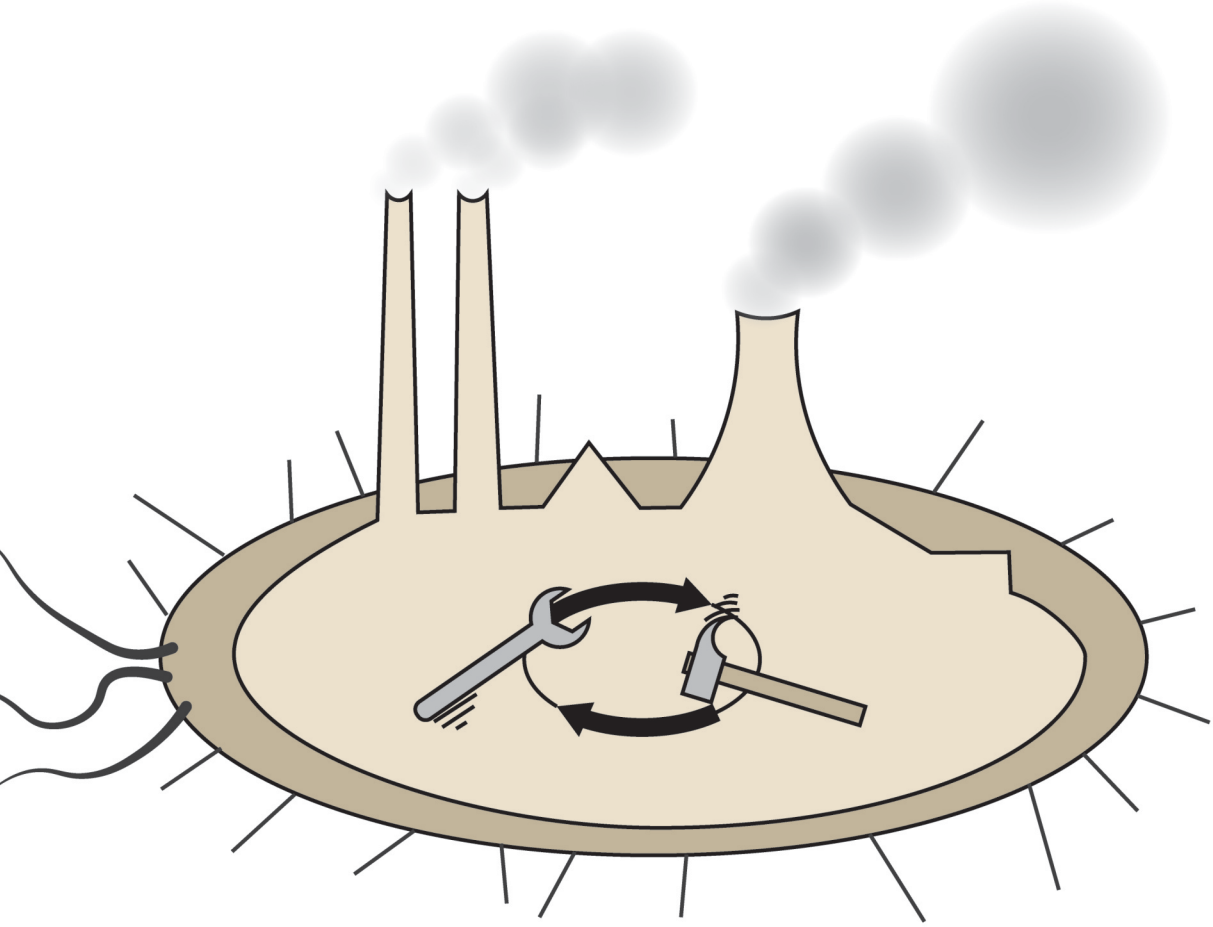
Stockholm University

Next generation tools for microbial cell factories

Alister James Cumming



Doctoral Thesis in Biochemistry at Stockholm University, Sweden 2023

Next generation tools for microbial cell factories

Alister James Cumming

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Wednesday 20 September 2023 at 14.00 in Magnelisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16 B.

Abstract

The bacterium E. coli is a popular choice for expressing recombinant proteins. It is easy to culture and many molecular tools for are available for directing protein production. When producing recombinant proteins, high yields and good quality soluble products are desirable. Poor yields or misfolded, aggregated and insoluble products are unwanted outcomes. In this thesis, existing tools have been evolved to enable increased yields of recombinant products from microbial cell factories. New tools have also been developed that help the user avoid bacterial stress and improve the quality of the recombinant proteins.

Keywords: *Microbial cell factories, Recombinant protein production, Bacteria, Plasmid, Genetic sensor, Inducible promoter, Genetic cassette.*

Stockholm 2023
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-219865

ISBN 978-91-8014-444-5
ISBN 978-91-8014-445-2

Department of Biochemistry and Biophysics

Stockholm University, 106 91 Stockholm

NEXT GENERATION TOOLS FOR MICROBIAL CELL FACTORIES

Alister James Cumming

Stockholm
University

Next generation tools for microbial cell factories

Alister James Cumming

©Alister James Cumming, Stockholm University 2023

ISBN print 978-91-8014-444-5
ISBN PDF 978-91-8014-445-2

Printed in Sweden by Universitetsservice US-AB, Stockholm 2023

Acknowledgments

I would like to thank:

My supervisor, Dan. Thank you for being an amazing mentor, giving me cool projects, and great guidance and advice! I am very grateful to have had the opportunity to do a PhD in your lab. I have learned a lot in the past 5 years and have really enjoyed nearly all of it! I will have fond memories of my time in here. Thank you!

My lab mate, Diana, who rescued my from solitude and isolation, by joining me in the lab, and creating a great atmosphere. You were a pleasure to work with, and have endless discussions with. I wish you the best of luck in your future! Thank you!

My mentor in the lab, Pat, who showed me how to 'biochemistry' for his final months here at the department. I learnt a lot with you and am very grateful for your help.

My lab mates, and former lab mates, Mati, Kiavash and Rageia who helped me progress as a scientist and collaborated with me on projects.

My former students, Sokeina, Ramona, Nicklas and HL! You guys were all a pleasure to supervise! I wish you all the best in the futures and am grateful for your contributions!

The GvH community, past and present, for all the great discussion in lab meetings and JCs and for general good environment! Ane, Justin, Felix, Daphne, Alex, Saira, Urska, Gunnar, Einar, Jan-Willem, Grant, Nir, and Jose, thank you!

My collaborators Leopold and Sara who helped me with mass

spectrometry analysis. And thank you to my collaborators at DTU, Morten and Carolyn.

And all of those who shared time with me here at DBB and made this a great place to work!

Contents

List of papers **1**

Author contributions to papers **2**

Summary **3**

Sammanfattning på svenska **4**

Protein biogenesis in bacteria **6**

Transcription **7**

Translation **12**

Translation initiation - Ribosome Assembly 12

Translation initiation – mRNA features 14

RNA thermometers 16

Riboswitches 16

Small RNAs (sRNAs) 17

Translation elongation 18

Translation termination and Ribosome recycling 19

Folding **20**

Folding in the cytoplasm 20

Folding in the periplasm 22

Protein misfolding and aggregation 25

Recombinant protein production **28**

E. coli as a production host 28

Plasmid copy number 28

Transcriptional control 29

Translational control 32

The TIR 32

Synthetic Riboswitches 33

Synthetic sRNAs 34

Metabolic burden 34

Biosensors.....36

Short summary of papers.....39

Future perspectives.....44

References.....48

List of papers

Paper I.
Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*.
Shilling PJ, Mirzadeh K, **Cumming AJ**, Widesheim M, Köck Z, Daley DO.
Commun Biol. 2020 May 7;3(1):214. doi: 10.1038/s42003-020-0939-8.

Paper II.
Signal amplification of araC pBAD using a standardized translation initiation region.
Shilling PJ, Khananisho D, **Cumming AJ**, Söderström B, Daley DO.
Synth Biol (Oxf). 2022 Jul 5;7(1):ysac009. doi: 10.1093/synbio/ysac009.

Paper III.
Antibiotic-Efficient Genetic Cassette for the TEM-1 β-Lactamase That Improves Plasmid Performance.
Cumming AJ*, Khananisho D*, Harris R, Bayer CN, Nørholm MHH, Jamshidi S, Ilag LL, Daley DO.
ACS Synth Biol. 2022 Jan 21;11(1):241-253. doi: 10.1021/acssynbio.1c00393.

*both authors contributed equally to this manuscript

Paper IV.
A genetic sensor for the production of proteins in the periplasm of *Escherichia coli*.
Cumming AJ, Khananisho D, Nicklas Liljestrand, Balka M, Daley DO.
In preparation.

Author contributions to papers

Paper I.
My contribution to this paper was minor. I collected data for Figures 4c, 4d and 5.

Paper II.
My contribution to this paper was minor. I collected data for Figures 2a and 2d.

Paper III.
My contribution to this paper was significant. I helped design experiments for Figures 2, 3, 4, 5 and 6. I collected all the data in Figures 2, 3, 6, S1, S3 and S4 and contributed data to figure 4. Finally, I helped write the manuscript.

Paper IV.
My contribution to this manuscript has been significant. I helped design experiments and collected all the data presented in Figures 1, 3, 4, and S2. I also collected data presented in Figure 2 and S1.

Summary

The bacterium *E. coli* is a popular choice for expressing recombinant proteins. It is easy to culture and many molecular tools for are available for directing protein production. When producing recombinant proteins, high yields and good quality soluble products are desirable. Poor yields or misfolded, aggregated and insoluble products are unwanted outcomes. In this thesis, existing tools have been evolved to enable increased yields of recombinant products from microbial cell factories. New tools have also been developed that help the user avoid bacterial stress and improve the quality of the recombinant proteins.

Sammanfattning på svenska

Bakterier är små enkla organismer. De används i stor utsträckning för att producera proteiner som vanligtvis är av intresse för akademin eller industrin. Dessa proteiner kan vara farmaceutiskt, industriellt eller miljömässigt relevanta. Exempel på några av dessa är antikroppsfragment, hormoner eller enzymer. För att producera ett protein av intresse (rekombinant protein) måste en gen som kodar för proteinet vara tillgänglig för bakterierna att uttrycka. Dessa gener introduceras till bakterierna i en plasmid som kodar för genen. En plasmid är en självreplikerande cirkulär DNA-molekyl som finns utanför bakteriens eget DNA (genom). Produktion av rekombinant proteinet utförs i tre huvudsteg: 1. Bakterierna producerar en budbärarmolekyl (så kallat mRNA) som innehåller information om genen av intresse, i en process som kallas Transkription. Transkriptionen kontrolleras av en promotor som sitter på plasmiden, som i sin tur rekryterar ett RNA-polymeras. Detta enzym skapar en mRNA-molekyl från DNA gen-mallen. 2. mRNA:t översätts till en polypeptid, en kedja av aminosyror, av en molekyl som kallas för ribosomen. 3. Polypeptiden vekar sig till sin korrekta proteinkonformation.

I denna doktorsavhandling har man arbetat med element av plasmider som påverkar deras prestation i proteinproduktion. I artikel I har promotorn på plasmiden förbättrats. Promotorn är området som rekryterar RNA-polymeraset. Den ligger uppströms av genen av intresse på en plasmid. Denna förbättring ledde till ökad kvantitet av det rekombinanta proteinet. I artiklar I och II utvecklades mRNA-molekylerna för diverse gener av intresse i syfte att effektivisera translationen av mRNA. Detta resulterade i ökade kvantiteter av rekombinant protein. I artikel I hade förändringarna i promotor regionen, samt i mRNA, en additiv effekt. Den resulterande plasmiden, vid namn pET28 TIR2 $T7p^{CONS}$, gav över 100 gånger ökade kvantiteter av protein.

Plasmidinnehållande celler selekteras genom uttryck av antibiotikaresistensgener. Detta görs för att möjliggöra överlevnad i medium som innehåller en annars dödligt mängd antibiotikum. Ett av de mest populära

antibiotikum som används är β-laktamantibiotika. Dock bryts dessa snabbt ned av enzymer som är kodat för i β-laktamasresistensgenerna. I artikel III förlängs β-laktamhalveringstiden genom att minska produktionen från β-laktamasresistensgenen. Ytterliga gynnsamma effekter av detta är att uppkomsten av en plasmidfri population av celler minskades.

När proteinet av intresse som uttryckt i en bakterie inte vekar sig korrekt till sin nativa struktur, kan detta leda till stress från felveckningen och påföljande aggregering. I artikel IV har en sensor för stress som uppkommer vid felveckning av proteiner utvecklats. Genom att använda denna stresssensor i samband med rekombinant proteinuttryck i bakterier så kan proteinuttrycket optimeras. Detta resulterar i förbättrad löslighet av protein av intresse.

Protein biogenesis in bacteria

E. coli are often used as factories to express recombinant genes and obtain recombinant proteins. Tools for recombinant protein production often take advantage of native transcription, translation, and folding pathways. This can affect the biogenesis of native proteins within the cell, as well as the cells physiology, which in turn may be detrimental to recombinant protein production. In this thesis I will first describe the basic steps involved in protein biogenesis in bacteria, and how they are regulated. I will then go on to describe how this knowledge has been applied to develop tools for microbial cell factories and how my work has contributed to the field.

Transcription

Transcription is the process where information stored in the genome is converted from DNA into mRNA. The process is driven by the RNA polymerase which is composed of 6 subunits, $\alpha_2\beta\beta'\omega$, forming a crab-claw like structure, and a sigma factor (1). Sigma factors are proteins that direct the RNA polymerase holoenzyme to promoters, specific regions on the genome upstream of coding sequences were transcription is directed to start (2). They recognise promoters in a sequence specific manner, by two regions: σ2 which recognises the -10 and the discriminator motif and the σ4 which recognises the -35 motif (3). Bacterial sigma factors differ structurally, for example the sigma70 (σ70) is composed of regions σ1.1, σ2, σ3, σ3.2 and σ4 (4), and the extracytoplasmic function (ECF), or group 4 sigma factors are composed of just two regions; σ2 and σ4 (5).

Sigma factors recognise a specific sets of promoters and thus regulate a specific gene or sets of genes (3,6). This set of genes is termed the regulon, and encompasses proteins required for a specific adaptation or coping with a specific stress. There are 7 known sigma factors in *E. coli* that compete for the core RNA polymerase (7).

Sigma factor competition is mediated by the affinity of the sigma factor for the polymerase, regulators of polymerase affinity, sigma factor abundance, sigma factor clearance, anti-sigma factors, and transcriptional regulators (7,8). In conditions of plenty it is the vegetative σ70 that dominates. This is due to its greater abundance and higher binding affinity for the core RNA polymerase (9). This affinity is modulated by ppGpp (magic spot) and Rsd; the anti-sigma factor for σ70 (7–10). The affinity of the core RNA polymerase for σ70 is decreased by ppGpp, which binds directly to the core (10). The ppGpp-RNA polymerase complex has been proposed to be stabilised by another protein, Dsk. The ppGpp-RNA polymerase complex has lower affinity for σ70 and thus raises the competitiveness of the alternative sigma factors for the core polymerase (7,10). The synthesis and turnover of ppGpp is performed by RelA and SpoT which are

activated by the absence of charged tRNAs - causing ribosome pausing or arrest due to starvation (11). Indeed ppGpp has been shown to be essential for the stationary phase responsive sigma factor σS and heat shock sigma factor, σ32 (10). Due to its role in facilitating competitiveness of alternative sigma factors ppGpp has been implicated in a wide variety of stress responses and is considered a general alarmone (12,13).

