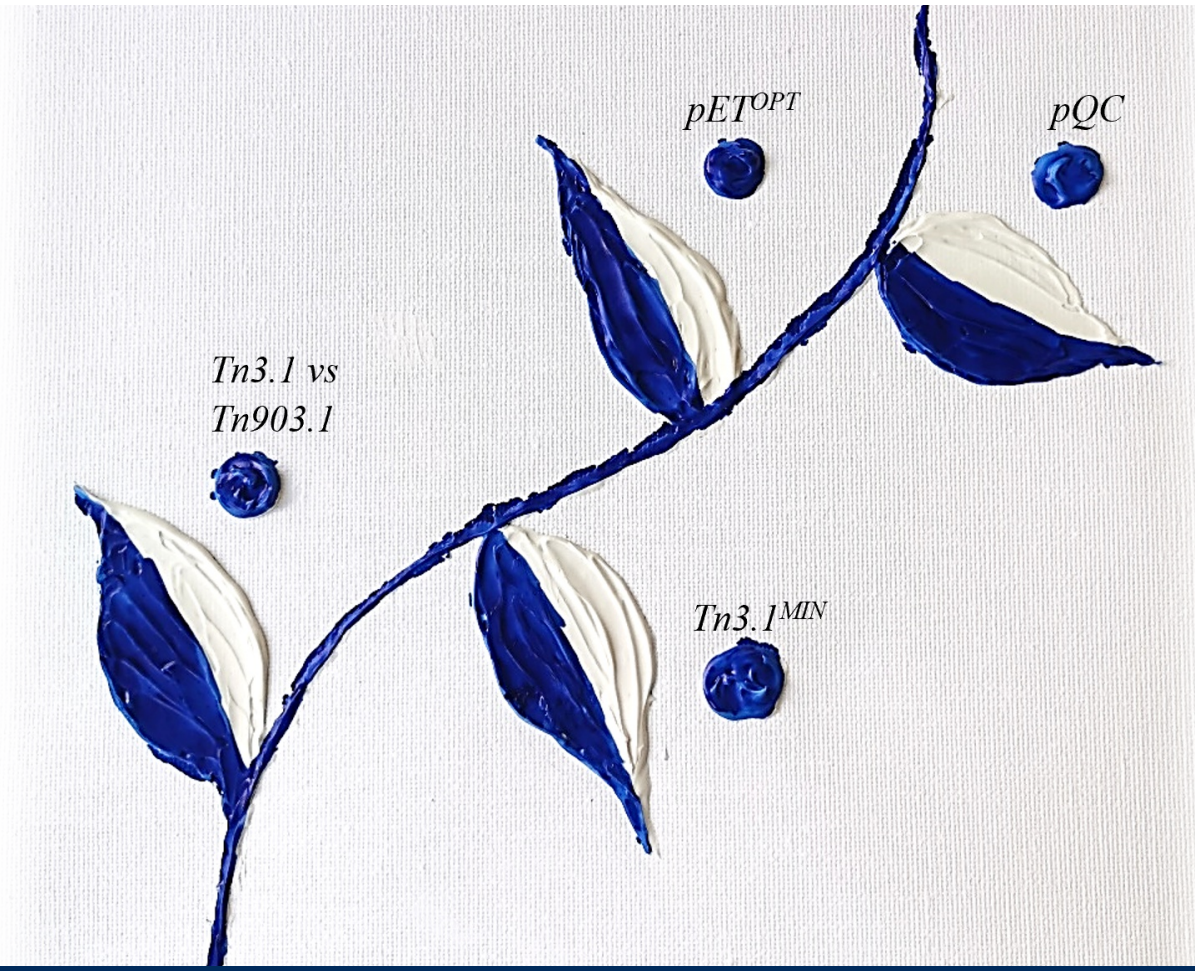
Stockholm University

New genetic modules for protein production in microbial cell factories

Diana Khananisho



Doctoral Thesis in Biochemistry at Stockholm University, Sweden 2024

New genetic modules for protein production in microbial cell factories

Diana Khananisho

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Wednesday 25 September 2024 at 14.00 in Magnélisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16 B.

Abstract

Escherichia coli (*E. coli*) is a commonly used microbial cell factory in recombinant protein production. Although extensive efforts have been made to optimize the production of soluble, functional recombinant proteins, sufficient yields are still not obtainable for all proteins. This doctoral thesis presents tools which address concerns of insufficient titres. These tools consist of 1) improvements to antibiotic resistance fragments 2) a new collection of pET expression plasmids containing re-designed antibiotic resistance fragments, transcription- and translation initiation modules and terminator module 3) description of a biosensor which can detect issues in quality of recombinant protein production in vivo. The use of these tools could possibly increase the yields of recombinant protein.

Keywords: *Microbial cell factories, recombinant protein, recombinant protein production, protein production system, E. coli, pET, pET system, protein production optimization, expression plasmid, biosensor, genetic modules.*

Stockholm 2024
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-232294

ISBN 978-91-8014-877-1
ISBN 978-91-8014-878-8

Department of Biochemistry and Biophysics

Stockholm University, 106 91 Stockholm

NEW GENETIC MODULES FOR PROTEIN PRODUCTION IN MICROBIAL CELL FACTORIES

Diana Khananisho

Stockholm University

New genetic modules for protein production in microbial cell factories

Diana Khananisho

©Diana Khananisho, Stockholm University 2024

ISBN print 978-91-8014-877-1
ISBN PDF 978-91-8014-878-8

Printed in Sweden by Universitetsservice US-AB, Stockholm 2024

To Jesus Christ, my Lord and Saviour

Abstract

Escherichia coli (*E. coli*) is a commonly used microbial cell factory in recombinant protein production. Although extensive efforts have been made to optimize the production of soluble, functional recombinant proteins, sufficient yields are still not obtainable for all proteins. This doctoral thesis presents tools which address concerns of insufficient titres. These tools consist of 1) improvements to antibiotic resistance fragments 2) a new collection of pET expression plasmids containing re-designed antibiotic resistance fragments, transcription- and translation initiation modules and terminator module 3) description of a biosensor which can detect issues in quality of recombinant protein production *in vivo*. The use of these tools could possibly increase the yields of recombinant protein.

Table of contents

List of publications 4

Authors contributions in the papers 5

Populärvetenskaplig sammanfattning på svenska 6

General introduction 7

Transcription in *E. coli* 8

Protein translation in *E. coli* 10

Protein folding and quality control in *E. coli* 13

Cytosolic folding and quality control systems 13

The heat shock response and inclusion body binding protein 14

Cpx envelope stress response 17

Replicons 20

Promoters 20

Translation initiation region on mRNA 23

Antibiotic selection fragments 25

Antibiotic-free selection fragments 26

Cell culture growth conditions 30

How to utilize innate E. coli stress responses to monitor and alleviate recombinant protein production stress 34

Plasmid maintenance in *E. coli* 37

Summary of publications 41

Paper I: Tips for efficiently maintaining pET expression plasmids 41

Paper II: Antibiotic-Efficient Genetic Cassette for the TEM-1 β-Lactamase That Improves Plasmid Performance 42

Paper III: The pET^{OPT} collection of expression plasmids: Better performing versions of the old favourites 44

Paper IV: Biosensor that detects stress caused by periplasmic proteins 45

Conclusions and prospects 47

Acknowledgements 51

References 53

List of publications

Paper I: Tips for efficiently maintaining pET expression plasmids
Diana Khananisho[1]*, Alister J. Cumming[1]*, Daria Kulakova[1], Patrick J. Shilling[1] and Daniel O. Daley[1]

Current Genetics 2023, 69(4-6), 277-287 *Authors contributed equally

Paper II: Antibiotic-Efficient Genetic Cassette for the TEM-1 β-Lactamase That Improves Plasmid Performance

Alister J. Cumming,[*] Diana Khananisho,[*] Ramona Harris, Carolyn N. Bayer, Morten H. H. Norholm, Sara Jamshidi, Leopold L. Ilag and Daniel O. Daley

ACS Synthetic Biology, 2022, 11(1), 241-253 *Authors contributed equally

Paper III: The pET^{OPT} collection of expression plasmids: Better performing versions of the old favourites
Diana Khananisho[1], Alister J. Cumming[1], Caroline Moberg[1], Magnus Widesheim[1], Patrick J. Shilling[1], Morten H.H. Norholm[2] and Daniel O. Daley[1]

Manuscript

Paper IV: Biosensor that detects stress caused by periplasmic proteins
Alister J. Cumming[1], Diana Khananisho[1], Nicklas Liljestrand[1], Mateusz Balka[1] and Daniel O. Daley[1]

ACS Synthetic Biology 2024, 13(5), 1477-1491

Additional papers

Signal amplification of araC pBAD using a standardised translation initiation region

Patrick J. Shilling[1,$],*, Diana Khananisho[1], James Cumming[1], Bill Söderström[2], and Daniel O. Daley[1,*]

Synthetic Biology 2022, 7(1), 1-9

Authors contributions in the papers

Paper I: Tips for efficiently maintaining pET expression plasmids

The author made a major contribution to this publication, which included experimental design, data collection and analysis, preparation of all figures and writing.

Paper II: Antibiotic-Efficient Genetic Cassette for the TEM-1 β-Lactamase That Improves Plasmid Performance

The author made a major contribution to this publication, which included the data collection and analysis presented in Table 1 and Figures 1, 4, 5.

Paper III: The pET ^{OPT} collection of expression plasmids: Better performing versions of the old favourites

The author made a major contribution to this publication, which included experimental design, data collection and analysis, preparation of all figures and writing.

Paper IV: Biosensor that detects stress caused by periplasmic proteins

The author made a minor contribution to this publication, which included cloning, data collection for Figure S2, data analysis and figure composition for Figures 1, 4, 5, S2 and S4-S6.

Populärvetenskaplig sammanfattning på svenska

Bakterien *Escherichia coli* (även känds som *E. coli*) är en gram-negativ bakterie som används till syfte att producera proteiner. Närmare bestämt, proteiner som inte alltid finns naturligt i *E. coli* utan som har ett annat ursprung (rekombinanta proteiner). Dessa anses ofta ha hög applicerbarhet i till exempel medicinsk behandling av diabetes (insulin), produktion av små molekyler (melanin) eller enzymer (detergenter i tvättmedel). *E. coli* är enkla och relativt billiga att kultivera, växer snabbt och är enkla att genmanipulera. Dock är inte produktion av proteiner alltid rättfram. För att ett protein ska vara av nytta så måste den anamma sin naturliga struktur, vilket kan åstadkommas genom att ändra på tillväxtmiljön eller mängd av proteiner som produceras i *E. coli*. På grund av detta så har mycket tid och ansträngning krävs för att optimera *E. coli* så att vi kan kontrollera deras miljö, plasmider och dess genetiska komponenter, samt när och hur starkt de genetiska komponenterna som ansvarar för proteinuttryck ska aktiveras då kemikalier tillsätts.

I denna avhandling har fokus lagts på att undersöka möjliga problem som kan uppstå då nyttjande av fel genetiska komponenter som ska tillåta ett visst urval av *E. coli* som enbart bär plasmider att växa men inte andra inte fungerar som de är tänkt, felaktig matchning av *E. coli* stammar med genetiska komponenter på plasmids som inte tillåter bakterier att växa eller tillverka tillräckligt med protein samt tillväxtmiljöer som stjälper *E. coli* genom att introducera stress i de olika delarna av *E. coli*: cytoplasman och periplasman. Denna avhandling ger även potentiella lösningar och verktyg för att undkomma problemen nämnda innan, så som minimala selektions fragment som nyttjar antibiotikaresistens, kombinationer av *E. coli* stammar med genetiska fragment i plasmider som tillåter olika mycket produktion av protein, samt en biosensor som möjliggör detektion av cellulär stress på grund av för hög produktion av protein.

General introduction

The bacterium *Escherichia coli* (*E. coli*) is used as a cell factory for production of recombinant proteins. The cells native machineries are utilized to not only make what they need to propagate, but also the recombinant protein. In this thesis the processes of protein biogenesis, from transcription, translation and quality control will be described. Next, the of choice of expression plasmid, recombinant bacterial strain, culturing conditions and their impact on recombinant protein yields will be explored. Finally, description on how alteration of these native processes or adaptation of principles from them can be used to craft genetic fragments will be explored. These fragments can then be combined into genetic tools which can improve recombinant protein yields in *E. coli*.

Transcription in *E. coli*

Transcription in *E. coli* is the process where RNA is transcribed from DNA. It starts by the recruitment of the RNA polymerase. The polymerase facilitates bonds formation between nucleotides which are complementary to a DNA template. This enzyme is composed of five different subunits: α, β' β, ω and σ (1). The σ subunit is subject to exchange upon different changes or requirements in the cell. It is not essential for the polymerases ability to catalyse the bond formation between nucleotides; however, it has been shown to be important in identification of the proper polymerase binding site on DNA (2). This is of utmost importance as successful RNA transcription starts upstream of a gene. Initiation of gene transcription begins with the RNA polymerase binding to the promoter in a so called Closed- promoter complex (3). It is so called as the DNA is not uncoiled yet. In the next state, which is the Open-promoter complex, the RNA polymerase unwinds approximately 15 nucleotide base pairs around the +1 site. In the Open-promoter state, the DNA is now single-stranded and can now be transcribed (4).

Free nucleotides form complementary pairs and during the binding of the first 10 nucleotides, the σ factor remains bound to the RNA polymerase to ensure continued fidelity of transcription. After this, the σ factor departs, being replaced by an elongation factor, NusA and the RNA polymerase can continue to uncoil the DNA downstream, transcribe the gene of interest and finishes by re-coiling the transcribed DNA. The transcription comes to a halt when the RNA polymerase encounters a termination signal. This signal is often composed of a GC rich RNA segment, followed by at least four sequential A's (5). This sequence of nucleotides allows formation of a stable stem loop to be formed in the RNA. This, in turn, sterically hinders the RNA from binding to DNA, which leads to the DNA being released from the polymerase, and the subsequent departure of the polymerase. Termination of transcription can also occur due to specific protein interactions with DNA.

There are both positive and negative regulators of RNA transcription. The presence of a certain metabolite can influence up- or downregulation of RNA transcription of certain genes (6). One example of transcriptional regulation is the regulation of the *E. coli* lac operon. In the absence of lactose, the LacI repressor binds to lac operator binding sites, lacO3 and lacO1 and coils DNA. This causes a steric hindrance and inhibits RNA polymerase from binding to the lac promoter (7,8). When lactose is present, it binds to the LacI repressor, causing a conformational change which makes it detach from the promoter, thus allowing for transcription initiation (9).

Protein translation in *E. coli*

The translation of an mRNA to a protein is aptly called protein translation. The process can be divided into three steps: 1) initiation, 2) elongation and 3) termination. Below the molecular details of each step are described in detail.

Translation initiation

The initiation of protein translation occurs when an mRNA, containing a start codon (AUG or other) is recognized by the 30S subunit of the ribosome. When this happens, a formylated methionine (f-Met) is recruited as well as one of the three initiation factors (IF1, 2 and 3) (10). IF2 is highly specific to the f-Met tRNA, and f-Met does not bind to the other IFs. All these components together form the so-called initiation complex. IF1 and IF3 are also considered to be part of the initiation complex, however, they do not have a direct role in the initiation of protein translation (11). Their primary function is to maintain a certain pool of 30S subunits available for initiation of translation. They achieve this by the IF1 facilitating dissociation of the 70S ribosome (fully formed ribosome) and IF3 moves in to keep the two ribosomal subunits (50S and 30S) separated by binding to the 30S ribosomal subunit (12). IF3 also has a proof-reading function as is hinders initiation complexes which contain non-start codon tRNA from initiating translation. The joint efforts of IF1- and 3 promote the formation of the f-Met-tRNA and IF2 complex to the 30S subunit. Due to the properties of the proteins that are involved in the initiation complex, it follows that all polypeptides in prokaryotes have an f-Met as the first amino acid in the polypeptide chain (13).

