

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Identification of the compendium of factors interacting with
translating ribosomes in *E.coli* - A novel proteomic approach**

Sneha Kumar

aus

Kolkata, Indien

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. F. Ulrich Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, am 14.11.2016

Sneha Kumar

Dissertation eingereicht am: 14.11.2016

1. Gutachter: Prof. Dr. F. Ulrich Hartl

2. Gutachter: Dr. Dietmar E. Martin

Mündliche Prüfung am: 18.01.2017

Contents

1. Summary	6
2. Introduction	8
2.1 Protein structure	9
2.2 Protein synthesis	10
2.2.1 Thermodynamics of protein folding	12
2.3 Chaperone network in <i>E.coli</i>	14
2.3.1 Ribosome Associated Chaperone: Trigger Factor.....	16
2.3.2 The Hsp70 system.	19
2.3.3 <i>E.coli</i> chaperonin GroEL/ES	22
2.3.4 Additional chaperone systems in <i>E.coli</i>	26
2.4 Protein folding on the Ribosome	28
2.5 Co-translational translocation and folding	30
2.5.1 Mechanisms to study co-translational protein folding.....	33
2.6 Aim of the project	39
3. Materials and Methods	40
3.1 Materials	40
3.2 Buffers and media	43
3.3 Instruments	48
3.4 Methods	53
3.4.1 Biochemical methods.....	53
3.4.2 Proteomics method	59
4. Results	63
4.1 Establishment of a rapid pull-down method to explore ribosome nascent chain (RNC) interactome in <i>E.coli</i>	63
4.1.1 Layout of experimental design to study factors associated with ribosome nascent-chain complexes (RNC)	65
4.1.2 SDS PAGE and Western blotting analysis validates ribosomal protein footprint isolated by pull-down procedure.....	67
4.1.3 SILAC quantification for discrimination of genuine RNC interactors from non-specific background binders.....	69
4.1.4 GO Enrichment analysis of biological processes of enriched proteins	71
4.2 Antibiotics to stabilize and destabilize ribosome nascent chain complexes	73
4.2.1 SDS PAGE and Western blotting analysis of lysates treated with CAM and PURO antibiotics to generate different species of ribosome nascent chain complex	73

4.3	Analysis of RNC complexes generated using antibiotics	75
4.3.1	Layout of SILAC experimental design to identify additional factors associated with translating ribosomes <i>in vivo</i>	76
4.3.2	PURO treatment does not disrupt ribosomal subunits and untreated RNC sample is less stable than CAM treated RNC sample	77
4.4	Additional ribosome-associated proteins can be identified under conditions that stabilize ribosome nascent chain complex	79
4.4.1	Chloramphenicol stabilizes nascent chains on ribosomes and is essential to generate consistent dataset of RNC interactors	79
4.5	Stabilization of RNC interactions by chemical crosslinking	81
4.5.1	Time and concentration scale optimization of DSP used for crosslinking the intact <i>E. coli</i> ..	81
4.5.2	Optimized conditions for chemical crosslinking using DSP integrated with antibiotic treatment of spheroplasts	82
4.6	Comparison of RNC interactome before and after crosslinking with DSP reveals groups of potential RNC interactors	85
4.6.1	Crosslinking stabilizes RNC interactome	85
4.6.2	GO ontology analysis of enriched protein groups before and after crosslinking with DSP ..	88
4.6.3	Analysis of the dynamic RNC interactome generated by using antibiotic treatment coupled with chemical crosslinking	90
4.7	Protein groups associated with translating ribosomes under crosslinked and non-crosslinked conditions	91
4.8	Quantification of TF & DnaK chaperone levels normalized to the ribosome nascent chain using stalled luciferase constructs	96
4.8.1	Design of stalled nascent chain constructs of different length using pBAD Luciferase gene construct and 17 amino acid SecM stalling sequence	96
4.8.2	Expression profiles for stalled luciferase nascent complexes of different length reveal weaker expression of shorter constructs stalled on ribosomes	97
4.9	Chaperone interaction profile with stalled nascent chains of different length	97
5.	Discussion	100
5.1	Capturing ribosome nascent chain interactors <i>in vivo</i>	100
5.2	Emergence of new tools to study interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes	103
5.3	Compendium of factors associated with quality control of nascent proteome	105
6.	References	110
7.	Appendix	118
7.1	Abbreviations	118
7.2	List of primers	119
7.3	List of enriched protein ratios obtained from crosslinked and non-crosslinked datasets.	120

7.3.1	List of 245 enriched protein ratios obtained from CK1 FLAG tag strain versus C wild type strain.	120
7.3.2	List of enriched protein ratios obtained from crosslinked and non-crosslinked datasets CAM versus PURO from CK1 strain.....	127
7.4	Curriculum Vitae	144

1. Summary

How newly synthesized proteins fold to their functionally active conformations remains one of the fundamental questions in biology. Studies *in vivo* and *in vitro*, conducted over the last decade, indicate that translation and folding are coupled. Various components, such as molecular chaperones, proteases and other factors, interact with nascent polypeptide chains emerging from ribosomes and assist in nascent chain folding, quality control, subcellular targeting and co-translational modifications. A complete compendium of these factors is still missing and our understanding of the cellular machinery that acts co-translationally on the ribosome is still incomplete.

In this study we sought to identify the factors that interact with translating ribosomes in *E. coli* cells using quantitative proteomics methods. Antibiotics, which inhibit translation in bacteria by interfering with protein synthesis at specific stages, were employed to generate ribosome nascent chain complexes amenable to isolation and unbiased analysis. Two examples of antibiotics used in this study are chloramphenicol and puromycin. Both act on the ribosome by targeting the ribosomal peptidyl transferase center.

By combining translation inhibition with chemical crosslinking in spheroplasts and analysis by SILAC-based mass spectrometry, we designed a method that can be used to generate an inventory of factors contributing towards efficient folding of nascent chains in the cellular environment. We identified approximately 378 factors that interact specifically with translating ribosomes and are depleted upon nascent chain release from the ribosome. These factors could be divided into the following main categories: Molecular chaperones, RNA quality control proteins, transcription factors, translation factors and translocation factors.

Chaperones of the small heat shock protein family, disaggregates and proteases, apparently interact with nascent chains in a concerted manner to support productive folding, reverse misfolding and remove aberrant protein chains. Various transcription factors also interacted

substantially with ribosome nascent chain complexes (RNCs), suggesting their additional function in translation and folding.

Our analysis of RNCs provides further insights into the mechanism of co-translational translocation and helps to elucidate the role of chaperones and translocon protein complex in guiding co-translational targeting and stabilizing secretory and membrane proteins. The data provide a valuable resource for future functional studies.

2. Introduction

The question of how a linear amino acid sequence of a polypeptide chain acquires its functionally active native conformation is one of the most puzzling problems of biology. Our current understanding about this process is mainly based on *in vitro* studies with isolated polypeptide chains by studying their refolding after denaturation. Studies conducted over the last two decades have helped us to understand the pathways and machineries involved in protein folding in the cell *in vivo* (Balchin et al. 2016). However, how protein folding is coupled to the vectorial translation of the polypeptide chain on the ribosome is not yet well understood and the cellular components acting on the nascent polypeptide are not completely defined.

All prokaryotes have 70S ribosomes (S stands for Svedberg units) while eukaryotes have larger 80S ribosomes in their cytosol. The 70S ribosome consists of 50S and 30S subunits. The 50S subunit comprises of the 23S and 5S *rRNA* and 33 ribosomal proteins while the 30S subunit consists of the 16S *rRNA* and 21 ribosomal proteins.

Compared to the bacterial ribosomes, eukaryotic ribosomes are larger due to insertions or extensions to the conserved ribosome core. Several additional proteins are found in the small and large subunits of eukaryotic ribosomes, which do not have prokaryotic homologs.

The 40S subunit contains 18S *rRNA* (homologous to the prokaryotic 16S *rRNA*) and 32 ribosomal proteins. The 60S subunit contains a 26S *rRNA* (homologous to the prokaryotic 23S *rRNA*) and 47 ribosomal proteins. In addition, it contains a 5.8S *rRNA* (corresponding to the 5' end of the 23S *rRNA*), and a short 5S *rRNA*.

Ribosomes account for as much as 30% of the total cell mass, with up to 10^5 and 10^6 ribosomes in bacteria and mammalian cells, respectively (Bashan & Yonath 2008). Most ribosomes in a growing cell are actively translating and synthesizing polypeptide chains at rates of about 20 amino acids per second in bacteria and 5–9 amino acids per second in eukaryotes. Ribosomes have a total molecular mass ranging between 2.4 MDa in bacteria and 4 MDa in eukaryotes (Kramer et al. 2009).

An *mRNA* transcript is translated by several ribosomes simultaneously and these structural organizations are called polysomes. A recent cryo-electron tomography analysis of active polysomes showed that ribosomes are arranged in a staggered or pseudohelical organization, with the polypeptide exit sites facing outward. This arrangement maximizes the distance between nascent chains on adjacent ribosomes, thereby reducing the probability of intermolecular interactions between emerging peptides. This arrangement prevents misfolding and aggregation (Brandt et al. 2009).

2.1 Protein structure

The building blocks of proteins are 20 different amino acid types, bound to each other through covalent peptide bonds to form a chain. The amino acid sequence is genetically determined and unique for every protein. It is called the primary structure of a protein.

During synthesis, most proteins begin to fold into local secondary structure elements, α -helices and β -sheets (Pauling and Corey, 1951a). Most proteins contain multiple helices and sheets. The arrangement of secondary structure elements in space is referred to as tertiary structure. The quaternary structure of a protein refers to the assembly of two or more protein subunits forming larger complexes. The native conformation of a protein is thermodynamically favorable.

As proteins fold, they undergo a variety of structural transitions before reaching their final conformation, which is unique and generally compact. Folded proteins are stabilized by a multitude of non-covalent bond between amino acids. In addition, several intermolecular forces between a protein and its immediate environment contribute to a stable protein conformation. For example, soluble proteins in the cell cytoplasm have hydrophilic chemical groups on their surfaces, whereas their hydrophobic domains tend to be placed inside. In contrast, proteins inserted into cell membranes exhibit some hydrophobic chemical groups on their surface, particularly in regions that are exposed to membrane lipids.

2.2 Protein synthesis

Ribosomes are highly conserved ribonucleoprotein organelles responsible for the translation of mRNA into proteins. The process of translation is highly conserved and is closely regulated by the assembly of ribosomal subunits on mRNA templates. Ribosomes are composed of two subunits, one large and one small, which assemble to perform the synthesis of polypeptide chains.

There are three tRNA binding sites in bacteria. The aminoacyl-tRNA in the A-site of the ribosome usually functions as acceptor for the growing polypeptide during protein synthesis. The ribosomal site most frequently occupied by peptidyl-tRNA, i.e. the tRNA carrying the growing peptide chain, is known as P-site. The P-site is also referred to as the puromycin sensitive site. The ribosomal site harboring deacylated tRNA on the transit site out from the ribosome is known as E-site.

During elongation, the ribosome moves one codon down towards the 3' end of mRNA and brings in a charged tRNA to the ribosomal A-site, transfers the growing polypeptide chain from the P-site tRNA to the carboxyl group of the A-site amino acid, and ejects the uncharged tRNA at the E-site. When a stop or nonsense codon (UAA, UAG, or UGA) is reached on the mRNA, the ribosome terminates translation. In this way, ribosomes guide assembly of amino acids into a protein molecule which then folds into a three dimensional functional entity.

In bacteria, ribosomes initiate translation on mRNAs during transcription. Translation and transcription are strongly coupled cellular processes in bacteria. During the translation initiation stage, mRNA interacts explicitly with tRNA as well as the 30S ribosomal subunit. The region of mRNA covered by the ribosome in the translation initiation stage is called the ribosomal binding site (RBS). It extends over about 30 nucleotides (Laursen et al. 2005). Bacterial mRNAs are usually polycistronic and they possess multiple signals for initiation and termination of protein synthesis. The beginning of an mRNA translation site is marked by the initiation codon, which is recognized by a *fMet-tRNA*. This precise matching is assisted by three initiation factors, IF1, IF2 and IF3, and results in the formation of the 30S initiation complex. Subsequently, this complex joins the 50S

ribosomal subunit and releases the initiation factors, rendering the 70S initiation complex active (Simonetti et al. 2009).

The 30S subunit binds to the mRNA template at a purine-rich region called the Shine-Dalgarno sequence upstream of the AUG initiation codon. The Shine-Dalgarno sequence is complementary to a pyrimidine rich region on the 16S *rRNA* component of the 30S subunit. During the formation of the initiation complex, these complementary nucleotide sequences pair to form a double stranded RNA structure, which binds the mRNA to the ribosome in such a way that the initiation codon is placed at the P site (G.-W. Li et al. 2012). At this stage, the peptidyl transferase center of the large 50S subunit initiates the formation of new peptide bonds (Voisset et al. 2008).

After its synthesis at the ribosome, a polypeptide chain is exposed to a number of non-specific interactions and is at risk of misfolding and aggregation a highly crowded cellular environment. However, cells have evolved mechanisms to ensure the correct and efficient folding of newly synthesized polypeptides into their unique, thermodynamically stable, fully functional native structures (Jahn & Radford 2008).

The understanding of protein folding is associated with a simple question "How does the primary structure of a protein determine its secondary and tertiary structure? ". In his Nobel-prize winning work, Christian Anfinsen in the 1950s discovered that small proteins, such as ribonuclease A, can assume their three-dimensional conformations spontaneously *in vitro*. He showed that the primary sequence of a polypeptide is sufficient to instruct the process of three-dimensional protein folding.

This insight established a link between genetic code and protein function. However, it soon became clear that spontaneous folding often fails for proteins that are more complex and in the late 1980s several lines of evidence emerged that protein folding in the crowded cellular environment cell is assisted by complex protein machineries called *molecular chaperones*.

Molecular chaperones shield hydrophobic surfaces of non-native proteins, preventing their aggregation, and promote correct folding. A molecular chaperone is defined as any protein that

interacts, stabilizes or helps a non-native protein to attain its native conformation, but is not part of its final functional structure (Hartl 1996; Hartl and Hayer-Hartl, 2009).

Chaperones have various functions in *de novo* protein folding, oligomeric assembly, intracellular protein transport, refolding of stress-denatured proteins, and in protein degradation. Chaperone binding prevents aggregation, whereas transient release of bound hydrophobic regions is important for folding to proceed. Chaperones act by optimizing the efficiency of protein folding (Vabulas et al. 2010).

2.2.1 Thermodynamics of protein folding

Many naturally occurring proteins fold rapidly and efficiently to their native state without chaperone assistance. However, if protein folding were to occur through random search among all possible conformations, the time required to sample through the entire conformational space would be beyond the time range of any biological process. This argument has become known as Levinthal's Paradox. Levinthal concluded that proteins must fold into their native conformation via specific defined folding pathways (Levinthal et al., 1968).

Proteins fold to reduce their Gibbs free energy (G), which can be achieved by either decreasing the protein's enthalpy (H), or by increasing the system's entropy (S). This is shown by the equation $G = H - (S \times T)$, where T is the temperature of the system. The Gibbs free energy is a thermodynamic potential – more stable protein states are characterized by lower values of the Gibbs free energy. While temperature under physiological conditions does not vary significantly, enthalpy can be decreased through the formation of stable interactions between amino acids of the protein chain. Entropy is defined as the number of possible conformations that the protein can adopt. Entropy lies at the heart of the Levinthal Paradox.

The attractive and repulsive forces between neighboring amino acid residues favors certain conformations of individual amino acids in the polypeptide chain, thereby significantly reducing the number of possible folding pathways available. The folding process usually involves several intermediate states. Folding intermediates can act as “stepping stones” for attaining the native state or represent kinetically stable states. Misfolded conformations require substantial

reorganization before the native state can be reached (Vabulas et al. 2010). The formation of metastable, non-native interactions during the folding process is illustrated as a funnel-shaped folding energy landscape (Figure 1) (Onuchic and Wolynes 2004; Jahn and Radford 2005; Lindberg and Oliveberg 2007).

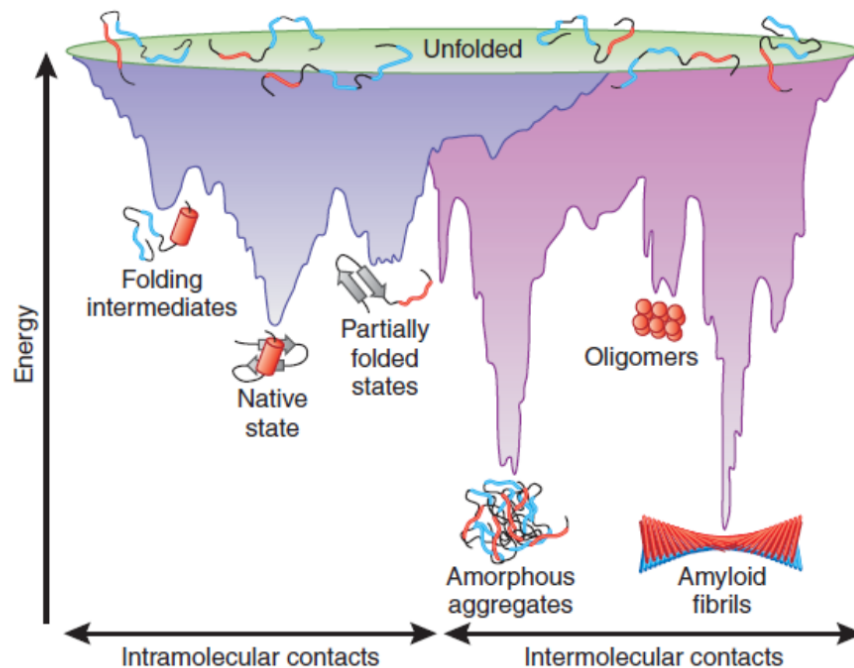


Figure 1. Energy landscape scheme of protein folding and aggregation. The purple surface shows the multitude of conformations ‘funneling’ to the native state via intramolecular contacts and the pink area shows the conformations moving toward amorphous aggregates or amyloid fibrils via intermolecular contacts. Both parts of the energy surface overlap. Aggregate formation can occur from intermediates populated during de novo folding or by destabilization of the native state into partially folded states and is normally prevented by molecular chaperones. Cell-toxic oligomers may occur as off-pathway intermediates of amyloid fibril formation. (Hartl & Hayer-Hartl 2009)

Partially folded or misfolded states tend to aggregate as they expose hydrophobic amino acid residues and regions of unstructured polypeptide backbone, which are largely buried in the native state. Aggregation can sometimes lead to the formation of ordered, fibrillar assemblies called

amyloids, in which beta strands run perpendicular to a long fibril axis (Figure 1). These thermodynamically very stable conformations are accessible to many proteins under denaturing conditions, independent of their sequence, suggesting that their formation is an inherent property of the polypeptide chain (Dobson 2003). Notably, this type of protein aggregation also involves the formation of less ordered, oligomeric, toxic intermediates, leading to neurodegenerative diseases, systemic amyloidosis, and other protein aggregation disorders.

2.3 Chaperone network in *E.coli*

Protein homeostasis in the cell is largely dependent on various classes of molecular chaperones functioning as a network. The concerted effort made by chaperones plays major role in maintaining the integrity of a nascent polypeptide emerging from the ribosome exit tunnel until the polypeptide attains its native functional structure. Molecular chaperones bind to their non-native substrate proteins by recognizing exposed hydrophobic surfaces. In correctly folded proteins, these hydrophobic patches are buried inside their globular structures. Chaperones promote correct folding of their substrate proteins by preventing or destabilizing incorrect polypeptide chain conformations, and, in some cases, by providing a sequestered environment in which correct protein folding can occur. The activity of chaperones often requires binding and hydrolysis of adenosine triphosphate (ATP) (Mayer & Bukau 2005).

Chaperone networks assisting protein folding in the cytosol follow a similar general organizational pattern in all three domains of life (Frydman 2001). A complex machinery is required to stabilize nascent polypeptides on ribosomes, while other components act downstream in a sequential manner to help complete the protein folding process (Langer, Lu, et al. 1992) (Figure 2).

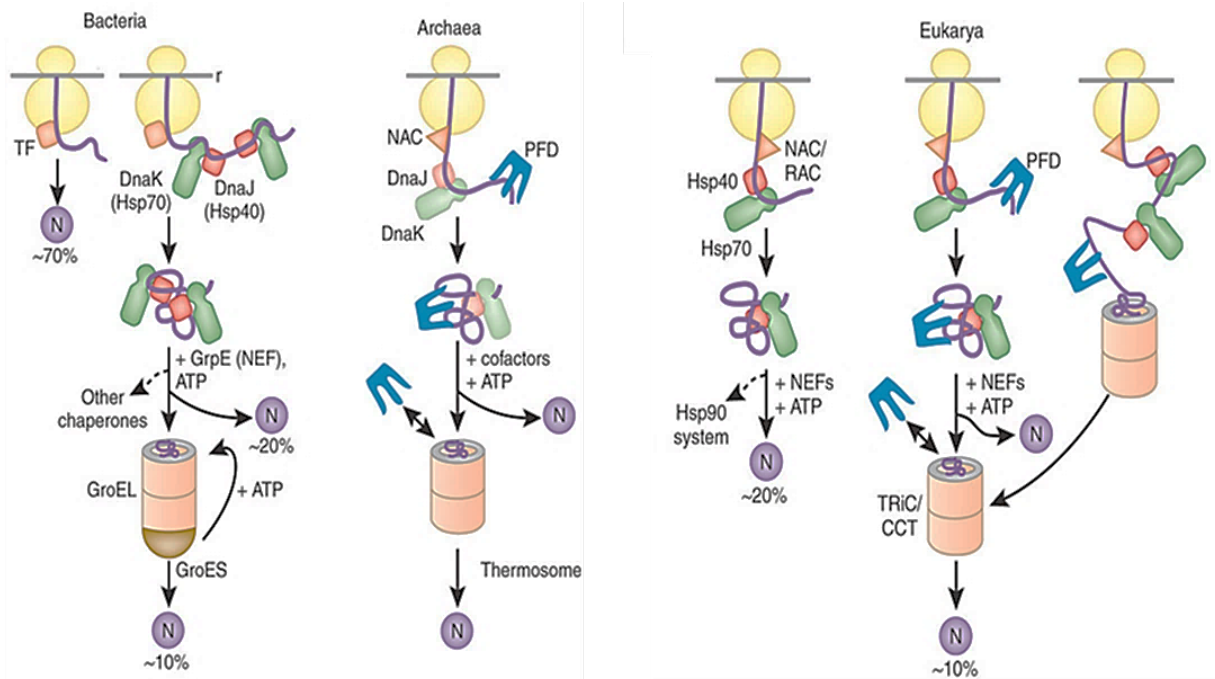


Figure 2: Protein folding in the cytosol. Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol. (a) Bacteria. Nascent chains probably interact generally with Trigger factor (TF), and most small proteins (70% of total) may fold rapidly upon synthesis without further assistance. Longer chains interact subsequently with DnaK and DnaJ (Hsp70 system) and fold upon one or several cycles of ATP-dependent binding and release. (b) Archaea. PFD, prefoldin; NAC, nascent chain-associated complex. Note that only some archaeal species contain DnaK and DnaJ. (c) Eukarya. TF, NAC probably interacts generally with nascent chains, but the role of NAC in folding is not yet clear. About 20% of chains reach their native states in a reaction assisted by RAC (ribosome-associated complex), Hsp70 and Hsp40. A fraction of these must be transferred to Hsp90 for folding. About 10% of chains are co- or post-translationally passed on to the chaperonin TRiC/CCT in a reaction mediated by Hsp70 and PFD, both of which interact directly with TRiC/CCT. PFD recognizes the nascent chains of certain TRiC substrates, including actin and tubulins. (Modified from Hartl & Hayer-Hartl, 2009)

Both these systems co-operate in coherent pathways. The number of interacting substrates for chaperone components decreases from upstream to downstream. To maintain the integrity of protein homeostasis (proteostasis) in *E. coli*, there exist three main chaperone systems: 1) the ribosome associated chaperone Trigger factor; 2) the DnaK-DnaJ-GrpE (KJE) system; 3) the GroEL/GroES chaperonin system (GroELS).

2.3.1 Ribosome Associated Chaperone: Trigger Factor

The nascent chain cannot undergo the highly co-operative process of folding, unless the synthesis of a domain or, in some cases, an entire protein is complete. Therefore, nascent chains emerging from the ribosome exit tunnel still exhibit non-native features, which renders them prone to aggregation. To maintain the homeostasis of nascent chains during their translation on the ribosome in *E.coli*, cells have evolved a ribosome associated chaperone, trigger factor (TF), which interacts with nascent chains and prevents them from misfolding or aggregation. This 48 kDa *E. coli* protein binds to a docking site on the L23 protein of the large 50S ribosomal subunit, in the vicinity of the ribosomal exit tunnel (Genevaux et al. 2004).

The 2.7Å crystal structure of *E.coli* TF together with the structure of its ribosome-binding domain in complex with the *Haloarcula marismortui* large ribosomal subunit revealed a unique elongated conformation comprising an amino-terminal ribosome-binding domain, a middle domain with peptidylprolyl isomerase (PPI) activity and the carboxy-terminal domain with two arm-like structures (Figure 3). From its attachment point at L23 protein on the ribosome, TF projects the extended domains over the exit tunnel of the ribosome, creating a sheltered folding space where nascent polypeptides may be shielded from degradation and aggregation (Ferbitz et al. 2004).

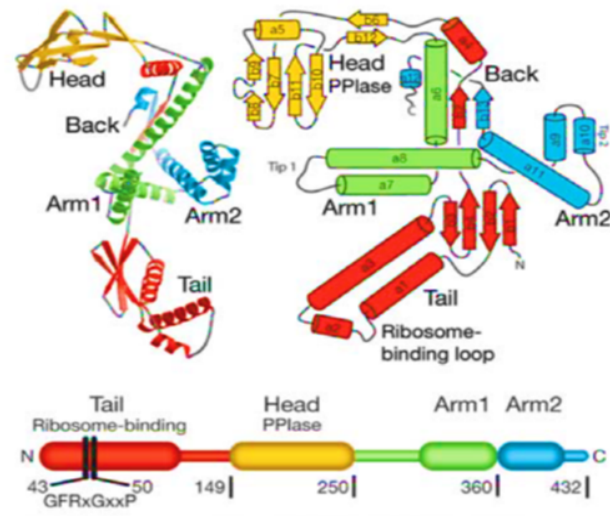


Figure 3: Structure of *E. coli* trigger factor and its N-terminal domain. Trigger factor adopts an extended fold. Left: ribbon diagram of the trigger factor fold. Right: schematic representation of the domain organization. Bottom: domain arrangement in sequence space. Positions of the ribosome-binding trigger factor signature (residues 43–50) and domain borders are indicated. The ribosome binding ‘tail’ is shown in red, the PPIase ‘head’ in yellow and ‘arm’ 1 and ‘arm’ 2 in green and blue, respectively. (Adapted from Ferbitz et al. 2004)

TF exhibits both peptidyl-prolyl cis/trans isomerase activity (PPIase) and chaperone-like function (Crooke & Wickner, 1987; Hesterkamp et al, 1996). The PPIase domain of TF is sufficient for the catalysis of peptidyl-prolyl isomerization but the catalysis is greatly enhanced by the presence of the two other domains, suggesting that cooperation between the PPIase domain and the chaperone-like functions is critical (Scholz et al, 1997). Additionally, crosslinking experiments have also shown that a TF fragment containing the PPIase domain linked to the ribosome via the N-terminal ribosome binding domain, is sufficient for interaction with nascent polypeptide substrates (Patzelt et al. 2001).

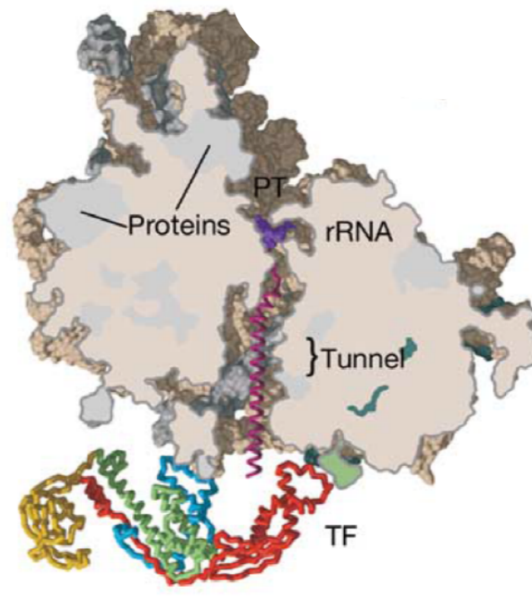


Figure 4: Structure of the trigger factor bound to the 50S ribosomal subunit. Overview of the trigger factor 50S complex. Full-length trigger factor positioned by superimposition onto the ribosome-bound fragment trigger factor 1–144 is shown as C α -trace together with a slice of 50S along the peptide exit tunnel with a modelled nascent chain in magenta, extending from the peptidyl transferase center (PT). (Adapted from Ferbitz et al. 2004)

TF is an abundant cytosolic protein with a concentration of $\sim 40\text{--}50\ \mu\text{M}$, exceeding that of ribosomes, which is $\sim 30\ \mu\text{M}$ (Bremer & Dennis 1996; Lill et al. 1988). Therefore, ribosome-bound TF is in equilibrium with free TF in the cytosol. It associates co-translationally very early with most nascent chains of cytosolic and secreted proteins independent of proline content (Valent et al, 1995; Patzelt et al, 2001). TF is ATP independent and rapid binding to and release of the substrate from TF suggests that it may facilitate the elongation of the polypeptide chains (Maier et al, 2001). The substrate binding motif of TF is a stretch of eight amino acids, enriched in basic and aromatic residues with a positive net charge, and TF requires ribosome association to create high local concentrations of nascent polypeptide substrates for productive interaction *in vivo* (Patzelt et al. 2001). The TF-substrate complex dissociates after release of polypeptide from the ribosome (Agashe et al. 2004; Hesterkamp et al. 1996).

The eukaryotic cytosol contains a heterodimeric complex called NAC (nascent polypeptide associated complex), which may be a functional TF homologue. NAC also binds to short nascent chains emerging from the ribosome and dissociates upon release from the ribosome (Beatrix et al. 2000; Hartl 1996). NAC consists of α NAC (33 kDa) and β NAC (22 kDa) (Shi et al. 1995; Wiedmann et al. 1994) and prevents mistargeting of non-secretory proteins to the endoplasmic reticulum (Martin Gamerding, Marie Anne Hanebuth, Tancred Frickey 2015). The docking site of NAC was mapped to the ribosomal protein L25, which is the eukaryotic homolog of L23 in bacteria (Grallath et al. 2006). NAC has been observed to prevent aggregation in experiments *in vitro* (Grallath et al. 2006), thus exhibiting certain chaperone-like properties.

In addition to NAC, the eukaryotic model organism *Saccharomyces cerevisiae* harbours another major ribosome associated chaperone system, RAC (ribosome associated complex). RAC consists of protein Ssz1, which is member of the Hsp70 family, and the DnaJ-related Hsp40 protein zuotin (Gautschi et al. 2001; Michimoto et al. 2000). Zuotin contains a ribosome binding domain and a J-domain for interaction with Hsp70 (Yan et al. 1998). In mammals, the constitutively expressed member of the Hsp70 molecular chaperone network is found assisting the maturation of emerging nascent polypeptides protein from ribosome (Beckmann et al. 1990; Frydman et al. 1994). Hsc70 was first reported to bind co-translationally to mammalian nascent chains, in cooperation with the Hsp40 homologs Hdj1 and Hdj2 (Nagata et al. 1998; Terada et al. 1997). Multifunctional Hsc70 is recruited to the ribosome by the recently identified human ortholog of yeast ribosome-associated J-protein Zuo (Mpp11 in humans). This discovery demonstrates that ribosome-tethered chaperones have been conserved throughout evolution (Hundley et al. 2005).

2.3.2 The Hsp70 system.

Members of the Hsp70 family of proteins are central components of the network of molecular chaperones and assist in a wide range of folding processes. These consist of folding and assembly of newly synthesized proteins, refolding of aggregated and misfolded proteins and membrane translocation of secretory proteins (Mayer & Bukau 2005). Members of the Hsp70 family of

proteins are located in the cytosol of bacteria and in membrane bound eukaryotic organelles such as endoplasmic reticulum and mitochondria.

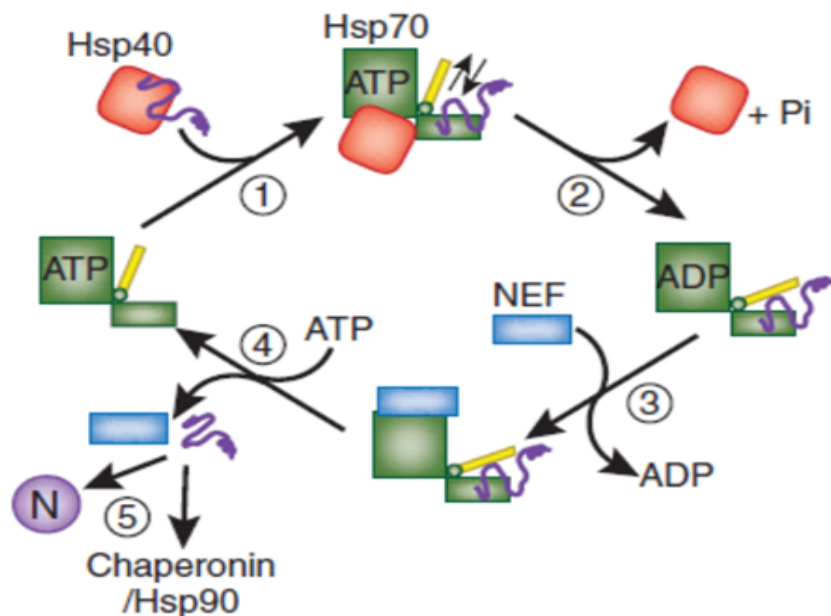


Figure 5: Hsp70 reaction cycle. (1) Hsp40-mediated delivery of substrate to ATP-bound Hsp70. (2) Hydrolysis of ATP to ADP, accelerated by Hsp40, results in closing of the α -helical lid and tight binding of substrate by Hsp70. Hsp40 dissociates from Hsp70. (3) Dissociation of ADP catalyzed by NEF. (4) Opening of the α -helical lid, induced by ATP binding, results in substrate release. (5) Released substrate either folds to native state (N) and is transferred to downstream chaperones or rebinds to Hsp70. (Modified from Hartl and Hayer Hartl, 2002)

The mechanism for Hsp70 chaperone function is well understood for the bacterial DnaK system (Mayer and Bukau, 2005). DnaK was identified in DNA replication and is found involved in various cellular pathways such as disaggregation, translocation, *de novo* folding and assembly of oligomeric complexes (Glover and Lindquist 1998; Teter et al. 1999). Like most Hsp70s, DnaK works with an Hsp40 co-chaperone, DnaJ and nucleotide-exchange factor GrpE that regulates ATP-dependent functional cycle of DnaK. In the ATP-bound state, the substrate-binding cleft of

DnaK is in an open conformation allowing rapid substrate association and dissociation. This renders a relatively low stability of chaperone-substrate complex. Hydrolysis of ATP to ADP leads to large structural rearrangements within DnaK, resulting in closing of the substrate binding pocket and stable substrate association. The cycling of DnaK between an ATP-bound state with low affinity and an ADP-bound state with high affinity for substrate is modulated by DnaJ and GrpE co-factors. The 41kDa protein DnaJ, binds to DnaK through its amino-terminal J-domain and stimulates the ATPase activity of DnaK, hence enhancing stable peptide binding. DnaJ delivers unfolded peptides to DnaK through binding of hydrophobic peptides to a carboxyl terminal binding site of DnaJ protein. In the meantime, the nucleotide exchange factor GrpE functions by rapidly releasing ADP from DnaK. The rebinding of ATP to DnaK leads to decreased affinity and release of the substrate, thus completing the DnaK cycle (Figure 6).

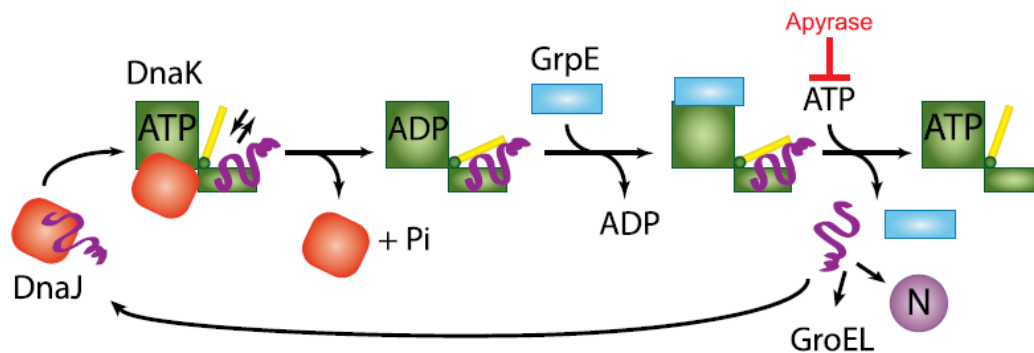


Figure 6 : Schematic representation of the DnaK reaction cycle. Upon DnaJ-mediated delivery of non-native protein substrate to ATP-bound DnaK, hydrolysis of ATP to ADP results in closing of the α -helical lid (yellow) and tight binding of substrate by DnaK. Stable DnaK-substrate complexes are accumulated by depleting ATP with apyrase upon cell lysis. (Calloni et al. 2012)

Recent quantitative proteomic studies have revealed DnaK-substrate complexes from wild-type, TF-deleted, or GroEL-depleted cells (Calloni et al. 2012). It has been noted on the basis of relative enrichment on DnaK that at least ~ 700 newly synthesized and pre-existing proteins interact with DnaK during folding. Individual deletion of either TF or depletion of GroEL/ES leads to distinct

changes in the DnaK interactome. These effects are highly informative about the functional cooperativity of chaperone modules. Therefore, DnaK is considered a central hub in the cytosolic *E. coli* chaperone network, acting downstream of TF and upstream of chaperonin. The functional interconnection of these major chaperone systems maintains a proteostasis control (Langer et al. 1992; Calloni et al. 2012).

