IMPROVING HETEROLOGOUS PROTEIN EXPRESSION IN *E. COLI* USING MOLECULAR CHAPERONES FROM *THERMUS* SPP.

Amélie Dukunde

June 2011

University of the Free State

Improving heterologous protein expression in *E. coli* using molecular chaperones from *Thermus* spp.

Amélie Dukunde

B.Sc. Hons (UFS)

Submitted in fulfilment of the requirements for the degree

MAGISTER SCIENTIAE

In the Faculty of Natural and Agricultural Sciences

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

Bloemfontein

South Africa

June 2011

Supervisor: Prof. J. Albertyn

Co-supervisor: Prof. D. Litthauer

Prof. E. van Heerden

Acknowledgments

"Many hands make light work." – English Proverb

- **Prof. J. Albertyn**: For providing me with a project that my restless mind never tired of, as well as giving me the independence and tools to explore it.
- **Prof. D. Litthauer** and **Prof. E. Van Heerden**: For taking an interest in this project and securing funds for it.
- **Dr. D. Opperman**: What can I say? Chapter 3 would not have been completed without his help.
- **Members of Lab 49**: For answering the 'stupid' questions that would otherwise have had me burning the lab down or releasing some Godzilla-inducing things down the drain.
- **Biomolecular Research Cluster Fund** and **TIA**: For funding this project.
- **Friends**: For the solidarity when we all ganged up against Evil Science if things didn't work out and for all the laughs when they actually did.
- **My family**: For everything.

Thank you all,

Amélie Dukunde

For my family, who cheered on even when they weren't quite sure why...

“**In the beginning there was nothing, which exploded.**”

– From Lords and Ladies, Terry Pratchett.

List of non-SI Abbreviations used in this study.

Abbreviation	Definition
AAA	ATPase associated with a variety of cellular activity
ADP	Adenosine diphosphate
Amp	Ampicillin
Arg	Arginine
ATCC	American type culture collection
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
BSE	Bovine spongiform encephalopathy
Cm	Chloramphenicol
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide phosphates
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
gDNA	Genomic DNA
GFP	Green fluorescent protein
HCl	Hydrochloric acid
Hsp	Heat shock protein
IB	Inclusion body
IPTG	Isopropyl β-D^{-1}-thiogalactopyranoside
Kan	Kanamycin
KJEA	DnaK, DnaJ, GrpE and DafA operon
LB	Luria-Bertani
Lys	Lysine
MCS	Multiple cloning site
MD	Middle domain
mRNA	Messenger ribonucleic acid
NBD	Nucleotide binding domain
NEF	Nucleotide exchange factor
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
P. falciparum	*Plasmodium falciparum*

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PGK	Phosphoglycerate kinase
Phe	Phenylalanine
PTC	Peptidyl transferase centre
S. cerevisiae	*Saccharomyces cerevisiae*
SBD	Substrate binding domain
SDS	Sodium dodecyl sulphate
sGFP	Superfolder GFP
T. aquaticus	*Thermus aquaticus*
T. halophilus	*Tetragenococcus halophilus*
T. scotoductus	*Thermus scotoductus*
T. thermophilus	*Thermus thermophilus*
Taq pol	*T. aquaticus* DNA polymerase
TF	Trigger factor
tRNA	Transfer ribonucleic acid
*Tth*PolI	*T. thermophilus* DNA polymerase I
X-gal	Bromo-chloro-indolyl-galactopyranoside
YFP	Yellow fluorescent protein

Table of contents

Chapter 1 . A review of protein folding and molecular chaperones in the cell.

1.1 Introduction

All living organisms have evolved highly organised mechanisms of self-replication and self-assembly that ensure the conservation of the complex genetic and physiological functions in their progeny. Macromolecules provide the building blocks required to assemble components that will, eventually, become a living entity or a structure that contributes to the organism's growth or metabolism; in this group of macromolecules, proteins are the most abundant and are the workhorses of the cell metabolome (Dobson, 2004). They are found in all the diverse facets of cellular function, from the start of the cell's life in the cell division cycle through to apoptosis, when the cell dies. In proteins, cells have also established channels of intra- and inter-cellular communication that allow the efficient transfer of information in signal transduction and as a consequence, multicellular organisms have evolved .

