Physiological aspects underpinning recombinant protein production in *Escherichia coli*

by

Sara Nuri Alfasi

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
The University of Birmingham
December 2010

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

ABSTRACT

Many biopharmaceutical projects require the production of recombinant protein in a bacterial host. Conventional procedures used for recombinant protein production (RPP) involve the rapid synthesis of the target protein. This results in the accumulation of unfolded protein, the induction of the heat shock stress response and bacterial growth arrest. More importantly, the target protein accumulates in inclusion bodies and hence is useless to determine its structure. The immediate impact of this is that both the yield and quality of the target protein are compromised.

This thesis reports two generically successful approaches that were developed to overcome this series of stress-induced events in *Escherichia coli*. Both strategies were developed during the production of a cytoplasmic protein and outer membrane lipoproteins using the pET expression system in the bacterial host, *E. coli* strain BL21(DE3)*. First, the induction protocol was modified to minimise the stress on the host bacterium. This method relies on the induction of very low levels of the T7 RNA polymerase in BL21* and thus the correspondingly slow synthesis of the target protein. Using this approach, growth and productivity of different types of correctly folded target proteins were sustained for at least 70 h.

Secondly, mutant hosts that significantly improve recombinant protein production during conventional protocols were isolated. These improved hosts are resistant to IPTG-induced stress and continue to accumulate high levels of the correctly folded target protein. Key to the stress resistance is the presence of mutations that downregulate the synthesis of T7 RNA polymerase. However, different improved hosts were able to enhance the production of different types of target protein, such as those requiring extensive post-translational modification. The potential for isolating a plethora of improved bacterial hosts that are tailored for the production of different types of recombinant protein is discussed in light of the challenges faced by bioindustry. Procedures enabling the isolation of mutant hosts during the production of GFP-tagged and untagged proteins are reported.

Dedicated with love and thanks

to Mum, Dad and my siblings

Acknowledgements

First and foremost, I extend sincerest thanks and gratitude to my supervisor, Jeff Cole, for all his support, help and guidance during this project. I am deeply honoured to have been given the opportunity to be part of your stellar research group, and am thoroughly grateful for all the help and constructive criticism provided in the writing of this thesis. My thanks also extend to my industrial supervisors Dr Richard Hall and Dr Jo Jones for their invaluable industrial advice and for making my work at GlaxoSmithKline a wholly enjoyable experience. Dr Yanina Sevastsyanovich deserves much appreciation for her friendship and exceptional contribution to this project. Central to acknowledgement is Lesley Griffiths, whom I thoroughly thank for her assistance with experimental techniques and particularly her unwavering mental support when the going got tough. A special thank you to Dr Tim Overton, Prof Chris Hewitt and Dr Eva Hyde for their insightful discussion of data including their assistance with flow cytometry work. To all members of the labs on the ground and first floor Biosciences, thank you for all your help, friendship and for making the last three years so memorable.

I would like to express my utmost gratitude to all staff of the 2nd floor teaching labs for their infinite assistance and boundless patience during fermentation practicals, especially in the events of accidents. To all BIO328 fermentation students, a part of whose data is included in this thesis, as well as to Ann Begum and Stephen Godfrey for their assistance with administration matters, I bestow my heartfelt thanks. Furthermore, I am deeply thankful to both the BBSRC and GSK, without whose funding, this project would not have been possible.

Last but not least, tremendous thanks go to my mum, dad and my siblings Hammam, Asia, Haret, Kawther and Mohammed for their continuous unending support and encouragement. Above all, I thank God for keeping me sane throughout this process.

Table of contents

List of figures

List of tables

Table

List of abbreviations

Cfu	Colony forming units
CCP	Cytochrome c peroxidase
DCW	Dry cell weight
EDTA	Ethylenediaminetetraacetic acid
FS	Forward scatter
GFP	Green fluorescent protein
GSK	GlaxoSmithKline
IBs	Inclusion bodies
IPTG	Isopropyl β-D-thiogalactosidase
LB	Luria Bertani
NA	Nutrient agar
OD	Optical density
PAC	Penicillin acylase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PI	Propidium iodide
PMT	Photomultiplier tube
RPP	Recombinant protein production
SDS	Sodium dodecyl sulphate
SS	Side scatter

CHAPTER 1

Introduction

Recombinant protein production (RPP) is necessary in scientific research and in the manufacture of therapeutics. The demand for recombinant protein is often very high, necessitating production of high yields of the target protein at a large scale. Occasionally, however, obtaining a high quality recombinant protein that is correctly folded becomes the overriding requirement. This is particularly important when protein structures are to be determined using NMR and X-ray crystallography and forms the basis for drug design and development. Despite that large-scale RPP has been underway for over 30 years, significant problems that are associated with it are still being investigated today. One of the main problems is centred on the conformational quality of the recombinant protein produced. This is because very often, the target protein that is accumulated is incorrectly folded, which aggregates in insoluble inclusion bodies (IBs). In addition, production of large quantities of recombinant protein causes stress, which has detrimental consequences on the physiology and metabolism of the production host. The bacterial response to this stress ultimately compromises yields of the target protein in prolonged production attempts (Sorensen and Mortensen, 2005). Strategies to improve the yields of correctly folded recombinant protein in bacteria have therefore long been the focus of many international meetings in the RPP scientific community, which have failed to define generic protocols for successful outcomes (Sevastsyanovich *et al*., 2009).

Conventional protocols for RPP

Escherichia coli remains the most used host for recombinant protein production as the target protein can account for up to 50% of the total protein content of the bacteria. Many strains are used in combination with different recombinant plasmids from which a

recombinant gene is expressed at high levels using a strong promoter (Terpe, 2007). One of the most popular is the pET expression system used in conjunction with *E. coli* BL21 (λDE3) first described by Studier and Moffatt (1986). In this system, the expression of the recombinant gene is directed from the recombinant plasmid by the T7 RNA polymerase, which is encoded on the chromosome of the BL21 host. The use of T7 RNA polymerase has the advantage of being seven times more efficient at transcription than the *E. coli* RNA polymerase and therefore guarantees the rapid transcription of the recombinant gene. Transcription of T7 RNA polymerase is induced from a *lac*UV5 promoter by the addition of IPTG. The T7 RNA polymerase then transcribes the recombinant gene from a T7 promoter (Studier and Moffatt, 1986; Studier *et al*., 1990). The induction of this system results in an amplification effect in which T7 RNA polymerase and the recombinant protein are continually synthesized and therefore ensuring the accumulation of a high yield of the recombinant protein.

The production of recombinant protein using the pET system according to the instructions of the manufacturer involves the initial growth of bacteria at 37°C and induction of expression of the recombinant gene using a high concentration of IPTG, typically 0.5 to 1 mM. The growth temperature is then decreased to 25°C in an attempt to facilitate correct folding of the recombinant protein. Indeed, a rapid burst of recombinant protein synthesis occurs immediately post-induction but its further accumulation soon stops (Jones, 2007; Dong *et al*., 2005). This protocol remains very successful for the production of high yields of the recombinant protein but is only sustainable for short-time periods. This is because the rapid induction of RPP induces the general stress response in bacteria that leads to

growth arrest (Hoffman and Rinas, 2004b; Gasser *et al*., 2008; Sevastsyanovich *et al*., 2010).

Despite their deleterious consequences to bacterial physiology, conventional protocols are still the main way by which recombinant protein is produced in *E. coli*. Regardless of the expression system used, RPP is induced at rapid rates using high inducer concentrations and high temperatures (Terpe, 2007). Not surprisingly therefore, the same problems are encountered repeatedly; the aggregation of the target protein in inclusion bodies, bacterial growth arrest and the loss of the recombinant plasmid from bacteria.

The first section of the introduction outlines the bacterial response to rapid RPP and its detrimental consequences to the host physiology. The impact of this on the conformational quality of the target protein, and in particular the problem of its aggregation in inclusion bodies, is discussed. The strategies currently available that aim to solve the problems associated with RPP, including previous attempts to isolate improved hosts, are summarised. The first objective of this project was to develop a generic strategy to improve RPP. Based on the underlying physiological responses to RPP, a generic approach that avoids the bacterial stress response and the consequent formation of inclusion bodies was required. The second objective of this study was to isolate improved host strains that are resistant to RPP stress and continue to accumulate high yields of correctly folded recombinant protein. The third objective was to assess the utility of the available improved hosts for the production of different types of recombinant protein. Finally, the potential of isolating different host strains that are 'specialised' for the accumulation of different types of recombinant protein, is discussed.

Physiological responses of bacteria during RPP

The production of recombinant protein leads to a fatal cascade of physiological responses within bacteria that undermines the prolonged accumulation of recombinant protein. Developing a correct understanding of how bacteria respond to RPP is the key to identifying factors that might potentially enhance or compromise the host's ability to accumulate recombinant protein. The cascade of physiological responses of bacteria during RPP is discussed below.

Metabolic burden and the stringent response

The requirement of large yields of recombinant protein entails the rapid expression of the cloned gene and synthesis of the protein upon induction. In cases where the pET expression system is used for RPP, this also involves the cumulative synthesis of the T7 RNA polymerase. It has been established that over-production of recombinant proteins generally induces a 'metabolic burden' on the host. This is defined as the amount of resources such as ATP and precursors that are withdrawn from the host metabolism for maintenance and expression of the foreign DNA (Bentley and Kompala, 1990). Indeed, it has been shown that the diversion of cellular resources from biosynthesis and cell division to synthesis of the RNA and protein required for RPP leads to growth arrest (Dong *et al*., 1995; Kurland and Dong, 1996; Bentley *et al*., 1990). Even when the target gene is not produced, the mere induction of the pET expression system in the presence of pET plasmid leads to cessation of bacterial growth and cell division (Andersson *et al*., 1996; Miroux and Walker, 1996; Soriano *et al*., 1999, 2002).

The metabolic burden of RPP leads to the 'stringent response' in *E. coli*. The stringent response is characterized by a rapid metabolic shift, from one that supports rapid growth to one that induces amino acid biosynthetic genes and allows for prolonged survival in stationary phase. The stringent response occurs when nutrients such as amino acids become limiting for growth and is triggered by the accumulation of the alarmone 5',3' guanosine pyrophosphate or ppGpp (Cashel *et al*., 1996). When amino acids are in short supply due to RPP, the concentration of uncharged tRNAs increases, causing the abrupt halting of translation by ribosomes. This provides a signal to the stringent factor RelA, which binds to stalled ribosomes and catalyses the synthesis of the alarmone ppGpp (Wendrich *et al*., 2002). ppGpp binds in the secondary channel close to the active site of RNA polymerase and inhibits transcription. The binding of ppGpp within the secondary channel is stabilized by the transcription factor DksA thus providing stronger inhibition of transcription (Paul *et al*., 2004; Perederina *et al*., 2004). Since around 60% of RNA polymerase molecules within growing bacterial cells are engaged in transcription of rRNA genes, the immediate consequence of ppGpp accumulation is the significant downregulation of transcription of stable RNAs, comprising ribosomal and transfer RNAs (Cashel *et al*., 1996). It is unclear how the accumulation of ppGpp precisely induces the stringent response. Many models as to how this might be achieved have been proposed. The first is that liberation of RNAP previously sequestered in stable RNA synthesis results in the increased availability for transcription of stringently induced promoters (Barker *et al*., 2001). Another proposed mechanism of action involves the direct downregulation of rRNA transcription. This is dependant on the observation that ppGpp, once bound to RNA polymerase, decreases the half-life of the open complex formation at promoters. This causes the significant downregulation of promoters with intrinsically short 'melting' half-

lives such as those required for the expression of genes encoding stable RNAs (Paul *et al.*, 2004). Since the transcription of ribosome genes is controlled by rRNA levels, the biosynthesis of ribosomes is therefore shut down (Paul *et al.*, 2004). However, recent evidence has emerged that many promoters of amino acid biosynthesis genes can be directly activated in the presence of both ppGpp and DksA (Paul *et al.*, 2005). This occurs through stimulating open complex formation and thus the rate of transcription initiation from these promoters.

The accumulation of ppGpp, and consequently the increased availability of RNA polymerase, soon results in large-scale adjustments of cellular metabolism as reported by several transcriptomics studies (Durfee *et al.*, 2008; Traxler *et al.*, 2008). This includes the upregulation of specific amino acid biosynthesis genes, as defined by limitations in the nutritional environment, and the inhibition of DNA replication and macromolecule biosynthesis such as membrane and cell wall components (Traxler *et al.*, 2008). Transcriptome analysis revealed that the stringent response involves changes in the expression of more than 30% of the bacterial genome by more than 2-fold (Traxler *et al.*, 2008). It was shown that a mutant unable to synthesise ppGpp failed to induce amino acid biosynthesis genes as part of the stringent response. This highlights the dependency of the stringent response on the availability of ppGpp. Expression of RpoS was also increased by approximately 7-fold, which induced more than 130 genes in its regulon members, resulting in the cessation of growth (Traxler *et* al., 2008). The upregulation of the *rpoH* regulon and the induction of genes involved in the heat-shock response was also evident (Durfee *et al.*, 2008).

The stringent response therefore allows bacteria to adapt their physiology in order to counteract the lethal effects of nutritional stress brought about by RPP and support prolonged survival in stationary phase. The degradation of rRNA and the destruction of ribosomes upon induction of RPP has also been reported (Dong *et al*., 1995). This paralleled a progressive loss in protein synthesis capacity, and was postulated to be part of the stringent response, perhaps as an attempt by the cells to recycle essential components and precursors required for its survival (Dong *et al*., 1995). The decreased capacity of cells to synthesise protein as part of the stringent response highlights major challenges regarding the sustainability of RPP. These challenges must be overcome if the prolonged accumulation of recombinant protein is to be maintained.

The heat shock response

The heat shock response has been reported as one of the main physiological responses that is most associated with growth arrest during rapid RPP. This was made apparent from findings of numerous transcriptomics and microarray studies. The *E. coli* heat shock response is induced when proteins become denatured at high temperature (Arsene *et al*., 2000). During growth at optimal temperatures, the low levels of the heat shock sigma factor, RpoH (also known as σ^H or σ^{32}), is bound to the heat shock chaperone proteins, DnaK, DnaJ/GrpE and GroEL/GroES, and is therefore transcriptionally inactive (Gamer *et al*., 1992; Guisbert *et al*., 2004). However, heat shock promotes the formation of denatured proteins that are bound preferentially by these chaperone complexes, so RpoH is released and activates transcription of the genes under its control (Arsene *et al*., 2000). This response is also triggered by the rapid accumulation of unfolded recombinant proteins, irrespective of whether the product is soluble or insoluble (Lesley *et al*., 2002). The

accumulation of unfolded protein intermediates depends in part on both specific properties of the target protein, and on rates of transcription and translation (reviewed by Hoffmann and Rinas, 2000; 2004).

The release of RpoH induces the synthesis of many proteins in its regulon that enhance folding of recombinant proteins and prevent their aggregation, most importantly chaperones such as GroEL, GroES and DnaK (reviewed in Baneyx and Mujacic, 2004; Young *et al.*, 2004; Hoffmann and Rinas, 2004). Proteases such as ClpP and Lon are also upregulated as part of the heat-shock response. Both of these proteases are involved in the degradation of insoluble protein, and in the case of ClpP, the soluble recombinant protein too (Gill *et al.*, 2000; Vera *et al.*, 2005). In addition, concentrations of the small heat shock proteins IbpA (inclusion body protein A) and IbpB are increased after high-level induction of heterologous protein production (Allen *et al.*, 1992; Han *et al.*, 2004). Synthesis of both these chaperones was found to be positively regulated by RpoH as part of the heat shock response (Allen *et al.*, 1992). IbpA and IbpB were found to be tightly associated with inclusion bodies indicating that they play a role in protecting the disaggregated proteins from degradation by directing folded and misfolded proteins into inclusion bodies (Allen *et al.*, 1992; Han *et al.*, 2004). Although this is disadvantageous to RPP, the aggregation of proteins into inclusion bodies might have strategically evolved to decrease stress on the host by decreasing levels of misfolded proteins within cells (Han *et al.*, 2004; Lethanh *et al.*, 2005).

In addition to the indirect depletion of availability of holo RNA plolymerase for other sigma factors, the release of RpoH also induces the transcription of the sigma factor, RpoS,

which facilitates the transition of bacteria into the stationary growth phase. A cascade of downstream responses involving the significant upregulation of the RpoH-regulated gene, *recA*, is also apparent (Gill *et al.* 2000). Apart from its involvement in critical DNA recombination and repair functions, the RecA protein is also a protease that destroys phage repressor proteins and hence induces phage-dependent lysis. It is therefore clear that the negative impact on continued protein accumulation, in addition to bacterial growth arrest, are direct consequences of a chain of reactions initiated by an excessive rate of RPP (**Figure 1.1**). This significantly limits yields of recombinant protein obtained per volume of culture and thus has huge implications in bioindustry. Therefore, in order to increase the financial viability of RPP in a commercial context, bacterial stress must be avoided.

Much confusion is centred on whether the heat-shock response is the mechanism by which cells counteract the stress caused by RPP or rather is the response by cells to help cope and adapt to RPP. It was the latter assumption that formed the basis from which strategies to improve RPP were inspired and further developed. In fact, attempts to exploit the heat shock response to improve protein folding during RPP was only recently demonstrated by exposing bacteria to a dual stress of thermal stress and the co-expression of the GroELS chaperone systems (Kim *et al.*, 2009). These measures resulted in a 2-fold increase in bacterial biomass to only 0.6 OD units (600 nm) and a slightly higher yield of recombinant protein at 50°C. However, this was not sustained for periods longer than 2 h after the dual stress. Despite such results, the authors optimistically concluded that co-expressing chaperones during thermal stress or the induction of the heat shock response in general is a successful strategy in RPP and is worth pursuing.
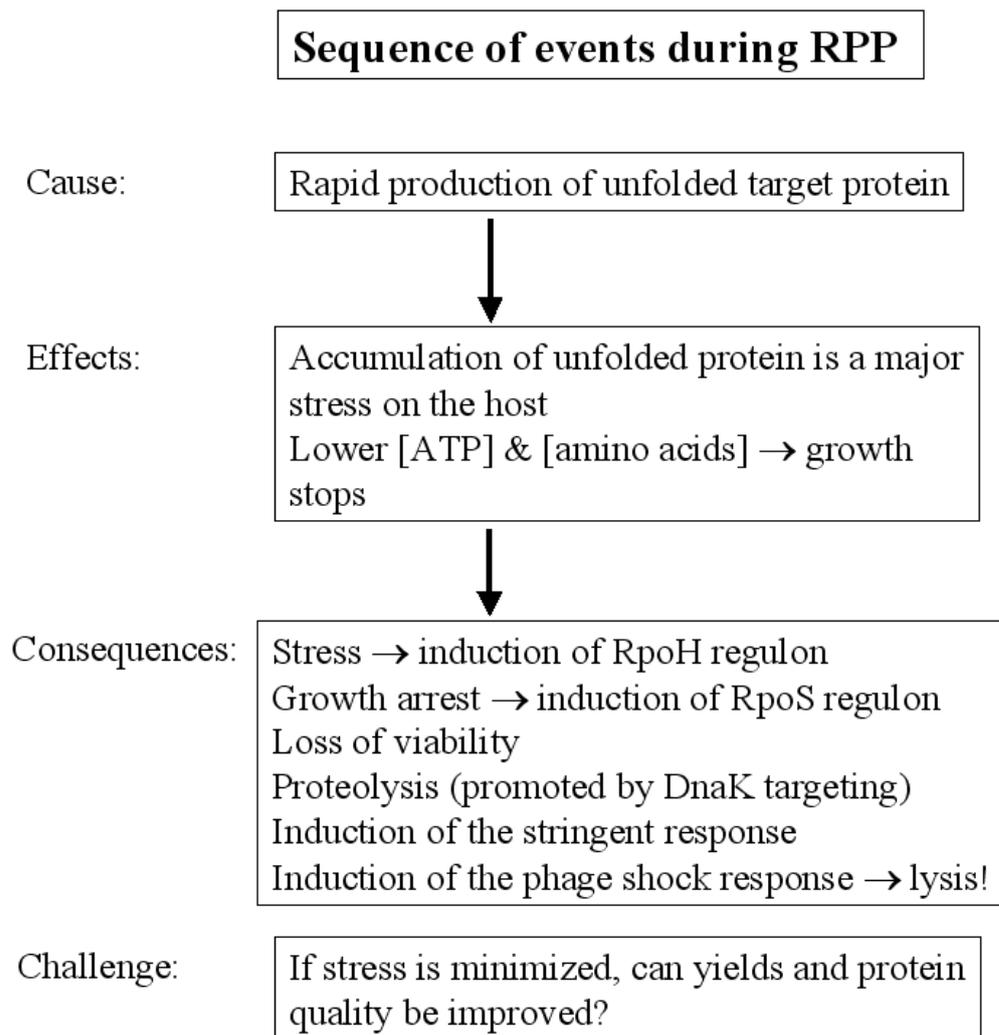
Figure 1.1: Physiological response of *E. coli* to a high rate of recombinant protein production. (*Sevastsyanovich et al., 2010*)

Transcriptomics studies of the general stress response during RPP

Understanding the physiological response of bacteria during RPP at the transcriptional level is an essential prerequisite for the development of generic strategies and mutant hosts that enhance RPP. The possibility of identifiying genes that affect RPP could also be used as a basis for the design of RPP protocols. Currently, approximately seven studies have reported attempts to analyse the transcriptional response of *E. coli* during RPP. However, in all experiments and regardless of the expression systems used, recombinant proteins were produced at rapid rates, which often included a decrease of temperature following induction amongst changes in other variables. Unsurprisingly, therefore, the findings from all these transcriptional studies were consistent. One of the most comprehensive studies was reported by Haddadin and Harcum (2005) who found that almost 23% of the 4404 genes in *E. coli* responded to RPP. Particularly notable were the upregulated regulons for RpoS, RpoH and the stringent response, which reflect stress of amino acid starvation during RPP. The very large response to RPP was due largely to the down-regulation of protein and RNA synthesis (including ribosome biosynthesis) as well as down-regulation of intermediary metabolism, energy-generating electron transfer pathways, and ATP synthesis. The main conclusion from this study was that RPP had "initiated a cascade of transcriptome responses that down-regulated the very genes needed to sustain productivity".

This was followed promptly by another transcriptomics study, designed to analyse the bacterial response during a dual stress of RPP induction by IPTG combined with thermal stress at 50°, in comparison with the corresponding individual stresses (Harcum and Haddadin, 2006). This was an attempt to identify possible genes or regulons that would be

beneficial to RPP in the dual stress. Also, the possibility of whether 'adapting' bacteria to heat shock using thermal stress could improve RPP was investigated. As expected, transcripts known to be regulated as part of the RpoH regulon or by growth rate were highly represented, but otherwise it was difficult to deduce clear patterns in the changes observed. However, on average, there was a 10-fold upregulation of 3.6% of genes in the dual stress conditions that had not previously been considered to be part of the heat shock response.

Results from transcriptomics studies were also mirrored in several other studies, all of which reported similar trends: down-regulation of intermediary metabolism and induction of the stringent and heat shock responses (Gill *et al*., 2000; Durrshmidt *et al*., 2008). The transcriptome response to the metabolic stress generated in cells over-expressing protein revealed at least two main classes of up-regulated transcripts. First, proteases such as DegP and Lon that degrade aggregated proteins in the periplasm and cytoplasm, respectively, were significantly upregulated. Secondly, chaperones, such as FtsH and ClpB that also degrade proteins involved in both prophage lambda regulation and the heat shock response, were induced (Gill *et al*., 2000).

The transcriptome of bacteria during accumulation of soluble or insoluble protein was also analysed. The aim of this was to identify genes that are differentially expressed in relation to the solubility of the target protein accumulated. It was apparent that genes of the TCA cycle and *ibpB* were upregulated during expression of exclusively insoluble protein, but there was no consistency with the expression of other genes (Jurgen *et al*., 2000). This was followed by a study that aimed to compare the responses to over-expression of soluble and

insoluble proteins (Smith, 2007). There was a difference of 55 genes that were upregulated during the accumulation of insoluble protein, 34 of which were components of the RpoH regulon. The transcriptome response was then compared with that during production of soluble protein. However, attempts to produce soluble protein, by altering the cultivation conditions such as temperature and media composition, failed completely and yielded a majority of insoluble protein. Despite this failure, the bacterial transcriptome response was analysed, which unsurprisingly produced very similar results, prompting the author to conclude that the upregulation of sigma 32 target genes represents a generalised cellular response to protein insolubility. This was followed by the optimistic suggestion that the "manipulation of the σ^{32} regulon might provide a general mechanism for improving protein solubility". The experimental design used to investigate the transcriptional response of bacteria during RPP highlights the fundamental question of whether the products of the RpoH regulon have been evolved to promote the synthesis of foreign proteins in a well adapted host, or whether they are simply the mechanism by which bacteria counteract the stress imposed by RPP.

The accumulation of recombinant protein using conventional protocols

Inclusion bodies (IBs) are intracellular deposits of aggregated polypeptides that usually occur upon the rapid production of heterologous protein. It has long been regarded that IBs were formed with unfolded or misfolded polypeptides that were devoid of any biological activity (Carrió and Villaverde, 2001). Hence, the occurrence of IBs was deemed problematic in a biotechnological context, which results in major economic impacts particularly when protein refolding from inclusion bodies becomes necessary (Carrió and

Villaverde, 2001; Yoon *et al*., 2010). This organized cell packaging of IBs has been regarded as a protective mechanism by cells in which harmful interactions between soluble protein and hydrophobic patches from misfolded protein is prevented (González-Montalbán *et al*., 2005, 2007).

The formation of IBs is usually the consequence of the rapid accumulation of newly synthesized polypeptides (Hoffman and Rinas, 2004; Baneyx and Mujacic, 2004). This occurs when rates of protein synthesis overwhelm the folding machinery or the cell's capacity for post-translational modification. The aggregation of unfolded polypeptides is facilitated by the presence of exposed hydrophobic patches, resulting in IBs either in the cytoplasm or periplasm, depending on the nature of the target protein produced (Baneyx and Mujacic, 2004; Arié *et al*., 2006).

The dynamics of inclusion body formation

The formation of inclusion bodies is a dynamic process resulting from an unbalanced equilibrium between the aggregation and disaggregation of polypeptides (Carrió *et al*., 1999; Carrió and Villaverde, 2001). Growth of inclusion bodies occurs during the rapid accumulation of unfolded polypeptides, which are prone to aggregation. This has been demonstrated with the fusion protein VP1LAC and a homotrimeric P22 TSP that undergoes a complex folding pathway. However, in the absence of *de novo* protein synthesis, for example when active protein production was stopped after the addition of chloramphenicol to bacteria, the disintegration of IBs was evident (Carrió and Villaverde, 2001). Concentrations of the fusion protein in the soluble fraction increased, which also coincided with a corresponding decrease in IBs (Carrió *et al*., 1999; Carrió and Villaverde, 2001).

This dynamic process was distinguished from indiscriminate proteolysis by the observation that full length target protein was liberated from IBs and was detected in the soluble fraction before degradation fragments were detected. Thus, the dissolution of the target protein from IBs precedes its degradation, after which the resulting fragments can aggregate back into the IBs.

A different view that IBs are transient deposits that act as reservoirs for polypeptide chains that are able to fold correctly was therefore adopted. IBs are deeply integrated in the quality control mechanism when levels of newly synthesised polypeptides overwhelm the chaperone folding pathway. Therefore polypeptides remain clustered until they can be released to re-enter the folding pathways.

Role of DnaK and GroEL in inclusion body formation

The extent of chaperone participation in inclusion body formation has been continuously investigated by Villaverde and co-workers. The effect of protein aggregation into IBs was explored in *E. coli* deficient in either DnaK or GroEL (Carrió and Villaverde, 2003; González-Montalbán *et al*., 2005). This was investigated with production of an aggregation-prone fusion protein VP1LAC in which between 50% and 100% of the protein had aggregated into IBs. The absence of functional DnaK resulted in around 2.5-fold more aggregated protein, leading to formation of much larger IBs. This was inevitable since the deficiency of DnaK leads to increased levels of unfolded polypeptides that will direct the equilibrium towards aggregation into IBs. The increased size of IBs was also attributed to the loss of the dissolution activity of IBs by DnaK. DnaK, in cooperation with ClpB, is efficient at the removal of protein aggregates from IBs (Mogk *et al*., 2003) (**Figure 1.2**).
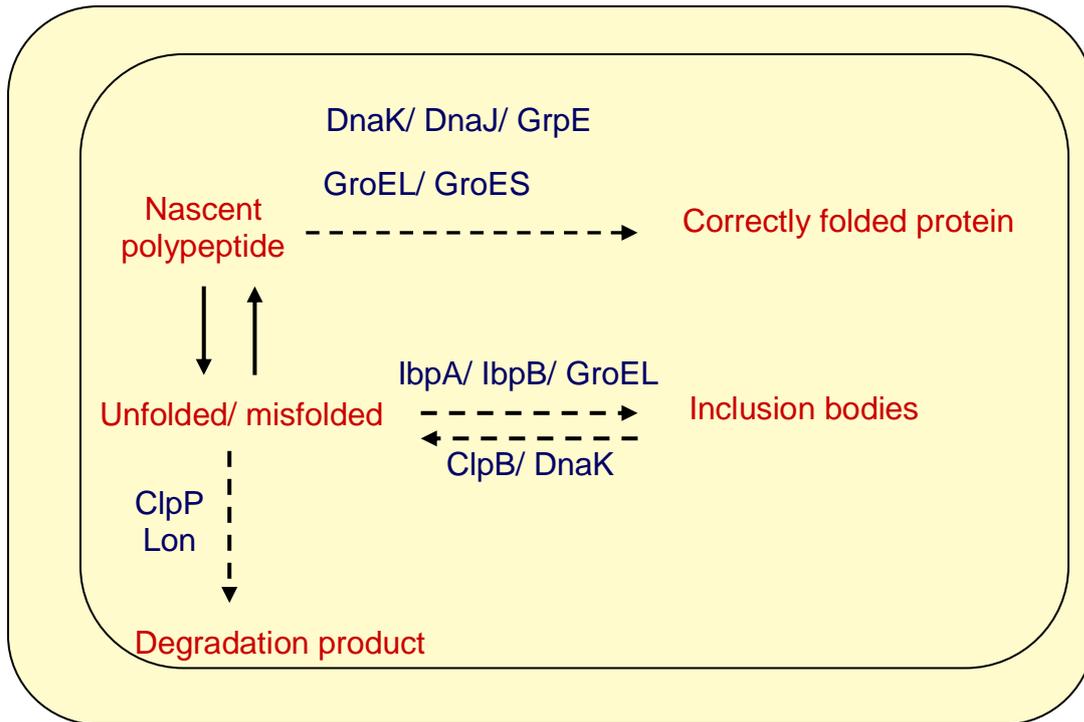
Figure 1.2: The protein folding pathway of recombinant protein in *E. coli*. The newly synthesized nascent polypeptide is folded with the aid of DnaK and GroEL/ES chaperone systems, producing correctly folded protein. However, unfolded or misfolded polypeptide is aggregated and directed into inclusion bodies by IbpA and IbpB. The dissolution of these unfolded intermediates from inclusion bodies is mediated by the chaperone ClpB (and DnaK), and their folding is attempted by the DnaK and GroEL chaperone systems. Unfolded protein intermediates that remain in the cytoplasm are degraded by the proteases ClpP and Lon, to avoid induction of the heat-shock response in the cytoplasm of *E. coli*.

Thus the loss of DnaK will inevitably result in the impaired removal of aggregates from IBs (Carrió and Villaverde, 2003). The significant consequences brought about by DnaK deficiency strongly imply that the conformational surveillance activity of DnaK cannot be completely complemented by other chaperones. This could be because the sets of substrates recognised by DnaK and other possible chaperones differ, or that other chaperones are inefficient folding modulators in comparison to DnaK.

The deficiency of functional GroEL in *E. coli* during RPP yielded a more surprising result. There was significantly less aggregation of the target protein but rather an increased occurrence of soluble protein by around 30% (Carrió and Villaverde, 2003; González-Montalbán *et al*., 2005). There were small and numerous deposits of aggregated protein within cells that were unlike the IBs obtained in the parental strain. This led to the suggestion that GroEL plays an important role in the nucleation of IBs or as a positive modulator of protein aggregation.

Evidence of the roles of DnaK and GroEL with regards to inclusion body formation has been supported by the location of these chaperones and their association with IBs (Carrió and Villaverde, 2005). DnaK has been detected only on the surface of IB, which reflects its function of IB dissolution, whereas GroEL has been found to be embedded within IBs, implying that it promotes the aggregation of IBs (Carrió and Villaverde, 2005). Thus, the divergent roles of DnaK and GroEL reflect that these chaperones are major antagonist controllers of inclusion body formation by preventing and promoting, respectively, the aggregation of misfolded protein (Carrió and Villaverde, 2003).

Effect of inclusion body formation on bacterial growth

The occurrence of inclusion bodies was reported to have cytotoxic effects in *E. coli* lacking functional DnaK and GroEL (González-Montalbán *et al*., 2005). Cell viability, in other words their ability to form colonies, had decreased by 30% and 50% in a GroEL and a DnaK mutant, respectively, 5 h after induction of VP1LAC synthesis, compared to uninduced cultures. This protein consists of an N-terminal fusion containing the VP1 capsid protein of the foot and mouth disease virus, which is prone to misfolding and aggregation into inclusion bodies. The effect of VP1LAC production on bacteria was also compared to that during the production of β-galactosidase alone that had not formed IBs despite being acummulated at a similar rate. It was therefore concluded that the growth arrest of bacteria is the result of a cytoplasmic stress response that is triggered by the accumulation of insoluble recombinant protein. Specifically, the toxicity of IBs is conferred by the misfolded protein on their surface, which could result in unfavourable interactions with cellular proteins (González-Montalbán *et al*., 2005; 2007a). Indeed, the surface area of the inclusion body was shown to correlate positively with growth arrest. Further investigation has demonstrated that the soluble recombinant protein produced in a DnaK mutant is much less stable and more prone to misfolding compared to that in the wild type strain (García-Fruitós *et al*., 2005a). Similar cytotoxic effects were reported during the production of a misfolding variant of MalE, called MalE31, that aggregated in IBs in the periplasm when produced at high levels at either 30°C or 37°C (Hunke and Betton, 2003). Production of MalE31 at 37°C resulted in the growth arrest of bacteria after only 3 h post-induction, as reflected by the loss of colony forming ability on non-selective nutrient agar by 2 orders of magnitude. When the periplasmic protease DegP was co-expressed, bacterial growth was not arrested. This coincided with a significant decrease in

MalE31 inclusion bodies. The physiological consequence for cells producing MalE31 was that expression of the periplasmic protease *degP* was activated, reflecting an extracytoplasmic stress response. However, it was shown that the production of a destabilized, misfolding form of MalE that did not aggregate also induced the Cpx stress response (Hunke and Betton, 2003). This provided a clear indication that it is the accumulation of misfolded proteins in the periplasm rather than the occurrence of inclusion bodies that activated the extracytoplasmic stress pathway. It was postulated that perhaps misfolded protein in the periplasm could interact with the sensor kinase, CpxA, which then induces the Cpx response.

The conformational status of polypeptides in inclusion bodies

The conformational quality of polypeptides within inclusion bodies is amongst the issues debated in the RPP community and for which a consensus has yet to be reached. A marked increase in research has taken place over the past 10 years that aimed to achieve a better understanding of the structure and composition of IBs. For a long time, it has been thought that inclusion bodies were formed by the non-specific aggregation of unfolded or partially folded polypeptides that have no biological activity (Georgiou and Valax, 1996; Carrió and Villaverde, 2001). Recent research has demonstrated that IBs are formed by the organized aggregation of polypeptides, which are ordered in cross β-sheet architecture. Thus IBs were shown to share similar architectural features to amyloids (Carrió *et al*., 2005). Aggregation of such polypeptides is the result of specific interactions between exposed hydrophobic patches of unfolded and partially folded polypeptide species (Carrió *et al*., 2005).

Further investigation revealed that bacterial inclusion bodies also incorporate 'significant amounts' of functional polypeptides (García-Fruitós *et al*., 2005b; 2007). This has been demonstrated by the inclusion bodies formed during the production of aggregation-prone proteins such as the Alzheimer peptide Aβ42 fused to either GFP or blue fluorescent protein (BFP), and the virus capsid protein VP1, fused to either β-galactosidase or GFP (García-Fruitós *et al*., 2005b; de Groot and Ventura, 2006a, b; García-Fruitós *et al*., 2007). Fluorescence microscopy revealed that the IBs formed by VP1GFP and Aβ42GFP/ BFP were highly fluorescent, indicating that aggregated protein retained biological activity. Both the fluorescence and size of IBs were inversely correlated with temperature (de Groot and Ventura, 2006; Vera *et al*., 2007). This was reasoned to be due to the increase in the proportion of correctly folded target protein in IBs. Of course, the increased fluorescence emitted from IBs could indirectly be due to higher yields of correctly folded protein from the soluble fraction that have adsorbed on the surface of IBs. In support of this proposition, it has been observed that fluorescent inclusion bodies formed at lower temperatures were less stable and more susceptible to solubilisation compared to IBs formed at higher temperatures (de Groot and Ventura, 2006). This indicated that IBs formed at low temperatures were the result of intermolecular interactions between correctly folded proteins on the IB surface.

In an attempt to ascertain the localization of these functional polypeptides, the fluorescence distribution within IBs was analysed using confocal microscopy. There was a clear gradient in fluorescence being lowest at the surface but increasing towards the core of IBs (García-Fruitós *et al*., 2007). This pattern was consistent over IBs formed at different suboptimal temperatures: 30, 25, 20 and 16°C. This was very surprising, particularly when

the growth of inclusion bodies is driven by the adsorption of polypeptides on the exterior surface. The authors concluded that since functional polypeptides were specifically localized and integrated at the core of IBs, this cannot be the result of passive entrapment of functional polypeptides. How this specific integration within IBs is possible was not proposed.

Further attempts to discern the localisation of functional polypeptide within IBs revealed that significant amounts of functional VP1LAC and VP1GFP were released upon resuspension of IBs in PBS (García-Fruitós *et al*., 2007). This indicated that polypeptides with biological activity were present at the surface of IB and not exclusively in the core. Very little biological activity remained associated with IBs 30 minutes after resuspension, whereas activity had increased significantly in the solvent, reflecting the loss of functional protein from the aggregates (García-Fruitós *et al*., 2007). The presence of functional protein within IBs could be due to the presence of misfolded stretches that facilitate aggregation in addition to properly folded domains that, if embracing an active site, would account for biological activity. The occurrence of active polypeptides as structural components of IBs has been used, rather prematurely, as there is evidence that solubility and functionality of recombinant protein are not synonymous (Gonzalez-Montalban *et al*., 2007b). It also led to the conclusion that IBs were a rich source of properly folded protein.

The divergence of solubility and conformational quality

Conventionally, solubility and conformational quality are considered to be unequivocally connected, such that co-expressing DnaK could increase both solubility and biological activity of the target protein (Sorensen and Mortensen, 2005 a,b; de Marco *et al*., 2007).

The divergence of solubility and functionality was demonstrated during the production of the aggregation-prone VP1 protein fused to a misfolding variant of GFP (mGFP) in *E. coli* strains that were deficient in chaperones such as DnaK and ClpB and proteases such as ClpP and Lon (García-Fruitós *et al*., 2007b). In these strains, the recombinant protein accumulated was significantly less soluble but unexpectedly more fluorescent, compared to the wild type strain. In the DnaK mutant, this was attributed to the loss of the dissolution of functional polypeptides by DnaK and their subsequent targeting to proteases (García-Fruitós *et al*., 2007b).

Further evidence supporting the divergence of protein solubility and activity was demonstrated by Arié *et al*. (2007). The inherently soluble periplasmic enzymes β-lactamase (Bla) and alkaline phosphatase (PhoA) were fused to an insoluble and misfolding variant of the maltose binding protein, MalE31. Both fusion proteins accumulated in inclusion bodies in the periplasm. However, bacteria that produced insoluble MalE31-Bla were able to form colonies in the presence of ampicillin, indicating that the Bla moiety was catalytically active in this fusion protein. Also, phosphatase activity was detected by bacteria that had accumulated insoluble MalE31-PhoA. No degradation products or soluble functional enzymes were detected that could have accounted for this activity, indicating that the enzymic activity was that from recombinant protein located in inclusion bodies. This clearly demonstrated that Bla and PhoA were active inside periplasmic inclusion bodies and that proteins within IBs can be correctly folded. However, the specific activity of alkaline phosphatase detected from permealised cells that have produced insoluble PhoA was much lower compared to when soluble PhoA was accumulated. This indicated that correct folding of PhoA in the insoluble fractions was

much slower and more difficult compared to the soluble fraction, and that protein aggregation compromises to a great extent biological function.

The most recent model for the accumulation of recombinant protein is that under conditions that favour folding such as low temperature, the yield of functional protein is increased in the soluble and insoluble fractions simultaneously. Hence, the solubility of protein as a universal indicator of conformational quality and biological function was therefore challenged (González-Montalbán *et al*., 2007b). Villaverde and co-workers have therefore proposed a model whereby the accumulation of correctly folded protein is localized non-preferentially in both the soluble fraction and IBs (González-Montalbán *et al*., 2007b). Therefore, the 'average conformational quality of inclusion body protein is representative of that found in the whole cell'. However, this is inconsistent with and contradicts results that have emerged from the same laboratory, which show that the soluble fraction comprises approximately 10-fold more fluorescent GFP fusion protein compared to the insoluble fraction in wild type *E. coli* at 37°C (García-Fruitós *et al*., 2007b; Martínez-Alonso *et al*., 2007, 2008; Vera *et al*., 2007). This indicates that correctly folded protein mainly exists as soluble protein but sometimes associates with aggregated proteins.

Other problems associated with conventional RPP

Proteolysis

Harcum and Bentley (1999) have reported that the heat shock and stringent responses have overlapping protease activity in *E. coli*. Although solubility can indeed be enhanced by high levels of DnaK and ClpB, this occurs at the expense of the yield of the target protein,

probably by generally stimulating proteolysis (Rinas *et al*., 2007). It has also been demonstrated from analysis of the bacterial transcriptome that many proteases that degrade the target protein are upregulated as part of the heat-shock response (Gill *et al*., 2000). Yields of the target protein accumulated by bacteria during the general stress response are therefore compromised.

Plasmid loss

The presence of a high copy number vector already creates a burden on the host due to overwhelming DNA replication resources. It has been reported that when rapid synthesis of a recombinant protein is induced by IPTG, the growth of plasmid-bearing cells deteriorates (Andersson *et al*., 1996; Bentley *et al*., 1990; Kurland and Dong, 1996; Miroux and Walker., 1996; Soriano *et al*., 1999; Sundström *et al*., 2004). This creates a selective advantage for a plasmid-free cell population to overgrow the plasmid-containing cells that are burdened with RPP (Jones, 2007; Sevastsyanovich et *al*., 2009). It has also been suggested that the extent of plasmid loss might be largely dependent on the nature of the protein being expressed in that bacteria are more likely to 'lose' plasmids that encode difficult target proteins (Corchero and Villaverde, 1998). The loss of plasmid from bacteria has been shown to coincide with cessation of protein accumulation. The problem of plasmid loss in bacteria must be overcome if continuous and sustainable protein accumulation is to be maintained.

The use of other expression systems in *E. coli*

Expression systems other than the T7 expression system for heterologous protein production in *E. coli* are available (reviewed in Terpe, 2007). These systems are based on the induction of heterologous gene expression from a plasmid using strong promoters. The hybrid promoters such as P*tac* and P*trc* are examples of strong promoters that drive the high-level expression of recombinant genes using the *E. coli* RNA polymerase (deBoer *et al*., 1983; Brosius *et al*., 1985). Levels of recombinant gene expression from the P*tac* promoter has been reported to be 11-fold higher compared to expression from the *lac*UV5 promoter (deBoer *et al*., 1983). Both *tac* and *trc* promoters allow the accumulation of the recombinant protein up to 15 -30% of total cell protein.

Perhaps the main disadvantage of plasmid vectors based on the *lac* promoter is that they are notoriously leaky. Plasmids have been therefore engineered to include the *lac* operator sequence downstream of the *lac* promoter to minimise expression of the recombinant gene in uninduced cells. Despite this, relatively high basal levels of expression from these vectors still occur even in the absence of the inducer. This is because these vectors lack additional 'pseudo-operator' sequences for the Lac repressor that are necessary for almost full repression of the *lac* operon (Oehler *et al*., 1990). The leaky basal expression from these vectors therefore renders them unsuitable for the production of toxic gene products. The need for tighter control of expression has therefore led to the development of expression plasmids based on the araBAD promoter (P*araBAD*) (Guzman *et al*., 1995). In this arabinose induction system, gene expression is activated by AraC upon induction by the addition of arabinose and is subjected to catabolite repression. This ensures a very

tight control of gene expression and renders the arabinose expression system ideal for the expression of toxic gene products. It was also reported that the levels of expression of the alkaline phosphatase gene *phoA* from the arabinose expression vectors was approximately 200- to 1,200-fold compared with a ratio of 50 fold from the P*tac* promoter (Guzman *et al*., 1995). The very high yields of recombinant protein that could be obtained using this system have made the arabinose expression system a popular choice for RPP. However, despite the availability and choice of different expression systems, the high level induction of recombinant protein from the P*tac*, P*trc* and the P*araBAD* systems result in severe physiological stress on the bacterial host. The upregulation of genes involved in the stringent and heat shock responses during RPP using these expression systems was evident from microarray studies (reviewed in Sevastsyanovich *et al*., 2010).

The use of other production hosts for RPP

Other hosts may be more suitable than *E. coli* depending on the nature of the protein to be produced. The Gram positive bacteria such as *Bacillus subtilis* and *Lactococcus lactis* are characterised by their natural tendency to secrete high-quality recombinant protein in the extracellular medium by the Sec pathway. The advantage of secretion of recombinant protein is that both the proteolytic degradation within the cell is avoided and the downstream processing is minimised. Not surprisingly, therefore, *B. subtilis* is currently the microbial host responsible for the overproduction of over 60% of commercially available enzymes. Ironically, however, protein secretion in both these strains remains the bottleneck in RPP, in which the many steps involved in secretion and translocation are rate limiting (Westers *et al*., 2004). This results in the accumulation of protein in the cytoplasm

and the increased aggregation of newly synthesised proteins in inclusion bodies (Jürgen *et al*., 2001). Similar to *E. coli*, the rapid overproduction of recombinant protein in *B. subtilis* results in severe physiological stress (Jürgen *et al*., 2001). Transcriptomic analysis of *B. subtilis* during production of the outer membrane protein PorA from *Neisseria meningitidis* revealed a significant upregulation of heat shock genes encoding the chaperones DnaK, GroEL and GrpE, in addition to the proteases ClpP and ClpC. Studies have reported the extreme negative consequences of RPP on bacterial growth, particularly as 'the relative synthesis rate of the general stress proteins constituted up to 40% of the total protein synthesis of stressed cells (Eymann *et al*., 2002; Bernhardt *et al*., 1997).

The need to produce proteins requiring special post-translational modifications such as glycosylation, proteolytic processing and disulphide bond formation also requires an expression host other than bacteria. For such proteins, lower eukaryotes such as yeast such as *Pichia pastoris* and filamentous fungi such as *Aspergillus niger* are the hosts of choice. During the last decade, yeast became increasingly popular for the production of antibodies and antibody fragments, whereas filamentous fungi have been widely used for large-scale production of industrial enzymes (reviewed in Gasser *et al*., 2008; Mattanovich *et al*., 2004). In these eukaryotic production hosts, the process of protein folding, post-translational modification and translocation for further secretion are accomplished in the endoplasmic reticulum. Excessive accumulation of misfolded proteins in the ER, however, leads to activation of a set of genes associated with protein folding and transport as well as ER-associated protein degradation (ERAD), in which misfolded polypeptides are translocated back into the cytosol for degradation. This activation is known as the unfolded protein response (UPR) and constitutes an ER stress response in these hosts. ER

stress is characterised by down-regulation of energy consuming biosynthetic pathways such as for vitamin, amino acid, carbohydrate and lipid metabolism, which ultimately inhibits the growth of these production hosts (reviewed in Graf *et al*., 2008). Early microarray studies of *P. pastoris* during production of trypsinogen has revealed the down regulation of genes related to ribonucleotide and ribosome synthesis as well as the tricarboxylic acid cycle, which are required to sustain RPP (reviewed in Sevastsyanovich *et al*., 2010).

It is clear therefore that the avoidance or minimisation of stress on the production host is essential when high final yields of the recombinant protein are required. This emphasises the need to review and most importantly redesign the methods by which recombinant proteins are produced by both bacteria and yeast.

Attempts to improve RPP

Attempts to improve RPP have focused mainly on addressing individual aspects such as the conformational quality of protein accumulated. Others were more concerned to improve yields of the target protein either directly by minimizing proteolysis and degradation or indirectly by improving plasmid retention in bacteria.

Protein folding

The folding of proteins into the native state is necessary for their function whether it be enzyme activity, fluorescence or binding to other proteins. Generally, protein folding is a spontaneous process that is often aided by chaperones. Chaperones are proteins that aid the

folding of newly synthesised or misfolded proteins and are upregulated as part of the heat-shock response in bacteria (Baneyx and Mujacic, 2004; reviewed in Hoffman and Rinas, 2004; Young *et al*., 2004; Gasser *et al*., 2008). They can actively assist in protein folding, prevention of aggregation, and the solubilisation of aggregates in the cytoplasm. Many chaperone systems are upregulated in response to the accumulation of misfolded proteins, most notably, DnaK and the GroEL/ES chaperones. DnaK has a broad spectrum of activities. Together with its cofactors DnaJ and GrpE, DnaK promotes the folding of proteins to a native or partially folded conformation. In addition, it plays a very important role in the removal of polypeptides from IBs and the targeting misfolded intermediates for proteolysis. The activity of the GroELS system, however, is brought about by GroEL structure in which 2 heptameric rings are stacked on top of each other forming a hollow hydrophobic cylinder cavity. GroEL acts by binding the unfolded polypeptide at either of the outer ends of its inner cavity through hydrophobic interactions. This is followed by the capping of the cavity by GroES, creating a closed hydrophobic environment, in which protein folding is favoured. Several studies have investigated the effect of co-expression of different combinations of chaperones on protein folding and solubility of the recombinant proteins (Nishihara *et al*., 1998; de Marco, 2007; Gupta *et al*., 2006). However, co-expression of chaperone systems to improve protein folding has only had limited success, and seems to be largely a trial and error process, which is dependent on different combinations of the chaperones and sometimes leads to unpredictable results. A different approach, which has become a standard procedure in industry, to increase yields of correctly folded target protein, is the addition of 2 to 3% ethanol to the culture around 5 hours post-induction (Jones, 2007). The addition of ethanol induces the heat shock response and thus the synthesis of both DnaK and GroEL/ES chaperone systems (Gerogiou

and Valax, 1996). This has been used with some success to increase yields of soluble mitochondrial membrane-bound protein, cytochrome P450 (Kusano *et al*., 1999).

In addition to chaperone co-expression, a well-known technique to limit the aggregation of unfolded proteins consists of cultivation of bacteria at low temperatures (Sorensen and Mortensen, 2005). Aggregation is favoured at high temperatures due to increased hydrophobic interactions between misfolded proteins (Kiefhaber *et al*., 1991; Baneyx and Mujacic, 2004). Yields of soluble recombinant proteins have been increased by 4-fold when temperatures were decreased from 37°C to 16°C (Vera *et al*., 2007). However, the RPP at decreased temperatures often occurs after the initial growth of bacteria at 37°C. Despite the increased potential for correct folding of proteins at low temperatures, the sudden decrease in temperature post-induction induces a cold-shock that is detrimental to prolonged RPP. It has been reported that this sudden decrease in temperature inhibits replication, transcription and translation (Shaw and Ingraham, 1967).

To understand why decreasing the temperature has beneficial effects on increasing the conformational quality of recombinant proteins, the transcriptome of *E. coli* has been analysed during RPP at different temperatures during growth with glucose as the primary source of carbon and energy (Gadgil *et al*., 2005). There was a general down-regulation of glycolysis and energy-generating enzymes as well as biosynthetic pathways at 28ºC and 33ºC compared with 37ºC. This suggests that the beneficial effects of lowering the growth temperature for biotechnological processes could be attributed to a general effect of reduced growth rate and metabolism.

Protein yields

Although yields of recombinant protein are initially very high using the standard protocol, these rapidly decline in prolonged production attempts mainly as a result of physiological stress. In addition to the physiological consequences associated with RPP and the heat-shock response, the significant upregulation of proteases was also documented in transcriptomics studies, most notably being Lon, ClpP and DegP (Gill *et al*., 2000). The ATP-dependent proteases Lon and ClpP participate in the physiological disintegration of cytoplasmic inclusion bodies and together they are responsible for over 70% of ATP-dependent proteolysis in the cell. This prompted the investigation of whether yields of recombinant proteins accumulated could be increased in protease deficient cells. Indeed, the yield of the recombinant protein VP1LAC was increased by approximately 60% in a *clpP* mutant relative to the wild type, but surprisingly decreased by about 50% in a *lon* background (Vera *et al*., 2005). It was therefore suggested that Lon might have a main role in the low specificity digestion or solubilisation by releasing proteins from inclusion bodies whereas ClpP has a higher proteolytic activity. Whether it would be beneficial to utilise Lon by co-expressing it with the recombinant protein was considered by the authors as a possible strategy to improve RPP. The production of Lon could prevent the bacterial cells from becoming 'choked' with inclusion bodies and enhance the amount of target protein in the soluble fraction. It was also revealed that yields of VP1LAC in the soluble fraction of a ClpP[-] mutant were considerably higher than in the wild type strain, suggesting that ClpP degrades soluble, conformationally active protein and was deemed very counterproductive to RPP (Vera *et al*., 2005).

Plasmid maintenance

The loss of the recombinant plasmid is very often associated with decreased yields of the recombinant protein during prolonged expression. Owing to the interpretation by many that the loss of plasmids from bacteria is a cause rather than an effect of the methodologies used for RPP, considerable progress has been made to overcome plasmid loss in its own right. This was achieved by including multimer resolution sites or partition elements to ensure accurate segregation during host cell division (Skogman *et al*., 1983). Attempts were also made to engineer a BL21(DE3) strain named BLR(DE3) that is deficient in *recA*, which successfully decreased the frequency of multimerisation and therefore the generation of plasmid-free bacteria in the culture (Zhao *et al*., 2007a and b). However, this was not accompanied by an increased yield of the target protein due to severe growth inhibition of plasmid bearing cells. It was reported that plasmid stability can be improved when production of the target protein is induced at lower temperatures (Zhang *et al*., 2003). This could be indirectly achieved by the slower bacterial growth rate, which ensures a more accurate segregation of plasmids during cell division. A more recent strategy was developed that does not rely on plasmids whatsoever (Streidner *et al*., 2010). This plasmid-free system relies on the overexpression of the recombinant gene, which is site-specifically integrated into the chromosome of the host, using the highly efficient T7/ pET expression system. The high levels of expression achieved from the pET system were exploited to compensate for the decreased gene dosage. This system clearly has the advantage of releasing the host cells from the metabolic burden of plasmid maintenance and eliminates problems of plasmid loss associated with conventional expression systems. However, attempts to use this system to produce the proteins GFPmut 3.1 and human superoxide dismutase in both *E. coli* HMS174(DE3) and BL21(DE3) have resulted in a very modest 2-

fold increase in the yields of recombinant protein accumulated relative to the plasmid-based system. The practicality of engineering different bacterial hosts for every recombinant protein is questionable and is highly unlikely to replace plasmids as expression vectors.

Decreasing stress

Improving the yields of recombinant protein, by minimizing the stress associated with RPP, has been one of the least studied strategies. Few have come to the realisation that the obvious way to achieve this is to slow down the rate of product synthesis so that folding pathways keep pace with protein synthesis, thus preventing the accumulation of unfolded products (Wagner *et al*., 2008). This can be achieved by decreasing the temperature of the culture following the addition of the inducer (reviewed in Sevastsyanovich *et al*., 2009) or decreasing the concentration of the inducer to decrease the expression level of the recombinant gene. One of the few studies to investigate this was reported by Narayanan *et al*. (2006) during the production of the periplasmic penicillin acylase (PAC). The production of PAC in its mature form was considered technically challenging. This is because many posttranslational steps are required such as translocation and periplasmic processing that includes the autoproteolysis of the PAC precursor (proPAC) into the mature form. It was found that when *pac* expression was induced from P*trc* or Pt7 promoters using IPTG, high levels of proPAC accumulated as insoluble aggregates in the periplasm. This resulted in severe growth inhibition of bacteria and cell lysis. Interestingly, it was observed that the addition of arabinose was also able to induce expression from *lac*-derived promoter systems, and prompted attempts to produce PAC by induction with arabinose rather than IPTG. It was reported that induction of *pac* expression using 0.015% arabinose,

rather than 0.1 mM IPTG, increased the yield of soluble and fully processed PAC in the periplasm by approximately 11-fold. Bacterial growth also continued to increase after induction.

The ability of arabinose to induce expression from *lac*-derived promoters suggests that it can bind to the Lac repressor. This possibility was supported by alignments of the primary sequence of the Lac repressor and the arabinose binding protein, which revealed that they share very similar secondary structures (Nichols *et al*., 1993). The success of arabinose induction of *lac*-derived promoter systems could therefore be attributed to the possible inefficiency of the induction of gene expression. Instead, the authors have attributed RPP outcomes to the nature of inducer, suggesting that arabinose, rather than IPTG, is able to 'mediate effective processing and folding' of the synthesised recombinant proteins. Although the authors recognise that a balanced flux between the different steps involved in the production of complex recombinant protein must be maintained, they fail to attribute the success of arabinose induction to the inefficiency in *pac* expression. This would result in a slower rate of proPAC synthesis and the elimination of bottlenecks involved with post-translational steps.

Similar approaches have been designed to regulate the expression of the recombinant gene. These include initiating the synthesis of the target protein at a slow rate, which is gradually increased by increasing the concentration of inducer (Trepod and Mott, 2002). Despite the use of this method in slowly 'priming' or adapting bacteria to RPP stress, it was not sustainable for prolonged RPP.

E. coli strains used to improve production of soluble recombinant protein

After the stress and consequences of RPP from the T7 expression system in BL21 had become apparent, it was essential to isolate improved *E. coli* strains that are able to withstand the stress of RRP. Miroux and Walker (1996) were the first to isolate improved hosts that were able to accumulate correctly folded membrane proteins without the detrimental effects of growth arrest following induction. The bacterial strain used in this work BL21, and the recombinant protein was produced by means of the T7 expression system. Improved host candidates were screened based on their growth in the presence of high levels of IPTG on solid media. Individual colonies of different sizes were screened for improved accumulation of the target proteins. Two strains called C41(DE3) and C43 (DE3) were isolated, the latter being much better at accumulating higher yields of membrane protein and growth to higher densities following induction. It was later revealed that both of these hosts had accumulated 3 identical mutations in the *lac*UV5 promoter and operator regions that govern the expression of the T7 RNA polymerase gene on the host chromosome (Wagner *et al*., 2008). These mutations resulted in the downregulation of T7 RNA polymerase to levels that were unable to be detected by western blotting. Further analysis of these strains, which were also termed the 'Walker strains', revealed that by swapping the promoter and operator regions of the T7 RNA polymerase from the Walker and wild type strains of BL21, a Walker strain can be transformed into a wild type BL21, and *vice versa*. This indicated that mutations that decreased the expression of T7 RNA polymerase were responsible for all of the improvement shown by these strains in accumulation of correctly folded membrane proteins. More importantly, this highlights the significance of slow expression rather than the rapid 'over-expression' in the prolonged

accumulation of native recombinant protein. The success of the Walker strains in their accumulation of membrane proteins such as F_1F_o ATP synthase subunit membrane protein was demonstrated (Arechaga et *al*., 2000; Sorensen and Mortensen, 2005).

Green fluorescent protein marker in RPP

GFP is a small protein of 238 amino acid residues (27kDa), which emits bright green fluorescence when exposed to blue light. GFP forms an 11- stranded *β*-barrel that is threaded by an α-helix running up through the axis of the cylinder (Yang *et al*., 1996; Cubitt *et al*., 1999). The fluorescence property of GFP is conferred by the presence of a covalently bound fluorophore, which is buried in the centre of the cylindrical structure. The fluorophore is assembled by the interaction of the side chains of three essential amino acids in the GFP protein, namely Ser65, Tyr66 and Gly67. The presence of oxygen is essential for the formation of the fluorophore. GFP has to be correctly folded before fluorophore formation is possible. The amino acid residues that span the fluorophore have been mutated to produce mutants with greater fluorescence or different spectral properties. For example, enhanced GFP (EGFP) is a form of GFP containing both F64L and S65T substitutions with increased fluorescence by 40- to 80-fold relative to the wild type form (Heim *et al*, 1995; Delagrave *et al*., 1995; Miller and Lindow, 1997). One of the main advantages of GFP compared to other fluorescent proteins such as luciferase is that it requires no cofactors for fluorescence. This has allowed the widespread use of GFP as a marker for RPP.

GFP could be exploited as a folding indicator of proteins.

