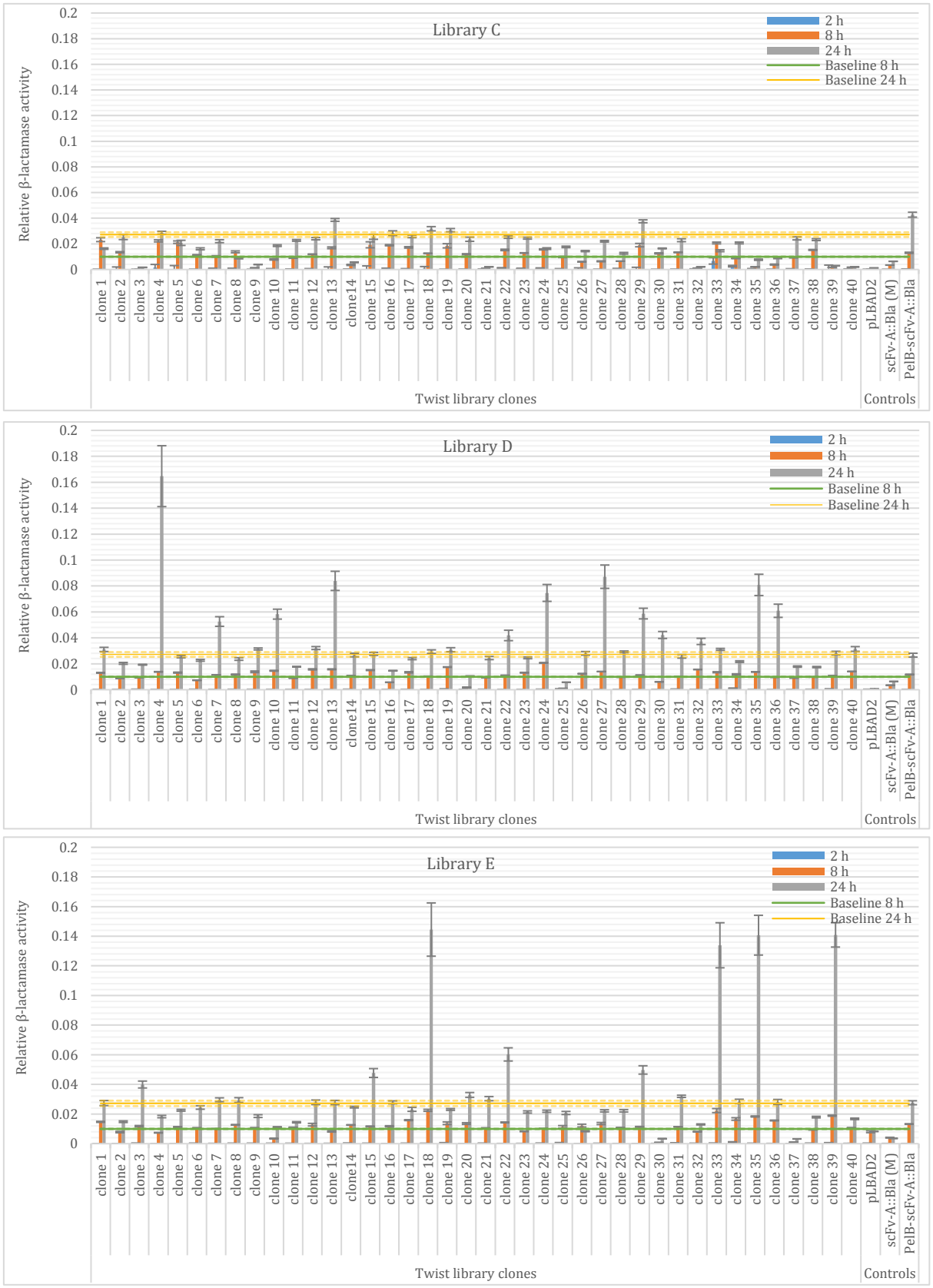


Figure 5.6. (Continuation)

concentration of β-lactamase is sufficient to confer resistance to β-lactam antibiotics and grow in ampicillin-supplemented agar plates. In order to evaluate whether the high β-lactamase activity exhibited by some of the clones obtained from the Ep-PCR and Twist libraries correlated with the production of the scFv163R4-A::β-lactamase; the production of scFv163R4-A::β-lactamase was further assessed by shake-flask growth experiments.

5.2.7 Evaluation of clone-to-clone variability

In order to evaluate whether the different levels of β-lactamase activity observed between clones were the result of the mutation diversity of the signal peptide library or the clone-to-clone variation, the β-lactamase enzymatic assay was also carried out using the PelB wild type signal peptide. *E. coli* BL21-A was transformed with the vector coding for the PelB-scFv163R4-A::β-lactamase, transferred to a shake-flask and grown overnight. Cultures were plated on M-H agar plates supplemented with ampicillin and 40 colonies were selected to be evaluated by β-lactamase enzymatic assay.

As expected, all the different clones containing the PelB wild type signal peptide showed a similar relative β-lactamase activity (**Figure 5.7**). No substantial differences in the level of β-lactamase activity were observed between different clones after 8 hours of growth; however, a higher variability in β-lactamase activity was observed after 24 hours of growth.

Overall, differences in β-lactamase activity between different clones carrying the PelB wild type signal peptide were not comparable to differences observed during the screening of the Ep-PCR and Twist signal peptides libraries. Hence, confirming that the variability in β-lactamase activity observed different clones was the result of the introduction of mutations in the signal peptide sequences and not physiological differences between different clones.

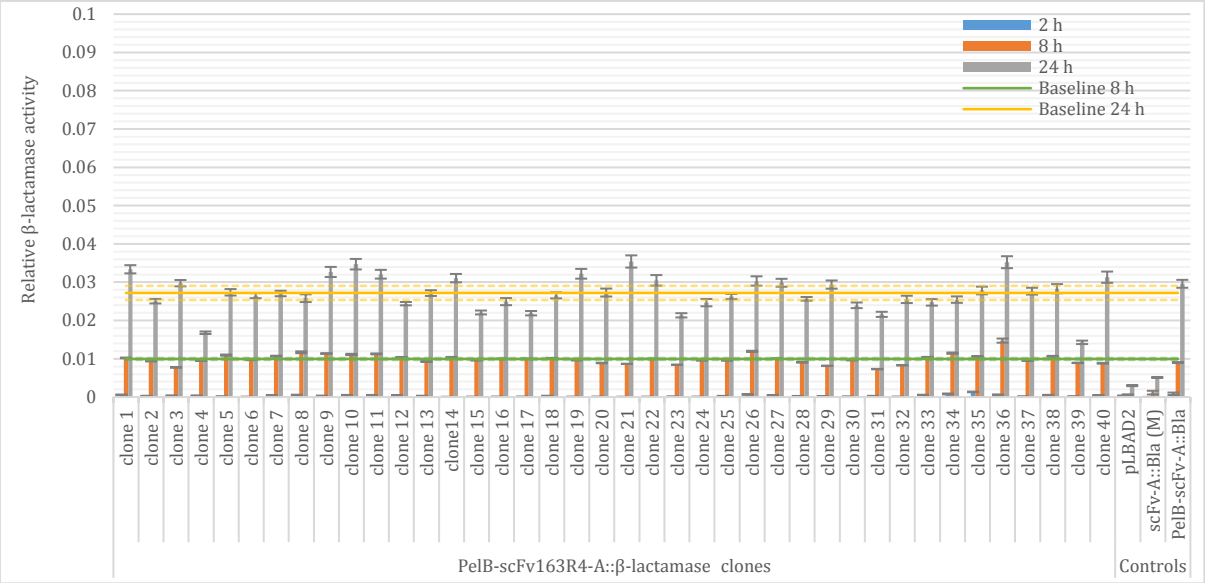


Figure 5.7. Evaluation of clone-to-clone variability using the wild type PelB signal peptide. *E. coli* BL21-A clones containing the vector coding for the PelB signal peptides coding for scFv163R4-A::β-lactamase were grown at 25 °C and induced with 0.02% arabinose, 2 hours post-inoculation. Cultures were evaluated for β-lactamase activity by quantification of the hydrolysis of the substrate nitrocefin by continuous measurement of the OD_{495} before (2 h) and after induction (8 h and 24 h post-inoculation). All the data was normalised by OD_{600}. The average relative β-lactamase activity after 8 hours (green line) and 24 hours (yellow line) of growth for the wild type PelB signal peptide fused to the scFv163R4-A::β-lactamase is represented. The dotted lines represent ±2 standard deviations. The initial velocity (slope) was calculated by linear regression, adapted to obtain an $R \geq 0.9$, and compared between the different samples. Error bars represent the 5% uncertainty in the calculation of the slopes.

5.2.8 Evaluation of clones selected from the Ep-PCR and the Twist libraries for the production of scFv163R4-A::β-lactamase

For further clarity, from this point forward clones selected from the Ep-PCR libraries will be referred as EP and from the Twist library as T, followed by a letter corresponding to the library were they were obtained from and the clone number. For example, a clone selected from the EP-PCR library C, clone 25, will be referred as clone EP-C25.

With the aim to evaluate whether the levels of β-lactamase activity detected by enzymatic assay were representative of the production of the scFv163R4-A::β-lactamase, a total of 10 clones were selected from the error-prone PCR library C, clones 25 to 34, and grown in shake-flasks. Clones evaluated included high productive clones, such as clones EP-C26, EP-C31 and EP-C33; clones with a similar β-lactamase activity to PelB wild type, as clones EP-C27, EP-C28, EP-C29 and EP-C30; and clones showing no or low β-lactamase activity, as clones EP-C25, EP-C32 and EP-C34 (**Figure 5.5**).

Cultures were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. Cultures where the translocation of the scFv163R4::β-lactamase was directed by the PelB wild type signal peptide were used as a control. Cell culturability was evaluated by serially diluted culture samples being plated onto non-selective LB agar.

A high variability between clones was observed in terms of growth, reaching a final OD_{600} between 15 and 25, after 24 hours of growth (**Figure 5.8 A**). All the clones reached a higher OD_{600} than the control culture (PelB WT), with the exception of clone EP-C31, which reached a final OD_{600} of 15 after 24 hours of growth. This clone, EP-C31, also showed a drastic decrease in the number of CFU between 9 and 24 hours of growth, reaching $1.5 \cdot 10^8$ CFU $\cdot$ mL^{-1}, in contrast with the control culture which reached $2.9 \cdot 10^9$ CFU $\cdot$ mL^{-1}, after 24 hours of growth (**Figure 5.8 B**).

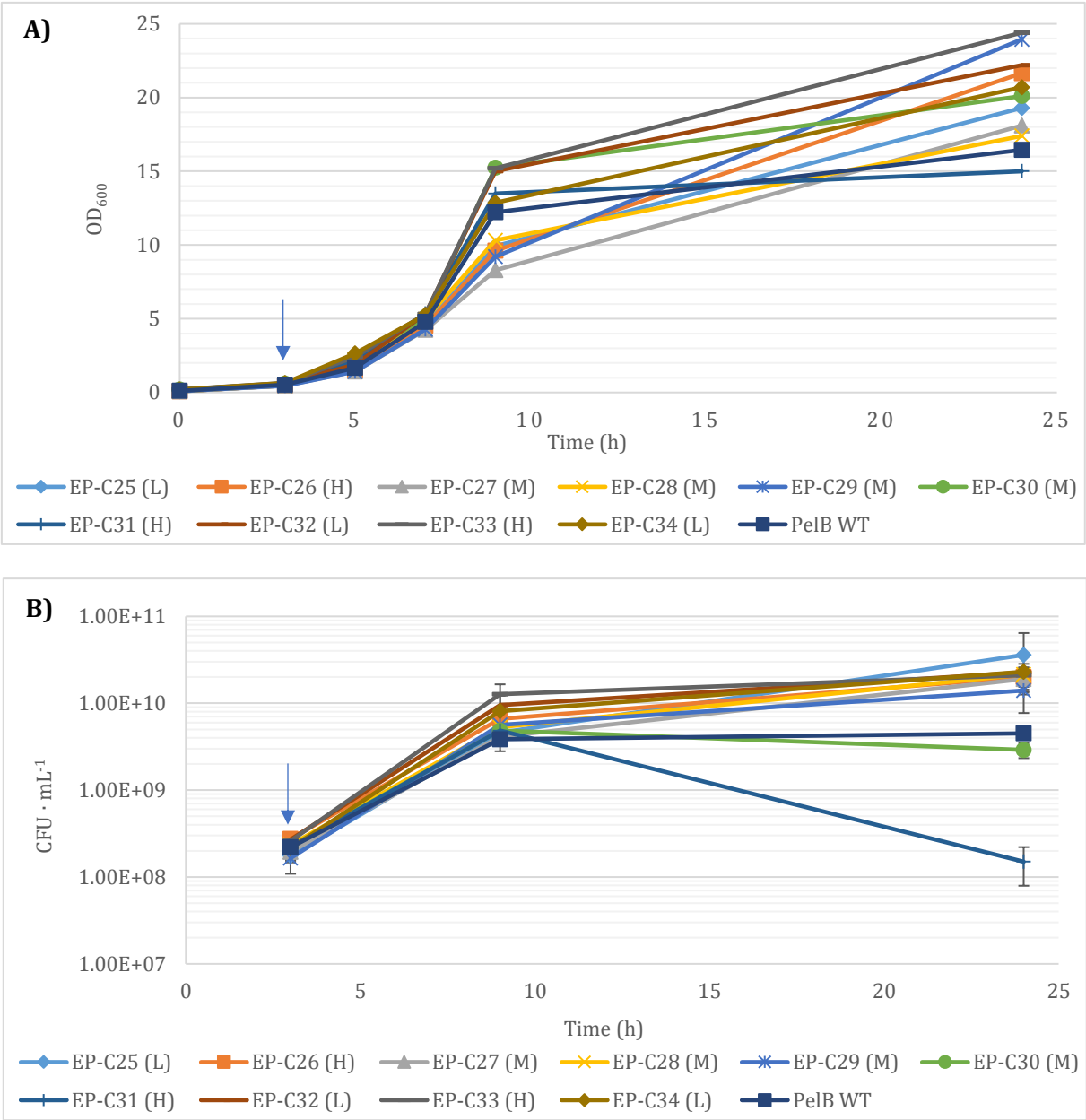


Figure 5.8. Growth and cell viability of clones selected from EP-PCR. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the error-prone PCR library C were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} ≈ 0.5 (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 3 h, 9 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Clones are labelled as high (H), medium (M) and low (L) producers depending on the level of β-lactamase activity detected in the enzymatic assay as a guidance. Data shown are mean values of two replica flasks and error bars are ±1 standard deviation.

In terms of the production of the scFv163R4-A::β-lactamase, results obtained from the shake-flask studies correlated with results obtained from the β-lactamase enzymatic assay (**Figure 5.9**). Clones which exhibited high levels of β-lactamase activity, such as clones EP-C26, EP-C31 and EP-C33, showed a high production of the scFv163R4-A::β-lactamase accounting for ≥ 20% of the total cell protein (TCP), after 24 hours of growth (**Figure 5.9 A**). Clone EP-C29 also showed a high production of scFv163R4-A::β-lactamase, accounting for 24% of the total cell protein. This result was unexpected, since this clone had not been identified as a high producer during the enzymatic assay (**Figure 5.9 C**). Clones EP-C27 and EP-C30 showed a similar level of protein production as PelB, accounting for 15% and 12% of TCP, respectively, 24 hours post-inoculation (**Figure 5.9 A**). Low accumulation of scFv163R4-A::β-lactamase was observed in samples obtained from clones EP-C25, EP-C32 and EP-C34, accounting for ≤ 2% of TCP protein corresponding to the scFv163R4-A::β-lactamase, correlating with the results obtained from the enzymatic assay.

A low concentration of scFv163R4-A::β-lactamase was detected in the periplasmic fraction. The fusion protein was found to be accumulated mainly in the spheroplast protein fraction for all the clones evaluated, independently of the production level of each clone (**Figure 5.9 B**).

In general terms, no correlation was obtained between growth, cell viability and protein production between clones selected from the EP-PCR library C. The level of production of the scFv163R4-A::β-lactamase observed during shake-flask experiments generally correlated with the levels of activity detected by the β-lactamase enzymatic assay.

As growth experiments only included signal peptides selected from the Ep-PCR library C, the shake-flask screening was expanded to include high productive clones selected from the Twist signal peptide libraries. In contrast with the previous shake-flask growth experiment, only high productive clones were selected from the Twist library, and a total of five clones, corresponding to clones T-A11, T-A19, T-B2, T-D4 and T-E18, were selected with the aim to further validate

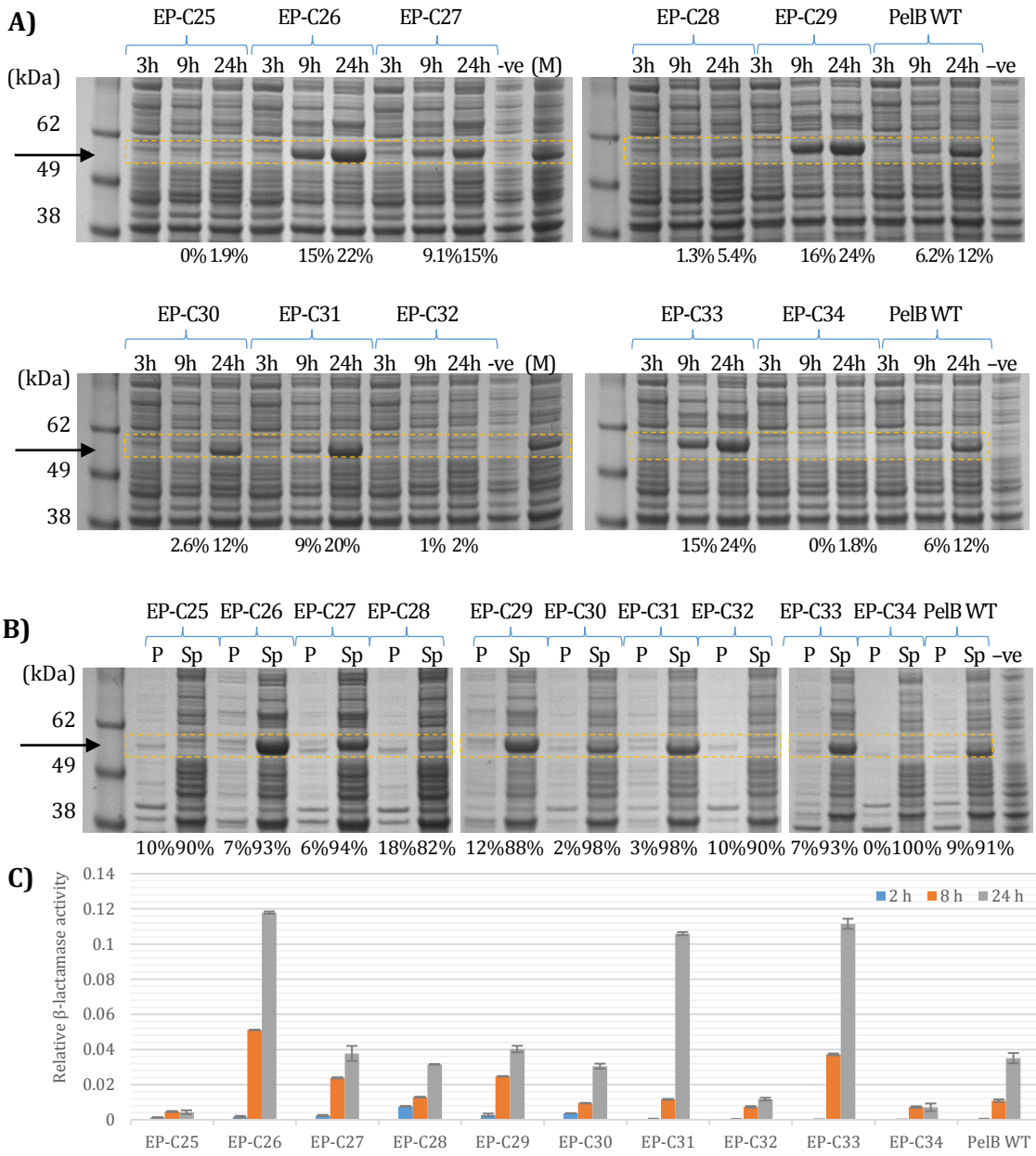


Figure 5.9. Production of scFv163R4-A::β-lactamase of clones selected from the Ep-PCR. SDS-PAGE gels showing the accumulation of scFv163R4-A::β-lactamase (black arrow) from whole cell lysates. Samples were obtained before (3 h) and after induction (9 and 24 hours post-inoculation) (A). The periplasm (P) and spheroplast (Sp) protein fractions were obtained by modified cold osmotic shock from samples after 9 hours of growth (B). Samples obtained from cultures carrying the empty vector (-ve) after 24 hours of growth were used as a control. The quantity of the scFv163R4-A is expressed as percentage of whole cell protein (A) and the ratio of protein accumulated in the periplasm and spheroplast protein fractions (B) are expressed as percentage at the bottom of the gel. The relative β-lactamase activity of each clone evaluated is represented as a guidance (C).

results obtained by the β-lactamase enzymatic assay. As in the previous shake-flask experiment, cultures were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. Cultures carrying the PelB wild type signal peptide were used as a control.