Rsd regulates σ70 by directly binding to the sigma factor (14). Bound by Rsd, σ70 is unable to form a complex with the core RNA polymerase and thus alternative sigma factor competitiveness is increased (14). Rsd has two promoters which are activated by σ70 itself, and by σS. ppGpp was found to increase expression of Rsd on entry into stationary phase (15). The concentration of Rsd has been shown to increase in stationary phase and reaches an almost equivalent concentration to that of σ70 in deep stationary phase (16). Regulation of sigma factors by anti-sigma factors is a common mechanism of control (σE by rcsA, σ32 by GroEL/S and DnaK/J/GrpE) (17). Another form of σ70 regulation is 6S RNA binding to the σ4 region of σ70 when in complex with of the RNA polymerase. The 6S RNA blocks binding to the -35 region of targeted promoters, inhibiting transcription. Unlike ppGpp or Rsd, 6S RNA does not modulate σ70 competition *per se*, but rather modulates the holo-RNA polymerase-σ70's ability to bind to promoters and thus initiate transcription.

To exemplify the complexity of sigma factor regulation I will briefly discuss the regulation of σ32 (Figure 1). The gene encoding σ32 is RpoH which has 5 annotated promoters and a plethora of nutritional and stress related transcription factor binding sites (regulonDB, (18)). Several of the promoters are σ70 dependent and transcription and translation of RpoH/σ32 is thus at least partly constitutive (19). Under normal growth conditions the concentrations of σ32 are kept low. This is achieved in three ways; by restriction of translation using an RNA thermometer, by chaperone binding, and by rapid degradation of the sigma factor (20). The signal recognition partial (SRP) plays a role in σ32 control, as it transports σ32 to the inner membrane where FtsH degrades it (21) (see Figure 1). In non-stressed conditions the σ32 half-life has been reported to be as short as 30 seconds (22), however upon heat shock, σ32 is stabilised. Upon a heat shock of 42ºC the free energy becomes sufficient to melt a hairpin in the RpoH mRNA, making the Shine-Dalgarno available for ribosome docking and assembly and allowing translation of σ32 to be increased (23). The chaperones that under non-stress conditions bind to σ32 are titrated away as their clients (misfolding/ unfolded proteins) increase (see Figure 1). The sigma factor is then

free to complex with the RNA polymerase and directs transcription at σ32 dependent promoters. The regulated promoters are rich in chaperones and proteases that help deal with a heat shock stress. As their abundance increases and the stress subsides these chaperones and proteases once again bind and degrade σ32 in a negative feedback loop (23).

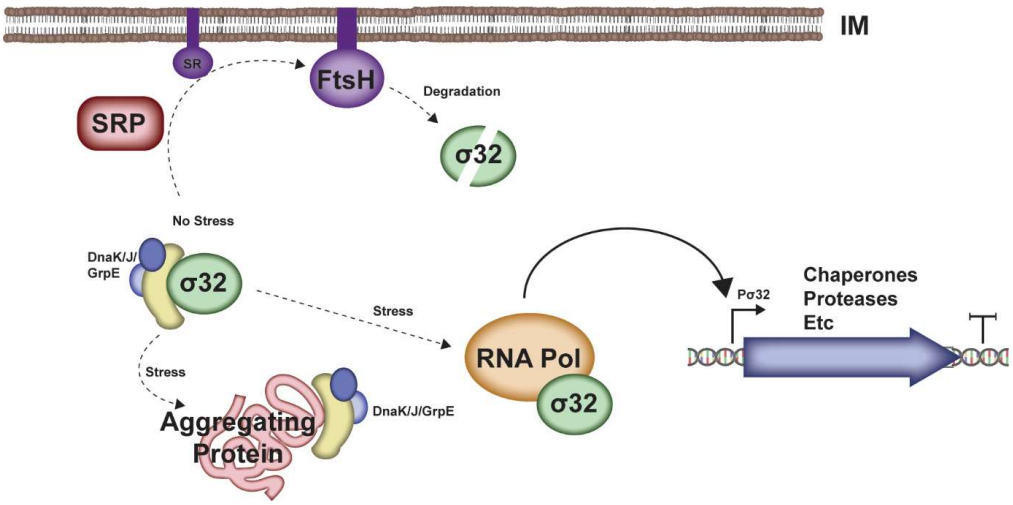


Figure 1. The complex regulation of σ32. *Under non-stressed conditions DnaK/J/GrpE sequester σ32 and with the aid of SRP it is transported to the IM and degraded. Under stress conditions, such as heat shock, proteins aggregate and DnaK/J/GrpE are titrated away. As a result σ32 is stabilised and can bind the RNA polymerase and regulate promoters. Figure inspired by* (21)

Transcriptional factors (TF) offer another layer of transcriptional control. They bind to specific regions of DNA and promote or repress RNA polymerase recruitment to promoters. They may act on a global level, affecting many major pathways of a bacterium, or on a local level, affecting limited sets of genes or operons (24). An example of a TF relevant to protein quality control and periplasmic homeostasis is CpxR, part of the Cpx two compartment system, or the Cpx response (see Figure 2).

The Cpx envelope stress response plays a key role in monitoring the state of the inner membrane (IM) and the periplasmic space (25). When activated, a number of periplasmic chaperones and proteases are upregulated that help fold or degrade misfolded periplasmic proteins (26,27) and it has been shown that Cpx plays a role in the regulation of the electron transport chain (28). Specific

sRNAs, such as CpxQ, are also produced which help control expression of regulated genes (29,30) (see Figure 2). The Cpx envelope stress response works alongside σE– which has key roles in monitoring the outer membrane (OM) and the transport of OM proteins to the OM (31). Interestingly it was found that Cpx regulates transcription of rpoE (the gene for σE) and inversely regulates key members of the σE regulon (32).

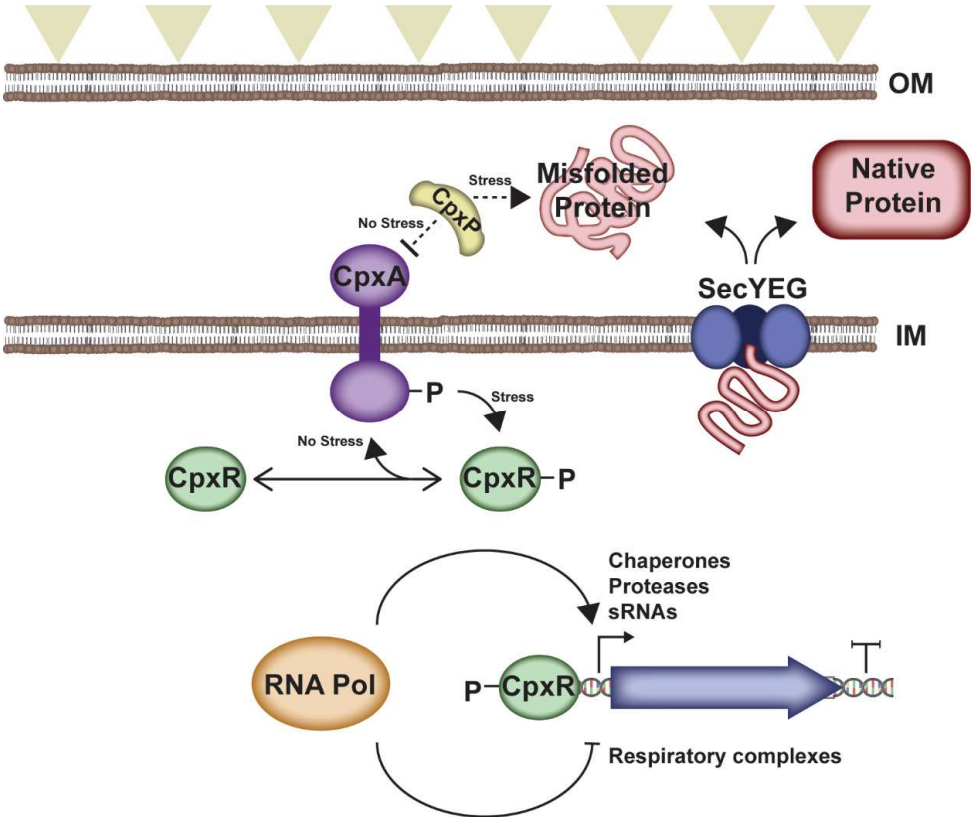


Figure 2. The Cpx two compartment system. *CpxA senses stress in the periplasm (sensor domain) and relays a message to the cytoplasm by phosphorylating its response regulator, CpxR. CpxA activity is suppressed by CpxP. CpxR-P is a transcription factor and promotes up- or down-regulation of genes. Figure inspired by* (36)

Cpx is activated when CpxA, a transmembrane protein with a sensor region in the periplasm and kinase region in the periplasm, phosphorylates its response

regulator CpxR (25,27,33). CpxR then binds to specific sequences resulting in activation or repression of sets of genes and operons (32,34) (Figure 2). NlpE is an OM protein that when in proximity to CpxA (for example, when it is mislocalised to the IM) strongly activates the Cpx response (35). Additionally a small periplasmic CpxP reduces cpxA activation, and is one of the most upregulated proteins in the Cpx regulon (32).

After the RNAP complex is assembled on the promoter, the polymerase typically goes through several cycles of abortive initiation, generating short polynucleotides of 3-15. After about 12 RNA polynucleotides have been generated, and occupy the RNA exit channel, affinity with the -35 region decreases and the polymerase can escape the promoter. Once this happens and the elongation phase begins the sigma factor dissociates and the elongation factor NusA associates (37). Transcription elongation continues until the RNAP reaches a termination sequence.
Termination can be Rho independent or dependent. In Rho independent termination, a tract of poly U enables easy melting of the DNA-RNA hybrid followed by a strong hairpin sequence. In Rho dependent termination Rho binds to the mRNA, and at transcriptional pause sites, terminates transcription in an ATP dependent manor by separating the RNAP from the mRNA (38).

Translation

The mRNA produced by transcription is translated by ribosomes into a polypeptide. Ribosome binding often occurs while the mRNA is still on the RNA polymerase, and a single mRNA may end up with many translating ribosomes at the same time, known as a polysome. The prosses of translation with be discussed below, with emphasis on translation initiation, and translation control.

Translation initiation - Ribosome Assembly

Ribosome assembly starts with the binding of the three initiation factors (IFs) to the 30S ribosomal subunit (Figure 3). These three initiation factors are called IF1, IF2 and IF3 (39,40). There is not believed to be an obligate order of assembly (41), but affinities of the IFs differ and thus a preferred order of assembly may exist. IF3 has the highest affinity for the 30S subunit and the measured arrival time was less than 1 ms. For IF2 the measured arrival time was 1.3 ms and for IF1 it was more than 30 ms (42). The functions of IF3 include interference with subunit association, regulation of tRNA binding in the P site, and discrimination of poor translation initiation regions (TIRs) on mRNAs. Its rapid binding also helps to protect against premature association of the 50S subunits. The function of IF2 is to recruit and bind the f-Met-tRNA to the 30S subunit, promote 50S subunit binding and 70S assembly and position the f-Met-tRNA correctly in the 70S ribosome prior to elongation. The function of IF1 is to stabilise the binding of IF2 and IF3 and to modulate correct mRNA and f-Met-tRNA selection (41). The arrival time of f-Met-tRNA was measured to be about 100 ms and would associate after the three initiation factors have already bound (42). The resulting complex is called the 30S pre- iniation complex (30S PIC) and structural re-arrangements mature the 30S PIC for correct tRNA and mRNA binding. 30S binding to mRNA can occur at any time during or after 30S PIC formation (39,41). IF1 and IF3 promote the rapid association of tRNAs into the P site to enable sampling and recruitment of the correct f-Met-tRNA (43). The formation of a 30S initiation complex (30S IC) occurs upon correct binding to

the start site of mRNA and pairing of the start codon with the anticodon of f-Met-tRNA (39,41). After formation of 30S IC the 50S subunit can bind. Upon correct loading of f-Met-tRNA to the 30S PIC, and upon 50 association, IF3 is released and initiation become irreversible. IF1 is also released upon 50S binding. GTP bound IF2 and f-Met-tRNA on the 30S subunit create a large interface for 50S binding. Upon binding GTP hydrolysis is triggered which results in a conformational shift, putting f-Met-tRNA into the P site. Once the ribosome has assembled, elongation factor (EF-Tu) can deliver an amino acylated-tRNA to the ribosome A-site, starting the elongation phase (44) (see Figure 3). Translation initiation is the rate limiting step in protein synthesis (45,46). Within this process the selection of appropriate mRNA is often the kinetic bottleneck. Features of mRNA and the Translation Initiation Region (TIR) which determine translation initiation efficiency are discussed below.

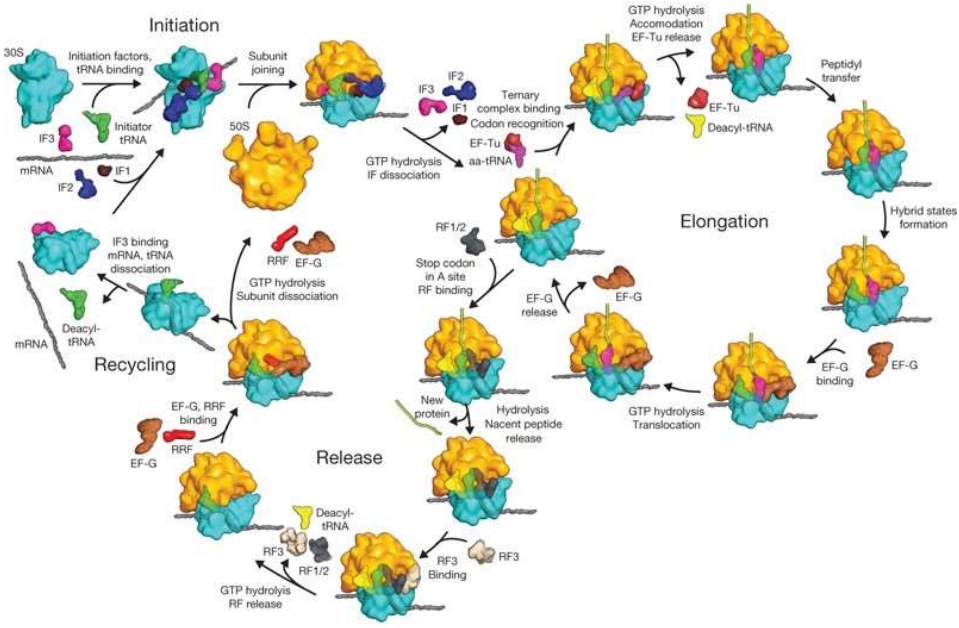


Figure 3. Translation initiation, elongation, release and recycling. *Each step in translation is detailed in this cartoon. Taken with permission from* (47).

