Rational design of improved host cell platforms for recombinant protein production by modulating the stress response in *Escherichia coli*

Thesis submitted to Jawaharlal Nehru University

in partial fulfillment for the

award of the degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By

RICHA GULERIA

SCHOOL OF BIOTECHNOLOGY
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI-110067
INDIA

2019

School of Biotechnology
Jawaharlal Nehru University
New Delhi – 110067
July 2019

CERTIFICATE

This is to certify that the work titled **"Rational design of improved host cell platforms for recombinant protein production by modulating the stress response in *Escherichia coli*"** submitted to the School of Biotechnology, Jawaharlal Nehru University, New Delhi, in fulfillment of the requirements for the award of the degree of Doctor of Philosophy, embodies a faithful record of original research work carried out by **Richa Guleria**. She has worked under my guidance and supervision. This work is original and has not been submitted so far in part or full for any other degree or diploma of any other university.

Richa Guleria
Ph.D. Candidate
School of Biotechnology
Jawaharlal Nehru University
New Delhi-110067

Prof. Pawan K. Dhar
Supervisor
School of Biotechnology
Jawaharlal Nehru University
New Delhi-110067

Professor
School of Biotechnology
Jawaharlal Nehru University
New Delhi-110067

Prof. Pawan K. Dhar
Dean
School of Biotechnology
Jawaharlal Nehru University
New Delhi-110067

PAWAN K. DHAR
Dean
School of Biotechnology
Jawaharlal Nehru University
New Delhi-110067

Dedicated to my Family

Acknowledgement

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

- Marie Curie

This journey has been the gift bestowed by the divine blessings. My first thanks go to Almigthy for all the beautiful and learning experiences throughout this journey.

This thesis work is a materialization of ideas and initiatives of my supervisor and mentor Prof. K.J. Mukherjee, who retired before I could finish my Ph.D research work in his lab. A sincere and heartfelt gratitude to him for the advices and guidance provided to conceive this area of research. As a teacher he always attempted to provide an environment, where we can learn to think, design and execute ideas which always shaped our scientific skills. I admire him for his great guidance, invaluable expertise, constant encouragement, constructive criticism and his extra-ordinary sense of humor. I consider myself fortunate to work under his supervision for four years. Thank you Sir, for all the support and valuable guidance during those years, which shaped this research work even after you left.

My special words of thanks to Prof. Pawan K. Dhar, who agreed to be my official supervisor after Prof. Mukherjee's retirement. I owe my deepest gratitude to him for extending his support in the best possible way he can, so that I can complete my Ph.D work on time. Thank you sir, for keeping me motivated during the course of my thesis. I would also like to convey my special thanks to Dr. Amulya K. Panda (NII) and Dr. Swati Tiwari for giving me valuable suggestions as doctoral committee members which helped me in improving my thesis greatly. I also owe a word of thanks to Prof. Rajiv Bhat for his guidance during the course work, which laid the foundation for my years as a Research Fellow. A warm thanks to Prof. Aparna Dixit for allowing me to set up RT-PCR experiments in her lab.

It's my pleasure to thank all the faculty members and office staff at the School of Biotechnology, JNU for sincere co-operation as and when required. A special thanks to Salil Sir and Tiwari sir for facilitating all the official work. No research is complete without a library, hence I owe thanks to Yadav ji for all their support related to use of library. I would like to extend my thanks to AIRF facility, JNU and CIF facility, NII for providing their instrumentation facility for conducting experiments. I am highly thankful to Priyank and Jayaram (NII Ph.D scholars) for their help during proteomics work.

Lab 110 of SBT has been a second home to me during all these years. I will always cherish the time spend here with seniors, friends and machines. Thanks to the friendly seniors Shubha

Mam, Vijeta Mam and Neetu Mam who made my life easy in the lab as well as in Godavari hostel, when I was a new comer here. Many times when I felt demotivated, Vijeta Mam remained the source of extra-motivation. Special thanks to Jaya Mam for being a wonderful senior who always cleared my doubts and helped me with scientific suggestions. My friend Priyanka, without her it would have been difficult to bring this work to an end. She always remained a helping hand to me in good and hard times. I will never forget the moments of bench work, and the pleasure during collective excursions with my labmates Neeraj, Neha Mam and Deepak. Among lab members, Geeta mam deserves a special mention for ensuring the smooth functioning with respect to all financial matters of the lab. I would also like to express my gratitude to my old colleagues Manjul Mam, Deepti Mam, Anita Mam, Arunagshu, Esha and Sarfaraz. My friend Ritu Som and Radha deserves a special mention here for always extending their help and making things workable.

I am thankful to Vinod bhaiya, Sanjay bhaiya, Ghanshyam bhaiya and other lab assistants who have been a great help by making our lives so much easier in lab.

I acknowledge the CSIR (Council of Scientific and Industrial Research), Government of India for providing me with SRF fellowship to pursue research at JNU. Department of Biotechnology is gratefully acknowledged for providing research funding.

My parents and siblings are my pillar of strength, who have formed my way of living and are my inspirations. They believed in my dream and nourished it with love, patience and support. Words can't express how grateful I am for their unconditional love and encouragement at every stage of my personal and professional life. Thank you so much to all of you for having strong belief in me, being patient and giving me liberty to choose what I desired. My acknowledgement would never be complete without the special mention of my new family members (Mumma, Shilpi and Saurabh) who remained supportive and understanding during my hard times.

Lastly and most importantly, none of these achievements would have been possible without the love and constant support of my husband Gaurav. A few lines in this page will be insufficient to value the unconditional love and immense care bestowed upon me by him. The completion of this work would never have been possible without his whole hearted encouragement in words and deeds. Thanks for your love and care. I cannot think of anything worth achieving without you.

Richa

Abbreviations

ATP	Adenosine triphosphate	**ml**	millilitre
AmpR	Ampicillin resistance	**µg**	microgram
APS	Ammonium persulphate	**mM**	millimolar
BSA	Bovine serum albumin	**MQ**	MilliQ water
bp	base pair	**mRNA**	messenger RNA
$CaCl_2$	Calcium chloride	**N**	Normality
CSR	Cellular stress response	**NaCl**	Sodium chloride
CBB	Coomassie Brilliant Blue	**NaOH**	Sodium hydroxide
cDNA	complementary DNA	**NAPs**	Nucleiod associated proteins
DCW	Dry cell weight	**NMR**	Nuclear magnetic resonance
DO	Dissolved oxygen	**ng**	nano gram
DNA	Deoxyribonucleic Acid	**nm**	nanometre
ddH_2O	Double distilled water	**$O.D._{600}$**	Optical Density at 600 nm
dNTP	deoxyribonucleotide triphosphate	**(p)pGpp**	Guanosine terta/penta-phosphate
DTT	Dithiothreitol	**PBS**	Phosphate buffered saline
EDTA	Ethylene diamine tetra-acetic acid	**PAGE**	Polyacrylamide gel electrophoresis
g	gram	**PCR**	Polymerase chain reaction
gm/l	gram/litre	**PCD**	Programmed cell death
GC MS	Gas chromatography-mass spectrometry	**PMSF**	Phenylmethylsulfonyl fluoride
HCDC	High cell density cultivation	**%**	Percentage
h	hour	**RNase**	Ribonuclease
IPTG	Isopropyl-β-D-thiogalactopyranoside	**RT-PCR**	Real Time PCR
IAA	Iodoacetamide	**RPM**	Revolutions per minute
kb	kilobase	**RT**	room temperature
L	Litre	**RPP**	Recombinant protein production
LA	Luria Agar	**sfGFP**	Super folding variant of GFP
LB	Luria Broth	**SDS**	Sodium dodecyl sulphate
L-asp	L-asparaginase	**sec**	second
LC MS	Liquid chromatography-mass spectrometry	**σ**	sigma
$MgCl_2$	Magnesium Chloride	**TAE**	Tris-acetate EDTA
mg/l	milligram/litre	**Tris**	Tris(hydroxylmethyl) aminomethane
M	Molarity	**TB**	Terrific Broth
mg	milligram	**TEMED**	N, N, N', N' tetramethyl ethylene diamine
µl	microliter	**tRNA**	transfer RNA
µM	Micro molar	**YE**	Yeast extract

Table of Contents

List of Figures

Figure 4.8(d): Gel picture showing amplification of *Ara* Promoter.

Figure 4.9(a): PCR confirmation *of glpD-araProm-glpK*insert (3.2 kb) by forward *glpD* and reverse *glpK* primers.

Figure 4.9(b): PCR confirmation of *araProm-glpK* insert (1.7 kb) by forward *araProm* and reverse *glpK* primers.

Figure 4.10(a): Growth profiles of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

Figure 4.10(b): Residual substrate profiles of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

Figure 4.10(c): Total volumetric productivity of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

Figure 4.10(d): Volumetric productivity per unit biomass of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

Figure 4.11(a): Growth profiles of control cultures expressing L-asp and test cultures co-expressing genes for glycerol uptake and L-asp production upon glycerol pulsing.

Figure 4.11(b): Residual glycerol profiles of control and test cultures induced with IPTG upon glycerol pulsing.

Figure 4.11(c): Total volumetric productivity for control and test cultures induced with IPTG at 12 and 24 hours post induction.

Figure 4.11(d): Volumetric productivity per unit biomass for control and test cultures induced with IPTG at 12 and 24 hours post induction.

Figure 4.11(e): SDS gel picture showing extracellular expression of L-Asp post-induction 6, 12 and 24 hours in BWΔ(*elaA+cysW)* and BWΔ(*elaA+cysW)glpDK* strains induced with IPTG only.

Figure 4.12(a): Growth profile for control and modified cells under non-producing conditions.

Figure 4.12(b): Viable cell count for control and modified cells under non-producing conditions.

List of Tables

1. Introduction

Two emergent and related disciplines have revolutionized the landscape of modern biology. The first constitutes the broad range of extremely powerful techniques and tools loosely grouped together as "omics" technologies - essentially genomics, transcriptomics and proteomics. Together they have generated enormous amounts of data on cellular physiology at the fundamental level of genes and proteins and how they seamlessly and synergistically interact to form the basic network of life. However, it has been the path-breaking developments in the second discipline i.e. systems biology that has enabled a deeper understanding of this complex network and also laid the theoretical foundations for analyzing and constructing these networks. This has helped in organizing the voluminous data generated by "omics" tools into meaningful and hierarchical structures bringing for the first time clarity in the relationship between genotype and phenotype. Thus whole genome sequencing today allows us to understand what the cell "can" and more importantly "cannot" do. Critically from the point of view of designing improved cell platforms for metabolite production, it compares the innate abilities of different hosts and identifies the chassis which will prove to be superior in production. Using basic stoichiometric principles and coupling them with optimizing functions like maximizing growth or yield, has led to predictions of metabolic fluxes having excellent match with experimental data. These optimizing functions essentially assume that the cell has evolved to its maximum efficiency in terms of utilizing substrate for growth though other equally valid efficiency criteria like minimizing resources have led to different optimizing functions with equally good predictive value. These tools, like flux balance analysis (FBA) or elementary mode analysis have been further fine-tuned using transcriptomic and proteomic data by appropriately changing the constraints on the flux cone (in FBA) to reflect the limitations imposed by the availability of different proteins or their cognate mRNAs. Thus from time course transcriptomic and proteomic data, snapshots of the dynamic changes in metabolic fluxes have also been captured using methods like iFBA or dFBA. These predicted changes in flux have also been validated using both theoretical and experimental metabolic flux analysis tools for example using pulses of C^{13}-labeled glucose and their capture using GC-MS, LC-MS and NMR. These

unprecedented developments have paved the way for not only a deeper insight into cellular functioning but also provided the tools and targets for rational host design. Today we can precisely engineer cells to perform at its peak capacity in terms of its productivity, yield, efficient substrate utilization etc., traits which have a critical impact on the viability and profitability of the biotechnology industry.

However much more needs to be done. Only a few industrially important microbes like *Escherichia coli, Pseudomonas, Bacillus* and yeasts like *Saccharomyces cerevisiae* have been analyzed in depth using the above tools and many prospective candidates still remain to be studied. They may well prove to be more suitable for specific applications but the lack of genetic engineering tools for many microbial and fungal systems means that the potential of these hosts may be difficult to exploit. It is for these historically contingent reasons that we continue to rely on traditional hosts like *E. coli* with their well-established tools for genetic engineering, multiplicities of vectors, expression systems and genome engineering. Even with such well-studied systems our knowledge is unfortunately far from complete. Thus typical FBA models used by researchers using EcoCyc and Metaflux software contain less than 1450 genes and equivalent number of metabolites (Weaver *et al*. 2014) and of the remaining more than 500 genes many still do not have any known function and hence cannot be incorporated into any model.

More importantly our attempts to correlate gene expression levels with transcription factor activities using Global regulatory networks (GRNs) have been a mixed bag of success (Fang *et al*. 2017; Larsen *et al*. 2019). Our knowledge of transcriptional regulatory networks, even in *E. coli*, is both incomplete and non-robust and predicting gene expression from these networks has only been moderately successful. Even more critical, our understanding of the environmental signals which trigger these transcriptional networks is woefully inadequate. To truly understand the relationship between genotype and phenotype we need to complete the circle; how the genotype leads to the phenotype is known, but how the micro-environment affects the phenotype and this in turn triggers an appropriate genomic response is unknown. Only when this knowledge is complete we can claim to have a complete dynamic picture of cellular behavior. For some specific cases like presence of antibiotics, toxins, heavy metals etc., or an

unfavorable environment in terms of pH or temperature the cell mounts a stress response which is reasonably well characterized. Thus, the transcriptional factors and the changes in gene expression associated with different stress responses like heat shock, osmotic stress etc., are well known. But even here the signaling molecules which trigger this stress have rarely been identified. With this large gap in information it is difficult to expediently manipulate the cellular physiology to serve specific goals.

In our case we approached this, as of now intractable, problem from the extremely practical aim of maximizing recombinant protein production (RPP). As will be elaborated in later sections, the problem of optimizing recombinant protein expression has been studied in depth over the past two decades. Given its relevance to the biotech industry, a comprehensive research program dedicated to the design of improved vectors, expression systems, improved *in-vivo* folding and even export has been undertaken and these efforts have been extremely successful. Today we have an extensive repertoire of genetic engineering tools which can be used for host cell engineering in *E. coli*. However, the over-expression of recombinant proteins leads to the diversion of metabolic fluxes towards product formation triggering a stress response within the cell. This stress response mimics many features of the heat shock response, osmotic stress and generalized stress response (Wick and Egli 2004; Singh *et al*., 2012; Carneiro *et al*., 2013). This in turn, down-regulates multiple pathways within the cell; these include critical energy generation and substrate uptake pathways which directly impact on protein production. Consequently in most cases protein production ceases within a few hours post induction leading to poor yields. In contrast metabolite production in *E. coli* can continue for fairly long periods leading to very high product concentrations. Thus, products like threonine have been produced at more than 100 g/l (Chen *et al*. 2009), whereas recombinant protein yields are typically in the range of 1-3 g/l. Clearly the bottleneck lies in our inability to sustain protein expression for longer time periods. If only we could understand the mechanism by which the cell mounts the stress response and somehow modulate it, then we should be able to design the next generation host platforms with significantly improved expression levels. These would no longer rely on incremental improvements in expression vector capabilities to increase production. Additionally they would provide useful insights on the genotype–phenotype linkage and

address the gap that exists in our knowledge of cellular dynamics. In the subsequent sections we will go into detail about the present state of knowledge on recombinant protein production, the existing problems and some of the interesting discoveries in our lab which paved the way for a more innovative and radically different approach to this problem.

1.1 Basic problems with recombinant protein production

Even as biotechnology has expanded into areas as diverse as pharmaceutical production, diagnostics, production of industrial biochemical, secondary agriculture etc., its life blood has remained developments in recombinant DNA technology and specifically the over expression of recombinant proteins. Since the last two decades the focus has been on the design of improved host systems with regulated gene expression in order to adapt to high cell density bioprocess conditions where typically micro anaerobic conditions and acetic acid accumulation are major issues (Eiteman and Altman 2006; Lara *et al*. 2006). Success in production of heterologous proteins has been achieved in last few decades by choosing an appropriate combination of host organism, plasmid system, promoter, selection markers, signal peptides and affinity tags for efficient purification (Rosano and Ceccarelli 2014). The primary focus of most of these studies has been on increasing the metabolic flux through RPP pathway which has been realized by incorporating stronger and tightly regulatable promoters, better ribosome binding sites etc., essentially focusing on increasing the rates of transcription and translation (Mahalik *et al*., 2014). Another set of related work has been on post translational steps essentially proper folding of the nascent polypeptide and its export to the periplasm or even in some cases to the extracellular medium (de Marco 2009; Albiniak *et al*. 2013; Matos *et al*. 2014)

It is well known that recombinant protein synthesis is an energy intensive process and utilizes a substantial amount of host's cellular energy in the form of amino acids, nucleotides, charged tRNAs and ribosomal machinery needed for efficient translation (Glick 1995; Mahalik *et al*., 2014). This metabolic drain forces the cell to reorient its metabolic fluxes and enzyme composition (Hoffmann and Rinas, 2004), which leads to significant alterations in biochemistry and physiology of host cells. Earlier these changes

in host physiology were viewed as "metabolic load or metabolic burden" which often results in growth retardation and hence, low product yields (Bentley *et al*., 1990; Glick 1995). However, it is more logical to view these changes as a "cellular stress response" (CSR) generated by the cell itself so that it can cope up with the specific stress conditions and prevent actual damage. The extent to which this cellular stress response is generated depends upon specific properties of the recombinant protein which is being expressed (Glick 1995; Panda *et al*, 1999) rather than on the amount of protein expressed. Thus in many cases no significant correlation is observed between growth retardation and high or low levels of recombinant protein expression (Bhattacharya *et al*. 2005; Srivastava and Mukherjee 2005; Vaiphei *et al*., 2009).

With the emergence of systems biology, the emphasis is now on gaining an in depth knowledge of host cell metabolism and its response to the induction of recombinant protein which manifests itself as cellular stress. However, if we look at the strategies that has been used to counter CSR, the main focus has been on improving individual steps like transcription (Giacalone *et al*. 2006; Y. J. Choi *et al*. 2010; Balzer *et al*. 2013), translation (Care *et al*. 2008; Salis *et al*., 2009; Schlesinger *et al*. 2017), protein folding (Hayer-Hartl 2002; Lesley *et al*. 2002; Jhamb and Sahoo 2012), export (Khushoo *et al*. 2004; Matos *et al*. 2012) and reducing by product accumulation (Dittrich *et al*. 2005; Eiteman and Altman 2006; De Mey *et al*. 2007). These extensive and diverse attempts all focuses on optimizing individual steps of recombinant protein synthesis instead of taking into account the larger systems biology picture that essentially enforces an upper limit on recombinant protein production. This is the primary reason why recombinant protein yields have plateaued to a maximum of around 3-4 grams per litre.

1.2 The regulatory control of recombinant protein production pathway

Eventually the regulatory control for RPP does not lie in any of these individual steps of the protein production pathway. Unlike metabolites, where manipulating a single pathway or blocking competing pathways and feedback steps improves production, recombinant protein expression consumes a major part of the cellular economy. Under recombinant protein producing conditions, the majority of cellular energy (approx 50%)

is directed towards protein synthesis and many highly expressed proteins comprise approximately 50-55% of the total cellular protein. That is why the cell has multiple regulatory mechanisms operating at different levels to govern the expression of single genes or operons making this regulation much more complex. Even with the data available after the integration of various disciplines, our understanding of this system level coordination is still in adequate. Therefore in addition to the present knowledge of the global regulation of the protein synthesis process, we also need to take into consideration a more comprehensive picture of cellular dynamics which would be helpful in unveiling some unknown factors that might have crucial role in disrupting or redirecting gene regulatory networks. This will further help in developing a truly dynamic model of cellular behavior under stress conditions, and facilitate the design of next generation host platforms with improved expression capabilities.

1.3 Analysis of recombinant protein mediated cellular stress response

Recent advancements in host cell engineering approaches and the application of various "omics" technologies like transcriptomics, proteomics, metabolomics and fluxomics powered by computational modeling has aided in our understanding of cellular dynamics allowing a better assessment of global cellular physiology with respect to various regulations at the transcriptional and translational level. Many transcriptomic studies have tried to address the issue of maximization of productivity by identifying the key regulatory molecules that get up and down regulated upon induction of foreign protein synthesis (Haddadin and Harcum 2005; Singh *et al*., 2012; Marisch *et al*. 2013). The genes that are down regulated in most of the cases belong to the carbon utilization, energy generation and ribosomal biosynthesis pathways (Ow *et al*. 2006; Dürrschmid *et al*. 2008). The problem of down regulation of critical genes, post-induction has been solved by co-expressing those genes either in plasmid based expression systems (Choi *et al*. 2003) or by making direct modifications into the host genome in order to increase their copy number in the host cell (Singh and Mukherjee 2013). However, the number of modifications required to supplement the expression of such a large number of genes severely limits the use of this strategy. The genes that get up-regulated due to recombinant protein mediated stress are the heat shock genes, stringent stress response

genes, oxidative stress and mRNA degradation genes (Oh and Liao 2000; Ow *et al*. 2006). Till date there are very limited studies that have focused on exploring the mechanisms via which these up-regulated genes mediate the CSR. In an attempt to identify the role of up-regulated genes, comparative transcriptomic profiling of high cell density post induction cultures of three different recombinant proteins (human interferon-β (IBs), xylanase and GFP (soluble)) was done previously in our lab and the genes that were commonly up-regulated across all the time points in three different cultures were identified (Sharma *et al*. 2011). It was hypothesized that many of these up-regulated genes could act as signaling messengers of the CSR which feedback controls recombinant protein expression (Sharma, 2015). This hypothesis was tested by knocking out some of the commonly up-regulated genes. These genes were non-essential and also did not have any downstream regulatees which ensured that there would be no cascading effects of these knock outs. The enhancement in protein expression levels in these knock out hosts was monitored using two model proteins viz. L-asparaginase and green fluorescent protein (GFP). Interestingly, some of these knock-outs showed more than 7 folds of improvements in expression levels compared to unmodified hosts and most of these genes whose deletion helped in improving expression levels have not been annotated and their function is as yet unknown. Subsequently, the synergistic effect of these modifications was studied by combining them in all possible combinations to generate double knock-outs. Results obtained showed that different combinations of knock-outs led to the maximum enhancements for different proteins, indicating that the gains obtained by these modifications were product specific. The leads obtained from this study can ideally be used to design a panel of genome engineered strains, each of which can be used for over expressing a different set of proteins. The major advantage of this strategy of making multiple knock-outs is that we can now utilize the synergistic effect of these genes to directly modulate the CSR and prevent down-regulation of a large number of genes. This is far more elegant than individually trying to up regulate all those critical genes which get down regulated due to the mounting of the CSR and are typically responsible for feedback inhibition of recombinant protein synthesis. These knock outs also provide key information of the critical gaps that exist in our knowledge of how the

cell responds to stress conditions and triggers global changes in gene expression patterns that helps it to maximize its chances of survival.

1.4 Stationary phase recombinant protein production

It has been reported earlier that stationary phase protein production is one of the fundamental properties of *E. coli* (Ou *et al*. 2004). Galloway *et al*., (2003) showed that late log phase induction is the most significant factor that contributes to increased soluble protein yields and low proteolytic rates. Enhancement in soluble protein expression and extracellular secretion of recombinant L-asparaginase-II upon late log phase induction has also been shown previously in our lab, where up to 3 folds of increase in specific productivity was obtained compared to cultures induced in mid-log phase (Khushoo *et al*. 2004). However, there are no reports available that have studied the reasons behind this improvement in production efficiency upon stationary phase induction. Comparison of transcriptomic profiles of soluble and insoluble forms of protein expression has revealed that the soluble protein expression imposes an additional stress to the cell in terms of down regulation of amino acid uptake and biosynthesis genes which is absent when the protein is expressed as inclusion bodies (IBs) (Sharma *et al*. 2011). These stress responses often mimic other physiological stress signals triggered upon stationary phase arrival like starvation stress, oxidative stress and stringent stress. Hence, it becomes very crucial to understand how the cell coordinates and integrates the large number of different signals originating from changes in the cellular environment. Also what contributes to the cell's adaptation in this changed environment which helps it in sustaining such high level of expression for long hours is still not known. Answering this question seems more complex than expected because of the complexities involved in transcriptional and translational regulations and their inter-relationship which is yet to be explored.

At a practical level, the present work focuses on designing improved host cell platforms for recombinant protein over-expression. Additionally, we have attempted to unravel the signaling pathways involved in generating this recombinant protein mediated cellular stress. For this purpose leads have been taken from previous work done in our lab

mostly based on analyses of transcriptomic profiles of recombinant protein producing cultures utilizing both control and modified host cells. Thus, the degree and extent to which the CSR was modulated by the knock out background provided useful hints on the way forward. For e.g., the down-regulation of substrate uptake genes post induction was observed in both control as well as modified cultures (Jain 2018) prompting us to independently augment their expression levels. Recombinant L-asparaginase was chosen as a model protein for our work. The results obtained from this study will be helpful in improving our knowledge of cellular physiology and concomitantly used for designing improved expression platforms with increased production efficiencies.

Aims and objectives

The overall aims and objectives of this work were:

1. Designing an improved host platform for enhancing recombinant protein yields using targets obtained from the transcriptomic analysis of control and modified cells post induction.

    a) Construction of expression vector for co-expression of substrate utilization genes (*glpDK*) for enhanced substrate uptake.

    b) To test double knock strains for improved substrate uptake and hence improved protein expression by co-expressing *glpDK* genes.

2) To improve the understanding of signaling pathways leading to CSR by analyzing the stress response.

    a) To identify the regulatory signals that play critical role in mounting stress under physiological as well as recombinant protein mediated cellular stress conditions.

    a) Proteomic profiling of control and modified cells to find key proteins involved in stress pathways.

3) To find the common regulatory pathways that mimic the features of recombinant protein mediated cellular stress. Hence, designing rational strategy for modified hosts for improved protein production.

2. Review of Literature

2.1 Recombinant Protein Production

Efficient strategies focused on enhanced recombinant protein production are gaining importance due to increasing requirement of high amounts of high quality proteins. With the emergence of the bio-similar market, enhancing the efficiencies of the production process has become a critical factor in commercial viability; similarly for industrial enzymes, bulk production of low cost enzyme is desired especially in areas like bio-fuels and metabolite production by enzymatic transformations. Recombinant DNA technology has become a novel tool in obtaining desired proteins yields with lower product costs needed for commercially viable processes. For production purposes, strain development is the most critical step which fixes all downstream processes for product recovery and purification. Various bioprocess strategies for increasing product yields are primarily focused on designing efficient microbial hosts or vector systems by employing various genetic manipulations using recombinant DNA technology. Since, recombinant protein expression is a multistep process involving transcription, translation, protein folding and export, the knowledge of expression limiting steps, protein secretion, post-translational issues, kinetics of soluble and insoluble protein expression, pathway construction etc. is required to realize the full potential of this technology.

2.2 *E. coli* as a host for recombinant protein production

The choice of appropriate host system depends upon the chemical properties of desired recombinant protein. Bacterial cells are efficient hosts for proteins that do not require extensive post-translational modifications like glycosylation because of their relative simplicity in terms of biochemistry and physiology. The process development with bacterial systems is generally cheaper than eukaryotic systems because of short process times and low media costs. *E. coli* has remained one of the extensively used host systems for recombinant protein production (RPP) due to various advantages over other expression hosts such as well studied genetics, simple growth requirements, fast high-cell density cultivation on inexpensive substrates, large number of molecular biology tools available for genetic manipulations, simple protein purification steps etc. The first

recombinant human protein produced in *E. coli* was somatostatin (Itakura K. *et al*, 1977). Human insulin, the first commercial recombinant drug approved by FDA was also synthesized in *E. coli* (Johnson 1983). Since then, *E. coli* has been used as host for production of many recombinant proteins approved for human use like interferons, antibody fragments, growth factors etc. (Walsh G 2010). Today, *E. coli* serves as a production host for approximately 30% of the approved therapeutic proteins (Baeshen *et al*, 2015) (Table 2.1).

Table 2.1: List of FDA approved therapeutic proteins produced in *E. coli.*

Biopharmaceutical products	Therapeutic indication	Year of approval	Company
Humulin (rh insulin)	Diabetes	1982 US	Eli Lilly
Glucagon	Hypoglycemia	1998 US	Eli Lilly
Lantus (long-acting insulin glargine)	Diabetes	2000 US	Aventis
Kineret (anakinra)	Rheumatoid arthritis	2001 US	Amgen
Natrecor (nesiritide)	Congestive heart failure	2001 US	Scios Inc
Calcitonin (recombinant calcitonin salmon)	Post menopausal osteoporosis	2005	Upsher-Smith Laboratories
Preotact (human parathyroid hormone)	Osteoporosis	2006 EU	Nycomed, Danmark
Nivestim (filgrastim, rhGCSF)	Neutropenia	2010	Hospira
Voraxaze (glucarpidase)	Lowering of toxic level of methotrexate conc. in patients with impaired renal function	2012	BTG International
Preos (parathyroid hormone)	Osteoporosis, hypoparathyroidism	2013 EU	NPS Pharmaceuticals

Based on Baeshen *et al*., (2015) J. *Microbiol. Biotechnol*., 25(7), 953–962

Recent advances in host cell engineering has made possible to produce smaller and less complex N-linked recombinant glycoproteins in *E. coli* by engineering its export system and reducing the rigidness of the bacterial oligosaccharyltransferase (Jaffé *et al*. 2014; Strutton *et al*. 2017).

There are countless reports available in literature on potential of *E. coli* cells as microbial cell factories. In spite of the detailed knowledge of *E. coli* genetics, the expression of every gene with high efficiency can't be guaranteed in this bacterium. This is because the factors such as the structural features of gene sequence, mRNA stability, codon usage, efficient protein folding, cytoplasmic degradation of recombinant protein by ATP-dependent proteases and toxicity of protein contribute to great extent to the expression level of a gene in addition to the choice of microbial host (Baneyx 1999; Swartz 2001).

2.3 Problems associated with recombinant protein production

The cellular restrictions caused due to over production of a recombinant protein can be grouped into two categories: metabolic limitations caused due to process of recombinant protein production; and cellular responses to the recombinant product synthesis. At metabolic level, cell experiences significant decline in growth rate upon recombinant protein synthesis due to sharing of cellular energy (ATP), reducing equivalents (NADH) and amino acids between growth and production processes, which leads to perturbations of flux rates in cellular networks. This metabolic limitation can be at the level of single enzyme catalyzed rate limiting step or due to limitation of multiple enzymes in the pathway. The intracellular restrictions due to cellular responses includes problems with respect to substrate uptake systems, changes in respiratory status of cell, down-regulation of energy metabolism or limitation of protein synthesizing machinery. However, these bottlenecks to recombinant protein production are not always stable and usually change with culture conditions (Klumpp *et al*., 2009). It is well known that introduction and expression of foreign protein imposes metabolic burden and impairs host cell metabolism. Several factors like size of cloning vector, plasmid copy number, composition of growth medium and metabolic state of cell determine the extent of metabolic burden imposed on host cell (Glick 1995). However, in recent times the introduction of various '-omics' technologies have made researchers more inquisitive in

their approach to understand the physiology of host cells producing recombinant protein. High level protein expression imbalances the flux rates through metabolic pathways which generates metabolic restrictions at several steps and makes cells sensitive to stress. The global control systems like stringent stress response gets activated during slow growth and stress conditions, limiting the synthesis of translational machinery and other components required for cellular growth.

2.3.1 Rate limiting steps in RPP

Like metabolite production, recombinant protein synthesis also occurs in multiple steps viz transcription, translation, protein folding and export that are intricately linked to each-other and to the host cell's cellular machinery. Due to this complex multi-step regulation, it is difficult to predict the optimal expression for a given recombinant protein even after the availability of efficient host backgrounds, high copy number vectors, selection markers, efficient tags for protein purification etc. The notion that transcription is a rate limiting step in recombinant protein production process led to huge amount of research done on vector development which includes (i) the development of high copy number plasmids like pUC series that contain 500-700 copies per cell (Minton 1984) and the pACYC and pBAD series with ori (p15A) compatible to pMB1 vectors (Chang and Cohen 1978; Guzman *et al*. 1995) for dual expression of proteins, (ii) strong promoters like *tac*, T7 (de Boer *et al*., 1983; Studier and Moffatt 1986) and promoters with tunable expression like araP_{BAD} (Guzman *et al*. 1995) (iii) affinity tags like poly-Arg-, poly-His-, c-Myc- etc. to aid purification (Terpe 2003) and (iv) introduction of coding sequences for tag removal like enterokinase, thrombin etc. (Jenny *et al*., 2003; Blommel and Fox 2007). But the enhancement in transcript levels do not necessarily increase protein expression levels, which has also been shown in by some research groups (Hou *et al*., 2013). This generally happens when cell fails to match the rate of translation in terms of supply of precursor molecules like amino acids, charged tRNA, ATPs etc. with the pace of transcript generation. This results in degradation of excess of mRNA transcripts in the cytoplasm (Kucharova *et al*. 2013). Equally important is the problem of protein folding and transport. The limited availability of molecular chaperons that are required to assist the folding of recombinant protein being over-produced results in formation of inclusion

bodies. Similarly, the export of the recombinant protein outside the cell is also a challenging task. The cytoplasmic accumulation of recombinant proteins sets an upper limit to its production and increases the steps in downstream processing. Above all, the recombinant protein production elicits a global cellular stress response which feedback inhibits each individual step of the protein production pathway and leads to declined growth and low product yields (Figure 2.1). Therefore, the knowledge of these rate limiting steps is crucial before designing any strategies to over produce recombinant proteins.

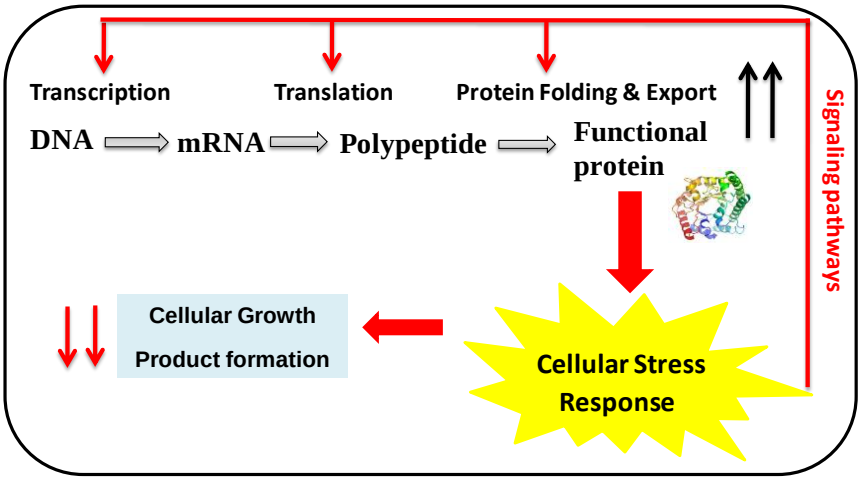

Figure 2.1 Schematic representation of recombinant protein production pathway and its inhibition by CSR.

2.3.2 Bacterial stress responses

Bacterial stress response can be defined as a cascade of alterations in global regulatory networks due to changes in gene expression patterns and protein activity in order to survive harsh and constantly changing conditions sensed by bacterial cells. The additional energy demand due to expression of plasmid based genes causes major perturbation of the host metabolism (F Hoffmann and Rinas 2001). Cell maintains homeostasis by various regulatory molecules that function as sensors and effectors. These molecules sense the changes in nutrient availability, temperature, pH, oxidation state, ribosome disruption etc. in terms of changed macromolecular composition (DNA and protein ratios) and the response of global regulatory networks of cell is proportional to the extent of these changes (Ron 2013). Normally, a bacterial cell is exposed to a multitude of stress

factors simultaneously which generally activates interconnected cellular mechanisms. But the response systems cell activates in order to establish homeostasis are specific to the kind of stress signal (Kültz 2005).

In bacteria, there are multiple control elements that regulate stress responses. These includes the control of transcription by alternative sigma factors and small RNAs, transcriptional control by repressor binding to DNA and transcriptional control by proteolysis i.e. SOS response (Ron 2013). The bacterial cells respond to stress in two ways: they counteract to these stress-induced changes by temporarily increases tolerance towards it; or they remove damaged cells by activating programmed cell death mechanism. The various signaling pathways that forms the part of bacterial stress response are discussed below.

2.3.2.1 The Stringent Stress Response

Bacterial cells respond to nutrient starvation stress by initiating complex protective responses that generates a stringent stress response by accumulation signaling alarmone guanosine tetra- and penta-phosphate, collectively called (p)ppGpp inside cells. The events that leads to intracellular accumulation of (p)ppGpp are triggered by amino acid starvation, phosphate, fatty acids, carbon or iron limiting conditions and variety of other stress signals (Battesti and Bouveret, 2009). The stringent response acts as control mechanism that reduces the expression of growth related genes on sensing the scarcity of substrate molecules for protein synthesis, and increases the expression of genes associated with cell maintenance and survival that economize the utilization of limited nutrients via transcriptional switching (Sharma and Chatterji 2010). The cellular pool of (p)ppGpp is regulated by members of RSH super-family and gram negative bacteria like *E. coli* contains two long-RSH syntheses known as RelA and SpoT (Irving and Corrigan 2018). Under nutrient rich conditions, RelA remain bounded to 70S ribosomes and catalyses (p)ppGpp synthesis only when ribosomal elongation stalls upon amino acid starvation due to accumulation of decylated tRNAs. *relA* encodes ppGpp-synthase I which synthesizes ppGpp upon sensing amino acid starvation specifically. However, *spoT* plays dual catalytic role; by encoding ppGpp-synthase II which synthesizes ppGpp under more diverse starvation conditions like carbon, phosphate or fatty acids limitation,

and by encoding ppGpp-hydrolase under favorable growth conditions which degrades the ppGpp accumulated inside cells (Figure 2.2) (Dalebroux and Swanson 2012).

The RelA/SpoT signaling pathway elevates the (p)ppGpp levels inside cells which in turn inhibits stable RNA synthesis and reduces the growth rate (Ramagopal and Davis 1974). Recently, (p)ppGpp has been shown to indirectly regulate the synthesis of ribosomal proteins upon nutritional shifts in log phase (Burgos *et al*. 2017). It is also known to positively regulate the transcription of ribosomal modulation factor (RMF) which reduces the cellular translational capacity by dimerizing 70S ribososmes to 100S form (known as sleeping 100S dimmers) without changing the total number of ribosomal proteins (Izutsu *et al*., 2001). (p)ppGpp acts as a key player in controlling *rpoS* mediated general stress response by inducing the anti-adaptor proteins IraP and IraD that blocks RssB (an adaptor protein) activity which otherwise directs σ^S to the action of ClpXP proteases (Bougdour and Gottesman 2007; Merrikh *et al*., 2009; Girard *et al*., 2017). The activity of σ^E regulon is also known to be signalled by (p)ppGpp which along with DksA protein directs σ^E mediated transcription of genes (Costanzo *et al*., 2008).