Albano *et al.* (1998) were amongst the first to demonstrate the potential of green fluorescent protein (GFP), when fused to the target protein, as a reporter of recombinant protein accumulation. This depends on the observation that GFP is essentially not fluorescent unless it has folded correctly (Waldo *et al.*, 1999). In a pioneering study by Waldo *et al.* (1999) using test panel of 20 proteins fused to GFP, it was demonstrated that the level of fluorescence emitted by bacteria was directly proportional to the productive folding of the upstream protein. These observations were mirrored in a study by Jones, 2007 during production of 3 different GFP protein fusions. This indicated that GFP fluorescence reports the folding of the fusion partner occurring co-translationally or soon after induction. Therefore, if GFP is correctly folded and fluoresces, then the upstream fusion partner is also correctly folded. This was tested further by Jones (2007), by attempting the production of GFP fusion proteins in which the GFP protein is fused upstream of the target protein. It was found that the ability of GFP to accurately report the folding of the target protein occurs only when GFP is fused to the C-terminus of the target protein. In contrast, N-terminal GFP fusions are poor reporters of the folded state of a C-terminal target (Jones, 2007). The discrepancy of reporter protein orientation to accurately report the folding and solubility of the target protein has also been demonstrated with β-galactosidase and maltose binding protein (MalE) (Carrió *et al.*, 1999; Aríe *et al.*, 2006). It was conclusively established that the solubility of fusion proteins is dictated by the intrinsic folding characteristics of the upstream fusion partner.

The level of fluorescence emitted from C-terminal GFP fusions was also used as an accurate measure of the yield of correctly folded target protein (Cha *et al.*, 2000; Waldo *et*

al., 1999; Jones, 2007). Linear positive correlation between GFP fluorescence and yields of correctly folded, biologically active recombinant proteins that were present in the soluble fraction, was also reported during the production of different GFP fusions (Cha *et al*., 2000; Jones, 2007; Waldo *et al*., 2009). However, proteins with no fluorescence were found to be incorrectly folded in inclusion bodies in the insoluble fraction. The utility of GFP fluorescence as a reporter to the production of the fusion protein was used in the development phase of an RPP project to determine the optimal conditions and accurately predict the outcome of the production of the untagged form of the heterologous protein (Jones, 2007). It was found that fermentation conditions optimised for production of proteins with GFP fused at the C-terminus were identical to those for the production of the untagged forms of the target protein. The advantage of this is that subsequent production rounds of the untagged protein will not require the GFP tag removal from the protein.

The ability of GFP fluorescence to report protein folding and solubility was used to evolve proteins that are normally prone to aggregation into closely related proteins that fold properly and are fully soluble and functional. This process was called 'directed evolution' and involved DNA shuffling of the misfolding protein sequence to generate and recombine mutations (Waldo *et al*., 1999). Fluorescence of GFP was then used as a basis to identify variants of protein that had improved folding and solubility by screening for increased fluorescence. These proteins were also shown to have retained their catalytic activity. Hence the utility of GFP, as an indicator of protein that is correctly folded and not aggregated, was established. Fortunately, the dependency for GFP fusions to be at the C-terminus to accurately report folding of target proteins leaves open possibilities of producing periplasmic or outer membrane proteins that have a signal sequence at their N-terminus for translocation.

In the majority of proteins tested, the fluorescence of GFP fusions was shown to correlate positively with the solubility of the non-fused counterpart of the target protein (Jones, 2007; Waldo *et al*., 1999). Production of GFP fusions is therefore currently being used by industry to make informed decisions about the optimal conditions that improve yields and solubility of the non-fused target proteins (Jones, 2007). This also has a bearing in determining the optimal harvest time by simply monitoring maximum fluorescence during process development (Cha *et al*., 2000; Jones, 2007).

The fluorescence of GFP as an indicator for the correct folding and assembly of inner membrane proteins was also investigated (Drew *et al*., 2001; Hedhammar *et al*., 2005). Production of GFP protein fusions that are inserted into the inner membrane (ProW, YidC, Lep-inv, KDEL-receptor) was compared to those that accumulate predominantly in inclusion bodies (GST-GPCR, M13 procoat). GFP was fused at the C-terminus of these proteins, such that it topologically remained in the cytoplasm when the proteins are inserted into the inner membrane. It was shown that when the membrane protein–GFP fusions were expressed as inclusion bodies in the cytoplasm, GFP was not fluorescent (Drew *et al*., 2001). This was because in inclusion bodies GFP had not folded properly into the correct β-barrel structure that is essential for fluorescence (Drew *et al*., 2001). However, when the membrane protein–GFP fusions were inserted into the cytoplasmic membrane, GFP was properly folded and became fluorescent. Thus, the fluorescence emitted from these fusion proteins could be used as a measure of the protein yield that had been inserted into the membrane rather than deposited in inclusion bodies (Drew *et al*., 2001). These findings have been supported by confocal microscopy data, which showed that aggregated GFP fusion proteins were non fluorescent (Hedhammar *et al*., 2005). Thus, agreement that GFP

fluorescence could be used to estimate whether the majority of the protein produced is soluble or insoluble was established.

The effect of protein translocation on fluorescence of GFP protein fusions

The main limitation of using GFP as a marker protein is that it fluoresces only in the cytoplasm of *E. coli* (Feilmeier *et al*., 2000). This is because GFP is incorrectly folded after being translocated to the periplasm via the general secretory system (the Sec pathway). When a GFP fusion protein is translocated to the periplasm by the Sec-dependent pathway, fluorescence is abolished. This has been demonstrated with studies in which GFP was fused to maltose-binding protein (MBP) at the C- terminus, with or without a signal sequence. Cell fractionation to separate the cytoplasm and periplasm showed that fluorescent MBP-GFP was localised only in the cytoplasm. In contrast, non-fluorescent MBP-GFP was localised to the periplasmic space (Feilmeier *et al*., 2000). However, the fluorescence of GFP could be restored upon treatment with acid and base in the presence of a reducing agent (Feilmeier *et al*., 2000). It was therefore proposed that the radically different redox environment of the periplasm might cause the formation of disulphide bonds between cysteines in GFP, and thus affect its overall structure. The lack of GFP fluorescence in the periplasm has been challenged by Thomas *et al*. (2001). They show that periplasmic GFP is as active as the cytoplasmic GFP when exported by the twin arginine translocase (Tat) pathway. Thus, the physical property of a translocated protein was attributed mainly to the transport system that is used. Since GFP must be threaded through the Sec translocon in the unfolded state, refolding might be difficult in a non-reducing environment such as the periplasm. The advantage of exporting proteins using

the Tat pathway is that a protein such as GFP could be translocated to the periplasm in a fully folded active state (Thomas *et al*., 2001).

However, this phenomenon still poses a problem when periplasmic or outer membrane proteins possessing Sec signal sequences are the focus of study, even though it has already been exploited to analyse the global topology of inner membrane proteins in *E. coli* (Daley *et al*., 2005). The fluorescence properties could facilitate the study of protein export in bacteria and serve as a new screening method for Sec transport defective mutants (Feilmeier *et al*., 2000).

Tools for monitoring accumulation of GFP protein fusions

Tools such as an online GFP sensor and flow cytometry have been successfully used for monitoring the real-time accumulation of GFP protein fusions. The value of an online GFP sensor is that it reveals the production rate and yield of the heterologous protein accumulated by bacteria over prolonged and continuous periods without the need for invasive sampling (DeLisa *et al*., 1999; Jones *et al*., 2004: Jones, 2007). The utility of this approach has been demonstrated in the development phase of an RPP project to screen for conditions that would affect the RPP response by measuring changes in the fluorescence of a culture. The main limitation with a GFP online sensor, however, is that it reflects an average RPP response from bacterial cells within a culture, regardless of their physiological state. The results obtained using this method is based on the false assumption that bacterial cultures are homogenous.

In contrast to an online GFP sensor, flow cytometry is a powerful tool for measuring single cell properties of thousands of cells, and has been used to identify subpopulations of bacteria at different stages of protein production (Borth *et al*., 1998; Hewitt and Nebe-Von Caron, 2001; 2004). As opposed to conventional techniques, which measure an average value for the entire cell population, flow cytometry provides information on cell-to-cell variability. The use of multi-parameter flow cytometry to analyse different properties of bacteria in the culture is integral in RPP research. The two main properties of bacteria that are frequently analysed are the light scattering profiles of individual cells and their fluorescence emission. Light scattering by cells can accurately indicate the size, shape and morphology of individual bacterial cells. The data obtained can provide an invaluable indication of the physiological state of bacteria in the culture during recombinant protein production. For example, cells that have become elongated or have adopted a coccoid morphology after induction of RPP could reflect physiological stress or inclusion body formation, respectively (Patkar *et al*., 2002; Soriano *et al*., 2002; Lewis *et al*., 2004; Hedhammar *et al*., 2005) (**Figure 1.3 A**).

A unique advantage of flow cytometry technique is that it can identify groups of cells with different physiological states and RPP profiles within a culture. An example of this is the 'viable but non-culturable' bacteria. Classical microbiology states that a cell is said to be viable is when bacterial cells are able to reproduce. This has become the 'golden standard' for evidence of cell viability (Hewitt and Nebe von Caron, 2004). However, the ability of bacteria to reproduce is largely dependent on the medium and artificial conditions used for growth. This has resulted in differences in counts of viable cells of up to several orders of magnitude between different studies. Stressed or injured cells are 'viable but non-

Figure 1.3: Schematic representation of flow cytometry output data. A) Light scattering parameters of bacterial cells. Laser beams are focussed on individual bacterial cells in the flow cytometer, causing the light to be diffracted by cells. The scattered light particles are detected by a forward scatter (FS) and side scatter (SS) detectors. The light scattering properties of approximately 10,000 individual cells is plotted on a scatter plot, generating a distribution pattern which indicates cell shape and morphology. The left panel shows the light scattering properties of cells that have a rod morphology whereas the scattering profile of cells with a coccoid morphology is represented in the right panel, the latter might indicate the presence of inclusion bodies after RPP has been induced. **B)** Fluorescence profile of bacterial cells. The fluorescence emitted by individual cells, such as green fluorescence in cases where GFP fusion proteins are accumulated, or red fluorescence emitted by dead bacteria that have taken up propidium iodide stain, is detected. The top left panel shows a cell population that emits no green or red fluorescence, indicating that GFP protein fusions are not being produced and that the cells are not dead. The top right panel shows a cell population that emits high levels of green fluorescence, indicating that a GFP protein fusion has accumulated. The bottom left panel shows a cell population that emits no green fluorescence but emits high levels of red fluorescence, indicating dead cells have not accumulated a GFP protein fusion. The bottom right panel shows depicts a cell population that emits both high levels of green and red fluorescence, indicating dead cells that have accumulated high levels of a GFP protein fusion.
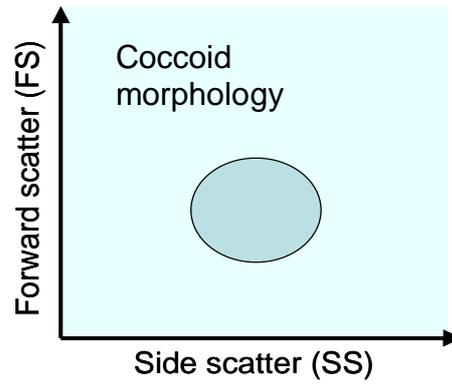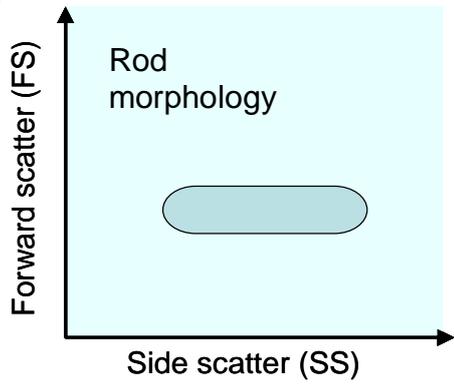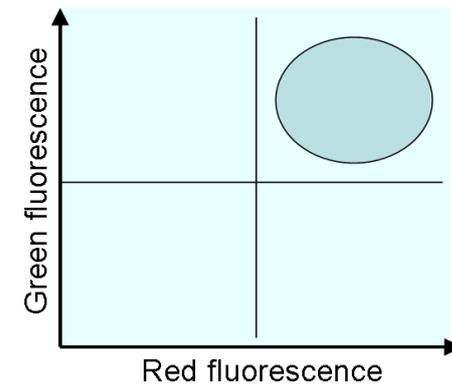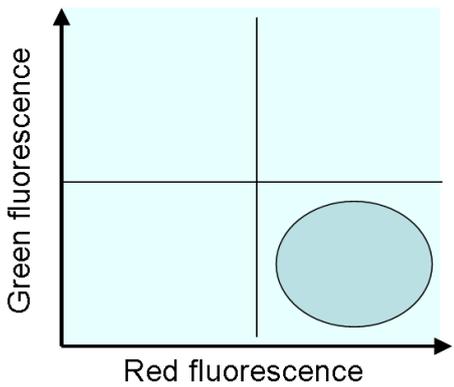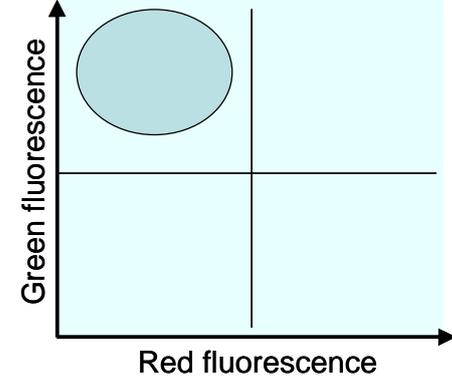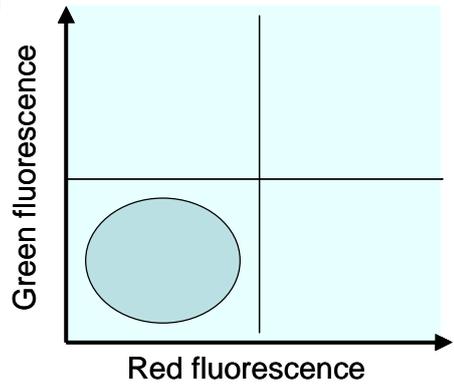
culturable' and can remain undetected by growth-based methods. However, groups of such bacteria that continued to grow in a bioreactor but ceased to form colonies on solid media after induction of RPP have been identified by flow cytometry and have been shown to produce high levels of protein (Andersson *et al*., 2006; Sundström *et al*., 2004). The value of these cells in RPP has not been thoroughly explored.

The fluorescence emission of individual bacterial cells can also be accurately measured. This depends on different fluorescence profiles emitted by bacteria, which could be intrinsic, such as during the production of fluorescent protein fusions. Alternatively, fluorescence of bacteria that have been positively stained with fluorescent dyes during a particular physiological response could be measured (Hewitt and Nebe-von-Caron., 2001; 2004). For example, propidium iodide is a red fluorescent dye that only stains dead bacteria with perforated membranes, whereas bis-oxonol is a green fluorescent dye that stains stressed cells with depolarised membranes. Thus the proportion of cells that have been differentially stained with either dye can be used as an accurate representation of the bacterial response in the context of RPP (Lewis *et al*., 2004). Perhaps the main application of flow cytometry during RPP is to monitor the accumulation of correctly folded fluorescent protein such as a GFP-protein fusion by the identification of cells with increased green fluorescence (Hedhammar *et al*., 2005; Patkar *et al*., 2002) (**Figure 1.3 B**). This way, and in combination with fluorescent stains, populations with elevated RPP levels and that respond differently to stress can be identified and physically separated for further analysis by fluorescence activated cell sorting (FACS). This could provide a platform for the isolation of improved hosts that are able to resist the stress associated with RPP and continue to accumulate high yields of correctly folded proteins. The full potential of multi-

parameter flow cytometry and FACS as tools for bioprocess development is yet to be fully utilised (Mattanovich and Borth, 2006).

The production of difficult recombinant proteins

In recent studies, the production of aggregation-prone proteins has been used to investigate various physiological responses during RPP and IB formation. However, production of difficult recombinant proteins that require many post-translational modifications might pose different challenges to bacteria. Therefore, knowledge derived on ways to improve RPP from production of soluble cytoplasmic proteins might be insufficient and inapplicable. Steps such as protein translocation and post-translational modifications are thought to significantly limit the rate and yield of production of the difficult proteins to be produced.

In most of this study, the production of two model proteins was investigated. These were an *E. coli,* cytoplasmic protein, CheY, and a difficult membrane protein from *Neisseria gonorrhoeae*, cytochrome *c* peroxidase that requires extensive post-translational modifications.

The chemotaxis protein, CheY

CheY is a member of the two component signal transduction membrane pathway in bacterial chemotaxis. The pathway consists of two proteins, an intracellular sensor kinase, CheA and a response regulator, CheY. In chemotaxis, the switch from a tumbling to smooth swimming of bacteria occurs when the rotation of the flagellum is switched from

clockwise to counterclockwise in response to an attractant chemical. This attractant creates a signal that triggers receptors in the membrane to phosphorylate the response regulator CheY. The phosphorylated CheY binds to a switching apparatus at the base of the flagellum and change its rotation, thereby controlling motility (Stock and Surrette, 1994).

Analysis of the structure of CheY has revealed that it undergoes a conformational change that enables it to interact directly with the flagellum. The structure of CheY has been elucidated; however the structure of the phosporylated form of CheY remains unknown (Silversmith and Bourret, 1999). The importance of CheY as a regulator of bacterial chemotaxis and in virulence is apparent and could be identified as a possible candidate for novel antibiotics. The production of CheY has been thoroughly investigated at pilot plant scale in industry in an attempt to define optimal conditions that would facilitate the accumulation of the protein in the soluble form (Jones, 2007). CheY was produced as a fusion with GFP and the accumulation of soluble fluorescent CheYGFP was monitored using an online GFP sensor. Although cultivation conditions such as pH, media composition and aeration were optimised to produce the highest yields possible of soluble CheYGFP, the majority of CheYGFP produced still accumulated in insoluble inclusion bodies. The overproduction of CheYGFP was therefore chosen for further investigation.

Cytochrome c peroxidase

Neisseria gonorrhoea is an obligate human pathogen. It has contributed to the significantly increasing cases of sexually transmitted infections, namely gonorrhoea, which if left untreated might cause bacteraemia. Key to maintaining infection, the gonococcus

possesses two very important outer membrane lipoproteins; cytochrome *c* peroxidase (CCP) and cytochrome *c'* (CycP). CCP and CycP enable the bacterium to evade the oxidative stress imposed by macrophage attack in the periplasm; namely hydrogen peroxide and nitrogen oxides, respectively (Turner *et al*., 2003; 2005). CCP protects bacteria from damage by reducing hydrogen peroxide to water, whereas CycP catalyses the conversion of potentially toxic nitric oxide to nitrous oxide.

CCP has been proposed as a promising candidate for use as an antigen in sub-unit vaccine development against gonorrhoea. Large quantities of the functional form of CCP however need to be obtained before it can be used in vaccinations. Many steps are involved in the production of mature CCP. First, CCP is synthesised as a pre-apo-protein with a signal sequence (Turner *et al*., 2003) (**Figure 1.4**). The signal sequence has to be cleaved by signal peptidase II as the protein is translocated across the cytoplasmic membrane. However, a covalent lipid attachment at the N-terminus of the protein is an essential pre-requisite for the cleavage of the signal peptide. Once the protein has been translocated into the periplasm, a haem group is covalently attached to the apo-protein (**Figure 1.4**). This is possible due to the combined activity of eight proteins encoded by the *ccm* (cytochrome *c* maturation) cluster of genes, which are essential for the assembly of all *E. coli c*-type cytochromes (Tanapongpipat *et al*., 1998). Other proteins such as DsbA/DsbB and DipZ also play an important role in reducing the cysteine residue to which the haem will be attached (Page *et al*., 1997, Metheringham *et al*., 1995, Crooke and Cole, 1995). A final step remains where the protein is loosely attached to the outer membrane by its lipid anchor. The complex post-translational steps involved in the production of mature CCP could significantly limit accumulation rates and yield of the functional recombinant protein
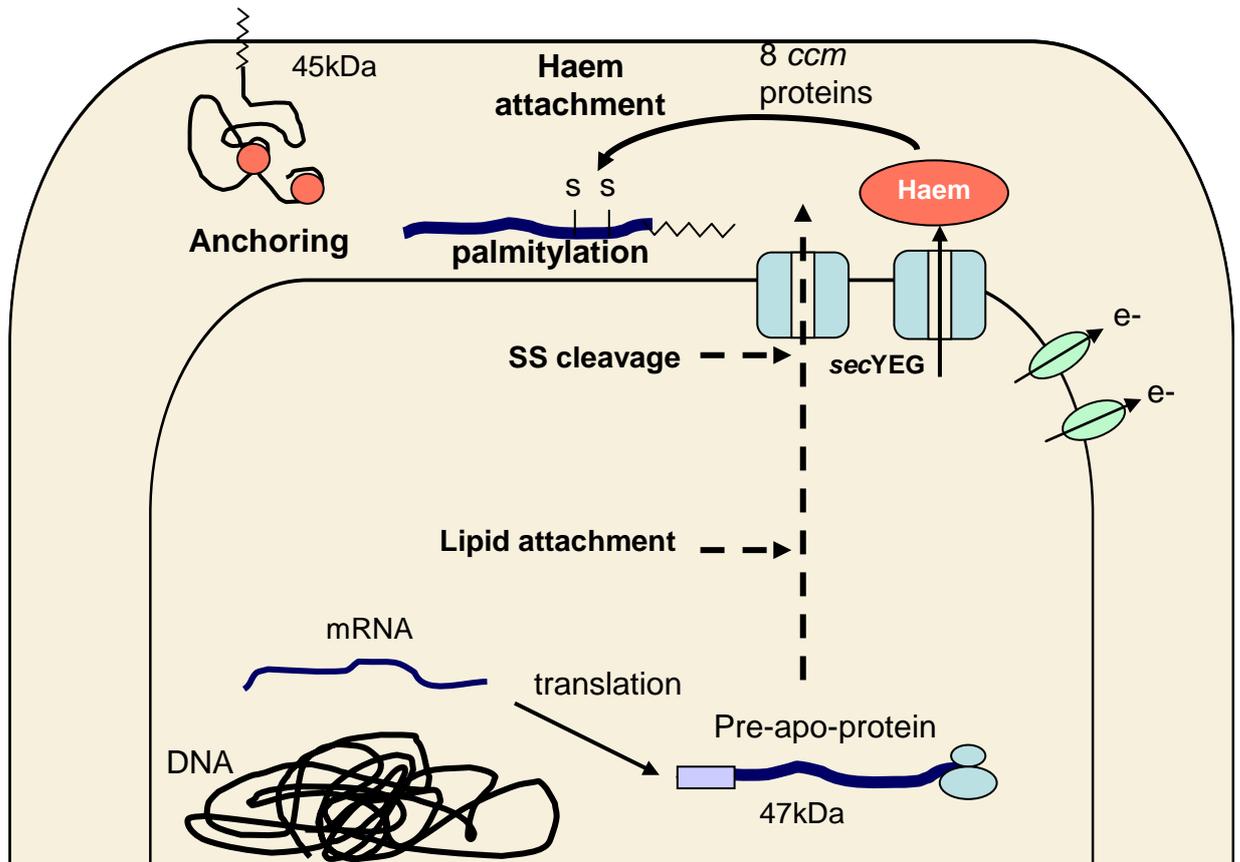
Figure 1.4: Production of mature CCP in *E. coli*. CCP is synthesised as a pre-apo-protein in the cell cytoplasm with a signal sequence at the N-terminus (47 kDa). Attachment of the lipid group precedes the translocation of the polypeptide across the cytoplasmic membrane. Cleavage of the signal sequence occurs as the lipid-modified CCP polypeptide is translocated by the sec system. Once in the periplasm, the cysteine groups in the haem attachment motif on the apo-protein are reduced by DsbA/DsbB and DipZ, to which the haem group is attached. A final step remains where the mature CCP protein (45 kDa) is loosely anchored to the outer membrane by its lipid group.

produced. This has been demonstrated by Turner *et al.* (2003) in that attempts to produce CCP in *E. coli* yielded mainly insoluble apo- form of CCP in inclusion bodies in the cytoplasm and very little had accumulated as mature form in the outer membrane. In conclusion, the cytochrome *c* proteins could be used to illustrate the production of difficult recombinant proteins and investigate their effects on bacterial physiology. Strategies that improve the production of functional forms of these proteins could then be developed based on the understanding of bacterial physiology and how it could be manipulated.

Conclusions and aims of this project

Conventional protocols for recombinant protein production entail the rapid synthesis of the recombinant gene in a heterologous host. In bacteria, this results in a severe metabolic burden and the induction of several stress responses such as the stringent and heat shock response which are detrimental on the host physiology. Soon after the induction of protein synthesis, bacterial growth is arrested and the recombinant plasmid is lost. Thus, the physiological responses of bacteria upon RPP using conventional methods compromise the yields of recombinant proteins that can ultimately be obtained. This highlights the unsustainability of conventional protocols for RPP. Furthermore, the recombinant protein accumulated using such protocols is often aggregated in insoluble inclusion bodies. This poses a problem if the prime objective is to achieve high quality recombinant protein, since extra downstream processing steps to refold the protein are required.

Strategies that aim to address one or more of the negative effects associated with conventional RPP protocols, such as plasmid loss or the quality of recombinant protein,

have been devised. These methods have seen partial success that resulted in a modest increase in the yields of correctly folded recombinant protein. However, generic protocols that result in the sustainable production of large yields of correctly folded recombinant proteins are unavailable and yet to be developed.

The first aim of this project was to exploit the findings from transcriptomics studies during RPP in order to develop generic strategies that do not result in physiological stress on the bacterial host and improve the accumulation of correctly folded target proteins. The second aim was to isolate improved strains of *E. coli* that resistant to RPP stress and continue to accumulate higher yields of correctly folded recombinant proteins. The third aim was to address the fundamental question of whether improved strains could be used as generic tools to enhance the production of different types of recombinant protein.

CHAPTER 2

Materials and Methods

Materials

Suppliers

Chemicals, reagents and media were from Sigma, BDH, Difco or Oxoid unless otherwise stated.

Growth media

Liquid and solid media were prepared by dissolving quantities specified below in distilled water and autoclaving at 121^oC and 6894 Pa for 15 minutes. Solutions and media used for growth were made in exceptionally cleaned glassware that had been rinsed thoroughly in distilled water to remove any traces of detergent. Other solutions such as antibiotics were sterilised by filtration through a sterile 0.2 µm filter (Millipore, England).

Solid media

For solid media, 11.2 g of nutrient agar were dissolved, by autoclaving in 400 ml of distilled water. The sterile molten agar was left to cool to 60^oC prior to addition of antibiotics and other supplements. For agar plates, approximately 20 ml of molten agar was poured into each Petri dish and allowed to solidify. Plates were dried for 20 minutes at 60^oC and stored at 4^oC for up to four weeks.

Liquid media for shake-flask experiments

Bacteria were aerobically grown in Lennox broth (LB) that contained 10 g Bacto-tryptone, 5 g yeast extract and 5 g NaCl dissolved in 1 L of distilled water. LB volume of 20 ml was dispensed in 100 ml shake flasks and autoclaved. Once cooled, this was supplemented

with glucose at 2% (w/v) final concentration in addition to the appropriate antibiotics. Glucose was added from a 40% (w/v) stock solution.

Liquid media used in batch and fed-batch fermentations

Two different models of fermenters were used for batch fermentations in this project; the Infors fermenter (3.6L) (article no. 26131) and the Biolafitte fermenter (20 L). For batch fermentations, the starting volume of media was 2 L or 2.8 L in a 3.6 L Infors fermenter and 15 L in the Biolafitte fermenter. For fed-batch fermentations, the starting volume of media was 1.5 L in the infors fermenter. Growth medium is based on LB, containing 0.1% silicone antifoam which was sterilised by autoclaving for 30 minutes at 121° C. Once cooled, this was supplemented with glucose at either 2% or 0.5% (w/v) final concentration for batch and fed-batch fermentations, respectively. The medium was also supplemented with the appropriate antibiotics. Prior to inoculation, the pH was adjusted to 6.3 by the addition of 5% HCl.

Antibiotics

Carbenicillin and chloramphenicol were added from freshly made stock solution of 100 mg/ml and 30 mg/ml, respectively. Carbenicillin or chloramphenicol were used to supplement liquid or solid media at a final concentration of 100 µg/ml or 30 µg/ml, respectively.

Inducer

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to induce protein expression from bacteria. This was added from a freshly made 0.1 M or a 0.5 M stock solution that had

been filter sterilised. The final concentration of IPTG used was varied according to the experiment.

Additives

Glucose was added from a 40% (w/v) stock solution, which had been sterilised by autoclaving for 10 minutes only. The amino acids serine, threonine and asparagine were added at a final concentration of 2% (w/v) by dissolving accurately weighed quantities in a minimal of distilled water with gentle heating before being filter sterilised using a 0.2 µm filter.

Buffers and solutions

Buffers for diluting bacterial cultures

Potassium phosphate buffer saline (PBS) (pH 7.4) was used to dilute bacterial cultures for accurate OD_{650} measurements and for plating onto agar plates (Sambrook *et al.*, 1989). PBS was prepared by dissolving 1 tablet (Oxoid) in 100 ml of distilled water and sterilised by autoclaving.

Buffers and solutions for electrophoresis of DNA

TBE was prepared as a 5× stock solution and was diluted five-fold to give a working solution. The 5× stock contained 0.445 M Tris (pH 8.0), 0.445 M boric acid and 0.01 M EDTA. Agarose solution contained 0.7% (w/v) agarose dissolved in TBE buffer supplemented with 250 ng/ml ethidium bromide. Sample buffer contained 0.025% (w/v) bromophenol blue, 10% (v/v) glycerol in 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA.

Buffers and solutions for SDS-PAGE

Ultrapure Protogel^{TM} (30% (w/v) acrylamide and 0.8% (w/v) NN'-methylenebisacrylamide) was used in the stacking and resolving gel. The resolving gel buffer was 0.75 M Tris HCl (pH 8.3) and the stacking gel buffer was 1.2 M Tris-HCl (pH 6.8). This was used to make 15% resolving gel and 5% stacking gels, respectively.

Sample buffer contained 2 g SDS (sodium dodecyl sulphate), 20 ml glycerol and 5 mg of bromophenol blue in 0.1× stacking gel buffer in a final volume of 92 ml. This was dispensed into 1 ml aliquots and 87 µl of β-mercaptoethanol was added immediately before use.

Stock electrode buffer (10×) contained 30 g Tris, 150 g of glycine per litre of distilled water. A working electrode buffer solution of a 1× concentration that contained 0.1% SDS was used.

Buffers and solutions for staining gels

Haem staining solution contained 0.113 g of 3,3',5,5'- tetramethylbenzidine dissolved in 75 ml methanol. Immediately before use, the stain solution was mixed with 175 ml of 0.25 M sodium acetate pH 5.0. Stop solution contained 175 ml 0.25 M sodium acetate pH 5.0 and 75 ml propan-2-ol.

Coomassie blue stain solution contained 0.2% (w/v) Coomassie brilliant blue in 50% (v/v) methanol and 10% (v/v) acetic acid in distilled water. This was filtered before use to remove any undissolved solid.

Fast destain solution contained 40% (v/v) methanol, 10% (v/v) acetic acid in distilled water. Shrink solution contained 48% (v/v) methanol, 2% (v/v) glycerol in distilled water.

Solutions for western blotting

TBS (Tris-buffered saline) (pH 7.6) was prepared as a $10 \times$ stock solution, which contained 24.2 g Tris base and 80 g NaCl in 1 L of distilled water. This was used in a $\times 1$ working concentration.

TBS/T (wash buffer) was composed of $1 \times$ TBS, 0.1% Tween-20.

Blocking buffer contained $1 \times$ TBS, 0.1 % Tween-20 with 5% (w/v) nonfat dry milk.

Bacterial strains and plasmids

Strains of *E. coli* and plasmids used in this work and their genotypes are listed in table 2.1. *E. coli* BL21(DE3)* was the parental strain used in the expression experiments for several recombinant proteins from pET vectors.

Bacteria were stored on nutrient agar plates at 4°C. For long term storage, bacteria were frozen at -80°C either in 15% glycerol-LB or as competent cells in a 15% glycerol 0.1 M CaCl$_2$ solution. Bacterial strains with and without plasmids were also frozen on Microbank™ beads with glycerol cryopreservative at -80°C. Single beads were used to inoculate 20 ml of sterile LB for overnight seed cultures.

Table 2.1: *E. coli* **strains and plasmids used in this work**

Strains or plasmids	Description	Reference or source
E. coli strain		
BL21*(DE3)	*F-ompT gal[dcm][ion]hsdsB rne131(rB -mB -;* an *E. coli* B strain); with DE3, a λ prophage carrying the T7 RNA polymerase gene	Invitrogen, GSK
C41	Derivative of BL21(DE3) with improved RPP properties of membrane protein production. Contains mutations in the *lac*UV5 promoter that controls expression of the T7 RNA polymerase	Miroux and Walker, 1996
C43	Derivative of BL21(DE3) with improved RPP properties	Miroux and Walker, 1996
P2	Derivative of BL21*(DE3) with improved production of recombinant CheY::GFP	This work
Mutant Pro	Derivative of BL21*(DE3) with improved production of recombinant cytochrome *c* peroxidase from *N. gonorrhoeae*	This work
JM109	*recA1, endA1, gyrA96, thi, hsdR17* (r_K^-, m_K^+), *relA1, supE44,* Δ(*lac-proAB*), [F', *traD36, proAB, lac^q ZΔM15*].	Promega
Plasmids		
pETCheYGFP	Derivative of pET20b series for the expression of the chemotaxis gene *cheY* from a T7 promoter, *bla*.	Jones *et al.*, 2004
pST203	Derivative of pET11c series, over-expresses the gonococcal *ccp* gene from a T7*lac* promoter, *bla*.	Turner *et al.*, 2003
pST2	Derivative of pACYC184 (an approximate copy number of 12). Constitutively expresses the *E. coli* cytochrome *c* maturation proteins CcmA-H from the upstream *ccmA* promoter, *cat*.	Turner *et al.*, 2003
pGEM® T-Easy	Commercial vector for cloning PCR products	Promega
pET20bhc	Derivative of pET20b series with additional restriction sites	Jones *et al.*, 2004

Methods

Recombinant protein production in aerated shake flasks

An overnight seed culture of bacteria that had been grown at 30°C for 12-14 hours was used to inoculate flasks containing LB-glucose medium at a final concentration of 2% (v/v). Bacteria were grown aerobically in linear shakers at 110 rpm. For the standard protocol, bacteria were grown at 37° to an OD_{650} of 0.5 before expression of the recombinant gene was induced with 0.5 mM IPTG. The growth temperature was decreased thereafter to 25°C. For the improved protocol, bacteria were grown at 25°C throughout, and the expression of the recombinant gene was induced with an optimised concentration of IPTG that had been empirically determined. For the production of CheYGFP and CCP, an IPTG concentration of 8 µM and 30 µM were used for induction, respectively. The amount of IPTG added was calculated relative to the volume of the remaining culture after sampling. Cultures were sampled for analysis immediately before induction and at intervals post-induction.

Recombinant protein production in batch and fed-batch fermentations

For batch fermentations, bacteria were grown in a 3.6 L Labfors fermenter (Infors) containing either 2 or 2.8 L of LB or in a 20 L Biolafitte fermenter containing 15 L of LB medium supplemented with 2% glucose. For fed-batch fermentations, bacteria were grown in the 3.6 L Labfors fermenter containing 1.5 L of LB medium supplemented with 0.5% glucose. The medium was also supplemented with the appropriate antibiotics and 0.1% silicone antifoam. Prior to inoculation, the sterile medium was aerated for 10 min and the oxygen probe was calibrated at 100%. The oxygen probe measures the dissolved oxygen

concentration (pO_2) throughout the fermentation. For batch fermentations using either the Infors or Biolafitte fermenters, the growth media was inoculated to 2% with a seed culture that had been grown aerobically at 30°C for 12-14 hours. Bacteria were grown aerobically with an aeration of 4 vvm (volume of air supplied per unit of liquid volume per minute) and a stirring speed of 700 or 900 rpm for batch and fed-batch fermentations, respectively. The pH of the culture was controlled at 6.3 with the automated addition of 5% HCl or 10% NH_4OH. The production of the recombinant protein was induced using either the standard or the improved protocols. Serine, threonine and asparagine were added 5 h post-induction at 1 mM final concentration. Feeding was commenced at approximately 8 h post-induction. The 1 L of feed contained 10× LB (without NaCl), 20% glucose, 10 mM each of serine, threonine and asparagine, 100 µg/ ml carbenicillin, 8 µM (for CheYGFP) or 0.5 mM IPTG (when using the improved or standard protocol, respectively), 1 ml of *E. coli* sulphur-free salts. 0.1% antifoam. An initial feed rate of 13.8 ml h^{-1} was used, followed by 20.6 ml h^{-1} 32.5 h post-induction, 27 ml h^{-1} 48.5 h post-induction and 39 ml h^{-1} 54.5 hours post-induction. The feed was depleted by 62 hours post-induction. Samples were independently withdrawn twice during each sampling point, the second sample being used for analysis. This is to ensure analysis of bacteria from the fermentation environment and not the residual culture remaining within the sampling tube.

Measuring optical density

The optical density of bacterial cultures was measured using a CamSpec spectrophotometer (M501 single beam scanning) at a wavelength of 650 nm. When required, cultures were diluted with PBS in 3 ml cuvettes such that the OD was between 0.2 and 0.5 units. The cuvettes were inverted to ensure proper mixing of the diluted culture.

Making calcium chloride competent bacteria

This method was used routinely to prepare *E. coli* for transformation with plasmids and ligation reaction products. A 20 ml culture of the desired strain was grown aerobically at $37^{o}C$ until the OD_{650} was between 0.45 and 0.6. Bacteria were incubated on ice for 20 minutes to stop further growth, then harvested in 10 ml test tubes at $5\ 000 \times g$ for 5 minutes. The pellet was resuspended in 5 ml of ice-cold 0.1 M calcium chloride and incubated on ice for 20 minutes. The centrifugation step was repeated and the pellet was resuspended in 1 ml of ice-cold 0.1 M calcium chloride and left on ice for at least 1 hour (Sambrook *et al*., 1989). For long-term storage, 15% (v/v) glycerol was added and bacteria stored at $-80^{o}C$.

Heat shock transformations of bacteria

Competent bacteria (20 µl) mixed with 2 µl of plasmid DNA (2 -5 ng/ µl) or 5 µl of a ligation reaction (containing 5ng DNA) were incubated on ice for at least 30 minutes. Bacteria were heat-shocked in a water bath at $42^{o}C$ for 2 minutes, supplemented with 0.5 ml LB and incubated aerobically at $30^{o}C$ for 1 hour. The recovered bacteria was then plated onto nutrient agar containing the appropriate antibiotics and incubated at $30^{o}C$ overnight.

Genetic manipulations

Polymerase Chain Reaction

The polymerase chain reaction was used to generate DNA fragments by amplification from a DNA template using specific primers. Single colonies of *E. coli*, plasmids or

chromosomal DNA were used as a template. Primers were synthesized by Alta Biosciences (University of Birmingham) and their melting temperatures were determined empirically.

DNA was amplified using a DNA polymerase enzyme from a High Fidelity PCR Super Mix ® (Invitrogen). The proof-reading of the exonuclease activity of this enzyme minimises chances of mutation to the PCR product. A denaturation temperature of 94°C and an extension temperature of 68°C was always used with this PCR supermix.

PCR reactions of 50 µl total volume contained 20 pmol of each primer, 1 µl of DNA template (10ng/ µl), distilled water and 45 µl of PCR supermix. A negative PCR control reaction with no DNA template was always included. The PCR cycle was started by initially denaturing the DNA template 94°C for 5 minutes; subsequent denaturation steps were of a duration of only 30 s. The reaction mixture was cooled to allow the primers to anneal to the target sites on the DNA. The annealing temperature was optimized empirically by setting a temperature gradient across the heating grid of the thermocycler, ranging from 50°C to 60°C. Synthesis of DNA by the DNA polymerase was initiated by increasing the extension temperature to 68°C. This was repeated for 25 cycles. The PCR was completed with a final extension step at 68°C for 10 minutes.

The PCR product(s) was resolved by electrophoresis and visualized by UV. PCR products were purified from the PCR supermix using a PCR purification kit (Qiagen).

Electrophoresis analysis of DNA

DNA fragments were analysed by electrophoresis on 0.7% agarose supplemented with 250 ng/ ml ethidium bromide. Molten agarose was poured onto a glass plate, and wells with a volume between 20-100 µl were created using a plastic comb. Once the gel had set, the comb was removed and the gel was placed in a 100 ml of TBE buffer in a horizontal tank. DNA samples were mixed at a ratio of 3:1 with sample buffer and loaded into the wells, and the gel was run at 100 volts. For analytical and preparation gels, the sample volumes loaded were 5 µl and 15-20 µl, respectively. DNA was visualized using a UV transilluminator at 366 nm.

Small-scale isolation of plasmid DNA (minipreps)

A single colony, freshly transformed with the plasmid to be isolated and purified, was used to inoculate 2 ml of LB supplemented with the appropriate antibiotic. The culture was grown aerobically by shaking at 25°C or 30°C, and the plasmid was extracted using a QIAprep Miniprep kit (Qiagen).

Cloning PCR products

To facilitate further cloning steps, PCR products were cloned into the commercially available vector pGEM®T-Easy (Promega).

Restriction digests

DNA inserts were digested from the pGEM T-Easy vector using the appropriate restriction enzymes (New England Biolabs). A 20 µl total reaction volume contained approximately 50 ng of pGEM T Easy vector, 1 µl (5 units) of each restriction enzyme (NEB), $\times$ 1 enzyme

buffer, BSA and sterile distilled water. Digests were incubated at 37$^{\text{o}}$C for 2 to 3 hours.

Digested DNA was resolved on a 0.7% agarose gel at 100 V for 1 hour, analysed using a

UV-transilluminator and purified from the gel using a Qiagen® kit.

Ligations

Ligation reactions were set up to join the digested DNA fragment and the digested DNA

vector. A total volume of 10 μl contained 50-100 ng of vector, 5 ng of insert, 1 μl T4 DNA

ligase, ligation buffer (NEB) in distilled water. Ligation reactions were incubated

overnight at 4°C overnight. The ligation reaction mix was used to transform competent *E.*

coli strain JM109.

Sequencing

DNA and primers at a concentration of 300 ng and 3.2 pmol, respectively, were used per

sequencing reaction. DNA was sequenced in the Functional Genomics Laboratory (School

of Biosciences, University of Birmingham) using the Plasmid-to-Profile service.

Sequencing of the *lac*UV5 promoter region from BL21*

The promoter region of *lac*UV5 in both wild type and P2 BL21* strains was amplified

using the lacUV5_KpnI_r and lacUV5_Nde_f primers listed in supplementary material of

Wagner *et al*. (2008). The melting temperature of the primers, which was found to be

55.1°C, was determined empirically using a gradient thermocycler. The PCR products

were purified from the PCR supermix and were immediately ligated into pGEM T-easy

vector by incubating overnight at room temperature. The ligation mix was transformed

into JM109 and plated onto NA supplemented with Xgal-IPTG (Sambrook *et al*., 1989).

Several candidates of positive transformants were purified first before the T-easy plasmid was isolated using a Qiagen miniprep kit. The presence of an insert in the T-easy vector was checked by digesting with *Eco*R1 before being resolved on an agarose gel. The *lac*UV5 fragment was then sequenced from the T-easy vector using T-easy forward and reverse universal primers.

Fractionation of *E. coli* using BugBuster®

Cells were lysed using BugBuster® extraction reagent, and soluble proteins were separated from insoluble proteins by centrifugation. Bacteria were pelleted from a 0.5 to 1 ml culture of a measured OD_{650}. The volume of BugBuster added was adjusted to the OD_{650} of the culture using the following equation.

$$\text{Volume of BugBuster per ml of culture} = \frac{OD_{650} \text{ of culture}}{0.015}$$

This achieves the same biomass concentration, thus allowing comparisons of protein concentration between bacterial samples. The pellet was gently resuspended and incubated at room temperature for 10 minutes whilst shaking gently to facilitate cell lysis. Inclusion bodies and cell debris were pelleted from the soluble protein fraction by centrifugation at 14000 g at $4^{o}C$ for 20 minutes. The resulting supernatant was retained in a separate tube and incubated on ice until further analysis. The pellet was washed in 160 μl potassium phosphate buffer (50 mM) to remove any contaminating soluble protein, and then centrifuged at 13000 g for 1 minute, and the supernatant discarded. The same volume of sample buffer (as that calculated for BugBuster) was added to both the insoluble fraction pellet and the soluble protein supernatant. Special care was taken to fully resuspend the

pellet containing the insoluble protein. Both fractions were boiled at 100$^{\text{o}}$C for 10 minutes before being loaded onto an SDS-PAGE. A 10 µl volume was loaded from the soluble fraction, and 5 µl were loaded from the insoluble fraction.

Analysis of recombinant protein accumulation by SDS-PAGE

Proteins were separated on 18 cm square SDS-PAGE gels using a 15% (w/v) polyacrylamide resolving gel and a 6% (w/v) stacking gel in a vertical electrophoresis tank. Ethanol-polished gel plates were separated by 0.75 mm spacers. The resolving gel contained 10 ml stock resolving gel buffer, 10 ml Ultrapure Protogel$^{\text{TM}}$ and 100 µl 20% (w/v) SDS. This was polymerized by the addition of 200 µl 80 mg/ ml ammonium persulphate (APS) and 10 µl N, N, N', N'-tetramethylethylenediamine (TEMED), and immediately poured between the assembled glass plates. Air was excluded from the separating gel by an overlay of 0.1% SDS. The resolving gel was allowed to polymerise for approximately 1 hour, after which the 0.1% overlay was poured off and replaced with the stacking gel. The stacking gel consisted of 1.5 ml stock stacking gel buffer, 3 ml 30% acrylamide stock, 10.5 ml water, 75 µl 20% SDS, and was polymerised with 150 µl APS and 7.5 µl TEMED. A comb was inserted into the stacking gel to create wells with a volume of approximately 60 µl. After polymerisation of the stacking gel, the wells were washed twice in distilled water before (whole) cell samples from the same biomass concentration were loaded (Equation. 1). The samples were heated to 100$^{\text{o}}$C for 10 minutes before loading. The volume of processed samples loaded was 5 µl and 20 µl when CheYGFP or CCP were produced, respectively. Protein bands were resolved at 160 volts in 1$\times$ Tris-glycine for approximately 1 h. Protein that was to be stained by haem dependent peroxidase activity (afterwards) was resolved overnight at 60 volts.

Proteins containing covalently attached haem groups were detected using a haem-dependent peroxidase activity stain as described by Thomas *et al*. (1976). Proteins were fixed by incubating the gel in freshly prepared 10% TCA (trichloroacetic acid) solution for 30 minutes. The stain solution was poured over the gel immediately before a 2 hour incubation in the dark. A volume of 0.3 ml hydrogen peroxide (98%) was then added, the solution was mixed and the stain allowed to develop. Proteins with covalently attached haem show up as blue/green bands on the gel. After the stain had developed, the stain solution was poured off and replaced with the stop solution to stop further reaction. The gel was scanned before being stained with Coomassie blue to reveal other non-haem cellular proteins.

Total protein was detected by staining the gel in Coomassie blue for 30 minutes at room temperature with gentle agitation. Background staining was removed by washing the gel twice in fast destain solution for 45 minutes each. The yield of total, soluble and insoluble recombinant protein accumulated was quantified from SDS-PAGE gels by densitometry using the volume analysis tool of the Quantity one® software (Bio-rad). The gel was then incubated for 1 hour in shrink solution prior to drying between two sheets of drying film.

Western blotting analysis

For Western analysis, culture samples were resuspended in sample buffer as described above (SDS-PAGE section) and loaded onto NuPAGE 4-12% Bis-Tris mini (8×8 cm) gel (Invitrogen). The NuPage electrode buffer ($\times 1$) was used to resolve proteins at 150 V until the blue dye reached the bottom of the gel. The gel was immersed in transfer buffer for 10 minutes. The NuPage transfer buffer was prepared according to manufacturers'

instructions and contained 10% methanol. A piece of Hybond-ECL nitrocellulose membrane (Amersham) of the same size as the gel was immersed in distilled water for 5 minutes then in transfer buffer for another 5 minutes. The membrane was handled with clean forceps at all times. Eight pieces of filter paper and blotting pads were also soaked in transfer buffer for 5 minutes. The Western transfer was set up in the Xcell II blot module (Invitrogen) as per manufacturer's instructions, such that the gel is closest to the cathode plate. Air bubbles were removed by rolling a 5 ml pipette tip over the surface of each layer. Care was taken to ensure that the number of blotting pads used were enough to ensure a snug fit into the blot module. Transfer buffer was poured inside the blot module till the top. Proteins were transferred at 30 V for approximately 2 hours at room temperature.

Following transfer, the nitrocellulose membrane was washed twice with 25 ml of $(1 \times)$ TBS for 5 minutes each at room temperature with vigorous shaking. The membrane was incubated in 25 ml of blocking buffer for 1 hour at room temperature. The membrane was then washed 3 times for 5 min each with approximately 25 ml of TBS/T with vigorous shaking. The membrane blots were incubated with T7 RNA polymerase monoclonal antibody (Novagen) at a dilution of 10^{-4} in 10 ml blocking buffer (in a sealed bag) with gentle agitation overnight at room temperature. The membrane was rinsed quickly TBS/T followed by 3 cycles of washing for 5 min each with about 25 ml of TBS/T with vigorous shaking. The membrane was incubated with a peroxidase-conjugated anti-mouse IgG (Amersham) in 10 ml blocking buffer at a dilution of 1 in 2500 in a sealed bag. This was gently agitated for 1 hour at room temperature. Again, the membrane was then washed in the same way as after the incubation with the primary antibody. The blots were developed

using EZ-ECL Chemiluminescence detection kit (Biological Industries) according to the provided protocol. The membrane was drained from excess TBS/T by holding it with a pair of tweezers and touching the edge against a tissue. The membrane was then placed (protein side up) on a clean piece of cling film, and a mix of 5 ml each of the detection reagents A and B was pipetted onto the surface and left for 5 min at room temperature. The membrane was wrapped very carefully with a clean piece of cling film, ensuring that all air bubbles were excluded. The bands were exposed on Hyperfilm ECL (Amersham) with varying times of exposure and the film developed using the X-ograph machine.

Fluorescence measurements

Fluorescence of GFP fused to the recombinant protein was exploited as an indicator for soluble protein production. Fluorescence was measured by a Perkin-Elmer benchtop fluorimeter in the University of Birmingham. The same dilutions of culture samples that were used for optical density measurements were also used to measure fluorescence. The fluorimeter was used with an excitation wavelength of 485 nm and emission wavelength of 510 nm.

Fluorescence of BL21* producing CheYGFP was measured at GSK using a plate reader. The same dilutions that were used for optical density measurements were also used to measure fluorescence.

Estimating yields of fluorescent CheYGFP

The concentration of fluorescent CheYGFP was determined from a standard purified solution of CheYGFP of 3.5 mg/ ml of a 94% purity (kindly provided by GSK).

Fluorescence was measured from serial dilutions of the standard (with PBS) using a Perkin-Elmer fluorimeter (model 203 – serial number 44272-7) at an excitation wavelength of 485 nm and an emission wavelength of 509 nm. The fluorescence of serial dilutions of the standard solution was measured from 3 ml polystyrene cuvettes ($10 \times 10 \times 45$ mm). A linear regression analysis from 2 independent calibration experiments revealed that a 1 mg/ ml concentration of CheYGFP corresponds to 2790 fluorescence units (**Figure 2.1**).

Attempts at calibrating the plate reader at GSK using serial dilutions of a purified standard solution of CheYGFP of 1.9 mg/ ml were unsuccessful due to the denaturation of the protein.

Determining bacterial dry cell weight

The dry cell weight (DCW) of bacteria after prolonged RPP using fed-batch fermentations was determined. A culture sample of 5 ml was diluted 30-fold using PBS that has been diluted 5-fold with sterile distilled water. This is necessary to dilute out the solutes and the dissolved components that are present in the growth medium, and allow accurate measurements of DCW. The diluted culture was mixed thoroughly. At least 10 ml of this diluted culture was centrifuged in a clean test tube for 5 minutes at 10, 000 rpm to pellet out the cells. The supernatant was carefully removed into a separate test tube for further analysis. Using forceps, small clean glass vials that were previously labelled with a marker pen, were weighed to the fourth decimal place and the weight recorded. A 1 ml sample from the diluted bacterial culture was carefully dispensed in to 10 separate glass vials. Also, a 1 ml sample from the diluted supernatant was carefully dispensed into 10 separate glass vials. The glass vials were carefully placed in a metal basket and incubated in a
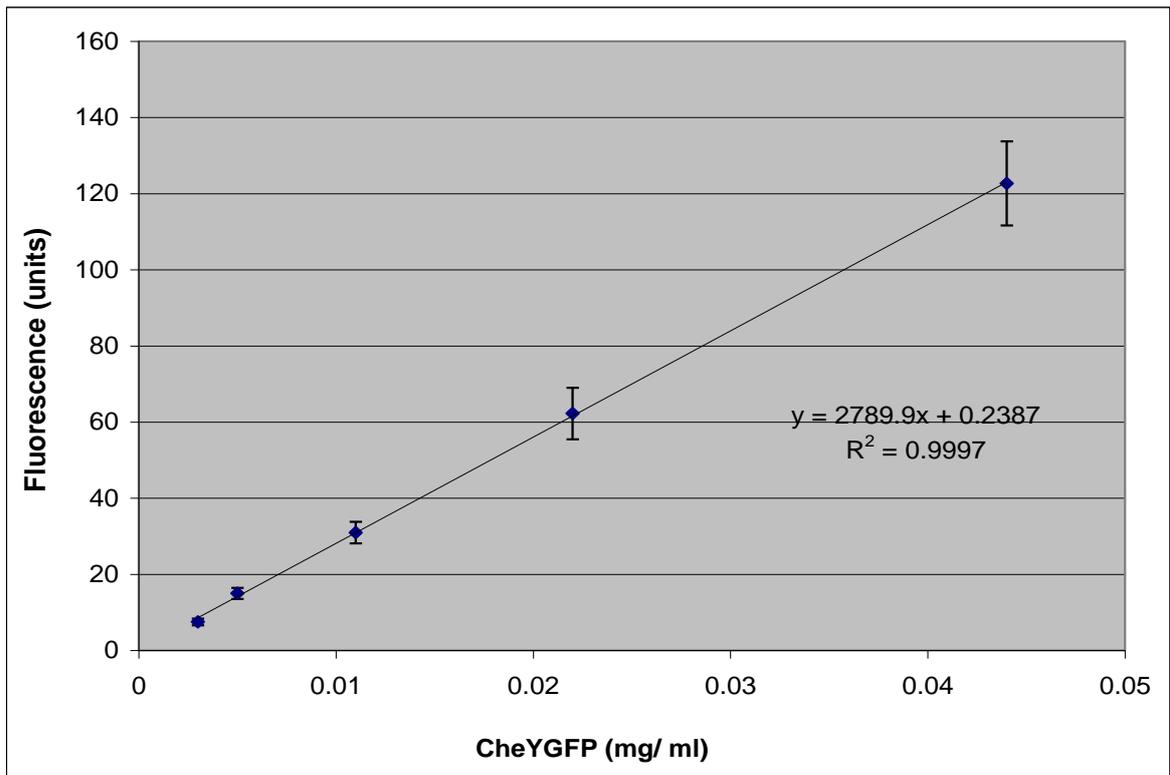
Figure 2.1. **Calibrating CheYGFP fluorescence**. The fluorescence of serially diluted CheYGFP standard solution of a concentration of 3.5 mg/ ml was measured in a Perkin-Elmer fluorimeter.   From the linear regression, it was revealed that a CheYGFP concentration of 1mg/ ml corresponds to 2790 fluorescence units. The calibration curve was generated from serial dilutions of two separate stocks of the CheYGFP standard solution. Standard error bars from three independent calibration experiments are presented.

sterilising oven overnight. After incubation, the glass vials were immediately placed in a desiccator containing blue silica desiccant, and left for a few hours to cool to room temperature. This is to prevent any condensation on the glass vials that might alter their weight. Once cooled, and using forceps, the vials were weighed. The weight of the dried culture and supernatant samples was determined from the difference from the pre-weighed vials, and an average value for 10 flasks was calculated. Finally, the difference between the weight of the culture samples and the supernatant corresponded to the bacterial dry cell weight in the starting culture.

Single-cell analysis using flow cytometry

This technique has been routinely used to analyse the different physiological states of bacteria during a fermentation experiment. Cells can be analysed in terms of their intrinsic fluorescence such as the production of a GFP-fusion protein, or by their uptake of fluorescent stains. Propidium iodide and bis-oxonol are red and green fluorescent stains, which are used to stain dead and stressed cells, respectively. The number and proportion of bacterial populations with different fluorescence properties could be identified (**Figure 2.2**).

Bacteria were analysed within 5 to 10 minutes of sampling. Bacteria were suspended in filtered PBS at a dilution of 10^{-3} or 10^{-4} depending on the density of the culture. Fluorescence stains were added at a final concentration of 1 mg /ml, and the samples were analysed. The flow cell was washed with filtered PBS to in between samples to prevent cross contamination of samples.
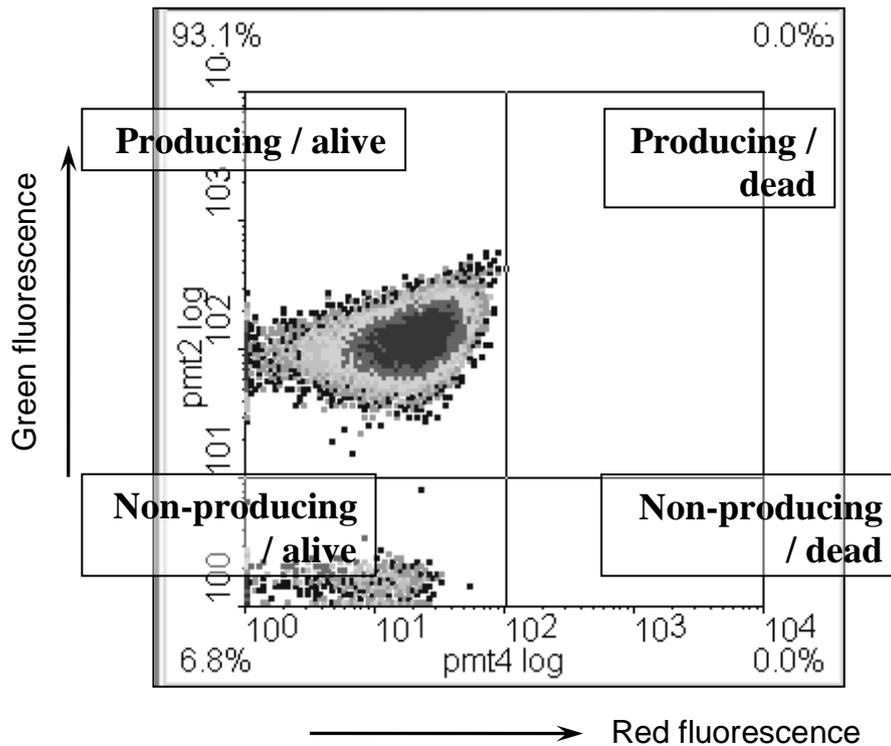
Figure 2.2: **Typical output from a flow cytometer**. The bacteria being analysed contain a recombinant plasmid from which production of the target green fluorescent protein has been induced. The green fluorescence emitted from the bacteria, reflecting the accumulation of the green fluorescent target protein, is measured by photomultiplier tube 2 (PMT2). The cells in the culture have also been stained with the red fluorescent dye propidium iodide, which specifically stains dead cells. The red fluorescence emitted by bacteria, reflecting dead cells, is measured by photomultiplier tube 4 (PMT4). Thus from this flow cytometer output, it is clear that a large bacterial population (93.1%) is emitting high levels of green fluorescence and not red fluorescence, indicating that it had accumulated the green fluorescent target protein and is still alive. However, a small population (6.8%) has basal levels of both green and red fluorescence, indicating that the green fluorescent target protein has not been accumulated and that the cells are still alive. The location of other possible bacterial populations with different physiological states has been annotated. (*Figure adapted from Dr Sevastsyanovich*).

74

To prepare an ethanol killed control, bacterial cultures were mixed with 100% ethanol in a 1 to 1 ratio, and incubated at room temperature for 5 minutes. The bacterial suspension was centrifuged to pellet out the cells. The pellet was re-suspended in 100% ethanol and incubated again for 5 minutes. Bacteria were pelleted again by centrifugation and re-suspended in 100 µl distilled water.