After all the above mentioned components have come together to assemble the initiation complex, the 50S subunit binds to the 30S initiation complex. This is allowed due to: GTP hydrolysis of IF2 which positions the f-Met-tRNA into the P-site, displacement of the IF3 and release of IF1 and IF2 (14). The hydrolysis of GTP is an

irreversible step and results in a conformational change in the 30S subunit which facilitates further interaction and bond formation between the ribosomal subunits, stabilizing the 70S ribosome. The now 70S ribosome loaded with the mRNA-fMet-tRNA is now ready to be elongated (15).

Translation elongation

After the fMet-tRNA has bound to the P site of the ribosome and a new aminoacyl-tRNA (aa-tRNA) enters the A site of the ribosome, initiation is turned into elongation of the polypeptide chain. Elongation of the polypeptide involves multiple processes which occur simultaneously. The ribosome decodes the mRNA and aminoacyl-tRNAs (aa-tRNAs) are delivered to the ribosome by the elongation factor (EF) Tu bound to GTP (15,16). The EF-Tu-GTP-aa-tRNA complex interacts with the bL12 stalk in the ribosome. Upon proper codon: anti-codon interaction, hydrolysis of GTP occurs, which allows for a conformational shift of the EF-Tu-GTP-aa-tRNA complex and subsequent release of the aa-tRNA. This allows for the aa-tRNA to properly position into the A-site and EF-Tu-GDP dissociates for the 70S ribosome (17).

Once the two aa-tRNAs are positioned in the A- and P-site, the ribosome can manipulate surrounding molecules and structure within itself for a nucleophilic attack of the amino group of the aa-tRNA on the carbonyl carbon on the ester bond in the peptidyl-tRNA. This reaction results in peptidyl bond formation. After the peptide bond is established, the ribosome undergoes a conformational change which primes for translocation of the free tRNA. However, the actual translation does not occur spontaneously, but at the cost of GTP hydrolysis by EF-G (18).

Translation termination

Once the polypeptide has been elongated and the ribosome encounters a stop-codon, release factors (RF) 1 or 2 (depending on the stop-codon) are recruited and hydrolysis of the peptide chain from the tRNA occurs in the peptidyl transferase centre in the ribosome and the conserved GGQ motif in RF1/RF2 (19). To have turnover of RF1/RF2, a third release factor is used, RF3 in a GDP bound form. It accomplishes this by binding to the RF1/RF2-ribosome complex after the peptide chain has been cleaved and a location shift of RF1/RF2. The RF3 releases GDP, binds to GTP, RF1/RF2 dissociates, and upon GTP hydrolysis RF3 leaves the ribosome for ribosome recycling to take place (20,21).

Protein folding and quality control in *E. coli*

An *E. coli* cell can fold, refold or degrade proteins. These processes occur in the two main compartments of the cell: 1) the cytosol and 2) the periplasm.

Cytosolic folding and quality control systems

Folding of native *E. coli* proteins can either take place after the protein is fully translated or while it is undergoing active translation, in the ribosomal exit tunnel (22,23). Some small proteins can be fully folded upon emergence from the ribosome exit tunnel (24–26). The ribosome has been shown to influence the folding of the protein with its negatively charged surface. It can act as a chaperone, stabilizing disordered regions of proteins (27,28). Proteins that fold into their native structure with little influence from the ribosome can sometimes fold on their own in the cytosol. This is not the case for all proteins, as many need folding assistance from chaperones. As to this day, we know that the three main chaperone-complexes are: GroEL/GroES (29), DnaK-DnaJ-GrpE (30) and Trigger factor (TF) (31). All these chaperones engage newly synthesised proteins (at different stages) as they emerge from the ribosome.

Another aspect of the quality control system in the *E. coli* cytoplasm are disaggregases, proteases, and holdases. Disaggregases, such as *E. coli* ClpB work together with chaperones DnaK/DnaJ and bind to agglomerations or aggregates of proteins and allow for their disentanglement (32). ClpB consists of two hexameric rings and is a member of the ATP dependent AAA+ family (33). Disentanglement of target proteins is achieved by binding to aromatic side chains in the target protein upon hydrolysis of ATP. This hydrolysis causes a conformational change within ClpB and results in ClpB ‘pulling’ the bound target protein into its core. This subsequently displaces the target protein from the larger protein aggregate (34). The pulling and threading actions of ClpB is continued until the target protein emerges from the other side of the

ClpB pore. There the protein can either re-fold (either by itself or with the aid of chaperones) or be degraded by proteases into smaller peptides which can be repurposed to amino acids. The ClpB disaggregase can therefore work in unison with proteases such as ClpXP, ClpAP for turn-over of disagreggated proteins (35).

Other proteases know to be important for protein homeostasis, especially membrane proteins, are membrane bound Lon (36,37) and FtsH protease (38,39). Disaggregases and proteases work together with holdases, such as IbpA/B, which will bind to proteins associated with aggregates and block them from interacting with other proteins not yet ensnared by the aggregates. All the aforementioned quality control players are central in maintenance of protein homeostasis. This is particularly important in the event of environmental stresses, such as a heat shock. A more detailed description of the heat shock response is below.

The heat shock response and inclusion body binding protein

Upon a heat shock event (42°C), many proteins in *E. coli* become denatured. This in turn leads to the formation of large, insoluble bodies of aggregates, called inclusion bodies. If native proteins are trapped in these bodies, they will not be repurposed fast enough to ensure cell viability. Moreover, the inclusion bodies will drag in other proteins, as their hydrophobic surfaces may not be protected by chaperones or properly folded. This leads to the size of the inclusion bodies increasing rapidly and compromising cellular function and integrity.

To deal with inclusion bodies, *E. coli* have a heat shock response (40), which is governed by the σ factor 32 of the RNA polymerase (41). The transcription of σ32 factor (RpoE) is under several promoters, one being another sigma factor, σ70 (42). This means that σ32 is present in low abundance in non-stressed conditions. These levels are post transcriptionally regulated as σ32 is

sequestered by the chaperone complex DnaK/DnaJ/GrpE (during non-stressed conditions) (43), and the complex is continually trafficked to the inner membrane protease FstH, where it is degraded (44). Upon a heat shock, the translated σ32 is released, as the DnaK/DnaJ/GrpE are targeted to aggregated proteins. σ32 can subsequently bind to the RNA polymerase (Figure 1). This facilitates the transcription of gene for chaperones, proteases and inclusion body binding proteins A/B (IbpA/B) (45,46). These proteins act as holdases, which bind to the inclusion bodies, blocking hydrophobic interactions with other proteins, allowing for downstream processing. In non-stressed conditions, IbpA negatively regulates its own translation, by binding to the mRNA, of σ32, thus blocking translation (40).

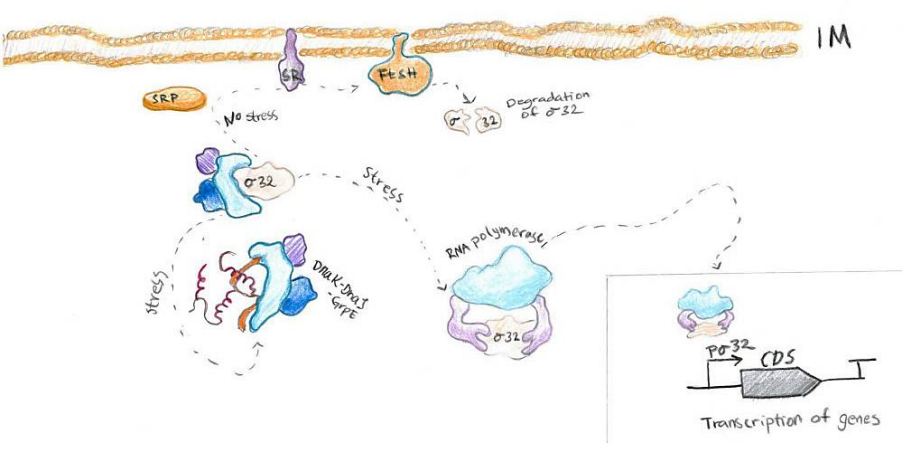


Figure 1. Illustration of the σ32- dependent heat shock response. In the absence of a heat shock, the DnaK-DnaJ-GrpE chaperone complex binds to σ32 factor, inhibiting it from binding to the RNA polymerase. The chaperone complex, together with Signal Recognition Particle (SRP) guilds the σ32 to the SRP receptor (SR) and FtsH is recruited to degrade σ32. Upon induction of heat shock, the DnaK-DnaJ-GrpE chaperone complex is diverted from the σ32 to instead bind to misfolding proteins. This enables σ32 to bind to RNA polymerase, subsequently leading to transcription of genes under the σ32 promoter on the genome. Image inspired by (47).

Periplasmic folding and quality control systems

Proteins destined for the periplasm need to be translocated over the inner membrane of *E. coli*. The post-translational translocation is performed primarily by the membrane spanning SecYEG translocon in co-operation with the soluble SecB chaperone. When the polypeptide beings to emerge from the ribosome, Trigger Factor (TF) will bind, blocking Signal recognition particle (SRP), and further facilitate SecB binding and priming the polypeptide for translocation by ATP dependent SecA. After translocation, LepB will cleave the signal sequence and the mature protein will be allowed to fold (as signal sequences are known to obstruct folding of pre-protein) (31,48).

Once the pre-protein has been translocated and the signal sequence cleaved, folding is facilitated by a plethora of chaperones depending on the needs of the proteins. For disulphide bond formation, the chaperone DsbA is the main chaperone. The conserved CXXC motif in the catalytic site allows for coordination of a cysteine on the substrate polypeptide with DsbA. Later, upon re-oxidation of DsbA by DsbB, the cysteine bond is transferred from DsbA to an internal cysteine in the polypeptide (49–52). If the polypeptide contains multiple cysteines in need of isomerization, incorrect cysteine bond formation or upon misfolding, DsbC/D in conjunction with NADPH-TrxA are employed for this task (53,54). The homodimeric DsbC is the main isomerase, it also has the conserved CXXC motif (55) and acts as a chaperone (56) Membrane spanning DsbD acts as the re-oxidiser for DsbC (53). There are other chaperones present in the periplasm besides the Dsb proteins, such as the PPIases PpiA, FkpA, PpiD and SurA (57–60). Upon misfolding of proteins in the inner membrane or the periplasmic space, a stress response may be activated, such as the Cpx envelope stress response.

Cpx envelope stress response

CpxA and CpxP have been shown to sense the overexpression of secreted lipoprotein, and influence the expression of genes encoding CpxP, Spy, DsbA, PpiA and DegP for the management of periplasmic stress (61–64). The Cpx response governs the expression of the σE RNA polymerase sigma factor, which is involved in the translocation of proteins to the outer membrane (65). When the integrity of inner membrane-or periplasmic proteins are threatened, the Cpx stress response is activated (66,67). This activation takes place by the CpxP protein, a chaperone-like protein located in the periplasmic space, leaving the transmembrane kinase CpxA to deal with misfolded proteins in the periplasm (44). This migration of CpxP from CpxA allows for phosphorylation of CpxR by the cytosolic part of CpxA (68,69). Upon phosphorylation, CpxR will bind to specific promoters on the genome and recruit RNA polymerase, thus initiating the production of chaperones and proteases to deal with the periplasmic/ inner membrane protein stress (Figure 2)(70).

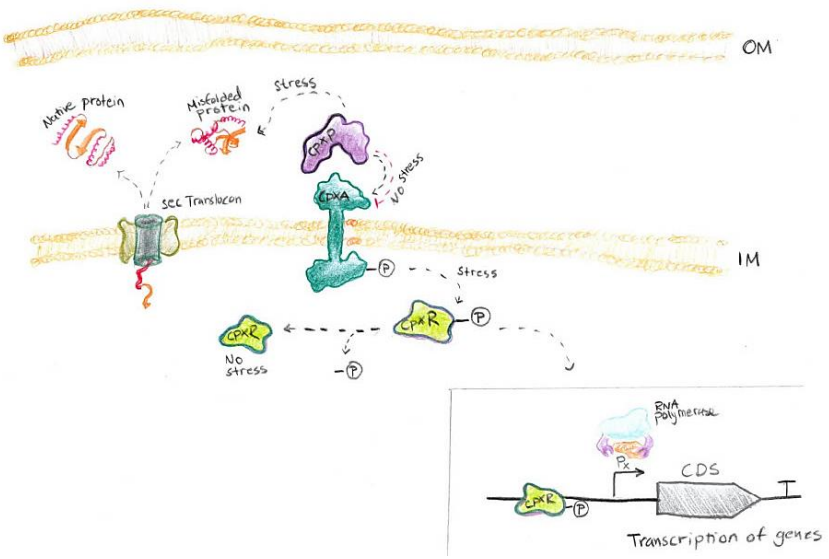


Figure 2. Illustration of the CpxP-CpxA-CpxR periplasmic stress response. In the absence of accumulated misfolded proteins in the periplasm the CpxP protein binds and inhibits the transmembrane kinase CpxA. When misfolded proteins are present, the CpxP chaperone is recruited to them. Without inhibition from CpxP, CpxA can phosphorylate CpxR, which becomes a transcriptional activator, binding to promoters on the genome and recruiting RNA polymerase to transcribe genes for components needed to combat the misfolded protein problem in the periplasm. Inspiration was taken from (71)

Recombinant protein production in *E. coli*

Production of non-native protein in single cell systems have long been a pursuit for both industry and academia. It may be for therapeutic use, enzymatic characterization, structural determination or other. Irrespective the reason, one thing is certain, larger amounts than what can be found in native sources are often needed. To achieve this, proteins are expressed recombinantly using genetic tools and strains. Plasmids are typically used to express recombinant proteins upon transformation into host cells. They are composed of multiple genetic modules with different purposes (Figure 3). Some of the essential genetic modules are explored below.

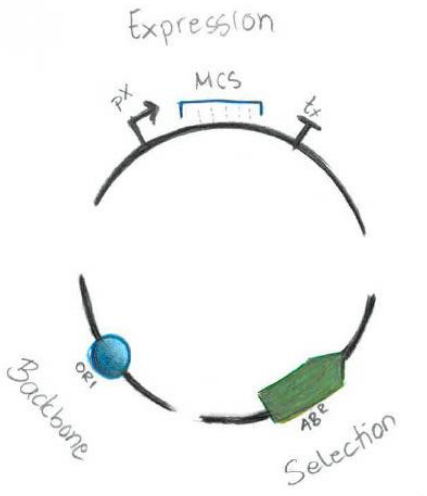


Figure 3. Illustration of the basic architecture of expression plasmids. Expression plasmids can be condensed to three main genetic fragments (clock-wise direction): 1) the Expression fragment, which contains the inducible promoter, the Multiple cloning site (MCS) and terminator. 2) The Selection module, which, in most cases, contain the antibiotic resistance fragment to allow for selection of plasmid carrying cells. 3) the Backbone fragment, which contains the origin of replication (ORI) and its cis-transfer elements.