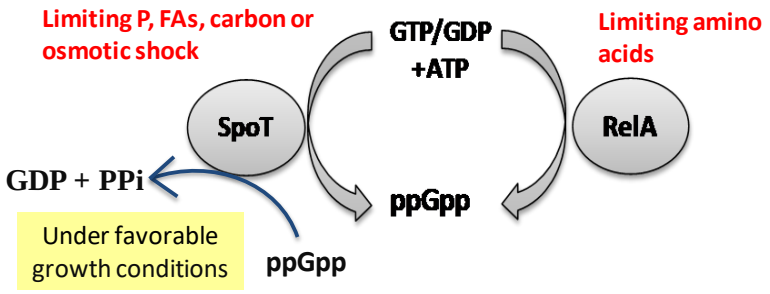

Figure 2.2 The pictorial representation of mechanism of (p)ppGpp accumulation and degradation inside cells.

2.3.2.2 σ^S (rpoS) mediated stress response

σ^S *(rpoS)* is the master regulator of the general stress response and starvation response in many Gram-negative bacteria. In *E. coli*, the stationary phase dependent specific gene expression is mainly controlled by this master regulator (Landini *et al*., 2014). The *rpoS* concentrations inside cells are dependent upon growth conditions and the nature of stress induced. The exponentially growing cells contain very low amounts of *rpoS*; the amounts

remain undetectable during growth on rich media but rise by several folds during entry to the stationary phase. The stress condition such as starvation, oxidative stress, shift to hyperosmolarity and exposure of cells to either low or high pH increases *rpoS* levels inside cells. *rpoS* induction has been shown to regulate the expression of 23% of *E. coli* genes, although to a different extent (Wong *et al*., 2017). In one study, the association of σ^S with the promoters regions of 63 genes during the early stationary phase has been identified by chromatin immunoprecipitation-sequencing (ChIP-seq) (Peano *et al*. 2015). A recent study conducted on *Salmonella* revealed that many *rpoS* concentration dependent genes are mainly regulated at protein level, indicating its major role in post-transcriptional regulation (Lago *et al*., 2017). Translation of the *rpoS* mRNA is positively regulated by signaling alarmone (p)ppGpp and some small non-coding RNAs; DsrA under the conditions of osmotic shock, RprA under low temperature stress and ArcZ under anaerobic growth conditions, and is negatively regulated by OxyS (Repoila *et al*., 2003; Mika and Hengge 2014). The activities of these all four small RNAs are dependent on their binding to chaperone Hfq (Gottesman 2005), which either activates the *rpoS* translation by stabilizing them (DsrA, RprA and ArcZ) or inhibits it by enhancing hydrolysis (oxyS) (Moon and Gottesman 2011; Henderson *et al*. 2013). Under normal growth situations, the degradation of *rpoS* mRNA is accomplished by ATP-dependent protease ClpXP in a RssB (regulator of σS B) dependent fashion (Zhou and Gottesman 1998; Y. Zhou *et al*. 2001). RssB directly recognizes σ^S, whose activity for σ^S degradation depends on phosphorylation of its receiver domain (Becker *et al*., 2000; Klauck *et al*., 2001). However, the presence of three proteins IraP, IraM and IraD negatively affects degradation of *rpoS* mRNA. These proteins acts as RssB inhibitors in response to phosphate starvation (IraP), magnesium starvation (IraM) and DNA damage (IraD) (Bougdour *et al*., 2006; Bougdour *et al*. 2008; Battesti *et al*., 2010, Battesti *et al*., 2015; Park *et al*. 2017). During transition from log phase to the stationary phase, the sigma factor σ^S competes with housekeeping sigma factor σ^D (rpoD) for binding with core RNAP, which is dependent on interaction of σ^D with Rsd protein (regulator of sigma D) and its binding with small RNA molecule 6S that is present in abundance in stationary phase mimics an open promoter complex structure (Patikoglou *et al*. 2008; Hofmann *et*

al., 2011; Park *et al*. 2017). The rpoS regulation at transcriptional and translational level is shown in Figure 2.3.

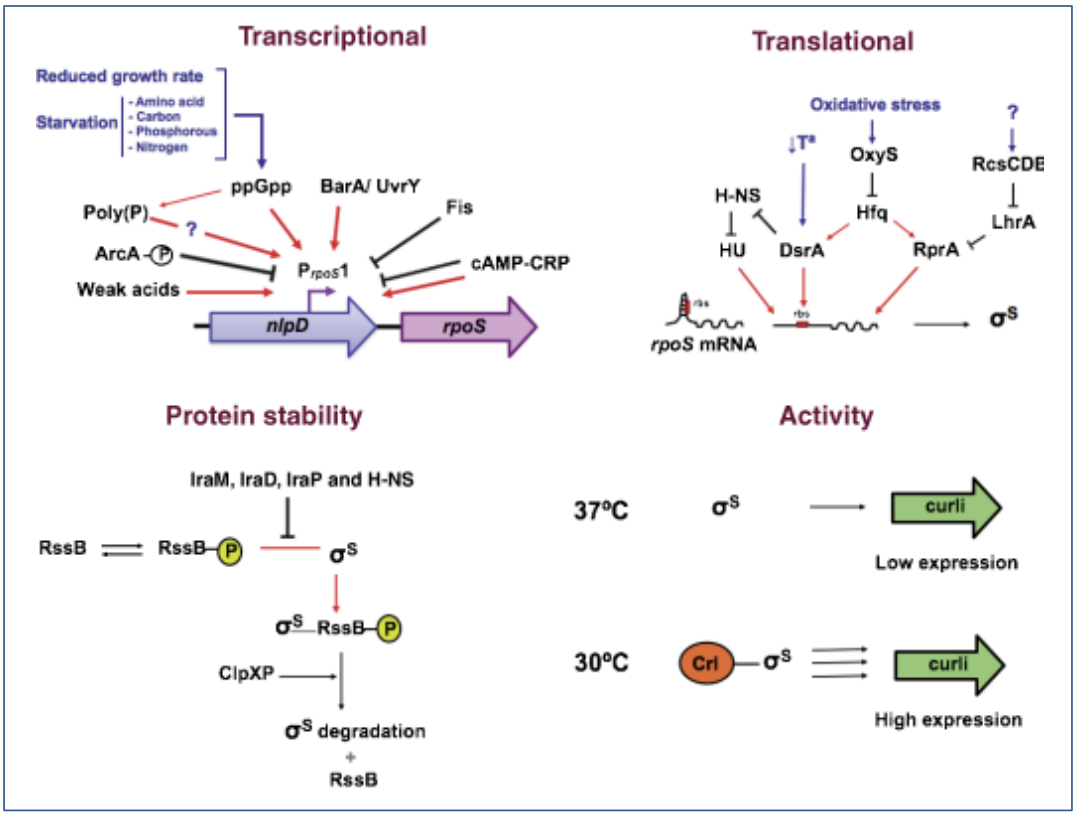

Figure 2.3 Pictorial representation of rpoS regulation in *E. coli.* (Llorens J M N *et al*., *FEMS Microbiol Rev* (2010) 34:476–495)

2.3.2.3 The Programmed Cell Death (PCD) mechanism

Programmed cell death (PCD) is a cell survival mechanism generally activated to eliminate superfluous cells from the population. Unlike other bacterial stress response mechanisms that generally operate at transcriptional level, the PCD mechanism operates at post-transcriptional or translational level and is sensitive to intensifying starvation (Moll and Engelberg-Kulka 2012). The PCD system utilizes unique 'addiction modules' or 'toxin-antitoxin (TA) modules' that comprises of stable toxic protein and its unstable antidote or antitoxin, which neutralizes the lethal activity of this toxic protein. Such genetic modules were initially discovered on *E. coli* low copy number plasmids due to

their tendency of exerting a post-segregational killing effect when bacterial cell loses a plasmid (Jensen and Gerdes 1995).

In *E. coli,* the TA module *mazEF* is one of the most well studied PCD mechanism containing two genes *mazE* and *mazF* both of which lie on the same operon downstream of *relA* gene (Metzger *et al*., 1988). The *mazF* gene codes for a stable cytotoxic protein and *mazE* encodes its labile antitoxin partner which inhibits the lethal action of this stable toxin. This antitoxin is degraded in vivo by ATP dependent ClpAP proteases. Under normal growth conditions, the expression of both genes results in inactivation of toxicity associated with *mazF* protein due to formation of *mazEF* complex. But the encounter of cell to stressful conditions like starvation, DNA damage, ppGpp accumulation etc. suppresses the transcription of this operon. These stringent conditions lead to a frequent degradation of labile antitoxin by ClpAP proteases thereby releasing the cytotoxic protein into the cells which inhibits the synthesis of majority of host cell proteins except those required for survival of very small population of viable cells, and thereby forcing the majority of population towards PCD by selective synthesis of cell death proteins (Engelberg-kulka *et al*., 2005) (Figure 2.4). Beside the role MazF play in degradation of host cell proteins, it also generates leaderless mRNAs by cleaving the ACA sites upstream of AUG start codon for some specific mRNAs. In parallel, MazF endoribonuclease cleaves 16S rRNA of the 30S ribosomal subunit, which results in loss of anti-Shine and Dalgarno (aSD) sequence required for initiating translation at canonical ribosome binding sites. This generates a special 'stress translational machinery (STM)' which translates the leaderless mRNA formed by MazF in a selective manner (Calogero *et al*., 2006; Vesper *et al*., 2011). Various stress conditions such as severe amino acid starvation, high temperatures, oxidative stress and DNA damage leads to the synthesis of this stress translational machinery which is beneficial from a physiological point of view as it help cells to tackle adverse conditions by enabling a quick response to them. The main advantage of this kind of stress response is that it does not require energy consuming steps of ribosome biogenesis and synthesis of ribosome bound factors that regulate translation process. Hence, this module plays both constructive and destructive roles in response to environmental changes.

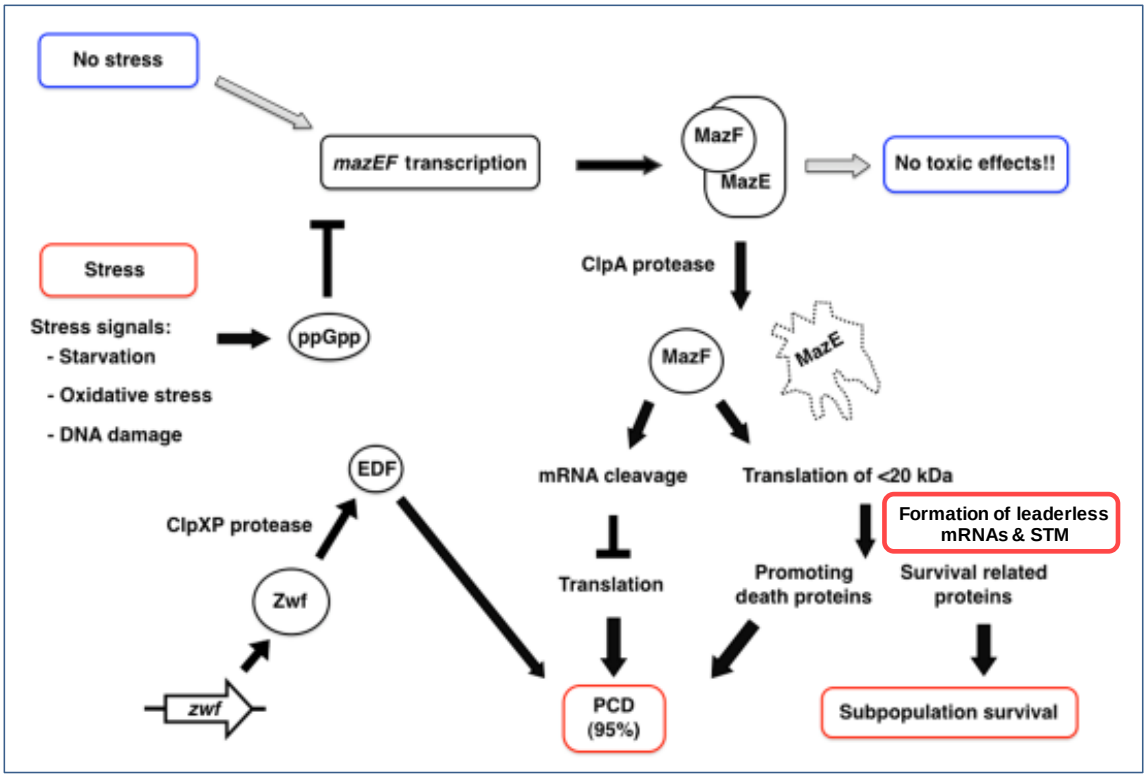

Figure 2.4 Schematic representation of *E. coli mazEF* mediated programmed cell death (PCD). Adapted from Llorens J M N *et al.*, *FEMS Microbiol Rev* (2010) 34:476–495, with slight modifications.

2.4 Recent advancements in the field of RPP

2.4.1 Removing transcriptional bottlenecks

The transcriptional strategies focused on improving recombinant protein production rates largely relies on the appropriate combination of host and vector system. A huge number of vectors are available for *E. coli* such as pET vectors based on strong T7 promoter system whose transcriptional efficiencies are known to be several folds higher than *E. coli* RNA polymerases (Ramos *et al.*, 2004; Choi *et al.*, 2018). These however can only be used for strains harboring lamda DE3 lysogen for expression of T7 RNA polymerase such as *E. coli BL21(DE3), Rosetta (DE3) strains* etc. In a recent study, the problem of leaky and non-tunable expression of T7 promoters has been resolved by using the bacterial hosts CD44(DE3) and CD45(DE3) which contains the stop codon in T7 RNA polymerase gene; thus induction under such conditions resulted in ten folds decline in its

transcript generate rate and hence slow build-up of target protein (Angius *et al*., 2018). The expression of membrane proteins under such conditions will minimize the problems associated with their toxicity and thus can lead to desired expression levels. The pBAD vectors based on $araP_{BAD}$ promoters also possess the versatile features for expression purposes like modulation of gene expression for a broad range of inducer concentrations and reduction of leaky expression (Guzman *et al*., 1995). The synthetic hybrid promoters *tac* combines the features of *lac* and *trp* (typtophan) promoter and are reported to be 10 times stronger than conventional lacUV5 promoter (de Boer and Comstock 1983). The commercial plasmids like pMAL series (NEB) are based on *tac* promoters. Plasmid instability is also another issue associated with heterologous gene expression and is the main cause for reduced protein productivity. Striedner *et al*., (2010) developed a plasmid free system for recombinant gene expression that is based on loci specific insertion of target gene into the genome of host cells, which led to only two fold increment in yield of recombinant protein in comparison to the plasmid based expression observed with GFP mutant version 3.1(*HMS174(DE3) strain)* and human superoxide dismutase (*BL21(DE3)* strain. Plasmid multimerization is the other reason for plasmid segregational instability. The attempts were made to resolve this issue by genetically modifying the *BL21(DE3)* strain to *recA* negative *BLR(DE3)* strain. *recA* mutation is associated with plasmid stability by reducing the frequency of multimerization and promoting plasmid monomer yields (Goffin and Dehottay 2017).

2.4.2 Removing translational bottlenecks

The translational efficiency is determined by two major factors: the starting codons at translation initiation site and the secondary structure of mRNA and ribosomal binding site. The presence of secondary structures in 5' UTR regions and at ribosomal binding sites might result in translation inhibition. The computational tool 'ExEnSo' (Expression Enhancer Software) solves this problem by allowing gene sequences to be designed on the basis of highest free energy which resolves the issue of translation inhibition. ExEnSo also creates a 5' primer based on these optimized gene sequences that can be used for amplifying the coding sequence by PCR (Care *et al*., 2008). Translation rate calculators are also available which allow users to estimate the strength of translation on the basis of

initial codon sequence of mRNA molecule and have shown to provide good predictive accuracy of actual protein expression (Reeve *et al*., 2014). The unavailability of charged tRNA species also slows down translation speed, which is indeed dependent on codon composition of the transcript. This problem has been solved by modifying the rare codons of the gene of interest to mimic the codon usage pattern of microbial host system (Burgess-Brown *et al*., 2008; Welch *et al*., 2009). But the previous tools for codon optimization such as ExpOptimizer (from NovoPro), GenSmart (from GenScript) have not taken into consideration the effect of changed coding sequence on mRNA secondary structure. A very recent study have applied the bicistronic design (BCD) to overcome the effect of codon usage on translation initiation and found 10-100 folds of increments in protein expression compared to the standard monocistronic design (MCD) (Nieuwkoop *et al*., 2019). Apart from this, the models like Ribosomal Flow Model (RFM) predicts the rates of translation based on its physical and dynamical nature (Reuveni *et al*., 2011). It considers the basic features of translation like ribosomal concentrations, translation rates, protein abundance levels and the interactions among these variables and provides more accurate predictions than contemporary approaches. Translational pause at a rare codon is another interesting feature that can be used for enhancing protein expression and folding process. The delay created by translational pause provides time for sequential folding of the defined portions of the nascent polypeptide originating from the ribosome (Komar 2009).

2.4.3 Improvements in protein folding

In recent years tremendous amount of efforts have been added to improve the protein folding process starting from design of genetically modified strains like Origami that provides oxidizing environment inside the cytoplasm due to *trxB gor* mutation (Bessette *et al*., 1999) to the co-expression of molecular chaperons of Hsp70 family (de Marco 2009; Nannenga and Baneyx 2011). Strains like Shuffle have been developed that contains chromosomal copy of *DsbC* gene in *trxB* and *gor* mutant and expresses DsbC protein that direct correct and stable disulphide formation in cytoplasm (Lobstein *et al*., 2012). The major problem encountered during protein folding is the high rates of generation of nascent polypeptide because of the use of high level expression systems,

which fail to match the availability of molecular chaperones inside cells and thereby results in protein misfolding. The most common solution to this problem is lowering of cultivation temperatures and use of low inducer concentrations along with weak and titratable promoters in order to slow down the rate of protein synthesis which provides sufficient time for protein folding. But the use of this approach also lowers the expression of molecular chaperons drastically that assists the folding of growing polypeptide chain. This problem has been overcome by design of *E. coli* strains like *Arctic Express* that co-expresses cold-adapted chaperonins *Cpn10* & *Cpn60* (from a psychrophilic bacterium) and retains their protein folding activity at low temperatures (4°-10°C) (Gopal and Kumar 2013). Many studies have addressed the issue of protein aggregation by co-expressing the target protein with folding attenuators such as protein-folding assisting chaperones (e.g., *DnaK, DnaJ, GroEL & GroES*) and disaggregating chaperones (*ClpB*) (Kolaj *et al*., 2009). The co-expression of groES/groEL chaperones has shown to improve the solubility of single chain antibody (scFv) against B-type natriuretic peptide to more than 64.9% of the total protein produced (Maeng *et al*., 2011).

The use of affinity tags has also been shown to enhance the solubility of recombinant protein. The pET vectors which are most extensively used for recombinant protein expression are also available with 6-histidine tags (Novagen) for protein purification using Ni-NTA chromatography, but these sometimes poses the problem of decreased solubility of fused protein. This problem can be solved by using affinity tags like maltose binding protein (MBP) and GST in pGEX or pMAL systems which enhances the solubility of a fused protein (Smith and Johnson 1988; di Guana *et al*., 1988; Hu *et al*., 2011) but have the associated problem of large size of these tags i.e. 44 kDa and 25 kDa. The tags available with small size of 1-11 kDa are His-tag, FLAG-tag, Strep-II tag, Trx. The larger size of fusion tags sometimes causes interference in the activity of recombinant protein, especially when tag is not removed properly. For efficient cleavage of these tags from the fusion protein expressed, different cleavage sites like Enterokinase, Thrombin, SUMO protease have been engineered into the expression vectors, which ensures their precise cleavage (Hefti *et al*., 2001; Lee *et al*., 2008; Shahravan *et al*., 2008).

2.4.4 Extracellular export of recombinant proteins

The export of a protein to the extracellular medium is a much more challenging task than expressing it in cytoplasm. The extracellular expression simplifies the downstream purification steps and also removes the constraint of upper limit of its accumulation inside the cell. Despite of the five protein secretory pathways (Type I, II, III, IV, V) present in *E. coli*, only Type I and Type II (sec or Tat pathway) are commonly used for recombinant protein expression (Choi and Lee 2004). Type I pathway is involved in the secretion of proteins to extracellular medium (Sugamata and Shiba 2005), whereas Type II pathway secretes proteins to the periplasmic chamber. Sec pathway which secretes unfolded proteins has been show to be overloaded in many studies resulting in unfolded protein accumulation (Pugsley 1993; Mergulhão *et al*., 2005; Su *et al*., 2013). The most frequently used Sec-dependent secretion pathway involves the use of various signal sequence such as *PelB, PhoA, OmpC, OmpF, OmpA, endoxylanase,* and *MalE*, where *pelB* is the most commercially exploited signal peptide in pET vectors. The properly folded proteins like GFP which are not secreted by sec pathway has been shown to be successfully get exported to the periplasm via tat pathway (Matos *et al*., 2012). Also, some proteins have been reported to be naturally gets secreted into the extracellular medium from cell's periplasm (Khushoo *et al*., 2004; Ni and Chen 2009). Many bioinformatics tools such as Phobius, Philius and SignalP have been designed that predicts whether a given amino acid sequence is eligible for sec signal peptide or not (Käll *et al*., 2004; Reynolds *et al*., 2008; Petersen *et al*., 2011). For prediction of Tat signal peptides, the available bioinformatics tools are TatP, TatFind and PRED-Tat (Dilks *et al*., 2003; Bendtsen *et al*., 2005; Bagos *et al*., 2010).

2.5 Systems metabolic engineering

Since early 90s, metabolic engineering strategies has been in use for manipulating and fine-tuning central carbon metabolic activities of microbial cell factories for enhancing recombinant product yield. Aristidou *et al*., (1995) engineered *E. coli* metabolically for efficient acetate utilization by heterologous expression of *alsS* gene (acetolactate synthase enzyme) from *Bacillus subtilis* which resulted in reduced accumulation of inhibitory by-product, acetate. This engineered strain grown in fed-batch mode of fermentation resulted

in significant enhancement of recombinant CadA/β- galactosidase fusion protein by marking increment of 60% in specific protein activity and 50% of volumetric productivity. Metabolic engineering strategies employed till date predominantly focused on reconstructing cellular metabolic network to overcome the negative effects of imposed metabolic burden during recombinant protein over-expression. This included alteration of carbon uptake and utilization machinery, optimization of flux going through the glycolytic, PP pathway and TCA cycle, over-expression of rate-limiting enzymes of central carbon metabolic pathway for increasing supply of biosynthetic precursors such as nucleotide, amino acid and ATP pool. For example, Flores *et al*. (2004) for fulfilling the increased demand of precursors molecules essential for both plasmid maintenance and recombinant protein expression, over-expressed *zwf* gene coding for foremost step of oxidative phase of PP pathway. Examples of successful engineered microbes by implementation of system-level approaches are involved in biofuel production (Colin *et al*., 2011), amino acids and derived chemicals (Wendisch 2014; Hirasawa and Shimizu 2016). However, now-a-days another powerful tool has emerged for generation of bioproducts is the synthetic minimal cell (Gibson *et al*. 2010; Zhang *et al*., 2010). By mere installing heterologous pathways in the existing microorganisms do not necessarily leads to higher production of desired product, for this host cell need to undergo rounds of metabolic engineering cycle. Therefore, reforming microbial host chassis involves the use of several strategies such as removing feedback control of pathways, increasing supply of precursor molecules & cofactors, blocking of competing pathways, manipulation of regulatory system and up gradation of the export machinery of the host cell (Mahalik *et al*., 2014). The two strategies employed for designing or reconstructing microbial platforms are rational or targeted engineering approach and inverse metabolic engineering approach.

2.5.1 Rational (or Targeted) Engineering

System level engineering using targeted approach leads to the generation of distinctive strain by making pinpoint genomic manipulation for achieving desired phenotype. This is a highly rational approach that is based on target prediction from existing genome scale computational models and implementing those modifications at genomic and molecular

level in the desired host. Development and advancement in transcriptomic analysis techniques from traditional DNA microarray to high throughput RNA sequencing (RNA-Seq) allowed us to analyze changes in mRNA abundance profiles as well as to detect and quantify novel transcripts of multiple samples simultaneously. Thus, information and knowledge obtained by comparing and analyzing transcriptomic data between different strains or strain cultured under distinct conditions or at different time points can be well exploited for identification of potential targets genes to be manipulated. Choi *et al*., (2003) have successfully implemented targeted engineering approach for improving the performance of microbial strain by analyzing transcriptomic profile of high cell density recombinant culture expressing human insulin-like growth factor I fusion protein (IGF-I_f). They demonstrated significant improvement in IGF-I_f protein productivity from 1.8 to 4.3 g/L by rational engineering of wild type cells through identification and supplementation of two key down-regulated genes *prsA* and *glpF* genes, encoding the phosphoribosyl pyrophosphate synthetase and glycerol transporter, respectively. Similarly, Singh and Mukherjee (2013) had significantly enhanced the expression and titer to 4.8 g/l of therapeutically important human interferon (rhIFN-β) protein by increasing the expression of down-regulated substrate utilization gene *glpK*, glycerol kinase through knocking in the genome under *ibpA* gene promoter. Likewise, Han *et al*., (2003) have rationally engineered *E. coli* strain by co-expression of *cysK* gene lead identified from proteomic profiling of culture expressing human leptin protein (serine rich protein) that not only enhanced production by four fold but also improved cell growth by two fold. Many of these above reported studies highlights the importance of directed approach for rational engineering of production host through identification of potential limiting metabolic pathway genes. However, engineering the production strain to next level requires better understanding of regulatory network that could be implemented in a targeted way for making alteration at regulatory level. Indeed, in collaboration with genome scale metabolic model, genome scale bacterial transcriptional regulatory network model is also paving their way out to capture the impact of gene expression regulation by global as well as local regulators of host cell (Faria *et al*., 2014).

2.5.2 Inverse Metabolic Engineering

Unlike rational engineering, the inverse engineering approach does not rely on pre-determined targets for genomic manipulation of cellular metabolism and regulatory network of host cell in order to achieve desired phenotypes. To deal with the complexity of cellular and regulatory networks, more global approach taking into picture both related and unrelated pathways is needed. The classical metabolic pathway building strategies for metabolite overproduction are comparatively straightforward, therefore rate-limiting steps, feedback regulation and by-product should be taken care of. However, in case of the heterologous protein expression which is inexplicably linked to the growth process; identification of genes that need to be knocked-out/knocked-in to get improved phenotype becomes an elusive task. Ghosh *et al*., (2012) have adopted an alternative inverse metabolic engineering strategy for screening those over-producers that are slow-growing but are metabolically active. *E. coli* genomic library was prepared and induced for the anti-sense RNA generation in cells where genomic fragment was cloned in reverse orientation for partially down-regulation of genes. Out of the 17 transcripts for slow growth phenotype identified through high throughput screening, two of the leads i.e. *ribB* gene (3, 4 dihydroxy-2- butanone-4-phosphate synthase) and *kdpD* gene (histidine kinase) increased specific product yields to 7 fold and 3.2 fold respectively. This phenotype would allow the cellular resources & metabolic flux to get directed more towards the recombinant product formation in spite of building biomass of the host cell. Similarly, Gialama *et al*., (2017) employed a reverse engineering strategy for identifying target gene that enhanced membrane protein expression by genome-wide screening using ASKA library, an ordered library of plasmids coding all known *E. coli* ORF under the control of the *T5lac* promoter. ASKA library (consisted of 4000 ORF plasmids) was transformed along with target membrane protein tagged with GFP for high-throughput screening by FACS for over-producers. After screening 4000 clones, two genes *djlA or rraA*, one belonging to the class of chaperone and other is the regulator of mRNA degrading enzyme RNasE improved the expression significantly, even exceeded the protein production level of commercially available strain such as *C41(DE3) and C43(DE3)* for membrane protein production.

2.6 Systems biology approaches

Systems biology analysis, powered by various '-omics' technologies allows the fast assessment of physiological behaviour of the cell to identify key engineering targets that would lead to designing of more robust and efficient microbial cell factories. To acquire a complete picture of the cell with respect to its regulatory mechanisms, the use of single x-ome data is not enough. This is because the level of mRNA, proteins and metabolite fluxes vary independently in a highly coordinated manner. Therefore, an integrated analysis becomes an absolute necessity, where global cellular information collected from various -omics data are then combined with genetic-relationship-matrix structural equation model (GSEM) to perform simulation studies for generation of predictive computational models of the microbial system**.** The physiological state of the recombinant protein expressing cell could be ascertained by analyzing the expression ratio of all coding and non-coding transcripts for all known 4600 *E. coli* genes. Indeed, with the recent advancement in high throughput genome sequencing and systems biology technologies along with the computational simulations it is now possible to generate large data seta for re-engineering & re-construction the existing microbial platforms in a more coordinated & comprehensive way. Initially, the metabolic stress and cellular health of host expressing recombinant protein was delineated by estimating the physiological parameters such as biomass profile or acetate build-up profile.

Although, these parameters are good indicators of metabolic status of the cell, they cannot predict the magnitude of metabolic perturbations occurring in the entire biochemical network of the cell. However, the advent in -omics techniques such as transcriptomics, proteomics and metabolomics has allowed the precise quantification of all cellular components encompassing the entire biochemical network of cell. Metabolic networks are highly interconnected and interwoven; even the slight perturbation at protein and metabolites level affects the entire cellular biochemical network, thus making these changes prediction difficult. Therefore, a system-level analysis is needed to monitor changes taking place at all level simultaneously and then integrating the information for locating the appropriate targets for rewiring the entire metabolic network of the cell. This analysis of high-throughput data obtained using latest "omics" technique provide system-

wide invaluable information about the phenotypic and metabolic characteristic of the cell under different environmental conditions for making alterations at large scale to get desired product in high concentrations.

2.6.1 Genomics

The ultimate goal of systems biology approach is to built models and programs using knowledge from large set of experimental high throughput omics data for predicting changes in cellular behavior or phenotype of host cell in a particular condition. Indeed, the large repertoire of omics datasets of microbial cell factories under various environmental conditions and the existing knowledge about biochemical reactions and network has allowed the construction of genome scale metabolic network models. These models have proven to be a powerful tool for predicting targets to metabolically engineer host cells subsequently for over-production of desired product and improved our understanding of complex metabolic regulatory network of the cell. Integration of multiple omics data with network based modeling approaches holds very promising strategy for unraveling the key regulatory elements of a metabolic & regulatory network resulting in particular phenotype of host cell under given condition (Rai and Saito 2016). However, to further enhance the prediction capabilities of this constraint based metabolic models; it need to be integrated with transcriptional regulatory information. FBA is the first constraint based models assuming steady state condition of the cell for computing the flux distribution rate and biomass production rate (Varma and Palsson 1994; Lewis *et al*., 2012). Nocon *et al*. (2014) have shown improvement in the expression of hSOD (human copper/zinc superoxide dismutase) in *Pichia Pastoris* by predicting gene knock-outs and over-expression targets of central carbon metabolism genes using Minimization of Metabolic Adjustment (MOMA) and Flux Scanning based on Enforced Objective Function (FSEOF) prediction tools respectively. The metabolism of an organism is the most well characterized biological process till date. However, to analyze the dynamic response of cellular metabolic perturbations there is a need to integrate genome scale metabolic network models with the information of regulatory as well as signaling pathway controlling the entire cellular machinery. Dynamic modeling is a propitious approach for linking the cellular processes in a single modeling framework. Such

networks could then be well connected to the existing genome scale metabolic models, and they could be utilized to quantitatively predict the cell dynamic behavior from extracellular perturbations to metabolic responses.

2.6.2 Transcriptomics

Out of all the omics techniques, transcriptomics is the most frequently used technique for characterizing changes in gene expression pattern of cells induced for recombinant protein over-expression that allows analyzing the cellular and physiological response generated towards the forced protein production. Transcriptomics provide the cellular snapshot of dynamic changes about the state of mRNA levels under different environmental conditions. It also offers the opportunity for identifying novel regulatory aspects of the host cell metabolism which is crucial for designing efficient microbial cell for metabolite or protein production. There are several methods for performing transcriptomics analysis starting from early EST, SAGE/CAGE, and microarray to latest RNA Seq technique. Most of the transcriptomics studies of *E. coli* recombinant cultures growing at high cell density and expressing recombinant proteins revealed that the characteristics of cellular stress response are analogous to that of heat sock, stringent and oxidative response. This similarity in stress response emanated from protein over production is reflected by the up-regulation of heat shock response genes such as *ftsH, clpP, lon, ompT, degP, groEL, aceA and ibpA*, oxidative stress response gene such as *soxS, soxR, oxyR*. By global analysis of whole cellular machinery gene expression profiles, it was well noted that recombinant protein expression had drastic effects on the host cell metabolism as well as replication and translational machinery of the host cell. Oh and Liao (2000) reported the molecular events that take place after induction of host cell for recombinant protein expression of LuxA soluble and non-toxic protein by microarray analysis. Genes whose expression got most affected and were repressed belongs to the biosynthetic pathways of amino acids (*lysC, leuA, metE, serA, and ilvC),* purine *(purF)* and fatty acid *(fabA),* central carbon metabolic genes of glycolysis, TCA cycle and PP pathways and the most important energy generating pathway genes. Similarly, Ow *et al*., (2006) reported the same trend of down-regulation of energy metabolism and biosynthetic genes by analyzing global transcriptome of plasmid bearing

E. coli DH5α cells in batch mode of fermentation. These observations implicates that recombinant protein over-expression imposes strong metabolic load and alters the metabolic infrastructure of the host cell to large extent which thus leads to severe decline in rates of both biomass as well as product formation. Jamnikar *et al*., (2015) with the application of microarray transcriptomic technique, identified the potential target genes that help CHO cell lines for stable production of recombinant protein by analyzing the transcriptomic profile of stable and unstable protein production clones. Additionally, Yu *et al*. (2017) have established and developed the methylotrophic yeast *Komagataella phaffii* (*Pichia pastoris*) as successful microbial cell factories for the various recombinant protein production by identification of three most up-regulated heat shock response genes (*CPR6*, *FES1*, and *STI1*). These target genes identified from transcriptomic analysis of recombinant strains expressing single copy gene for Phospholipase protein (PLA_2) to the strain expressing its 12 copies were then co-expressed, which resulted in increased production of phospholipase and prolyl endopeptidases upto 1.41 fold.

2.6.3 Proteomics

Majority of the studies has primarily and effectively used transcriptomics technique for understanding the regulation behind the metabolic and physiological changes that take place in response to changing environmental conditions. However, the need for integration with proteomic approach arises due to number of reasons: (i) non-linearity between transcriptomic and proteomic analysis (ii) apart from transcriptional regulatory control, post-translational regulatory mechanism for protein modifications and degradation also exist, which cannot be analyzed and visualized at the transcriptomic level (Farrell *et al*., 2014; Kim *et al*., 2016). Therefore, studying the changes at transcriptome level is not enough for complete understanding or prediction of the regulatory mechanism of the cell. The commonly used methods for proteomic profiling of recombinant cultures are gel based approaches such as two-dimensional gel electrophoresis (2DE) & two-dimensional difference in-gel electrophoresis (2D-DIGE) and non-gel based approaches such as reversed-phase liquid chromatography coupled with tandem mass spectrometry (LC−MS/MS). For quantitative proteomics, the approaches such as stable isotope labelling by/with amino acids (SILAC), isotope-coded

affinity tag (ICAT) labeling and isobaric tag for relative and absolute quantitation (iTRAQ) are in current use (Aslam *et al*., 2017). Label-free protein quantification using LC-MS has also gained tremendous interest in past decade due to simplicity and low cost associated with sample preparation steps than label based techniques (Levin *et al*., 2007; Clough *et al*., 2009; Cox *et al*., 2014). Yoon *et al*., (2012) did the combined transcriptomic and proteomic profiling of *E. coli* cells in order to envisage the metabolic and physiological changes when cells were grown at high cell density culture in fed-batch fermentation. This combined -omics study data analysis provides us the snapshots of the global changes at transcriptomic and proteomic level and allowed us to design better fermentation strategies and locating targets for efficient metabolic engineering of host cell for overcoming the limitation of HCDC. Similarly, Han *et al*., (2003) examined the proteomic profile of recombinant culture over-expressing human leptin protein & enhanced its productivity by *cysK* gene co-expression coding for cysteine synthase A enzyme. Proteomic analysis revealed the down-regulation of pathway genes responsible for synthesis of serine amino acid thus limiting the production of serine rich human leptin recombinant protein. Thus, transcriptomics and proteomics are systematic approaches that allow simultaneous and fast identification of more than hundreds of genes/proteins responsive to different types of stress response in up and down-regulated manner (Haddadin and Harcum 2005; Weber *et al*., 2005; Franchini *et al*., 2015). However, most critical step is the selection of appropriate target gene for manipulation of cellular metabolism and regulatory network of microbial cell factories rationally for over-production of our desired product.

2.6.4 Metabolomics

Metabolomics is the next level up approach for complementing the data obtained from transcriptomics and proteomics approaches and used for metabolites quantification that are the end-product of cellular processes and regulation. Intracellular metabolic fluxes are quantified using substrates that are isotopically labeled such as ^{2}H, ^{13}C, ^{14}C, ^{15}N to determine the change in flux rate by calculating the amount of isotopic label integrated into the subsequent down-stream products. Thus, metabolomics by allowing the quantification of the relative levels of intracellular metabolite concentration over time

reflects the cell's physiological status of the cell. With the development of advanced high-throughput quantitative techniques such as HPLC, NMR, liquid & gas chromatography-mass spectrometry (LC-MS & GC-MS) coupled with time-of-flight mass analyzers have made the metabolic analysis much more sophisticated. The information from various -omics approaches can be well used for unraveling the key regulatory nodes of cellular metabolism by estimating the abundances of RNA, protein and metabolites. The composition of intracellular metabolites is perturbed when cell is induced for production of recombinant protein by redirecting the normal metabolic pathway flux to meet the additional demands for precursors. Therefore, by comparing and analyzing the metabolic status of cells expressing recombinant protein up to maximal levels with the strain expressing either nil or very low recombinant protein would give us insights for engineering the host cell to achieve that altered metabolic state responsible for maximizing specific cellular productivity. However, most of the studies have analyzed the effect on metabolic burden on physiological parameters such as growth rate, carbon utilization rate, by-products formation while there are very few studies depicting the changes in intracellular metabolite concentration due to the effect of recombinant protein over-production. Recently, Chae, *et al*., (2017) using two-dimensional NMR spectroscopy have tried to map the metabolites level changes in relation to the recombinant protein production, the metabolite composition that favors the production of recombinant proteins to a higher level. Similarly, Muhamadali *et al*. (2016) used both metabolic footprinting & profiling approach for investigating the altered biochemical composition of *BL21(DE3)* strain while over-expressing recombinant GFP protein in comparison to the *BL21(IL3)* strain containing orthogonal riboswitch for controlling T7 RNA polymerase gene expression at the transcriptional & translational level. Metabolomics & fluxomics analysis provide insights into hierarchical metabolic regulation of central carbon metabolic reactions by capturing the actual metabolite levels and determining the flux rate of reacting leading to that metabolite.