In terms of growth, differences were observed between different clones selected from the Twist library. Clones T-A11, T-A19 and T-B2 showed a similar growth profile as the control culture (PelB WT), reaching a final OD_{600} between 11 and 13, after 24 hours of growth (**Figure 5.10 A**). In contrast, clones T-D4 and T-E18 showed signs of growth arrest, reaching an OD_{600} of 3 and 6, 24 hours post-inoculation. Nevertheless, growth variability was lower than in clones selected from the Ep-PCR, especially between 3 and 9 hours post-inoculation (**Figure 5.8 A**). Loss of cell culturability was observed in cultures corresponding to clones T-D4 and T-E18, showing a decrease in CFU between 9 and 24 of growth (**Figure 5.10 B**). A lower cell culturability was also observed for clones T-A19 and T-B2 when compared to PelB WT, after 24 hours of growth.

In terms of protein production, all the clones evaluated during this experiment showed higher or similar production of the scFv163R4::β-lactamase than the PelB wild type signal peptide, accounting for 11% to 19% of the total cell protein (**Figure 5.11 A**). However, the scFv163R4::β-lactamase protein yields were lower than for high producer clones selected from the Ep-PCR signal peptide library, as EP-C26, EP-C29, EP-C31 and EP-C33 (**Figure 5.9**), even though β-lactamase activities of Twist library clones were higher than Ep-PCR clones.

A high accumulation of the fusion protein in the periplasm was detected in samples obtained from clones T-A11, T-A19 and T-B2, from 35% to 40% of the total scFv163R4::β-lactamase, after 9 hours of growth (**Figure 5.11 B**). Low concentration of scFv163R4::β-lactamase was detected in periplasmic protein samples obtained from clones T-D4 and T-E18, which did not correlate with results obtained from the β-lactamase enzymatic assay. The high levels of β-lactamase activity exhibited by these two clones could be the result of cell lysis, as growth arrest and low cell viability were observed during shake-flask experiments (**Figure 5.10**).

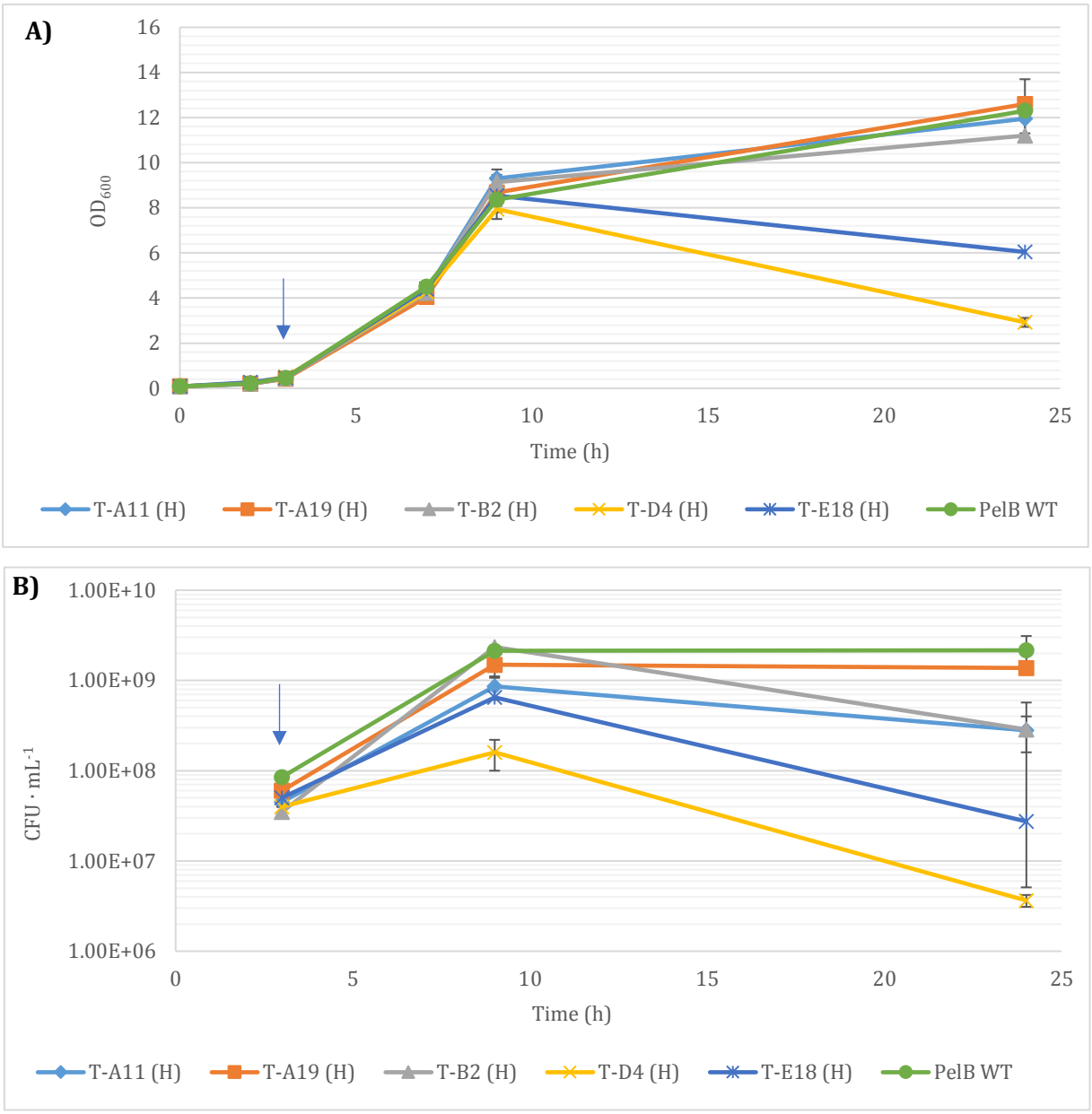


Figure 5.10. Growth and cell viability of clones selected from Twist library. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the Twist libraries were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} ≈ 0.5 (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at 3 h, 9 h and 24 h post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Clones are labelled as high (H), medium (M) and low (L) producers depending on the level of β-lactamase activity detected in the enzymatic assay as a guidance. Data shown are mean values of two replica flasks and error bars are ±1 standard deviation.

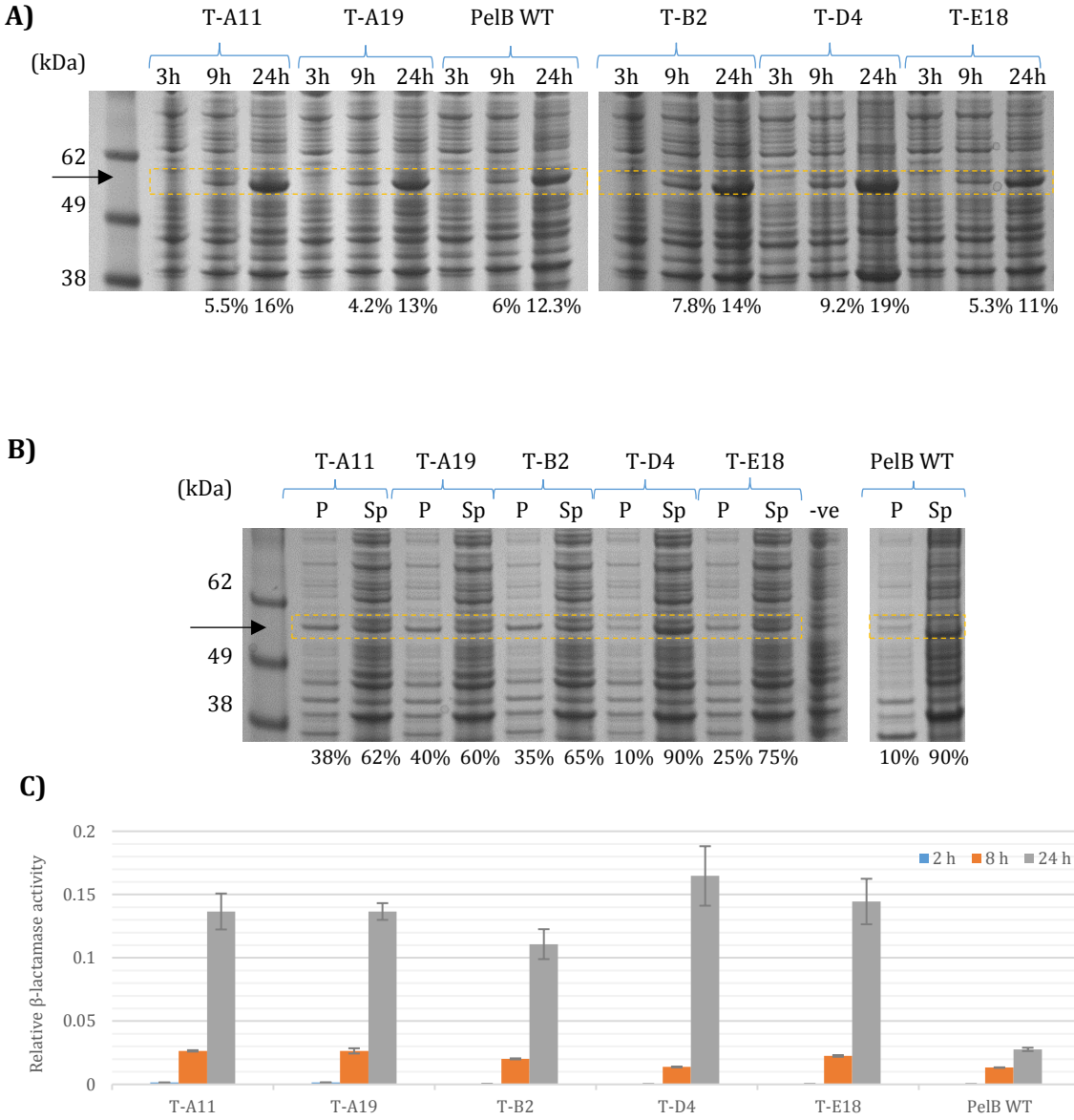


Figure 5.11. Production of scFv163R4-A::β-lactamase of clones selected from the Twist library. SDS-PAGE gels showing the accumulation of scFv163R4-A::β-lactamase (black arrow) from whole cell lysates. Samples were obtained before (3 h) and after induction (9 and 24 hours post-inoculation) (A). The periplasm (P) and spheroplast (Sp) protein fractions were obtained by cold osmotic shock from samples after 9 hours of growth (B). Samples obtained from cultures carrying the empty vector (-ve) after 24 hours of growth were used as a control. The quantity of the scFv163R4-A is expressed as percentage of whole cell protein (A) and the ratio of protein accumulated in the periplasm and spheroplast protein fractions (B) are expressed as percentage at the bottom of the gel. The relative β-lactamase activity quantified for each clone evaluated is represented as a guidance (C).

In general, higher production of the scFv163R4-A::β-lactamase was detected in clones obtained from the Ep-PCR signal peptide library than in the Twist library (**Figure 5.12**). Nevertheless, clones T-A11, T-A19 and T-B2 showed an enhanced production of the scFv163R4-A::β-lactamase in the periplasmic space (**Figure 5.11 B**). In general, clones EP-C26, EP-C33, from the Ep-PCR libraries, and T-A11, T-A19 and T-B2, from the Twist libraries, showed the most promising results in terms of growth, cell viability and production of the scFv163R4-A::β-lactamase. These clones showed a higher production of scFv163R4-A::β-lactamase when compared with the PelB wild type after 24 of growth, without compromising growth and cell viability (**Figure 5.12**). Therefore, further experiments would be required to evaluate if any of these clones showed an enhanced production of periplasmic scFv163R4-A β-lactamase when compared with the PelB signal peptide.

Overall, the correlation between the results obtained from the enzymatic assay and the shake-flask growth experiments validates the use of β-lactamase enzymatic assay as a high-throughput screening method, allowing the screening of a large number of clones and the selection of those with improved β-lactamase activity. However, growth studies are still required to evaluate the effect of the production of the scFv163R4-A::β-lactamase on growth and cell viability, since high levels of production and translocation of recombinant proteins to the periplasm may result in growth arrest and cell lysis. In addition, high levels of β-lactamase activity could be the result of the release of cytoplasmic β-lactamase caused by cell lysis, leading to misleading results from the enzymatic assay. In addition, further evaluations would be required to understand the relationship between the mutations introduced in the signal peptide sequence and the physiological and protein production differences observed between different clones.

Analysis of mutations of signal peptides

With the main objective to expand the knowledge regarding the structure of the signal peptide and to identify which mutations resulted in a higher production of the scFv163R4-A::β-lactamase,

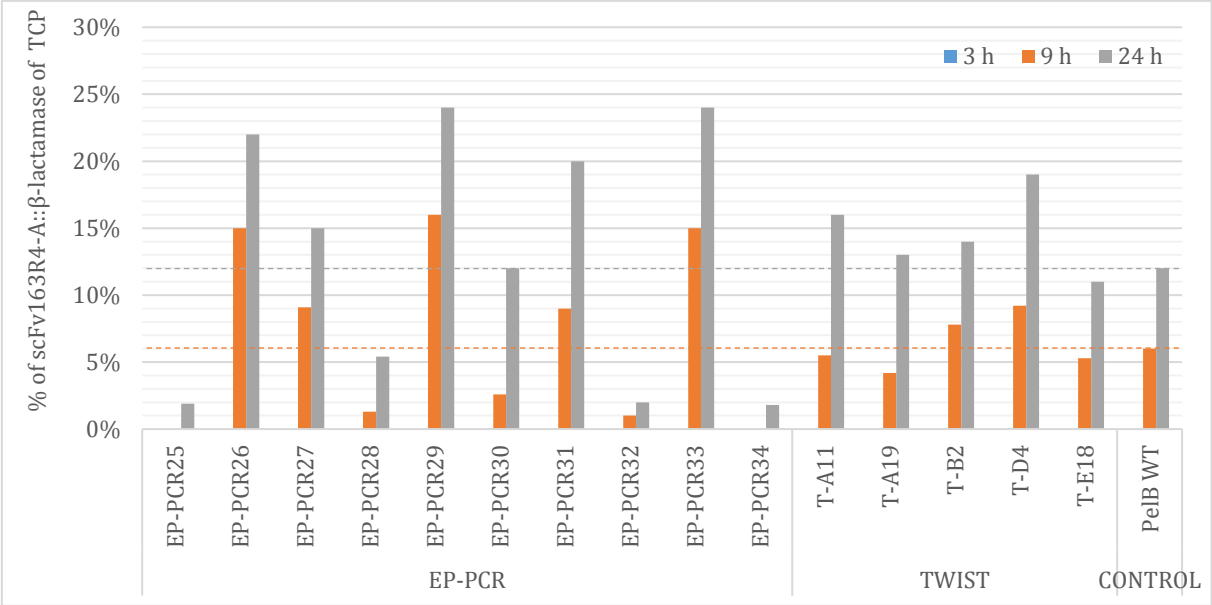


Figure 5.12. Evaluation of the production of scFv163R4-A::β-lactamase. Production of the scFv163R4-A::β-lactamase in cultures grown at 25 °C, induced with 0.02% arabinose at an OD_{600} of 0.5. The production of the scFv163R4-A::β-lactamase is represented as a percentage of the total cell protein (TCP) obtained after 3, 9 and 24 hours of growth. The production of PelB-scFv163R4-A::β-lactamase, used as control culture, after 9 (orange dotted line) and 24 hours of growth (grey dotted line) is represented as a guidance. Data shown here are single values obtained by quantifying the accumulation of the scFv163R4-A::β-lactamase in whole cell lysates by densitometry from SDS-PAGE gels.

plasmid DNA samples corresponding to 10 clones obtained from Ep-PCR signal peptide libraries (EP-C25, EP-C26, EP-C27, EP-C28, EP-C29, EP-C30, EP-C31, EP-C32, EP-C33 and EP-C34) and 5 clones obtained from the Twist libraries (T-A11, T-A19, T-B2, T-D4 and T-E18), were sequenced.

Alignments of the DNA and protein sequences obtained from clones selected from the error-prone PCR can be found in the **Figure 5.13** and from the Twist library in the **Figure 5.14**. In general, a high variability within the number of mutations was found in sequences obtained from the error-prone PCR, with sequences containing only 1 mutation to sequences with up to 10 mutations (**Figure 5.13**). DNA sequencing results showed that two clones with a low or non-detectable β-lactamase activity (**Figure 5.5**), EP-C25 and EP-C34, had insertions or deletions, which altered the open reading frame (ORF). EP-C25 presented an insertion which introduced a stop codon at the end of the signal peptide sequence, which could have resulted in the generation of a truncated protein, whereas EP-C34 showed an alteration in the ORF which could have led to generation of a cytoplasmic version of the scFv163R4-A::β-lactamase. The methionine located at -2 position of the signal peptide could act as an initiation codon, generating a cytoplasmic form of the scFv163R4-A::β-lactamase containing the last two amino acids of the signal peptide, an extra methionine and alanine residues at position +1 and +2. The alteration of the ORF of the EP-C25 and EP-C34 could explain the low level of β-lactamase activity and production of the scFv163R4-A::β-lactamase showed by these clones (**Figures 5.5 and 5.12**). In general, clones with a similar level of β-lactamase activity (**Figures 5.5 and 5.6**) and protein production level (**Figures 5.12**) as the wild type PelB signal peptide showed a small number of mutations, including silent mutations, which introduced changes in the codon usage but not in the amino acid sequence, such as EP-C28 and EP-C30 (**Figures 5.13**).