CHAPTER 3

Production of the CheYGFP

recombinant protein in *E. coli* BL21(DE3)*

The production of soluble CheY as a recombinant protein fused to green fluorescent protein (GFP) has been exploited industrially as a model protein to understand RPP in *E. coli* within batch and fed-batch fermentations (Jones, 2007). More importantly, however, the development of novel strategies that allow the production of the soluble and high quality form of recombinant proteins is the major concern. Factors affecting bacterial growth such as temperature, aeration, media and supplements used during the overproduction of CheYGFP were investigated in an attempt to improve yields of the soluble protein. It was reported that a GFP tag fused to the C-terminus of the protein reliably reported the production of the soluble recombinant protein in a fermentation process. The increased fluorescence emitted by the recombinant protein was therefore used as an indicator of soluble protein accumulation. In contrast, an N-terminal GFP tag failed to report the production of correctly-folded CheY (Jones, 2007). In this project, the production of CheYGFP was first attempted in shake-flasks and batch fermentation using the optimised conditions as defined by Jones (2007). Bacterial physiology, in addition to the yields of soluble CheYGFP produced, were subsequently analysed throughout the production process.

Production of soluble CheYGFP in shake-flasks using the 'standard protocol'

Recombinant protein production (RPP) in industry is often based on the T7 expression system in *E. coli* BL21*. In this system, the recombinant gene is expressed from a pET vector by T7 RNA polymerase, which is chromosomally encoded and regulated by an IPTG-inducible promoter. CheYGFP was expressed from a pET20 vector. The overproduction of CheYGFP using the standard protocol was attempted and the physiological state of the bacteria during protein accumulation was investigated.

BL21(DE3)*pETCheYGFP was subcultured from an overnight inoculum into fresh medium supplemented with glucose and carbenicillin. Bacteria were grown aerobically at 37°C to an OD_{650} of approximately 0.5, and *cheY::gfp* expression was induced by the addition of 0.5 mM IPTG. The growth temperature was decreased to 25°C thereafter to increase the yield of soluble recombinant protein. Samples were taken before induction and hourly up to 12 hours post-induction and at 27.5 hours post-induction to assay growth, fluorescence, plasmid retention and protein accumulation.

The biomass concentration continued to increase post-induction, reaching an OD_{650} of 14.2, 27 hours post-induction (**Figure 3.1 A**). Similarly the total fluorescence of the culture increased, reaching an average of 560 units 27 hours post-induction, indicating an accumulation of CheYGFP within an increased culture density. However, the specific fluorescence, defined as the fluorescence per OD unit of culture, rapidly decreased from 180 immediately after induction to only 40 units 27 hours post-induction, suggesting that soluble CheYGFP protein was not being accumulated within a large proportion of cells (**Figure 3.1 B**).

Culture samples taken throughout the experiment were serially diluted and 0.1 ml from suitable dilutions was plated onto non-selective nutrient agar (NA) to estimate the number of colony forming units (cfu) per ml. After overnight incubation, the phenotype of the colonies was recorded, and the colonies were replica plated onto selective agar supplemented with carbenicillin to assess plasmid retention. The total number of cfu decreased sharply from 0.72×10^8/ ml immediately post-induction to less than 0.07×10^8/ ml 7 h post-induction, and had increased to only 0.62×10^8/ ml at 12 hours after induction
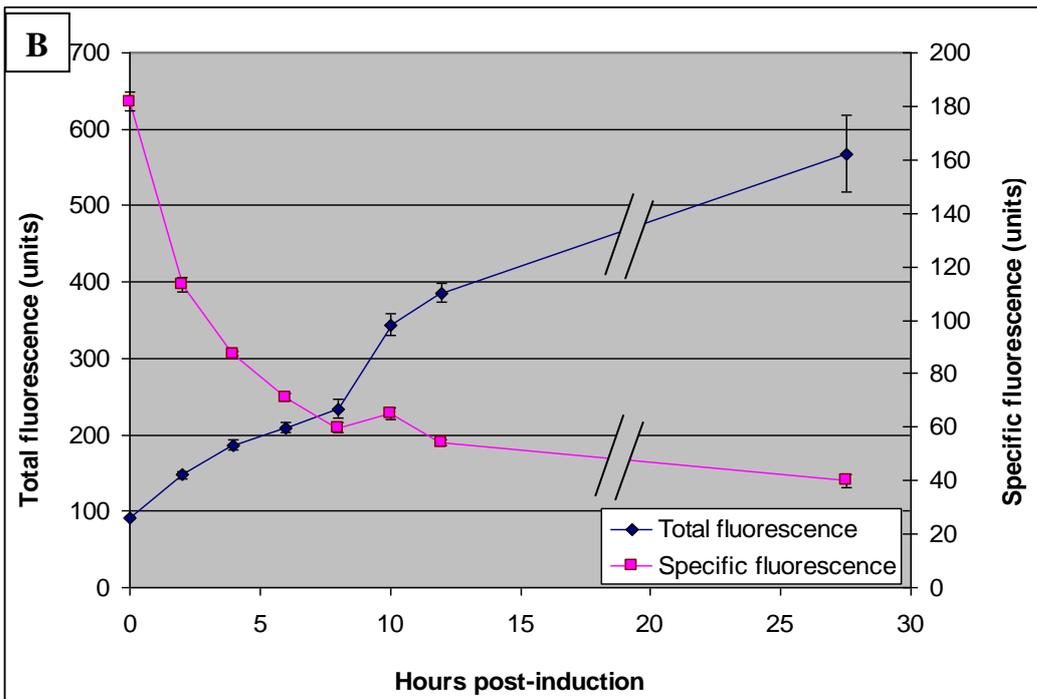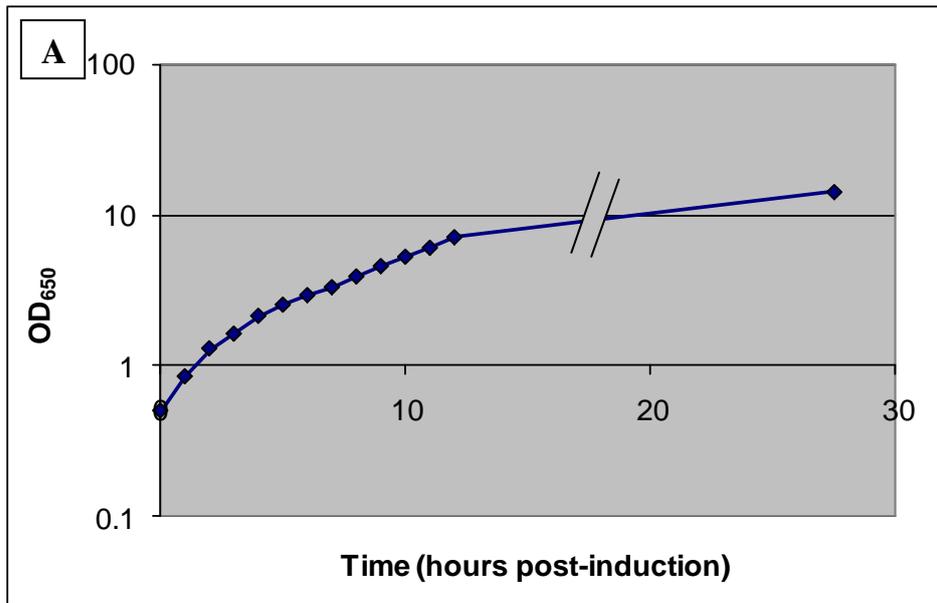
Figure 3.1: The production of CheYGFP in BL21* using the standard protocol. A)
The optical density at 650 nm of the culture was measured at intervals post-induction. B)
The total fluorescence of the culture and its corresponding specific fluorescence, shown in
arbitary units, were measured until 27 hours post-induction using a calibrated fluorimeter.
Standard error bars for triplicate repeats from an experiment using 3 different restreaked
colonies are shown.

(**Figure 3.2 A**). The loss of cfu post-induction varied between 1 and 2 orders of magnitude in different experiments. The number of colony forming units had recovered 27 hours post-induction, reaching 84.3×10^8/ ml.

All colonies formed on nutrient agar were phenotypically green immediately before and after induction (**Figure 3.2 B**). This is probably due to the basal level of expression of fluorescent CheYGFP from the pET vector used, which is also characteristic of this expression system. The majority of colonies, however, appeared white after only 5 hours post-induction (when low serial dilutions were plated), and remained so until 27 hours post-induction. This was readily shown by replica plating to be due to the overgrowth of plasmid-free bacteria in the culture. The distinction between fluorescent and non-fluorescent colonies was therefore exploited as a scorable phenotype to distinguish between plasmid-retaining and plasmid-free bacteria, and therefore provided a method for differentiating between bacteria that have produced CheYGFP and those that have not.

Total protein from whole cell samples that were normalised for cell biomass was resolved by SDS-PAGE and stained with Coomassie blue. A very prominent band at 42 kDa corresponding to CheYGFP was present before induction. This significantly increased in amount soon after induction, reflecting a rapid burst of CheYGFP synthesis, but there was little increase after a further 2 h (**Figure 3.3 A**). The yield of CheYGFP per unit of biomass had decreased significantly by 27 h post-induction. Bacteria were lysed chemically and the soluble and insoluble fractions were separated, after which proteins from within these fractions were resolved by SDS-PAGE. CheYGFP was detected before and immediately after induction, mostly in the soluble fraction due to the leaky expression
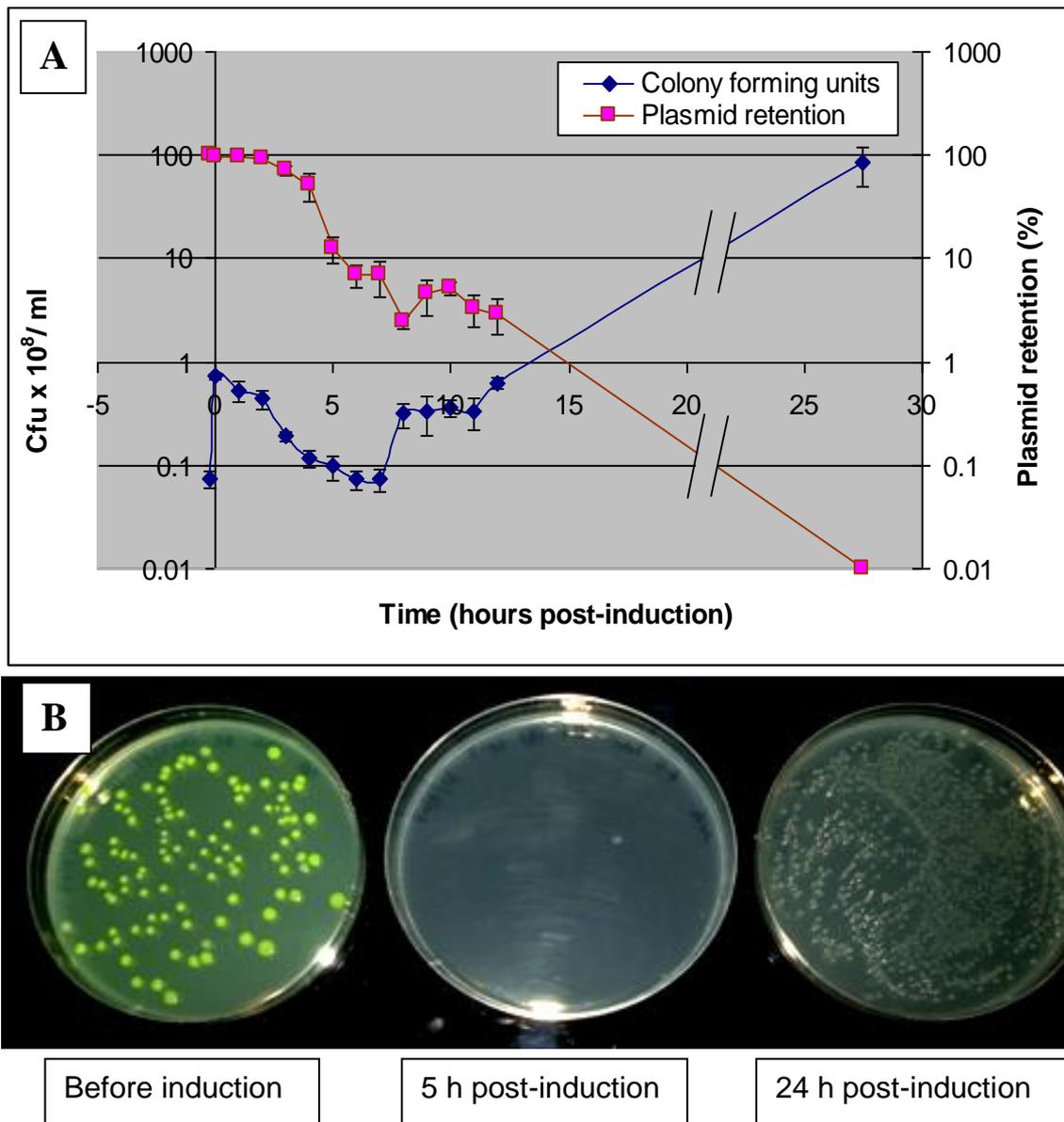
Figure 3.2: The effect of induction of BL21*CheYGFP using the standard protocol on colony forming units and plasmid retention. A) Serial dilutions of culture samples were plated on to non-selective nutrient agar and the cfu/ ml were calculated. Colonies formed were replica plated on to nutrient agar containing carbenicillin and the % plasmid retention was calculated. Standard error from triplicate experiments completed within one time series using 3 different colonies is shown. B) The typical phenotype of the colony forming units recovered on nutrient agar before induction, 5 h and 27 h post-induction are shown. Samples of the same serial dilution were plated for comparison.
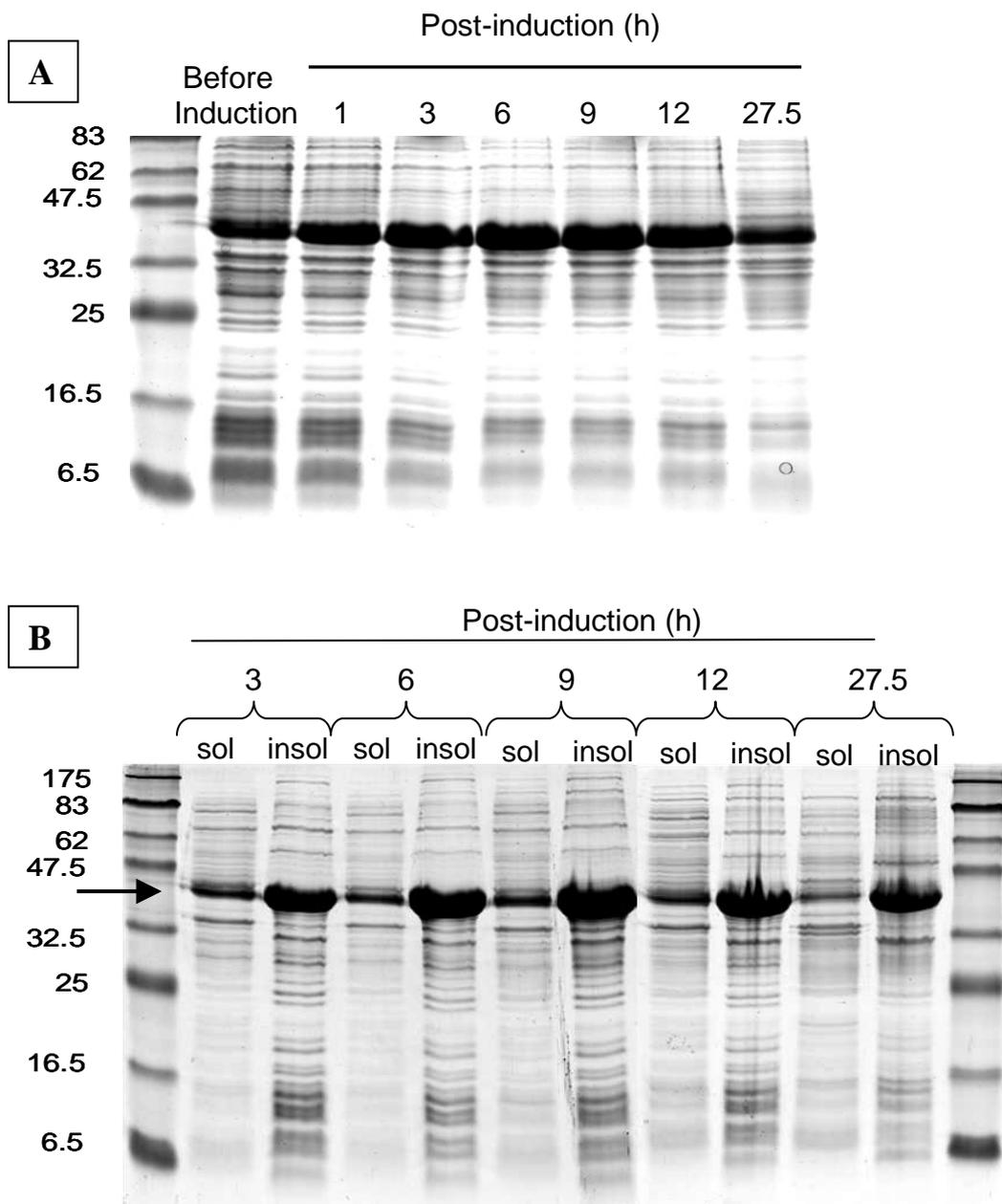
Figure 3.3: SDS-PAGE analysis of CheYGFP accumulation in BL21* induced using the standard protocol. A) Total proteins from whole cell samples before induction and at intervals post-induction were resolved by SDS-PAGE and stained with Coomassie blue to reveal the total yield of CheYGFP accumulated (indicated by arrow). B) Protein from soluble (sol) and insoluble (insol) cell fractions at intervals post-induction were resolved by SDS-PAGE to reveal proportion of soluble to insoluble CheYGFP accumulated.

from the pET vector (**Figure 3.3 B**). However, CheYGFP accumulated mainly in the insoluble fraction at all intervals after induction, indicating that the majority of the recombinant protein had accumulated in insoluble inclusion bodies. The soluble and insoluble cell fractions were separately analysed for fluorescence, and compared to a non-fractionated whole cell sample. It was revealed that at least 83% of the fluorescence was emitted from the soluble fraction (**Figure 3.4**). This indicates that GFP was mainly incorrectly folded in insoluble inclusion bodies.

CheYGFP production using the standard protocol in a bench-top fermenter

The overproduction of CheYGFP in BL21* using the standard protocol was repeated in a pH-controlled batch fermentation and all results obtained from shake-flask experiments were reproduced in batch fermentation. The dissolved oxygen in the culture was monitored throughout the fermentation using an oxygen probe (**Figure 3.5**). Dissolved oxygen decreased immediately after induction but increased again after 1.5 hours of induction (**Figure 3.5**). This indicated that the rate of metabolism and perhaps the rate of protein synthesis had increased upon induction. However, the dissolved oxygen steadily decreased from around 10 hours post-induction until it was approximately 50% of the starting concentration 24 h post-induction (**Figure 3.5**). The increased oxygen consumption suggests an increase in bacterial growth rate and biomass at around 10 hours post-induction. It is worth noting that the aeration of 4 vvm used was the highest possible that could be supplied by the air pump available, and hence, it was not possible to increase aeration further.
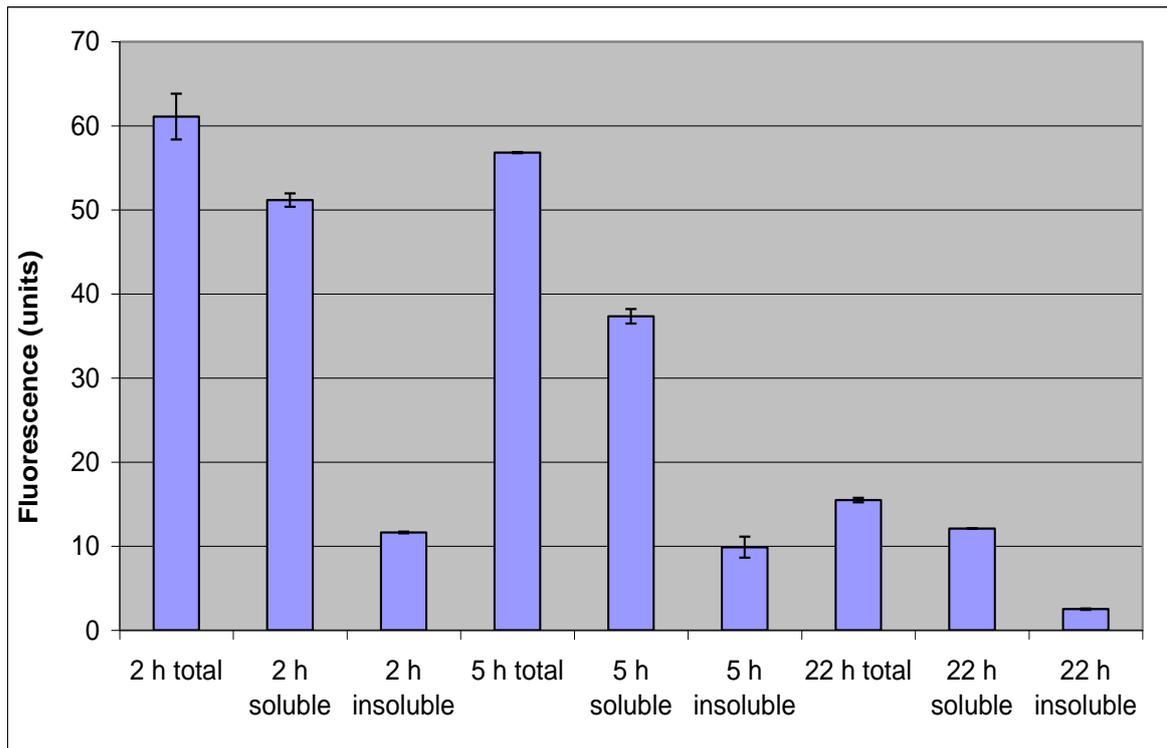
Figure 3.4: Fluorescence of total, soluble and insoluble fractions of BL21*CheYGFP post-induction with the standard protocol. Production of CheYGFP was induced in BL21* using the standard protocol in a batch fermentation in which the pH was controlled at 6.3. The total fluorescence of the culture 2 h, 5 h and 22 hours post-induction was measured. Cells from these samples were chemically lysed using BugBuster® reagent and the soluble and insoluble cell fractions were separated. The GFP fluorescence of both fractions was measured. Results were reproduced during the analysis of different samples from other batch fermentations and shake-flask experiments. Standard error bars from independently processed triplicate samples of bacteria induced using the standard protocol in a fermenter are presented.
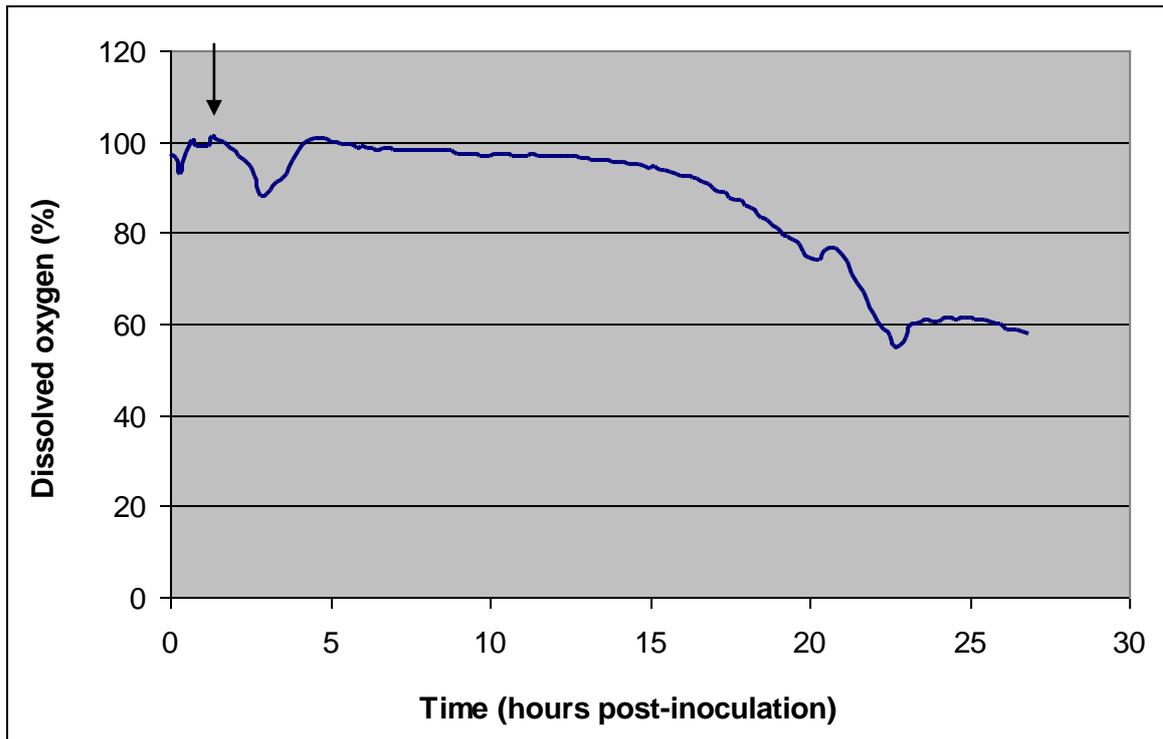
Figure 3.5: The dissolved oxygen concentration during production of CheYGFP with the standard protocol in a benchtop fermenter. Strain BL21*CheYGFP was grown and induced with the standard protocol in a fermenter. The dissolved oxygen concentration post-inoculation was measured using an oxygen probe. The oxygen probe was calibrated at 100% before inoculation and the oxygen consumption was monitored throughout the fermentation. The point of induction is indicated by an arrow.

The physiology and population dynamics of bacteria post-induction were also analysed by flow cytometry. This revealed the number of different populations as depicted by the light scatter and the proportion of cells overproducing fluorescent recombinant protein as shown by the fluorescence plot (**Figure 3.6**). Culture samples taken before induction and at intervals post-induction were analysed. The red-fluorescent stain, propidium iodide (PI), was used to identify dead cells, as measured by the pmt4 axis. An ethanol-killed control of non-fluorescent BL21* cells and stained with PI produced 90% PI$^+$ cells (**Figure 3.6 A**). Before induction, the majority of the cell population (94.5%) were highly fluorescent did not take up propidium iodide (**Figure 3.6 B**). This indicates that these cells were alive and contained high levels of CheYGFP due to the expression from a leaky pET promoter. After 5 hours of induction, the formation of a non-fluorescent bacterial population of 8.7% was detected from both light scatter and fluorescence plots (**Figure 3.6 C**). Cells of this non-fluorescent population were not stained positive for PI, and were therefore considered alive. The proportion of the non-fluorescent different shaped population continued to increase, as depicted by both light scatter and fluorescence plots, until 8 hours after induction, when it became almost equal to the original fluorescent population (**Figure 3.6 D**). By 24 hours post-induction, two main bacterial populations were present in the culture: a non-fluorescent majority of 95%; and a minor fluorescent population of 3.9% (**Figure 3.6 E**). There was no uptake of PI by either population, both of which were therefore alive. Consistent with plasmid loss data, these results indicate that the productive bacteria had become almost completely overgrown by plasmid-free non-productive variants. This indicates physiological stress on the bacterial host during the overproduction of CheYGFP.

Figure 3.6: Flow cytometry analysis of BL21(DE3)*CheYGFP during the production of CheYGFP using the standard protocol. Culture samples before and at intervals post-induction were analysed by flow cytometry to reveal the number of different populations present, and their proportions, as depicted by light scatter and their corresponding fluorescence profiles. Light scatter analysis combines data from the forward scatter (FS) and side scatter (SS) profiles of individual cells to identify different bacterial sub-populations. The forward and side scatter were detected separately using FS and SS (or photomultiplier tube 1 (PMT1) detectors, respectively. Whereas forward scatter is roughly proportional to the cell size, the side scatter is a function of cell morphology and internal complexity. Fluorescence, a second parameter for analysis, was also used to distinguish between fluorescent and non-fluorescent bacterial populations. Green fluorescence due to the accumulation of CheYGFP was detected by photomultiplier tube 2 (PMT2). The red-fluorescent dye propidium iodide (PI) was used to stain dead cells in the samples, and the red fluorescence was detected by PMT4. An ethanol-killed sample taken before induction and stained with PI is shown as a control for PI function. The proportions of each subpopulation within separated quadrants were calculated as percentages from a total of 25000 events.

A) Ethanol-killed control
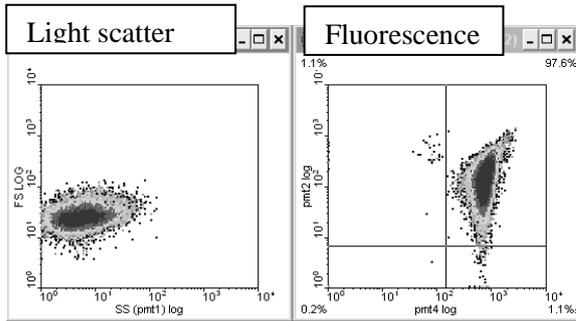
B) Before induction
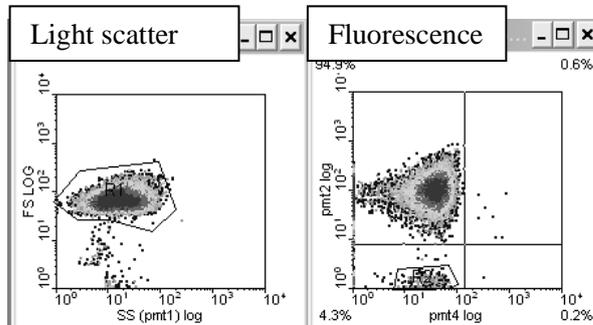
C) 5 h post-induction
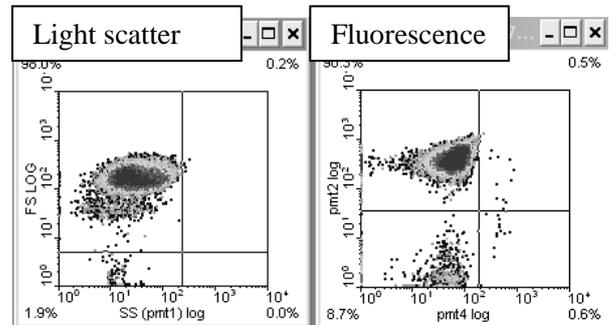
D) 8 h post-induction
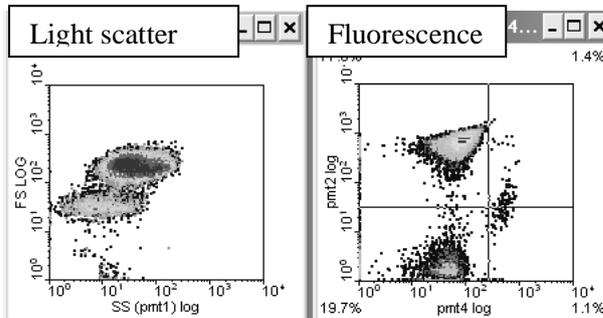
E) 24 h post-induction
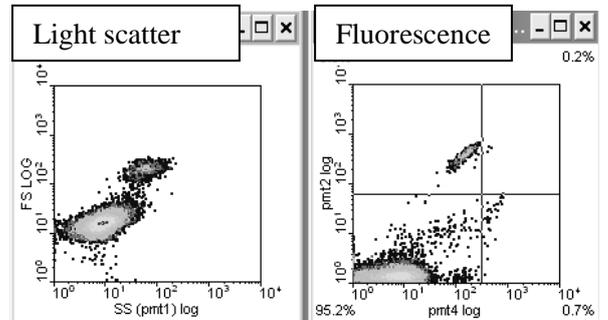
In conclusion, the standard protocol for the production of CheYGFP in BL21(DE3)* introduces a significant stress factor that results in the accumulation of mainly insoluble protein and the overgrowth of plasmid-free non-productive bacteria by 24 hours post-induction.

The effect of avoiding a decrease in temperature post-induction on accumulation of soluble CheYGFP

One factor that might have resulted in the deleterious effects on BL21(DE3)* during the production of CheYGFP under the standard conditions is the temperature decrease immediately post-induction. Although a lower temperature can help increase yields of soluble recombinant protein, the sudden decrease in temperature induces a cold-shock effect on bacteria that involves the general stress response. In an attempt to investigate this, the production of CheYGFP was induced as previously with 0.5 mM IPTG, in a benchtop fermenter, but the bacteria were grown at 25°C before and after induction, so that the sudden temperature change upon induction was avoided.

Culture samples taken before and at intervals post-induction were serially diluted and plated onto non-selective agar and the cfu/ ml in the culture was calculated. There was a 20-fold decrease in colony formation from 2.6 x 10^8/ ml before induction to 0.13 x 10^8/ ml 7 h post-induction. The number of colony forming units had recovered 24 hours post-induction with 24 x 10^8/ ml white colonies, of which, when replica plated, only 0.67% grew with carbenicillin selection. These results were mirrored by data from flow cytometry, revealing that only 7.5% of the bacteria had overproduced CheYGFP 24 h post-induction.

This suggests that the loss of colony formation is not solely due to a cold-shock effect post-induction and that the source of bacterial stress still remains.

Total proteins from whole-cell samples taken up to 7 hours post-induction were resolved by SDS-PAGE. Similar to previously, there was a rapid burst of relatively high yields of CheYGFP soon after induction, with little further accumulation 7 hours post-induction (**Figure 3.7 A**). There was also a marked decrease in the yield of CheYGFP by 24 hours post-induction (not shown). Bacteria were chemically lysed to separate soluble and insoluble fractions, and protein within these fractions was resolved by SDS-PAGE (**Figure 3.7 B**). CheYGFP was present in the soluble fraction before and immediately post-induction, reflecting the advantage of the slow expression from the leaky pET over-expression system at 25°C for solubility. However, the majority of CheYGFP was present in the insoluble fraction post-induction, indicating that the protein continues to accumulate mainly in inclusion bodies, even though there was slightly more soluble protein produced relative to previously by 7 hours post-induction. Nevertheless, despite decreasing the growth kinetics to minimise possible bottlenecks of protein folding, the quality of CheYGFP accumulated remained very poor. In conclusion, the immediate decrease of bacterial growth temperature post-induction was not the main source of stress on bacterial physiology that limited the recombinant protein quality.
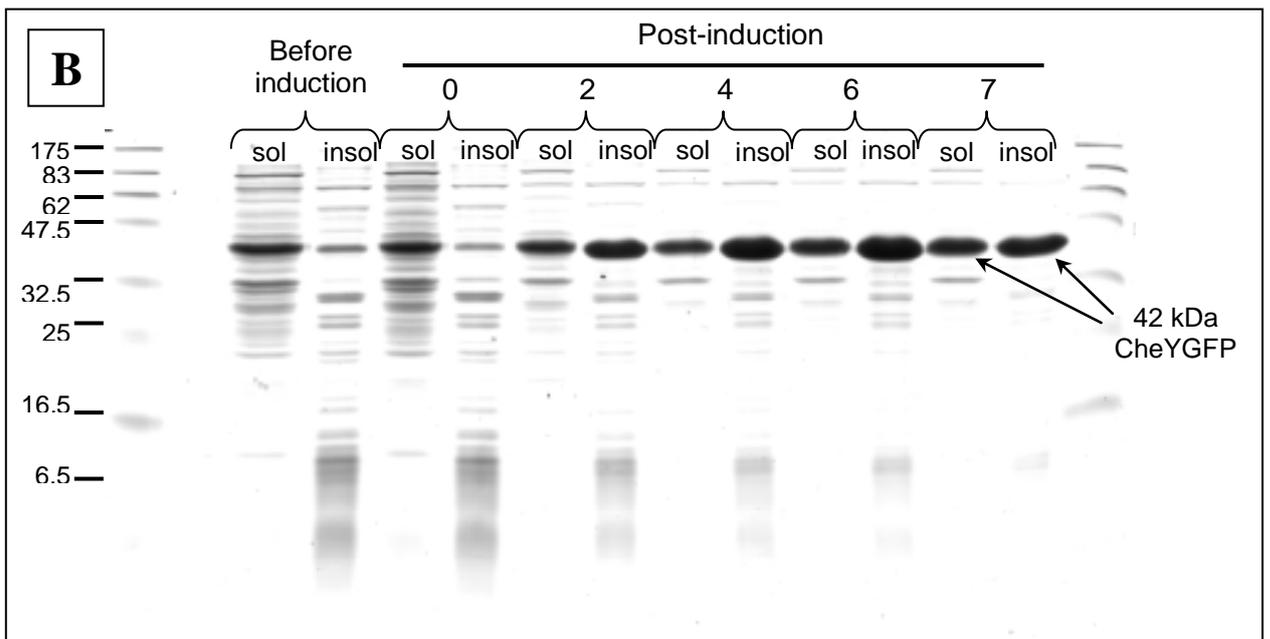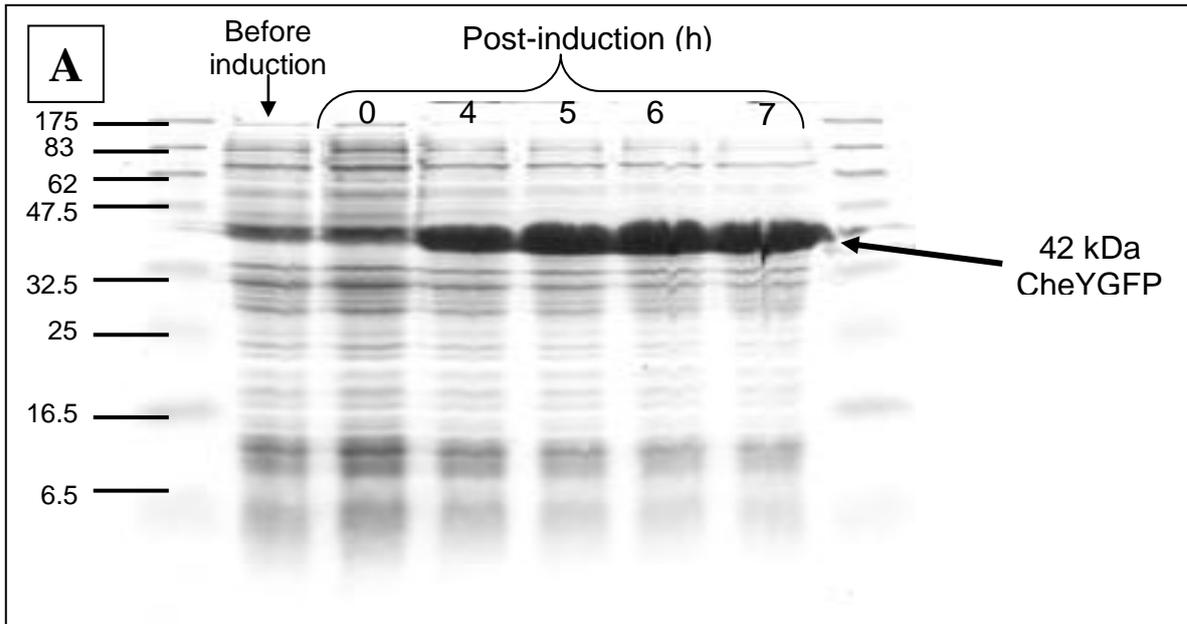
Figure 3.7: SDS-PAGE analysis of CheYGFP accumulation after induction with 0.5 mM IPTG at 25°C. **A)** Total protein from whole-cell samples before induction and at intervals post-induction. **B)** Protein from the soluble (sol) and insoluble (insol) fractions from cells before induction and at intervals post-induction were also resolved by SDS-PAGE.

The physiological source of stress on BL21* during CheYGFP production using the T7 expression system

It is evident, from the changes in bacterial physiology, that the standard protocol for induction introduces a huge source of stress on the BL21* host during the over-expression of CheYGFP. The sudden decrease in temperature post-induction had only a slight negative effect on the yield of soluble CheYGFP and does not account for the significant stress effect on the bacterial host that is manifested physiologically in the loss of cell culturability. It was therefore imperative to discover the main source of stress that remains associated with the standard protocol.

The over-expression of *cheY::gfp* is achieved by the use of the T7 expression system in *E. coli*. This involves a two-step process in which the chromosomally-encoded T7 RNA polymerase is first produced, which in turn transcribes the recombinant CheYGFP gene from the T7 promoter on the pET vector. This creates a cascade that ultimately results in high production yields of the recombinant protein. It was important therefore to investigate whether it is the continuous production of the T7 RNA polymerase in this expression system post-induction or the production of CheYGFP itself that results in significant levels of stress on the bacterial host.

Fresh transformants of BL21(DE)* containing the pET20CheYGFP construct (positive control) or an empty pET20bhc vector only (negative control) were induced according to the standard protocol in shake flasks. Serial dilutions of cultures were plated onto non-selective nutrient agar to measure colony formation ability. The colonies formed were

subsequently replica plated onto selective nutrient agar to calculate % plasmid retention.

The ability of cells containing the pET20CheYGFP plasmid to form colonies had decreased slightly 3 h post-induction but had increased a little by 6 h post-induction (**Figure 3.8 A**). Numbers of colony forming units of bacteria containing the empty vector, however, were low even before induction, and remained at 1.06×10^8/ ml until 6 hours post-induction despite a little increase in the previous sample. Colony formation by both strains had fully recovered by 24 hours post-induction, being double for BL21*CheYGFP compared to BL21*pET20bhc (**Figure 3.8 A**). Plasmid retention in both strains decreased from 100% before induction to less than 1% after only 6 hours post-induction reflecting deleterious consequences of significant bacterial stress (**Figure 3.8 B**). The percentage of plasmid-bearing cells of BL21*CheYGFP and BL21*pET20bhc continued to decrease for 24 hours post-induction to less than 0.1% and 0.001%, respectively. These results indicate that the physiological impact of bacterial stress is associated with induction of the T7 expression system in BL21* rather than a direct result of recombinant protein production per se, which is manifested in the decrease or loss of colony forming ability and the overgrowth of a plasmid-free bacterial population. These data are in agreement with previous studies that report similar physiological responses during the expression of only protein tags in BL21 (Miroux and Walker, 1996; Soriano *et al*., 1999).

Optimisation of IPTG concentration for CheYGFP production in BL21*

The physiological impact associated with the induction of the T7 expression system in BL21* according to the standard protocol highlighted the need for the development of an
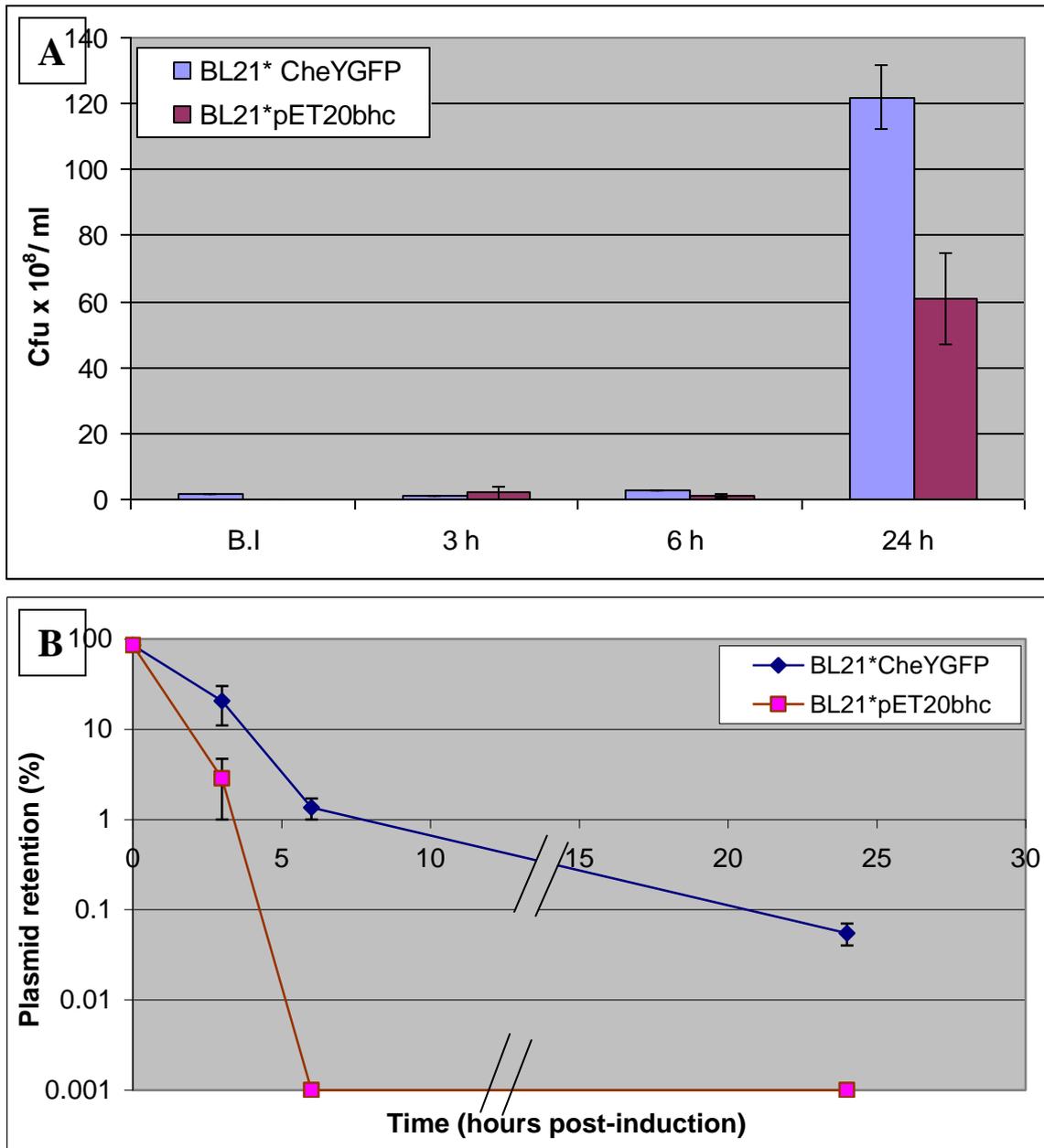
Figure 3.8: The effect of induction of BL21* using the standard protocol on colony formation and plasmid retention of pETCheYGFP and the empty vector pET20bhc. A) Serial dilutions of culture samples before induction (B.I), 3 h, 6 h and 24 hours post-induction were plated on non-selective nutrient agar and the colony forming units per ml were calculated. B) Colonies were replica plated on to carbenicillin-supplemented NA to reveal % plasmid retention. The error bars represent the standard error from two different clones of each strain from a single experiment.

induction protocol that minimises stress and its consequences on the yield and quality of the recombinant protein produced. This involved, most importantly, the determination of an optimal inducer concentration of IPTG that would both maintain colony formation ability and allow maximal production of soluble CheYGFP.

The overproduction of CheYGFP was induced in BL21* with different concentrations of IPTG at a constant temperature of 25°C in shake flasks. Preliminary IPTG titration experiments indicated that the optimal [IPTG] for induction was less than 0.1 mM and therefore, the expression of CheYGFP was induced with a narrow range of [IPTG] (0, 0.008 mM, 0.01 mM, 0.02 mM, 0.04 mM and 0.08 mM). Samples were taken before induction, 3 h, 6 h and 24 h post-induction to assay growth and fluorescence using the same culture dilutions, and for protein and colony formation analysis.

Bacterial growth in all cultures was very similar 3 h and 6 h post-induction reaching densities of around 1.9 and 5.2, respectively (**Figure 3.9**). Different final densities between cultures were reached 24 h post-induction, tending to correlate positively with the inducer concentration and suggesting that IPTG in low concentrations is beneficial to cell growth. Cultures induced with 0.008 and 0.01 mM IPTG reached similar optical densities of around 14, whereas those induced with 0.02 and 0.04 reached optical densities of 15 and 17, respectively (**Figure 3.9**). However, the final cell density of the culture induced with 80 µM IPTG was slightly lower than the culture induced with 0.04 mM IPTG, suggesting that an [IPTG] of 0.08 mM was high enough to stress the culture.
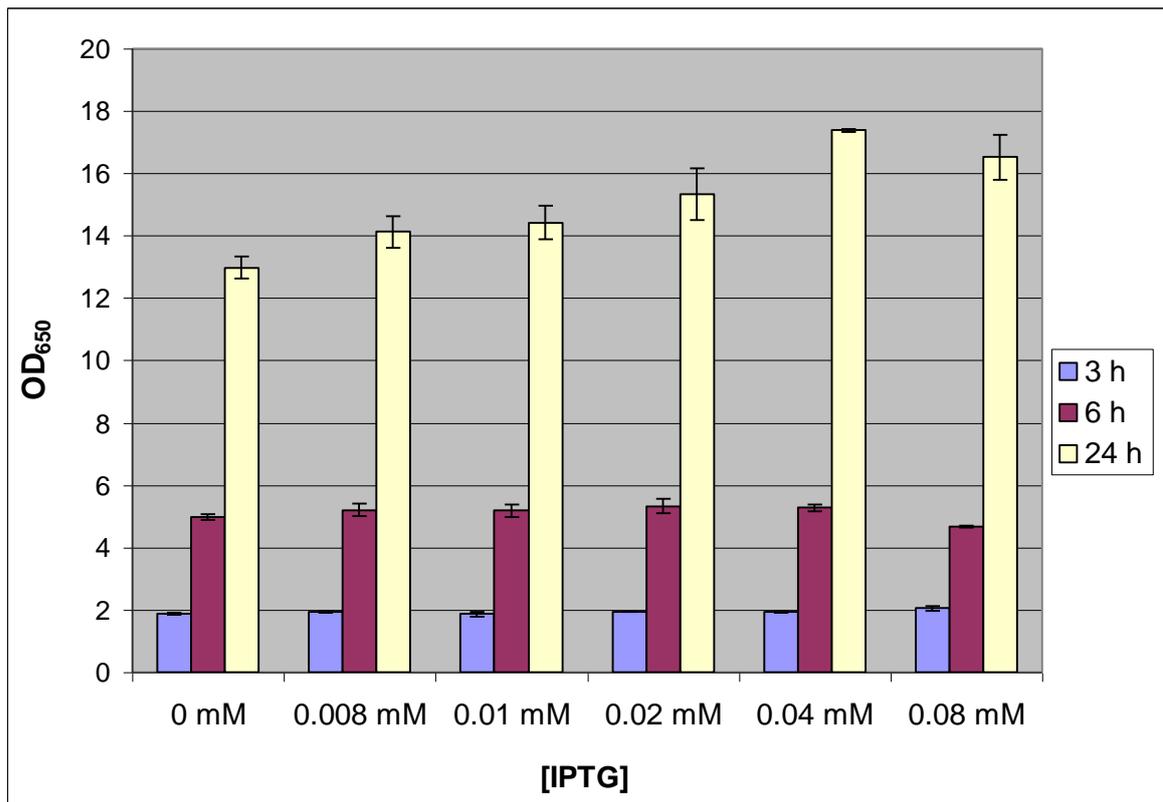
Figure 3.9: The effect of inducing concentrations of IPTG on growth densities of BL21*CheYGFP. BL21*CheYGFP was grown at 25°C in shake flasks and induced with different IPTG concentrations. Growth was assayed from dilutions of culture samples at 3 h, 6 h and 24 hours post-induction. An un-induced control culture was also assayed. Standard error bars from 3 independent experiments within different time series are shown.

The number of cfu was highest in the uninduced culture and that induced with the lowest IPTG concentration at both 6 h and 24 hours post-induction (**Figure 3.10**). This strongly suggests that the IPTG concentration is directly associated with the physiological stress imposed on the bacterial host. There was a decrease in colony formation by 24 h post-induction with 0.01 mM IPTG compared to the uninduced culture, which decreased further with increasing inducer concentrations. These data highlight the negative impact of even moderately low concentrations of IPTG on cell physiology (**Figure 3.10**).

The phenotype of colonies formed was scored by colour and the % plasmid retention within each culture was calculated (**Figure 3.11**). A negative correlation between the IPTG concentration used for induction and the % plasmid retention was apparent at 6 and 24 h post-induction. Plasmid retention was highest in the un-induced culture followed by the culture induced with 8 μM IPTG. This suggests that even very low concentrations of IPTG can impose low levels of stress and create selective advantages for the growth of plasmid-free bacteria. Replicate experiments revealed a % plasmid retention of 90-100% for un-induced cultures and in the range of 85-100% in the cultures induced with 8 μM IPTG.

The specific fluorescence, which reflects the accumulated yield of soluble CheYGFP per OD of culture 24 h post-induction, was calculated (**Figure 3.12**). The highest specific fluorescence was achieved in cultures induced with 0.008 and 0.01 mM, indicating that high yields of soluble CheYGFP were accumulated with low [IPTG]. However, there was a marked negative correlation between the IPTG concentrations used for induction and the specific fluorescence of cultures, indicating that less soluble CheYGFP was being accumulated. These results indicate that the optimal IPTG concentration for the production
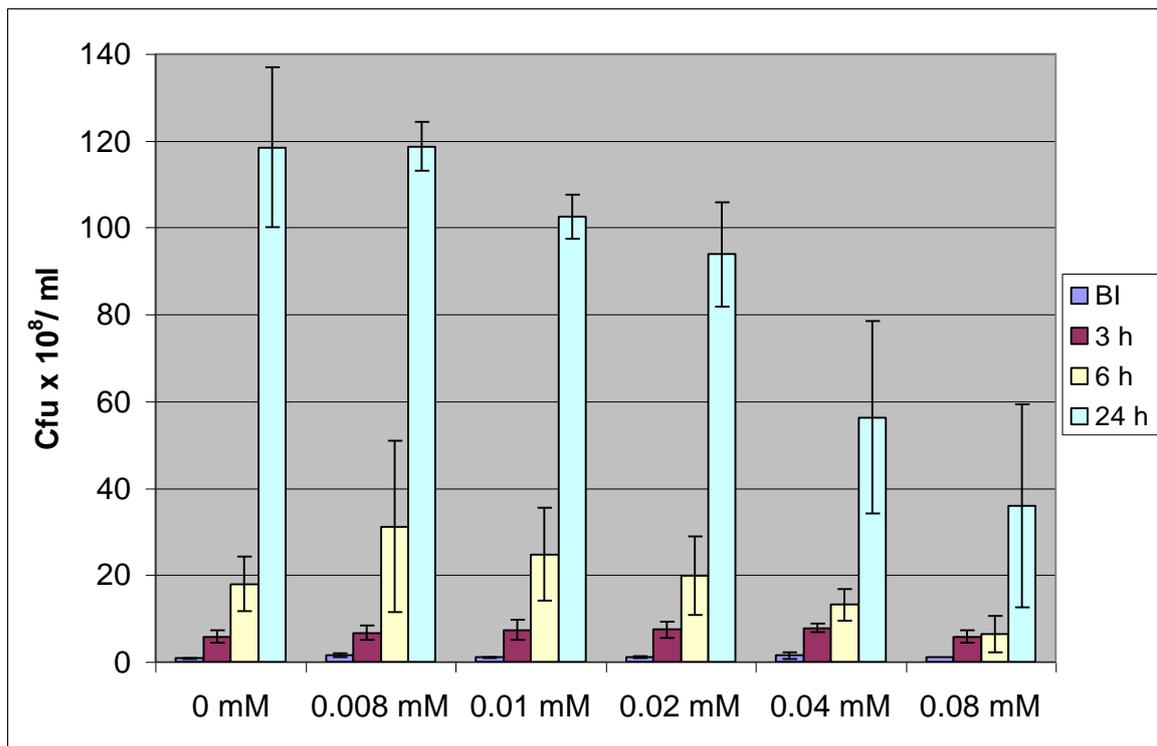
Figure 3.10: The effect of inducing concentrations of IPTG on colony formation of BL21*CheYGFP. BL21*CheYGFP was grown at 25°C in shake flasks and induced with different concentrations of IPTG. Serial dilutions of culture samples taken before induction (B.I) and at 3, 6 and 24 hours post-induction were plated on to non-selective nutrient agar, and the number of cfu was calculated. An un-induced control culture was also analysed. Standard error bars from 3 independent experiments, 2 of which were completed on the same day, are shown.
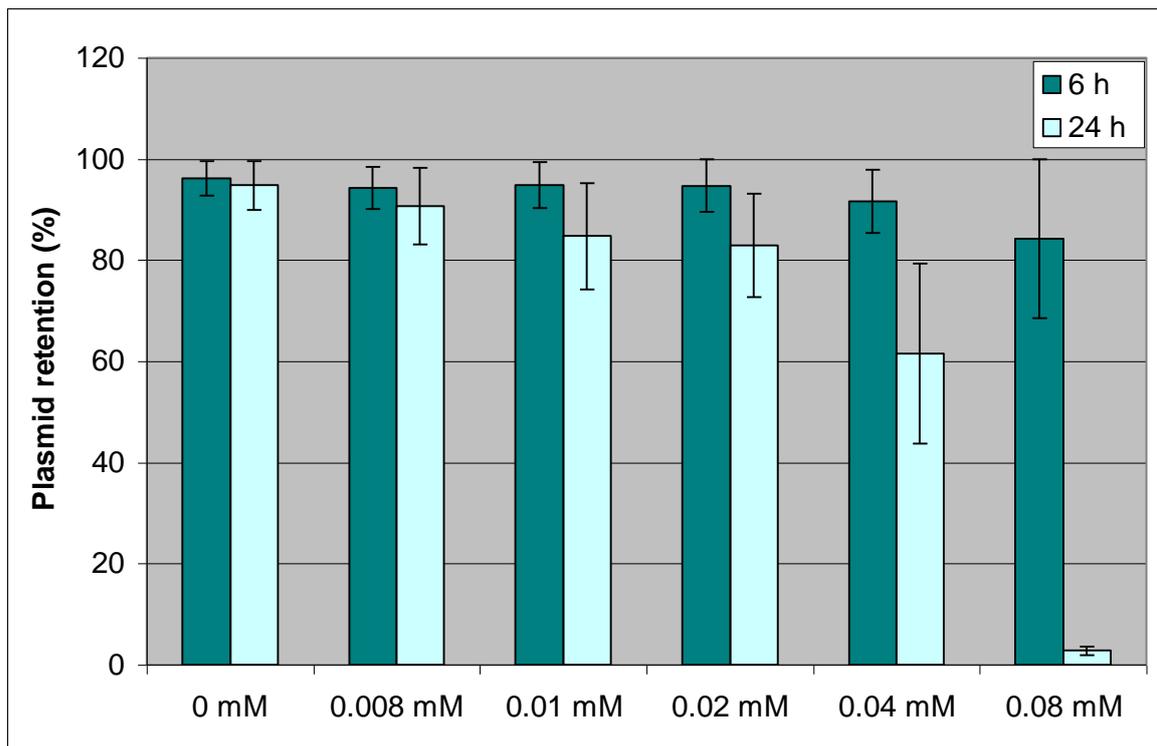
Figure 3.11: The effect of different inducing concentrations of IPTG on plasmid retention. BL21*CheYGFP was grown at 25°C in shake flasks and induced with different concentrations of IPTG. Plasmid retention from the different cultures at 6 and 24 h post-induction was calculated from colour phenotypes of colonies formed on nutrient agar. A non-induced control culture was also analysed. Standard error bars from 3 independent experiments, 2 of which were completed on the same day, are shown.
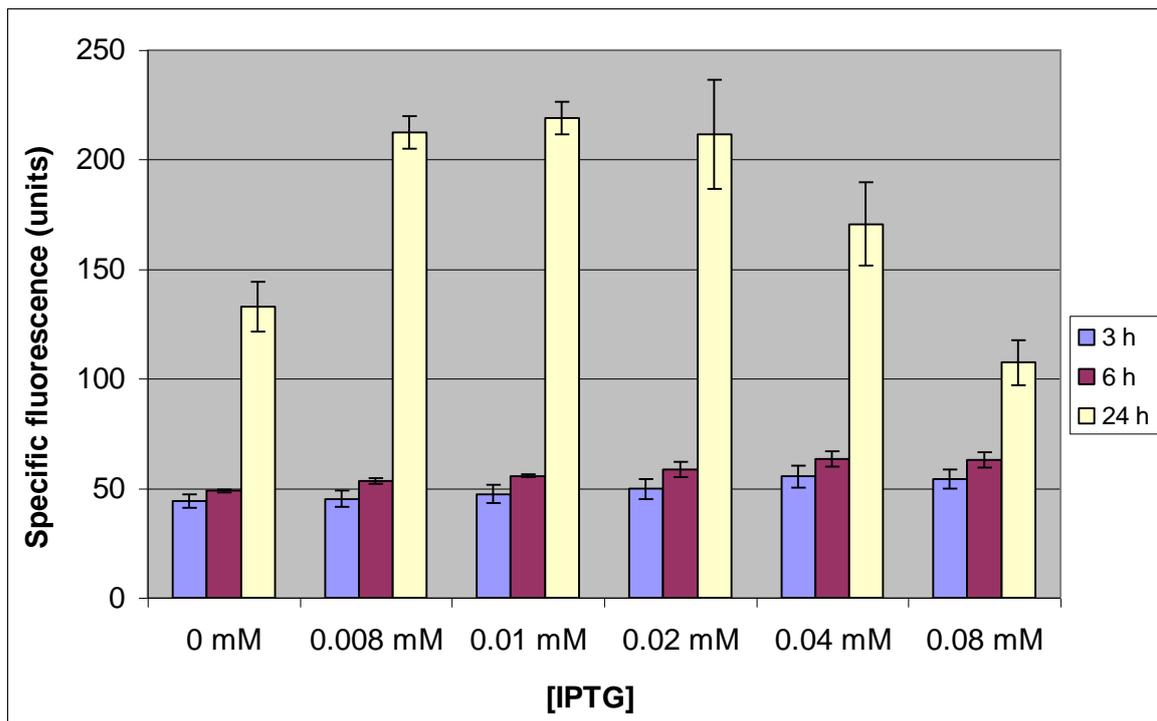
Figure 3.12: The specific fluorescence of BL21*CheYGFP induced with different IPTG concentrations at 25°C. Specific fluorescence of cultures at 3 h, 6 h and 24 h post-induction was calculated from fluorescence and biomass measurements obtained from the same culture samples. An uninduced control culture was also analysed. Standard error bars from 3 independent experiments, 2 of which were completed on the same day, are shown.

of soluble CheYGFP is in the range of 8 to 10 μM.