2.7 The model protein: L-asparaginase

The secretory expression of recombinant proteins to the extracellular medium has many advantages over cytoplasmic localization such as (i) simplified downstream processing

steps (ii) no limitation of space for recombinant protein accumulation (iii) presence of only few contaminating proteins as *E. coli* secretes only few proteins to culture medium and (iv) protease deficient environment of extracellular compartment. The proteins secreted outside are mostly soluble and retains their biological activity (Mergulhão *et al*., 2005). The passage of protein through periplasmic space provides it a more favourable redox potential which ensures the proper protein folding, specially the cases that require disulphide bond formation. The recombinant proteins are either produced in soluble form or as inclusion bodies (IBs) which are insoluble protein aggregates. Some reports have shown the leakage of soluble proteins from periplasm to the extracellular medium (Cornelis 2000; Shokri *et al.*, 2003). Although the mechanism behind such secretion is not completely known, it is believed that it involves selective passage of protein through the outer membrane (Shokri *et al.*, 2003). There are several other factors that influence the secretion efficiency of host cells such as protein size & its nature, signal peptide amino acid composition, co-expression of facilitator proteins, the capacity of export machinery, the regulatory system associated with protein secretion system and genetic properties of the host strain. Recombinant L-asparaginase secretion to the extracellular medium has been shown previously in our lab by optimizing the media composition and induction strategy, which resulted in enhanced efficiency of secretion of recombinant protein containing pelB leader sequence in biologically active form (Khushoo *et al*. 2004).

L-asparaginase-II (E.C. 3.5.1.1) also known as L-asparagine amidohydrolase is encoded by ansB gene in *E. coli* and is the most extensively studied enzyme because of its use as a theraupeutic agent in acute lymphoid malignancy (Bonthron 1990). It is a periplasmic enzyme synthesized as a tetramer of approximately 140 kDa composed of four identical subunits (Epp *et al*., 1971) and catalyzes the deamination of L-asparagine into aspartate and ammonia (Bonthron 1990). This homo-tetramer is composed of intimate dimmers, but the active form of protein is always tetrameric (Swain *et al*., 1993). Due to its unique anti-cancerous activity, the multidrug chemotherapy utilizing L-asparaginase has resulted in complete remission of acute lymphoblastic leukemia (ALL) in 90% of patients (Piatkowska-Jakubas *et al*. 2008). Other than ALL, its utility has also been shown in Hodgkin's disease, chronic lymphocytic leukemia, acute myelocytic leukemia, melanosarcoma, lymphosarcoma and reticulosarcoma (Stecher *et al*. 1999).

3. Materials and Methods

3.1 Chemicals and reagents

Media and bulk chemicals were purchased from local manufacturers, Himedia, Sigma, Qualigens, and Merck. Media used was LB (Luria Broth – Casein enzymatic hydrolysate 10 g/L, yeast extract 5 g/L and NaCl 5 g/L, final pH (at 25°C) 7.0 ± 0.2) and TB (Terrifc Broth – yeast extract 24 g, tryptone 12 g/L, potassium dihydrogen phosphate 2.2 g/L, dibasic potassium phosphate 9.4 g/L, final pH (at 25°C) 7.0 ± 0.2). IPTG from Ameresco, USA, Arabinose from HiMedia & antibiotics such as Ampicillin and Kanamycin were from HiMedia, India. Restriction and modifying enzymes were purchased from NEB. Reverse transcriptase was purchased from Thermo Fischer Scientific Pvt. Ltd, India. All other chemicals were of analytical grade and obtained from local manufacturers.

3.2 Bacterial strains and plasmids

Table 3.1 Strains used in the study

***E. coli* strains**	**Genotype/ Characteristics**	**Source/ Reference**
DH5α	(*supE44_lac*U169 (*_80 lacZ _M15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*)	Amersham Biosciences, USA
BW25113	F$^-$,Δ*(araD-araB)567, ΔlacZ4787(::rrnB-3),* λ^-*, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Yale University and the Department of Molecular, Cellular & Developmental Biology
W3110	F$^-$ lambda$^-$ IN(*rrnD-rrnE*)1 rph-1	Dr. David Summers, Cambridge,UK

Table 3.2 Plasmids used in the study

Plasmid	**Source/ Reference**	**Description**
sfGFP-pBAD	Dr. Devpriya Choudary, SBT, JNU	Size: 5429 bp, AmpR , *sfGFP* gene under arabinose inducible promoter, N-terminal 6x His tag, f1 ori
pMAL-p2x-S1-Asp	Lab/ Shubhashree Mahalik, Ph.D thesis 2014	Size: 6645 bp, AmpR , *ans*B gene under P_{tac} promoter (IPTG inducible), pBR322 ori.
pPROLar.A122	Prof. Hermann Bujard, Heidelberg	kanR, lac/arabinose promoter PRO expression vector, size 2.6 kb

Table 3.3 List of Primers used in our study and for clone confirmation

Name	**Sequence (5' to 3')**
glpK Fw	CCAAGCTTATGACTGAAAAAAAATATATCG
glpK Rv	GCTCTAGATTATTCGTCGTGTTCTTC
glpF Fw	CGGGATCCATGAGTCAAACATCAACC
glpF Rv	GCTCTAGATTACAGCGAAGCTTTTTG
glpD Fw	GCAAGCTTATGGAAACCAAAGATCTGATTG
glpD Rv	AATCTAGATTACGACGCCAGCGATAACC
1_FW pri glpD_vec_overhang	TGGAGATGACGATGACAAGGTGGTCGACAAGCTT ATGGAAACCAAAGATCTGATTGTGATAGG
2_RV pri glpD_overhang	CGCTAATCTTATGGATAAAAATGCTATGCTCGATT ACGACGCCAGCGATAACCTCT
3_FW pri AraProm_overhang	TATACGCAGCAGAGGTTATCGCTGGCGTCGTAATC GAGCATAGCATTTTTATCCATAAGATTAG
4_RV pri AraProm_overhang	TGGTCGAGCGCAACGATATATTTTTTTTCAGTCATG GGTACCTTTCTCCTCTTTAATGAATTCT
5_FW pri glpK_overhang	TCACACAGAATTCATTAAAGAGGAGAAAGGTACC CATGACTGAAAAAAAATATATCGTTGCGCTC

6_RV pri glpK_overhang	CCGCATCGATCGGGCCCTGAGGCCTGCAGGGATCC TTATTCGTCGTGTTCTTCCCACGC
7_pPRO_Fw_glpK overhang	AACGCGCGATGGCGTGGGAAGAACACGACGAATA AGGATCCCTGCAGGCCTCAGGGCC
8_pPRO_Rv_glpD overhang	CGCCCCCTATCACAATCAGATCTTTGGTTTCCATAA GCTTGTCGACCACCTTGTCATCGTCATCTC

3.3 Preparation of competent cells

The *E. coli* competent cells were prepared with a slight modification in the standard protocol (Maniatis, Fritsch, & Sambrook, 1982). A glycerol stock of *E. coli* cells was streaked on LB agar plate in a laminar hood. A single colony was picked and inoculated in 5 ml of LB broth and incubated overnight at 37°C with shaking. After 16 hours, 500 µl of the culture was used as inoculum (1% final concentration) for 50 ml LB broth in a 500 ml flask. Cells were grown to an $O.D._{600}$ of 0.3-0.4. The cells were chilled on ice and then transferred to a pre-chilled sterile oakridge tube under aseptic conditions. The cells were centrifuged at 4000 rpm for 10 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 10 ml of chilled 100 mM $CaCl_2$ and incubated on ice for 30 minutes. The cells were centrifuged again at 4000 rpm for 10 minutes at 4°C. The pellet was resuspended in 2 ml of 100 mM $CaCl_2$ and 50% glycerol was added to it to a final concentration of 15%. The cells were kept on ice for ~16 hours and were finally stored at –80°C as 200 µl aliquots. Next day one aliquot was used for transformation with 10ng DNA of a standard plasmid in order to check the efficiency of the competent cells as number of transformants per µg of supercoiled plasmid DNA.

3.4 Transformation of *E. coli*

A 200 µl aliquot of competent cells was thawed on ice. Plasmid DNA (20 ng) was added to the thawed cells and incubated on ice for 30 minutes. Cells were given heat shock by keeping them in a water bath which was set at 42°C for 90 seconds. The cells were immediately transferred onto ice and 800 µl of autoclaved LB broth was added. The cells were kept on an incubator shaker set at 37°C, for 1 hour. Out of 1 ml culture, 100 µl of cells were plated on a LB agar plate containing an appropriate antibiotic for selection of

transformants. The LB plate was incubated at 37°C for 12-16 hours in order to obtain the transformed colonies.

3.5 Small-scale plasmid DNA isolation

Mini preparation was done to obtain plasmid DNA in reasonable amounts for purification, restriction digestion and ligation. For this a total of 5 ml of the overnight grown culture in LB media was pellet in microcentrifuge tubes at 12,000 x 'g' for 2 minutes. The supernatant was discarded and 250 µl of solution I (50 mM glucose, 25 mM Tris HCl pH 8.0, 10 mM EDTA pH 8.0) was added to the cell pellet and resuspended by vigorous vortexing. 400 µl of freshly prepared solution II (0.2 N NaOH, 1% SDS) was added and mixed gently by inverting the contents of the tube. Finally 350 µl of ice-cold solution III (5 M potassium acetate, glacial acetic acid) was added, mixed well and stored on ice for 10 minutes. The microcentrifuge tube was centrifuged at 12,000 x 'g' for 10 minutes at 4°C and the supernatant was transferred to a fresh tube. 0.6 volumes of isopropanol was added at room temperature and mixed well. The mixture was centrifuged again at 12,000 x 'g' for 30 minutes at room temperature. The pellet was washed twice with 400 µl of 70% ethanol and dried in a dry bath. The dried pellet was finally suspended in 50 µl of TE buffer and treated with RNaseA.

3.6 DNA Quantification

The concentration of DNA was determined by spectrophotometery in the UV range. Optical density (OD) of DNA solution was measured at 260 nm and 280 nm with appropriate blank (the solvent in which DNA was dissolved) using quartz cuvettes. An OD at 260 nm = 1 was considered equivalent to a concentrations of 50 µg/ml for double stranded DNA and 20 µg/ml for single stranded oligonucleotides. The ratio of $O.D._{260}/O.D._{280}$ was checked to determine the purity of the DNA preparation. The ratio for a protein free pure DNA sample should be 1.8 to 2.0.

3.7 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out in a horizontal gel apparatus with 1X TAE as electrophoresis buffer. As per requirement, 1.0 to 1.5% agarose was dissolved in 1X TAE

buffer by heating it in a boiling water bath. After allowing it to cool to 40-45°C, ethidium bromide was added to it to a final concentration of 1 µg/ml. The gel was set by pouring it into a casting tray in which a comb of desired tooth size was inserted at one end to form wells. After the gel had set, the comb was removed and the gel was transferred to the gel tank filled with the electrophoresis buffer. Samples mixed with loading buffer were loaded into the wells. The gel was run at a constant voltage of 4-5 V/cm. The DNA bands were visualized under 260 nm UV light on a trans-illuminator.

3.8 Polymerase chain reaction (PCR) amplification

The PCR was set up in 20 µl volume approximately in a chamber that was used for PCR entirely, to avoid contamination. Genomic DNA (50 ng) or plasmid DNA (10 ng) was used with ~20 picomoles of each primer. dNTPs and $MgCl_2$ were used at the final concentration of 200 µM each and 1.5 mM respectively. Commercially available stock of reaction buffer 10X was added such that final concentration was 1X in the reaction. 1 unit of Taq DNA polymerase, or phusion polymerase enzyme was used depending on the need of experiment. The total volume was made up with sterile MQ water. The programme of the thermal cycler was set accordingly with changing annealing temperature and number of cycles to run for the particular reaction. In the presence of Taq DNA polymerase the denaturation temperature (94°C) and extension temperature (72°C) were same for all the programmes and in the presence of phusion polymerase enzyme the denaturation temperature was set at 98°C.

3.9 DNA extraction from agarose gel

DNA was extracted from agarose gel using Agarose Gel DNA Extraction Kit (Qiagen, USA). Nucleic acids bind specifically to the surface of glass or silica materials in the presence of a chaotropic salt (sodium perchlorate). The binding reaction occurs due to the disruption of the organized structure of water molecules and the interaction with the nucleic acids. Thus the adsorption to the specifically pre-treated spherical silica matrix is favored. Since the binding process is specific for nucleic acids, the bound material can be separated and purified from impurities e.g. salts and proteins, by a simple washing step. Nucleic acids elute from the matrix in a low salt buffer or water.

3.9.1 Ligation of DNA fragments

The DNA fragments digested with restriction enzyme(s) were mixed with the vector digested with the same restriction enzyme(s) or vector having compatible ends. Usually 30 to 50 ng of vector DNA, 3 to 6 fold molar excess of insert and 0.5 units of T4 DNA ligase were used in ligation set up. The reaction was set up in final volume of 10 µl and the reaction mix was incubated at 22°C for 16 hours. The reaction mixture was heat inactivated at 65°C for 20 minutes prior to transformation of ligase mix in high efficiency competent cells.

3.9.2 DNA sequencing

The sequences of the DNA fragments cloned were confirmed by sequencing of the constructs by automated sequencing at the commercial DNA sequencing facility.

3.10 Growth and expression studies in TB media

Selected clones were grown overnight with shaking at 37°C either in 5 ml TB media in 50 ml tubes or in 10 ml TB media in 100 ml flask. Secondary inoculation was done by adding 500 µL of overnight grown culture in 50 ml of TB medium in 500 ml flask. After 1.5- 2.0 hours when the $O.D._{600}$ of 1.5-2.0 was attained, cultures were induced by adding 1 mM IPTG or 0.2% L-arabinose respectively (as per experiment requirement). After induction, the $O.D._{600}$ was monitored at 600 nm at regular intervals and samples were pelleted and stored at -20^0C for further analysis.

3.11 Genomic DNA isolation from *E. coli* cells

E. coli culture was inoculated from glycerol stock in 50 ml and grown overnight at 37°C with constant shaking at 200 rpm. Overnight grown culture was pelleted down by centrifugation at 5000 rpm for 10 minutes at 4°C. The pellet was resuspended in 4 ml TE buffer pH 8 (10 mM Tris HCl, 1 mM EDTA) to which 0.5 ml of 10% SDS was added. 30 µl of proteinase K (20 mg/ml) was added to the resuspended culture which was incubated at 37°C for 1 hour. After complete lysis of cells, 1 ml of 5 M NaCl was added and mixed gently. 750 µl CTAB NaCl mixture was added to the lysate and incubated at 65°C for 20 minutes. Later equal volume of chloroform: isoamyl alcohol mix (approximately 7.5ml)

was added and mixed gently. It was centrifuged at 12,000 rpm at 4°C for 30 minutes. To the aqueous phase containing genomic DNA, 12.5 µl of RNase (2 mg/ml) was added. The supernatant was incubated at 37°C for 1.5 hours; equal volume of phenol: chloroform: isoamyl alcohol mixture was added and mixed properly. The tubes were centrifuged at 12,000 rpm, for 30 minutes at 4°C. The supernatant was again extracted with equal volume of phenol: chloroform: isoamyl alcohol mixture. The supernatant (aqueous phase) was collected in Korex tube and 0.6 volume of isopropanol was added and mixed properly. Korex tube was centrifuged at 10,000 rpm for 20 minutes at 4°C. The pellet was washed with 70% ethanol and kept for drying in room temperature. Finally the pellet was dissolved in 0.5 ml autoclaved water and run on 0.8% agarose gel.

3.12 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Electrophoresis (SDS-PAGE) Polyacrylamide gel electrophoresis under denaturing condition (in the presence of 0.1% SDS) was performed according to the method described by Laemmli (1970). The stacking gel containing 4% acrylamide, 0.106% N,N'-methylene bisacrylamide, 0.1% SDS and 0.125 M Tris-HCl (pH 6.8) was mixed and polymerized. The separating gel had 12% acrylamide depending on the size of the proteins being separated and 0.1% SDS. Running buffer consisted of 0.025 M Tris-base, 0.192 M glycine, pH 8.3 containing 0.1% SDS. The protein samples were prepared in sample buffer (0.0625 M Tris-HCl, pH 6.8; 2% SDS, 10% glycerol and 5% β-mercaptoethanol) and immersed in a boiling water bath for 3-5 minutes. Standard protein molecular weight markers (MBI, Fermentas) were run simultaneously to calculate the molecular size of the proteins.

3.12.1 Preparation of sample for loading on SDS-PAGE

Since L-Asparaginase was secreted to extracellular media, samples from each time point were pelleted down (1 ml culture) at 13,000 rpm for 5 minutes. The supernatant was filtered through 0.2µm filters and mixed with 6X SDS loading dye in 4:1 ratio followed by boiling for 5 minutes. The solution was spun down for 8 minutes at 10,000 rpm to remove any residual debris. Equal volume of supernatant (10-20µl) for each sample was used for loading on SDS-PAGE.

3.12.3 Coomassie Brilliant Blue staining

SDS-Polyacrylamide gels containing more than 2µg protein concentration were visualised by standard Coomassie Brilliant Blue (CBB) staining (0.1% (w/v) CBB dissolved in 25% (v/v) methanol and 10% (v/v) acetic acid in water), followed by destaining in 25% (v/v) methanol and 10% (v/v) acetic acid in water.

3.13 Enzymatic assay and volumetric productivity calculation for L-asparaginase

Asparaginase activity was measured in terms of rate of hydrolysis of L-asparagine by measuring the amount of ammonia released in the reaction. Sample supernatant and purified enzyme samples were mixed with 10mM L-asparagine dissolved in 50mM Tris–HCl, pH 8.6. The enzyme substrate mixtures were incubated at 37°C for 10 min, after which the reaction was stopped by addition of 100 µl of 1.5M TCA. Samples were centrifuged and then used for estimation of ammonia. The amount of ammonia released was determined by Nessler's reagent using ammonium sulphate solution as standard. An international unit (UI) of L-asparaginase is defined as the amount of enzyme required to release one micromole of ammonia per minute under the conditions of the assay at saturating substrate concentration (Wriston 1985). Volumetric productivity (mg/ml) was calculated by dividing enzymatic activity (U/ml) by specific productivity of L-asparaginase.

3.14 Total RNA isolation from *E. coli* cells for transcriptomic profiling

For RNA isolation, samples (1 ml culture volume equivalent to 2 $O.D._{600}$) were collected at specified time intervals from *E. coli* BW25113 and BW*Δ(elaA+ΔcysW)* cells under protein producing and non-producing conditions for the batch runs (as per experimental requirement). To stop the mRNA degradation, 100µL of ice cold EtOH/Phenol stop solution (5% water saturated phenol (pH<7.0) in ethanol) was immediately added to the 1ml culture. The culture was spun down at 8000 rpm for 2 min at 4ºC. The pellet was stored at -80 ºC after removing the media until RNA extraction. Total RNA isolation was done from cell pellet using Trizol method. The stored pellet was resuspended in 80µL of 0.5mg/ml lysozyme, TE pH 8.0 to lyse the cells. 8µL of 10% SDS was added, mix and

kept at 64ºC for 1-2 minutes, till the solution became clear. After incubation 8.8µL of 1M sodium acetate (pH 5.2) was added and mixed. Equal volume of Trizol reagent was added and the culture tubes were inverted 10 times followed by incubation at 64ºC for 6 min. The tubes were placed in ice to chill for 5 minutes and then spun at 14000 rpm for 10 min at 4ºC. The resulting aqueous layer was then transferred to the fresh 1.5ml microfuge tube. 3M sodium acetate equivalent to one-tenth of aqueous layer volume, 1mM EDTA and 2-2.5 volumes of cold 100% ethanol was added to the microfuge tube. The mixture was incubated at -80ºC for 20 minutes and then centrifuged at 14000 rpm for 25 min at 4ºC. A small white pellet containing RNA became visible. The ethanol was carefully removed without losing the pellet. The pellet was washed with 100µL 80% cold ethanol and again centrifuged at 14000 rpm for 25 min at 4ºC. The ethanol was removed and the pellet was air dried for 15-20 minutes. The pellet was resuspended in 22µL RNase-free DEPC treated H_2O, 1µL RNase Inhibitor (RiboLock), 6µL of 5X DNase-I buffer and 1µL of RNase-free DNase-I. The reaction mix was incubated at 37ºC for 30 min for DNase I treatment. This was followed by addition of the equal volume of water saturated phenol. The tubes were inverted 8-10 times and centrifuged at room temperature (RT) for 2-3 min at 12000 rpm. Equal volume of phenol/chloroform (1:1) was added and mixed followed by spin at 12000 rpm for 2-3 min at RT. The aqueous layer was transferred to the fresh microfuge tube and equal volume of chloroform was added and mixed. The resulting mixture was centrifuged at 12000 rpm for 2-3 min at RT. To the aqueous layer one-tenth volume of 3M sodium acetate and two volumes of ice-cold ethanol were added. The mix was incubated at -80ºC for 20 minutes and then centrifuged at 14000 rpm for 25 min at 4ºC. The ethanol was removed carefully and the pellet was dried for 15-20 minutes. The pellet containing RNA was resuspended in 30µL RNase-free DEPC treated H_2O.

3.15 Real Time PCR

For real time PCR, the DNase I treated RNA samples (equal amount) were reverse transcribed into cDNA using Thermo Scientific RevertAid Reverse Transcriptase and random hexamer (Thermo Fisher Scientific, India) as a primer. Primer3 software was used for designing gene-specific primers for the cDNA amplification and qRT-PCR was carried out in 7500 Fast Real-time PCR System (Applied Biosystem, Foster, USA). The

reaction mix contained 5 pmol primers (Table 3.4), 10µl Fast Universal SYBR Green (2x) Master Mix (Kapa Biosystems), 0.2 µl of 20 µM primer mix, 0.4 µl of ROX Reference Dye Low (50x), 1 µl of a diluted 1:5 template cDNA and nuclease free water to make final volume 20µl. The "Quantitaive PCR" method was used to determine the relative quantities of the mRNA expression level. The ribosomal-protein-L12-serine-acetyltransferase (rimL) was used as the endogenous control. A validation curve using serial dilutions of cDNA was used to ensure that the replication efficiencies of the tested genes. After initial denaturation for 10 min at 95°C, the amplification cycle (repeated 40 times) was as follows: 15 s at 95°C and 1 min at 60°C. Samples were assayed in triplicate and PCR reactions without the template were performed as negative controls. The relative mRNA levels were expressed against the reference sample of each run which was 4h in case of non-producing cultures and 0h (induction point) in case of protein producing cultures. The comparative Ct method ($2^{-\Delta\Delta CT}$) (also known as Delta-Delta Ct method) was used to calculate the changes in gene expression levels as relative log2 fold difference between reference gene and test genes (Livak and Schmittgen 2001).

Table 3.4: List of primers used for RT-PCR

Primer name	**Sequence (5' to 3')**
rimL RTF	CGGTGCAGGGTAATGTGATG
rimL RTR	ATGAGATTCGTCCAGCCAGT
fis RTF	CGTACTGACCGTTTCTACCG
fis RTR	CCACGGGTGTATTGCATCAC
lrp RTF	TGGCAAAGATCTCGACCGTA
lrp RTR	GGTTAAGCAGCGCCGTATAG
dps RTF	GCGCTAACTTCATTGCCGTA
dps RTR	CCTGAACGTTGTGGATGTCC
hns_Fw	TGCTGCTGAAGTTGAAGAGC
hns_Rv	CCAGGTTTTAGTTTCGCCGT
ihfB_Fw	CAGGGCGAGCGTATTGAAAT

ihfB_Rv	GATCGCGCAGTTCTTTACCA
cbpA_Fw	TTCATATTGCGCCACATCCG
cbpA_Rv	TTTGCTCACCAGACCTTTGC
rpoS_Fw	GTCGCGCACTGCGTGGAGAT
rpoS_Rv	CGCGGATCAGCCCCAGGTTG
rpoH_Fw	TCTGGAAGCAGCTAAAACGC
rpoH_Rv	GCAACTTTGACGATACGCCA
rpoE_Fw	ATTCGTTCCGGGGAGATAGC
rpoE_Rv	CCATGCGTAAATCTTCCGGG
RelA_Fw	GAAGATGTGCTGCGTGAGAG
RelA_Rv	CCGCCAGTTTGATGACTACG
SpoT_Fw	CTGACACCTGTTATCGCGTG
SpoT_Rv	GGTTTCGCCGTGCTCTTTAT
mazE_Fw	GGCTACGTTAATGCAGGCG
mazE_Rv	GGCTCTCCCCAGTCGATATT
mazF_Fw	CGACAAAAGGTAGCGAGCAA
mazF_Rv	CTGGGGCAACTGTTCCTTTC
ompA_Fw	GGTGCATACAAAGCTCAGGG
ompA_Rv	GTGATCGCGTACTCAACACC
rybB_Fw	GCCACTGCTTTTCTTTGATGTC
rybB_Rv	ACAAAAAACCCATCAACCTTGAAC
oxyS_Fw	GAAACGGAGCGGCACCTC
oxyS_Rv	GGAGATCCGCAAAAGTTCACG
rprA_Fw	TTATAAGCATGGAAATCCCCTGA
rprA_Rv	AAAAAAAGCCCATCGTGGGAG

3.16 Sample preparation for proteome analysis

Protein was extracted by employing a cell lysis protocol for *E. coli* described by EMBL, Heidelberg, Germany with slight modifications. Briefly, the cell pellets kept at -80ºC for proteomic studies were washed three times in 10 ml ice-cold 1X phosphate buffered saline (PBS, pH 7.4) and then resuspended in chilled lysis buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM EDTA and 1 mM PMSF. The ratio of cell wet weight to buffer volume was kept 1:4. The suspension was kept on ice for 10 min and 10µl of 1X Protease Arrest (Comapny) per ml of culture was added. The suspension was sonicated for 10 min with 10 short burst of 10 sec followed by intervals of 30 sec for cooling. The cell debris was removed by ultracentrifugation at 4°C for 30 min at 45000 rpm and the supernatants were recovered. Protein concentration was determined by using BSA as standard and samples were either diluted or concentrated to the final concentration of 1 µg/µl in 100mM ammonium bicarbonate buffer. Total 20µg of protein was taken for tryptic digestion. Tryptic digestion was performed based on the textbook protocol (Kinter and Sherman 2005) with slight modifications. The tryptic digests were desalted and concentrated using ZipTip (Pierce C18 Tips, Thermo) by following the manufacturer's protocol.

3.17 Label free LC MS/MS protein quantification

Samples prepared for proteomic analysis were given to the CIF facility of NII, New Delhi in lyophilized form for label free LC-MS/MS procedure. The digested peptides were dissolved in solvent A (5% acetonitrile containing 0.1% formic acid) and loaded for reverse phase chromatography using C-18 New objective Picofrit analytical column in a Thermo-scientific Proxeon Nano LC. Samples were run at flow rate of 300 nl/min using linear gradient of solvent B (95% acetonitrile containing 0.1% formic acid); 70 mins in 5-40% solvent B, 10 mins in 40-80% solvent B, 10 mins in 80% solvent B, 5 mins in 80-5% solvent B followed by 25 mins in 5% solvent B. Mass spectrometry was performed in Orbitrap Velos mass spectrometer and data was analyzed using Thermo Proteome Discoverer software (1.3.0.339 DBV version). Spectra of peptides were queried against Uniprot knowledgebase for *E. coli K-12* lineage containing decoy database using a target false discovery rate of 1% for strict and 5% for relaxed condition. The algorithm used for

protein quantification was based on measurement of peptide peak area changes in chromatography and spectral counting of proteins identified after MS/MS analysis. Data were acquired under dynamic exclusion mode and the initial mass tolerance was kept 20 ppm. MaxQuant software was used for performing data normalization and analysis. Further analysis was done by sorting out data based on peptide spectral matches (PSMs) which denotes the total number of identified peptide sequences for a particular protein.

3.18 HPLC monitoring of glycerol

Glycerol was monitored by Agilent 1260 Infinity HPLC system having an RID detector using 300 X 7.8 mm Aminex HPX-87H column (Biorad Laboratories). The mobile phase used was 5 mM H_2SO_4 at a flow rate of 0.5 ml/min. The column temperature was maintained at 50°C during the run. 20 µl of supernatant extracted from the samples was injected for every run.

3.19 Bioreactor studies

3.19.1 Microbioreactor studies for online GFP monitoring

Micro fermentation was done in a BioLector (m2p-labs GmbH, Aachen, Germany) (Kensy *et al*. 2009), which is meant for high-throughput micro fermentations and allow online monitoring of bioprocess parameters (biomass, GFP fluorescence, pH, DO and NADH levels) in 48 well format microtiter plate. We used the FlowerPlate format of the microtiter plate for our studies which ensures maximum oxygen transfer rate (OTR). Temperature was set at 37°C and humidity at 85%. The maximum volumetric capacity of each microflower plate well is 1500µl, but we worked with 900µl of culture volume because of the higher OTR at this volume with a constant shaking of 1400 rpm. The pH, DO and biomass concentration (arbitrary units) was measured online. The biomass concentrations were measured via scattered light at 620 nm excitation with an emission filter at Gain 10. The online GFP monitoring was done using an excitation filter of 485 nm and an emission filter of 520 nm at gain 20.

3.19.2 Batch studies

Batch cultivation using *E. coli* control and modified cells with and without recombinant

plasmid pMAL-p2X-S1Asp (see Appendix Figure A2) was done to check cell viability and collect samples for transcriptomic and proteomic analysis. A standard 2 litre bench top reactor (Electrolab Biotech Limited, UK) was set up with a 1 litre working volume. Batch media was TB (with 0.4 % glycerol and 10mM $MgSO_4$). The media was aseptically inoculated with 100 ml of overnight grown secondary culture. Ampicillin was used at a concentration of 100µg/ml (1X) as per requirement of the run. The temperature, pH and the dissolved oxygen (DO) was set at 37^{o}C, 7.0 and 100% respectively with an initial stirrer speed of 250 rpm. The airflow was set at 1 vvm and was not changed during the fermentor run. The pH was controlled automatically with the dosing pumps using 1N NaOH and 1N HCl. Automatic DO control was set at 40% until the agitation rate reached 1000 rpm. Excess foaming was controlled by addition of antifoam.

4. Results and Discussion

The innovative design of genetically engineered host cells have led to a significant increase in bioprocess yields at industrial scale with minimized production costs. Although these genetic manipulations has resulted in accumulation of recombinant proteins sometimes up to 40~50% of the total cellular protein (TCP), at the same time they lead to changes that either alters or deteriorates cellular physiology and limits productivity. It is well known that cell growth and expression of foreign protein are related biological processes that compete for utilization of various intracellular resources like amino acids, nucleotides, charged tRNAs etc. which serve as key energy molecules in driving various biosynthetic reactions in recombinant cells. Therefore, an in-depth understanding of the cellular response which triggers changes at the molecular level is crucial forthe design of superior host platforms.

Previous attempts of analyzing the CSR in our lab has aided in our understanding of host cell physiology in response to foreign protein induction. The idea of designing host platforms for superior expression by simply creating double knock-outs has brought us one step forward and helped in the discovery of signaling genes, which were previously considered to be non-essential, as important players of the CSR in *E. coli*. Results obtained from previous study showed that the double knock-out combination **Δ(*elaA*+*cysW*)** improved L-asp production by 2.3 folds compared to unmodified host strain *E. coli* BW25113 (Sharma, 2015). These *elaA* and *cysW* genes are functionally non-defined and their involvement in the protein production pathway has not been reported elsewhere. Although various information databases are available for *E. coli*, there still exist many genes that are predicted to have some regulatory function but have not been completely annotated. Ghatak *et al.* (2019) assembled information from *E. coli* databases Ecocyc, UniProt, EcoGene and RegulonDB and concluded that 34.6% of the *E. coli* genes are not supported by experimental evidence of function with mechanisms that lead to changed phenotypes and coined the term 'y-ome' for them. This gap in information made us curious about how the simple deletion of one or two genes of unknown function could enhance protein expression by several folds. Comparative transcriptomic studies of 6 hour post induction cultures of control and modified host

BW25113*Δ(elaA+cysW)* revealed the significant up-regulations of genes of central carbon metabolism, global regulator *lrp* and *dps* genes (which are related to the transcriptional and translational machinery of the cell), increased levels of starvation sigma factor *rpoS* and unchanged levels of house-keeping sigma factor *rpoD*, which could be the reasons behind the improved performance of these knock out strains (Jain 2018). Other than these positive changes, these genomic modifications also resulted in up-regulation of *rmf* and *yqjD* genes that are known to inhibit the translational machinery (Yamagishi *et al*. 1993; Yoshida *et al*. 2012) and a severe down-regulation of substrate uptake genes (*glpABC*, *glpF*, *glpK*, *glpD*, *glpT*, *glpQ*) compared to the recombinant protein expressing control cultures (Figure 4.1). This indicated that the modified cells were not able to fully recover from all the deleterious effects of the CSR. We also realized that the information gained from the above study was not sufficient to fill the gap in our understanding of how the cell actually triggers the CSR.

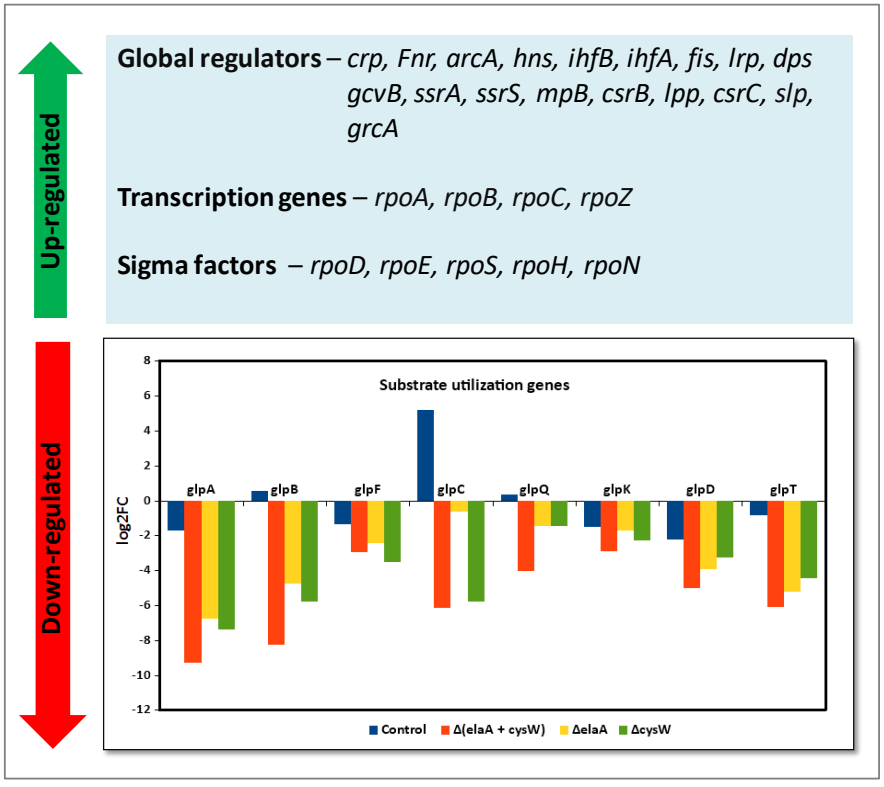

Figure 4.1 Transcriptomic analysis of cellular stress response in 6 hour post induction cultures of modified host BW25113Δ*(elaA+cysW)* and control expressing L-asp.

This study led to two key questions: (a) whether production capabilities of these modified hosts can be further improved by augmenting the substrate uptake capabilities? (b) Would these modifications allow us to elucidate the signaling pathways that lead to the generation of CSR? To answer these questions we proposed an improved strategy which is slightly different from the strategy proposed previously to counter the negative effect of the CSR on recombinant protein expression (Figure 4.2).

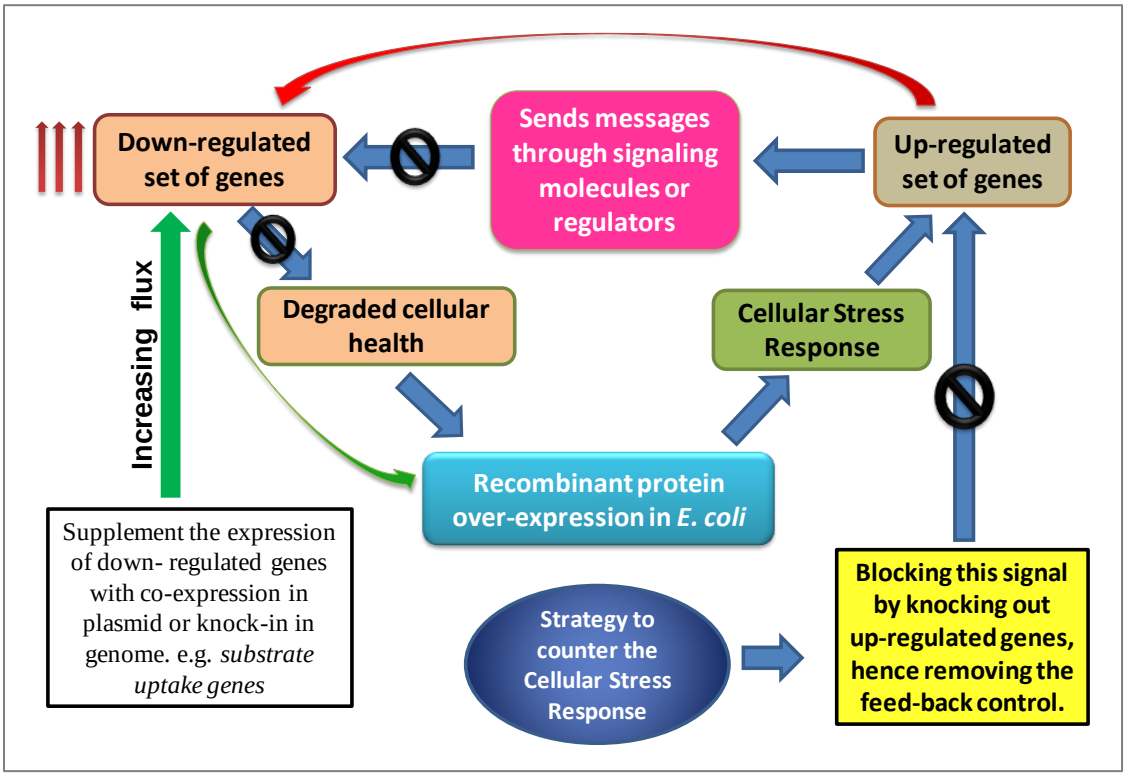

Figure 4.2 Improved strategy to ameliorate the negative effect of the CSR on recombinant protein expression.

Our strategy is based on our understanding of the CSR which postulates that there are multiple regulatory signals which sends messages to the cell forcing it to down-regulate a critical set of genes associated with cellular health and protein production. When some of these messengers are blocked, it partially ameliorates the negative effects of the CSR and some critical pathways are no longer down-regulated. However, other up-regulated genes are still able to communicate a partial signal via some unknown mechanism and these are capable of down regulating other critical genes like those associated with substrate

uptake. This degrades cellular health, and even though its extent is not as severe as in the wild type strain it still limits protein production. Therefore, our strategy is to knock-out the up-regulated genes while simultaneously supplementing the flux of down-regulated genes with plasmid based coexpression or by knock-in in the host genome. This two pronged approach would be a more effective strategy and comprehensively counter the harmful effects of the CSR.

We therefore, started our work with further modifications on the previously constructed knock-outs. Recombinant L-asparaginase (**L-asp**) was chosen as a model protein in this study because its extracellular localization ensures that there is no problem of protein accumulation inside the cell which could set an upper limit to the level of production (Khushoo *et al*. 2004).