Replicons

All plasmids contain a replicon, which contains the origin of replication (ORI). This genetic module is responsible for propagating the plasmid and determines the copy number of the plasmid within the host cell (72). In the pET system plasmids, the ColE1 derived pMB1 ORI was incorporated, which gives a plasmid copy number of 15-60 plasmids per cell (73). Wild type ColE1 ORI provides a copy number of 15-20 copies per cell (74) (found in pQE plasmids). Due to similarity of ORI and utilization of the same parts of *E. coli* replication mechanism, pET and pQE plasmids are incompatible. There are other ORI which can be used simultaneously with pMB1 and ColE1 ORIs, such as the 10-12 copy number p15A ORI (75), carried in the pACYC plasmid. For expression of highly toxic proteins where low plasmid copy numbers are needed (~5 copies per cell), the pSC101 ORI can be used (76).

Depending on the ORI used, the recombinant protein yields will change as well as the amount of antibiotics needed to sustain selection of plasmid bearing cells. In *Paper II* a minimal version of the Tn3.1 resistance fragment was constructed ($Tn3.1^{MIN}$) and tested in vectors with different ORIs. It was observed that $Tn3.1^{MIN}$ was compatible with medium and high copy number ORIs, but incompatible with low copy number ORIs.

Promoters

To obtain mRNA of the gene of interest and to have transcriptional control of mRNA production, the expression plasmid needs a genetic module called a promoter. This is a site recognized by RNA polymerase. The promoter has two binding sites, six nucleotides each, located -35 and -10 nucleotides upstream from the transcription initiation site (defined as +1) (77). Although not all promoters have the same nucleotide sequences in the -10 and -35 initiation sites, they have been shown to have high sequence similarity, thus designated as ‘consensus sequences’. These have

been shown to influence the transcription efficiency. Promoters void of consensus sequences have a lowered transcription efficiency. Due to the sequence dependency of promoters, mutations in the consensus sequences can strongly impact the "strength" of the transcriptional output and mRNA yields. The specificity of RNA polymerase binding to the DNA is achieved by Sigma factors, which guide the polymerase to the -35 and -10 consensus sequences. In the absence of Sigma factors, the polymerase can bind to DNA in a non-specific and low affinity manner (6).

In expression plasmids which are designed on genetic modules from the lac operon, the induction of gene transcription is govern by presence of lactose (or Isopropyl-β-D-thiogalactopyranoside, IPTG) and absence of glucose (78). Repression of the lac operon is achieved by the LacI repressor binding to lac operator binding sites (lacO3 and lacO1) and coiling DNA, thus making the promoter unavailable to the RNA polymerase. The operators are essential for the proper repression of the system pre-induction (7,8). Upon binding of inducer (such as lactose or IPTG) to the repressor, it detaches from the promoter, thus allowing for transcription initiation (9). The T7plac used in the pET system (Figure 4) harbours genetic modules from the lac operon. Dubendorff J. W. et al 1991 introduced the φ10 promoter of the T7 bacteriophage in the early pET vectors. Downstream of the T7p, a lacO1 was positioned to allow for transcriptional repression of the T7p by LacI in the absence of inducer (79). A common architecture of T7plac is shown in Figure 4.

Upon introduction of the lacO1 sequence, four nucleotides were removed by restriction enzyme digestion. These four nucleotides were re-introduced by Shilling et al. 2020 (80), thus reconstituting the consensus T7 promoter ($T7p^{CONS}$). This increased the strength of the promoter and increased the yield of recombinant protein. Although being the most widely used system to express recombinant proteins, the T7 system has some bottlenecks. One is the result of the design being based on the lac operon. Pre-induction leakiness is often observed in high copy number plasmids and is particularly

problematic when used to express proteins which pose a threat to host cell integrity.

To circumvent these issues, numerous attempts have been made to make the T7-system more regulatable. Site directed mutagenesis in the placI increased the amount of LacI mRNA and increased repression of T7p, thus making the placI^{q} promoter (81). The pLysS plasmid encoding the T7 RNA polymerase inactivating lysozyme was made to supress the T7 system in absence of inducer (82). Wycuff D. R. et al 2000 (83) placed the T7 RNA polymerase under the control of L-arabinose inducible promoter to allow for tighter regulation of recombinant protein expression in the T7 system. In the BL21AI stain, the T7 RNA polymerase was introduced into the *araB* locus of *E. coli*, thus making the expression of T7 RNA polymerase titratable by L-arabinose induction (84). Du F. et al. 2021 (85) introduced $P_{araBAD,}$ P_{rhaBAD} and P_{tet} in the *λDE3* fragment of BL21(*DE3*). This made it possible to have T7 RNA polymerase was under the control of other promoter/inducers systems. They then compared the leaky expression (pre induction), protein production (post induction) and plasmids stability in the different BL21(*DE3*) strains. Leaky expression of EGFP was reduced in all but not BL21(*DE3::* P_{T7plac}). Expression of EGFP was the highest in BL21(*DE3::*P_{tet}) and BL21(*DE3::*P_{rha}). Plasmid maintenance efficiency post induction of EGFP was the lowest in BL21(*DE3*:: P_{T7plac}) compared to the other three constructs. Protein production of four membrane proteins showed that BL21(*DE3::*P_{rha}) was comparable or more efficient than BL21(*DE3::* P_{T7plac}). These changes presented are but a few of many attempts to improve and expand the applicability of the pET system which was first developed by Studier F. W. in the 1980s.

The araBAD promoter used in (83–85) is taken directly from *E. coli* native operon (75). In the absence of L-arabinose, the DNA in the vicinity of the promoter is curved by the AraC regulator protein due to binding to the I_2 and O_2 sites. Upon binding of L-arabinose to the AraC protein, a conformational shift takes place. AraC binds to I_1 and I_2 sites upstream of the promoter, straightening the DNA and allowing RNA polymerase to access the promoter region (86). The

most common design found in expression plasmids (Figure 4) is that the AraBAD promoter is placed upstream of the multiple cloning site and downstream of the AraC regulator protein upstream of the (which is orientated in the opposite direction).

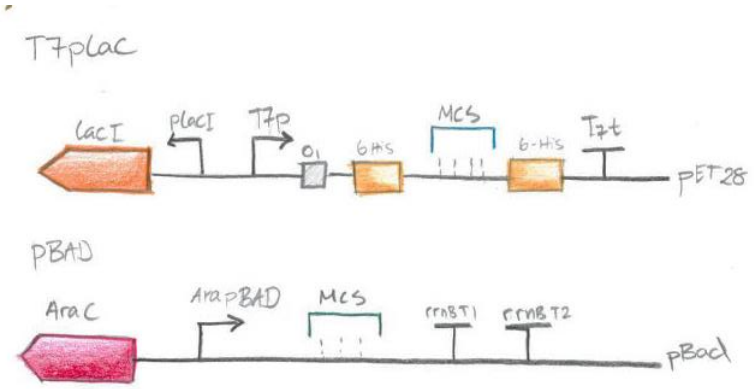


Figure 4. Illustration of the $T7_{plac}$ genetic module (top) present in the pET28 expression plasmid and the P_{araBAD} genetic module (bottom) present in the pBAD expression plasmid.

Translation initiation region on mRNA

mRNA contains the genetic code of a protein, but also elements which are important for initiation of protein translation. One such element is the Shine-Dalgarno (SD) sequence, located upstream of the start codon (87) (Figure 5). In *E. coli*, the SD spacer has been observed to be located five to nine nucleotides away from the Start codon for optimal efficiency, with variations depending on the gene in question (88,89). Translation initiation has been shown to be less efficient when the spacing between the SD and Start codon is smaller than four nucleotides and greater than ten nucleotides (90). The nucleotide composition is usually rich in adenosine and guanidine (91), and a conserved consensus sequence is found to be 5'-AGGA-3' (92).

The SD sequence allows for initiation of translation by anchoring an anti-SD in the16S rRNA located in the 30S ribosomal subunit to the mRNA. The SD sequence allows for the mRNA to keep an unfolded state and facilitates the ribosome-mRNA alignment of the Start codon to the ribosomal A-site. The SD: anti-SD interaction is maintained even after the first peptidyl bond has been formed and can sometimes be re-initiated in certain stages during the elongation process to restabilize the ribosome-mRNA interaction (13,15).

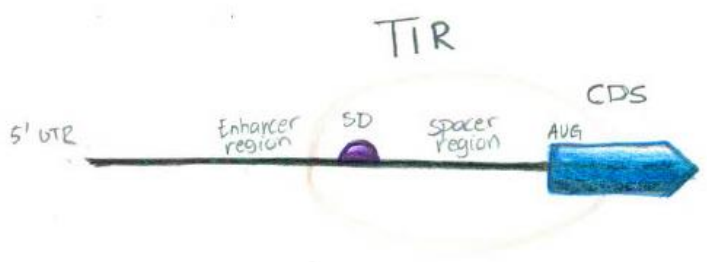


Figure 5. Illustration of the linearized mRNA and the features present in the upstream region of the Coding sequence (CDS).

The Translation Initiation Region (TIR), describes the ribosomal footprint; The stretch of nucleotides located -20 to +15 nucleotides around the translation Start codon. By manipulating the TIR region, it is possible to alter the translational output of a mRNA, without changing the amount of transcribed mRNA (80,93–95). In *Papers II, III* and *IV* manipulations of the TIR were performed to either decrease the protein yields of antibiotic resistance proteins (as in the case of Tn3.1^{MIN} (*Paper II*) and Tn903.1^{MIN} (*Paper III*) or increase protein yields (as in the case of TIR^{EVOL} (*Paper III*) and P_{CpxP}-mCherry$_{ASV}^{OPT}$ (*Paper IV*)).

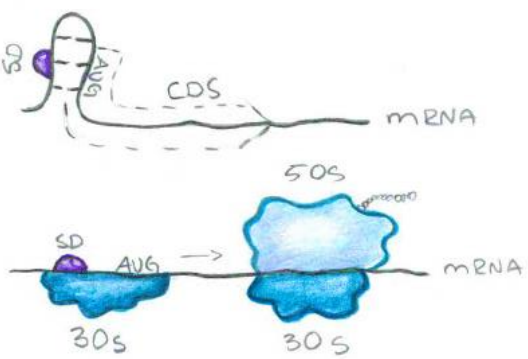


Figure 6. Illustration of the folding of mRNA around the Translation Initiation Region. If mRNA has a secondary structure- known as hair pin, which sequesters the SD and the AUG, translation efficiency is compromised (top). If mRNA secondary structure is minimized, initial 30S ribosome subunit binding is enabled. This is followed by 70S ribosome assembly at AUG, translation initiation and ultimately leads to increased translation efficiency (bottom).

Antibiotic selection fragments

Upon transformation of plasmid into the host cell, the host cells must be prompted to propagate the plasmid to future generations. This can be achieved by using an antibiotic selection fragment in the plasmid. These are DNA fragments excised from other organisms and that allow for growth in the presence of antibiotics which are normally lethal to *E. coli*.

One of the most used antibiotic resistance cassettes is the Tn3 fragment (Figure 7) discovered in the 1960s (96). It was isolated from the Tn3 transposon on the R1 plasmid in *Salmonella paratyphi* B. This fragment harbours the *bla* gene, encoding the TEM-1 β-lactamase enzyme (97), which inactivates β-lactam antibiotics by cleaving the amide bond in the β-lactam ring (98). A commonly known issue when using this fragment is that the β-lactamase enzyme (which is targeted to the periplasm) leaks into the culture media. This leads to inactivation of the β-lactam antibiotics (99).

The long-term consequence is that selection in the culture is lost within 3h (100,101). Another commonly used antibiotic resistance fragments is the Tn903 fragment from *Streptomyces kanamyceticus* (102), harbouring the *aph* gene which encodes the aminoglycoside-3'- phosphotransferase type Ia enzyme (Figure 7). The inactivation of aminoglycoside antibiotics is achieved by phosphorylation of the 3' carbon in the glycoside ring. This leads to steric hindrance and lowered binding affinity of mRNA to the A-site of the small ribosomal subunit (103). Plasmids harbouring this fragment allows for selection using kanamycin (104). In *Paper I*, the plasmid maintenance ability of Tn3.1 and Tn903.1 resistance fragments were compared against each other to determine which one provides the most reliable plasmid maintenance in *E. coli* cultures.

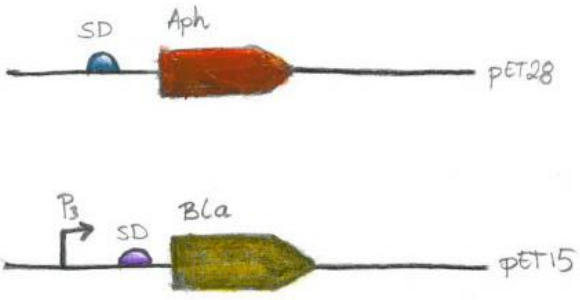


Figure 7. Illustration of antibiotic resistance fragment Tn903 encoding for the aminoglycoside-3'- phosphotransferase type Ia enzyme (Aph) protein (kanamycin resistance) in pET28 (top) and antibiotic resistance fragment Tn3 encoding for the TEM-1 β-lactamase (Bla) enzyme (ampicillin resistance) in pET15.

Antibiotic-free selection fragments

The emergence of antibiotic-resistant bacterial strains has placed a demand on society to decrease the use of antibiotics to minimize emergence of new antibiotic-resistant strains. The World Health Organization has flagged that the leading cause of death by 2050 will be infection by antibiotic resistant microbes (105). Due to this,

biomanufacturing (which includes recombinant protein production in microbial cell factories) has been placed under strong control and legislations within the European union. All biomanufacturing processes are discouraged and sometimes prohibited from containing antibiotics. In cases where antibiotics are permitted to be added, downstream processing becomes more laborious as removal of antibiotics from the final product is often a requirement (106,107). Use of kanamycin and tetracycline selection fragments in strains producing recombinant biopharmaceuticals are currently accepted by health authorities. However, trace amounts of β-lactams can induce allergic reactions (108) and are thus strictly prohibited. The antibiotic-free selection systems rely on other types of mechanism to ensure that only plasmid carrying cells, encoding for the recombinant protein are viable in the cell culture. Some antibiotic-free selection strategies are presented below.

Essential gene complementation is, as the name suggests, when an essential gene is deleted or suppressed in the host genome. The essential gene is reintroduced on the expression plasmids and used as a selection marker. The *dapD* gene complementation been a preferred method by many researchers. The pathway is well known, and mutation strains absent plasmid encoded *dapD* don't survive, as the absence of *dapD* make cells deficient in the Lysine precursor diaminopimelate. This leads to defective cell wall synthesis and cell lysis upon cell division (109). The essential gene complementation selection is the most common antibiotic-free selection system.

Separate-component-stabilization or the Poison/antidote selection mechanism for antibiotics-free selection presented by Szpirer C. Y. et al. 2005 is based on introduction of the poison gene, *ccdB*, encoding for the protein binding gyrase into BL21(*DE3*) host genome. The binding of CcdB to DNA gyrase subunit A forces it into a complex with broken double stranded DNA, resulting in cell death. The plasmid encoding the antidote gene product of the *ccdA*, allows for inhibition of the CcdB. The CcdA will block CcdB by forming a complex with it, as well as removing CcdB from complex with GyrA and facilitating complex formation with itself (110).