2.3.3 *E. coli* chaperonin GroEL/ES

Chaperonins are oligomeric assemblies with a molecular mass of 800kDa and consisting of two stacked ring structures, in which each ring encloses a central cavity (Bukau and Horwich 1998, Hartl and Hayer-Hartl, 2002). Group I chaperonins referred to as Hsp60 in eukaryotes and GroEL in the bacterial cytosol are also present in eukaryotic organelles such as mitochondria matrix and chloroplast stroma (Cpn60) (Horwich et al. 2007). They consist of two homo-heptameric rings which form a barrel-shaped structure for substrate encapsulation. Group I chaperonins depend on the presence of a lid-like cofactor GroES in bacteria, Hsp10 in mitochondria and Cpn10/Cpn20 in chloroplasts (Hayer-hartl & Hartl 2015). Their functional reaction cycle involves opening and closing of the central cavity through an association and dissociation cycle of the respective cofactor in an ATP dependent fashion (Hartl et al. 2011). In the closed state, an encapsulated substrate protein can fold inside the central cavity. Group II chaperonins include TRiC/CCT in the eukaryotic cytosol and thermosome in archaea and consist of two octa- or nonameric rings. They are topologically similar to Group I chaperonins but do not share distinct sequence homology. In contrast to group I chaperonins, group II chaperonins have an in-built lid in the form of an apical protrusion.

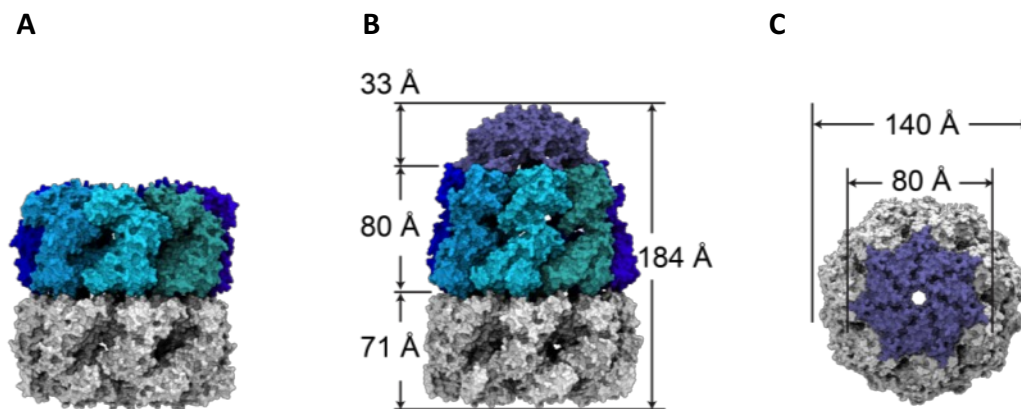


Figure 7: Structure and dimensions of apo GroEL and the GroEL/ES complex(A) Apo conformation of GroEL in absence of nucleotide (PDB 1XCK), with one ring colored in grey and the subunits in the other ring colored in shades of blue. (B) Structure and dimensions of the ADP/GroES bound GroEL complex (PDB 1AON). The trans-ring is colored in grey, the subunits of the cis-ring in different shades of blue and GroES is colored in purple. (C) Top view of the GroEL/ES complex as shown in (B). GroEL is colored in grey and GroES in purple. Dimensions from (Xu, Horwich, and Sigler 1997) (Adapted from Amit Gupta's thesis)

Opening and closing of the central cavity in the Group II chaperonins also needs ATP and ATP hydrolysis triggers closing of the cage by structural rearrangements in an iris-like fashion (Michael J. Kerner et al. 2005). All chaperonins act by capturing non-native polypeptides through hydrophobic interactions.

The *E. coli* cytosolic chaperonin GroEL is the most studied group I chaperonin. The cylindrical GroEL consists of two heptameric rings of identical ~57 kDa subunits stacked back to back in a staggered conformation (Figure 7) (Hayer-hartl & Hartl 2015).

GroEL monomers are identical and it contains an equatorial ATPase domain, an apical domain and a intermediate hinge domain. The equatorial domain forms an inter-subunit contact between the two GroEL rings and contains ATP or ADP binding site. The apical domains outline the entrance to the central cavity and contain hydrophobic segments for binding of non-native substrate proteins. The intermediate domain serves as a linker, conducting structural changes from

equatorial to apical domains. GroES, a heptameric co-factor of ~10 kDa subunits acts as a lid, binding to the apical domains of GroEL.

a. The reaction cycle of the GroEL/GroES-system

The reaction cycle of GroEL-mediated protein folding is strongly linked to the GroEL ATPase function. GroEL acts like a two-stroke engine because of allosteric regulation of the GroEL ATPase cycle. Binding of ATP to GroEL results in large structural rearrangements, preparing the apical domains for binding to GroES. Binding of GroES to the apical domains then results in displacement of the substrate protein into the hydrophilic cage. This results in the formation of an asymmetric complex called the cis-ring. The unliganded ring is called trans-ring. In the cis-ring, ATP hydrolysis takes ~10 sec at 25°C when the substrate is absent and ~2.2 sec at 37°C when the substrate is present (Gupta et al. 2014). The functional cycle is finally completed by the binding of ATP and GroES to the trans-ring. This triggers the release of GroES, ADP and the substrate from the former cis-ring. Substrate proteins that were unable to fold during one round of encapsulation are promptly recaptured and subjected to additional folding attempts. (Figure 8). The structure of GroEL in its various nucleotide-bound conformations and in the presence and absence of GroES has been extensively studied by electron microscopy (Saibil et al. 1991; Langer, Pfeifer, et al. 1992; Braig et al. 1993) and crystallography (Braig et al. 1994; Xu et al. 1997).

b. GroEL/GroES mediated protein folding

Current understanding of the pathway of GroEL/GroES mediated protein folding is summarized in Figure 8. When a non-native polypeptide binds to an asymmetric GroEL-GroES complex, a dynamic species is formed. In this conformation, the GroEL central channel is at one end of the cylinder and is capped by GroES, whereas the channel at the other end remain accessible to the polypeptides. This conformational rearrangement happens in the presence of physiological levels of ATP or ADP (Fenton and Horwich 1997). Depending on whether GroEL is in the ADP or ATP bound state, GroES binds to either one of the two GroEL rings. The co-factor GroES is a single ring

dome-shaped heptamer of identical 10kDa subunits. It contains β -strands with one exceptionally long β -hairpin loop through which GroES contacts GroEL.

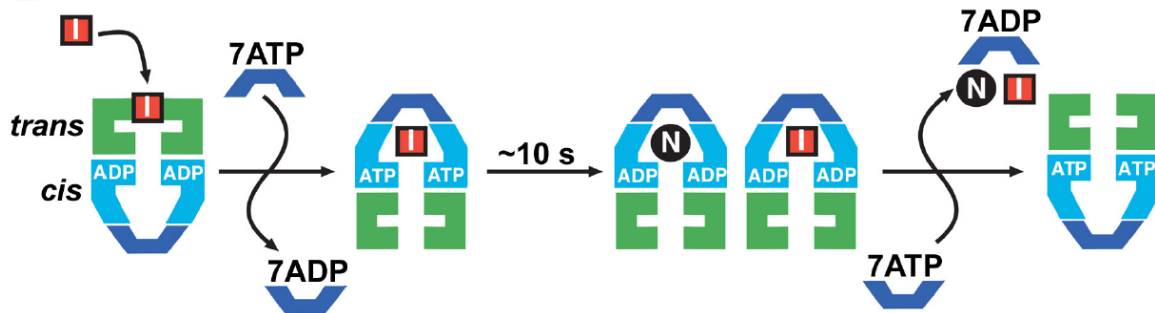


Figure 8: The reaction cycle of the GroEL/GroES-system. I: folding intermediate; N: natively folded protein. The GroEL trans ring is shown in green, the cis ring in light blue. GroES is colored dark blue. Binding of the folding intermediate at the apical domains of the Trans ring is followed by binding of GroES and seven molecules of ATP to the same ring, leading to release of the substrate into the GroEL/GroES cavity. Simultaneously, ADP and GroES dissociate from the opposite ring. The substrate is allowed to fold inside the now hydrophilic cavity for ~ 10 s, before ATP hydrolysis leads to binding of GroES and ATP to the opposite ring and release of ADP, GroES and the substrate. Within one cycle, both N and I accumulate, and I rapidly rebinds. (Adapted from Hartl & Hayer-Hartl, 2009)

Since the hydrophobic binding region of GroEL overlaps with that of GroES (Fenton et al. 1994) the binding of GroES causes the elevation of the hydrophobic binding surface. Subsequently, the substrate is released from the GroEL apical domain into a cage-like cavity for folding into its native conformation. Although the domain rearrangement results in burying hydrophobic residues and changing the environment inside the GroEL-GroES cavity to hydrophilic, the cage volume is enlarged by two-fold, to accommodate polypeptides up to 60kDa size (Hartl & Hayer-Hartl 2009).

c. Substrates of GroEL

The GroEL interactome was determined by a quantitative proteomics approach. Around ~ 250 of the ~ 2400 cytosolic *E. coli* proteins were identified as interactors of GroEL (Kerner et al. 2005). Identified GroEL interactors were further grouped into three different classes. Class I proteins interact with GroEL but do not require GroEL for folding. Further analysis revealed that less than

1% of class I proteins need the assistance from GroEL (Michael J. Kerner et al. 2005). Class II proteins require the assistance of GroEL for efficient folding at 37°C but fold spontaneously at 25°C. Further, class II proteins can fold via the Hsp70 system and therefore, are only partially dependent on GroEL for folding. However, the ~ 84 class III groups of proteins are obligate GroEL substrates and occupy 75%-80% of the total cellular chaperonin capacity. Depletion of GroEL showed aggregation or depletion of class III proteins from the cellular proteome. Since class III contains 14 essential proteins, GroEL is critical for cell viability. Most class III proteins are between 20kDa and 50kDa with complex topologies and distinct enrichment of the $(\beta\alpha)_8$ -TIM barrel fold (Kerner et al. 2005). An extensive analysis of GroEL obligation for protein solubility showed an enrichment of metabolic enzymes amongst obligate GroEL interactors (Fujiwara et al. 2010). Thus, from an evolutionary perspective, GroEL might reduce the aggregation propensity of structurally destabilizing mutations in enzymes, by either promoting the folding of partially misfolded species of enzymatic proteins or preventing their aggregation (Fenton & Horwich 1997).

2.3.4 Additional chaperone systems in *E.coli*

In addition to all the cytosolic chaperone systems described above, a large number of other cellular chaperones are found to assist the folding of newly-synthesized polypeptides or denatured proteins (due to stress), often in co-operation with the Dnak/DnaJ system or the GroEL/GroES chaperonin system. Compartments other than the cytosol, such as the mitochondria, chloroplasts and endoplasmic reticulum in eukaryotes or the periplasm in bacteria, harbor their own exceptional inventory of specialized molecular chaperones. For example, the periplasmic group of chaperones SurA, DegP, and Skp in gram-negative bacterium *E. coli* are thought to function as chaperones in the outer membrane protein targeting pathways. SurA is a member of the peptidyl-prolyl isomerase family and also has general chaperone activity (Sklar et al. 2007). DegP displays both protease and general chaperone activity regulated in a temperature dependent fashion (Lipinska et al. 1990). Skp interacts with denatured outer membrane proteins but not with denatured periplasmic or cytosolic proteins (Chen & Henning 1996).

Small heat-shock proteins often found associated with inclusion bodies in *E. coli*, are stress-inducible molecular chaperones. These bind to unfolded proteins, preventing their aggregation and enabling their refolding by ATP-dependent DnaK, DnaJ, GrpE and ClpB chaperones (Ehrnsperger et al. 1997; Montfort et al. 2001). Small heat shock proteins of the eukaryotic cytosol include Hsp12, Hsp42, and the mammalian α -crystallins whereas in eubacteria, this class of proteins include IbpA and IbpB. It was also shown that during heat stress, IbpA/B localize to the fraction of denatured and aggregated *E. coli* proteins (Matuszewska et al. 2008).

The HSP100/Clp proteins constitute a subfamily of AAA proteins and participate in the resolubilization of aggregated proteins together with the DnaK chaperone system or the protein degradation machinery (Glover & Lindquist 1998; Mogk & Bukau 2004; Schirmer et al. 1996; Weibezahn et al. 2005). They are a recently discovered family of chaperones with a great diversity of functions including transcriptional regulation, increased tolerance to high temperatures and proteolysis of specific cellular substrates. HSP100/Clp proteins are localized to different subcellular compartments. ClpA protein in *E. coli* was discovered as a component of an ATP-dependent protease and was subsequently shown to induce DNA-binding activity (Schirmer et al. 1996). Another chaperone, Hsp33, is seen activated upon exposure of cells to peroxide stress at elevated temperatures. The specific activation of Hsp33 by the oxidative unfolding of its redox-switch domain makes this chaperone suitable for oxidative stress conditions that lead to protein unfolding. Hsp 33 acts as a holding chaperone (Saper 2001).

Hsp90 is another essential ATP-dependent molecular chaperone that associates with numerous client proteins. Hsp90 molecules are conserved from bacteria to mammals and occur in many eukaryotic organelles as well as in the cytosol. HtpG is a prokaryotic homolog of Hsp90 and is essential for thermotolerance in cyanobacteria and *in vitro*, it efficiently suppresses the aggregation of denatured proteins (Nakamoto et al. 2014). Table 1 summarizes additional chaperones found in bacteria.

Table 1: Additional Chaperone Proteins in *E.coli*

Name	Cofactors	Function	Substrate specificity	ATP requirement
ClpB		Disaggregase	Segments enriched in aromatic and basic residues	+
HscA	HscB	Iron-sulfur cluster protein assembly	LPPVK motif in iron-sulfur cluster protein assembly IscU	+
HscC	YbeV, YbeS	regulation	Unknown	+
Hsp33		Holding chaperone	Unknown	-
SecB		Secretory chaperone	Nine amino acid motif enriched in aromatic and basic residues	-
HtpG		Possible folding/secretory chaperone	Unknown	+
IbpA, IbpB		Holding chaperone	Unknown	-

2.4 Protein folding on the Ribosome

Protein folding generally occurs in a stepwise manner. In the first step, the newly synthesized polypeptide chain emerges from the ribosome and short segments fold into secondary structural units that provide local regions of organized structure. Successful folding now depends on selection of an appropriate arrangement of this relatively small number of secondary structural elements (Levinthal paradox). In the second step, the forces that drive the hydrophobic regions onto the interior of the protein away from solvent predominate and the modules of secondary structure rearrange to arrive at the mature conformation of the protein. In general, each element of secondary or super secondary structure facilitates proper folding by directing the folding process towards the native conformation and away from unproductive and sometimes toxic alternatives.

Significant progress has been made in elucidating the role of the ribosome in the folding process of a polypeptide chain. A number of *in vivo* and *in vitro* experiments have shown that nascent peptides fold through specific interactions of some amino acids with the nucleotides in the peptidyl transferase center (PTC) of the large ribosomal subunit (Das et al. 2008). Chaperones such as DnaK/J and trigger factor are known to associate with nascent proteins on ribosomes before their release. However, there are proteins that cannot be folded with the help of specific chaperones or are intrinsically unstructured and do not show interactions with ribosome bound chaperones. Some of the experiments done in the past that showed the role of chaperones on the folding of proteins still bound to the ribosome in *E.coli* involved the termination of translation by puromycin or chilling the translating complex to delay the release of the nascent full-length proteins (Hoffmann et al. 2012).

To understand the process of protein folding that is expected to begin at the ribosomal exit tunnel and involve the ribosomal surface, a number of molecular biology approaches have been identified which use stalling sequences both in prokaryotic and eukaryotic model systems to generate a species of Ribosome Nascent Chain (RNC) complexes. A protein of interest can be stalled on ribosome and its folding can be studied through crosslinking, NMR spectroscopy, fluorescence, and mass spectrometry, both *in vivo* and *in vitro*. Our current understanding about the process of protein folding is mainly based on the *in vitro* studies done on isolated polypeptide chains by studying the refolding of denatured proteins. However, to understand protein folding *in vivo*, it is important to concentrate on how the nascent chain starts emerging from the ribosome in an N-terminal to C-terminal vectorial fashion and the interactions during its emergence from the ribosome exit tunnel that helps it to acquire its native functional conformation.

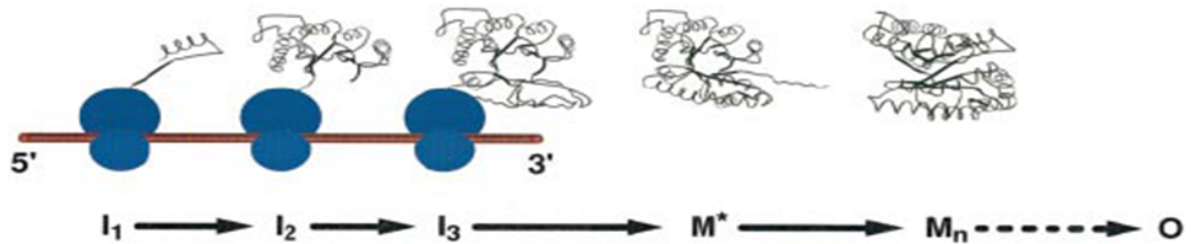


Figure 9: Cartoon depiction of co-translational folding of a polypeptide. The nascent polypeptide chain is shown assuming secondary structure as it emerges from the ribosome during the process of biosynthesis. The earliest intermediate, I_1 is not well stabilized by extensive tertiary interactions and is in equilibrium with multiple conformations. The second intermediate shown, I_2 , is the N terminal domain; more extensive tertiary interactions would allow this intermediate to be more stable. The final intermediate, I_3 , depicts the structure of the full length polypeptide immediately prior to release from the ribosome with the C terminal domain not yet fully packed. Chaperones and /or folding catalysts may interact with either the nascent chain intermediate structures or with full length product (M^*) following release from the ribosome. The final stage of folding release M^* to native monomer, M_n , occur following release. Association of monomeric units into oligomeric units may occur post-translationally as depicted above. The above cartoon representation of co-translational protein folding process has been taken from a Minireview by Thomas O. Baldwin (Alexey N. Fedorov and Thomas Baldwin, 1997).

This process of folding on the surface of the ribosome is known as co-translational protein folding. (Halic et al., 2006) This early stage of protein folding, which includes secondary structure formation and compaction, requires less than 1 second (Roder. H and Colon W. 1997). A number of biochemical and biophysical experiments coupled with computer simulations have improved our understanding of co-translational protein folding considerably.

2.5 Co-translational translocation and folding

The bacterial cytoplasmic membrane contains a huge variety of proteins involved in important enzymatic processes. It is estimated that a bacterial cell has 5×10^5 protein molecules in its inner membrane that are embedded in a bilayer of 2×10^7 lipid molecules. Most proteins span the membrane multiple times and therefore, the proteins and lipids approximately cover an equal surface of the membrane (Facey & Kuhn 2004). The molecular events underlying newly

synthesized membrane protein folding, targeting, translocation and insertion have been investigated using a number of different model proteins (Saraogi & Shan 2011). In recent years, the Sec protein translocation pathway has been extensively characterized. This is because of its role in translocating many membrane proteins, in particular those containing larger hydrophilic domains in the periplasm (Du Plessis et al. 2011). Only very few proteins have been described that use the Tat translocase, a system that is able to translocate a folded protein chain (Mori & Ito 2001).

Recently, a new component, YidC, has been identified in *E. coli*, which helps newly synthesized proteins to fold into the membrane layer. Bacterial pre-protein secretion occurs in three distinct stages, targeting (I) translocation (II) and release (III). The pre-protein crosses the membrane plane through the pre-protein translocase. The cellular machinery involved in the three stages of protein secretion in *E. coli* is shown in Figure 10, demonstrating the step-wise movement of a model outer membrane protein. Several housekeeping chaperones also contribute to efficient membrane targeting. SecA and SecYEG compose the core of the pre-protein translocase, whereas SecD and SecF are regulatory subunits (Economou 1999). This process is initiated by binding of the translocase to a hydrophobic region called as a signal sequence. The Sec component then catalyzes the translocation of the adjacent hydrophilic domain across the membrane (von Loeffelholz et al. 2011). In contrast, membrane insertases interact with hydrophobic regions of a newly synthesized protein and catalyze their folding into transmembrane domains. In order to unravel the mechanism of operation of these systems, model proteins have been used extensively to address specific aspects of each of these systems (Woldringh 2002). One approach to understand cellular processes underlying co-translational translocation is based on reconstituted systems, where the function of purified components can be investigated at the molecular level. With the transport systems discussed here, these reconstituted systems consists of the lipid bilayer, the purified transport system and a purified substrate protein.

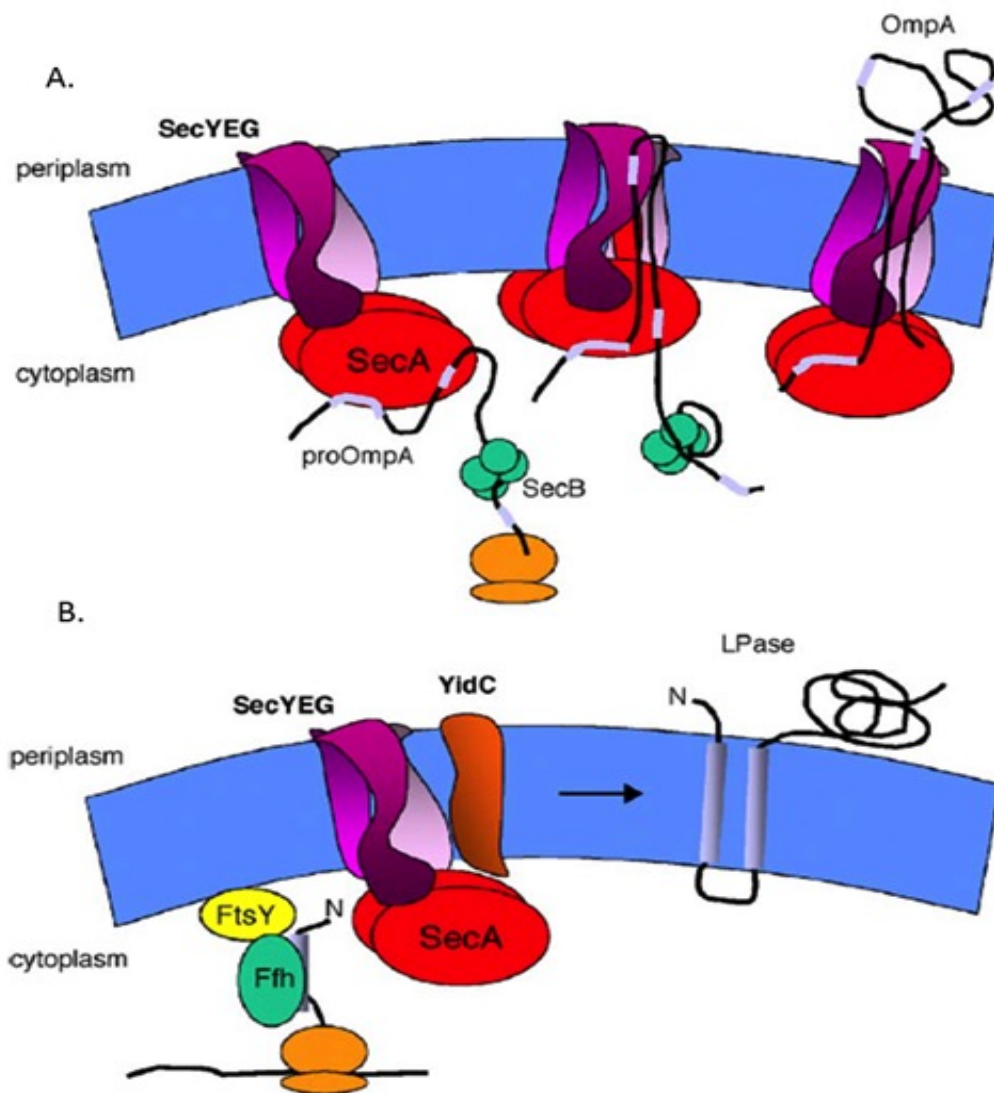


Figure 10: Model for the stepwise movement of proOmpA during translocation across the cytoplasmic membrane of *E. coli*. Upon initiation of translocation, the mature part of the pre-protein is transferred from SecB to SecA, which then transfers the polypeptide chain into the membrane-embedded translocase. Shorthydrophobic segments (marked in blue) in the mature domain of proOmpA participate in the stepwise translocation of proOmpA through the Sec translocon. (b) Model for the cotranslational targeting of the leader peptidase (LPase) by the bacterial SRP and its receptor to the SecYEG translocon and YidC insertase in the plasma membrane. This pathway requires Ffh for targeting the protein to its receptor FtsY. The SRP – nascent chain – ribosome complex is then targeted to the plasma membrane with the help of FtsY. SecA is required for translocation of the large periplasmic domain of LPase. Translocation and integration into the membrane requires SecYEG and YidC (Facey & Kuhn 2004).

To study the essential features that enable proteins to insert into the membrane, model proteins have been characterized with a wide variety of sequence variations. These investigations also allow scientists to explore the role of the ribosome in the translocation of membrane proteins (Lin et al. 2012). RNCs are an important component of co-translational targeting and translocation processes and therefore, understanding the role of ribosomes associated chaperone and other binding factors in translocation processes is essential. Cryo-EM technology has provided initial structural insights into the understanding of co-translational translocation (Frauenfeld et al. 2012).

2.5.1 Mechanisms to study co-translational protein folding

It was first proposed by Ito and Nakatogawa that the constricted part of the ribosome exit tunnel acts as a gate by interacting with nascent chains (Nakatogawa & Ito 2002). Previously, the ribosome tunnel was considered only a passive conduit for nascent polypeptide chains but recent Cryo EM studies (Villa et al. 2009) have revealed that the nascent chain makes a number of interactions with the ribosome exit tunnel, thus suggesting that these interactions could regulate pausing or further attenuate the rate of translational processing. Fluorescence resonance energy transfer (FRET) studies have also revealed that a transmembrane protein sequence within a nascent membrane protein starts folding into a compact conformation near the peptidyl transferase center (PTC) and remains folded as the sequence moves through a membrane bound ribosome into the translocon (Woolhead CA, McCormick PJB et al., 2004). Thus, data from past experimental trials on ribosome bound nascent chain studies have shown that proteins start folding in the ribosome tunnel. To further understand the process of co-translational protein folding on the ribosomal surface, a molecular biology approach has been identified which leads to the stalling of the translation machinery via a signal from a model secretory protein, SecM. The resulting translation-arrested, ribosome bound nascent chains (RNCs) of varying length generated using residues 150 to 166 of SecM (FSTPVWISAQGIRAGP) help in elucidating the interaction of translating protein with the surface of ribosome and also its interaction with chaperones like DnaK and TF. SecM is a leader peptide that induces transitional stalling in response to the

presence or absence of an effector molecule and in doing so it regulates the translation of a downstream gene (Tenson, T. & Ehrenberg, M., 2002). Other examples of well characterized of leader peptides that show translational stalling are TnaC and ErmC. Mutations within the leader peptide sequence or ribosome tunnel components can relieve the translational arrest (Nakatogawa & Ito 2002). Similarly, eukaryotes have also evolved stalling sequences. Table 2 below summarizes the well-studied stalling sequences in both prokaryotes and eukaryotes.

Table 2: Stalling sequences in both prokaryotes and eukaryotes

	Gene	Amino Acid Sequence	Refs
Prokaryote			
Bacteria	SecM	FSTPVWISQAQGIRAGP	(Nakatogawa et al. 2005)
	TnaC	MNILHICVTSKWFNIDNKIVDHRP	(Gong & Yanofsky 2002)
Eukaryote			
Virus	CMV	MQPLVLSAKKLSLLTCKYIPP	(Bhushan et al. 2010)
Fungi	AAP	MNGRPSVFTSQDYLSDHLWRALNA	(Spevak et al. 2010)
Mammals	AdoMedDC	MAGDIS	(Raney et al. 2002)

a. SecM mediated stalling

Since the SecM arrest motif tightly interacts with the ribosomal tunnel, the stability of the RNCs were found to increase when associated with nascent chains containing the stalling sequence. In a study by Nakatogawa and Ito (Nakatogawa & Ito 2002), it was shown that Pro166 of the arrest sequence is the element that causes arrest from within the ribosome during translation. There is a specific interaction of the SecM arrest sequence with the ribosomal components shown by mutations in both RNA (23S rRNA) and protein (L22) components of the ribosome that enabled completion of translation beyond Pro166. Nakatogawa and Ito in 2002 demonstrated that a SecM amino acid sequence produces a ribosome-bound protein and showed that the motif FXXXXWIXXXGIRAGP is critical for elongation arrest. In *E. coli*, the SecM sequence is found upstream from the SecA sequence. In between SecM and SecA is a hairpin structure. This hairpin structure prevents the translation of SecA. The translation of SecM initiates the alleviation of the hairpin structure (Nakatogawa et al. 2005). While SecM stalls the ribosome, SecA targets a signal recognition peptide, upstream from the SecM, to the SecA translocase. In the cell, the SecM protein can be released from the ribosome tunnel by SecA protein at the expense of ATP. The hydrophobic export signal of the SecM protein is exposed at the ribosome exit region and SecA interacts with it to release the arrest in vivo (Gumbart et al. 2012). The principle of SecM stalling has been compared to action of macrolide antibiotic (troleandomycin), which has shown to induce conformational changes in L22 (ribosomal protein) leading to occlusion of the ribosome tunnel (Berisio et al. 2003).

One of the classical methods of generating RNCs is by removing the stop codon from the gene construct by restriction enzyme digestion and PCR. Another method of RNC generation is by using the translational/transcriptional systems (IVT/T), which has several limitations (Sunohara et al. 2004). One of the major drawbacks of this system is that only small fractions of ribosome are actively involved in protein biogenesis and therefore additional purification steps are required to obtain homogenous samples. This results in low concentrations of stalled RNCs which are inadequate for carrying out various biophysical experiments. Another disadvantage of the system

is that it generates polysomes, which require another step of sucrose gradient centrifugation for separation (Qin & Fredrick 2013).

SecM technology can avoid many of these pitfalls (Hoffmann et. al., 2005). Thus, SecM used with a translational system has been utilized to stall various lengths of a polypeptide chains and shown that stalled nascent protein degrades quickly without Trigger factor.

This SecM technology has been shown to have advantages such that it allows the presence of the stop codon downstream of the arrest point, therefore *in vitro* transcription and translation can be performed without the necessity of generating truncated mRNA. Another major advantage of using SecM technology is that it enables the production of large quantities of homogenous and highly concentrated, stably-arrested nascent chain complexes (RNCs). To study co-translational protein folding on the ribosome, one of the most suitable methods is Nuclear Magnetic Resonance Spectroscopy (NMR), which provides high resolution structural information on flexible nascent chains exposed at ribosome. NMR requires a high concentration of stably arrested RNCs, which can be obtained through SecM technology (Rutkowska et al. 2009).

There are many factors that may affect stalling *in vivo*. The SsrA directs the stalled, nascent polypeptide to degradation. Sunohara et al. studied the impact of SsrA on protein degradation. They used SecM to achieve stalling. With active SsrA, there was a low level of stalling species in compared to the SsrA mutants. With this in mind, SsrA mutants are found to be more efficient for the production of stalled species. In the long term, SecM can be used to diagnose the structure of disease-related proteins. Researchers can fuse the stalling sequence with any particular protein to study misfolded proteins (Dobson 2004).

b. Antibiotic mediated generation of RNCs

There are several drugs that inhibit peptide bond synthesis during translation, such as cycloheximide in eukaryotes or chloramphenicol in prokaryotes, and are known to bring about stabilization of many or perhaps all cellular mRNAs (Lopez et al. 1998). This effect can also be mimicked in the absence of these translation inhibitors by mutations in the translational apparatus that slows or blocks ribosome movement (Caponigro and Parker 1996). Various currently used antibiotics target the bacterial ribosome apparatus. However, the mechanism by which growth arrest or cell death is mediated differs greatly for these various agents.

Chloramphenicol and puromycin target the ribosomal peptidyl transferase center (PTC) (Croons et al. 2008). Chloramphenicol acts by occupying the position of the amino acid attached to the A-site tRNA and prevents peptide bond formation. As a result, elongation of translating protein is arrested and the peptidyl-tRNA is trapped on the ribosome.

Thus, chloramphenicol stabilizes ribosome-nascent chain complexes. Puromycin mimics the amino-acylated end of the aa-tRNA and participates in peptide bond formation. Its non-hydrolyzable amide bond cannot be cleaved and as a result, the peptidyl-puromycin chain dissociates from the ribosome. Thus, puromycin leads to premature release of the nascent chain from ribosome and strips mRNAs from ribosomes. In prokaryotes, it is mostly assumed that translation inhibitors affect mRNA stability by altering the packing or activity of translating ribosomes, also known as a 'cis' effect (Azzam & Algranati 1973). This theory is supported by two strong pieces of evidence. First, drugs affect mRNA stability in different ways depending on how they inhibit translation. Thus, whereas chloramphenicol, tetracycline and fusidic acid stall ribosomes on mRNAs yielding stabilization, puromycin and kasugamycin, conversely, strip mRNAs off ribosomes leading to destabilization. Second, ribosomes can also be stalled or pulled off mRNA in the absence of inhibitors and in these cases, the changes affect mRNA stability like the corresponding classes of inhibitors (Vince et al. 1975). Since these antibiotics bring about translation inhibition by distinct mechanisms, they are exploited in studies of stalled ribosome nascent polypeptide complexes as explained in Figure 11.

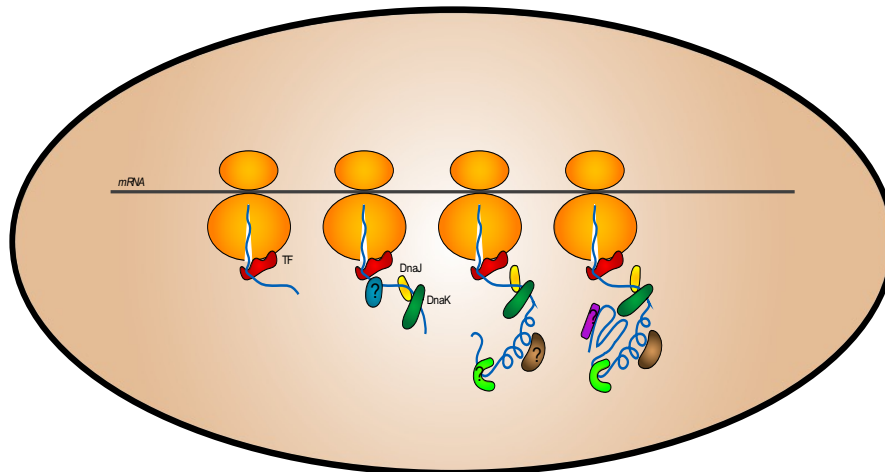
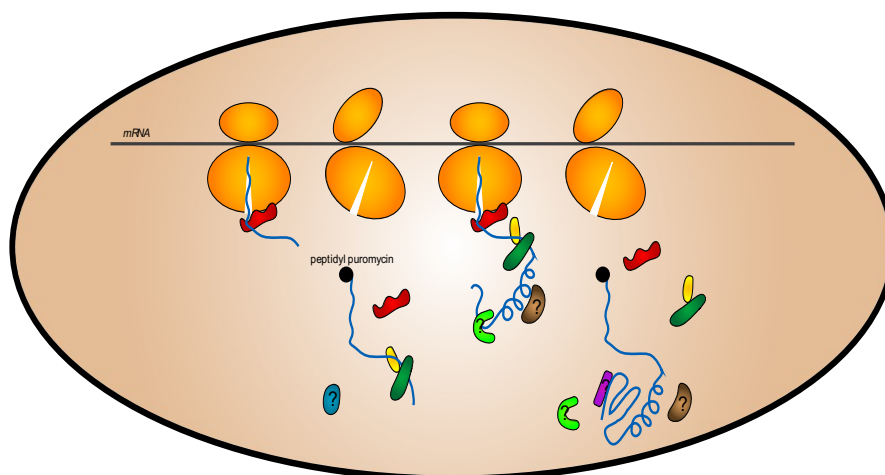
A. Chloramphenicol (CAM) induced translational stalling**B. Puromycin (PURO) induced premature release of Nascent chains**

Figure 11: A. Chloramphenicol mediated stalling of nascent chains on the ribosome by stabilization of mRNAs. This subsequently leads to the stabilization of various binding factors and chaperones that assist the nascent chains in folding. **B.** Puromycin mediated premature nascent chain release results in the production of vacant ribosomal species devoid of nascent chains. Thus binding partners and chaperones that stabilise the nascent chain on the ribosome are also pulled off

2.6 Aim of the project

Nascent polypeptide chains emerging from the ribosome exit tunnel interact with numerous chaperones, including the well studied TF, DnaK/J and GroEL, to attain their biologically active states. These stable and transient interactions have been studied in isolated systems, such as cell-free translation mixtures (Evans et al. 2005), in order to decipher the role of chaperones in assisting a nascent polypeptide in folding to the correct native conformation. The concerted activities of these factors in maintaining the integrity of nascent polypeptide on the ribosome and in folding the polypeptides into three-dimensional conformations have not been studied in detail. Specifically, studies in the natural, highly crowded environment of the cell need further investigations.

Here, we set out to analyze the dynamic chaperone repertoire that interacts with nascent chains translating on the ribosome in a bacterial system. To investigate this, we generated ribosome nascent chain complexes (RNCs) by utilizing antibiotics, which stall or release nascent chains from the ribosome and enlisted the factors that interact with ribosomes in a nascent chain-dependent manner. Since interactions of various factors with RNCs are expected to be dynamic, we employed *in vivo*-crosslinking to stabilize the RNC interactome (Becker et al. 2013).

The RNCs were isolated via co-immunoprecipitation utilizing ribosomes carrying a FLAG-tag at L17 protein of 50S ribosome subunit, integrated into the genome of a bacterial strain. To obtain real-time snapshots of RNC interactors, we performed stable isotope labeling of amino acids in cell culture (SILAC) coupled with mass spectrometry (MS). This study will enhance the understanding of interactions of different binding factors with nascent chains bound to ribosomes and thus provide insight into the process of co-translational protein folding.

3. Materials and Methods

3.1 Materials

Unless indicated otherwise, chemicals and reagents used were of pro analysis (p.A.), ACS quality or comparable assay grade and were mostly purchased from Sigma-Aldrich (St. Louis, USA).