The ubiquitous nature of proteins raises the question: where do proteins come from? In response, Francis Crick formulated the central dogma that DNA makes RNA which, in turn makes protein (Crick, 1970). This holds for most proteins but when viewed in light of evolution, on what came first, it becomes a 'chicken or egg' dilemma which evolutionary biology is only beginning to understand.

The classic model of protein synthesis proposed by Crick also excludes a crucial step in the production of biologically active proteins: that of protein folding. The folded state of proteins holds the key to their function. It will determine how stable the protein is within the cell, dictate its activity as well as where it will carry out its functions; whether extracellular or intracellular, cytosolic or membrane-bound. It is through folding that proteins are able to achieve such a great diversity in structure, substrate range and enzymatic selectivity (Dobson, 2004). The importance of protein folding within the cell is further illustrated by the stringent quality control methods that living organisms have developed to ensure that errors in protein folding are minimised if not wholly avoided (Wickner *et al.*, 1999), the increasing number of disorders where protein misfolding is implicated such as Creutzfeldt-Jacob disease,

Parkinson's and Alzheimer's in humans (Dobson, 2001) and the difficulty of recovering soluble, heterologous proteins in industry, from bacteria (de Marco *et al.*, 2007).

Folding can be spontaneous or mediated by folding modulators known as molecular chaperones. Chaperones increase the chance of bringing compatible protein domains together, thus preventing unfavourable associations that would lead to protein misfolding (Hartl & Hayer-Hartl, 2002). This review will focus on bacterial molecular chaperones, particularly those of *Escherichia coli*, how they achieve the correct folding of proteins within the bacterial cell and how this function has been exploited in industry to improve soluble, recombinant protein recovery.

1.1.1 Hypothesis and aim of this project

A large body of literature exists regarding overexpression of molecular chaperones for the production of highly soluble and active recombinant proteins in *E. coli*. Recombinant proteins are of particular importance in the biotechnology industry. Expression of heterologous molecular chaperones has also been carried out, but this has not been for purposes of overexpression, with the exception of the *P. falciparum* Hsp70-based vector constructed by Stephens *et al.*, (2011). In addition, heterologous chaperones appear to confer a higher degree of tolerance to environments that would normally be stressful to the cell, such as thermotolerance and halotolerance compared to their mesophilic counterparts.

In this project, a single plasmid system for the simultaneous co-expression of molecular chaperones and heterologous proteins will be constructed. Two DnaK chaperone systems, *Tt*DnaK and *Ts*DnaK, from *Thermus thermophilus* and *Thermus scotoductus*, respectively will be co-expressed with the *T. thermophilus* DNA polymerase I protein, *Tth*PolI, in *E. coli*. This will be carried out in two parts:

1) Cloning of *Thermus* spp. DnaK operon genes and *TthPolI* gene from their native host and constructing a single expression vector for the co-expression of both chaperone and polymerase.
2) Evaluation of the co-expression vector by carrying out heterologous expression to determine the effect of thermophilic chaperones on the solubility, activity and thermostability of *Tth*PolI. As a negative control, the polymerase will also be expressed in wild-type *E. coli* host for accurate determination of the folding capacity of thermophilic chaperones in comparison to endogenous *E. coli* chaperones.

Such a study will reveal new sources of novel chaperones from thermophilic environments that could be used as alternative folding enhancers in heterologous protein expression.

1.2. A protein is born: Fundamental concepts of protein folding *in vitro* and *in vivo*.

Proteins are essential biopolymers of the living cell; required to provide structural support or catalytic activity to the cell. They are the most abundant macromolecules in the cell and they oversee all the metabolic processes carried out by the cell (Dobson, 2004). There are numerous metabolic pathways and each step is supervised by a specific protein or group of proteins, indicating that, despite their diversity, their structures and functions are not an arbitrary assignment but a careful selection of features that have undergone stringent evolutionary selection to reach their current state (Dobson, 2004).