Translation initiation – mRNA features

Translation initiation is the rate limiting step in protein synthesis (41,45,48). Formation of the 30S IC requires the association of many factors ranging in speed from less than 1 ms (IF3) to more 100 ms (f-Met-tRNA) (42). The speed of association (and stability once associated) of mRNA is influenced by structural features around the AUG start site (49). Structure in this region can vary from tight, compact, globular structures, to unstructured stretches of mRNA (Figure 4). The speed of 30S IC formation, or translation initiation, is influenced greatly for any particular gene by structure around the start site in an area called the translation initiation region (TIR). Differences in the TIR can result in many fold differences in expression rate (50). The TIR comprises the regions of approximately -20 upstream and +15 downstream of the start codon, and includes the Shine Dalgarno sequence (41). Structure of the TIR is considered to have the greatest effects on mRNA translation efficiency (39–41).

Translation initiates at start codons, which are usually AUG. In some cases GUG and UUG are used. Codon choice affects translation efficiency and UUG is about 10-fold less effective than AUG (49). mRNA can be divided into three categories with regard to translation initiation; SD lead, non-SD lead, and leaderless mRNA. In leaderless mRNA the start site, typically an AUG, is located near the 5' end of the mRNA and it directly binds the 70S ribosomes. Leaderless mRNA, while common in some bacteria is rare in *E. coli*. In non-SD lead mRNA a 5' UTR upstream of the start codon exists. In general these mRNAs possess a local lack of structure and a presence of an adenosine rich tract upstream of the start codon (51,52). This adenosine rich tract is thought to interact and recruit the 30S ribosome through an interaction with the S1 protein (41). In *E. coli* and other fast-growing organisms, leaderless and non-SD lead mRNAs are not common in highly expressed genes and therefore will not be discussed further here. SD lead mRNA contains a Shine-Dalgarno (SD) upstream of the AUG. The SD has a consensus sequence GGAGGU as identified by Shine and Dalgarno (53). The spacing between the start codon and the SD varies with optimal spacing between 5 and 9 nucleotides (49,54). The 30S rRNA contains an aSD sequence positioned to facilitate f-Met-tRNA interactions. Upon SD : aSD formation with mRNA with optimal spacing, the start codon is positioned in the P site to facilitate codon-anti-codon pairing (55). Upstream of the SD sequence an adenosine rich tract often exists. As with non-SD lead mRNA, this helps recruit the ribosome by interaction with the S1 protein. In fast growing bacteria SD lead mRNA constitutes up to 90% of the mRNA pool (57% for *E. coli*) (54).

Features that affect 30S IC formation and therefore translation initiation speed include the SD sequence, the SD – start codon spacing, the presence of an adenosine rich upstream tract, start codon selection, and importantly, mRNA structure. Again, the strength of the TIR, often dictated by secondary structure formation, is one of the key features in determining translation rate. Rare codons downstream of the AUG have also been suggested to increase translation elongation rate, however this may be because they cause less structure in the TIR, rather than tRNA abundance (56). Expression of recombinant proteins can be enhanced by changing the SD sequence, randomising spacer sequences, or synonymous codon changes in the TIR (57). This is believed to affect structure and there for 30S ribosome binding kinetics.

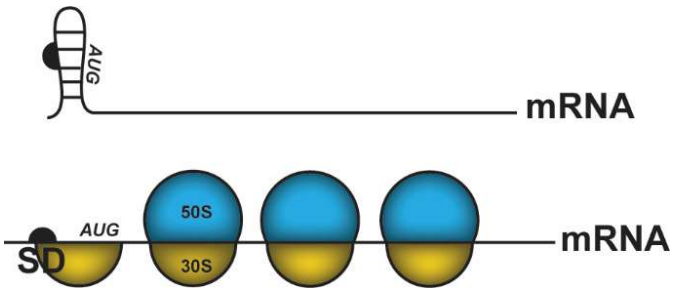


Figure 4. Structure in the TIR region of mRNA affects the translation initiation rate. *Structure in the TIR reduces the ability of ribosome binding and assembly. As illustrated here, a structured region results in low translation initiation rates, while an unstructured mRNA results in high translation initiation rates.*

Many efforts have been put into developing synthetic tools for predicting TIR strength, however so far they have not surpassed experimental approaches for determining structure (58), or predicting efficacy (59). Besides the TIR the 5' untranslated region (5' UTR) may contain other regulatory features such as RNA thermometers, protein binding sequences, or small molecule binding sequences. These features will be discussed here below.

RNA thermometers

RNA thermometers are structures of RNA, which typically occlude the SD or the TIR. They are molten and can become unstructured by shifts in temperature. An example is the ROSE like element in the 5'UTR of the mRNA of *ibpAB* which includes the genes for the inclusion body binding proteins A and B, part of the heat shock response in the σ32 regulon (60). At 30ºC a stable hairpin occludes the SD and limits translation but when the temperature shifts to 42ºC the thermometer melts, allowing ribosome binding and subsequent translation. The thermometer is made of 4 complementary base pairs that form a tight hairpin. In the centre of the hairpin there is a single unpaired G present on one side. Removal of this G has resulted in a more stable hairpin and elevated melting point making the thermometer insensitive to a 42ºC heat shock (60). Another example of an RNA thermometer is in the *rpoH* mRNA, encoding σ32. The 5'UTR of *rpoH* forms a hairpin with the start of the coding sequence, occluding the TIR (61). Again, upon temperature shift to 42ºC the hairpin melts and translation is increased.

Riboswitches

Riboswitches are ligand binding motifs, typically found in the 5' UTR of mRNAs that regulate translation initiation (but sometimes also termination) by ligand binding (62). They are structurally diverse and abundant throughout all prokaryotes, believed to be a relic from the RNA world (63). They function independently of protein factors and are involved in many cellular presses, with similar homologs found in diverse species. Ligand interactions with conserved structural motifs in riboswitches leads to structural rearrangements which result in modulated activity (62). Often such structural rearrangements are irreversible. Modes of action include regulation of structure in the RBS and inhibiting hairpin formation in terminators. Common ligands include cofactors, signalling molecules, and amino acids but the variety of ligands are broad. Ligand binding often controls import, synthesis or metabolism of the ligand. An example, and one of the first riboswitches identified in bacteria regulates vitamin B12 biosynthesis. The btuB gene, responsible for import or synthesis of vitamin B12 in bacteria is controlled by the concentration of B12 itself (64). When B12 is low or absent an anti-terminator sequence inhibits the characteristic formation of a terminator upstream of a poly-U track upstream of btuB, and as a result the gene is transcribed. When B12 is present in adequate concentrations it binds to the

riboswitch, resulting in formation of the terminator, and no further transcription. At a translational level B12 also regulates its own synthesis. In its absence the RBS is free for ribosome binding and assembly, however upon B12 binding a hairpin occluding the RBS is formed and translation is inhibited (62,64).

Small RNAs (sRNAs)

Another way of modulating translation is with sRNAs. sRNAs can modulate translation positively or negatively. Their mode of action varies widely and includes; binding to an RBS or upstream elements to directly inhibit translation, binding to structured regions of mRNA resulting in access or occultation of regulatory regions by restructuring the region, and binding to promote or retard mRNA cleavage by occluding or exposing cleavage sites (65). sRNAs can function unaided, however most require the protein Hfq to facilitate and stabilise sRNA-mRNA interactions (65,66). Rho independent terminator structures in sRNA consist of poly-U rich tails following strong hairpin structures which facilitate Hfq binding. Hfq is an RNA binding protein that forms a hexametric circular structure, binding RNA at both proximal and distal sites (67). These 2 sites have differing physiochemical properties and while the proximal site preferentially binds U residue tails, which are nearly always found in sRNAs, the distal site has stronger affinity of ARN_N residues that are often present in mRNAs (66).

Stress responsive regulons, such as that of Cpx and σE benefit from the action of sRNAs in broadening their regulons and dynamically regulating expression levels (29,68). There are many examples of sRNA regulating membrane and periplasmic proteins during stress responses. For example, upon Cpx stress activation, a small 3' derived sRNA is cleaved from CpxP mRNA, and regulates many targets. This sRNA is called cpxQ and it negatively regulates the expression of inner membrane proteins, which relieves IM stress (29). The sigma factor, σE also regulates a number of sRNAs that regulate a number of outer membrane proteins. By downregulating these outer membrane proteins in a sRNA dependent manor stress is relieved (65,68).

Proteins can also interact with mRNA to modulate translation. IbpA expression can again be used as an example here (Figure 5). The ROSE like element in the 5'UTR of *ibpA* mRNA (mentioned previously) has 2 loops upstream of the thermometer (69). It was shown that IbpA binds these loops, and in doing so hinders translation. As the mRNA is not protected by ribosomes it is degraded rapidly.

Upon protein aggregation, IbpA is titrated away allowing for translation and stabilisation of the mRNA by polysome formation, leading to greater protein synthesis.

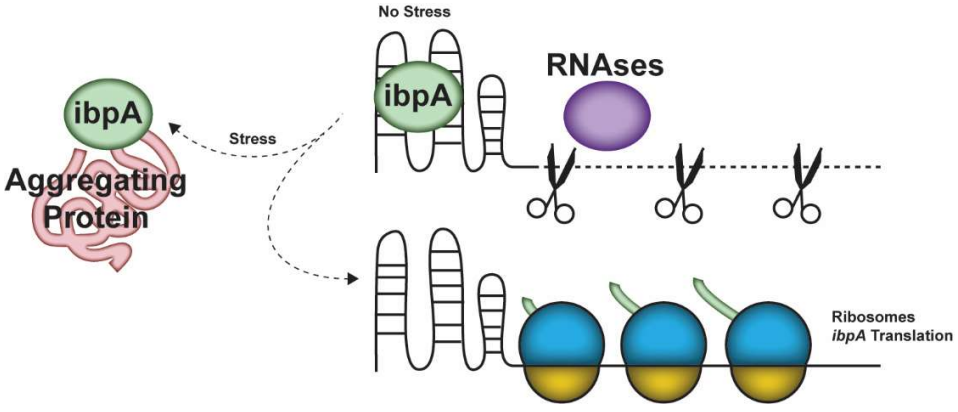


Figure 5. IbpA self-sequesters its own mRNA in a feedback loop. *In unstressed states IbpA binds its own mRNA and inhibits translation. Under stressed conditions IbpA is titrated away by aggregating proteins allowing translation initiation to occur and make IbpA. Figure inspired by* (69)

Translation elongation

Once translation initiation is complete and the second aa-tRNA is loaded in the A site translation elongation proceeds. tRNA incorporation is driven by the G-proteins EF-Tu and EF-G (see Figure 3). EF-Tu brings a charged tRNA into the A site of the ribosome (47,70). Upon correct codon-anticodon recognition GTP is hydrolysed to GDP and Pi as a result of proximity to the GTPase activating SRL in the 50S subunit (44). After GTP hydrolysis EF-Tu diffuses out of the ribosome and with the addition of a small rotation the aa-tRNA relaxes in the A site into its accommodated state (44,47). In the accommodated state both the P and A site amino acids are brought into close proximity. By stabilising intermediates the peptidyl transferase centre facilitates the condensation reaction, where through nucleophilic attack the nascent chain is extended by one, onto the A site tRNA (47,71). Subsequently small spontaneous tilts and rotations in the 30S subunit relative to the 50S subunit (about 3.7° and 11° respectively) promote

the formation of A/P and P/E tRNAs as the tRNAs tilt into the neighbouring sites (72). These rotations facilitate binding of EF-G-GTP in the A site. EF-G-GTP binds, and by its hydrolysis and a ratcheting movement of about 15° of the 30S head domain, and 5° rotation of the 30S subunit relative to the 50S subunit, the ribosome slides the mRNA forward 3 nucleotides (70,72). This results in the tRNAs shifting from the A and P sites into the P and E sites respectively. The E site tRNA then defuses out and the ribosome is ready for the next cycle.

Translation termination and Ribosome recycling

When a stop codon is encountered, either UAA, UAG or UGA, translation is terminated. UAA and UAG are recognised by RF1 and UAA and UGA are recognised by RF2 (73) (see Figure 3). Upon loading to the A site and correct position, a hydrolysis reaction releases the newly synthesised polypeptide (47). RF1 or RF2 dissociate from the ribosome spontaneously, but their turnover is promoted by RF3 which stimulates structural rearrangements by GTP hydrolysis, promoting release (74,75). RF3 releases RF1 or RF2. This is achieved by an induced rotation of the 30S subunit which causes RF1 (or RF2) to dislodge (76). The ribosome is recycled by EF-G with the ribosome recycling factor (RRF), which promotes dissociation of the 50S complex, which can rapidly re-associate. Once dissociated IF3 binds which inhibits re-association, and primes the ribosome for the next round of translation initiation (77).

Folding

Folding in the cytoplasm

One third of the *E. coli* proteome is predicted to start folding whilst still on the ribosome (78). This mostly occurs for multi-domain proteins, whereas single domain proteins typically fold off the ribosome. Secondary structures such as α-helixes can also start forming in the 15 Å space of the ribosome exit tunnel and small domains have been observed in the vestibule of the exit tunnel (79,80). Once the nascent chain emerges from the tunnel exit it becomes free to sample more conformations. The surface of the ribosome may limit the possible conformations in space, and the negative charge of the ribosome may have interactions with the nascent chain. The ribosome has been shown to stabilise disordered regions, and delay folding of domains while in close proximity (80).

Many proteins spontaneously fold *de-novo*, or refold after chemical denaturation with, for example, concentrated urea. However many require chaperones to achieve there correct final confirmations. For cytosolic proteins the chaperone network can be sequential. It includes 3 main players: Trigger factor (TF), DnaK (and its co-factors DnaJ/GrpE) and GroEL (and its co-factor GroES) (81) (see Figure 6). TF binds to nascent chains emerging from the tunnel exit (80,82,83). It has a PPIase activity and speeds up proline isomerisation. Approximately 70% of the cytosolic proteome folds spontaneously upon release from TF (84) (see Figure 6). The remainder may be clients of downstream chaperones. About 25% of cytosolic proteins were found to interact with DnaK when investigated by pull down experiments (83) (see Figure 6). These interacting proteins varied in enrichment factor, and the most enriched proteins where generally low abundance proteins and tended to be large, or hetro-oligomeric. In general the relative enrichment correlated with aggregation propensity. Most of the DnaK interactome was found to be newly synthesised proteins (but not exclusively), and many ribosomal proteins were represented (83). A number of chaperone and heat shock proteins where found to associate with DnaK including GroEL, SecB and IbpA suggesting possible hand-over events and a central role for DnaK in the chaperone network. Indeed by mass 30 % of DnaK clients were found to be

GroEL substrates (152 proteins) and 42 proteins that where dependent on GroEL for folding. Upon deletion of GroEL most of these proteins aggregated or were degraded (83). GroEL, together with GroES forms a ~800 kDa barrel, consisting of two heptomeric rings, in which substrates can fold. By trapping substrates in the barrel and identifying them by HRMS about 250 proteins were identified (see Figure 6), and a size limit of approximately 60 kDa could be established (85). 13 of these proteins are essential. Substrates were in general of lower net charge than average cytosolic proteins, and TIM barrel proteins (consisting of $(\alpha\beta)_8$) were specifically enriched.