Chemicals

Suppliers	Chemicals
Biomol GmbH (Hamburg, Germany)	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (Hepes)
BioRad (Hercules, USA)	Ethidium bromide Bradford Protein Assay
BD (Franklin Lakes, USA)	Bacto™ Tryptone Bacto™ Yeast Extract
Cambridge Isotope Laboratories (MA,USA)	arginine- ¹³ C ₆ , lysine-4,4,5,5-d ₄ , arginine- ¹³ C ₆ , ¹⁵ N ₄ and lysine- ¹³ C ₆ , ¹⁵ N ₂
Gibco, Life technologies	Puromycin
Invitrogen	Nu-page 4-12% Bis Tris gel (10 wells) Nu-page 4-12% Bis Tris gel (15 wells)

Promega (Wisconsin, USA)	Trypsin ProteaseMAX surfactant
Roche (Basel, Switzerland)	Complete protease inhibitor EDTA-free protease inhibitor cocktail DNase I, RNase free water
Thermo Fisher Scientific (Waltham, USA)	Dithiothreitol (DTT) Isopropyl β -D-1-thiogalactopyranoside (IPTG)
Serva Electrophoresis GmbH (Heidelberg, Germany)	Phenylmethylsulfonyl fluoride (PMSF) Acrylamide/Bis Solution, 37.5:1 (30% w/v) Dodecylsulfate sodium salt in pellets
Schleicher & Schuell (Dassel, Germany)	Protran Nitrocellulose Transfer Membrane
Sigma-Aldrich (St. Louis, USA)	Chloamphenicol Tetracycline hydrochloride Magnesium acetate tetrahydrate Potassium acetate Apyrase

	<p>Dithiobis(succinimidyl propionate)</p> <p>N,N,N',N'-Tetramethylethylenediamine (TEMED)</p> <p>ANTI-FLAG® M2 Affinity Gel</p> <p>Anti-FLAG® M2 Magnetic Beads</p>
<p>VWR (Radnor, USA)</p>	<p>Lysozyme from chicken egg white</p> <p>2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris)</p> <p>2-mercaptoethanol</p> <p>3',3'',5',5''-Tetrabromophenolsulfonephthalein (Bromphenol Blue)</p> <p>Triton X-100</p> <p>Hydrochloric acid, 37%</p> <p>Sodium chloride (NaCl)</p> <p>Methanol</p> <p>Dimethylsulfoxide (DMSO)</p>

3.2 Buffers and media

Buffers were prepared with deionized water (electrical resistance 18.2 M·cm).

Antibiotic solutions (1000×):	34mg/ml chloramphenicol 5 mg/ml tetracycline
Cell resuspending buffer:	Hepes 50mM 0.5M Sucrose
Lysis buffer (HMN)	50mM Hepes 50mM K(oAc) 20mM Mg(oAC) ₂ 1× EDTA free Complete Protease Inhibitor 1mM PMSF
Washing buffer	50mM Hepes 50mM K(oAC) 20mM Mg(oAC) ₂ 1× EDTA free Complete Protease Inhibitor 1mM PMSF 0.1% Triton
FLAG-tag pull-down elution buffer	100 mM Tris-HCl pH 6.8 4 % (w/v) SDS 20 % (v/v) glycerol

SDS-loading buffer (2×)	100 mM Tris-HCl pH 6.8 4 % (w/v) SDS 200 mM DTT 20 % (v/v) glycerol 0.2 % (w/v) bromphenolblue
TBST (10×):	200 mM Tris-HCl pH 7.5 1.37 M NaCl 1 % (v/v) Tween 20
SDS-running buffer (10×)	250 mM Tris 2.5 M glycine 1 % (w/v) SDS
SDS Stripping buffer:	2% (w/v) SDS 62.5 mM Tris-HCl, pH 6.8 100 mM β-mercaptoethanol
TBS (10×)	500 mM Tris-HCl pH 8.0 1.37 M NaCl 27 mM KCl
Western Blot buffer:	50 mM Tris 20 % (v/v) methanol 192 mM Glycine

Ponceau S stain:	0.2 % (w/v) Ponceau S 3 % (v/v) trichloro acetic acid
Blocking buffer	5% (w/v) Milk powder in TBST
LB medium:	10 g/l Bacto Tryptone 5 g/l Bacto Yeast Extract 10 g/l NaCl
Solid LB medium:	10 g/l Bacto Tryptone 5 g/l Bacto Yeast Extract 10 g/l NaCl 15 g/l Agar-Agar
M63 medium	2 g/l (NH ₄) ₂ SO ₄ , 13.6 g/l KH ₂ PO ₄ , 0.5 mg/l FeSO ₄ x 7 H ₂ O. Before use, 1 ml MgSO ₄ (1 M), 10 ml glucose (20 % w/v), L-amino acid mix (to 0.5 mM of each amino acid final) were added per 1 l medium and filter sterilized

SILAC medium:	To prepare L, M, and H media, the respective amino acids in M63 medium were exchanged by respective labelled amino acids, for L: Arg0 and Lys0 (arginine and lysine, Sigma), for M: Arg6 and Lys4 (arginine-13C6 and lysine-4,4,5,5-d4, Cambridge Isotope Laboratories), for H: Arg10 and Lys8 (arginine-13C6, 15N4 and lysine-13C6, 15N2, Cambridge Isotope Laboratories).
ECL solution I:	100 mM Tris-HCl pH 8.8 2.5 mM luminol (3-aminophthalhydrazid; stock: 250 mM in DMSO, dark, 4 °C) 0.4 mM p-coumaric acid (stock: 90 mM in DMSO, dark, 4 °C)
ECL solution II	100 mM Tris-HCl pH 8.5 5.4 mM H ₂ O ₂ PBS (10×): 92 mM Na ₂ HPO ₄ ·2H ₂ O 1.5 M NaCl adjust pH 7.2 with NaOH 16 mM NaH ₂ PO ₄ ·H ₂ O
Coomassie destaining solution:	10 % (v/v) ethanol 10 % (v/v) acetic acid

Coomassie staining solution:	0.1 % (w/v) Serva Coomassie Blue R250 40 % (v/v) ethanol 10 % (v/v) acetic acid
DNA-loading buffer (6×):	0.25 % (w/v) bromophenol blue 0.25 % (w/v) xylene cyanol FF
TAE buffer (50×):	42 M Tris-acetate 50 mM EDTA pH 8.00 % (w/v) sucrose in H ₂ O
Digestion buffer:	50 mM NH ₄ HCO ₃
Digestion destaining buffer	50 mM NH ₄ HCO ₃ 50% ethanol
Extraction buffer	3% trifluoroacetic acid 30% acetonitrile
Denaturation buffer	10 mM HEPES pH8.0 6 M urea 2 M thiourea 1% n -octylglucoside (w/v)
C18 Equilibration solution	0.1% TFA

C18 Elution buffer	0.1% TFA, 80% acetonitrile.
Digestion alkylation buffer	55 mM iodoacetamide 50 mM NH ₄ HCO ₃
UA	8 M urea (Sigma, U5128) in 0.1 M Tris/HCl pH 8.5. Prepare 1 ml per 1 sample.
IAA solution	0.05 M iodoacetamide in UA. Prepare 0.1 ml per 1 sample
ABC	0.05M NH ₄ HCO ₃ in water. Prepare 0.25 ml per 1 sample

3.3 Instruments

Supplier	Instrument
Beckman Coulter GmbH (Krefeld, Germany):	Centrifuge GS-6R DU640 UV/VIS Spectrophotometer Biometra (Göttingen, Germany): T3 Thermocycler
Bio-Rad (München, Germany):	Mini Protean® Tetra Cell (SDS gel) Mini Protean™ II (WB) Mini Trans-Blot® Module Power supply Power Pac Basic 300

<p>Eppendorf (Hamburg, Germany):</p>	<p>Centrifuge 5415C, 5415R Thermomixer comfort Pipettes 1.5 ml tubes</p>
<p>Fisher Scientific (Schwerte, Germany):</p>	<p>2 ml tubes</p>
<p>Accumet Basic pH meter Forma Scientific (Waltham, USA):</p>	<p>CO2 Water Jacketed Incubator</p>
<p>Fuji/Raytest (Straubenhardt, Germany):</p>	<p>Fuji LAS3000 Gel imaging software AIDA (version 4.15.025) Shaker Duomax 1030</p>
<p>Merck Milipore</p>	<p>Centrifugal filter units</p>
<p>Josef Peske GmbH & Co KG (Karlsruhe, Germany):</p>	<p>1.5 ml tubes 2 ml tubes</p>

LG (Seoul, South Korea):	Microwave
Wavedom Melter Toledo (Gießen, Germany):	Balances (AB265-S, FACT)
Millipore (Schwalbach, Germany):	Millex-SV syringe filter unit, 0.22 µm Millipore Milli Q Plus PF water purification system Scepter™ Handheld Automated Cell Counter Gel documentation system Steritop GP Filter Units, pore size 0.22 µM
Heraeus Instruments (Hanau, Germany):	Cell culture hood Microbiological Incubator Function Line B12
Melter Toledo (Gießen, Germany):	Melter Toledo (Gießen, Germany):
Empore TM	Solid phase extraction disk
Qiagen (Hilden, Germany):	QIAshredder columns
Sarstedt (Nümbrecht, Germany):	15 ml tubes 50 ml tubes Filter tips Pipettes

	Serological pipette, 1 ml Tube, 5ml, 75x12 mm
Greiner (Frickenhausen, Germany)	Heatable magnetic stirrer
Schott (Mainz, Germany):	Laboratory glass ware
Scientific Industries (Bohemia, USA):	Vortex-Genie® 2
Sorensen™ BioScience, Inc. (Salt Lake City, USA):	Safe Seal Microcentrifuge Tubes (0.65 ml, 1.7 ml)
Störktronic (Stuttgart, Germany):	Heating block
Thermo Scientific (Waltham, USA):	LTQ-Orbitrap mass spectrometer Q-Exactive mass spectrometer EASY-nLC 1000 Microscopy slides Nanodrop 1000 Spectrophotometer Orbital shaker Waters
Whatman™ (Dassel, Germany)	Nitrocellulose membrane
Fermentas (St. Leon-Rot, Germany)	6x DNA Loading Dye GeneRuler™ 100bp DNA Ladder GeneRuler™ 1kb DNA Ladder PageRuler™ Prestained

Abimed (Langenfeld, Germany):	Gilson Pipetman 2, 10, 20, 100, 200 and 1000
Amersham Pharmacia Biotech (Freiburg, Germany):	EPS 300 electrophoresis power supply FPLC chromatography systems ÄKTA Explorer 100 chromatography system HiTrap Chelating column, 5ml HiPrep Desalting column NAP-5 Sephadex G25 desalting columns
Amersham Pharmacia Biotech (Freiburg, Germany):	Concentration devices (Centriprep, Centricon)
Avestin (Mannheim, Germany):	EmulsiFlex C5 homogenizer
Sigma Aldrich	EZview™ Red Anti-FLAG M2 Affinity Gel

3.4 Methods

3.4.1 Biochemical methods

a. Preparation of cell extracts / Soft lysis

To prepare whole cell extracts for western blotting and immunoprecipitation, cells were grown to OD_{600} : 0.7 and then pelleted at 3,000 g for 10 min at 4°C. To prepare soluble cell fractions, cells were re-suspended in cell resuspension buffer followed by preparation of spheroplasts using 0.5mg/ml concentration of lysozyme treatment for 7mins. Cells were then pelleted at 1090g, suspended in cell lysis buffer (HMK) with 0.1% triton. After that, crude cell lysate was centrifuged at 20,000 g for 10 min at 4°C. The supernatant fraction was used for protein quantification analysis.

b. Generation of various RNC species using Chloramphenicol and Puromycin

Antibiotic mediated translation inhibition is used in this project for the generation of stalled nascent chain and released nascent chain ribosomes. This treatment is administrated at two steps during lysis. First, antibiotics are added after the preparation of spheroplasts when the cells reach and OD_{600} of ~0.6-0.7. Second, during the preparation of spheroplasts (as described above), pellets are dissolved in cell lysis buffer without detergent and with 150 μ g/ μ l Chloramphenicol or 5mM Puromycin. The treatment is performed at 37°C on the thermomixer for 10mins. Puromycin is most effective at 37°C and therefore, both antibiotic treatments are carried out at 37°C. Final lysis is performed as described earlier.

c. SDS-PAGE

For cell extract analysis, samples were separated by discontinuous Tris-glycine sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) according to their electrophoretic mobility. The compositions of the resolving and stacking gels are shown in Table 3.

Table 3: Compositions of resolving and stacking gels.

Solution component	12% resolving gel	15% resolving gel	5% stacking gel
H ₂ O	5.9 ml	4.9 ml	5.5 ml
30% Acrylamide/ 0.8% bisacrylamide	8 ml	10 ml	1.3 ml
0.5 M Tris (pH 6.8)	-	-	1ml
1.5 M Tris (pH 8.8)	3.8 ml	3.8 ml	-
10% SDS	150 μ l	150 μ l	80 μ l
10% APS	150 μ l	150 μ l	80 μ l
TEMED	6 μ l	6 μ l	80 μ l

d. Western blotting

If proteins were to be detected by specific antibodies after native PAGE or SDS-PAGE, they were transferred electrophoretically to a Protran™ nitrocellulose membrane (Whatman). Transfer of proteins from SDS gels was done using a Mini Trans-Blot Cell (Bio-Rad) filled with transfer buffer. Proteins from SDS and 4% to 12% gradient gels were transferred by semi-dry transfer for 90mins using 0.8 mA per cm² of gel. After transfer, the nitrocellulose membranes were blocked by incubation in blocking buffer, generally for approximately 30 min at room temperature. Primary antibody was diluted into blocking buffer prior to incubation with the nitrocellulose membrane. The incubation of the membrane with the respective antibodies was done in a rotating chamber to ensure even coverage of the membrane with the antibody solution. Incubation with the primary antibody was done for approximately 16 h at 4°C or alternatively for approximately 2 h at room temperature. The nitrocellulose membrane was washed three times with TBS-T after incubations with primary antibody followed by incubation in secondary antibody diluted in 1X TBS buffer. Typically, the membrane was incubated with secondary antibody for approximately 2 h at

room temperature. Prior to detection, the membrane was washed three times with TBS-T. Secondary antibodies coupled to horseradish peroxidase were used. For detection, the membrane was incubated with the USB Rodeo™ ECL Detection Reagents (Affymetrix). The emitted light was detected using the LAS-3000 imaging system (Fujifilm).

Primary antibodies used were rabbit anti-L23 serum (from Andreas Kolbeck, MPI for Biochemistry, Martinsried, Germany; diluted 1:5,000) and the monoclonal anti-FLAG M2 antibody (Sigma; diluted 1:2000). If the anti-FLAG M2 antibody was used, TBS was used instead of TBST, omitting the Tween 20 from the buffer. HRP-coupled goat anti-rabbit IgG (Transduction Laboratories) or goat anti-mouse IgG (Sigma) were used as secondary antibodies at a dilution of 1:2,000.

e. Ribosome tagging

In order to introduce a FLAG-His6 (FH) epitope tag into the large ribosomal subunit of *E. coli* MC4100, a recombination system for gene deletions in *E. coli* was utilized (Datsenko & Wanner 2000). This system is based on the phage λ Red recombinase system and consists of the genes γ , β and *exo*. These genes were cloned into a plasmid pKD46 under the control of an arabinose inducible promoter to allow their regulated expression in *E. coli* (plasmid pKD46, Datsenko & Wanner 2000). To generate *rpl*-FH-tetR cassettes, a synthetic oligonucleotide encoding the FH epitope tag (FHfw) was mixed with a complementary oligonucleotide (FHrv) at equimolar concentrations of 60 μ M each. Christian Kaiser, a former graduate member of our laboratory designed the oligonucleotides in such a way that in addition to the complementary central sequence, they contained overhangs that resulted in the formation of cohesive ends after annealing. For annealing, the mixed oligonucleotides were heated to 96 °C for 5 min and allowed to cool down to room temperature. Through this procedure, a double stranded oligonucleotide was generated encoding the FH tag with cohesive ends at the 5'- and the 3'-termini compatible with cohesive ends resulting from Not I and Xho I restriction digestion, respectively. pET22b (+) was then digested with Not I and Xho I endonucleases and dephosphorylated with Shrimp Alkaline

Phosphatase (SAP). The annealed oligonucleotide encoding the FH epitope tag was ligated into this backbone as described above, yielding pFH plasmid. The open reading frame (ORF) of the tetracycline resistance gene (*tetR*) was amplified by PCR from plasmid pBR322. Since the resistance cassette was intended to be inserted into existing operons, a promoter for *tetR* -gene expression was not considered necessary. The PCR product was gel purified, digested with Hind III and Nco I endonucleases and ligated into plasmid pFH that had been digested with the same enzymes and dephosphorylated with SAP. The resulting plasmid is referred to as pFH-tet. Stretches of ~500 bp downstream of the *rpl*-ORF of interest were amplified by PCR from *E. coli* K12 genomic DNA using primer pairs *rplP3* and *rplP4*, where *rpl* is L13, L17, L20 or L22. The PCR products were digested and inserted into the plasmid pFH-tet downstream of the *tetR* -ORF via Nco I and Xba I endonuclease sites, yielding plasmids pFH-tet-13, pFH-tet-17, pFH-tet-20 and pFH-tet-22. Positive clones were verified by DNA sequencing. Using primer pairs *rplP1* and *rplP2*, the ORFs encoding L13, L17, L20 and L22 together with ~500 bp of upstream flanking sequences were amplified by PCR from *E. coli* K12 genomic DNA. Products for L13, L17 and L20 were digested with Xba I and Eco RI, the product for L22 was digested with Xba I and Sac I, and the digested products were ligated via the same sites into pBAD18 (Guzman et al. 1995). No positive clones could be obtained for L22 protein, indicating that there was interference with survival of the host cell. The resulting plasmids were pL13, pL17 and pL20. The inserts of plasmids pFH-tet-13, pFH-tet-17 and pFH-tet-20 were released by restriction digestion with Xho I and Xba I endonucleases, gel-purified and ligated through the same sites into pL13, pL17 and pL20, respectively. The resulting plasmids are referred to as pL13FH, pL17FH and pL20FH. *E. coli* cells transformed with pL13FH showed impaired growth rates.

Chromosomal integration of rpl-FH-tetR cassettes

The insert of plasmid pL17FH was amplified by PCR using the primer-pair L17P1 – L17P4. The template was digested with Dpn I endonuclease, and the PCR product was gel-purified twice and eluted in 50 µl of sterile water. *E. coli* M4100 cells were transformed with pKD46 and plated on LBamp-agar. Transformants were allowed to form colonies overnight at 30 °C. A single colony

was used to inoculate 50 ml LBamp. A culture was grown at 30 °C to an OD₆₀₀ of ~0.1 before arabinose was added to a final concentration of 0.02% (w/v). The culture was further grown at 30 °C to an OD₆₀₀ of 0.8, arabinose was added to a final concentration of 0.2% and the culture was grown for another 10 min. Cells from 5 ml of this culture were harvested by centrifugation and rendered electrocompetent as described above. The cell pellet was re-suspended in 50 µl of the purified PCR product and incubated 15 min on ice. Cells were subjected to electroporation and re-suspended in SOC medium. After incubation at 37 °C with shaking at 300 rpm for 1 h, one third of the cells was plated on LBtet-agar and incubated at 37 °C overnight. A single colony was picked and re-suspended in 50 µl LB medium. 10 µl of the suspension was plated on either LBamp- or LBtet-agar. After overnight incubation at 37 °C, colonies appeared only on LBtet-, not on LBamp-agar and a single colony was used to inoculate 5 ml of LBtet medium. The culture was grown overnight at 37 °C to saturation. A glycerol stock of this culture was prepared by mixing 600 µl of the culture with 300 µl sterile glycerol. The stock was flash-frozen in liquid nitrogen and then stored at –80 °C. The resulting strain is referred to as CK1 (*E. coli* MC4100 rplQ::rplQ-FH-tetR). Growth was observed upon selection for tetracycline resistance and cells failed to grow on ampicillin selective medium. (Adapted from Christian Kaiser's thesis)

f. Sedimentation analysis of *E. coli* ribosomes

To confirm the incorporation of FH-tagged ribosomal proteins into the 70S ribosome, ribosomes from lysates of different *E. coli* strains were isolated by sedimentation through a sucrose cushion. Ribosomes were further analyzed by separation on a 10–40% sucrose gradient by the polysome profiling method: A primary culture is incubated overnight with CK1 and C strain. ~30 mL of secondary culture is inoculated with 1% of primary inoculum. 150 µL of 5 mg/mL of tetracycline stock is added to reach a final tetracycline concentration of 25 µg/mL and the culture is incubated at 37 °C with 350 rpm in a shaker incubator. The secondary culture takes approximately 4 hours to grow and when OD₆₀₀ ~0.7 is reached the sample is harvested.

g. Preparation of *E. coli* lysates for analytical ribosome isolations

For the preparation of lysates, 30ml cultures of different *E. coli* strains were grown in LB medium or later isotopically labelled media at 37 °C to an OD₆₀₀ of 0.6 – 0.7. For preparation of ribosomes enriched in ribosome-nascent chain complexes, the culture was supplemented with 150 µg/ml chloramphenicol before it was poured into flash frozen 50ml falcon tubes. Further cells were harvested by centrifugation at 3000g and all subsequent steps were performed at 4 °C or on ice. Cell pellet is suspended in 750 µl of cell suspension buffer and lysozyme was added to a final concentration of 500 µg/µl for 7 minutes for the preparation of the spheroplasts. The samples were pelleted in ultracentrifuge at 1090g for 8 minutes. The pellet was dissolved in 1ml lysis buffer without Triton and for the preparation of different species of nascent chains antibiotic administration was performed at 37°C for 10 minutes. Final lysis step included addition of 0.1% Triton and 5mins incubation in the rotating wheel in the cold room followed by clearing of the lysate with high speed spin on cooling centrifuge. The soluble fraction of the protein was harvested at 20,000g for 10 minutes and the top three fourths of the cleared lysate were recovered. The protein concentration estimation was done using standard Bradford quantification.

h. Protein quantification

Protein concentrations of the soluble *E.coli* lysates from CK1 strain were measured calorimetrically using the Bio-Rad protein assay reagent according to the manufacturer's recommendations (Bradford 1976). 5µl of a 1:9 dilution of the protein extracts were mixed with 795 µl ddH₂O and 200 µl Bradford solution. The mixture was mixed briefly and the absorption was measured at 595 nm using a spectrophotometer (Beckman).

i. Ribosome pull-downs using anti-FLAG agarose

For the pull-down of 50S ribosome subunit tagged with the FLAG peptide sequence DYKDDDDK at the N terminal of L17 protein, the agarose-based anti-FLAG M2 Affinity Gel (Sigma) was used. The matrix was equilibrated in Buffer HMN and washed two times. 60µl of beads were incubated

next with 300µg of cleared lysate for 60 min at 4 °C with constant gentle agitation on rotating wheel at 4°C beads. The matrix was sedimented at 2,000 g for 1 min and washed three times with 500 µl of lysis buffer containing 0.1% triton. Bound material was eluted with pre-heated loading dye without bromophenol blue and DTT.

j. Cloning of Luciferase Sec M stalled different chain length expression constructs

c-myc tagged Luciferase SecM GFP gene cloned in a pBAD33 vector under the arabinose promotor was generated in our laboratory. The GFP gene was cloned out of the vector by introducing NotI restriction site 5' end of the GFP gene. Different lengths of Luciferase SecM constructs were designed by introducing a BglII restriction site at 3' end of luciferase gene on basis of hydrophobicity plot. This is done by use of following reverse primers 77 Rev, 171 Rev and 330 Rev and SecM forward primer (as listed in appendix). The PCR amplified product is digested with BglII restriction enzyme and further ligated. The specificity of the reaction and the correct size of the products were confirmed by agarose gel electrophoresis. All constructs were verified by DNA sequencing.

k. Immunoprecipitation of *c-myc* tagged Sec M stalled luciferase constructs.

The immunoprecipitation experiment was optimized using µMAC magnetic beads coupled with anti-myc antibody. The experiment is carried out by first growing the primary culture at 37°C. The secondary was grown at 30°C and induced with 0.2% arabinose. The cells were lysed using lysozyme (0.5mg/ml) and 0.1% triton. Mild lysis is done in order to prevent the loss of transient interactions of binding factors with nascent chains on ribosomes. The lysate concentration is calculated using Bradford assay. 500µg of protein is loaded on 100µl of magnetic beads and eluted with pre-heated loading buffer.

3.4.2 Proteomics method

a. Sample preparation for mass spectrometry: Filter Aided Sample Preparation (FASP)

The protocol termed *Filter Aided Sample Prep* (FASP) (add literature) is used to generate tryptic peptides from crude lysates for LC-MS/MS analysis. Aliquots containing 12 μ l of each immunoprecipitated protein extract were mixed with 200 μ l of UA buffer (8.0 M urea in 0.1M Tris/HCl, pH 8.5) in the centrifugal ultrafiltration units with a nominal molecular weight cut-off of 10,000, and then centrifuged at 14,000 g, at 20 °C, for 15 min. The flow-through was discarded. 200 μ l of UA was pipetted into the filter and centrifuged again. Then, 200 μ l of 0.02M DTT was added into the filter. After incubation at room temperature for 45 min, the filter unit was centrifuged. Subsequently, 200 μ l 0.05 M IAA solution in UA was added and the mixtures were incubated in the dark for 30 min. Filters were washed twice with 200 μ l UA followed by another two washes with 0.05 M Tris/HCl (pH 8.5). Proteins were digested overnight with the addition of 50 μ l 0.05M Tris/HCl (pH 8.5) with trypsin (enzyme to protein ratio 1:100(m/m)). After incubation in a wet chamber at 37°C overnight, the tryptic peptides were collected in new tubes by centrifuging the filter units at 14,000 x g for 10 min. 40 μ l 0.05M Tris/HCl was added, and the filter was centrifuged again. The flow-through was combined, acidified with 50 μ l of 25% CF₃COOH, fractionated by SCX into six fractions and desalted using the C18 stage-tip method (Rappsilber et al. 2007). The samples were dried in a vacuum centrifuge concentrator and re-suspended in 10 μ l of 5% FA solution for LC-MS/MS analysis.

b. LC-MS/MS Analysis

Chromatographic peptide separation was performed on a reverse phase column (75 μ m inner diameters, New Objective) packed with Reprosil-Pur 1.9 μ m C18 material (Dr. Maisch). Column temperature was maintained at 50°C by a column oven (Sonation). The peptide mixtures were analyzed by a Q-Exactive mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific) via a nanoelectrospray interface. The samples were loaded in buffer A (0.1% formic acid) and eluted with a 120 min gradient of 5%-60% buffer B (80% acetonitrile and 0.1% Formic acid) at a flowrate of 250 nL/min. The Q-Exactive was operated in data-dependent mode with survey acquired at a mass range of 300 to 1750 m/z and a resolution setting of 70,000 (FWHM). Up to the ten most abundant isotope patterns from the survey scan

were selected and fragmented by higher energy collisional dissociation. MS/MS spectra were acquired with a resolution of 17,500 (FWHM), maximum injection times of 120 ms, and a target value of 2e5 charges.

c. Data analysis

Raw MS data were analyzed with MaxQuant version 1.4.1.4 (Cox and Mann, 2008) with the false discovery rate of 0.01 for both peptides and proteins. MS/MS spectra were searched against *E.Coli* K12 proteins in Uniprot. For the Andromeda search, trypsin allowing for cleavage N-terminal to proline was chosen as enzyme specificity. Arg6 and Lys4 were selected for medium labels while Arg10, Lys8 were selected for heavy labels. Cysteine carbamidomethyl (C) was set as fixed modifications, while protein N-terminal acetylation and methionine oxidation were selected as variable modifications. Maximally two missed cleavages and three labels were allowed. Initial mass tolerance of 7ppm for precursor ions and 20 ppm mass accuracy was applied for HCD MS/MS spectra. Proteins and peptides (minimal seven amino acids) were identified using a target-decoy approach with a reversed database. The “re-quantify” option was enabled for protein quantification. To correct the ribosomal pull-down efficiency due to the different antibiotic treatment, the SILAC ratios were normalized based on the average ratios of all the quantified ribosomal 50S proteins. SILAC based control experiment was performed to establish cut-off value for significantly enriched and diminished proteins. This is performed by treating the light and heavy isotope labeled CK1 cells equally with CAM antibiotic. Subsequently, immunoprecipitated RNCs from different labeled samples were mixed and analyzed by mass spectrometry. All the quantified protein in 2 of 3 experiments were ranked in decreasing order of SILAC ratio for selecting significantly enriched proteins and increasing order for selecting significantly diminished proteins. A ratio cut-off of 1.8 in 2 of 3 experiments resulted in an error rate of 1 % and was applied to select significantly enriched proteins. Similarly, a ratio cut-off of 0.45 in 2 of 3 experiments resulted in a error rate of 1 % was applied to significantly diminished protein. Functional annotation enrichment was performed using DAVID (Huang et al. 2009).

- d. Determination of the factors associated with FLAG-tagged ribosome (CK1 strain) versus the background wild type non-FLAG-tagged cells (C strain)

Three sets of biological replicates of FLAG-tag pull-down SILAC experiments, were performed with the following designs: double-SILAC, CK1 strain (H) versus C strain(L) (3h of growth for C strain and 4 hours for CK1

strain). RNC interactors were determined from the experiment (i) by selecting the pulled-down proteins enriched a minimum of 1.8-fold over the background in at least two of the biological replicates. All protein ratios were normalized by the average of all detected large subunit ribosomal protein ratios from CK1 strain. In the RNC interactome experiment CAM and PURO were used to determine changes in the RNC interactome. Three sets of biological replicates of FLAG-tag pull-down SILAC experiments, were performed with the following designs: triple SILAC, CAM treated CK1 strain (H) versus PURO treated CK1 strain(M) versus Untreated CK1 strain (L) (4 hours growth for CK1 strain).

4. Results

4.1 Establishment of a rapid pull-down method to explore ribosome nascent chain (RNC) interactome in *E.coli*

In addition to TF, other chaperones are also involved in assisting nascent proteins towards reaching their native active conformation in bacteria. A functional overlap between TF and the DnaK/DnaJ/GrpE system has already been described (Nishihara et al. 2000), but additional groups of unexplored factors acting concertedly during early stages, possibly at the stage of exit of a nascent polypeptide chain from the ribosome, need to be identified.

The classical way of isolating ribosomes and its associated factors is sedimentation of these large particles, which allows their enrichment by successive cycles of centrifugation through a dense media such as a gradient of sucrose solutions. Recent studies have given structural insights into the dynamic nature of nascent chains emerging out of ribosomes by applying sensitivity optimized NMR technology on SecM-stalled RNCs isolated using metal affinity chromatography and sucrose density gradient centrifugation method. Though this method gives structural snapshots of RNC, it still shows several drawbacks such as poor separation of 70S ribosomes from contaminants during sucrose gradient, low yields of RNCs due to poor expression of stalled construct or poor recovery during preparative steps, loss of the integrity of RNC samples due to proteolysis or dissociation during handling and lower resolution of RNCs (Cassaignau et al. 2016). However, ribosome-associated factors may not co-migrate with ribosomes through density gradients because they bind with relatively low affinities and show transient interactions with the RNC complex.

To overcome the challenges of loss of RNC integrity and ensure a global analysis of RNC interactome, an alternative approach for ribosome purification circumventing this problem was developed in our studies by optimizing a rapid pull-down method. Our method focuses on ensuring the stability of RNC complex by developing a gentle lysis protocol.

Ribosomal protein L17 has solvent exposed carboxyl terminal, and is at a considerable distance from the TF docking site on protein L23 as shown in Figure 12. Thus, an affinity tag was genetically introduced into selected L17 large subunit ribosomal protein as described in method section, to allow for the efficient isolation of ribosomes and their associated factors. Applied to different strain backgrounds, this technique proved useful to identify factors that either act in concert with ribosome-associated proteins, such as TF or DnaK/J, or compensate for the loss of function in their absence.

We chose the FLAG-His6-tag (FH-tag) as an affinity tag. Due to its small size (less than 2.5 kDa, including a flexible linker), compared to larger epitopes or proteins such as streptavidin, MBP or GST, it seemed unlikely that this tag would interfere with the ribosome structure or function. The L17-tagged *E. coli* strain was developed by a former graduate member, Christian Kaiser, of our laboratory.

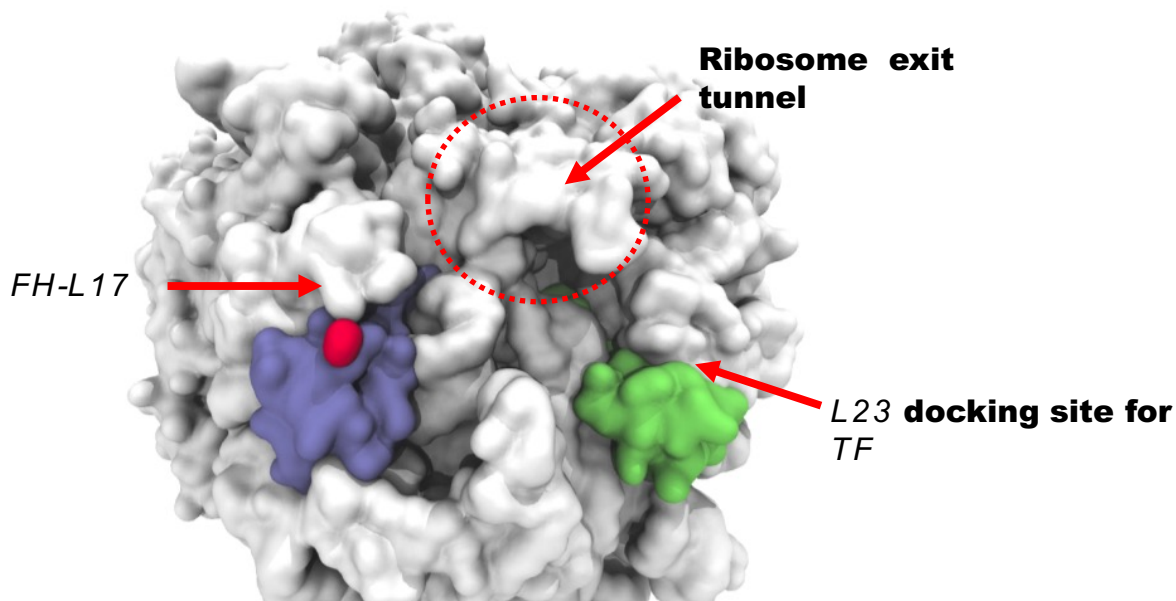


Figure 12: Surface representation of the *E. coli* 50S ribosomal subunit. Ribosomal RNA and most proteins are shown in grey. Ribosomal proteins L17 and L23 are colored purple and green, respectively. The carboxyl-terminal residues of L17 protein are colored red to indicate the position for placement of the FH-

tag. Protein L23 shown in green is the docking site for TF chaperone and is in close proximity to the ribosomal exit tunnel (circled by red dots). The model was generated with VMD .1.9.1 and is based on the coordinates from a previous study (Vila-Sanjurjo et al., 2004), and the image was rendered with Tachyon.

Through chromosomal integration of the FH-tag with the tetracycline-resistance gene for the ribosomal protein L17, an *E. coli* strain with the genotype MC4100 rplQ:: rplQ-FH-tetR designated as strain CK1 was obtained. The incorporation of L17FH tag in the ribosome was assessed by proteomic and biochemical methods. It was compared with the wild type untagged *E. coli* strain MC4100 designated as C strain further in this document.

4.1.1 Experimental design to study factors associated with ribosome nascent-chain complexes (RNC)

Quantitative proteomic analysis using SILAC (stable isotope labeling with amino acids in cell culture) and biochemical analysis was carried out using CK1 and C bacterial strains. C strain was used as a control for eliminating non-specific interactions of proteins with the anti-FLAG affinity matrix.

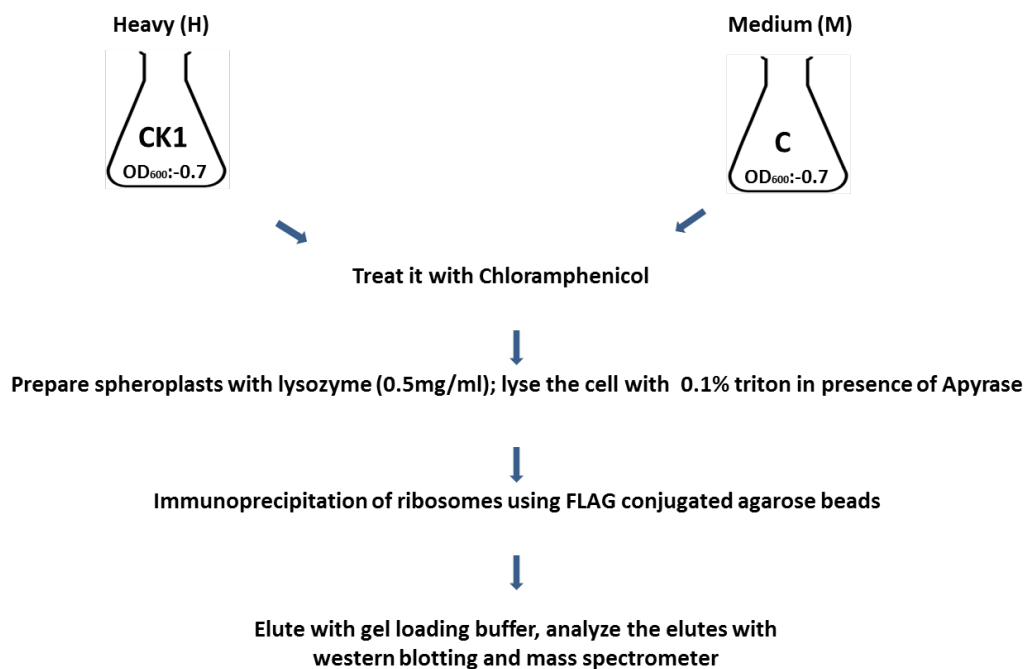


Figure 13: Experimental design for SILAC-based mass spectrometry and biochemical sample preparation for comparing the FLAG-tagged CK1 strain with the wild type untagged C strain.

As shown in the experimental design (Figure 13), the heavy and medium isotope labeled medium is incubated overnight with CK1 strain and C strain, respectively, to obtain primary cultures. Secondary cultures are further inoculated with respective bacterial strains as shown in Figure 13. A low concentration of lysozyme was used to remove the outer bacterial membrane to produce spheroplasts. This mild lysis of outer membranes of cells results in viable spheroplasts where the translation machinery is intact (Netzer & Hartl 1997).

The spheroplasts are then treated with low concentration of CAM (150 µg/µl), and finally lysis is carried out with lower percentage of nonionic detergent Triton (0.1%). This is followed by incubation of soluble protein lysate harboring FLAG-tagged ribosomes with anti-FLAG affinity matrix for a short time frame of 60 min. Bound material is eluted with pre-heated loading dye devoid of the reducing agent DTT.