The native three dimensional structure of a protein is an inherent property of the amino acid sequence and within it (Anfinsen, 1973), there are several possible native conformations (Samanta *et al.*, 2009); however, according to Dobson, (2003), computer simulations and *in vitro* protein folding studies have shown that, within a cell, only specific conformations are adopted for each protein. Samanta *et al.*, (2009), speculate that random sampling of all these conformations during folding events would take even a relatively small protein several million years. Not only does this demonstrate the diversity of protein conformations, it also raises the question of how rapid and efficient folding by one protein into a unique and conserved shape is achieved, consistently (Wang *et al.*, 2005). In order to reach this final state, the nascent polypeptide transitions from a decidedly disordered energy state to an energetically stable one as shown in the 'folding funnel' model (Dinner *et al.*, 2000) in Fig. 1.1. The model adopts the standard that a stable protein conformation is achieved and maintained at the lowest possible energy state (Dinner *et al.*, 2000; Dobson 2004). As folding progresses, amino acid residues interact and make contacts that are favourable and lead to a permanent association and lower the enthalpy, or total energy of the polypeptide. At this stage, these stable regions within the polypeptide will be core centres for which other residues can associate, giving rise to a native-like, partially folded intermediate. The intermediate is now a structure that exhibits low conformational energy and it is evident that there are fewer possible conformations which the protein can adopt (Dinner *et al.*, 2000). The number of native-like contacts within a molecule will increase proportionally with the size of

the protein and in turn, this will increase the number of partially folded intermediate (Dinner *et al.*, 2000). Yet, whatever the case may be, they serve to reduce the number of possible conformations that a protein would have to pass through to reach its native structure.