The hydrophobicity of the signal peptides was evaluated by calculating the grand average of hydropathicity (GRAVY), as it is known that hydrophobicity of the signal peptide plays an

Figure 5.13. Alignment of mutant signal peptides obtained from the Ep-PCR libraries. Alignment of DNA sequence and amino acid sequence of signal peptides obtained from the error-prone PCR library C and evaluated by shake-flask experiments. Amino acid colour code: polar positive (light blue), polar negative (red), polar neutral or hydrophilic (green), non-polar or hydrophobic (yellow) and special conformation (dark blue for proline and pink for glycine). The grand average of hydropathicity (GRAVY) was calculated for each signal peptide.

```
CLUSTAL O (1.2.3) multiple sequence alignment: DNA sequence

PelB wild type          ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCC
T-A11 (H)               ATGTTATACCTGACAAGGACCGCTATAGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCC
T-A19 (H)               ATGAAATACCTGCTGAAGACCGCTTTGCTGGTCTGCTGATCCTCGCTGCCCCCCGGCGATGGCC
T-B2  (H)               ATGAAATACCTGCTGCGGACCGCTGCTGCTGCACTTCTGCTCCTCGCTGCCCAGCCGGCGATGGCC
T-D4  (H)               ATGAAATACCTGCTGATGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCTGGGTAGCGATGGCC
T-E18 (H)               ATGAAATACCTGCTGCAAACCGCTGCTGCTTATCTGCTACTCCTCGCTGCCCAGCCGATTATGGCC
                        ***  *******      ******   ***   ** **  ***********         ******
```

```
CLUSTAL O (1.2.3) multiple sequence alignment          GRAVY          Amino acid changes

PelB wild type          MKYLLPTAAAGLLLLAAQPAMA          1.19                  -
T-A11 (H)               MLYLTRIAIAGLLLLAAQPAMA          1.33                  4
T-A19 (H)               MKYLLKIAFAGLLILAAPPAMA          1.25                  4
T-B2  (H)               MKYLLRIAAAALLLLAAQPAMA          1.16                  2
T-D4  (H)               MKYLLMIAAAGLLLLAAWVAMA          1.73                  3
T-E18 (H)               MKYLLQIAAAYLLLLAAQPIMA          1.19                  3
                        * **  ** * ** ***   **
```

Figure 5.14. Alignment of mutant signal peptides obtained from the Twist libraries. Alignment of DNA sequence and amino acid sequence of signal peptides obtained from the Twist library and evaluated by shake-flask experiments. Amino acid colour code: polar positive (light blue), polar negative (red), polar neutral or hydrophilic (green), non-polar or hydrophobic (yellow) and special conformation (dark blue for proline and pink for glycine). The grand average of hydropathicity (GRAVY) was calculated for each signal peptide.

important role in protein translocation. It has been reported that highly hydrophobic signal peptides have a higher affinity for the translocation machinery (Izard *et al.,* 1996). Nevertheless, highly hydrophobic signal peptides, as T-D4, with a GRAVY score of 1.73 (**Figure 5.14**), did not correlate with high levels of accumulation of the scFv163R4-A::β-lactamase in the periplasm (**Figure 5.11 C**). On the other hand, several signal peptides showed lower hydrophobicity than PelB wild type (average hydrophobicity of 1.19), obtaining hydrophobicity score lower than 1, such as EP-C26, EP-C27, EP-C29 and EP-C33 (**Figure 5.13**). None of these signal peptides showed a high concentration of protein in the periplasm either. This example helps to illustrate how difficult is to elucidate which mutations within the signal peptide sequence caused the high variability between clones observed in terms of growth, cell viability and protein production.

In general terms, it has been proved to be difficult to identify which mutations favour a higher production and translocation of the scFv163R4-A::β-lactamase based on results obtained from the β-lactamase enzymatic assay, shake-flask experiments and the DNA sequencing results. As for example, a certain mutation in the N-terminus may have a negative effect, which could be alleviated by a second mutation that increases the average hydrophobicity of the signal peptide (Izard and Kendall, 1994). This could mask the effect of each independent mutation, hampering the understanding of relationship between the signal peptide mutations and the production of the precursor or the mature form of the protein and its effect on cell physiology. Further studies will be required with the aim to understand the effect of different individual mutations on the production and translocation of recombinant proteins to the periplasmic space.

5.2.9 Fed-batch fermentation of clones selected from the Ep-PCR and Twist libraries for the production of scFv163R4-A::β-lactamase (early induction)

As some mutant signal peptides selected from the error-prone and Twist signal peptide libraries showed a higher level of production of the scFv163R4-A::β-lactamase than the wild type signal

peptide, the next step was the evaluation by fed-batch fermentation. For this purpose, the Ambr® 250 modular, a benchtop bioreactor system that allows performing up to 8 parallel fermentations, was used for evaluation of the clones selected by the β-lactamase screening assay. As a proof-of-concept, 5 clones with the best performance in terms of growth, cell viability and production of the scFv163R4-A::β-lactamase were selected to be evaluated by fed-batch fermentation: EP-C26, EP-C33, T-A11, T-A19 and T-B2. The PelB wild type signal peptide fused to the scFv163R4-A::β-lactamase was used as control.

Fed-batch fermentations were carried out at 250 mL scale and cultures were grown in 150 mL of the semi-defined culture medium obtained from Want *et al.* (2009). The feed solution, composed of glycerol and magnesium sulphate, was started 10 hours post-inoculation following an exponential feeding profile to achieve a specific growth rate of 0.1 h^{-1}. Cultures were grown were at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5. The dissolved oxygen was maintained at 20% by continuous increments of the stirrer speed and pure oxygen was added when required. The pH was kept at 7.0 by the addition of 1 M NH_4OH or 1 M HCl.

In terms of growth, only two of the clones, EP-C26 and EP-C33 were able to grow to an OD_{600} above 12 during fed-batch fermentation, reaching a final OD_{600} of 94 and 82, after 48 hours of growth (**Figure 5.15 A**). Clones T-A11, T-A19, T-B2 and the control culture (PelB WT) showed signs of growth arrest after 12 hours of growth, reaching a final OD_{600} lower than 10, after 48 hours. A similar profile was observed in terms of cell viability, where a decrease in the CFU was observed 12 hours post-inoculation, for clones T-A11, T-A19, T-B2 and the control culture, PelB WT (**Figure 5.15 B**). In contrast, clones EP-C26 and EP-C33 showed a continuous increase in the CFU during through the course of the fed-batch fermentation without loss of cell culturability.

The production of the scFv163R4-A::β-lactamase was evaluated for EP-C26 and EP-C33 and showed a high concentration of the recombinant protein in the spheroplast protein fraction (Data not shown). No accumulation of the scFv163R4-A::β-lactamase was detected in the periplasmic

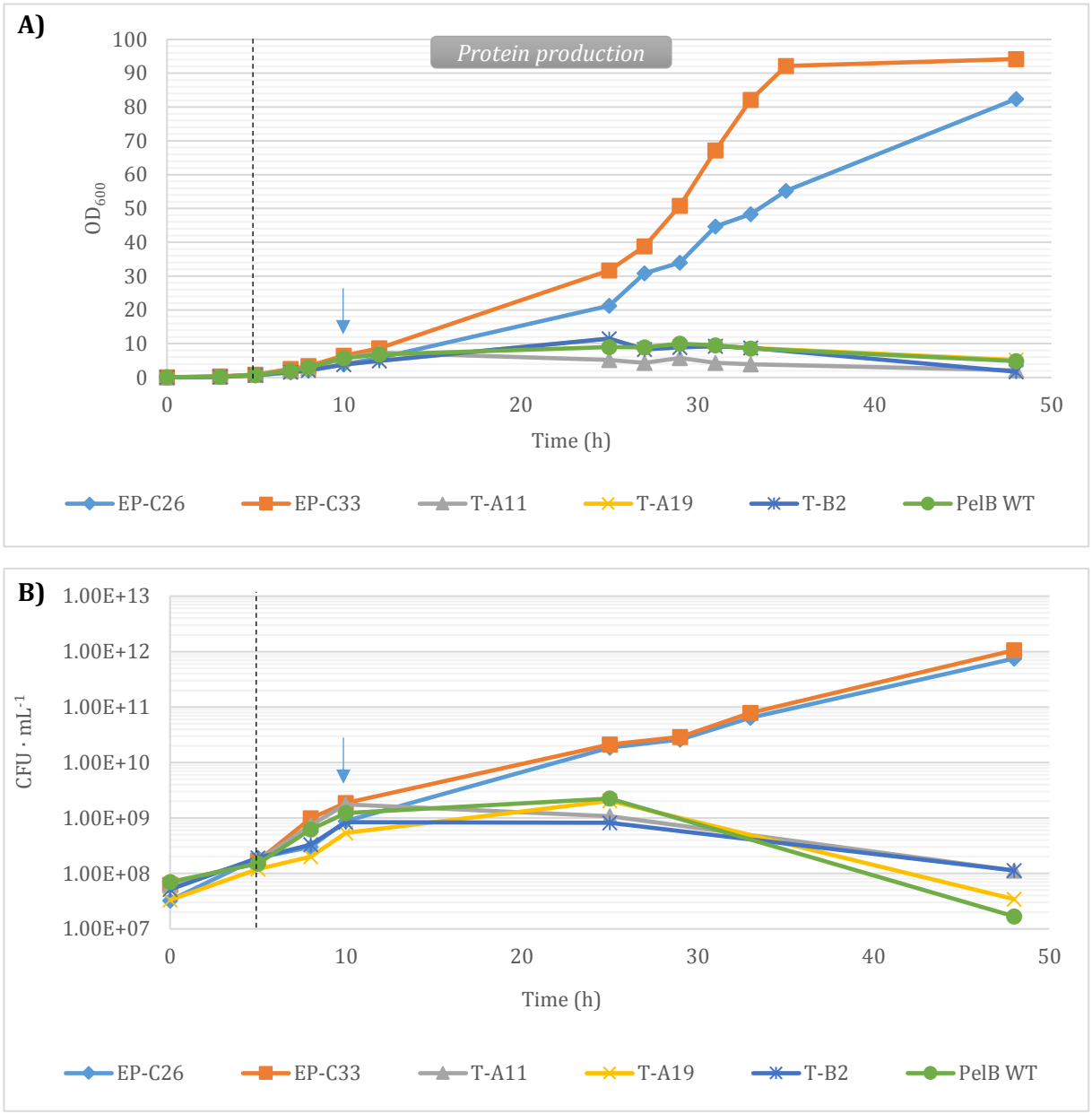


Figure 5.15. Fed-batch fermentation for the production of scFv163R4-A::β-lactamase induced at low cell density. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the error-prone PCR and Twist signal peptide libraries fused to the scFv163R4-A::β-lactamase were grown at 25 °C and induced with 0.02% arabinose at an OD_{600} of 0.5 (dotted line). The feed was started after 10 hours of growth (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at intervals post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data shown are single values for culture samples.

fraction, correlating with results previously obtained from shake-flask experiments (**Figure 5.9 B**). The superior growth and cell culturability observed for clones EP-C26 and EP-C33 could be the result of the accumulation of the scFv163R4-A::β-lactamase in inclusion bodies, which helped to minimise the metabolic burden imposed to the host cells. The possible overload of the SecB-dependent translocation pathway could had been the cause of the growth arrest and cell lysis observed for clones T-A11, T-A19 and T-B2 and the control culture PelB WT. Nevertheless, this result led to the conclusion than the growth conditions for the production of the scFv163R4-A::β-lactamase were not optimal, as the control culture carrying the PelB wild type signal peptide was not able to grow by fed-batch fermentation. The cultivation conditions need to be optimised, aiming to avoid the possible overload the SEC translocation pathway and favour the production of the scFv163R4-A::β-lactamase in the periplasmic space.

5.2.10 Fed-batch fermentation of clones selected from the Ep-PCR and Twist libraries for the production of scFv163R4-A::β-lactamase (late induction)

With the aim to develop a fermentation protocol which allows the evaluation of growth and production scFv163R4-A::β-lactamase from clones selected from the error-prone PCR and the Twist libraries, a late induction protocol was evaluated. Fed-batch fermentations were carried out as described previously, but the induction point was shifted to a target OD_{600} of 70-80, to allow the separation of the biomass and protein production stages. The cultivation temperature was 30 °C, and upon induction, shifted to 25°C, to favour a slower protein synthesis rate, avoiding the overload of the translocation machinery. The feed solution was started 10 hours post-inoculation, and adjusted based on growth (OD_{600}) after 16 hours of inoculation, following an exponential feeding profile to achieve a specific growth rate of 0.1 h^{-1}.

In general, the modification of the cultivation conditions had a positive effect in terms of growth and all the clones showed a similar growth profile, reaching a final OD_{600} ranging between 140

and 160, after 28 hours of growth (**Figure 5.16 A**). Loss of cell culturability in samples taken post-induction was observed (**Figure 5.16 B**); however, the loss of colony forming units was not as severe as in the previous fermentation (**Figure 5.15 B**).

However, the high biomass obtained at the end of the fermentation did not correlate with a high production of the scFv163R4-A::β-lactamase (**Figure 5.17 A**). The accumulation of scFv163R4-A::β-lactamase was lower in samples taken post-induction when compared with results obtained during shake-flask experiments (**Table 5.1**). Clones EP-C26 and EP-C33 showed a production of the scFv163R4-A::β-lactamase, accounting for 7% and 9% of total cell protein (TCP), after 26 hours of growth. An even lower accumulation was detected 28 hours post-inoculation, possibly as a result of plasmid loss, proteolysis or protein degradation. The concentration of scFv163R4-A::β-lactamase was higher in samples obtained from clones T-A11 and T-B2 after 28 hours of growth, reaching a 9% of TCP (**Figure 5.17 A**). Samples obtained from clone T-A19 showed a similar production of the scFv163R4-A::β-lactamase, accounting for 7% of TCP, to the control culture carrying the PelB signal peptide.

No accumulation of the protein in the periplasm was observed from samples obtained from clones EP-C26 and EP-C33 (**Figure 5.17 B**). Nevertheless, the large molecular weight of the scFv163R4-A::β-lactamase (≈60 kDa) could impede the extraction of the protein, and as some protein can be found accumulated in the spheroplast protein fraction. In contrast, samples obtained from clones T-A11, T-A19 and T-B2 and the wild type PelB signal peptide showed a high concentration of the scFv163R4-A::β-lactamase in the periplasmic fraction, accounting for 35% to 44% of the total scFv163R4-A::β-lactamase produced.

A summary of data obtained from shake-flask experiments (**section 5.2.8**) and fed-batch fermentation experiments (**section 5.2.10**) in terms of scFv163R4-A:: β-lactamase can be found in **Table 5.1**. Data corresponding to OD_{600} was measured and percentage of scFv163R4-A::β-

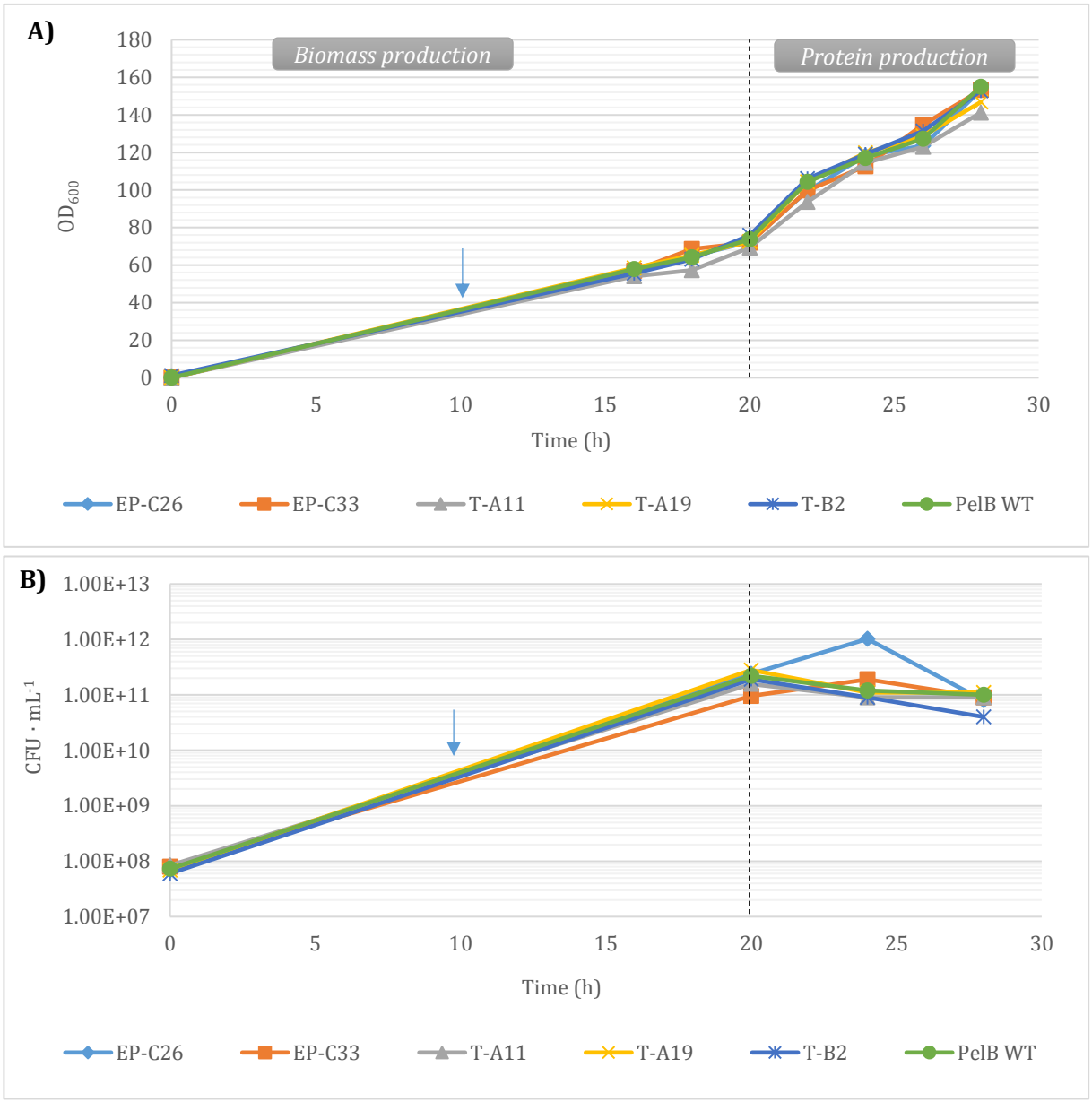


Figure 5.16. Fed-batch fermentation for the production of scFv163R4-A::β-lactamase. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the error-prone PCR and Twist signal peptide libraries fused to the scFv163R4-A::β-lactamase were grown at 30 °C, and upon induction at 25 °C. Cultures were induced with 0.02% arabinose at an OD_{600} ≈ 70 – 80 (dotted line). The feed was started after 10 hours of growth (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at intervals post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data shown are single values for culture samples.