SDS Polyacrylamide-Gel Electrophoresis and Western blotting analyze the eluates from individual pull-downs from each strain. Further, the total eluate fractions are analyzed by mass spectrometry. The procedures are described in details in the methods section. This protocol is optimized by taking into consideration the dynamic and fragile nature of RNC fractions. Mild lysis protocol is developed and care is taken to avoid physical and mechanical stress to the cells during lysis. Therefore, the method developed in this study ensure stability of RNCs during the sample preparatory phase.

4.1.2 SDS PAGE and Western blotting analysis validates ribosomal protein footprint isolated by pull-down procedure

Molecular weights of isolated proteins from the 30S and 50S ribosomal subunits of *E. coli* were determined by two independent methods: polyacrylamide gel electrophoresis with sodium dodecylsulphate and immunoblotting with antibodies. Additionally, the presence of ribosome associated chaperone TF was also confirmed using the TF antibody.

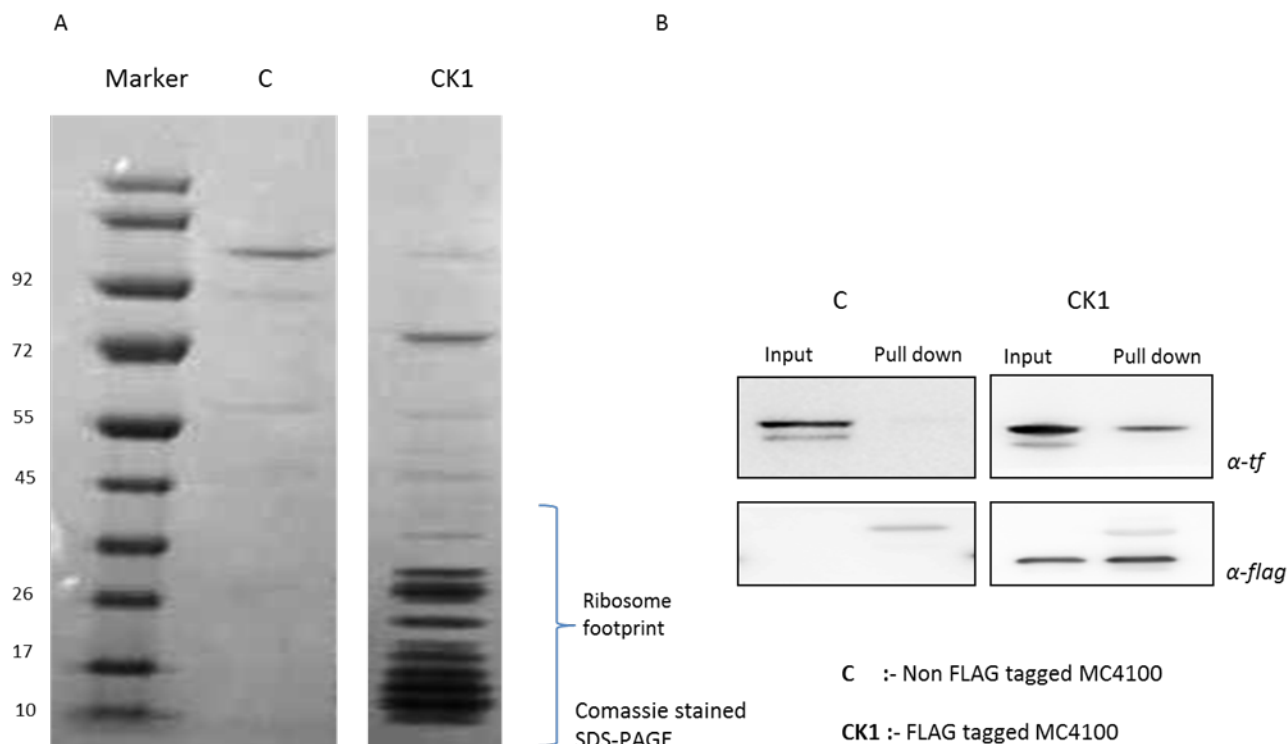


Figure 14: Ribosome-nascent chain associated complexes are successfully isolated using the Flag-tagged L17 protein of *E.coli* ribosome. **A:** Coomassie stained SDS-PAGE analysis of samples prepared from *E. coli* strains CK1 or C strain (MC4100; control). **B:** Immunoblotting analysis of equivalent amounts of lysates and eluates from the pull-down fraction of C and CK1 strain using the ribosomal L17 protein and trigger factor (TF) antibodies.

In Figure 14A, a Coomassie-blue stained SDS-PAGE gel from CK1 pull-down fractions showed the typical ribosomal protein pattern called ribosomal footprint, containing all the large and small ribosomal subunit proteins, while only a very weak background was visible in the pull-down fractions from C strain. The presence of immunoprecipitated FLAG-tagged ribosomal protein L17 was indicated by using anti-flag antibody (Figure 14B). This result validates incorporation of the FLAG tag into the L17 ribosomal protein (Dzionara et al. 1970).

The presence of trigger factor (TF) associated with immunoprecipitated FLAG-tagged ribosomes indicated successful isolation of stable ribosome nascent chain (RNC) complex. The above shown results indicate that L17 is a suitable target for the introduction of an affinity tag into the

ribosome. At this position, the tag can be introduced at the chromosomal level without compromising cellular viability. Furthermore, the presence of TF as shown in Figure 14B indicates that this system is suitable for the co-purification of other less known ribosome-associated factors. Gentle washing steps are recommended during the pull-down of RNC complexes with our established protocol. This avoids any mechanical stress, and thus ensures the integrity of the RNC samples and prevents the loss of transient, less abundant interactions of binding factors with RNC complexes. Design of this experimental protocol would provide insight into the concerted manner in which these binding factors bring about folding of the nascent chain translating on the ribosome.

4.1.3 SILAC quantification for discrimination of genuine RNC interactors from non-specific background binders

Understanding the nature of this transient nascent chain-ribosome complex is vital to delineate the concerted pathway of how the chaperones assist nascent chain folding on the ribosomes. Proteome wide analysis of isolated ribosomal proteins and associated factors was performed to further reveal the nascent chain-ribosome interactome by comparing CK1 to C strain. As a result, 1165 proteins were quantified in at least 2 out of 3 experiments and 245 proteins from this list were significantly enriched (Appendix, Table 7.3.1), as shown in Venn diagram below (Figure 15A). High enrichment shown for all quantified 52 ribosomal proteins indicates the successful isolation of the entire FLAG-tagged ribosome complex (Figure 15B).

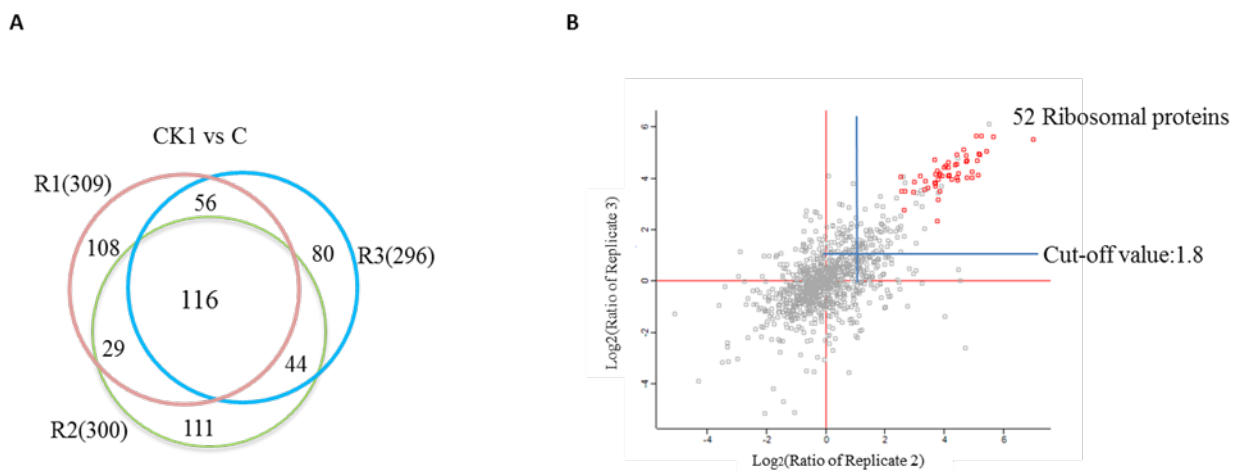


Figure 15: A: Venn diagram indicating the enriched overlapped interactome of ribosome nascent chain (RNC) complexes among the three biological replicates (CK1 vs C strain). **B:** Standard scatterplots of Log₂ SILAC ratios between replicate 2 and 3. The ribosomal proteins (marked in red) are among the highest enriched proteins, indicating that the ribosomal complex has been successfully isolated through immunoprecipitation

TF was among the 245 (Figure 15A) proteins that were found to be enriched, as expected (Merz et al. 2008). This result further validates our method of an *in vivo* approach for generating RNCs. Using this method, we then attempted to decipher interplay of downstream factors associated with TF-RNC complex.

4.1.4 GO Enrichment analysis of biological processes of enriched proteins

To better understand various protein components that make up the RNC complex, we used the online-tool REVIGO to summarize a gene ontology-based (GO) functional categorization of the 245 enriched proteins found in the ribosome pull-down fraction of FLAG tagged-CK1 strain. (Appendix, Table 7.3.1).

GO term analysis of molecular functions of the enriched protein group showed that structural components of the ribosomal proteins, proteins with structural molecular activity, and RNA and rRNA binding terms were significantly enriched. This result confirms our approach of isolating ribosomes by using an endogenous FLAG tag. In addition, enrichment analysis of biological processes revealed that the top-scoring biological functional group of genes encode SRP-dependent co-translational proteins and proteins involved in posttranscriptional gene regulation of gene (Figure 16B). This result suggests that the translational unit along transcription, protein targeting and post-transcriptional factors are also isolated in our ribosome pull-down. The enrichment of these factors in FLAG-tagged ribosome pull-down fractions, when compared to non-FLAG tagged control cells, validates our experimental design. By tagging L17 ribosomal protein and isolating ribosomes nascent chains, we can capture a real-time snapshot of translating *E.coli*. This enables us to identify a compendium of both known and unknown binding partners that play an important role in the translation and co-translation processes.

This data intrigued us to further analyze the nascent chains associated factor in FLAG tagged CK1 strain by introducing two antibiotics: Chloramphenicol (CAM) and Puromycin (PURO), which are used to generate subsets of RNC fractions with stalled nascent chains and released nascent chains respectively.

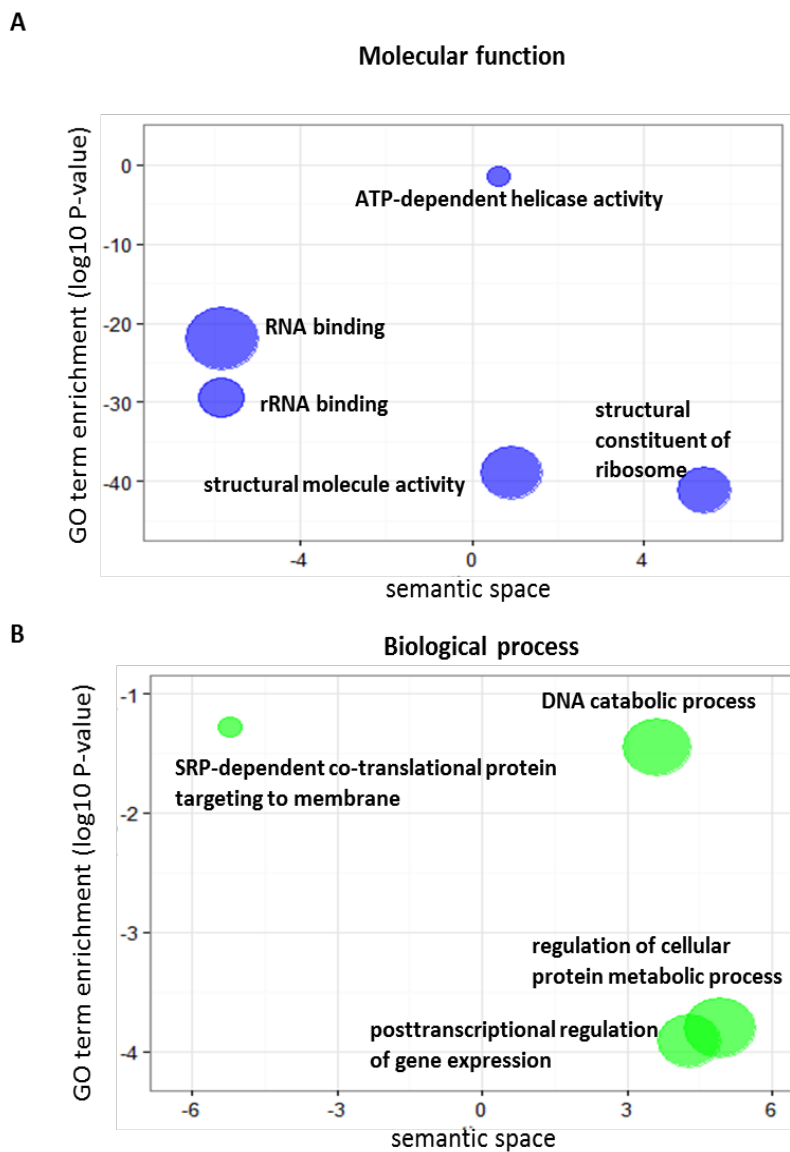


Figure 16 : The GO terms for molecular function (A) and biological processes (B) of all the enriched factors (CK1 VS C strains) were summarised and visualized by the semantic similarity based scatterplot using REVIGO. The semantic space in x-axis indicates the likeliness of meaning between GO terms, the Bubble size indicates frequency of the GO terms. The statistic significance (P-value) of GO enrichment terms was obtained by using online-tool DAVID as all identified proteins are set as the background.

4.2 Antibiotics to stabilize and destabilize ribosome nascent chain complexes

After substantiating our experimental approach using a FLAG-tagged ribosomal protein to isolate RNC complexes for analysis using proteomic and biochemical approaches, we next compared RNC fractions with arrested or released nascent chains.

Puromycin (PURO) mimics the amino-acylated end of the tRNA and participates in peptide bond formation. Its non-hydrolysable amide bond cannot be cleaved, and the resulting peptidyl-puromycin chain dissociates from the ribosome causing premature release of nascent polypeptide chains, thus producing vacant ribosome species (Azzam & Algranati 1973). Chloramphenicol (CAM), on the other hand, acts by occupying the position of the amino acid attached to the A-site tRNA and prevents peptide bond formation. As a result, elongation is arrested and the peptidyl-tRNA is trapped on the ribosome. Thus, chloramphenicol stabilizes ribosome-nascent chain complexes, whereas puromycin leads to premature release of the nascent chain.

4.2.1 SDS PAGE and Western blotting analysis of lysates treated with CAM and PURO antibiotics to generate different species of ribosome nascent chain complex

Different ribosome-associated proteins are pulled down using antibiotics that stabilize or destabilize ribosome-nascent chain complexes in *E.coli* and compared by two independent methods: polyacrylamide gel electrophoresis with sodium dodecylsulphate and equilibrium sedimentation as depicted in Figures 17.

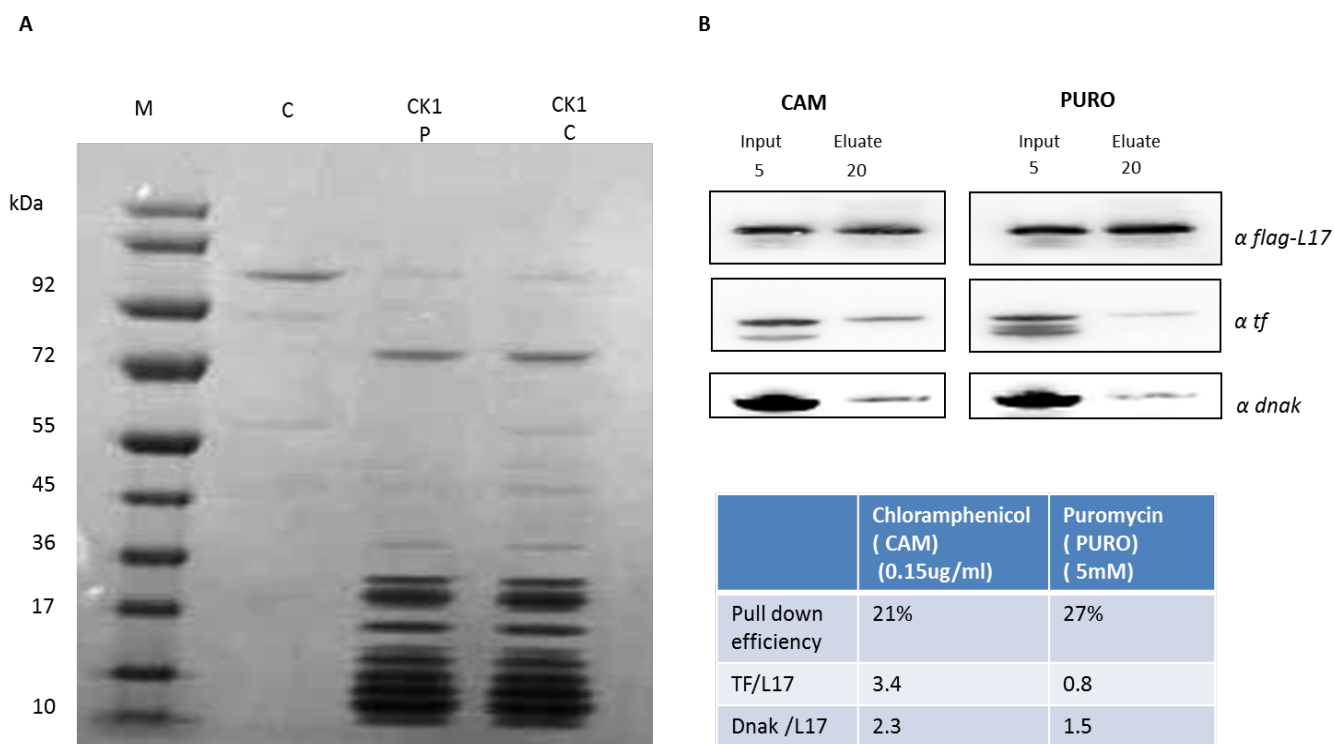


Figure 17: Ribosome nascent chain complexes (RNCs) from a FLAG-tagged CK1 strain of bacteria is isolated after treatment with antibiotics chloramphenicol or puromycin to stabilize and destabilize the complexes, respectively. A: Coomassie blue stained SDS-PAGE gel. Lysates prepared from *E. coli* strains MC4100 C strain (control, lane 2) or strain CK1 (lanes 3 and 4) were subjected to a pull-down using an anti-FLAG affinity matrix, and bound material was eluted with pre-heated loading dye devoid of DTT. Spheroplasts were treated with Puromycin (PURO, lane 3) to release the nascent chains from translating ribosomes or with Chloramphenicol (lane 4) to stabilize the RNCs. The typical ribosomal protein pattern was detected in both treatments from strain CK1 but not in the control sample from C strain. The arrowhead shown in the CAM-treated lane when compared to the PURO-treated lane indicates additional bands. These are proteins bound to the stabilized RNC due to Chloramphenicol (CAM) administration. Lane 1 shows the marker. **B:** Equivalent amounts of lysates and elutes from the pull-down fraction differentially treated CK1 strain was analyzed by immunoblotting using trigger factor (TF), DnaK, and GroEL antibodies. Immunoprecipitated FLAG-tagged ribosomes were visualized using an anti-FLAG antibody (acts as a loading control). The table shows the western blot 2D densitometry analysis of chaperone intensity seen in different treatments.

A2D densitometry analysis (AIDA software) of the TF band associated with immunoprecipitated FLAG-tagged ribosomes from Chloramphenicol treated CK1 strain (CK1_{CAM}) compared to TF band signal from PURO treated CK1 strain (CK1_{PURO}) revealed a 3-to 4-fold enrichment in the CK1_{CAM}

sample. This result clearly indicates stabilization of nascent chains on ribosome by CAM treatment using the lysis method developed here. Simultaneously, the release of nascent chain from translating ribosomes generating a population of vacant ribosomes is required for setting up a negative control. The enrichment of DnaK is less than that of TF (Figure 15B).

Furthermore, the method of using antibiotic mediated translation inhibition to generate stable ribosome-nascent chain complexes provides an opportunity to investigate additional proteins that show transient interactions with ribosome nascent chain complex. Some of these proteins may have a function in chaperoning the nascent chain as it is emerging from the ribosome exit tunnel. However, the transient nature of some protein-protein interactions, seen in the formation of these RNC complexes, may change in response to stimuli such as changes in pH, temperature and osmolality. Thus, further verification of the method developed here is required.

4.3 Analysis of RNC complexes generated using antibiotics

So far, our approach towards the development of a new method to study RNC complexes *in vivo* has shown promising results. The immunoprecipitation technique coupled with functional and quantitative SILAC mass spectrometry analysis facilitates assurance of specific known RNC interactors and potential revelation of weak unknown interactions. In the triple labeled SILAC experimental design one cell population functions as a control, whereas the other two are perturbed by different antibiotics and the sub-proteome of interest is isolated. The SILAC approach itself is simple, but identifying the most sophisticated experimental design and control requires ingenuity (Mann 2006). Keeping our target in focus, we designed the experimental layout shown in Figure 18 that would help us to decipher the binding factors associated with translating ribosomes.

4.3.1 Layout of SILAC experimental design to identify additional factors associated with translating ribosomes *in vivo*

In the experimental design shown in Figure 18, the heavy and medium labeled SILAC media as well as light medium, which uses natural amino acids, are inoculated with the CK1 FLAG-tagged strain. The differential treatment with antibiotic mentioned above is performed after the preparation of spheroplasts with a low concentration of lysozyme treatment. The spheroplasts from each medium are re-suspended in HMK (Hepes, Mg (OAc)₂, K (OAc)) buffer and respective antibiotic treatments are carried out at 37°C in the thermomixer for 10 mins. The heavy labeled sample is subjected to 0.15µg/ml chloramphenicol (CAM); the medium labeled sample is subjected to 5mM of puromycin and the light labeled sample is not treated with any antibiotic in order to compare the stability of untreated RNC with CAM treated RNC.

CAM treatment is administered to the cells as soon as they reach an OD₆₀₀: 0.7 and then the cells are poured into flash frozen vials (50ml falcons pre-cooled in liquid nitrogen). Further, the spheroplasts are also treated with CAM at 37°C for 10 min. On the contrary, the negative control is untreated and the only attempt to capture translating ribosomes is made by pouring the cells into the flash frozen vials to arrest translation. No antibiotic treatment is administered and the untreated sample is processed together with other samples. The positive control (i.e. the puromycin (PURO) control) is treated similarly, except that Puromycin is administered to the cells instead of chloramphenicol.

Analyzing the untreated sample as a negative control enables to decipher any antibiotic induced effects. Equal amounts of eluates from different samples are combined and treated as a single sample in all subsequent steps for FASP-dependent sample preparation. This enables the use of this protocol to compare the pool of RNC interactors from CAM treated samples to the destabilized puromycin treated samples. This experiment is performed three times independently to check the reproducibility of the experimental set-up before validating the dataset.

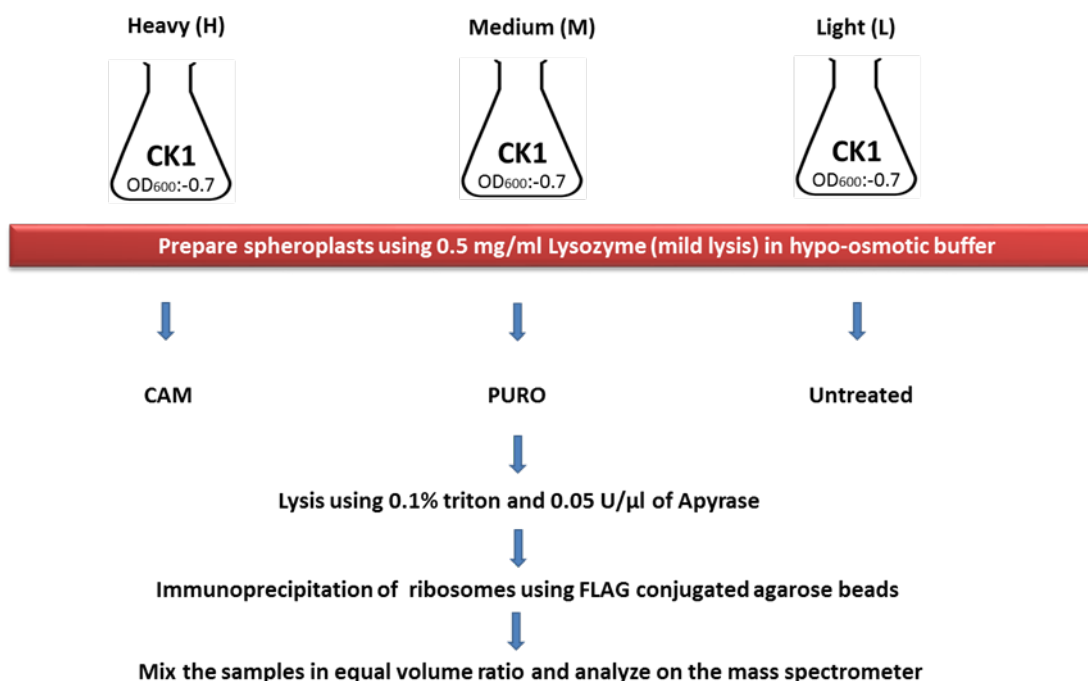


Figure 18: Experimental design for triple labeled SILAC based mass spectrometry.

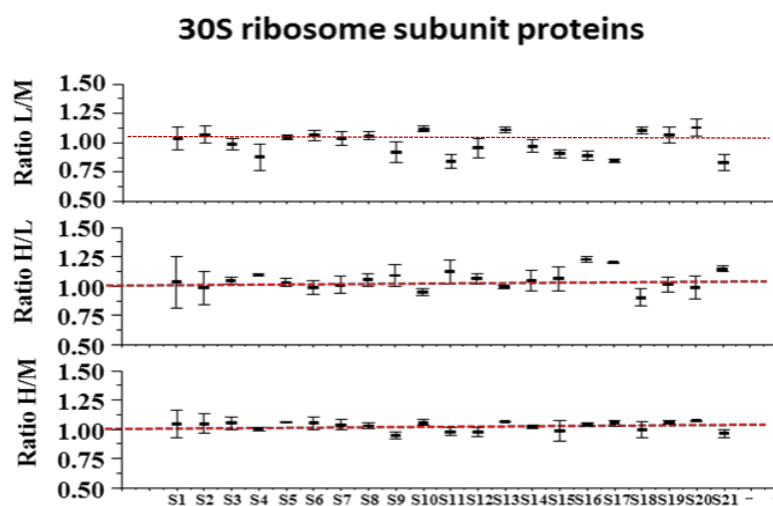
The scheme depicts the biochemical sample preparation for comparing the FLAG-tagged CK1 strain under CAM and PURO treatment with introduction of a negative control (i.e. untreated CK1 strain).

4.3.2 PURO treatment does not disrupt ribosomal subunits and untreated RNC sample is less stable than CAM treated RNC sample

First, the whole dataset is normalized using an average of SILAC ratios of all quantified 50S proteins to correct for pull-down efficiency variations of ribosomal pull-down experiments. Then, the relative abundance of small and large subunit ribosomal proteins is compared to check if the differential treatment with antibiotics induces any specific effect, such as interference with the stability of assembled ribosome subunits during translation. The exact mechanism of puromycin antibiotic action is still unknown but it is understood that the 3' position contains an amide linkage instead of the ester linkage of *tRNA*. This enables the peptidyl puromycin molecule to become resistant to hydrolysis and stops the ribosome from translating further. As shown in Figure 19, when the ratios of ribosomal proteins from puromycin treated sample (M)

are compared with CAM treated sample (H), they are found to be approximately in the same range of SILAC ratios (~ 1). This result suggests that all the smaller 30S subunit proteins are also precipitated along with all the larger 50S subunit proteins and that the release of nascent chains from the ribosome by puromycin treatment does not disassemble the ribosomal subunits.

A



B

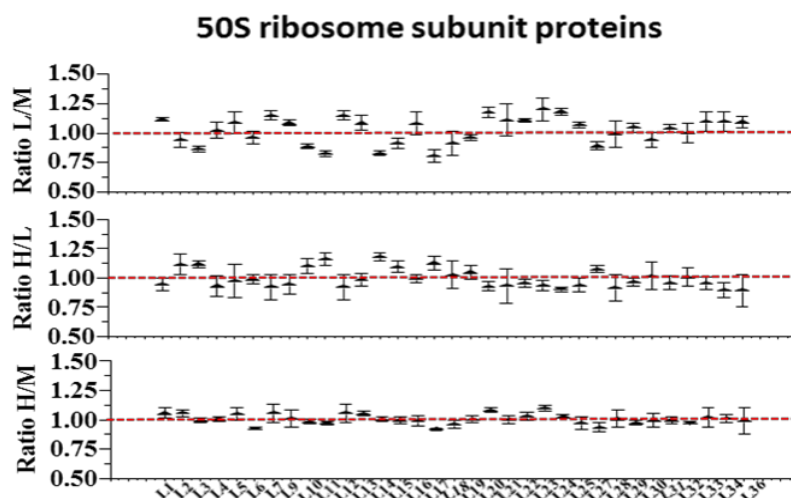


Figure 19: Relative abundance of all quantified ribosomal proteins in ribosomal subunits. A: 21 small subunit 30S proteins (labeled S1 to S22) **B:** 36 large subunit 50S proteins (labeled L1 to L36).

The typical S1 protein required for translation of most natural mRNAs except the leaderless mRNA is an indicator of translating ribosomes and is quantified in all three samples in similar ratios (~1). (Delvillani et al. 2011). It has also been shown that the protein S1 has an RNA chaperone activity. It is vital for the binding and unfolding of structured mRNAs and correct positioning of the initiation codon for translation (Duval et al. 2013). This indicates that puromycin stalls the translating ribosome on mRNA and only nascent chain associated factors should be released from the translating complex upon release of the pool of nascent protein chains from ribosome exit tunnel. On the other hand, the pull-down efficiency of the ribosomal proteins from the untreated sample (L) varies more compared with the CAM treated sample (H) (Ratios L/M, H/L in Figure 19). Thus, indicating that in the absence of a nascent chain stalling agent like CAM, the RNC complex is susceptible to dissociation during lysis, therefore affecting reproducibility of the samples during each individual pull-down. Chloramphenicol treatment ensures the stability of the RNC complex. By exploiting the translation inhibition mechanism of CAM and PURO antibiotics, we hereby establish a method for global analysis of RNC complexes *in vivo* and overcome one of the technical challenges of studying RNC complexes. Further analyzes of protein dataset obtained by this method ensures capturing of unknown binding factors which interact with RNC complex.

4.4 Additional ribosome-associated proteins can be identified under conditions that stabilize ribosome nascent chain complex

The validation of experimental design acquired through initial biochemical and proteomic analysis of ribosomal proteins and identification of known RNC interactors prompted us to further analyze protein dataset obtained from the SILAC experimental designs.

4.4.1 Chloramphenicol stabilizes nascent chains on ribosomes and is essential to generate consistent dataset of RNC interactors

Proteome-wide analysis of RNC complexes generated with the use of antibiotics revealed a group of specific RNC interactors. A total of 1089 proteins were quantified in two out of three biological

replicates of CAM vs PURO and of those, 378 proteins were found to be enriched (Figure 20A). Almost the same number of proteins, i.e. 1091, were quantified in Untreated vs PURO, but fewer were enriched (259) (shown in appendix 7.3.2) in two out of three biological replicates (Figure 20B).

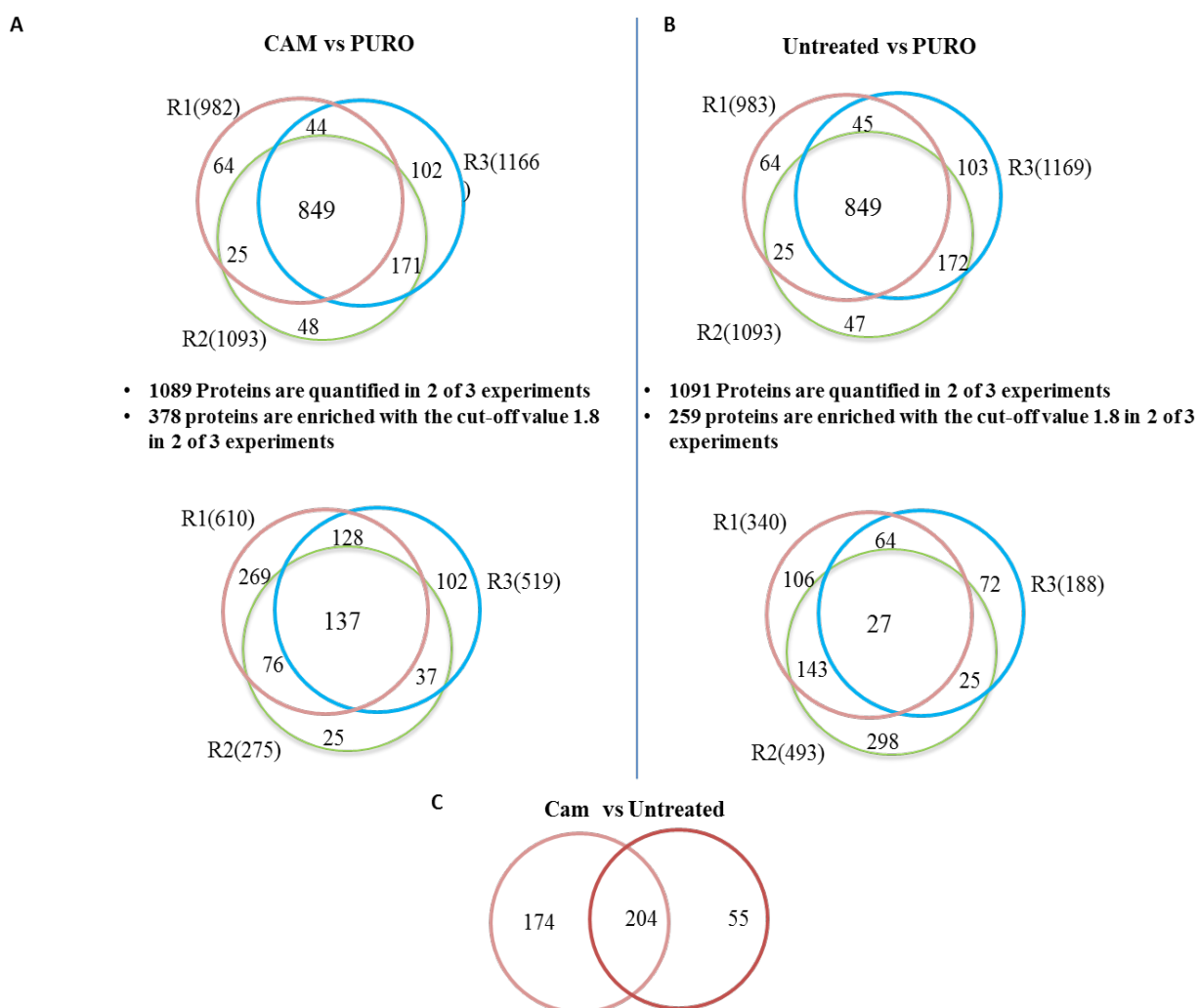


Figure 20: Venn diagram indicating overlap between interactome of ribosome-complexes upon antibiotic treatment (CAM vs PURO) and between untreated vs. PURO-treated samples in three biological replicates from the CK1 strain. A: Overlap of CAM and PURO treated proteins in three independent biological replicates. B: Overlap of Untreated RNC sample with PURO treated samples. C: A nearly 50% overlap of ribosome nascent chain (RNC) interactome from untreated sample was found in CAM treated sample. Furthermore, 174 proteins were stabilised by CAM treatment.

When the group of enriched proteins from each replicate of Untreated vs PURO experiment was overlapped, only 27 proteins were found to be enriched in all 3 replicates. This is five times less than the 137 proteins stringently enriched in CAM vs PURO experiment (Figure 20A and 20B). It is known that CAM irreversibly binds to a receptor site on the 50S subunit of the bacterial ribosome, inhibiting peptidyl transferase and stabilizing the ribosome nascent chain interactome (Dunkle et al. 2010) . This suggests that CAM provides a line of defense against the instability of RNCs due to any mechanical or physiological stress triggered by the lysis procedure or external stimuli.

This result complements our previous finding of pull-down efficiency of all the ribosomal subunit proteins to be stable in CAM treated fraction of RNC (Figure 19). Hence, our method ensures the production of a homogenous population of stable RNC interactomes for understanding early-stage protein folding *in vivo*.

4.5 Stabilization of RNC interactions by chemical crosslinking

The RNC interactome analysis described in this work provides a protocol for enlisting a plethora of factors involved in maturation of a nascent polypeptide chain in bacteria. Owing to the short half-lives of the RNC interactors, most factors interact only transiently with translating ribosomes. Upon affinity purification of RNCs only the factors whose substrate binding is ATP or GTP-dependent, such as Hsp70 chaperones and SRP, can be stabilized by rapid ATP depletion using apyrase treatment. The ATP-independent RNC binding partners can be stabilized *in vivo* by using a chemical cross-linker that must be specific enough to avoid the formation of artificial complexes. We have incorporated chemical crosslinking in the lysis protocol using the antibiotic treatment, at the spheroplast preparation stage.

4.5.1 Time and concentration scale optimization of DSP used for crosslinking the intact *E.coli*

Recent studies indicated that the use of chemical cross-linker could successfully stabilize binding factor-RNC interactions *in vivo* (Becker et al. 2013, Oh et al.2013). Selective ribosome profiling

(SeRP) is a technique that identifies substrate pools and points of RNC engagement for a number of factors associated with co-translational protein folding on the ribosome. SeRP is based on sequencing mRNA fragments covered by translating ribosomes combined with a procedure to selectively isolate RNCs whose nascent polypeptides are associated with the factor of interest.