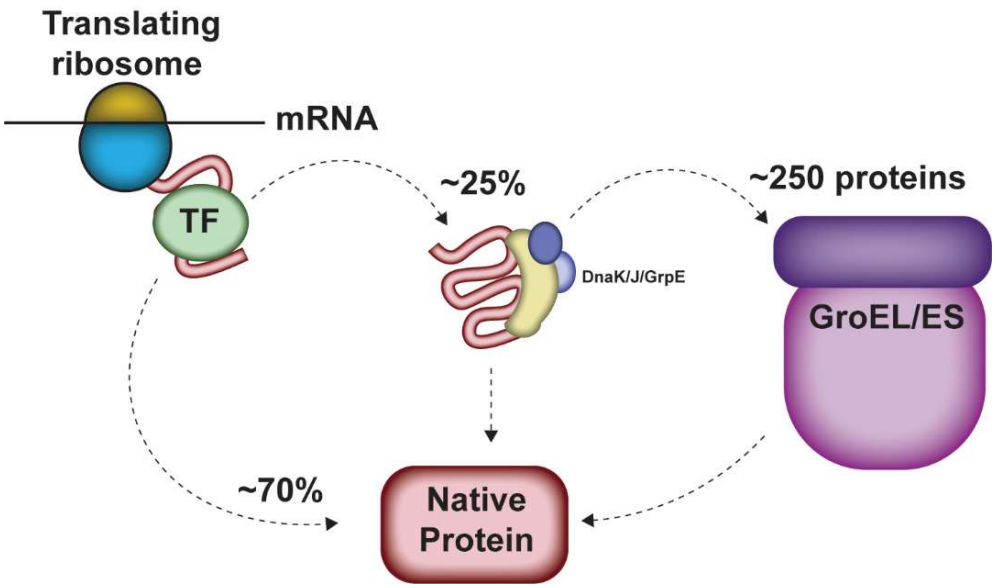


Figure 6. The cytoplasmic chaperone network. *Trigger factor (TF) binds proteins as they imerge from the exit tunnel. About 70% of cytoplasmic proteins fold off TF. About 25% of cytoplasmic proteins require DnaK/J/GrpE for folding. About 250 proteins fold with GroEL/ES, of which 13 are essential. Figure inspired by* (83).

Folding in the periplasm

Periplasmic proteins are translocated through the IM into the periplasm. This is facilitated for most periplasmic proteins via the SecYEG translocon. Proteins are translocated either co-, or post-translationally, and the predominant pathways are SRP or SecB dependent, respectively (86). Generally SRP dependent co-translation is the pathway used by IM proteins, and SecB dependent post-translation pathway is the pathway used by periplasmic and OM proteins (86). TF binds the nascent chain as they emerge from the tunnel exit. It has specificity for sites enriched with aromatic side chains and positive charges (84). SRP interacts with emerging nascent chains on the ribosome. Recruitment of SRP is thought to happen when a signal peptide is still in the exit tunnel (80). Experiments in the Müller lab showed that SecB binding to nascent chains is mediated by TF which binds to emerging nascent chains and inhibits SRP from binding (82). In these experiments, once dissociated from the ribosome, TF was replaced SecA/SecB. SRP brings nascent chain-ribosome complexes to the SRP receptor, FtsY, (SR) protein and shuttles them to the holotranslocon were they are inserted. For periplasmic proteins, SecB plays an important role in keeping the peptide in a translocation competent state (87). The force for translocation is provided by SecA, which in an ATP dependent manor pushes the client polypeptide through the translocon (88). By exerting mechanical force it can unfold structures to make them competent for translocation (89). For membrane proteins translocated via the SRP pathway, SRP docks the ribosome and on the Sec translocon and insertion is spontaneous (90), and transmembrane helixes insert into the membrane through the lateral gate of SecY (91). The signal peptidase 1, LepB, cleaves the signal peptide upon insertion (92).

Many proteins can fold spontaneously on a sequential pathway. As an example, MBP refolding has be studied *in vitro* using hydrogen-deuterium exchange-mass spectrometry (HX-MS). In HX-MS experiments proteins denatured with high concentrations of urea are allowed to refold by dilution with H_2O or D_2O. Subsequent pulse labelling requires exposure to D_2O or H_2O respectively, prior to quenching, digestion, and analysis by high-resolution mass spectrometry

(HRMS) (93,94). These experiments have indicated that folding starts at specific areas called foldons, where folding is nucleated. When MBP was studied these foldons where at a specific solvent exposed patches (surprisingly not the hydrophobic core) and occurred rapidly in about 7 s. The remaining domains folded on the 100 s timescales and displayed a variety of folding speeds, suggesting a sequential pathway (93).

The signal sequence disrupts and slows folding prior to export. HX-MS Studies on PpiB and PpiA found that folding happened in a sequential manor for both proteins, however with a different order. As PpiA is periplasmic and PpiB is cytoplasmic the effect of a signal peptide on protein folding could be examined. Here PpiA folding was dramatically retarded in a signal sequence dependant manor. Intriguingly PpiB folded slower, and in a different sequence when the PpiA signal sequence was transplanted onto its N-terminal (94).

Protein folding in the periplasm is facilitated by a number of chaperones. For example, the oxidation of cysteines into disulphide bridges is catalysed by DsbA/B (95) (Figure 7). Disulphide formation is often a rate limiting step in protein folding. In *E. coli* DsbA is the main catalyst of disulphide bond formation. A conserved motif CXXC catalyses the reaction. The cysteine's sulphur in the substrate is a nucleophile which attacks the electron accepting oxidised cysteine pair in DsbA, forming new mixed disulphide between DsbA and the substrate (95). Subsequently this disulphide bond is attacked by a free cysteine on the substrate resulting in a new disulphide bond between internal cysteines in the substrate. This results in the reduced form of CXXC in DsbA. DsbA is recycled by DsbB, a membrane protein capable of shuttling electrons to ubiquinol. Nucleophilic attack by cysteines of DsbA on a disulphide pair of cysteines on periplasmic loops of DsbB result in a DsbA-DsbB disulphide hybrid. Subsequent nucleophilic attack by the second free cysteine of DsbA in the hybrid pair results in recycled DsbA (oxidised CXXC). An internal disulphide pair of DsbB is able re-oxidise the periplasmic loop, and then re-oxidise itself by reduction of quinol (95). Some proteins require multiple disulphide bonds, or contain multiple cysteines and require isomerases to correct erroneous disulphide formation. This isomerisation is catalysed by the homodimer DsbC (95). Unstable disulphides form between the reduced cysteines in CXXC of DsbC and cysteines in the substrate. These hybrids are are subsequently attached by other cysteines in the substrate, resulting in an isomerisation (attack by a cysteine in DsbC results in reduction). To maintain DsbC in a reduced state electrons are shuttled from NADPH (by TrxA) to DsbC via DsbD (95) (See Figure 7).

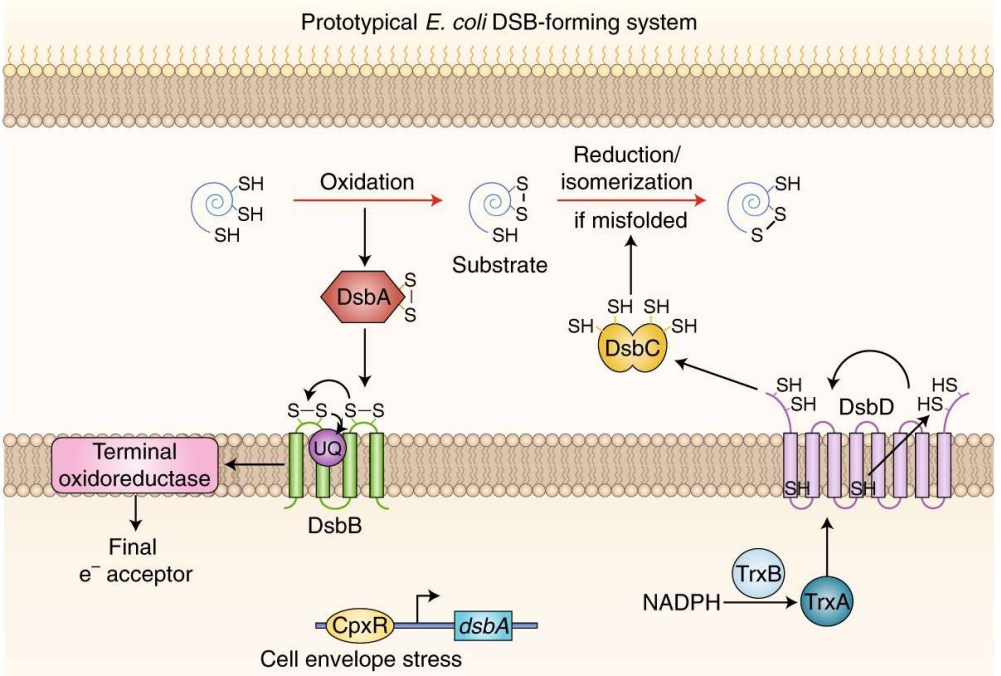


Figure 7. Disulphide bond formation in the periplasm. *DsbA catalyses oxidation and DsbC catalyses isomerisation. DsbB recycles DsbA and DsbD recycles DsbC. Figure taken from* (96) *with permission.*

Another potentially rate limiting step in folding in the periplasm is proline isomerisation (97). There are two well characterised PPIases in the periplasm: SurA and FkpA (97,98). The folding of a number of OMPs has been shown to be at least partly dependent on SurA. FkpA has a role in folding of periplasmic proteins and OMPs and has been shown to increase solubility of mutant *E. coli* proteins such as MalE31, and heterologous proteins such as antibody fragments (99,100). Other periplasmic chaperones include (but not limited to) Skp which forms a jellyfish like structure with holdase activity, DegP which as well as being a protease, acts as a chaperone, Spy, and the acid responsive HdeA/B (98). As well as folding in the periplasm, chaperones have an important function in the shuttling of OMPs across the periplasm (101).

Protein misfolding and aggregation

Protein misfolding and subsequent aggregation is problematic for a bacterium. If the aggregating protein forms unspecific mixed aggregates or aggregates with essential proteins, the result can be cytotoxic (102). The inducers of misfolding are often environmental conditions such as a change in temperature, pH or salt. The mechanisms of misfolding differ and for example; excess heat provides energy to overcome weak van der Waals forces or hydrophobic interactions that can lead to loss of native structure and the exposure of hydrophobic patches (102), while changes in pH cause the charge distribution over the surface of the protein to change, affecting its interaction with the bulk solvent and other proteins. Many of these conditions are taken advantage of in sample preparation, such as Salting out proteins, or TCA precipitations (103). Recombinant protein production places an additional stress on the host. Proteins derived from higher eukaryotes are often large and multi-domain with more complex folding pathways, and have evolved with slower translation speeds (80,104). Codon context differs between organisms, and a rare codon in a human cell for example, is not necessarily rare in *E. coli*. This can have deleterious effect on the folding outcome if slow regions of translations, that are evolved to give a domain time to fold, are no longer slow (104). Additionally loss of energy, or ATP, is associated with increased aggregation as the key chaperons GroEL/ES and DnaK/J/GrpE need ATP to perform their functions (102).

The result of misfolding or aggregating proteins can be amorphous aggregates driven by unspecific hydrophobic interactions, amyloid fibrils, driven β-interactions, or inclusion body (IB) formation, which can result from both hydrophobic and β-interactions (102). Studies of inclusion body formation using GFP as a model (due to its activity in aggregates) provided insights into their early stages of formation. Kopp and colleagues found that GFP formed small rod-shaped aggregates that distributed through the cytoplasm. As the amount of GFP expressed increased these agglomerated into globular IBs. The proposed mechanism of IB formation was concentration dependent interactions which eventually resulted in large aggregates of aggregates forming sponge like structures. These structures formed due to steric restrains of the cell (increased chance of collision) rather than by an active process (105). In *E. coli* IBs are typically located at the poles of the cell. This has been proposed to occur as a result from exclusion from the nucleoid rather than an active process, and differs in species with a different number of chromosomes (102).

Upon protein aggregation, cellular stress responses are upregulated. The specific stress response(s) depends of the nature of the aggregating protein, but the most represented responses are the heat-shock (σ32) response in the cytoplasm (102), and the Cpx and sigmaE responses in the periplasm (31,106,107). The first line of defence against protein aggregation in the cytoplasm are the small heat shock proteins IbpA/B (69,108) (see Figure 8). These proteins act as holdases, binding to, and stabilising inclusion bodies. A plethora of proteases and chaperones, including DnaK/J/GrpE are upregulated as they are titrated away from the sigma factor, allowing it to bind RNA polymerase and direct transcription (21). Subsequent binding of DnaJ recruits DnaK, which protects the aggregate from proteases and recruits ClpB (109) (Figure 8). ClpB is a disaggregase that threads polypeptides through its core to disaggregate them for re-folding or degradation. ClpB is a member of the AAA+ family of proteins and is made up of two stacked hexametric rings. ATP hydrolysis in a monomer triggers aromatic residues that associate with the aggregated peptide to rearrange exerting a pulling force. By adjacent subunits sequentially hydrolysing ATP and activating, a polypeptide can be pulled through the channel (109). The pulling force generated can reach up to 50 pN. To ensure that ClpB only pulls on aggregates and does not unfold soluble proteins, its action is strictly regulated by DnaK/J (109).

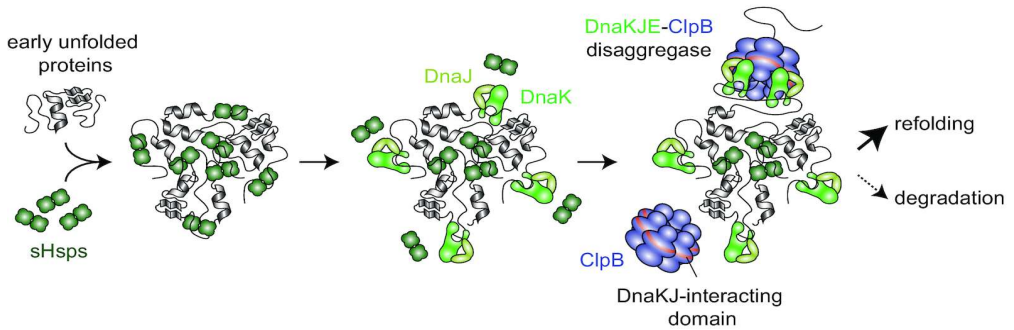


Figure 8. Cytoplasmic response to aggregation. *sHsps (such as IbpA/B) are initially recruited. Subsequently DnaK/J bind. The disaggregase, ClpB, is recruited and acts to pull polypeptides out of the aggregate. Figure taken from* (102) *with permission.*

When chaperones fail to meet the demand of protein folding, proteases help relieve the problem by destroying the misfolded proteins. Many proteases are under the control of σ32 promoters, and they are expressed under conditions

where this sigma factor is upregulated, for example, by aggregating proteins (110).

AAA+ proteases, like ClpB, are made up of hexametric AAA+ domains that use an ATP energised power stroke to pull peptides through a central channel. The AAA+ domain is coupled (sometime as a single polypeptide) to a protease domain that degrades the peptide. AAA+ proteases are present in all domains of life and are powerful degradation machines. The AAA+ domains recognise degron sequences which gives them specificity. Each protease will recognise specific degradation signals, which vary from peptide tags, to exposed hydrophobic or aromatic residues. In *E. coli* the Lon proteases, ClpXP, ClpAP, and the membrane bound FtsH play key roles in protein homeostasis and response to aggregation stress (111). Of these proteases, only FtsH is essential (110), and is important for regulation of the sigma factor, σ32 (21).