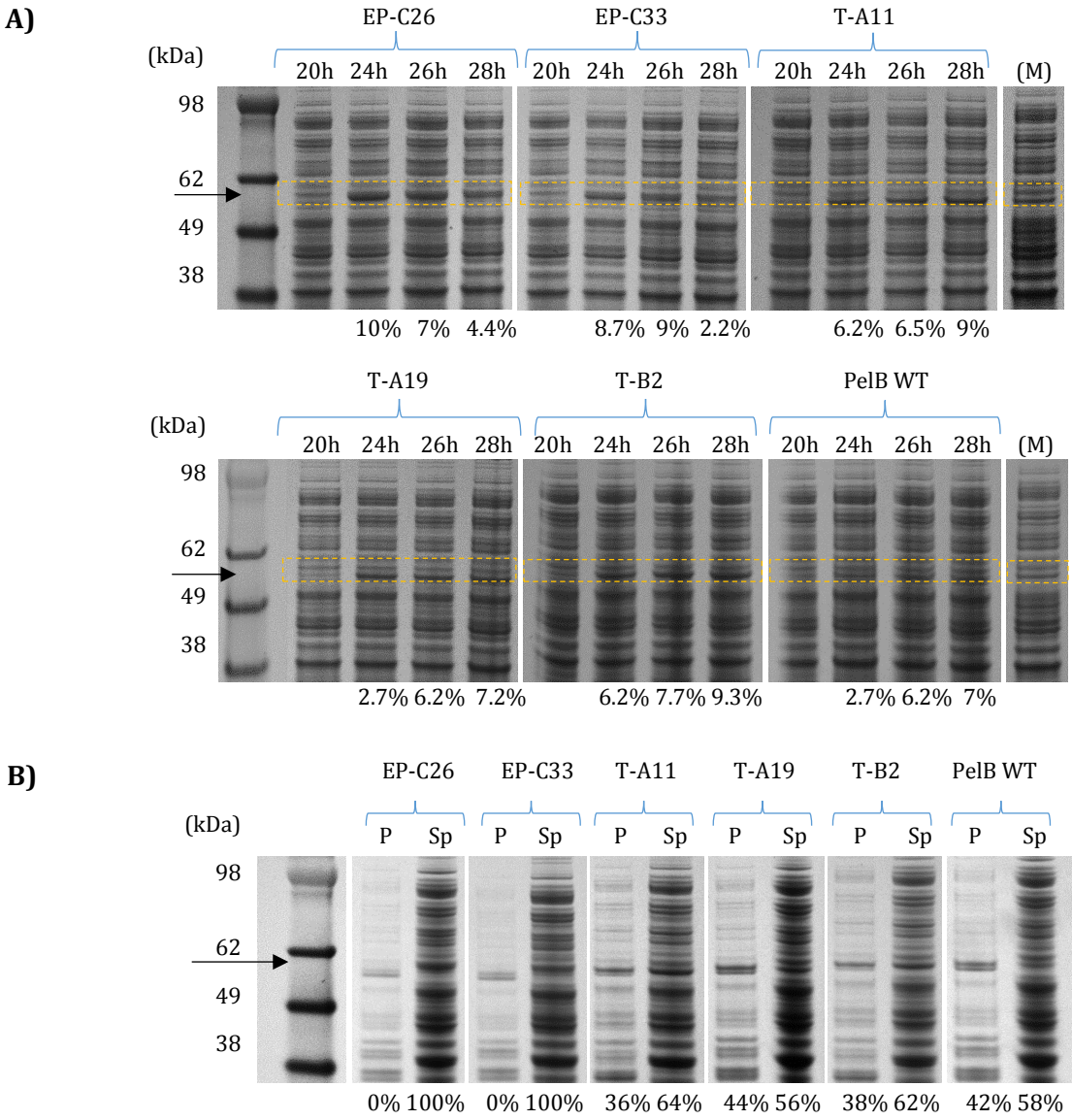


Figure 5.17. Production of scFv163R4-A::β-lactamase by fed-batch fermentation. SDS-PAGE gels showing the accumulation of scFv163R4-A::β-lactamase (black arrow) from whole cell lysates obtained from cultures grown at 30 °C, and upon induction at 25 °C. Cultures were induced with 0.02% arabinose at an $OD_{600} \approx 70 - 80$. Samples were obtained before (20 h) and after induction (24, 26 and 28 hours post-inoculation) (A). The periplasm (P) and spheroplast (Sp) protein fractions were obtained by modified cold osmotic shock from samples obtained 28 hours post-inoculation (B). Samples obtained from cultures carrying the scFv163R4-A::β-lactamase produced in the cytoplasm (M) after 24 hours of growth were used as a control. The quantity of the scFv163R4-A is expressed as percentage of whole cell protein (A) and the ratio of protein accumulated in the periplasm and spheroplast protein fractions (B) are expressed as percentage at the bottom of the gel.

Table 5.1. Comparison of data obtained from shake-flask and fed-batch fermentation experiments at harvest in terms of production of scFv163R4-A::β-lactamase.

	Shake-flask data						***Fed-batch fermentation data***					
Clone name	EP-C26	EP-C33	T-A11	T-A19	T-B2	PelB wt	EP-C26	EP-C33	T-A11	T-A19	T-B2	PelB wt
Growth temperature	25 °C						30 °C → 25 °C					
Inducer concentration	0.02%						0.02%					
Induction time	OD_{600} ≈ 0.5						OD_{600} ≈ 70-80					
Cultivation time	24 h						28 h					
Final OD_{600}	21.7	24.4	12	12.6	11.2	12	153.2	153.2	141.2	146.8	152.8	155
scFv163R-A::Bla as % of TCP	22%	24%	16%	13%	14%	12.3%	4.4%	2.2%	9%	7.2%	9.3%	7%
% of periplasmic scFv163R-A::Bla	7%	7%	38%	40%	35%	10%	0%	0%	36%	44%	38%	42%
% of spheroplast scFv163R-A::Bla	93%	93%	62%	60%	65%	90%	100%	100%	64%	56%	62%	58%
Total scFv163R-A::Bla (g · L⁻¹)	1.15[1]	1.42[1]	0.46[1]	0.39[1]	0.38[1]	0.36[1]	1.6[1]	0.8[1]	3.1[1]	2.5[1]	3.4[1]	2.6[1]
Periplasmic scFv163R-A::Bla (g · L⁻¹)	0.08[2]	0.1[2]	0.17[2]	0.16[2]	0.13[2]	0.04[2]	0[2]	0[2]	1.12[2]	1.1[2]	1.29[2]	1.09[2]
Spheroplast scFv163R-A::Bla (g · L⁻¹)	1.07[3]	1.32[3]	0.29[3]	0.23[3]	0.25[3]	0.32[3]	1.6[3]	0.8[3]	1.98[3]	1.4[3]	2.11[3]	1.51[3]

[1] Estimated from DCW assuming 1 OD_{600}=0.4 g · L^{-1} and % of scFv163R4-A::Bla of total cell protein assuming protein comprises 60% *E. coli* dry cell weight (g · L^{-1}) (based on 50-61% estimates from Valgepea *et al.* (2013)).
[2] Estimated from DCW and % of periplasmic scFv163R4-A::Bla or [3] % of spheroplast scFv163R4-A::Bla of total cell protein calculated from the estimated total yield of scFv163R4-A::Bla (g · L^{-1}).

lactamase from total cell protein was obtained from SDS-PAGE and quantified by densitometry. The scFv163R4-A::β-lactamase yields are estimated, shown here to illustrate the comparison between the data obtained from shake-flask and fermentation experiments.

In general terms, the production level of scFv163R4-A::β-lactamase was not comparable between shake-flask and fermentation experiments (**Table 5.1**). In shake-flask experiments, the production of the scFv163R4-A::β-lactamase for clones EP-C26 and EP-C33 was over 20% of the total cell protein, comparing with less than 5% obtained by fed-batch fermentation. The accumulation of the scFv163R4-A::β-lactamase was higher for clones EP-C26 and EP-C33 between 24 and 26 hours of growth reaching between 9% and 10% of TCP (**Figure 5.17**), but only data obtained at the harvest was included in **Table 5.1**. The difference in protein production between shake-flask and fermentation experiments was lower in clones T-A11, T-A19 and T-B2, with over 12% of TCP corresponding to scFv163R4-A::β-lactamase in samples obtained from shake-flask experiments and less than 10% in samples obtained from fermentation experiments. Similarly, results obtained from the control culture carrying the PelB signal peptide showed almost a 2-fold reduction in the total production of scFv163R4-A::β-lactamase. In terms of the concentration of scFv163R4-A::β-lactamase in the periplasmic and spheroplast fractions, results obtained from shake-flask and fed-batch fermentation had a higher degree of similarity.

Overall, large differences between the total scFv163R4-A::β-lactamase yields obtained from shake-flask experiments and fed-batch fermentation were observed. Nevertheless, it is important to highlight the differences in the induction point, induction at an $OD_{600} \approx 0.5$ for shake-flask experiments and $OD_{600} \approx 70\text{-}80$ for fed-batch fermentation, which could have a great impact in recombinant protein production. The "production window" varied between shake-flasks, with an induction window of approximately 20 hours, and fed-batch fermentation, where cultures were induced for a period of 12 hours after induction. Also, differences in the growth profile and final

OD_{600}, a maximum final OD_{600} of 25 for shake-flasks and 155 for fed-batch fermentations, could affect protein production.

In general terms, the late induction favoured the growth of all the bioreactor cultures reaching a high cell density (**Figure 5.15 and 5.16**). However, the low production of the scFv163R4-A::β-lactamase suggested that further optimisation of the fermentations conditions will be required with the aim to improve the yield of scFv163R4-A::β-lactamase.

5.2.11 Fed-batch fermentation of clones selected from the Ep-PCR and Twist libraries for the production of scFv163R4-A

With the aim to evaluate whether the signal peptides selected from the EP-PCR and Twist libraries could show an improvement in the production of the scFv163R4-A, the reporter protein, β-lactamase, was removed from the vector. Fed-batch fermentations were carried out using the same conditions described in **Section 5.2.10**. The feed was started 14 hours post-inoculation, upon depletion of glycerol in the batch medium (determined by a DO spike) following an exponential feeding profile to achieve a specific growth rate of 0.1 h^{-1}.

As observed in the previous fermentation, all the clones showed an excellent growth profile and the production window was extended to 12 hours post-induction to favour the production of scFv163R4-A. Cultures reached a final OD_{600} ranging from 155 to 186, after 30 hours of growth (**Figure 5.18 A**). No loss of cell viability was observed in samples obtained post-induction, but the number of CFU did not increase with the increment of cell biomass (**Figure 5.18 B**).

As in the previous fermentation, the production of the scFv163R4-A was lower than expected and did not correlate with the production levels observed in shake-flask experiments (**Figure 5.19**). The low concentration of the scFv163R4-A did not allow an accurate quantification and the production of the scFv163R4-A was confirmed by western blot using an anti-myc antibody.

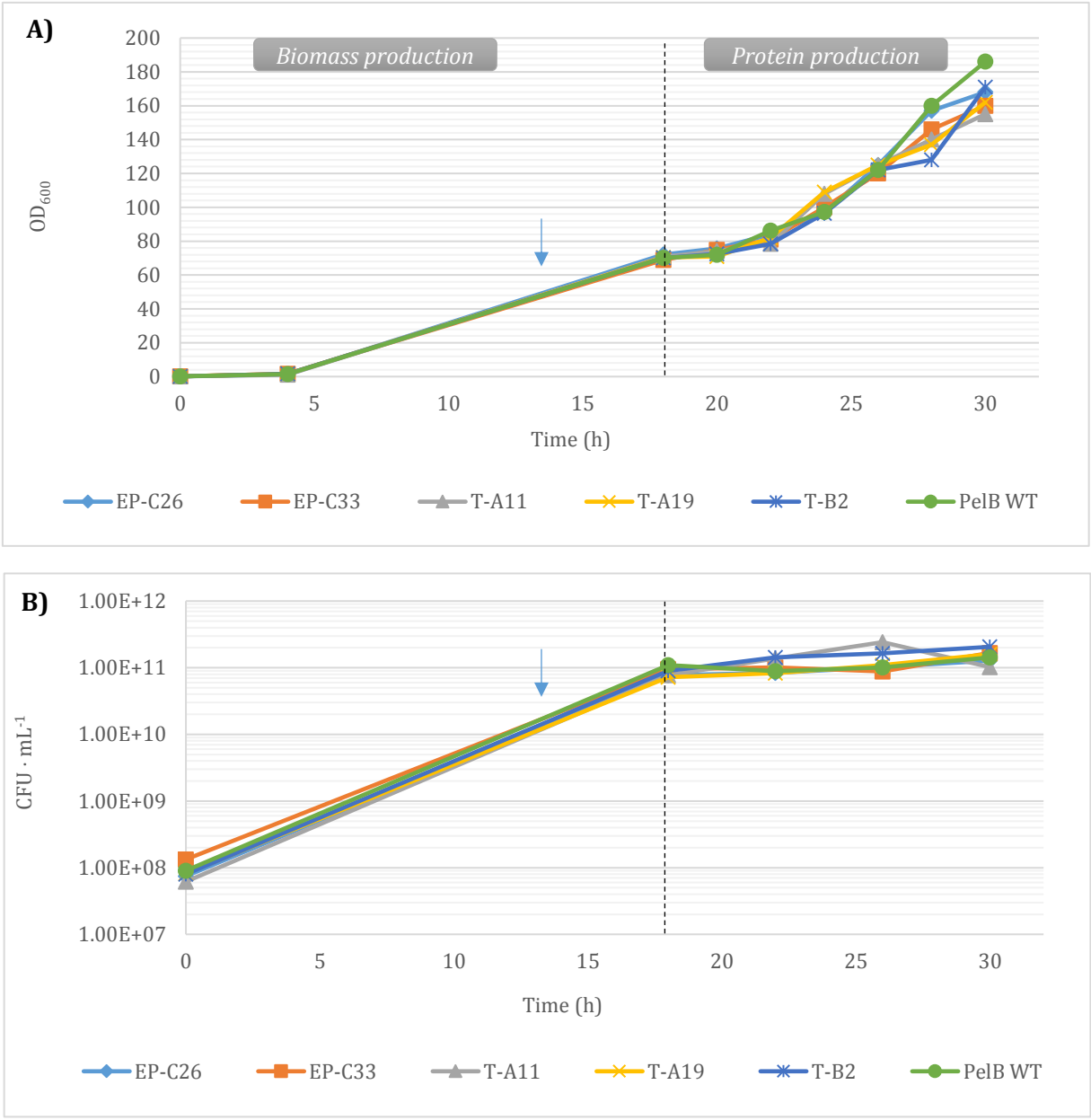


Figure 5.18. Fed-batch fermentation for the production of scFv163R4-A. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the error-prone PCR and Twist signal peptide libraries fused to the scFv163R4-A were grown at 30 °C, and upon induction at 25 °C. Cultures were induced with 0.02% arabinose at an OD_{600} ≈ 70 - 80 (dotted line). The feed was started after 14 hours of growth (blue arrow). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at intervals post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data shown are single values for culture samples.

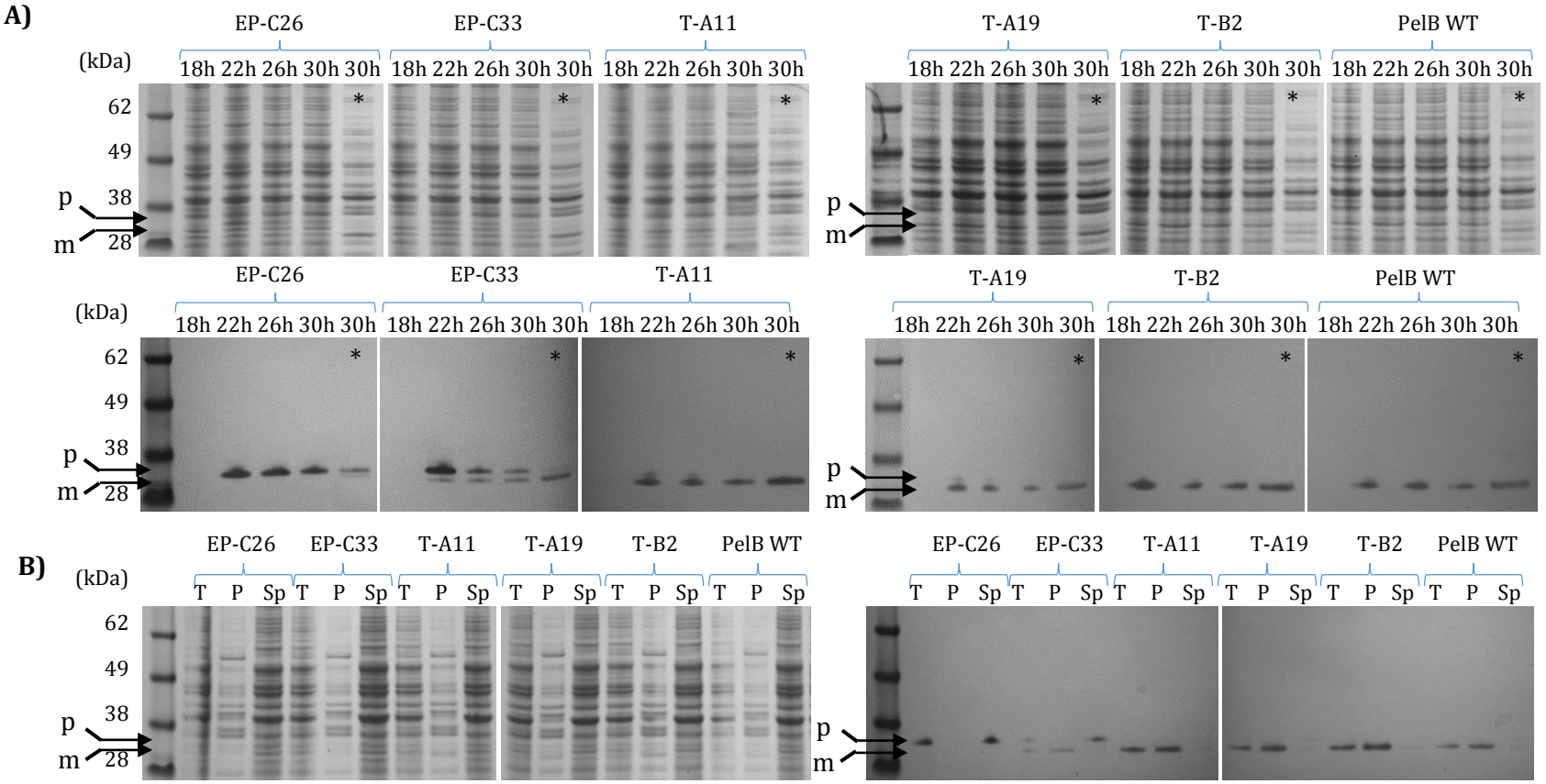


Figure 5.19. Production of scFv163R4-A by fed-batch fermentation. SDS-PAGE gels showing the accumulation of the precursor (p) and mature form (m) of the scFv163R4-A from whole cell lysates and culture medium samples (*) obtained from cultures induced with 0.02% arabinose at an OD_{600} ≈ 70 - 80. Samples were obtained before (18 h) and after induction (22, 26 and 30 h post-inoculation) (A). The periplasm (P) and spheroplast (Sp) protein fractions were obtained by modified cold osmotic shock from samples obtained 30 hours post-inoculation (B). The scFv163R4-A was detected by western blot using an anti-myc antibody.

The precursor form of the scFv163R4-A was detected in samples obtained from EP-C26, being only detected in the samples corresponding to whole cell lysates or the spheroplast protein fraction (**Figure 5.19 B**). The two forms of the protein, the precursor and the mature, could be identified in samples obtained from clone EP-C33. In samples obtained from EP-C33, a high accumulation of the precursor form of the scFv163R4-A was observed in samples taken 22 hours post-inoculation and the ratio between the precursor and the mature form of the scFv163R4-A seems to shift to higher accumulation of the mature form of the scFv163R4-A over the course of the fermentation (**Figure 5.19 A**). The mature form of the protein was only detected in the periplasmic protein fraction (**Figure 5.19 B**) and culture medium (**Figure 5.19 A**). The high accumulation of the precursor form of the scFv163R4-A correlates with the results obtained from initial fermentation experiments, when under early induction conditions, only clones EP-C26 and EP-C33 were able to grow by fed-batch fermentation (**Figure 5.15**).

The production of the scFv163R4-A was also detected in whole cell lysates obtained from clones T-A11, T-A19 and T-B2 and the control culture carrying the PelB signal peptide (**Figure 5.19 A**). The mature form of the scFv163R4-A was the predominantly observed, as confirmed by western blot, being detected in samples corresponding to the whole cell lysates, periplasmic protein fraction and the culture medium (**Figure 5.19 B**). Only traces of the scFv163R4-A were detected in the spheroplast protein fraction, probably as the result of residual periplasmic content present in the spheroplast fraction.

The scFv163R4-A was detected in samples obtained from the culture medium (**Figure 5.19 A**). Since *E. coli* does not naturally secrete recombinant proteins to the culture medium, the release of the scFv163R4-A into the culture medium could be the result of outer membrane damage and subsequent release of the protein to the culture medium (Georgiou and Segatori, 2005; Hsu *et al.,* 2016).

In general terms, it seems that signal peptides obtained from clones T-A11, T-A19 and T-B2 allowed the translocation of the protein to the periplasmic space. The optimisation of the fermentation process allowed high cell densities (OD_{600} ≥ 150) to be achieved and the extension of the "production window" to 12 hours, with no detrimental effect on cell viability. However, the production of scFv163R4-A seems to be lower than observed in shake-flask experiments (**Figure 5.9 and 5.11**), being not possible to quantify the production of the scFv163R4-A from SDS-PAGE gels, confirming the production of the antibody fragment by western blot. It is important to highlight the difference in the molecular weight between the scFv163R-A::β-lactamase (≈60 kDa) and the scFv163R4-A (≈30 kDa), which could explain the differences in the protein yields obtained by fed-batch fermentation. The smaller size of the scFv163R4-A also facilitates its translocation to the periplasm, and its extraction by osmotic shock. Further fermentation optimisation would be required to achieve a higher protein production without comprising cell viability.