4.1 Growth and substrate utilization profiles of the modified host producing L-asp

Since transcriptomic studies of the modified hosts showed down-regulation of genes for glycerol uptake, we decided to monitor its glycerol consumption rates in comparison to control cultures. For this we conducted shake flask studies with L-asp producing *E. coli* BW25113*Δ(elaA+cysW)* (modified) and unmodified *E. coli* BW25113 (control) cells. A single colony of freshly transformed control and modified cells was used to separately inoculate 100 ml flasks containing 10 ml of TB medium with 100µg/ml of amphicillin and grown overnight at 37°C with constant shaking at 200 rpm. 500µL of this primary culture was inoculated into secondary flasks containing 50 ml of TB media supplemented with 0.4% v/v glycerol and 10 mM $MgSO_4$. After the OD_{600} reached 1.5~2.0 the cultures were induced with 1mM IPTG and samples were collected till 24 hours post-induction. One control flask containing an un-induced culture of BW25113 strain was run in parallel to check the glycerol uptake capability in the absence of recombinant protein induced cellular stress. The growth and substrate profiles are shown in Figure 4.3. The control flask which was left un-induced for L-asp synthesis showed a normal growth pattern and the cells reached a maximum OD of 26 after 12 hours.

Both the induced flasks showed slower growth compared to uninduced cells, with the modified hosts showing a significantly higher decline in growth in comparison to the unmodified cells (Figure 4.3). This indicated that these genomic modifications had a direct impact on the growth rate of the cells. As expected the residual substrate profiles also showed a slower consumption of glycerol for the L-asp producing modified host cells when compared to the unmodified cultures. Thus while the normal L-asp producing control cells were able to consume glycerol within 10 hours post induction, significant amounts of glycerol were left unconsumed in the modified cells even after 14 hours post-induction. These results confirmed the phenotypic effect of the down-regulation of glycerol uptake genes found in transcriptomic analysis of this modified strain.

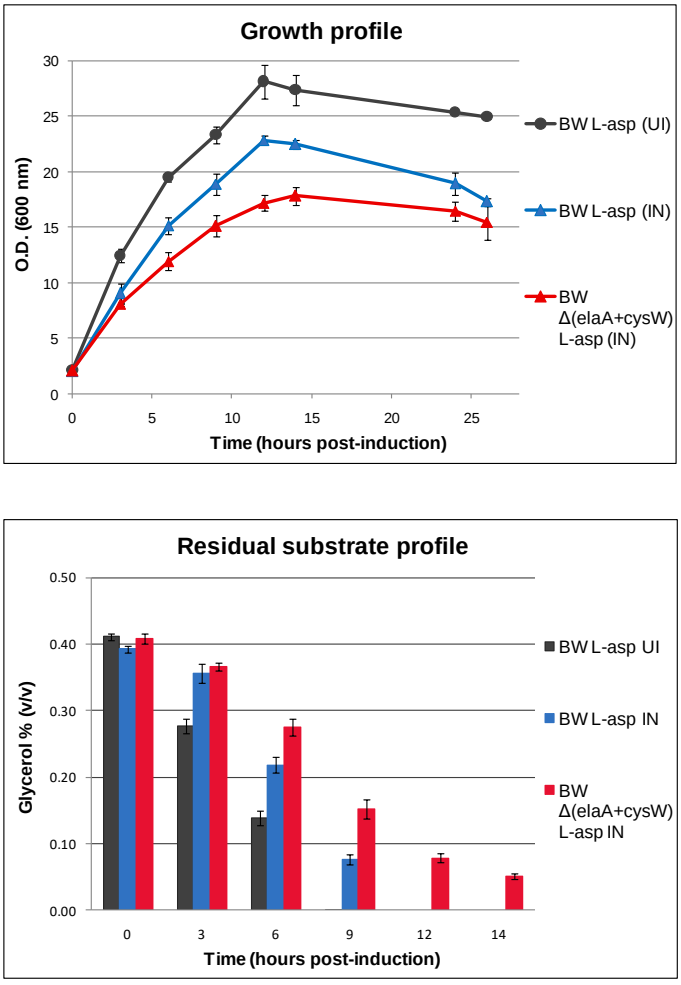

Figure 4.3 Growth and substrate utilization profiles of control and modified host cells.

4.2 Cloning and expression studies of glycerol uptake genes in modified host

It is well known that cellular growth rate and protein production share common pathways, and therefore the generation of a CSR impacts on both these processes. It would be logical to view this reduced growth rate as a defense mechanism mounted by the cell in response to the diversion of metabolic fluxes towards foreign protein synthesis. One of the mechanisms by which this growth retardation takes place post induction is by the reduction of substrate uptake and respiration capacity of the cells (Oh and Liao 2000; Neubauer, Lin, and Mathiszik 2003). We chose glycerol as a carbon source for our studies since it eliminates the problem of carbon catabolite repression and reduces acetate accumulation in comparison to cultures growing on glucose (Wang *et al*. 2015). Studies on the respiratory metabolism of *E. coli* has revealed that three global regulators namely cAMP-CRP, *Cra* and *ArcA* play an important role in promoting balanced aerobic growth on glycerol, where cAMP-CRP acts as main inducer of the glycerol catabolic regulon (*glpF, glpK and glpD* genes), *Cra* regulates genes of gluconeogenesis pathway and *ArcA* serves as a regulator of central metabolic genes (Iuchi *et al*., 1990; Weissenborn *et al*;, 1992). *E. coli* has a glycerol catabolic regulon for both aerobic and anaerobic conditions. The aerobic module contains *glpF, glpK* and *glpD* genes that encodes for a glycerol facilitator, glycerol kinase and glycerol-3-phosphate dehydrogenase *(glpD*), while the anaerobic module contains *glpABC* genes (Schryvers and Weiner 1982; Iuchi *et al*., 1990). *glpF* gene is an integral membrane protein that catalyzes the facilitated diffusion of glycerol across the inner membrane (Sweet *et al*. 1990), where *glpK* phosphorylates it to sn-glycerol-3-phosphate (G3P) using ATP as a phosphoryl donor. G3P acts as an inducer of the *glpD* gene (Beijer*et al* 1993), which is an inner membrane enzyme which catalyzes the oxidation of G3P into dihydroxyacetone phosphate (DHAP) with a simultaneous reduction of (FAD) to $FADH_2$ that passes electrons to the electron transport chain (Yeh *et al*., 2008). Figure 4.4 shows a schematic representation of the respiratory metabolism of glycerol in *E. coli* under aerobic conditions. Up-regulation of the substrate utilization gene *glpK* has been shown to increase the expressions levels of recombinant IFNβ by 35% in high cell density fed batch cultures (Singh and Mukherjee 2013).

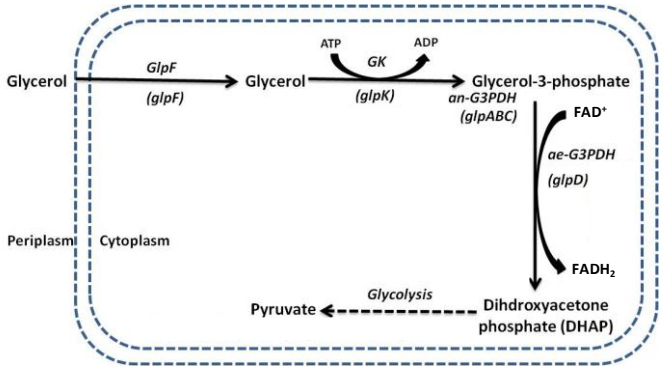

Figure 4.4 Schematic representation of glycerol metabolism under aerobic conditions in *E. coli.*

Our transcriptomic studies had also shown down regulation of glycerol uptake genes in modified hosts (shown in previous chapter). Since recombinant protein production at industrial scale is carried out under aerobic conditions, we focused on *glpD* that belongs to the aerobic pathway, along with *glpK* and *glpF* genes.

The plasmid we chose for co-expression studies was pPROLAR.A122 (Clontech, Mountain View, CA), which is known to be a highly inducible, lac/arabinose-regulated expression system (Link, Phillips and Church 1997). It contains the hybrid $P_{lac/ara-1}$ promoter, a ribosome binding site (RBS), a multiple cloning site (MCS) and a kanamycin resistance marker. This plasmid system is highly repressible, allowing fora greater degree of regulation than the P_{lac} promoter and is inducible with L-arabinose. Since the *ansB* gene which produces L-asp was on a pMAL-p2X vector that contains the P_{tac} promoter, a pMB1 ori from pBR322 and ampicillin resistance marker, these two vectors were compatible and suitable for co-expression studies.

4.2.1 Cloning of glycerol uptake genes: *glpF, glpD* and *glpK*

All the three genes *glpF*, *glpD* and *glpK* were amplified by PCR from the genome of *E. coli* BW25113 strain using forward and reverse primers which also contained the appropriate restriction sites as listed in Table 3.3. The PCR was done using Phusion polymerase (protocol described in Material and Methods section) which possesses 5´ to 3´ polymerase activity and 3′ to 5′ exonuclease proof reading activity. It is a high fidelity

DNA polymerase and generates blunt-ended products. The PCR steps resulted in amplified fragments of 1506 bp for *glpD*, 1509 bp for *glpK* and 846 bp for the *glpF* gene. These amplified gene products were subjected to restriction digestion using Hind III/BamH1 and XbaI enzymes and the digested fragments were eluted from 1% agarose gel. The eluted *glpD* and *glpK* fragments were ligated to the pPROLAR.A122 vector which was also restriction digested with same enzymes while the *glpF* gene was ligated under BamH1 and XbaI restriction sites (Figure 4.5(a)). Clone confirmation was done by restriction digestion where a fall out corresponding to the size of the three genes was seen respectively in all three cases (Figure 4.5(b)).

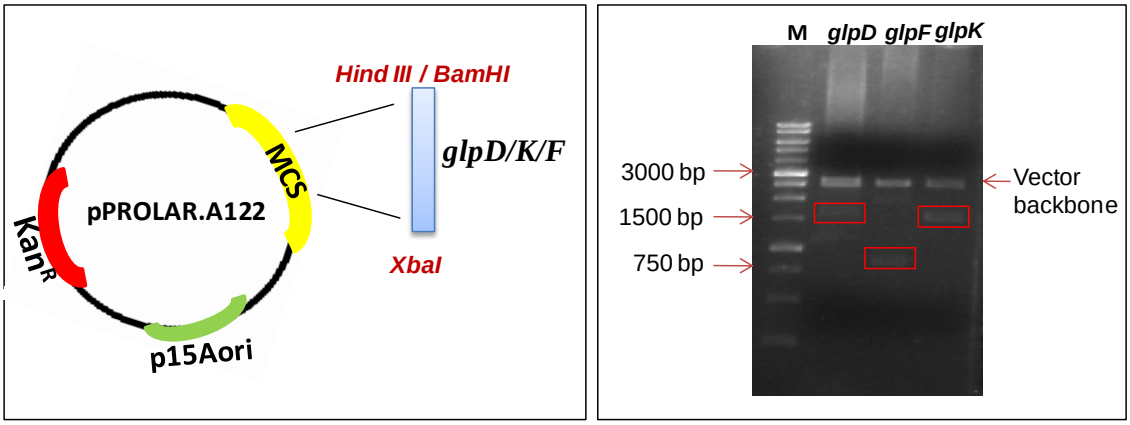

Figure 4.5 (a) Schematic representation of cloning of *glpD*, *glpK* and *glpF* genes in pPROLAR.A122 vector. (b) Gel picture showing clone confirmation by double digestion where Lane M is 1kb DNA molecular weight ladder.

4.2.2 Co-expression studies of glycerol uptake genes with L-asp

E. coli utilizes glycerol as a carbon source by the combined action of glycerol facilitator (*glpF*), glycerol kinase (*glpK*) and sn-glycerol-3-phosphate dehydrogenase (*glpD*) genes. We co-transformed our modified cells with plasmid containing L-asp and pPROLAR.A122 vector containing either the *glpD*, *glpK* or *glpF* gene. Modified cells containing only the L-asp gene was kept as control for this study. A single colony from the LB agar plates was inoculated in 50 ml flask containing 5 ml TB medium with 50µg/ml of ampicillin and 25µg/ml of kanamycin with the control culture containing only ampicillin. The cultures were grown overnight at 37ºC with constant shaking at 200 rpm. The secondary flasks of 500 ml volume containing 50 ml of TB media supplemented with

0.4% v/v glycerol, 10mM $MgSO_4$ and respective antibiotics were inoculated with 500µl of these overnight grown primary cultures. When the OD_{600} reached ~ 1.0-1.2, the cultures were induced with 1mM IPTG and 0.2% L-arabinose to co-express both genes for substrate uptake and L-asp production. The control flask was induced with IPTG only. Samples were collected for 24 hours post induction. The residual substrate profiles were monitored by subjecting the supernatant to HPLC analysis.

The co-expression of *glpD* and *glpK* genes started showing improvement in glycerol uptake from the medium within 4 hours post induction compared to control cultures (Figure 4.6). The complete consumption of glycerol was seen within 12 hours post induction for modified cells co-expressing the *glpD* and *glpK* genes. However the co-expression of *glpF* gene did not show any improvement in substrate consumption rates which was in fact slightly hampered due to this co-expression. It has been previously reported that *glpK* is the key enzyme of the *glpFK* operon that plays a major role in glycerol utilization (Zwaig *et al.*, 1970). A probable explanation is that the *glpF* gene cannot work in isolation and hence without supplementation of the other *glp* operon genes (which remained down-regulated in our modified cells) it is not able to counter the problem of down-regulation of substrate uptake rates.

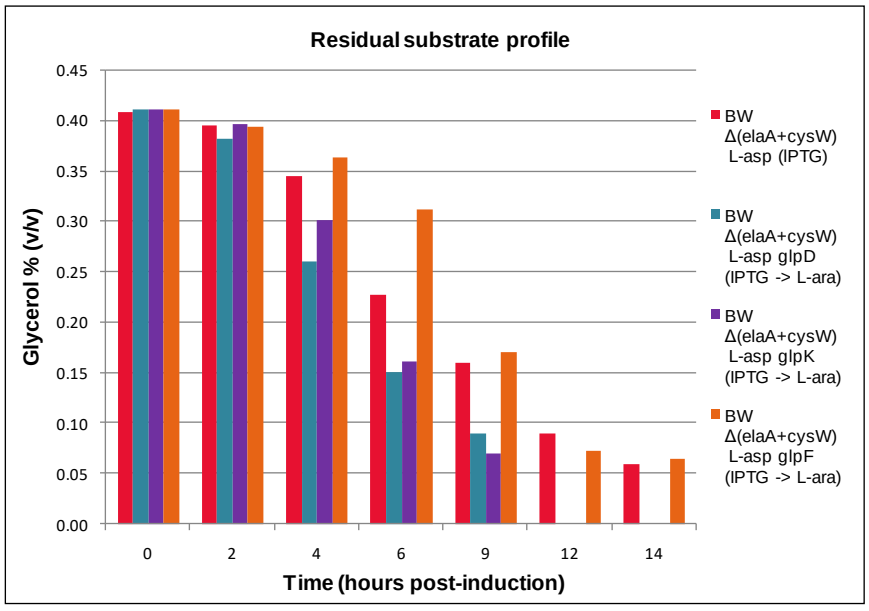

Figure 4.6 Residual substrate uptake profiles of modified cells upon co-expression of *glpD, glpK* and *glpF* genes.

4.2.3 Construction of expression vector *glpDK*

Since we found improvement in substrate uptake capabilities upon co-expression of *glpK* and *glpD* genes, we decided to check the synergistic effect of co-expression of both these genes in our modified host cells. Availability of Duet vectors (Novagen) that allow cloning of two genes in the same plasmid would have been ideal for this task. But these vectors contain MCS regions under the T7 promoter system, which require *E. coli* BL21 (DE3) host backgrounds for expression of proteins. Therefore, Duet vectors were not suitable for our study. Also, we wished to have medium level expression of these substrate uptake genes so as to not overload the protein synthesis machinery of the cells. Therefore, we decided to clone both *glpD* and *glpK* genes in the same plasmid system pPROLAR.A122 by introducing a tandem $P_{lac/ara1}$ promoter region into the MCS of this vector. The clone was constructed using the principle of homologous recombination demonstrated by Jacobus *et al* (2015) in *E. coli.* For this, the vector was linearized by PCR using primers that contained 30-35 bp overhangs (listed in Table 3.3) matching the ends of the gene fragments that needed to be cloned in respective directions (Figure 4.8(a)). Concurrently, the *glpK* and *glpD* genes were amplified from the *E. coli* genome by PCR with primers containing overlapping ends that matched these overlapping fragments in the vector so that they could be cloned in forward and reverse directions (Figure 4.8(b) and (c)). The $P_{lac/ara-1}$ promoter was cloned from the pPROLAR.A122 vector (Figure 4.8(d)). The cloning of all the genes fragments is shown as a schematic representation in Figure 4.7.

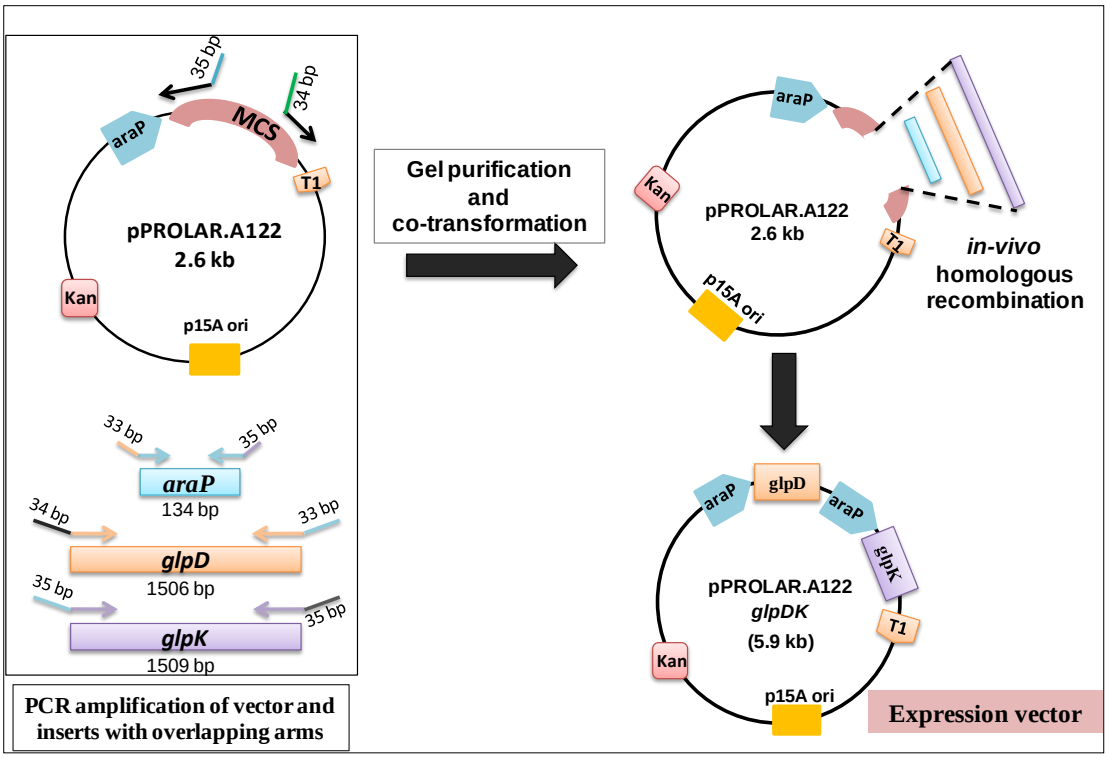

Figure 4.7 Pictorial representation of simultaneous cloning multiple PCR products into the pPROLAR.A122 vector using principle of homologous recombination in *E.coli* DH5α strain.

The PCR products containing the amplified vector and $P_{lac/ara-1}$ promoter were digested with DpnI to remove the traces of the original plasmid template in order to reduce the number of false colonies after transformation. The amplified gene products were eluted from 1% agarose gel and co-transformed into *E. coli DH5α* cells in a vector insert ratio of 2:1. The transformed cells were plated on LB agar plates containing kanamycin at a concentration of 50µg/ml and the resulting colonies were screened by colony PCR using forward *glpD* and reverse *glpK* primers (Figure 4.9(a)). The positive clones were also confirmed for $P_{lac/ara1}$ integration using forward araProm and reverse *glpK* primers (Figure 4.9(b)).

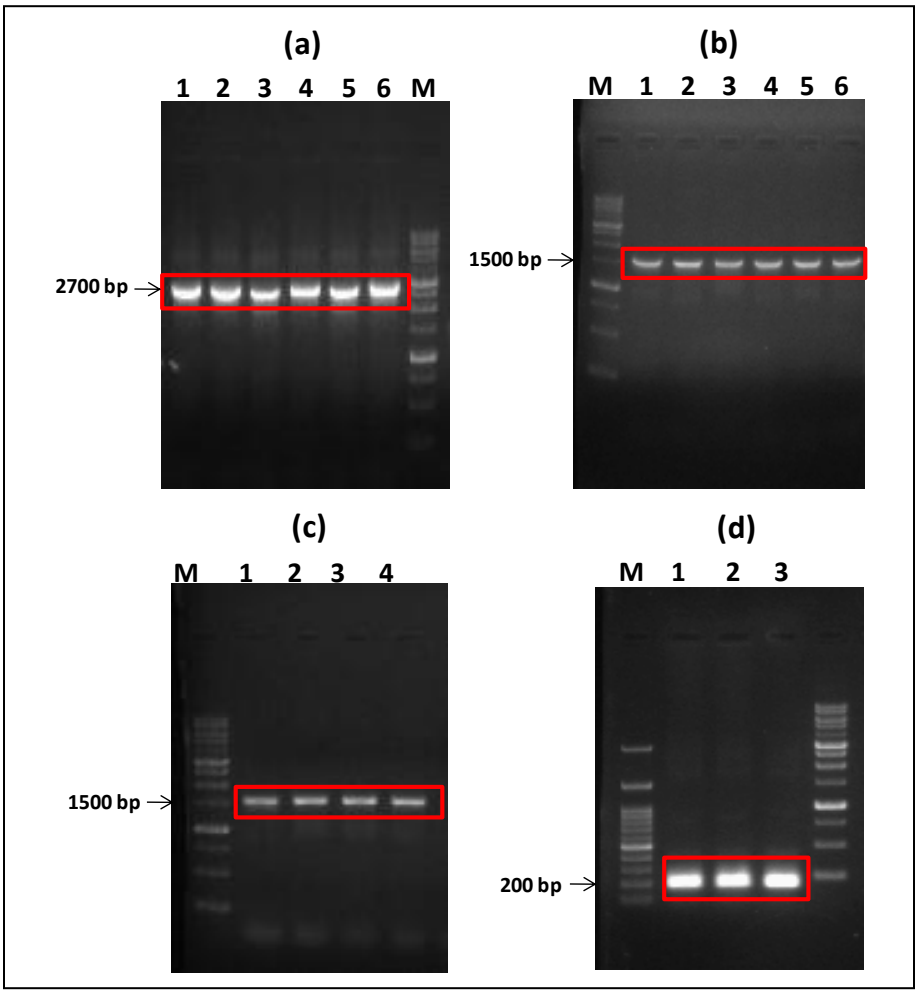

Figure 4.8 (a) Gel picture showing amplification of pPROLAR.A122 vector. Lane M: 1 kb DNA ladder, Lane 1-6: pPROLAR.A122 linear vector with overlapping ends (2.7 kb). (b) Gel picture showing amplification of *glpD* gene. Lane M: 1 kb DNA ladder, Lane 1-6: *glpD* gene with overlaps (1573 bp). (c) Gel picture showing amplification of *glpK* gene. Lane M: 1 kb DNA ladder, Lane 1-4: *glpK* gene with overlaps (1577 bp). (d) Gel picture showing amplification of *Ara* Promoter. Lane M: 100 bp DNA ladder, Lane 1-3: Amplified *Ara* promoter with overlaps (202 bp).

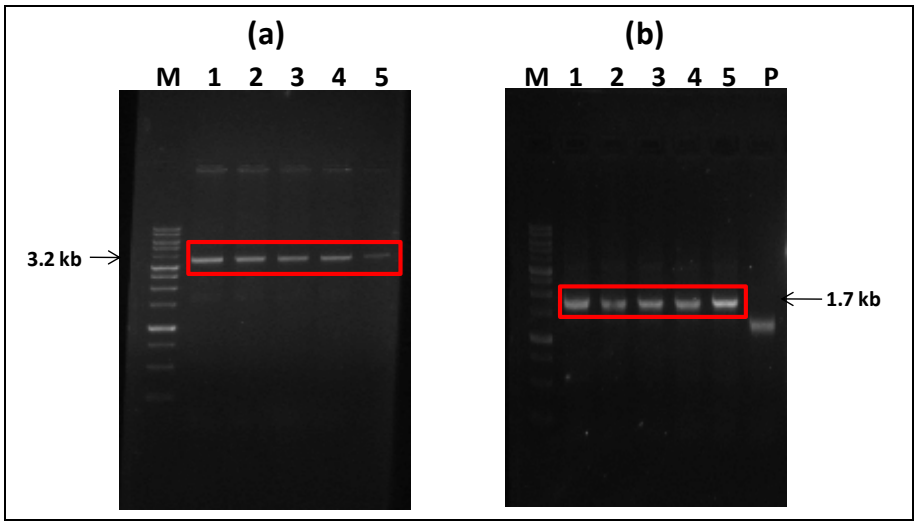

Figure 4.9 Agarose gel picture showing clone confirmation by colony PCR. (a) PCR confirmation *of glpD-araProm-glpK* insert (3.2 kb) by forward *glpD* and reverse *glpK* primers. (b) PCR confirmation of *araProm-glpK* insert (1.7 kb) by forward araProm and reverse *glpK* primers.

4.2.4 Effect of *glpDK* co-expression on growth and substrate utilization rates of L-asp producing modified cells

To study the effect of co-expression of *glpDK* genes on glycerol uptake rates, the modified host cells were co-transformed with plasmids containing both genes for glycerol uptake as well as L-asp production. Modified cells containing plasmid only for L-asp production were kept as control for this experiment. A single colony was inoculated into 50 ml flask containing 5 ml of TB medium containing 25µg/ml of kanamycin and 50µg/ml of ampicillin and grown overnight at 37°C with constant shaking at 200 rpm. We decided to supplement the TB media with 0.2% v/v glycerol rather than 0.4% v/v glycerol for further studies since it was found to diminish the problem of drop in pH after 5-6 hours of growth at shake flask level because of the very slow build of acetate especially when no automated external control for pH could be applied, like those for bioreactors (See appendix Figure A1).

500µl of primary culture was used to inoculate three secondary flasks containing 50 ml of TB media supplemented with 0.2% v/v glycerol, 10mM $MgSO_4$and 100µg/ml of ampicillin in case control culture (BW25113*Δ(elaA+cysW)* L-asp) and 25µg/ml of

kanamycin and 50µg/ml of amphicillin for the test culture (BW25113*Δ(elaA+cysW)* L-asp *glpDK)*. The test flasks that had modified cells containing both plasmids (pMAL-p2X*ansB* and pPROLAR.A122*glpDK*) were induced at OD ~1.5-2 with either 1mM IPTG for L-asp production only or with both 1mM IPTG and 0.2% L-arabinose to co-express both plasmids. The control flask was induced at $O.D._{600}$ ~1.5 with 1mM IPTG. Both control and test flasks were run in duplicates. Samples were collected till 26 hours post induction and cell growth was monitored by measuring $O.D._{600}$. Residual glycerol concentration was measured using HPLC and protein concentrations were determined by measuring the enzymatic activity for L-asparaginase as described by Khushoo *et al.* (2004).

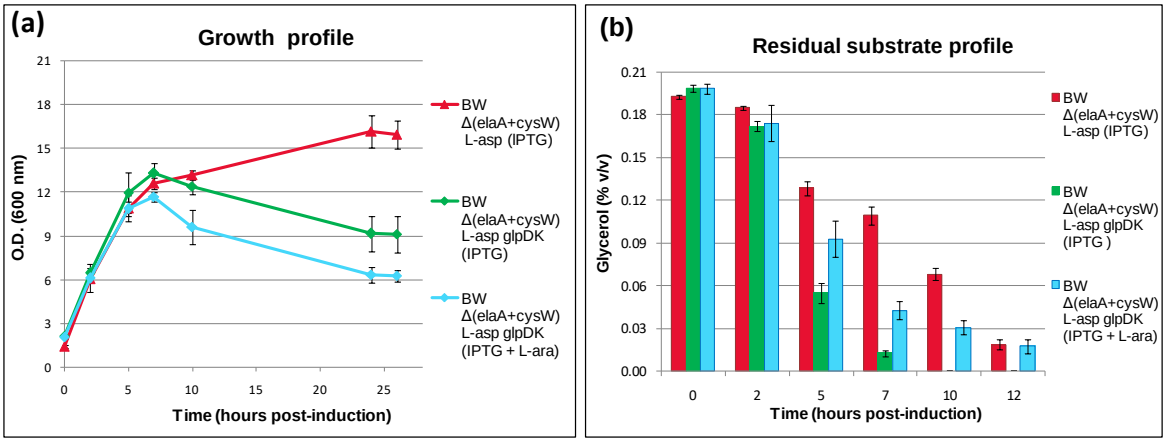

Figure 4.10(a) Growth profiles and (b) Residual substrate profiles of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

The co-expression of genes from plasmids in test cultures resulted in significant fall in growth rate 7 hours post induction compared to the control cultures (Figure 4.10(a)). This may be due to the additional burden imposed on the host cells due to the presence of second plasmid which competes for the utilization of cellular machinery for expression of their genes. This decline in growth rate was less for the test culture induced for L-asp only compared to the test culture induced for both plasmids. It has been reported previously that the plasmid systems based on *lac* promoter have the problem of basal level of expression due to incomplete repression offered by the natural *lac* operon

(Oehler *et al*. 1990). Since our pPROLAR.A122 vector contains a $P_{lac/ara1}$ hybrid promoter which contains repression operator sites from the *lacZYA* promoter, it has the associated problem of leaky expression which could be the reason for the decline in growth rate of the test culture induced for protein production only. Residual substrate profiles showed improvement in glycerol uptake rates 5 hours post induction for the test cultures induced for both plasmids. However the improvement in glycerol uptake rates was significantly better for the test culture that contained both plasmids but induced only for protein production (Figure 4.10(b)). The volumetric productivity per unit biomass at 24 hours post induction for the same test culture was 3.3 fold higher than control (Figure 4.10(c) and (d)). This led us to conclude that basal level expression of *glpD* and *glpK* genes was sufficient to improve the substrate uptake rates of modified cells. Also the extra burden imposed on the cells due to the presence of a second plasmid is insignificant as long as it is not induced and the low basal level expression obtained is ideal for countering the problem of reduced substrate uptake rates thereby leading to high volumetric productivities with these modified host backgrounds.

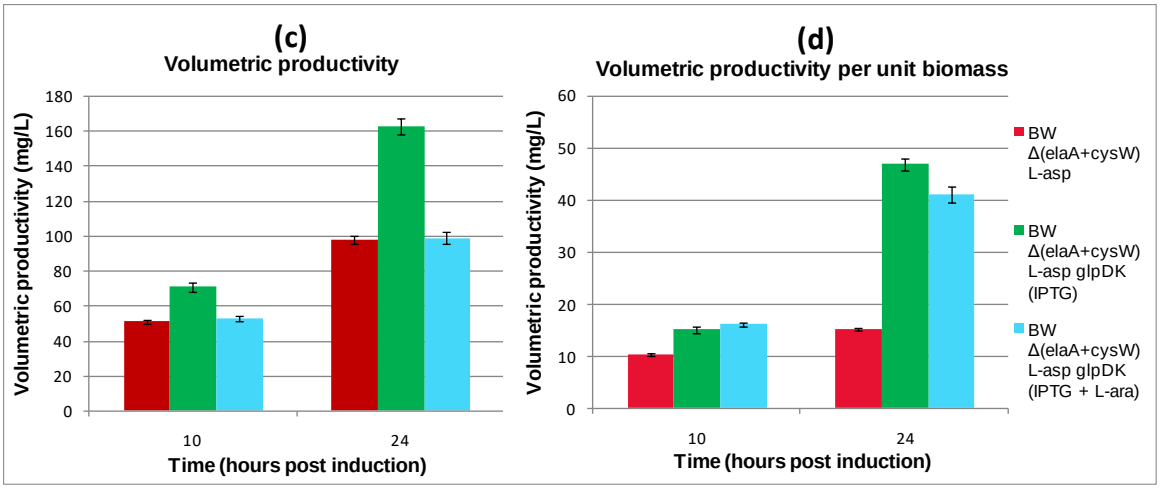

Figure 4.10(c) Total volumetric productivity and (d) volumetric productivity per unit biomass of control cells expressing L-asp and modified cells co-expressing genes for substrate uptake and L-asp production.

4.2.5 Effect of glycerol pulsing on substrate uptake rates and productivity

It has been reported earlier that *glpK* co-expression with protein production led to an increment in productivity only when glycerol was present in the medium (Singh and Mukherjee 2013). Therefore, we decided to add pulses of glycerol to the growing cultures beginning from the time-point when the residual glycerol got exhausted in the culture. The culture medium used was TB media supplemented with 0.2% v/v glycerol, 10mM $MgSO_4$ and 100µg/ml of amphicillin in case of control culture (BW25113*Δ(elaA+cysW)* L-asp) and 25µg/ml of kanamycin and 50µg/ml of ampicillin for the test culture (BW25113*Δ(elaA+cysW)* L-asp *glpDK)*. Since, we found that basal level expression of the glycerol uptake genes was sufficient to improve substrate uptake and productivity, we decided to induce the test culture for protein production only. The control and test flasks were run in duplicates. Induction was done at $O.D._{600}$ ~1.5-2 with 1mM IPTG for both control and test cultures. After 6 hours post induction, a glycerol pulse equivalent to raising the concentration of glycerol by 0.3% v/v was given to both control and test flasks. Glycerol pulsing was repeated at intervals of three hours till 12 hours post induction. Another feed pulse of 0.2% v/v glycerol was given 21 hours post induction to check the substrate uptake capacity of the control and test cultures during late hours, after the onset of stationary phase. Growth profiles were monitored by measuring $O.D._{600}$ and substrate uptake profiles were monitored using HPLC.

The growth rates of test cultures containing both plasmids was lower in comparison to control, but unlike previous experiment no sharp fall in the $O.D._{600}$ was observed for the test flasks 6 hours post induction (Figure 4.11(a)). This fall in $O.D._{600}$in the previous case can therefore be attributed to the depletion of glycerol in the medium post-induction (Figure 4.10(a)) implying that co-expression of glycerol uptake genes has a beneficial effect on growth rates only when glycerol is present in the medium. Residual glycerol profiles showed a significant improvement in substrate uptake capability of the test cultures upon glycerol pulsing. The glycerol uptake rates remained high till 12 hours post induction and then declined gradually (Figure 4.11(b)). The accumulation of glycerol after 21 hours post induction showed that cells were not able to consume glycerol in later hours after the onset of stationary phase.

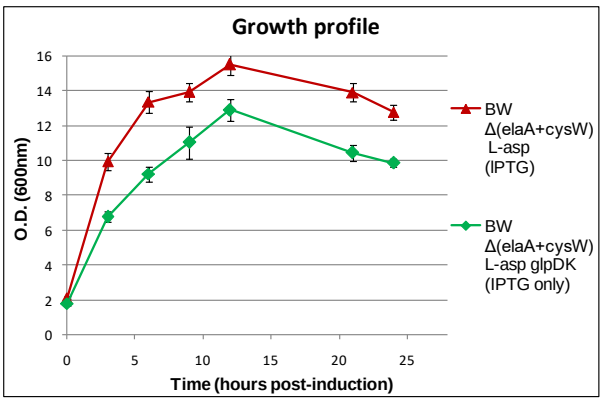

Figure 4.11(a) Growth profiles of control cultures expressing L-asp and test cultures co-expressing genes for glycerol uptake and L-asp production upon glycerol pulsing.

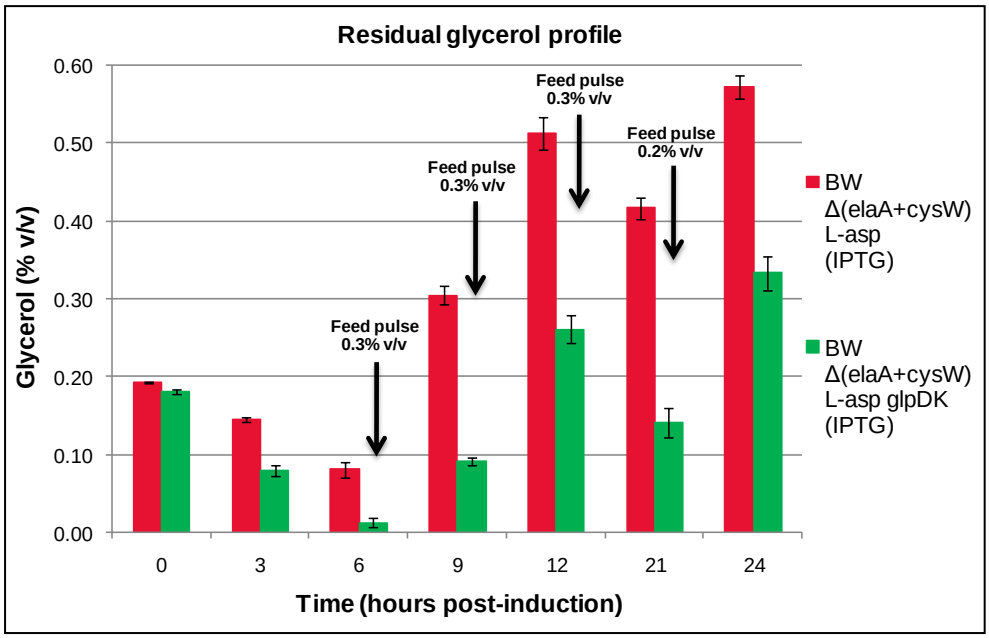

Figure 4.11(b) Residual glycerol profiles of control and test cultures induced with IPTG upon glycerol pulsing.

The increase in glycerol uptake rates of modified cells also resulted in an increased volumetric productivity compared to control cultures (Figure 4.11(c)). The volumetric productivity per unit biomass was 2.3 folds higher for modified cells carrying extra copies of glycerol uptake genes compared to control cells (Figure 4.11(d)). Equal volumes of supernatant run on 12% SDS gel showed higher expression of L-asp in modified cells supplemented with substrate uptake genes than control (Figure 4.11(e)).

This increment in glycerol uptake and volumetric productivity for the modified cells with basal level expression for *glpD* and *glpK* genes demonstrated the proof of principle that supplementation of substrate uptake genes resolved a critical issue which had remained unaddressed by the knock out strategy, thereby helping in the further improvement of L-asp productivity.