Although antibiotics-free selection allows for better plasmid retention in cell cultures, it is evident that it is less flexible than the antibiotic dependent selection systems, as there is an enforced strain dependence. Furthermore, there is a dependence of plasmid encoding the anti-toxin or essential gene prior to introduction of the recombinant plasmid. Nevertheless, development of systems which have efficient plasmid maintenance is important, as plasmid maintenance can, and often is, linked to recombinant protein production yields.

Bacterial host strains

After obtaining a plasmid with appropriate genetic modules, the host strain needs to be chosen. Production of recombinant protein and small molecule compounds (such as melanin (111), terpenoids (112) and amino acids (113)) can take place in different bacterial strains.

E. coli are gram-negative prokaryotes which live in the gut of mammalian and reptilian animals (114). The first strain, which was to become a lab strain, was isolated from stools samples belonging to a patient with diphtheria (115). The discovery of this stain led to the development of the strains most commonly used in industry for the production of biotherapeutics, K12 strains. Some of these strains were made by the National Institution of Health (NIH) under safety guidelines. One of the common K12 derivates is the W3110 strain (116).

One downside with these strains is that they produce high amounts of acetate, which stunts cell growth, especially in large scale fermentations using glucose as a carbon source (117). To circumvent this problem, other strains have been developed which mitigate this issue, such as the B strain (118). Besides the benefit of lower acetate production (119,120), the B strain is used as a model for testing restriction modification systems and recombinant protein

expression in both industry and academia. These strains were further engineered to contain the *λDE3* fragment from the T7 phage in the BL21 genome (121). The *λDE3* fragment encodes for the powerful phage T7 RNA polymerase, under the P_{UVlac5} promoter. The T7 RNA polymerase is approximately five times faster than *E. coli*'s own RNA polymerase (101). Furthermore, it is selective to the T7 promoter in plasmids, which have also been adapted to mimic the features of the native lac operon. The whole system is inducible by the lactose homologue IPTG. As the lac promoter, it is not titratable, and upon induction acts as an on/off- switch (122). The intended use of the T7 promoter was to produce large amounts of recombinant proteins in *E. coli* (101,123) and upon overexpression of non-cytosolic proteins cellular stress becomes imminent (124–126) .

As B-strains are often used in recombinant protein expression, they have been developed to contain lower amounts of proteases in the inner- and outer membranes to lower the probability of degradation of recombinant protein and higher membrane permeability. They have also been further developed for better expression of membrane proteins, such as in C41(*DE3*) and C43(*DE3*) (127). In the C41(*DE3*) strain the strong P_{UVLac5} has reverted to a weaker P_{lac}, due to mutations in the -10 region (124). In the C43(*DE3*) strain, amino acid mutations in the LacI repressor increased binding affinity of the LacI repressor to the lacO1 operator. This hampers T7 RNA polymerase expression upon IPTG induction, as revealed by genomic analysis (128). The alterations result in a weaker transcriptional output which is suitable for production of unstable proteins prone to aggregate at high titres or membrane proteins (127). Further BL21(*DE3*) derivatives have been isolated which are more suitable for overexpression of non-cytosolic proteins. Some of these strains, though different in genetic profiles, are C44(*DE3*) and C45(*DE3*) (129). BL21(*DE3*) pLysS, containing the T7RNA polymerase inactivating lysozyme gene (82). Lemo21(*DE3*) which allows for titration of T7 lysozyme by placing it under a Rhamnose-inducible promoter (P_{rha}) in the pLemo plasmid (130). And Evo21(*DE3*) which has truncations in the RNAse E, allowing for increased RNA stability and subsequently, ability to produce toxic

proteins (131). These are but some the strains which have been developed by the community out of necessity.

As the origin and subsequent complexity of the proteins of interest have increased, the need for strains which have a wide array of properties has increased. The B-strains have been engineered to be able to fold proteins with disulphide bonds in the cytosol (i.e. the Origami strain) (84,132–134). The titratability of IPTG has been improved in the Tuner strain (84). The Rosetta strain has been supplemented with tRNAs not common to *E. coli* for production of proteins of eukaryotic origin (84,135). These are some examples to name a few.

The wide array of host strains available in the pET system alone (not to mention non- pET system strains) imposes a requirement on the user to know the properties of the protein of interest prior to embarking on protein production in *E. coli*, as an ill-suited strain can result in insufficient protein yields depending on the expression plasmid and downstream application. In *Paper III* strains matching with pET28 and derivatives harbouring combinations of genetic modules in the T7plac and TIR was performed to understand which genetic modules were compatible with which strain depending on the desired outcome.

Cell culture growth conditions

After the expression plasmid and host strain has been chosen, the culturing conditions need to be considered, as they also have an influence on the protein yields. As was explained in the previous section, the BL21(*DE3*) based strains have been engineered to produce less acetate (119,120), but they still produce of metabolic bioproducts as cells are growing.

Depending on the composition of cell culturing media, more

precisely the carbon source used, different bioproducts will be produced which can change the properties of the media. For example, growth of *E. coli* cells using glucose as the main carbon source leads to acidification of the media, which hampers the cellular growth (136–138). The acidification is the result of acetate production and overflow metabolism due to acetate excretion (139). During acidification of the media, *E. coli* activate stress responses to combat the acidification, such as the general stress response (governed by *rpoS*, encoding σS) (140,141). Growth in alkaline environments also cause stress to cells (140). Alkalinization of cellular media occurs when peptones (amino acids) are used as a carbon source, such as in the case of Luria Broth (LB) (142). LB was first developed to support rapid growth of bacteriophages without depletion of the host cell resources (143). Although LB supports rapid growth of *E. coli* cultures, it leads to low growth viability, high death rates within the culture and low adaptability in starvation conditions (144). The alkalinization of LB media lead to increased glycosylation of proteins, which negatively affected cell viability (145–147). Having said this, there are benefits to culturing cells in LB, as the primary carbon source is not sugar, but instead peptones (amino acids). Work presented by Baez A. et al. 2019. (148) has shown that the addition of amino acids in culturing media redirected the metabolic pathways to support higher density growth.

To mitigate some of the issues described for LB, alternative rich media have been developed. Terrific broth (TB) is a richer medium than LB and was developed to support greater plasmid yields by increasing the steady state duration (149). Super broth (SB) is another rich media, which contains soy hydrolysate instead of tryptone (150). Both TB and SB are buffered by phosphate salts and are supplemented with glycerol. The buffering of media allows for stabilization of pH, which decreases the glycosylation of proteins, thus decreasing cell death (145).

Not only does the culturing media effect the outcomes of recombinant protein production experiments, but also growth vessel. A commonality in biochemical and molecular biology research is to cultivate cell cultures in shaker flasks either

exclusively or as a pre-test before moving on to fermentation-based technologies (which will not be mentioned in this thesis). The shaker flasks are often limited in aeration capacity as new oxygen is not supplemented throughout the experiments. Users often must manipulate the hardware used to culture recombinant *E. coli* to obtain better aeration. This entails 1) shape of the shaker flask (baffled- or smooth bottomed), 2) setting revolution speed of the incubator (speeds from 250-450 depending on the shaker flask used) (151,152), 3) use of gas-permeable lids (153) or 4) limiting media volumes to not exceed 10-20 % of the total capacity of the shaker flask (151,154). According to Kram E. K et al. 2014 (155) the volume of the culture and volume of the growth vessel had an impact on the amount of oxidative stress, estimated by quantification of β-Galactosidase production from P_{oxyR}-β-Galactosidase sensor. During the first 2 days of cultivation, the larger the culture media volume, the lower the activity of P_{oxyR} was. The authors link this to the slower pH alkalinization rate in the larger volumes.

As the culturing media and culturing vessel affect the yields of recombinant protein, so does temperature during growth. Culturing temperatures are for many protein production experiments the factor which leads to either abundance of target protein, or full growth arrest and host cells death (127,156). Manipulation of growth temperature leads to the slowing down of *E. coli* growth and all the molecular processes within the cell. This gives the recombinant protein time to fold, and lowers the probability of inclusion body formation (157–160). In *Paper IV*, it was observed that the growth temperatures influenced the activation of the two responses used in the biosensor upon induction of MalE and derivatives. Induction at 20 °C showed that the activation of P_{IbpA}-$GFPmut3_{ASV}$ (secretion limitation) occurs first followed by the activation of P_{CpxP}-$mcherry_{ASV}^{OPT}$ (periplasmic misfolding), as compared to induction at 37 °C where the stresses were inseparable.

Lastly, the inducer influences the outcome of a recombinant protein production experiment. If the inducer concentrations are low, the

yields of the recombinant protein often are insufficient. The same applies when inducer concentrations are high, but instead, the amount of recombinant protein produced overwhelms the host. This results in protein assimilating into inclusion bodies (161). Furthermore, issues with cellular health and plasmid instability are associated with higher induction protocols. Using IPTG can sometime be harsh. To mitigate this, autoinduction media, such as was done by Studier W. et al. 2005 (162) can be used. It consists of a mixture of glucose, glycerol and lactose. Glucose and glycerol are utilized to obtain higher cell densities and lactose for the induction of protein expression. In *Paper IV*, the influence of induction strength on cellular health was observed when titrating expression of hGH and $scFv^{Her2}$. The higher the inducer concentration, the more metabolic stress was put on the cell, resulting in poor biomass, protein production yields and plasmid maintenance. In *Paper I,* the effects of induction strength on plasmid maintenance in cell cultures were observed as a proxy by switching the host strain. The strength of the response determined the plasmid maintenance fidelity in cells cultures.

Growth conditions have not been explored in detail in the work presented in the thesis. However, the importance of the growth conditions cannot be ignored as a large body of literature has been accumulated. Furthermore, the impact of improper growth conditions on the recombinant protein yields are significant.

How to utilize innate E. coli stress responses to monitor and alleviate recombinant protein production stress

Since recombinant protein production often leads to misfolding of target protein, it is crucial to understand what the recombinant protein production setup does to the host cell. Using conventional methods, such as fractionating cells and analysing via immunoblotting, can be arduous and time consuming.

As was presented in a previous section, *E. coli* respond to environmental stressors, such as heat shock by inducing the heat-shock response. If a researcher understand the native processes that *E. coli* use to combat environmental stress and how to activate them, they can either 1) manipulate *E. coli* to activate protein quality control to keep the recombinant protein from misfolding, thus increasing protein yields or 2) take genetic fragments and implement them in tools which can be used to monitor when cells are stressed to guild users away from those conditions. If a researcher wishes to manipulate the growth conditions to force *E. coli* to activate stress responses to produce proteins in protein quality control this can be done in several ways. One way is to rapidly change temperatures (approx. 12°C differences) or grow in temperatures below steady state (20-37°C), as it will induce the cold shock response. The major protein in this response is the CspA protein, which has been shown to transcriptionally regulate other genes. CspA works together with other cold shock proteins CspB-D (163). Upon expression of cold-shock proteins, production of aggregation-prone protein in *E. coli* cell-free synthesis have been observed to solubilize insoluble proteins and increase yields (164). Another way to prime *E. coli* for recombinant protein production is the opposite to cold shock, by inducing heat shock. As described previously in the thesis the heat shock response is activated when temperatures reach 40-42 °C, which will increase the expression of molecular chaperones. Having said this, it becomes cumbersome the larger the culture volumes are to get efficient heating and subsequent cooling of the cultures for the sake of inducing chaperone production. Heat shock chaperones can also be achieved chemically by addition of benzyl alcohol (165).

Induction of stressful growth conditions to combat misfolded recombinant protein may seem counter-intuitive and monitoring of when stress is induced may be more appealing. As was previously mentioned, conventional detection methods to estimate when misfolding of recombinant protein occurs are time consuming. Therefore, a quick and *in vivo* screening method is useful for optimization of culturing conditions which promote lower cellular stress. This can be obtained by using a biosensor (Figure 8). Biosensors are fusion of genetic modules, such as promotes together with a protein gene which either emits fluorescence or has an enzymatic function with can be detected.

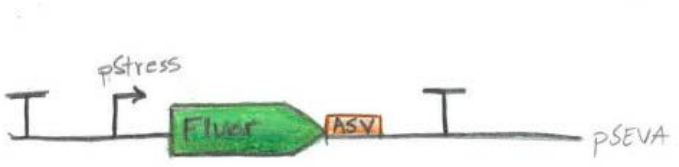


Figure 8. Illustration of the basic architecture of a biosensor. A promoter which is activated upon an environmental stress (p_{Stress}) can be fused to the coding sequence of a fluorescence protein (Fluor). To ensure constant turn-over and time-independent signal of the fluorescence protein, a degradation tag (ASV) can be used. The biosensor is insulated by engineering two terminator sequences on either side of the biosensor to mitigate transcriptional run-through and thus unspecific activation. Using the pSEVA backbone architecture (166)*, low copy number ORI can be chosen to keep the metabolic load of the plasmid minimal.*

Biosensor construction has been done by many labs. Compilation of a library of annotated promoters in *E. coli* fused to GFPmut2 harboured within a low copy number plasmid (either pUA66 or pUA139) was done by Zaslaver A. et al 2006 (167). These biosensor plasmids can be transformed into *E. coli* and be used to detect activation of transcriptional responses to a plethora of physiological changes within *E. coli.* Creation of biosensors detecting protein folding stress and protein translation efficiency was developed by Zuts A. et al. 2021 (168). They linked P_{IbpA} to a destabilized GFP

mutant, GFP-mut3$_{ASV}$. To detect inclusion body formation either due to heat shock or recombinant protein production, the GFPmut3$_{ASV}$ whole-cell fluorescence could be quantified. For detection of protein translation efficiency of a gene, the protein translation sensor was developed. The target protein gene is linked via translational coupling sequence to mCherry, which allows for quantification of red fluorescence from whole cells.

Besides the use of fluorescent proteins, biosensor for detection of cellular stress using enzymes have been developed. Fusions of promoters to β-Galactosidase have been employed for detection of stress in *E. coli* due to acidification or alkalinization of media (169) and insoluble protein formation in the cytosol (170). Fusions of promoters to bioluminescent luciferases have also been employed to detect cellular stress *in vivo* (171).

If neither manipulation of the growth temperature nor use of biosensor are applicable to increase yield of recombinant protein production, co-expression of chaperones encoded on separate plasmids can be used. Takara Bio Inc., Japan created a collection of inducible plasmids encoding for chaperones to alleviate misfolding of recombinant proteins under the induction of L-arabinose. The plasmids contain either GroEL/GroES, DnaK/DnaJ/GrpE (172) or Trigger Factor (173). deMarco A. et al. 2005 (165) observed that recombinant protein yields increased when co-expressing chaperones DnaK/DnaJ/GrpE and GroEL/GroES, disaggragase ClpB, and holdases IbpA/B.

Plasmid maintenance in *E. coli*

In the previous section, factors affecting the outcome of a recombinant protein production experiment were discussed. The same factors have also been seen to affect the plasmid maintenance in an *E. coli* culture. If the plasmid is poorly maintained within the culture, protein production could be compromised (as cells that don't have plasmid can't produce recombinant protein). In the following section, plasmid maintenance in recombinant protein production will be explored.

Plasmid maintenance within a culture is dependent on the propagation of plasmids from mother cells to daughter cells upon division. Approximately 1 in every 10,000 divisions results in a daughter cell devoid of plasmids (Figure 8), and if culturing conditions do not have selection against plasmid-free cells, they will survive (174). This phenomenon is most noticeable when using plasmids which have no active partition locus (such as pBR322 and derivatives). The partition (*par*) locus allows for equal distribution of plasmid in the nodes of daughter cells following division of *E. coli* (175,176). Plasmid-free daughter cells are only an issue, if they have a growth advantage compared to plasmid-carrying cells. This generally occurs upon induction of recombinant protein production, as plasmid-carrying cells have the metabolic burden of transcribing recombinant mRNA (177) and producing recombinant protein (178).