5.2.12 Improved fermentation conditions for the production scFv163R4-A

It was considered that a higher concentration of arabinose may be required to induce the production of the scFv163R4-A at higher cell densities, since the number of arabinose molecules per cell would be lower than for shake-flask cultures induced at lower cell densities. A 10-fold higher concentration of arabinose, 0.2%, than in previous experiments was used to induce the culture at an OD_{600} ≥ 70, with the aim of increasing the production of scFv163R4-A without having a deleterious effect on growth and cell viability. Fed-batch fermentations were carried out as described in previous sections and the feed was started 14 hours post-inoculation following an exponential feeding profile to achieve a specific growth rate of 0.1 h^{-1}.

In terms of growth, even when a concentration of arabinose of 0.2% was used to induce the production of the scFv163R4-A, no signs of growth arrest were observed and the cultures reached a OD_{600} ranging from 140 to 175, after 30 hours of growth (**Figure 5.20 A**). In terms cell viability,

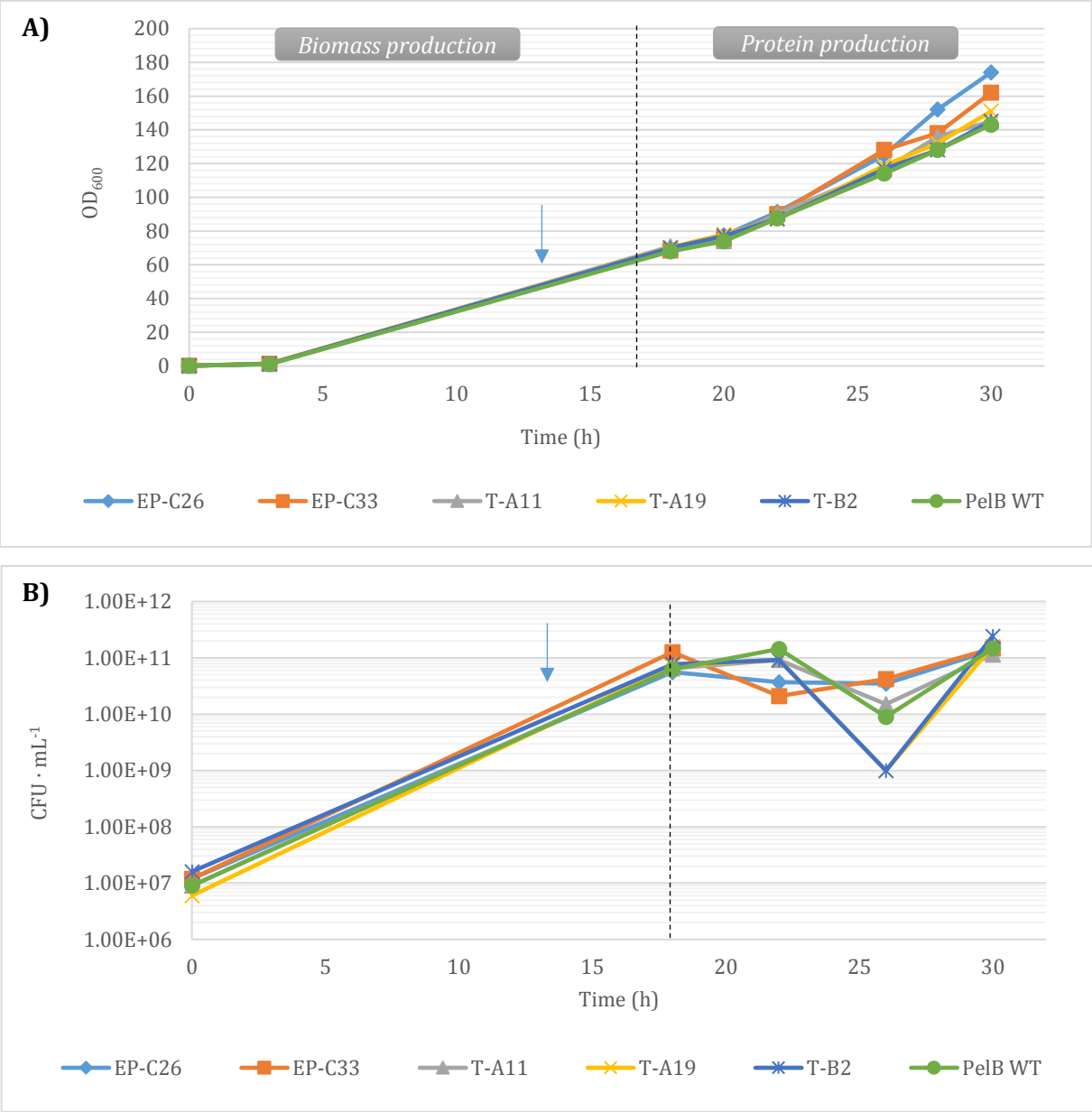


Figure 5.20. Fed-batch fermentation for the production of scFv163R4-A induced with 0.2% arabinose. *E. coli* BL21-A carrying the vector coding for the different signal peptides selected from the error-prone PCR and Twist signal peptide libraries fused to the scFv163R4-A were grown at 30 °C, and upon induction at 25 °C. Cultures were induced with 0.02% arabinose at an OD_{600} ≈ 70 - 80 (dotted line). The OD_{600} of the culture was measured at intervals post-inoculation (A). Serial dilutions of culture samples at intervals post-inoculation were plated onto non-selective LB agar to assess the cell viability after induction (B). Data shown are single values for culture samples.

a decrease in the CFU was observed after 26 hours of growth, which recovered at the end of the fermentation reaching between 1 and $2.5 \cdot 10^{11}$ CFU $\cdot$ mL^{-1} (**Figure 5.20 B**).

The higher concentration of arabinose used to induce protein expression favoured the production of the scFv163R4-A for all the cultures evaluated by fed-batch fermentation. As in previous fermentations, only the precursor form of the scFv163R4-A was detected in samples obtained from clones EP-C26 and EP-C33, accumulated in the spheroplast protein fraction (**Figures 5.21 and 5.22**). Clones T-A11 and T-B2 showed a high level of production of the scFv163R4-A in samples corresponding to whole cell lysates after 30 hours of growth, accounting for 4% to 5% of the total cell protein (**Figure 5.21**). Clone T-A19 showed a similar level of production of the scFv163R4-A to PelB wild type, accounting for at least 3% of the TCP in whole cell lysates, after 30 hours of growth. A high concentration of the scFv163R4-A was observed in the culture medium samples obtained from clones T-A11, T-A19, T-B2 and PelB-scFv163R4-A, accounting for at least 6% of the protein content in the culture medium.

A high concentration of the scFv163R4-A was detected in the periplasmic protein fraction obtained from clones T-A11, T-A19, T-B2 and PelB wild type, accounting for 24% to 30% of the total scFv163R4-A produced; being the second most abundant protein in the periplasm (**Figure 5.22**). However, the protein was detected mainly in the culture medium, accounting for 43% to 52% of the total scFv163R4-A produced. A low concentration of scFv163R4-A was detected in the spheroplast protein fraction.

Non-specific periplasmic protein release had been reported during the production of antibody fragments, typically when PelB signal peptide is used to direct the translocation to the periplasm (Georgiou and Segatori, 2005). Therefore, it is not surprising to observe the scFv163R4-A being released to the medium when a high level of protein is detected in the periplasmic space. The high accumulation of recombinant proteins may exceed the maximum capacity of the periplasm, favouring the leakage of the periplasmic proteins to the culture medium (Newton *et al.,* 2016).

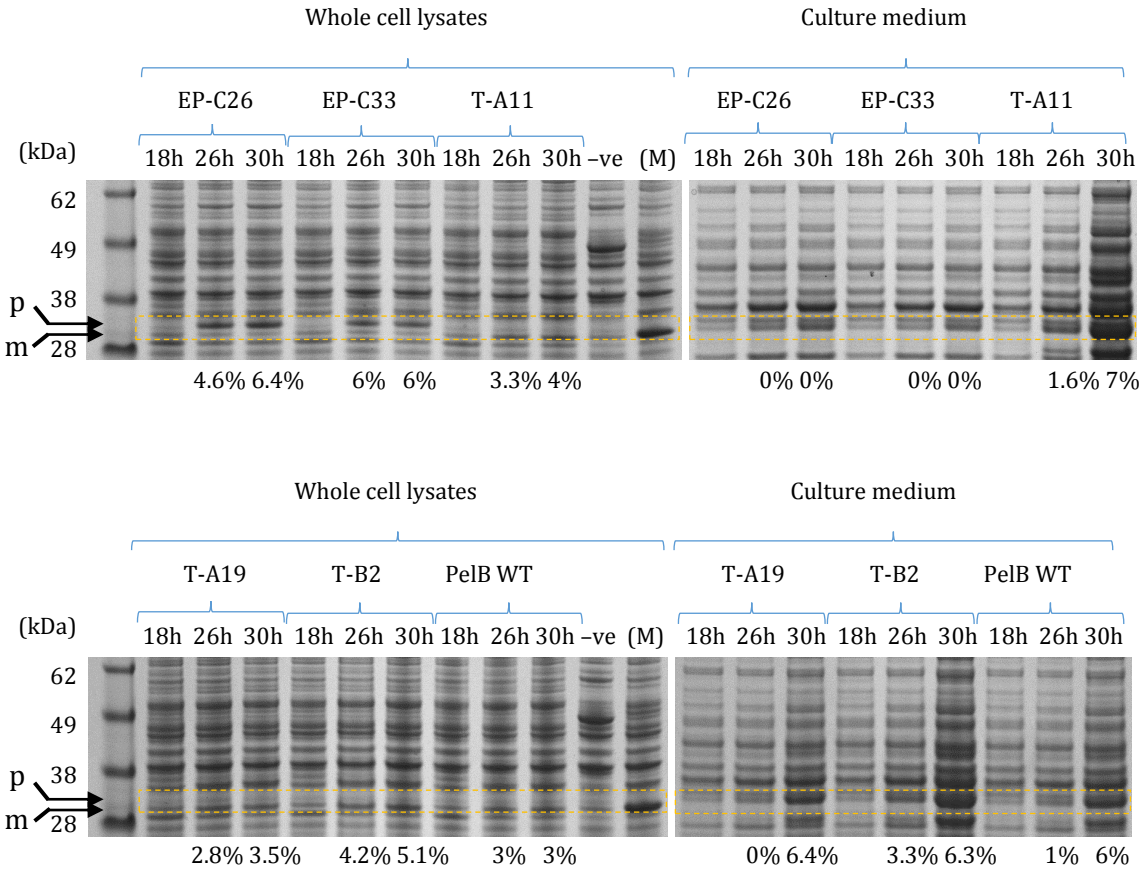


Figure 5.21. Production of scFv163R4-A by fed-batch fermentation induced with 0.2% arabinose. SDS-PAGE gels showing the accumulation of the precursor (p) and mature (m) form of the scFv163R4-A from whole cell lysates and culture medium samples obtained from cultures induced with 0.2% arabinose at an OD_{600} ≈ 70 - 80. Samples were obtained before (18 h) and after induction (26 and 30 h post-inoculation). Samples obtained from cultures carrying the empty vector (-ve) and cultures producing the cytoplasmic scFv163R4-A (M) after 24 hours of growth were used as a control. The quantity of the scFv163R4-A is expressed as percentage of whole cell protein or the culture medium protein at the bottom of the gel.

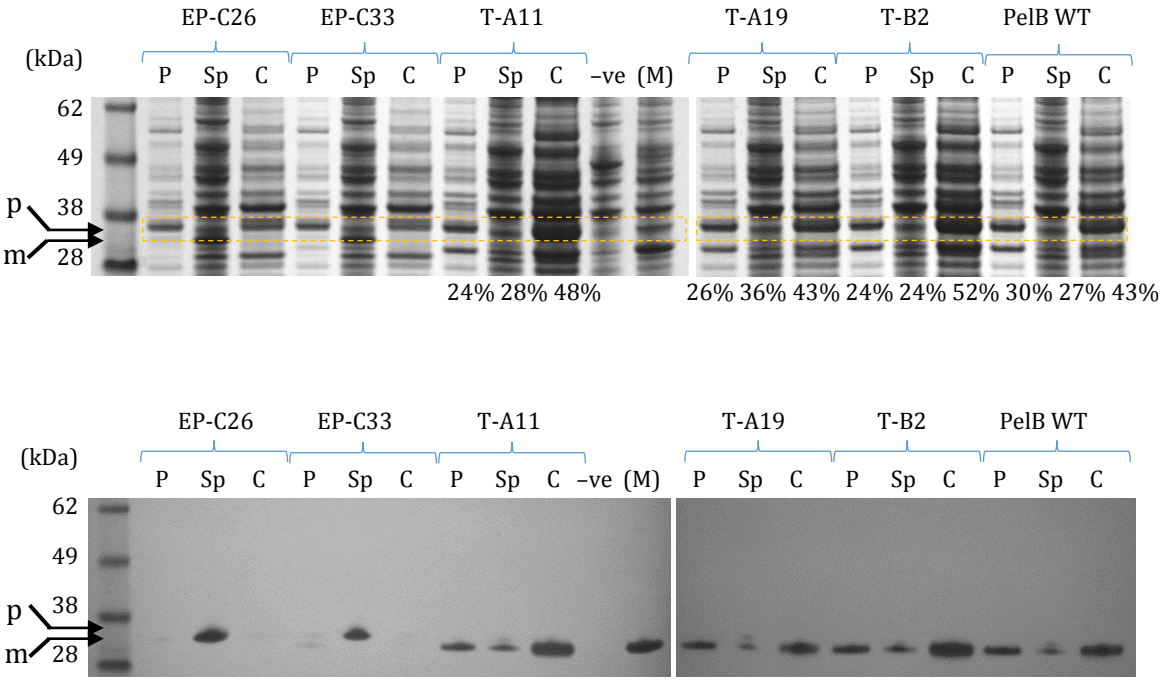


Figure 5.22. Production of scFv163R4-A in the periplasm, spheroplast and culture medium. SDS-PAGE gels showing the accumulation of the precursor (p) and mature (m) form of the scFv163R4-A from periplasm (P), spheroplast (Sp) and culture medium (C) samples obtained from cultures induced with 0.2% arabinose at an OD_{600} ≈ 70 – 80, after 30 hours of growth. Samples obtained from cultures carrying the empty vector (-ve) and cultures producing the cytoplasmic scFv163R4-A (M) after 24 hours of growth were used as a control. The western blot was developed using a mouse anti-myc antibody. The ratio of scFv163R4-A accumulated in the periplasm and spheroplast protein fractions are expressed as percentage at the bottom of the gel.

Overall, it seems that mutations introduced in the signal peptides of clones EP-C26 and EP-C33 diminished the translocation ability of the signal peptide, as slow or no processing of the scFv163R4-A was observed, the protein mainly being accumulated as precursor in the spheroplast fraction. In contrast, mutations introduced in the signal peptide sequence of clones T-A11, T-A19 and T-B2 allowed the translocation of the scFv163R4-A in the periplasm, and only the mature form of the protein was detected in protein samples. The production of mature scFv163R4-A of clones T-A11, T-A19 and T-B2 was slightly better or similar to production levels observed with the PelB wild type signal peptide.

In general terms, the optimisation of the fermentation conditions, late induction protocol in combination with a higher concentration of inducer, favoured the production of the scFv163R4-A. However, the production of the scFv163R4-A in the periplasm does not necessarily correlate with the production of properly folded and active product, and therefore, it would be required to evaluate the binding activity of the scFv163R4-A.

5.2.13 Purification of scFv163R4-A

As reference material was required for quantification and the development of an activity assay, the scFv163R4-A obtained from a culture carrying the PelB signal peptide was purified by Nickel-affinity chromatography. A culture medium sample obtained from a fed-batch fermentation carried out under identical conditions as those described in **section 5.2.12** was used as starting material for the purification process, as only the mature form of the scFv163R4-A was detected in this sample. Clarified culture medium was loaded onto a HisTrap FF column. The HisTrap FF column was washed by using a PBS pH 7.2 buffer containing 40 mM imidazole and the scFv163R4-A was eluted by using a PBS pH 7.2 buffer containing 100 mM imidazole (**Figure 5.23**). The final concentration of the purified scFv163R4-A material was 0.1 mg · mL^{-1}, determined by densitometry and measurement of the absorbance at 280 nm.

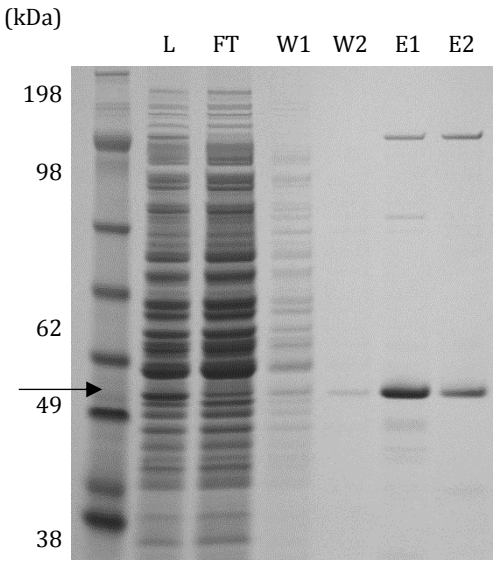
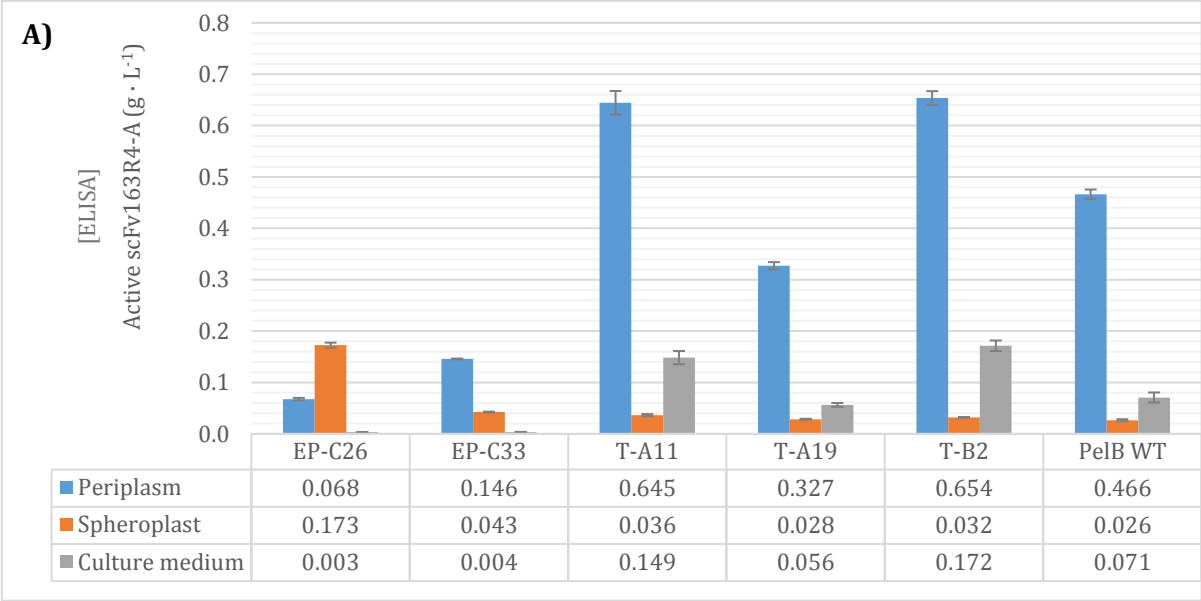


Figure 5.23. Purification of scFv163R4-A from PelB-scFv163R4-A. SDS-PAGE gel showing the clarified culture medium loaded (L) into the HisTrap FF column, the flow-through (FT) from the column, the material obtained after the washing steps (W1, W2) and the eluted material (E1 and E2). The scFv163R4-A was eluted in a PBS pH 7.2 buffer containing 100 mM imidazole (E1). A second elution with PBS pH 7.2 buffer containing 500 mM imidazole (E2).