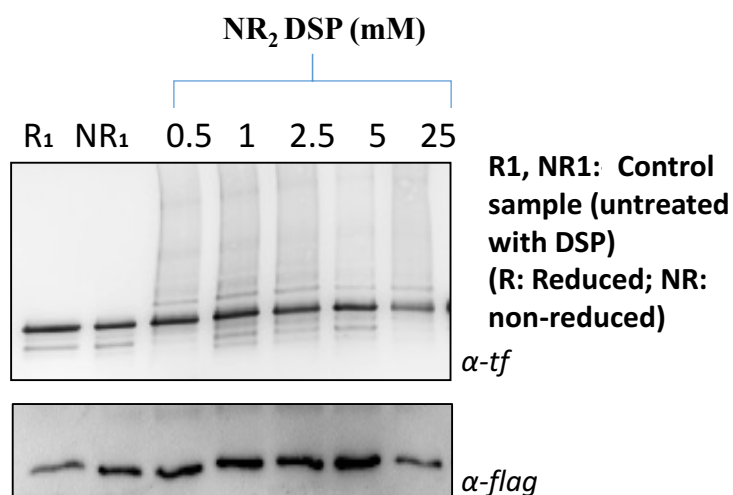


Figure 21: Stabilization of Nascent Chain interactions by chemical cross-linking using membrane-permeable Dithiobissuccinimidyl propionate (DSP) prior to cell lysis and immunoprecipitation. Validating the concentration of DSP to demonstrate that 2.5mM is optimal for crosslinking the cells. As shown in Figure 21, the presence of the upper molecular weight smear above the trigger factor band indicates the extent of crosslinked TF-RNC substrates.

4.5.2 Optimized conditions for chemical crosslinking using DSP integrated with antibiotic treatment of spheroplasts

Puromycin treatment does not yield desired comparative profiles of RNC binding partners when the cross-linker is administered to intact *E.coli* cells. Therefore, we attempted to employ crosslinking treatment on antibiotic treated spheroplasts. The design of this experiment includes re-suspension of spheroplasts in lysis buffer without triton and administration of CAM and PURO

in respective differentially labelled SILAC samples. An untreated sample is added as a control to eliminate any antibiotic induced artifact. DSP dissolved in DMSO is added to the sample. It is allowed to react for a short time frame of 30 sec on thermomixer at 37°C, before quenching with 50mM tris (pH 8). Then triton (0.1%) is added and incubated for 5 min, followed by a pre-clearing step. The concentration of the lysate is then measured by bradford method by loading 300µg protein on 75µl of agarose conjugated FLAG beads, followed by elution in 60 µl of pre-heated loading dye without DTT.

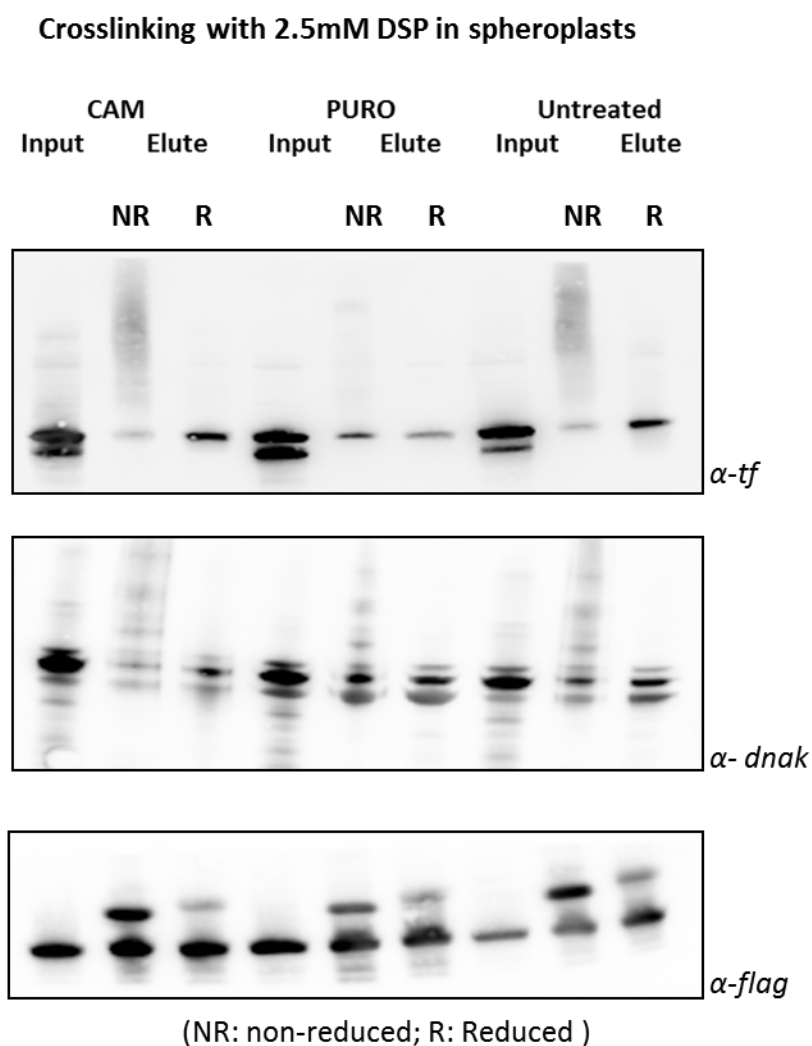


Figure 22: Cross-linking CK1 strain proteins with 2.5mM DSP for 30 sec after administration of antibiotics in the spheroplasts and subsequent western blotting analysis with TF, DnaK and FLAG antibodies.

The advantage of using a reversible cross-linker such as DSP is that the disulfide bond can be cleaved using DTT and thus the cross-linked interacting protein partners can be cleaved and disintegrated before processing the mass spectrometry sample. In addition, the extent of cross-linking and its effect coupled with antibiotic treatment can be tested prior to the mass spectrometry analysis by immunoblotting against TF and DnaK antibody.

Immunoblotting with the trigger factor antibody revealed a high molecular weight smear above the TF band (non-reduced band) on crosslinking after CAM treatment and in the untreated sample. The smear disappeared upon reversing the cross-linking by treatment with thiols, such as dithiothreitol (reduced band, Figure 22). The upper molecular weight smear seen above the TF band is indicative of the TF-associated substrates cross-linked with TF-RNCs. The absence or very weak upper molecular weight smear seen upon puromycin treatment in the non-reduced fraction of the eluate demonstrates release of nascent chain from translating ribosomes upon PURO treatment. The lower panel in Figure 22 shows the loading control on the immunoblot with FLAG antibody. This result suggests that crosslinking of RNC fractions would reveal subsets of potential RNC interactors *in vivo*.

Factor-RNC interactions are stabilized by crosslinking but care must be taken to stabilize and purify factors associated with ribosomes and nascent chains without introducing a bias due to long purification protocols that facilitate either loss or formation of new interactions. Alternatively, our global analysis of *in vivo* RNCs through isolation of tagged ribosomes, where the nascent chains and mRNAs stabilized by CAM antibiotic are compared to vacant ribosomes having nascent chains replaced by peptidyl puromycin analogue, establishes a less biased approach to identify the compendium of protein interactors, which play significant roles in translation of nascent proteins on the ribosome. Further, our short lysis protocol and immunoprecipitation techniques coupled with the stabilization of transient interactors by use of a cross-linker make our approach versatile. Figure 22 demonstrates the coupling of our protocol with crosslinking *in vivo*.

4.6 Comparison of RNC interactome before and after crosslinking with DSP reveals groups of potential RNC interactors

RNCs-binding partner interactions play indispensable catalytic and regulatory roles within the cell. Understanding their physical association *in vivo* provides valuable insight into the process of co-translational protein folding, protein targeting and translocation as well as post-transcriptional regulation in the cellular milieu. Having combined the use of highly reactive, reversible cross-linker DSP with the high-stringency immunoprecipitation procedure developed in this study, we next sought to identify specific protein interactors associated with translating ribosomes in *E.coli* using mass spectrometry.

4.6.1 Crosslinking stabilizes RNC interactome

The Venn diagram in Figure 23A below shows overlap of enriched groups of proteins in crosslinking experiments from three independent biological replicates. A total of 999 proteins were quantified in two out of three sets of experiments and a total of 374 proteins (also listed in Table 7.3.2 in the Appendix section) were found to be enriched in two out of three sets of experiments.

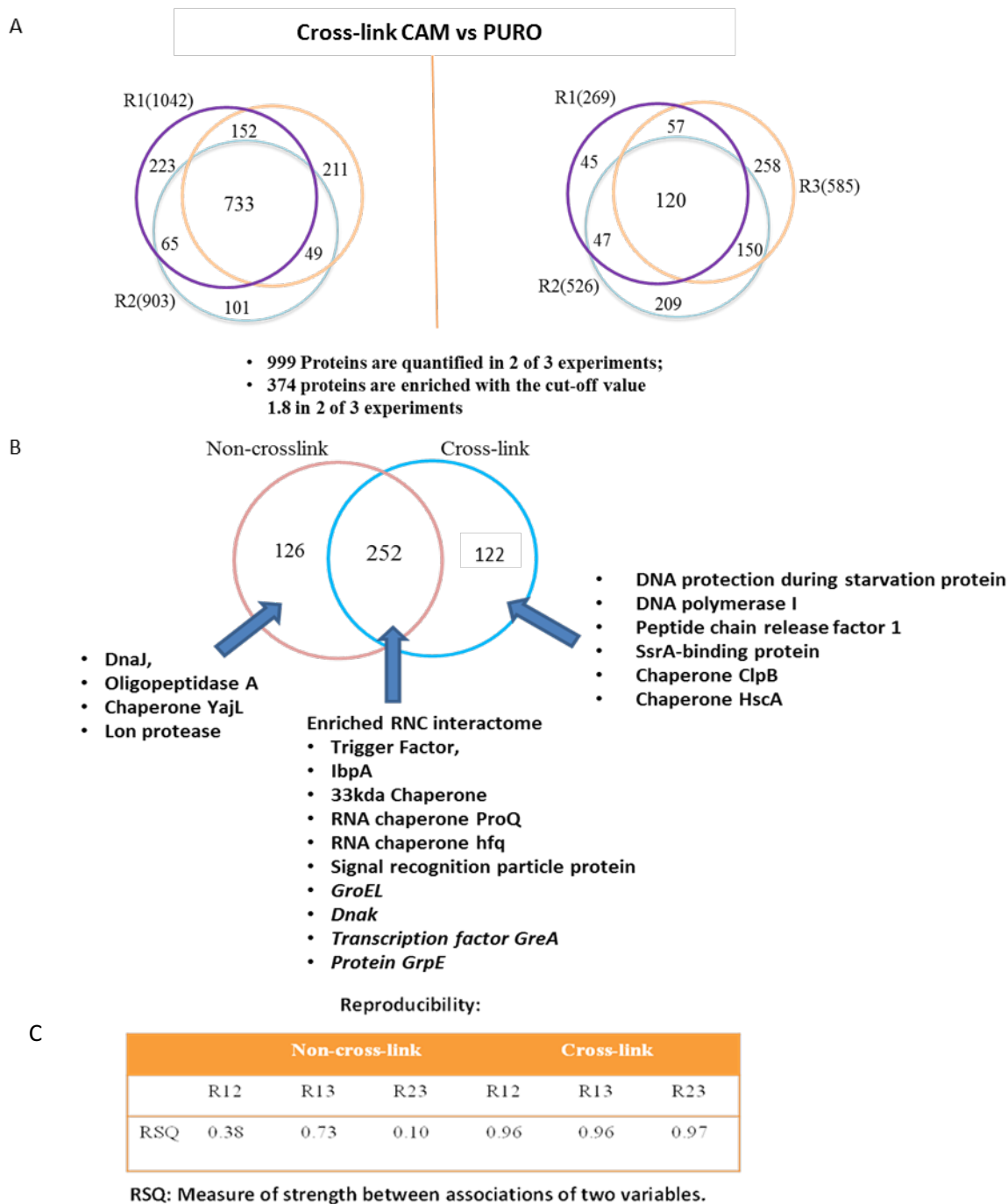


Figure 23: A. Venn diagram indicating the extent of overlap of the ribosome nascent chain interactome upon cross-linking and antibiotic treatment (CAM vs PURO) between the three biological replicates from CK1 strains. **B.** Venn diagram indicating the extent of overlap of the interactome of the ribosome nascent chain complex between the crosslinked and non-crosslinked datasets. **C.** The reproducibility of three

independent dataset improves upon crosslinking as indicated in the table above. RSQ (also called determination coefficient) is a measure for the correlation of ratios between independent biological replicates (R).

The Figure 23B showed 67% overlap of enriched proteins between crosslinked and non-crosslinked dataset including some known factors TF, Dnak and signal recognition particle (SRP) (Appendix 7.3.2 shows overlapping proteins between the crosslinked and non-crosslinked databases). The bacterial SRP receptor (SR) FtsY was also found to be enriched upon CAM treatment of RNCs. Co-translational membrane targeting of proteins by the bacterial signal-recognition particle (SRP) requires specific interaction of the SRP-ribosome nascent chain complex with SRP receptor (FtsY) (Saraogi & Shan 2011). Therefore, some inner membrane proteins were also identified in our final dataset, which further validates the method developed in this study. Interestingly, some less known chaperones such as GroEL, IbpA and 33kDa redox regulated molecular chaperone were also found enriched on RNC. Their fold-enrichment varies in each independent experiment, presumably reflecting their dynamic and transient interaction with a set of specific nascent chains based on their length, sequence composition and point of engagement. These novel factors require further *in vitro* experimental validation to decipher their role in nascent chain protein homeostasis.

Incorporating crosslinking in our protocol reduced the enrichment fold of TF (Appendix 7.3.2, yellow highlight). This indicates that DSP crosslinker may also crosslink the binding partners of RNC in puromycin treated sample due to their close proximity to ribosome exit site. This reduces the enrichment score of CAM vs PURO for some of the binding partners. There are binding partners found enriched only in the non-crosslink dataset such as Lon protease and chaperone YajC. In addition, there were some proteins found enriched only in the crosslinked database such as chaperone ClpB and chaperone HscA. The presence of these chaperones indicates that crosslinking can be a potential tool for capturing some transient interactions between chaperones and RNCs. As a result, the reproducibility of proteomic dataset is greatly improved by further stabilizing the RNC complex upon crosslinking (Figure 23C).

4.6.2 GO ontology analysis of enriched protein groups before and after crosslinking with DSP

To understand the biological process and molecular functions of enriched proteins found in both crosslinked and non-crosslinked RNC interactomes, we performed GO ontology and represented it in Figure 24A and 24B.

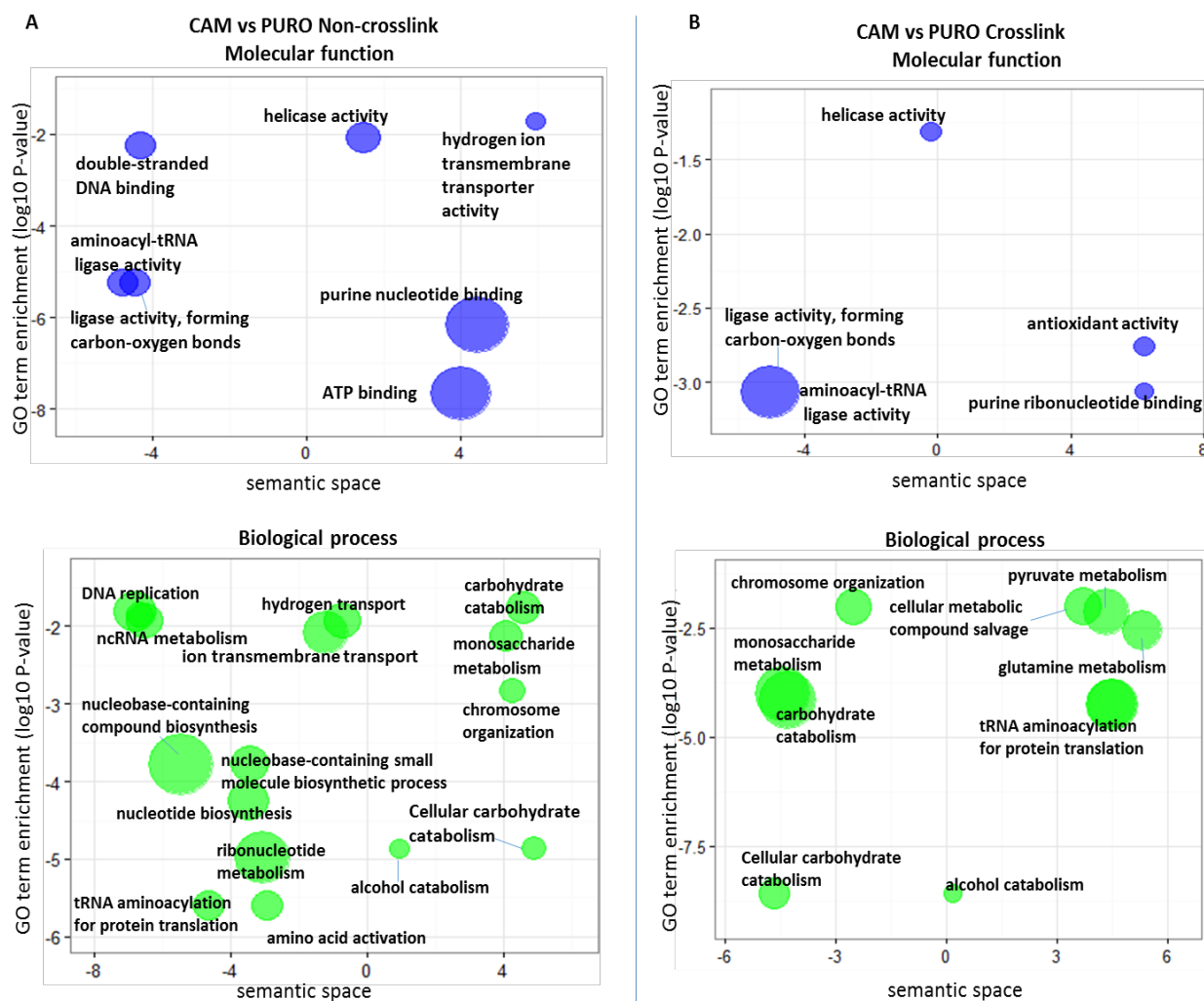


Figure 24: A: The GO terms molecular function (upper panel) and biological process (lower panel) of enriched proteins found in the non-crosslinked dataset (CAM vs PURO) of RNC interactome was summarised and visualized by the semantic similarity based scatterplot using REVIGO **B:** Molecular and biological process of the enriched proteins groups found in cross-linked dataset (CAM vs PURO) of RNC interactome. The semantic space in x-axis indicates the likeliness of meaning between GO terms; the

Bubble size indicates frequency of the GO terms. The statistical significance (P-value) of GO enrichment terms was obtained by using online-tool DAVID as all identified proteins are set as the background.

The GO terms of molecular function for the enriched protein group showed predominantly genes with encoding proteins with amino acyl tRNA ligase activity. These proteins were identified both in crosslinked and non-crosslinked dataset. They function as catalyzing agents for the formation of aminoacyl-tRNA from amino acid, tRNA and ATP with the release of diphosphate and AMP. Some examples of enriched proteins from this group is aspS (aspartate tRNA ligase) and alaS (alanyl-tRNA synthetase).

From the GO ontology analysis of genes regulating different biological processes, the most interesting groups were chromosome organization and tRNA amino acylation for protein translation. DNA-binding protein H-NS (hns) was found enriched in both the crosslinked and non-crosslinked datasets and is responsible for chromosome organization and compaction in *E.coli*. Since transcription and translation are coupled in bacteria, we expected to identify proteins involved in both these biological processes. H-NS protein, however, is reported to show transcription repression, especially at low temperatures (Atlung et al. 1996). The presence of this factor indicates the translational stalling by CAM and PURO antibiotic stalls the transcription process as well.

The predominant presence of proteins involved in tRNA aminocyclation, forming an ester bond between the 3'-hydroxyl group of the 3' adenosine of the tRNA, shows that we isolate translating ribosomes. This process is also known as tRNA charging and it is an important step of the ribosome mediated polypeptide synthesis. This data further prompted us to analyze the crosslinked dataset and compare it with the results from non-crosslinked experiments in order to identify groups of factors associated with RNC in *E.coli*.

4.6.3 Analysis of the dynamic RNC interactome generated by using antibiotic treatment coupled with chemical crosslinking

The scatter plot in Figure 25 shows the Log₂ SILAC ratios of CAM/PURO enrichment plotted against the protein abundance (iBAQ values of whole proteome). The SILAC ratios have been normalized against all 50S ribosomal proteins to correct for pull-down efficiency.

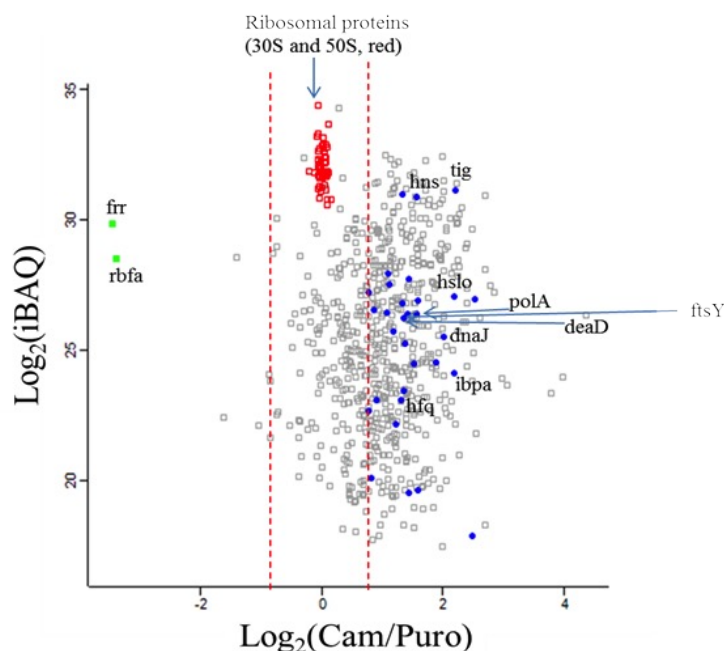


Figure 25: Standard scatterplot with Log₂ (CAM/PURO) ratios/log₂ (iBAQ, whole proteome). Proteins of interest are marked with the following color code: green, significantly diminished; blue, significantly enriched; red, the ribosomal proteins. The abundance distribution of all the quantified proteins is shown. The red dotted line shows the cut-off.

Alongside enriched factors described earlier in Figure 23B, the proteins with diminished CAM/PURO ratios (≥ 0.45) are also significant for understanding the mechanism of antibiotic action on RNC. These factors are found to be enriched in PURO treated samples indicating that the ribosome is destabilized upon release of nascent chains. Among notable members of our dataset is Ribosome Binding Factor A (RbfA), also known as the cold shock protein, which associates with 70S monosomes and free ribosomal subunits in *E.coli*. It is shown in the literature that a downshift in temperature causes increase in non-translatable ribosomes for most mRNAs

and this signal induces the cold shock response. This signaling mechanism is also expected to occur in the presence of inhibitors of translation (Jones & Inouye 1996). Therefore, the enrichment of this factor in PURO treated RNC fraction indicates that our method generates a homogenous population of non-translating ribosomes by the use of puromycin antibiotic. Another protein enriched in PURO treated RNC fraction was ribosome-recycling factor (frr), which is known to disassemble the post-termination ribosomal complex. At the termination step of protein synthesis, nascent polypeptide is released from tRNA by release factor, thus forming the post-termination complex. This complex is then disassembled by ribosome recycling factor (RRF). On binding of RRF, the ribosome is released from mRNA. This process is also called unscheduled translation (Hirokawa et al. 2004). Our results indicate that the puromycylated nascent chain induces the formation of post termination ribosomal complex. The presence of RRF and Rbfa protein factors further confirms our experimental design and reveals that puromycin is a correct negative control to compare translating and non-translating ribosomes.

4.7 Protein groups associated with translating ribosomes under crosslinked and non-crosslinked conditions

Analyzing the RNCs generated by this protocol led us to identify inventory of factors contributing towards well-organized folding of nascent chains in the cellular environment. These factors are divided into the following main categories: translocation factors, nascent chain modifying enzymes, proteases, molecular chaperones, translation factors and RNA quality control proteins. Here we compare the enrichment of these categories of proteins from the cross-linked and non-crosslinked datasets.

Studies in the past had shown the interplay between these groups of factors during the process of co-translational translocation process (Sandikci et al. 2013). We identified proteins involved in co-translational targeting of membrane proteins in bacteria such as SRP (ffh) and the bacterial SRP receptor, FtsY. These proteins are known to interact functionally and physically with the Sec YEG translocon (Angelini et al. 2005) and our study demonstrates enrichment of SEC proteins on

CAM stabilized RNC complex (Figure 26A). These proteins were found also enriched upon crosslinking (the red bar graph). Recent studies on global profiling of SRP interaction with RNC complex by selective ribosome profiling method has revealed most substrates of *E.coli* SRP are inner membrane proteins (Schibich et al. 2016). Our proteomic quantification also identified some of these substrates (shown in green highlight in Table 7.3.2 of Appendix), such as YajC protein, which is a small integral membrane protein found as the heterotrimeric complex with SecDFyajC. It is a translocase-associated protein with various functions (Veenendaal et al. 2004). Another member of Sec protein family, the secretion chaperone SecB also shows enrichment on RNC fraction both in crosslinked and non-crosslinked dataset. SecB holds newly synthesized Sec substrates in an translocation competent, unfolded conformation by wrapping them around SecB tetramer thus helping *E.coli* to cope with the challenge of protein homeostasis posed by post-translational translocation (Randall et al. 2005). SecB is a multitasking chaperone in the cytosol and it exhibits strong anti-folding activity by keeping the secretory proteins in secretion competent state and promoting their targeted delivery to SecA ATPase (Huang et al. 2016). This result (Figure 26A) shows that we are able to capture snapshots of co-translational translocation process in *E.coli* and proteins of the chaperone group play most essential role in mainting nascent proteome homeostasis.

Our data also revealed enrichment of other ribosome associated biogenesis factors such as peptidyl deformylase (def) and methionine aminopeptidase (map). These enzymes are nascent chain modifying enzymes and their docking sites are ribosomal proteins L17 and L22. Peptidyl deformylase (PDF) is the first enzymatic factor that processes the nascent chain by removing the formyl group as polypeptides emerge out of exit tunnel and before the nascent polypeptides bury their N-termini to adopt their native folds. Following this step, the N-terminal methionine is excised by methionine aminopeptidase. It has been shown *in vitro* that PDF and MAP bind concurrently to TF or SRP but do not influence their binding affinities to RNCs. PDF and TF are known to act in a concerted manner and data from selective ribosome profiling indicated TF and PDF do not compete for binding to RNCs (Becker et al. 2013). MAP, on the other hand, was found

not to displace TF or SRP from RNCs but to compete with PDF for binding to RNCs (Bornemann et al. 2014). The presence of these factors as enriched proteins on translating ribosomes (Figure 26B) indicates that our method is valuable in enlisting binding partners of RNC complex *in vivo*.

In prokaryotes, TF is the first known interactor of RNCs and it shows highest enrichment amongst the group of chaperone found on RNC fraction of CAM vs PURO enriched protein in both the crosslinked and non-crosslinked datasets. We also found other known downstream chaperones such as DnaK and GroEL that cooperate with TF in maintaining nascent chain integrity on the ribosome. It has been shown in the past that DnaK and TF have distinct but overlapping substrates (Deuerling et al. 2003) as do DnaK and GroEL (Endo & Kurusu 2007). The Hsp70/DnaK system and the GroEL cooperate to promote folding. They are known to interact with their substrates both post-translationally and co-translationally (Teter et al. 1999). The identification of these chaperones validates our method of studying RNC complex *in vivo*.

Further analysis of our dataset revealed interesting less known chaperones such as HscA, ClpB and HtpG (Figure 26C). The pool of these chaperones are expected to interact with nascent chains in a concerted manner to support productive folding and reverse misfolding in a cellular environment. Among notable interactors identified by our method we also found two other chaperones, small heat shock protein IbpA and the Hsp33 chaperone, which have not been previously reported to be associated with the ribosome nascent chain complex. These are cytoplasmically localized. IbpA in complex with another chaperone from the same family, IbpB, had been reported to protect non-native proteins produced by heat, freeze thawing and other stress conditions (Kitagawa et al. 2002). During the nascent chain translation and folding on ribosomes, the cytosolic environment is under enormous stress and such conditions lead to activation of heat shock proteins and disaggregating chaperones. Identification of these chaperones by the method established here to study the early stage protein folding on ribosomes requires further investigation in order to decipher new functions of these chaperones with respect to RNCs.

In addition to the factors discussed above, we also observed enrichment of translational protein factors such as elongation factor Ts (tsf) and elongation factor 4 (lepA) in CAM stalled RNCs, both before and after crosslinking (Figure 26D). This data indicates isolation of integral RNC complex from *E.coli*. We also demonstrated the enrichment of ATP dependent proteases such as HslU and ClpX. Both these proteases show sequence similarity and they belong to a two-component system, i.e HslVU and ClpXP. HslU recognizes the N-terminal part of its protein substrate and unfolds them (Yoo et al. 1996). The nascent chains while translating on the ribosomes are susceptible to degradation or misfolding and therefore, bacteria has evolved powerful quality control system consisting of proteases and chaperones (Mogk et al. 2011). The presence of these proteases (Figure 26E) in our RNC fraction indicates that they cooperate with chaperones to monitor the folding states of proteins and eliminate misfolded conformers through degradation or refolding.

Other than chaperones, translation factors and proteases, RNA quality control group of proteins were also found enriched (Figure 26F). Identification of the sigma factor rpoD, which is the primary sigma factor during exponential growth, indicates that we isolate the actively translating RNC fractions. Another interesting factor in this group was transcription elongation factor greA. It has been shown recently that GreA in *E.coli* effectively promotes the refolding of denatured proteins and inhibits aggregation of several substrate proteins under the heat shock conditions. These facts reveal that GreA exhibits chaperone like activity. Distinctive from other chaperones, GreA does not form stable complexes with unfolded substrates and its overexpression confers the host cells resistance to heat shock and oxidative stress (K. Li et al. 2012). The enrichment of these proteins involved in transcription, post-transcriptional regulation, co-translation folding of proteins shows that the method described in this study can be effectively utilized to study the process of co-translational protein folding, targeting and translocation process.

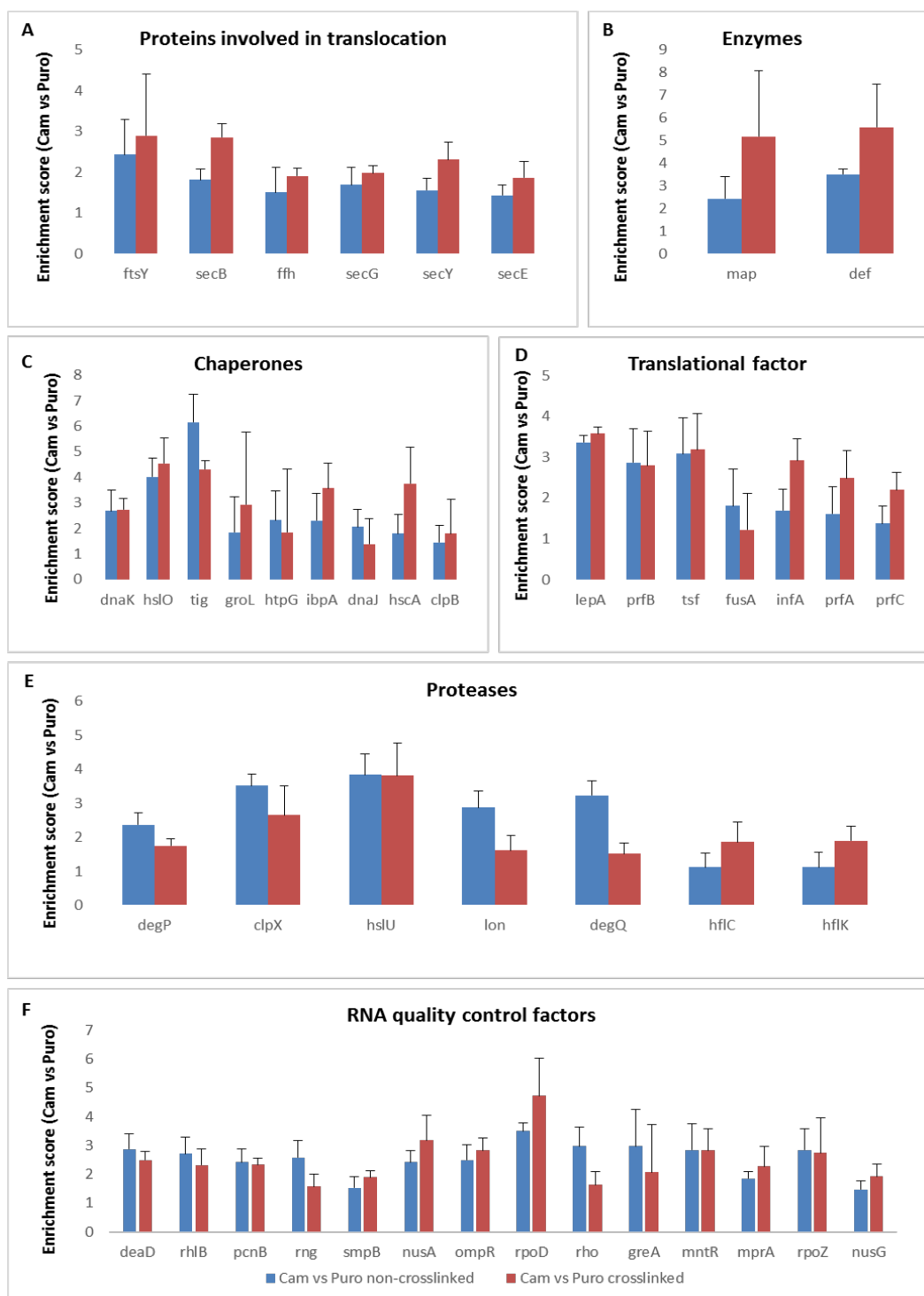


Figure 26: A-F Enrichment scores for various protein categories compared under crosslinked and non-crosslinked conditions. Median values of three biological replicates with standard error of the mean values are shown.

4.8 Quantification of TF & DnaK chaperone levels normalized to the ribosome nascent chain using stalled luciferase constructs

The global analysis of RNCs generated using the method described above gave us insight into a set of novel protein factors that potentially interact with RNCs *in vivo*. This further instigated us to investigate a specific model protein stalled on the ribosome by the classical method of using a stalling sequence and immunoprecipitation of a specific RNC complex using an affinity tag on the model protein. Herein, we investigated the intracellular chaperone network associated with a specific ribosome nascent chain complex in prokaryotes. In order to generate stable RNCs, the 17 amino acid long SecM stalling sequence from *E. coli* was used.

4.8.1 Design of stalled nascent chain constructs of different length using pBAD Luciferase gene construct and 17 amino acid SecM stalling sequence

These constructs, as shown in Figure 27, are designed on the basis of different hydrophobic patches in the firefly luciferase protein (Conti et al. 1996). Firefly luciferase (550 aa) was used as a model protein as it shows high chaperone dependence for folding.

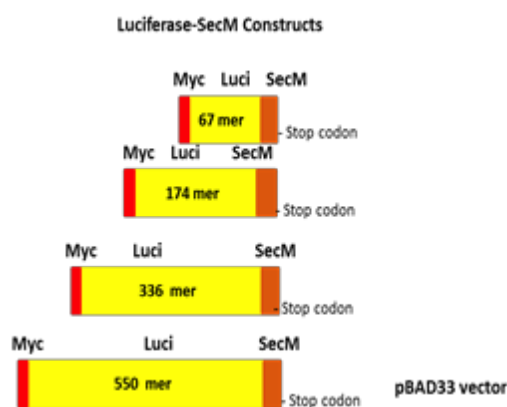


Figure 27: Schematic representation of SecM-stalled luciferase constructs harboring different lengths of the gene based on the hydrophobicity profile.

The RNC complex of these constructs was isolated by the pull-down of the *c-myc* tag present in the N-terminal region of the respective length luciferase nascent chain.

4.8.2 Expression profiles for stalled luciferase nascent complexes of different length reveal weaker expression of shorter constructs stalled on ribosomes

Nascent chain stalled constructs shorter than 100 amino acids show weak expression levels (Figure 28) and their presence is detected by western blotting only at a higher exposure time of chemiluminescence.

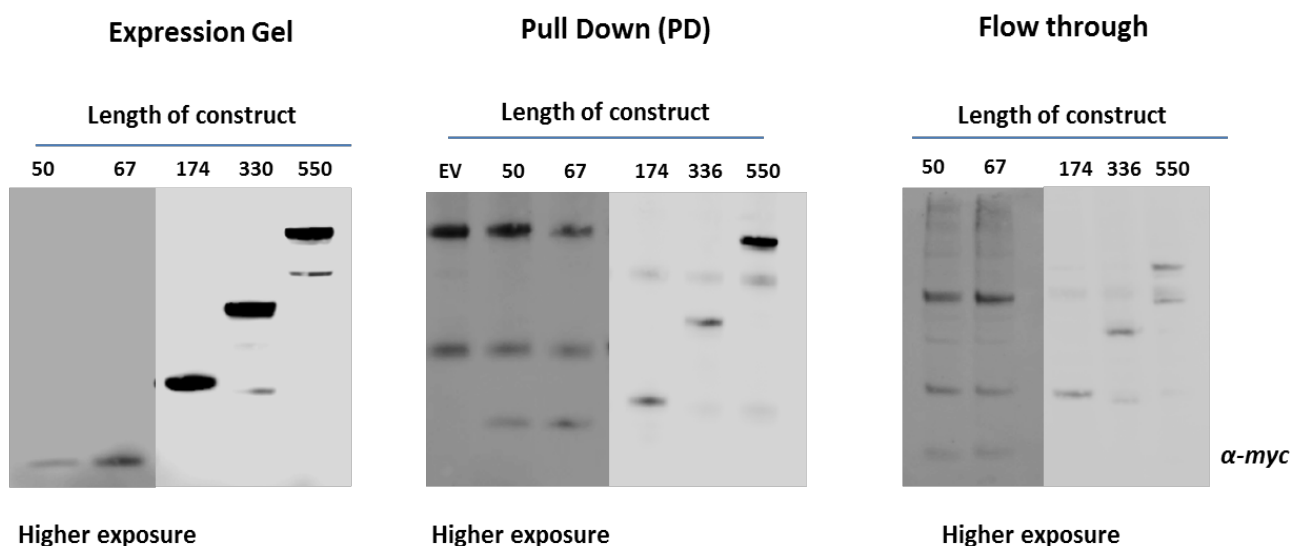


Figure 28: Western blotting analysis of immunoprecipitated stalled complex of different length luciferase RNCs isolated via *c-myc* tag.