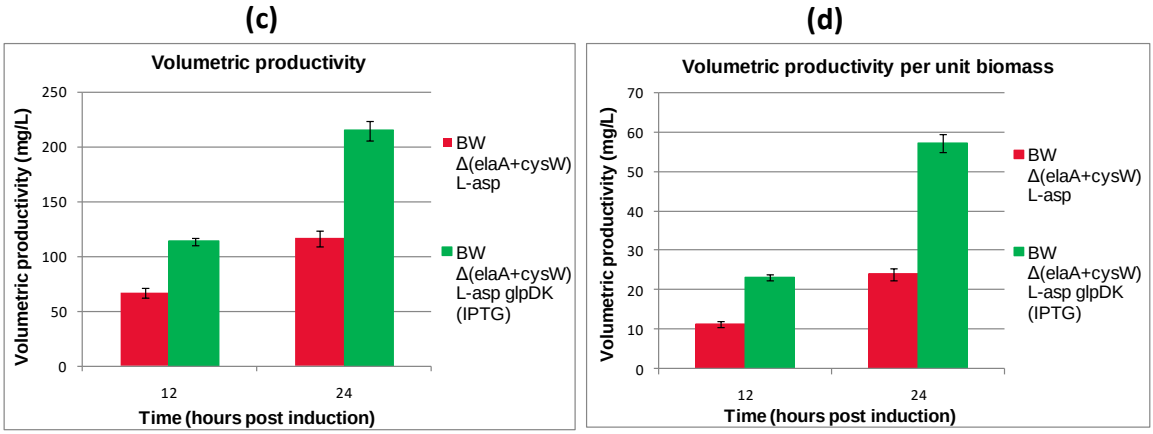

Figure 4.11(c) Total volumetric productivity and (d) volumetric productivity per unit biomass for control and test cultures induced with IPTG at 12 and 24 hours post induction.

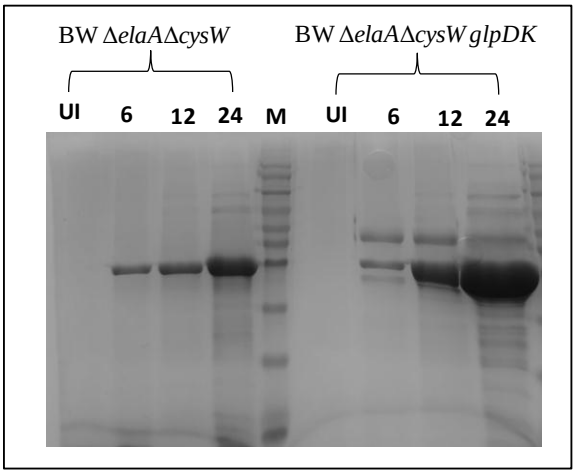

Figure 4.11(e) SDS gel picture showing extracellular expression of L-Asp post-induction 6, 12 and 24 hours in BWΔ(*elaA+cysW)* and BWΔ(*elaA+cysW)glpDK* strains induced with IPTG only. UI represents un-induced culture and M represents protein molecular weight ladder.

4.3 Analysis of cellular stress response

The second related goal of this work was to analyze the changes that took place at the molecular level which not only improved but also sustained the expression levels of L-Asp for longer time-periods. The enhancement in productivity levels of the modified cells indicates a disruption of some of the signaling pathways leading to CSR, thereby making them better in terms of their ability to counter the negative effects of the CSR. In order to gain a deeper in-sight of the effects of these modifications, an investigation of the modified host cell at the regulatory and physiological levels needed to be carried out under high cell density cultivation conditions. The idea was to discover the relationship between signaling pathways that are disrupted by these genomic modifications and recombinant protein synthesis, and how this relationship is mediated under bioprocess conditions. This would help in detecting the unidentified changes that occur during the process of recombinant protein synthesis. In the previous chapter, we showed an enhancement in expression levels of L-asp in modified host cells by following up on the leads obtained from transcriptomic studies. Here we extend this work to study the impact of recombinant protein production on the dynamic changes taking place inside the cell and its effect on cellular health.

4.3.1 Cell viability studies of control and modified cells

Cell viability is one of the parameter for measuring cellular health as it determines the number of living cell present in the culture medium. As shown in the previous chapter, modified cells were able to sustain higher expression levels even after entering the stationary phase, which indicated a better cellular health of the modified strain. Therefore, we decided to measure the cell viability of control (*E. coli* BW25113) and modified (*E. coli* BW*Δ(elaA+cysW)*) cells under non-producing conditions. Batch cultivations were carried out in an Electrolab fermenter with 750 ml of TB media supplemented with 0.4% v/v glycerol and 10mM $MgSO_4$. The secondary culture grown till mid-log phase was used as inoculum to the batch reactor. The temperature, pH and dissolved oxygen (DO) was set at 37°C, 7 and 100% respectively. Antifoam was added to the culture as required. The initial stirrer speed, automatic pH and DO control was kept same for both runs. The airflow rate was kept 1 vvm till the culture remained in log phase

and thereafter it was reduced gradually. The batch run had duration of 36 hours and samples were collected till late stationary phase.

The growth rate of control cells was higher than modified cells under non-producing conditions and they reached a maximum $O.D._{600}$ of 47 in 12 hours whereas the modified cells grew till an $O.D._{600}$ of 40 in 8 hours and then entered stationary phase (Figure 4.12(a)). Cell viability was monitored by a colony formation assay, which is based on the ability of single cells to form colonies. Samples collected at different time-points were diluted to equal $O.D._{600}$ for respective time-points as shown in Table 4.1. The number of serial dilutions was reduced according to the growth phase of culture. 100µl of sample from these serial dilutions was plated into petridishes containing LB agar medium and grown overnight at 37°C. The colonies formed on each plate were counted on the next day. To our surprise, cell viability of modified cells was five folds higher than control cells at 24 hours (Figure 4.12(b)) which suggests that these genomic modifications might be responsible for enhancing cellular survival in the stationary phase and this was a critical factor that probably helped these modified cells to sustain high levels of expression for longer time periods.

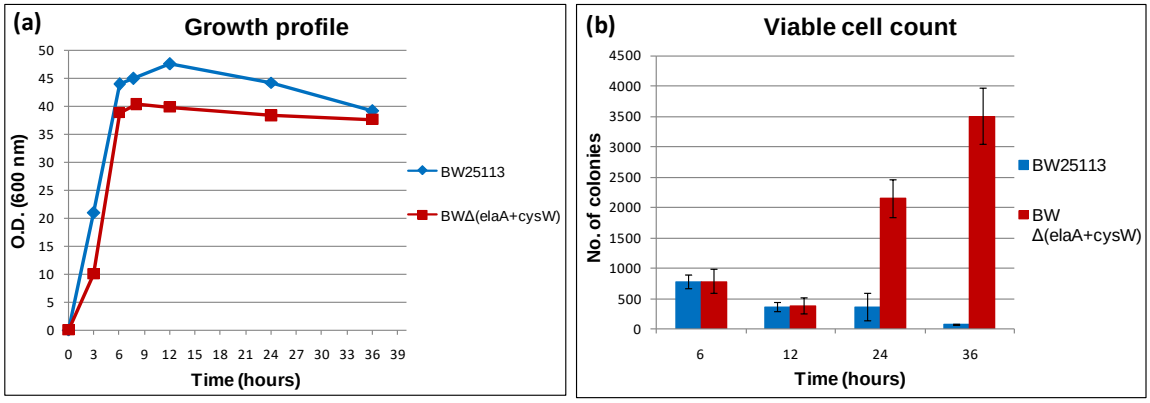

Figure 4.12(a) Growth profile and (b) Viable cell count for control and modified cells under non-producing conditions.

Time (hours)	Volume plated	O.D. (600 nm) after serial dilution
6	100 µl	0.0001
12	100 µl	0.0001
24	100 µl	0.001
36	100 µl	0.1

Table 4.1 Serial dilutions of respective time-points taken for colony formation assay.

4.3.2 Confirmation of cellular health by induction for protein synthesis in stationary phase

Because of these surprising results obtained by cell viability studies, we decided to further check cellular health by imposing a metabolic burden of recombinant protein synthesis during stationary phase. We wanted to analyze whether this improved cellular health was associated with enhanced production capability of these modified cells in stationary phase. We chose sfGFP (a superfolding variant of GFP) as a model protein for this purpose, which is expressed as a soluble protein inside the cells and can be monitored online. This experiment was done in a microbioreactor (M2P Labs, Germany) which is meant for high-throughput micro fermentations and allows online monitoring of biomass, GFP fluorescence and other bioprocess parameters like pH, DO and NADH levels in a 48 well format. Both control and modified cells were transformed with *pBAD24* plasmid containing the sfGFP gene. 900µl of TB media supplemented with 0.4% v/v glycerol and 10mM $MgSO_4$ was inoculated with a 1% inoculum taken from an overnight grown primary culture. The culture was allowed to grow under controlled bioprocess conditions and induced for sfGFP synthesis at early-log phase (3 hours), late-log phase (6 hours) and stationary phase (12 hours). As expected the sfGFP production per unit biomass was 1.7 folds higher for modified cells compared to control for the culture induced in early-log phase. The sfGFP production per unit biomass was 2.3 folds higher for modified cells when induction was done in late-log phase (Figure 4.13). The induction in stationary phase did not result in sfGFP accumulation in control cells; however modified cells were still able to produce fair amounts of sfGFP upon induction in stationary phase. Also, a sharp decline in growth rate was observed for control cells after 12 hours of cultivation which could be reason behind no sfGFP build-up in these cultures upon stationary phase

induction. These results confirmed that the improved performance of the modified cells was primarily due to the fact that they could sustain better cellular health as well as productivity in comparison to control cells.

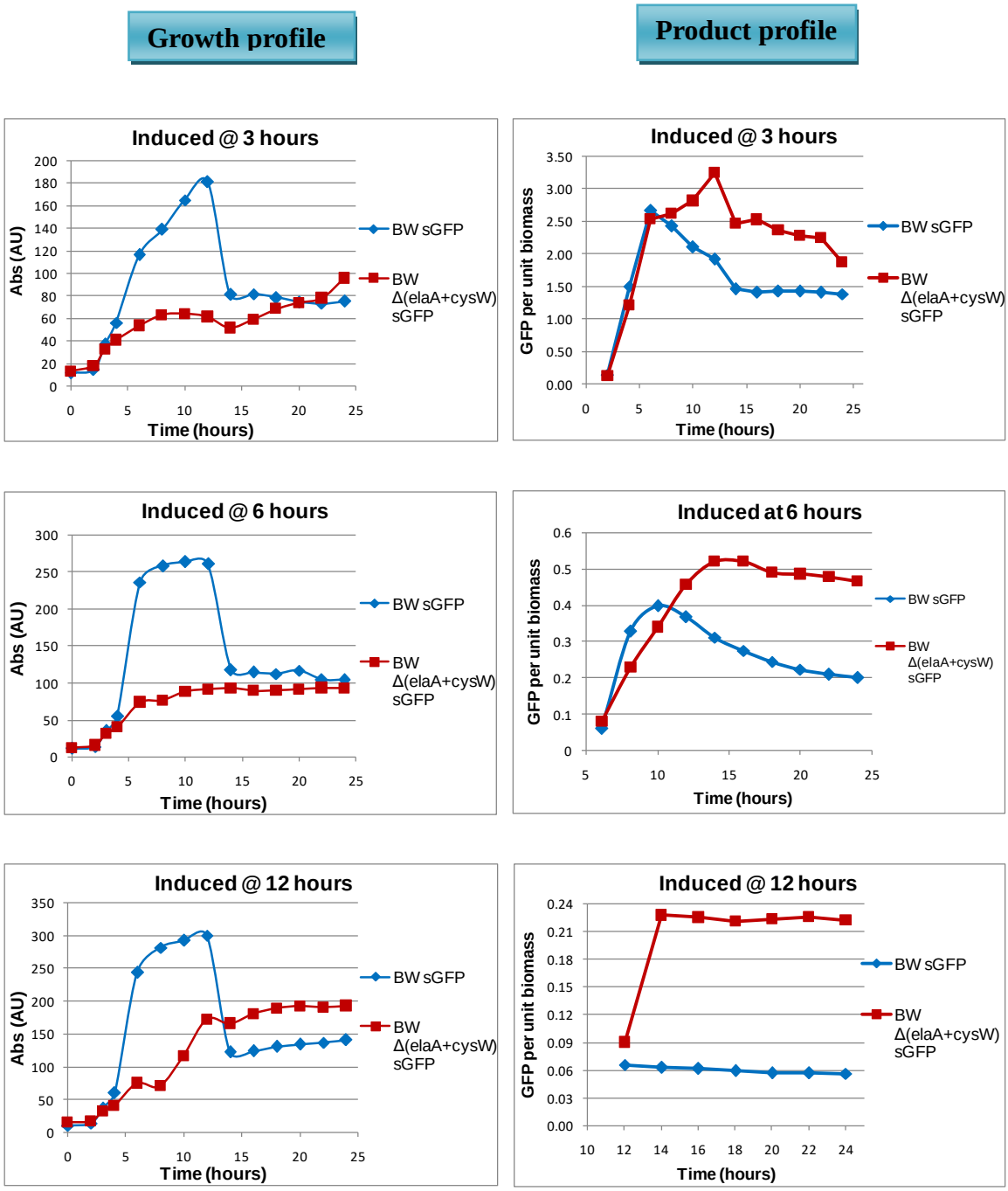

Figure 4.13 Growth and Product profiles for control and modified cells when induced for sfGFP production at early-log phase (3 hours), late-log phase (6 hours) and stationary phase (12 hours).

4.4 Study of metabolic regulation during stationary phase

The results obtained from the studies on cell viability and cellular health led to two key questions; (a) Is the stationary phase gene expression higher in modified cells compared to control cells? (b) Can we identify similarities between recombinant protein mediated CSR and CSR generated due to physiological stress? To answer these questions, we realized the need to study the CSR generated during stationary phase in terms of metabolic regulation in both control and modified cells. The regulation of stationary phase is primarily mediated by transcriptional regulators or transcription factors (TFs), a majority of which are global regulators involved in maintaining nucleoid structure (*fis, dps, ihf* and *hns*) and adaptation of cellular metabolism to perturbed conditions (*crp* and *lrp*), and sigma factors which are the protein sequences that initiate the binding of RNA polymerase to the promoter region of the gene (Mar *et al*. 2010). The changes in interaction patterns of RNA polymerases with seven sigma subunits and about 300 TFs control this metabolic regulation (Yamamoto, Watanabe, and Ishihama 2014).

We identified some key stress signaling molecules involved in stationary phase survival from reports available in literature (Marta *et al*. 1992; Lange and Hengge-Aronis 1994; Azam *et al*. 1999; Nair and Finkel 2004) and categorized them under four categories namely stringent stress response which involves two regulatory molecules *relA* and *spoT*, the programmed cell death (PCD) mechanism based on toxin-antitoxin (TA) module, global regulators that are involved in maintaining nucleoid structure and sigma factors (Figure 4.14).

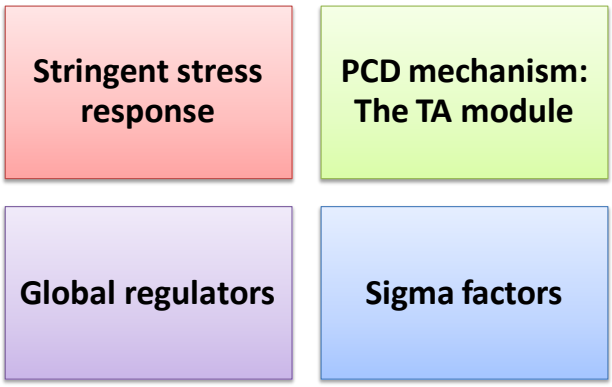

Figure 4.14 Grouping of key stress signaling molecules involved in stationary phase survival into four major categories.

4.4.1 Transcriptomic profiling to identify the regulatory molecules involved in stationary phase survival

Earlier transcriptomic studies conducted in our lab had analyzed the post induction response of log phase cultures expressing recombinant protein (Ashish K. Sharma *et al*. 2011). Since we were interested in analyzing the reasons behind sustained expression levels in stationary phase, we decided to extend our analysis till late stationary phase. To identify the similarities and differences between recombinant protein mediated CSR and CSR generated due to physiological stress, we analyzed the control and modified cells under both induced and non-induced conditions. Samples were collected at various time points as shown in Table 4.2. For non-induced cultures, the mid-log phase sample was taken as a reference point in order to capture the changes occurring during the cell's transition to stationary phase. For recombinant protein producing cultures, the un-induced samples were taken as a reference point to capture the changes that take place after the cell encounters this stress.

Strain	Reference point	Time-points
BW25113	4 h	10 h, 24 h, 36 h
BW25113 *Δ(elaA + cysW)*	4 h	10 h, 24 h, 36 h
BW25113 L-asp	0 h (Induction point)	4 h, 10 h, 24 h (post-induction)
BW25113 *Δ(elaA + cysW)* L-asp	0 h (Induction point)	4 h, 10 h, 24 h (post-induction)

Table 4.2 Details of samples collected for transcriptomic studies.

It is well known that recombinant protein synthesis elicits a cellular stress response (CSR) which causes growth retardation. However, many aspects of this CSR are not specific to the stress inducer (the trigger factor which actually induces this stress) because the cell detects stress based on changes in its macromolecular composition irrespective of what kind of stress has led to those changes (Kültz 2005). Entry into stationary phase is the first line of defense of the cell in response to such changes. However, the cellular responses initiated in order to re-establish homeostasis are specific to the kind of stress

induced and mostly activated in parallel to the CSR (Kültz 2005). This is one of the reasons why recombinant protein mediated stress often mimics the features of a stationary phase stress response. Therefore, under recombinant protein producing conditions, we assumed that the stress response would be fully active by 4 hours post induction and therefore, the cellular characteristics should be somewhat analogous to the stationary phase.

4.4.2 qRT-PCR (Real time RT-PCR) experiment

For obtaining a comprehensive picture of metabolic regulation inside the cell, we decided to monitor the relative transcript levels of key stress signaling molecules which we have grouped into four categories as shown earlier (Figure 4.14). Transcriptomic profiling was performed using qRT-PCR, where relative mRNA levels of 17 major regulators involved in stationary phase survival were measured. The samples for RNA isolation were collected from high cell density batch cultures (1L volume) of control and modified strain with and without L-asp production. For relative quantification, the changes in mRNA levels are generally measured by choosing a 'reference gene' or 'calibrator gene' gene that should possess no or minimal changes in gene transcript level for the different time points of the particular condition. We selected five reference genes for screening namely *gapA, cysM, rimL, recC, uxuR* based on results of previous RNAsec analysis conducted in our lab, out of which *rimL* (ribosomal-protein-serine acetyltransferase) showed minimal variations across all the time points and was selected for further studies. All experiments were run in triplicate with positive control (PC) using *E. coli BW25113* genomic DNA for each set of primer and NTC (i.e. non-template control) to check any non-specific signal amplification from primer-dimer formation or from cross-contamination.

4.4.3 Analyzing the transcriptomic data

4.4.3.1 Analysis of stringent stress response

It is well known that (p)ppGpp accumulation inside bacterial cells triggers a stringent stress response (Bouveret 2009), which in turn induces the stationary phase sigma factor RpoS (σ^S) to modify the cellular metabolism towards growth retardation, sustenance of

cellular maintenance and activation of stress survival genes. Two classes of enzymes belonging to the RelA and SpoT family are known to control cellular concentrations of (p)ppGpp (Battesti and Bouveret 2009). We therefore compared the transcript levels of *relA* and *spoT* which play a major role in eliciting this stringent stress response during stationary phase in both control and modified cells. As expected, both *relA* and *spoT* were found to be 1.2~1.5 folds upregulated in later hours of the stationary phase in both control and modified cells under non-producing conditions (Figure 4.15). Downregulation of *spoT* at the 10^{th} hour of growth when cells had just entered stationary phase indicated that no (p)ppGpp accumulation took place via this pathway beyond this time point. However, the trend was quite different for recombinant protein producing cultures. The transcript levels of *relA* were found to be downregulated (1.2 folds) in post induction stationary phase cultures of L-asp producing control cells. A decline in *relA* expression levels with declining growth rate has also been observed previously in L-asp expressing *E. coli W3110* strain carrying hnsΔ93-1 mutation (Mahalik *et al*. 2017). However, *spoT* was respectively 4.2 and 1.3 folds upregulated in 10 hour and 24 hour post induction samples.

These results indicate that only the *spoT* mediated pathway is responsible for activation of stringent stress response in L-asp producing control cells. Interestingly both *relA* and *spoT* were found to be 2~3.3 fold downregulated in post induction cultures of modified cells expressing L-asp indicating no ppGpp accumulation and hence no activation of the stringent stress response pathway in this case. This means that the genomic modifications *Δ(elaA+cysW)* had disrupted the signaling pathways that enables cells to sense nutrient deprivation upon recombinant protein synthesis. It should be noted that pathways leading to the stringent stress response in modified cells are inactivated only when recombinant protein mediated stress is present. These modifications are not able to counter the effect of stringent stress response under the conditions of general physiological stress.

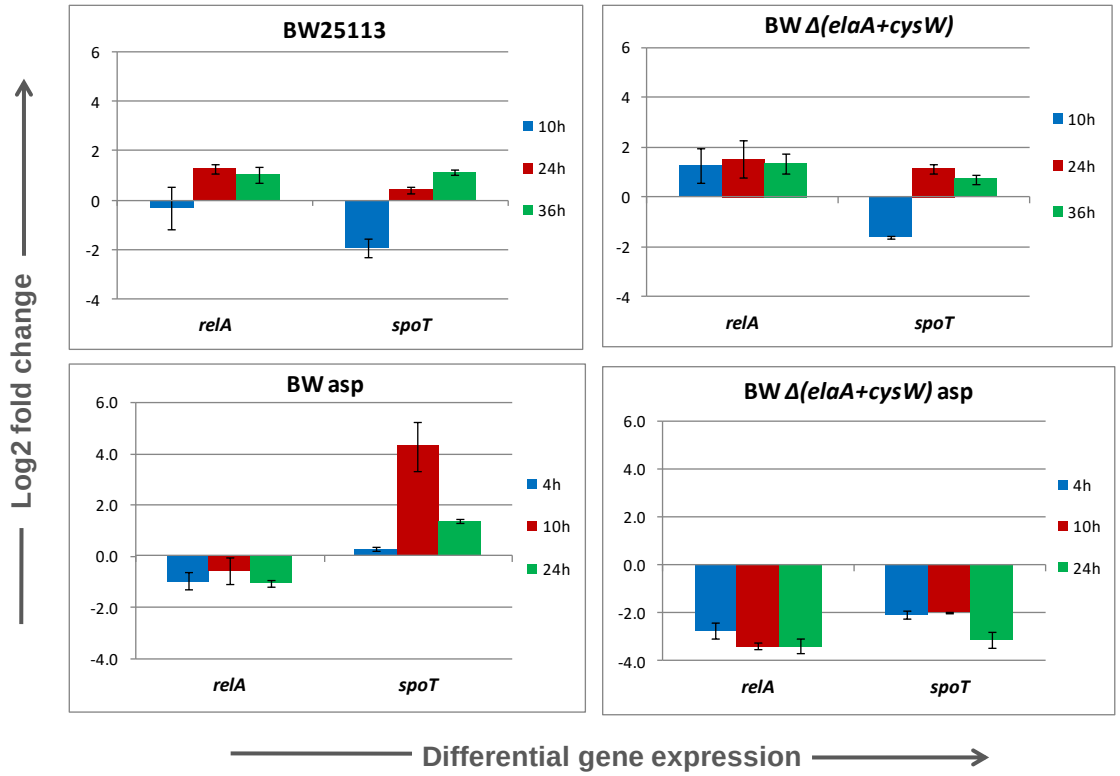

Figure 4.15 Differential mRNA expression levels of stringent stress response pathway genes *relA* and *spoT* in control and modified cells under protein producing and non-producing conditions.

4.4.3.2 Analysis of the programmed cell death (PCD) signaling pathway

Programmed cell death (PCD) is one of the long term survival strategies that bacterial cells adopt by terminating a majority of the population in response to the complete disappearance of nutrients and accumulation of toxic waste products in its surroundings. In *E. coli*, the TA module *mazEF* is the most extensively studied PCD mechanism, which consists of two genes *mazF* and *mazE*; the former encodes a toxic protein and the latter its corresponding antitoxin. An analysis of the differential mRNA expression profiles of these proteins did not show any significant changes in control (non-modified) strain, with only a mild up-regulation of *mazE* and *mazF* in the 10^{th} and 24^{th} hour of growth (Figure 4.16). We had also previously observed a significant decline in cell viability for the control strain in the 24^{th} and 36^{th} hour of cultivation (Figure 4.12(b)), which can be correlated with these mild changes obtained in mRNA expression levels of *mazE* and

mazF. The transcript levels of *mazF* were found to be upregulated in L-asp producing control cultures, whereas a slight downregulation of *mazE* which produces a labile antitoxin was seen during the 4th and 10th hour post induction indicating the possibility of *mazF* mediated activation of PCD in later hours. However, significant differences were observed for the modified cells, where *mazF* was found to be 1.7 fold down-regulated at the 10th hour under non-producing conditions. For the modified cells producing L-asp, *mazF* was found to be 2~2.5 folds down-regulated at the 4th, 10th and 24th hour post induction. The *mazE* levels were correspondingly 2~2.6 fold downregulated in these cultures indicating the inactivation of *mazEF* mediated PCD pathway in modified cells. These results elucidated the reason behind the high cell viability observed for the modified cells during stationary phase in comparison to control.

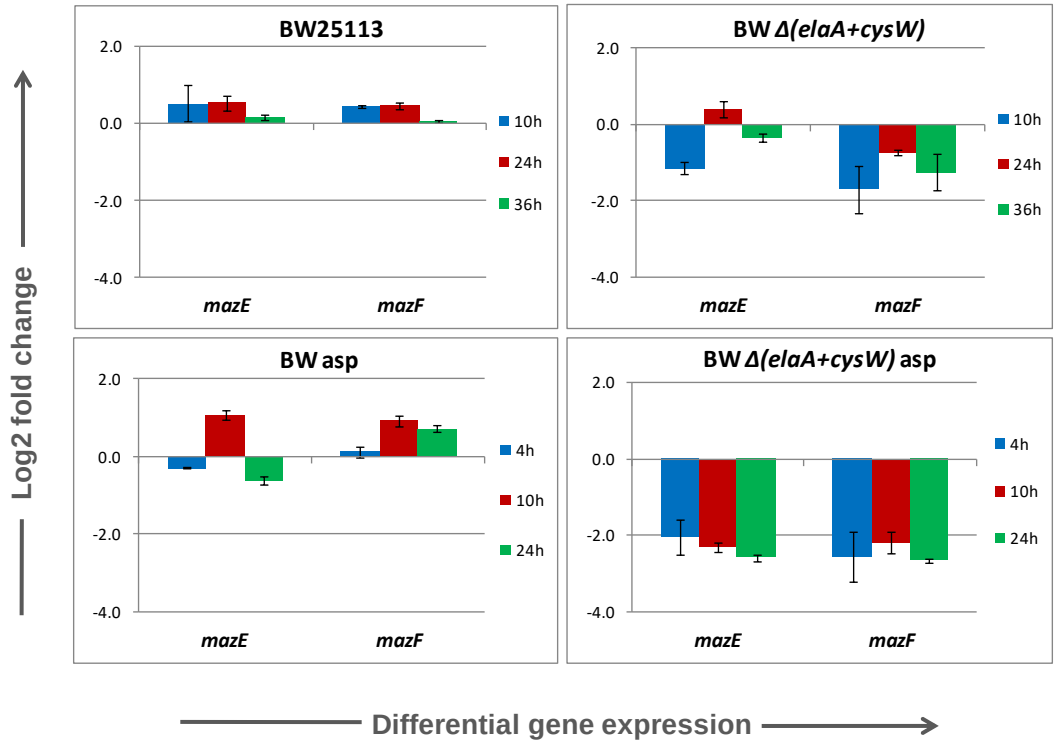

Figure 4.16 Differential mRNA expression levels of the toxin antitoxin module *mazEF* in control and modified *E. coli* cells under protein producing and non-producing conditions.

4.4.3.3 Analysis of global regulators

Global regulators control the expression levels of a large number of genes and hence belong to the top most hierarchy in the gene regulatory network (Julio *et al*. 2010). We studied the global regulators belonging to three categories: nucleoid associated proteins (NAPs) which are involved in maintain nucleoid structure, regulators involved in adaptation of cellular metabolism and small RNAs that have regulatory role and alters gene expression in stationary phase. The first group includes *fis, lrp, dps, hns, cbpA* and *ihfB* whose expression levels are correlated with the growth phase of bacterial cells. In addition to their structural roles, these nucleoid proteins are also involved in global regulation of essential cellular functions like replication, recombination and transcription.

The factor for inversion stimulation (fis) is a DNA binding protein which is known to play an important role in site-specific DNA inversion reactions (Xu and Johnson 1995). In addition to this fis also plays a major role in trans activating many stable RNA operons (Nilsson *et al*. 1990). The expression levels of *fis* mRNA and protein peak during early exponential phase, and gradually decline thereafter, becoming undetectable during starved conditions (Hirsch and Elliott 2005; Mallik *et al*. 2006). The same pattern for *fis* mRNA was reflected in our studies under both protein producing and non-producing conditions (Figure 4.17).

The global regulatory protein lrp is one of the major regulators of the stationary phase which controls the expression of genes involved in the biosynthesis of amino acids, carbon source utilization, transport system of nutrients etc. (Tani *et al*. 2002). Lrp senses nutrient limitation and helps the cell in adapting to changes in nutritional state by increasing amino acid anabolism (Landgraf, Jingcai, and Calvo 1996). *lrp* mRNA levels were found to be down-regulated during the stationary phase in our control and modified cells with maximal down-regulation (4 folds) seen at 24 hours post induction in the modified strain producing L-asp (Figure 4.17). This decline in *lrp* expression in the stationary phase is probably related to growth stoppage and subsequent slowdown of anabolic reactions inside the cells. It is important to note that we used complex growth medium (Terrific Broth) for cultivation unlike minimal media used in some previous studies (Newman and Lin 1995), where *lrp* levels were found to be up-regulated.

Dps is a starvation induced DNA binding protein present in *E. coli* which protects DNA from damage by formation of DNA–Dps crystalline structures during stationary phase (Minsky *et al*. 1999). Dps is also known to provide resistance to cells from oxidative damage and its levels have been shown to peak near late stationary phase in *E. coli* cells (Almiron *et al*. 1992). Upregulation of the *dps* gene was observed in the 4th (4.8 folds) and 10th hour post induction (3.4 folds) in control cells producing L-asp indicating the presence of starvation stress possibly induced due to recombinant protein synthesis. However, in modified cells producing L-asp, *dps* levels remained near marginal values indicating that the genomic modifications were somehow preventing the cells from sensing starvation (in the form of diversion of metabolic fluxes towards recombinant protein synthesis). The down-regulation of *relA* and *spoT* observed earlier is also in accordance with these results. Interestingly, *dps* also got down-regulated in the late stationary phase of the control culture which could be one of the reasons behind the loss of viability in later hours of culture.

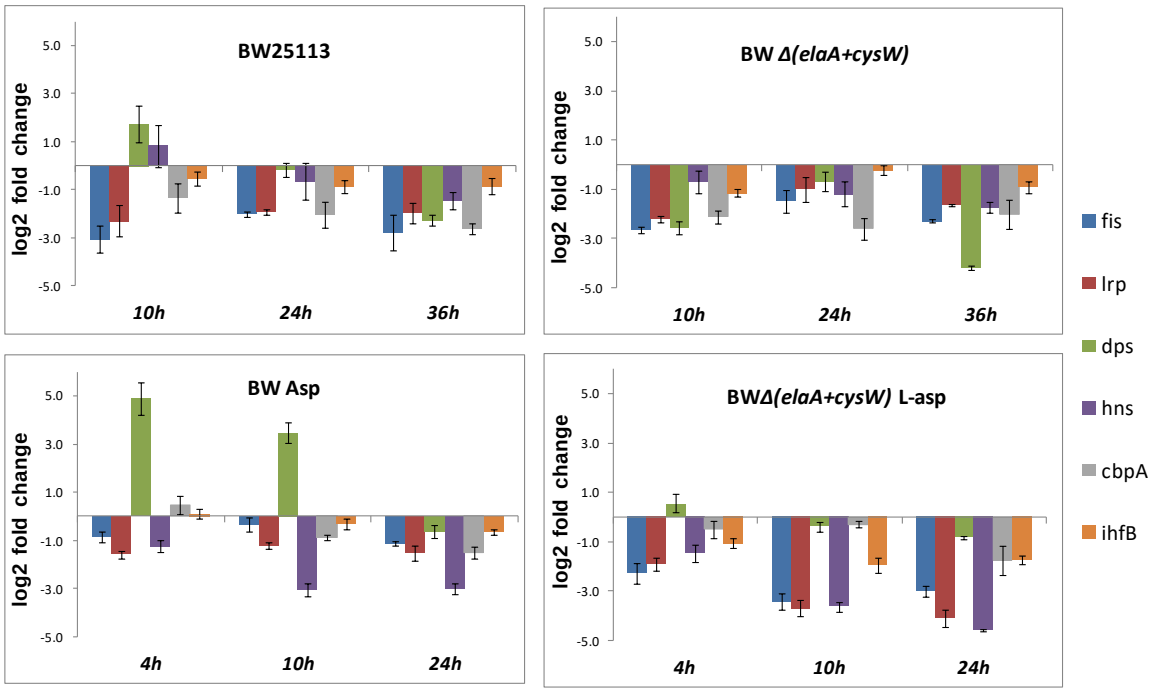

Figure 4.17 Differential mRNA expression levels of global regulators in control and modified *E. coli* cells under protein producing and non-producing conditions.

H-NS (histone-like nucleoid structuring protein) was originally identified as one of the most abundant nucleoid associated protein which acts as a global repressor for the transcription of a large number of *E. coli* genes that are generally regulated by various environmental stimuli (Atlung and Ingmer 1997). It has been reported that *hns* mRNA levels decline at the onset of stationary phase due to blockage of DNA synthesis in cells (Free and Dorman 1995). Except for the early stationary phase time point in control cells, *hns* was found to be down-regulated in all conditions with the maximum decline (3~4.5 fold) in expression levels seen for control and modified cells in the 10^{th} and 24^{th} hour post induction (Figure 4.17).

IHF (the integration host factor) is a small DNA binding heterodimeric protein with two subunits (ifhA and ihfB) and is known to regulate various bacterial processes like DNA packaging and partition, replication, recombination and gene expression (Freundlich *et al*. 1992). The *ihfB* gene is known to be transcribed from multiple promoters and it has been reported that expression levels of polycistronic transcripts of *ihfB* decrease as cells enter stationary phase, whereas the monocistronic *ihfB* levels becomes abundant upon growth retardation (Weglenska *et al*. 1996). Similarly, cbpA is another DNA binding protein which is known to be expressed only during stationary phase. The mRNA expression levels of *cbpA* gene are regulated by σ^{38} subunit of RNAP in stationary phase (Azam *et al*. 1999). Both ihf and cbpA proteins are predicted to be involved in organization of nucleoid structure in stationary phase. The *ihfB* expression levels remained significantly unchanged in L-asp producing control, but downregulated by ~1.95 folds in modified cultures producing L-asp. *ihfB* mRNA expression also remained downregulated in control and modified cells under non-producing conditions. A steady decline in expression levels of *cbpA* gene and trend observed for *ihfB* in stationary phase cultures (Figure 4.17) is somewhat contradictory to the reports available in literature and we could not assign any reason behind this trend.

The next category of global regulators analyzed was small RNAs that regulate the gene expression in stationary phase by stimulating translation and influencing the stability of specific mRNAs. One among them is *rybB*, an Hfq-binding small RNA which is involved in regulating outer membrane permeability. The induction of *rybB* sRNA results in

decreased synthesis of outer membrane porin proteins OmpC and OmpW and allow the cells to withstand stress periods by fine tuning the composition of the outer membrane (Johansen *et al*. 2006). *rybB* was found to be up-regulated (1.1~3.6 fold) in stationary phase cultures under all conditions except for control expressing L-asp (Figure 4.18), indicating the decreased synthesis of porins and induction of envelope stress response.

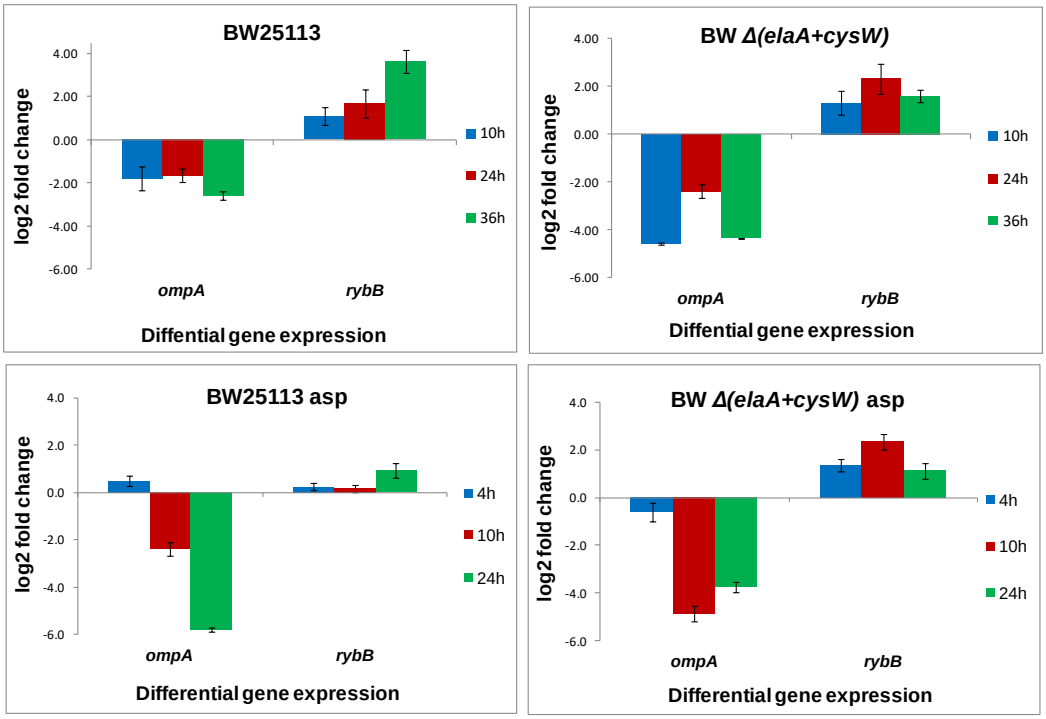

Figure 4.18 Differential mRNA expression levels of small RNAs associated with cell envelope stress in control and modified *E. coli* cells under protein producing and non-producing conditions.

OmpA, one of the three major bacterial outer membrane proteins, plays a significant role in maintaining integrity of the bacterial cell envelope and is negatively regulated by another Hfq-binding small RNA *micA* (Johansen *et al*. 2006). It has been reported that an increase in *micA* levels upon entry to stationary phase and concomitant decrease in *ompA* expression levels inhibits ribosome binding at the ompA translation start site which facilitates RNaseE cleavage and affects the stabilities of several RNAs (Udekwu *et al*. 2005). A sharp decline in *ompA* sRNA expression levels was seen at 10 hours and 24 hours post induction in both control and modified cells (Figure 4.18). *ompA* was also

down-regulated in stationary phase cultures of control (1.6~2.5 folds) and modified cells (2.34.6 folds) under non-producing conditions. The similar behavior of both *rybB* and *ompA* under induced and non-induced conditions confirmed that their regulation is indeed part of the generalized stress response.