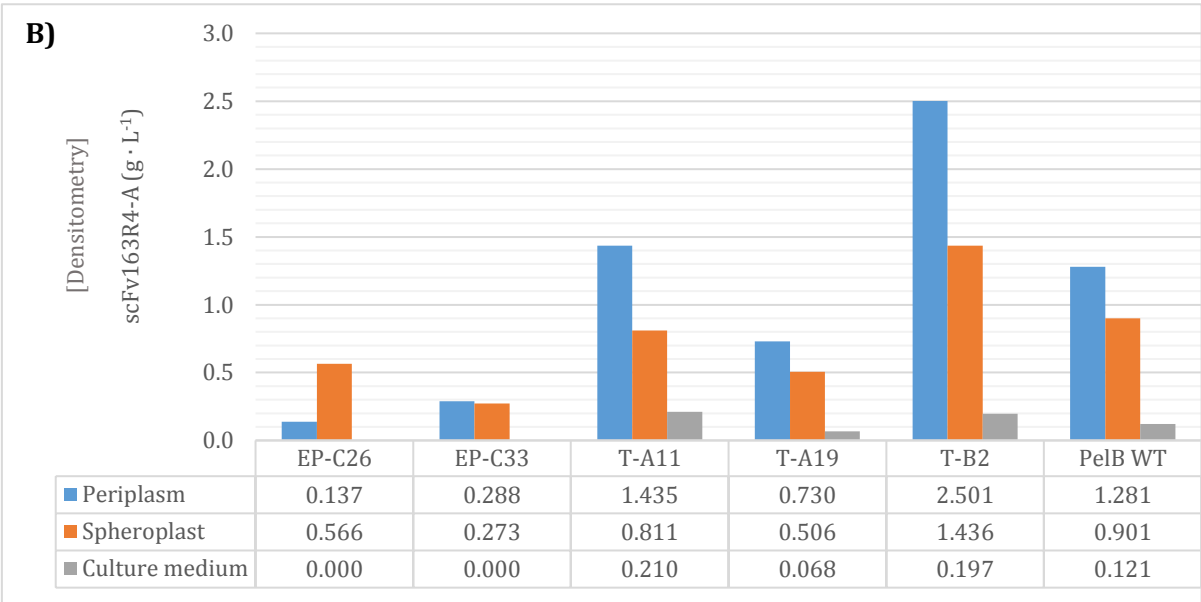
5.2.14 Evaluation of the activity of scFv163R4-A by "sandwich" ELISA

The primary function of an antibody fragment is the recognition of the target molecule that it was designed to bind, and the scFv163R4-A was selected for its ability to recognise and bind the product of the *lacZ* gene, the protein β-galactosidase (Martineau *et al.*, 1998). A "sandwich" ELISA was developed to evaluate the binding properties of the scFv163R4-A produced by fed-batch fermentation. This assay was used an indication of activity, and does not provide information with regards to the folding or the formation of disulphide bonds. The cell extracts, corresponding to the periplasmic, spheroplast and culture medium protein fractions, obtained from fed-batch fermentation experiments described in **section 5.2.12**, were evaluated for activity with the aim to compare the quantity of scFv163R4-A and its activity (**Figure 5.24 A**). The concentration of scFv163R4-A in each protein fraction was also quantified by densitometry from SDS-PAGE gels (**Figure 5.24 B**). The purified scFv163R4-A (**Figure 5.23**) was used as a reference material. Nevertheless, it is important to highlight that densitometry data could result in an overestimation of the protein yields and is shown here to aid the comparison between total concentration of scFv163R-A produced and the concentration of active scFv163R4-A.

As expected, low levels of active scFv163R4-A were detected in samples obtained from fed-batch fermentations carrying the signal peptide corresponding to clones EP-C26 and EP-C33 (**Figure 5.24 A**). The scFv163R4-A was mainly detected in the spheroplasts (**Figure 5.24 B**), correlating with results previously obtained by SDS-PAGE and western blot (**Figures 5.20 to 5.22**). A low concentration of scFv163R4-A in the spheroplast protein fraction was able to bind the β-galactosidase, 0.17 and 0.04 $g \cdot L^{-1}$, when compared with the estimated concentration of scFv163R4-A quantified by densitometry, 0.57 and 0.27 $g \cdot L^{-1}$, corresponding to clones EP-C26 and EP-C33 respectively (**Figure 5.24**). The total yield of active scFv163R4-A produced by fed-batch fermentation of clones EP-C26 and EP-C33 was 0.244 $g \cdot L^{-1}$ and 0.193 $g \cdot L^{-1}$, respectively (**Figure 5.24 A**).



	EP-C26	EP-C33	T-A11	T-A19	T-B2	PelB WT
Periplasm	0.068	0.146	0.645	0.327	0.654	0.466
Spheroplast	0.173	0.043	0.036	0.028	0.032	0.026
Culture medium	0.003	0.004	0.149	0.056	0.172	0.071



	EP-C26	EP-C33	T-A11	T-A19	T-B2	PelB WT
Periplasm	0.137	0.288	1.435	0.730	2.501	1.281
Spheroplast	0.566	0.273	0.811	0.506	1.436	0.901
Culture medium	0.000	0.000	0.210	0.068	0.197	0.121

Figure 5.24. Quantification of scFv163R4-A and detection of anti-β-galactosidase activity of scFv163R4-A. The binding activity of scFv163R4-A present in the periplasmic, spheroplast and culture medium protein fractions, produced by fed-batch fermentation from clones obtained from the error-prone PCR and Twist signal peptide libraries. The binding activity was detected by ELISA using an anti-myc antibody HRP-conjugated (A). Estimation of the concentration of scFv163R4-A in the periplasmic, spheroplast and culture medium protein samples by densitometry (B). Data are shown are mean values from ELISA experiments and single values for the quantification of scFv163R4-A, error bars are ±1 standard deviation.

Clones T-A11, T-A19, T-B2 and PelB WT showed a higher binding activity of the protein, mainly detected in samples corresponding to the periplasmic protein fraction or the culture medium (**Figure 5.24 A**). Results obtained from the ELISA showed that clones T-A11 and T-B2 produced high concentrations of active scFv163R4-A, with 0.645 and 0.654 g · L^{-1} corresponding to periplasmic fraction, and 0.149 and 0.172 g · L^{-1} corresponding to the culture medium. Samples obtained from clone T-A19 showed a lower scFv163R4-A binding activity, similar to the culture carrying the PelB WT, and the total yield of active scFv163R4-A produced for clones T-A19 and PelB WT was 0.41 g · L^{-1} and 0.56 g · L^{-1}, respectively (**Figure 5.24 A**). In addition, the concentration of scFv163R4-A in the spheroplast did not correlate with high levels of active protein, as the absence of disulphide bonds could have a negative effect on activity and stability of the scFv163R4-A. This could also explain the low concentration of active scFv163R4-A in samples obtained from clones EP-C26 and EP-C33, as the protein was mainly accumulated as a precursor in the cytoplasmic space.

In general, results obtained from the ELISA binding assay are approximated, helping to quantify the scFv163R4-A produced with antigen recognition and binding properties. The highest yield of active scFv163R4-A was found to be produced by clones T-A11 and T-B2, which yielded almost 1 g · L^{-1} of active scFv163R4-A, mainly detected in the periplasmic protein fraction and culture medium. The enhanced production of mature and active scFv163R4-A of clones T-A11 and T-B2 when compared to PelB wild type signal peptide confirms the success in the development of the β-lactamase screening and in the optimisation of the fermentation conditions for the production of the scFv163R4-A.

5.3 Discussion

In *E. coli*, more than 90% of the proteins translocated across the cytoplasmic membrane are targeted to the SecYEG translocon, either via the SecB-dependent pathway or the SRP-mediated

translocation pathways. The majority of translocated proteins can be found in the periplasmic space or associated with the inner or outer membranes, and only a small number of proteins, mostly virulence factors, are secreted to the culture medium (Georgiou and Segatori, 2005). For recombinant protein production, the production of the proteins targeting the periplasm or the culture medium is often desirable, since it allows the generation of post-translational modifications, such as disulphide bonds, which are essential for the stability of antibody fragments.

With the aim to develop a screening assay, β-lactamase was used as a reporter assay to facilitate the screening of signal peptides generated by random mutagenesis. In order to evaluate the possibilities of this screening assay, a model protein, the scFv163R4-A, was introduced between the N-terminal signal peptide and C-terminal marker protein, β-lactamase (**Figure 5.1**). The resistance to β-lactam antibiotics was used a selectable phenotype for periplasmic protein production, allowing the high-throughput screening of signal sequences obtained from mutagenized signal peptide libraries, by directly screening for antibiotic resistance levels of *E. coli* colonies (Mansell *et al.,* 2010).

5.3.1 High-throughput screening of clones by evaluation of β-lactamase activity

The selection of β-lactamase as a reporter protein was based on the essential requirement of this protein to be localised in the periplasmic space to confer protection against β-lactam antibiotics (McCann *et al*., 2007). Therefore, the fusion of a given protein to β-lactamase could be used to identify which signal peptides allowed the translocation of the recombinant protein to the periplasmic space by the selection of β-lactam-resistant colonies, as opposed to a more time-consuming screening approach. This approach has been previously exploited in other bacteria, such as *Mycobacterium tuberculosis*, where the TEM 1-β-lactamase was used as a model protein for protein export. McCann (2007), showed how a truncated version of β-lactamase, lacking the

signal peptide, yielded no colonies on agar containing carbenicillin, confirming that periplasmic localisation of β-lactamase was essential to confer antibiotic resistance. Nevertheless, cytoplasmic β-lactamase can show a certain degree of enzymatic activity (**Chapter 4, Figure 4.21**), but it does not confer resistance to β-lactam antibiotics, as observed by MIC experiments, where cultures carrying a vector coding for β-lactamase or the scFv163R4-A::β-lactamase without the signal peptide, were unable to grow at the lowest concentration of ampicillin evaluated, 3 $\mu g \cdot mL^{-1}$ (**Chapter 4, Figures 4.19 and 4.20**).

It has been reported that the selection of colonies expressing a protein of interest fused to β-lactamase based on its resistance to ampicillin could be used as an indication of the translocation of the fusion protein to the periplasmic space (Mansell *et al.,* 2010). Therefore, once the mutagenized signal peptide libraries were introduced in the expression vector (**Figure 5.3**), ampicillin selection was used to favour the enrichment and selection of only those colonies able to produce the scFv163R4-A::β-lactamase in the periplasmic space . In addition, the availability of enzymatic assays, as the nitrocefin assay, allowed the screening of colonies expressing the scFv163R4-A::β-lactamase based on the quantification of the different levels of β-lactamase activity detected in clones containing different mutant signal peptides (**Figures 5.5 and 5.6**).

A high variability in β-lactamase activity was observed between clones, possibly as the result of different levels of translocation favoured by each mutant signal peptide (**Figures 5.5 and 5.6**), since these differences in β-lactamase activity were not observed in different clones carrying the PelB wild type signal peptide (**Figure 5.7**). Unfortunately, cytoplasmic expression of the scFv163R4-A::β-lactamase could also result in β-lactamase activity, since the reporter protein exhibits activity when localised in the cytoplasmic space (**Chapter 4, Figure 4.21**). Cell lysis induced during the resuspension of cells in the assay buffer could also release scFv163R4-A::β-lactamase produced in the cytoplasm, resulting in a false positive result.

5.3.2 Evaluation of selected signal peptides for periplasmic protein production

The structure of the signal peptide plays an important role in the secretion of proteins to the periplasmic space, and extensive studies have been carried out to understand its structure. There are three main distinct regions in the structure of a signal peptide: the N-terminal region constituted by at least two positively charged amino acids, such as lysine and arginine; the central hydrophobic core composed by highly uncharged hydrophobic residues, such as leucine or alanine; and the hydrophilic cleavage region formed by amino acid residues with neutral small side-chains at position −1 and −3, often preceded by an helix-breaker amino acid residues, such as proline or a glycine (Jonet *et al.,* 2012; Izard and Kendall, 1994).

The presence of highly hydrophobic amino acids is considered the signature of signal peptides and the disruption of this region with charged residues may prevent protein secretion (Izard and Kendall, 1994). Rusch *et al.* (1994) showed how the increase in the hydrophobicity of the signal peptide favoured a higher translocation and processing rate of alkaline phosphatase (PhoA), by replacing the core region of the PhoA signal peptide with different combinations of leucine and alanine residues. However, the use of highly hydrophobic core sequences, constituted only by bulky hydrophobic residues as leucine, could compete and inhibit the transport of other host proteins (Rusch *et al.,* 1994). A similar result was observed in clone T-D4, with a GRAVY score of 1.73 (**Figure 5.14**), this highly hydrophobic signal peptide showed the highest β-lactamase activity detected during the β-lactamase screening (**Figure 5.6**). However, impaired growth and low cell viability was observed in cultures grown under induced conditions during shake-flask experiments (**Figure 5.10**). This could be the result of the high affinity of this signal peptide for the Sec translocation machinery outcompeting host proteins required for cell growth (Rusch *et al.,* 1994). In contrast, when Rusch *et al.* (1994) replaced the core region of the PhoA signal peptide solely with alanine residues, generating signal peptides with low average hydrophobicity, resulted in slow or no processing of into the mature form of alkaline phosphatase. It was therefore not surprising that the two signal peptides with the lowest average hydrophobicity (**Figure 5.13**),

clones EP-C26 and EP-C33, showed none or very slow processing of the scFv163R4-A, being accumulated as a precursor protein in the spheroplasts (**Figure 5.19**).

The generation of mutant signal peptides allowed the evaluation of the impact of different mutations in the translocation efficiency of recombinant proteins. Nevertheless, there is a requirement of a balance of the physicochemical properties in combination of the chosen secretory protein for the design and selection of the optimal signal peptide, as the negative effects of the replacement of a feature may be alleviated by the improvement of another (Izard and Kendall, 1994). This effect could mask the effect of each independent mutation, being difficult to understand the relationship between specific mutations within the signal peptide sequence and its effect on growth, cell physiology and protein production.

5.3.3 Development of fermentation strategies for periplasmic protein production

Nowadays, the fed-batch fermentation technique dominates the industry for the production of recombinant proteins. The development of fermentation protocols to achieve an optimal growth profile that favours high production rates is often desired during high cell density fermentation. Induction at lower OD_{600} is often preferable, as it has been associated with high protein production and cell viability through the course of the fermentation (Sevastsyanovich *et al.,* 2009; Wyre and Overton, 2014). However, during the production of the fusion scFv163R4-A::β-lactamase, growth arrest and low cell culturability was observed for all the clones carrying a signal peptide which allowed the translocation of the protein to the periplasmic space when induced at low cell densities (**Figure 5.15**). Cultures showed signs of growth arrest early after induction, even under "stress minimisation conditions", such as low concentration of inducer, 0.02% arabinose, in combination with a low cultivation temperature, 25 °C.

Many RPP protocols often separate the biomass generation and the recombinant protein stages by inducing protein production at high cell density, which is often carried out when the induction

of protein production could induce stress in the host organism. The separation of the biomass generation and protein production stages often helps to reduce the metabolic burden caused by the high cellular requirements demanded for both biomass and protein production. The length of the induction period, commonly known as production window, is determined by the amount of time that the cells can produce the recombinant protein without a negative impact on cell viability (Overton and Wyre, 2014).

Induction at high-cell-densities, at an OD_{600} ≥ 70, was evaluated for the production of the scFv163R4-A::β-lactamase. The separation of the biomass and protein production stages allowed also the use of different cultivation temperatures during each stage; cultures were grown at 30 °C to an OD_{600} ≥ 70, and upon induction, the temperature was reduced to 25 °C. This approach favoured the rapid growth under non-induced conditions and the decrease in the temperature during protein production, with the aim to decrease the protein synthesis rate and favour the translocation of the protein to the periplasmic space. Following this protocol, all the cultures reached a high cell density, with a final OD_{600} ≥ 140 after 28 hours of growth (**Figure 5.16**). Unfortunately, the yield of antibody fragment was lower than expected, but showing a high accumulation of the scFv163R4-A::β-lactamase in the periplasm for clones T-A11,T-A19 and T-B2 (**Figure 5.17**).

With the aim to understand whether the behaviour observed from the selected mutagenic signal peptides and evaluate if the production scFv163R4-A::β-lactamase could be used as an indication of the production of the scFv163R4-A, the sequence coding for the reporter protein, β-lactamase, was removed from the expression vector. Cultures expressing the scFv163R4-A reached a high-cell-density, with an OD_{600} ranging from 155 to 186 after 30 hours of growth (**Figure 5.18 A**), with a minimal loss of cell viability after protein production induction (**Figure 5.18 B**). In addition, the production window was extended to 12 hours during this process, considering the maximum induction time period before the culture showed signs of growth arrest, loss of cell viability and

cell lysis. Unfortunately, low levels of production of the scFv163R4-A were observed (**Figure 5.19**), mainly accumulated as mature scFv163R4-A in the periplasm, with the exception of samples obtained from clones EP-C26 and EP-C33, where the protein was primarily accumulated as a precursor in the spheroplast protein fraction. In general, no major discrepancies were obtained in terms of growth profile, cell culturability and protein production were observed during fed-batch fermentations carried out for the production of the scFv163R4-A::β-lactamase (**Figures 5.16 and 5.17**) and the scFv163R4-A (**Figures 5.18 and 5.19**). Therefore, the fusion of recombinant proteins to the reporter protein β-lactamase, could be used for high-throughput screening of signal peptides for the production of recombinant proteins in the periplasmic space.

As large differences had been previously observed in the production levels of the scFv163R4-A::β-lactamase between shake-flask and fed-batch fermentation experiments (**Table 5.1**), it was considered that a concentration of arabinose of 0.02% could be not sufficient to induce the production of the scFv163R4-A at high cell densities. This effect was also observed by Wilms *et al.* (2001) that described how a 20-fold increase in the concentration of rhamnose was required for the induction of the production of carbomylase at high cell densities, considering that the amount of rhamnose present in the culture medium could be limiting for the induction of *rhaBAD* promoter. As the concentration of arabinose used could be not enough for the induction of the *araBAD* promoter when induced at an $OD_{600} \geq 70$, a 10-fold increase in the concentration of arabinose was evaluated for the production of the scFv163R4-A. Cultures reached a high cell density, with an $OD_{600} \geq 140$ after 30 hours of growth (**Figure 5.20 A**), and the concentration of the scFv163R4-A produced was higher (**Figure 5.21**) than in previous fermentations induced with 0.02% arabinose (**Figure 5.19**). The mature form of the scFv163R4-A was detected in the periplasmic fraction of clones T-A11, T-A19, T-B2 and the PelB wild type (**Figure 5.22**). Only the accumulation of the precursor form of the scFv163R4-A was detected in samples obtained from EP-C26 and EP-C33. In addition, a high concentration of antibody fragment was found in the culture medium of samples obtained from T-A11, T-A19, T-B2 and PelB wild type (**Figure 5.21**).

Since *E. coli* naturally does not secrete proteins into the culture medium, extracellular secretion could be correlated with non-specific leakage and cell lysis, resulting from the pressure build-up in the periplasmic space (Jonet *et al.,* 2012). The capacity of the periplasm is limited, and it has been reported leakage of Fab fragments when its accumulation exceeds 6% of the volume of the periplasm (Newton *et al.,* 2016). In addition, scanning electron microscopy (SEM) obtained by Newton and collaborators (2016) during the production of antibody fragments by fed-batch fermentation, showed mostly large intact cells in pre-induction samples, which turned into a mixture of swollen cells, empty shells of lysed cells and some still healthy cells at late post-induction stage.