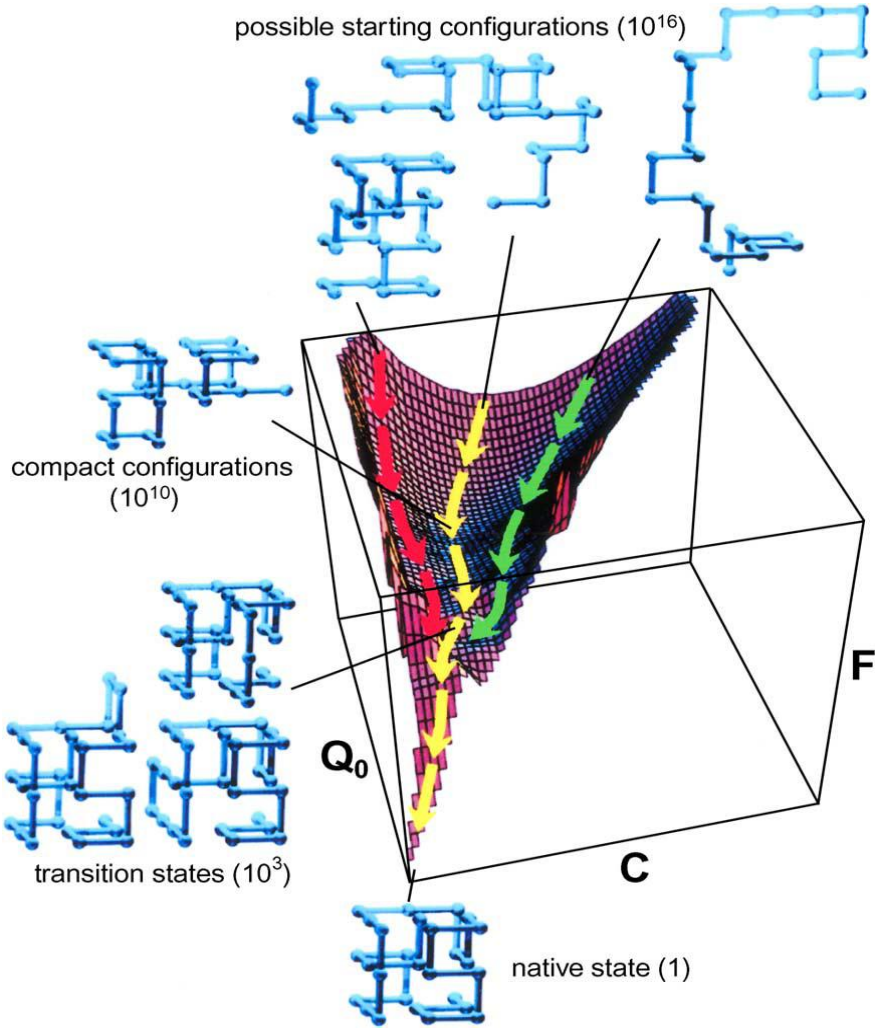

Figure 1.1. This 'funnel' model of the energy landscape employed in protein folding demonstrates how a polypeptide chain is transformed from a random coil to a highly ordered and structurally conserved three dimensional structure. As the amino residues within the molecule make contact (C), some associations are stable than others and cause the molecule to have a structure resembling the native conformation (Q_0), this in turn reduces the total energy of the system and promotes the folding of the protein into its native structure with the lowest possible energetic conformation. Source: Dinner *et al.*, 2000.

This model also provides an explanation of how evolutionary selection has produced proteins that can rapidly fold – according to Dobson, (2003), small helical bundles can fold as fast as 50 μs – into a specific three dimensional structure with relatively high precision. The global

distribution of protein network within the cell indicates that they are required to work quickly and accurately. Therefore, their synthesis must be achieved with equal speed and fidelity by eliminating unnecessary off-pathway intermediates that would not only waste the cell's resources but possibly, lead to misfolding and aggregation, which will be discussed in greater detail later.

1.2.1 Protein folding within the cell

Anfinsen and co-workers demonstrated how a polypeptide will spontaneously fold *in vitro* using bovine pancreatic ribonuclease A (RNAse A) denatured in urea and refolded with full catalytic activity (Anfinsen *et al.*, 1961). Since that time, this phenomenon has been documented by a number of laboratories employing techniques such as computational simulations, NMR spectroscopy and atomic force microscopy (Vendruscolo & Paci, 2003). The pioneering work of this group also initiated intense research activity focused on the physics, chemistry, thermodynamics and physiological properties of proteins and how folding affects them. Anfinsen's assumptions were that the polypeptide behaves as a randomly coiled polymer in a free solution (Dobson, 2003) and therefore, only the intrinsic properties, i.e. amino acid sequence, of the chain would allow it to fold into its native state (Anfinsen, 1973). But how do these relatively 'ideal' experimental conditions translate in the cellular environment (Baker, 2000)?

There is no clear determinant of what affects protein folding and tertiary structure conformation; a debate based on the nature versus nurture concept (Itzhaki & Wolynes, 2010) shows that in a complex system such as a living organism, it is not only impossible to have a central feature of protein function determined by only one factor, it is practical for the cell to rely on multiple channels that will act as rescue centres should one of them fail, so as not to compromise cellular metabolism. The following section examines some of the key factors that are known to influence *in vivo* protein folding.

1.2.1.1 Codon usage and folding on the ribosome

Within the cytosol, protein folding is initiated by the ribosome during translation of mRNA (Zhang & Ignatova, 2010). As mRNA is 'fed' through the peptidyl transferase centre (PTC) of the small subunit in the ribosome, at the 5' start codon, a polypeptide emerges from the large subunit tunnel (Samanta *et al.*, 2009), with the N-terminal being translated first and the C-terminal last. This process of co-translational folding of polypeptides, whereby a

polypeptide is folded as mRNA is translated (Hardesty *et al.*, 1999), has been documented in all eubacteria, archaea and eukaryotes (Zhang & Ignatova, 2010) and provides the first platform for protein folding. As the ribosome advances along the mRNA strand, it encounters codons that are rare (Kramer *et al.*, 2009; Zhang & Ignatova, 2010), the corresponding tRNA molecules are in low abundance within the cell; consequently, the fluctuating supply of tRNA to the ribosome causes it to stall at certain regions (Kramer *et al.*, 2009) and results in slowing down the entire process of translation, as shown in Fig. 1.2. In addition, it allows the N-terminal residues exiting the large subunit enough time to associate and form partially folded and stable native-like structures that would not form if the ribosome were to translate the whole strand at one time and present the whole molecule for folding as reported by (Evans *et al.*, 2008) on the co-translational folding of *Salmonella* P22 tailspike protein.