This is also indicated in the pull-down fraction of nascent chain constructs. The pull-down efficiency for higher length nascent chains (>174mer) is above 70% when compared to the flow through fraction. Empty vector was designed to eliminate background binding.

4.9 Chaperone interaction profile with stalled nascent chains of different length

The RNC interactors known from previous studies were used in this experimental set-up to evaluate their interaction profile *in vivo* with luciferase stalled constructs as shown in Figure 27.

We observed a linear decrease in recruitment of TF as the length of stalled nascent chain increased on the ribosome (Figure 29).

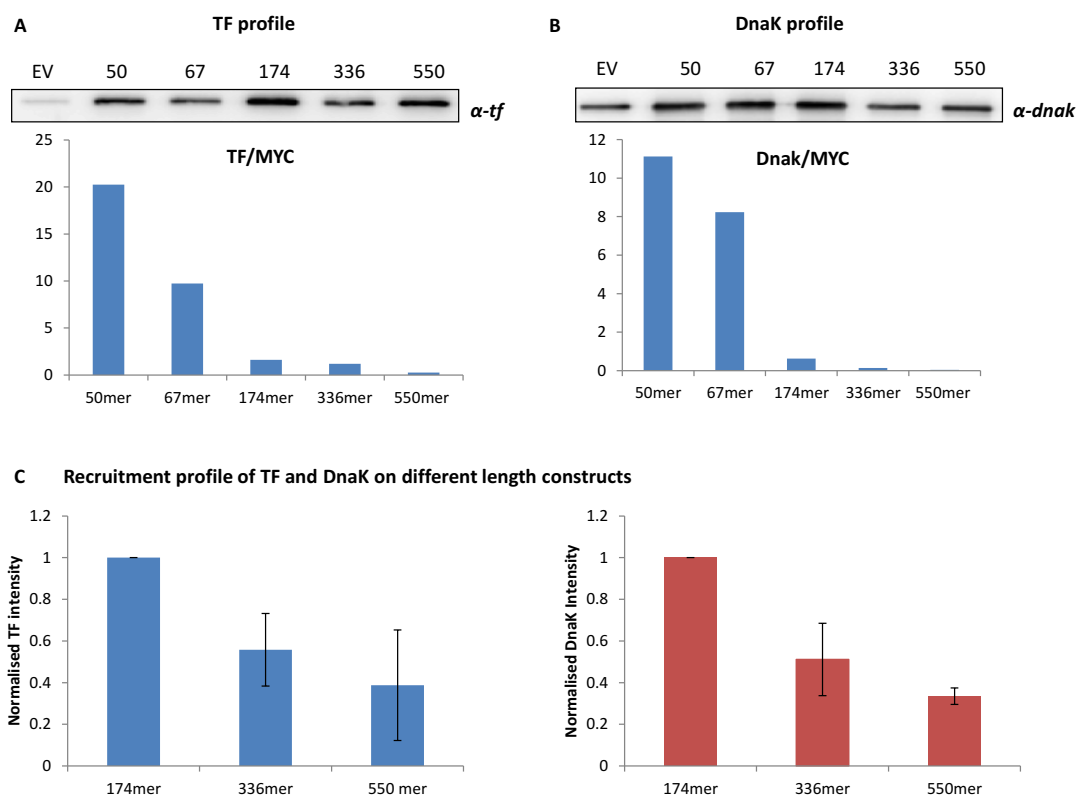


Figure 29: Western blotting analysis of TF and DnaK association with stalled luciferase constructs normalized to nascent chain expression levels. **A:** Chaperone recruitment profile of TF interaction with a linear decrease in association shown with increasing nascent chain length stalled on the ribosome, corroborating the data available from *in vitro* studies (Rutkowska et al. 2008) on higher affinity of TF to nascent chains associated with ribosomes. **B:** Chaperone recruitment profile of DnaK interaction showing a linear decrease in association with increasing nascent chain length stalled on the ribosome, which indicates its role in folding of nascent peptides translating on the ribosome surface. **C:** Standard deviations from at least three independent experiments are shown. The error bar is lowest for intermediate chain length RNC complex (174mer construct), which suggests that upstream chaperones, i.e. TF and DnaK, exhibit the most stable interaction with nascent chains of intermediate length.

The *in vivo* data presented in Figure 29 is in agreement with the previous study where luciferase nascent chains up to 280 amino acids were shielded by trigger factor from protease digestion (Tomic et al. 2006). This suggests that trigger factor efficiently binds to nascent chains, which have structurally incomplete domains, delays their premature folding, and prevents them from

aggregation or misfolding. This observation also correlates with the *in vivo* ribosome-profiling data which suggest that trigger factor stably engages with nascent chains of an average length of 100 amino acids (Oh et al. 2011). The nascent chains of this length may undergo partial chain collapse and would require efficient binding to avoid non-specific compaction. This could be part of a potential cellular surveillance mechanism to reduce the possibility of misfolding events for structurally incomplete domains by delaying the folding of multi-domain proteins and to improve their folding yields (Agashe et al. 2004). Both the ribosome and other groups of molecular chaperones can affect the folding of translating nascent chains. The structure of co-translational intermediates depends on the length and sequence composition of the nascent chain emerging from the ribosome tunnel. The results presented here provide insight into the formation of stable co-translational structures on the ribosome.

Further investigation of these stalled nascent constructs expressed in the CK1 strain would give an in-depth insight into relative quantitative changes in different binding factors based on the length of the translating nascent chain. This study broadens our understanding of interactions of novel binding factors (identified here) with specific nascent chains and gives us insight into the process of co-translational protein folding in the presence of various ribosome binding partners.

5. Discussion

5.1 Capturing ribosome nascent chain interactors *in vivo*

Research over the last five decades has yielded many great contributions towards our understanding of how mRNAs are translated into functional polypeptides by ribosomes. With the recent advent of high throughput screening technologies, it has become possible to capture global snapshots of active, translating ribosomes. A large part of this wealth of information is derived from studies conducted in cell-free translation systems (Nirenberg & Matthaei 1961). The control of the efficiency of translation, modification of reaction conditions, suitability for high-throughput strategies, reduced reaction volumes and process time makes this system highly advantageous for studying *in vitro* protein folding (Katzen et al. 2005). The exogenous supply of chaperones to cell-free protein synthesis reactions has provided additional information about chaperone-assisted protein folding. This has been elucidated by studying the enzymatic activity of newly synthesized firefly luciferase in *E.coli* (Kolb et al. 2000). However, much is still unknown about the translation of mRNA fragments in their natural habitat: a living cell. A variety of cellular components, including the ribosome, molecular chaperones, proteases and crowded intracellular milieu modulates the folding mechanisms in physiologically relevant environments (Kuznetsova et al. 2014). In our study, we have utilized the wealth of knowledge on the current state of co-translational protein folding from both *in vitro* cell-free systems and *in vivo* systems (Chadani et al. 2016). This led us to develop a protocol for deciphering a compendium of protein factors involved in the early stage of protein folding, as the nascent polypeptide traverses and emerges out of the ribosome exit tunnel (Voisset et al. 2008).

An *in vivo* approach for generating RNCs presents a unique set of challenges compared to *in vitro* strategies, as it is more difficult to manipulate translational arrest *in vivo*. The main challenge of the *in vivo* approach is to keep the enormous protein complex consisting of ribosomes, translating

nascent chains, chaperones, and transcription and translation protein factors intact (Gething & Sambrook 1992).

The polypeptide chain emerging from the ribosome is highly dynamic and may interact with various protein factors in the cytosol. As soon as a nascent chain emerges out of the ribosome exit tunnel, it is acted upon by a series of well investigated chaperone systems such as TF, DnaK/J and GroEL. The physical properties of the nascent chains, including length and sequence composition, are responsible for the recruitment of these chaperones and other binding factors at various stages of emergence from the ribosome. *In vivo* experiments have demonstrated that TF and the DnaK system improve the folding yield of certain multi-domain proteins by stimulating a post-translational folding mechanism, resulting in a substantial deceleration of folding relative to translation (Agashe et al. 2004). The affinity of DnaK for unfolded proteins is regulated by the binding and hydrolysis of ATP, resulting in nucleotide-dependent substrate binding and release. This nucleotide-dependent cycling is controlled by the co-factors DnaJ and GrpE. A different mechanism exhibited by GroEL, which shows ATP-regulated substrate binding and release regulated by its co-factor GroES, also aids protein folding (Ying et al. 2005). These established and well-investigated chaperone groups act as control points to ensure the nascent chain folding and protection of unfolded hydrophobic patches from misfolding (Figure 30).

Our protocol design for RNC isolation combined with proteome wide analysis of RNC fraction using SILAC mass spectrometry approach, gives us the advantage of observing the role of TF, DnaK/J and GroEL chaperones in housekeeping of nascent proteome in *E.coli*. In addition, we have shown in this study that the nascent chain transiently interacts with several other potential binding partners of the protease and chaperone families or proteins which have chaperone like function, such as GreA transcription factor. (Figure 26E and 26F). These factors make contact with translating nascent chain on ribosome, thus stabilizing the RNCs, preventing aggregation of nascent polypeptides and initiating the folding process (Hoffmann et al. 2012).

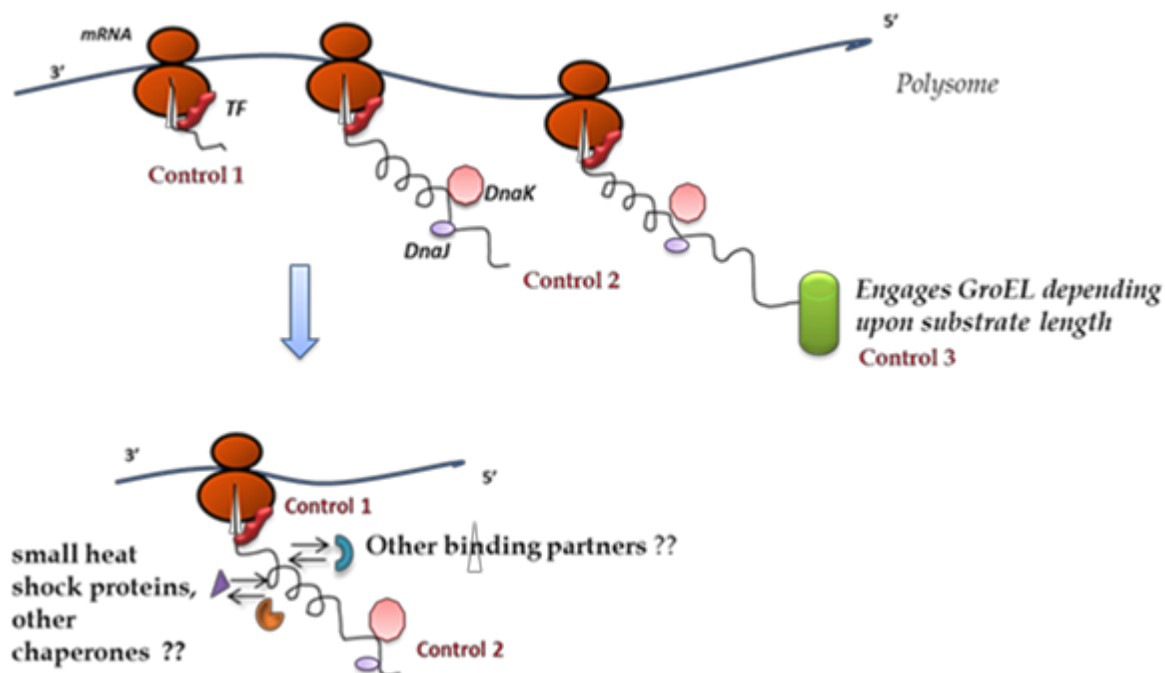


Figure 30: Model of engagement sites for novel RNC interactors that keep the nascent proteome composition intact during translation on ribosomes.

Other known chaperone groups also assist in the process of nascent protein folding on the ribosome, yet a detailed molecular understanding of the underlying interplay between chaperones and other binding factors in the nascent chain folding pathway remains to be explored. In this study, we have shed light on potential interactions between different protein groups such as chaperones, proteases, nascent chain modifying enzymes and proteins involved in translocation of membrane proteins with newly emerging nascent chains on the ribosome. It is essential to understand the role of these proteins since they contribute towards keeping the nascent proteome composition intact while nascent polypeptides traverse and emerge out of the ribosomal exit tunnel. The results of this study provide qualitative insights into the organization of proteomic factors associated with the co-translational protein folding network *in vivo*.

5.2 Emergence of new tools to study interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes

As nascent polypeptide chains exit the ribosome, they are engaged by a series of processing, targeting and folding factors. Recently, several methods have been developed to globally monitor the stages when these factors engage with polypeptides in the complex cellular environment. One of the most commonly used methods is ribosome profiling developed by Nick Ingolia and Jonathan Weissman (Ingolia et al. 2012). In this method, a specialized messenger RNA sequencing is employed to identify the actively translated mRNAs. Ribosome profiling involves a sequencing, library preparation and data analysis strategy similar to the RNA-Seq, but unlike RNA-Seq, this technique allows to distinguish the actively translated mRNAs regions and measure translation levels in each of these regions, to gain insights into global gene expression (Ingolia et al. 2009). On the contrary, our method focuses on proteome-wide analysis of binding factors associated to RNC complex by pulling down FLAG-tagged ribosomes from *E.coli* and using antibiotics. SILAC based mass spectrometry analysis of RNC complexes generated using affinity tagged purification coupled with antibiotic treatment and *in vivo* crosslinking emerges as a potential tool to study the interaction of chaperones and targeting factors with the RNC complex at a proteomic level. In this study, we have used *E.coli* as a model system in which translation and transcription is coupled. This method therefore reveals factors involved both in the transcription and translation processes. Since translocation is a co-translational event in prokaryotes, we anticipated to observe proteins from the translocation complex co-isolated with RNCs. The dataset obtained in this study (Appendix 7.3.2) correlates with the findings obtained using the ribosome profiling technique.

Prior to its development, efforts to measure translation *in vivo* included microarray analysis on the RNA isolated from polysomes, and translational profiling through the affinity purification of epitope-tagged ribosomes. These were useful and complementary methods, but none of them allowed high sensitivity or positional information. However, combining ribosome profiling with affinity purification of ribosomes whose nascent chains are associated with a factor of interest,

for example TF, it becomes possible to quantitatively identify substrate pools and points of RNC engagement (Becker et al. 2013). This method is called selective ribosome profiling (SeRP). Selective ribosome profiling has been adapted to monitor the action of co-translational chaperones and to measure the consequence of their deletion. Further, footprints from complexes of ribosomes and nascent polypeptides that are cross-linked to trigger factor showed that trigger factor engages nascent proteins well after their emergence from the ribosome (Oh et al. 2011). Another study using SeRP in a mammalian system measured the folding of the nascent chain itself and identified some structures that formed progressively in emerging proteins and others that formed in a concerted manner (Han et al. 2012).

However, there are several limitations of this method. The most important has been its focus on the rate of protein production and not on the total abundance of a protein. Unbiased proteomic measurements have advanced remarkably in recent years and can provide information that complements the limitations of ribosome profiling. Mass spectrometry is the simplest way to measure total protein abundance, since translation directly corresponds to protein synthesis, which is the change in protein abundance over time (Vogel & Marcotte 2012).

Therefore, our study focuses on establishing a dataset of protein factors associated with the process of translation in a bacterial cell. As shown in Figures 18, 19 and 20 of the Results section, we use a unique combination of antibiotic-mediated translation inhibition mechanisms together with chemical cross-linking. Further, quantitative proteomic analysis is done by stable isotope labeling of amino acids in cell culture (SILAC)-based mass spectrometry. Strong bioinformatics tools and biochemical assays conjointly formulate the background of our co-translational protein folding project. Apart from the use of a 'global monitoring' approach towards understanding of translation factors or translome, such as the methods of SeRP and mass spectrometry, there are other molecular biology and structural tools to further strengthen the information obtained from the global analysis. To enable homogeneous translational arrest *in vivo*, introduction of the 17 amino acid motif derived from SecM into the gene of interest has proved to be instrumental in the development of RNC studies by NMR and FRET (Cabrita et al. 2009). Selective stalling of

nascent chains on *E. coli* ribosomes for structural studies using cryoEM has also been achieved using sequences derived from the *E. coli* tryptophanase operon, TnaC (Wilson et al. 2011). Figure 31 gives an overview of approaches currently available to gain insights into co-translational protein folding mechanisms. In-depth experimentation with ribosome nascent chain complexes, both *in vivo* and *in vitro*, by employing the latest methodologies gives us insight into the convoluted pathway of the early stages of protein folding on the ribosomes surface. Therefore, understanding the principles of this basic biological phenomenon would yield long-term benefits in the field of protein misfolding and diseases.

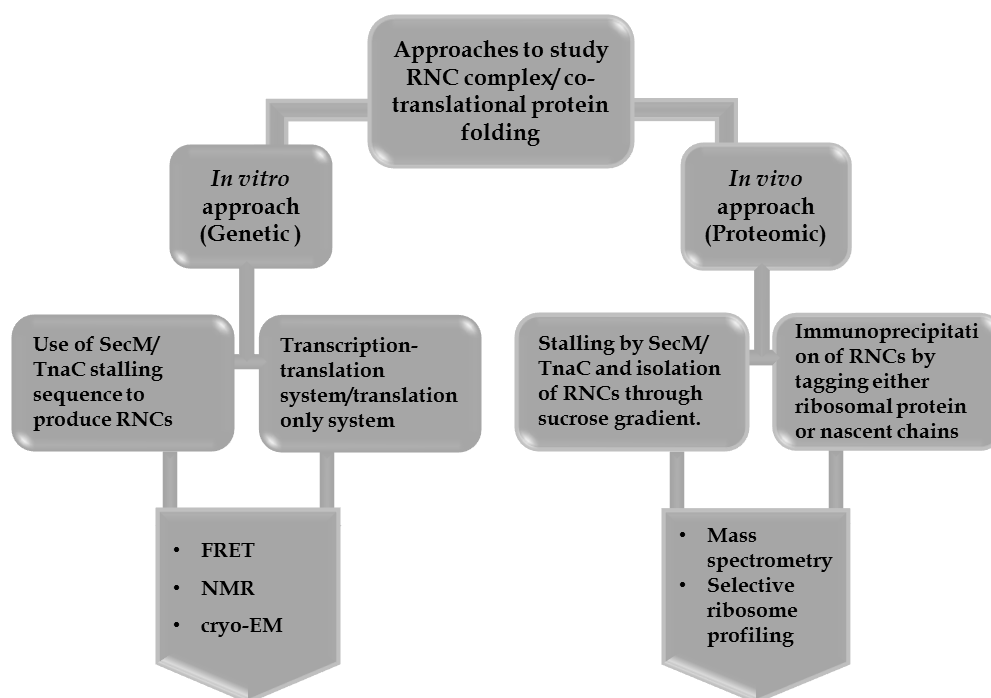


Figure 31: Overview of methods and tools available to study RNCs

5.3 Compendium of factors associated with quality control of nascent proteome

Co-isolation of factors associated with the RNC complex reveals nascent chain associated binding factors of hitherto uncharacterized proteins. The classical way of isolating ribosomes by repeated steps of centrifugation includes large dilutions of the sample, which could result in the loss of

factors that bind with low affinities. Therefore, the method developed in the present study allows isolation of ribosomes at high concentrations through a pull-down procedure of FLAG affinity tagged ribosomes and further stabilization of the complex using chloramphenicol. Utilizing the sophisticated SILAC experimental design and further qualitatively analyzing the dataset using bioinformatics tools, we have been able to identify groups of protein factors co-operating in cotranslational process in *E.coli* (Figure 32). These factors are involved in the process of chaperoning, translocation and RNA quality control, which potentially play a significant role in maintaining the composition of nascent chain integrity on intact ribosomes during translation in *E.coli*. Nascent polypeptides emerging from the ribosomal exit tunnel are susceptible to misfolding and aggregation in the crowded cellular milieu. To circumvent this problem, cells have evolved a complex machinery of molecular chaperones. Trigger factor, as we know, is the first to make contact with the nascent chains in bacteria. This factor prevents misfolding and assists the co- and post-translational folding of nascent polypeptides to their functional state. Enrichment of TF on RNCs stabilised by CAM and crosslinker justifies the use of our novel alternative method for RNC isolation. Further, identification of downstream chaperones Dnak/J and GroEL, which are known to co-operate with upstream chaperone TF and keep the proteins in unfolded translating state on ribosomes, corroborates our experimental approach (Ying et al. 2006). In this study, we have developed a unique approach to understand the early folding events in the life of the nascent polypeptide, before it is released from the ribosome. Our dataset shows transiently binding partners of RNCs in *E.coli* such as ClpB chaperone, HscA and HtpG. The enrichment of these chaperones on translating ribosomes indicate that nascent chains are susceptible to misfolding and require intensive chaperoning to maintain their integrity while translating. The role of these factors as RNC binding partners have not been explored yet and further investigation using the *in vitro* cell-free translation system would give new insights into their functional activity and time of engagement by nascent proteins.

Transcription and translation are coupled in *E.coli* and therefore, immunoprecipitation of FLAG peptide tagged ribosomes from *E.coli* CK1 strain results in the isolation of translating complexes

as well as transcription-associated proteins. The global analysis of translating complexes in *E. coli* using a quantitative mass spectrometry approach gives us insight into these factors that play an important role in the mRNA quality control, such as transcription regulatory protein ompR. The N-terminus of this protein controls the transcriptional expression of both major outer membrane protein genes, ompF and ompC. We also found enrichment of outer membrane proteins ompC (Appendix 7.3.2, blue highlight) in both the crosslinked and non-crosslinked datasets. These results indicate that our method explores the inventory of factors associated with transcription regulation and translation processes.

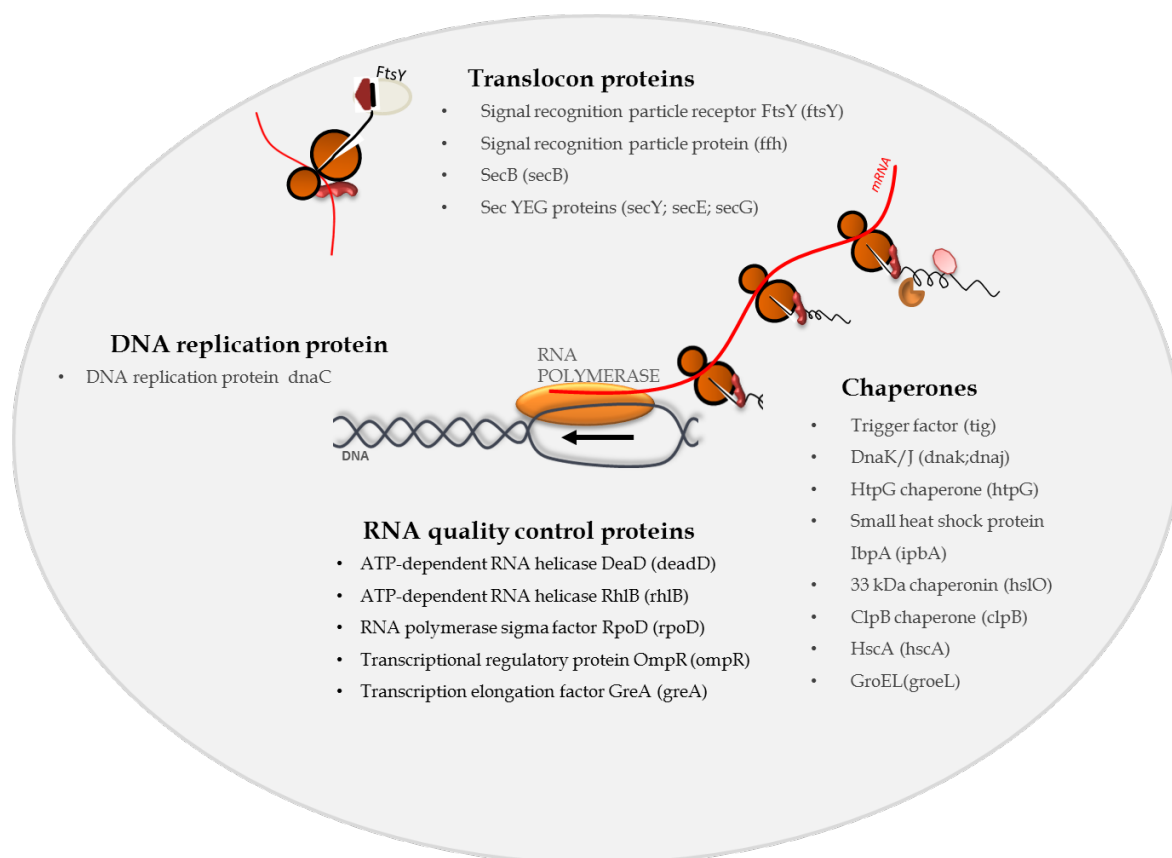


Figure 32: Groups of notable protein factors associated with RNC complex deciphered through global analysis of RNC by quantitative mass spectrometry.

It has recently been shown that nascent chains are co-translationally engaged by various ribosome-associated factors (Becker et al. 2013). These factors, such as PDF and MAP, SRP, SRP

receptor and SEC complex, are competently functional in binding to the ribosome near the ribosomal exit tunnel. These factors participate in the processing and modification of nascent chain N-termini and they assist in co-translational protein folding and targeting of the nascent chain or exported proteins to the translocation machinery. Thus, they cooperate functionally in keeping the nascent proteome homeostasis. However, a detailed *in vivo* understanding of the mechanistic aspects of their function and their interaction with the ribosome has not yet been entirely explored. To understand these processes and their dynamic interplay, it will be essential to use interdisciplinary experimental approaches, combining *in vivo* biochemical studies with detailed structural studies ranging from crystallography and EM to global proteomic analysis using mass spectrometry and ribosome profiling. Owing to the complexity of these processes and the large size of the ribosome, these experiments present formidable technical challenges. Our study contributes towards overcoming technical challenges and shows stepwise protocol development for stable RNC isolation, which is specific, rapid and avoids strong dilution of the samples prior to elution. Batch sample handling allows the analysis of multiple samples simultaneously. When ribosomes isolated from *E.coli* were treated with antibiotics to either stabilize or destabilize ribosome nascent chain complexes, marked differences in the protein patterns of the isolated ribosomal samples were observed. Thus, by selecting the appropriate conditions as shown here, it is possible to categorize factors that specifically co-fractionate with RNCs. Ribosomal proteomes can be defined under different conditions. In future, it will be of interest to see how these proteomes change in response to stress conditions that affect general protein folding, such as elevated temperatures, nutrient depletion and lack of specific chaperones. In addition, the role of ribosome-nascent chain associated proteins can also be studied during the translation of any specific substrate for a chaperone of interest, since this system is amenable to expression of any model. The gene constructs of the luciferase model protein and methods described here for *in vivo* studies may prove useful for identifying factors that function during the earliest steps of protein biogenesis. The experiments performed in this study have provided a detailed real-time view of the dynamic nature of co-translational protein folding in bacteria.

6. References

- Agashe, V.R. et al., 2004. Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. *Cell*, 117(2), pp.199–209.
- Angelini, S., Deitermann, S. & Koch, H.-G., 2005. FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *EMBO reports*, 6(5), pp.476–481.
- Atlung, T. et al., 1996. The histone-like protein H-NS acts as a transcriptional repressor for expression of the anaerobic and growth phase activator AppY of *Escherichia coli*. *Journal of Bacteriology*, 178(12), pp.3418–3425.
- Azzam, M.E. & Algranati, I.D., 1973. Mechanism of puromycin action: fate of ribosomes after release of nascent protein chains from polysomes. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12), pp.3866–3869.
- Balchin, D., Hayer-Hartl, M. & Hartl, F.U., 2016. In vivo aspects of protein folding and quality control. *Science*, 353(6294), pp.aac4354–aac4354. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.aac4354>.
- Bashan, A. & Yonath, A., 2008. Correlating ribosome function with high-resolution structures. *Trends in Microbiology*, 16(7), pp.326–335.
- Becker, A.H. et al., 2013. Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nature protocols*, 8(11), pp.2212–39. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24136347>.
- Berisio, R. et al., 2003. Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nature structural biology*, 10(5), pp.366–370.
- Bhushan, S. et al., 2010. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Molecular Cell*, 40(1), pp.138–146. Available at: <http://dx.doi.org/10.1016/j.molcel.2010.09.009>.
- Bornemann, T., Holtkamp, W. & Wintermeyer, W., 2014. biogenesis factors on the ribosome. *Nature Communications*, 5(May), pp.1–8. Available at: <http://dx.doi.org/10.1038/ncomms5180>.
- Braig, K. et al., 1993. A polypeptide bound by the chaperonin groEL is localized within a central cavity. *Proceedings of the National Academy of Sciences of the United States of America*, 90(9), pp.3978–82.
- Braig, K. et al., 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature*,

- 371(6498), pp.578–86.
- Brandt, F. et al., 2009. The Native 3D Organization of Bacterial Polysomes. *Cell*, 136(2), pp.261–271. Available at: <http://dx.doi.org/10.1016/j.cell.2008.11.016>.
- Cabrita, L.D. et al., 2009. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), pp.22239–22244.
- Calloni, G. et al., 2012. DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Reports*, 1(3), pp.251–264. Available at: <http://dx.doi.org/10.1016/j.celrep.2011.12.007>.
- Cassaignau, A.M.E. et al., 2016. A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nature Protocols*, 11(8), pp.1492–1507. Available at: <http://www.nature.com/doi/10.1038/nprot.2016.101>.
- Chadani, Y. et al., 2016. Integrated in vivo and in vitro nascent chain profiling reveals widespread translational pausing. *Proceedings of the National Academy of Sciences of the United States of America*, 113(7), pp.E829–38.
- Chen, R. & Henning, U., 1996. A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. *Molecular Microbiology*, 19(6), pp.1287–1294.
- Croons, V. et al., 2008. Differential effect of the protein synthesis inhibitors puromycin and cycloheximide on vascular smooth muscle cell viability. *J Pharmacol Exp Ther*, 325(3), pp.824–832.
- Das, D. et al., 2008. Role of the ribosome in protein folding. *Biotechnology Journal*, 3(8), pp.999–1009.
- Delvillani, F. et al., 2011. S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Research*, 39(17), pp.7702–7715.
- Deuerling, E. et al., 2003. Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Molecular Microbiology*, 47(5), pp.1317–1328.
- Dobson, C.M., 2004. Principles of protein folding, misfolding and aggregation. *Seminars in Cell and Developmental Biology*, 15(1), pp.3–16.
- Dunkle, J.A. et al., 2010. Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proceedings of the National Academy of Sciences*, 107(40), pp.17152–17157. Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20876128&retmode=ref&cmd=prlinks\npapers3://publication/doi/10.1073/pnas.1007988107>.
- Duval, M. et al., 2013. Escherichia coli Ribosomal Protein S1 Unfolds Structured mRNAs Onto the Ribosome for Active Translation Initiation. *PLoS Biology*, 11(12), pp.12–14.
- Dzionara, M., Kaltschmidt, E. & Wittmann, H.G., 1970. Ribosomal Proteins, XIII. Molecular Weights of Isolated Ribosomal Proteins of Escherichia coli. *Proceedings of the National*

- Academy of Sciences*, 67(4), pp.1909–1913.
- Economou, A., 1999. Following the leader: Bacterial protein export through the Sec pathway. *Trends in Microbiology*, 7(8), pp.315–320.
- Ehrnsperger, M. et al., 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. , 16(2), pp.221–229.
- Endo, A. & Kurusu, Y., 2007. Identification of in vivo substrates of the chaperonin GroEL from *Bacillus subtilis*. *Bioscience, biotechnology, and biochemistry*, 71(4), pp.1073–1077.
- Evans, M.S. et al., 2005. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nature methods*, 2(10), pp.757–762.
- Facey, S.J. & Kuhn, A., 2004. Membrane integration of E. coli model membrane proteins. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp.55–66.
- Fenton, W. a & Horwich, a L., 1997. GroEL-mediated protein folding. *Protein science : a publication of the Protein Society*, 6(4), pp.743–760.
- Frauenfeld, J. et al., 2012. Membrane Environment. , 18(5), pp.614–621.
- Frydman, J., 2001. F OLDING OF N EWLY T RANSLATED P ROTEINS I N V IVO : The Role of Molecular Chaperones.
- Fujiwara, K. et al., 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. *The EMBO journal*, 29(9), pp.1552–64.
- Genevaux, P. et al., 2004. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO reports*, 5(2), pp.195–200.
- Gething, M.J. & Sambrook, J., 1992. Protein folding in the cell. *Nature*, 355(6355), pp.33–45.
- Glover, J.R. & Lindquist, S., 1998. Hsp104, Hsp70, and Hsp40: A novel chaperone system that rescues previously aggregated proteins. *Cell*, 94(1), pp.73–82.
- Gong, F. & Yanofsky, C., 2002. Instruction of translating ribosome by nascent peptide. *Science (New York, N.Y.)*, 297(5588), pp.1864–1867.
- Gumbart, J. et al., 2012. Mechanisms of SecM-mediated stalling in the ribosome. *Biophysical Journal*, 103(2), pp.331–341.
- Gupta, A.J. et al., 2014. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *Journal of molecular biology*, 426(15), pp.2739–54.
- Han, Y. et al., 2012. Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proceedings of the National Academy of Sciences*, 109(31), pp.12467–12472.
- Hartl, F.U., Bracher, A. & Hayer-Hartl, M., 2011. Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356), pp.324–32.
- Hartl, F.U. & Hayer-Hartl, M., 2009. Converging concepts of protein folding in vitro and in vivo.