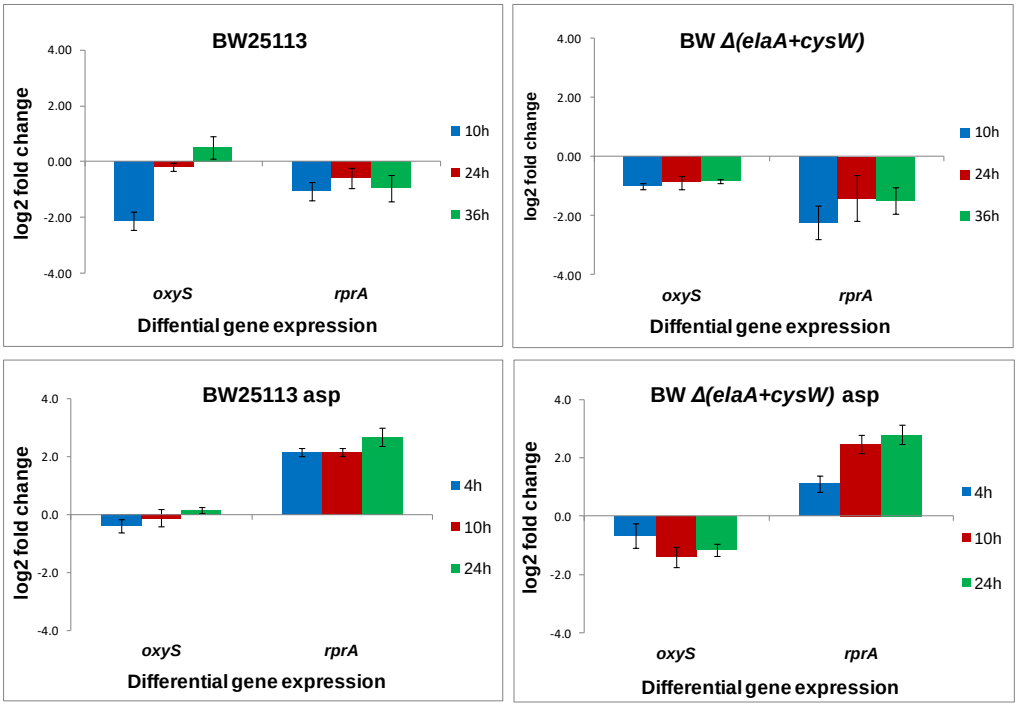

Figure 4.19 Differential mRNA expression levels of small RNAs associated with *rpoS* translation in control and modified *E. coli* cells under protein producing and non-producing conditions.

The small RNAs *rprA* and *oxyS* are involved in regulation of σ^S factor translation. *rprA* is known to stimulate *rpoS* translation by interacting with *rpoS* mRNA (Majdalani, Hernandez, and Gottesman 2002), whereas oxyS which appears under the conditions of oxidative stress represses RpoS translation by competitive binding of RNA chaperone Hfq (A. Zhang *et al*. 1998). There was no significant change in *oxyS* sRNA expression levels in control cells during the later hours of stationary phase under both induced and non-induced conditions. In contrast, modified cells showed down-regulation of *oxyS* in stationary phase cultures under both induced and non-induced conditions (Figure 4.19). Possibly the modified cells have a heightened ability to block the oxidative stress

response even though it seems to be not directly related to the stress induced by recombinant protein expression. The small RNA *rprA* was found to be 2~2.7 fold up-regulated under recombinant protein producing conditions which indicates the stimulation of *rpoS* translation via this route and hence, activation of the RpoS mediated stress response.

4.4.3.4 Analysis of sigma factors

Bacterial cells have evolved signaling cascades for regulation of their gene expression in order to respond adequately to changing physiological conditions and hence enhance their chances of survival. The major players of this regulation are sigma factors which play a key role in promoter recognition and translational initiation (Gruber and Gross 2003). The entry of a bacterial cell into stationary phase is mediated by the interchange of these sigma factors and their interaction with many other global regulators (Ishihama 2000). It is well known that the *E. coli* genome synthesizes seven different sigma factors (Table 4.3), each of which is involved in recognition of a different group of promoters. The exposure of cells to various stress conditions also influences the action of sigma factors and transcriptional regulators which bring about changes in gene expression. In order to understand the differences in the gene expression pattern of the modified strain in comparison to control, we analyzed the fold changes in expression levels of three sigma factors (*rpoS, rpoH and rpoE*) which are known to get expressed during various stress conditions.

Sigma Factors	Function
RpoD (σ70)	Housekeeping gene expression
RpoS (σ38)	Initiation of stationary phase stress response
RpoF (σ28)	Synthesis of flagella and chemotaxis
RpoN (σ54)	Activation of nitrogen metabolism
RpoH (σ32)	Response to heat shock
RpoE (σ24)	Response to membrane damage associated stress
FecI (σ19)	Nitrogen regulation

Table 4.3 List of sigma factors encoded by *E. coli* genome, along with their functions.

***rpoS* (σ^S):** *rpoS* is stationary phase sigma factor which is known as a master regulator of the general stress response in *E. coli*. Genome wide analysis of RpoS mediated gene expression has revealed its direct or indirect involvement in regulating the expression of around 10% of *E. coli* genes (Weber *et al*. 2005). *rpoS* plays a critical role in cell adaptation under stress conditions and is also known for negatively regulating the expression of TCA cycle and flagellum biosynthesis genes in stationary phase (Patten *et al*. 2004). RpoS expression is nearly undetectable in the logarithmic phase and generally gets induced by changes in metabolism that leads to growth reduction. RpoS levels have a complex regulation at the transcriptional, translational and post-translational levels that are tightly coordinated in response to various stress signals (Hengge-Aronis 2002). Multiple factors induce the expression of RpoS like nitrogen starvation, accumulation of signaling alarmone (p)ppGpp, low pH etc. and it is negatively regulated by the global regulator *fis* and ClpXP (that degrades RpoS under normal growth conditions), cyclic AMP and UDP-glucose (Ishihama Akira 2000). In L-asp producing control cultures, the *rpoS* levels were found to be 1.7 and 1.2 fold upregulated initially at the $4^{th}$and 10^{th} hour but later got downregulated at the 24^{th} hour post induction (Figure 4.20). *rpoS* was also upregulated in L-asp expressing modified cells except at the 10^{th} hour post induction. Given the complexities involved in *rpoS* function, it is difficult to understand whether this *rpoS* up-regulation is hindering or promoting cell survival under recombinant protein induced stress conditions. A contradictory behavior was observed for *rpoS* under non-protein producing conditions, where it was found to be down-regulated in stationary phase cultures of both control and modified cells.

***rpoH*(σ^H):** The *rpoH* sigma factor governs the heat shock response in *E. coli* by initiating transcription of heat-shock promoters. It is known to get activated in response to the stress induced in cytoplasmic compartment (Johansen *et al*. 2006). *rpoH* was found to be 1.5~3.5 fold downregulated in control and modified cells during stationary phase (Figure 4.20). A similar observation was made for post induction cultures of modified cells expressing L-asp indicating the absence of heat shock type or cytoplasmic stress response. However, a minimal change in *rpoH* expression levels on transition to stationary phase was observed for L-asp producing control cells.

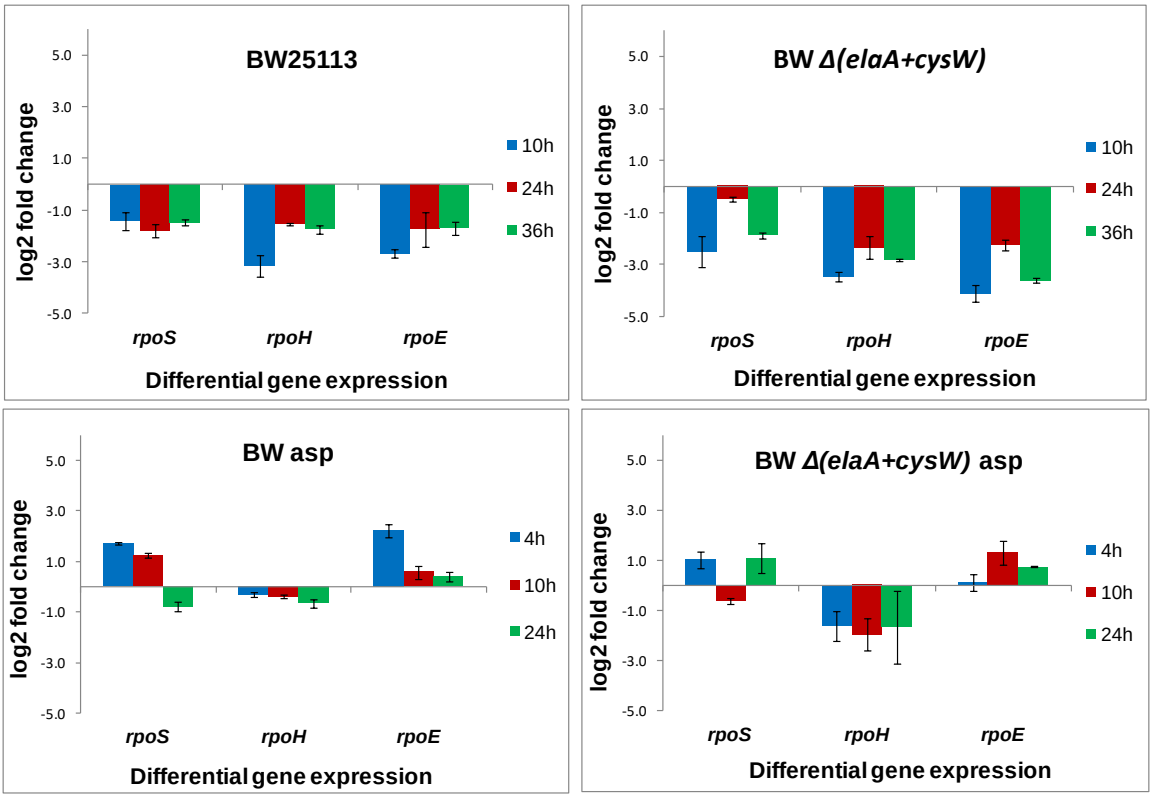

Figure 4.20 Differential mRNA expression levels of *E. coli* sigma factors in control and modified *E. coli* cells under protein producing and non-producing conditions.

***rpoE* (σ^E):** *rpoE* sigma factor is known to get induced in response to extra-cytoplasmic stress by sensing the stress associated with cell wall, cell membrane or change in oxidation state (Roncarati and Scarlato 2017). It also governs the expression of heat shock genes induced in response to accumulation of misfolded or unfolded proteins in *E. coli* cell envelope (Dartigalongue, Missiakas, and Raina 2001). *rpoE* was found to be 1.7 to 4.1 fold downregulated in stationary phase cultures of control and modified cells (Figure 4.20). However, the upregulation of *rpoE* up to 2.2 folds was seen in post induction cultures of L-asp producing control and modified cells, which indicates the induction of extra-cytoplasmic stress possibly due to L-asp accumulation in the periplasmic compartment followed by its export outside the cell.

4.5 Impact of genomic modifications on physiological stress response

System biology tools have allowed us to recognize that the cellular stress response is the key factor responsible for reduced growth rates and compromised cellular health, but this problem has not been addressed comprehensively. Some efforts to identify the features of the physiological stress response were made in this study by mapping the changes in mRNA expression levels of control and modified cultures in stationary phase (Figure 4.21). As shown in this study, the genomic deletions of *ΔelaA* and *ΔcysW* genes improved cellular health and cell viability of *E. coli* BW25113 cells. The positive impact these modifications had on cellular metabolism that resulted in improved phenotype are summarized here:

(i) downregulation of TA module '*mazEF'* that directs cells towards programmed cell death by initiating the synthesis of special stress translational machinery and cell death proteins in modified cells, which is possibly helping these cells to remain in viable state till longer time periods.

(ii) the inability of modified cells to sense starvation induced stress as revealed by the downregulated levels of starvation induced DNA binding protein 'dps' throughout the cultivation period.

(iii) the inability of modified cells to sense oxidative stress revealed by unchanged levels of sRNA *oxyS* which appears only during the conditions of oxidative stress response.

The negative impact of these modifications was seen in terms of increased downregulation of *ompA* in modified cells (4 folds) than control cells (2 folds). The declined levels of *ompA* sRNA are known to affect the stabilities of several RNAs by facilitating RNase E mediated cleavage.

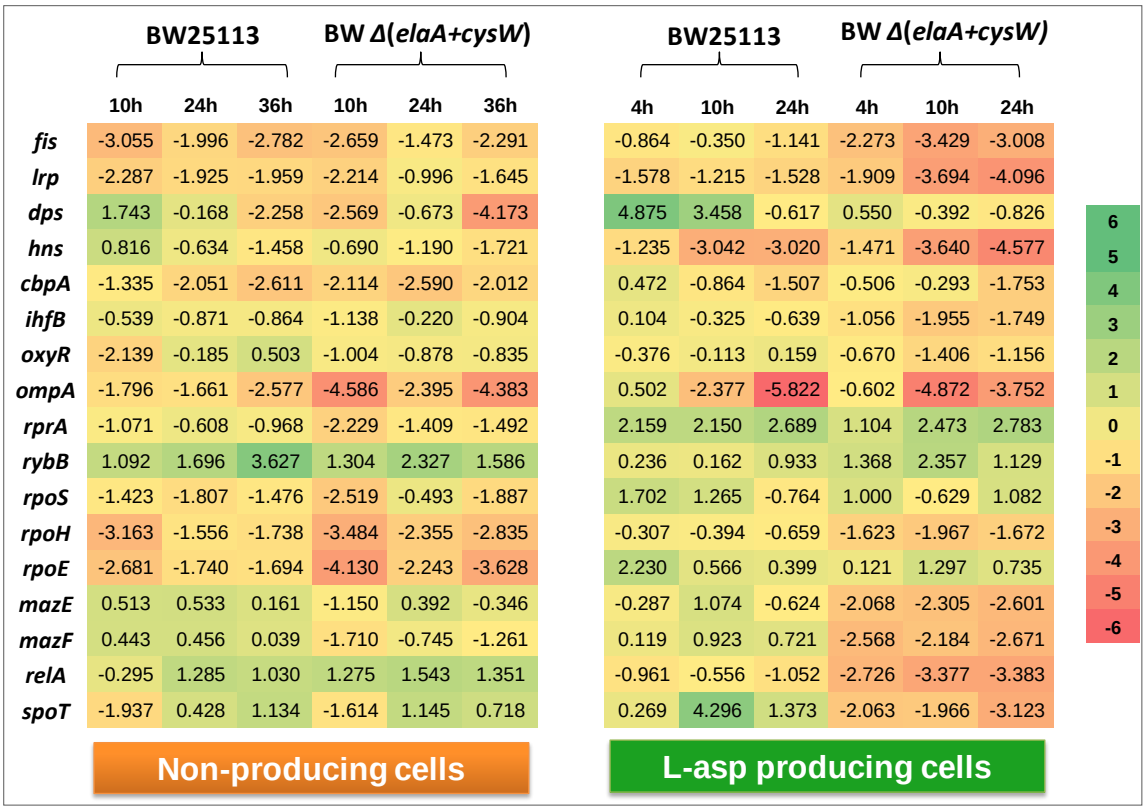

Figure 4.21 Heat maps of differentially expressed genes in control and modified strain under protein producing and non-producing conditions.

4.6 Impact of genomic modifications on recombinant protein induced cellular stress response

The impact of these genomic modifications was greater under recombinant protein producing conditions. Since, the modified cells were better producers of L-asp than control cells; we attempted to find out the reasons behind their enhanced production capability and sustained expression till longer periods in terms of their ability to counter the harmful effects of CSR in production environment. The positive impacts these changes had on countering the effect of CSR are summarized here:

(i) downregulation of *relA* and *spoT* genes of stringent stress response pathway resulting in very low or no (p)pGpp accumulation and hence, absence of nutrient starvation induced stress in L-asp secreting modified cells.

(ii) downregulation of *mazE* and *mazF* genes of programmed cell death pathway indicating the better cellular health and cell viability of modified cells even under protein producing conditions.

(iii) the marginal changes in the levels of starvation induced DNA binding protein 'dps' till 10 hours post induction indicating the absence of starvation induced stress in modified cells.

(iv) downregulated levels of *oxyS* sRNA indicating the absence of oxidative stress.

(v) downregulated levels of *rpoH* sigma factor indicating the absence of heat shock response in modified cells.

The negative effect of these modifications was similar to that found under the conditions of physiological stress i.e. increased downregulation of sRNA *ompA* in modified cells than control cells under protein producing conditions (Figure 4.21). All changes observed here in modified strain clearly showed that cell is not able to mount the cellular stress response when *elaA* and *cysW* genes are knocked-out.

4.7 Proteomic profiling (label free LC-MS/MS approach)

Proteome profiling is a powerful tool for deciphering the underlying mechanisms behind cellular processes because the metabolic fluxes inside cells are directly regulated by differential levels of protein expression. The transcriptomic and proteomic profiles can differ significantly from each-other due to differences in translational efficiencies and stabilities of proteins within the cell. Therefore, transcriptomic analysis does not always provide the accurate picture of cellular physiology. We performed proteomic investigations of stationary phase samples for control and modified cells under protein producing and non-producing conditions as shown in Table 4.4. For uninduced cultures, the proteomic levels of early stationary phase (10 hour) were compared with late stationary phase samples (24 hour & 36 hour), and for induced cultures the 4 hour post induction cultures were compared with 10 hour and 24 hour post induction cultures. As done for transcriptomic studies, we focused on changes in the protein expression levels of NAPs, transcriptional regulator Lrp and sigma factors, which are key players of stationary phase metabolic regulation. Since we were interested in unraveling the reasons behind improved cellular health of modified cells in addition to their enhanced protein production capability, we extended this analysis to proteins belonging to the DNA damage sensing/repair and protein degradation pathways. The mass spectrometry (MS) based label-free quantitative proteomic approach was used for studying differential protein expression in our biological samples. Sample preparation and LC-MS/MS analysis was done as described in materials and methods.

Strain	Reference point	Time-points
BW25113	10 h	24 h, 36 h
BW *Δ(elaA+cysW)*	10 h	24 h, 36 h
BW L-asp	4 h (post induction)	10 h, 24 h (post induction)
BW *Δ(elaA+cysW)* L-asp	4 h (post induction)	10 h, 24 h (post induction)

Table 4.4 Details of samples collected for proteomic studies.

4.7.1 Proteomic profiling of uninduced cultures

The economics of protein synthesis during prolonged stationary phase becomes important as cells prioritize those biological processes which are essential for sustaining cellular growth and survival under adverse conditions. The analysis of stationary phase cultures of control cells revealed significant upregulation of NAPs during transition from early stationary phase to late stationary phase. The DNA binding protein Dps which help in nucleoid condensation was found to be 6 to 8 folds upregulated at 24th and 36th hour (Figure 4.22). An increase in Dps concentrations during stationary phase (30 hour to 4 days) has also been reported previously in proteome studies of *E. coli* BW25113 strain (Soufi *et al*. 2015). The CbpA and IhfA protein expression levels increased by 2 to 4 fold during prolonged stationary phase. A sharp increase in IhfB levels at 24th hour (2.7 fold) and 36th hour (7.4 fold) compared to 10th hour culture was observed. However, H-NS levels remained unchanged between 10th and 24th hour followed by a 3.2 fold increase at 36 hours. Lrp (leucine responsive regulatory protein) which helps cells in adapting to nutrient starvation conditions was also 4 fold and 6 fold upregulated at 24th and 36th hour compared to early stationary phase cultures. The relative increase in expression levels of proteins involved in maintain nucleoid structure and cellular adaptation during stationary phase indicates the presence of starvation induced stress that generally leads to alterations in DNA structure and topology. The upregulation of NAPs is probably a defense mechanism of cell to prevent DNA damage.

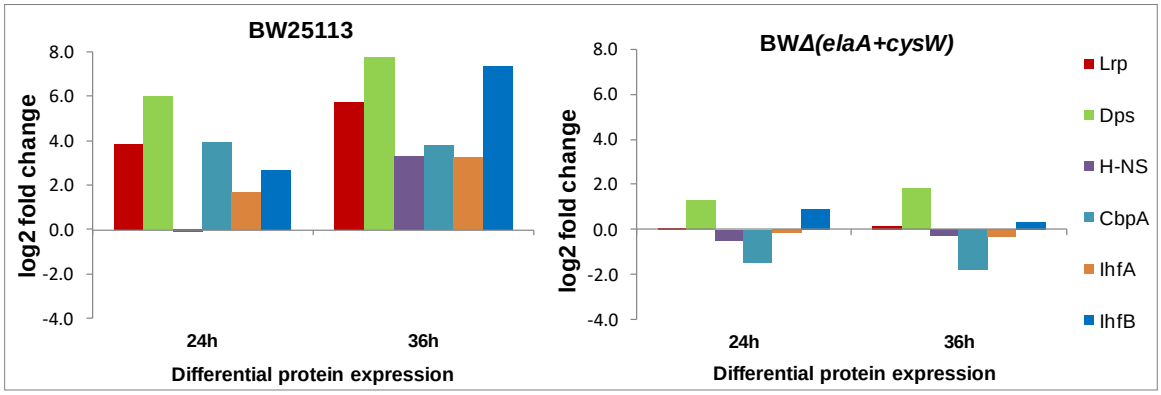

Figure 4.22 Differential protein expression levels of global regulators in control and modified *E. coli* cells.

Interestingly the levels of NAPs did not change significantly in modified cells during the entire stationary phase. Only a slight upregulation in Dps levels (1.2-1.8 fold) was seen at 24 hour and 36 hour. The downregulation of CbpA protein expression levels by 1.4 to 1.8 folds was also observed in stationary phase. There was no change in protein expression levels of Lrp during the stationary phase. These findings suggests that modified cells are able to counter the effect of starvation induced stress and able to maintain themselves in same physiological state similar to the early stationary phase. The downregulation of DNA repair protein RecA and unchanged levels of Topoisomerase I at 24 hour and 36 hour in modified cells, which are otherwise upregulated (2.5-5 fold) in control culture (Figure 4.23), is also in accordance with these findings.

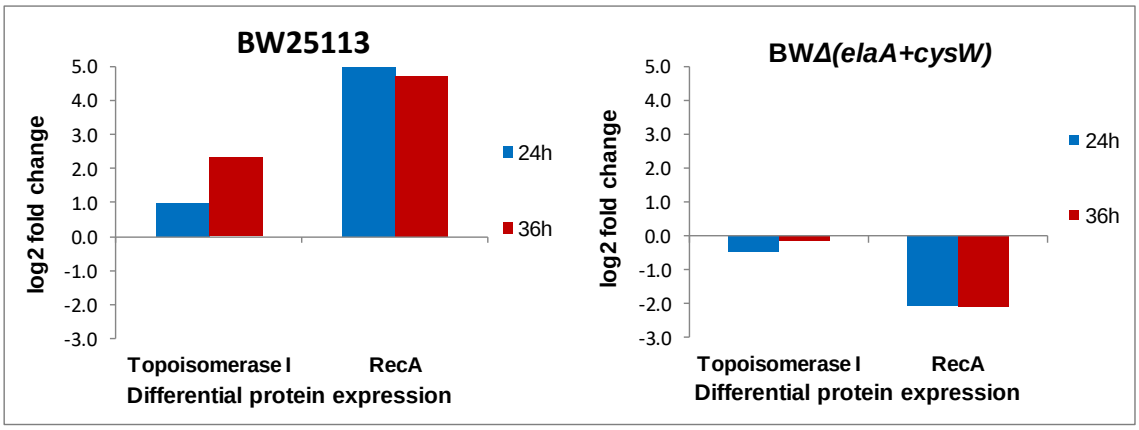

Figure 4.23 Differential protein expression levels of DNA damage sensing/repair proteins in control and modified *E. coli* cells.

Another set of proteins that play a major role in generating stationary phase stress response is shown in Figure 4.24. The RNA polymerase sigma factor RpoS was found to be 2.3 fold upregulated at 24th hour in comparison to 10th hour in control cells (Figure 4.24). RpoS levels were not detected at 36 hours in control cultures, whereas RpoD, which promotes the expression of genes related to cell survival, was found to be 2.8 fold upregulated in control culture at 36 hours. The binding of RpoS, which competes with RpoD for binding with core RNAP, is dependent on interaction of RpoD with another protein called Rsd protein (regulator of sigma D). Rsd protein is known to get accumulated during stationary phase, which by sequestering RpoD increases the levels of

RNA polymerase containing RpoS (Jishage and Ishihama 1998; Mitchell *et al*. 2007). Rsd was found to be 1.8 fold upregulated in control cells at 36th hour. The RpoS factor-binding protein Crl, which activates the expression of RpoS related genes, was also detected in control culture at 36th hour. The presence of Rsd and Crl at 36 hours in control cells confirms the presence of RpoS mediated stationary phase stress response. However, in modified cells both RpoS and Rsd remained 2 to 3 fold downregulated in stationary pahse. The unchanged levels of RpoD in modified cells indicate the similar expression of house-keeping genes, as it was during early stationary phase i.e. at 10 hours. These results are confirming the absence of RpoS mediated stationary phase stress in modified cells.

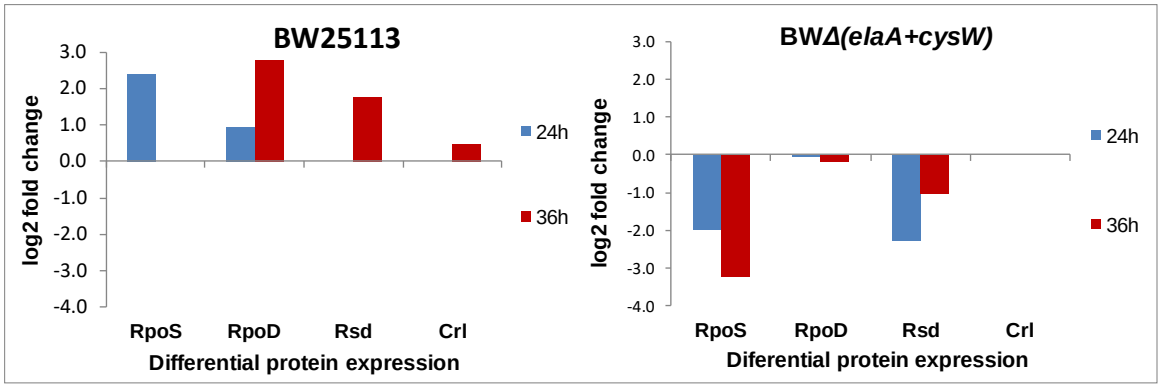

Figure 4.24 Differential protein expression levels of transcriptional regulators in control and modified *E. coli* cells.

The cellular rate of protein synthesis is not the only detrimental factor of protein expression levels inside cells. Instead, the cellular rates of protein degradation also contribute to major extent in controlling the levels of protein inside cells. Most of the regulatory proteins like TFs have rapid turnover inside cells facilitated by protein degradation pathways, which allow them to change their concentrations rapidly in response to various stress signals. We observed the relative increase in expression levels of a group of proteins that play a major role in protein degradation in response to a stress signals. The proteins that were upregulated at 24 h in control cells were periplasmic serine endoproteases (DegP: 2.86 fold; DegQ: 5 fold) and lon protease (2 fold) in control cells (Figure 4.25 (a)). The proteins of Clp protease system, aromatic amino acid

aminotransferase and aminobutyrate aminotransferase, periplasmic serine endoproteases and lon proteases were found to be upregulated at 36 hours in control cells indicating the activation of most of the protein degradation pathways in control cells during late stationary phase. In contrast to this, the modified showed downregulation of all these proteins throughout the stationary phase (Figure 4.25(b)) which is again pointing towards their ability to not sense physiological stress signals. This could be one among the reasons for better cellular health of the modified strain.

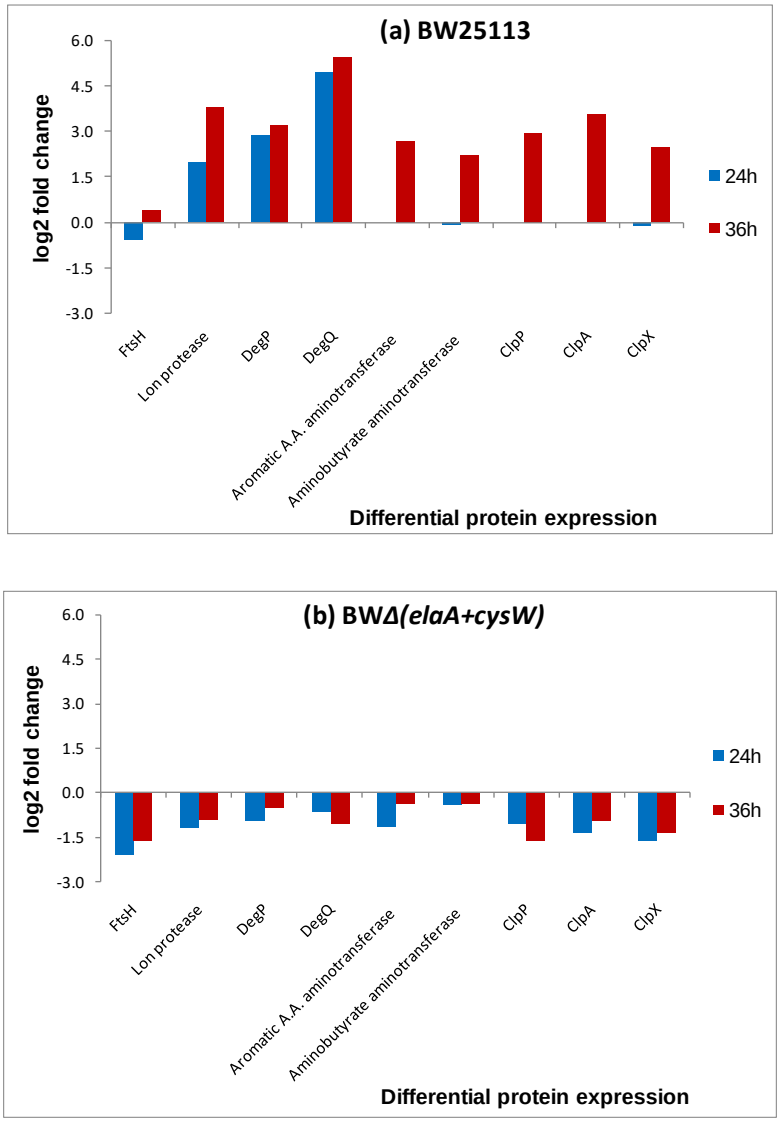

Figure 4.25 Differential expression levels of proteins belonging to the protein degradation pathway in (a) control and (b) modified *E. coli* cells.

4.7.2 Proteomic profiling of induced cultures

As discussed in previous sections, recombinant protein production elicits a global cellular stress response which has a widespread effect on cellular physiology. Since we were interested in finding out the reasons behind better cellular health and sustained expression capability of modified cells till longer hours, we compared the proteome changes for later hours of production with 4 hour post induction cultures (Table 4.5). Unlike uninduced cultures, the changes in protein expression levels were not very significant for L-asp producing control cultures, apart from Dps whose protein levels raised by 1.9 fold at 24 hours (Figure 4.26). In L-asp expressing modified cells, the proteins that were upregulated by 1.5-2 folds at 24 hours were Dps, CbpA and IhfB. The trend was different from the transcriptomic profiles of global regulators, where other than Dps and CbpA the increased downregulation in mRNA expression levels at 10 and 24 hours compared to 4 hour post induction sample was seen (Figure 4.17). The relative protein expression levels of RecA and Topoisomerase I, which are required for DNA repair and maintenance were 1.5 folds upregulated in modified cells, however they did not change very much in control cells. This implies that recombinant protein mediated stress perhaps acts via some different mechanism than that of physiological stress.

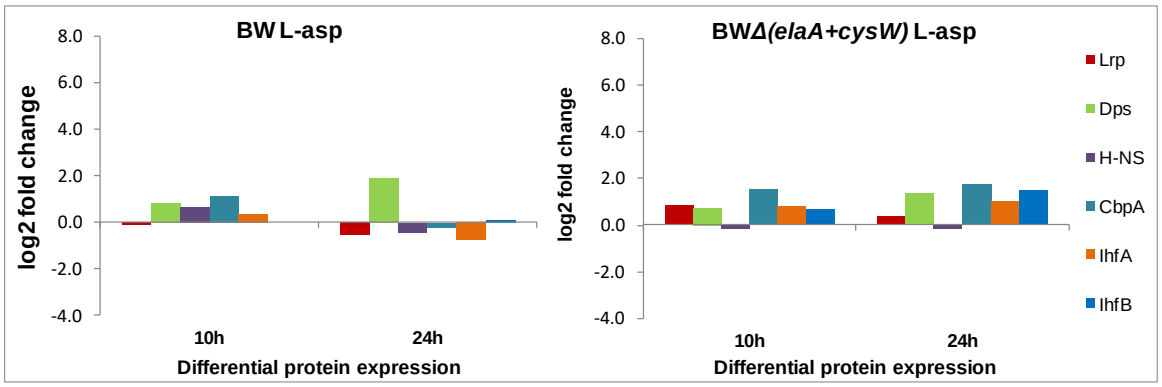

Figure 4.26 Differential protein expression levels of global regulators in control and modified *E. coli* cells induced for L-asp expression.

The RNA polymerase sigma factor RpoS was found to be downregulated at 24th hour in L-asp expressing control cells (Figure 4.27). Somewhat similar trend for *rpoS* downregulation was seen previously in transcriptomic profiles, where it was

downregulated at 24th hour but was upregulated in 4 hour post induction cultures of L-asp expressing control. However, in modified cells RpoS was upregulated by 1.7 folds in 24 hour post induction cultures. Given the complex regulation of RpoS and RpoD, we could not assign any reason to its downregulated levels in recombinant protein producing control cells.

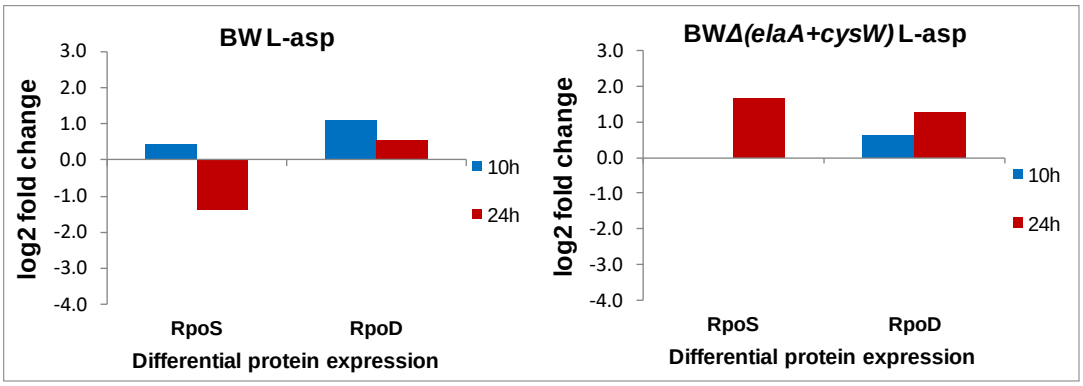

Figure 4.27 Differential protein expression levels of transcriptional regulators in control and modified *E. coli* cells induced for L-asp expression.

The proteins Lon protease, periplasmic serine endoproteases and ClpX proteases were found to be downregulated in L-asp expressing control cells in stationary phase. For modified cultures expressing recombinant protein, marginal changes in protein expression levels of lon protease and Clp protease family were found. The positive change observed in L-asp expressing cultures of modified cells was 3 folds upregulation of aminobutyrate aminotransferase (Figure 4.28). This enzyme plays a key role in gamma-aminobutyrate (GABA) degradation pathway by facilitating the amino group transfer from GABA to α-ketaglutarate to yield glutamate and succinic semialdehyde, thereby enabling cells to use GABA as a nitrogen source for their growth (Schneider *et al*. 2002; Liu *et al*. 2005). The upregulation of aminobutyrate aminotransferase in modified cells indicate that they are able to utilize the amino acids generated by GABA degradation for their growth and maintenance.

The marginal differences in expression levels of proteins between control and modified cells under recombinant protein producing conditions prevented us from drawing firm conclusions from proteomic analysis. Perhaps the proteomic investigation of energy

metabolism and redox regulation pathways might be able to locate key reasons behind enhanced and sustained expression capability of modified cells compared to control cultures.

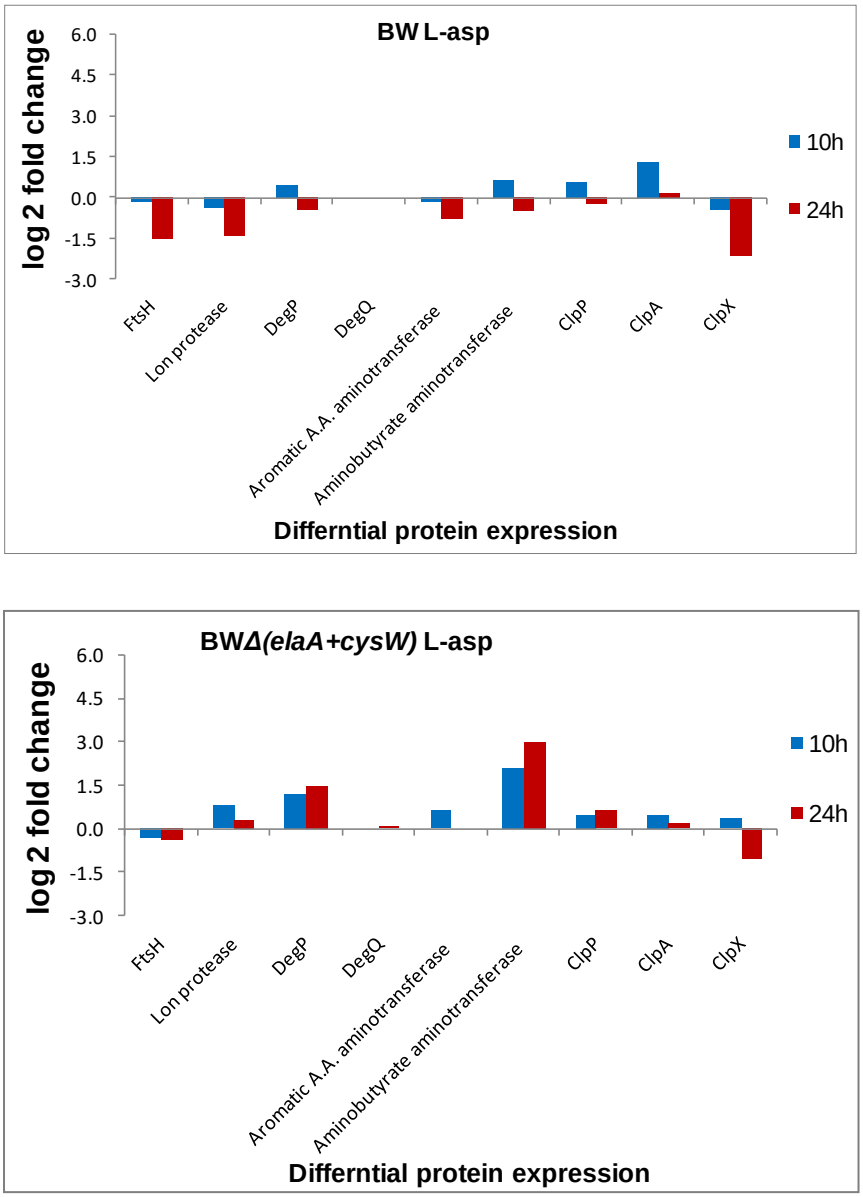

Figure 4.28 Differential expression levels of proteins belonging to the protein degradation pathway in control and modified *E. coli* cells induced for L-asp expression.