If the strain background is not compatible with the expression plasmid and the protein of interest harboured in the plasmid is 1) aggregation prone, 2) requires translocation into periplasm or 3) requires insertion into the cellular membrane, it often leads to cellular growth arrest (85,125,130,179–181), plasmid retention issues (179,181–183), or inability to get transformants, as observed by Dumon-Seignover L. et al. 2004 (182). In the study, they transformed C41(*DE3*) and C43(*DE3*) strains with the pRSETA plasmid harbouring 12 different protein genes. They observed that some recombinant protein constructs were toxic, thus resulting in

lack of viable transformants. In a study done by Sieben M. et al 2016. (184) they observed that the pRSET plasmid (amp^R) encoding YFG-IL-6, a yellow florescent protein tagged unto an Interleukin-6 protein, allows for high levels of leakiness, even in the BL21(*DE3*) pLysS strain. Upon induction of protein production for 24h at 30 ℃, the plasmid maintenance of the pRSET plasmid was 40%. These finding emphasises the importance on proper stain-plasmid matching for reliable plasmid propagation and plasmid maintenance. Besides the issues that recombinant protein genes harboured in the plasmid can cause issues to host cells, the plasmid copy number can affect the efficiency of plasmid maintenance, as higher copy number plasmids can cause metabolic stress on the host cell (178). This metabolic stress saps resources from the translation machinery and lowers yields of recombinant protein (185).

To ensure that the plasmid is propagated throughout the culture, plasmid partitioning modules, such as the *par* locus previously described or plasmid stabilizing genetic modules can be implemented in expression plasmids to stabilize the plasmid. The *cer* locus propagates inheritance of ColE1 and related plasmids by suppressing occurrence of plasmids in multimeric states. This is achieved by the actions of a XerCD site specific recombinase action at the *cer* site (186,187). Plasmid propagation can also be improved by optimization of the selection fragments by engineering a minimally expressed resistance enzyme, thus improving the plasmid maintenance in the culture and increase yields of recombinant protein (188,189). In *Papers II* and *III*, the Tn3.1 and Tn903.1 resistance fragments were TIR optimized to obtain minimally expressed variants which supported better plasmid retention compared to the original selection fragments.

One final alternative to obtain faithful plasmid retention in cell cultures can be to switch from antibiotics-based selection to antibiotics-free selection. Peubez I. et al 2010. (108) engineered the PM1800 plasmid, a derivative of pET28c, in which they placed the *cer* locus. Comparison of the plasmids maintenance of BL21(*DE3*) harbouring pET28c-AlpA and BL21(*DE3*) PM1800-AlpA. they saw that the *cer* locus allowed for superior plasmid maintenance in cells

cultures cultivated in absence of kanamycin but producing AlpA protein. The plasmid retention of pET28c-AlpA plasmid harbouring cell cultures 23h post induction was 3%, and PM1800-AlpA plasmid harbouring cell cultures 79-98%. They further compared the protein yields between cells cultures grown in kanamycin and cells selected with the poison-antidote (both induced with 1mM IPTG) and observed that the poison/antidote cell system outperformed the antibiotic selection cell system in both plasmid maintenance and protein production (100% and 350 mg/mL, respectively) vs a cell system relying on selection with kanamycin (5% and 36 mg/mL, respectively). The findings presented in literature and in *Papers I* to *IV* highlight possible flaws and solutions to maintain reliable plasmid propagation in recombinant bacteria.

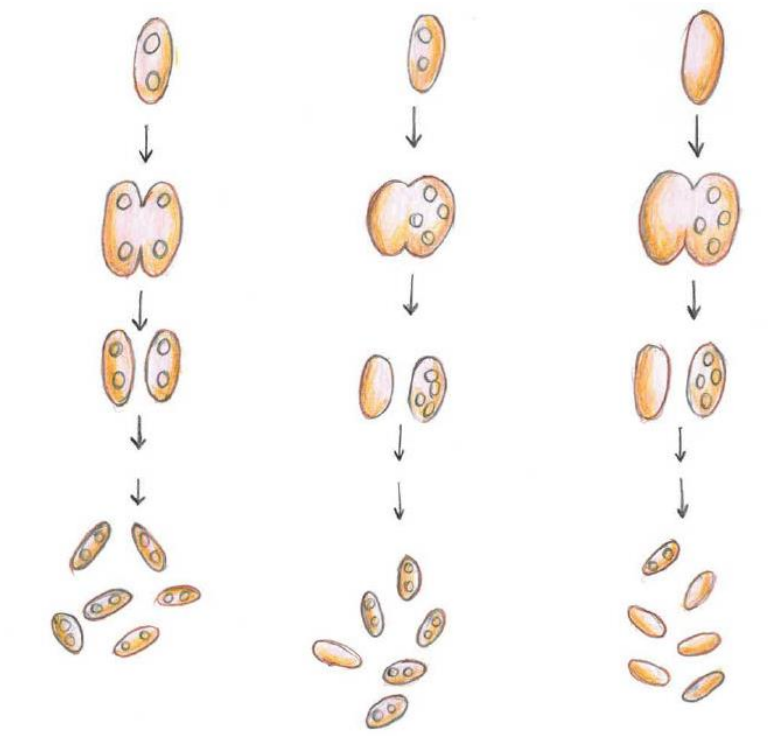


Figure 9. Illustration of plasmid maintenance efficiency in different conditions. If selection pressure is present only cells which have plasmid will be able to further propagate and the culture will contain only plasmid-carrying cells (left). If selection is not present and no stress is induced, which provides plasmid-free cells with a growth advantage, cell cultures will contain majority plasmid-bearing cells (middle). If selection pressure is absent and stress is induced, plasmid-free cells will outgrow their plasmid-carrying sisters and take over the culture (right).

Summary of publications

Paper I: Tips for efficiently maintaining pET expression plasmids

As noted previously in this thesis, the stability of plasmids is important for obtaining recombinant proteins. Therefore, plasmid maintenance efficiency in the two most commonly used pET expression plasmids, pET28 and pET15 were investigated.

In *Paper I*, plasmid retention of BL21(*DE3*) harbouring either pET15b Tn3.1 or pET28a Tn903.1 was compared. The pET15 plasmid harbours the Tn3.1 fragment (containing the *bla* gene) which confers resistance towards β-lactam antibiotics (amp^R). This resistance fragment is known for causing unreliable plasmid maintenance in cell cultures (99–101). The pET28 plasmid harbours the Tn903 antibiotic resistance fragment (containing the *aph* gene) from *Streptomyces kanamyceticus* which provides cells with resistance against aminoglycosides, such as kanamycin (kan^R). The pET15b and pET28a plasmids both harboured the sfGFP gene for protein production. In the absence of recombinant protein production, both resistance fragments allowed for propagation of cells cultures harbouring plasmids. Upon induction of recombinant protein production with 0.5 mM IPTG for 20h, BL21(*DE3*) cells harbouring the Tn903.1 fragment maintained plasmid to a greater extent than the BL21(*DE3*) cells harbouring the Tn3.1 fragment. Plasmid retention efficiency of the cultures also influenced sfGFP yields as BL21(*DE3*) harbouring pET28a Tn903.1 produced more sfGFP compared to pET15-Tn3.1.

Plasmid maintenance was assessed in the C41(*DE3*) strain (with lower levels of the T7 RNA polymerase) to elucidate if metabolic burden from protein production was part of the plasmid maintenance issues. Plasmid maintenance efficiency of C41(*DE3*) cells harbouring either pET15b Tn3.1-sfGFP or pET28a Tn903.1-sfGFP were compared in the presence of inducer (0.5 mM IPTG) to the plasmis maintenance in BL21(*DE3*) with the same plasmids and induction conditions. It was shown that C41(*DE3*) cells cultures

retained >90 % plasmid 20h post induction with 0.5 mM IPTG irrespective of presence or absence of antibiotics selection or plasmid used. BL21(*DE3*) cell cultures harbouring the pET15b Tn3.1-sfGFP plasmid retained plasmid <10 % upon induction with 0.5 mM IPTG. BL21(*DE3*) cells cultures harbouring pET28a Tn903.1-sfGFP plasmid retained plasmid <50%. In the absence of selection, both pET15b-sfGFP and pET28a-sfGFP cell cultures harboured <30% plasmid 20h post induction with 0.5 mM IPTG. The results showed that the strain culture less burdened by protein production was more efficient in retaining the plasmid independent of the presence or absence of antibiotics or plasmid used.

The resistance fragments Tn903.1 and Tn3.1 were also compared in *Paper IV*. It was observed that upon expression of human growth hormone (hGH) and single-chain antibody against Her2 ($scFv^{Her2}$) from pBAD vectors harbouring either Tn903.1 or Tn3.1 resistance fragments, the Tn903.1 resistance fragment facilitated plasmid maintenance to a greater extent than Tn3.1. The combined findings in *Paper I* and *Paper IV* suggests that there is a need to choose the antibiotic resistance fragment with care, as it can influence the outcome of the recombinant protein production experiment.

Paper II: Antibiotic-Efficient Genetic Cassette for the TEM-1 β-Lactamase That Improves Plasmid Performance

Faithful plasmid retention in recombinant cell cultures is one of the pre-requisites for successful recombinant protein production. It is well known in literature that the use of antibiotics resistance cassettes encoding for resistance against β-lactams (such as $Tn3/amp^{R}$) are problematic. This is due to the high amount of resistance enzyme released into the culturing media that inactivates the antibiotic. In *Paper II* the amount of TEM-1 β- Lactamase enzyme, from the Tn3 transposon of *Salmonella paratyphi B*, was decreased by altering the TIR of the *bla* mRNA. The new TIR variant (called $Tn3.1^{MIN}$), was obtained using PCR with synonymous codon primers. Cells harbouring plasmid with Tn3.1

^{MIN} were more sensitive to ampicillin and carbenicillin, compared to cells harbouring plasmid with the original Tn3.1 fragment.

As β-lactamases produced by the Tn3 resistance fragments are known to leak out to the media and break down β-lactam antibiotics, determination of the half time ($t_{1/2}$) of ampicillin and carbenicillin in media from cells harbouring plasmid with either Tn3.1 or Tn3.1^{MIN} was performed using Mass-spectrometry. It was observed that the Tn3.1^{MIN} resistance fragment increased the $t_{1/2}$ of antibiotics compared to Tn3.1. Breakdown of ampicillin in culture media was further verified by testing if spent over-night media from cells harbouring plasmid with either Tn3.1 or Tn3.1^{MIN} contained enough β-lactamase to inactivate ampicillin in fresh media. Growth of contaminant cultures (not resistant to ampicillin) was followed. The cells were cultivated in media supplemented with ampicillin and spend media from BL21(DE3) harbouring either Tn3.1 or Tn3.1^{MIN}. It was again observed that Tn3.1^{MIN} outperformed Tn3.1, as the spent over-night media from Tn3.1^{MIN} harbouring cells did not allow for contaminant culture to grown in the presence of ampicillin. Plasmid retention fidelity in BL21(*DE3*) cell cultures harbouring pET15-sfGFP with either Tn3.1 or Tn3.1^{MIN} resistance were assessed. BL21(*DE3*) cells harbouring pET15-sfGFP with Tn3.1^{MIN} resistance was observed to allow for a higher percentage of plasmid maintenance after 20h post induction with 0.5 mM IPTG both in the presence or absence of ampicillin selection.

Lastly, the Tn3.1^{MIN} resistance fragment was assessed for applicability into plasmid backbones with different origin of replication which dictates plasmid copy number). The fragment was able to be used in high-to medium copy number plasmids, but not low copy number plasmids. The results in *Paper II* further emphasizes the notion that caution needs to be taken when choosing resistance cassettes, which was also presented in *Papers 1* and *IV*. *Paper II* also provides an improved resistance fragment for those who wish to use resistance cassettes encoding for β- lactam antibiotics or who are limited to them, due to use of strains which are resistant to other antibiotics.

Paper III: The pET^{OPT} collection of expression plasmids: Better performing versions of the old favourites

Paper III focused on improving the pET expression plasmids. The original T7plac genetic module, which is used in most pET plasmids, is known to allow for pre-induction expression of recombinant protein (leakiness). This can in some instances cause a plethora for problems, such as insufficient yields of soluble recombinant protein and poor cellular health. These problems become more pronounced upon induction of recombinant protein. To explore and understand which combinations of genetic modules cause different levels of leakiness, pET28 plasmids harbouring sfGFP were adapted with different combinations of T7plac genetic modules which: I) repressed transcriptional efficiency (combination of lacO3 and lacO1 around the promoter (7,8) and placI^{q} (81), II) increased transcriptional efficiency (T7p^{CONS}) (80) III) increased translational efficiency (TIR$^{EVOL/TIR2}$) (80) or IV) a combination of the genetic modules previously mentioned. Leakiness of different combinations of genetic modules were tested by measuring *in cell* sfGFP fluorescence.

In an attempt to understand which genetic modules were compatible with with which type of strain, pET28 and derivatives were transformed into six different strains. These strains either contained high or low levels of T7 RNA polymerase. It was observed that the amount of recombinant sfGFP produced pre-induction could either be dampened or increased depending on the genetic module combinations. In some instances, the leakiness impaired cellular growth to a point of growth arrest. Furthermore, it was observed that the implementation of repressive genetic modules resulted in a decreased amount of sfGFP fluorescence in some instances, while in other instances it increased the sfGFP fluorescence.

With the fluorescence data attained from the strain matching, three outcomes were presented: 1) combinations of T7plac genetic modules and strains which had no leakiness of sfGFP, 2) combinations of T7plac genetic modules and strains which increased the yields of sfGFP upon induction, 3) a compromise of

different combination of T7plac genetic modules in both high and low T7 RNA polymerase strains which decreased leakiness, but also increased yields of sfGFP upon induction. The compromise of T7plac genetic modules, one for each strain groups was chosen for implementation of improved terminator (Tz^{OPT}) and antibiotic resistance cassettes: $Tn901^{MIN}$ for kanamycin resistance and $Tn3.1^{MIN}$ for ampicillin resistance (developed in *Paper II*). These plasmids were renamed pET^{OPT} plasmids and benchmarked against pET28. The pET^{OPT} yielded higher amounts of recombinant protein compared to pET28. Implementation of 5 other multiple cloning sites of the 5 other most used pET plasmids into pET^{OPT} plasmids was done to obtain the pET^{OPT} collection. Lesson learnt in *Papers I*, *II* and *IV* about plasmid maintenance were taken into consideration when creating the pET^{OPT} plasmids. The aim of *Paper III* was to provide useful tools for protein scientists to use.