Total proteins from whole cell samples taken 24 hours post-induction were resolved by SDS-PAGE. Samples for these were normalised for cell biomass prior to loading. There was an incremental increase in the yield of recombinant CheYGFP with increasing IPTG inducer concentration, reaching maximal levels in cultures induced with 10 to 20 μM IPTG (**Figure 3.13 A**). However, yields of CheYGFP progressively decreased with increasing [IPTG] above 20 μM. The majority of CheYGFP was present in the soluble fraction when low [IPTG] was used for induction. The highest ratio of soluble to insoluble CheYGFP was present in the culture that had been induced with 8 μM IPTG, consistent with the highest specific fluorescence value of this culture (**Figure 3.13 B**). The ratio of soluble to insoluble CheYGFP progressively decreased with increased inducer concentrations to approximately 50% and 20% when 40 μM IPTG and 80 μM IPTG were used for induction, respectively. It was therefore concluded that the optimum IPTG concentration that both minimises the stress response on BL21* and produces the highest yield of soluble CheYGFP per gram of cell biomass is 0.008 mM. The collectively modified conditions used for the production of CheYGFP have been named the 'improved protocol'.

The role of T7 RNA polymerase accumulation on the stress response during RPP

It has been demonstrated that inducing the T7 expression system in BL21* with relatively high concentrations of IPTG results in detrimental consequences on cell physiology. It remains unclear, however, whether it is the accumulation of T7 RNA polymerase itself that accounts for the stress response in BL21* rather than an effect of inducing the T7
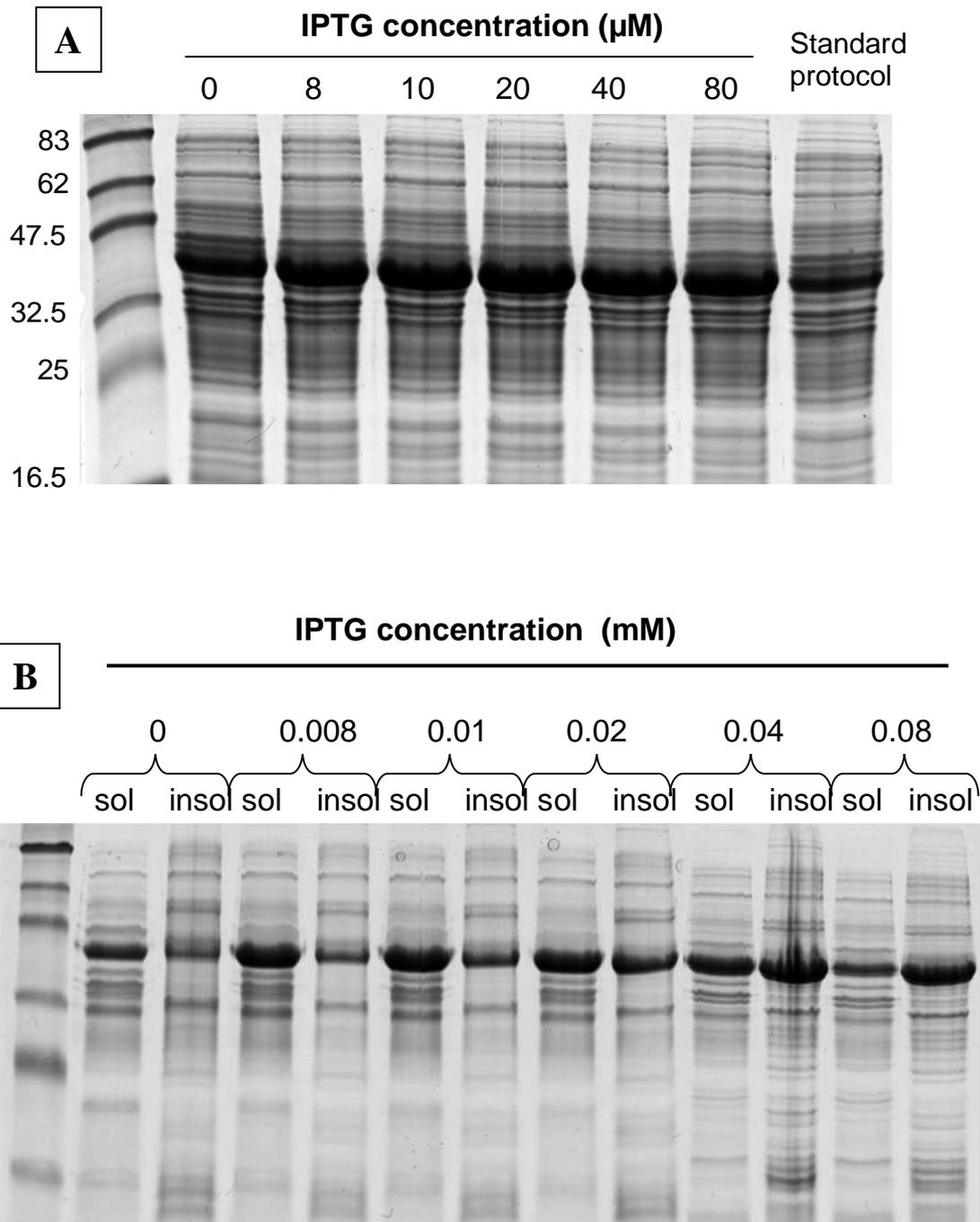
Figure 3.13: The effect of [IPTG] titration on the accumulation of CheYGFP in BL21*. BL21*CheYGFP was grown at 25°C in shake flasks and induced with different IPTG concentrations. Proteins from samples taken 24 hours post-induction were resolved by SDS-PAGE, in which samples were normalised for biomass prior to loading. A) Total proteins from whole cell samples. The yield of CheYGFP accumulated 24 h post-induction using the standard protocol is shown for comparison. B) Yields of CheYGFP present in the soluble (sol) and insoluble (insol) cell fractions.

expression system. This prompted an attempt to analyse the T7 RNA polymerase levels within cells using Western blotting, and determine whether it positively correlates with cellular stress.

Western blots of samples of BL21*CheYGFP after 24 hours of induction with a narrow-range of IPTG (8, 20, 40 and 80 µM) were probed with anti-T7 RNA polymerase antibody (**Figure 3.14**). A relatively strong signal of cross-reacting antigen was detected in the sample induced with 80 µM, whereas a significantly lower signal was detected from the sample induced with 40 µM. This is consistent with the stress response manifested on cell physiology, and defines IPTG concentrations of 40 µM or above as being counterproductive to RPP. There was no visible signal detected in samples induced with either 8 or 20 µM IPTG, which correspond with the relatively low stress response when such inducer levels were used. It was therefore concluded that the low concentration of IPTG simply limits the accumulation of the T7 RNA polymerase, which combined with a lower and constant temperature, explains why the improved protocol was successful.

The production of CheYGFP using the improved protocol in a fermenter

The scale-up of recombinant protein production processes in fermenters is a fundamental process in industry as well as being a vital step for many downstream applications. It was therefore imperative to test the reliability of the improved protocol in a fermenter. BL21*CheYGFP was grown aerobically at a constant temperature of 25°C and induced at an OD of 0.5 with 8 µM of IPTG. The pH was controlled at 6.3 by the automated addition of 10% NH_4OH and 5% HCl. Threonine, serine and asparagine were added at 1 mM final
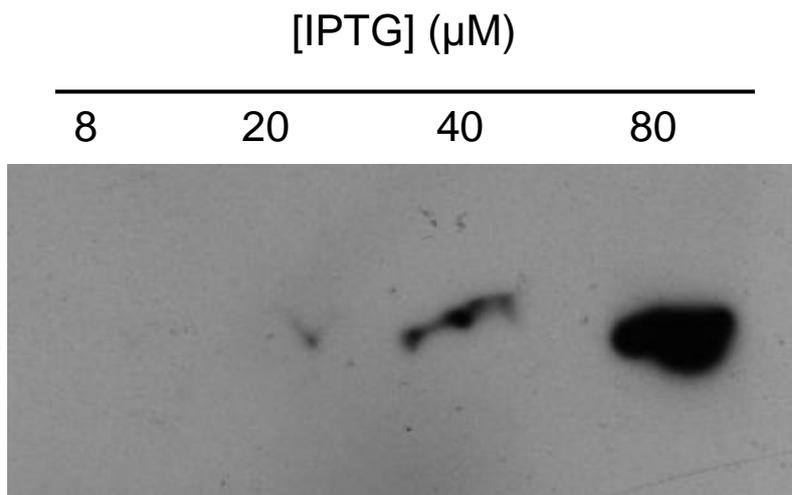
Figure 3.14: The correlation of T7 RNA polymerase levels with physiological stress in bacteria during RPP. Western blotting analysis of T7 RNA polymerase in BL21*CheYGFP 24 hours post-induction with 8, 20, 40 and 80 μM IPTG. Samples analysed were normalised for biomass to allow for reliable comparisons.

concentration each around 5 h post-induction since these amino acids are utilised rapidly during RPP (Jones, 2007). Bacterial growth, fluorescence, colony forming units and protein accumulation were analysed at intervals post-induction. Flow cytometry was also used to analyse bacterial populations within the culture at different stages of the fermentation.

Bacterial biomass progressively increased post-induction, reaching an optical density of around 25 units by 45 hours post-induction after which it had plateaued (**Figure 3.15 A**). The specific fluorescence also continued to increase post-induction reaching approximately 250 units towards the end of the fermentation, indicating a constant accumulation of soluble CheYGFP (**Figure 3.15 A**). Serial dilutions of cultures at different stages revealed a constant increase of green colony forming units reaching a maximum of 0.49×10^{11}/ ml with 99.5% plasmid retention 49 hours post-induction (**Figure 3.15 B**). These data reflect the successful minimisation of the stress response and its positive effects on recombinant protein production in large scale fermentations. There was a slight decrease in the colony forming units 68 hours post-induction, most likely due to nutrient limitations.

The dissolved oxygen concentration of the culture in the fermenter was measured from the point of inoculation (**Figure 3.16**). The dissolved oxygen concentration was highest immediately before induction, but had rapidly decreased to less than 60% of the starting concentration by 6 h post-induction. This indicated the rapid growth of bacteria in addition to the overproduction of the recombinant protein. This contrasts with the production of CheYGFP using the standard protocol in a fermenter where there was very little decrease in the level of oxygen soon after induction. The dissolved oxygen continued to decrease until
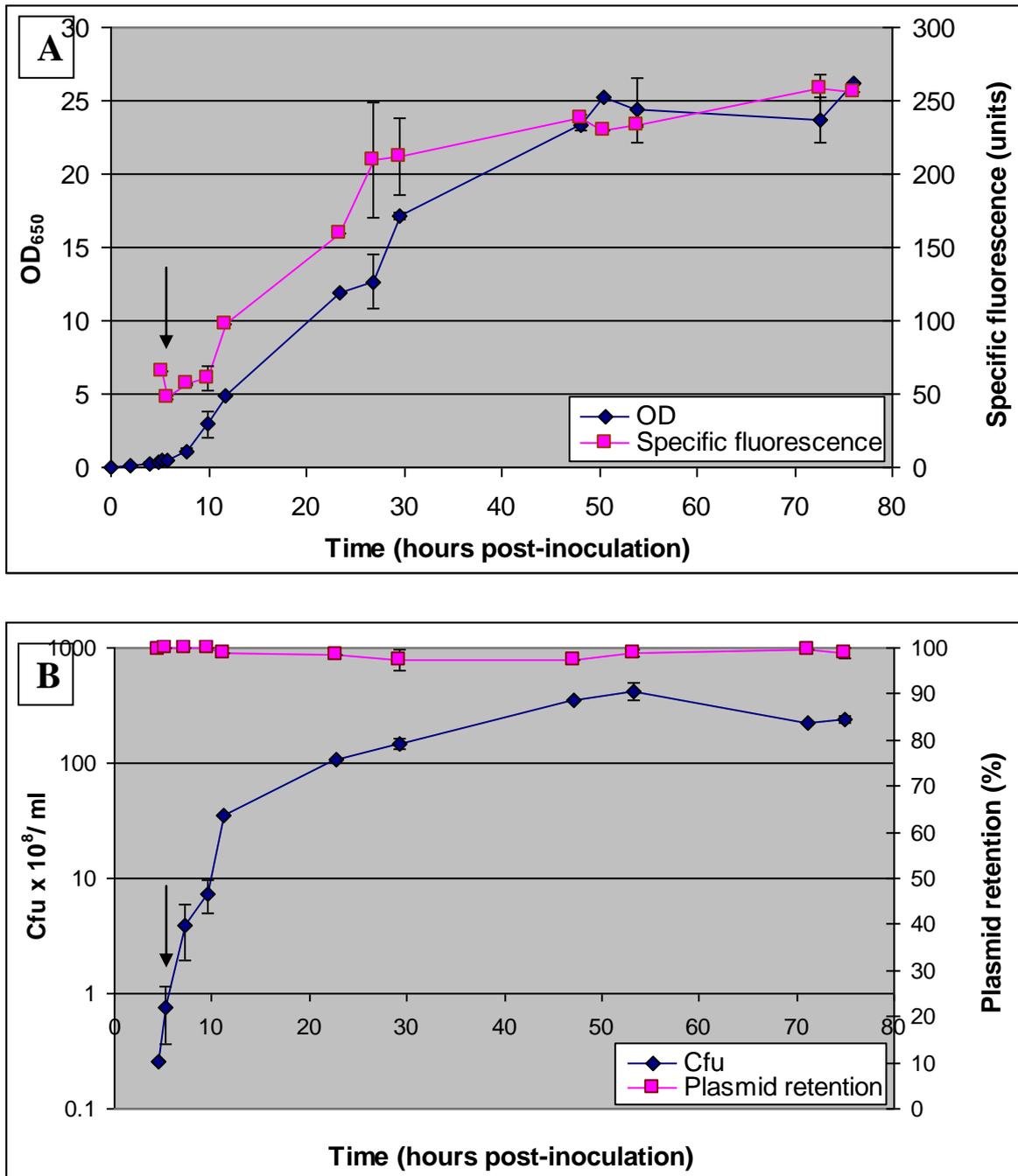
Figure 3.15: Physiological responses of BL21*CheYGFP post-induction with the improved protocol in a benchtop fermenter. **A)** Growth and specific fluorescence post-inoculation were assayed from the same culture sample. The point of induction is shown by an arrow. **B)** The number of colony forming units and plasmid retention post-induction (shown by an arrow). Serial dilutions of culture samples were plated on to non-selective nutrient agar, and the cfu/ ml was calculated. The colour of the colonies formed was used to calculate % plasmid retention. The standard error of samples from triplicate fermentation experiments is presented.
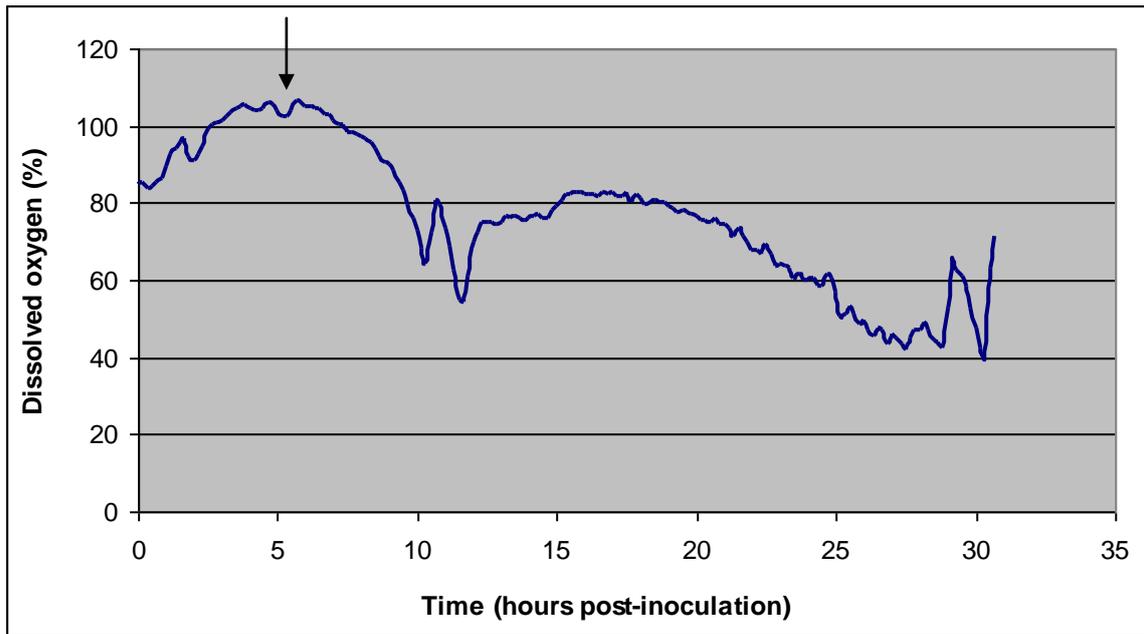
Figure 3.16: Dissolved oxygen concentration during production of CheYGFP with the improved protocol in a fermenter. BL21*CheYGFP was grown and induced using the improved protocol in a benchtop fermenter. The dissolved oxygen concentration of the culture was measured using an oxygen probe from the point of inoculation. The point of induction is shown by an arrow.

24 h post-induction when it became less than 40% of the starting concentration.

Culture samples before induction and at intervals post-induction were analysed by flow cytometry. Cells with green fluorescence that have accumulated CheYGFP were detected by PMT2, whereas cells with red fluorescence, due to uptake of propidium iodide, were detected by PMT4. Before induction, a majority cell population with high green fluorescence, due to the leaky expression of CheYGFP, was detected (**Figure 3.17 A**). A minor population of bacteria with relatively no green fluorescence was also detected. Cells of this population have likely lost the recombinant plasmid. There was no uptake of propidium iodide by cells of either population, reflecting intact outer membranes of healthy living bacteria. A highly fluorescent bacterial population, comprising approximately 98% of cells in the culture, was detected 4 h post-induction and maintained up to 22 h post-induction, which was consistent with colony phenotype after plating (**Figure 3.17 B**). Minor populations of dead bacteria (0.2%) and living plasmid-free bacteria (3.2%) that had no green fluorescence were also detected.

Proteins from samples of whole and fractionated cells taken throughout the fermentation and corrected for biomass were analysed by SDS-PAGE (**Figure 3.18 A**). There was progressive increase in total yield of CheYGFP accumulated within cells up to 70 hours post induction. Also, the majority of CheYGFP had accumulated in the soluble fraction, indicating that high quality protein had been produced (**Figure 3.18 B**).

A

Before induction   4 h post-induction   22 h post-induction

B
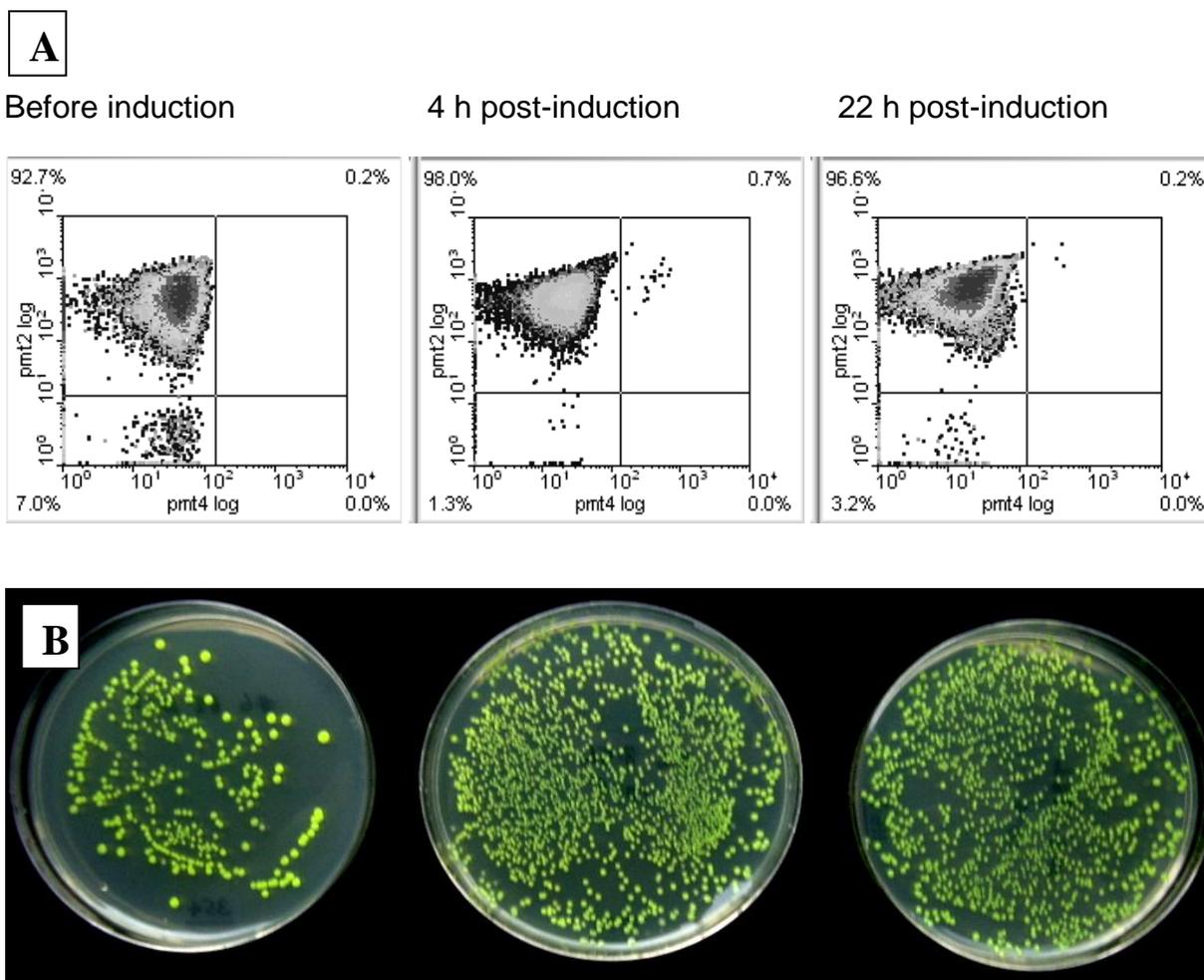
Figure 3.17:   Physiology of BL21*CheYGFP post-induction using the improved protocol in a batch fermenter.   A)   Flow cytometric analysis of the fluorescence of subpopulations and the corresponding colony phenotype.   Bacteria with green GFP fluorescence were detected by PMT2, and dead red-fluorescent cells were detected by PMT4. B) Serial dilutions of the same samples were plated on non-selective nutrient agar.
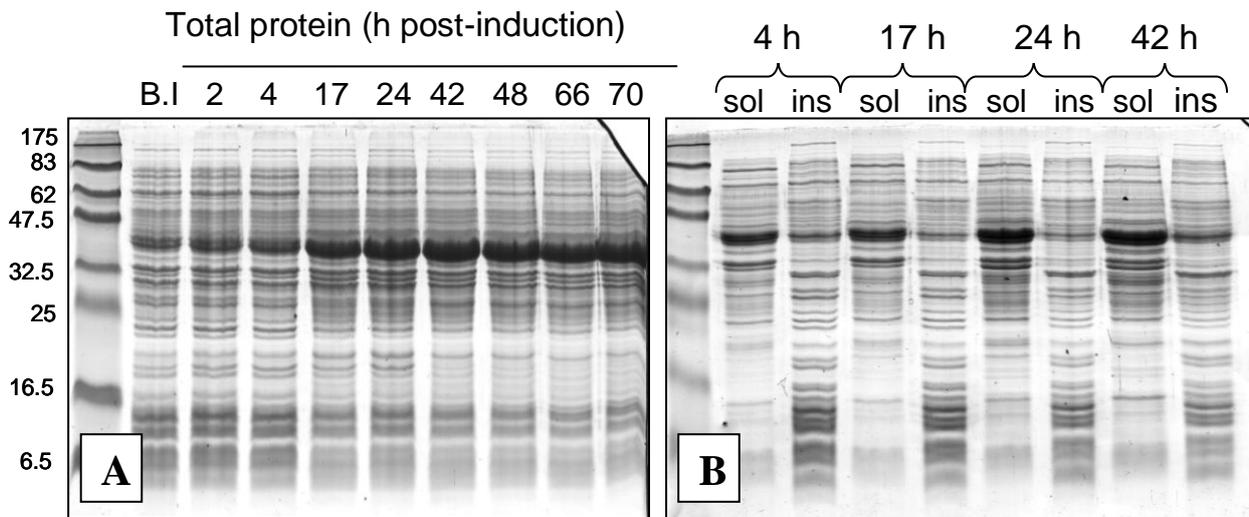
Figure 3.18: **SDS-PAGE analysis of CheYGFP production with the improved protocol in a batch fermenter**. A) Total protein accumulated in BL21*CheYGFP before induction (B.I) and at intervals post-induction. B) Protein accumulated in the soluble (sol) and the insoluble (insol) fractions of BL21*CheYGFP.

Comparison of levels of T7 RNA polymerase in the standard and improved protocols

Samples of BL21*CheYGFP before and post-induction with the standard and improved protocols were analysed by Western blotting and probed with anti-T7 RNA polymerase antibody (**Figure 3.19**). Strong bands of cross-reacting antigen with increasing intensity post-induction were clear in samples induced by the standard protocol, indicating the accumulation of high levels of T7 RNA polymerase. In contrast, there was no visible signal detected from samples induced using the improved protocol, indicating that little T7 RNA polymerase was produced. Quantification by gel densitometry revealed approximately a 200-fold difference in the levels of T7 RNA polymerase between the two protocols. It was therefore concluded that the magnitude of difference in the T7 RNA polymerase levels accounted for the extremely contrasting effects that the two protocols have on bacterial physiology during recombinant protein production.

The role of antibiotic selection in the improved protocol during CheYGFP production

Many pharmaceutical applications and product manufacturers require the production of recombinant protein without any presence of antibiotic in the medium. This introduces problems of plasmid loss and overgrowth by plasmid-negative bacteria in the culture due to bacterial stress, especially if the standard protocol is being used for the protein production. Using a method that minimises bacterial stress, which is the underlying key aspect of the improved protocol, is therefore imperative to overcome such problems and could therefore be used as a solution to such demands by pharmaceutical companies. To explore the possibility of exploiting the improved protocol for such applications, attempts were made

Hours post-induction

Standard protocol | Improved protocol

0 1 3 6 9 12 27   0 3 6 9 24 31 48
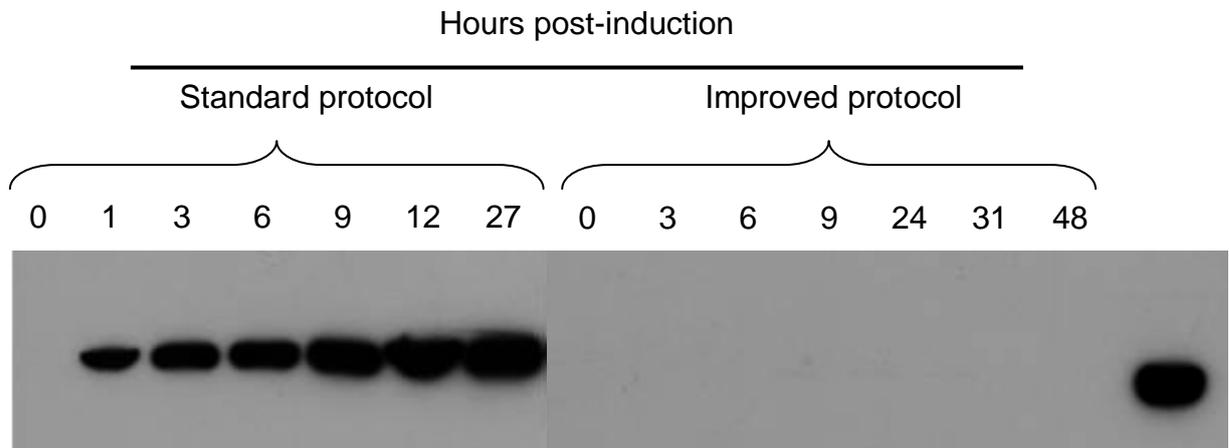
Figure 3.19: Western blotting analysis of T7 RNA polymerase in BL21*CheYGFP induced using the standard and improved protocols. CheYGFP was produced using both protocols in pH-controlled batch fermentations. Samples were taken at intervals post-induction and were normalised for cell biomass prior to analysis. A very faint trace of signal was detected from the 9 h, 24 h and 31 h post-induction samples of the improved protocol.

to scale-up CheYGFP production by using the improved protocol but without using carbenicillin to maintain the plasmid.

In such an experiment in a 3.1 L fermenter, bacterial growth and specific fluorescence continued to increase post-induction, indicating the continuous accumulation of CheYGFP (**Figure 3.20 A**). The highest optical density measured was approximately 26 with a specific fluorescence of 260 after 70 hours post- induction. These values are comparable to fermentation experiments in which antibiotics were present during protein overproduction.

Colony forming units increased post-induction to a maximum of 3.47 x 10^{10}/ ml 42 hours post-induction (**Figure 3.20 B**). As before, colony forming units had decreased by 70 hours post-induction, probably due to nutrient limitation in the growth medium at this point. The majority of colonies formed were green with 100% plasmid retention 70 hours post-induction, indicating that the low stress profile of the improved protocol, rather than the presence of antibiotic, is in itself sufficient to maintain a healthy majority of plasmid-containing bacteria.

The dissolved oxygen concentration (pO_2) was monitored throughout the fermentation (**Figure 3.21**). Similar to the production of CheYGFP using the improved protocol in the presence of antibiotics, the oxygen concentration decreased after induction to around 60% of the initial concentration by 6 h post-induction. This coincided with bacterial growth and the overproduction of CheYGFP. The pO_2 continued to decrease to less than 20% by 24 h post-induction. Oxygen levels within the culture gradually increased thereafter, which coincided with a plateau in bacterial growth. This suggests that the plateau in bacterial
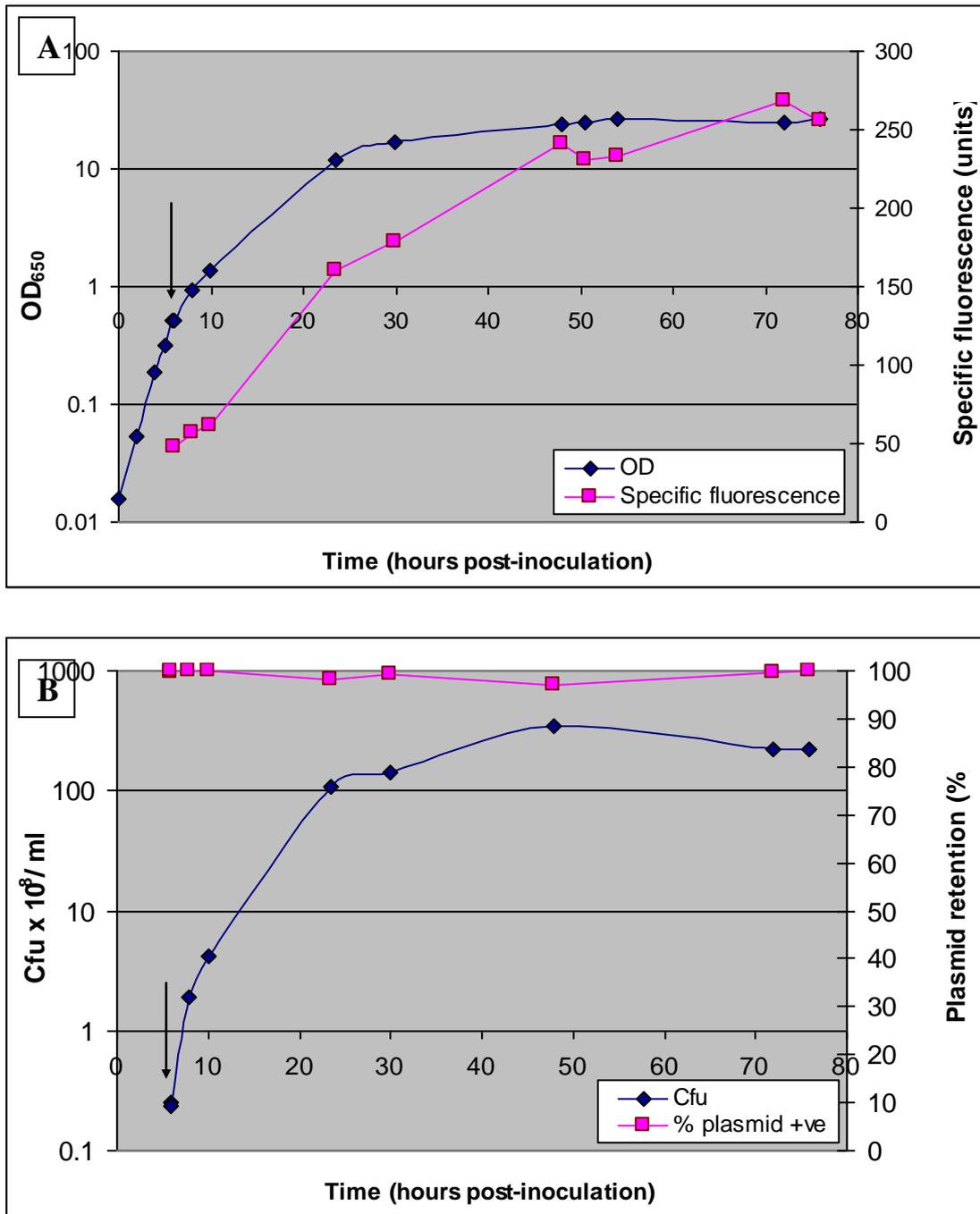
Figure 3.20: Induction of BL21*CheYGFP with the improved protocol without antibiotic selection in a batch fermenter. A) Growth and specific fluorescence post-induction (induction point indicated by an arrow). B) Colony forming units (cfu) and plasmid retention post-induction (induction point indicated by an arrow). Data from a single fermentation experiment is presented.
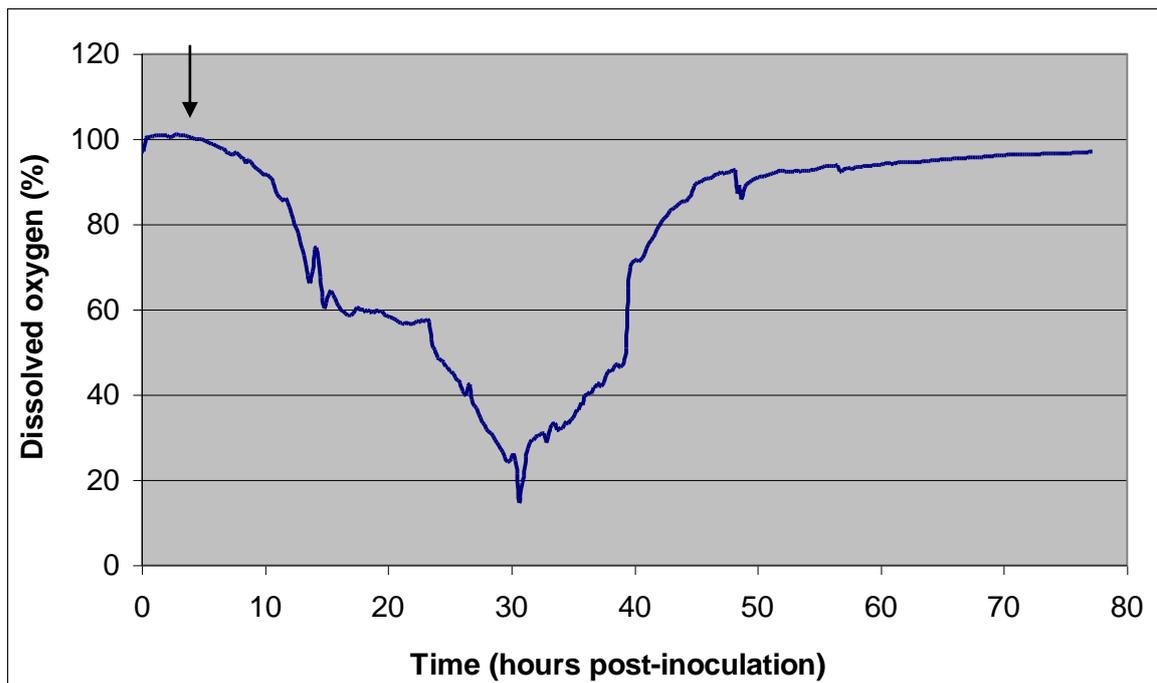
Figure 3.21: Dissolved oxygen concentration during production of CheYGFP with the improved protocol without antibiotics. BL21*CheYGFP was grown and induced using the improved protocol without carbenicillin in a benchtop fermenter. The dissolved oxygen concentration of the culture was measured using an oxygen probe from the point of inoculation. The point of induction is shown by an arrow.

growth is not due to oxygen limitation *per se* but rather to limitations in nutrient components in the media. These results highlight the prospect of intensification of the improved protocol for RPP by improving fermentation engineering processes.

In conclusion, these preliminary data suggest that the improved protocol could be applied for the overproduction of protein without the need for antibiotics.

Production of CheYGFP using the improved protocol in a fed-batch fermentation

Pharmaceutical companies often require large quantities of recombinant protein in a relatively short time. This usually involves the generation of high dry cell biomass in excess of 50 g/ L. Fed-batch fermentations are routinely used to maximise the biomass of the recombinant host by feeding with 40% glucose at 37°C before the recombinant protein is induced with a high concentration of IPTG, typically 0.5 – 1 mM. Owing to the deleterious consequences of the standard protocol on bacterial physiology and plasmid retention, the induction period is not prolonged and bacteria are harvested after only a few hours of induction. The prospect of whether the improved protocol could be combined with process intensification using fed-batch fermentations was therefore investigated. In particular, the possibility of achieving high cell biomass whilst maintaining high yields of soluble recombinant protein was explored.

An overnight culture of freshly transformed BL21*CheYGFP was used to inoculate at 2% (v/v) a 3.6 L fermenter containing 1.5 L LB supplemented with 2% (v/v) glucose and carbenicillin (100 μg/ ml). Bacteria were grown and induced according to the improved

protocol. An aeration of 6 L/ min (4vvm) was used, and the pH was maintained at 6.3. Serine, threonine and asparagine were added at 1 mM final concentration each after 5 hours of induction, and the feed pump was turned on 7 h post-induction. An incremental feeding regime was used, whereby the feed rate is manually increased when the specific growth rate of bacteria decreases. The incremental feed rate was based on indications from the dissolved oxygen concentration in the culture. When the dissolved oxygen concentration in the culture began to increase, it provided an indication that the specific growth rate of bacteria had decreased. This is because further growth and biomass of bacteria can no longer be supported with the feed rate used. Thus, the feed rate was increased to ensure the support of continued bacterial growth at higher biomass concentrations. Therefore, an initial feed rate of 13.8 ml h^{-1} was used, followed by 20.6 ml h^{-1} 32.5 h post-induction, 27 ml h^{-1} 48.5 h post-induction and 39 ml h^{-1} 54.5 hours post-induction. Feed was depleted by 62 hours of induction.

The dissolved oxygen concentration decreased at a constant rate upon induction due to the rapid growth of bacteria and the overproduction of recombinant protein, until it was only 20% of the initial concentration by 27 h post-induction (**Figure 3.22**). The pO$_2$ continued to decrease to less than 0% by around 30 h post-induction. However, concentrations of dissolved oxygen began to increase soon afterwards, indicating a slower specific growth rate or possibly even death of bacteria. This provided an indication to increase the rate of the feed pump in order to support a higher bacterial biomass in the culture. The concentration of dissolved oxygen decreased almost immediately in response, indicating the recovery of bacterial growth (**Figure 3.22**). The pO$_2$ was used to monitor bacterial growth and the feed rate was increased accordingly. The dissolved oxygen concentration
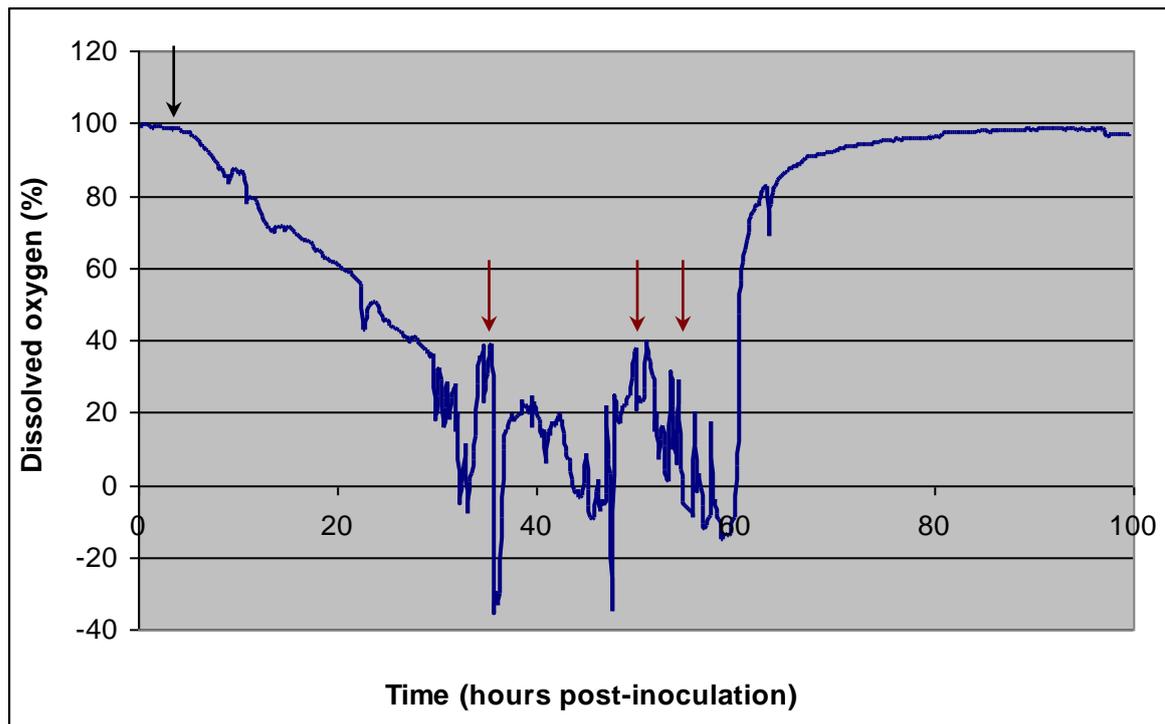
Figure 3.22: **Dissolved oxygen concentration during production of CheYGFP with the improved protocol in a fed-batch fermentation.** BL21*CheYGFP was grown and induced using the improved protocol in a fed-batch fermentation. The dissolved oxygen concentration of the culture was measured using an oxygen probe from the point of inoculation. The point of induction is shown by a black arrow. The pO2 was used as an indicator of bacterial growth in the culture. The feed rate was manually increased when oxygen levels increased above 20% (shown by red arrows).

increased to almost 100%, corresponding with the cessation of bacterial growth.

Bacterial density increased steadily after induction, reaching a recorded maximum of 150 units 60 hours post-induction (**Figure 3.23**). An optical density of 144 was maintained until 93 hours of induction, generating a dry cell biomass of 43.9 g/ L. The accumulation of soluble CheYGFP per unit of cell biomass, as reflected by the specific fluorescence, increased continually until a maximum of 361 units. The total fluorescence of the culture therefore increased too, reflecting both the increase in cell biomass and the accumulation of recombinant protein, reaching 5.2×10^4 fluorescence units 93 hours post-induction (**Figure 3.23**). This indicates that the improved protocol is suitable for process intensification, both at the recombinant protein production and cell biomass levels.

Serial dilutions of culture samples were plated on to non-selective nutrient agar and the colony forming units and plasmid retention were calculated. The number of colony forming units increased steadily reaching a maximum of 0.166×10^{12} 44 hours post-induction, after which it decreased (**Figure 3.24**). The plasmid retention remained at a minimum of 96% by 60 hours of induction, but had slightly decreased to 86% towards the end of the fermentation.

Total proteins from whole cells (normalised for cell biomass) were resolved by SDS-PAGE. A gradual increase in yield of CheYGFP was apparent post-induction up to 94 hours (**Figure 3.25**). Quantification by gel densitometry revealed that the accumulated CheYGFP accounts for approximately 35 % of the total cellular protein. It was concluded from these data that the improved protocol is ideal for process intensifications industrially,
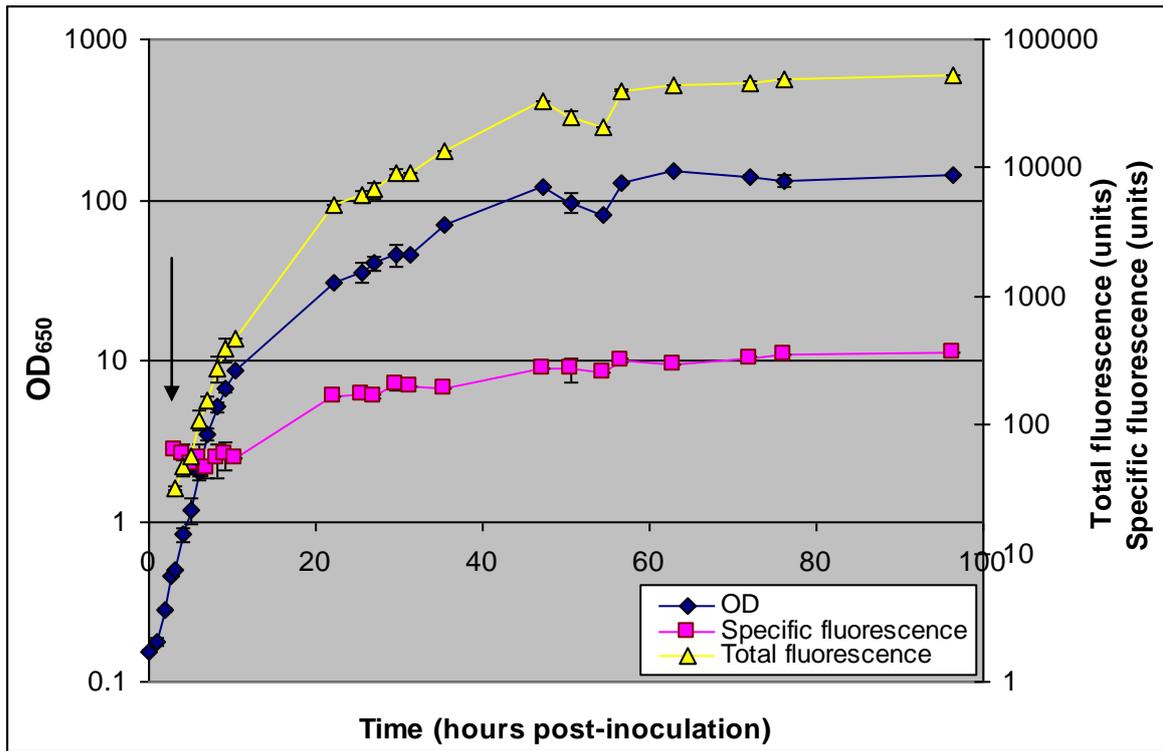
Figure 3.23: Production of CheYGFP in BL21* using the improved protocol in a fed-batch fermentation. The optical density, total and specific fluorescence were measured post-induction (indicated by an arrow). Feeding was commenced 8 hours post-induction with incremental rate. Standard error of each parameter was calculated from duplicate fermentation experiments.
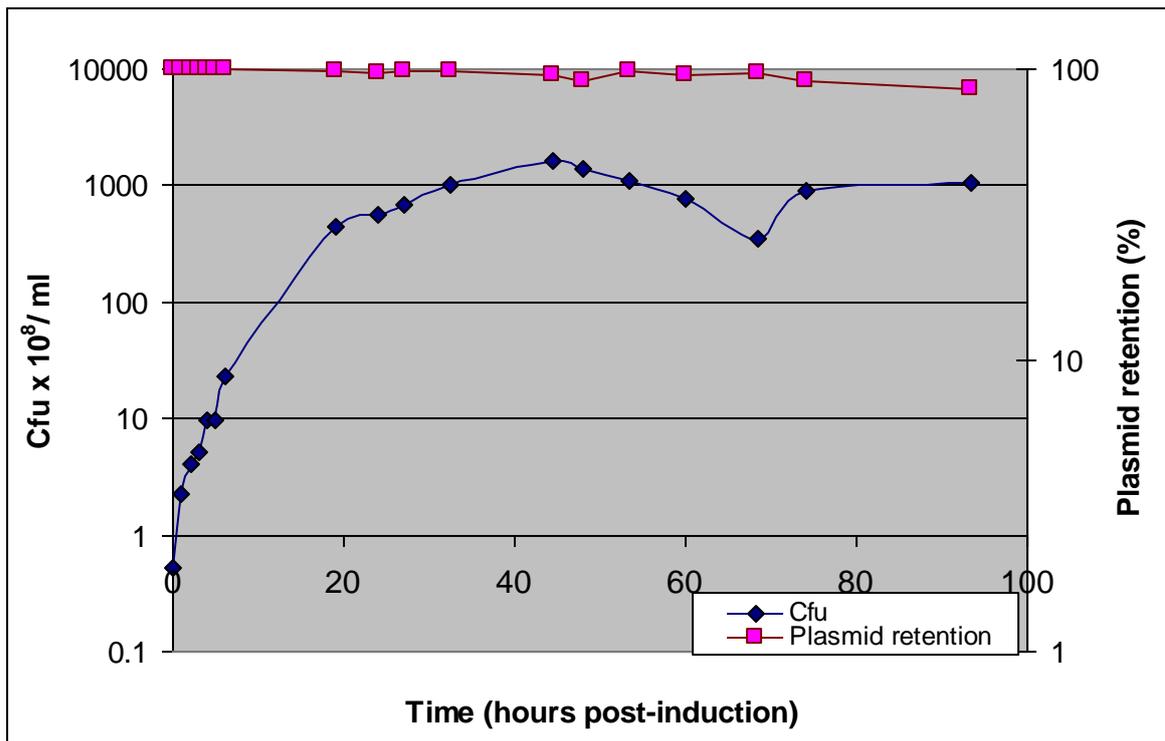
Figure 3.24: Colony forming units and plasmid retention in BL21* during CheYGFP production using the improved protocol in a fed-batch fermentation. Serial dilutions of culture samples taken before induction (0 h) and at intervals post-induction were plated onto NA and the % plasmid retention was calculated.
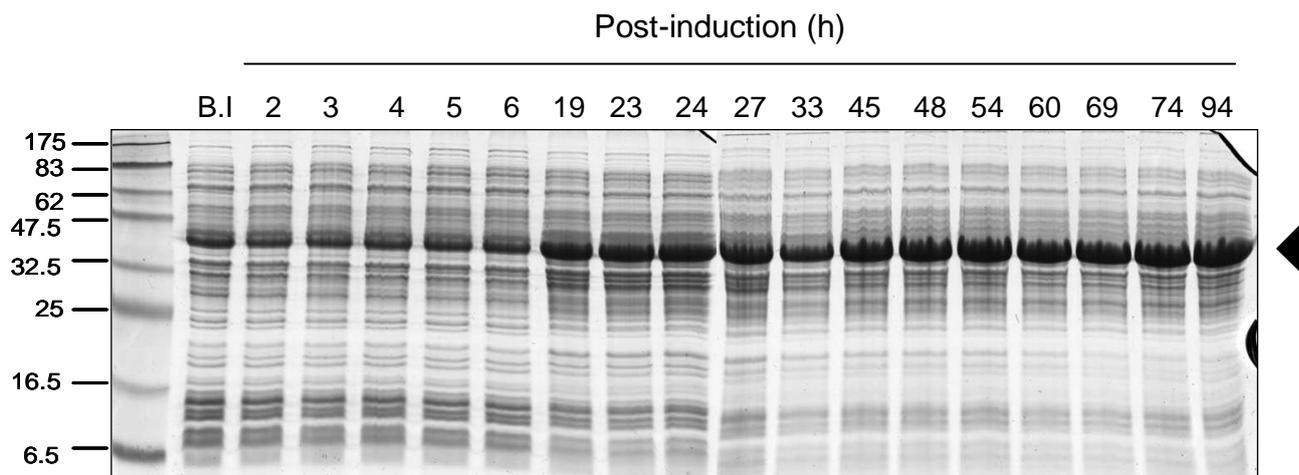
Figure 3.25: Yields of CheYGFP accumulated in BL21* using the improved protocol in a fed-batch fermentation. Total proteins from BL21*CheYGFP of samples taken before induction (B.I) and at intervals post-induction were resolved by SDS-PAGE. The accumulated CheYGFP is shown by an arrow.

and perhaps the only compromise is the length of time involved.

CHAPTER 4

The isolation of a mutant strain of *Escherichia coli* for enhanced recombinant protein production

The improved protocol has been developed to prolong the accumulation of recombinant proteins through minimising the stress on the bacterial host. An alternative approach to overcome the problems associated with bacterial stress during RPP is to identify a mutant *E. coli* host that is tolerant of or resistant to that stress. Several projects have focussed on developing host strains by specifically targeting genes that are known to affect RPP yields. For example, strains deficient in proteases such as ClpP and DegP have resulted in an increase in yields of the recombinant protein by minimising protein degradation (Vera *et al*., 2007). Other strains that are deficient in the recombinase gene *recA* have also been developed in an attempt to increase the RPP yields indirectly by decreasing the frequency of plasmid multimerisation and therefore the occurrence of plasmid-free segregants that accumulate towards the late stages of a fermentation (Zhao *et al*., 2003). Whereas these methods comprise 'local' approaches to solving the problems associated with RPP, little attention has been focussed on addressing the underlying physiological factors that globally affect RPP. This prompted the isolation of a mutant host that is primarily adapted to resist the stress associated with RPP.

The isolation of a stress-resistant BL21*strain during CheYGFP production

The high levels of bacterial stress associated with the standard protocol during RPP were evidently sufficient to create a selective advantage for the overgrowth of plasmid-free bacteria after the induction of *cheY::gfp*. However, rare occurrences of green, plasmid-containing colonies were also detected on non-selective nutrient agar 24 h post-induction. This immediately raised the question of whether those *E. coli* BL21* cells might be stress-

resistant mutants that continued to accumulate recombinant protein, and hence could account for much of the protein detected by SDS-PAGE after prolonged expression.

Four green colonies, some brighter than others, that were identified 24 h post-induction on non-selective nutrient agar, were purified on nutrient agar with carbenicillin selection. All candidates formed colonies, indicating that the cells still contained the pETCheYGFP plasmid. In addition, the colour intensity of the colonies was reproduced. The production of CheYGFP in these strains, using the standard protocol in shake flask experiments, was investigated. Only one candidate (that was 'pale' green and therefore named P2) was able to overproduce CheYGFP. This strain, P2 BL21*, was further investigated during the overproduction of CheYGFP.

The production of CheYGFP in wild type and P2 BL21*

Since P2 BL21* was isolated 24 hours after induction using the standard protocol, it was predicted that it contained mutations that rendered this strain capable of being resistant to the induction stress during RPP. The physiological response of P2 BL21* during the overproduction of CheYGFP in the standard protocol was investigated and compared to that of wild type BL21*, using either the standard or improved protocol. Bacterial growth, fluorescence and colony formation ability post-induction were measured, and protein accumulation was analysed.

Optical densities of all 3 cultures continued to increase until 24 hours post-induction, being highest for P2 BL21* at an OD of 16.3 (**Figure 4.1**). There was a little further increase in
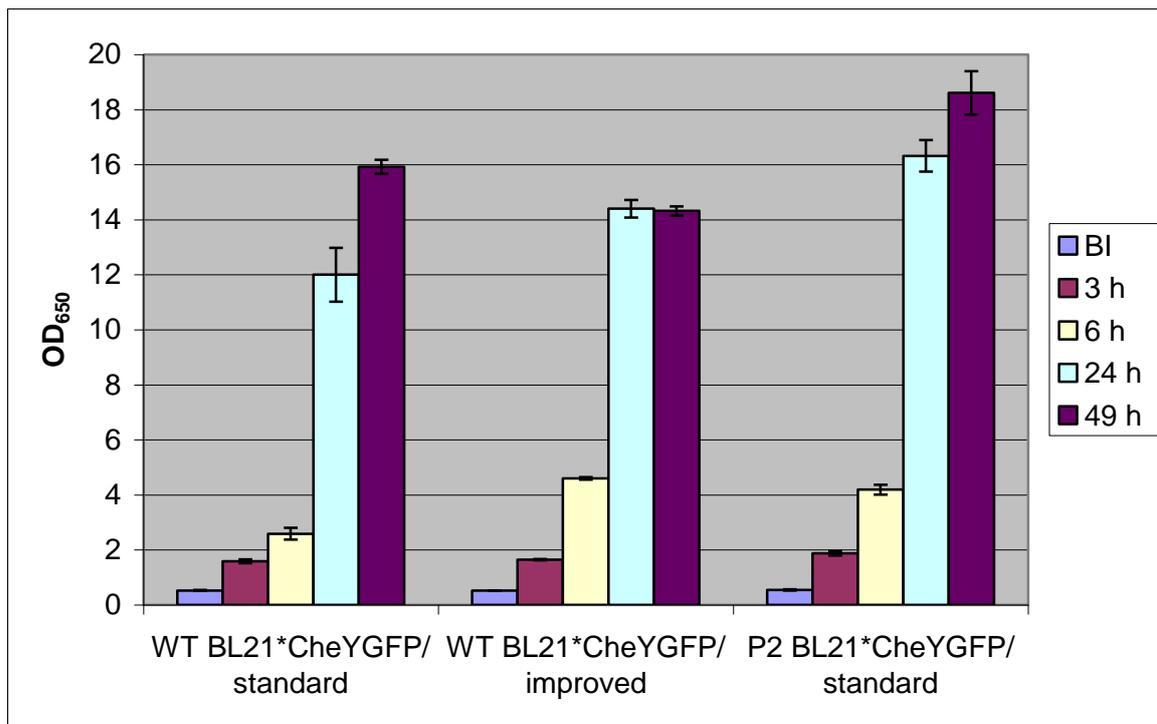
Figure 4.1: Comparing growth of wild type and P2 BL21* strains during CheYGFP production. Production of CheYGFP by the stress-resistant host, P2 BL21*, was induced using the standard protocol. This was compared to control cultures of the parental strain that were induced with either the standard or the improved protocols. Optical densities were measured before induction (B.I) and at intervals post-induction. Standard error bars from triplicate repeats using 3 different colonies of each strain, and within one time series, are presented.

the optical density of P2 BL21* at 49 h post-induction whereas there was no further increase in the wild type host after 24 h when the improved protocol was used for induction. This indicates that the P2 BL21* strain is better able to continue growing during prolonged CheYGFP production compared to the wild type strain.

The specific fluorescence of all cultures at intervals post-induction was calculated (**Figure 4.2**). The specific fluorescence significantly increased 24 hours post-induction reaching 190 units when wild type was induced with improved protocol compared to 230 units when P2 BL21* was induced with the standard protocol. In contrast, there was little increase in the specific fluorescence of the wild type strain that had been induced with the standard protocol. There was no further increase in the specific fluorescence 49 hours post-induction of the wild type strain that had been induced with the improved protocol, whereas there was a notable further increase of P2 BL21* by at least 24%.