5. Conclusion

The inherent complexity of the *E. coli* transcriptional regulatory networks and the gaps in our knowledge precludes the possibility of a perfectly rational host design. Clearly much more clarity is needed before we can accurately predict the dynamic response of the cell to stressful conditions, and more importantly modulate this CSR to achieve our desired goals. We not only need a global view of the changes in gene metabolic networks that accompany the mounting of a CSR upon induction of recombinant protein synthesis but also understand how these changes trigger the CSR in the first place. Explicating this bi-directional information flow within the cell would truly enable us to design superior host platforms which would enhance as well as sustain recombinant protein expression for longer periods.

We initiated this work with leads obtained from previous transcriptomic studies of a recombinant protein expressing double knock-out strain that was shown to have higher protein expression capabilities compared to its unmodified counterpart. These improved expression capabilities of the double knock-out was associated with increased expression levels of genes that belonged to energy metabolism, transcription, translation and some global regulators like *lrp* and *dps*. These knock-outs also imposed some negative effects on the host in terms of lowered expression levels of substrate utilization genes, thus adding a cost associated with these modifications as well as imposing a constraint on the upper limit to which protein production can be achieved. Therefore, the first objective of this study was to remove this constraint by designing a better host expression platform with augmented substrate uptake capabilities so that protein expression could be enhanced and also sustained for longer periods. The second goal was to identify the signaling pathways leading to CSR that are disrupted in these double knock-out mutants, resulting in better performance of these modified strains.

We engineered the double knock-out strain for improved substrate uptake by supplementing the expression levels of glycerol uptake genes (*glpD* and *glpK)* in a single plasmid based system. Major improvements in protein production were obtained simply with basal level expression of these glycerol uptake genes resulting in a 2.3 fold increase

in productivity per unit biomass. This was achieved by glycerol pulsing in the culture medium to ensure the continuous availability of glycerol. These results demonstrated the potential of synergistically combining two different strategies viz. knock-outs along with supplementation of gene copies for the critically down-regulated genes.

The other interesting result was the five-fold improvement in cell viability of the double knock-out mutants in stationary phase (24 hours) compared to its native counterpart hinting towards an amelioration of the CSR leading to an improvement in cellular health. Also, the double knock-out strain retained its stationary phase protein production capability compared to the control which was conclusively demonstrated by successful induction of sfGFP expression after entering stationary phase. As expected the control strain was not able to produce sfGFP upon induction in the stationary phase.

Since, the double knock-out mutant turned out to be a better performer not only in terms of recombinant protein production but also in terms of enhanced stationary phase survival, we did transcriptomic and proteomic analysis of stationary phase cultures to find out the factors that are responsible for eliciting the CSR but got disrupted by these genomic deletions. The key signaling pathways that were found to be deactivated upon these modifications were: (a) stringent stress response pathway due to down-regulation of *relA* and *spoT* genes that leads to (p)pGpp accumulation inside cells upon sensing nutrient starvation and (b) programmed cell death pathway due to inactivation of TA module *mazEF*. The double knock-outs were also not able to sense heat shock type stress which was reflected as increased down-regulation of *rpoH* inside the cells and the marginal changes in level of global regulator Dps also showed their inability to sense stationary phase starvation conditions. Changes at the proteomic level included the ability to utilize nitrogen generated from the GABA degradation pathway for sustaining cellular metabolism and increased levels of DNA repair protein recA which initiates DNA repair mechanisms in the stationary phase.

From an analysis of the changes at regulatory level, it can be concluded that the competition for intracellular resources upon induction of recombinant protein synthesis has dramatic effects on cellular health of normal unmodified cells. However, the double

knock-out strain was partly able to balance this trade-off of resources between growth and production processes. The inability of the double knock-out strain to sense various stress responses clearly demonstrates the positive effect of these modifications which prevents these cells from fully mounting the CSR upon recombinant protein synthesis. This study thus enabled us to establish the role of these non-essential genes in generating the CSR and thus, despite having no causal connection with the recombinant protein synthesis pathway, it opens up an array of possibilities that will allow us to find new and novel strategies for efficient host design.

Bibliography

Albiniak, Anna M., Cristina F R O Matos, Steven D. Branston, Robert B. Freedman, Eli Keshavarz-Moore, and Colin Robinson. 2013. "High-Level Secretion of a Recombinant Protein to the Culture Medium with a Bacillus Subtilis Twin-Arginine Translocation System in Escherichia Coli." *FEBS Journal* 280 (16): 3810–21. https://doi.org/10.1111/febs.12376.

Almiron, M., A. J. Link, D. Furlong, and R. Kolter. 1992. "A Novel DNA-Binding Protein with Regulatory and Protective Roles in Starved Escherichia Coli." *Genes and Development* 6 (12 B): 2646–54. https://doi.org/10.1101/gad.6.12b.2646.

Angius, Federica, Oana Ilioaia, Amira Amrani, Annabelle Suisse, Lindsay Rosset, Amélie Legrand, Abbas Abou-Hamdan, Marc Uzan, Francesca Zito, and Bruno Miroux. 2018. "A Novel Regulation Mechanism of the T7 RNA Polymerase Based Expression System Improves Overproduction and Folding of Membrane Proteins." *Scientific Reports* 8 (1): 8572. https://doi.org/10.1038/s41598-018-26668-y.

Anna Weglenska, Biji Jacob, Agnieszka Sirko. 1996. "Transcriptional Pattern of Escherichia Coli IhfB (HimD) Gene Expression" 181: 85–88.

AP, Pugsley. 1993. "The Complete General Secretory Pathway in Gram Negative Bacteria." *Trends in Microbiology* 57 (1): 50–108. https://doi.org/10.1016/0966-842X(93)90043-Q.

Aristidou, Aristos A., Ka -Yiu San, and George N. Bennett. 1995. "Metabolic Engineering of Escherichia Coli To Enhance Recombinant Protein Production through Acetate Reduction." *Biotechnology Progress* 11 (4): 475–78. https://doi.org/10.1021/bp00034a019.

Aslam, Bilal, Madiha Basit, Muhammad Atif Nisar, Mohsin Khurshid, and Muhammad Hidayat Rasool. 2017. "Proteomics: Technologies and Their Applications." *Journal of Chromatographic Science* 55 (2): 182–96.

https://doi.org/10.1093/chromsci/bmw167.

Atlung, Tove, and Hanne Ingmer. 1997. "H-NS: A Modulator of Environmentally Regulated Gene Expression." *Molecular Microbiology* 24 (1): 7–17. https://doi.org/10.1046/j.1365-2958.1997.3151679.x.

Azam, Talukder A L I, Akira Iwata, Akiko Nishimura, and Susumu Ueda. 1999. "Growth Phase-Dependent Variation in Protein Composition of the Escherichia Coli Nucleoid Protein" 181 (20): 6361–70.

Bagos, Pantelis G., Elisanthi P. Nikolaou, Theodore D. Liakopoulos, and Konstantinos D. Tsirigos. 2010. "Combined Prediction of Tat and Sec Signal Peptides with Hidden Markov Models." *Bioinformatics* 26 (22): 2811–17. https://doi.org/10.1093/bioinformatics/btq530.

Balzer, Simone, Veronika Kucharova, Judith Megerle, Rahmi Lale, Trygve Brautaset, and Svein Valla. 2013. "A Comparative Analysis of the Properties of Regulated Promoter Systems Commonly Used for Recombinant Gene Expression in Escherichia Coli." *Microbial Cell Factories* 12 (1): 1. https://doi.org/10.1186/1475-2859-12-26.

Baneyx, Francois. 1999. "Recombinant Protein Expression in Escherichia Coli." *Current Opinion in Biotechnology* 10: 411–21.

Battesti, Aurelia, Nadim Majdalani, and Susan Gottesman. 2010. " The RpoS-Mediated General Stress Response in Escherichia Coli ." *Annual Review of Microbiology* 65 (1): 189–213. https://doi.org/10.1146/annurev-micro-090110-102946.

Battesti, Aurelia, Nadim Majdalani, Susan Gottesman. 2015. "Stress Sigma Factor RpoS Degradation and Translation Are Sensitive to the State of Central Metabolism." *Proceedings of the National Academy of Sciences* 112 (16): 5159–64. https://doi.org/10.1073/pnas.1504639112.

Becker, Gisela, Eberhard Klauck, and Regine Hengge-Aronis. 2000. "The Response Regulator RssB, a Recognition Factor for S." *Molecular Microbiology* 35 (3):

657–66.

Bendtsen, Jannick Dyrløv, Henrik Nielsen, David Widdick, Tracy Palmer, and Søren Brunak. 2005. "Prediction of Twin-Arginine Signal Peptides." *BMC Bioinformatics* 6: 167. https://doi.org/10.1186/1471-2105-6-167.

Bentley, W E, N Mirjalili, D C Andersen, R H Davis, and D S Kompala. 1990. "Plasmid-Encoded Protein: The Principal Factor in the 'Metabolic Burden' Associated with Recombinant Bacteria. Biotechnology Bioengineering, 1990." *Biotechnol Bioeng* 35:668–81. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19215036.

Bessette, P H, F Aslund, J Beckwith, and G Georgiou. 1999. "Efficient Folding of Proteins with Multiple Disulfide Bonds in the Escherichia Coli Cytoplasm." *Proceedings of the National Academy of Sciences of the United States of America* 96(24):13703–8. http://www.ncbi.nlm.nih.gov/pubmed/10570136%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC24128.

Bhattacharya, Palash, Gaurav Pandey, Poonam Srivastava, and Krishna Jyoti Mukherjee. 2005. "Combined Effect of Protein Fusion and Signal Sequence Greatly Enhances the Production of Recombinant Human GM-CSF in Escherichia Coli." *Molecular Biotechnology* 30 (2): 103–15. https://doi.org/10.1385/MB:30:2:103.

Blommel, Paul G., and Brian G. Fox. 2007. "A Combined Approach to Improving Large-Scale Production of Tobacco Etch Virus Protease." *Protein Expression and Purification* 55 (1): 53–68. https://doi.org/10.1016/j.pep.2007.04.013.

Bonthron, David T. 1990. "L-Asparaginase II of Escherichia Coli K-12: Cloning, Mapping and Sequencing of the AnsB Gene." *Gene* 91 (1): 101–5. https://doi.org/10.1016/0378-1119(90)90168-Q.

Bougdour, A., and S. Gottesman. 2007. "PpGpp Regulation of RpoS Degradation via

Anti-Adaptor Protein IraP." *Proceedings of the National Academy of Sciences* 104 (31): 12896–901. https://doi.org/10.1073/pnas.0705561104.

Bougdour, Alexandre, Christofer Cunning, Patrick Jean Baptiste, Thomas Elliott, and Susan Gottesman. 2008. "Multiple Pathways for Regulation of ΣS (RpoS) Stability in Escherichia Coli via the Action of Multiple Anti-Adaptors." *Molecular Microbiology* 68 (2): 298–313. https://doi.org/10.1111/j.1365-2958.2008.06146.x.

Bougdour, Alexandre, Sue Wickner, and Susan Gottesman. 2006. "Modulating RssB Activity: IraP, a Novel Regulator of ΣS Stability in Escherichia Coli." *Genes and Development* 20 (7): 884–97. https://doi.org/10.1101/gad.1400306.

Bouveret, Aurelia Battesti and Emmanuelle. 2009. "Bacteria Possessing Two RelA / SpoT-Like Proteins Have Evolved a Specific Stringent Response Involving the Acyl Carrier." *Journal of Bacteriology* 191 (2): 616–24. https://doi.org/10.1128/JB.01195-08.

Burgess-Brown, Nicola A., Sujata Sharma, Frank Sobott, Christoph Loenarz, Udo Oppermann, and Opher Gileadi. 2008. "Codon Optimization Can Improve Expression of Human Genes in Escherichia Coli: A Multi-Gene Study." *Protein Expression and Purification* 59 (1): 94–102. https://doi.org/10.1016/j.pep.2008.01.008.

Burgos, Hector L., Kevin O'Connor, Patricia Sanchez-Vazquez, and Richard L. Gourse. 2017. "Roles of Transcriptional and Translational Control Mechanisms in Regulation of Ribosomal Protein Synthesis in Escherichia Coli." *Journal of Bacteriology* 199 (21): 1–14. https://doi.org/10.1128/jb.00407-17.

Calogero, R. A., C. L. Pon, M. A. Canonaco, and C. O. Gualerzi. 2006. "Selection of the MRNA Translation Initiation Region by Escherichia Coli Ribosomes." *Proceedings of the National Academy of Sciences* 85 (17): 6427–31. https://doi.org/10.1073/pnas.85.17.6427.

Care, S., C. Bignon, M. C. Pelissier, E. Blanc, B. Canard, and Bruno Coutard. 2008. "The Translation of Recombinant Proteins in E. Coli Can Be Improved by in Silico Generating and Screening Random Libraries of a -70/+96 MRNA Region with Respect to the Translation Initiation Codon." *Nucleic Acids Research* 36 (1): 2–7. https://doi.org/10.1093/nar/gkm1097.

Carneiro, Sónia, Eugénio C. Ferreira, and Isabel Rocha. 2013. "Metabolic Responses to Recombinant Bioprocesses in Escherichia Coli." *Journal of Biotechnology* 164 (3): 396–408. https://doi.org/10.1016/j.jbiotec.2012.08.026.

Chad A Galloway, Mark P Sowden, Harold C Smith. 2003. "Increasing the Yeild of Soluble Recombinant Protein Expressed in E. Coli by Induction during Late Log Phase." *Biotechniques* 34: 524–30.

Chae, Young Kee, Seol Hyun Kim, and John L. Markley. 2017. "Relationship between Recombinant Protein Expression and Host Metabolome as Determined by Two-Dimensional NMR Spectroscopy." *PLoS ONE* 12 (5): 1–12. https://doi.org/10.1371/journal.pone.0177233.

Chang, A. C Y, and S. N. Cohen. 1978. "Construction and Characterization of Amplifiable Multicopy DNA Cloning Vehicles Derived from the P15A Cryptic Miniplasmid." *Journal of Bacteriology* 134 (3): 1141–56.

Chen, Ning, Jin Huang, Zhi Bin Feng, Lei Yu, Qing Yang Xu, and Ting Yi Wen. 2009. "Optimization of Fermentation Conditions for the Biosynthesis of L-Threonine by Escherichia Coli." *Applied Biochemistry and Biotechnology* 158 (3): 595–604. https://doi.org/10.1007/s12010-008-8385-y.

CHIEN-DER LEE, HUI-CHIEN SUN, SU-MING HU, CHING-FENG CHIU, ATTHACHAI HOMHUAN, SHU-MEI LIANG, CHIH-HSIANG LENG, AND TING-FANG WANG. 2008. "An Improved SUMO Fusion Protein System for Effective Production of Native Proteins." *Protein Science* 12: 1241–48. https://doi.org/10.1110/ps.035188.108.6.

Choi, J. H., and S. Y. Lee. 2004. "Secretory and Extracellular Production of Recombinant Proteins Using Escherichia Coli." *Applied Microbiology and Biotechnology* 64 (5): 625–35. https://doi.org/10.1007/s00253-004-1559-9.

Choi, Jong Hyun, Sang Jun Lee, Seok Jae Lee, and Sang Yup Lee. 2003. "Enhanced Production of Insulin-like Growth Factor I Fusion Protein in Escherichia Coli by Coexpression of the down-Regulated Genes Identified by Transcriptome Profiling." *Applied and Environmental Microbiology* 69 (8): 4737–42. https://doi.org/10.1128/AEM.69.8.4737.

Choi, Tae-jin, and Temesgen Tola. 2018. "High Level Expression and Purification of Recombinant Flounder Growth Hormone in E . Coli." *Journal of Genetic Engineering and Biotechnology* 16 (2): 347–55. https://doi.org/10.1016/j.jgeb.2018.03.006.

Choi, Young J., Lyne Morel, Teffanie Le François, Denis Bourque, Bourget Lucie, Denis Groleau, Bernard Massie, and Carlos B. Miguez. 2010. "Novel, Versatile, and Tightly Regulated Expression System for Escherichia Coli Strains." *Applied and Environmental Microbiology* 76 (15): 5058–66. https://doi.org/10.1128/AEM.00413-10.

Clough, Timothy, Melissa Key, Ilka Ott, Susanne Ragg, Gunther Schadow, and Olga Vitek. 2009. "Protein Quantification in Label-Free LC-MS Experiments." *Journal of Proteome Research* 8 (11): 5275–84. https://doi.org/10.1021/pr900610q.

Colin, Vernica Leticia, Analía Rodríguez, and Héctor Antonio Cristbal. 2011. "The Role of Synthetic Biology in the Design of Microbial Cell Factories for Biofuel Production." *Journal of Biomedicine and Biotechnology* 2011 (601834). https://doi.org/10.1155/2011/601834.

Cornelis, Pierre. 2000. "Expressing Genes in Different Escherichia Coli Compartments." *Current Opinion in Biotechnology* 11 (5): 450–54.

Costanzo, Alessandra, Herve Nicoloff, Sarah E. Barchinger, Amy B. Banta, Richard L.

Gourse, and Sarah E. Ades. 2008. “PpGpp and DksA Likely Regulate the Activity of the Extracytoplasmic Stress Factor ΣE in Escherichia Coli by Both Direct and Indirect Mechanisms.” *Molecular Microbiology* 67 (3): 619–32. https://doi.org/10.1111/j.1365-2958.2007.06072.x.

Cox, Jürgen, Marco Y Hein, Christian A Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. 2014. “Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ.” *Molecular and Cellular Proteomics* 13 (9): 2513–26. https://doi.org/10.1074/mcp.M113.031591.

Dalebroux, Zachary D., and Michele S. Swanson. 2012. “PpGpp: Magic beyond RNA Polymerase.” *Nature Reviews Microbiology* 10 (3): 203–12. https://doi.org/10.1038/nrmicro2720.

Dartigalongue, Claire, Dominique Missiakas, and Satish Raina. 2001. “Characterization of the Escherichia Coli ΣE Regulon.” *Journal of Biological Chemistry* 276 (24): 20866–75. https://doi.org/10.1074/jbc.M100464200.

Dilks, Kieran, R. Wesley Rose, Enno Hartmann, and Mechthild Pohlschröder. 2003. “Prokaryotic Utilization of the Twin-Arginine Translocation Pathway: A Genomic Survey.” *Journal of Bacteriology* 185 (4): 1478–83. https://doi.org/10.1128/JB.185.4.1478-1483.2003.

Dittrich, Cheryl R., Ravishankar V. Vadali, George N. Bennett, and Ka Yiu San. 2005. “Redistribution of Metabolic Fluxes in the Central Aerobic Metabolic Pathway of E. Coli Mutant Strains with Deletion of the AckA-Pta and PoxB Pathways for the Synthesis of Isoamyl Acetate.” *Biotechnology Progress* 21 (2): 627–31. https://doi.org/10.1021/bp049730r.

Dürrschmid, Karin, Helga Reischer, Wolfgang Schmidt-Heck, Thomas Hrebicek, Reinhard Guthke, Andreas Rizzi, and Karl Bayer. 2008. “Monitoring of Transcriptome and Proteome Profiles to Investigate the Cellular Response of E. Coli towards Recombinant Protein Expression under Defined Chemostat

Conditions." *Journal of Biotechnology* 135 (1): 34–44. https://doi.org/10.1016/j.jbiotec.2008.02.013.

Eiteman, Mark A., and Elliot Altman. 2006. "Overcoming Acetate in Escherichia Coli Recombinant Protein Fermentations." *Trends in Biotechnology* 24 (11): 530–36. https://doi.org/10.1016/j.tibtech.2006.09.001.

Engelberg-kulka, Hanna, Ronen Hazan, and Shahar Amitai. 2005. "MazEF : A Chromosomal Toxin-Antitoxin Module That Triggers Programmed Cell Death in Bacteria." https://doi.org/10.1242/jcs.02619.

Epp, Otto, Wolfgang Steigemann, Helmut Formanek, and Robert Huber. 1971. "Crystallographic Evidence for the Tetrameric Subunit Structure of L-Asparaginase from Escherichia Coli." *European Journal of Biochemistry* 20 (3): 432–37. https://doi.org/10.1111/j.1432-1033.1971.tb01410.x.

Fang, Xin, Anand Sastry, Nathan Mih, Donghyuk Kim, Justin Tan, James T. Yurkovich, Colton J. Lloyd, Ye Gao, Laurence Yang, and Bernhard O. Palsson. 2017. " Global Transcriptional Regulatory Network for Escherichia Coli Robustly Connects Gene Expression to Transcription Factor Activities ." *Proceedings of the National Academy of Sciences* 114 (38): 10286–91. https://doi.org/10.1073/pnas.1702581114.

Faria, J. P., R. Overbeek, F. Xia, M. Rocha, I. Rocha, and C. S. Henry. 2014. "Genome-Scale Bacterial Transcriptional Regulatory Networks: Reconstruction and Integrated Analysis with Metabolic Models." *Briefings in Bioinformatics* 15 (4): 592–611. https://doi.org/10.1093/bib/bbs071.

Farrell, Amy, Niaobh McLoughlin, John J. Milne, Ian W. Marison, and Jonathan Bones. 2014. "Application of Multi-Omics Techniques for Bioprocess Design and Optimization in Chinese Hamster Ovary Cells." *Journal of Proteome Research* 13 (7): 3144–59. https://doi.org/10.1021/pr500219b.

Flores, Salvador, Ramón De Anda-Herrera, Guillermo Gosset, and Francisco G. Bolívar.

2004. “Growth-Rate Recovery of Escherichia Coli Cultures Carrying a Multicopy Plasmid, by Engineering of the Pentose-Phosphate Pathway.” *Biotechnology and Bioengineering* 87 (4): 485–94. https://doi.org/10.1002/bit.20137.

Franchini, Alessandro G., Julian Ihssen, and Thomas Egli. 2015. “Effect of Global Regulators RpoS and Cyclic-AMP/CRP on the Catabolome and Transcriptome of Escherichia Coli K12 during Carbon- and Energy-Limited Growth.” *PLoS ONE* 10 (7): 1–24. https://doi.org/10.1371/journal.pone.0133793.

Free, Andrew, and Charles J. Dorman. 1995. “Coupling of Escherichia Coli Hns MRNA Levels to DNA Synthesis by Autoregulation: Implications for Growth Phase Control.” *Molecular Microbiology* 18 (1): 101–13. https://doi.org/10.1111/j.1365-2958.1995.mmi_18010101.x.

Garrett T. Wong, Richard P. Bonocora, Alicia N. Schep, Suzannah M. Beeler, Anna J. Lee Fong, Lauren M. Shull, Lakshmi E. Batachari, Moira Dillon, Ciaran Evans, Carla J. Becker, Eliot C. Bush, Johanna Hardin, Joseph T. Wade, Daniel M. Stoebela. 2017. “Genome-Wide Transcriptional Response to Varying RpoS Levels in Escherichia Coli K-12.” *Journal of Bacteriology* 199 (7): e00755-16.

Ghatak, Sankha, Zachary A. King, Anand Sastry, and Bernhard O. Palsson. 2019. “The Y-Ome Defines the 35% of Escherichia Coli Genes That Lack Experimental Evidence of Function.” *Nucleic Acids Research* 47 (5): 2446–54. https://doi.org/10.1093/nar/gkz030.

Ghosh, Chaitali, Rashmi Gupta, and Krishna J. Mukherjee. 2012. “An Inverse Metabolic Engineering Approach for the Design of an Improved Host Platform for Over-Expression of Recombinant Proteins in Escherichia Coli.” *Microbial Cell Factories* 11: 1–9. https://doi.org/10.1186/1475-2859-11-93.

Giacalone, Matthew J., Angela M. Gentile, Brian T. Lovitt, Neil L. Berkley, Carl W. Gunderson, and Mark W. Surber. 2006. “Toxic Protein Expression in Escherichia Coli Using a Rhamnose-Based Tightly Regulated and Tunable Promoter System.” *BioTechniques* 40 (3): 355–64. https://doi.org/10.2144/000112112.

Gialama, Dimitra, Kalliopi Kostelidou, Myrsini Michou, Dafni Chrysanthi Delivoria, Fragiskos N. Kolisis, and Georgios Skretas. 2017. "Development of Escherichia Coli Strains That Withstand Membrane Protein-Induced Toxicity and Achieve High-Level Recombinant Membrane Protein Production." *ACS Synthetic Biology* 6 (2): 284–300. https://doi.org/10.1021/acssynbio.6b00174.

Gibson, Daniel G, John I Glass, Carole Lartigue, Vladimir N Noskov, Mikkel A Algire, Gwynedd A Benders, Michael G Montague, et al. 2010. "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome." *Science* 329 (May): 52–56. https://doi.org/10.1126/science.1190719.

Glick, Bernard R. 1995. "Metabolic Load and Heterologous Gene Expression." *Biotechnology Advances* 13 (2): 247–61. https://doi.org/10.1016/0734-9750(95)00004-A.

Gopal, Gopal Jee, and Awanish Kumar. 2013. "Strategies for the Production of Recombinant Protein in Escherichia Coli." *Protein Journal* 32 (6): 419–25. https://doi.org/10.1007/s10930-013-9502-5.

Gottesman, Susan. 2005. "Micros for Microbes: Non-Coding Regulatory RNAs in Bacteria." *Trends in Genetics* 21 (7): 399–404. https://doi.org/10.1016/j.tig.2005.05.008.

Gruber, Tanja M., and Carol A. Gross. 2003. "Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space." *Annual Review of Microbiology* 57 (1): 441–66. https://doi.org/10.1146/annurev.micro.57.030502.090913.

Guana, Chu di, Ping Lib, Paul D. Riggsa, and Hiroshi Inouyeb. 1988. "Vectors That Facilitate the Expression and Purification of Foreign Peptides in Escherichia Coli by Fusion to Maltose-Binding Protein." *Gene* 67 (1): 21–30. https://doi.org/10.1016/0378-1119(88)90004-2.

Guzman, L M, D Belin, M J Carson, and J Beckwith. 1995. "Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose

PBAD Promoter." *Journal of Bacteriology* 177 (14): 4121–30. http://www.ncbi.nlm.nih.gov/pubmed/7608087%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC177145.

H A de Boer, L J Comstock, and M Vasser. 1983. "The Tac Promoter: A Functional Hybrid Derived from the Trp and Lac Promoters." *Proceedings of the National Academy of Sciences of the United States of America* 80 (01): 21–25. http://www.pnas.org/content/80/1/21.short.

Haddadin, Fu'ad T., and Sarah W. Harcum. 2005. "Transcriptome Profiles for High-Cell-Density Recombinant and Wild-Type Escherichia Coli." *Biotechnology and Bioengineering* 90 (2): 127–53. https://doi.org/10.1002/bit.20340.

Han, Mee Jung, Ki Jun Jeong, Jong Shin Yoo, and Sang Yup Lee. 2003. "Engineering Escherichia Coli for Increased Productivity of Serine-Rich Proteins Based on Proteome Profiling." *Applied and Environmental Microbiology* 69 (10): 5772–81. https://doi.org/10.1128/AEM.69.10.5772-5781.2003.

Hayer-Hartl, F. Ulrich Hartl and Manajit. 2002. "Molecular Chaperones in the Cytosol : From Nascent Chain to Folded Protein." *Science* 295 (March): 1852–59. https://doi.org/10.1126/science.1068408.

Hefti, Marco H., Caroline J.G. Van Vugt-Van Der Toorn, Ray Dixon, and Jacques Vervoort. 2001. "A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site." *Analytical Biochemistry* 295 (2): 180–85. https://doi.org/10.1006/abio.2001.5214.

Henderson, Charlotte A., Helen A. Vincent, Alessandra Casamento, Carlanne M. Stone, Jack O. Phillips, Peter D. Cary, Frank Sobott, Darren M. Gowers, James E. Taylor, and Anastasia J. Callaghan. 2013. "Hfq Binding Changes the Structure of Escherichia Coli Small Noncoding RNAs OxyS and RprA, Which Are Involved in the Riboregulation of RpoS." *Rna* 19 (8): 1089–1104. https://doi.org/10.1261/rna.034595.112.

Hengge-Aronis, R. 2002. "Recent Insights into the General Stress Response Regulatory Network in *Escherichia Coli*." *Molecular Microbiology* 4 (3): 341–46. https://doi.org/10.1186/s40001-015-0181-x.

Hirasawa, Takashi, and Hiroshi Shimizu. 2016. "Recent Advances in Amino Acid Production by Microbial Cells." *Current Opinion in Biotechnology* 42: 133–46. https://doi.org/10.1016/j.copbio.2016.04.017.

Hirsch, Matthew, and Thomas Elliott. 2005. "Stationary-Phase Regulation of RpoS Translation In." *Microbiology* 187 (21): 7204–13. https://doi.org/10.1128/JB.187.21.7204.

Hoffmann, F, and U Rinas. 2001. "On-Line Estimation of the Metabolic Burden Resulting from Synthesis of Plasmid-Encoded and Heat-Shock Proteins in High-Cell Density Cultures of Recombinant Escherichiacoli." *Biotechnol Bioeng* 76 (4): 333–40. https://doi.org/10.1002/bit.10098.

Hoffmann, Frank, and Ursula Rinas. 2004. "Stress Induced by Recombinant Protein Production in Escherichia Coli." *Advances in Biochemical Engineering/Biotechnology* 89: 73–92. https://doi.org/10.1007/b93994.

Hofmann, Nina, Reinhild Wurm, and Rolf Wagner. 2011. "The E. Coli Anti-Sigma Factor Rsd: Studies on the Specificity and Regulation of Its Expression." *PLoS ONE* 6 (5). https://doi.org/10.1371/journal.pone.0019235.

Hou, Jixian, Yangqiu Liu, Qiang Li, and Jichu Yang. 2013. "High Activity Expression of D-Amino Acid Oxidase in Escherichia Coli by the Protein Expression Rate Optimization." *Protein Expression and Purification* 88 (1): 120–26. https://doi.org/10.1016/j.pep.2012.11.020.

Hu, Jian, Huajun Qin, Fei Philip Gao, and Timothy A Cross. 2011. "A Systematic Assessment of MAture MBP in Membrane Protein Production." *Protein Expr Purif* 80 (1): 34–40. https://doi.org/10.1016/j.pep.2011.06.001.A.

Irving, Sophie E., and Rebecca M. Corrigan. 2018. "Triggering the Stringent Response:

Signals Responsible for Activating (p)PpGpp Synthesis in Bacteria." *Microbiology (United Kingdom)* 164 (3): 268–76. https://doi.org/10.1099/mic.0.000621.

Ishihama Akira. 2000. "Functional Modulation of Escherichia Coli RNA Polymerase." *The Dictionary of Genomics, Transcriptomics and Proteomics* 54: 499–518. https://doi.org/10.1002/9783527678679.dg04662.

Iuchi, S., S. T. Cole, and E. C.C. Lin. 1990. "Multiple Regulatory Elements for the GlpA Operon Encoding Anaerobic Glycerol-3-Phosphate Dehydrogenase and the GlpD Operon Encoding Aerobic Glycerol-3-Phosphate Dehydrogenase in Escherichia Coli: Further Characterization of Respiratory Control." *Journal of Bacteriology* 172 (1): 179–84. https://doi.org/10.1128/jb.172.1.179-184.1990.

Izutsu, Kaori, Akira Wada, and Chieko Wada. 2001. "Expression of Ribosome Modulation Factor (RMF) in Escherichia Coli Requires PpGpp." *Genes to Cells* 6 (8): 665–76. https://doi.org/10.1046/j.1365-2443.2001.00457.x.

Jacobus, Ana Paula, and Jeferson Gross. 2015. "Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia Coli." *PLoS ONE* 10 (3): 1–17. https://doi.org/10.1371/journal.pone.0119221.

Jaffé, Stephen R.P., Benjamin Strutton, Zdenko Levarski, Jagroop Pandhal, and Phillip C. Wright. 2014. "Escherichia Coli as a Glycoprotein Production Host: Recent Developments and Challenges." *Current Opinion in Biotechnology* 30: 205–10. https://doi.org/10.1016/j.copbio.2014.07.006.

Jain, Priyanka. 2018. "Designing Next-Generation Platforms for Recombinant Protein Production by Genome Engineering of Escherichia Coli."

Jamnikar, Uros, Petra Nikolic, Ales Belic, Marjanca Blas, Dominik Gaser, Andrej Francky, Holger Laux, Andrej Blejec, Spela Baebler, and Kristina Gruden. 2015. "Transcriptome Study and Identification of Potential Marker Genes Related to the Stable Expression of Recombinant Proteins in CHO Clones." *BMC Biotechnology*

15: 98. https://doi.org/10.1186/s12896-015-0218-9.

Jenny, Richard J, Kenneth G Mann, and Roger L Lundblad. 2003. "A Critical Review of the Methods for Cleavage of Fusion Proteins with Thrombin and Factor Xa." *Protein Expression and Purification* 31 (1): 1–11. https://doi.org/10.1016/S1046-5928(03)00168-2.

Jensen, Rasmus Bugge, and Kenn Gerdes. 1995. "Programmed Cell Death in Bacteria: Proteic Plasmid Stabilization Systems." *Molecular Microbiology* 17 (2): 205–10. https://doi.org/10.1111/j.1365-2958.1995.mmi_17020205.x.

Jhamb, Kamna, and Debendra K. Sahoo. 2012. "Production of Soluble Recombinant Proteins in Escherichia Coli: Effects of Process Conditions and Chaperone Co-Expression on Cell Growth and Production of Xylanase." *Bioresource Technology* 123: 135–43. https://doi.org/10.1016/j.biortech.2012.07.011.

Jishage, M, and A Ishihama. 1998. "A Stationary Phase Protein in Escherichia Coli with Binding Activity to the Major Sigma Subunit of RNA Polymerase." *Proceedings of the National Academy of Sciences of the United States of America* 95 (9): 4953–58. http://www.ncbi.nlm.nih.gov/pubmed/9560209%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC20194.

Johansen, Jesper, Anders Aamann Rasmussen, Martin Overgaard, and Poul Valentin-Hansen. 2006. "Conserved Small Non-Coding RNAs That Belong to the ΣE Regulon: Role in Down-Regulation of Outer Membrane Proteins." *Journal of Molecular Biology* 364 (1): 1–8. https://doi.org/10.1016/j.jmb.2006.09.004.

Johnson, Irving S. 1983. "Human Insulin from Recombinant DNA Technology Author (s): Irving S . Johnson." *Science* 219 (4585): 632–37.

Julio, By, Augusto Freyre, Ph D Ctr, National Autonomous, Luis Gerardo, Trevino Quintanilla, and Ph D Universidad. 2016. "Analyzing Regulatory Networks in Bacteria Understanding Genetic Organization : The First Steps," 1–8.

K Itakura, T Hirose, R Crea, AD Riggs, HL Heyneker, F Bolivar, HW Boyer. 1977. "Expression in Escherichia Coli of a Chemically Synthesized Gene for the Hormone Somatostatin." *Science* 198 (4321): 1056–63.

Käll, Lukas, Anders Krogh, and Erik L.L. Sonnhammer. 2004. "A Combined Transmembrane Topology and Signal Peptide Prediction Method." *Journal of Molecular Biology* 338 (5): 1027–36. https://doi.org/10.1016/j.jmb.2004.03.016.

Khushoo, Amardeep, Yogender Pal, Bhairab Nath Singh, and K. J. Mukherjee. 2004. "Extracellular Expression and Single Step Purification of Recombinant Escherichia Coli L-Asparaginase II." *Protein Expression and Purification* 38 (1): 29–36. https://doi.org/10.1016/j.pep.2004.07.009.

Kim, Minseung, Navneet Rai, Violeta Zorraquino, and Ilias Tagkopoulos. 2016. "Multi-Omics Integration Accurately Predicts Cellular State in Unexplored Conditions for Escherichia Coli." *Nature Communications* 7 (13090): 1–12. https://doi.org/10.1038/ncomms13090.

Kinter, Michael, and Nicholas E Sherman. 2005. "The Preparation of Protein Digests for Mass Spectrometric Sequencing Experiments." In *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, 147–65. John Wiley & Sons, Inc. https://doi.org/10.1002/0471721980.ch6.

Klauck, E., M. Lingnau, and R. Hengge-Aronis. 2001. "Role of the Response Regulator RssB in (ΣS Recognition and Initiation of (ΣS Proteolysis in Escherichia Coli." *Molecular Microbiology* 40 (6): 1381–90. https://doi.org/10.1046/j.1365-2958.2001.02482.x.

Klumpp, Stefan, Zhongge Zhang, and Terence Hwa. 2009. "Growth Rate-Dependent Global Effects on Gene Expression in Bacteria." *Cell* 139 (7): 1366–75. https://doi.org/10.1016/j.cell.2009.12.001.Growth-rate.

Kolaj, Olga, Stefania Spada, Sylvain Robin, and J. Gerard Gerard. 2009. "Use of Folding Modulators to Improve Heterologous Protein Production in Escherichia Coli."

Microbial Cell Factories 8 (9): 1–17. https://doi.org/10.1186/1475-2859-8-9.

Komar, Anton A. 2009. "A Pause for Thought along the Co-Translational Folding Pathway." *Trends in Biochemical Sciences* 34 (1): 16–24. https://doi.org/10.1016/j.tibs.2008.10.002.

Kucharova, Veronika, Trine Aakvik Strand, Eivind Almaas, Adrian E. Naas, Trygve Brautaset, and Svein Valla. 2013. "Non-Invasive Analysis of Recombinant MRNA Stability in Escherichia Coli by a Combination of Transcriptional Inducer Wash-Out and QRT-PCR." *PLoS ONE* 8 (6): e66429. https://doi.org/10.1371/journal.pone.0066429.

Kültz, Dietmar. 2005. "Molecular and Evolutionary Basis of the Cellular Stress Response." *Annual Review of Physiology* 67 (1): 225–57. https://doi.org/10.1146/annurev.physiol.67.040403.103635.

Laemmli, U. K. 1970. "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." *Nature* 227 (5259): 680–85. https://doi.org/10.1038/227680a0.

Lago, Magali, Véronique Monteil, Thibaut Douche, Julien Guglielmini, Alexis Criscuolo, Corinne Maufrais, Mariette Matondo, and Françoise Norel. 2017. "Proteome Remodelling by the Stress Sigma Factor RpoS/ΣS in Salmonella: Identification of Small Proteins and Evidence for Post-Transcriptional Regulation." *Scientific Reports* 7 (1): 1–15. https://doi.org/10.1038/s41598-017-02362-3.

Landgraf, Jeffrey R., W. U. Jingcai, and Joseph M. Calvo. 1996. "Effects of Nutrition and Growth Rate on Lrp Levels in Escherichia Coli." *Journal of Bacteriology* 178 (23): 6930–36. https://doi.org/10.1128/jb.178.23.6930-6936.1996.

Landini, Paolo, Thomas Egli, Johannes Wolf, and Stephan Lacour. 2014. "SigmaS, a Major Player in the Response to Environmental Stresses in Escherichia Coli: Role, Regulation and Mechanisms of Promoter Recognition." *Environmental Microbiology Reports* 6 (1): 1–13. https://doi.org/10.1111/1758-2229.12112.