Paper IV: Biosensor that detects stress caused by periplasmic proteins

Recombinant protein production is an endeavour which needs to be embarked upon with great care and understanding of the recombinant bacteria. If culturing conditions are optimized, recombinant bacteria can be utilized to produce functional protein of interest. In *Paper IV* a biosensor was constructed which can detect recombinant protein production stress *in vivo*. This sensor was named plasmid for Quality Control (pQC), as it allows for *in-vivo* detection of recombinant protein misfolding both in the cytoplasm and periplasm, without the need for fractionation and immunoblotting. This was achieved by fusing the IbpA promoter to a de-stabilized GFP mutant ($GFPmut3_{ASV}$) to create P_{ibpa}-$GFPmut3_{ASV}$ and the CpxP promoter fused to a de-stabilized mCherry mutant ($mCherry_{ASV}$). If insufficient secretion reached above a certain threshold in the cytoplasm due to sub-optimal culturing conditions, P_{IbpA} will be activated, and green fluorescence can be quantified. The same principle holds for periplasmic stress; activation of P_{CpxP} will occur, production of $mCherry_{ASV}$ will increase and red fluorescence and be quantified.

Evaluation of pQC was performed by expression of MalE and derivatives which mis-fold in the cytoplasm and periplasm. Insufficient secretion for MalE and derivatives could be detected by quantification of GFP fluorescence and misfolding in the periplasm by quantification of mCherry fluorescence. pQC was further tested by producing two industrially relevant periplasmic proteins: single-chain antibody against Her2 ($scFv^{Her2}$) and human Growth Hormone (hGH). The pQC plasmid was used to estimate the state of the stress cells were under by the fluorescence profile. This was further validated by combining the fluorescence profile with other metrics of cellular health such as biomass accumulation. The higher the concentration of inducer L-arabinose, the higher the fluorescence output was in the culture, combined with lower biomass and plasmid maintenance and soluble protein production, which induced stress in both cellular compartments. It was thus possible to optimize the functional protein yields of $scFv^{Her2}$ by using the fluorescence profile as a sole indicator.

Paper IV allows for *in-vivo* detection of stress in two cellular compartments. This simplifies the culturing optimization process and allows for elimination of culturing conditions which are unproductive and allows for screening of culturing conditions where minor changes in experimental set-up could yield major results in terms of recombinant protein yields.

Conclusions and prospects

The success of a recombinant protein production experiment is dependent on many factors. Some of these factors have been discussed in the thesis and explored in *Paper I-IV*.

In *Paper I* and *Paper II* the influence of antibiotic resistance fragments on plasmid maintenance efficiency and recombinant protein production yields was investigated. The findings in *Paper I* highlight that not all resistance cassettes are equally efficient in maintaining plasmids. The work in *Paper II* shows that certain antibiotics fragments lead to rapid inactivation of antibiotics in culture media. Optimization of those fragments can mitigate the problem and aid in dampening the plasmid maintenance issues. The findings in *Paper II* expands upon the findings in *Paper I* and highlight the problem of using improper antibiotic resistance cassettes. In both *Paper I* and *Paper IV* it was demonstrated that inefficient plasmid maintenance can have a direct consequence on protein production yields.

Other aspects of optimizing plasmid retention and general selection of strains could be considered. Although plasmid maintenance efficiency can be improved by exchanging the antibiotic resistance fragments or by optimizing the expression of the resistance enzyme, the pBR322 based plasmids (such as the pET plasmids) do not contain an active partition (*par*) locus (174). A *par* locus that can be found in pSC101 plasmids allow for active partitioning of the plasmids within the cells and increased plasmid retention within the culture, even in absence of antibiotic selection (190). To improve maintenance of pET plasmids, integration of a *par* locus into the backbone module could be done. This would require additional testing to verify plasmid maintenance fidelity under multiple induction regimes, temperatures, growth vessels, host strains and using CDS of proteins from different origins and cellular compartment localizations.

If antibiotics-based selection is not an option and reliance on the *par* locus in the absence of antibiotics is not sufficient, antibiotics free selection fragments can be implemented. Studies performed by others on antibiotics-free selection presented in the thesis indicate that plasmid retention can be optimized to be close to 100%. This is due to the nature of the selection, as plasmid-free cells become unable to propagate when the plasmid is absent, because the plasmid often contains a gene which is essential to cell survival. This is the idea with antibiotic selection fragments, but the risk for bacterial subcultures which have obtained mutations in efflux proteins, (which would allow for increased resistance of antibiotics) is high (191). Development of the pET system with antibiotics-free fragments and suitable host strains could provide users with increased yields of recombinant protein, especially in industrial setting, where the use of antibiotics is sometimes prohibited. Downsides could be the dependence on specific strains and need to cure out the plasmid which allows for survival prior to transformation with the recombinant plasmid. This process could contain antibiotic selection in the beginning, but not during the production of recombinant protein.

The development of the stress sensor plasmid pQC presented in *Paper IV* will allow for screening of multiple culturing conditions at once, as experiments can be set up in culturing plates and whole-cell fluorescence can be measured. Unfavourable conditions can be eliminated early, which allows optimization of culturing conditions and increases the chances of obtaining functional protein. This work has followed the rational of others, such as Zuts A. et al. 2021 (168) where a folding sensor was made using the IbpA promoter linked to a destabilized GFPmut3. The work performed in *Paper IV* has expanded upon the idea of biological sensor by implementing two distinct sensors, allowing for detection of stresses in two separate cellular compartments within in the same plasmid backbone.

Further testing to see if pQC can be utilized to detect misfolded proteins in the cytosol and if it can be used to detect membrane

protein insertion problems is to be done. The Cpx stress response has been shown to activate upon membrane protein perturbations (192,193). The pQC could be further adapted with promoters governed by σS for the detection of alkaline, acidic and starvation stress during culturing, and P_{OxyR} to monitor oxygen depletion stress (155). These promoters could be fused to other fluorescent proteins which are not currently employed in the plasmid. Additionally, a collection of quality control plasmids with specific utility could be constructed, such as promoters involved in growth condition monitoring or recombinant proteins production stress.

Beyond reporting the utility of pQC, findings in *Paper IV* additionally to aspects of protein production discussed further emphasises that poor cellular health has direct impact on plasmid maintenance as presented in *Paper I* and alluded to in *Paper II*. Work done by (124–126) to name a few, strengthen the hypothesis that improper culturing conditions result in unhealthy recombinant cells and impact production yields. Results in *Paper IV* also verify that culturing temperature play a role in recombinant protein production and cellular processes.

In *Paper III*, different improved genetic modules scattered in literature were collected and placed into one plasmid backbone. The new pET^{OPT} plasmids were compared to the original plasmids and increased protein yields were observed. A limitation of the pET^{OPT} plasmids is that they have been developed for globular proteins. Further testing and implementation of genetic modules developed for membrane proteins, such as in (95,194) could to be done to expand the pET^{OPT} plasmids. The pET^{OPT} plasmids have so far been tested in two growth temperatures and one IPTG induction regime. Different growth media, temperatures and inducer concentrations for selected proteins from different origins and cellular compartments could be explored. To eliminate leakiness of pET system without impacting the total yield of proteins proved to be challenging.

Use of the strains developed in (85) together with the pET^{OPT} plasmids could provide the scientific community with an even better expression system as the induction of T7 RNA polymerase can be titrated and thus regulated. The aim of *Paper* III was to gather design improvements to the pET system plasmids from literature and my work and implementing them in a collection of plasmids which could improve yields of recombinant proteins. The pET^{OPT} plasmids could facilitates use of plasmids with updated modules. The pET^{OPT} collection also brings multiple peoples work together into one place, instead of having it dispersed.

An analysis of the literature in preparation for the writing of this thesis revealed that knowledge of how to set up a recombinant protein production experiment is sometimes difficult to find. In the spirit of the work performed by (195) and (196) and expanding upon it, one could create a digital database of growth conditions, expression plasmids, expression strains, recombinant proteins and everything involved in the recombinant protein production experiment. This would be a valuable tool for all interested in expressing a recombinant protein. This could be done by creating a website which is open to the public to access and upload information to. This would provide the opportunity to not only insert protocols, information about proteins and results, but also to connect external resources based on certain criteria. An example could be a protein from mammalian origin and located in the membrane. Through navigation of relevant links, protocols could be accessed from people having expressed similar proteins and uploaded successful and non-successful conditions, which could be used as a guide.

Work performed and presented in this thesis has shown that the protein expression tools currently most widely used are by no means ideal. Work done in the thesis addressed some of the issues and provided options to mitigate these issues, but there are other aspects which are still not addressed. This could partially be due to the information being scattered and not common knowledge.

Acknowledgements

First and foremost, I would like to give thanks to my Lord and Saviour Jesus Christ for everything. You are my everything, my love and my life. Thank you for getting me through this PhD, it has been tough and filled with many obstacles, but You made a way. It has certainly been eventful, educational, and character developing. Through the pain I'm grateful. I give all the glory to The Father, The Son and The Holy Spirit. Amen.

I would like to thank my parents, Almas and Sergon, for their never-failing support and love through this time and through-out my life. I have been truly blessed by God to have you as my parents. Thank you for all the prayers and your relentless encouragement. I would like to thank my brother, Daniel for all the support and the laughs. My lovely family, I hold you dear to my heart, cherish and love you. God bless you, keep you and increase you.

Thank you to my fiancée and future husband, Joakim Lindström for your support and the fact that you had the energy to deal with me during this time. It has not been easy, I'm sure of it. But if we can make it through this, the rest of our lives together will be a blast. You are a blessing from God given to me, one which I treasure more than words can express. After this it is my turn to support you, as your future wife, which I look forward to, with God by our side. I love you until death do us apart. May God bless you always Thank you to my extended family for your prayers and support. My Atoo Atoor and Amo Dadisho, for your prayers and always opening your home to me. I love you and may God bless you. Thank you to Joakim's extended family for your love and support. God bless you.

Thank you to my supervisor Daniel O Daley for taking me on as a PhD student. For allowing me to finish my PhD, although I started somewhere else. And thank you for these eventful, 5-odd years. You have taught me many lessons and opened my eyes to how we do science. This is insight which I would not have had, had I not finished this journey. Thank you, all you have taught me will not go to waste. I promise. God bless you.

I would also like to thank my first supervisor, Martin Ott for allowing me to start my PhD with you and for teaching me many useful lessons.

I would like to thank my current group members, Mati and Maurice and previous group member, James. Maurice and Mati, I wish you the best, that you acquire wisdom and further knowledge in all your endeavours. God bless you.

James, it was you and me in the lab for approximately 2-3 years and man, did we accomplish much during those years. It was such a pleasure working with you. We had so many laughs and insightful discussions. You have helped me grow intellectually and as a person. God bless you and may you be successful in your endeavours.

Thank you, my Bea for your encouragement and all the lovely words and hugs. You are like a big sister to me. I will miss you. God bless you.

Thank you Juliane John for taking the time and proof-reading part of my thesis and giving me helpful comments and constructive criticism. God bless you.

Thank you to all my students, shared students and lab students throughout the years: Natalia, Nicole, Daria, Caroline, Tiago, Hanna-Louise, Nicklas and Christina. You have become a part of me and help me to communicate, kept me humble and to not forget that I was once a student. Thank you to all my students I had while I was a teaching assistant, you have taught me how to communicate better and kept me humble. I'm honoured to have been given the opportunity to (I hope) help you understand biochemistry.

Thank you to all the DBB people on the 4:th and 5:th floor. Thank you for all the times we were teaching assistants together, we laughed, discussed topics between heaven and earth. Thank you to the administration and technical team at DBB. I wish you all the best, I will miss you all and may God bless you all.

References

1. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. Cell. 1999 Sep;98(6):811–24.

2. Gruber TM, Gross CA. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annu Rev Microbiol. 2003 Oct;57(1):441–66.

3. Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. Annu Rev Microbiol. 2014 Sep 8;68(1):357–76.

4. Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. eLife. 2015 Sep 8;4:e08504.

5. Nedialkov YA, Opron K, Assaf F, Artsimovitch I, Kireeva ML, Kashlev M, et al. The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. Biochim Biophys Acta BBA - Gene Regul Mech. 2013 Feb;1829(2):187–98.

6. Cooper GM. Transcription in Prokaryotes. In: The Cell: A Molecular Approach 2nd edition. Sunderland (MA): Sinauer Associates; 2000.

7. Oehler S, Amouyal M, Kolkhof P, Von Wilcken-Bergmann B, Müller-Hill B. Quality and position of the three lac operators of *E. coli* define efficiency of repression. EMBO J. 1994 Jul;13(14):3348–55.

8. Becker NA, Peters JP, Lionberger TA, Maher LJ. Mechanism of promoter repression by Lac repressor–DNA loops. Nucleic Acids Res. 2013 Jan 1;41(1):156–66.

9. Gilbert W, Müller-Hill B. ISOLATION OF THE LAC REPRESSOR. Proc Natl Acad Sci. 1966 Dec;56(6):1891–8.

10. Gualerzi CO, Brandi L, Caserta E, La Teana A, Spurio R, Tomšic J, et al. Translation Initiation in Bacteria. In: Garrett RA, editor. The Ribosome. Washington, DC, USA: ASM Press; 2014. p. 475–94.

11. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of Protein Synthesis in Bacteria. Microbiol Mol Biol Rev. 2005 Mar;69(1):101–23.

12. Milón P, Rodnina MV. Kinetic control of translation initiation in bacteria. Crit Rev Biochem Mol Biol. 2012 Aug;47(4):334–48.

13. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999 Jul;234(2):187–208.

14. Rodnina MV, Gromadski KB, Kothe U, Wieden HJ. Recognition and selection of tRNA in translation. FEBS Lett. 2005 Feb 7;579(4):938–42.

15. Rodnina MV. Translation in Prokaryotes. Cold Spring Harb Perspect Biol. 2018 Sep;10(9):a032664.

16. Schmeing TM, Ramakrishnan V. What recent ribosome structures have revealed about the mechanism of translation. Nature. 2009 Oct 29;461(7268):1234–42.

17. Pape T. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E.coli ribosome. EMBO J. 1998 Dec 15;17(24):7490–7.

18. Carbone CE, Loveland AB, Gamper HB, Hou YM, Demo G, Korostelev AA. Time-resolved cryo-EM visualizes ribosomal translocation with EF-G and GTP. Nat Commun. 2021 Dec 13;12(1):7236.

19. Scolnick E, Tompkins R, Caskey T, Nirenberg M. Release factors differing in specificity for terminator codons. Proc Natl Acad Sci. 1968 Oct;61(2):768–74.

20. Freistroffer DV, Pavlov MYu, MacDougall J, Buckingham RH, Ehrenberg M. Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. EMBO J. 1997 Jul 1;16(13):4126–33.

21. Klaholz BP, Myasnikov AG, Van Heel M. Visualization of release factor 3 on the ribosome during termination of protein synthesis. Nature. 2004 Feb;427(6977):862–5.

22. Milligan RA, Unwin PNT. Location of exit channel for nascent protein in 80S ribosome. Nature. 1986 Feb;319(6055):693–5.

23. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. Science. 2000 Aug 11;289(5481):905–20.

24. Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, Johansson M, et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. Cell Rep. 2015 Sep;12(10):1533–40.

25. Liutkute M, Samatova E, Rodnina MV. Cotranslational Folding of Proteins on the Ribosome. Biomolecules. 2020 Jan 7;10(1):97.

26. Kudva R, Tian P, Pardo-Avila F, Carroni M, Best RB, Bernstein HD, et al. The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. eLife. 2018 Nov 26;7:e36326.

27. Kudlicki W, Chirgwin J, Kramer G, Hardesty B. Folding of an Enzyme into an Active Conformation While Bound as Peptidyl-tRNA to the Ribosome. Biochemistry. 1995 Nov 1;34(44):14284–7.

28. Cassaignau AME, Cabrita LD, Christodoulou J. How Does the Ribosome Fold the Proteome? Annu Rev Biochem. 2020 Jun 20;89(1):389–415.