The OM and periplasm have evolved to deal with a greater number of environmental threats than the IM or cytoplasm. When the thermal stability of the *E. coli* proteome was examined, Mateus and colleagues found OM proteins to be the most thermally stable. The periplasmic and IM proteomes were the next most stable and the cytoplasmic proteome was found to be the least thermostable (112). Upon stress a large number of periplasmic chaperones and proteases are upregulated by Cpx, σE and other stress responses. Upregulated proteins include DegP, FkpA, Spy, SurA and DegS (32,113,114). Intriguingly, it was recently found that for aggregated uOmpX, both Skp and SurA could work together as disaggregases (115).

Recombinant protein production

E. coli as a production host

Bacteria have been used as a workhorse in recombinant protein production since the early days of molecular biology in the 70s (116). They benefit from easy culturing, fast growth, simple manipulations and a variety of available tools for cloning and expression (117). *Escherichia coli* (*E. coli*) is by far the most used and studied (118). As a gram negative bacteria that naturally resides in the intestines of host organisms, *E. coli* contains two compartments with differing physiochemical properties. Moreoever, it can grow in a variety of nutritional and environmental conditions and can have a doubling time of as little as 20 minutes (118).

A plethora of compatible expression plasmids exist for producing proteins in *E. coli*. Some plasmids are strain specific, such as the pET vectors, as they require the T7 RNA polymerase in the host strain. Other plasmids are universal, such as the pBAD vectors (although ΔaraD stains are preferable). Plasmids are typically constructed from genetic modules that have been identified in nature, such as an origin of replication, an antibiotic selection marker, and an inducible promoter 5' of the coding sequence of a gene of interest. In this section I will review the design of commonly used genetic models in plasmids, and describe how they control plasmid copy number, transcription and translation.

Plasmid copy number

The number of plasmids per cell, called the copy number, is determined by the origin of replication. For the commonly used pBR322 backbone the copy number is controlled by the priming RNA1, the antisense RNA2 and Rop which promotes RNA1 and RNA2 interaction (119). Most medium to high copy number plasmids are maintained through divisions by random distribution in the

cytoplasm (120). The probability of a daughter cell inheriting no plasmids is very small with a medium copy number vector and antibiotic selection.

Transcriptional control

There are several features that make a promoter desirable for recombinant protein production (RPP) in *E. coli*. These include response to an inducer, titratability over a useful range of inducer concentrations, low off rate, high maximal on rate, prolonged (temporal) activation, and orthogonality. Promoters from metabolite responsive operons have been used extensively. Of these the lactose operon has been the most used. The genes for lactose metabolism are encoded under the lacZYA (lac) promoter. The promoter is activated by the presence of lactose, which binds to the lactose inhibitor protein (LacI) and releases it from the $LacO_1$, $LacO_2$ and $LacO_3$ operator sites, freeing the promoter for transcription. In addition the RNA polymerase is recruited to the lac promoter by the catabolite repressor protein (CRP) which when bound to cAMP, acts as a transcription factor. Levels of cAMP production decrease with preferred carbons sources, and in the absence of glucose, are sufficiently high to promote CRP binding. Interestingly sugar metabolism in *E. coli* is regulated by cAMP levels (via CRP) with decreasing sensitivity with less preferred sugars (121,122).

Of the promoters used for RPP, the T7 promoter is by far the most frequently used. It is derived from the T7 bacteriophage and was cloned into the pBR322 vector in 1986 by Studier and Moffatt (123,124). The T7 promoter is recognised by its orthogonal T7 RNA polymerase which specifically transcribes the downstream coding sequences. The gene encoding the T7 RNA polymerase was cloned into the genome of BL21 under the control of the lac promoter and was therefore regulated by lactose (or the artificial homolog IPTG) (124) (Figure 9). When lactose or IPTG is added, the gene encoding the T7 RNA polymerase is expressed and the T7 RNA polymerase is produced. This powerful polymerase then transcribes the genes downstream of the T7 promoter (see Figure 9). To improve titratability and reduce the off rate, $lacO_1$, the binding site for the lac repressor protein, LacI, was cloned downstream of the T7 promoter on many of the pET based plasmids (truncating the last 4 nucleotides) (123). Subsequent developments to reduce the off rate and increase treatability included the co-expression of lysozyme, which inhibits the T7 RNA polymerase (125). The T7 promoter has been used in more publications than any other promoter system. It benefits from high on rates, and good orthogonality. It is however limited by leakiness (high off rates) and poor titratability (126,127).

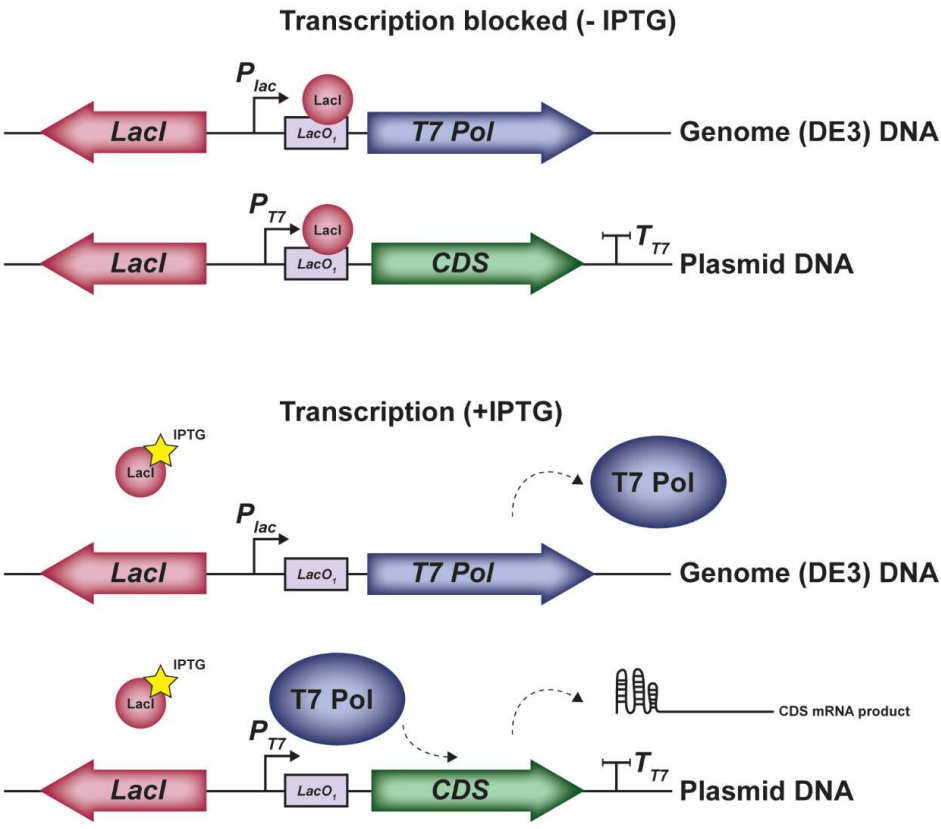


Figure 9. An overview of the regulation of T7 promoters. *In the absence of IPTG the LacI protein binds to the operator sequence and inhibits transcription of the T7 RNA polymerase (and the CDS). When IPTG is added, LacI no longer binds the operator, and the T7 polymerase is produced. The T7 polymerase then binds to, and transcribes genes cloned downstream from the T7 promoter.*

Another commonly used promoter is the pARA^{BAD} (pBAD) promoter. Like the lac promoter, the pBAD promoter is a native promoter to *E. coli,* which is located upstream of the genes encoding proteins for arabinose metabolism; AraB, A and D. It is regulated by AraC which forms a dimer and regulates transcription. AraC binds arabinose when it is present, and forms a compact dimer which binds at the two half sites upstream of pBAD, I_1 and I_2. Bound at I_1 and I_2 AraC

promotes transcription when cAMP associated CRP is also bound to its site adjacent to I_1. AraC at I_1 and I_2 together with CRP recruit the RNA polymerase to pBAD promoter (Figure 10). In the absence of arabinose the conformation of the AraC dimer changes and it binds at I_1 and O_2 in an extended conformation. In this conformation it is unable to promote transcription, and the resulting DNA curvature sterically hinders access of RNA polymerase (128) (Figure 10). The gene encoding araC, and the pBAD promoter where cloned into expression vectors including pBR322 in 1995 to create the first series of pBAD vectors (129). pBAD based expression systems have good titratability and low off rates, but do not produce yields as high as the T7 based expression systems (126,127).

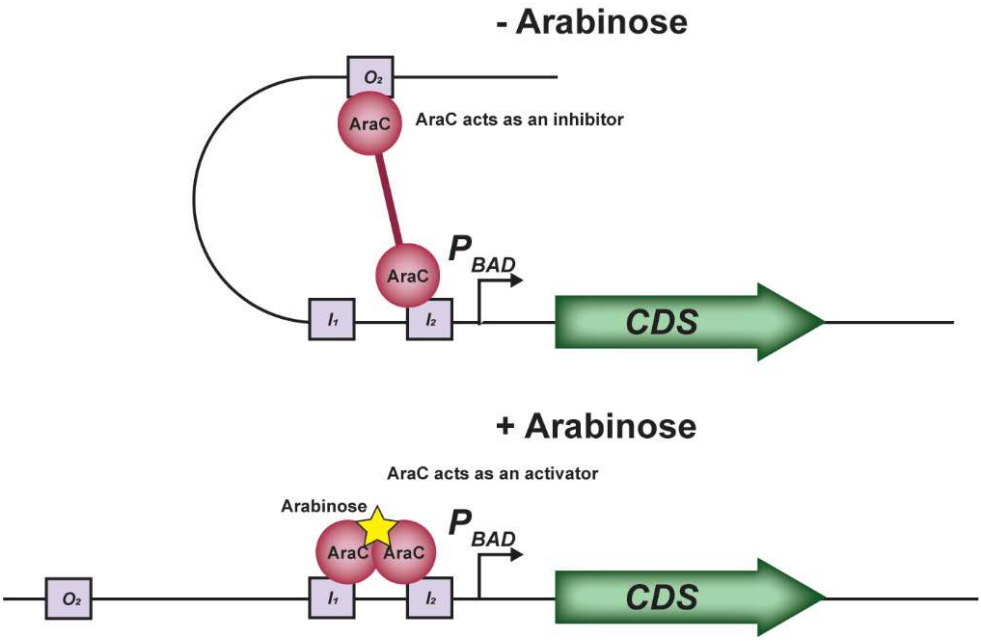


Figure 10. Overview of the regulation of the pBAD promoter. *In the absence of arabinose AraC binds to the I_2 and O_2 sites, inhibiting transcription. When arabinose is present it binds to AraC and acts as a transcriptional activator. Figure inspired by* (130).

Other commonly used promoters are the $pRha^{BAD}$ and pTetA promoters (127). The choice of promoter is often dictated by the desired experimental outcome. For example, for high level of production the T7 promoter may be the top choice, while for titratability the pBAD or RhaPBAD may be preferred.

It is often desirable to express 2 or more gene products simultaneously with different promoters which makes orthogonal systems extremely important. Work from the Voigt lab took metabolite sensitive promoter/ repressors, and by directed evolution reduced crosstalk to make each orthogonal to another and improve response (131). Future promoters might be derived from mutation of existing promoters in to new sequence space (as done with PhoQP) (132) or the uses of orthogonal sigma factors, and promoters from other bacterial species (133,134).

Translational control

The TIR

The sequence of the recombinant mRNA determines its structure, which in turn partly determines how efficiently it is translated. The TIR is often low in structure at the expense of optimal codons (TIRs are often enriched in rare codons at the start of the coding sequence (56)). Low structure in the TIR is one of the greatest determinants translation initiation rate.

In nature, a TIR co-evolves with a specific open reading frame and the ribosome of the host organism. However in recombinant plasmids most TIRs are the result of ad-hoc fusions between genetic elements and are thus often not optimal. For example, expression plasmids made in the 80s, 90, and 00s where constructed by restriction cloning, where the cut site had to be taken opportunistically and often left cloning scars. Importantly there was no design of TIRs, selection for more optimal expressers, and no evolution process to optimise mRNA structure. In recent years *in silico* tools have been developed that predict RNA structure in the TIR of genes of interest and have be used with to improve translation efficiency of genes (58) or even optimisation of metabolic pathways (135). Yet these predictors do not yet outperform experimental selection methods (58).

In papers I and II, this design flaw has been addressed in two of the most commonly used expression plasmid families, namely pET and pBAD. After correcting the 4 nucleotide truncation that resulted from cloning the $LacO_1$ site downstream of T7 promoter, named $T7p^{CONS}$ (which resulted in a threefold increase in expression of His-Thrombin-sfGFP), the 6 nucleotides upstream of the ATG

were randomised, and the 2 codons downstream of the ATG were either synonymously (TIR1) or fully randomised (TIR2) resulting 13 or 47 fold changes in fluorescence respectively. When $T7p^{CONS}$ and TIR1 where combined a 33 fold increase was observed, and when $T7p^{CONS}$ and TIR2 where combined a 121 fold increase was observed. When tested with His-Thrombin-MTH1, and His-Thrombin-Neil3 at least two fold increases were observed. In paper II, a similar synthetic evolution on the TIR of pBAD-HisB-sfGFP was able to select a TIR that increased the activation: repression ratio from 43 to 9025.

Synthetic Riboswitches

Riboswitches provide an efficient means of controlling translation rates in response to ligands. An important property of riboswitches for synthetic biology is orthogonality. The ligand should be specific for its RNA aptamer, and the aptamer should not bind to other ligands. In 1990 the first SELEX experiment was described where random RNA sequences where selected for ligand (Dye) binding (136). In the 2000s, using fluorescent nucleic acid aptamers and *in vitro* randomisation of sequences, aptamers that changed fluorescence upon binding of ATP (and other adenosine based substrates) was selected, indicating the first synthetic structure shifting RNAs (137) and DNAs (138). Engineering *in vitro* riboswitches has been successful, but a low number have been used *in vivo* (139,140), limited by *in vivo* environmental differences, cross-talk and stand-alone functions. Typically the aptamer must bind its ligand dynamically, with induced structural changes to make a good riboswitch. This is one of the limitations of such screening techniques such as SELEX. Synthetic aptamers have been used to crease riboswitches by insertion upstream of control regions such as RBSs. A good example of this was the generation of theophylline riboswitches by inserting the theophylline aptamer in the 5' UTR of β-galactosidase, randomising the RBS, and selecting theophylline dependent expressers. Using a flow-cytometry screen (and mutating the TIR) highly responsive riboswitches were identified, capable of 96 fold activation (141). To address the problem of a the limited number of orthogonal riboswitch-ligand pairs, Dixon and colleagues re-engineered the natural *add* riboswitch (activated by adenine) by rational design, specifically mutating nucleotides in the substrate binding pocket and stabilising a hairpin (142). By placing the riboswitch upstream of a chloramphenicol resistance gene 80 potential substrates were screened. Riboswitches M6" and M6C" where selected and showed good dynamic range and induction ratios for Amm and AzC respectively. A further screen with the same riboswitches identified new substrates PPDA and PPAO for M6" and M6C" respectively with improved induction properties (143). Using the M6"

riboswitch between the lac promoter and the T7 RNA polymerase on the genome, and between the T7 promoter and the gene of interest on the expression plasmid, the expression levels could be very tightly regulated and controlled (144). Using the same system, but with additional regulatory elements on the genome (O_3 and flipped $LacI^Q$) the expression and the secretion of an antibody fragment could be tightly coupled, resulting in an increased, folded, periplasmic population (145).