The total fluorescence (per unit volume), which combines factors of optical density and specific productivity, was measured for all cultures (**Figure 4.3**). Total fluorescence 24 h post-induction was highest in P2 BL21*CheYGFP, approximately 36% more than the wild type strain that had been induced using the improved protocol. This increased by a further 41% 49 hours post-induction. In contrast, the total fluorescence of the wild type control culture induced using the standard protocol was very low, with a maximum value of 1000 units 24 h post-induction. These data suggest that the P2 BL21* strain accumulated higher yields of CheYGFP, and for longer periods, compared to the normal host, regardless of the protocol used for induction of the wild type strain.
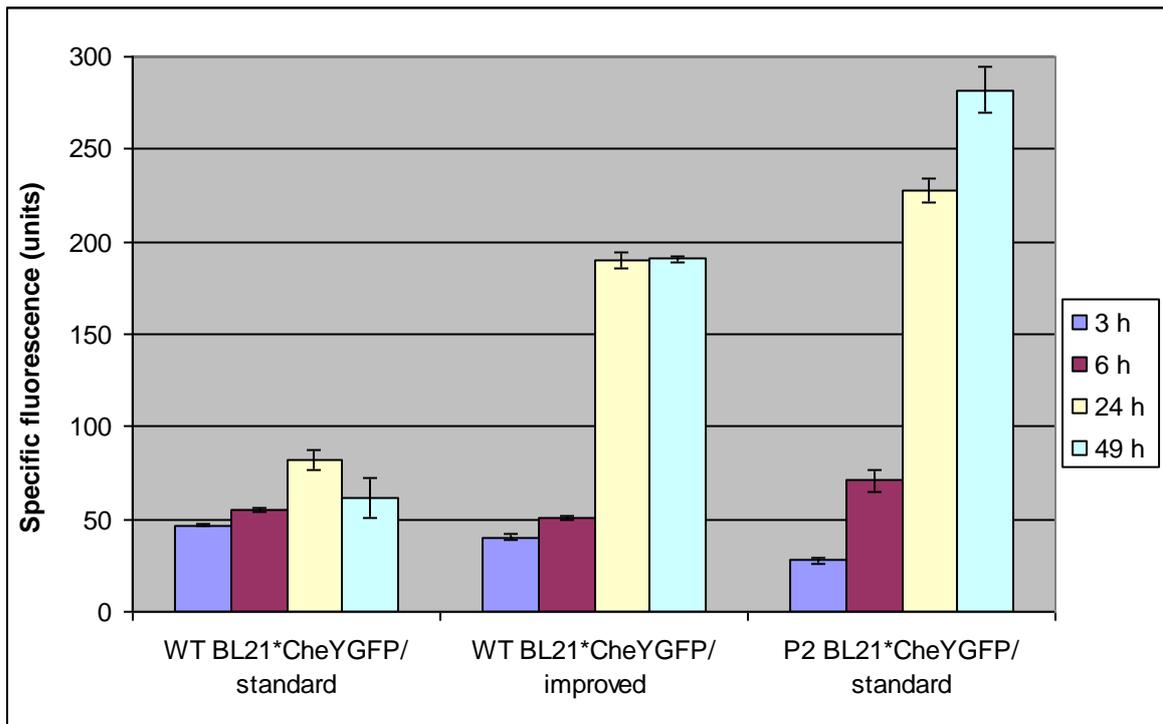
Figure 4.2: Comparison of the specific fluorescence of P2 BL21* and the parental strain during CheYGFP production. The production of CheYGFP by wild type and P2 BL21* was induced using the improved or the standard protocols, respectively. These were compared to control cultures of the parental strain that were induced using the standard protocol. The fluorescence per unit of biomass of all cultures was calculated at intervals post-induction. Standard error bars from triplicate repeats using 3 different colonies for each strain, and within one time series, are presented.
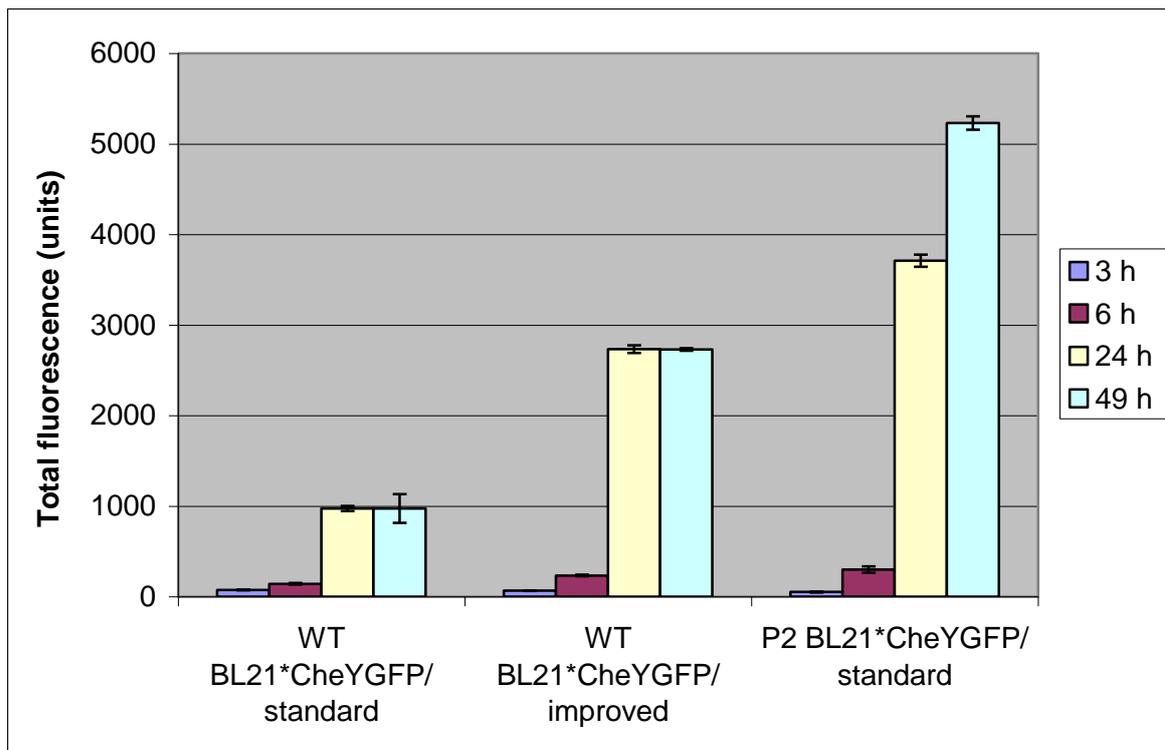
Figure 4.3: Total fluorescence of wild type and P2 BL21* strains during CheYGFP production. The production of CheYGFP was induced in wild type and P2 BL21* using the improved and standard protocol, respectively. These were compared to control cultures of the parental strain that were induced with the standard protocol. The total fluorescence (per volume of culture) was measured at intervals post-induction. Units are the specific fluorescence multiplied by the concentration of biomass in the culture. Standard error bars from triplicate repeats using 3 different colonies for each strain, and within one time series, are presented.

Serial dilutions of samples from all cultures were plated onto non-selective nutrient agar, and the cfu were calculated (**Figure 4.4**). There was a characteristic loss of colony formation of wild type BL21* by about 2 orders of magnitude when the standard protocol was used for induction, reflecting physiological stress. In contrast, there was a continued increase of cfu of the P2 BL21* strain after induction using the standard protocol to 73×10^8/ ml, 24 hours post-induction. This is similar to the increase in cfu of the wild type strain that had been induced by the improved protocol. This suggests that the P2 BL21* strain, unlike the parental strain, is resistant to the stress that is introduced by the standard protocol.

The colonies formed were phenotypically scored as green or white to reflect the presence or loss of the recombinant plasmid, respectively, in addition to replica plating on to nutrient agar supplemented with carbenicillin. There was a continued decrease in the proportion of plasmid bearing cells of wild type BL21* post-induction using the standard protocol, as shown previously, due to the overgrowth of plasmid-free cells in response to the stress involved (**Figure 4.5**). In contrast, the % plasmid retention by P2 BL21* decreased only slightly and was maintained at 85% 24 hours after induction. This again suggests that P2 BL21* is adapted to tolerate the stress involved with this induction protocol. Replicate experiments show percentage plasmid retention in the range of $85 - 98\%$ by P2 BL21* 24 h post-induction. Using the improved protocol to induce recombinant protein expression in the normal host resulted in plasmid retention of about 100%. This, together with the higher number of cfu generated, suggests that the stress during RPP using the improved protocol with the wild type strain was less compared using the improved host P2 BL21*.
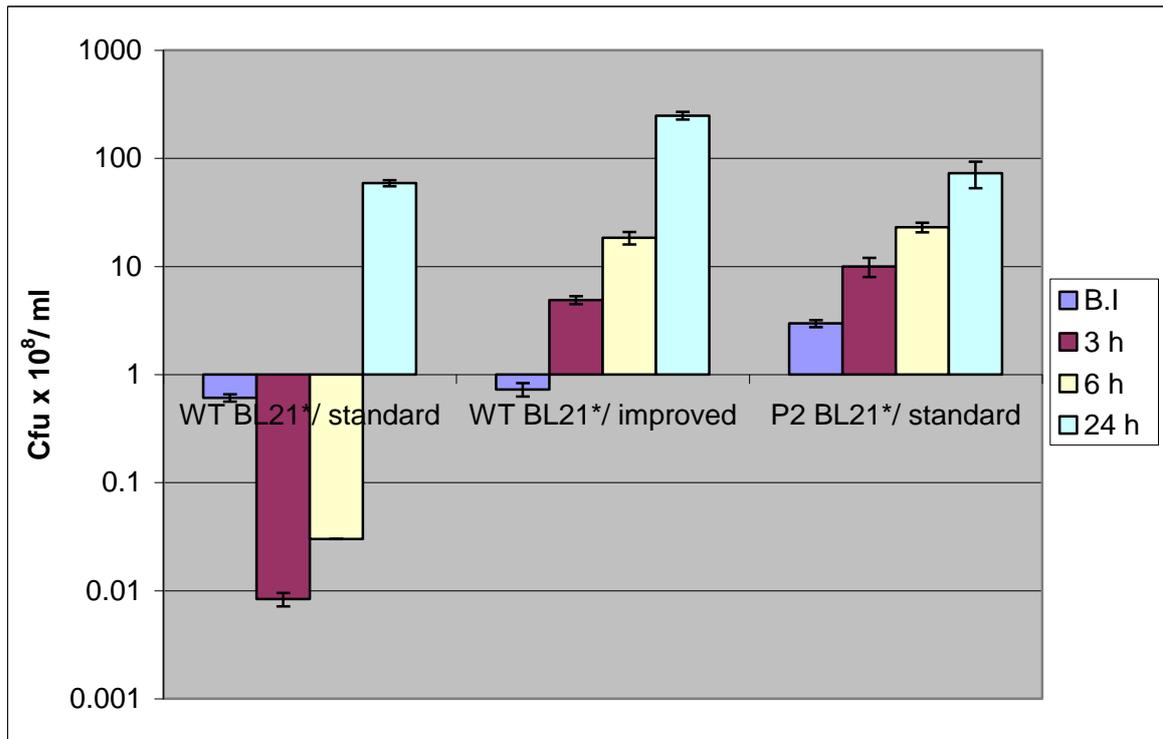
Figure 4.4: **Colony forming units of wild type and P2 BL21* strains during CheYGFP production**. The production of CheYGFP was induced in wild type BL21* using either the standard or improved protocols, and in P2 BL21* using the standard protocol only. Serial dilutions from all cultures were plated onto non-selective nutrient agar before induction (B.I) and at intervals post-induction, and the cfu were calculated. Data shown were obtained using several serial dilutions from triplicate repeats. Standard error bars from triplicate repeats using 3 different colonies for each strain, and within one time series, are presented.
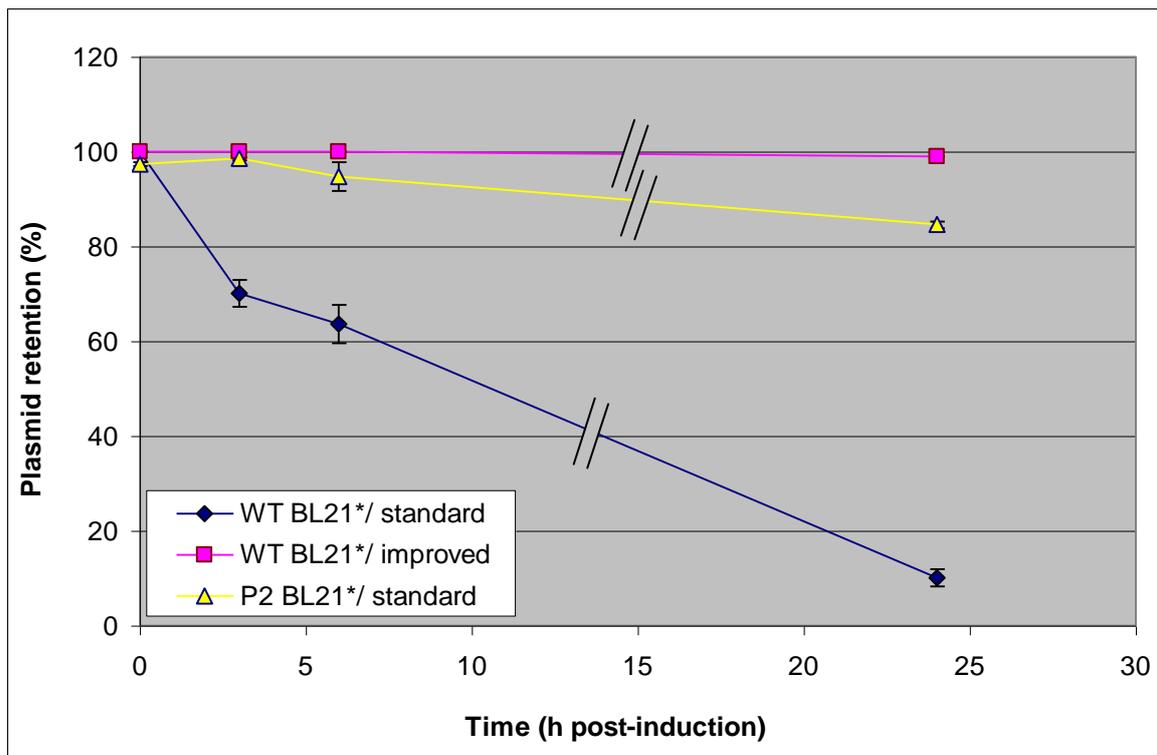
Figure 4.5: The effect of CheYGFP production by wild type and P2 BL21* strains on plasmid retention. The production of CheYGFP was induced in wild type BL21* using either the standard or the improved protocols, or in P2 BL21* using only the standard protocol. Colonies formed before induction and at intervals post-induction were replica plated onto nutrient agar containing carbenicillin and the % plasmid retention was calculated. These data were reproduced when colonies were phenotypically scored as white or green and the % green colonies calculated. Standard error bars from triplicate repeats using 3 different colonies for each strain, and within one time series, are presented.

Proteins from whole cells were analysed by SDS-PAGE (**Figure 4.6**). There was a rapid burst of CheYGFP synthesis in the wild type strain that had been induced using the standard protocol after only 3 hours post-induction but the protein concentration had decreased 24 hours post-induction, presumably due to the overgrowth by non-productive plasmid-free bacteria in the culture. In contrast, there was slow accumulation of CheYGFP after induction in both P2 BL21* using the standard protocol and in the wild type strain using the improved protocol, but with high yields generated 24 and 49 hours post-induction. These data suggest that the physiological stress-resistant phenotype shown by P2 BL21* might be associated with the decreased rate of recombinant protein synthesis, as with the improved protocol.

Proteins from soluble and insoluble cell fractions of both strains were analysed by SDS-PAGE (**Figure 4.7**). Using the standard protocol, the majority of CheYGFP was present in the insoluble fraction of the wild type host. In contrast, the majority of CheYGFP had accumulated in the soluble fraction of P2 BL21*, indicating that high yields of soluble CheYGFP had been produced. Similarly, the majority of CheYGFP was present in the soluble fraction of wild type BL21* when the improved protocol was used for induction.

It was therefore evident that an adapted *E. coli* strain, P2 BL21*, had been isolated that tolerates the stress inflicted by the standard protocol and is able to sustain production of soluble CheYGFP for prolonged periods.
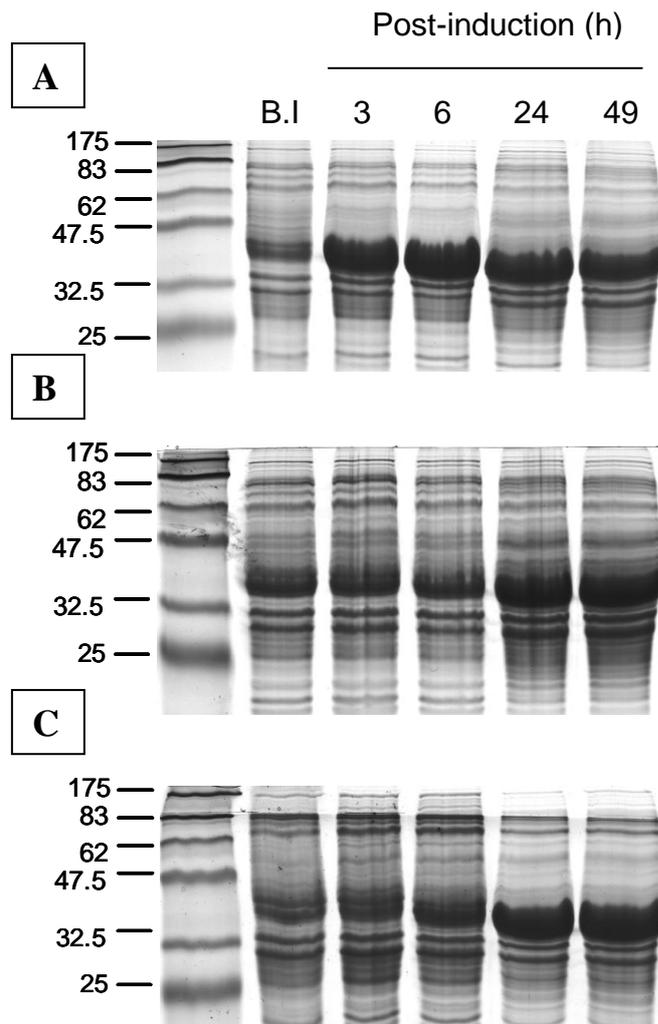
Figure 4.6: Comparison of the yields of CheYGFP accumulated by wild type and P2 BL21* strains. The accumulation of CheYGFP in wild type BL21* that was induced using either the standard (A) or the improved protocol (B) was compared to P2 BL21* that had been induced using the standard protocol (C). Proteins from whole cell samples taken before induction (B.I) and at intervals post-induction were resolved by SDS-PAGE. Equal biomass was loaded into each lane to enable fair comparison of recombinant protein yields. The sizes of marker proteins are shown.
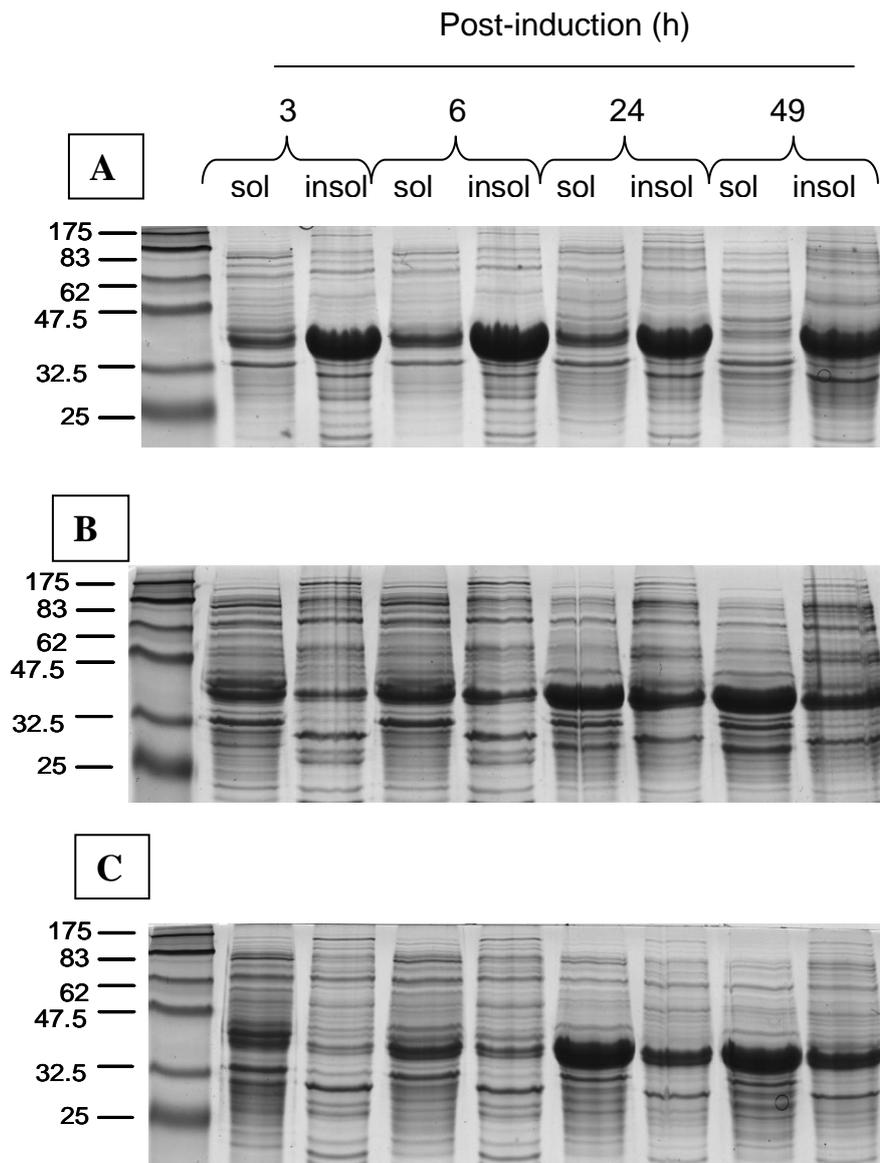
Figure 4.7: Yields of soluble and insoluble CheYGFP accumulated by wild type and P2 BL21* hosts. Production of CheYGFP was induced in wild type BL21 using either the standard (A) or the improved protocol (B) was compared to P2 BL21* that had been induced using the standard protocol (C). Protein from the soluble (sol) and insoluble (insol) cell fractions from samples taken at intervals post-induction were resolved by SDS-PAGE. Equal quantities of biomass were loaded into each lane. The sizes of marker proteins are shown.

The location of mutation in P2 BL21* that improves CheYGFP production

The different physiological response and improvement of CheYGFP production shown by P2 BL21* prompted the investigation of whether this was due to mutation(s) on the *E. coli* chromosome or on the pET plasmid.

P2 BL21* was first cured from CheYGFP. This was achieved by plating serial dilutions of P2 BL21*CheYGFP 24 h post-induction onto non-selective nutrient agar. Non-fluorescent derivatives were identified amongst the majority of fluorescent colonies, and the absence of the expression plasmid from such candidates was confirmed subsequently by the lack of growth in the presence of carbenicillin selection. Several cured candidates of P2 BL21* were transformed with the original construct of CheYGFP. All P2 BL21*CheYGFP transformants were pale green compared to transformants of BL21*, which were bright green. This indicates that a mutation is present on the chromosome of the P2 BL21* strain that affects phenotype. Fresh transformants of both strains were purified and the production of CheYGFP was attempted using the standard protocol.

The optical densities of P2 BL21*CheYGFP and the wild type strain were significantly different 24 hours post-induction, reaching 8 and 14 units, respectively. Serial dilutions of culture samples were also plated on nutrient agar before induction, 3.5 h and 24 h post-induction. Colony formation by the wild type control decreased by almost 300-fold 3.5 h post-induction, but had increased to 28.3×10^8/ ml of 95% white colonies by 24 h post-induction. In contrast, colony formation by the P2 BL21* strain continued to increase until 24 h post-induction to 41×10^8/ ml of 92% green, plasmid positive cells. The

reproducibility of these results compared to earlier ones suggested that mutation(s) were present on the chromosome of P2 BL21* that confer stress resistance as reflected by the physiological response.

CheYGFP accumulation by both strains was analysed by SDS-PAGE. There was a rapid burst of CheYGFP production by transformants of wild type BL21* compared to slow accumulation by P2 BL21* until 24 h post-induction. Further analysis of protein solubility revealed that the majority of CheYGFP that had accumulated by the parental strain was insoluble, whereas the majority accumulated by P2 BL21* was soluble (not presented).

The possibility that mutations have accumulated on the pETCheYGFP plasmid, whilst in P2 BL21*, which might have affected the overproduction of CheYGFP, was also investigated. The plasmid pETCheYGFP was isolated from transformants of P2 BL21* after 24 hours post-induction using the standard protocol, and was used to transform naive wild type BL21*. The production of CheYGFP in wild type BL21* using either this plasmid or the original plasmid construct was investigated using the standard protocol. There was a loss of colony forming units of all transformants post-induction regardless of the source of the pETCheYGFP plasmid. This indicates that a mutation that improved RPP in P2 BL21*CheYGFP was not present on the plasmid, but rather on the chromosome of the P2 BL21* host.

Effect of combining both the improved host and protocol on CheYGFP accumulation

It has been shown that using the improved protocol and the improved host P2 BL21*, independently, can enhance the yield of soluble CheYGFP produced. This immediately raised the question of whether combining both the improved protocol and the improved host would further improve yields of soluble CheYGFP produced. The yield of CheYGFP in P2 BL21* using the improved protocol was therefore investigated and compared to a P2 BL21* control when using the standard protocol.

The optical density of P2 BL21*CheYGFP induced using the standard protocol was almost double that when the improved protocol was used 24 h post-induction, suggesting that using the standard protocol for induction enhances growth of P2 BL21* (**Figure 4.8**). There was no loss of colony formation when either protocol was used for induction. However, numbers of colony forming units 24 h post-induction were more than 2-fold higher at 1.06×10^{10}/ ml when the improved protocol was used compared to only 0.48×10^{10}/ ml when the standard protocol was used. This suggests that there was significantly less stress when the improved protocol is used for induction even in the improved host compared to when the standard protocol is used. However, the discrepancy between the optical density obtained and the corresponding colony forming units in the different production protocols could be due to an artefact of cell size.

Phenotypic scoring revealed that 99% of colonies formed were green and therefore plasmid positive when the improved protocol was used for induction. This was similar to approximately 96% green colonies generated when the standard protocol was
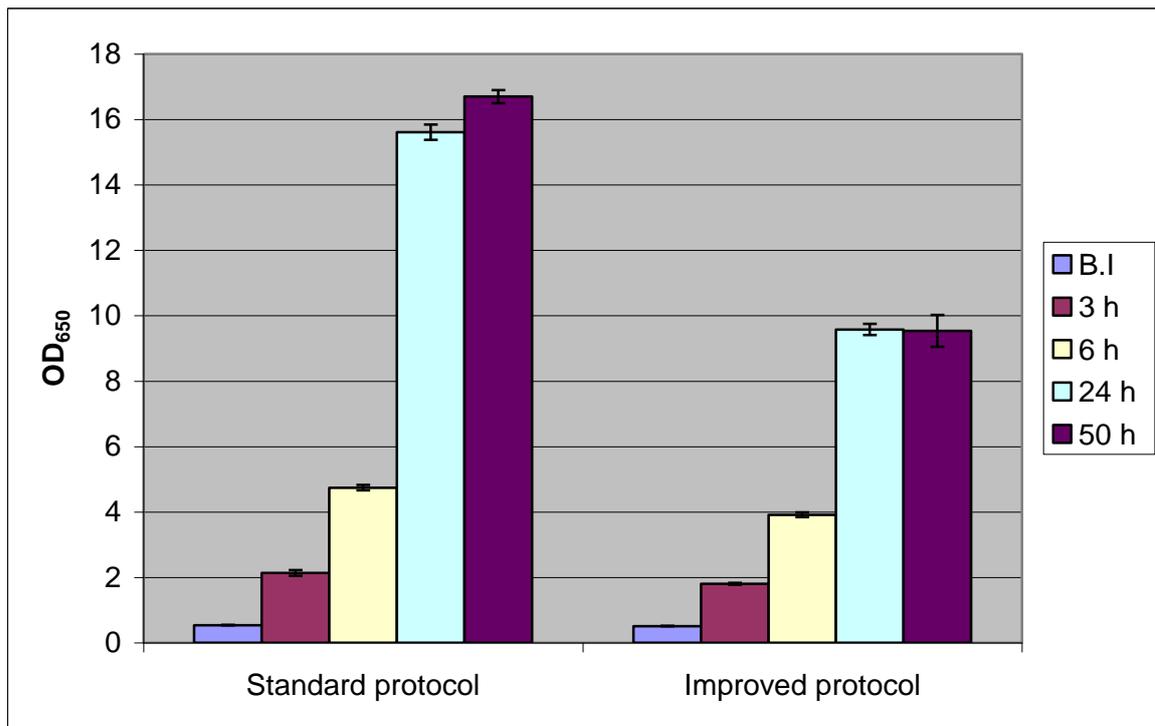
Figure 4.8: Comparison of growth of P2 BL21* during CheYGFP production using both protocols. Production of CheYGFP was induced in P2 BL21* using either the improved or the standard protocol. The optical density of both cultures was measured before induction (B.I) and at intervals post-induction. Standard error bars from triplicate repeats using 3 different colonies of each strain, and within one time series, are presented.

used for induction, and indicates that there was no significant physiological stress on P2 BL21* from either induction protocol.

The fluorescence of both cultures was also measured during the experiment. Total fluorescence 24 h after induction was only 400 units when the improved protocol was used for induction, suggesting that little soluble CheYGFP had been produced (**Figure 4.9**). This was compared to at least a 10-fold higher fluorescence when the standard protocol was used for induction, which had further increased 49 h after induction.

Total proteins from whole-cell samples that were processed relative to biomass were resolved by SDS-PAGE (**Figure 4.10**). There was no accumulation of CheYGFP when the improved protocol was used for induction, whereas continuous accumulation was detected when the standard protocol was used. These data suggest that combining both strategies of improved host and improved protocol has counterproductive effects on RPP.

Western blotting analysis of the T7 RNA polymerase in P2 BL21*CheYGFP

It has been demonstrated that P2 BL21* continues to accumulate soluble CheYGFP for up to 49 hours post-induction using the standard protocol. Unlike wild type BL21*, little physiological stress was shown by P2 BL21* post-induction using the same protocol, suggesting that the P2 host is resistant to the stress imposed. Since the source of stress in wild-type BL21* during RPP was due to the overproduction of T7 RNA polymerase, this was also investigated in P2 BL21* during CheYGFP production.
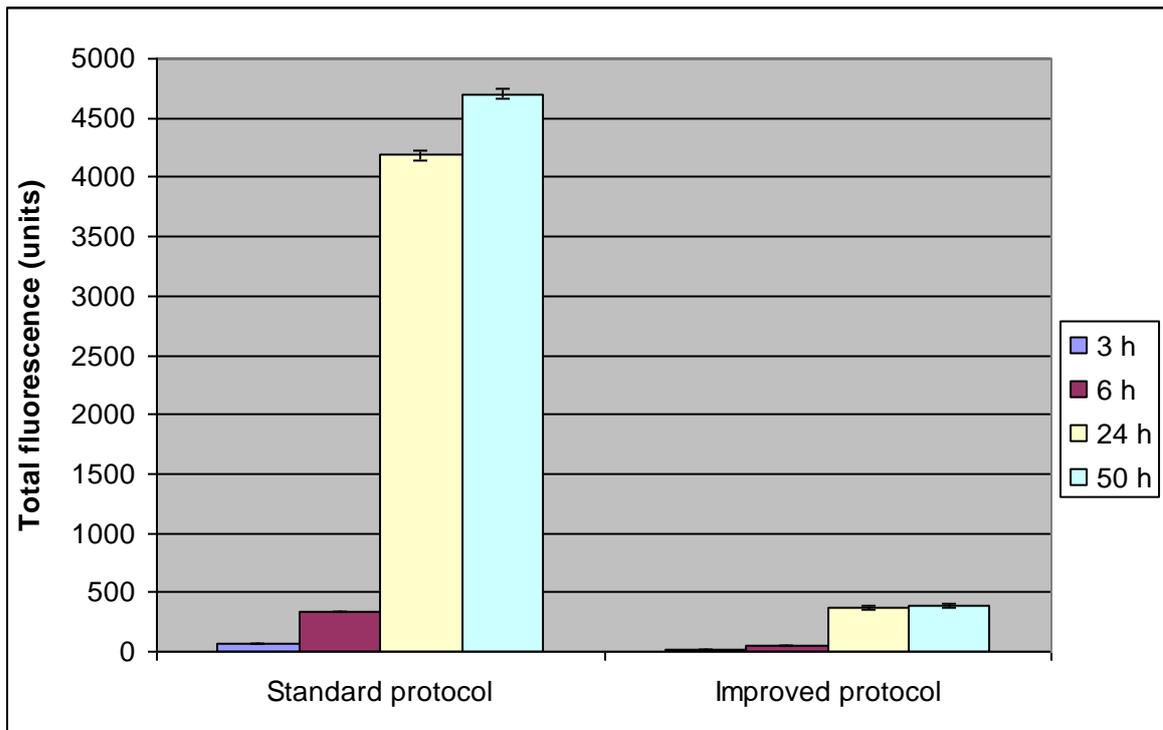
Figure 4.9: Comparison of total fluorescence of P2 BL21*CheYGFP post-induction using both protocols. Production of CheYGFP was induced in P2 BL21* using either the improved or the standard protocol. Total fluorescence was measured at intervals post-induction. Standard error bars from triplicate repeats using 3 different colonies of each strain, and within one time series, are presented.
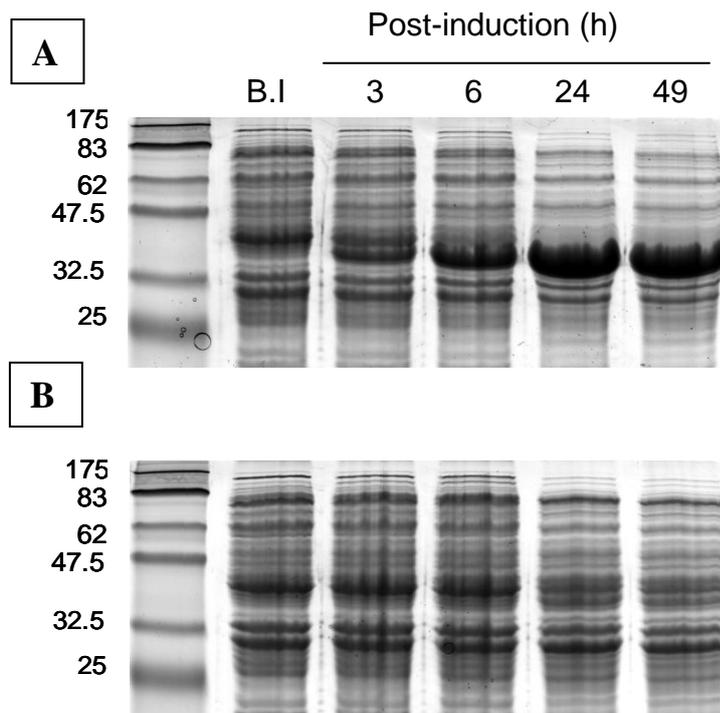
Figure 4.10: Comparison of yields of CheYGFP accumulated by P2 BL21* using both production protocols. Proteins from whole cell samples of P2 BL21*CheYGFP that was induced with either the standard protocol (A) or the improved protocol (B) were resolved by SDS-PAGE. Equal quantities of biomass from samples taken before induction (B.I) and at intervals post-induction were loaded into each lane. The sizes or marker proteins are shown.

Samples of P2 BL21*CheYGFP before and post-induction using the standard protocol were analysed by Western blotting and probed with anti-T7 RNA polymerase antibody (**Figure 4.11**). Control samples of wild type BL21*CheYGFP that had been induced with the standard protocol were included as a positive control. Strong bands of cross-reacting antigen with increasing intensity were evident post-induction in wild type BL21*, indicating the accumulation of high yields of T7 RNA polymerase. In contrast, there were no visible signals in samples of P2 BL21*, indicating that the expression of T7 RNA polymerase has been significantly down-regulated beyond detection point (**Figure 4.11**). It was therefore concluded that the down-regulation of T7 RNA polymerase expression in P2 BL21* removes the main stress source that is associated with the standard protocol and therefore minimises the negative impact on bacterial physiology during RPP. Hence, it is unsurprising that using the improved protocol with its very low [IPTG] for induction in P2 BL21* was counterproductive.

Production of CheYGFP in P2 BL21* in fermentations at pilot plant scale

The production of recombinant protein in shake flasks can be significantly compromised by the accumulation of acetic acid in the medium as glucose is metabolised. Since BL21* is more tolerant to acetic acid compared to other *E. coli* strains, it becomes difficult to distinguish between physiological responses of acid tolerance and those specific to recombinant protein production. Hence, it is important to investigate physiological responses of P2 BL21* to RPP in a pH-controlled fermentation.
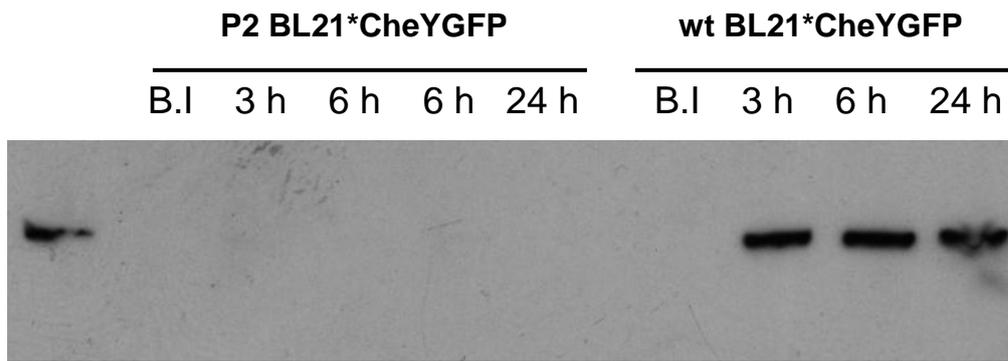
Figure 4.11: Western blotting analysis of P2 BL21*CheYGFP after induction using the standard protocol. Samples of P2 BL21*CheYGFP before induction (B.I) and at intervals post-induction were probed with anti-T7 RNA polymerase antibodies, and compared to control samples of the parental strain that had been induced with the standard protocol

The production of CheYGFP in P2 BL21* using the standard protocol was investigated in a pH-controlled batch 20 L fermenter at GlaxoSmithKline's pilot plant (**Figure 4.12**). This was compared to the production of CheYGFP in wild type BL21* using the improved protocol in a parallel fermentation. The fluorescence of the culture was monitored in real-time using an online GFP sensor and offline using a fluorescence plate reader. Unfortunately due to working hour constraints, the fermentation experiments could not be extended beyond 25 hours post-induction.

The final culture optical densities of P2 and wild-type BL21* 24 h post-induction was 26.1 and 16.4, respectively. Real-time fluorescence 24 h post-induction, as measured by the online GFP sensor was 35% greater for the P2 BL21*CheYGFP culture (2725 units) compared to when the wild type strain using the improved protocol (2010 units) (**Figure 4.13**). However, the offline fluorescence of P2 BL21*CheYGFP and the wild-type cultures reflected an almost 100% increase in the yield of soluble CheYGFP in P2 BL21* (**Figure 4.14; Table 4.1**). The discrepancy between the online and offline fluorescence measurements reflects the inaccuracy of online GFP measurements from undiluted cultures using the online GFP sensor. This suggests that the P2 BL21* host is better at the prolonged accumulation of soluble CheYGFP in pH-controlled fermentations than when using the improved protocol alone with the parental strain.

In an attempt to reproduce the above data, the same fermentation experiments were repeated using bench-top fermenters in which the pH was controlled, with the advantage of being able to extend the expression phase for up to 48 hours post-induction. There was almost a 100% increase in the total fluorescence 24 h post-induction in P2 BL21*

Figure 4.12: CheYGFP production using a 20 L Biolafitte fermenter with a 15 L working volume. The photograph shows the colour of the P2 BL21*CheYGFP culture 24 h post-induction using the standard protocol (*photograph courtesy of Dr Jo Jones, GlaxoSmithKline*).
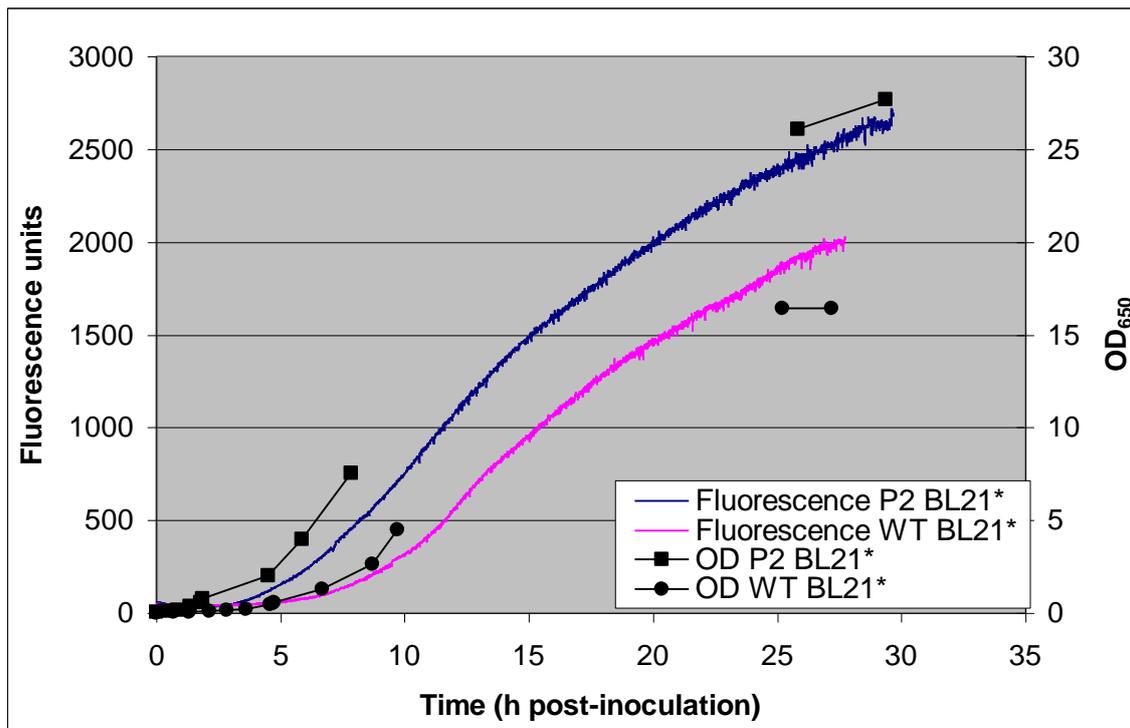
Figure 4.13: Online GFP fluorescence during CheYGFP accumulation when using either the improved host or improved protocol. Production of CheYGFP was induced using either the improved protocol in wild type BL21* or the standard protocol in P2 BL21* in pH-controlled 15 L batch fermentations. P2 BL21* and the wild type were induced at an OD of 0.5, which correspond to 2 h and 5 h post-inoculation, respectively. Total fluorescence of the cultures was measured online using a GFP sensor to reflect the real-time accumulation of soluble CheYGFP. Optical densities at 650 nm were measured before induction and at intervals post-induction.
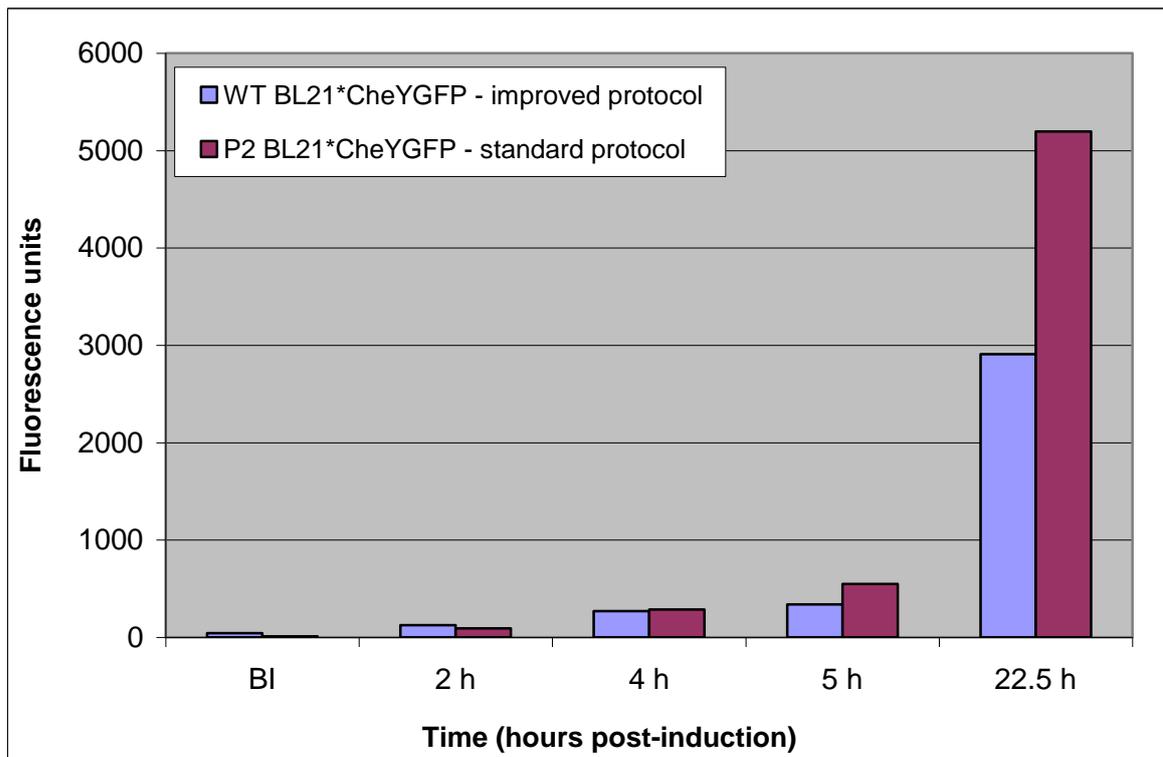
Figure 4.14: Comparing offline fluorescence of wild type and P2 BL21* strains during CheYGFP production in 15 L fermentation. Production of CheYGFP was induced in wild type and P2 BL21* using the improved and standard protocols, respectively, in pH-controlled batch fermentations. Total fluorescence was measured before induction (B.I) and at intervals post-induction using a plate-reader.

Table 4.1: Summarised data of CheYGFP production using the improved host or improved protocol in batch fermentations

		Wild type BL21*CheYGFP/ improved protocol				P2 BL21*CheYGFP/ standard protocol			
		OD	Total fluo (ml^{-1})	Sp fluo	Fluorescent CheYGFP $(g\ L^{-1})$	OD	Total fluo (ml^{-1})	Sp fluo	Fluorescent CheYGFP $(g\ L^{-1})$
3.6 litre bench top fermenters	24 h	17.2	3200	185	1.14	24.8	5300	215	1.90
	48 h	24.4	5680	235	2.04	38.6	11300	295	4.05
20 litre Biolaffite fermenters	24 h	16.4	2910*	180*	-	26.1	5720*	220*	-

CheYGFP was produced in bench-top and pilot plant scale fermenters. Average data of OD, total fluorescence (Total fluo) and specific fluorescence (Sp fluo) from duplicate benchtop fermenters are presented. * Fluorescence data from bench-top and Biolafitte fermentations were obtained using different fluorimeters. Only the Birmingham fluorimeter was accurately and reproducibly calibrated. The yields of fluorescent CheYGFP accumulated in benchtop fermenters were estimated using the GFP calibration value of 1mg/ ml = 2790 fluorescence units.

compared to using the improved protocol with the parental strain, which was due to a combined increase in bacterial density and the specific productivity of soluble CheYGFP (**Table 4.1**). This compares with only a 36% increase in shake-flask experiments and highlights the advantage of using pH control in batch fermentations. The total fluorescence in both cultures in batch fermentations continued to increase for 48 h post-induction, with the 2-fold difference maintained (**Table 4.1**).

The number of cfu of P2 BL21* and wild type BL21* progressively increased, reaching 5.9 $\times 10^{10}$/ ml of 93% plasmid positive cells and 1.95×10^{10}/ ml of 97% plasmid positive cells, respectively. This was almost identical to results obtained from shake-flask experiments.

The dissolved oxygen concentration (pO2) in the culture was monitored throughout the fermentation (**Figure 4.15**). The dissolved oxygen decreased upon induction to nearly 40% of the starting concentration by 6.5 h post-induction, and remained in the region between 40%-60% until 35 h post-induction. This indicated steady oxygen consumption as bacteria continue to grow and overproduce CheYGFP. However, the dissolved oxygen concentration in the culture remaining at around 40% for most of the fermentation indicates that other factors such as nutrient limitation (rather than oxygen availability) could be limiting a higher bacterial growth rate. The dissolved O_2 concentration increased to around 100% by the end of the fermentation indicating a cessation of bacterial growth.

Culture samples before induction and at intervals post-induction were analysed by flow cytometry (**Figure 4.16**). Propidium iodide was used to stain dead cells. Cells with green fluorescence, due to the production of CheYGFP were detected by PMT2, whereas cells
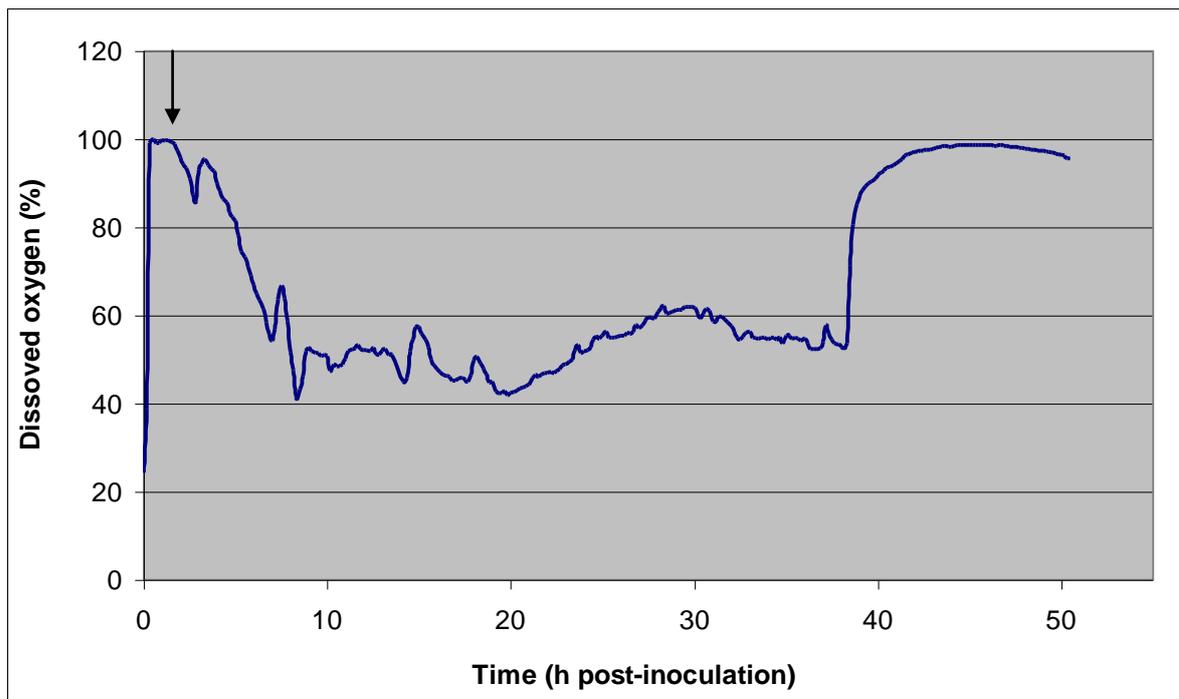
Figure 4.15: Dissolved oxygen concentration during production of CheYGFP in strain P2 BL21* in a fermenter. Strain P2 BL21*CheYGFP was grown and induced using the standard protocol in a benchtop fermenter. The dissolved oxygen concentration of the culture was measured using an oxygen probe from the point of inoculation. The time of induction is indicated by an arrow.
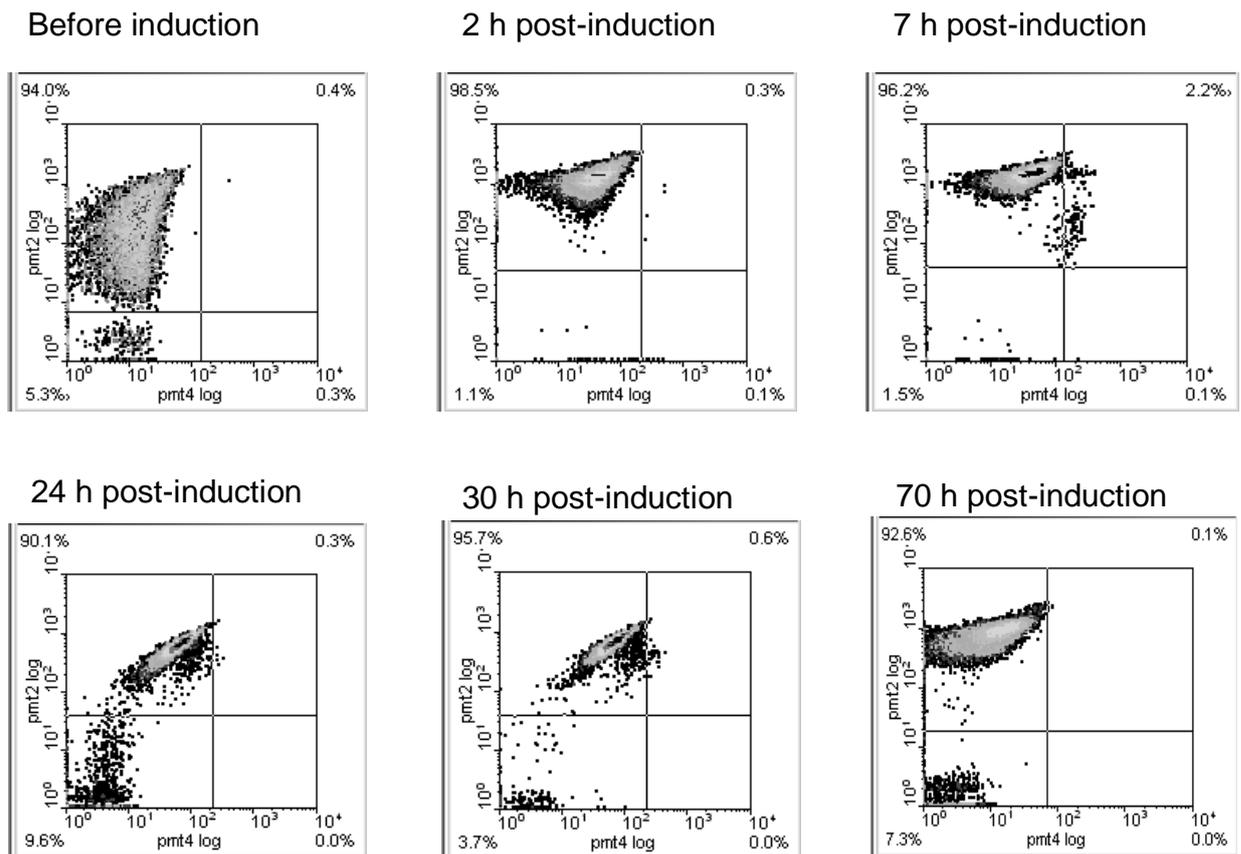
Figure 4.16: Flow cytometric analysis of P2 BL21*CheYGFP post-induction using the standard protocol. Production of CheYGFP was induced in P2 BL21* using the standard protocol in a bench-top fermenter. Culture samples were taken before induction and at intervals post induction. The red-fluorescent dye, propidium iodide (PI) was used to stain dead cells in the samples. Green fluorescence emitted from cells due to the accumulation of CheYGFP was detected by PMT2 whereas red fluorescence emitted from dead cells due to the uptake of PI was detected by PMT4. The proportions of different populations were calculated from a total sample of 25 000 bacteria.

with red fluorescence, due to uptake of propidium iodide, were detected by PMT4. Before induction, most of the bacteria were fluorescent due to the leaky expression of CheYGFP. There was also a minor population of bacteria with relatively little green fluorescence. There was no uptake of propidium iodide by cells of either population, reflecting intact outer membranes of healthy living bacteria. Shortly after induction, a population with increased green fluorescence intensity was detected, indicating that more soluble CheYGFP had accumulated within the cells. The intensity of green fluorescence of the population, as detected by PMT2, continued to increase until 30 h post-induction (**Figure 4.16**). This indicated that P2 BL21* had continued to accumulate soluble CheYGFP up to 30 h post-induction. This population remained at a majority of 92% even after 70 h post-induction.

It was concluded that fermentation experiments in which pH was controlled is important to accurately investigate RPP outcomes. In addition, since the basis for the improved RPP using the improved protocol and in P2 BL21* is the decreased levels of T7 RNA polymerase, then the increased growth density and specific productivity shown by P2 BL21* in batch fermentations could be attributed to additional chromosomal mutations. It is also apparent from the oxygen availability in fermentation experiments using strain P2 BL21* that there is scope for further improvement of RPP in fed-batch fermentations.

The production of CheYGFP in other improved BL21 hosts

The inefficiency of the BL21/ pET system for prolonged RPP led Miroux and Walker (1996) to isolate mutant hosts that were resistant to stress and therefore continued to accumulate recombinant protein far longer than their parent strain. Two of these strains

were called C41 and C43, with C43 reportedly being superior to C41 in the accumulation of recombinant membrane protein. It was subsequently shown that the basis for the improvement was due to mutations in the regulatory region of the promoter of the T7 polymerase gene that downregulated its expression. It was therefore interesting to compare yields of CheYGFP from strains P2 BL21*, C41 and C43 using the standard protocol with that of the wild type generated using the improved protocol.

The strain C41 had highest growth density at 9 h post-induction, and remained the best strain until 49 h post-induction with an average optical density of 19 units (or 7.6 g biomass/ L) (**Figure 4.17**). This was significantly more than that of the P2 BL21* host, indicating that strain C41 was more tolerant to stress. In contrast, the growth density of C43 and the wild type control were both around 13 OD units (5.2 g biomass/ L) 24 h post-induction, but had increased slightly by 49 h post-induction.

Specific fluorescence was highest in strain P2 BL21* at 260 units 24 h and 49 h post-induction (**Figure 4.18**). This was approximately 20% more than that of C41 and the wild type control, indicating that the P2 host accumulates higher yields of soluble CheYGFP. This also highlights the possibility of the presence of (other) mutations in P2 BL21* that distinguish it from the C41 strain. Interestingly, specific fluorescence was lowest in strain C43, indicating that this host, unlike C41, is a rather poor producer of soluble cytoplasmic CheYGFP. The combined effects of differences in growth density and specific fluorescence mean that the total yields of soluble CheYGFP obtained per volume of culture in P2 BL21* and C41 were similar, and significantly higher compared to using the improved protocol with BL21*, which in turn depends on the decreased expression of T7
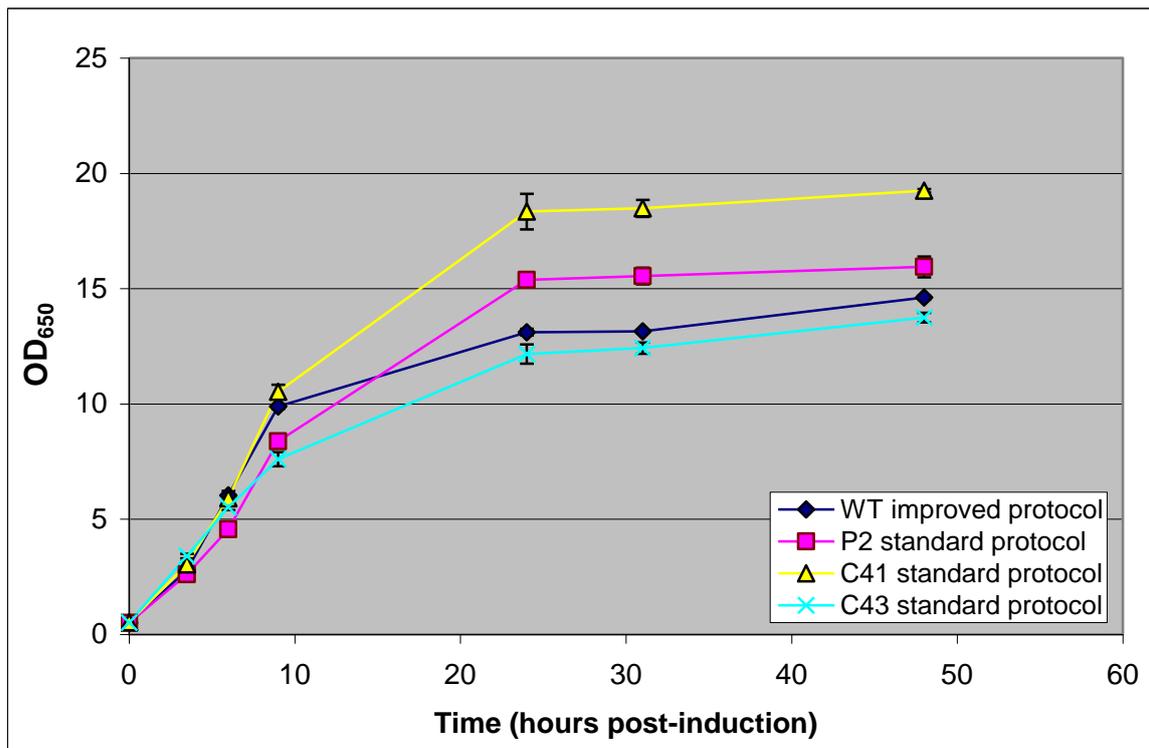
Figure 4.17: Comparing growth of different improved hosts during CheYGFP production. Production of CheYGFP was induced in P2 BL21* and the commercially available C41 and C43 improved hosts using the standard protocol. These were compared to control cultures of the parental strain that were induced with the improved protocol. Optical densities before induction and at intervals post-induction were measured. Standard errors from data obtained from triplicate cultures comprising different candidates of each strain, are shown.
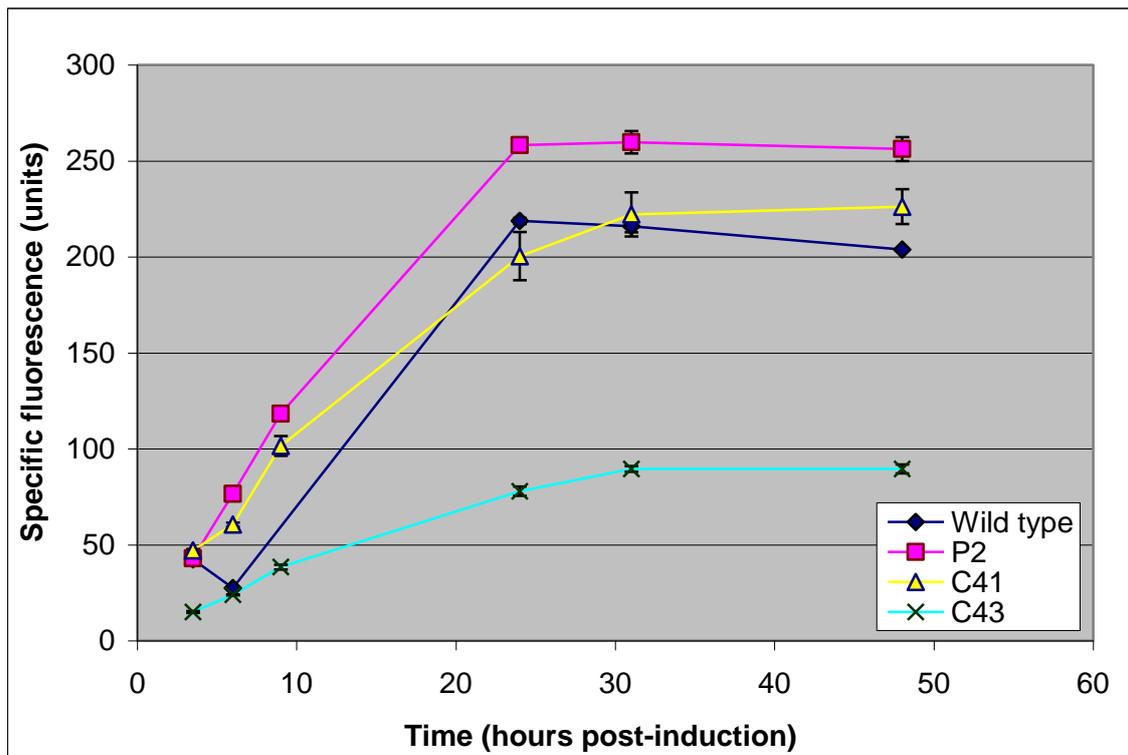
Figure 4.18: Specific fluorescence during CheYGFP production in different improved hosts. Production of CheYGFP was induced in P2 BL21*, C41 and C43 strains using the standard protocol and compared to cultures of the parental strain that were induced using the improved protocol. Specific fluorescence at intervals post-induction was calculated to reveal the specific productivity of soluble CheYGFP between different hosts. Standard errors from data obtained from triplicate cultures comprising different candidates of each strain, are shown.