29. Hayer-Hartl M, Bracher A, Hartl FU. The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. Trends Biochem Sci. 2016 Jan;41(1):62–76.

30. Santra M, Farrell DW, Dill KA. Bacterial proteostasis balances energy and chaperone utilization efficiently. Proc Natl Acad Sci. 2017 Mar 28;114(13).

31. Hoffmann A, Bukau B, Kramer G. Structure and function of the molecular chaperone Trigger Factor. Biochim Biophys Acta BBA - Mol Cell Res. 2010 Jun;1803(6):650–61.

32. Katikaridis P, Bohl V, Mogk A. Resisting the Heat: Bacterial Disaggregases Rescue Cells From Devastating Protein Aggregation. Front Mol Biosci. 2021 May 4;8:681439.

33. Uchihashi T, Watanabe Y hei, Nakazaki Y, Yamasaki T, Watanabe H, Maruno T, et al. Dynamic structural states of ClpB involved in its disaggregation function. Nat Commun. 2018 Jun 1;9(1):2147.

34. Sauer RT, Baker TA. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. Annu Rev Biochem. 2011 Jul 7;80(1):587–612.

35. Gottesman S, Wickner S, Maurizi MR. Protein quality control: triage by chaperones and proteases. Genes Dev. 1997 Apr 1;11(7):815–23.

36. Micevski D, Dougan DA. Proteolytic Regulation of Stress Response Pathways in Escherichia coli. In: Dougan DA, editor. Regulated Proteolysis in Microorganisms. Dordrecht: Springer Netherlands; 2013. p. 105–28. (Subcellular Biochemistry; vol. 66).

37. Laskowska E, Kuczyńska-Wiśnik D, Skórko-Glonek J, Taylor A. Degradation by proteases Lon, Clp and HtrA, of *Escherichia coli* proteins aggregated *in vivo* by heat shock; HtrA protease action *in vivo* and *in vitro* . Mol Microbiol. 1996 Nov;22(3):555–71.

38. Kihara A, Akiyama Y, Ito K. FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. Proc Natl Acad Sci. 1995 May 9;92(10):4532–6.

39. Langklotz S, Baumann U, Narberhaus F. Structure and function of the bacterial AAA protease FtsH. Biochim Biophys Acta BBA - Mol Cell Res. 2012 Jan;1823(1):40–8.

40. Miwa T, Taguchi H. *Escherichia coli* small heat shock protein IbpA plays a role in regulating the heat shock response by controlling the translation of σ^{32}. Proc Natl Acad Sci. 2023 Aug 8;120(32):e2304841120.

41. Cowing DW, Bardwell JC, Craig EA, Woolford C, Hendrix RW, Gross CA. Consensus sequence for *Escherichia coli* heat shock gene promoters. Proc Natl Acad Sci. 1985 May;82(9):2679–83.

42. Karp PD, Paley S, Caspi R, Kothari A, Krummenacker M, Midford PE, et al. The EcoCyc Database (2023). Lovett ST, editor. EcoSal Plus. 2023 Dec 12;11(1):eesp-0002-2023.

43. Zhou YN, Kusukawa N, Erickson JW, Gross CA, Yura T. Isolation and characterization of *Escherichia coli* mutants that lack the heat shock sigma factor sigma 32. J Bacteriol. 1988 Aug;170(8):3640–9.

44. Hews CL, Cho T, Rowley G, Raivio TL. Maintaining Integrity Under Stress: Envelope Stress Response Regulation of Pathogenesis in Gram-Negative Bacteria. Front Cell Infect Microbiol. 2019 Sep 4;9:313.

45. Zhao K, Liu M, Burgess RR. The Global Transcriptional Response of *Escherichia coli* to Induced σ32 Protein Involves σ32 Regulon Activation Followed by Inactivation and Degradation of σ32 *in vivo*. J Biol Chem. 2005 May;280(18):17758–68.

46. Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a multifaceted cellular response to heat stress. Genes Dev. 2006 Jul 1;20(13):1776–89.

47. Lim B, Miyazaki R, Neher S, Siegele DA, Ito K, Walter P, et al. Heat Shock Transcription Factor σ32 Co-opts the Signal Recognition Particle to Regulate Protein Homeostasis in *E. coli*. Horwich AL, editor. PLoS Biol. 2013 Dec 17;11(12):e1001735.

48. Oswald J, Njenga R, Natriashvili A, Sarmah P, Koch HG. The Dynamic SecYEG Translocon. Front Mol Biosci. 2021 Apr 15;8:664241.

49. Darby NJ, Creighton TE. Catalytic Mechanism of DsbA and Its Comparison with That of Protein Disulfide Isomerase. Biochemistry. 1995 Mar 21;34(11):3576–87.

50. Dailey FE, Berg HC. Mutants in disulfide bond formation that disrupt flagellar assembly in *Escherichia coli*. Proc Natl Acad Sci. 1993 Feb;90(3):1043–7.

51. Bardwell JC, Lee JO, Jander G, Martin N, Belin D, Beckwith J. A pathway for disulfide bond formation *in vivo*. Proc Natl Acad Sci. 1993 Feb;90(3):1038–42.

52. Kobayashi T, Kishigami S, Sone M, Inokuchi H, Mogi T, Ito K. Respiratory chain is required to maintain oxidized states of the DsbA-DsbB disulfide bond formation system in aerobically growing *Escherichia coli* cells. Proc Natl Acad Sci. 1997 Oct 28;94(22):11857–62.

53. Rietsch A, Belin D, Martin N, Beckwith J. An *in vivo* pathway for disulfide bond isomerization in *Escherichia coli*. Proc Natl Acad Sci. 1996 Nov 12;93(23):13048–53.

54. Kadokura H, Katzen F, Beckwith J. Protein Disulfide Bond Formation in Prokaryotes. Annu Rev Biochem. 2003 Jun;72(1):111–35.

55. McCarthy AA, Haebel PW, Törrönen A, Rybin V, Baker EN, Metcalf P. Crystal structure of the protein disulfide bond isomerase, DsbC, from *Escherichia coli*. Nat Struct Biol. 2000 Mar 1;7(3):196–9.

56. Chen J, Song J li, Zhang S, Wang Y, Cui D fu, Wang C chen. Chaperone Activity of DsbC. J Biol Chem. 1999 Jul;274(28):19601–5.

57. Liu J, Walsh CT. Peptidyl-prolyl cis-trans-isomerase from Escherichia coli: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. Proc Natl Acad Sci. 1990 Jun;87(11):4028–32.

58. Horne SM, Young KD. *Escherichia coli* and other species of the enterobacteriaceae encode a protein similar to the family of Mip-like FK506-binding proteins. Arch Microbiol. 1995 May;163(5):357–65.

59. Dartigalongue C, Raina S. A new heat-shock gene, ppiD, encodes a peptidyl–prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. EMBO J. 1998 Jul 15;17(14):3968–80.

60. Rouvière PE, Gross CA. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. Genes Dev. 1996 Dec 15;10(24):3170–82.

61. Snyder WB, Davis LJ, Danese PN, Cosma CL, Silhavy TJ. Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. J Bacteriol. 1995 Aug;177(15):4216–23.

62. Danese PN, Silhavy TJ. The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. Genes Dev. 1997 May 1;11(9):1183–93.

63. Danese PN, Silhavy TJ. CpxP, a Stress-Combative Member of the Cpx Regulon. J Bacteriol. 1998 Feb 15;180(4):831–9.

64. Danese PN, Snyder WB, Cosma CL, Davis LJ, Silhavy TJ. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. Genes Dev. 1995 Feb 15;9(4):387–98.

65. Ruiz N, Silhavy TJ. Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. Curr Opin Microbiol. 2005 Apr;8(2):122–6.

66. Raivio TL, Silhavy TJ. The σE and Cpx regulatory pathways: Overlapping but distinct envelope stress responses. Curr Opin Microbiol. 1999 Apr;2(2):159–65.

67. Raivio TL, Laird MW, Joly JC, Silhavy TJ. Tethering of CpxP to the inner membrane prevents spheroplast induction of the Cpx envelope stress response. Mol Microbiol. 2000 Sep;37(5):1186–97.

68. DiGiuseppe PA, Silhavy TJ. Signal Detection and Target Gene Induction by the CpxRA Two-Component System. J Bacteriol. 2003 Apr 15;185(8):2432–40.

69. Raivio TL, Silhavy TJ. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. J Bacteriol. 1997 Dec;179(24):7724–33.

70. Pogliano J, Lynch AS, Belin D, Lin EC, Beckwith J. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. Genes Dev. 1997 May 1;11(9):1169–82.

71. Tsviklist V, Guest RL, Raivio TL. The Cpx Stress Response Regulates Turnover of Respiratory Chain Proteins at the Inner Membrane of Escherichia coli. Front Microbiol. 2022 Jan 28;12:732288.

72. Del Solar G, Espinosa M. Plasmid copy number control: an ever-growing story. Mol Microbiol. 2000 Aug;37(3):492–500.

73. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, et al. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene. 1977;2(2):95–113.

74. Lee C, Kim J, Shin SG, Hwang S. Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. J Biotechnol. 2006 May;123(3):273–80.

75. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995 Jul;177(14):4121–30.

76. Nordström K. Plasmid R1—Replication and its control. Plasmid. 2006 Jan;55(1):1–26.

77. Feklistov A, Darst SA. Structural Basis for Promoter −10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. Cell. 2011 Dec;147(6):1257–69.

78. Aggarwal RK, Narang A. Inducer exclusion, by itself, cannot account for the glucose-mediated lac repression of *Escherichia coli*. Biophys J. 2022 Mar;121(5):820–9.

79. Dubendorf JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol. 1991 May;219(1):45–59.

80. Shilling PJ, Mirzadeh K, Cumming AJ, Widesheim M, Köck Z, Daley DO. Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*. Commun Biol. 2020 May 7;3(1):214.

81. Calos MP. DNA sequence for a low-level promoter of the lac repressor gene and an ‘up’ promoter mutation. Nature. 1978 Aug;274(5673):762–5.

82. Studier FW. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol. 1991 May;219(1):37–44.

83. Wycuff DR, Matthews KS. Generation of an AraC-araBAD Promoter-Regulated T7 Expression System. Anal Biochem. 2000 Jan;277(1):67–73.

84. Novagen. pET System Manual 11th Edition. EMD Biosciences Inc; 2005.

85. Du F, Liu YQ, Xu YS, Li ZJ, Wang YZ, Zhang ZX, et al. Regulating the T7 RNA polymerase expression in *E. coli* BL21 (*DE3*) to provide more host options for recombinant protein production. Microb Cell Factories. 2021 Dec;20(1):189.

86. Schleif R. Regulation of the l-arabinose operon of *Escherichia coli*. Trends Genet. 2000 Dec;16(12):559–65.

87. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009 Oct;27(10):946–50.

88. Steitz JA, Jakes K. How ribosomes select initiator regions in mRNA: base pair formation between the 3’ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proc Natl Acad Sci. 1975 Dec;72(12):4734–8.

89. Munson LM, Stormo GD, Niece RL, Reznikoff WS. lacZ translation initiation mutations. J Mol Biol. 1984 Aug;177(4):663–83.

90. Vellanoweth RL, Rabinowitz JC. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus*

subtilis and *Escherichia coli in vivo*. Mol Microbiol. 1992 May;6(9):1105–14.

91. McCarthy JEG, Gualerzi C. Translational control of prokaryotic gene expression. Trends Genet. 1990;6:78–85.

92. Shine J, Dalgarno L. The 3′-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. Proc Natl Acad Sci. 1974 Apr;71(4):1342–6.

93. Bhattacharyya S, Jacobs WM, Adkar BV, Yan J, Zhang W, Shakhnovich EI. Accessibility of the Shine-Dalgarno Sequence Dictates N-Terminal Codon Bias in *E. coli*. Mol Cell. 2018 Jun;70(5):894-905.e5.

94. Mirzadeh K, Shilling PJ, Elfageih R, Cumming AJ, Cui HL, Rennig M, et al. Increased production of periplasmic proteins in *Escherichia coli* by directed evolution of the translation initiation region. Microb Cell Factories. 2020 Dec;19(1):85.

95. Mirzadeh K, Martínez V, Toddo S, Guntur S, Herrgård MJ, Elofsson A, et al. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. ACS Synth Biol. 2015 Sep 18;4(9):959–65.

96. Datta N, Kontomichalou P. Penicillinase Synthesis Controlled By Infectious R Factors In Enterobacteriaceae. Nature. 1965 Oct;208(5007):239–41.

97. Sutcliffe JG. Complete Nucleotide Sequence of the *Escherichia coli* Plasmid pBR322. Cold Spring Harb Symp Quant Biol. 1979 Jan 1;43(0):77–90.

98. Fröhlich C, Chen JZ, Gholipour S, Erdogan AN, Tokuriki N. Evolution of β-lactamases and enzyme promiscuity. Protein Eng Des Sel. 2021 Feb 15;34:gzab013.

99. Overton TW. Recombinant protein production in bacterial hosts. Drug Discov Today. 2014 May;19(5):590–601.

100. Korpimäki T, Kurittu J, Karp M. Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. J Microbiol Methods. 2003 Apr;53(1):37–42.

101. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 1986 May;189(1):113–30.

102. Oka A, Sugisaki H, Takanami M. Nucleotide sequence of the kanamycin resistance transposon Tn903. J Mol Biol. 1981 Apr;147(2):217–26.

103. Kotra LP, Haddad J, Mobashery S. Aminoglycosides: Perspectives on Mechanisms of Action and Resistance and Strategies to Counter Resistance. Antimicrob Agents Chemother. 2000 Dec;44(12):3249–56.

104. Umezawa H. Studies on aminoglycoside antibiotics: enzymic mechanism of resistance and genetics. Jpn J Antibiot. 1979 Dec;32 Suppl:S1-14.

105. World Health Organization. Global action plan on antimicrobial resistance. Geneva: World Health Organization; 2015. 28 p.

106. Aguilera J, Gomes AR, Olaru I. Principles for the risk assessment of genetically modified microorganisms and their food products in the European Union. Int J Food Microbiol. 2013 Oct;167(1):2–7.

107. Fraiture MA, Deckers M, Papazova N, Roosens NHC. Are antimicrobial resistance genes key targets to detect genetically modified microorganisms in fermentation products? Int J Food Microbiol. 2020 Oct;331:108749.

108. Peubez I, Chaudet N, Mignon C, Hild G, Husson S, Courtois V, et al. Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. Microb Cell Factories. 2010 Dec;9(1):65.

109. Degryse E. Stability of a host-vector system based on complementation of an essential gene in *Escherichia coli*. J Biotechnol. 1991 Apr;18(1–2):29–39.

110. Szpirer CY, Milinkovitch MC. Separate-Component-Stabilization System for Protein and Dna Production Without the Use of Antibiotics. BioTechniques. 2005 May;38(5):775–81.

111. Chávez-Béjar MI, Balderas-Hernandez VE, Gutiérrez-Alejandre A, Martinez A, Bolívar F, Gosset G. Metabolic engineering of *Escherichia coli* to optimize melanin synthesis from glucose. Microb Cell Factories. 2013 Dec;12(1):108.

112. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol. 2003 Jul;21(7):796–802.

113. Park JH, Lee SY. Fermentative production of branched chain amino acids: a focus on metabolic engineering. Appl Microbiol Biotechnol. 2010 Jan;85(3):491–506.

114. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol. 2010 Mar;8(3):207–17.

115. Delbrück M. Interference Between Bacterial Viruses: III. The Mutual Exclusion Effect and the Depressor Effect. J Bacteriol. 1945 Aug;50(2):151–70.

116. Huang CJ, Lin H, Yang X. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. J Ind Microbiol Biotechnol. 2012 Mar 1;39(3):383–99.

117. Nakano K, Rischke M, Sato S, Märkl H. Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. Appl Microbiol Biotechnol. 1997 Nov 25;48(5):597–601.

118. Campbell JL, Richardson CC, Studier FW. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc Natl Acad Sci. 1978 May;75(5):2276–80.

119. Shiloach J, Rinas U. Glucose and Acetate Metabolism in E. coli – System Level Analysis and Biotechnological Applications in Protein Production Processes. In: Lee SY, editor. Systems Biology and Biotechnology of *Escherichia coli*. Dordrecht: Springer Netherlands; 2009. p. 377–400.

120. Yoon S, Han MJ, Jeong H, Lee C, Xia XX, Lee DH, et al. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. Genome Biol. 2012;13(5):R37.

121. Demerec M, Fano U. BACTERIOPHAGE-RESISTANT MUTANTS IN *ESCHERICHIA COLI*. Genetics. 1945 Mar 15;30(2):119–36.

122. Silverstone AE, Arditti RR, Magasanik B. Catabolite-Insensitive Revertants of *Lac* Promoter Mutants. Proc Natl Acad Sci. 1970 Jul;66(3):773–9.

123. Chamberlin M, Mcgrath J, Waskell L. New RNA Polymerase from *Escherichia coli* infected with Bacteriophage T7. Nature. 1970 Oct;228(5268):227–31.

124. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, et al. Consequences of Membrane Protein Overexpression in *Escherichia coli*. Mol Cell Proteomics. 2007 Sep;6(9):1527–50.

125. James J, Yarnall B, Koranteng A, Gibson J, Rahman T, Doyle DA. Protein over-expression in *Escherichia coli* triggers

adaptation analogous to antimicrobial resistance. Microb Cell Factories. 2021 Dec;20(1):13.

126. Schlegel S, Genevaux P, de Gier JW. De-convoluting the Genetic Adaptations of *E. coli* C41(*DE3*) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields. Cell Rep. 2015 Mar;10(10):1758–66.

127. Miroux B, Walker JE. Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. J Mol Biol. 1996 Jul;260(3):289–98.

128. Kwon SK, Kim SK, Lee DH, Kim JF. Comparative genomics and experimental evolution of *Escherichia coli* BL21(*DE3*) strains reveal the landscape of toxicity escape from membrane protein overproduction. Sci Rep. 2015 Nov 4;5(1):16076.

129. Angius F, Ilioaia O, Amrani A, Suisse A, Rosset L, Legrand A, et al. A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. Sci Rep. 2018 Jun 5;8(1):8572.

130. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci. 2008 Sep 23;105(38):14371–6.

131. Heyde SAH, Nørholm MHH. Tailoring the evolution of BL21(*DE3*) uncovers a key role for RNA stability in gene expression toxicity. Commun Biol. 2021 Aug 12;4(1):963.

132. Stewart EJ. Disulfide bond formation in the *Escherichia coli* cytoplasm: an invivo role reversal for the thioredoxins. EMBO J. 1998 Oct 1;17(19):5543–50.

133. Bessette PH, Åslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the

Escherichia coli cytoplasm. Proc Natl Acad Sci. 1999 Nov 23;96(24):13703–8.

134. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J. Conversion of a Peroxiredoxin into a Disulfide Reductase by a Triplet Repeat Expansion. Science. 2001 Oct 5;294(5540):158–60.

135. Drott D. Overcoming the codon bias of *E. coli* for enhanced protein expression. inNovations. 2001 Jan 1;12.

136. Luli GW, Strohl WR. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl Environ Microbiol. 1990 Apr;56(4):1004–11.

137. Bernal V, Castaño-Cerezo S, Cánovas M. Acetate metabolism regulation in *Escherichia coli*: carbon overflow, pathogenicity, and beyond. Appl Microbiol Biotechnol. 2016 Nov;100(21):8985–9001.

138. Lara AR, Caspeta L, Gosset G, Bolívar F, Ramírez OT. Utility of an *Escherichia coli* strain engineered in the substrate uptake system for improved culture performance at high glucose and cell concentrations: An alternative to fed-batch cultures. Biotechnol Bioeng. 2008 Mar;99(4):893–901.

139. Phue J, Noronha SB, Hattacharyya R, Wolfe AJ, Shiloach J. Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: Differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. Biotechnol Bioeng. 2005 Jun 30;90(7):805–20.

140. Small P, Blankenhorn D, Welty D, Zinser E, Slonczewski JL. Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of rpoS and growth pH. J Bacteriol. 1994 Mar;176(6):1729–37.

141. Tao H, Bausch C, Richmond C, Blattner FR, Conway T. Functional Genomics: Expression Analysis of *Escherichia coli* Growing on Minimal and Rich Media. J Bacteriol. 1999 Oct 15;181(20):6425–40.

142. Sánchez-Clemente R, Igeño MI, Población AG, Guijo MI, Merchán F, Blasco R. Study of pH Changes in Media during Bacterial Growth of Several Environmental Strains. In: Environment, Green Technology, and Engineering International Conference. MDPI; 2018. p. 1297.

143. Miller JH. Experiments in molecular genetics. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1972. 466 p.

144. Biselli E, Schink SJ, Gerland U. Slower growth of *Escherichia coli* leads to longer survival in carbon starvation due to a decrease in the maintenance rate. Mol Syst Biol. 2020 Jun;16(6):e9478.

145. Kram KE, Finkel SE. Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. Kelly RM, editor. Appl Environ Microbiol. 2015 Jul;81(13):4442–50.

146. Farrell MJ, Finkel SE. The Growth Advantage in Stationary-Phase Phenotype Conferred by *rpoS* Mutations Is Dependent on the pH an dNutrient Environment. J Bacteriol. 2003 Dec 15;185(24):7044–52.

147. Nyström T. Stationary-Phase Physiology. Annu Rev Microbiol. 2004 Oct 1;58(1):161–81.

148. Baez A, Kumar A, Sharma AK, Anderson ED, Shiloach J. Effect of amino acids on transcription and translation of key genes in *E. coli* K and B grown at a steady state in minimal medium. New Biotechnol. 2019 Mar;49:120–8.

149. KD T. Improved media for growing plasmid and cosmid clones. Bethesda Res Lab Focus. 1987;9:12.

150. BD Bionutrients (TM) Technical Manual. BD Bionutrients(TM); BD bionutrhttp://www.bd.com/ds/technicalcenter/misc/br_3_2547.pdf

151. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014 Apr 17;5.

152. Aguilar Lucero D, Cantoia A, Ceccarelli EA, Rosano GL. Starting a new recombinant protein production project in *Escherichia coli*. In: Methods in Enzymology.

153. Sieben M, Giese H, Grosch J, Kauffmann K, Büchs J. Permeability of currently available microtiter plate sealing tapes fail to fulfil the requirements for aerobic microbial cultivation. Biotechnol J. 2016 Dec;11(12):1525–38.

154. Rosano GL, Bruch EM, Ceccarelli EA. Insights into the CLP/HSP100 Chaperone System from Chloroplasts of Arabidopsis thaliana. J Biol Chem. 2011 Aug;286(34):29671–80.

155. Kram KE, Finkel SE. Culture Volume and Vessel Affect Long-Term Survival, Mutation Frequency, and Oxidative Stress of Escherichia coli. Appl Environ Microbiol. 2014 Mar;80(5):1732–8.

156. Dong H, Nilsson L, Kurland CG. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J Bacteriol. 1995 Mar;177(6):1497–504.

157. Katipoglu-Yazan T, Dev S, Desmond-Le Quéméner E, Bouchez T. Data on the influence of temperature on the growth of Escherichia coli in a minimal medium containing glucose as the sole carbon source for the joint computation of growth yields and rates at each temperature from 27 to 45°C. Data Brief. 2023 Jun;48:109037.

158. Schein CH, Noteborn MHM. Formation of Soluble Recombinant Proteins in *Escherichia Coli* is Favored by Lower Growth Temperature. Nat Biotechnol. 1988 Mar;6(3):291–4.

159. Vasina JA, Baneyx F. Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the *Escherichia coli* cspA and tac Promoter Systems. Protein Expr Purif. 1997 Mar;9(2):211–8.

160. Vera A, González-Montalbán N, Arís A, Villaverde A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng. 2007 Apr 15;96(6):1101–6.

161. Marschall L, Sagmeister P, Herwig C. Tunable recombinant protein expression in *E. coli*: enabler for continuous processing? Appl Microbiol Biotechnol. 2016 Jul;100(13):5719–28.

162. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif. 2005 May;41(1):207–34.

163. Jones PG, Inouye M. The cold-shock response — a hot topic. Mol Microbiol. 1994 Mar;11(5):811–8.

164. Higuchi K, Yabuki T, Ito M, Kigawa T. Cold shock proteins improve *E. coli* cell-free synthesis in terms of soluble yields of aggregation-prone proteins. Biotechnol Bioeng. 2020 Jun;117(6):1628–39.

165. De Marco A, Vigh L, Diamant S, Goloubinoff P. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol–overexpressed molecular chaperones. Cell Stress Chaperones. 2005;10(4):329.

166. Martínez-García E, Aparicio T, Goñi-Moreno A, Fraile S, de Lorenzo V. SEVA 2.0: an update of the Standard European

Vector Architecture for de-/re-construction of bacterial functionalities. Nucleic Acids Res. 2015 Jan 28;43(D1):D1183–9.

167. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat Methods. 2006 Aug;3(8):623–8.

168. Zutz A, Hamborg L, Pedersen LE, Kassem MM, Papaleo E, Koza A, et al. A dual-reporter system for investigating and optimizing protein translation and folding in *E. coli*. Nat Commun. 2021 Oct 19;12(1):6093.

169. Kumar S, Tiwari V, Doerrler WT. Cpx-dependent expression of YqjA requires cations at elevated pH. FEMS Microbiol Lett [Internet]. 2017 Jul 3 ;364(12).

170. Kraft M, Knüpfer U, Wenderoth R, Pietschmann P, Hock B, Horn U. An online monitoring system based on a synthetic sigma32-dependent tandem promoter for visualization of insoluble proteins in the cytoplasm of *Escherichia coli*. Appl Microbiol Biotechnol. 2007 May;75(2):397–406.

171. Schultz T, Martinez L, De Marco A. The evaluation of the factors that cause aggregation during recombinant expression in *E. coli* is simplified by the employment of an aggregation-sensitive reporter. Microb Cell Factories. 2006 Dec;5(1):28.

172. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. Appl Environ Microbiol. 1998 May;64(5):1694–9.

173. Nishihara K, Kanemori M, Yanagi H, Yura T. Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia coli*. Appl Environ Microbiol. 2000 Mar;66(3):884–9.

174. Chiang CS, Bremer H. Stability of pBR322-derived plasmids. Plasmid. 1988 Nov;20(3):207–20.

175. Skogman G, Nilsson J, Gustafsson P. The use of a partition locus to increase stability of tryptophan-operon-bearing plasmids in *Escherichia coli*. Gene. 1983 Aug;23(2):105–15.

176. Miller CA, Tucker WT, Meacock PA, Gustaf Sson P, Cohen SN. Nucleotide sequence of the partition locus of *Escherichia coli* plasmid pSC101. Gene. 1983 Oct;24(2–3):309–15.

177. Li Z, Rinas U. Recombinant protein production associated growth inhibition results mainly from transcription and not from translation. Microb Cell Factories. 2020 Dec;19(1):83.

178. Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng. 1990 Mar 25;35(7):668–81.

179. Sun X, Zhang Z, Wang L, Wang J, Liang Y, Yang H, et al. Downregulation of T7 RNA polymerase transcription enhances pET-based recombinant protein production in *Escherichia coli* BL21 (*DE3*) by suppressing autolysis. Biotechnol Bioeng. 2021 Jan;118(1):153–63.

180. Kuipers G, Karyolaimos A, Zhang Z, Ismail N, Trinco G, Vikström D, et al. The tunable pReX expression vector enables optimizing the T7-based production of membrane and secretory proteins in *E. coli*. Microb Cell Factories. 2017 Dec;16(1):226.

181. Neubauer P, Lin HY, Mathiszik B. Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. Biotechnol Bioeng. 2003 Jul 5;83(1):53–64.

182. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of

overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expr Purif. 2004 Sep;37(1):203–6.

183. Pan S hao, Malcolm BA. Reduced Background Expression and Improved Plasmid Stability with pET Vectors in BL21 (*DE3*). BioTechniques. 2000 Dec;29(6):1234–8.

184. Sieben M, Steinhorn G, Müller C, Fuchs S, Ann Chin L, Regestein L, et al. Testing plasmid stability of *Escherichia coli* using the Continuously Operated Shaken BIOreactor System. Biotechnol Prog. 2016 Nov;32(6):1418–25.

185. Birnbaum S, Bailey JE. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. Biotechnol Bioeng. 1991 Apr 5;37(8):736–45.

186. Balding C, Blaby I, Summers D. A mutational analysis of the ColE1-encoded cell cycle regulator Rcd confirms its role in plasmid stability. Plasmid. 2006 Jul;56(1):68–73.

187. Summers DK, Sherratt DJ. Multimerization of high copy number plasmids causes instability: Cole 1 encodes a determinant essential for plasmid monomerization and stability. Cell. 1984 Apr;36(4):1097–103.

188. Kallehauge TB, Li S, Pedersen LE, Ha TK, Ley D, Andersen MR, et al. Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. Sci Rep. 2017 Jan 16;7(1):40388.

189. Panayotatos N. Recombinant protein production with minimal-antibiotic-resistance vectors. Gene. 1988 Dec;74(2):357–63.

190. Eguia FAP, Ramos HR, Kraschowetz S, Omote D, Ramos CRR, Ho PL, et al. A new vector for heterologous gene expression in *Escherichia coli* with increased stability in the absence of antibiotic. Plasmid. 2018 Jun;98:22–30.

191. Roberts MC. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol Rev. 1996 Oct;19(1):1–24.

192. Shimohata N, Chiba S, Saikawa N, Ito K, Akiyama Y. The Cpx stress response system of *Escherichia coli* senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site. Genes Cells. 2002 Jul;7(7):653–62.

193. Guidi C, De Wannemaeker L, De Baets J, Demeester W, Maertens J, De Paepe B, et al. Dynamic feedback regulation for efficient membrane protein production using a small RNA-based genetic circuit in *Escherichia coli*. Microb Cell Factories. 2022 Dec 15;21(1):260.

194. Mirzadeh K, Toddo S, Nørholm MHH, Daley DO. Codon Optimizing for Increased Membrane Protein Production: A Minimalist Approach. In: Mus-Veteau I, editor. Heterologous Expression of Membrane Proteins. New York, NY: Springer New York; 2016. p. 53–61. (Methods in Molecular Biology; vol. 1432).

195. Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, et al. Protein production and purification. Nat Methods. 2008 Feb;5(2):135–46.

196. Meyer AJ, Segall-Shapiro TH, Glassey E, Zhang J, Voigt CA. *Escherichia coli* "Marionette" strains with 12 highly optimized small-molecule sensors. Nat Chem Biol. 2019 Feb;15(2):196–204.