Synthetic sRNAs

Another method for regulating translation is by using sRNAs, typically binding to the RBS of a gene of interest. They usually function by repressing ribosome binding, or promoting mRNA degradation. Synthetic sRNA have been widely used in metabolic engineering where they can be used to modulate many genes simultaneously. A synthetic sRNA requires 3 domains typically; a U rich Hfq binding domain, a transcriptional terminator, and a region complementary to the target sequence (146). sRNAs pair with unperfected base pairings, and a challenge is to ensure that the sRNA is specific to its target (147). Many existing scaffolds have been used to create synthetic sRNAs such as MicC.

Metabolic burden

Upon induction of recombinant protein production, a decrease in growth is often observed. This is a consequence of metabolic burden, a multifaceted phenomenon that can include plasmid replication, transcription, translation, folding, misfolding, and resource allocation. Metabolic burden is different with different recombinant proteins. In one study, independent component analysis of transcriptomics data in response to the expression of 45 diverse recombinant proteins showed large variability in cellular responses (148). Some of the conclusions the authors could offer where that there was a trade-off between the stress response (rpoS iModulon) and growth (translational iModulon), and the activation of heat shock genes (rpoH iModulon).

Whether metabolic load originates mainly from the recombinant protein or the recombinant mRNA is unclear, and contradicting work would suggest that it can be both. A recent study by Li and co-workers showed transcription was the main factor affecting growth inhibition when hFGF-2 was produced in BL21(*DE3*) cells (149). However, early studies of metabolic burden pointed to

the recombinant protein as the main source (150). This is in line with a recent study from the Doyle lab who observed a massive drop in CFUs when sfGFP or MdfA where over expressed in BL21(*DE3*) cells with IPTG concentrations above 0.2 mM but not when IPTG concentrations where below 0.1 mM. In this study, colonies on plates could be described as large white, large and pale green, or small and variable. When the large white and large pale green colonies where sequenced it was found that they contain mutations in the T7 RNA polymerase gene, suggesting that bacteria who deactivated the T7 RNA polymerase had a survival advantage and could outgrow those with an intact T7 RNA polymerase (151). Previous studies also observed that the proportion of plasmid containing cells decreased rapidly after induction by almost three orders of magnitude (152). This growth of plasmid free cells was probably only possible because the antibiotic used was ampicillin which is degraded rapidly resulting in no selection pressure for cells with a plasmid (153) allowing them to rapidly overtake the culture (see Figure 11).

Small reductions in metabolic load can have significant benefits. Construction of minimum-antibiotic expression plasmids, by mutation of the kanamycin resistance gene promoter, resulted in an increased capacity to produce some proteins (154). This may be due to the reduced metabolic load put on a bacteria by the lower expression of an antibiotic resistance gene (as kanamycin is not rapidly degraded). Monitoring and reducing small changes in metabolic load has helped increase yields of recombinant products. The Ellis lab was able to optimise the expression of an operon by monitoring decreases in fluorescence from a GFP gene under a σ70 dependent promoter. The capacity monitor reflected resource availability and was more sensitive than monitoring growth (155). This suggests that lowering the metabolic load by, for example, reducing expression of an antibiotic resistance gene, could result in increased performance of microbial cell factories.

In paper III we report that reducing levels of the TEM1 β-lactamase reduced metabolic load and still permitted selection on ampicillin. As a result antibiotics lasted longer cultures and the proportion of cells harbouring the plasmid increased.

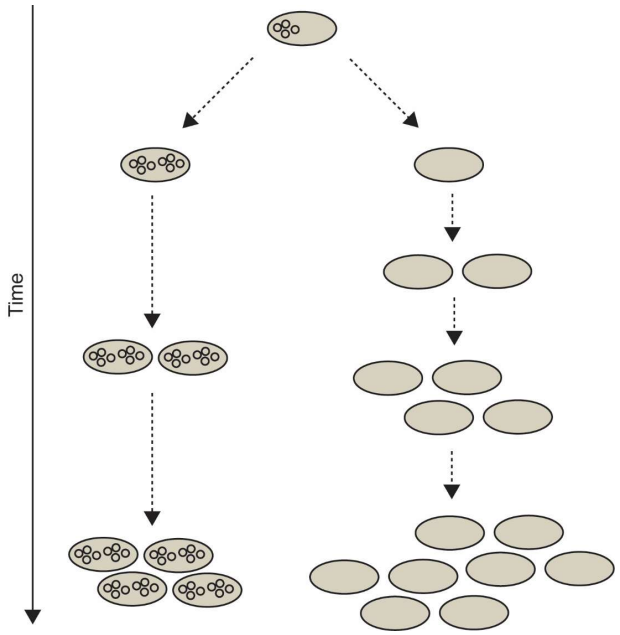


Figure 11. Plasmid free cells quickly overgrow the culture in the absence of antibiotic selection. *If a daughter cell does not inherit a plasmid, the doubling time to will be reduced and the plasmid free cells will eventually overtake the culture (if there is enough time in growth phase)*

Biosensors

Organism have evolved to sense and respond to environmental stimuli, such as nutrient availability or stress. The response almost always results in a change of gene expression at the transcriptional or translational level. Biosensors for microbial cell factories can take advantage of this, by reprograming the response into a desirable OUTPUT, typically an optical signal such as a fluorescent protein or an enzymatic activity such as β-galactosidase. For example, a library of all the annotated promoters of *E. coli* were cloned upstream of the coding sequence for the unstable GFPmut2 in the low copy number plasmids pUA66 or pUA139. This resulted in the construction of reporters for 95% of annotated promoters with a predicted accuracy of 99% (5% of the plasmids sequenced contained errors and were removed) (156). This biosensor library is a valuable resource and was used by the authors to investigate expression profiles in diauxic shift.

In essence any inducible promoter is a biosensor, for example the presence of arabinose is detected by the pBAD promoter by modulating the repressor activator activity of AraC in a concentration dependent manor (in the absence or glucose or lactose). Promoter based biosensors also have the potential to be expanded into new sequence space in order to respond to other small molecules. An example of this is the re-engineering of the PcaV operator, by targeted mutations in the substrate binding site, to be responsive to vanillin (and other aromatic aldehydes) rather than its original substrate PCA (157).

For recombinant protein production purposes it is valuable to be able to monitor the homeostasis of the cell. Zutz and colleagues created a system for simultaneously monitoring protein production and protein folding (158). By translationally coupling the gene of interest to mCherry the rate of protein production could be followed. Folding was monitored by the expression of the destabilised GFP_{ASV} under the control of the ibpA promoter. Here it was found that gene dosage was important, and only when the authors used a medium copy number plasmid (as opposed to a high copy number plasmid), was a clear difference in GFP_{ASV} expression distinguishable.

In characterising the periplasmic inclusion body forming MalE31, Hunke and Betton used biosensors consisting of lacZ downstream of the promoters for DegP and CpxP to identify stresses activated by this periplasmic protein (159). They were able to show that DegP was essential for the clearance of MalE31 at 42°C and that it was the Cpx response, and not the σE response that sensed periplasmic MalE31 and other misfolding variants. Recently Guidi and colleagues created a biosensor consisting of the CpxP promoter upstream of GFP in the process of developing a feedback loop for membrane insertion. The cpxP promoter was used to drive the expression of a CpxQ sRNA where the seed region was been mutated to specifically bind the 5'UTR of the gene of interest (160).

In paper IV, we have used the ibpA promoter upstream of the coding sequence for GFP_{ASV} and the cpxP promoter upstream of the coding sequence for $mCherry_{ASV}$ to make a duel reporter for optimising periplasmic protein production. Upon stress caused by protein export the ibpA promoter is activated (161) and a green fluorescence signal can be detected. Upon stress caused by proteins misfolding in the periplasm the cpxP promoter is activated and a cherry signal can be detected. This dual biosensor allows the user to optimise induction conditions

to stay below the stress threshold. We show that the yields of soluble $scFv^{HER2}$ increase by a aproximately 40% using the biosensor.

Short summary of papers

Paper I.
Improved designs for pET expression plasmids increase protein production yield in Escherichia coli.

The pET expression plasmids remain the most popular choice for producing recombinant proteins. Of the pET vectors, pET28a is the most commonly used. The plasmid contains a LacI inhibiter protein, T7 promoter, lacO operator, an N-terminal 6-his tag and thrombin cleave site upstream of a MCS and a T7 terminator, as well as an APH kanamycin resistance gene, and pBR322 origin of replication. The expression levels obtained with pET28a are generally good, but are limited by the truncatation of the T7 promoter by 4 nucleotides, and an unoptimised TIR region formed by ad-hoc fusion of genetic parts.

To address the truncation of the T7 promoter, the 4 nucleotides where re-introduced by PCR to restore the full length T7 promoter (T7pCONS). When this was tested by expressing GFP with 0.5mM IPTG a three-fold increase in fluorescence was observed in BL21(*DE3*)pLysS cells.

To address the sub-optimal TIR a synthetic evolution experiment was performed. Using PCR, a library of constructs was created with 6 completely randomised nucleotides upstream of the start ATG and 2 completely, or synonymously randomised codons downstream of the ATG. The library was created with the coding sequence for sfGFP translationally coupled to β-lactamase. Top performing TIRs from the randomised and synonymous pools where selected by screening for survivors on high concentrations of ampicillin. Two TIRs where selected; TIR1 from the synonymous pool increased sfGFP expression by 13-fold and TIR2 form the randomised pool increased sfGFP expression by 47-fold.

When T7pCONS and TIR2 where combined (pET28 TIR2 T7pCONS) a 121-fold increase in sfGFP fluorescence was observed. Increases were also observed, albeit more modestly, when MTH1 and Neil3 where expressed. To investigate if these improvements affected solubility sfGFP, MTH1 and Neil3 were expressed from either pET28a or pET28 TIR2 T7pCONS and samples where fractionated into soluble and insoluble and analysed by SDS-PAGE and Western blotting. In all cases the solubility profiles remained the same but the yields increased.

Paper II.
Signal amplification of araC pBAD using a standardized translation initiation region.

The $pARA^{BAD}$ (pBAD) promoter and inhibitor/activator protein AraC make the arabinose sensitive genetic sensor. The pBAD promoter has been widely used for the production of recombinant proteins, and in synthetic biology, as it has a very low level of background expression and is titratable. Having successfully improved production yields by synthetically evolving the TIR of pET28a (Paper I), a similar approach was used to improve production yields from pBAD plasmid .

Using pBAD / HisB-sfGFP, 6 nucleotides upstream of the ATG and 2 codons downstream of the ATG in the TIR region were randomised by PCR. The library was transformed into MC1061 cells and plated onto agar plates containing arabinose. 50 bright colonies where chosen and further screened in liquid culture. The top performing clones where sequenced. The top 2 clones; TIR^{22} (later renamed to TIR^{EVOL}) and TIR^{23} where compared to the original TIR^{STD} with arabinose titrations producing GFP. Fluorescence was improved by 13.5 fold from TIR^{STD} with TIR^{EVOL} and the activation:repression ratios increased to 9025 from 43. A similar improvement was observed for a further 3 fluorescent proteins with no sequence homology.

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

The TEM-1 β-Lactamase catalyses the breakdown of β-Lactam antibiotics in the periplasm of gram negative bacteria such as *E. coli*. It is secreted to the periplasm where it catalyses the hydrolysis of the antibiotics. It is encoded by the *bla* gene in the Tn3 fragment, found in *pBR322* based plasmids as well as

many others used for recombinant protein production. It is known that TEM-1 β-Lactamase can rapidly degrade ampicillin in cultures leaving the bacteria with no selection for those that contain a plasmid. This is exacerbated by the leaking of TEM-1 β-Lactamases into the extracellular media. As a result plasmid free cells may arise and overgrow the culture.

To address these problems we performed a synthetic evolution experiment on the *bla* gene within the Tn3.1 fragment of pET15b, a commonly used expression vector. A library was created where 6 nucleotides upstream of the ATG of *bla* in Tn3.1 were completely randomised and 2 codons downstream were synonymously randomised. The resulting library was transformed into BL21(*DE3*) cells. From the transformants 48 colonies where selected and grown on increasing concentrations of ampicillin. Colonies that could only grow on low concentrations of ampicillin where selected and sequenced. The mutation giving the lowest levels of resistance was named Tn3.1^{MIN} and was further characterised. BL21(*DE3*) cells containing pET15b (Tn3.1) had a MIC of approximately 800 μg/mL of ampicillin, but using pET15b (Tn3.1^{MIN}) this was reduced to approximately 30 μg/mL of ampicillin. The decay of ampicillin in cultures was investigated using a semi-quantitative mass spectrometry approach. It was found that the concentration of ampicillin remained above 1 μg/mL (the concentration that was needed for selection against cells without a plasmid) for just 70 minutes when using pET15 (Tn3.1). This increased to approximately 180 minutes when using pET15b (Tn3.1^{MIN}) after a 1: 100 back dilution.

The proportion of cells that had a plasmid after culturing was also assessed. When no inducer (IPTG) was added, the plasmid was well maintained. However, when recombinant protein production was induced with IPTG for 8 hours, about 50% of BL21(*DE3*) had the pET15 sfgfp (Tn3.1) plasmid. When Tn3.1^{MIN} was used, no plasmid loss was observed after 8 hours of culturing. Interestingly this was irrespective of presence of ampicillin.

Finally the Tn3.1^{MIN} cassette was introduced into 9 plasmids with diverse origins of replications: pSEVA111 to 191. In all cases the resistance to ampicillin was dramatically decreased. Only the low copy number plasmids pSEVA15^{MIN}1 and pSEVA17^{MIN}1 could not survive a minimum concentration of ampicillin for selection suggesting the Tn3.1^{MIN} is most applicable for medium and high copy number vectors.

Paper IV.
A genetic sensor for the production of proteins in the periplasm of Escherichia coli.

Several important protein products such as hormones or antibody fragments can only be produced in the periplasm as they require disulphide bonds. Producing proteins in the periplasm presents two major limitations. Firstly, should the titers of the recombinant protein exceed the capacity of the Sec translocon, a heat shock response will occur and cell growth and health may suffer. Secondly the recombinant protein may exceed the capacity of the periplasmic chaperones and misfold, aggregate or require proteases for its removal. A strategy to circumvent this problem is to adjust the inducer concentration and produce less of the recombinant protein. However to determine suitable production conditions, one must understand the fate of their protein. This is traditionally done by cell lysis, fractionation, SDS-PAGE, and Coomassie staining, or Western blotting. Thus determining protein fate is time and resource expensive. To address this problem, we built a sensor that could monitor both protein export and protein aggregation in the periplasm.