RNA polymerase in the bacterial host (**Figure 4.19**).

Total protein from (whole cell) samples that were normalised for biomass were resolved by SDS-PAGE (**Figure 4.20**). There was a continuous accumulation of CheYGFP post-induction in both strains C41 and P2, with slightly higher yields accumulated in strain C41. The total yields of CheYGFP produced in either P2 or C41 strain were markedly higher than when using the improved protocol and strain BL21* alone, indicating that there are additional mutations in these improved hosts that lead to increased production of recombinant protein. In contrast, there was very little CheYGFP produced in strain C43.

To asses the quality of the CheYGFP accumulated, proteins from the soluble and insoluble cell fractions of cells taken 24 hours post-induction were resolved by SDS-PAGE (**Figure 4.21**). The majority of the CheYGFP produced using P2 BL21 or the improved protocol was present in the soluble fraction. However, approximately 30%, of the CheYGFP produced by C41 was present in the insoluble fraction. This might be an indirect consequence of the high yields of CheYGFP yields produced, which causes it to aggregate into inclusion bodies. In contrast, the majority of the low yield of CheYGFP produced by strain C43 was present in the soluble fraction.

In conclusion, there are clear differences between the abilities of the improved hosts to accumulate CheYGFP. This suggests that extra mutations, apart from the down-regulation of T7 RNA polymerase are present in improved host strains, which either enhance RPP as in P2 and C41 strains, or compromise RPP as in strain C43.
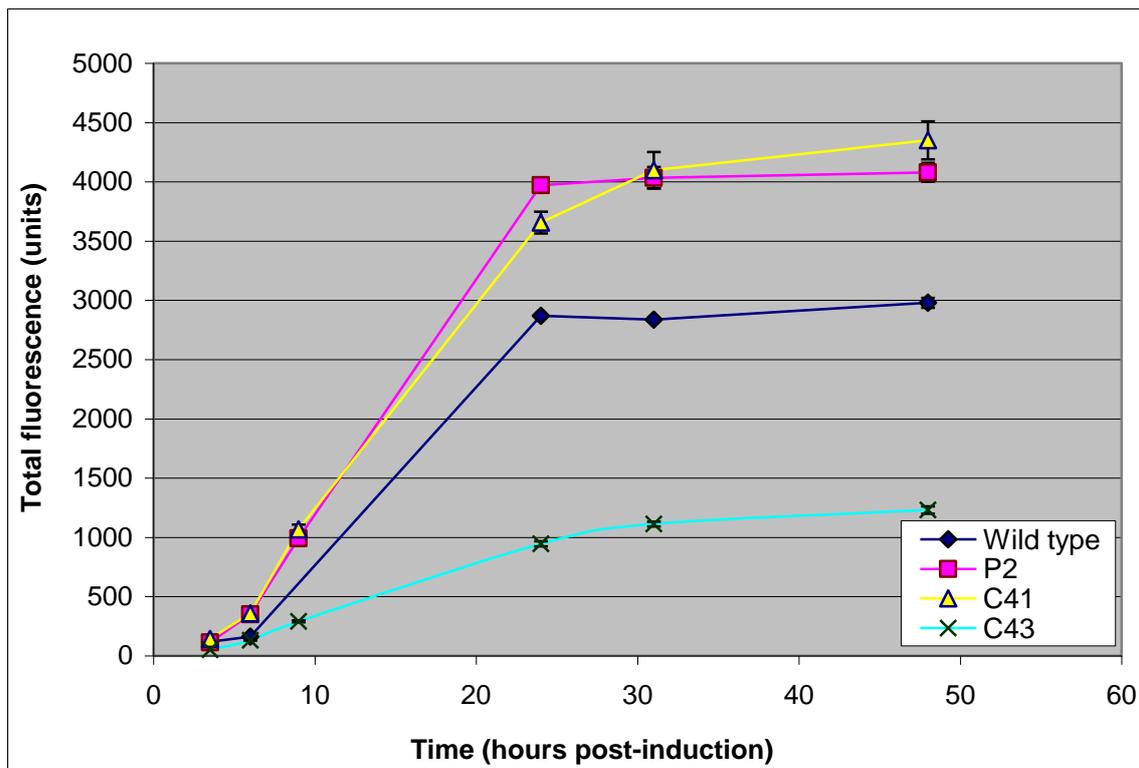
Figure 4.19: **Total fluorescence of different improved hosts during CheYGFP overproduction.** Production of CheYGFP was induced in P2 BL21*, C41 and C43 strains using the standard protocol and was compared to using the improved protocol with the parental strain. Total fluorescence was measured at intervals post-induction. Standard errors from data obtained from triplicate cultures comprising different candidates of each strain are shown.
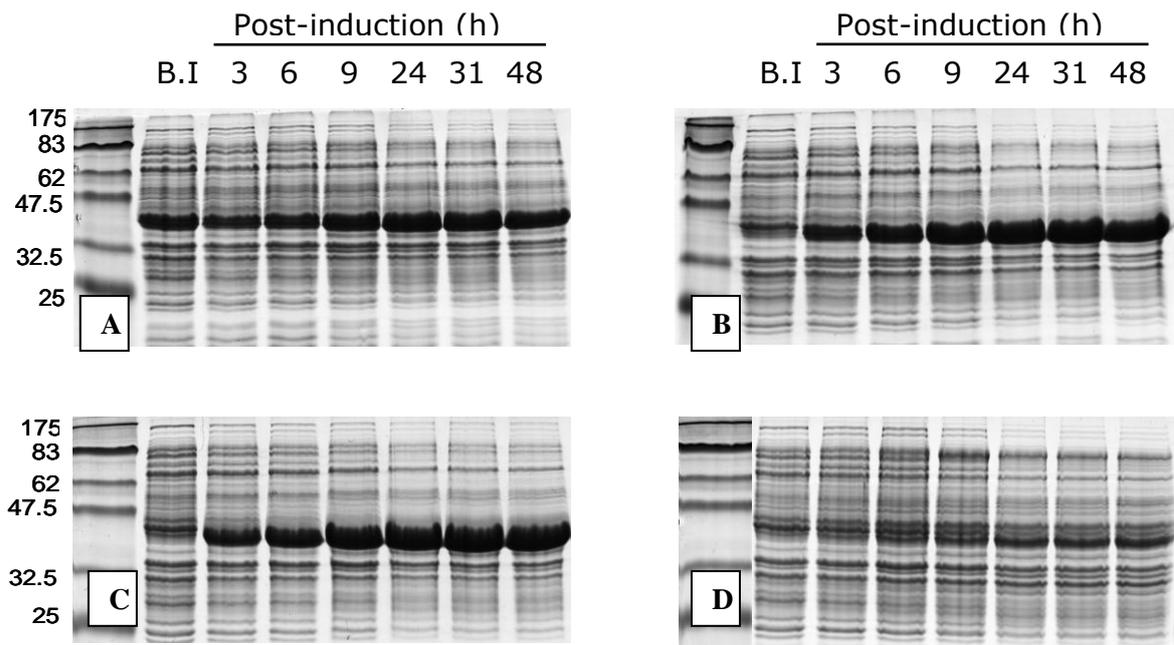
Figure 4.20: **Comparison of the yields of CheYGFP accumulated by different improved hosts or using the improved protocol alone.** Production of CheYGFP from strains P2 BL21* (B), C41 (C) and C43 (D) was induced using the standard protocol and compared to that of the parental strain that had been induced using the improved protocol (A). Total proteins from whole cell samples taken before induction (B.I) and at intervals post-induction were resolved on SDS-PAGE. Equal quantities of biomass were loaded into each lane. The sizes of marker proteins are shown.
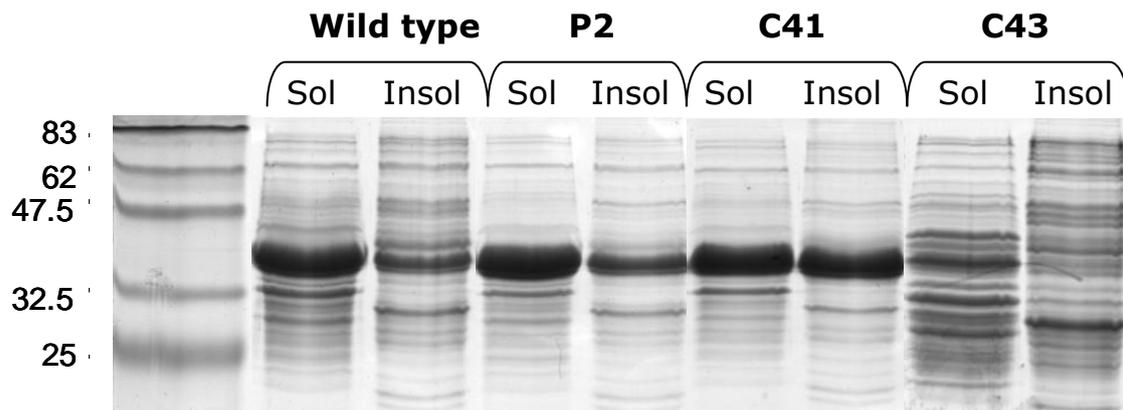
Figure 4.21: **Comparison of the yields of soluble and insoluble CheYGFP accumulated by different improved hosts.** Production of CheYGFP was induced in P2 BL21*, C41 and C43 strains using the standard protocol, and in the parental strain using the improved protocol as a control. Cell samples taken 24 h post-induction were fractionated into soluble (sol) and insoluble (insol) fractions, and proteins from within these fractions were resolved by SDS-PAGE. Equal quantities of cell biomass were loaded for analysis.

The production of protein D-GFP by wild type and P2 BL21* strains

It was essential to test whether the improved protocol or the improved host were generally applicable to the overproduction of different recombinant proteins. We were challenged by GlaxoSmithKline to produce a non-*E. coli* cytoplasmic protein that was GFP-tagged called protein D-GFP (45 kDa) (intellectual property; GSK). The production of protein D-GFP was therefore attempted using the improved protocol and the improved host P2 BL21*.

Fresh transformants of wild type and P2 BL21* with the pETprotein D-GFP fusion plasmid were induced using the improved and standard protocols, respectively. These were compared to control cultures of wild type BL21* transformants that were induced using the standard protocol.

The P2 BL21* transformants had the highest OD 24 hours post-induction at 12 units, followed by the wild type strain that was induced with the improved protocol at 10.5 units (**Figure 4.22**). This indicates that there was less stress when using an improved host compared to when using an improved protocol with the normal host. In contrast, bacterial growth of the control culture using BL21* under standard conditions was significantly lower at 24 hours post-induction, reflecting the stress imposed by the standard protocol.
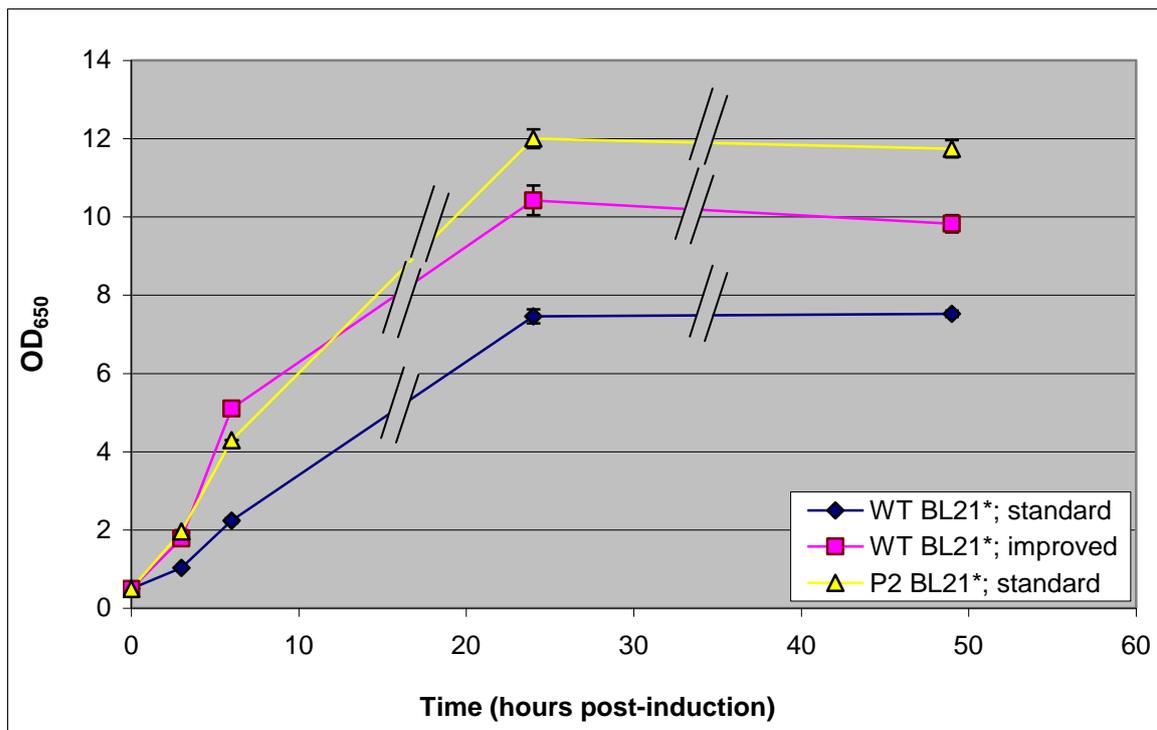
Figure 4.22: The effect of protein D-GFP production on growth of wild type and P2 BL21*. The accumulation of protein D-GFP was induced in P2 BL21* and the parental strain using the standard and improved protocols, respectively. These were compared to control cultures of the parental strain that had been induced using the standard protocol. Optical densities were measured before induction and at intervals post-induction. Standard errors from data obtained from duplicate cultures comprising different candidates of each strain, are shown.

The specific fluorescence was highest in the P2 BL21* transformants at approximately 300 units 24 hours post-induction and up to 49 hours post-induction (**Figure 4.23**). In comparison, using only the improved protocol, the specific fluorescence 24 h post-induction was significantly lower (by approximately 70%). This suggests that higher yields of soluble protein D-GFP had accumulated when using P2 BL21* than when using the improved protocol with the parental strain. There was a significant decrease in the specific fluorescence 49 hours post-induction when using the improved protocol with BL21*, indicating that the concentration of IPTG used was not optimal. Unsurprisingly, the specific fluorescence of the control culture progressively decreased post-induction, indicating that less soluble protein D-GFP was being accumulated within cells. This reflects the consequences of using the standard protocol for RPP.

Proteins from whole cell samples taken before induction and at intervals post-induction were resolved by SDS-PAGE (**Figure 4.24**). Production of a 45 kDa protein D-GFP was evident when either improved approaches were used for RPP, with high yields being accumulated until 49 h post-induction. Unlike the wild type strain, there was almost no D-GFP produced before induction in P2 BL21* reflecting tighter control of expression of the recombinant gene. There was a burst of protein D-GFP synthesis in the wild type host 3 h and 6 h post-induction when the standard protocol was used, but levels significantly decreased 24 h post-induction. Further analysis revealed that the majority of CheYGFP accumulated by P2 BL21* was present in the soluble fraction (not presented).
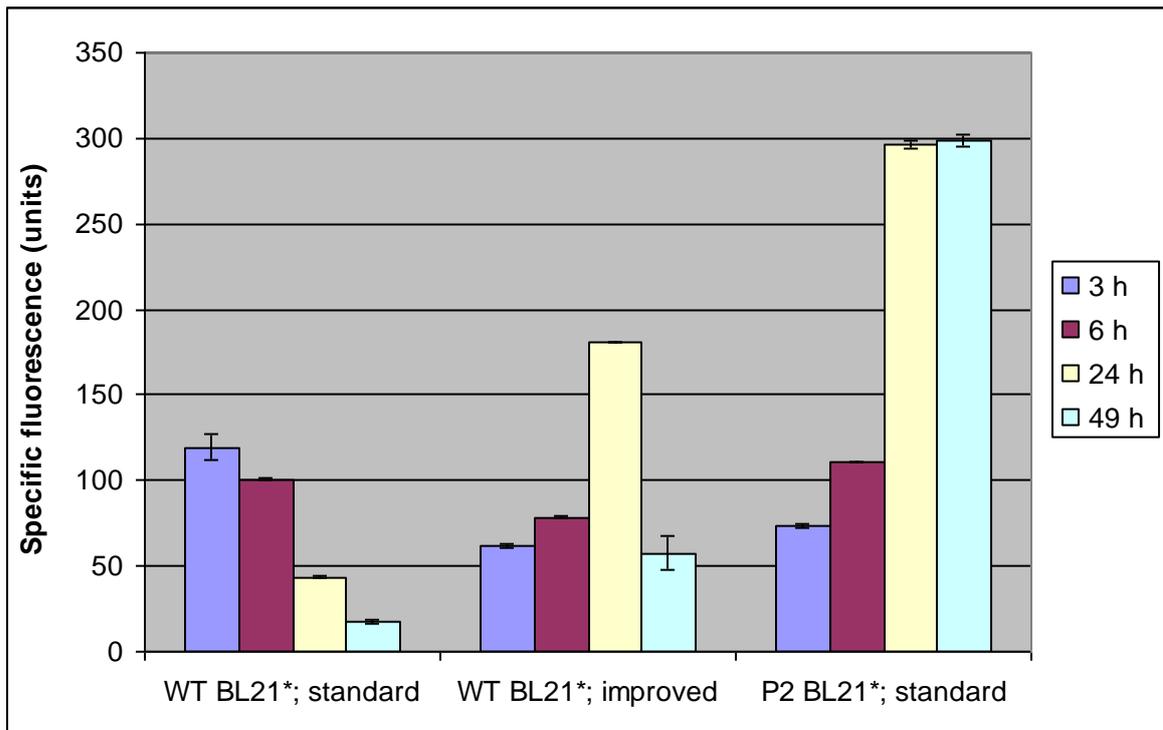
Figure 4.23: Specific fluorescence of wild type and P2 BL21* during protein D-GFP production. The production of protein D-GFP was induced in P2 BL21* and the parental strain using the standard and improved protocols, respectively. These were compared to control cultures of the parental strain that were induced using the standard protocol. The specific fluorescence was calculated at intervals post-induction. Standard errors from data obtained from duplicate cultures comprising different candidates of each strain are shown.
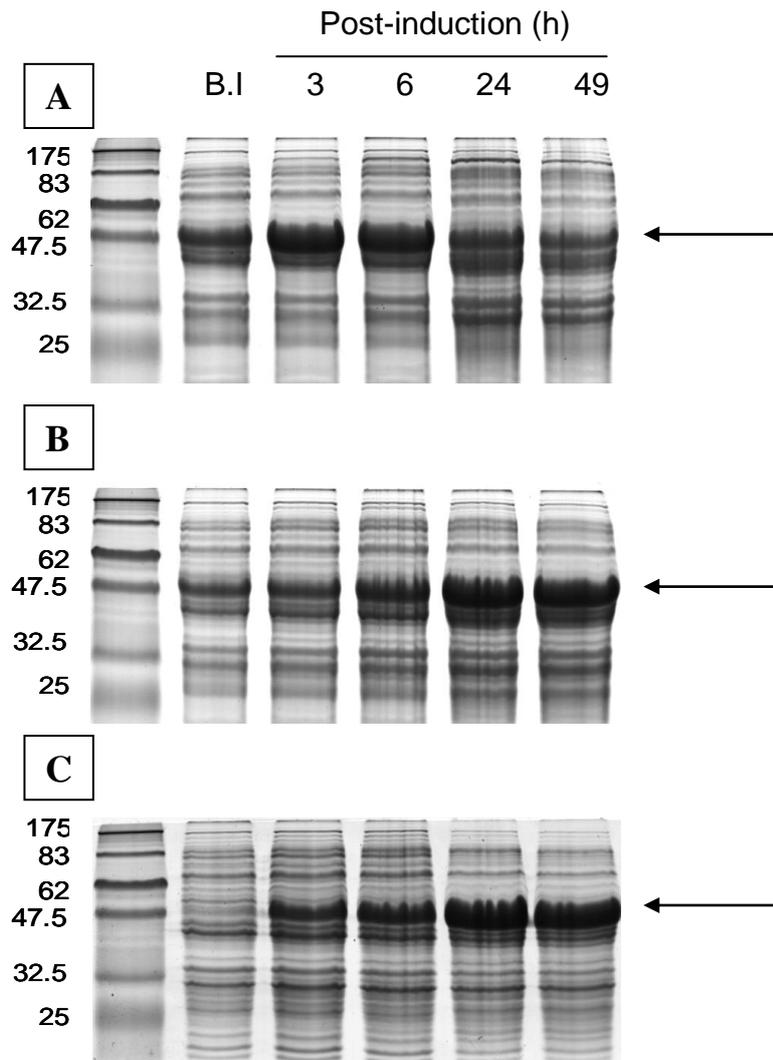
**Figure 4.24: Yields of protein D-GFP accumulated by wild type and P2 BL21*
strains.** Production of protein D-GFP was induced in P2 BL21* (C) and the parental strain
(B) using the standard and improved protocols, respectively. These were compared to a
control of the parental strain that was induced using the standard protocol (A). Total
proteins from whole cell samples taken before induction (B.I) and at intervals post-
induction were resolved by SDS-PAGE. Protein D-GFP is indicated by an arrow.

In conclusion, using either the improved protocol or an improved host significantly increases the efficiency of recombinant protein production than when using the standard protocol. In addition, using a suitable improved host for the over-production of a recombinant protein is often better than using an improved protocol.

Sequencing of the *lac*UV5 promoter region of the T7 RNA polymerase gene in the improved host, P2 BL21*

The down-regulation of T7 RNA polymerase has been shown to be key in the success of both the improved protocol and the stress-resistant host, P2 BL21*, the latter being due to a chromosomal mutation. It has been shown recently that mutations in the *lac*UV5 promoter and its regulatory region, which governs the expression of the T7 RNA polymerase, were present in C41 and C43 improved strains, resulting in improved characteristics in the production of recombinant proteins (Wagner *et al*., 2008). This prompted the sequencing of the *lac*UV5 promoter region in P2 BL21* to see whether the same mutations are present that also result in improved overproduction of cytoplasmic recombinant proteins.

Sequencing of the *lac*UV5 promoter revealed the presence of three mutations, two in the -10 element and one in the *lac* operator just upstream of the symmetric part of the *lac* repressor binding site and the transcription start site (**Figure 4.25**) These mutations are identical to those present in both C41 and C43, and suggests this to be a common way for a bacterial strain to evolve stress-resistance during RPP. However, mutations in the *lac*UV5 promoter do not eliminate the possibility of the presence of additional chromosomal mutations that result in improved RPP *per se* rather than just stress-resistance, as
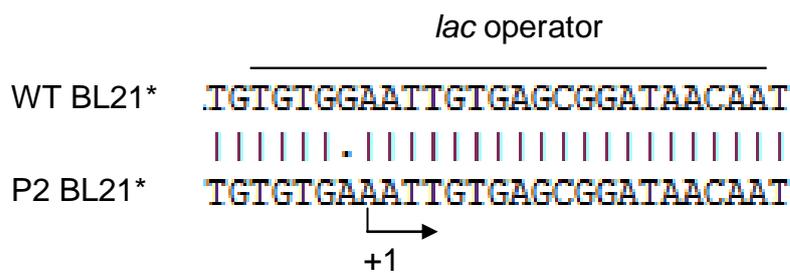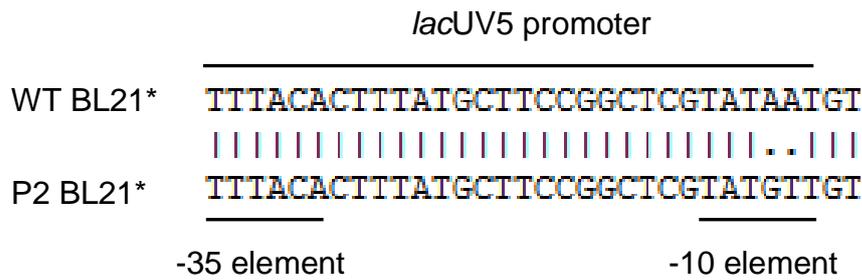
Figure 4.25: Nucleotide sequence of the _lac_UV5 promoter in P2 BL21*. The _lac_UV5 promoter fragment was sequenced from both P2 BL21* and the parental strain. Pair-wise alignment of sequences from the promoter and operator regions of _lac_UV5 in both strains is presented.

highlighted by differences in CheYGFP overproduction using different improved hosts.

The P2 BL21* was therefore sent to our collaborators at GSK for whole-genome sequencing. This is currently still in progress, so the results will not form part of this thesis.

CHAPTER 5

Production of the membrane lipoprotein cytochrome *c* peroxidase from *Neisseria gonorrhoeae* in *Escherichia coli*

Gonorrhoea is characterized by the inflammation of the sexual organs, which can lead to complications of infertility in both sexes and a possible development of bacteremia in approximately 1% of cases. The long incubation periods in males and its asymptomatic nature in females increases the prevalence and epidemiological risks within populations, accounting for an estimated global incidence of 62 million people infected annually (world health organisation, 2009). Despite efforts to increase awareness of the disease, its persistent prevalence heightens demands for vaccine development.

Several attempts at vaccine production were made using antigens such as porin proteins from the outer membrane or pili, which are essential for the attachment of gonococci to mucosal cell surfaces. Unfortunately, considerable antigenic variation of the pilus structure and the weak immune response elicited from using porin protein antigens have vastly decreased the efficacy and wide-spread use of the vaccine, which ultimately hampered vaccine development against gonorrhoea.

The outer membrane lipoprotein, cytochrome *c* peroxidase (CCP), from *N. gonorrhoeae* has been shown to play a key role in protecting the gonococcus against reactive oxygen species, which it encounters as a consequence of localized inflammatory responses (Turner *et al*., 2003). CCP was therefore proposed as a promising candidate for use as an antigen in subunit vaccine development against gonorrhoea. Research by Turner *et al*. (2003) revealed that the gonococcal CCP is first synthesised as a pre-apo lipoprotein with a signal peptide sequence and a total molecular weight of 47kDa. The signal peptide is cleaved by signal peptidase II as the pre-apo-protein is translocated across the cytoplasmic membrane to the periplasm (**Figure 1.4**). This results in the accumulation of an apo- form of CCP in

the periplasm. Once in the periplasm, two haem prosthetic groups are covalently attached to the apo-CCP before it is anchored loosely by its lipid group to the outer membrane. This mature form of CCP has a molecular weight of approximately 45kDa, and can be clearly distinguished from the 47 kDa pre-apo-protein on SDS-PAGE. Owing to the numerous post-translational modifications required for the production of its mature, native form, CCP is considered a 'difficult' protein that poses challenges to current procedures of RPP. Turner *et al*. (2003) successfully expressed recombinant *ccp* in *E. coli*, but the majority of the recombinant protein accumulated as an insoluble aggregate in inclusion bodies. In this chapter, the production of CCP and its effect on physiological responses on *E. coli* were investigated.

The production of pre-apo-CCP in *E. coli* BL21*

The production of CCP, as a pre-apo protein, was attempted using the T7/ pET expression system in BL21*, and its effects on bacterial physiology were investigated. Production of pre-apo-CCP was attempted using the standard and improved protocols. Expression of *ccp* was induced from a T7*lac* promoter on a pST203 plasmid, which is a derivative of the pET-11c expression vectors. The presence of a *lac* operator site downstream of the T7 promoter in pST203 provides additional regulation of gene expression compared to the T7 promoter alone in pET20hbc. Therefore, similar to the overproduction of CheYGFP in the improved protocol, it was necessary to find an optimal IPTG concentration range that will induce RPP, but have little consequences on bacterial physiology.

Freshly transformed cultures of BL21* pST203 were grown aerobically at 25°C in shake flasks and induced at an OD of 0.5 with different IPTG concentrations (0, 0.01, 0.03, 0.05 and 0.1 mM). These were compared to control cultures that were induced using the standard protocol. Samples were taken to monitor bacterial growth, cfu, plasmid retention and protein accumulation.

Bacterial growth was similar in all cultures 3 h after induction with optical densities of around 2, but there was a slightly lower optical density when the standard protocol was used for induction, reflecting the impacts of physiological stress associated with this protocol (**Figure 5.1**). Optical densities when bacteria were either not induced or induced with 0.01 mM IPTG had both increased to approximately 9 by 24 h post-induction, indicating that very little physiological stress was imposed. However, there was a significant decrease in the optical density following prolonged induction with 0.03 mM IPTG, suggesting that considerable physiological stress had been introduced that compromised growth. Higher final optical densities of cultures were obtained when 0.05 and 0.1 mM IPTG were used for induction (compared to 0.03 mM), indicating the recovery of bacterial growth. Growth of the control cultures also continued to increase, reaching optical densities of around 11 units by the end of the experiment (**Figure 5.1**).

Numbers of cfu were similar in all cultures 3 h post-induction (**Figure 5.2**). However, the number of cfu decreased by more than 2-fold 24 h post-induction when bacteria were induced with 0.01 mM IPTG compared to the uninduced control, reflecting significant physiological stress despite the very low levels of IPTG used for induction. This suggests that the accumulation of CCP is more stressful than that of CheYGFP on the host. There
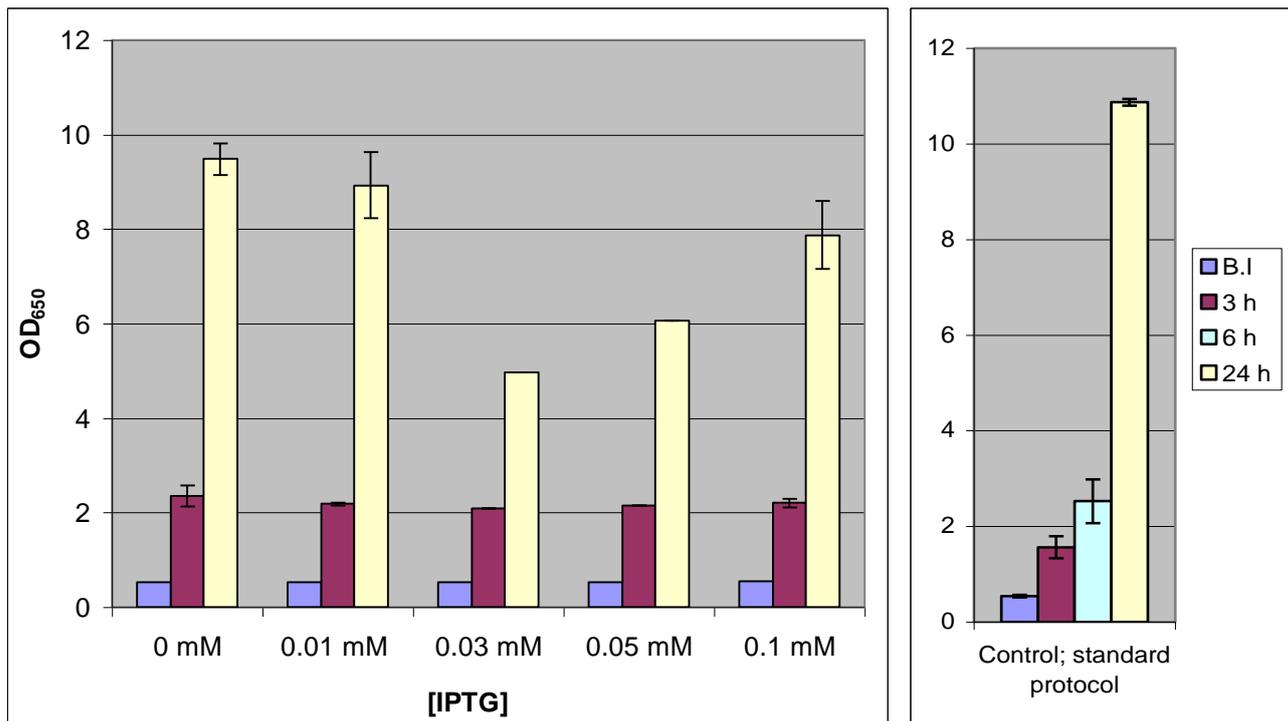
Figure 5.1: **The effect of IPTG inducer concentrations on growth densities of BL21*pST203**. Freshly transformed BL21*pST203 was grown at 25°C and the production of CCP was induced with different concentrations of IPTG. These were compared to control cultures that were induced using the standard protocol. Samples were taken before induction (B.I), 3 h, 6 h (for control only) and 24 h post-induction and the optical density was measured. Standard error from duplicate repeats of the titration experiment and triplicate repeats of the control, using different candidates of the strain, is presented.
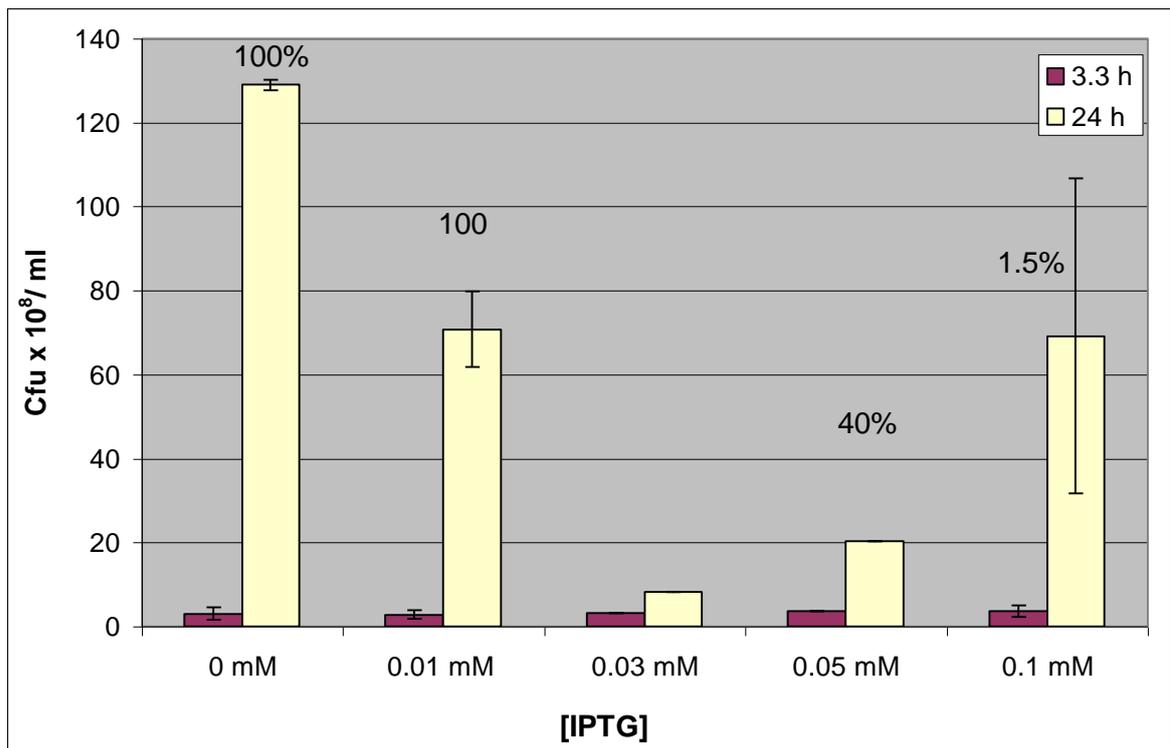
Figure 5.2: The effect of IPTG concentration on cfu of BL21*pST203. Production of

pre-apo-CCP was induced in freshly transformed BL21* pST203 with different

concentrations of IPTG at 25°C. Serial dilutions of culture samples taken before induction

(B.I), 3.3 h and 24 h after induction were plated on to nutrient agar, and the number of

colony forming units was calculated. Colonies obtained 24 h post-induction were replica

plated on to carbenicillin-supplemented NA and the % plasmid retention (shown above

bars of cfu data) was calculated. Standard error from duplicate repeats using different

candidates of the strain is presented.

was a decrease in cfu by a further 6-fold when bacteria were induced with 0.03 mM IPTG, suggesting a positive correlation between the induction of pre-apo-CCP synthesis and levels of physiological stress. However, there was an increase in the number of cfu by 2-fold 24 h post-induction with either 0.05 mM or 0.1 mM IPTG. When the standard protocol was used for the production of pre-apo-CCP, the number of cfu had decreased by more than 10-fold after only 3 hours post-induction, highlighting the significant physiological stress that is characteristic of this protocol (**Figure 5.3**). Compared to 3 h after induction, there was an increase in the number of cfu 6 h and 24 h post-induction in these control cultures.

Colonies formed 24 h post-induction were replica plated on to selective medium and the % plasmid retention was calculated. There was 100% plasmid retention by bacteria when the culture was either uninduced or induced with only 0.01 mM IPTG, despite the significant decrease in the number of cfu in the latter (**Figure 5.2**). However, only 40% of bacteria that had been induced with 0.05 mM were plasmid-positive, suggesting that the recovery shown by the increased number of cfu was due to the overgrowth by non-productive plasmid-free cells. There was complete plasmid loss when bacteria were induced with either 0.1 mM IPTG or with the standard protocol, re-enforcing the negative consequences of induction by high concentrations of IPTG on bacterial physiology (**Figure 5.2; 5.3**). These data suggest that the [IPTG] that has little negative impact on bacteria is around 0.01 mM.

Proteins from samples taken 3 h and 24 h post-induction were resolved by SDS-PAGE to assess the accumulation of pre-apo-CCP (**Figure 5.4**). There was no accumulation of pre-
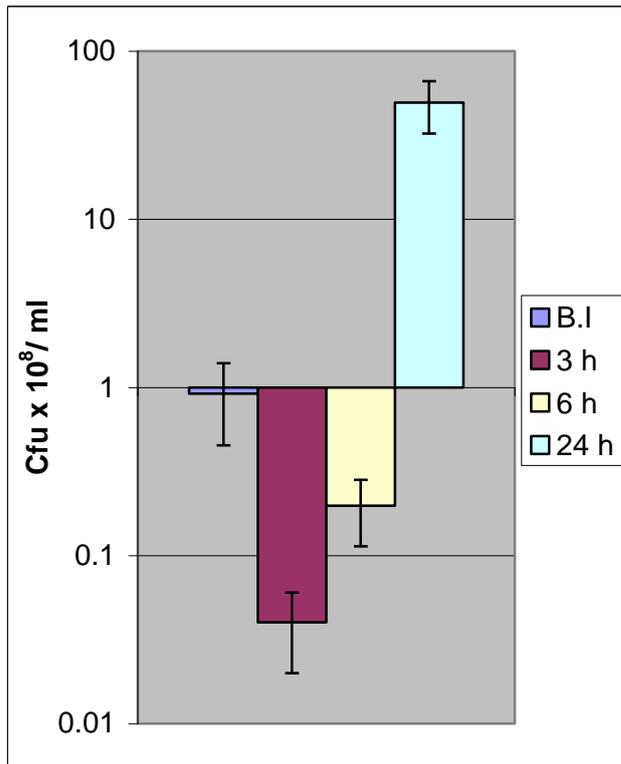
Figure 5.3: **The effect of pre-apo-CCP production on cfu of BL21* using the standard protocol.** Production of pre-apo-CCP was induced from freshly transformed BL21* pST203 using the standard protocol. Serial dilutions of samples taken before induction (B.I), 3 h, 6 h and 24 h were plated on to NA, and the cfu were calculated. Standard error from triplicate repeats using different candidates of the strain is presented.
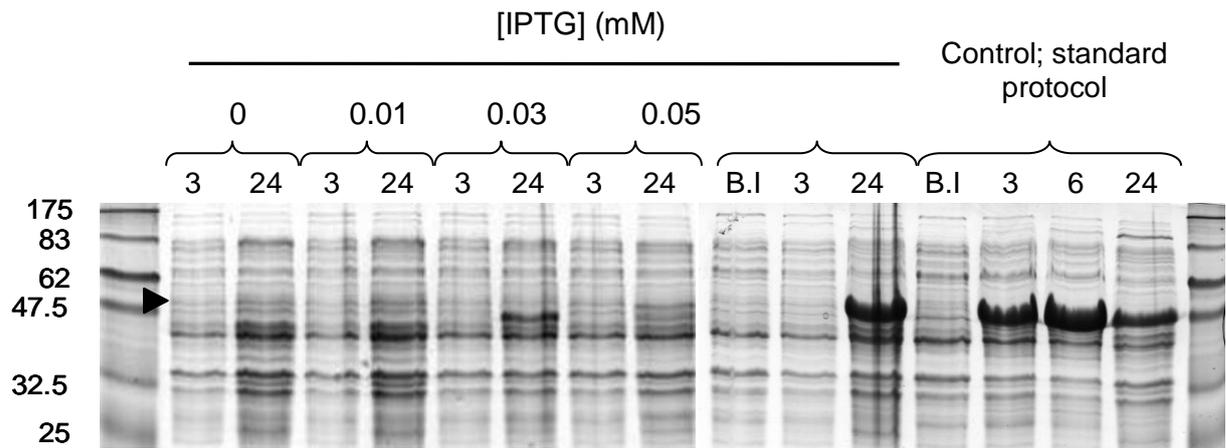
Figure 5.4: The accumulation of pre-apo-CCP using the standard and improved protocols. Production of pre-apo-CCP was induced in freshly transformed BL21* pST203 using either the improved protocol with different IPTG concentrations or the standard protocol as a negative control. Total proteins from whole cell samples taken 3 h and 24 h post-induction were resolved by SDS-PAGE. Equal quantities of bacterial biomass were loaded and the accumulated pre-apo-CCP is indicated.

apo-CCP when bacteria were either uninduced (not shown) or induced with 0.01 mM. However, a very faint band of 47 kDa was detected 24 h post-induction when bacteria were induced with 0.03 mM IPTG. Relatively high yields of pre-apo-CCP accumulated when bacteria were induced with 0.1 mM IPTG 24 h post-induction. The increasing accumulation of pre-apo-CCP positively correlated with the physiological responses of bacteria shown by colony formation ability and plasmid retention. No pre-apo-CCP was detected in all cultures induced with ≤ 0.1 mM IPTG 3 h post-induction, indicating a very slow accumulation rate. In contrast, there was a rapid burst of pre-apo-CCP synthesis when bacteria were induced using the standard protocol, with a clear band of pre-apo-CCP visible after only 3 h post-induction that had slightly increased in yield 6 h post-induction (**Figure 5.4**). However, yields of pre-apo-CCP per unit of biomass had decreased 24 h post-induction, which is characteristic of the overgrowth of plasmid-free cells.

To assess the quality of the accumulated pre-apo-CCP, proteins from the soluble and insoluble cell fractions were resolved by SDS-PAGE. The majority of pre-apo-CCP was detected in the insoluble fraction when either 0.05 mM, 0.1 mM IPTG, or when the standard protocol was used for induction (**Figure 5.5**). However, very low yields of pre-apo-CCP were also detected in the soluble fractions in bacteria that had been induced with 0.03 and 0.05 mM IPTG, reflecting the difficulty associated with the production of this protein in its soluble form.

These data confirm the negative impacts of using the standard protocol for RPP with pre-apo-CCP and suggest that the [IPTG] range that offers a compromise between minimising
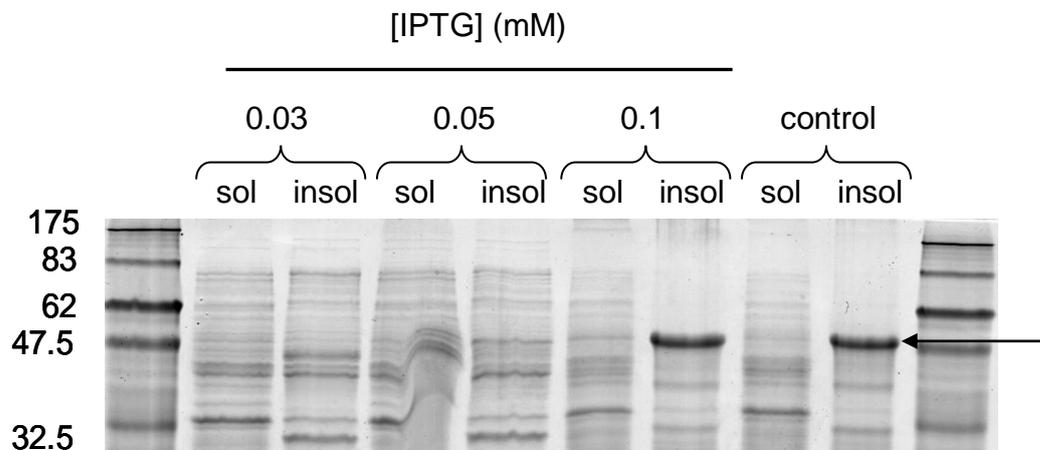
Figure 5.5: Yields of soluble and insoluble pre-apo-CCP accumulated by BL21* **using the improved and standard protocols.** The production of pre-apo-CCP was induced from pST203 in BL21* using either the improved protocol with different IPTG concentrations or the standard protocol as a negative control. Proteins from the soluble (sol) and insoluble (insol) cell fractions from samples taken 24 h post-induction were resolved by SDS-PAGE. Equal quantities of biomass were loaded into each well, and the accumulated pre-apo-CCP is indicated by an arrow.

the physiological impact on bacteria and inducing RPP is in the range of 0.01 and 0.03 mM.

Optimisation of the inducer concentration for the production of mature CCP using the improved protocol

The production of pre-apo-CCP has been shown to have considerable impact on bacterial physiology despite the very low IPTG concentrations used for induction. However, it is the accumulation of gonococcal CCP in its mature form that is required. This entails many steps of post-translational modifications, including the attachment of haem in the periplasm, which will inevitably result in even higher levels of stress on the host. The accumulation of mature CCP may therefore be facilitated by the addition of a second plasmid, pST2, which constitutively expresses genes required for cytochrome *c* maturation, *ccm*, from the endogenous *ccmA* promoter. The production of mature CCP was therefore attempted using the improved protocol, with a range of low IPTG inducer concentrations. This was compared to using the standard protocol for mature CCP production.

Fresh transformants of BL21* with plasmids pST203 and pST2 were grown at 25ºC and induced at an OD of 0.5 with different IPTG concentrations (0.01, 0.025, 0.04 and 0.05 mM). Samples were taken before induction and 5.5 h and 24 h post-induction. Growth densities were similar 5.5 h post-induction in most cultures at around 3.5 units, but slightly lower when bacteria were induced with 0.05 mM IPTG (**Figure 5.6**). However, the optical density 24 h post-induction was highest at around 7.6 units when bacteria were induced with only 0.01 mM IPTG, but decreased to approximately 5 units when IPTG
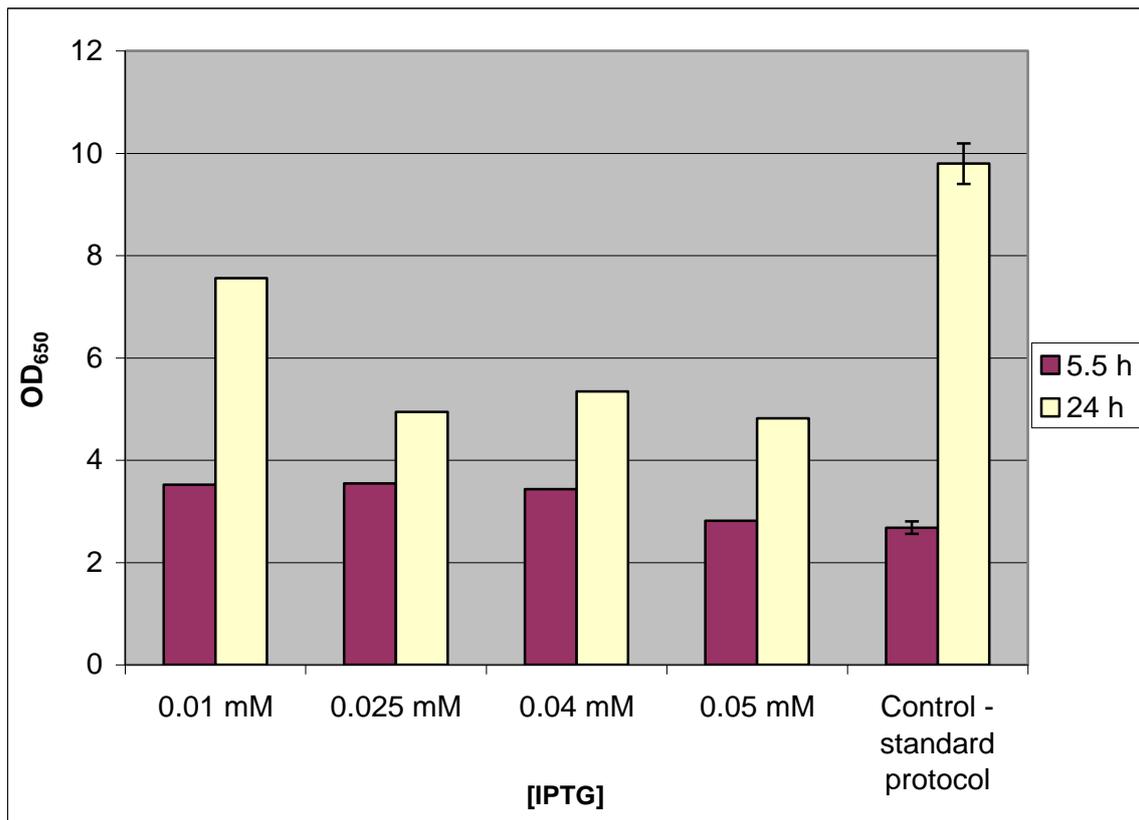
Figure 5.6: The effect of mature CCP production on growth of BL21*. The production of mature CCP was induced from freshly transformed BL21* with plasmids pST203 and pST2, using either the improved protocol with different [IPTG], or the standard protocol. The synthesis of pre-apo-CCP was induced from pST203 and its maturation was facilitated using pST2, which constitutively expresses genes for cytochrome *c* maturation. Optical densities 5.5 h and 24 h post-induction were measured. Standard error from triplicate repeats of the control, using different candidates of the strain, is presented.

concentrations of 0.025 mM or more were used for induction. This indicates that the physiological stress is present when mature CCP accumulation is induced with [IPTG] higher than 0.01 mM. However, the final optical density was almost 10 when the standard protocol was used for induction.

Numbers of colony forming units were calculated from serial dilutions of culture samples taken 5.5 h and 24 h post-induction (**Figure 5.7**). There was a slight increase in the number of cfu 5.5 h post-induction when higher [IPTG] were used for induction. In contrast, there was a decrease in the number of cfu when bacteria were induced using the standard protocol, as compared to before induction. The number of cfu 24 h post-induction, however, was highest at 84×10^8/ ml when bacteria were induced with the lowest [IPTG]. The number of cfu decreased by approximately 50% when 0.025 mM IPTG was used for induction. There was a further decrease in cfu with increasing IPTG inducer concentrations, indicating the positive correlation between increased induction and physiological stress on the host. In contrast, the number of cfu recovered when bacteria were induced using the standard protocol was approximately 55×10^8/ ml.

Colonies from all cultures were replica plated on to nutrient agar supplemented with either carbenicillin or chloramphenicol to calculate the % retention of each of the over-expression plasmid pST203 and pST2, respectively. There was complete retention of both plasmids 5 h post-induction in all cultures. The retention of pST203 by bacteria 24 h post-induction remained at 100% when 0.01, 0.025 or 0.04 mM IPTG were used for induction (**Figure 5.7**). This, however, had decreased to approximately 31% with a slightly higher IPTG concentration of 0.05 mM, suggesting that this IPTG concentration was sufficient to
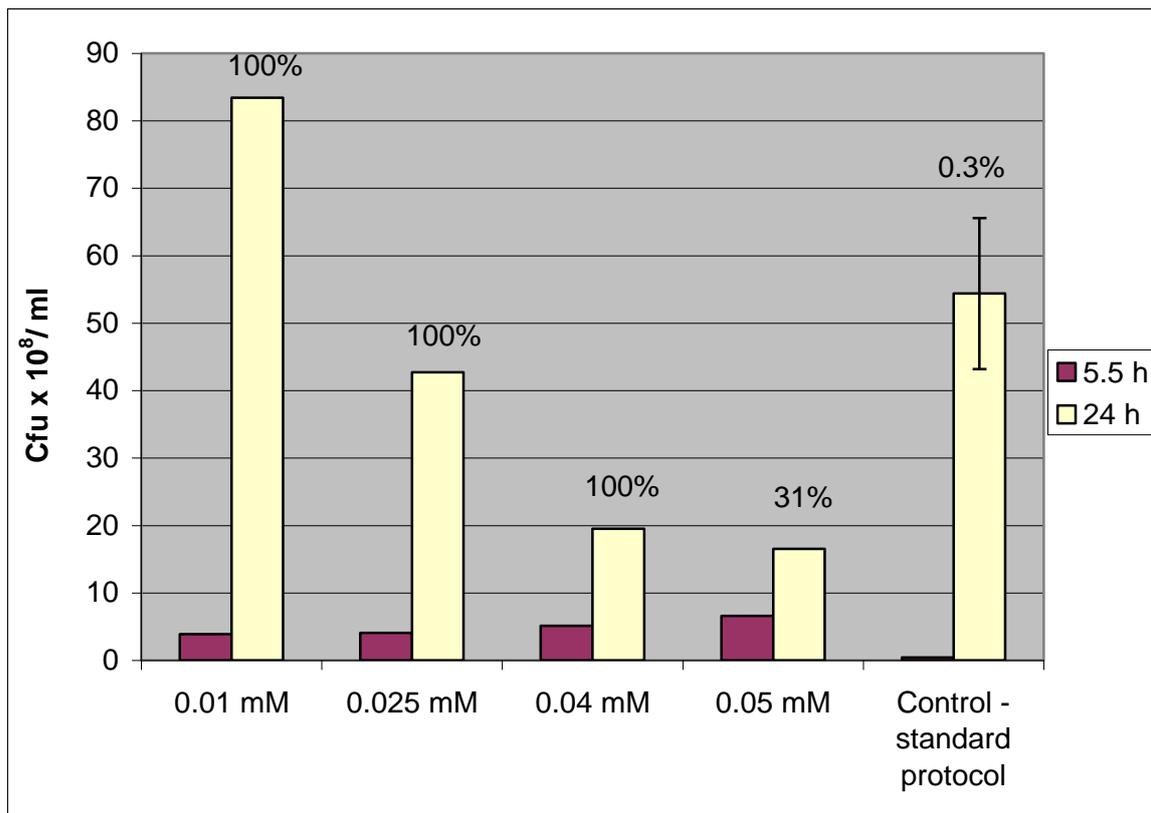
Figure 5.7: The effect of mature CCP accumulation on colony forming units of BL21*. The accumulation of mature CCP was induced in freshly transformed BL21* pST203 pST2 using either the improved protocol with different concentrations of IPTG or the standard protocol. Serial dilutions of culture samples taken 5.5 h and 24 h post-induction were plated on to non-selective NA and the cfu were calculated. Colonies obtained 24 h post-induction were replica plated on to carbenicillin-supplemented NA and the % retention of pST203 was calculated (presented above cfu data). Standard error from triplicate repeats of the control, using different candidates of the strain, is presented.

introduce stress on the host that induced the overgrowth by plasmid-free cells in the culture. These data also indicate a very small margin between effective inducer concentrations and those that have counter-productive effects on bacterial physiology. There was almost complete loss of pST203 from bacteria induced using the standard protocol, reflecting the inevitable overgrowth by plasmid-free bacteria (**Figure 5.7**). The pST2 plasmid was retained by bacteria until 24 h post-induction regardless of the IPTG inducer concentration.

Total proteins from whole cell samples taken before induction and at intervals after induction were resolved by SDS-PAGE. The gel was first stained for haem-dependent peroxidase activity to detect the accumulation of mature CCP (**Figure 5.8 A**). A band of approximately 45 kDa corresponding to mature CCP was detected 5.5 h post-induction, and increased in intensity by 24 h when 0.01, 0.025 or 0.04 mM IPTG were used for induction, reflecting the prolonged accumulation of mature CCP. The highest yields of mature CCP at both 24 h and 48 h post-induction were detected when bacteria were induced with 0.04 mM IPTG. However, yields of mature CCP were considerably lower when bacteria were induced with 0.05 mM IPTG. Yields had decreased further by 48 h post-induction (**Figure 5.8 A**). This indicates that the physiological stress which resulted in the overgrowth by plasmid-free bacteria had ultimately compromised yields of mature CCP accumulated. Additional bands with peroxidase activity that might be degradation products from the accumulated recombinant protein were also detected.