The activity of the scFv163R4-A produced by fed-batch fermentation was confirmed by ELISA, aiming to evaluate the antigen-binding activity of the scFv163R4-A produced by the different signal peptides selected from the β-lactamase screening. In general, high levels of activity were detected in samples obtained from clones T-A11, T-A19 and T-B2 (**Figure 5.24**), mainly corresponding to the scFv163R4-A in the periplasm and the culture medium protein samples. The bulk of the mature scFv163R-A was localised in these two protein fractions, correlating with the results observed by SDS-PAGE and western blot (**Figures 5.21 and 5.22**). The highest yields of active scFv163R4-A were obtained with the signal peptide obtained from clones T-A11 and T-B2, with a yield of almost 1 $g \cdot L^{-1}$ of active scFv163R4-A produced by fermentation, which outperformed when compared with the control culture carrying the PelB signal peptide.

5.4 Conclusions and future work

The development of a screening assay using β-lactamase as a reporter protein was proven to be an excellent tool, to allow the rapid selection of signal peptides obtained from large signal peptide libraries. Signal peptide libraries were obtained by random mutagenesis of the PelB signal peptide, a commonly used signal peptide for the secretion of antibody fragments. Mutant signal

peptides were easily and rapidly screened by their ability to confer resistance to ampicillin, being the level of production of the scFv163R4-A::β-lactamase relatively quantified by the β-lactamase enzymatic assay. Results obtained from the screening assay showed that signal peptides can be improved by the introduction of mutations in the sequence and optimised for the efficient production and translocation of recombinant protein into *E. coli*'s periplasmic space. The genetic approach used to develop the signal peptide libraries could be applied to any other wild type signal peptide sequence, improving the translocation efficiency of signal peptides already available in nature (Heggeset *et al.*, 2013).

Chapter 6:

Conclusions

6.1 Conclusions

6.1.1 Development of approaches for cytoplasmic recombinant protein production

The optimisation of the fermentation conditions for the production of certain recombinant protein therapeutics remains a challenge for the biotechnology industry. Strong promoters to achieve high expression and protein production levels are often used during the development of processes for RPP, however, the underlying effects on the host physiology are often unseen. This approach frequently fails, as plasmid loss, accumulation of misfolded products and even cell lysis, often leads to the concomitant reduction in the final protein yield. Good growth and cell viability profiles are as important as high protein expression; one must not underestimate how these factors affect RPP for the development of a successful production process.

"Stress-minimisation" conditions aim to reduce the metabolic burden imposed on the host organism during RPP, maximise plasmid retention and reduce protein synthesis rate to enhance protein folding and minimise inclusion body formation. Several fermentation conditions, including culture medium, inducer concentration, induction point and cultivation temperature were evaluated by shake-flask experiments, resulting in the accumulation of a soluble and active rhTNFα. A fully cGMP compliant fermentation process was developed, which achieved a final OD_{600} of 111 and yielded 5.35 $g \cdot L^{-1}$ of rhTNFα, of which 70% was accumulated as a soluble product.

The evaluation of the cultivation conditions was carried following a trial-and-error approach, a labour-intensive process which only allows the screening of small number of variables (Holmes et *al.,* 2009). In addition, the one-factor-at-a-time method used for the optimisation of the fermentation conditions for the production of rhTNFα, only allowed the screening of one single variable at the time, assuming all variables are independent. This simplistic approach does not allow the identification of interactions between different variables, and often fails to identify the true optimal conditions for RPP (Islam *et al.*, 2007). An alternative approach could be the use of

Design of Experiments (DoE), where statistical and modelling tools are used to explore the optimal production conditions with the minimum number of experiments, helping to more rapidly determine the relationship between multiple screened variables for maximum soluble protein yield (Hsu *et al.,* 2016). The implementation of DoE approaches with high-throughput fermentation equipment, such as the Ambr250, could facilitate the development of fermentation processes.

6.1.2 Development of approaches for periplasmic recombinant protein production

Results obtained using scFv163R4-A as a model protein showed the importance of balancing the protein synthesis rate with the translocation machinery capability. Several approaches were evaluated, such as the modulation of activity of the arabinose promoter by catabolite repression or the optimisation of the cultivation conditions, to achieve the fine-tuning between protein synthesis and periplasmic translocation. Also, different signal peptides were evaluated including STII, PelB, Bla (SecB-dependent pathway) and DsbA (SRP-mediated pathway). Even though all of them resulted in the production of periplasmic scFv163R4-A, great differences were observed in terms of growth, cell viability and protein production kinetics. Signal peptides allowing a high and rapid production and translocation of scFv163R4-A often resulted in the overload of translocation pathways, leading to growth arrest and cell lysis. Signal peptides which allowed a slower production showed no detrimental effect on the host organism, which could potentially result in higher yields at fermentation scale.

The use of β-lactamase protein fusions helped to elucidate the subcellular localisation of the scFv163R4-A, as results obtained from SDS-PAGE and western blot were often inconclusive. The use of β-lactamase as a reporter protein was found to be key during the screening of mutant signal peptide libraries, helping to reduce the number of clones selected with none or poor translocation efficiency.

6.1.3 β-lactamase screening assay as a high-throughput screening tool for periplasmic recombinant protein production

The work presented in this thesis included: (1) the generation of two mutant signal peptide libraries, one generated by error-prone PCR and the other chemically synthesised by Twist Biosciences, both using the PelB signal peptide as template. (2) A cloning method based on type IIs restriction endonucleases developed to allow seamless cloning of the signal peptide libraries. (3) A screening method based on the use of β-lactamase as a reporter protein, for the selection of clones carrying mutant signal peptides with enhanced translocation activity of the model protein. (4) The selection of mutant signal peptides with higher expression than the wild type signal peptide. (5) The optimisation of the fermentation conditions for the production of the model protein, scFv163R4-A, in the periplasm. The enhanced production of mature and active scFv163R4-A of clones T-A11 and T-B2, when compared to PelB wild type signal peptide, confirmed the success in the development of the β-lactamase screening. Both clones yielded almost 1 g · L^{-1} of mature and active scFv163R4-A, achieving a 1.5-fold increase over the wild type. Overall, the β-lactamase screening assay was proven to be a promising tool for the selection of the optimal signal peptide for the production of recombinant proteins in the periplasm.

Limitations of β-lactamase screening assay

Results obtained from the β-lactamase screening showed promising results, however, two mutant signal peptides with no or reduced production of the protein in periplasm were selected during the screening, EP-C26 and EP-C33. Firstly, only a minimal amount of periplasmic β-lactamase is required to allow the growth in ampicillin-supplemented plates, which resulted in the overgrowth of these two clones. Secondly, it is unclear if nitrocefin could be able to diffuse through the cytoplasmic membrane, resulting in the misleading detection of high levels of cytoplasmic scFv163R4-A::β-lactamase, as periplasmic β-lactamase activity. In addition, cell lysis caused

during the cell resuspension could lead to the release of cytoplasmic fusion protein. To avoid this outcome, the methodology employed during the development of the β-lactamase screening should be reviewed and further optimised. Clone selection needs to be carried out under stringent conditions, to favour the selection of only those clones able to grown at highest ampicillin concentrations, and therefore, with enhanced translocation of the fusion protein to the periplasm. In addition, the β-lactamase screening is limited in the number of clones that can be evaluated (100 clones/week). A higher throughput could be achieved by using a fluorescent β-lactamase substrate in combination with fluorescent-activated cell sorting (FACS), potentially resulting in rapid screening and selection of a larger number of clones.

6.2 Future work

6.2.1 Improvement of arabinose-inducible expression systems

The tight repression of arabinose inducible expression systems has shown to be critical, allowing cell growth to high cell densities with minimal protein expression under non-induced conditions. However, further development is required to achieve a dose-dependent induction, in contrast with the "all-or-nothing" induction effect observed with the arabinose expression system. This can be achieved by engineering of host strain targeting the arabinose transport system, resulting in a more homogenous cell induction if the arabinose uptake is catalysed by arabinose independent transporters from a constitutive promoter or an arabinose-independent inducible promoter (Khlebnikov *et al.*, 2000).

6.2.2 Optimisation of the fermentation conditions for RPP

Flow cytometry enables single cell analysis allowing the study of bacterial populations during high-cell-density fermentation, being an essential tool for bioprocessing monitoring. A few

fluorescent dyes are available to evaluate cell physiological state, including the evaluation of membrane polarisation and permeability, enzyme and efflux pump activity (Silva *et al.,* 2012). Two dyes are routinely used for the assessment of cell viability during fermentation, bis-oxonol (BOX) and propidium iodide (PI), revealing the proportion of healthy, damaged and dead cells (Hewitt and Nebe-Von-Caron, 2001), valuable information for the determination of the optimal conditions for the production of recombinant proteins, such induction point and harvest time.

As certain periplasmic extraction methods, such as osmotic shock, are not practical at large scale, other methods such as periplasmic protein release methods derived from a temperature increase (45 °C to 95 °C) for a variable length of time (12 hours to seconds) are the most credible alternative at industrial scale (Weir and Bailey, 1997). However, the release of the periplasmic content is highly variable and dependent on the cell metabolic state. Flow cytometry could be used to evaluate damaged cells with the depolarised cytoplasmic membrane, to determine the optimal conditions for the release of periplasmic proteins to the culture medium.

6.2.3 Optimisation of β-lactamase screening assay for periplasmic protein production

Initial experiments carried out with the STII, DsbA, PelB and Bla signal peptides offered an insight of which signal peptide could be the optimal for the production of the scFv163R4-A. This screening resulted in the selection of PelB, which allowed a lower production of scFv163R4-A when compared with other signal peptides, but without entailing a negative effect on cell growth and cell viability. This was an essential requirement for the development of a fermentation production process, being the signal peptide used as a template for the generation of mutant signal peptides, aiming to select mutant versions showing an enhancement in the production of periplasmic scFv163R4-A.

Nevertheless, the preliminary evaluation was limited to three SEC signal peptides and one SRP signal peptide. A larger number of signal peptides targeting both pathways needs to be evaluated

to expand our knowledge regarding periplasmic protein production. In addition, PelB may not be the optimal signal peptide for the production of other recombinant proteins, as proteins which fold too quickly in the cytoplasm may fail to be translocated by the SEC pathway. A different signal peptide library could be generated with a second signal peptide targeting the SRP-mediated translocation pathway, which could expand the capabilities of the β-lactamase screening assay to a broader range of recombinant proteins.

In general terms, the β-lactamase screening assay showed promising results for the screening of a large number of signal peptide libraries in a rapid and cost-effective manner. Nevertheless, results obtained from the β-lactamase screening assay were limited to one single model protein, being an essential requirement to validate the screening assay with a broad range of different model proteins.

Future work could also include the generation of a signal peptide database containing all the mutations introduced in the signal peptide sequence and their effect on growth and cell viability, as well as its impact on protein productivity and translocation to the periplasmic space. This would help to expand the knowledge regarding the structure of the signal peptide.

Development of a "toolbox" for the production of disulphide bonded recombinant proteins

Results included in this PhD thesis could be directly applied to the production of disulphide bonded protein by the biotechnology industry. A suggested approach for the production and optimisation of production conditions for a target protein is detailed in **Figure 6.1**. Briefly, this approach includes an initial screening using wild type signal peptides targeting both the SecB-dependent and the SRP-mediated pathway, and if this approach fails, the evaluation using β-lactamase screening assay. Mutant signal peptides could potentially include, a mutant library

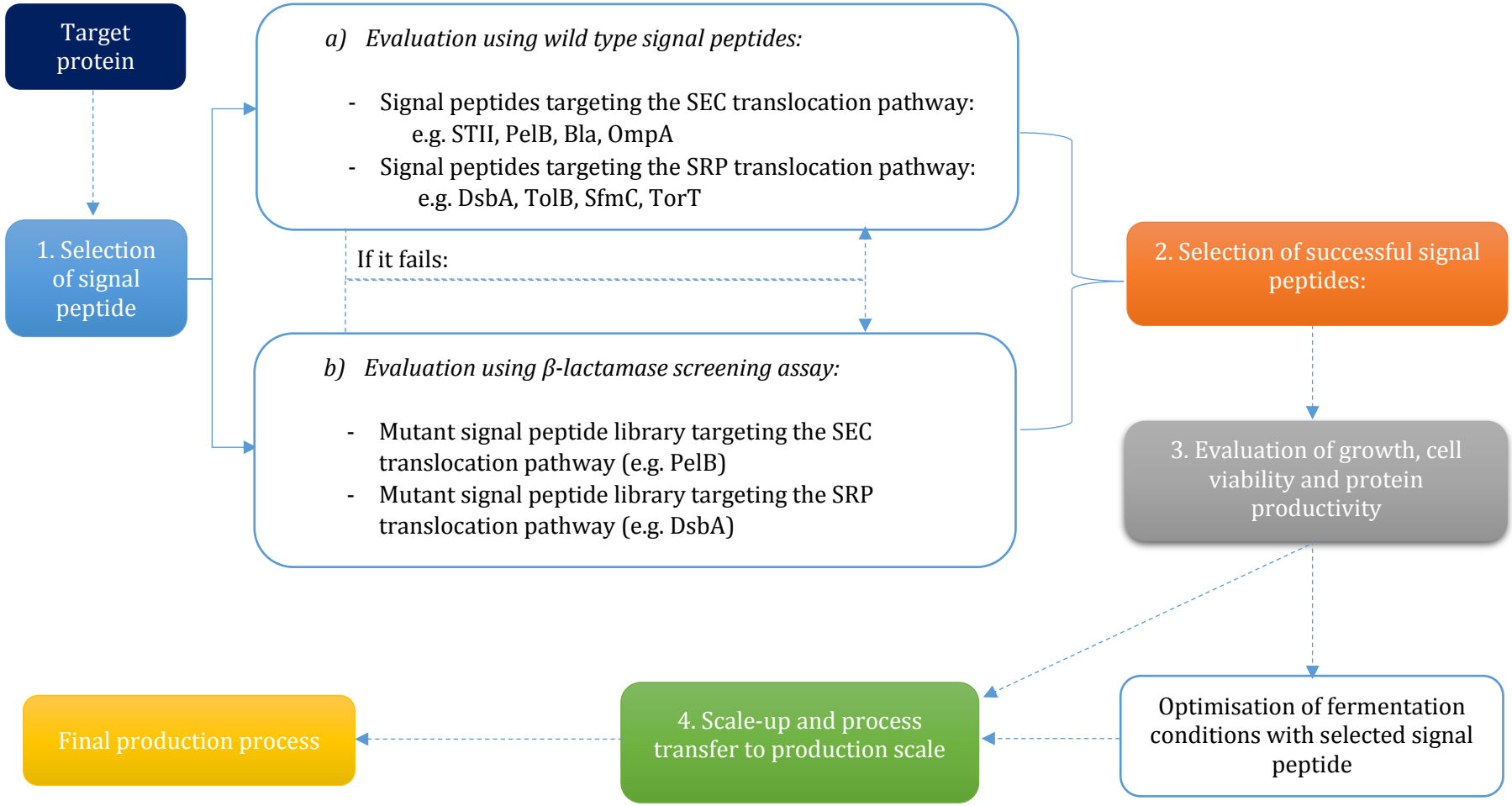


Figure 6.1. Suggested approach for the production of disulphide bonded proteins and its application for the biotechnology industry.

targeting the SEC pathway and a second library targeting the SRP-mediated pathway. Successful signal peptides would be evaluated for growth, cell viability and protein production. The production conditions using the most promising signal peptide could be optimised using a DoE approach combined with a screening fermentation system, such as the Ambr250. Once the fermentation process has been optimised, it would be transferred to a larger scale, for the development of the final production process. This suggested approach helps to illustrate possible industrial applications of the work developed in this PhD thesis, and how they could be applied to the production of recombinant proteins for the biotechnology industry.

References

Aggarwal, S.R. (2014) What's fueling the biotech engine — 2012 to 2013. **Nature biotechnology,** 32 (1): 32-39.

Aidelberg, G., Towbin, B.D., Rothschild, D., *et al.* (2014) Hierarchy of non-glucose sugars in *Escherichia coli*. **BMC systems biology,** 8 (1): 1.

Albiniak, A.M., Matos, C.F., Branston, S.D., *et al.* (2013) High-level secretion of a recombinant protein to the culture medium with a *Bacillus subtilis* twin-arginine translocation system in *Escherichia coli*. **FEBS journal,** 280 (16): 3810-3821.

Alfasi, S.N. (2011) Physiological aspects underpinning recombinant protein production in *Escherichia coli.* PhD Thesis. University of Birmingham.

Andrews, J.M. (2001) Determination of minimum inhibitory concentrations. **The Journal of antimicrobial chemotherapy,** 48 Suppl 1 5-16.

Angus, B.L., Carey, A.M., Caron, D.A., *et al.* (1982) Outer membrane permeability in *Pseudomonas aeruginosa*: comparison of a wild-type with an antibiotic-supersusceptible mutant. **Antimicrobial Agents and Chemotherapy,** 21 (2): 299-309.

Babaeipour, V., Shojaosadati, S., Robatjazi, S., et al. (2007) Over-production of human interferon-γ by HCDC of recombinant *Escherichia coli*. **Process Biochemistry,** 42 (1): 112-117.

Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli*. **Nature biotechnology,** 22 (11): 1399-1408.

Bebrone, C., Moali, C., Mahy, F., *et al.* (2001) CENTA as a chromogenic substrate for studying beta-lactamases. **Antimicrobial Agents and Chemotherapy,** 45 (6): 1868-1871.

Bibi, E. (2007) "Co- and posttranslational protein targeting to the SecYEG translocon in *Escherichia coli*" In Ehrmann, M. (ed.) **The periplasm.** United States: American Society for Microbiology. pp. 3.

Brandenberger, R., Burger, S., Campbell, A., *et al.* (2011) Cell therapy bioprocessing. **BioProcess Int,** 9 (Suppl 1): 30-37.

Broome-Smith, J., Tadayyon, M. and Zhang, Y. (1990) β-lactamase as a probe of membrane protein assembly and protein export. **Molecular microbiology,** 4 (10): 1637-1644.

Brosius, J., Erfle, M. and Storella, J. (1985) Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. **The Journal of biological chemistry,** 260 (6): 3539-3541.

Buchholz, K. (2007) Science–or not? The status and dynamics of biotechnology. **Biotechnology journal,** 2 (9): 1154-1168.

Buchholz, K. and Collins, J. (2013) The roots—a short history of industrial microbiology and biotechnology. **Applied Microbiology and Biotechnology,** 97 (9): 3747-3762.

Carter, P.J. (2011) Introduction to current and future protein therapeutics: a protein engineering perspective. **Experimental cell research,** 317 (9): 1261-1269.

Choi, J.H., Keum, K.C. and Lee, S.Y. (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. **Chemical engineering science,** 61 (3): 876-885.

Choi, J., Jeong, K., Kim, S., *et al.* (2000) Efficient secretory production of alkaline phosphatase by high cell density culture of recombinant *Escherichia coli* using the Bacillus sp. endoxylanase signal sequence. **Applied Microbiology and Biotechnology,** 53 (6): 640-645.

Choi, J. and Lee, S. (2004) Secretory and extracellular production of recombinant proteins using Escherichia coli. **Applied Microbiology and Biotechnology,** 64 (5): 625-635.

Chou, C.P. (2007) Engineering cell physiology to enhance recombinant protein production in Escherichia coli. **Applied Microbiology and Biotechnology,** 76 (3): 521-532.

Chusacultanachai, S. and Yuthavong, Y. (2004) Random mutagenesis strategies for construction of large and diverse clone libraries of mutated DNA fragments. **Parasite Genomics Protocols,** 319-333.

Danese, P.N. and Silhavy, T.J. (1998) Targeting and assembly of periplasmic and outer-membrane proteins in Escherichia coli. **Annual Review of Genetics,** 32 (1): 59-94.

Demain, A.L. and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. **Biotechnology Advances,** 27 (3): 297-306.

Egan, S.M. and Schleif, R.F. (1993) A regulatory cascade in the induction of rhaBAD. **Journal of Molecular Biology,** 234 (1): 87-98.

Elvin, C.M., Thompson, P.R., Argall, M.E., *et al.* (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. **Gene,** 87 (1): 123-126.

Enfors, S., Jahic, M., Rozkov, A., *et al.* (2001) Physiological responses to mixing in large scale bioreactors. **Journal of Biotechnology,** 85 (2): 175-185.