- Nature structural & molecular biology*, 16(6), pp.574–581.
- Hayer-hartl, M. & Hartl, F.U., 2015. Accelerates the Rate of TIM-Barrel Domain Folding. , 157(4), pp.922–934.
- Hirokawa, G. et al., 2004. In vivo effect of inactivation of ribosome recycling factor - Fate of ribosomes after unscheduled translation downstream of open reading frame. *Molecular Microbiology*, 54(4), pp.1011–1021.
- Hoffmann, A. et al., 2012. Concerted Action of the Ribosome and the Associated Chaperone Trigger Factor Confines Nascent Polypeptide Folding. *Molecular Cell*, 48(1), pp.63–74.
- Horwich, A.L. et al., 2007. Two families of chaperonin: physiology and mechanism. *Annual review of cell and developmental biology*, 23, pp.115–145.
- Huang, C. et al., 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature*, pp.1–20.
- Huang, D.W., Sherman, B.T. & Lempicki, R. a, 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), pp.44–57.
- Hundley, H. a et al., 2005. Human Mpp11 J protein: ribosome-tethered molecular chaperones are ubiquitous. *Science (New York, N.Y.)*, 308(5724), pp.1032–1034.
- Ingolia, N.T. et al., 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, 324(5924), pp.218–223.
- Ingolia, N.T. et al., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, 7(8), pp.1534–1550.
- Jahn, T.R. & Radford, S.E., 2008. Folding versus aggregation: Polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics*, 469(1), pp.100–117. Available at: <http://dx.doi.org/10.1016/j.abb.2007.05.015>.
- Jones, P.G. & Inouye, M., 1996. RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. *Molecular microbiology*, 21, pp.1207–1218. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8898389>.
- Katzen, F., Chang, G. & Kudlicki, W., 2005. The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3), pp.150–156.
- Kerner, M.J. et al., 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell*, 122(2), pp.209–220. Available at: <http://dx.doi.org/10.1016/j.cell.2005.05.028>.
- Kerner, M.J. et al., 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell*, 122(2), pp.209–20.
- Kolb, V. a., Makeyev, E. V. & Spirin, A.S., 2000. Co-translational folding of an eukaryotic

- multidomain protein in a prokaryotic translation system. *Journal of Biological Chemistry*, 275(22), pp.16597–16601.
- Kramer, G. et al., 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature structural & molecular biology*, 16(6), pp.589–597.
- Kuznetsova, I., Turoverov, K. & Uversky, V., 2014. *What Macromolecular Crowding Can Do to a Protein*, Available at: <http://www.mdpi.com/1422-0067/15/12/23090/>.
- Langer, T., Pfeifer, G., et al., 1992. Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *The EMBO journal*, 11(13), pp.4757–4765.
- Langer, T., Lu, C., et al., 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, 356(6371), pp.683–689.
- Laursen, B.S. et al., 2005. Initiation of protein synthesis in bacteria. *Microbiology and ...*, 236(3), pp.747–771. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1996.00747.x/full> \n<http://mibr.asm.org/content/69/1/101.short>.
- Li, G.-W., Oh, E. & Weissman, J.S., 2012. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484(7395), pp.538–541. Available at: <http://dx.doi.org/10.1038/nature10965>.
- Li, K. et al., 2012. Transcription Elongation Factor GreA Has Functional Chaperone Activity. *PLoS ONE*, 7(12), pp.1–8.
- Lin, K.F. et al., 2012. Cotranslational protein folding within the ribosome tunnel influences trigger-factor recruitment. *Biophysical Journal*, 102(12), pp.2818–2827. Available at: <http://dx.doi.org/10.1016/j.bpj.2012.04.048>.
- von Loeffelholz, O., Botte, M. & Schaffitzel, C., 2011. *Escherichia coli* Cotranslational Targeting and Translocation. *Encyclopedia of Life Sciences*, pp.1–8. Available at: <http://doi.wiley.com/10.1002/9780470015902.a0023170>.
- Mann, M., 2006. Functional and quantitative proteomics using SILAC. *Nature reviews. Molecular cell biology*, 7(12), pp.952–958.
- Martin Gamerding, Marie Anne Hanebuth, Tancred Frickey, E.D., 2015. The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science*, 348(6231), pp.201–207.
- Matuszewska, E. et al., 2008. *Escherichia coli* heat-shock proteins IbpA/B are involved in resistance to oxidative stress induced by copper. *Microbiology*, 154(6), pp.1739–1747.
- Mayer, M.P. & Bukau, B., 2005. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6), pp.670–684.
- Mogk, A., Huber, D. & Bukau, B., 2011. Integrating protein homeostasis strategies in

- prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 3(4), pp.1–19.
- Montfort, R.L.M. Van et al., 2001. letters Crystal structure and assembly of a eukaryotic small heat shock protein. , 8(12), pp.1025–1030.
- Mori, H. & Ito, K., 2001. The Sec protein-translocation pathway. *Trends in Microbiology*, 9(10), pp.494–500.
- Nakamoto et al., 2014. Physical interaction between bacterial heat shock protein (Hsp) 90 and Hsp70 chaperones mediates their cooperative action to refold denatured proteins. *Journal of Biological Chemistry*, 289(9), pp.6110–6119.
- Nakatogawa, H. et al., 2005. SecM facilitates translocase function of SecA by localizing its biosynthesis. *Genes and Development*, 19(4), pp.436–444.
- Nakatogawa, H. & Ito, K., 2002. The ribosomal exit tunnel functions as a discriminating gate. *Cell*, 108(5), pp.629–636.
- Nc, M. et al., 2004. Letters To Nature. *Nature*, 431(September), pp.9396–9401.
- Netzer, W.J. & Hartl, F.U., 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*, 388(6640), pp.343–349.
- Nirenberg, M.W. & Matthaei, J.H., 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 47(10), pp.1588–1602. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC223178/>.
- Nishihara, K., Kanemori, M. & Yanagi, H., 2000. Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia coli* Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia coli*. *Applied and Environmental Microbiology*, 66(3), pp.884–889.
- Oh, E. et al., 2011. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, 147(6), pp.1295–1308.
- Patzelt, H. et al., 2001. Binding specificity of *Escherichia coli* trigger factor. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25), pp.14244–14249.
- Du Plessis, D.J.F., Nouwen, N. & Driessen, A.J.M., 2011. The Sec translocase. *Biochimica et Biophysica Acta - Biomembranes*, 1808(3), pp.851–865.
- Qin, D. & Fredrick, K., 2013. *Analysis of polysomes from bacteria* 1st ed., Elsevier Inc. Available at: <http://dx.doi.org/10.1016/B978-0-12-420037-1.00008-7>.
- Randall, L.L. et al., 2005. Asymmetric binding between SecA and SecB two symmetric proteins: Implications for function in export. *Journal of Molecular Biology*, 348(2), pp.479–489.
- Raney, A. et al., 2002. Regulated translation termination at the upstream open reading frame in S-adenosylmethionine decarboxylase mRNA. *Journal of Biological Chemistry*, 277(8),

- pp.5988–5994.
- Rappsilber, J., Mann, M. & Ishihama, Y., 2007. Protocol for micro-purification , enrichment , pre-fractionation and storage of peptides for proteomics using StageTips. , (Step C).
- Rutkowska, A. et al., 2008. Dynamics of trigger factor interaction with translating ribosomes. *Journal of Biological Chemistry*, 283(7), pp.4124–4132.
- Rutkowska, A. et al., 2009. Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Letters*, 583(14), pp.2407–2413. Available at: <http://dx.doi.org/10.1016/j.febslet.2009.06.041>.
- Saibil, H. et al., 1991. Binding of chaperonins. *Nature*, 353(6339), pp.25–26.
- Sandikci, A. et al., 2013. Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding. *Nature Publishing Group*, 20(7), pp.843–850. Available at: <http://dx.doi.org/10.1038/nsmb.2615>.
- Saper, M. a, 2001. ° Crystal Structure of Hsp33: The 2.2 Å Heat Shock Protein with Redox-Regulated Chaperone Activity. *Shock*, 9(01), pp.367–375.
- Saraogi, I. & Shan, S., 2011. Molecular mechanism of co-translational protein targeting by the signal recognition particle. *Traffic (Copenhagen, Denmark)*, 12(5), pp.535–542.
- Schibich, D. et al., 2016. Global profiling of SRP interaction with nascent polypeptides. *Nature*, pp.1–16. Available at: <http://www.nature.com/doi/10.1038/nature19070>.
- Schirmer, E.C. et al., 1996. HSP100/Clp proteins: A common mechanism explains diverse functions. *Trends in Biochemical Sciences*, 21(8), pp.289–296.
- Simonetti, a. et al., 2009. A structural view of translation initiation in bacteria. *Cellular and Molecular Life Sciences*, 66(3), pp.423–436.
- Sklar, J.G. et al., 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes and Development*, 21(19), pp.2473–2484.
- Spevak, C.C., Ivanov, I.P. & Sachs, M.S., 2010. Sequence requirements for ribosome stalling by the arginine attenuator peptide. *Journal of Biological Chemistry*, 285(52), pp.40933–40942.
- Sunohara, T. et al., 2004. Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. *RNA (New York, N.Y.)*, 10(3), pp.378–386.
- Teter, S.A. et al., 1999. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell*, 97(6), pp.755–765.
- Tomic, S. et al., 2006. Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. *FEBS Letters*, 580(1), pp.72–76.
- Vabulas, R.M. et al., 2010. Heat Shock Response. , pp.1–18.
- Veenendaal, A.K.J., Van Der Does, C. & Driessen, A.J.M., 2004. The protein-conducting channel

-
- SecYEG. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp.81–95.
- Villa, E. et al., 2009. Translational Stalling. *Metro*, 1621(May), pp.1412–1415.
- Vince, R. et al., 1975. Chloramphenicol binding site with analogues of chloramphenicol and puromycin. *Antimicrobial agents and chemotherapy*, 8(4), pp.439–443.
- Vogel, C. & Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4), pp.227–232. Available at: <http://dx.doi.org/10.1038/nrg3185>.
- Voisset, C. et al., 2008. Tools for the study of ribosome-borne protein folding activity. *Biotechnology Journal*, 3(8), pp.1033–1040.
- Wilson, D. et al., 2011. Nascent polypeptide chains within the ribosomal tunnel analyzed by cryo-EM. *Ribosomes: Structure, Function, and Dynamics*, 23, pp.393–404. Available at: <Go to ISI>://WOS:000299455400031.
- Woldringh, C.L., 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Molecular Microbiology*, 45(1), pp.17–29.
- Xu, Z., Horwich, A.L. & Sigler, P.B., 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)₇ chaperonin complex. *Nature*, 388(6644), pp.741–50.
- Ying, B.W. et al., 2005. Co-translational involvement of the chaperonin GroEL in the folding of newly translated polypeptides. *Journal of Biological Chemistry*, 280(12), pp.12035–12040.
- Ying, B.W., Taguchi, H. & Ueda, T., 2006. Co-translational binding of GroEL to nascent polypeptides is followed by post-translational encapsulation by GroES to mediate protein folding. *Journal of Biological Chemistry*, 281(31), pp.21813–21819.
- Yoo, S.J. et al., 1996. Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in *Escherichia coli*. *Journal of Biological Chemistry*, 271(24), pp.14035–14040.

7. Appendix

7.1 Abbreviations

aa	Amino acid	mRNA	Messenger RNA
ADP	Adenosine 5'-diphosphate	NAC	Nascent chain associated complex
Amp	Ampicillin	PAGE	Polyacrylamide Gel Electrophoresis
ATP	Adenosine 5'-triphosphate	PBS	Phosphate buffer saline
BSA	Albumin bovine serum	PCR	Polymerase chain reaction
CAM	Chloramphenicol	PDB	Protein data bank
DMSO	Dimethyl-sulfoxide	PPIase	Peptidyl prolyl isomerase
DNA	Deoxyribonucleic acid	PURO	Puromycin
DNase	Desoxyribonuclease	RAC	Ribosome associated complex
DSP	Dithiobis succinimidyl propionate	RecA	Recombinase A
DTT	Dithiothreitol	RNA	Ribonucleic acid
IAA	Idoacetamide	RNC	Ribosome nascent chain complex
FA	Formic acid	rRNA	Ribosomal RNA
E.coli	Escherichia coli	tRNA	Transfer RNA
EB	Elution buffer	SCX	Strong cation exchange
EDTA	Ethylenediaminetetraacetic acid	SDS	Sodiumdodecylsulfate
FL	Firefly luciferase	TCA	Trichloroacetic acid
FRET	Fluorescence resonance energy transfer	TCEP	Tris-(2-carboxyethyl)phosphine
HCD	High energy collisional dissociation	TEMED	N,N,N',N'-tetramethylethylenediamine
Hsp	Heat shock protein	Tet	Tetracycline
IPTG	Isopropyl- β -D-1-thiogalactopyranoside	TF	Trigger factor
Kan	Kanamycin	tig	Gene encoding TF
LB	Luria bertani	TRiC	TCP1 ring complex Tris(hydroxymethyl)aminomethane
LC-MS	Liquid chromatography-mass spectroscopy	Tris-HCl	hydrochloride
Luc	Luciferase	WT	Wild type
MreB	Rod shape determining protein		

7.2 List of primers

BglII incorporation and formation of LucSecM constructs		T_m (°C)
50 Rev	GGAAGATCTTGCATCTGTAAAAGCAATTGTTCCA	56
67 Rev	GAGAGTTTTCACTGCATACAGATCTGACGATTCTGTGATTTG	60
174 Rev	CAAAATCGTATTCATTAAGATCTAAACCGGGAGGTAGATG	58
336 Rev	CGTATCCCTGGAAGAAGATCTTGGGAAGCGTTTTGC	61
SecM Fwd	GCGGATCTGGTACTAGAGATCTTAAATTCAGCACGCCC	63
Prokaryotic Fluc primers for reading the Luciferase gene.		
Ara Fwd	ATCACGGCAGAAAAGTC	55
Fluc I fwd	ACATATCGAGGTGAACATCACGTACGCGGAATACT	59
Fluc II fwd	AATTATTATCATGGATTCTAAAACGGATTACCAG	52
Fluc III fwd	CCATCACGGTTTTGGAATGTTTACTACT	54
Fluc IV fwd	TCCAGGGATACGACAAGGATATGGGCTCAC	59
Fluc V fwd	TTACAACACCCCAACATCTTCGACGCGGGCG	62

7.3 List of enriched protein ratios obtained from crosslinked and non-crosslinked datasets.

7.3.1 List of 245 enriched protein ratios obtained from CK1 FLAG tag strain versus C wild type strain.

Ribosome-nascent chains (RNCs) interactors sorted according to alphabetical order of gene name; CK1/C ratio, enrichment of the RNCs interactors in FLAG-tagged L17 CK1 strain compared to the background non-FLAG tagged C strain as measured by SILAC MS; GRAVY means the average hydrophobicity of protein sequence (Kyte & Doolittle 1982); EG is abbreviation of essential gene (Gerdes et al. 2003).

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
P0A9G7	aceA	47.521	1.866	-0.2241	5.16		homo-oligomer
Q1RF28	allR	29.285	2.883	0.0934	5.77		homo-oligomer
P0AE13	amn	53.994	3.845	-0.2947	5.89		
P05637	apaH	31.296	8.254	-0.2456	5.2		Monomer
P0A9Q3	arcA	27.292	2.026	-0.4869	5.21		
Q8X824	aroB	38.884	12.46	0.1702	5.58		Monomer
P33363	bglX	83.459	1.979	-0.2599	5.85		
P06610	btuE	20.469	4.892	-0.1305	4.81	E	
P61517	can	25.096	2.446	-0.2313	6.16	E	homo-oligomer
P0A9H8	cfa	43.909	8.422	-0.313	5.69		
P31801	chaA	39.168	25.57	0.9077	6.14		
P0ABI1	clpA	84.206	2.187	-0.2978	5.91		homo-oligomer
Q8XCH5	cmoB	37.052	9.169	-0.2507	6.13	E	
P0A9Y1	cspA	7.4032	2.136	-0.2513	5.57		
P22255	cysQ	27.176	3.844	-0.2511	5.59		
P63944	dapA	31.284	1.93	0.0668	5.98	E	homo-oligomer
Q8X9Y6	ddlB	32.853	2.611	0.0837	4.77	E	Monomer
Q8XA87	deaD	70.575	4.09	-0.5862	8.74		hetero-oligomer
P09549	dedD	22.938	3.224	-0.3058	7.8		
Q1R260	deoB	44.392	1.958	-0.2699	5.15		
P37349	dhaM	51.448	2.763	-0.0173	4.61		homo-oligomer
Q8FD98	diaA	21.09	5.514	0.0628	5.44		homo-oligomer

Appendix

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
Q8XBZ3	dnaA	52.54	3.099	-0.362	8.77	E	hetero-oligomer
P08622	dnaJ	41.1	2.168	-0.6342	7.97		homo-oligomer
Q8FJM0	dps	18.753	2.724	-0.2328	5.52		homo-oligomer
P0ABU5	elbB	22.981	3.767	0.2673	4.67		
POC960	emtA	22.226	27.5	-0.2033	9.16		
Q1R7A6	epd	37.329	3.075	-0.0486	6.26		homo-oligomer
Q1R3A2	epmA	36.976	3.291	-0.2913	5.07	E	homo-oligomer
P0A9R5	fdx	12.331	3.633	-0.3567	4.49	E	
P0AGD9	ffh	49.787	2.494	-0.2434	9.51	E	hetero-oligomer
P28861	fpr	27.751	11.58	-0.1943	6.17		Monomer
Q8X5J4	frmA	39.332	12.35	-0.0319	5.84		homo-oligomer
P10121	ftsY	54.513	1.803	-0.3345	4.47	E	
Q8X4P8	fumA	60.297	1.816	-0.3554	6.11		homo-oligomer
Q8X769	fumC	50.461	5.411	-0.1273	6.12		homo-oligomer
P0A9C3	galM	38.19	1.825	-0.4661	4.84		
P09148	galT	39.645	1.821	-0.5243	6		homo-oligomer
P31120	glmM	47.543	2.154	0.0133	5.71		
POAC82	gloA	14.92	3.178	-0.3458	4.95		homo-oligomer
P0A9D0	glpX	35.852	2.105	-0.0163	5.32		homo-oligomer
Q8XD88	gmk	23.65	3.938	-0.4448	5.9	E	homo-oligomer
P77695	gnsB	6.5476	2.786	-0.4806	8.96		
Q8FKM7	gpt	16.969	2.729	-0.0479	5.77		homo-oligomer
POACB6	hemG	21.226	2.535	-0.5872	9.67	E	hetero-oligomer
Q8X736	hflD	22.946	2.07	0.1249	9.68		Monomer
P25519	hflX	48.327	5.82	-0.3323	5.68		Monomer
Q1R388	hfq	11.166	5.311	-0.5048	6.96	E	homo-oligomer
P0ACE8	hinT	13.241	1.863	-0.0621	5.73		homo-oligomer
P0ACG0	hns	15.539	2.426	-0.751	5.43	E	homo-oligomer
P08956	hsdR	134.09	1.86	-0.5032	5.66		hetero-oligomer
P0A6Y7	hslo	32.534	2.558	-0.3129	4.35		
Q0TB20	ibpA	15.774	2.356	-0.5247	5.57		Monomer
P00894	ilvH	17.977	1.82	-0.0545	8		hetero-oligomer
P76071	insH5;	37.777	1.931	-0.5757	9.55		
Q0TEV7	iscA	11.556	2.643	-0.1027	4.74		homo-oligomer
P77338	kefA	127.21	1.923	-0.0311	8.04		
P00803	lepB	35.96	6.808	-0.1558	6.84	E	

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
P37339	lhgO	46.081	2.328	-0.0852	8.66		
Q8XBK2	Int	57.093	2.09	0.2697	9.51		
P75957	lolD	25.438	2.065	-0.1823	7.85	E	
P75958	lolE	45.344	3.895	0.4111	8.92	E	
P0AB39	lpoB	22.515	17.97	-0.2145	6.41		
P69778	lpp	8.3234	3.762	-0.314	9.3		
P23917	mak	32.499	3.636	-0.2019	5.47		
Q8X701	malT	103.02	7.821	-0.2638	6.14		Monomer
Q8FBM1	metE	84.69	1.802	-0.2632	5.61		Monomer
Q1R389	miaA	35.064	2	-0.1578	5.8	E	Monomer
P64602	miaB	10.68	9.76	0.032	4.74		
P64605	miaD	19.576	2.231	-0.0677	4.78		
P65383	moaA	37.304	3.568	-0.3513	7.67		Monomer
P0A9G9	modE	28.281	3.16	-0.0209	5.22		homo-oligomer
P0AF04	mog	21.222	2.977	-0.0132	4.97	E	homo-oligomer
P37773	mpl	49.874	2.802	-0.0507	5.53		
Q8XE45	mgo	60.229	4.072	-0.2559	6.74		
P24205	msbB	37.41	2.679	-0.4497	9.6		
P76270	msrC	18.121	2.043	0.1364	4.65		
Q1R6R1	mug	18.673	2.819	-0.2981	9.16		Monomer
P08373	murB	37.851	1.944	-0.2034	5.77		Monomer
Q8X9Y8	murG	37.8	4.279	0.0392	9.73	E	
P63633	murl	31.074	2.503	0.1758	5.16	E	monomer
P00393	ndh	47.358	2.779	-0.1285	8.96		
Q8XBE7	nudK	21.744	68.84	-0.2941	4.87		homo-oligomer
Q1R9D0	nuoB	25.056	2.384	-0.3204	5.58		hetero-oligomer
P0AFD2	nuoE	18.59	4.475	-0.3915	5.4		hetero-oligomer
P42641	obgE	43.285	1.816	-0.4514	4.74	E	Monomer
P0A913	pal	18.824	2.618	-0.4537	6.29		
P0AFI3	parC	83.83	3.829	-0.3688	6.24	E	hetero-oligomer
P0ABF3	pcnB	53.87	2.207	-0.5304	9.66		Monomer
P29745	pepT	44.923	6.17	-0.2151	5.36		homo-oligomer
P0AFJ3	phnA	12.345	2.195	-0.5323	4.97		
P0AA48	plaP	49.537	3.296	0.7303	8.8		
P0AB69	pntB	48.722	3.368	0.6024	5.72		hetero-oligomer
P00582	polA	103.12	1.871	-0.264	5.39		Monomer
P43671	pqiB	60.52	10.63	-0.2801	7.73		
Q1RCM7	prfA	40.517	4.873	-0.6599	5.14		
P27298	prlC	77.166	1.926	-0.4162	5.15		

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
P45577	proQ	25.892	1.92	-0.7762	9.66		
P23830	pssA	52.801	2.062	-0.4294	9.05	E	hetero-oligomer
Q0TF00	purC	26.995	2.527	-0.3695	5.07		homo-oligomer
P31466	purP	46.865	3.617	0.9966	8.43		
Q1R309	pyrB	34.427	2.282	-0.1044	6.11		hetero-oligomer
P0AFU5	qseF	49.148	11.65	-0.1562	5.99		
P64555	queE	25.029	3.691	-0.2595	5.7		
P0AD51	raiA	12.784	12.78	-0.5353	6.2		
Q8XAX4	ravA	56.387	3.151	-0.2152	6.3		hetero-oligomer
P0AGI3	rbsC	33.451	3.707	0.9243	9.81		
P14376	rccC	106.51	2.167	-0.1098	5.95		
P27129	rfaJ	39.04	1.83	-0.1523	8.93		
P25742	rfaQ	38.73	2.433	-0.1641	6.82		
Q8XAT4	rhIB	47.092	1.974	-0.3436	7.28		hetero-oligomer
P25888	rhIE	49.989	2.786	-0.5061	10.05		Monomer
P63179	rlmB	26.556	1.995	-0.0332	6.17		homo-oligomer
Q8X9F0	rluD	37.121	3.031	-0.4177	6.31	E	
B1LEQ7	rnfC	73.299	2.143	-0.3645	8.7		hetero-oligomer
P0A9J2	rng	55.364	2.315	-0.2784	5.6		
Q1R8L7	rodZ	36.126	3.03	-0.3942	5.52		
P0AG09	rpe	24.554	4.103	0.0662	5.12	E	
Q1R5U9	rplA	24.729	13.21	-0.1084	9.64		ribosome
Q1R607	rplB	29.86	17.05	-0.6992	10.93	E	ribosome
Q1R602	rplC	22.243	21.19	-0.2348	9.9	E	ribosome
Q1R604	rplD	22.086	17.49	-0.2352	9.72	E	ribosome
Q1R620	rplE	20.301	9.429	-0.282	9.49		ribosome
Q1R624	rplF	18.904	19.22	-0.227	9.7	E	ribosome
Q1R357	rplI	15.769	28.06	0.0906	6.16		ribosome
Q1R5V0	rplJ	17.711	17.78	0.0455	9.03	E	ribosome
Q1R5U7	rplK	14.875	27.39	-0.0612	9.64		ribosome
Q1R5V1	rplL	12.295	14.41	0.295	4.6	E	ribosome
Q1R6A9	rplM	16.018	16.27	-0.54	9.9	E	ribosome
Q1R617	rplN	13.541	26.54	-0.1275	10.43	E	ribosome
P02413	rplO	14.98	49.89	-0.2499	11.18	E	ribosome
Q1R613	rplP	15.281	17.57	-0.3837	11.22		ribosome
Q1R638	rplQ	14.364	50.37	-0.5645	11.05		ribosome

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
Q1R626	rplR	12.769	30.13	-0.3948	10.41	E	ribosome
Q1R8B9	rplS	13.133	16.63	-0.5269	10.61	E	ribosome
Q1RB79	rplT	13.497	30.97	-0.3363	11.47	E	ribosome
Q1R6F0	rplU	11.564	19	-0.3688	9.84		ribosome
Q1R610	rplV	12.226	13.01	-0.349	10.23		ribosome
Q1R606	rplW	11.199	12.97	-0.3729	9.93	E	ribosome
P60625	rplX	11.316	21.48	-0.3807	10.21	E	ribosome
P68919	rplY	10.693	11.22	-0.468	9.59	E	ribosome
Q1R6F2	rpmA	9.1244	34.66	-0.6375	10.58		ribosome
Q1R4V5	rpmB	9.0064	22.87	-0.6499	11.42		ribosome
Q1R615	rpmC	7.2734	10.7	-0.657	9.98		ribosome
Q1R629	rpmD	6.5417	25.49	-0.1372	10.95		ribosome
Q0TAC8	rpmE	7.871	6.692	-0.6485	9.46		ribosome
Q1RD68	rpmF	6.4463	22.38	-0.9736	11.03		ribosome
Q1R4V6	rpmG	6.3715	16.4	-0.8035	10.24		ribosome
Q1R4N3	rpmH	5.3803	49.62	-1.1108	13		ribosome
P0AG69	rpsA	61.157	10.68	-0.3003	4.89		ribosome
Q1RG21	rpsB	26.743	15.56	-0.2675	6.61	E	ribosome
Q1R612	rpsC	25.983	14.56	-0.4218	10.27		ribosome
Q1R636	rpsD	23.469	45.22	-0.664	10.05		ribosome
Q1R627	rpsE	17.603	7.595	-0.1005	10.1	E	ribosome
Q1R360	rpsF	15.187	17.51	-0.7449	5.25	E	ribosome
P02359	rpsG	20.019	16.88	-0.4597	10.36		ribosome
Q1R622	rpsH	14.126	25.64	-0.0999	9.44	E	ribosome
Q1R6B0	rpsI	14.856	24	-0.6868	10.94	E	ribosome
Q1R601	rpsJ	11.735	11.65	-0.3523	9.68	E	ribosome
Q1R635	rpsK	13.845	17.57	-0.4456	11.32		ribosome
C4ZUJ7	rpsL	13.765	21.37	-0.6273	11	E	ribosome
Q1R633	rpsM	13.099	17.02	-0.4236	10.78	E	ribosome
Q1R621	rpsN	11.58	22.86	-0.7949	11.16	E	ribosome
Q1R6H3	rpsO	10.269	23.61	-0.6729	10.4	E	ribosome
Q0TEN0	rpsP	9.1904	10.95	-0.3292	10.54	E	ribosome
Q1R616	rpsQ	9.7043	36	-0.2928	9.64	E	ribosome
Q1R358	rpsR	8.9863	4.981	-0.7746	10.59		ribosome
Q1R609	rpsS	10.43	10.9	-0.6195	10.52	E	ribosome
Q1RGH9	rpsT	9.6843	18.99	-0.7045	11.18	E	ribosome

Appendix

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
Q1R6R5	rpsU	8.4999	18.15	-1.09	11.15		ribosome
P0AAT7	rsfS	11.582	2.079	0.0343	4.51		
Q8X510	rsmC	37.601	2.222	-0.1119	5.92		Monomer
P13458	sbcC	118.72	6.391	-0.6345	5.46		hetero-oligomer
P07014	sdhB	26.77	2.474	-0.2991	6.31		hetero-oligomer
P0AGA4	secY	48.511	1.98	0.6025	9.89	E	hetero-oligomer
P14081	selB	68.867	3.93	-0.3538	6.1		
Q1RE21	serS	48.413	1.993	-0.4473	5.33	E	homo-oligomer
Q1R8A4	smpB	18.269	10.55	-0.5812	9.89		
P21169	speC	79.415	2.001	-0.1615	5.66		
Q1RBT8	sra	5.0958	4.055	-1.2688	11.04		hetero-oligomer
P21507	srmB	49.914	2.986	-0.5573	9.27		
P0AGE2	ssb	18.975	2.938	-0.7078	5.44		homo-oligomer
P76423	thiM	27.339	2.199	0.2225	5.57		
P0A851	tig	48.192	4.09	-0.4274	4.82		homo-oligomer
Q8X8H6	tmk	23.816	2.783	-0.2065	5.35	E	homo-oligomer
Q8XB34	tnaA	52.789	3.019	-0.2218	5.87		homo-oligomer
P0ABV0	tolQ	25.597	3.411	0.2261	6.52		
P23003	trmA	41.966	1.838	-0.3218	5.7		
Q8XA83	trmJ	26.919	2.204	-0.1187	5.82		homo-oligomer
P0A877	trpA	28.724	3.603	0.1228	5.31		hetero-oligomer
Q1R3P6	ubiA	32.545	2.411	0.6866	9.07		
Q8FFP0	ubiG	26.538	4.63	-0.1991	6.17	E	homo-oligomer
P10908	ugpQ	27.409	3.081	-0.2125	6.09		
P0A8G2	uvrC	68.187	1.888	-0.3438	9.11		Monomer
Q1R2J3	uxuA	44.823	1.949	-0.4027	5.39		
P39160	uxuB	53.58	10.75	-0.1408	5.3		
P0AC77	waaA	47.29	3.656	0.0221	9.79	E	
P27242	waaU	41.729	4.072	-0.1758	8.76		
P37749	wbbI	37.757	2.548	-0.1969	5.84		
P37750	wbbJ	21.675	2.906	0.1026	9.54		
P37751	wbbK	43.188	2.751	-0.158	9.01		
Q8X4B4	wrbA	20.834	7.906	-0.0777	5.9		
Q1RFB8	xseB	8.9518	3.049	-0.7349	4.42		hetero-oligomer
P09030	xthA	30.969	3.566	-0.5551	5.79		Monomer
P0AFP5	ybbO	29.41	3.109	-0.043	8.97		

Appendix

Majority protein IDs	Gene names	Mol. weight [kDa]	Ratio CK1/C	GRAVY	pI	EG	oligomeric state
P0A9K5	ybeZ	39.038	4.355	-0.3259	5.71		
P30177	ybiB	35.048	3.704	-0.1652	6.37		
Q1RE59	ybjQ	11.437	18.03	0.0336	4.9	E	
P75829	ybjX	38.357	2.319	-0.3451	9.43		
P75874	yccU	14.701	2.527	0.0964	6.72	E	
P28306	yceG	38.247	6.566	-0.2361	9.42		
P08245	yciH	11.396	4.501	-0.4184	9.37		
P77748	ydiJ	113.25	1.978	-0.1599	6.68		
Q8X5W3	ydiU	54.532	5.385	-0.4263	5.6		
P0ACY2	ydjA	20.059	2.093	-0.1677	6.31	E	homo-oligomer
P76220	ydjY	24.125	21.91	-0.2421	6.21		
P76256	yeaZ	25.181	2.002	-0.0705	5.01		homo-oligomer
Q8XCN8	yebO	10.806	3.558	-0.041	4.97		
P31063	yedD	14.983	2.402	-0.1948	4.86		
P46125	yedI	32.19	2.438	0.8036	7.85		
P76418	yegU	35.62	7.518	0.0308	5.14		
P65558	yfcD	20.376	20.78	-0.4271	4.69		
P67097	yfcE	20.122	3.001	0.0239	5.63		
P64541	yfcL	9.9999	1.982	-0.1477	4.25		
P76537	yfeY	20.897	2.145	-0.3512	5.21		
P0AD42	yfhB	24.439	2.035	-0.0402	9.14		
P76578	yfhM	181.58	4.563	-0.321	5.25		
P0ADE7	ygaU	16.063	2.112	-0.4583	5.7	E	
P38521	yggL	12.88	8.318	-0.6777	4.9	E	
P30871	ygiF	48.388	5.456	-0.2602	5.73		
Q8FDK9	ygiQ	83.52	3.254	-0.5224	9.32		
P63419	yhbS	18.534	3.113	-0.0748	4.56	E	
P25536	yhdE	21.515	1.833	-0.1872	5.55		
P0ADK2	yiaF	25.663	3.96	-0.1512	6.07		
P0A8Y6	yidA	29.721	6.029	-0.0018	5.1		homo-oligomer
P27837	yifK	50.398	7.416	0.8432	9.49		
P0A8Y3	yihX	22.732	2.056	-0.1582	5.17		
Q8FAF3	yjgA	21.373	4.551	-0.9054	5.3		
P0AF94	yjgF	13.611	3.31	0.0859	5.36		
Q8FGT6	yobH	8.514	2.317	0.6013	9.5		
P64431	ypfJ	31.46	6.184	-0.5117	5.57		

Q46868	yqiC	11.276	2.576	-0.8176	5.89		
P37617	zntA	76.839	3.174	0.2817	5.69		

7.3.2 List of enriched protein ratios obtained from crosslinked and non-crosslinked datasets CAM versus PURO from CK1 strain.

Ribosome-nascent chains (RNCs) interactors sorted according to alphabetical order of gene name; Untreated / PURO ratio, enrichment of the RNCs interactors in FLAG-tgagged L17 CK1 strain Untreated compared to Puromycin treated CK1 strain as measured by SILAC MS; CAM/PURO enrichment in noncrosslinked CK1 strain; CAM/PURO enrichment in crosslinked CK1 strain. GRAVY means the average hydrophobicity of protein sequence (Kyte & Doolittle 1982); EG is abbreviation of essential gene (Gerdes et al. 2003).

N-CR: Non Crosslinked

CR: Crosslinked

O: Overlapping

EG: Essential genes (E)

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (N-CR/CR)	GRAVY	pl	EG	oligomeric state
accA	0.68	1.85	1.25			-0.2482	5.76	1	hetero-oligomer
aceE	1.16	1.9	2.26		O	-0.4368	5.45	0	homo-oligomer
aceF	1.27	1.76	2.26			0.0044	5.09	1	hetero-oligomer
acnB	2.5	4.55	4.72	O	O	-0.1174	5.23	0	Monomer
acpP	0.73	1.9	2.07		O	-0.3461	3.97	0	
acul	1.92	1.48	1.77			-0.0039	5.63	0	homo-oligomer
adhE	2.3	2.72	1.75	O		-0.066	6.32	0	homo-oligomer
adk	0.67	1.9	1.7			-0.3578	5.75	1	Monomer
ahpC	2.49	2.09	2.41	O	O	-0.2774	5.02	0	homo-oligomer
ahpF	0.82	1.14	6.48			-0.1225	5.46	0	homo-oligomer
alaS	2.04	3.01	3.48	O	O	-0.303	5.53	1	homo-oligomer
allR	1.19	1.41	2.27			0.0934	5.77	0	homo-oligomer
ampH	1.47	1.84	1.42			-0.195	9.49	0	
apt	0.51	NaN	4.44			0.0077	5.25	1	homo-oligomer
argS	0.73	1.64	5.84			-0.2555	5.31	1	Monomer
aroC	1.16	1.42	1.99			-0.2669	5.81	0	homo-oligomer
aroG	2.76	3.8	7.02	O	O	-0.1822	6.13	0	homo-oligomer

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
aroK	1.72	2.54	2.72		O	-0.6213	5.26	1	Monomer
asd	1.27	NaN	6.84			-0.0367	5.36	1	homo-oligomer
asnA	1.83	3.12	2.24	O	O	-0.2893	5.45	0	homo-oligomer
asnC	1.27	2.13	NaN			0.0618	6.28	0	
aspA	NaN	NaN	10.05			-0.066	5.19	0	homo-oligomer
aspC	1.94	3.31	3.18	O	O	-0.2014	5.54	0	homo-oligomer
aspS	0.62	2.99	2.99		O	-0.2645	5.47	1	homo-oligomer
atpA	1.84	2.58	2.17	O	O	-0.0447	5.79	0	hetero-oligomer
atpC	1.94	1.96	1.44	O		-0.0949	5.45	1	hetero-oligomer
atpD	1.41	2.69	2.25		O	-0.1177	4.9	0	hetero-oligomer
atpE	1.17	2.29	1.23			1.2671	4.43	0	hetero-oligomer
atpF	1.27	1.96	1.71			-0.2134	5.99	1	hetero-oligomer
atpG	1.28	2.36	1.81		O	-0.2271	8.83	0	hetero-oligomer
atpH	1.85	2.06	1.22	O		0.0627	4.94	0	hetero-oligomer
azoR	3.29	1.74	1.87			0.0667	5.06	1	homo-oligomer
bamA	0.57	0.62	2.86			-0.4829	4.92	1	
barA	1.34	2.1	NaN			-0.1093	5.4	0	
bdcR	NaN	NaN	2.69			-0.1669	6.59	1	
bioB	1.87	2.28	3.49	O	O	-0.3823	5.32	0	homo-oligomer
can	5.24	5.4	5.75	O	O	-0.2313	6.16	1	homo-oligomer
carA	0.79	2.2	1.93		O	-0.1989	5.91	0	hetero-oligomer
carB	1.78	2.45	2.43		O	-0.1619	5.22	0	hetero-oligomer
clpB	0.47	1.45	1.8			-0.3574	5.36	0	homo-oligomer
clpX	1.97	3.52	2.65	O	O	-0.2383	5.23	0	hetero-oligomer
cmk	1.11	1.84	2.23		O	-0.0003	5.55	1	
coaBC	2.77	1.99	1.68	O		0.03	7.05	1	homo-oligomer
corA	3.13	1.34	NaN			-0.1556	4.63	1	
corC	1.4	4.62	NaN			-0.3718	4.55	0	
cpxR	0.69	1.94	1.45			-0.3443	5.39	0	
cra	0.94	1.37	2.25			-0.3463	6.47	0	homo-oligomer
crr	3.44	5.05	2.12	O	O	0.0568	4.73	0	hetero-oligomer
cspA	2.42	4.22	3.26	O	O	-0.2513	5.57	0	
cspB	3.33	3.97	3.14	O	O	-0.09	6.53	0	
cspC	1.84	2.62	3.09	O	O	-0.1898	6.53	0	

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
cspD	1.63	3.76	4.21		O	-0.1972	5.8	0	homo-oligomer
cspE	1.75	2.38	3.34		O	-0.2129	8.08	0	
csrD	1.9	2.1	NaN	O		-0.1919	8.79	0	
cyaA	1.47	2.17	1.07			-0.2886	5.81	0	
cyoB	1.98	1.17	0.53			0.6582	6.69	0	hetero-oligomer
cysB	1.92	2.2	2.1	O	O	-0.0768	6.87	0	homo-oligomer
cysG	0.65	1.51	1.9			-0.0382	5.76	0	
cysK	4.06	2.27	2.03	O	O	-0.0782	5.82	0	homo-oligomer
cysS	1.08	2.67	NaN			-0.4055	5.32	1	Monomer
dacB	1.09	2.63	0.73			-0.0504	8.94	0	
dapA	0.58	1.59	6.76			0.0668	5.98	1	homo-oligomer
dapD	3.72	5.87	3.11	O	O	-0.0539	5.55	1	homo-oligomer
dapF	2.03	3.34	3.41	O	O	-0.0813	5.85	1	
deaD	2.78	2.5	2.49	O	O	-0.5862	8.74	0	hetero-oligomer
dedD	2.07	1.41	1.27			-0.3058	7.8	0	
def	1.32	3.5	5.57		O	-0.2898	5.22	1	Monomer
degP	2.53	2.36	1.75	O		-0.0672	8.64	0	homo-oligomer
degQ	2.35	3.23	1.51	O		0.0552	5.76	1	
deoB	NaN	NaN	4.28			-0.2699	5.15	0	
deoD	2.74	3.39	8.19	O	O	0.069	5.41	0	homo-oligomer
dgkA	0.91	1.97	NaN			0.8811	6.05	0	
dgt	1.98	2.43	1.75	O		-0.451	7.04	0	homo-oligomer
dhaM	0.49	3.93	3.52		O	-0.0173	4.61	0	homo-oligomer
dkgB	2.05	3.77	2.77	O	O	-0.1138	5.49	0	Monomer
dnaA	0.84	1.32	1.87			-0.3637	8.77	1	hetero-oligomer
dnaB	0.78	1.88	1.79			-0.4201	4.94	1	homo-oligomer
dnaC	1.97	2.23	3.62	O	O	-0.4807	9.09	1	
dnaJ	2.32	2.06	1.39	O		-0.6278	7.97	0	homo-oligomer
dnaK	1.9	2.68	2.73	O	O	-0.4085	4.83	1	
dnaX	0.91	1.53	1.95			-0.2576	6.38	1	hetero-oligomer
dps	0.61	1.47	3.26			-0.2328	5.52	0	homo-oligomer
dusB	4	3.1	NaN	O		-0.2457	6.26	0	
ecnB	2.3	2.4	NaN	O		0.4188	7.93	0	
eno	2.62	3.73	3.95	O	O	-0.1575	5.32	0	homo-oligomer

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
entB	1.72	2.84	2.15		O	-0.2876	5.04	0	hetero-oligomer
envC	1.8	1.82	0.92	O		-0.688	9.9	0	
erpA	1.42	1.71	2.23			-0.0885	4.1	0	homo-oligomer
etk	1.82	1.84	1.6	O		-0.3122	6.22	0	
evgS	0.76	2.02	0.72			-0.1993	5.79	0	
exbB	2.13	1.3	1.04			0.2648	7.88	0	hetero-oligomer
exbD	2.19	1.45	NaN			0.0128	4.69	0	hetero-oligomer
exoX	3.99	1.48	3.94			-0.219	7.09	0	
fabB	2.28	3.75	3.16	O	O	0.0079	5.34	1	homo-oligomer
fabD	1.65	2.28	1.97		O	0.1544	4.94	1	
fabF	0.64	3.09	4.15		O	0.0017	5.71	0	homo-oligomer
fabG	0.51	1.68	2.28			0.0996	6.76	1	homo-oligomer
fabH	0.4	2.48	0.46			0.1429	5.07	1	homo-oligomer
fabI	0.61	2.43	1.9		O	0.163	5.57	1	homo-oligomer
fabR	4.55	4.07	2.34	O	O	-0.2832	9.33	0	
fabZ	1.75	2.18	2.24		O	0.0536	6.83	1	homo-oligomer
fadR	NaN	NaN	4.5			-0.2827	6.5	0	homo-oligomer
fbaA	2.65	4.18	2.87	O	O	-0.2239	5.52	1	homo-oligomer
fdhE	3.37	2.38	NaN	O		-0.2886	5.03	0	
ffh	1.07	1.51	1.91			-0.2434	9.51	1	hetero-oligomer
fis	1.98	1.91	1.36	O		-0.5305	9.34	1	homo-oligomer
flkI	0.85	8.21	6.31		O	-0.0805	4.84	0	homo-oligomer
fkpA	2.42	3.09	2.84	O	O	-0.4673	8.38	0	
fmt	1.98	2.32	1.98	O	O	-0.0586	5.55	1	Monomer
folE	0.66	0.98	8.66			-0.1211	6.79	1	homo-oligomer
folM	3.15	3.68	1.36	O		-0.0741	8.4	0	
ftnA	0.34	4.94	5.68		O	-0.5684	4.76	0	homo-oligomer
ftsA	1.97	1.36	1.39			-0.0244	5.84	1	Monomer
ftsE	1.98	2.33	1.61	O		-0.0774	9.36	1	homo-oligomer
ftsH	2.44	1.02	1.62			-0.2629	5.9	0	
ftsI	1.13	1.93	1.58			-0.1502	9.64	1	
ftsN	1.83	1.96	1.04	O		-1.0234	10.16	0	
ftsX	1.22	2.08	1.12			0.1918	9.07	0	
ftsY	1.52	2.44	2.9		O	-0.3345	4.47	1	