To screen for suitable responsive promoters three variants of MalE (MBP) where used that misfolded to varying extents. The reporter [Pr^{IbpA}-gfp_{ASV}] from the heat-shock response could detect an export defect. To find a promoter that was responsive to misfolding periplasmic proteins several promoters where selected from literature. The selected promoters covered many of the bacteria's periplasmic stress responses, however only [Pr^{cpxP}-gfp_{ASV}] from the Cpx response was activated. The GFP was replaced with mCherry and a TIR optimisation was performed to identify reporters with a better translational response, resulting in [Pr^{cpxP}-$mcherry_{ASV}$]^{OPT}. [Pr^{IbpA}-gfp_{ASV}] and [Pr^{cpxP}-$mcherry_{ASV}$]^{OPT} where cloned into the same plasmid which we call pQC (plasmid for Quality Control). pQC could detect aggregation events at high induction titers, but failed to detect misfolding at low titers, suggesting that it was responsive to stress caused by aggregation in the periplasm rather than aggregation itself.

pQC was used to monitor the production of two proteins; the human growth hormone (hGH), and the single chain antibody fragment, $scFv^{HER2}$. Tests where set up in small, 5mL cultures and induced with a range of L-arabinose concentrations. After 3 hours of induction fluorescence was measured and the resulting stress fingerprints where used to determine optimal and sub optimal culture conditions. Large cultures where subsequently set up, induced with optimal and sub optimal concentrations of L-arabinose. After 3 hours and 20 hours cells where

isolated from the cultures, lysed, separated into soluble and insoluble fractions and analysed by SDS-PAGE and Western blotting. After 3 hours of induction the highest (sub optimal) concentration of inducer resulted in the highest yields of protein, both total and soluble. However after 20 hours of culture the optimal concentrations resulted in 2 to 20 fold increases in yield of soluble protein.

Future perspectives

In papers I, II and III, design flaws in commonly used genetic modules where identified and corrected. In each case this resulted in improvements to cell factory performance. Many genetic modules exist for microbial cell factories but many are still sub-optimally designed or not performing to their full potential. It is therefore important to address the questions: what genetic modules are sub-optimally designed and not performing to their full potential? And how can these modules be improved?

An example from paper III is improving plasmid performance (antibiotic efficiency and plasmid stability) by optimising the expression of the resistance cassette. By optimising the expression of β-lactamase both the stability of the expression plasmid and longevity of the antibiotic increased. The observed improvement in plasmid stability when using $Tn3.1^{MIN}$ was irrespective of antibiotic presence. This would suggest that the differences in stability are a result of metabolic load. β-lactamases are just one class of commonly used selection markers. Other examples include the amino glycosylase (APH), commonly used to to select on kanamycin, and chloramphenicol acyl transferase (CAT), commonly used to select on chloramphenicol. By decreasing the expression of these genes, it seems reasonable to speculate that similar benefits may be observed. Decreases in metabolic load may free up expression potential and allow for increased protein titters and improved plasmid stability. In addition, the amount of antibiotics used maybe brought down, and plasmid stability improved. A similar experimental approach to that used in paper III could be applied.

In paper I and II a design flaw in the T7 promoter was identified and corrected and the TIRs of two commonly used expression plasmids where evolved. Less attention however has been put into the regulatory proteins such as LacI. By synthetically evolving the TIR of this protein, it is possible that its translation rate could be increased or decreased, leading to tighter regulation or more leakiness. The pET vectors have high basal expression rates which can cause issues prior to induction. By increasing expression of the regular protein, the leakiness

may be reduced. TIR evolutions would have advantages over promoter mutations (such as $lacI^{Q}$) as improvements would result in no additional transcriptional load, and no changes in specificity would occur.

Another important component of an expression cassette that does not receive it due attention are transcriptional terminators. Terminators are important for insulating genetic modules and to inhibit cross talk with adjacent genes. Inefficient terminators lead to expression of downstream genes by RNA polymerase read through, cross talk at the RNA or protein level, and an increased metabolic load from additional mRNA (or even polypeptide) synthesis. For a robust biosensor, or component of a circuit of metabolic pathway, insulation from surrounding metabolic noise is important. For example, a biosensor should not activate when an upstream gene is transcribed.

Rho independent terminators are simplistically composed of a RNA hairpin followed by a poly U tract. The formation of the hairpin is important for termination efficiency. Due to the structure forming nature of mRNA, the sequence of mRNA upstream of the hairpin is important. Should the hairpin not form properly as it is sequestered into an alternative structure then termination efficiency would decrease. Terminators used in synthetic biology have been incorporated ad hoc downstream of coding sequences where there sequence context has not had any optimisation or adaption. Adapting terminators to their sequence context would involve ensuring there is no undesired interaction with the hairpin from adjacent sequence. Improving termination efficiency would be particularly useful when designing adjacent synthetic operons, and circuits. Or to improve termination after strong promoters, such as the T7. Current solutions to the T7 terminator read-through involve multiple tandem terminators which are bulky and can make cloning and synthesis complicated.

Could the sequence upstream of a terminator effect hairpin formation and ultimately termination efficiency? Synthetic evolution of sequence regions could be applied to address this question. By creating libraries of random sequence 5' of the terminators, improved termination efficiency could be screened for and characterised.

In paper IV, a sensor was created for periplasmic protein production using two biosensors of cellular stress. This was applied to monitor how stress differed at different induction conditions when expressing an antibody fragment and a hormone. The resulting stress fingerprint was then used to direct production of the

recombinant proteins, resulting in increased yield of soluble product. This work demonstrates the usefulness of biosensors in the design of production experiments. The application of biosensors in optimising recombinant protein production has not been applied to its full potential. A greater understanding of biosensors for recombinant protein production is needed to realise how they could be best applied.

Could pQC be improved? The cpxP promoter (Pr^{cpxP}) on pQC was found by searching literature and screening a small number of promoters. There may however be promoters that are more responsive to periplasmic stress or aggregation of recombinant proteins than Pr^{cpxP}. A genome wide screen to assess the promoters most activated in response to stresses (such as aggregation) caused by recombinant proteins could potentially enable the discovery of improved biosensors.

Biosensors could also be applied to understand the roles of stress and capacity in metabolic load. Stress involves the response to a stimuli, such as aggregation, while capacity involves the saturation of transcriptional or translational machinery. Both these events many result in decreased growth rates, and reduced fitness as a result of recombinant protein production.

Monitoring stress and capacity using biosensors could enable metabolic load to be characterised as a cellular response to a stress where the cell slows its own growth rate, or as a consequence of reduced capacity which is needed for recombinant protein production. To deconvolute these two phenomena, growth phenotypes could be correlated with signal from stress and capacity biosensors. Stress monitoring could be archived by using such plasmids as pQC and capacity monitored with the capacity sensor developed in the Ellis lab.

Understanding metabolic load in more detail could enable predictive strategies to reduce it. It could also help understand the observation in paper III of increased plasmid stability with Tn3.1MIN when compared to Tn3.1 containing vectors.

Another potential application of the biosensors on pQC could be to create simple negative feedback loops to self-regulate recombinant protein production. For example, sRNAs or guide RNAs (gRNAs) and dCAS based feedback loops. A sRNA auto regulating system has been described using the cpxP promoter and a modified CpxQ sRNA (160). The authors synthetically evolved the sRNA

to sequester the RBS of the coding sequence. Using this system, yields of several membrane proteins where increased. In paper IV, we observed that with the proteins tested, the ibpA based secretion sensor was activated along side the cpxP periplasmic aggregation sensor. By building a feedback loop based on both ibpA and cpxP promoters, stress from secretion defects, membrane proteins, and/or periplasmic aggregation could potentially all be avoided.

In paper IV biosensors were used to monitor stress of secretion and aggregation of periplasmic proteins. However biosensors have been, and could also be applied to monitor many other cellular processes in recombinant protein production. For example, translational fusions would allow the expression of a protein to be coupled to a fluorescence signal and to monitor production titers. Protein fusions with fluorescent proteins can be used to monitor correct membrane insertion and topology. And split reporter proteins could be used to study dimerization or oligomerization. By combining biosensors for different cellular events it may be possible to create a holistic view of a recombinant protein production experiments with simple fluorescence outputs.

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. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, et al. Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. Mol Cell. 2002 Mar;9(3):527–39.

5. Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, et al. Crystal Structure of Escherichia coli σE with the Cytoplasmic Domain of Its Anti-σ RseA. Mol Cell. 2003 Apr;11(4):1067–78.

6. Burgess RR, Travers AA, Dunn JJ, Bautz EKF. Factor Stimulating Transcription by RNA Polymerase. Nature. 1969 Jan;221(5175):43–6.

7. Sharma UK, Chatterji D. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of σ^{70} activity. FEMS Microbiol Rev. 2010 Sep;34(5):646–57.

8. Mauri M, Klumpp S. A Model for Sigma Factor Competition in Bacterial Cells. Rao CV, editor. PLoS Comput Biol. 2014 Oct 9;10(10):e1003845.

9. Maeda H. Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res. 2000 Sep 15;28(18):3497–503.

10. Jishage M, Kvint K, Shingler V, Nyström T. Regulation of sigma factor competition by the alarmone ppGpp. Genes Dev. 2002 May 15;16(10):1260–70.

11. Magnusson LU, Farewell A, Nyström T. ppGpp: a global regulator in Escherichia coli. Trends Microbiol. 2005 May;13(5):236–42.

12. Åberg A, Fernández-Vázquez J, Cabrer-Panes JD, Sánchez A, Balsalobre C. Similar and Divergent Effects of ppGpp and DksA Deficiencies on Transcription in *Escherichia coli*. J Bacteriol. 2009 May 15;191(10):3226–36.

13. Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. Proc Natl Acad Sci USA. 2019 Apr 23;116(17):8310–9.

14. Patikoglou GA, Westblade LF, Campbell EA, Lamour V, Lane WJ, Darst SA. Crystal Structure of the Escherichia coli Regulator of σ70, Rsd, in Complex with σ70 Domain 4. J Mol Biol. 2007 Sep;372(3):649–59.

15. Jishage M, Ishihama A. Transcriptional Organization and In Vivo Role of the *Escherichia coli rsd* Gene, Encoding the Regulator of RNA Polymerase Sigma D. J Bacteriol. 1999 Jun 15;181(12):3768–76.

16. Piper SE, Mitchell JE, Lee DJ, Busby SJW. A global view of Escherichia coli Rsd protein and its interactions. Mol BioSyst. 2009;5(12):1943.

17. Treviño-Quintanilla L, Freyre-González J, Martínez-Flores I. Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. CG. 2013 Sep 1;14(6):378–87.

18. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. Nucleic Acids Res. 2019 Jan 8;47(D1):D212–20.

19. Erickson JW, Vaughn V, Walter WA, Neidhardt FC, Gross CA. Regulation of the promoters and transcripts of rpoH, the Escherichia coli heat shock regulatory gene. Genes Dev. 1987 Jul;1(5):419–32.

20. Roncarati D, Scarlato V. Regulation of heat-shock genes in bacteria: from signal sensing to gene expression output. FEMS Microbiology Reviews. 2017 Jul 1;41(4):549–74.

21. 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.

22. Herman C, Thévenet D, D'Ari R, Bouloc P. Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. Proc Natl Acad Sci USA. 1995 Apr 11;92(8):3516–20.

23. Yura T. Regulation of the heat shock response in *Escherichia coli*: history and perspectives. Genes Genet Syst. 2019 Jun 1;94(3):103–8.

24. Seshasayee ASN, Sivaraman K, Luscombe NM. An Overview of Prokaryotic Transcription Factors: A Summary of Function and Occurrence in Bacterial Genomes. In: Hughes TR, editor. A Handbook of Transcription Factors. Dordrecht: Springer Netherlands; 2011. p. 7–23. (Subcellular Biochemistry; vol. 52).

25. Cosma CL, Danese PN, Carlson JH, Silhavy TJ, Snyder WB. Mutational activation of the Cpx signal transduction pathway of Escherichia coli suppresses the toxicity conferred by certain envelope-associated stresses. Mol Microbiol. 1995 Nov;18(3):491–505.

26. Raivio TL. Everything old is new again: An update on current research on the Cpx envelope stress response. Biochim Biophys Acta Mol Cell Res. 2014 Aug;1843(8):1529–41.

27. Vogt SL, Raivio TL. Just scratching the surface: an expanding view of the Cpx envelope stress response. FEMS Microbiol Lett. 2012 Jan;326(1):2–11.

28. Guest RL, Wang J, Wong JL, Raivio TL. A Bacterial Stress Response Regulates Respiratory Protein Complexes To Control Envelope Stress Adaptation. DiRita VJ, editor. J Bacteriol. 2017 Oct 15;199(20).

29. Chao Y, Vogel J. A 3′ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. Mol Cell. 2016 Feb;61(3):352–63.

30. Vogt SL, Evans AD, Guest RL, Raivio TL. The Cpx Envelope Stress Response Regulates and Is Regulated by Small Noncoding RNAs. J Bacteriol. 2014 Dec 15;196(24):4229–38.

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

32. Price NL, Raivio TL. Characterization of the Cpx Regulon in *Escherichia coli* Strain MC4100. J Bacteriol. 2009 Mar 15;191(6):1798–815.

33. Mileykovskaya E, Dowhan W. The Cpx two-component signal transduction pathway is activated in Escherichia coli mutant strains lacking phosphatidylethanolamine. J Bacteriol. 1997 Feb;179(4):1029–34.

34. 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.

35. Otto K, Silhavy TJ. Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. Proc Natl Acad Sci USA. 2002 Feb 19;99(4):2287–92.

36. 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.

37. Yang X, Lewis PJ. The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. Transcription. 2010 Sep;1(2):66–9.

38. Peters JM, Vangeloff AD, Landick R. Bacterial Transcription Terminators: The RNA 3′-End Chronicles. J Mol Biol. 2011 Oct;412(5):793–813.

39. Gualerzi CO, Pon CL. Initiation of mRNA translation in bacteria: structural and dynamic aspects. Cell Mol Life Sci. 2015 Nov;72(22):4341–67.

40. 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.

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

42. Milón P, Maracci C, Filonava L, Gualerzi CO, Rodnina MV. Real-time assembly landscape of bacterial 30S translation initiation complex. Nat Struct Mol Biol. 2012 Jun;19(6):609–15.

43. Antoun A, Pavlov MY, Lovmar M, Ehrenberg M. How Initiation Factors Maximize the Accuracy of tRNA Selection in Initiation of Bacterial Protein Synthesis. Mol Cell. 2006 Jul;23(2):183–93.

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

45. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science. 2009 Apr 10;324(5924):218–23.

46. Jacques N, Dreyfus M. Translation initiation in Escherichia coli: old and new questions. Mol Microbiol. 1990 Jul;4(7):1063–7.

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

48. Reeve B, Hargest T, Gilbert C, Ellis T. Predicting Translation Initiation Rates for Designing Synthetic Biology. Front Bioeng Biotechnol. 2014;2.

49. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA. Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res. 2013 Jan 1;41(1):474–86.

50. Balakrishnan R, Mori M, Segota I, Zhang Z, Aebersold R, Ludwig C, et al. Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. Science. 2022 Dec 9;378(6624):eabk2066.

51. Mortimer SA, Kidwell MA, Doudna JA. Insights into RNA structure and function from genome-wide studies. Nat Rev Genet. 2014 Jul;15(7):469–79.

52. Scharff LB, Childs L, Walther D, Bock R. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. Casadesús J, editor. PLoS Genet. 2011 Jun 23;7(6):e1002155.

53. 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 USA. 1974 Apr;71(4):1342–6.

54. Ma J, Campbell A, Karlin S. Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures. J Bacteriol. 2002 Oct 15;184(20):5733–45.

55. Kaminishi T, Wilson DN, Takemoto C, Harms JM, Kawazoe M, Schluenzen F, et al. A Snapshot of the 30S Ribosomal Subunit Capturing mRNA via the Shine-Dalgarno Interaction. Structure. 2007 Mar;15(3):289–97.