The accumulation of mature CCP by bacteria using the standard protocol was also assessed (**Figure 5.9 A**). A very faint 45 kDa band was detected 5.5 h post-induction, indicating
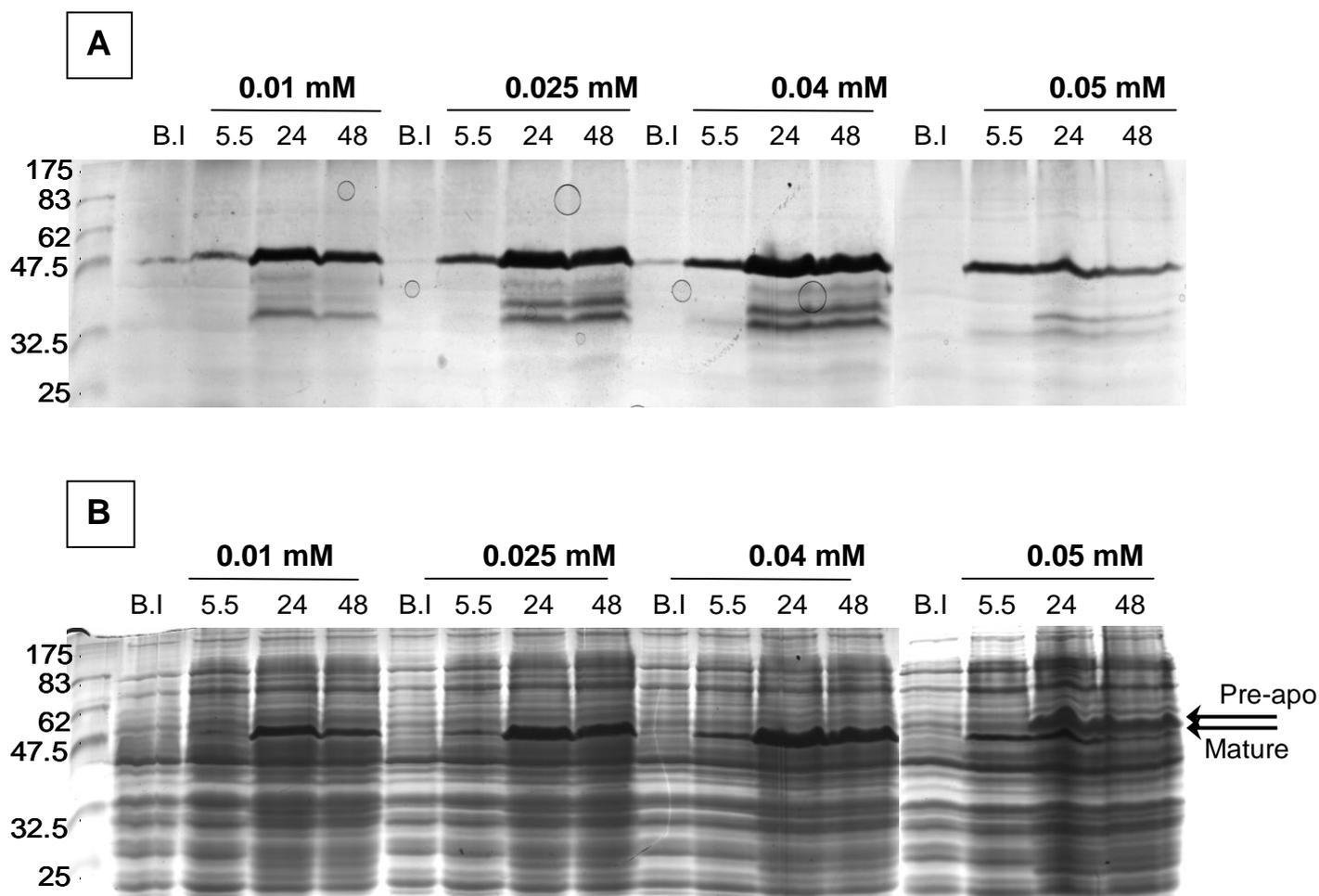
Figure 5.8: The accumulation of mature CCP in BL21* using the improved protocol.
Overproduction of mature CCP was induced in freshly transformed BL21* pST203 pST2 using the improved protocol with different concentrations of IPTG. Total proteins from whole cell samples taken before induction (B.I), 5.5 h, 24 h and 48 h post-induction were resolved by SDS-PAGE. The gel was stained first for haem-dependent peroxidase activity to detect the accumulation of the 45 kDa mature CCP (A); and with Coomassie blue to reveal the total cellular protein, including the 47 kDa pre-apo-CCP (B).
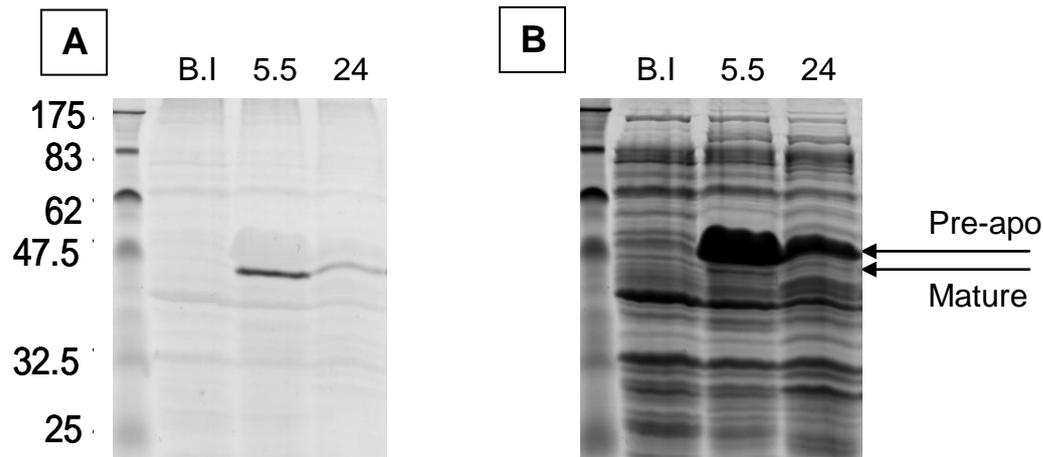
Figure 5.9: The accumulation of mature CCP in BL21* using the standard protocol.
Production of mature CCP was induced in freshly transformed BL21* pST203 pST2 using
the standard protocol. Total proteins from whole cell samples taken before induction (B.I),
5.5 h and 24 h post-induction were resolved by SDS-PAGE. The gel was stained first for
haem-dependent peroxidase activity to detect the accumulation of the 45 kDa mature CCP
(A); and with Coomassie blue to reveal all cellular protein, including the 47 kDa pre-apo-
CCP (B). The accumulated pre-apo and mature forms of CCP are indicated.

poor yields of the mature gonococcal protein, and had decreased further by 24 h post-induction presumably due to the overgrowth by plasmid-free cells. These data indicate that the optimal [IPTG], which results in the accumulation of the highest levels of mature CCP despite small levels of stress on bacteria was 0.04 mM. This concentration was therefore used as the inducer concentration for the improved protocol.

The SDS-PAGE gels were stained with Coomassie blue to reveal the total cellular proteins, and in particular, any accumulation of the 47 kDa pre-apo-CCP protein (**Figure 5.8 B**). The higher molecular weight of pre-apo-CCP relative to the mature form is due to the uncleaved leader peptide sequence of the protein before translocation to periplasm (Turner *et al*., 2003). A very high yield of pre-apo-CCP was detected 24 h post-induction when bacteria were induced with 0.05 mM IPTG. This suggested that the relatively high [IPTG] had induced the synthesis of pre-apo-CCP at a rate that was too rapid for the pre-apo-protein to be secreted and post-translationally modified. In contrast, there was no accumulation of pre-apo-CCP when lower [IPTG] were used for induction, suggesting that all of the apo-CCP that had been produced was converted into the mature form. However, production of very high yields of pre-apo-CCP was evident when bacteria were induced using the standard protocol (**Figure 5.9 B**). Despite the production of very high yields of pre-apo-CCP, very little was converted into the mature form. Thus the contrasting effects of the improved and standard protocols can ultimately be illustrated by the reversal of the ratio between the accumulation of the pre-apo- and mature forms of CCP.

Production of other proteins such as the periplasmic cytochrome c_2 from *N. gonorrhoeae* and the human kinase Protein S-GFP (intellectual property; GSK) were also investigated

using both the standard and improved protocols. Cytochrome c_2, like CCP, is translocated across the cytoplasmic membrane into the periplasmic space and is further modified by covalent attachment of a haem group, whereas Protein S-GFP is a kinase protein from human origin that requires insertion into the cytoplasmic membrane. Both proteins were inherently difficult to produce using the standard protocol, but yields of the proteins significantly increased when using the improved protocol (Sevastsyanovich *et al.*, 2009).

In conclusion, the improved protocol is a robust method for the production of higher yields of recombinant proteins in their native forms despite requirements of post-translational modifications, which is achieved indirectly by minimising physiological stress on the host. In addition, the accumulation of either pre-apo- or mature CCP could be obtained by using either the standard or the improved protocols, respectively.

The production of mature CCP in batch fermentations

The accumulation of mature CCP using the improved protocol in a pH-controlled batch fermentation was compared to yields in shake-flasks. Freshly transformed BL21* pST203 pST2 was grown and induced using the improved protocol in a 1.5 L fermenter in which the pH was controlled at 6.3, and in shake flasks in which there was no pH control. An IPTG concentration of 0.04 mM was used for induction of *ccp* expression in both cultivation conditions. Samples were taken before induction and at intervals post-induction to assess bacterial growth, numbers of colony forming units and accumulation of mature CCP.

Optical densities of bacteria in shake flasks and fermenters continued to increase after induction, reaching approximately 8 and 18 units, respectively, 24 h post-induction and were maintained until 44 h post-induction (**Figure 5.10**). The dissolved oxygen of the culture in the fermenter decreased at a constant rate soon after induction to approximately 0% of the initial concentration by 18 h of induction (**Figure 5.11**). This suggests that there was continuous bacterial growth and recombinant protein production upon induction. It is also apparent from the pO2 data that the consumption of oxygen by bacteria during mature CCP overproduction is greater than that during the production of CheYGFP by BL21* using the improved protocol (since levels of pO2 decreased to only 40% of the initial concentration during production of CheYGFP) (**Figure 3.16., 5.11**). This indicates that the production of mature CCP places a higher metabolic burden on bacteria. The concentration of dissolved oxygen remained at around 0% until 22 h post-induction, thus creating an anaerobic environment. The oxygen limitation in the culture could have both limited further bacterial growth and increased acetate production, which is detrimental to RPP. The dissolved oxygen rapidly increased to approximately 100% by 24 h post-induction and remained at this concentration until 45 h post-induction, coinciding with the cessation of further bacterial growth.

There was a gradual increase in the number of cfu after induction using both cultivation conditions, but the number of cfu 24 h post-induction when bacteria were grown in fermenters was at least 5-fold higher than when bacteria were grown in shake-flasks (**Figure 5.12**). Colonies were replica plated onto selective medium and the retention of both plasmids was calculated. There was 100% retention of the pST2 plasmid by bacteria after induction regardless of the cultivation conditions used. However, there was
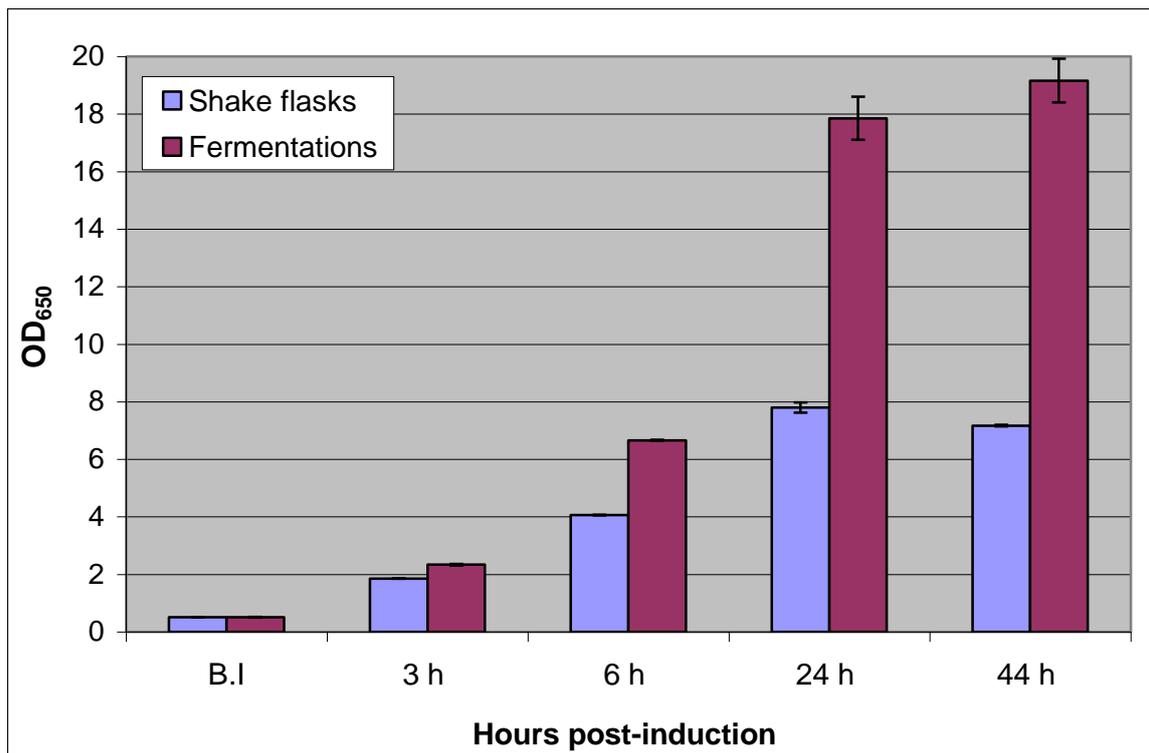
Figure 5.10: The effect of cultivation conditions on growth of wild type BL21* during mature CCP accumulation. The production of mature CCP was induced in freshly transformed BL21* pST203 pST2 using the improved protocol with 0.04 mM IPTG in either shake-flasks or batch fermentations. Optical densities from both cultivation conditions were measured before induction (B.I), 3 h, 6 h, 24 h and 44 h post-induction. Average results from two independent fermentation experiments are presented. Standard error from duplicate repeats of both shake flask and fermentation experiments, using different candidates of the strain, is presented.
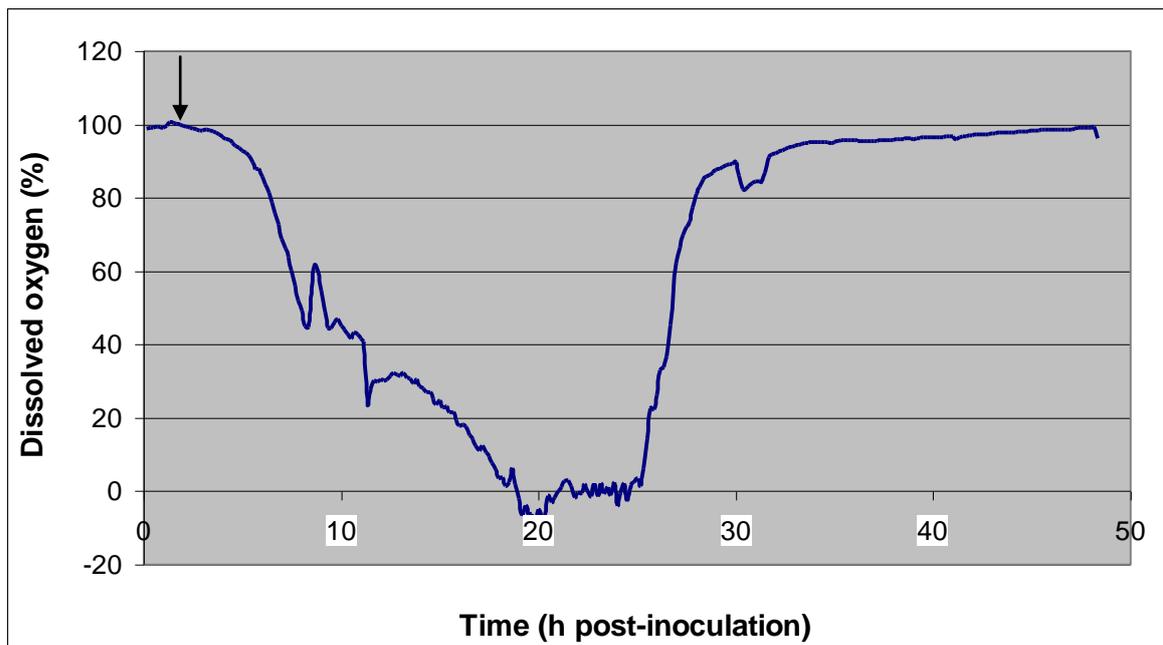
Figure 5.11: The dissolved oxygen concentration during mature CCP overproduction by BL21* using the improved protocol in a fermenter. BL21*pST203 pST2 was grown and induced using the improved protocol in a fermenter. An IPTG concentration of 0.04 mM was used to induce CCP production. The dissolved oxygen concentration of the culture was monitored from the point of inoculation using an oxygen probe. The point of induction is indicated by an arrow.
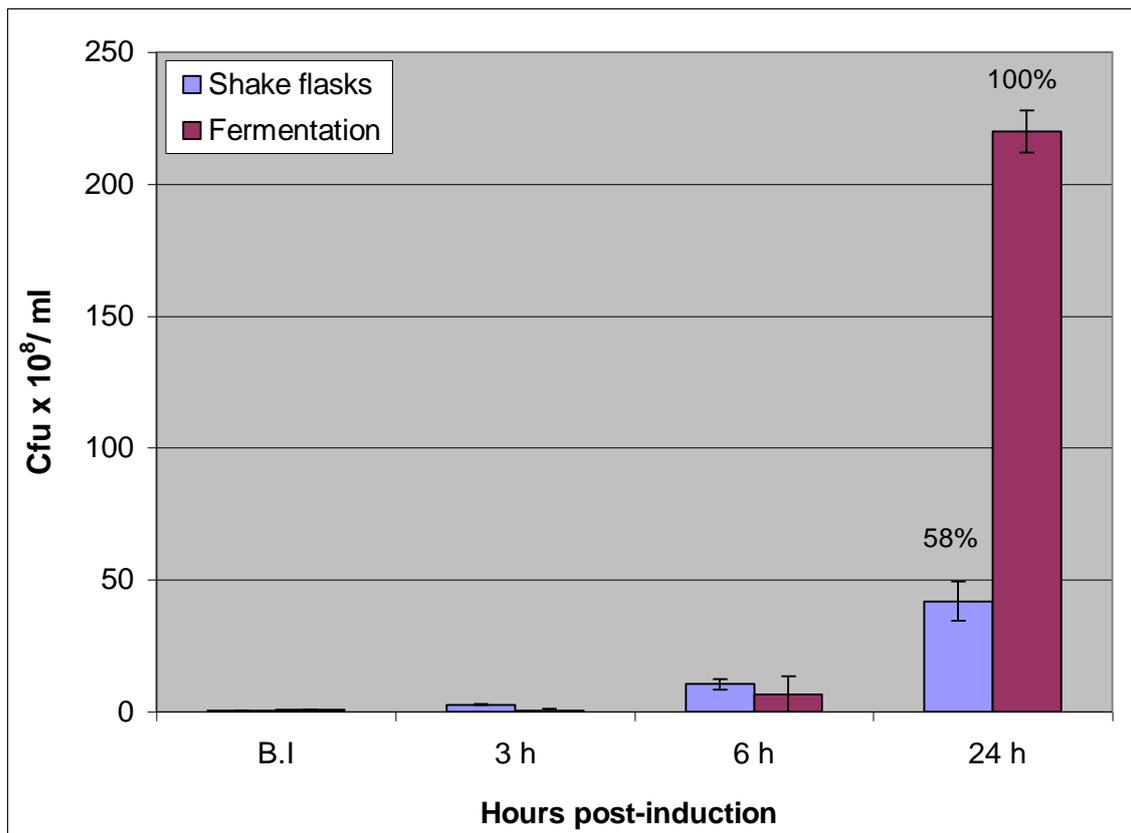
Figure 5.12: The effect of cultivation conditions on colony forming units of wild type BL21* during mature CCP production. The accumulation of mature CCP was induced in BL21* pST203 pST2 using the improved protocol in either shake-flasks or pH-controlled batch fermentations. Serial dilutions of culture samples taken immediately before induction (B.I), 3 h, 6 h and 24 h post-induction were plated on to non-selective NA and the cfu was calculated. Colonies obtained 24 h post-induction were replica plated on to carbenicillin-supplemented NA and the % retention of pST203 was calculated (shown above the corresponding cfu bars). Standard error from duplicate repeats of both shake flask and fermentation experiments, using different candidates of the strain, is presented.

approximately 60% and 100% retention of pST203 when bacteria were grown in shake-flasks and fermenters, respectively (**Figure 5.12**). This suggests that levels of physiological stress were much lower when bacteria were grown in a pH-controlled fermenter than in shake-flasks. There was a gradual increase in the accumulation of mature CCP after induction when bacteria were grown in either shake flasks or fermenters (**Figure 5.13**). However, yields of mature CCP 24 h and 44 h post-induction were higher by approximately 20% when bacteria were grown in fermentation conditions, indicating that factors associated with fermentation conditions such as pH control and aeration can enhance further accumulation of correctly-folded recombinant protein.. However, despite this, cultivation conditions such as aeration and media composition must be optimised in order to maintain higher levels of bacterial growth and RPP, particularly during production of difficult recombinant proteins.

The accumulation of mature CCP by the improved host strain P2 BL21*

The minimisation of physiological stress on the bacterial host by using the improved protocol to decrease levels of T7 RNA polymerase has been shown to indirectly increase yields of mature CCP. This prompted the investigation of whether the accumulation of mature CCP could also be enhanced when using the P2 BL21* host. Yields of mature CCP accumulated by the wild type and P2 BL21* host when using the improved and standard protocols, respectively, were therefore compared.

Fresh transformants of wild type and P2 BL21* pST203 pST2 were grown in shake flasks and induced using the improved and standard protocols, respectively. An IPTG
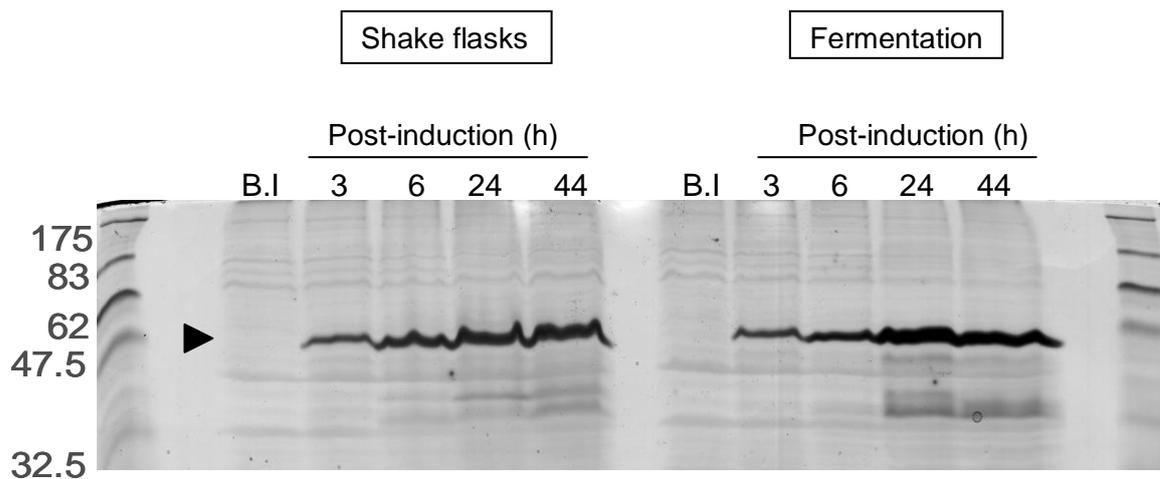
Figure 5.13: The effect of cultivation condition on yields of mature CCP accumulated by wild type BL21*. Production of mature CCP was induced in freshly transformed BL21* pST203 pST2 using the improved protocol in either shake flasks or pH-controlled batch fermentations. Total proteins from whole cell samples taken before induction (B.I) and at intervals after induction were resolved by SDS-PAGE. The gel was stained for haem-dependent peroxidase activity to detect the accumulation of the 45 kDa mature CCP (shown by arrow).

concentration of 0.03 mM was used for induction in the improved protocol to avoid possibilities of variations of bacterial physiological responses to stress.   Growth and colony forming units were monitored up to 48 h post-induction, and the accumulation of mature CCP was assessed.

Growth of both strains increased steadily up to 9 h post-induction, reflecting the low levels of stress introduced using either the improved host or the improved protocol (**Figure 5.14 A**). The growth of P2 BL21* was slightly higher than the parental strain 24 h and 49 h post-induction. There was a continuous increase in the number of colony forming units of P2 BL21* up to 24 h post-induction, reaching approximately 1.08×10^{10}/ ml, after which it had decreased to only 7×10^{9}/ ml by 31 h post-induction (**Figure 5.14 B**). Colony forming units of the parental strain, however, peaked at around only 3×10^{9}/ ml by 9 h post-induction, and remained so until 48 h post-induction. The higher number of cfu of the P2 BL21* strain compared to the wild type suggests that the levels of stress in the improved host were lower than in the improved protocol. Colonies obtained 24 h and 49 h post-induction were replica plated on to carbenicillin-supplemented NA and the % retention of pST203 was calculated. Approximately 61% of cells of the parental strain had retained the pST203 plasmid 24 h post-induction, which increased to approximately 83% 49 h post-induction (**Figure 5.14 B**). This suggested that bacteria had recovered from a slight stress due to a sub-optimal concentration of IPTG being used for induction, but nevertheless a majority of the population had retained the plasmid. In contrast, almost 100% of P2 BL21* cells had retained pST203 24 h and 48 h post-induction. These data, in addition to the numbers of cfu, indicate that P2 BL21* is better adapted to stress-resistance than when using a low-stress-inducing protocol alone with the parental strain.
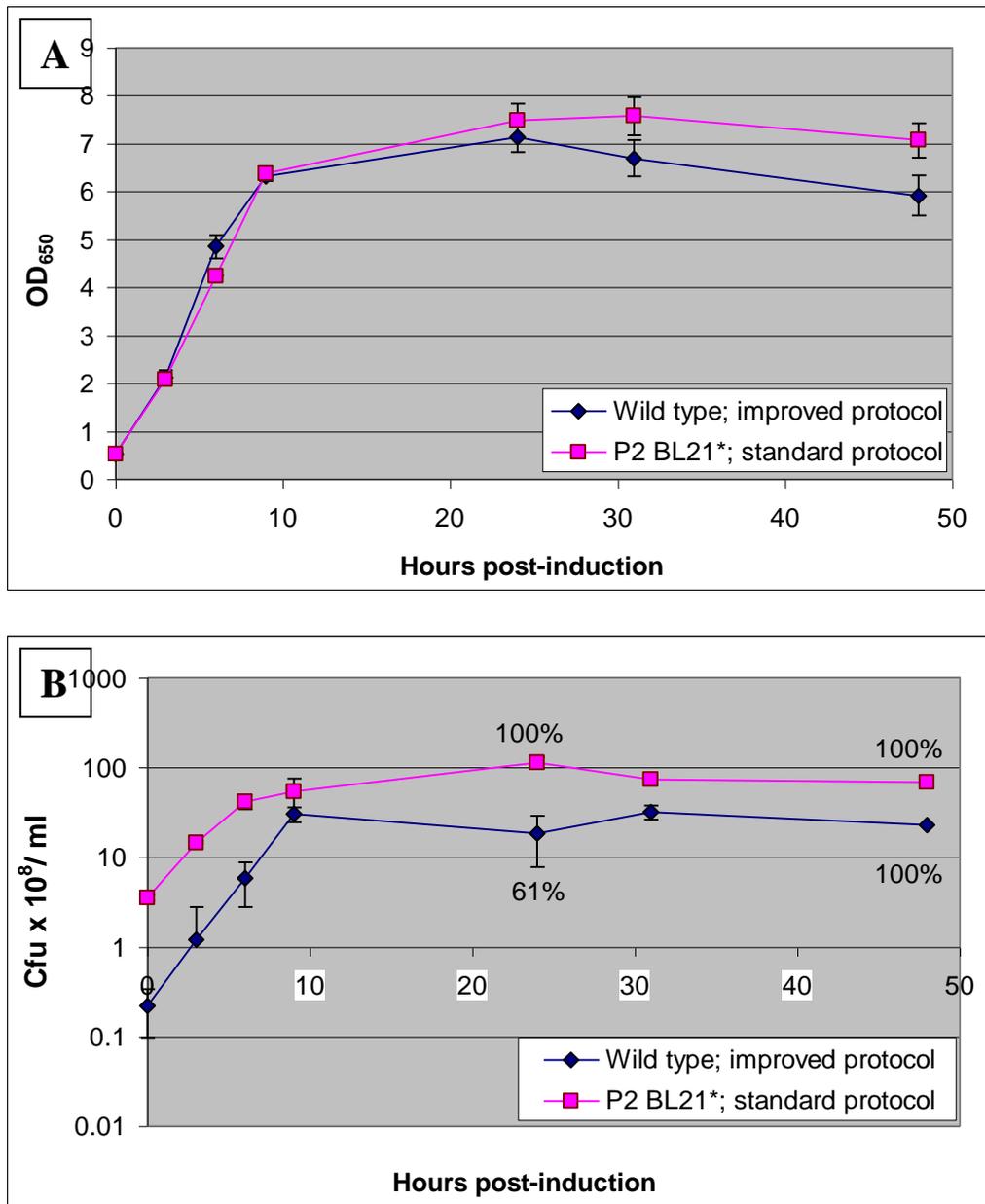
Figure 5.14: The effect of mature CCP production on growth and cfu of wild type and P2 BL21*. Production of mature CCP was induced in freshly transformed wild type and P2 BL21* pST203 pST2 using the improved and standard protocols, respectively. Samples were taken before induction and at intervals post induction for OD measurement (A) and cfu analysis (B). Colonies obtained 24 h and 49 h post-induction were replica plated on to carbenicillin-supplemented NA and the % retention of pST203 was calculated (shown adjacent to cfu data). Standard error from triplicate repeats using different candidates of each strain is presented.

Proteins from whole-cell samples taken before induction and at intervals post-induction were resolved by SDS-PAGE. The gel was first stained for haem-dependent peroxidise activity to reveal yields of mature CCP that had accumulated (**Figure 5.15 A**). Production of mature CCP was evident after induction in the parental strain, with increasing yields post-induction. In contrast, poor yields of mature CCP were detected in P2 BL21* with little further accumulation beyond 6 h post-induction. Final yields of mature CCP 24 h post-induction in P2 BL21* were comparable to those obtained only 6 h post-induction when using the improved protocol for RPP using BL21*, despite the plasmid loss in a large proportion of bacteria in BL21*. This indicated that P2 BL21* host was not suitable for the accumulation of relatively high yields of mature CCP, despite the decreased levels of T7 RNA polymerase. Hence, compared with the improved protocol, which relies mainly on the decreased levels of T7 RNA polymerase, it seems likely that P2 BL21* contains additional chromosomal mutations that decrease or compromise the accumulation of mature CCP. The gel was also stained with Coomassie blue to assess any accumulation of pre-apo-CCP (**Figure 5.15 B**). There was no accumulation of pre-apo-CCP by either P2 BL21* or the parental strain, indicating that the poor yields of mature CCP accumulated by P2 were not due to physiological stress that is associated with the accumulation of large quantities of pre-apo-CCP.

These data suggest that the improved protocol is an inherently more generic method for enhancing RPP than using a randomly chosen improved host. Following this result, more candidates of P2 BL21* pST203 pST2 were screened for improved accumulation of mature CCP without positive results. It was concluded therefore that P2 BL21* was
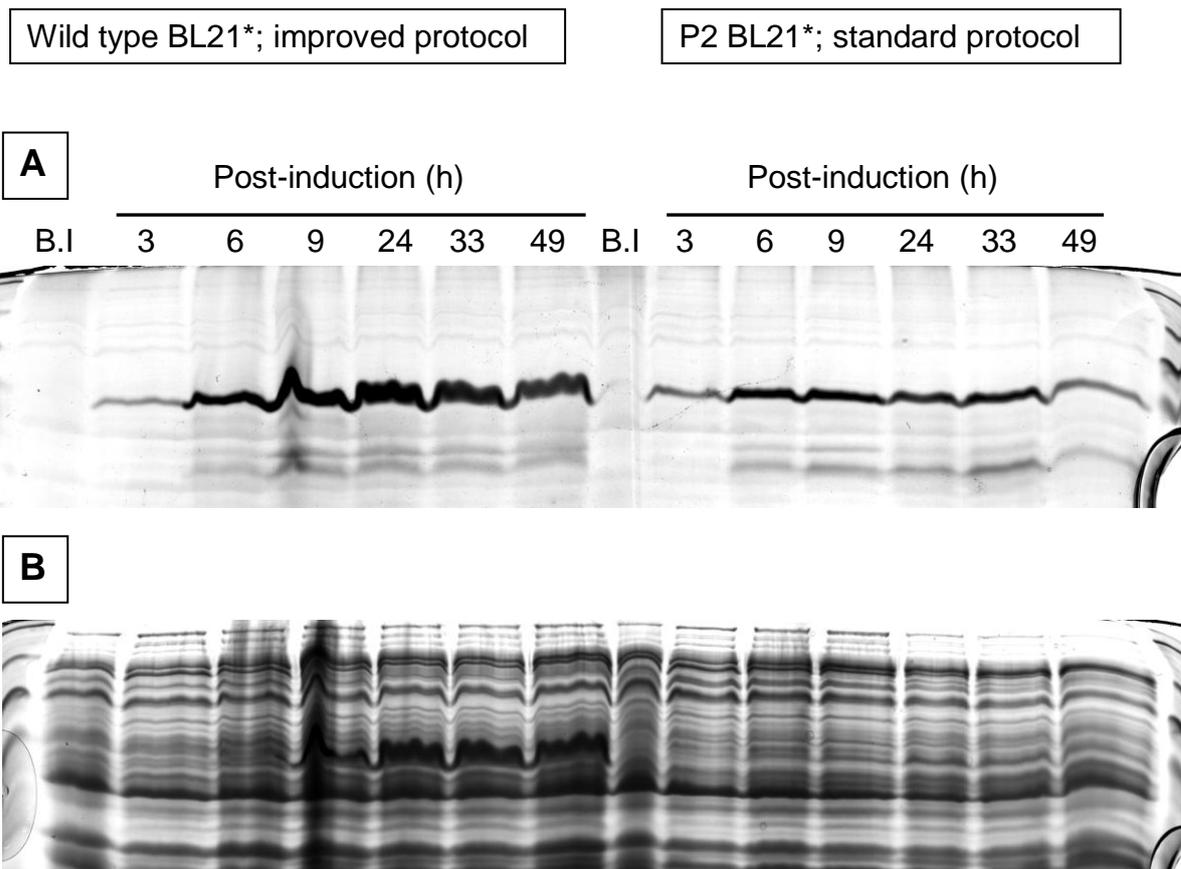
Figure 5.15: The accumulation of mature CCP using either the improved protocol or the improved host P2 BL21*. Accumulation of mature CCP was induced in fresh transformants of wild type and P2 BL21* pST203 pST2 using the improved or standard protocols, respectively. Total proteins from whole cell samples taken before induction (B.I) and at intervals post-induction were resolved by SDS-PAGE. A) The gel was first stained for haem-dependent peroxidase activity to reveal yields of the 45 kDa mature CCP accumulated. B) The gel was then stained with Coomassie blue to reveal background proteins particularly any accumulation of the 47 kDa pre-apo-CCP.

generically an unsuitable host for the accumulation of high yields of mature gonococcal CCP.

The isolation of an improved BL21* host for mature CCP overproduction

The poor accumulation of mature CCP by P2 BL21* has led to the suggestion that it might not be possible to isolate a 'universal' improved host that can increase the yields of all recombinant proteins, and perhaps it is necessary to isolate different mutant hosts, individually, based on the protein to be produced. This prompted attempts to isolate an improved host to enhance the accumulation of mature CCP.

CCP production was induced in wild type BL21* pST203 pST2 using the standard protocol, and the rare stress-resistant mutants within the culture 24 h post-induction were positively selected by plating on either NA supplemented with carbenicillin alone, or with the addition of 0.5 mM IPTG. Both methods have been previously used in this work and by Miroux and Walker (1996) to successfully isolate the improved hosts P2 BL21* or C41 and C43, respectively. Multiple candidate colonies from each isolation procedure were purified and individually screened for improved accumulation of mature CCP using the standard protocol. They were also compared to a negative control of the parental strain induced in the same way.

Proteins from whole-cell samples taken 21 h post-induction were resolved by SDS-PAGE and stained for haem-dependent peroxidase activity to reveal yields of mature CCP accumulated (**Figure 5.16 A**). Only one candidate showed an enhanced accumulation of
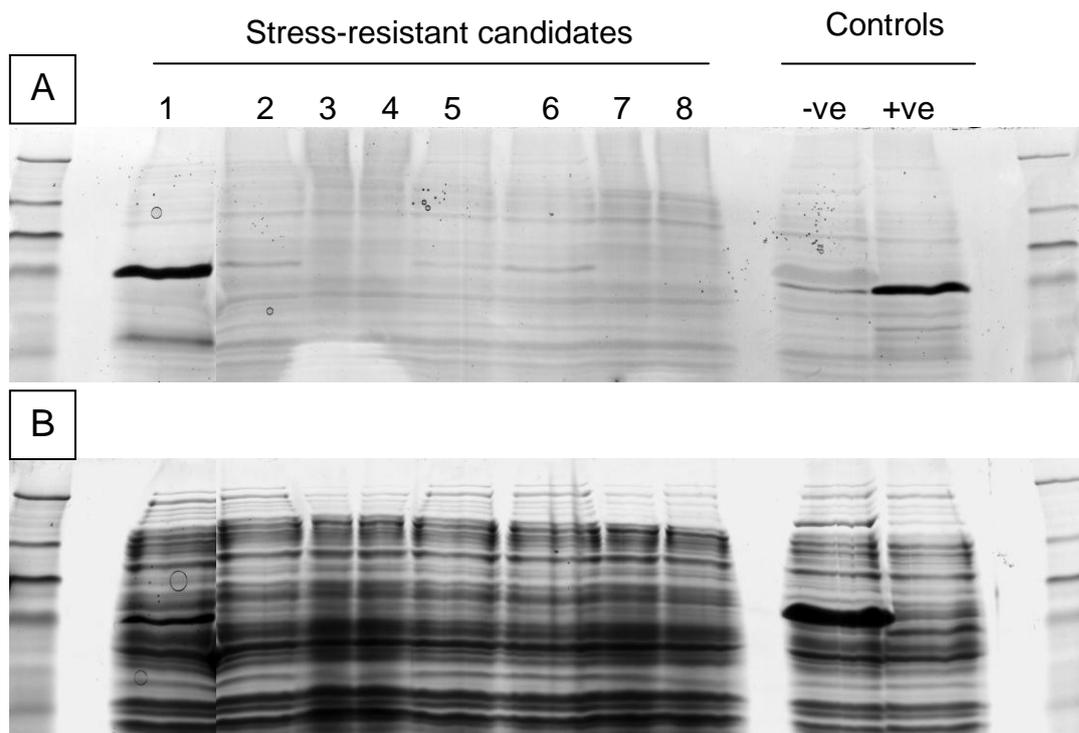
Figure 5.16: The accumulation of mature CCP in stress-resistant BL21* candidates.
Production of CCP was induced in wild type BL21* pST203 pST2 using the standard
protocol. Culture samples taken 24 h post-induction were plated on to NA supplemented
with either carbenicillin alone or in addition to 0.5 mM IPTG to positively select bacteria
that have retained the over-expression plasmid and were therefore 'stress-resistant'. Eight
stress-resistant candidates obtained from NA with carbenecillin selection alone were
screened for improved accumulation of mature CCP using the standard protocol in order to
identify a potential improved host. These were compared with positive and negative
control cultures of the parental strain that had been induced with either the improved
(marked +ve) or the standard (-ve) protocol. Proteins from whole cell samples taken 24 h
post-induction were resolved by SDS-PAGE and stained for haem-dependent peroxidase
activity to reveal yields of mature CCP accumulated (A). Proteins were then stained with
Coomassie blue to reveal any possible accumulation of pre-apo-CCP (B).

mature CCP relative to the positive control of the wild type strain that had been induced using the improved protocol. No accumulation of pre-apo-CCP was detected in any of the candidate strains, whereas it was evident in the parental strain that had been induced using the standard protocol (**Figure 5.16 B**). Subsequent replica plating revealed almost 100% retention of the pST203 plasmid in all of the candidate strains, suggesting that all were resistant to the induction stress of the standard protocol, but most without improvements in CCP accumulation.

The candidate strain that showed improved accumulation of mature CCP was cured of both plasmids and re-transformed with the original constructs. The production of mature CCP was reproduced (data not shown). This established that the improved ability to accumulate CCP was a result of a chromosomal mutation rather than to mutations on the expression plasmid. This strain was named 'Mutant Pro' BL21*.

The production of mature CCP using different improved hosts

The difference in yields of mature CCP accumulated by P2 BL21*, and provisionally by Mutant Pro, led to the hypothesis that different hosts need to be independently isolated such that they are 'tailored' for the production of different groups of recombinant proteins. Given this assumption, it was also of interest to investigate the production of mature CCP by the improved hosts C41 and C43 that had been isolated during recombinant membrane protein production. The accumulation of mature CCP in Mutant pro BL21*, P2 BL21*, C41 and C43 was therefore investigated using the standard protocol and compared to that

generated by the wild type strain using the improved protocol and induction with 0.03 mM IPTG..

Bacterial growth was very similar in all cultures 3 h, 6 h and 9 h post-induction (not shown). However, the optical density 24 h post-induction of P2 BL21* and the parental strain that had been induced using the improved protocol, were approximately 7 OD units, whereas optical densities of C41, C43 and Mutant Pro were between 9 to 10 OD units, indicating that these improved hosts grow better than strains BL21* or P2 after induction of membrane recombinant proteins (**Figure 5.17**). There was little change in the optical densities of any of the cultures in the subsequent 24 h.

The number of cfu 24 h post-induction was lowest when wild type BL21* was induced using the improved protocol at around only 0.25×10^{10}/ ml, reflecting the physiological burden on the normal host due to the accumulation of recombinant membrane protein (**Figure 5.18**). However, numbers of cfu of P2 BL21*, C41 and Mutant Pro were similar at around 1.05×10^{10}/ ml, whereas the cfu of C43 was highest at 1.7×10^{10}/ ml. This reflects the 'stress-resistance' of the improved hosts compared to the wild type during the overproduction of mature CCP. The number of cfu of all improved hosts, except C41, had decreased 49 h post-induction, suggesting that C41 might be the best host for the accumulation of mature CCP.
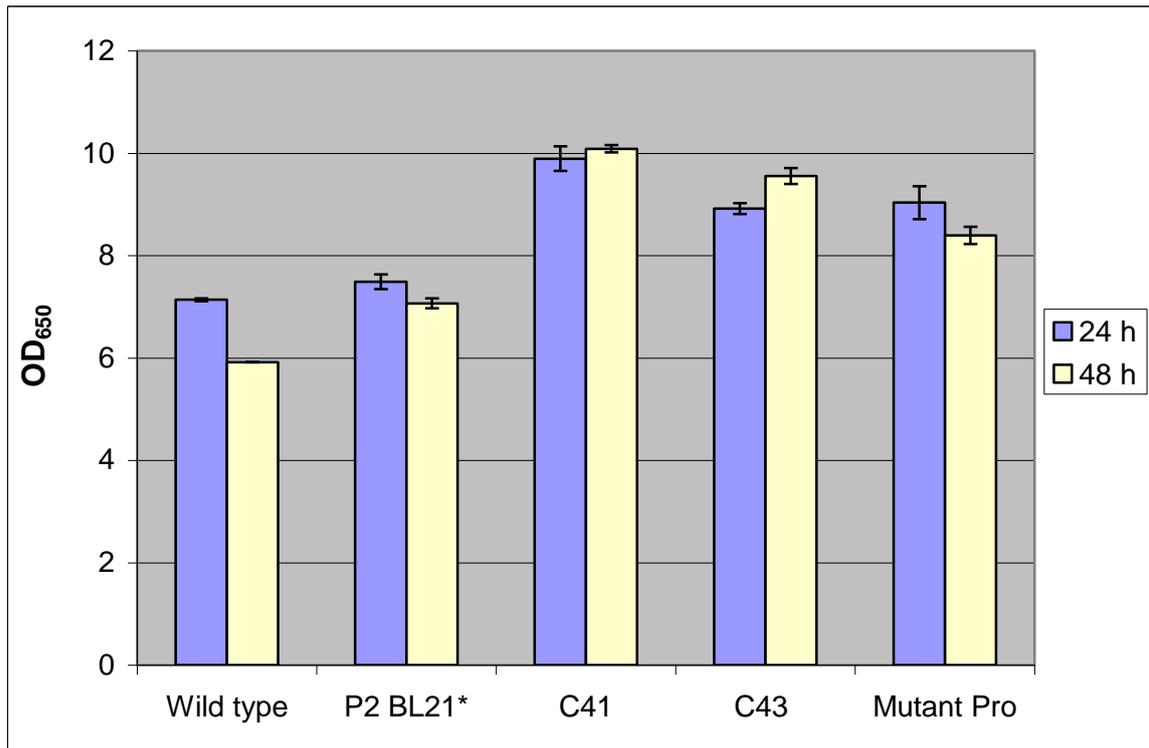
Figure 5.17:   The growth of different improved hosts during CCP production. Production of mature CCP was induced in improved host strains P2 BL21*, C41, C43 and Mutant Pro using the standard protocol, and compared to the parental strain that had been induced using the improved protocol. Optical densities of culture samples taken 24 h and 48 h post-induction were measured. Standard error from triplicate repeats using different candidates of each strain is presented.
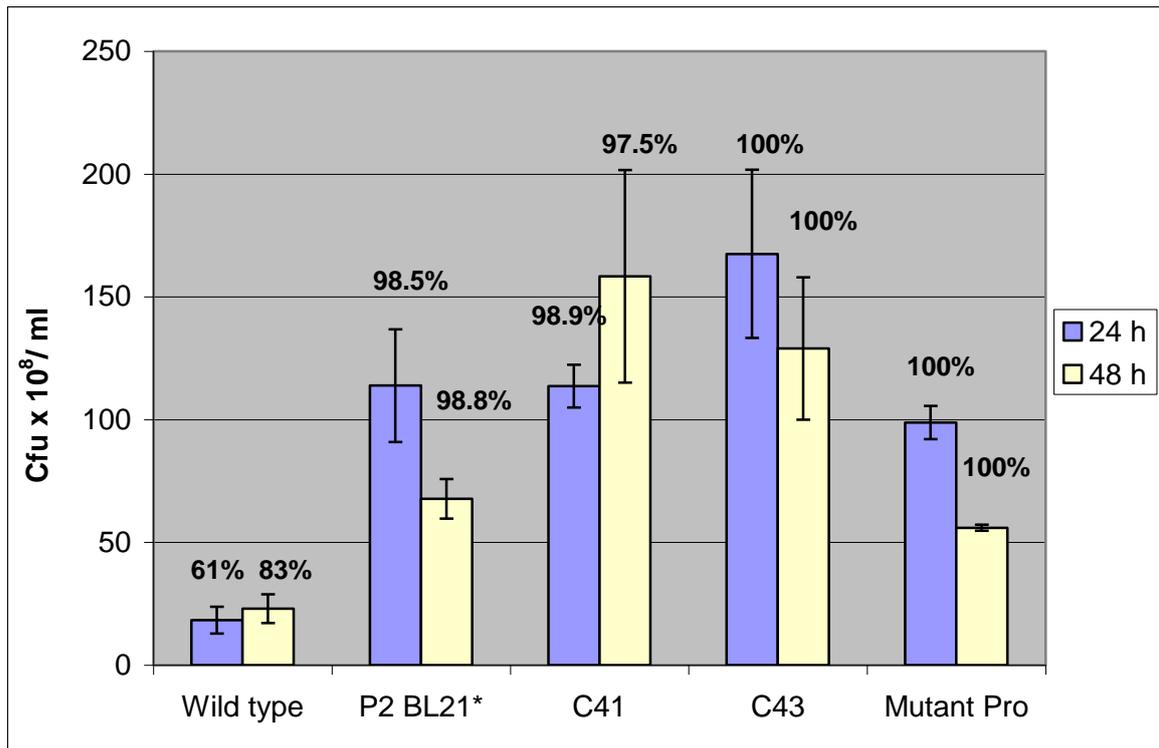
Figure 5.18: The effect of CCP production on the cfu of different improved hosts. Production of mature CCP was induced in improved host strains P2 BL21*, C41, C43 and Mutant Pro using the standard protocol, and compared to the parental strain that had been induced using the improved protocol. Serial dilutions from culture samples taken 24 h and 48 h post-induction were plated on to non-selective NA and the cfu was calculated. Colonies obtained were replica plated onto carbenicillin-supplemented NA and the % retention of pST203 was calculated (presented above corresponding cfu data). Standard error from triplicate repeats using different candidates of each strain is presented.

Plasmid retention of pST203 in the normal host using the improved protocol 24 h post-induction was approximately 61%, which had increased to 83% by 49 h post-induction (**Figure 5.18**). Plasmid retention in all improved hosts was approximately 100%, indicating that there was virtually no physiological stress associated with the accumulation of the membrane recombinant protein using the standard protocol.

The accumulation of mature CCP in all hosts was assessed before induction and at intervals post-induction (**Figure 5.19**). The highest yields of mature CCP were produced in strains C41 and C43, with continuous accumulation up to 49 h post-induction. In contrast, the lowest yields of mature CCP were produced by P2 BL21*, providing evidence to support the hypothesis that the enhanced production of a recombinant protein in a host could be related to how the host was originally isolated. Similar yields of mature CCP were accumulated by Mutant Pro and by the wild type host when using the improved protocol alone, with maximum yields obtained 24 h post-induction. This suggests that Mutant Pro contains mutations that decrease the stress impact of RPP, with similar effects to the improved protocol that ultimately enhance yields of the correctly folded recombinant protein. It seems unlikely, however, that Mutant Pro contains additional mutations that significantly and specifically improve the accumulation of recombinant membrane proteins, as compared to C41 or C43.

It is important to highlight that although very little CheYGFP was produced by strain C43, this host accumulated high yields of mature CCP. Conversely, P2 accumulated high yields of soluble CheYGFP, but poor yields of mature CCP. The common mutations in both these hosts are those that result in decreased levels of T7 RNA polymerase, but it seems

likely that there are additional different mutations between these two hosts that contribute to the discrepancy in their production of different recombinant proteins. This led to the proposal of a second hypothesis that a mutation that improves the production of one type of recombinant protein might compromise the overproduction of other different types of recombinant protein. However, strain C41 had produced good yields of both correctly folded cytoplasmic CheYGFP and the membrane lipoprotein, CCP. The production of many different recombinant proteins in different improved hosts needs to be thoroughly investigated.

Figure 5.19: Yields of mature CCP accumulated using either the improved protocol or different improved hosts. Production of mature CCP was induced in improved hosts P2 BL21*, C42, C43 and Mutant Pro using the standard protocol and compared to the parental strain that had been induced using the improved protocol. Total proteins from whole cell samples taken before induction (B.I) and at intervals post-induction were resolved by SDS-PAGE and stained for haem-dependant peroxidase activity to reveal yields of mature CCP accumulated. To account for the slight variation in the haem-staining procedure between gels, samples taken from all cultures 24 h post-induction were stained in one gel. The accumulation of mature CCP in the parental strain using the standard protocol was also included as a negative control (marked as –ve).

Wild type BL21*; improved protocol

P2 BL21*

C41

C43

Mutant Pro
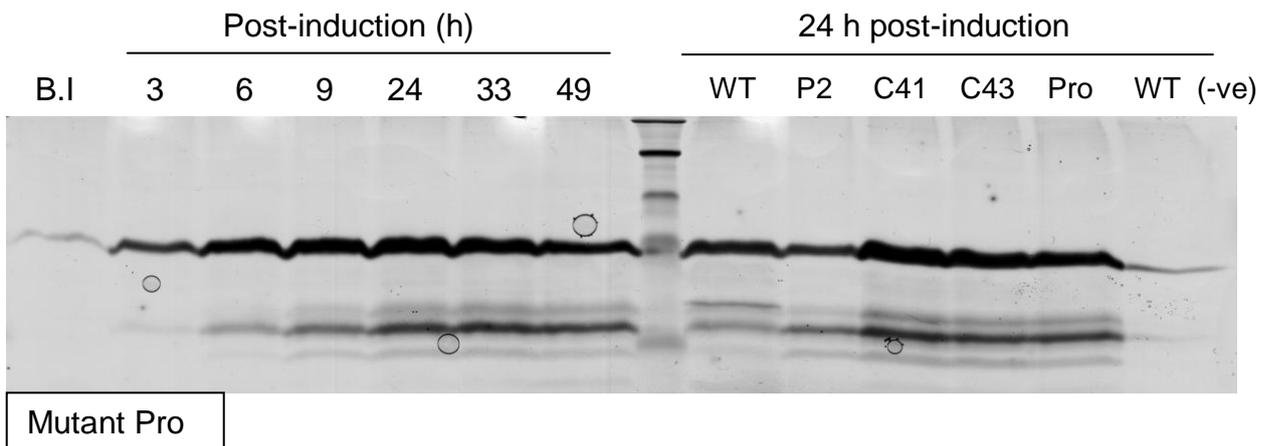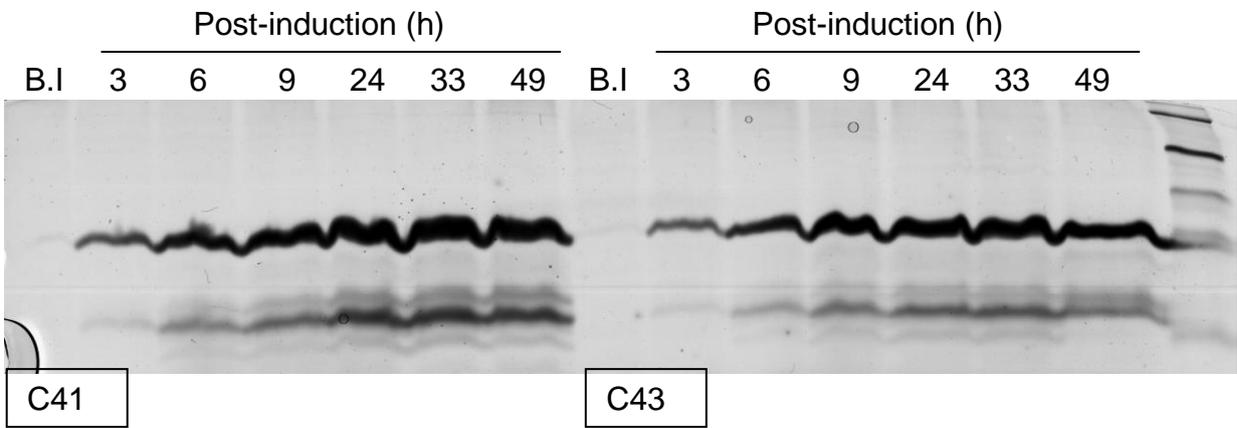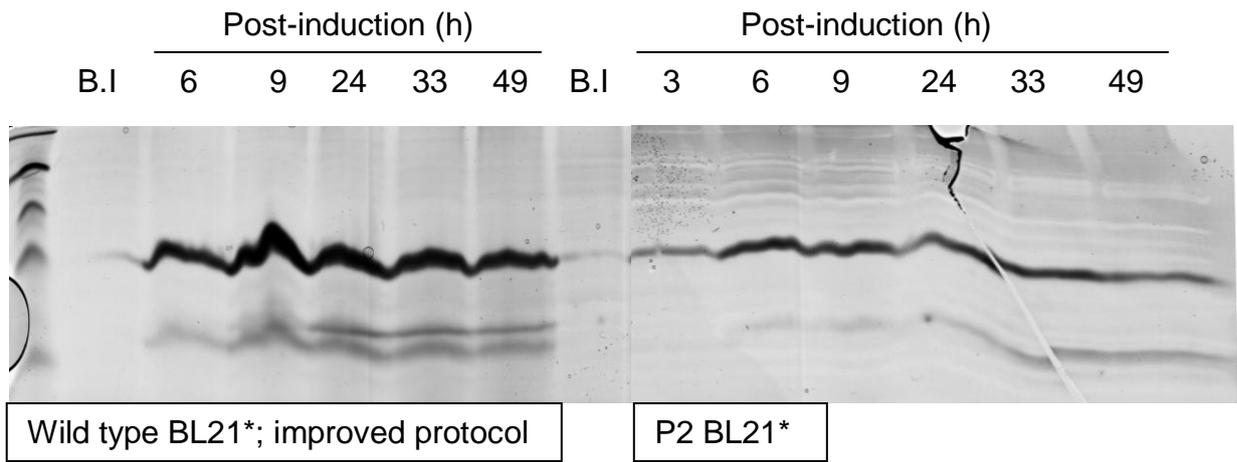
CHAPTER 6

Discussion

Improving RPP by decreasing the rate of protein synthesis

The primary cause of failure to accumulate correctly folded protein in high yields has been attributed to the rapid accumulation of misfolded target protein after induction (Hoffmann and Rinas, 2004; Gasser *et al*., 2008). This is mainly due to rates of protein synthesis overwhelming post-translational modifications such as protein folding and secretion. Ultimately, this leads to a cascade of physiological responses in bacteria, the most important being growth arrest and the cessation of further protein accumulation. A growing consensus that has emerged from multiple transcriptomics studies seeking to understand underlying physiological responses to RPP is that stress responses such as the stringent and heat shock responses are induced in bacteria following induction by standard RPP protocols (Gill *et al*., 2000; Haddadin and Harcum, 2005; Dürrschmidt et al., 2008; reviewed in Sevastyanovich *et al*., 2009). Despite this, agreement has not yet extended to the development of strategies that generically improve RPP. Instead, individual trial and error attempts that exploit certain aspects of heat shock response such as decreased temperatures alone or the co-expression of chaperones continue to be explored (reviewed in Sevastsyanovich *et al*., 2010; Gasser *et al*, 2008).

In this project, two generic strategies that greatly improve the accumulation of correctly folded protein have been developed. The 'improved protocol' is a generic protocol that seeks to minimise the physiological stress of bacteria even before induction by combining decreased growth temperatures and decreased rates of RPP synthesis. Lower temperatures have the advantage of decreasing the specific growth rate of bacteria, which indirectly facilitates correct protein folding (Hoffman and Rinas, 2004; Gadgil *et al*., 2005).

Interestingly, the final biomass of bacteria during RPP using the improved protocol was almost twice that when the standard protocol was used. This could be due to a more efficient use of resources during slow growth.

The development of the improved protocol

The development of the improved protocol was based on the observation that bacterial colony forming units are decreased by several orders of magnitude following high-levels of RPP induction using standard protocols (Andersson, *et al*., 1996; Kurland and Dong, 1996; Sundström *et al*., 2004). Hence, by titrating lower concentrations of IPTG to decrease protein synthesis levels, the loss of colony forming units is minimised. The unique feature of this approach is that it was possible to identify the induction level at which RPP-induced stress becomes apparent by monitoring the loss of colony formation. The improved protocol relies on finding the optimum IPTG concentration that decreases physiological stress and increases specific productivity. Thus, the best possible compromise that would allow the highest yields of correctly folded target protein could be reached. Specific productivity was monitored either by increased GFP fluorescence or the peroxidase activity during the production of either CheYGFP or mature CCP, respectively. The advantage of this empirical approach is that it identifies the best conditions for a sustainable accumulation of different types of recombinant proteins, which includes difficult proteins that require extensive post-translational modifications. Decreasing the physiological stress of bacteria has been successfully used to optimise production of correctly folded CheYGFP using other expression systems (Sevastsyanovich; personal communication).

Another important facet of the improved protocol is the growth of bacteria at continuous low temperatures. Decreased temperatures have been shown to increase correctly folded protein by decreasing general metabolism and growth rates (Gadgil *et al*., 2005). This is reflected by the rapid decrease in expression levels of many enzymes involved in glycolysis. The indirect effect of this is a decreased rate of protein synthesis. This provides better opportunities for protein folding by the available chaperones and for subsequent modifications of proteins (Hoffman and Rinas, 2004; Baneyx and Mujacic, 2004). Apart from indirect effects of decreased growth, expression of genes encoding enzymes involved in acetate formation are decreased at 28°C relative to 37°C (Gadgil *et al*., 2005). This results in the decreased production of acetate, a feature that is desirable for RPP. Amongst other effects of decreased temperature are an increased expression of translation initiation factor 1 (*infA*) and of the chaperone involved in the maturation of iron-sulphur cluster-containing proteins, *hscA*. Decreased temperatures can also directly increase yields of periplasmic or outer membrane proteins. In cases where a periplasmic recombinant protein is produced, this is also due to the function of the periplasmic protease, DegP, switching from a protease to a folding chaperone (Speiss *et al*., 1999). This was recently shown do be due to binding of misfolded proteins transforms hexameric DegP into large, catalytically active 12-meric and 24-meric multimers (Krojer *et al*., 2008). Further elucidation of the structure of DegP revealed that it contains an inner binding cavity that serves antagonistic functions depending on the nature of the bound protein substrate. The immediate consequence of this can be recognised in the context of RPP.

Downregulation of T7 RNA polymerase is key to improving RPP in pET expression systems

Key to the success of the improved protocol and indeed the improved host P2 BL21*, is the decreased synthesis of T7 RNA polymerase, which results in slower accumulation of recombinant protein. In fact, the induction of high levels of T7 RNA polymerase alone results in the significant loss of colony forming units, suggesting it is the source of stress that is inherently associated with the high level induction of the T7 expression system. Despite similar findings reported by Miroux and Walker (1996), different conclusions were made by these authors -that the over-expression plasmids themselves are intrinsically toxic to *E. coli* BL21(DE3). Evidence supporting this was reported by Soriano *et al*. (1999), who showed that there was no growth arrest post-induction of naive *E. coli* HMS174(DE3) or with *E. coli* JM105 strains carrying a plasmid that contains the target gene under the control of the *tac* promoter. It was also shown that growth arrest of *E. coli* HMS174(DE3), which contained either a pET3a plasmid with or without the target gene, had occurred after induction with 0.1 M IPTG. Certainly these findings reflect the metabolic burden of the pET plasmids on the host cell upon induction (Bently *et al*., 1990; Kurland and Dong, 1996; Soriano *et al*., 1999). The continued growth of bacteria after inducing the expression of the target gene using the *tac* but not the T7 expression system clearly reflects the negative impact of the T7 expression system on host physiology. In an attempt to identify the source of stress associated with induction of the pET system in BL21*, the effect of inducing T7 RNA polymerase with 0.5 mM IPTG on the number of colony forming units of naïve BL21* was compared to that of uninduced naive BL21* and induced P2 BL21* strain (preliminary data; this study). Preliminary data revealed that the

number of colony forming units of wild type BL21* 3.5 h post-induction was 3-fold and 10-fold lower than the uninduced BL21* and the induced P2 BL21* strains. This indicates a cessation of cell division in wild type BL21* upon induction. However, compared from the results presented by Soriano et al. (1999), it appears that the presence of an empty pET plasmid, when T7 RNA polymerase is induced, increases further the consequences of growth arrest. This indicates the presence of binding sites for the T7 RNA polymerase on the pET vector such that toxic effects are exerted upon their binding of the polymerase. Indeed it was reported by Studier and Moffatt (1986) that due to both the high efficiency of transcription by the T7 RNA polymerase and specificity for its own promoter, in addition to the inefficiency of the T7 terminator, the T7 RNA polymerase is able to completely transcribe around the plasmid approximately 3 times, producing very long mRNA species. Thus, it becomes apparent as to why the presence of even an 'empty' pET plasmid, containing a T7 promoter, results in severe growth retardation in bacterial after the induction of the T7 RNA polymerase.