Lange, Roland, and Regine Hengge-Aronis. 1994. “The Cellular Concentration of the ??S Subunit of RNA Polymerase in Escherichia Coli Is Controlled at the Levels of Transcription, Translation, and Protein Stability.” *Genes and Development* 8 (13): 1600–1612.

Lara, Alvaro R., Consuelo Vazquez-Limón, Guillermo Gosset, Francisco Bolívar, Agustín López-Munguía, and Octavio T. Ramírez. 2006. “Engineering Escherichia Coli to Improve Culture Performance and Reduce Formation of By-Products during Recombinant Protein Production under Transient Intermittent Anaerobic Conditions.” *Biotechnology and Bioengineering* 94 (6): 1164–75. https://doi.org/10.1002/bit.20954.

Larsen, Simon J., Richard Röttger, Harald H.H.W. Schmidt, and Jan Baumbach. 2019. “E. Coli Gene Regulatory Networks Are Inconsistent with Gene Expression Data.” *Nucleic Acids Research* 47 (1): 85–92. https://doi.org/10.1093/nar/gky1176.

Lesley, Scott A, Jim Graziano, Charles Y Cho, Mark W Knuth, and Heath E Klock. 2002. “Gene Expression Response to Misfolded Protein as a Screen for Soluble Recombinant Protein.” *Protein Engineering* 15 (2): 153–60. http://www.ncbi.nlm.nih.gov/pubmed/11917152.

Levin, Yishai, Emanuel Schwarz, Lan Wang, F. Markus Leweke, and Sabine Bahn. 2007. “Label-Free LC-MS/MS Quantitative Proteomics for Large-Scale Biomarker Discovery in Complex Samples.” *Journal of Separation Science* 30 (14): 2198–2203. https://doi.org/10.1002/jssc.200700189.

Link, Andrew J, and George M Church. 1997. “Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type.” *Journal of Bacteriology* 179 (20): 6228–37.

Liu, Wenshe, Peter E. Peterson, James A. Langston, Xueguang Jin, Xianzhi Zhou, Andrew J. Fisher, and Michael D. Toney. 2005. “Kinetic and Crystallographic Analysis of Active Site Mutants of Escherichia Coli γ-Aminobutyrate

Aminotransferase." *Biochemistry* 44 (8): 2982–92. https://doi.org/10.1021/bi048657a.

Livak, Kenneth J., and Thomas D. Schmittgen. 2001. "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method." *Methods* 25 (4): 402–8. https://doi.org/10.1006/meth.2001.1262.

Lobstein, Julie, Charlie A. Emrich, Chris Jeans, Melinda Faulkner, Paul Riggs, and Mehmet Berkmen. 2012. "SHuffle, a Novel Escherichia Coli Protein Expression Strain Capable of Correctly Folding Disulfide Bonded Proteins in Its Cytoplasm." *Microbial Cell Factories* 11 (1): 1. https://doi.org/10.1186/1475-2859-11-56.

Maeng, Bo Hee, Dong Hyun Nam, and Yong Hwan Kim. 2011. "Coexpression of Molecular Chaperones to Enhance Functional Expression of Anti-BNP ScFv in the Cytoplasm of Escherichia Coli for the Detection of B-Type Natriuretic Peptide." *World Journal of Microbiology and Biotechnology* 27 (6): 1391–98. https://doi.org/10.1007/s11274-010-0590-5.

Mahalik, Shubhashree, Ashish K. Sharma, and Krishna J. Mukherjee. 2014. "Genome Engineering for Improved Recombinant Protein Expression in Escherichia Coli." *Microbial Cell Factories* 13 (1): 1–13. https://doi.org/10.1186/s12934-014-0177-1.

Mahalik, Shubhashree, Ashish Kumar Sharma, Priyanka Jain, and Krishna Jyoti Mukherjee. 2017. "Identifying Genomic Targets for Protein Over-Expression by 'Omics' Analysis of Quiescent Escherichia Coli Cultures." *Microbial Cell Factories* 16 (1): 1–11. https://doi.org/10.1186/s12934-017-0744-3.

Majdalani, Nadim, David Hernandez, and Susan Gottesman. 2002. "Regulation and Mode of Action of the Second Small RNA Activator of RpoS Translation, RprA." *Molecular Microbiology* 46 (3): 813–26. https://doi.org/10.1046/j.1365-2958.2002.03203.x.

Mallik, Prabhat, Brian J. Paul, Steven T. Rutherford, Richard L. Gourse, and Robert

Osuna. 2006. “DksA Is Required for Growth Phase-Dependent Regulation, Growth Rate-Dependent Control, and Stringent Control of Fis Expression in Escherichia Coli.” *Journal of Bacteriology* 188 (16): 5775–82. https://doi.org/10.1128/JB.00276-06.

Mar, Juana, Navarro Llorens, Antonio Tormo, and Esteban Mart. 2010. “Stationary Phase in Gram-Negative Bacteria” 34: 476–95. https://doi.org/10.1111/j.1574-6976.2010.00213.x.

Marco, Ario de. 2009. “Strategies for Successful Recombinant Expression of Disulfide Bond-Dependent Proteins in Escherichia Coli.” *Microbial Cell Factories* 8. https://doi.org/10.1186/1475-2859-8-26.

Marisch, Karoline, Karl Bayer, Theresa Scharl, Juergen Mairhofer, Peter M. Krempl, Karin Hummel, Ebrahim Razzazi-Fazeli, and Gerald Striedner. 2013. “A Comparative Analysis of Industrial Escherichia Coli K-12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Level.” *PLoS ONE* 8 (8). https://doi.org/10.1371/journal.pone.0070516.

Marta, A, J Link, Deirdre Furlong, and Roberto Kolter. 1992. “A Novel DNA-Binding Protein with Regulatory and Protective Roles in Starved Escherichia Coli,” 2646–54.

Mary E. Girard, Saumya Gopalkrishnan, Elicia D. Grace, Jennifer A. Halliday, Richard L. Gourse, Christophe Herman. 2017. “DksA and PpGpp Regulate the σS Stress Response by Activating Promoters for the Small RNA DsrA and the Anti-Adapter Protein IraP.” *Journal of Bacteriology* 200 (2): e00463-17. https://doi.org/10.1128/jb.00463-17.

Matos, Cristina F R O, Steven D. Branston, Anna Albiniak, Arjun Dhanoya, Robert B. Freedman, Eli Keshavarz-Moore, and Colin Robinson. 2012. “High-Yield Export of a Native Heterologous Protein to the Periplasm by the Tat Translocation Pathway in Escherichia Coli.” *Biotechnology and Bioengineering* 109 (10): 2533–42. https://doi.org/10.1002/bit.24535.

Matos, Cristina F R O, Colin Robinson, Heli I. Alanen, Piotr Prus, Yuko Uchida, Lloyd W. Ruddock, Robert B. Freedman, and Eli Keshavarz-Moore. 2014. "Efficient Export of Prefolded, Disulfide-Bonded Recombinant Proteins to the Periplasm by the Tat Pathway in Escherichia Coli CyDisCo Strains." *Biotechnology Progress* 30 (2): 281–90. https://doi.org/10.1002/btpr.1858.

Mergulhão, F. J.M., D. K. Summers, and G. A. Monteiro. 2005. "Recombinant Protein Secretion in Escherichia Coli." *Biotechnology Advances* 23 (3): 177–202. https://doi.org/10.1016/j.biotechadv.2004.11.003.

Merrikh, Houra, Alexander E. Ferrazzoli, and Susan T. Lovett. 2009. "Growth Phase and (p)PpGpp Control of IraD, a Regulator of RpoS Stability, in Escherichia Coli." *Journal of Bacteriology* 191 (24): 7436–46. https://doi.org/10.1128/JB.00412-09.

Metzger, Shula, Iris Ben Dror, Einat Aizenman, Gideon Schreiber, Mark Toone, James D Friesens, Michael Cashel, and Gad Glaserll. 1988. "The Nucleotide Sequence and Characterization of The" 263 (30): 15699–704.

Mey, Marjan De, Sofie De Maeseneire, Wim Soetaert, and Erick Vandamme. 2007. "Minimizing Acetate Formation in E. Coli Fermentations." *Journal of Industrial Microbiology and Biotechnology* 34 (11): 689–700. https://doi.org/10.1007/s10295-007-0244-2.

Mika, Franziska, and Regine Hengge. 2014. "Small RNAs in the Control of RpoS, CsgD, and Biofilm Architecture of Escherichia Coli." *RNA Biology* 11 (5): 494–507. https://doi.org/10.4161/rna.28867.

Minsky, Abraham, Sharon G. Wolf, Daphna Frenkiel, Talmon Arad, Steven E. Finkel, and Roberto Kolter. 1999. "DNA Protection by Stress-Induced Biocrystallization." *Nature* 400 (6739): 83–85. https://doi.org/10.1038/21918.

Minton, Nigel P. 1984. "Improved Plasmid Vectors for the Isolation of Translational Lac Gene Fusions." *Gene* 31 (1–3): 269–73. https://doi.org/10.1016/0378-1119(84)90220-8.

Mitchell, Jennie E., Taku Oshima, Sarah E. Piper, Christine L. Webster, Lars F. Westblade, Gouzel Karimova, Daniel Ladant, et al. 2007. "The Escherichia Coli Regulator of Sigma 70 Protein, Rsd, Can up-Regulate Some Stress-Dependent Promoters by Sequestering Sigma 70." *Journal of Bacteriology* 189 (9): 3489–95. https://doi.org/10.1128/JB.00019-07.

Mohammed N. Baeshen, Ahmed M. Al-Hejin, Roop S. Bora, Mohamed M. M. Ahmed, Hassan A. I. Ramadan, Kulvinder S. Saini, Nabih A. Baeshen, and and Elrashdy M. Redwan. 2015. "Production of Biopharmaceuticals in E. Coli: Current Scenario and Future Perspectives." *J. Microbiol. Biotechnol* 25 (7): 953–62. https://doi.org/10.4014/jmb.1405.05052.

Moll, Isabella, and Hanna Engelberg-Kulka. 2012. "Selective Translation during Stress in Escherichia Coli." *Trends in Biochemical Sciences* 37 (11): 493–98. https://doi.org/10.1016/j.tibs.2012.07.007.

Moon, Kyung, and Susan Gottesman. 2011. "Competition among Hfq-Binding Small RNAs in Escherichia Coli." *Molecular Microbiology* 82 (6): 1545–62. https://doi.org/10.1111/j.1365-2958.2011.07907.x.

Muhamadali, Howbeer, Yun Xu, Rosa Morra, Drupad K. Trivedi, Nicholas J.W. Rattray, Neil Dixon, and Royston Goodacre. 2016. "Metabolomic Analysis of Riboswitch Containing E. Coli Recombinant Expression System." *Molecular BioSystems* 12 (2): 350–61. https://doi.org/10.1039/c5mb00624d.

Nair, Sudha, and Steven E Finkel. 2004. "Dps Protects Cells against Multiple Stresses during Stationary Phase." *Journal of Bacteriology* 186 (13): 4192–98. https://doi.org/10.1128/JB.186.13.4192-4198.2004.

Nannenga, Brent L., and François Baneyx. 2011. "Reprogramming Chaperone Pathways to Improve Membrane Protein Expression in Escherichia Coli." *Protein Science* 20 (8): 1411–20. https://doi.org/10.1002/pro.669.

Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. 2012. "Constraining the

Metabolic Genotype Phenotype Relationship Using a Phylogeny." *Appl Microbiol Biotechnol* 10 (4): 291–305. https://doi.org/10.1038/nrmicro2737.Constraining.

Neubauer, P., H. Y. Lin, and B. Mathiszik. 2003. "Metabolic Load of Recombinant Protein Production: Inhibition of Cellular Capacities for Glucose Uptake and Respiration after Induction of a Heterologous Gene in Escherichia Coli." *Biotechnology and Bioengineering* 83 (1): 53–64. https://doi.org/10.1002/bit.10645.

Newman and Lin. 1995. "Leucine-Responsive Regulatory Protein: A Global Regulator of Gene Expression in E. Coli." *Annual Review of Microbiology* 49 (1): 747–75. https://doi.org/10.1146/annurev.micro.49.1.747.

Ni, Ye, and Rachel Chen. 2009. "Extracellular Recombinant Protein Production from Escherichia Coli." *Biotechnology Letters* 31 (11): 1661–70. https://doi.org/10.1007/s10529-009-0077-3.

Nieuwkoop, Thijs, Nico J. Claassens, and John van der Oost. 2019. "Improved Protein Production and Codon Optimization Analyses in Escherichia Coli by Bicistronic Design." *Microbial Biotechnology* 12 (1): 173–79. https://doi.org/10.1111/1751-7915.13332.

Nilsson, L., A. Vanet, E. Vijgenboom, and L. Bosch. 1990. "The Role of FIS in Trans Activation of Stable RNA Operons of E. Coli." *The EMBO Journal* 9 (3): 727–34. https://doi.org/10.1002/j.1460-2075.1990.tb08166.x.

Nocon, Justyna, Matthias G. Steiger, Martin Pfeffer, Seung Bum Sohn, Tae Yong Kim, Michael Maurer, Hannes Rußmayer, et al. 2014. "Model Based Engineering of Pichia Pastoris Central Metabolism Enhances Recombinant Protein Production." *Metabolic Engineering* 24: 129–38. https://doi.org/10.1016/j.ymben.2014.05.011.

Oehler, S., E.R. Eismann, H. Krämer, and B. Müller-Hill. 1990. "The Three Operators of the Lac Operon Cooperate in Repression." *The EMBO Journal* 9 (4): 973–79. https://doi.org/10.1002/j.1460-2075.1990.tb08199.x.

Oh, Min Kyu, and James C. Liao. 2000. "DNA Microarray Detection of Metabolic Responses to Protein Overproduction in Escherichia Coli." *Metabolic Engineering* 2 (3): 201–9. https://doi.org/10.1006/mben.2000.0149.

Ou, Jingxing, Lei Wang, Xiaolan Ding, Jianyang Du, Yuan Zhang, Huiping Chen, and Anlong Xu. 2004. "Stationary Phase Protein Overproduction Is a Fundamental Capability of Escherichia Coli." *Biochemical and Biophysical Research Communications* 314 (1): 174–80. https://doi.org/10.1016/j.bbrc.2003.12.077.

Ow, Dave Siak Wei, Peter Morin Nissom, Robin Philp, Steve Kah Weng Oh, and Miranda Gek Sim Yap. 2006. "Global Transcriptional Analysis of Metabolic Burden Due to Plasmid Maintenance in Escherichia Coli DH5α during Batch Fermentation." *Enzyme and Microbial Technology* 39 (3): 391–98. https://doi.org/10.1016/j.enzmictec.2005.11.048.

Panda, Amulya K., R. H. Khan, K. B.C. Appa Rao, and S. M. Totey. 1999. "Kinetics of Inclusion Body Production in Batch and High Cell Density Fed-Batch Culture of Escherichia Coli Expressing Ovine Growth Hormone." *Journal of Biotechnology* 75 (2–3): 161–72. https://doi.org/10.1016/S0168-1656(99)00157-1.

Park, Hongmarn, Louise C. McGibbon, Anastasia H. Potts, Helen Yakhnin, Tony Romeo, and Paul Babitzke. 2017. "Translational Repression of the RpoS Antiadapter IraD by CsrA Is Mediated via Translational Coupling to a Short Upstream Open Reading Frame." *MBio* 8 (4): 1–14. https://doi.org/10.1128/mbio.01355-17.

Patikoglou, Georgia A, Lars F Westblade, Elizabeth A Campbell, Valérie Lamour, J Lane, and Seth A Darst. 2008. "Crystal Structure of the Escherichia Coli Regulator of Σ70, Rsd, in.Pdf" 372 (3): 649–59.

Patten, C. L., M. G. Kirchhof, M. R. Schertzberg, R. A. Morton, and H. E. Schellhorn. 2004. "Microarray Analysis of RpoS-Mediated Gene Expression in Escherichia Coli K-12." *Molecular Genetics and Genomics* 272 (5): 580–91. https://doi.org/10.1007/s00438-004-1089-2.

Peano, Clelia, Johannes Wolf, Julien Demol, Elio Rossi, Luca Petiti, Gianluca De Bellis, Johannes Geiselmann, Thomas Egli, Stephan Lacour, and Paolo Landini. 2015. "Characterization of the Escherichia Coli ΔS Core Regulon by Chromatin Immunoprecipitation-Sequencing (ChIP-Seq) Analysis." *Scientific Reports* 5 (May): 1–15. https://doi.org/10.1038/srep10469.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. "SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions." *Nature Methods* 8 (10): 785–86. https://doi.org/10.1038/nmeth.1701.

Philippe Goffin, Philippe Dehottay. 2017. "Complete Genome Sequence of Escherichia Coli BLR(DE3), a RecA-Deficient Derivative of E. Coli BL21(DE3)." *American Society for Microbiology* 5 (22): e00441-17.

Piatkowska-Jakubas, Beata, Małgorzata Krawczyk-Kuliś, Sebastian Giebel, Maria Adamczyk-Cioch, Anna Czyz, Ewa Lech Marańda, Monika Paluszewska, Grazyna Pałłynyczko, Jarosław Piszcz, and Jerzy Hołowiecki. 2008. "Use of L-Asparaginase in Acute Lymphoblastic Leukemia: Recommendations of the Polish Adult Leukemia Group." *Polskie Archiwum Medycyny Wewnetrznej* 118 (11): 664–69.

Rai, Amit, and Kazuki Saito. 2016. "Omics Data Input for Metabolic Modeling." *Current Opinion in Biotechnology* 37: 127–34. https://doi.org/10.1016/j.copbio.2015.10.010.

Ramagopal, S, and B D Davis. 1974. "Localization of the Stringent Protein of Escherichia Coli on the 50S Ribosomal Subunit." *Proceedings of the National Academy of Sciences of the United States of America* 71 (3): 820–24. http://www.ncbi.nlm.nih.gov/pubmed/4595574%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC388106.

Ramani, M Freundlich N, E Mathew, A Sirko, and P Tsuf. 1992. "MicroReview The Role of Integration Host Factor In Gene Expression in Escherichia Coli" 6: 2557–63.

Ramos CR, Abreu PA, Nascimento AL, Ho PL. 2004. "A High-Copy T7 Escherichia Coli Expression Vector for the Production of Recombinant Proteins with a Minimal N-Terminal His-Tagged Fusion Peptide" 37 (8): 1103–9.

Reeve, Benjamin, Thomas Hargest, Charlie Gilbert, and Tom Ellis. 2014. "Predicting Translation Initiation Rates for Designing Synthetic Biology." *Frontiers in Bioengineering and Biotechnology* 2 (January): 1–6. https://doi.org/10.3389/fbioe.2014.00001.

Repoila, F., N. Majdalani, and S. Gottesman. 2003. "Small Non-Coding RNAs, Co-Ordinators of Adaptation Processes in Escherichia Coli: The RpoS Paradigm." *Molecular Microbiology* 48 (4): 855–61. https://doi.org/10.1046/j.1365-2958.2003.03454.x.

Reuveni, Shlomi, Isaac Meilijson, Martin Kupiec, Eytan Ruppin, and Tamir Tuller. 2011. "Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model." *PLoS Computational Biology* 7 (9): e1002127. https://doi.org/10.1371/journal.pcbi.1002127.

Reynolds, Sheila M., Lukas Käll, Michael E. Riffle, Jeff A. Bilmes, and William Stafford Noble. 2008. "Transmembrane Topology and Signal Peptide Prediction Using Dynamic Bayesian Networks." *PLoS Computational Biology* 4 (11): e1000213. https://doi.org/10.1371/journal.pcbi.1000213.

Ron, Eliora Z. 2013. "Bacterial Stress Response." In *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, 589–98. https://doi.org/10.1007/978-3-642-30141-4.

Roncarati, Davide, and Vincenzo Scarlato. 2017. "Regulation of Heat-Shock Genes in Bacteria: From Signal Sensing to Gene Expression Output." *FEMS Microbiology Reviews* 41 (4): 549–74. https://doi.org/10.1093/femsre/fux015.

Rosano, Germán L., and Eduardo A. Ceccarelli. 2014. "Recombinant Protein Expression in Escherichia Coli: Advances and Challenges." *Frontiers in Microbiology* 5

(APR): 1–17. https://doi.org/10.3389/fmicb.2014.00172.

Rutberg, Lars. 2019. "The GlpP and GlpF Genes of the Glycerol Regulon in Bacillus Subtilis," no. 1993: 349–59.

S. Hesam Shahravan, Xuanlu Qu, I-San Chan, Jumi A. Shin. 2008. "Enhancing the Specificity of the Enterokinase Cleavage Reaction to Promote Efficient Cleavage of a Fusion Tag." *Protein Expression and Purification* 59 (2): 314–19. https://doi.org/10.1038/mp.2011.182.doi.

Salis, Howard M., Ethan A. Mirsky, and Christopher A. Voigt. 2009. "Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression." *Nature Biotechnology* 27 (10): 946–50. https://doi.org/10.1038/nbt.1568.

Schlesinger, Orr, Yonatan Chemla, Mathias Heltberg, Eden Ozer, Ryan Marshall, Vincent Noireaux, Mogens Høgh Jensen, and Lital Alfonta. 2017. "Tuning of Recombinant Protein Expression in Escherichia Coli by Manipulating Transcription, Translation Initiation Rates, and Incorporation of Noncanonical Amino Acids." *ACS Synthetic Biology* 6 (6): 1076–85. https://doi.org/10.1021/acssynbio.7b00019.

Schneider, Barbara L, Barbara L Schneider, Stephen Ruback, Stephen Ruback, Hillary Kasbarian, Hillary Kasbarian, Christine Pybus, Christine Pybus, Larry Reitzer, and Larry Reitzer. 2002. "The Escherichia Coli GabDTPC Operon: Specific Gamma-Aminobutyrate Catabolism and Nonspecific Induction." *Journal of Bacteriology* 184 (24): 6976–86. https://doi.org/10.1128/JB.184.24.6976.

Schryvers, Anthony, and Joel H. Weiner. 2009. " The Anaerobic Sn -Glycerol-3-Phosphate Dehydrogenase: Cloning and Expression of the Glp A Gene of Escherichia Coli and Identification of the Glp A Products ." *Canadian Journal of Biochemistry* 60 (3): 224–31. https://doi.org/10.1139/o82-027.

Sharma, Ashish K., Shubhashree Mahalik, Chaitali Ghosh, Anuradha B. Singh, and Krishna J. Mukherjee. 2011. "Comparative Transcriptomic Profile Analysis of

Fed-Batch Cultures Expressing Different Recombinant Proteins in Escherichia Coli." *AMB Express* 1 (1): 1–12. https://doi.org/10.1186/2191-0855-1-33.

Sharma, Ashish Kumar. 2015. "Metabolic Engineering Strategies for Enhancing Recombinant Protein Expression in Escherichia Coli."

Sharma, Umender K., and Dipankar Chatterji. 2010. "Transcriptional Switching in Escherichia Coli during Stress and Starvation by Modulation of Σ70 Activity." *FEMS Microbiology Reviews* 34 (5): 646–57. https://doi.org/10.1111/j.1574-6976.2010.00223.x.

Shokri, A., A. M. Sandén, and G. Larsson. 2003. "Cell and Process Design for Targeting of Recombinant Protein into the Culture Medium of Escherichia Coli." *Applied Microbiology and Biotechnology* 60 (6): 654–64. https://doi.org/10.1007/s00253-002-1156-8.

Singh, Anuradha B., and Krishna J. Mukherjee. 2013. "Supplementation of Substrate Uptake Gene Enhances the Expression of RhIFN-β in High Cell Density Fed-Batch Cultures of Escherichia Coli." *Molecular Biotechnology* 54 (2): 692–702. https://doi.org/10.1007/s12033-012-9611-y.

Singh, Anuradha B., Ashish K. Sharma, and Krishna J. Mukherjee. 2012. "Analyzing the Metabolic Stress Response of Recombinant Escherichia Coli Cultures Expressing Human Interferon-Beta in High Cell Density Fed Batch Cultures Using Time Course Transcriptomic Data." *Molecular BioSystems* 8 (2): 615–28. https://doi.org/10.1039/c1mb05414g.

Smith, Donald B., and Kevin S. Johnson. 1988. "Single-Step Purification of Polypeptides Expressed in Escherichia Coli as Fusions with Glutathione S-Transferase." *Gene* 67 (1): 31–40. https://doi.org/10.1016/0378-1119(88)90005-4.

Soufi, Boumediene, Karsten Krug, Andreas Harst, and Boris Macek. 2015. "Characterization of the E. Coli Proteome and Its Modifications during Growth and Ethanol Stres." *Frontiers in Microbiology* 6 (103): 1–11.

https://doi.org/10.3389/fmicb.2015.00103.

Srivastava, Poonam, and K. J. Mukherjee. 2005. "Kinetic Studies of Recombinant Human Interferon-Alpha (RhIFN-α) Expression in Transient State Continuous Cultures." *Biochemical Engineering Journal* 26 (1): 50–58. https://doi.org/10.1016/j.bej.2005.06.004.

Stecher, A. L., P. Morgantetti De Deus, I. Polikarpov, and J. Abrahão-Neto. 1999. "Stability of L-Asparaginase: An Enzyme Used in Leukemia Treatment." *Pharmaceutica Acta Helvetiae* 74 (1): 1–9. https://doi.org/10.1016/S0031-6865(99)00009-6.

Striedner, Gerald, Irene Pfaffenzeller, Luchner Markus, Sabine Nemecek, Reingard Grabherr, and Karl Bayer. 2010. "Plasmid-Free T7-Based Escherichia Coli Expression Systems." *Biotechnology and Bioengineering* 105 (4): 786–94. https://doi.org/10.1002/bit.22598.

Strutton, Benjamin, Stephen R P Jaffé, Jagroop Pandhal, and Phillip C Wright. 2017. "Generation of Recombinant N-Linked Glycoproteins in E. Coli." In *Methods in Molecular Biology*, 1586:233–50. https://doi.org/10.1007/978-1-61737-967-3.

Studier, F. William, and Barbara A. Moffatt. 1986. "Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes." *Journal of Molecular Biology* 189 (1): 113–30. https://doi.org/10.1016/0022-2836(86)90385-2.

Su, Lingqia, Chenhua Xu, Ronald W. Woodard, Jian Chen, and Jing Wu. 2013. "A Novel Strategy for Enhancing Extracellular Secretion of Recombinant Proteins in Escherichia Coli." *Applied Microbiology and Biotechnology* 97 (15): 6705–13. https://doi.org/10.1007/s00253-013-4994-7.

Sugamata, Yasuhiro, and Toshikazu Shiba. 2005. "Improved Secretory Production of Recombinant Proteins by Random Mutagenesis of HlyB, an Alpha-Hemolysin Transporter from Escherichia Coli." *Applied and Environmental Microbiology* 71

(2): 656–62. https://doi.org/10.1128/AEM.71.2.656-662.2005.

Swain, Amy L, Mariusz Jask6lski, Dominique Houssett, J K Mohana Rao, and Alexander Wlodawert. 1993. "Crystal Structure of Escherichia Coli L-Asparaginase, an Enzyme Used in Cancer Therapy (Amidohydrolase/Leukemia/Active Site/Aspartate)." *Proc. Natl. Acad. Sci. USA* 90 (4): 1474–78. http://www.pnas.org/content/pnas/90/4/1474.full.pdf.

Swartz, James R. 2001. "Advances in Escherichia Coli Production of Therapeutic Proteins." *Current Opinion in Biotechnology* 12 (2): 195–201. https://doi.org/10.1016/S0958-1669(00)00199-3.

Tani, T. H., A. Khodursky, R. M. Blumenthal, P. O. Brown, and R. G. Matthews. 2002. "Adaptation to Famine: A Family of Stationary-Phase Genes Revealed by Microarray Analysis." *Proceedings of the National Academy of Sciences* 99 (21): 13471–76. https://doi.org/10.1073/pnas.212510999.

Terpe, K. 2003. "Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems." *Applied Microbiology and Biotechnology* 60 (5): 523–33. https://doi.org/10.1007/s00253-002-1158-6.

Truniger, V, J Beuerle, C Gandor, N Wittekindt, G Sweet, R Voegele, E C Lin, and W Boos. 2017. "Glycerol Facilitator of Escherichia Coli: Cloning of GlpF and Identification of the GlpF Product." *Journal of Bacteriology* 172 (1): 424–30. https://doi.org/10.1128/jb.172.1.424-430.1990.

Udekwu, Klas I., Fabien Darfeuille, Jörg Vogel, Johan Reimegård, Erik Holmqvist, and E. G H Wagner. 2005. "Hfq-Dependent Regulation of OmpA Synthesis Is Mediated by an Antisense RNA." *Genes and Development* 19 (19): 2355–66. https://doi.org/10.1101/gad.354405.

Vaiphei, S. Thangminlal, Gaurav Pandey, and K. J. Mukherjee. 2009. "Kinetic Studies of Recombinant Human Interferon-Gamma Expression in Continuous Cultures of E. Coli." *Journal of Industrial Microbiology and Biotechnology* 36 (12): 1453–58.

https://doi.org/10.1007/s10295-009-0632-x.

Varma, A, and B O Palsson. 1994. "Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type Escherichia Coli W3110." *Applied and Environmental Microbiology* 60 (10): 3724–31. http://www.ncbi.nlm.nih.gov/pubmed/7986045%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC201879.

Vesper, Oliver, Shahar Amitai, Maria Belitsky, Konstantin Byrgazov, Anna Chao Kaberdina, Hanna Engelberg-Kulka, and Isabella Moll. 2011. "Selective Translation of Leaderless MRNAs by Specialized Ribosomes Generated by MazF in Escherichia Coli." *Cell* 147 (1): 147–57. https://doi.org/10.1016/j.cell.2011.07.047.

Walsh G. 2010. "Biopharmaceutical Benchmarks 2010." *Nature Biotechnology* 28 (9): 917–24. https://doi.org/http://dx.doi.org/10.1038/nbt0706-769.

Wang, Jian, Bing Wen, Qingyang Xu, Xixian Xie, and Ning Chen. 2015. "Optimization of Carbon Source and Glucose Feeding Strategy for Improvement of L-Isoleucine Production by Escherichia Coli." *Biotechnology and Biotechnological Equipment* 29 (2): 374–80. https://doi.org/10.1080/13102818.2015.1006899.

Weaver, Daniel S., Ingrid M. Keseler, Amanda Mackie, Ian T. Paulsen, and Peter D. Karp. 2014. "A Genome-Scale Metabolic Flux Model of Escherichia Coli K-12 Derived from the EcoCyc Database." *BMC Systems Biology* 8 (79): 1–24. https://doi.org/10.1186/1752-0509-8-79.

Weber, Harald, Tino Polen, Johanna Heuveling, Volker F Wendisch, Regine Hengge, Forschungszentrum Ju, and Weber E T Al. 2005. "Genome-Wide Analysis of the General Stress Response Network In." *Society* 187 (5): 1591–1603. https://doi.org/10.1128/JB.187.5.1591.

Weissenborn, Deborah L, Nicola Wittekindtn, and Timothy J Larsonsii. 1992. "Structure and Regulation of the GZpFK Operon Encoding Glycerol." *Biochemistry* 267 (9):

6122–31.

Welch, Mark, Sridhar Govindarajan, Jon E. Ness, Alan Villalobos, Austin Gurney, Jeremy Minshull, and Claes Gustafsson. 2009. "Design Parameters to Control Synthetic Gene Expression in Eschorichia Coli." *PLoS ONE* 4 (9): e7002. https://doi.org/10.1371/journal.pone.0007002.

Wendisch, Volker F. 2014. "Microbial Production of Amino Acids and Derived Chemicals: Synthetic Biology Approaches to Strain Development." *Current Opinion in Biotechnology* 30: 51–58. https://doi.org/10.1016/j.copbio.2014.05.004.

Wick, Lukas M., and Thomas Egli. 2004. "Molecular Components of Physiological Stress Responses in Escherichia Coli." *Adv Biochem Engin/Biotechnol* 89: 1–45. https://doi.org/10.1007/b93957.

Wriston, John C. 1985. "Asparaginase." In *Methods in Enzymology,* 113:608–18.

Xu, J., and R. C. Johnson. 1995. "Identification of Genes Negatively Regulated by Fis: Fis and RpoS Comodulate Growth-Phase-Dependent Gene Expression in Escherichia Coli." *Journal of Bacteriology* 177 (4): 938–47.

Yamagishi, M, H Matsushima, A Wada, M Sakagami, N Fujita, and A Ishihama. 1993. "Regulation of the Escherichia Coli Rmf Gene Encoding the Ribosome Modulation Factor: Growth Phase- and Growth Rate-Dependent Control." *The EMBO Journal* 12 (2): 625–30. http://www.ncbi.nlm.nih.gov/pubmed/8440252%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC413246.

Yamamoto, Kaneyoshi, Hiroki Watanabe, and Akira Ishihama. 2014. "Expression Levels of Transcription Factors in Escherichia Coli: Growth Phase- and Growth Condition-Dependent Variation of 90 Regulators from Six Families." *Microbiology (United Kingdom)* 160 (PART 9): 1903–13. https://doi.org/10.1099/mic.0.079889-0.

Yeh, J.I., U. Chinte, and S. Du. 2008. "Structure of Glycerol-3-Phosphate Dehydrogenase, an Essential Monotopic Membrane Enzyme." *Acta Crystallographica Section A Foundations of Crystallography* 64 (a1): C283–84. https://doi.org/10.1107/s0108767308090934.

Yoon, Sung H., Mee Jung Han, Haeyoung Jeong, Choong H. Lee, Xiao Xia Xia, Dae Hee Lee, Ji H. Shim, Sang Y. Lee, Tae K. Oh, and Jihyun F. Kim. 2012. "Comparative Multi-Omics Systems Analysis of Escherichia Coli Strains B and K-12." *Genome Biology* 13 (5): R37. https://doi.org/10.1186/gb-2012-13-5-r37.

Yoshida, Hideji, Yasushi Maki, Shou Furuike, Akiko Sakai, Masami Ueta, and Akira Wada. 2012. "YqjD Is an Inner Membrane Protein Associated with Stationary-Phase Ribosomes in Escherichia Coli." *Journal of Bacteriology* 194 (16): 4178–83. https://doi.org/10.1128/JB.00396-12.

Yu, Xiao Wei, Wei Hong Sun, Ying Zheng Wang, and Yan Xu. 2017. "Identification of Novel Factors Enhancing Recombinant Protein Production in Multi-Copy Komagataella Phaffii Based on Transcriptomic Analysis of Overexpression Effects." *Scientific Reports* 7 (1): 16249. https://doi.org/10.1038/s41598-017-16577-x.

Zhang, Aixia, Shoshy Altuvia, Anita Tiwari, Liron Argaman, Regine Hengge-Aronis, and Gisela Storz. 1998. "The OxyS Regulatory RNA Represses RpoS Translation and Binds the Hfq (HF-I) Protein." *EMBO Journal* 17 (20): 6061–68. https://doi.org/10.1093/emboj/17.20.6061.

Zhang, Liu yan, Su hua Chang, and Jing Wang. 2010. "How to Make a Minimal Genome for Synthetic Minimal Cell." *Protein and Cell* 1 (5): 427–34. https://doi.org/10.1007/s13238-010-0064-4.

Zhou, Y., S. Gottesman, J. R. Hoskins, M. R. Maurizi, and S. Wickner. 2001. "The RssB Response Regulator Directly Targets ΣS for Degradation by ClpXP." *Genes and Development* 15 (5): 627–37. https://doi.org/10.1101/gad.864401.

Zhou, Yanning, and Susan Gottesman. 1998. "Regulation of Proteolysis of the Stationary-Phase Sigma Factor RpoS." *Journal of Bacteriology* 180 (5): 1154–58.

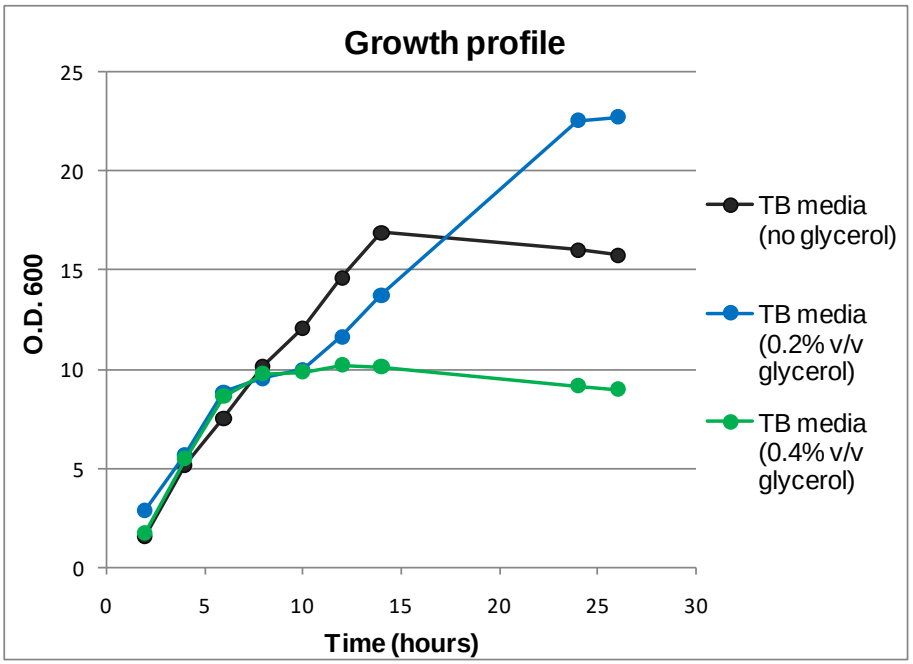

Figure A1: *E. coli* K-12 W3110 growth profiles with 0.2% and 0.4% v/v glycerol and without glycerol supplementation in TB media.

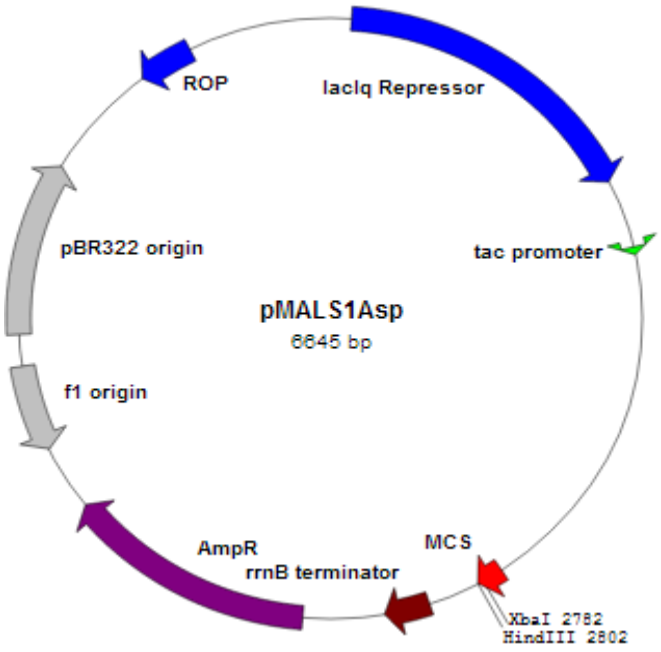

Figure A2: The pMAL-p2X L-asp vector map.