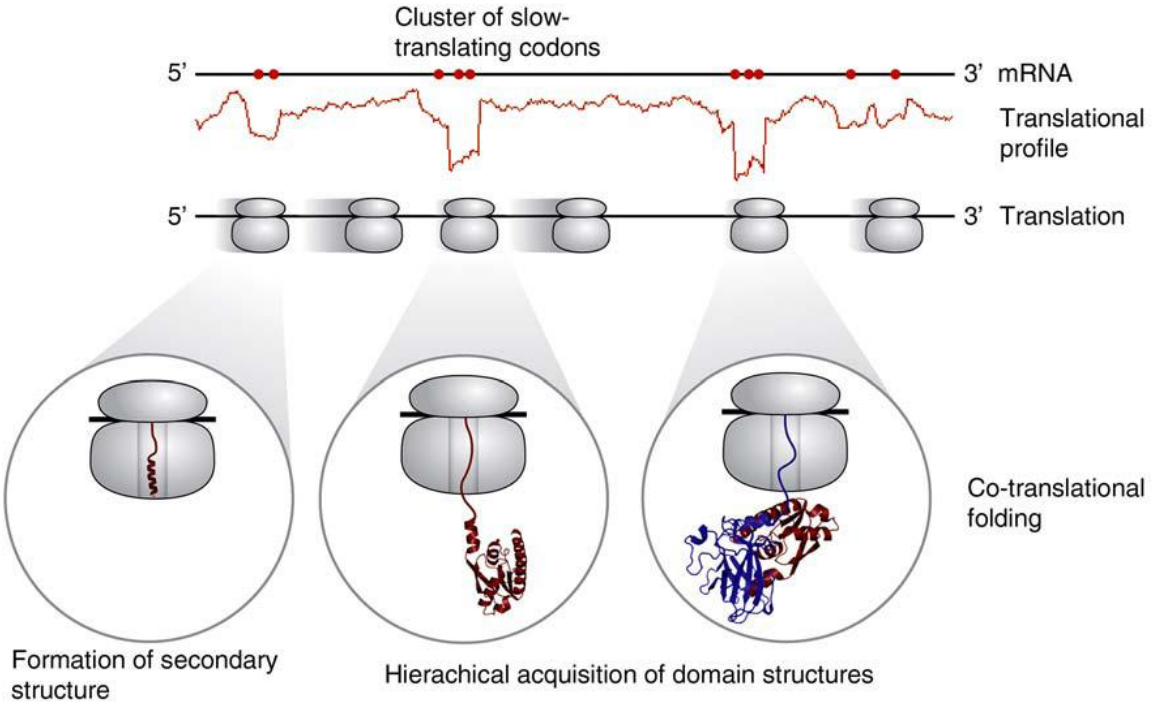

Figure 1.2. Schematic representation of discontinuous translation of mRNA with subsequent folding of the nascent polypeptide at the N-terminal exit. Note the clusters of codons (marked in red), downstream of the 5' mRNA, where translation is retarded due to their low abundance; this causes a pause in translation and allows the N-terminal of the polypeptide to fold to a native-like state. Source: Zhang & Ignatova, 2010.

It is interesting to note that these slow translating areas are often located in the interdomain regions along a polypeptide chain which indicates that folding takes place after an entire

domain is translated (Kramer *et al.*, 2009). Thus, the ribosome is offers a small degree of protection against the association of non-native residues in different domains of the protein.