The rationale of the improved protocol is to decrease the synthesis levels of T7 RNA polymerase to a concentration so low that it is not even detected by western blotting. This is to ensure the slow production of recombinant protein and its accumulation mainly in a correctly folded state. Despite the reported toxicity of the T7 expression system, at least two approaches were developed earlier, which are present in the literature and that claimed generic improvements to RPP (Wagner *et al*., 2008; Streidner *et al*., 2010). Both strategies were based on the assumption that production of the target protein, rather than the high-level synthesis of T7 RNA polymerase, is the main cause of physiological stress in bacteria.

Wagner *et al*. (2008) proposed a strategy similar to the 'improved protocol', which involved the decreased synthesis of the target protein by inhibiting the activity, rather than the synthesis of T7 RNA polymerase, in BL21(DE3). This was achieved by titrating T7 lysozyme that was encoded on a plasmid under the control of an inducible L-rhamnose promoter. Hence, adding different concentrations of rhamnose provided a way of 'tuning' RPP rates and increasing yields of correctly folded recombinant membrane protein. In this strategy, however, high level synthesis of T7 RNA polymerase is induced by 0.5 mM IPTG, but its activity is dampened afterwards. Compared to the improved protocol, this strategy is highly inefficient for the accumulation of correctly folded target protein. This is because cellular resources have already been wasted in the production of otherwise 'redundant' T7 RNA polymerase, which places an unnecessary metabolic burden on the bacterial host.

A more recent strategy that aimed to decrease the stress effects of RPP using the T7 expression system, by decreasing the levels of RPP, has been reported. This involved the development of a plasmid-free system in which the target gene is transcribed from a single copy on the host's chromosome (Striedner *et al*., 2010). Production of the target protein is achieved using highly levels of the induced chromosomally-encoded T7 RNA polymerase. Again, compared to the improved protocol, this approach is highly inefficient in that cellular resources are wasted on the production of T7 RNA polymerase, much of which will be redundant due to the very low dosage of the target gene. Unsurprisingly, there was only a modest two-fold increase in the specific production of the target protein. This strategy had resulted in increased biomass but not specific productivity.

The broad-range applicability and utility of the improved protocol has been demonstrated for the production of cytoplasmic proteins CheYGFP and protein D-GFP in addition to more difficult membrane proteins that require extensive post-translational modifications such as CCP from *Neisseria gonorrhoeae* and the human protein kinase S-GFP, compared to the standard protocol (Dr Sevastsyanovich, personal communication). The accumulation of these proteins in a correctly folded state has been increased several-folds when using the improved protocol compared to standard methods. The sustainability of the improved protocol has also been demonstrated in that minimal stress of RPP resulted in the retention of plasmid by bacteria without any overgrowth of plasmid-free bacteria up to at least 48 h post-induction.

Use of improved hosts to overcome problems associated with RPP

In this project, two improved host strains P2 BL21* and Mutant Pro, have been isolated during production of CheYGFP and mature CCP, respectively. Both hosts were 'stress resistant' in that they were able to form colonies even after high induction levels of the T7 expression system. Also, the majority of bacteria still contained the expression plasmid after prolonged induction periods. Key to the improvement in RPP shown by P2 BL21* is the presence of three chromosomal mutations in the *lac*UV5 promoter that down-regulate expression of the T7 RNA polymerase. These mutations are identical to those in the Walker strains C41 and C43. The two mutations in the -10 sequence revert the promoter to a much weaker, wild type *lac* promoter and therefore expression of the T7 polymerase gene is significantly decreased. The third mutation in the operator region increases the binding affinity for the *lac* repressor. The decreased synthesis of T7 RNA polymerase

inevitably decreases toxicity associated with induction, and in turn results in the slow production of correctly folded recombinant protein. Unlike the improved protocol, in which a very narrow IPTG concentration range is required for optimal accumulation of correctly folded target protein, improved hosts such as P2 BL21* are able to sustain prolonged RPP even after induction with concentrations of 1 mM IPTG (data not presented). This is mainly due to the (high) inefficiency of transcription of T7 RNA polymerase from the mutated *lac*UV5 promoter. Unsurprisingly therefore, a combination of both these approaches was unable to induce sufficient levels of T7 RNA polymerase for the overproduction of the recombinant protein CheYGFP.

As revealed by SDS-PAGE, levels of CheYGFP accumulated before induction by the improved hosts P2 BL21*, C41 and C43 using the standard protocol were noticeably lower compared to the parental strain when the improved protocol was used. This is most likely due to the lower basal levels of T7 RNA polymerase within improved hosts, which would inevitably result in lower levels of RPP. It is known that the wild type *lac* promoter is more susceptible to catabolite repression compared to *lac*UV5 (Silverstone *et al*., 1970), which might account for the tighter control of expression in the presence of glucose as the main carbon source. Thus, the lower basal expression of T7 RNAP in improved hosts will inevitably correspond to lower physiological stress compared to the parental strain even before induction. Hence, the production of recombinant protein using bacterial strains that harbour mutations in *lac*UV5 will be more sustainable compared to using the improved protocol alone. Another advantage of using improved hosts with tighter basal expression of T7 RNAP becomes apparent when the target protein is toxic. This intrinsic advantage in improved hosts abolishes the need for co-transforming with pLysS plasmids that inhibit

T7 RNA polymerase activity, which would otherwise add a further metabolic burden on the host.

In this project, and in collaboration with Dr Y. Sevastsyanovich, more BL21* improved host strains were isolated during production of CheYGFP, and all have mutations that downregulate the accumulation of T7 RNA polymerase. This raises the important question of whether slow synthesis of recombinant protein is the main requirement for the prolonged accumulation of correctly folded target protein. In order to investigate this, an improved host that produces successful RPP outcomes, without dependence on the down regulation of T7 RNA polymerase or the inefficient transcription of the target gene, should be isolated. Alternatively, the production of recombinant protein using different promoters could be investigated.

Balancing *de novo* protein synthesis and maturation is necessary for efficient accumulation of correctly folded recombinant protein

The rapid production of recombinant protein results in the accumulation of high yields of insoluble protein. This protein often aggregates into inclusion bodies in either the cytoplasm or the periplasm, depending on the type of target protein being produced (Carrió and Villaverde, 1999; 2003, Aríe *et al.*, 2006). Certainly, the majority of the cytoplasmic proteins CheYGFP and protein D-GFP always accumulated in insoluble inclusion bodies after production by the standard protocol. The aggregation of protein was detected as early as 1 h post-induction, supporting reports that relate this to the rapid rates of protein synthesis, which results in the accumulation of unfolded newly synthesized

polypeptides (Baneyx and Mujacic, 2004; Hoffman and Rinas, 2004). Production of 'difficult' proteins that require further modification or secretion in soluble fractions is even more challenging. This has been demonstrated during production of cytochrome proteins CCP and c_2 from the gonococcus (Sevastsyanovich et al., 2009). Large quantities of apo-cytochrome still containing the signal sequence were always detected in the cytoplasm soon after induction with the standard protocol. Indeed, decreasing protein synthesis rate alone using the improved protocol resulted in increased yields of mature CCP produced, indirectly, by decreasing physiological stress. However, to increase the efficiency and the potential of obtaining the highest possible yields of mature protein, the rate of protein synthesis must match the capacity of cells for post-translational modifications (Sevastsyanovich et al., 2009; Yoon et al., 2010). The secretion process involved in transporting apo-CCP once the signal peptide is cleaved has been identified previously as the rate limiting step in the accumulation of mature CCP (Turner et al., 2003). It has been shown in this work that the capacity for BL21* for an effective rate of mature CCP production is achieved when 0.04 mM IPTG is used to induce expression from a pET11c vector. A slightly higher concentration of IPTG resulted in greater synthesis of the pre-apo form of the target protein, which becomes aggregated in inclusion bodies in the cytoplasm (**Figure 5.8**). This illustrates the importance of balanced rates of protein synthesis and maturation for the accumulation of maximal yields of correctly folded protein.

The slow synthesis of recombinant protein is an important aspect for improved protein folding (Baneyx and Mujacic, 2004). However, in the case of difficult recombinant proteins, an increased capacity for post-translational modifications is also required to

accumulate high yields of the protein. This is exemplified by the different yields of mature CCP accumulated by strains P2 BL21* and C41.

The problem of inclusion bodies

The accumulation of misfolded intermediates even before the formation of inclusion bodies was identified as the main factor contributing to bacterial growth arrest (Hunke and Betton, 2003). The presence of misfolded proteins was able to induce the extracytoplasmic stress response in the periplasm of *E. coli*. This has also been illustrated after the production of mature CCP was induced using different concentrations of IPTG. Numbers of colony forming units after induction were highest when CCP production was induced with 0.01 mM IPTG. However slightly higher inducer concentrations that increased yields of mature CCP resulted in decreased cfu by almost 80%. Despite an increased synthesis rate of CCP, which was within the maturation capacity of the host cell (since there was no accumulation of pre-apo CCP in the cytoplasm), it is possible that unfolded polypeptides, which escape the folding chaperones or misfold, are continuously accumulated within cells. An increased concentration of misfolded protein intermediates might lead to unfavourable non-specific interactions with cellular proteins or intermediates, and is the key trigger for induction of the heat-shock response (Baneyx and Mujacic, 2004; González-Montalbán *et al*., 2005; 2007a)

The conformational status of protein in inclusion bodies

It has long been considered, and largely remains the consensus today, that inclusion bodies are mainly composed of unfolded and misfolded polypeptides that have aggregated as a result of non-specific interactions. There is growing evidence, however, that inclusion bodies are composed of proteins that are correctly folded and with biological activity (Garcia-Fruitos *et al*., 2005 a, b; 2007a; Aríe *et al*., 2006). This body of evidence has been obtained from production of proteins that are especially prone to aggregation, such as the viral capsid protein VP1 that had been fused to either GFP or β-galactosidase. The fluorescence emission or β-galactosidase activity was used as an indicator of correct protein folding. Most of the evidence that inclusion bodies comprised significant quantities of correctly folded protein was that high levels of fluorescence were measured from inclusion bodies that contained VP1GFP.

In this thesis, the conformational status of CheYGFP in the soluble and insoluble fraction was investigated by measuring fluorescence emission from these fractions at different periods post-induction. A clear correlation between the protein solubility and fluorescence was identified. In contrast, very little fluorescence was associated with the insoluble fraction. This was despite the fact that very high yields of protein had accumulated, as shown by SDS-PAGE analysis, reflecting that it had aggregated in inclusion bodies. Clearly, this indicates that the majority of the CheYGFP present in the soluble fraction is fluorescent and hence correctly folded. In contrast, the majority of CheYGFP that has aggregated in inclusion bodies is non-fluorescent and therefore incorrectly folded. The results presented here are in strong agreement with results reported from previous

research, which demonstrate the link between fluorescence of GFP protein fusions and solubility (Waldo *et al*., 1999; Cha *et al*., 2000; Drew *et al*., 2001; Hedhammar *et al*., 2006; Jones, 2007). In fact, results that have emerged from the laboratory of Villaverde and co-workers show that in wild type *E. coli*, the fluorescence of GFP protein fusions present in the soluble fraction is approximately 10-fold higher than that in inclusion bodies (García-Fruitós *et al*., 2007b; Martínez-Alonso *et al*., 2007, 2008; Vera *et al*., 2007). This provides strong evidence that the incorrectly folded recombinant protein is 'preferentially' located or directed by IbpA and IbpB into inclusion bodies (Allen *et al*., 1992; Han *et al*., 2004). These observations are therefore inconsistent with the recent model proposed by Villaverde and co-workers that folded protein is distributed equally between the soluble and insoluble fractions (González-Montalbán *et al*., 2007b). The very low levels of fluorescence emitted from the insoluble fraction could be due to contamination by or the entrapment of correctly folded recombinant protein. The possibility that this accounts for the activity of inclusion bodies has not yet been excluded. The contamination of inclusion bodies with other proteins such as chaperones has already been documented (Carrió and Villaverde, 2003; Han *et al*., 2004; Rinas *et al*., 2007).

Studies that report inclusion bodies contain correctly folded protein have been obtained with proteins that are inherently prone to aggregation. For example, a misfolding variant of GFP (mGFP) and a maltose binding protein (MalE31) fusion that aggregate in inclusion bodies (García-Fruitós *et al*., 2007b; Arié *et al*., 2007). Of course, biological activity from aggregated proteins is possible if the recombinant protein has both a correctly folded domain(s) that embodies the active site and unfolded stretches of polypeptide that facilitate aggregation. In the case of GFP, many GFP variants with slight changes in the

amino acid sequence are used in research today. Changes in the amino acid sequence often result in changes to the emission spectrum of the fluorophore in GFP, and this is used as a basis for the development of different forms of GFP (Jones, 2007). However, mutations that increase the stability of GFP have been reported (Pedelacq *et al*., 2006). This variant of GFP is called superfolding GFP as it is able to fold properly and fluoresce even if it is fused to misfolding-prone proteins that become aggregated in inclusion bodies as a consequence. Superfolding GFP was therefore shown to report reliably the total yield of the fused target protein, regardless of whether it is soluble or insoluble. It is possible that the (misfolding) variant of GFP used by Villaverde and co-workers could have harboured 'superfolding' mutations, which would be consistent with the results reported (García-Fruitós *et al*., 2007b).

The accumulation of CheYGFP in P2 BL21* and C41

The production of CheYGFP in improved hosts C41 and P2, which are derivatives of BL21(DE3) and BL21(DE3)*, respectively, was investigated. There is one major difference between those two hosts in that unlike BL21, BL21* contains a mutation in RNaseE that renders it dysfunctional. RNaseE is one of the major sources of mRNA degradation. Its loss therefore increases the stability of the mRNA transcripts and hence improves yields of the target protein. However, C41 accumulated much higher yields of total CheYGFP compared to P2 BL21* (**Figure 4.19**). In addition, approximately only half of the CheYGFP produced by C41 was soluble and the rest had accumulated into inclusion bodies (**Figure 4.20**). In contrast, the majority of CheYGFP produced by P2 was soluble and only a little was found to be insoluble. This was also reflected by the

higher specific fluorescence in P2 BL21* compared to C41 24 h post-induction (**Figure 4.17**). This indicates that the additional protein that had accumulated in the insoluble fraction of C41 did not contribute to fluorescence, providing further evidence that protein solubility is a good indicator of conformational quality.

It was surprising that C41 accumulated much higher yields of CheYGFP compared to P2. Clearly, the accumulation of higher protein yields did not correspond with good conformational quality, indicating that protein folding by chaperones, primarily DnaK and GroEL, must have been overwhelmed. Since approximately half of the CheYGFP protein produced by C41 had accumulated in inclusion bodies, then presumably there is also a paralleled increase in the concentration of unfolded intermediates in the cytoplasm. Unfolded proteins are known to be the signals that induce the heat shock response, if present in the cytoplasm, or extracytoplasmic response if triggered in the periplasm (Gill *et al*., 2000; Lesley *et al*., 2002; Hunke and Betton, 2003). As a consequence of the heat shock response, the accumulation of unfolded recombinant protein would therefore lead to growth arrest. Nevertheless, C41 continued to grow to higher optical densities compared to P2 BL21*. This raises the possibility that C41 might have additional mutations that interfere with the induction of the RpoH regulon and silences the heat-shock response. Alternatively, due to the presence of higher yields of inclusion bodies in C41, then perhaps mutations are present in DnaK chaperone that might decrease its efficiency for protein folding. This would be consistent with findings that in a *dnaK* mutant, higher yields of protein aggregate into much larger inclusion bodies (Carrió and Villaverde, 2003; González-Montalbán *et al*., 2005).

The accumulation of CheYGFP using the improved protocol or P2 BL21*

This study has shown the positive correlation between specific fluorescence and yields of soluble GFP protein fusions that have been accumulated. It reaffirms results shown in previous studies (Cha *et al*., 2000; Waldo *et al*., 1999). Higher fluorescence corresponds to higher yields of correctly folded GFP fusion protein accumulated. Interesting trends were observed, however, when comparing CheYGFP production by wild type and P2 BL21* hosts in shake flasks and batch fermentations (**Table 6.1**). It is clear that in both shake flasks and fermenters, P2 BL21* accumulates higher yields (%) of CheYGFP per gram of cell biomass compared to the parental strain. However, in fermentation experiments, the % soluble CheYGFP that had accumulated in P2 BL21* was significantly lower than the parental strain. This indicates that the higher capacity shown by P2 BL21* for RPP, perhaps due to an increased efficiency of protein synthesis, compromises to a great extent the solubility of the protein accumulated. Presumably, due to a limited capacity of chaperone-assisted folding within cells, the 'surplus' newly synthesized protein is consequently aggregated and deposited into inclusion bodies. Indeed, it has been reported that the yield of soluble recombinant protein accumulated within cells is governed by its total yield and perhaps also by the chaperone-folding capacity in bacteria (Martinez-Alonso *et al*., 2007). However, similar to the improved protocol, the accumulation of CheYGFP occurs at a very slow rate in P2 BL21*. Hence, it is unlikely that chaperones would have been saturated with unfolded polypeptides. An alternative explanation for the high levels of insoluble CheYGFP accumulated within P2 BL21* is that a deficiency in a component of the folding machinery, perhaps DnaK, could have occurred in P2 BL21*. Certainly, it has been reported that in the absence of DnaK, a

Table 6.1. The production of CheYGFP in wild type and P2 BL21

Host	Cultivation conditions	OD$_{650}$	Specific fluorescence (units)	Total fluorescence (units)	% CheYGFP	% soluble CheYGFP	Fluorescent CheYGFP (g L^{-1})
BL21* (improved protocol)	Shake flasks ($n = 8$)	14.3 (±0.3)	223 (±12)	3190 (±200)	22.5% (±4)	73% (±3)	14.5 (±1)
	Batch fermenter ($n = 3$)	17.1 (±0.4)	210 (±26)	3610 (±430)	21% (±2)	86.5% (±2)	16.4 (±2)
P2 BL21* (standard protocol)	Shake flasks ($n = 9$)	15.8 (±0.2)	250 (±7)	3960 (±75)	28.4% (±5)	72% (±5)	18 (±0.3)
	Batch fermenter ($n = 2$)	23.2 (±1.6)	210 (±0)	4950 (±350)	33% (±2)	55% (±3)	22.5 (±1.6)

The % CheYGFP of the total proteome was estimated from SDS-PAGE gels by densitometry. The proportion of soluble CheYGFP accumulation (% soluble) was also estimated from SDS-PAGE gels. The yield of fluorescent CheYGFP (g/ L) was estimated from the total fluorescence using the GFP calibration curve in which 1g/ L is equivalent to 2790 fluorescence units. Average data with standard error from replicates (*n*) are presented.

target protein that had accumulated in the soluble fraction was much less stable and more prone to aggregation compared to the parental strain. (García-Fruitós *et al*., 2005a).

Attempts to increase accumulation of soluble recombinant protein by the co-expression of chaperones were successful when whole repertoires of chaperones were co-expressed with target proteins (reviewed in Gasser *et al*., 2008; Sevastsyanovich *et al*., 2010). However, this approach adds the inevitable disadvantage of increased metabolic burden on the bacterial host (de Marco *et al*., 2007). A similar route to increase production of soluble recombinant protein, and with huge potential, would be to co-express variants of molecular chaperones that have a higher affinity and efficiency for protein folding. A study by Wang *et al*. (2002) has reported the use of successive rounds of DNA shuffling to evolve variants of GroEL that have an enhanced ability to fold GFP. Such methods of 'directed evolution' could be used to generate better variants of the molecular chaperone, DnaK, with a higher folding efficiency, or ClpB, with a better dissolution activity of inclusion bodies.

The differences in the % solubility of the CheYGFP accumulated by wild type BL21* and P2 BL21* were most apparent when optical densities were different. In shake flasks, optical densities of both strains and the % soluble CheYGFP accumulated were similar. However, in fermentations, optical densities of both strains 24 h post-induction and the % soluble CheYGFP accumulated within cells were significantly different. It has been reported that the specific growth rate of bacteria is inversely correlated with the quality of recombinant protein produced (Kim *et al*., 1996; Ihssen and Egli, 2004; Chou, 2007). Also, it was shown that a diluted growth medium that supported less bacterial biomass

resulted in decreased levels of insoluble CheYGFP (Jones, 2007). Increased growth rates therefore would result in increased protein aggregation. This has also been illustrated during production of CheYGFP by C41. Optical densities and total yields of CheYGFP accumulated by C41 were both higher than P2 BL21*, resulting in approximately only 50% of CheYGFP being soluble on C41 compared to around 80% in P2 BL21*. Further research is necessary to find ways in which both growth and solubility of protein accumulated could be maximised.

Evidence for the presence of multiple mutations in different improved BL21 hosts

Indications showing that improvements from the improved protocol and the improved host P2 BL21* were different in origin was revealed when CheYGFP was produced in batch fermentations. Yields of fluorescent CheYGFP (g/ L) produced by P2 BL21* host using the standard protocol were almost double that when the improved protocol with BL21* alone was used. This was due to a combined effect of a higher specific productivity and biomass of the improved host. The higher specific productivity of P2 BL21* compared to the parental strain during the improved protocol clearly suggest the presence of extra mutations that further enhance RPP in strain P2 BL21*, in addition to those causing the downregulation of T7 RNA polymerase,. The increased biomass might be due to an increased efficiency in the utilization of glucose. It has been previously documented that increased biomass and improved specific productivity of recombinant proteins cannot be achieved simultaneously (Chou, 2007). This leads to the suggestion that P2 BL21 might contain two (extra) mutations that independently enhance biomass and specific productivity.

The biggest indication of the presence of multiple mutations in improved hosts came from comparing production of CheYGFP and mature CCP in P2 BL21* and C43. The stress resistance of these two hosts can be attributed to the presence of the same mutations in the promoter of the T7 polymerase gene, yet P2 can accumulate CheYGFP whereas C43 cannot. Conversely, more mature CCP was accumulated in C43 than in P2. Clearly, therefore, these strains carry one or more additional mutations, at least one of which specifically determines which target protein or proteins can be accumulated. The difference in the ability of C41 and C43 to accumulate CheYGFP also indicates the possibility of the presence of additional mutations that accounts for this discrepancy. Wagner *et al* (2008) have attempted to test the whether the mutations in *lac*UV5 promoter were responsible for the improvement in RPP of membrane proteins. The *lac*UV5 promoter fragments of C43(DE3) and the parental strain BL21(DE3) were swapped, and production of a target protein, YidC::GFP, was tested. Results revealed that the wild type strain had been transformed to a Walker strain and *vice versa* with regards to the accumulation of the membrane protein YidC::GFP. This implies that the only mutation in the Walker strains are those that downregulateT7 RNA polymerase. Indeed, the mutations in the *lac*UV5 might account for the 'stress resistance' properties of the Walker strains, which is essential for improving RPP. However, indications of the presence of different additional mutations between strains might only become apparent during the production of different groups of recombinant protein.

Role of DnaK in enhancing solubility and conformational quality of target proteins

The chaperone DnaK aids the folding of newly synthesized polypeptides. Co-production of DnaK enhanced yields of both the conformational quality and solubility of the recombinant protein (Caspers *et al*., 1994; Chen *et al*., 2003; deMarco *et al*., 2007; Ikura *et al*., 2002; Rinas *et al*., 2007). Different results were reported by Villaverde and co-workers when DnaK was co-expressed with the target protein. They show that co-production of DnaK with an aggregation-prone fusion protein VP1GFP, using moderately low IPTG concentrations (0.02 mM), resulted in a slight increase in fluorescence and therefore an enhanced yield of correctly folded protein (Martínez-Alonso *et al*., 2007). However, this did not correspond to an increased yield of recombinant protein in the soluble fraction. The conclusion was therefore made that conformational quality and protein solubility are inconsistent parameters, and that conformational quality is a more accurate indicator of correct protein folding. It was also shown that production of VP1GFP alone, using high IPTG concentrations such as 0.1 and 1 mM, rather than 0.02 mM, resulted in significantly decreased solubility and fluorescence (Martínez-Alonso *et al*., 2007). This is consistent with findings from this study and previous reports that the production level of a recombinant protein is the overriding factor in determining its solubility. Unsurprisingly therefore, the solubility of VP1GFP remained very low when DnaK was co-produced in the same way. It is likely that the rapid synthesis of DnaK itself had resulted in its own aggregation.

The observation that DnaK increased yields of correctly folded VP1GFP during slow protein synthesis indicates that DnaK is more effective during slower accumulation of the

target protein. Given this observation, the role of DnaK during CheYGFP production in wild type BL21* using the improved protocol should be investigated. It is still unclear whether the slow rate of protein synthesis alone, which is key to the improved protocol, rather than the presence of DnaK, is sufficient to enhance yields of soluble recombinant protein. This could be investigated during production of recombinant protein using the improved protocol in a mutant strain of BL21* that is deficient in DnaK. In addition, the effect of DnaK co-expression on the conformational quality and solubility of CheYGFP in P2 BL21* and C41 strains, and whether both parameters can be simultaneously improved, is worthy of investigation.

The use of GFP for the production of outer membrane proteins

In this project, the production of both cytoplasmic and outer membrane proteins was investigated. The accumulation of cytoplasmic GFP fusion proteins such as CheY and Protein D were monitored by GFP fluorescence. However, this was not possible when GFP was fused with the outer membrane protein CCP or the human kinase protein S-GFP (Dr Sevastsyanovich; personal communication). Despite the accumulation of S-GFP and mature CCP within bacteria, this was not reflected by an increase in GFP fluorescence. This defines clear drawbacks of the use of GFP reporter system when producing proteins that are secreted by the Sec pathway (Feilmeier *et al*., 2001). However, it has been demonstrated that GFP can fluoresce in the periplasm of *E. coli* if it is exported by the Tat system (Thomas *et al*., 2002). Thus, it is possible to engineer outer membrane protein-GFP fusions with a tat signal sequence and exploit the tat export pathway. This could

facilitate the isolation of bacterial strains that accumulate higher yields of periplasmic or membrane proteins.

Methods for the isolation of improved hosts

The method used to isolate the first improved *E. coli* Walker strains was based on the effect of IPTG toxicity on colony forming units of bacteria containing the expression plasmid (Miroux and Walker, 1996). Mutant hosts were selected based on colony size on NA supplemented with 1 mM IPTG. This showed that bacterial clones were indeed resistant to IPTG induction stress. It remained unclear, however, as to whether those candidates could also produce recombinant protein. In this thesis, a better approach was developed for the selection of improved BL21* mutants. This method exploits the leakiness of the T7/ *lac* expression system during the production of GFP fusion proteins. By plating samples on NA alone 24 h after induction of the target gene, it was possible to identify stress resistant candidates that also have a pale green colony phenotype. This provided a quick screening advantage bacterial clones that have retained the recombinant plasmid. Thus, this also provides a quick indication of the possibility that those candidates are stress resistant and accumulate the target protein for prolonged periods. Using this method, approximately 50% of the pale green stress-resistant candidates were indeed improved hosts that accumulated high yields of correctly-folded recombinant protein.

In this project, improved strains were also isolated by Dr Sevastsyanovich using fluorescent activated cell sorting. Bacteria were screened for increased fluorescence by flow cytometry during production of CheYGFP 24 h post-induction using the standard protocol. Bacteria with increased fluorescence had accumulated higher yields of

fluorescent, correctly folded CheYGFP, and were therefore considered to be possible mutants with improved RPP characteristics. Several rounds of cell sorting were required to enrich the culture with improved bacterial strains. The advantage of using cell sorting to isolate improved hosts is that yields of correctly folded GFP-fusions within individual cells can be monitored by flow cytometry, and used as a basis to reflect their productive potential before cell sorting takes place. This differs from the plating method that exploits the leaky *lac*UV5 promoter to screen for pale green colonies of plasmid-containing bacteria. Two mutant BL21* strains, Y5 and Y8 were isolated in this way. They showed increased stress resistance and improved accumulation of soluble CheYGFP. The efficacy of this approach is yet to be exploited at its full potential.

The advantage of using GFP fusions to monitor the accumulation of correctly folded protein, and the ease of cell sorting, could facilitate the isolation of many more improved bacterial hosts that are tailored for the production of different types of recombinant protein. Thus, a repertoire of bacterial mutants that encompasses every single type of protein could be generated (**Figure 6.1**). The applicability of methods used here to isolate improved hosts could be extended to yeast and mammalian host cells, thereby increasing production of different recombinant proteins such as those that require glycosylation.
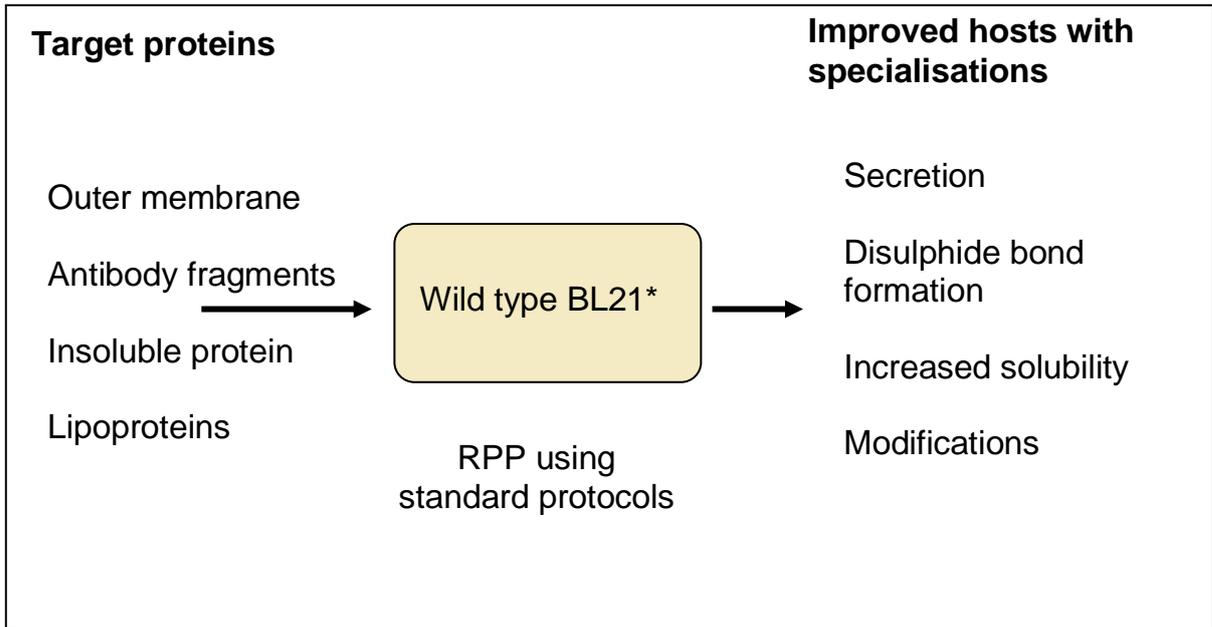
Figure 6.1: The isolation of different 'specialised' bacterial hosts. Different target proteins that require post-translational modification are produced in wild type BL21* using standard procedures. Mutant hosts that are stress resistant are isolated 24 h and screened for improved RPP properties. A catalogue of improved hosts that are specialised in the production of different classes of proteins could be generated.

Future work and prospects of RPP

Identification of genes that specifically enhance RPP

Seven studies that aim to analyse the transcriptional response of bacteria during RPP have been reported (reviewed in Sevastsyanovich *et al*., 2010). All of these studies share similar findings such as the upregulation of the heat shock response, unsurprisingly, due to the use of the same standard conditions to induce RPP. The improved protocol can be used as a platform for further investigation of the transcriptional response of bacteria during RPP without the consequence of physiological stress. Microarray analysis of bacteria during RPP using the improved protocol could reveal genes that specifically respond to RPP, which are either helpful or detrimental to the accumulation of correctly folded target protein. In addition, given that there are strong indications to the presence of different mutations in different improved hosts, transcriptomics could also be used to identify genes that are differentially expressed during production of different types of recombinant proteins.

Protein secretion into medium rather than accumulation within cells

Increasing yields of soluble, correctly folded recombinant protein within the cytoplasm is a main objective in bioindustry. A main obstacle to achieving this is that soluble protein within the cytoplasm is susceptible to proteolysis, mainly by ClpP and Lon, hence compromising yields obtained (Baneyx and Mujacic, 2004; Vera *et al*., 2007; Rinas *et al*., 2007). One possible solution to this problem is that recombinant protein could be produced in conditions that would favour their accumulation in inclusion bodies. This is because proteins retained within inclusion bodies are resistant to proteolysis, and that

protein will be obtained in a relatively high purity. However, the main drawback with this approach is the necessary costly procedures involving the denaturation and refolding of proteins obtained from inclusion bodies. On the other hand, the recent reports from the Villaverde laboratory, showing that inclusion bodies contain 'significant' amounts of correctly folded protein, might make this an appealing route to obtaining recombinant protein (Garcia-Fruitos *et al*., 2005; 2007; Carrio *et al*., 2005; Arie *et al*., 2006; de Groot and Villaverde., 2005).

Another, more promising approach of obtaining higher yields of the correctly folded target protein is by having it secreted to the (outer) medium as soon as it is produced. The secretion of the recombinant protein decreases the likelihood of degradation by proteases within the cell cytoplasm (Yoon *et al*., 2010). Currently, in collaboration with industrial partners, research is being directed towards developing methods that improve the downstream processing of recombinant protein. This can be achieved through secretion of the target protein directly into the growth medium such that purification and downstream processing becomes simpler and commercially advantageous. The secretion of recombinant protein in Gram negative bacteria can be achieved by means of the type V secretion system, which encompasses the autotransporter proteins (reviewed in Henderson *et al*., 2004). This secretion system enables the translocation of protein through the inner and outer membranes in a temporally dependent manner. Autotransporter proteins are synthesised as precursor proteins containing three functional domains; an N-terminal leader sequence, a passenger domain and a C-terminal helper or 'autotransporter' domain (reviewed in Henderson *et al*., 2004; Wells *et al*., 2007). The N-terminal leader sequence mediates export of the protein across the cytoplasmic membrane. The passenger domain

is the target protein to be secreted. The C-terminal domain adopts a β-barrel tertiary structure when embedded in the outer membrane and serves as a pore, facilitating translocation of the passenger domain in to the extracellular medium. Current prospects are to utilise the autotransporter secretion system by engineering target proteins with autotransporter domains to facilitate their secretion (Henderson *et al*., 2004).

Amongst the target proteins required in high quality are antibodies. The development of better procedures or the isolation of improved hosts that efficiently produce and secrete large amounts of antibody protein into the outer medium would be invaluable to bioindustry. Ongoing research has revealed that the improved strain P2 BL21* can also efficiently secrete recombinant protein in large quantities into the growth medium (Dr Sevastsyanovich*, personal communication*). The implications of obtaining bacterial hosts that combine capabilities of improved production yields of different types of protein, secretion, translocation and post-translational modifications could open up new avenues and possibilities that lead to a revamping of the recombinant protein production scene.

Sequencing of improved hosts

In this project, the improved hosts P2 BL21* and Mutant Pro were isolated. Both strains continued to form colonies after recombinant protein production was induced using the standard protocol. However, high yields of mature CCP were accumulated by Mutant Pro but not by P2 BL21*. The difference in their ability to accumulate different proteins was originally attributed to how those strains were isolated. In other words, strains isolated during production of different recombinant proteins might accumulate mutations that render them better able to accumulate those groups of proteins (**Figure 6.1**). Evidence for

the directed evolution of bacterial strains has been obtained from the genome sequence of BL21. It was revealed that the largest deletion of *ompT*, which has been desirable for recombinant protein production, arose spontaneously. OmpT is an outer membrane protease that can degrade protein during purification. The improved strains isolated in this project were obtained from different starter stocks of BL21* strain, kindly donated by different laboratories. The genome sequence of BL21 from two independent studies has been recently published (Jeong *et al*., 2009). Surprisingly, a difference of 62 genes between both genome sequences was found. The significant divergence of the genome sequence between parental strains of BL21 indicates high plasticity of the BL21 genome. However, the ease with which BL21 genome can accumulate mutations will have implications on the reliability of conclusions made with regards to physiological responses during RPP. At present, the naïve BL21* strain from which P2 was isolated is being sequenced at GlaxoSmithKline. Once completed, it would be possible to compare improved strains of BL21* to the parental strain by genome hybridisation, in addition to facilitating future microarray work using this strain.

Already it is obvious that one is not limited to traditional systems and local solutions to solving the problems associated with RPP. With the availability of whole genome sequences of bacterial hosts, strategies based on a systems biology approach will inevitably generate rational improvements to RPP.

REFERENCES

Albano, C.R., Randers-Eichhorn, L., Bentley, W.E. and Rao, G. (1998) Green fluorescent protein as a real time quantitative reporter of heterologous protein production. *Biotechnol Prog*. **14**: 351-4.

Allen, S.P., Polazzi, J.O., Gierse, J.K. and Easton, A.M. (1992) Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J Bacteriol*. **174**: 6938-47.

Andersson, L., Yang, S., Neubauer, P. and Enfors, S.V. (1996) Impact of plasmid presence and induction on cellular responses in fed batch cultures of *Escherichia coli*. *J Biotechnol*. **46**: 255-63.

Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M.J. and Walker, J.E. (2000) Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale overproduction of the b subunit of F_1F_o ATP synthase. *FEBS Lett*. **482**: 215-9.

Arie, J., Miot, M., Sassoon, N. and Betton, J. (2006) Formation of active inclusion bodies in the periplasm of *Escherichia coli*. *Mol Microbiol*. **62**: 427-37.

Arsène, F., Tomoyasu, T. and Bukau, B. (2000) The heat shock response of *Escherichia coli*. *Int J Food Microbiol*. **55**: 3-9.

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

Barker, M.M., Gaal, T. and Gourse, R.L. (2001) Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *J Mol Biology*. **305**: 689-702.

Barkley, M. D., Riggs, A. D., Jobe, A. and Bourgeois, S. (1975) Interaction of effecting ligands with *lac* repressor and repressor-operator complex. *Biochemistry*. **14**: 1700-12.

Bentley, W.E. and Kompala, D.S. (1990) Optimal induction of protein synthesis in recombinant bacterial cultures. *Ann N Y Acad Sci*. **589**: 121-138.

Bentley, W.E., Mirjalili, N., Andersen, D.C., Davis, R.H. and Kompala, D.S. (1990) Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng*. **35**: 668-681.

Bernhardt, J., Völker, U., Völker, A., Antelmann, H., Schmid, R., Mach, H. and Hecker, M. (1997) Specific and general stress proteins in *Bacillus subtilis*--a two-deimensional protein electrophoresis study. *Microbiology*. **143**: 999-1017.

Borth, N., Mitterbauer, R., Mattanovich, D., Kramer, W., Bayer, K. and Katinger, H. (1998) Flow cytometric analysis of bacterial physiology during induction of foreign protein synthesis in recombinant *Escherichia coli* cells. *Cytometry*. **31**: 125-9.

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. *J Biol Chem*. **260**: 3539-41.

Carrió, M.M., Corchero, J.L. and Villaverde, A. (1999) Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms. *Biochim Biophys Acta*. **1434**: 170-6.

Carrió, M.M. and Villaverde, A. (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett*. **489**: 29-33.

Carrió, M.M. and Villaverde, A. (2003) Role of molecular chaperones in inclusion body formation. *FEBS Lett*. **537**: 215-21.

Carrió, M.M. and Villaverde, A. (2005) Localisation of chaperones DnaK and GroEL in bacterial inclusion bodies. *J Bacteriol*. **187**: 3599-601.

Carrió, M.M., González-Montalbán, N., Vera, A., Villaverde, A. and Ventura, S. (2005) Amyloid-like properties of bacterial inclusion bodies. *J Mol Biol.* **347**: 1025-37.

Cashel, M.D.R., Gentry, V.J., Hernandez and Vinella. (1996) The stringent response. In *Escherichia coli* and *Salmonella*: cellular and molecular biology. Neidhardt, F.C., Curtiss R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., et al. (eds). Washington, DC: American Society for Microbiological Press, pp. 1458-96.

Caspers, P., Stieger, M. and Burn, P. (1994) Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*. *Cell Mol Biol*. **40**: 635-44.

Cha, H.J., Wu, C.F., Valdes, J.J., Rao, G. and Bentley, W.E. (2000) Observations of green fluorescent protein as a fusion partner in genetically engineered *Escherichia coli*: monitoring protein expression and solubility. *Biotechnol Bioeng*. **67**: 565-74.

Chen, Y., Song, J., Sui, S.F. and Wang, D.N. (2003) DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in *Escherichia coli*. *Protein Expr Purif*. **32**: 221-31.

Chou, C.P. (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl Microbiol Biotechnol*. **76**: 521-32.

Corchero, J.L. and Villaverde, A. (1998) Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily and specifically influenced by features of the coded proteins. *Biotechnol Bioeng*. **58:** 625-32.

Crooke, H. and Cole, J. (1995) The biogenesis of *c*-type cytochromes in *Escherichia coli* requires a membrane-bound protein, DipZ, with a protein disulphide isomerase-like domain. *Mol Microbiol*. **15**: 1139-50.

Cubitt, A.B., Wollenweber, L.A. and Heim, R. (1999) Understanding structure-function relationships in the *Aequorea Victoria* green fluorescent protein. *Methods Cell Biol*. **58**: 19-30.

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D. and von Heijne, G. (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science*. **308**: 1321-3.

deBoer, H.A., Comstock, L.J. and Vasser, M. (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci USA*. **80**: 21-5.

De Groot, N.S. and Ventura, S. (2006a) Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett*. **580**: 6471-6.

De Groot, N.S. and Ventura, S. (2006b) Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J Biotechnol*. **125**: 110-3.

Delagrave, S., Hawtin, R.E., Silva, C.M., Yang, M.M. and Youvan, D.C. (1995) Red-shifted excitation mutants of the green fluorescence protein. *Biotechnol*. **13**: 151-4.

DeLisa, M.P., Li, J., Rao, G., Weigand, W.A. and Bentley, W.E. (1999) Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. *Biotechnol Bioeng*. **65**: 54-64.

De Marco (2007) Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. *Nat protoc*. **2**: 2632-9.

Dong, H., Nilsson, L. and Kurland, C.G. (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol.* **177**: 1497-1504.

Drew, D.E., Von Heijne, G., Nordlund, P. and de Gier, J.W. (2001) Green fluorescent protein as an indictor to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett*. **507**: 220-4.

Durfee, T., Hansen, A.M., Zhi, H., Blattner, F.R. and Jin, D.J. (2008) Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol*. **190**: 1084-96.

Dürrschmidt, K., Reischer, H., Schmidt-Heck, W., Hrebicek, T., Guthke, R., Rizzi, A and Bayer, K. (2008) Monitoring of transcriptome and proteome profiles to investigate the cellular response of *E. coli* towards recombinant protein expression under defined chemostat conditions. *J Biotechnol*. **135:** 34-44.

Eymann, C., Homuth, G., Scharf, C. and Hecker, M. (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J Bacteriol*. **184**: 2500 – 20.

Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, H. and Phillips, G.J. (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J Bacteriol*. **182**: 4068-76.

Gadgil, M., Kapur, V. and Hu, W.S. (2005) Transcriptional response of *Escherichia coli* to temperature shift. *Biotechnol Prog*. **21**: 689-699.

Gamer, J., Bujard, H. and Bukau, B. (1992) Physical interaction between heat shock proteins DnaK, DnaJ and GrpE and the bacterial heat shock transcription factor sigma 32. *Cell* **69**: 833-42.

García-Fruitós, E, Carrió, M.M., Arís, A. and Villaverde, A. (2005a) Folding of a misfolding-prone β-galactosidase in absence of DnaK. *Biotechnol Bioeng*. **90**: 869-75.

García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R.M., Arís, A., Ventura, S. and Villaverde, A. (2005b) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb Cell Fact*. **4**: 27.

García-Fruitós, E., Arís, A. and Villaverde, A. (2007a) Localisation of functional polypeptides in bacterial inclusion bodies. *Appl Environ Microbiol*. **73**: 289-94.

García-Fruitós, E., Martínez-Alonso, M., González-Montalbán, N., Valli, M., Mattanovich, D. and Villaverde, A. (2007b) Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. *J Mol Biol*. **374**: 195-205.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D. and Villaverde, A. (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Micro Cell Fact*. **7**:11.

Georgiou, G. and Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr Opin Biotechnol.* **7**: 190-7.

Gill, R.T., Valdes, J.J. and Bentley, W.E. (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabol Eng.* **2**: 178-89.

González-Montalbán, N., Carrió, M.M., Cuatrecasas, S., Arís, A. and Villaverde, A. (2005) Bacterial inclusion bodies are cytotoxic in vivo in absence of functional chaperones DnaK and GroEL. *J Biotechnol*. **118**: 406-12.

González-Montalbán, N., Villaverde, A. and Arís, A. (2007a) Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies. *Biochem Biophys Res Commun.* **355**: 637-42.

González-Montalbán, N., García-Fruitós, E. and Villaverde, A. (2007b) Recombinant protein solubility – does more mean better. *Nat Biotechnol*. **27**: 718-20.

Graf, A., Gasser, B., Dragosits, M., Sauer, M., Leparc, G.G., Tüchler, T., Kreil, D.P. and Mattanovich, D. (2008) Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics*. **19**: 390.

Grossman, A.D., Taylor, W.E., Burton, Z.F., Burgess, R.R. and Gross, C.A. (1985) Stringent response in *Escherichia coli* induces expression of heat shock proteins. *J Mol Biol*. **186**: 357-365.

Guisbert, E., Herman, C., Lu, C.Z. and Gross, C.A. (2004) A chaperone network controls the heat shock response in *E. coli*. *Genes Dev*. **18**: 2812-2209.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol.* **177:** 4121-30.

Haddadin, F.T and Harcum, S.W. (2005) Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. *Biotech Bioeng*. **90**: 127-153.

Han, M.J., Park, S.J., Park, T.J. and Lee, S.Y. (2004) Roles and applications of small heat shock proteins in the production of recombinant proteins in *Escherichia coli*. *Biotech Bioeng*. **88**: 426-36.

Harcum, S.W. and Haddadin, F.T. (2006) Global transcriptome response of recombinant *Escherichia coli* to heat-shock and dual heat-shock recombinant protein induction. *J Ind Microbiol Biotechnol*. **33**: 801-814.

Hedhammar, M., Stenvall, M., Lonneborg, R., Nord, O., Sjolin, O., Brismar, H., Uhlen, M., Ottosson, J. and Hober, S. (2005) A novel flow cytometry-based method for analysis of expression levels in *Escherichia coli*, giving information about precipitated and soluble protein. *J Biotechnol*. **119**: 133-46.

Heim, R., Cubitt, A.B. and Tsien, R.Y. (1995) Improved green fluorescence. *Nature.* **373**: 663-4.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. and Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev*. **68**: 692-744.

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**: 179 – 87.

Hewitt, C.J. and Nebe-Von-Caron, G. (2004) The application of multi-parameter flow cytometry to monitor individual microbial cell physiological state. *Adv. Biochem. Eng. Biotechnol.* **89**: 197-223.

Hoffmann, F. and Rinas, U. (2000) Kinetics of heat-shock response and inclusion body formation during temperature-induced production of basic fibroblast growth factor in high-cell-density cultures of recombinant *Escherichia coli. Biotechnol Prog*. **16**: 1000-7.

Hoffmann, F. and Rinas, U. (2004a) Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli. Adv Biochem Eng Biotechnol*. **89**: 143-161.

Hoffmann, F. and Rinas, U. (2004b) Stress induced by recombinant protein production in *Escherichia coli. Adv Biochem Eng Biotechnol*. **89**: 73-92.

Hunke, S. and Betton, J.M. (2003) Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. Mol Microbiol. **50**: 1579-89.

Ihssen, J. and Egli, T. (2004) Specific growth rate and not cell density controls the general stress response in *Escherichia coli*. *Microbiol*. **150**: 1637-48.

Ikura, K., Kokubu, T., Natsuka, S., Ichikawa, A., Adachi, M., Nishihara, K., Yanagi, H. and Utsumi, S. (2002) Co-overexpression of folding modulators improves the solubility of the recombinant guinea pig liver transglutaminase expressed in *Escherichia coli*. *Prep Biochem Biotechnol*. **32**: 189-205.

Jones, J.J., Bridges, A.M., Fosberry, A.P., Gardner, S., Lowers, R.R., Newby, R.R., James, P.J., Hall, R.M. and Jenkins, O. (2004) Potential of real-time measurement of GFP-fusion proteins. *J Biotechnol.* **109**: 201-11.

Jones, J.J. (2007) PhD Thesis submission. *The University of Birmingham*.

Jürgen, B., Hanscke, R., Sarvas, M., Hecker, M. and Schweder, T. (2001) Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Appl Microbiol Biotechnol*. **55**: 326-332.

Jürgen, B., Lin, H. Y., Riemschneider, S., Scharf, C., Neubauer, P., Schmidt, R., Hecker, M. and Schweider, T. (2000) Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations. *Biotechnol Bioeng*. **70**: 217-224.

Kiefhaber, T., Rudolph, R., Kohler, H.H. and Buchner, J. (1991) Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (N Y),* **9**: 825-829.

Kim, S.S., Kim E.K. and Rhee, J.S. (1996) Effects of growth rate on the production of *Pesudomonas fluorescens* lipase during the fed-batch cultivation of *Escherichia coli*. *Biotechnol Prog*. **12**: 718-22.

Kim, S.Y., Ayyadurai, N., Heo, M.A., Park, S., Jeong,Y.J. and Lee, S.G. (2009) Improving the productivity of recombinant protein in *Escherichia coli* under thermal stress by co-expressing GroELS chaperone system. *J Microbiol Biotechnol.* **19**: 72-7.

Krojer, T., Sawa, J., Schäfer, E., Saibil, H.R., Ehrmann, M. and Clausen, T. (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature.* **453:** 885-890.

Kurland, C.G. and Dong, H.J. (1996) Bacterial growth inhibition by overproduction of protein. *Mol Microbiol*. **21**: 1-4.

Kusano, K., Waterman, M.R., Sakaguchi, M., Omura, T. and Kawaga, N. (1999) Protein synthesis inhibitors and ethanol selectivity enhance heterologous expression of P450s and related proteins in *Escherichia coli*. *Arch Biochem Biophys*. **367**: 129-36.

Lesley, S.A. Graziano, J., Cho, C.Y., Knuth, M.W. and Klock, H.E. (2002) Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Prot Eng*. **15**: 153-60.

Lethanh, H., Neubauer, P. and Hoffmann, F. (2005) The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant *Escherichia coli* and delay degradation of inclusion bodies. *Microb Cell Fact*. **4**: 6.

Lewis, G., Taylor, I.W., Nienow, A.W. and Hewitt, C.J. (2004) The application of multi-parameter flow cytometry to the study of recombinant *Escherichia coli* batch fermentation processes. *J Indust Micobiol Biotechnol*. **31**: 311-22.

Martínez-Alonso, M., Vera, A. and Villaverde, A. (2007) Role of the chaperone DnaK in protein solubility and conformational quality in inclusion-body forming *Escherichia coli cells*. *FEMS Microbiol Lett*. **273**: 187-195.

Martínez-Alonso, M., García-Fruitós, E. and Villaverde, A. (2008) Yield, solubility and conformational quality of soluble proteins are not simultaneously favoured in recombinant *Escherichia coli*. *Biotechnol Bioeng*. **101**: 1353-7.

Mattanovich, D., Gasser, B., Hohenblum, H. and Sauer, M. (2004) Stress in recombinant protein producing yeasts. *J Biotechnol*. **113**: 121-35.

Mattanovich, D. and Borth, N. (2006) Applications of cell sorting in biotechnology. *Microb Cell Fact*. **5**: 12.

Miller, W.G. and Lindow, S.E. (1997) An improved GFP cloning cassette designed for prokaryotic transcriptional fusions. *Gene*. **191**: 149-153.

Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E. and Bukau, B. (2003) Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol Microbiol*. **50**: 585-95.

Metheringham, R., Griffiths, L., Crooke, H., Forsythe, S. and Cole, J. (1995) An essential role for DsbA in cytochrome *c* synthesis and formate-dependent nitrite reduction by *Escherichia coli* K-12. *Arch Microbiol*. **164**: 301-7.

Narayanan, N., Hsieh, M., Xu, Y. and Chou, C.P. (2006) Arabinose-induction of *lac*-derived promoter systems for penicillin acylase production in *Escherichia coli. Biotechnol Prog*. **22**: 617-625.

Nichols, J. C., Vyas, N. K., Quiocho, F. A. and Matthews, K. S. (1993) Model of lactose repressor core based on alignment with sugar-binding proteins is concordant with genetic and chemical data. *J. Biol. Chem.* **268**: 17602-12.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. and Yura, T. (1998) Chaperone co-expression plasmids: differential and synergistic roles DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen Cryj2 in *Escherichia coli. Appl Environ Microbiol.* **64**: 1694-9.

Oehler, S., Eismann, E.R., Krämer, H. and Müller-Hill, B. (1990) The three operators of the *lac* operon cooperate in repression'. *EMBO J*. **9:** 973-9.

Patkar, A., Vijayasankaran, N., Urry, D.W. and Srienc, F. (2002) Flow cytometry as a useful tool for process development: rapid evaluation of expression systems. *J Biotechnol*. **93**: 217-29.

Paul, B.J., Berkmen, M.B. and Gourse, R.L. (2004) rRNA transcription in *Escherichia coli*. *Annu Rev Genet*. **38**: 749-770.

Paul, B.J., Berkmen, M.B. and Gourse, R.L. (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci USA*. **102**: 7823-8.

Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol*. **24**: 79-88.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I. and Vassylyev, D.G. (2004) Regulation through the secondary channel – structural framework for ppGpp-DksA synergism during transcription. *Cell*. **118**: 297-309.

Rinas, U., Hoffmann, F., Betiku, E., Estape, D. and Marten, S. (2007) Inclusion body anatomy and functioning of chaperone-mediated *in vivo* inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. *J Biotechnol*. **127**: 244-57.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd ed. Plainview, NY: Cold Spring Harbour Laboratory Press.

Sevastsyanovich, Y., Alfasi, S. and Cole, J. (2009) Recombinant protein production: a comparative view on host physiology. *N Biotechnol*. **25**: 175-80.

Sevastsyanovich, Y., Alfasi, S., Overton, T., Hall, R., Jones, J., Hewitt, C. and Cole, J. (2009) Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. *FEMS Microbiol Lett*. **299**: 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. *Biotehcnol Appl Biochem*. **55**: 9-28.

Shaw, M.K. and Ingraham, J.L. (1967) Synthesis of macromolecules by *Escherichia coli* near the minimal temperature for growth. *J Bacteriol*. **94**: 157-64.

Silversmith, R.E. and Bourret, R.B. (1999) Throwing the switch in bacterial chemotaxis. *Trends Microbiol*. **7:** 16-22.

Silverstone, A.E., Ariditti, R.R. and Magasanik, B. (1970) Catabolite-insensitive revertants of *lac* promoter mutatnts. *Proc Natl Acad Sci*. **66**: 773-9.

Skogman, G., Nilsson, J., and Gustafsson, P. (1983) The use of a partition locus to increase stability of tryptophan-operon-bearing plasmids in *Escherichia coli*. *Gene* **23:** 105-15.

Smith, H. E. (2007) The transcriptional response of *Escherichia coli* to recombinant protein insolubility. *J Struct Funct Genomics.* **8**: 27-35.

Sorensen, H.P and Mortensen, K.K. (2005) Advanced genetic strategies for recombinant expression in *Escherichia coli*. *J Biotechnol.* **115**: 113-128.

Soriano, E., Borth, N., Katinger, H. and Mattanovich, D. (1999) Flow cytometric analysis of metabolic stress effects due to recombinant plasmids and proteins in *Escherichia coli* production strains. *Metabol Eng*. **1**: 270-4.

Soriano, E., Borth, N., Katinger, H. and Mattanovich, D. (2002) Optimization of recombinant protein expression level in *Escherichia coli* by flow cytometry and cell sorting. *Biotechnol Bioeng*. **80**: 93-9.

Spiess, C., Beil, A. and Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell*. **97**: 339-47.

Stock, J. and Surette, M. (1994) Bacterial chemotaxis. The motor connection. *Curr Biol*. **4:** 143-4.

Striedner, G., Pfaffenzeller, I., Markus, L., Nemecek, S., Grabherr, R. and Bayer, K. (2010) Plasmid-free T7-Based *Escherichia coli* expression systems. *Biotechnol Bioeng*. **105**: 786-794.

Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct the selective high-level expression of cloned genes. *J Mol Biol*. **189**: 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*. **185**: 60-89.

Sundström, H., Wållberg, F., Ledung, E., Norrman, B., Hewitt, C.J. and Enfors, S.O. (2004) Segregation to non-dividing cells in recombinant *Escherichia coli* fed-batch fermentation processes. *Biotechnol Lett*. **26**: 1533-9.

Tanaponqpipat, S., Reid, E., Cole, J.A. and Crooke, H. (1998) Transcriptional control and essential roles of the *Escherichia coli ccm* gene products in formate-dependent nitrite reduction and cytochrome *c* synthesis. *Biochem J*. **334**: 355-65.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*. **72**: 211-222.

Thomas, J.D., Daniel, R.A., Errington, J. and Robinson, C. (2001) Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. *Mol Microbiol*. **39**: 47-53.

Traxler, M.F., Summers, S.M., Nguyen, H.T., Zacharia, V.M., Hightower, G.A., Smith, J.T. and Conway, T. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol*. **68**: 1128-48.

Turner, S., Reid, E., Smith, H. and Cole, J. (2003) A novel cytochrome *c* peroxidase from *Neisseria gonorrhoeae*: a lipoprotein from a Gram–negative bacterium. *Biochem J*. **373**: 865-73.

Turner, S.M., Moir, J.W.B., Griffiths, L., Overton, T.W., Smith, H. and Cole, J.A. (2005) Mutational and biochemical analysis of cytochrome *c'*, a nitric oxide- binding lipoprotein important for adaptation of *Neisseria gonorrhoeae* to oxygen-limited growth. *Biochem. J*. **388**: 545-53.

Vera, A., Arìs, A., Carrio, M., Gonzalez-Montalban, Villaverde, A. (2005) Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. *J Biotechnol*. **119**: 163-171.

Vera, A., González-Montalbán, N., Arìs, A. and Villaverde, A. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng*. **96**: 1101-6.

Waldo, G.S., Standish, B.M., Brendzen, J. and Terwilleger, T.C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol*. **17**: 691-5.

Wang, J.D., Herman,C., Tipton, K.A., Gross, C.A. and Weismann, J.S. (2002) Directed evolution of substrate-optimised GroEL/S chaperonins. *Cell*. **111**: 1027-39.

Wells, T.J., Tree, J.J., Ulett, G.C. and Schembri, M.A. (2007) Autotransporter proteins: novel targets at the bacterial cell surface. *FEMS Microbiol Lett*. **274**: 163-72.

Wendrich, T.M., Blaha, G., Wilson, D.N., Marahiel, M.A. and Nierhaus, K.H. (2002) Dissection of the mechanism for the stringent factor RelA. *Mol Cell*. **10**: 779-88.

Westers, L., Westers, H. and Quax, W.J. (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimise the host organism. *Biochim Biophys Acta*. **1694**: 299-310.

Yang, F., Moss, L.G. and Phillips, G.N. (1996) The molecular structure of green fluorescent protein. *Nat Biotech*. **14**: 1246-51.

Yoon, S.H., Kim, S.K. and Kim, J.F. (2010) Secretory production of recombinant proteins in *Escherichia coli*. *Rec Pat Biotechnol*. **4**: 23-9.

Young, J.C., Agashe, V.R., Siegers, K. and Hartl, F.U. (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev*. **5**: 781-791.

Zhang, Y., Taiming, L. and Liu, J. (2003) Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in *Escherichia coli*. *Prot Expression purif*. **29**: 132-9.

Zhao, J.B., Wei, D.Z. and Tong, W.Y. (2007) Identification of *Escherichia coli* host cell for high plasmid stability and improved production of antihuman ovarian carcinoma x antihuman CD3 single-chain bispecific antibody. *Appl Microbiol Biotechnol.* **76**: 795-800.

Zhao, J.B. and Wei, D.Z. (2007) Use of derivative of *Escherichia coli* BL21(DE3) for efficient production of three different recombinant proteins. *Biotechnol Prog*. **23**: 1043-8.