56. 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.

57. 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 Fact. 2020 Dec;19(1):85.

58. Zhang J, Fei Y, Sun L, Zhang QC. Advances and opportunities in RNA structure experimental determination and computational modeling. Nat Methods. 2022 Oct;19(10):1193–207.

59. Levy L, Anavy L, Solomon O, Cohen R, Brunwasser-Meirom M, Ohayon S, et al. A Synthetic Oligo Library and Sequencing Approach Reveals an Insulation Mechanism Encoded within Bacterial σ 54 Promoters. Cell Rep. 2017 Oct;21(3):845–58.

60. Waldminghaus T, Gaubig LC, Klinkert B, Narberhaus F. The *Escherichia coli* ibpA thermometer is comprised of stable and unstable structural elements. RNA Biol. 2009 Sep;6(4):455–63.

61. Morita M, Kanemori M, Yanagi H, Yura T. Heat-Induced Synthesis of ς ^{32} in *Escherichia coli* : Structural and Functional Dissection of *rpoH* mRNA Secondary Structure. J Bacteriol. 1999 Jan 15;181(2):401–10.

62. Mandal M, Breaker RR. Gene regulation by riboswitches. Nat Rev Mol Cell Biol. 2004 Jun;5(6):451–63.

63. Kavita K, Breaker RR. Discovering riboswitches: the past and the future. Trends Biochem Sci. 2023 Feb;48(2):119–41.

64. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic Control by a Metabolite Binding mRNA. Chem Biol. 2002 Sep;9(9):1043–9.

65. Hör J, Matera G, Vogel J, Gottesman S, Storz G. Trans-Acting Small RNAs and Their Effects on Gene Expression in *Escherichia coli* and *Salmonella enterica*. Lovett ST, Hinton D, editors. EcoSal Plus. 2020 Mar 20;9(1):ecosalplus.ESP-0030-2019.

66. Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol. 2011 Aug;9(8):578–89.

67. Link TM, Valentin-Hansen P, Brennan RG. Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proc Natl Acad Sci USA. 2009 Nov 17;106(46):19292–7.

68. Guisbert E, Rhodius VA, Ahuja N, Witkin E, Gross CA. Hfq Modulates the σ^{E} -Mediated Envelope Stress Response and the σ^{32} -Mediated Cytoplasmic Stress Response in *Escherichia coli*. J Bacteriol. 2007 Mar;189(5):1963–73.

69. Miwa T, Chadani Y, Taguchi H. *Escherichia coli* small heat shock protein IbpA is an aggregation-sensor that self-regulates its own expression at posttranscriptional levels. Mol Microbiol. 2021 Jan;115(1):142–56.

70. Dunkle JA, Cate JHD. Ribosome Structure and Dynamics During Translocation and Termination. Annu Rev Biophys. 2010 Apr 1;39(1):227–44.

71. Beringer M. Modulating the activity of the peptidyl transferase center of the ribosome. RNA. 2008 May;14(5):795–801.

72. 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.

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

74. Grentzmann G, Kelly PJ, Laalami S, Shuda M, Firpo MA, Cenatiempo Y, et al. Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex. RNA. 1998 Aug;4(8):973–83.

75. 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.

76. Graf M, Huter P, Maracci C, Peterek M, Rodnina MV, Wilson DN. Visualization of translation termination intermediates trapped by the Apidaecin 137 peptide during RF3-mediated recycling of RF1. Nat Commun. 2018 Aug 3;9(1):3053.

77. Peske F, Rodnina MV, Wintermeyer W. Sequence of Steps in Ribosome Recycling as Defined by Kinetic Analysis. Mol Cell. 2005 May;18(4):403–12.

78. Ciryam P, Morimoto RI, Vendruscolo M, Dobson CM, O'Brien EP. In vivo translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome. Proc Natl Acad Sci USA. 2013 Jan 8;110(2).

79. 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.

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

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

82. Beck K. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. EMBO J. 2000 Jan 4;19(1):134–43.

83. Calloni G, Chen T, Schermann SM, Chang H chun, Genevaux P, Agostini F, et al. DnaK Functions as a Central Hub in the E. coli Chaperone Network. Cell Rep. 2012 Mar;1(3):251–64.

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

85. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, et al. Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. Cell. 2005 Jul;122(2):209–20.

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

87. Sala A, Bordes P, Genevaux P. Multitasking SecB chaperones in bacteria. Front Microbiol. 2014 Dec 5;5.

88. Zimmer J, Nam Y, Rapoport TA. Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature. 2008 Oct;455(7215):936–43.

89. Gupta R, Toptygin D, Kaiser CM. The SecA motor generates mechanical force during protein translocation. Nat Commun. 2020 Jul 30;11(1):3802.

90. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005 Jan;433(7024):377–81.

91. Egea PF, Stroud RM. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. Proc Natl Acad Sci USA. 2010 Oct 5;107(40):17182–7.

92. Tsirigotaki A, De Geyter J, Šoštaric´ N, Economou A, Karamanou S. Protein export through the bacterial Sec pathway. Nat Rev Microbiol. 2017 Jan;15(1):21–36.

93. Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW. Folding of a large protein at high structural resolution. Proc Natl Acad Sci USA. 2013 Nov 19;110(47):18898–903.

94. Smets D, Tsirigotaki A, Smit JH, Krishnamurthy S, Portaliou AG, Vorobieva A, et al. Evolutionary adaptation of the protein folding pathway for secretability. EMBO J. 2022 Dec;41(23):e111344.

95. Denoncin K, Collet JF. Disulfide Bond Formation in the Bacterial Periplasm: Major Achievements and Challenges Ahead. Antioxid Redox Signal. 2013 Jul;19(1):63–71.

96. Landeta C, Boyd D, Beckwith J. Disulfide bond formation in prokaryotes. Nat Microbiol. 2018 Feb 20;3(3):270–80.

97. De Geyter J, Tsirigotaki A, Orfanoudaki G, Zorzini V, Economou A, Karamanou S. Protein folding in the cell envelope of Escherichia coli. Nat Microbiol. 2016 Jul 26;1(8):16107.

98. Merdanovic M, Clausen T, Kaiser M, Huber R, Ehrmann M. Protein Quality Control in the Bacterial Periplasm. Annu Rev Microbiol. 2011 Oct 13;65(1):149–68.

99. Sonoda H, Kumada Y, Katsuda T, Yamaji H. Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in Escherichia coli. J Biosci Bioeng. 2011 Apr;111(4):465–70.

100. Ow D, Lim D, Morin Nissom P, Camattari A, Wong V. Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. Microb Cell Fact. 2010;9(1):22.

101. Costello SM, Plummer AM, Fleming PJ, Fleming KG. Dynamic periplasmic chaperone reservoir facilitates biogenesis of outer membrane proteins. Proc Natl Acad Sci USA. 2016 Aug 16;113(33).

102. Schramm FD, Schroeder K, Jonas K. Protein aggregation in bacteria. FEMS Microbiol Rev. 2020 Jan 1;44(1):54–72.

103. Polson C, Sarkar P, Incledon B, Raguvaran V, Grant R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography–tandem mass spectrometry. J Chromatogr B. 2003 Mar 5;785(2):263–75.

104. Liu Y, Yang Q, Zhao F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. Annu Rev Biochem. 2021 Jun 20;90(1):375–401.

105. Kopp J, Bayer B, Slouka C, Striedner G, Dürkop M, Spadiut O. Fundamental insights in early-stage inclusion body formation. Microb. Biotechnol. 2023 May;16(5):893–900.

106. Rowley G, Spector M, Kormanec J, Roberts M. Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat Rev Microbiol. 2006 May;4(5):383–94.

107. Grabowicz M, Silhavy TJ. Envelope Stress Responses: An Interconnected Safety Net. Trends Biochem. Sci. 2017 Mar;42(3):232–42.

108. Veinger L, Diamant S, Buchner J, Goloubinoff P. The Small Heat-shock Protein IbpB from Escherichia coli Stabilizes Stress-denatured Proteins for Subsequent Refolding by a Multichaperone Network. J Biol Chem. 1998 May;273(18):11032–7.

109. 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.

110. Gottesman S. Proteases and their targets in *Escherichia coli*. Annu Rev Genet. 1996 Dec;30(1):465–506.

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

112. Mateus A, Bobonis J, Kurzawa N, Stein F, Helm D, Hevler J, et al. Thermal proteome profiling in bacteria: probing protein state *in vivo*. Mol Syst Biol. 2018 Jul;14(7):e8242.

113. Bury-Moné S, Nomane Y, Reymond N, Barbet R, Jacquet E, Imbeaud S, et al. Global Analysis of Extracytoplasmic Stress Signaling in Escherichia coli. Matic I, editor. PLoS Genet. 2009 Sep 18;5(9):e1000651.

114. Mutalik VK, Nonaka G, Ades SE, Rhodius VA, Gross CA. Promoter Strength Properties of the Complete Sigma E Regulon of *Escherichia coli* and *Salmonella enterica*. J Bacteriol. 2009 Dec;191(23):7279–87.

115. Chamachi N, Hartmann A, Ma MQ, Svirina A, Krainer G, Schlierf M. Chaperones Skp and SurA dynamically expand unfolded OmpX and synergistically disassemble oligomeric aggregates. Proc Natl Acad Sci USA. 2022 Mar;119(9):e2118919119.

116. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, et al. Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin. Science. 1977 Dec 9;198(4321):1056–63.

117. Rosano GL, Morales ES, Ceccarelli EA. New tools for recombinant protein production in *Escherichia coli* : A 5-year update. Protein Science. 2019 Aug;28(8):1412–22.

118. Blount ZD. The unexhausted potential of E. coli. Elife. 2015 Mar 25;4:e05826.

119. Balbás P, Bolívar F. Back to Basics: pBR322 and Protein Expression Systems in E. coli. In: Recombinant Gene Expression. New Jersey: Humana Press; 2004. p. 077–90.

120. Hernández-Beltrán JCR, San Millán A, Fuentes-Hernández A, Peña-Miller R. Mathematical Models of Plasmid Population Dynamics. Front Microbiol. 2021 Nov 4;12:606396.

121. Ammar EM, Wang X, Rao CV. Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. Sci Rep. 2018 Jan 12;8(1):609.

122. Aidelberg G, Towbin BD, Rothschild D, Dekel E, Bren A, Alon U. Hierarchy of non-glucose sugars in Escherichia coli. BMC Syst Biol. 2014 Dec;8(1):133.

123. William Studier F, Rosenberg AH, Dunn JJ, Dubendorff JW. [6] Use of T7 RNA polymerase to direct expression of cloned genes. In: Meth Enzymol Elsevier; 1990. p. 60–89.

124. 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.

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

126. Balzer S, Kucharova V, Megerle J, Lale R, Brautaset T, Valla S. A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. Microb Cell Fact. 2013 Dec;12(1):26.

127. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2006 Sep;72(2):211–22.

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

129. 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.

130. Goldstein MA, Doi RH. Prokaryotic promoters in biotechnology. In: Biotechnol. Annu. Rev. Elsevier; 1995. p. 105–28.

131. 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.

132. McClune CJ, Alvarez-Buylla A, Voigt CA, Laub MT. Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. Nature. 2019 Oct 31;574(7780):702–6.

133. Bervoets I, Van Brempt M, Van Nerom K, Van Hove B, Maertens J, De Mey M, et al. A sigma factor toolbox for orthogonal gene expression in Escherichia coli. Nucleic Acids Res. 2018 Feb 28;46(4):2133–44.

134. Tripathi L, Zhang Y, Lin Z. Bacterial Sigma Factors as Targets for Engineered or Synthetic Transcriptional Control. Front Bioeng Biotechnol. 2014 Sep 3;2.

135. Ng CY, Farasat I, Maranas CD, Salis HM. Rational design of a synthetic Entner–Doudoroff pathway for improved and controllable NADPH regeneration. Metab Eng. 2015 May;29:86–96.

136. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990 Aug;346(6287):818–22.

137. Jhaveri S, Rajendran M, Ellington AD. In vitro selection of signaling aptamers. Nat Biotechnol. 2000 Dec;18(12):1293–7.

138. Nutiu R, Li Y. In Vitro Selection of Structure-Switching Signaling Aptamers. Angew Chem Int Ed. 2005 Feb 4;44(7):1061–5.

139. Etzel M, Mörl M. Synthetic Riboswitches: From Plug and Pray toward Plug and Play. Biochem. 2017 Mar 7;56(9):1181–98.

140. Vinkenborg JL, Karnowski N, Famulok M. Aptamers for allosteric regulation. Nat Chem Biol. 2011 Aug;7(8):519–27.

141. Lynch SA, Gallivan JP. A flow cytometry-based screen for synthetic riboswitches. Nucleic Acids Res. 2009 Jan;37(1):184–92.

142. Dixon N, Duncan JN, Geerlings T, Dunstan MS, McCarthy JEG, Leys D, et al. Reengineering orthogonally selective riboswitches. Proc Natl Acad Sci USA. 2010 Feb 16;107(7):2830–5.

143. Robinson CJ, Vincent HA, Wu MC, Lowe PT, Dunstan MS, Leys D, et al. Modular Riboswitch Toolsets for Synthetic Genetic Control in Diverse Bacterial Species. J Am Chem Soc. 2014 Jul 30;136(30):10615–24.

144. Morra R, Shankar J, Robinson CJ, Halliwell S, Butler L, Upton M, et al. Dual transcriptional-translational cascade permits cellular level tuneable expression control. Nucleic Acids Res. 2016 Feb 18;44(3):e21–e21.

145. Horga LG, Halliwell S, Castiñeiras TS, Wyre C, Matos CFRO, Yovcheva DS, et al. Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact. 2018 Dec;17(1):199.

146. Yoo SM, Na D, Lee SY. Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli. Nat Protoc. 2013 Sep;8(9):1694–707.

147. Xie WH, Deng HK, Hou J, Wang LJ. Synthetic small regulatory RNAs in microbial metabolic engineering. Appl Microbiol Biotechnol. 2021 Jan;105(1):1–12.

148. Tan J, Sastry AV, Fremming KS, Bjørn SP, Hoffmeyer A, Seo S, et al. Independent component analysis of E. coli's transcriptome reveals the

cellular processes that respond to heterologous gene expression. Metabolic Engineering. 2020 Sep;61:360–8.

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

150. 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.

151. 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 Fact. 2021 Dec;20(1):13.

152. 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 inEscherichia coli. Biotechnol Bioeng. 2003 Jul 5;83(1):53–64.

153. 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.

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

155. Ceroni F, Algar R, Stan GB, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods. 2015 May;12(5):415–8.

156. 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.

157. F. M. Machado L, Currin A, Dixon N. Directed evolution of the PcaV allosteric transcription factor to generate a biosensor for aromatic aldehydes. J Biol Eng. 2019 Dec;13(1):91.

158. 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.

159. Hunke S, Betton JM. Temperature effect on inclusion body formation and stress response in the periplasm of Escherichia coli: Temperature effect on inclusion body formation. Mol Microbiol. 2003 Nov 11;50(5):1579–89.

160. 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 Fact. 2022 Dec 15;21(1):260.

161. Wild J, Walter WA, Gross CA, Altman E. Accumulation of secretory protein precursors in Escherichia coli induces the heat shock response. J Bacteriol. 1993 Jul;175(13):3992–7.