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NC R/C R)	GRAVY	pI	EG	oligomeric state
ftsZ	1.26	1.9	1.98		O	-0.0046	4.64	1	homo-oligomer
fur	1.09	1.11	1.97			-0.5594	5.68	0	homo-oligomer
fusA	2.8	1.8	1.21	O		-0.2854	5.23	1	
galF	0.93	1.88	2.04		O	-0.0922	5.73	1	
gapA	0.83	2.49	2.65		O	-0.1331	6.6	1	homo-oligomer
garR	0.47	1.2	2.33			0.2619	5.58	0	
gatY	NaN	NaN	3.67			-0.0936	5.93	0	hetero-oligomer
gcp	2	2.59	1.39	O		-0.0005	5.75	0	homo-oligomer
gcvA	1.76	2.08	1.31			-0.1432	9.06	0	
gcvP	0.7	0.59	5.89			-0.0594	5.62	0	hetero-oligomer
gcvT	NaN	NaN	2.6			-0.1801	5.45	0	hetero-oligomer
ghrA	2.45	2.57	1.3	O		-0.2579	6.31	1	
ghrB;tkrA	2.24	3.41	3.78	O	O	0.0052	5.5	0	
glf	1.09	2.25	2.12		O	-0.5449	6.61	0	homo-oligomer
glgB	2.11	1.49	1.39			-0.5135	5.91	0	Monomer
glmM	2.85	3.59	2.71	O	O	0.0133	5.71	0	
glmS	6.13	2.67	2.18	O	O	-0.1012	5.56	1	homo-oligomer
glnA	0.79	4.65	1.91		O	-0.292	5.26	0	homo-oligomer
glnB	2.34	1.61	2.34			0.0089	5.17	0	homo-oligomer
glnD	0.55	1.5	1.83			-0.2944	6.22	1	Monomer
glnH	NaN	NaN	4.26			-0.2334	8.43	0	
glnS	2.13	3.61	3.83	O	O	-0.5178	5.88	1	Monomer
glrK	1.44	1.72	2.39			-0.1403	7.67	0	
gltB	2.87	2.51	2.05	O	O	-0.2143	6.14	0	hetero-oligomer
gltD	2.76	2.32	1.64	O		-0.3132	5.53	0	hetero-oligomer
gltX	2.3	4.68	3.17	O	O	-0.5275	5.59	1	Monomer
glyA	2.16	2.48	3.29	O	O	-0.2272	6.02	1	homo-oligomer
glyQ	0.65	1.72	2.56			-0.3742	4.94	1	hetero-oligomer
glyS	0.55	1.9	2.85		O	-0.2191	5.29	1	hetero-oligomer
gmk	1.44	1.89	2.16		O	-0.4299	6.06	1	homo-oligomer
gmr	2.8	3.17	4.05	O	O	-0.1899	6.28	0	
gnd	1.96	2.86	3.29	O	O	-0.1997	5.05	0	homo-oligomer
gnsA	NaN	NaN	6.91			-0.5455	5.39	0	

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
gntR	0.89	2.13	4.16		O	-0.1038	6.4	0	
gpmA	0.63	0.82	4.02			-0.5679	5.85	0	homo-oligomer
gpmB	2.79	2.55	3.39	O	O	-0.3706	5.45	0	
gpml	0.8	2.3	4.6		O	-0.1937	5.17	0	Monomer
gpsA	1.53	2.08	2.27		O	-0.0247	6.08	0	
grcA	5.86	5	5.47	O	O	-0.3322	5.08	1	
greA	0.78	2.3	2.06		O	-0.3664	4.71	0	
groL;groL1	0.95	1.84	2.93		O	-0.0087	4.84	0	
grpE	1.12	4.26	2.05		O	-0.372	4.68	1	homo-oligomer
grxD	1.14	1.79	2.02			-0.2721	4.74	1	homo-oligomer
gshA	0.74	0.62	4.93			-0.3103	5.26	0	
gsk	0.47	1.57	2.84			-0.3458	5.5	0	
gsp	0.95	NaN	8.78			-0.3417	5.13	0	
gstA	2.04	1.09	2.21			-0.2203	5.84	0	homo-oligomer
guaA	0.56	3.36	4.67		O	-0.0963	5.23	0	homo-oligomer
gyrA	2.05	2.82	2.82	O	O	-0.2398	5.08	1	hetero-oligomer
gyrB	1.94	2.06	2.57	O	O	-0.4539	5.72	0	hetero-oligomer
hemB	NaN	NaN	3.27			-0.1249	5.24	1	homo-oligomer
hemD	3.76	2.48	0.65	O		-0.1235	5.98	1	Monomer
hemE	2.03	1.55	1.11			-0.1491	5.88	1	homo-oligomer
hflC	2.59	1.11	1.86			-0.3646	6.29	0	hetero-oligomer
hflK	2.74	1.12	1.89			-0.6316	6.18	0	hetero-oligomer
hfq	1.33	1.8	1.99		O	-0.5048	6.96	1	homo-oligomer
hinT	0.43	NaN	5.68			-0.0621	5.73	0	homo-oligomer
hisS	0.46	1.92	1.54			-0.2716	5.64	1	homo-oligomer
hldD	0.75	1.29	2.03			-0.2834	4.8	0	homo-oligomer
hns	2.38	2.01	2.97	O	O	-0.751	5.43	1	homo-oligomer
hoIA	1.82	2.44	1.82	O	O	-0.0984	6.51	1	hetero-oligomer
hpt	0.43	NaN	4.26			-0.0684	5.09	0	homo-oligomer
hrpA	2.07	2.17	2.03	O	O	-0.4409	7.88	0	
hscA	1.98	1.79	3.76			-0.0278	4.98	1	
hsdM	3.31	3.76	3.81	O	O	-0.5275	5.05	0	hetero-oligomer
hsdS	2.21	3.1	2.63	O	O	-0.2605	9.59	0	hetero-oligomer

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
hslO	1.98	4.02	4.55	O	O	-0.3129	4.35	0	
hslU	3.01	3.83	3.81	O	O	-0.3568	5.23	0	hetero-oligomer
htpG	2.49	2.32	1.84	O	O	-0.5235	5.08	0	homo-oligomer
hupA	3.23	2.33	3.19	O	O	-0.2277	9.57	0	hetero-oligomer
hupB	2.83	2.41	3.22	O	O	-0.0421	9.69	0	hetero-oligomer
ibpA	1.97	2.3	3.58	O	O	-0.5247	5.57	0	Monomer
icd	2.79	4.19	3.62	O	O	-0.1535	5.15	0	homo-oligomer
iciA	2.82	4.31	5.18	O	O	-0.1726	6.4	0	homo-oligomer
ileS	2.17	3.35	2.33	O	O	-0.2561	5.65	1	Monomer
ilvC	1.29	NaN	8.82			-0.2048	5.2	0	
ilvH	2.05	1.9	2.09	O	O	-0.0545	8	0	hetero-oligomer
infA	2.02	1.69	2.92			-0.3541	9.22	0	
insL3	3.99	3	2.74	O	O	-0.2221	8.76	0	
intA	1.03	2.73	2.37		O	-0.4307	9.6	0	
iscR	2.23	1.86	2.17	O	O	-0.2561	6.82	0	
ispG	0.59	2.64	2.39		O	-0.1572	5.87	1	
ivy	2.01	0.91	0.97			-0.1024	6.26	0	
katG;katG1	0.56	1.33	2.91			-0.3809	5.13	0	
kefA	3.26	2.83	5.59	O	O	-0.0311	8.04	0	
kgtP	1.97	1.87	1.33	O		0.6171	9.47	0	
lasT	2.1	2.1	3.1	O	O	0.0053	5.57	0	
ldcA	2.51	1.73	NaN			0.022	5.98	0	
ldhA	2.9	2.5	3.07	O	O	-0.1005	5.28	0	
lepA	0.94	3.36	3.57		O	-0.1951	5.46	0	
leuS	1.96	2.22	4.81	O	O	-0.3675	5.15	1	Monomer
lon	1.68	2.87	1.61			-0.3123	6.01	0	homo-oligomer
lpdA	1.34	1.91	2.38		O	-0.0109	5.78	1	homo-oligomer
lptB	0.85	1.67	2.51			-0.1244	5.63	1	hetero-oligomer
lysS	1.16	2	1.46			-0.3963	5.1	0	homo-oligomer
lysU	1.1	1.86	1.66			-0.4011	5.14	0	homo-oligomer
maeA	0.89	2.57	0.99			-0.178	5.24	0	homo-oligomer
malT	1.56	1.22	3.26			-0.2727	6.02	0	Monomer
manX	3.52	3.68	2.91	O	O	-0.0581	5.74	0	

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
manZ	1.2	1.83	1.47			0.4066	9.17	0	
map	1.74	2.43	5.18		O	-0.162	5.64	1	Monomer
mdh	2.71	2.32	2.48	O	O	0.1936	5.6	0	homo-oligomer
mdoG	1.1	2.06	1.38			-0.5361	6.69	0	
metC	1.81	1.72	NaN			0.0238	6	0	homo-oligomer
metG	1.62	2.13	2.34		O	-0.3017	5.56	1	homo-oligomer
metJ	0.74	NaN	3.61			-0.7475	5.39	0	homo-oligomer
metN	1.9	1.74	1.33			-0.0719	5.94	0	
mfd	1.24	0.2	5.2			-0.2916	5.78	0	Monomer
mgtA	1.84	1.12	0.81			0.1592	5.63	0	
miaB	0.59	1.57	1.8			-0.3782	5.19	0	Monomer
minC	0.7	NaN	5.37			-0.051	6.38	0	homo-oligomer
minD	1.97	3.14	3.13	O	O	-0.118	5.24	1	Monomer
minE	2.15	3.08	1.6	O		-0.2942	5.14	0	
mIaB	2.01	1.68	NaN			0.032	4.74	0	
mIaC	1.92	2.69	1.38	O		-0.4155	9.35	0	
mIaD	2.03	1.9	1.57	O		-0.0677	4.78	0	
mIaF	2.3	1.93	1.63	O		0.1335	6.15	0	
mngR	2.37	2.24	2.66	O	O	-0.5628	6.25	0	
mnmE	0.79	2.17	1.01			-0.0933	4.89	1	homo-oligomer
mntR	2.17	2.54	2.83	O	O	-0.3767	6.31	1	homo-oligomer
moaA	1.83	1.78	1.95			-0.3513	7.67	0	Monomer
moaB	1.7	2.58	NaN			-0.2252	5.72	0	homo-oligomer
mprA	1.95	1.87	2.27	O	O	-0.6374	5.81	0	
mrcB	1.01	1.85	0.98			-0.3178	9.39	0	hetero-oligomer
mrdB	2.11	1.41	NaN			0.9081	9.34	1	
mreB	2.64	3.46	2.98	O	O	0.0573	5.19	1	
msbA	1.17	1.95	0.93			0.11	8.62	1	homo-oligomer
mScS	1.9	2.39	2.67	O	O	0.4944	7.89	0	homo-oligomer
mukB	0.99	1.8	1.87		O	-0.5989	5.23	1	homo-oligomer
mukE	0.8	3.06	NaN			-0.4932	4.97	1	hetero-oligomer
mukF	0.99	1.97	2.43		O	-0.3692	4.75	1	Monomer
murA	2.25	2.13	1.13	O		0.0511	5.8	0	
murC	1.25	1.92	2.84		O	-0.1606	5.52	1	

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
murl	2.35	2.68	3.21	O	O	0.1758	5.16	1	monomer
mutL	1.23	1.96	1.22			-0.2566	6.3	0	
mutS	0.59	1.98	1.32			-0.2403	5.35	0	
mutY	1.65	2.07	1.86		O	-0.1462	8.6	0	
nadB	1.82	2.84	2.69	O	O	-0.2319	5.88	0	Monomer
nagE	2.31	0.8	NaN			0.5843	5.78	0	
narP	1.57	2.43	2.58		O	-0.0139	4.81	0	
nei	2.5	1.59	1.77			-0.3375	8.31	0	
nfi	1.21	3.05	1.99		O	-0.0636	8.42	0	Monomer
nfnB	0.12	2.78	6.56		O	-0.2105	5.8	0	homo-oligomer
nfsA	2.09	2.05	1.35	O		-0.1545	6.44	0	
nfuA	1.86	1.89	2.41	O	O	-0.2559	4.52	0	homo-oligomer
nifU	1.89	2.09	2.02	O	O	-0.3554	4.81	0	hetero-oligomer
nikE	NaN	NaN	11.16			-0.1253	10.1	0	
nlpD	2.3	1.38	NaN			-0.4031	9.51	0	
nnr	1.01	1.19	1.82			-0.0463	5.98	0	
nrdA	0.91	1.89	1.81		O	-0.3372	5.79	1	hetero-oligomer
nsrR	0.83	1.86	2.07		O	0.0546	7.7	0	
nudC	0.72	1.5	2.04			-0.3373	5.51	0	homo-oligomer
nuoB	1.41	2.15	2.56		O	-0.3204	5.58	0	hetero-oligomer
nusA	1.05	2.16	3.17		O	-0.2779	4.52	0	Monomer
nusB	0.79	1.77	1.91			-0.2193	6.6	1	Monomer
nusG	0.92	1.39	1.92			-0.3911	6.33	0	Monomer
ompC	2.71	2.98	2.52	O	O	-0.5331	4.58	0	
ompR	2.26	1.96	2.83	O	O	-0.3455	6.03	0	Monomer
oppA	2.3	2.47	3.83	O	O	-0.4364	6.05	0	
orn	1.04	NaN	9.42			-0.3546	5.04	0	homo-oligomer
paaX	1.25	NaN	1.95			0.0044	7.09	1	
panC	2.3	2	1.69	O		-0.108	5.91	0	homo-oligomer
panE	1.16	2.74	2.48		O	-0.1045	5.51	0	Monomer
parC	5.04	2.9	2.45	O	O	-0.3688	6.24	1	hetero-oligomer
parE	0.8	7.09	5.34		O	-0.3451	5.43	1	hetero-oligomer
pbpG	2.19	2.48	1.2	O		-0.056	9.94	0	
pckA	0.92	3.66	5.25		O	-0.3145	5.45	0	Monomer

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
pcm	2.07	2.32	2.99	O	O	-0.1576	6.52	0	Monomer
pcnB	1.15	2.23	2.33		O	-0.5304	9.66	0	Monomer
pdxB	1.72	1.98	2.41		O	-0.0364	6.22	0	homo-oligomer
pdxY	2.75	2.33	NaN	O		0.0962	6.04	0	homo-oligomer
pepA	3.04	1.68	1.4			-0.2172	6.81	0	homo-oligomer
pepB	2.83	2.9	3.01	O	O	-0.214	5.6	0	homo-oligomer
pepD	0.71	1.5	3.68			-0.1455	5.2	0	
pepQ	NaN	NaN	6.46			-0.2349	5.53	0	
pfkA	1.98	2.07	3.4	O	O	0	5.47	0	homo-oligomer
pflB	2.69	4.83	2.82	O	O	-0.3799	5.69	0	homo-oligomer
pgi	1.59	2.54	4.98		O	-0.2691	5.92	0	homo-oligomer
pgk	0.63	2.45	2.34		O	0.0718	5.07	1	Monomer
pgm	0.78	4.16	5.53		O	-0.1677	5.42	0	
pheT	3.02	4.13	4.38	O	O	-0.0971	5.16	0	hetero-oligomer
phnA	2.22	2.67	4.91	O	O	-0.5323	4.97	0	
phoP	0.72	4.38	5.29		O	-0.2282	5	0	
pncB	2.27	3.98	4.71	O	O	-0.2982	6.2	0	
pnp	2.12	2.6	2.93	O	O	-0.1448	5.08	1	homo-oligomer
pntA	1.92	2.12	0.43	O		0.2727	5.64	0	hetero-oligomer
pntB	2.34	1.85	0.69	O		0.6024	5.72	0	hetero-oligomer
polA	0.4	1.69	2.65			-0.264	5.39	0	Monomer
polB	2.18	1.08	1.34			-0.4505	6.39	0	
potD	NaN	NaN	3.88			-0.4453	5.24	1	
ppa	2.83	3.67	4.82	O	O	-0.2692	5.02	1	homo-oligomer
ppc	0.95	1.62	1.95			-0.2767	5.51	0	homo-oligomer
ppiD	1.51	2.21	3.08		O	-0.4054	4.93	0	homo-oligomer
ppk	1.01	2	1.09			-0.3394	8.96	0	homo-oligomer
prfA	1.82	1.61	2.49			-0.6599	5.14	0	
prfB	2.43	2.86	2.79	O	O	-0.6829	4.64	0	
prfC	0.93	1.38	2.2			-0.3262	5.65	0	
priA	1.91	1.77	1.25			-0.1864	9.05	1	hetero-oligomer
prlC	1.79	5.09	1.76			-0.4162	5.15	0	
proQ	1.62	1.91	2.13		O	-0.7762	9.66	0	
proS	0.85	1.29	2.21			-0.2427	5.12	1	homo-oligomer

Appendix

Gene names	Untreated/PURO	CAM /PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NC R/C R)	GRAVY	pI	EG	oligomeric state
proV	1.84	1.72	1.8			-0.1342	5.38	0	
prs	1.13	2.83	1.88		O	0.0962	5.23	1	
psd	1.91	1.93	1.57	O		-0.0934	5.5	0	hetero-oligomer
pspA	2.13	2.05	4.34	O	O	-0.722	5.38	0	hetero-oligomer
pspB	1.2	1.97	NaN			-0.2229	6.72	0	Monomer
pstA	1.13	1.23	2.27			0.7084	9.94	0	
pstB	1.87	1.83	1.26	O		-0.2699	6.12	0	hetero-oligomer
ptsH	NaN	NaN	7.09			-0.1658	5.64	0	
ptsl	1.88	3.99	4.13	O	O	-0.1202	4.77	0	homo-oligomer
ptsP	0.92	2.22	2.64		O	-0.0942	5.57	0	
purB	0.34	1.87	0.94			-0.2431	5.67	0	homo-oligomer
purC	0.71	2.09	2.61		O	-0.3695	5.07	0	homo-oligomer
purD	1.09	NaN	5.64			-0.071	4.95	0	Monomer
purE	NaN	NaN	8.47			0.0864	6.03	0	homo-oligomer
purF	0.42	0.94	1.91			-0.2425	5.32	0	homo-oligomer
purH	2.33	3.32	3.14	O	O	-0.1396	5.52	0	
purK	2.2	3.21	NaN	O		-0.1393	5.59	0	homo-oligomer
purL	0.79	2.73	0.35			-0.2349	5.22	0	Monomer
purM	3.41	3.79	5.58	O	O	0.0299	4.78	0	homo-oligomer
purN	4.66	2.8	2.51	O	O	-0.1235	5.51	0	Monomer
purR	2.64	2.85	1.4	O		-0.2374	6.27	0	homo-oligomer
puuA	0.56	227.09	301.65		O	-0.3552	4.82	0	
pykA	0.82	1.55	2.08			-0.0245	6.23	0	homo-oligomer
pykF	2.25	3.1	4.02	O	O	-0.0818	5.76	0	homo-oligomer
pyrB	0.93	4.62	5.01		O	-0.1044	6.11	0	hetero-oligomer
pyrC	1.04	4.59	NaN			-0.122	5.96	0	homo-oligomer
pyrD	1.92	2.19	1.31	O		-0.11	7.66	0	Monomer
pyrG	1.83	1.95	2.79	O	O	-0.1388	5.63	1	homo-oligomer
pyrH	1.78	NaN	2.15			0.1328	6.85	1	homo-oligomer
qorA	0.82	1.79	2.04			-0.0589	8.35	0	homo-oligomer
radA	2.65	2.09	1.28	O		0.0385	6.8	0	
rapA	0.92	2.48	2.99		O	-0.3553	5.02	0	hetero-oligomer
rbsB	NaN	NaN	2.69			-0.0306	6.84	0	
rbsB	1.53	2.1	2.88		O	0.0213	6.84	0	

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
rscC	1.86	2.26	3.24	O	O	-0.1098	5.95	0	
recA	1.99	2.78	2.48	O	O	-0.1753	5.09	0	
recF	2.09	1.51	1.52			-0.4231	6.77	0	
recG	1.81	2.28	2.07	O	O	-0.1265	6.77	0	
recJ	1.06	1.51	2			-0.0692	5.38	0	
recQ	1.97	1.7	2.47			-0.2689	6.9	0	
rep	2.87	2.56	2.01	O	O	-0.4081	6.76	1	homo-oligomer
rfaB	1.23	1.46	1.83			-0.1584	6.31	0	
rfaJ	1.9	1.29	0.87			-0.1523	8.93	0	
rfaS	2.09	1.71	1.52			-0.3385	8.9	0	
rfaY	3.09	2.23	2.31	O	O	-0.5167	9.68	0	
rfaZ	1.83	1.8	1.94	O	O	-0.2483	9.24	0	
rfbD	0.46	1.47	2.22			-0.0929	5.54	0	homo-oligomer
rhlB	2.49	2.53	2.32	O	O	-0.3436	7.28	0	hetero-oligomer
rho	2.26	2.43	1.63	O		-0.2978	6.75	0	homo-oligomer
ribB	NaN	NaN	5.36			-0.1474	4.88	1	homo-oligomer
ribC	0.75	NaN	5.92			-0.0999	5.64	1	homo-oligomer
rimJ	1.53	3.14	1.32			-0.5236	9.2	0	
rimN	NaN	NaN	6.6			-0.0573	4.94	0	
rimP	0.97	1.5	1.87			0.0467	4.59	1	hetero-oligomer
rlmF	1.44	2.22	2.68		O	-0.3824	9.29	0	
rlmG	0.88	1.72	2.53			-0.1785	6.31	0	
rlmM	1.53	2.14	1.26			-0.2841	7.03	0	Monomer
rlmN	1.28	1.78	1.88			-0.2851	6.56	0	
rlpA	1.2	1.87	NaN			-0.2447	5.51	0	
rluB	0.79	2.27	1.24			-0.5744	10.02	1	
rluC	1.58	2.82	2.44		O	-0.4607	9.84	0	
rluD	1.09	1.5	1.84			-0.4155	6.31	1	
rluE	1.37	3.46	2.53		O	-0.6801	10.02	0	
rluF	1.84	2.18	1.6	O		-0.5216	9.68	0	Monomer
rmlA1	0.38	1.04	2.76			-0.1927	5.38	0	homo-oligomer
rnc	1.64	2.31	1.75			-0.4486	6.4	1	homo-oligomer
rne	2.32	2.39	2.1	O	O	-0.6445	5.47	1	hetero-oligomer
rnfC	1.07	1.89	1.25			-0.4158	8.88	0	hetero-oligomer

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
rng	1.81	2.26	1.57	O		-0.2784	5.6	0	
rnhB	3.03	3.71	3.52	O	O	-0.0055	7.14	1	
rob	3	2.81	2.9	O	O	-0.3549	6.66	0	
rph	0.99	1.3	1.98			-0.0835	5.53	0	homo-oligomer
rpoA	0.82	1.83	2.17		O	-0.2276	4.98	0	homo-oligomer
rpoB	0.98	1.76	1.99			-0.3944	5.14	1	hetero-oligomer
rpoC	0.98	1.84	2.17		O	-0.2396	6.66	1	hetero-oligomer
rpoD	2.54	3.49	4.73	O	O	-0.5888	4.68	1	
rpoZ	0.55	2.69	2.73		O	-0.5241	4.87	0	hetero-oligomer
rppH	2.42	2.84	2.93	O	O	-0.6743	10.04	0	Monomer
rraB	1.29	1.83	1.86		O	-0.605	3.63	0	
rsfS	0.95	1.87	1.12			0.0343	4.51	0	
rsmA	1.08	2.01	2.48		O	-0.0816	8.52	0	
rsmB	1.15	1.65	1.82			-0.3938	7.16	0	
rsmG	0.75	1.55	2.49			-0.1535	6.06	0	
rsuA	1.85	2.18	2.94	O	O	-0.3627	5.74	0	Monomer
ruvB	1.97	3.28	3.72	O	O	-0.0743	5.02	0	homo-oligomer
secB	1.49	1.82	2.85		O	-0.1876	4.26	1	homo-oligomer
secE	0.71	1.44	1.87			0.8764	10.55	0	hetero-oligomer
secG	2.29	1.7	1.99			0.3755	6.09	0	hetero-oligomer
secY	0.9	1.56	2.31			0.6025	9.89	1	hetero-oligomer
selA	2.51	3.32	5.03	O	O	-0.047	6.04	0	homo-oligomer
selD	1.38	2.9	1.92		O	0.0683	5.29	1	Monomer
serA	5.66	1.46	3.54			0.0251	5.92	0	homo-oligomer
serC	1.07	1.58	2.69			-0.1123	5.36	0	homo-oligomer
serS	3.56	6	5.69	O	O	-0.4473	5.33	1	homo-oligomer
sfsA	1.19	2.2	2.14		O	-0.3375	6.31	0	
slmA	1.63	2.28	3.15		O	-0.3095	8.78	0	homo-oligomer
slyD	1.49	2.28	2.48		O	-0.3826	4.85	1	Monomer
smpB	1.02	1.33	1.89			-0.5812	9.89	0	
smtA	0.73	1.35	1.88			-0.4022	7.72	0	
sodA	3.08	2.27	4.98	O	O	-0.4198	6.44	0	homo-oligomer
sodB	3.06	4.22	3.88	O	O	-0.1688	5.58	0	homo-oligomer
soxR	NaN	NaN	1.93			-0.4109	9.17	0	homo-oligomer

Appendix

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
soxS	1.81	1.46	NaN			-0.5186	9.88	0	
speG	2.89	3.15	1.39	O		-0.4391	6.2	0	homo-oligomer
spoT	2.82	2.46	2.14	O	O	-0.3321	8.88	0	Monomer
sspA	0.7	1.25	3.01			-0.265	5.22	0	
sucA	1.65	2.31	4.05		O	-0.441	6.03	1	homo-oligomer
sucB	0.66	2.19	1.9		O	-0.2167	5.57	1	hetero-oligomer
talB	2.17	2.45	3.98	O	O	-0.2097	5.1	0	homo-oligomer
tdh	1.3	2.28	2		O	0.0636	5.81	0	homo-oligomer
tdk	1.96	1.35	1.78			-0.3204	5.97	1	homo-oligomer
tgt	1.16	2.08	1.62			-0.4119	5.97	0	homo-oligomer
thiC	0.92	2.47	2.08		O	-0.3869	5.69	0	homo-oligomer
thiE	1.72	2.8	2.57		O	0.0322	5.53	0	
thiF	1.94	2.95	2.37	O	O	-0.0389	4.69	0	homo-oligomer
thiG	7.63	3.93	NaN	O		0.3004	5.35	0	homo-oligomer
thiI	1.15	1.86	1.38			-0.2219	6.16	0	
thiM	1.22	1.61	2.62			0.2225	5.57	0	
thrC	2.4	2.2	2.23	O	O	-0.0966	5.23	0	
thrS	1.87	2.24	1.18	O		-0.493	5.8	1	homo-oligomer
tig	3.79	6.15	4.31	O	O	-0.4274	4.82	0	homo-oligomer
tktA	2.97	3.66	4.32	O	O	-0.2275	5.42	1	homo-oligomer
tolA	3.06	2.98	NaN	O		-0.6094	9.09	0	Monomer
topA	3.39	2.69	3.06	O	O	-0.6085	8.67	1	Monomer
topB	1.66	1.8	1.85		O	-0.3702	8.44	0	
tpiA	1.34	0.97	7.89			0.0102	5.63	0	homo-oligomer
tpx	2.93	3.49	4.04	O	O	0.256	4.74	0	
trmA	1.58	3.65	3.65		O	-0.3218	5.7	0	
trmL	1.81	2.21	2.67	O	O	-0.2426	6.19	0	homo-oligomer
trpS	0.6	3.16	5.11		O	-0.3197	6.26	1	homo-oligomer
truA	1.16	1.73	1.9			-0.314	8.68	0	homo-oligomer
truD	1.31	1.7	2.52			-0.269	6.15	0	Monomer
trxA	2.3	3.16	3	O	O	0.0248	4.67	0	Monomer
tsaA	0.79	1.66	2.04			-0.1484	7.85	0	homo-oligomer
tsf	2.76	3.08	3.18	O	O	-0.1179	5.21	1	hetero-oligomer
ttcA	1.44	1.85	1.72			-0.4552	5.95	0	

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NCR/CR)	GRAVY	pI	EG	oligomeric state
typA	1.65	3.16	3.27		O	-0.2978	5.15	0	
tyrR	8.44	4.57	1.59	O		-0.2083	5.54	0	homo-oligomer
ubiD	2.71	4.61	2.14	O	O	-0.2501	5.37	0	homo-oligomer
ubiG	1.77	1.87	1.18			-0.2274	6.11	1	homo-oligomer
ubiX	1.04	2.22	2.8		O	0.2439	6.28	1	
udk	0.98	1.94	1.91		O	-0.2581	6.39	0	homo-oligomer
ugpC	2.61	2.1	NaN	O		-0.2041	7.82	0	
ung	NaN	NaN	2.15			-0.3051	6.66	0	Monomer
upp	0.76	1.58	1.89			0.0731	5.31	0	homo-oligomer
uppS;ispU	0.69	1.82	1.65			-0.3956	6.45	0	
uspF	1.05	3.2	NaN			0.0222	5.6	0	homo-oligomer
wrbA	NaN	NaN	6.57			-0.0817	5.59	0	
uup	3.07	4.21	3.99	O	O	-0.4667	5.35	0	
uvrB	2.07	2.29	2.27	O	O	-0.3576	5.13	0	hetero-oligomer
uvrC	1.36	2	1.62			-0.3438	9.11	0	Monomer
uvrD	1.99	2.07	1.24	O		-0.4578	5.85	0	
valS	0.77	2.63	3.46		O	-0.3802	5.19	1	Monomer
waaA	1.53	2.03	1.37			0.0221	9.79	1	
waaU	2.58	1.77	2.39			-0.1758	8.76	0	
wcaK	1.11	2.81	NaN			-0.018	6.59	0	
wecC	0.74	3.65	2.38		O	0.0655	5.22	0	
wzzB	1.33	2.07	0.68			-0.2646	5.42	0	homo-oligomer
wzzE	1.05	1.93	NaN			-0.3539	6.24	0	
xthA	1.48	2.15	1.88		O	-0.5551	5.79	0	Monomer
yadG	1.86	1.51	1.95			-0.2372	8.43	0	
yajC	3.17	3.53	3.18	O	O	0.3955	9.57	0	hetero-oligomer
yajL	1.33	2.82	1.37			0.3235	5.24	0	homo-oligomer
yajO	2.43	2.13	1.62	O		-0.4042	5.18	0	
ybaB	0.95	1.69	1.82			-0.4733	5.01	0	homo-oligomer
ybbN	NaN	NaN	5.09			-0.2418	4.49	0	
ybeZ	2.05	2.08	2.26	O	O	-0.3259	5.71	0	
ybiB	1.03	3.49	3.13		O	-0.1652	6.37	0	
ybjD	1.3	2.84	2.46		O	-0.3258	9.35	0	

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NC R/C R)	GRAVY	pI	EG	oligomeric state
ybjX	2.06	1.76	1.76			-0.3451	9.43	0	
ycaO	1.88	1.63	12.23			-0.1956	4.37	0	
ycbB	1.41	2.83	1.5			-0.2757	8.63	0	
yceM	0.68	1.45	2.27			-0.0976	6.83	0	
ycfD	0.58	NaN	4.79			-0.4358	4.69	0	
ycfH	0.92	2.2	2.23		O	-0.2116	5.19	0	
ychF	0.57	1.18	4.49			-0.1627	4.87	1	
yciK	1.9	2.49	1.5	O		-0.3209	7.66	1	
yciT	0.86	2.16	1.2			-0.0915	5.98	0	
ydbK	1.17	2.34	1.06			-0.1914	5.52	0	
ydcl	1.9	1.98	3.16	O	O	-0.0224	7.82	0	
ydcP	1.9	2.33	1.22	O		-0.3112	6.68	0	
ydfG	1.37	2.23	1.9		O	-0.1745	5.65	0	homo-oligomer
ydgA	1.08	1.95	2.65		O	-0.2985	5.06	1	homo-oligomer
ydhF	0.95	1.47	2.34			-0.1204	5.76	0	
ydjG	1.81	1	1.39			-0.0925	5.12	0	
yebC	6.63	7.83	5.28	O	O	-0.3828	4.71	1	
yebO	1.62	0.82	4.11			-0.041	4.97	0	
yecE	1.99	1.34	1.5			-0.3506	7.23	0	
yedE	1.83	1.66	NaN			0.5863	10.04	1	
yeeN	1.57	5.05	5.42		O	-0.2587	4.71	0	
yeeZ	0.83	2.65	2.76		O	-0.0999	5.27	0	
yejL	7.97	5.58	7.59	O	O	-0.1799	5.49	0	
yfeA	1.26	2.04	NaN			0.3056	8.58	0	
yfgM	1.58	2.11	7.04		O	-0.2217	5.06	0	
yfiF	1.04	1.74	1.98			-0.6816	8.93	0	
ygiQ	1.88	2.53	1.54	O		-0.5092	9.18	0	
ygjR	2.51	2.45	NaN	O		0.0613	5.68	0	
yhbJ	2.22	2.21	2.57	O	O	-0.3157	6.72	0	homo-oligomer
yhbY	2.06	1.66	NaN			-0.1824	9.41	0	
yhcM	0.78	2.47	2.5		O	-0.2884	6.36	0	
yheS	2.62	3.79	2.73	O	O	-0.463	5.5	0	
yhgF	1.16	2.2	2.69		O	-0.3342	5.91	0	
yhil	1.59	1.39	2.06			-0.1605	6.89	0	

Gene names	Untreated/PURO	CAM/PURO (N-CR)	CAM vs PURO (CR)	Overlap (Untreated/CAM)	Overlap (NC R/C R)	GRAVY	pI	EG	oligomeric state
yhjK	2.3	1.91	1.06	O		0.1828	6.06	0	
yiaF	2.03	1.62	NaN			-0.1512	6.07	0	
yibL	2.75	2.19	1.62	O		-0.6282	9.48	0	
yicC	1.36	NaN	4.49			-0.4421	5.09	0	
yifE	5.84	4.81	6.87	O	O	-1.0687	6.1	0	
yihI	3.11	2.89	2.07	O	O	-1.3768	6.36	0	homo-oligomer
yjhU	1.46	2.22	2.63		O	-0.0249	6.32	0	
yjjK	3.21	4.36	3.91	O	O	-0.4777	5.43	0	Monomer
yliF	2.06	1.74	1.1			0.1457	8.25	0	
ynjA	0.65	3.9	NaN			0.0368	8.79	1	
yoaE	2.83	1.71	NaN			0.4519	4.88	0	
ypeA	1.91	2.93	1.79	O		-0.4049	4.84	0	
yqiC	2.05	2.31	NaN	O		-0.8176	5.89	0	
yqjG	2.14	1.62	0.54			-0.3682	5.98	0	homo-oligomer
yrbL	1.81	1.83	1.79	O		-0.557	8.98	0	
ytfP	1.99	1.95	2.75	O	O	-0.5141	6.39	0	
zapA	NaN	NaN	6.12			-0.7329	5.21	0	homo-oligomer
zapB	2.07	1.63	2.21			-1.0202	4.68	1	homo-oligomer
zapD	1.42	1.92	2.87		O	-0.142	6.3	0	
zinT	3.48	2.58	2.91	O	O	-0.5828	5.91	0	
zipA	1.34	1.91	1.45			-0.6029	5.63	1	hetero-oligomer
znuA	3.45	4.25	5.32	O	O	-0.2105	5.6	0	
znuC	2.93	2.38	1.71	O		-0.1055	9.43	0	hetero-oligomer
zwf	3.19	4.8	4.58	O	O	-0.3734	5.55	1	

7.4 Curriculum Vitae

Sneha Kumar

PhD student, Department of Cellular Biochemistry
Max Planck Institute of Biochemistry
Am Klopferspitz 18
82152 Martinsried / Munich, Germany
Office: +49 089/8578-2232
snkumar@biochem.mpg.de

Education

Degree: **M.Sc.** in Biomedical Genetics 2008 to 2010
Institution: Vellore Institute of Technology Tamil Nadu,
India Grade: Distinction (8.9 CGPA)

Degree: **B. Sc.** triple majors in Chemistry, Microbiology & Zoology 2005 to 2008
Institution: Mount Carmel College Bangalore,
India Grade: Distinction (82 Percentile)

Research Experience

PhD candidate Sept 2010 – Present
Institution: Max Planck Institute of Biochemistry (MPIB), Martinsried, Germany.
Project: Identification of the compendium of factors interacting with translating ribosomes in *E.coli* - A novel proteomic approach
Supervisor: Prof. F.U. Hartl

Masters Dissertation Jan. to July 2010
Institution: Department of Chemistry, University of Wisconsin-Madison, USA.
Project: Studying co-translational protein folding using NMR technology.
Supervisor: Prof. Silvia Cavagnero

Summer Research Fellow May to July 2009
Institution: Saha Institute of Nuclear Physics, Kolkata, India
Project: Expression and Purification Huntington binding protein (Htt) and subsequent Biophysical analysis.
Supervisor: Dr. N.P.Bhattacharya