Fahner, B. Lilie, H. and Neubauer, P. (2004) "Inclusion bodies: formation and utilisation" In Scheper, T. and Enfors, S.O. (eds.) **Physiological stress responses in bioprocess** 1st ed. Berlin, Heidelberg: Springer. pp. 93.

Feilmeier, B.J., Iseminger, G., Schroeder, D., *et al.* (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. **Journal of Bacteriology,** 182 (14): 4068-4076.

Fernandez, J.M. and Hoeffler, J.P. (1999) **Gene Expression Systems: Using Nature for the Art of Expression.** United States: Academic Press.

Fu, Z., Ng, K., Lam, T., *et al.* (2005) Cell death caused by hyper-expression of a secretory exoglucanase in Escherichia coli. **Protein expression and purification,** 42 (1): 67-77.

Gao, W., Xing, B., Tsien, R.Y., *et al.* (2003) Novel fluorogenic substrates for imaging β-lactamase gene expression. **Journal of the American Chemical Society,** 125 (37): 11146-11147.

Garcia-Fruitos, E., Gonzalez-Montalban, N., Morell, M., *et al.* (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. **Microbial Cell Factories,** 4 (1): 1.

Gasser, B., Saloheimo, M., Rinas, U., *et al.* (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. **Microbial cell factories,** 7 (1): 11.

Georgiou, G. and Segatori, L. (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. **Current opinion in biotechnology,** 16 (5): 538-545.

Georgiou, G., Telford, J.N., Shuler, M.L., *et al.* (1986) Localization of inclusion bodies in Escherichia coli overproducing beta-lactamase or alkaline phosphatase. **Applied and Environmental Microbiology,** 52 (5): 1157-1161.

Giepmans, B.N., Adams, S.R., Ellisman, M.H., *et al.* (2006) The fluorescent toolbox for assessing protein location and function. **Science (New York, N.Y.),** 312 (5771): 217-224.

Gleave, M.E. and Monia, B.P. (2005) Antisense therapy for cancer. **Nature Reviews Cancer,** 5 (6): 468-479.

Glick, B.R. (1995) Metabolic load and heterologous gene expression. **Biotechnology Advances,** 13 (2): 247-261.

Gottesman, S. (1996) Proteases and their targets in *Escherichia coli*. **Annual Review of Genetics,** 30 (1): 465-506.

Gronenborn, B. (1976) Overproduction of phage lambda repressor under control of the lac promotor of *Escherichia coli.* **Molecular and General Genetics MGG,** 148 (3): 243-250.

Gründemann, D. and Schömig, E. (1996) Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. **BioTechniques,** 21 (5): 898-903.

Guzman, L.M., Belin, D., Carson, M.J., *et al.* (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. **Journal of Bacteriology,** 177 (14): 4121-4130.

Hagemeyer, C.E., von Zur Muhlen, C., von Elverfeldt, D., *et al.* (2009) Single-chain antibodies as diagnostic tools and therapeutic agents. **Thromb Haemost,** 101 (6): 1012-1019.

Haldimann, A., Daniels, L.L. and Wanner, B.L. (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 (5): 1277-1286.

Hanson-Manful, P. and Patrick, W.M. (2013) Construction and analysis of randomized protein-encoding libraries using error-prone PCR. **Protein Nanotechnology: Protocols, Instrumentation, and Applications, Second Edition,** 251-267.

Heggeset, T.M., Kucharova, V., Naerdal, I., *et al.* (2013) Combinatorial mutagenesis and selection of improved signal sequences and their application for high-level production of translocated heterologous proteins in Escherichia coli. **Applied and Environmental Microbiology,** 79 (2): 559-568.

Heras, B., Shouldice, S.R., Totsika, M., *et al.* (2009) DSB proteins and bacterial pathogenicity. **Nature Reviews Microbiology,** 7 (3): 215-225.

Hewitt, C.J. and Nebe-Von-Caron, G. (2001) An industrial application of multiparameter flow cytometry: assessment of cell physiological state and its application to the study of microbial fermentations. **Cytometry,** 44 (3): 179-187.

Hodgson, I.J., Lennon, C.D.J. and Kara, V.L. (2011) **Expression system.** EP2695943 (A2) edn, Europe.

Hoffmann, F. and Rinas, U. (2004) "Stress induced by recombinant protein production in *E. coli*" In Sheper, T. and Enfors, S.O. (eds.) **Physiological stress responses in bioprocess** 1st ed. Berlin, Heidelberg: Springer. pp. 73.

Holmes, W.J., Darby, R.A., Wilks, M.D., *et al.* (2009) Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime. **Microbial cell factories,** 8 (1): 1.

Hsu, C., Thomas, O.R. and Overton, T.W. (2016) Periplasmic expression in and release of Fab fragments from *Escherichia coli* using stress minimization. **Journal of Chemical Technology and Biotechnology,** 91 (3): 815-822.

Huang, C., Lin, H. and Yang, X. (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. **Journal of industrial microbiology & biotechnology,** 39 (3): 383-399.

Humphreys, D.P. (2007) "Periplasmic expression of antibody fragments" In Ehrmann, M. (ed.) **The periplasm.** United States: American Society for Microbiology. pp. 361.

Islam, R., Tisi, D., Levy, M., *et al.* (2007) Framework for the rapid optimization of soluble protein expression in Escherichia coli combining microscale experiments and statistical experimental design. **Biotechnology progress,** 23 (4): 785-793.

Izard, J.W., Doughty, M.B. and Kendall, D.A. (1995) Physical and conformational properties of synthetic idealized signal sequences parallel their biological function. **Biochemistry,** 34 (31): 9904-9912.

Izard, J.W. and Kendall, D.A. (1994) Signal peptides: exquisitely designed transport promoters. **Molecular microbiology,** 13 (5): 765-773.

Jonasson, P., Liljeqvist, S., Nygren, P., *et al.* (2002) Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. **Biotechnology and applied biochemistry,** 35 (2): 91-105.

Jonet, M.A., Mahadi, N.M., Murad, A.M., *et al.* (2012) Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in Escherichia coli. **Journal of Molecular Microbiology and Biotechnology,** 22 (1): 48-58.

Kadonaga, J.T., Gautier, A.E., Straus, D.R., *et al.* (1984) The role of the beta-lactamase signal sequence in the secretion of proteins by Escherichia coli. **The Journal of biological chemistry,** 259 (4): 2149-2154.

Kay, M.A. (2011) State-of-the-art gene-based therapies: the road ahead. **Nature Reviews Genetics,** 12 (5): 316-328.

Khlebnikov, A., Risa, O., Skaug, T., *et al.* (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. **Journal of Bacteriology,** 182 (24): 7029-7034.

Kosinski, M.J., Rinas, U. and Bailey, J.E. (1992) Isopropyl-β-D-thiogalactopyranoside influences the metabolism of *Escherichia coli*. **Applied Microbiology and Biotechnology,** 36 (6): 782-784.

Krzeslak, J., Braun, P., Voulhoux, R., *et al.* (2009) Heterologous production of *Escherichia coli* penicillin G acylase in *Pseudomonas aeruginosa*. **Journal of Biotechnology,** 142 (3): 250-258.

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

Lee, Y. and Jung, K. (2007) Modulation of the tendency towards inclusion body formation of recombinant protein by the addition of glucose in the araBAD promoter system of *Escherichia coli*. **Journal of microbiology and biotechnology,** 17 (11): 1898.

Lee, Y.J. and Jeong, K.J. (2013) Enhanced production of antibody fragment via SRP pathway engineering in *Escherichia coli*. **Biotechnology and bioprocess engineering,** 18 (4): 751-758.

Lodish, H. Berk, A. Matsudaira, P. *et al.* (2003) **Molecular cell biology.** 5th ed. Freeman.

Low, K.O., Mahadi, N.M. and Illias, R.M. (2013) Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. **Applied Microbiology and Biotechnology,** 97 (9): 3811-3826.

Lu, Q. (2005) Seamless cloning and gene fusion. **Trends in biotechnology,** 23 (4): 199-207.

Luirink, J. and Sinning, I. (2004) SRP-mediated protein targeting: structure and function revisited. **Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research,** 1694 (1): 17-35.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. **Microbiological reviews,** 60 (3): 512-538.

Manoil, C., Mekalanos, J.J. and Beckwith, J. (1990) Alkaline phosphatase fusions: sensors of subcellular location. **Journal of Bacteriology,** 172 (2): 515-518.

Mansell, T.J., Linderman, S.W., Fisher, A.C., *et al.* (2010) A rapid protein folding assay for the bacterial periplasm. **Protein Science,** 19 (5): 1079-1090.

Martineau, P., Jones, P. and Winter, G. (1998) Expression of an antibody fragment at high levels in the bacterial cytoplasm. **Journal of Molecular Biology,** 280 (1): 117-127.

McCann, J.R., McDonough, J.A., Pavelka, M.S., *et al.* (2007) β-Lactamase can function as a reporter of bacterial protein export during Mycobacterium tuberculosis infection of host cells. **Microbiology,** 153 (10): 3350-3359.

McGee, Z.A. and Clemens, C.M. (1994) Effect of bacterial products on tumor necrosis factor production: quantitation in biological fluids or tissues. **Methods in enzymology,** 236 23-31.

Meager, A., Leung, H. and Woolley, J. (1989) Assays for tumour necrosis factor and related cytokines. **Journal of immunological methods,** 116 (1): 1-17.

Mergulhao, F., Summers, D. and Monteiro, G. (2005) Recombinant protein secretion in Escherichia coli. **Biotechnology Advances,** 23 (3): 177-202.

Meyer, H. and Schmidhalter, D.R. (2014) "The History and Economic Relevance of Industrial Scale Suspension Culture of Living Cells" In **Industrial Scale Suspension Culture of Living Cells.** Wiley-VCH Verlag GmbH & Co. KGaA. pp. 1-38.

Miroux, B. and Walker, J.E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. **Journal of Molecular Biology,** 260 (3): 289-298.

Miyake, T., Oka, T., Nishizawa, T., *et al.* (1985) Secretion of human interferon-alpha induced by using secretion vectors containing a promoter and signal sequence of alkaline phosphatase gene of *Escherichia coli*. **Journal of Biochemistry,** 97 (5): 1429-1436.

Naylor, L.H. (1999) Reporter gene technology: the future looks bright. **Biochemical pharmacology,** 58 (5): 749-757.

Nelson, A.L. and Reichert, J.M. (2009) Development trends for therapeutic antibody fragments. **Nature biotechnology,** 27 (4): 331-337.

Newton, J.M., Schofield, D., Vlahopoulou, J., *et al.* (2016) Detecting cell lysis using viscosity monitoring in *E. coli* fermentation to prevent product loss. **Biotechnology progress,** 32 (4): 1069-1076.

Ni, Y. and Chen, R. (2009) Extracellular recombinant protein production from *Escherichia coli*. **Biotechnology Letters,** 31 (11): 1661.

Norrander, J., Kempe, T. and Messing, J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. **Gene,** 26 (1): 101-106.

O'Callaghan, C.H., Morris, A., Kirby, S.M., *et al.* (1972) Novel method for detection of beta-lactamases by using a chromogenic cephalosporin substrate. **Antimicrobial Agents and Chemotherapy,** 1 (4): 283-288.

Oliver, D. (1985) Protein secretion in *Escherichia coli.* **Annual Reviews in Microbiology,** 39 (1): 615-648.

Oshima, T. and Tanaka, S. and Matsukura, S. (1989) **Expression vector for human TNF.** US4871663 edn, United States.

Overton, T.W. (2014) Recombinant protein production in bacterial hosts. **Drug discovery today,** 19 (5): 590-601.

Palmer, T. and Berks, B.C. (2012) The twin-arginine translocation (Tat) protein export pathway. **Nature Reviews Microbiology,** 10 (7): 483-496.

Palomares, L.A., Estrada-Moncada, S. and Ramírez, O.T. (2004) Production of recombinant proteins. **Recombinant Gene Expression: Reviews and Protocols,** 15-51.

Papanikou, E., Karamanou, S. and Economou, A. (2007) Bacterial protein secretion through the translocase nanomachine. **Nature Reviews Microbiology,** 5 (11): 839-851.

Passarinha, L., Bonifacio, M. and Queiroz, J. (2009) Application of a fed-batch bioprocess for the heterologous production of hSCOMT in Escherichia coli. **Journal of microbiology and biotechnology,** 19 (9): 972-981.

Patel, R., Smith, S.M. and Robinson, C. (2014) Protein transport by the bacterial Tat pathway. **Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,** 1843 (8): 1620-1628.

Petsch, D. and Anspach, F.B. (2000) Endotoxin removal from protein solutions. **Journal of Biotechnology,** 76 (2): 97-119.

Pharmaceutical Research and Manufacturers of America (PhRMA) (2013) **The biopharmaceutical pipeline: evolving science, hope for patients**: United States: PhRMA's Communications & Public Affairs Department.

PharmaCompass (2016) **Top drugs by sales revenue in 2015: Who sold the biggest blockbuster drugs?** [Online]. Available from: http://www.pharmacompass.com/radio-compass-blog/top-drugs-by-sales-revenue-in-2015-who-sold-the-biggest-blockbuster-drugs [Accessed 12/26 2016].

Plückthun, A., Krebber, A., Krebber, C., *et al.* (1996) "Antibody Engineering: Producing antibodies in *Escherichia coli*: From PCR to fermentation" In McCafferty, J., Hoogenboom, H. R., and Chiswell, D. J. (eds) 1st ed. Oxford: IRL Press. Pp. 203.

Plückthun, A. (1991) Strategies for the expression of antibody fragments in *Escherichia coli*. **Methods,** 2 (2): 88-96.

Rodríguez-Carmona, E., Cano-Garrido, O., Dragosits, M., *et al.* (2012) Recombinant Fab expression and secretion in *Escherichia coli* continuous culture at medium cell densities: Influence of temperature. **Process Biochemistry,** 47 (3): 446-452.

Rosenberg, A.H., Lade, B.N., Dao-shan, C., *et al.* (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. **Gene,** 56 (1): 125-135.

Rusch, S.L., Chen, H., Izard, J.W., *et al.* (1994) Signal peptide hydrophobicity is finely tailored for function. **Journal of cellular biochemistry,** 55 (2): 209-217.

Sahdev, S., Khattar, S.K. and Saini, K.S. (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. **Molecular and cellular biochemistry,** 307 (1-2): 249-264.

Samuelson, J.C. (2011) "Recent developments in difficult protein expression: a guide to *E. coli* strains, promoters, and relevant host mutations" In Evans, T.C. and Xu, M. (eds) **Heterologous Gene Expression in *E. coli*. Methods in Molecular Biology 705.** 1st ed. Humana Press: Springer Science+Business Media. pp. 195.

Steiner, D., Forrer, P., Stumpp, M.T., *et al.* (2006) Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. **Nature biotechnology,** 24 (7): 823-831.

Schmidt, F. (2004) Recombinant expression systems in the pharmaceutical industry. **Applied Microbiology and Biotechnology,** 65 (4): 363-372.

Schumann, W. and Ferreira, L.C.S. (2004) Production of recombinant proteins in *Escherichia coli*. **Genetics and Molecular Biology,** 27 (3): 442-453.

Sevastsyanovich, Y., Alfasi, S., Overton, T., *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.

Sevastsyanovich, Y.R., Alfasi, S.N. and Cole, J.A. (2010) Sense and nonsense from a systems biology approach to microbial recombinant protein production. **Biotechnology and applied biochemistry,** 55 (1): 9-28.

Silva, F., Queiroz, J.A. and Domingues, F.C. (2012) Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. **Biotechnology advances,** 30 (3): 691-708.

Skerra, A. (1994) Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli.* **Gene,** 151 (1): 131-135.

Sørensen, H.P. and Mortensen, K.K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. **Journal of Biotechnology,** 115 (2): 113-128.

Spadiut, O., Capone, S., Krainer, F., *et al.* (2014) Microbials for the production of monoclonal antibodies and antibody fragments. **Trends in biotechnology,** 32 (1): 54-60.

Studier, F.W., Davanloo, P., Rosenberg, A.H., *et al.* (1990) **Cloning and expression of the gene for bacteriophage T7 RNA polymerase,** US4952496 edn, United States.

Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. **Journal of Molecular Biology,** 189 (1): 113-130.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. **Applied Microbiology and Biotechnology,** 72 (2): 211-222.

Valdez-Cruz, N.A., Caspeta, L., Pérez, N.O., *et al.* (2010) Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters. **Microbial cell factories,** 9 (1): 1.

Valgepea, K., Adamberg, K., Seiman, A., *et al.* (2013) Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. **Molecular BioSystems,** 9 (9): 2344-2358.

Vizcaino Caston, I. (2012) Monitoring bacterial physiology during recombinant protein production using reporter gene technology. PhD Thesis. Uiversity of Birmingham.

Vizcaino-Caston, I., Wyre, C. and Overton, T.W. (2012) Fluorescent proteins in microbial biotechnology—new proteins and new applications. **Biotechnology Letters,** 34 (2): 175-186.

Walsh, G. (2014) Biopharmaceutical benchmarks 2014. **Nature biotechnology,** 32 (10): 992-1000.

Walsh, G. (2010) Biopharmaceutical benchmarks 2010. **Nature biotechnology,** 28 (9): 917-924.

Walsh, G. (2005) Biopharmaceuticals: recent approvals and likely directions. **Trends in biotechnology,** 23 (11): 553-558.

Walsh, G. (2006) Biopharmaceutical benchmarks 2006. **Nature biotechnology,** 24 (7): 769-U5.

Want, A., Thomas, O.R., Kara, B., *et al.* (2009) Studies related to antibody fragment (Fab) production in *Escherichia coli* W3110 fed-batch fermentation processes using multiparameter flow cytometry. **Cytometry Part A,** 75 (2): 148-154.

Wei, H., Hu, J., Wang, L., *et al.* (2012) Rapid gene splicing and multi-sited mutagenesis by one-step overlap extension polymerase chain reaction. **Analytical Biochemistry,** 429 (1): 76-78.

Weiner, L.M., Surana, R. and Wang, S. (2010) Monoclonal antibodies: versatile platforms for cancer immunotherapy. **Nature Reviews Immunology,** 10 (5): 317-327.

Westers, L., Westers, H. and Quax, W.J. (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. **Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,** 1694 (1): 299-310.

Weir, A.N.W. and Bailey, N.A (1997) **Process for obtaining antibodies utilizing heat treatment.** US5665866 (A) edn. United States.

Wilms, B., Hauck, A., Reuss, M., *et al.* (2001) High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the *Escherichia coli* rhaBAD promoter. **Biotechnology and bioengineering,** 73 (2): 95-103.

Wurm, F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. **Nature biotechnology,** 22 (11): 1393-1398.

Wycuff, D.R. and Matthews, K.S. (2000) Generation of an AraC-araBAD promoter-regulated T7 expression system. **Analytical Biochemistry,** 277 (1): 67-73.

Wyre, C. J. (2015). Recombinant protein production in *Escherichia coli*: optimisation of improved protocols. PhD Thesis. University of Birmingham.

Wyre, C. and Overton, T.W. (2014) Use of a stress-minimisation paradigm in high cell density fed-batch *Escherichia coli* fermentations to optimise recombinant protein production. **Journal of industrial microbiology & biotechnology,** 41 (9): 1391-1404.

Xing, B., Khanamiryan, A. and Rao, J. (2005) Cell-permeable near-infrared fluorogenic substrates for imaging β-lactamase activity. **Journal of the American Chemical Society,** 127 (12): 4158-4159.

Yin, H., Kanasty, R.L., Eltoukhy, A.A., *et al.* (2014) Non-viral vectors for gene-based therapy. **Nature Reviews Genetics,** 15 (8): 541-555.

Zhang, D., Sweredoski, M.J., Graham, R.L., *et al.* (2012) Novel proteomic tools reveal essential roles of SRP and importance of proper membrane protein biogenesis. **Molecular and cellular proteomics: MCP,** 11 (2): M111.011585.