1.2.1.2 Macromolecular crowding contributes to protein compaction

The cytosol is not an inert medium to transport molecules to and fro but a highly dynamic environment (Mittal & Best, 2010) whose composition can change fundamental folding pathways of a protein. Macromolecules, also known as crowders (Mittal & Best, 2010) such as lipids, carbohydrates, nucleic acids and proteins are found in concentrations of 300 to 400 mg ml^{-1} in the cytosol (Dobson, 2004), around 10 - 40% of the total fluid volume (Engel *et al.*, 2008). These densely packed molecules leave very little room for a protein to fold freely (Mittal & Best, 2010). By using the 'funnel model' of the energy landscape described before, it is evident that a crowded environment further reduces the conformational freedom of a polypeptide strand. Therefore, macromolecular crowding will also act as a funnel in reducing the number of possible conformations a protein can adopt.

One of the most important features of crowding, however, is the excluded volume effect (Zhou, 2008). This is based on the principle of spatial volume that two molecules or solutes cannot occupy the same space at the same time (Minton, 1992) and in a crowded biological fluid, the volume in which additional solutes may occupy is reduced by those already present (Minton, 2006). According to Engel and co-workers, (2008), the reduced volume encourages the disordered and highly unstable polypeptide chain to adopt a more compact state that leads to native folding.

Crowding makes the cytosol more viscous, a property which boosts protein folding by maintaining interdomain units in close proximity (Gershenson & Gierasch, 2011). It was demonstrated by Dhar *et al.*, (2010), when they reported correct folding and enhanced activity of phosphoglycerate kinase (PGK) using Ficoll 70 as the crowding agent; as well as Zimmerman & Harrison, (1987), on improving the activity of *E. coli* DNA polymerase I using PEG 8000, Dextran T-70 and bovine serum albumin (BSA).

All this serves to show that crowding promotes native folding but this is under conditions where crowders behave as inert bodies (Engel *et al.*, 2008), while *in vivo*, there might be instances of chemical interaction between proteins and other macromolecules, where the effects of crowding lead to misfolding and aggregation.

1.3 Protein misfolding and aggregation

Having detailed features of *de novo* synthesis and folding of proteins, it is now possible to explore protein misfolding, a major occurrence which can be fatal to the cell and as a whole, the organism.

1.3.1 Principles of misfolding, aggregation and the formation of inclusion bodies

The transition of a polypeptide from a random polymer to a structurally defined protein involves one or more partially folded intermediates; this is particularly true for large, multi-domain proteins as they cannot be folded in a single step as smaller proteins (< 100 residues)(Dobson, 2003; Hartl & Hayer-Hartl, 2009). It is during the folding of these intermediates that misfolding is encountered. Often, the ribosome offers little in the way of shielding non-native domains from each other, allowing for the formation of secondary structures like simple α-helices and β-sheets (Evans *et al.*, 2008; Kramer *et al.*, 2001). Translation proceeds at a rate of 15-20 residues/s, a 'slow' process as reported by (Hartl & Hayer-Hartl, 2009; Zhang & Ignatova, 2010); as a result, non-contiguous domains on the nascent chains are brought into close proximity with each other while the rest of the molecule is tethered to the ribosome, which causes them to associate for longer than necessary (Hartl & Hayer-Hartl, 2009). Most of these interactions give unstable proteins but there are those that might be energetically favourable and fold into an off pathway conformation. This process that results in a stable but structurally abnormal proteins is misfolding (Dobson, 2003; Kopito, 2000), a characteristic of both *in vitro* and *in vivo* protein synthesis (Marquardt & Helenius, 1992) for a range of prokaryotic and eukaryotic proteins that may be cytosolic or membrane bound (de Groot *et al.*, 2009).

Misfolded intermediates may have their hydrophobic regions exposed in the aqueous environment *in vitro* or *in vivo* (Vabulas *et al.*, 2010). In a bid to internalise these non polar groups, they cluster together and form insoluble, crystalline precipitates (González-Montalbán *et al.*, 2007) with an extremely low conformation energy that makes them highly stable. These precipitates are collectively known as inclusion bodies (IBs) (Rinas *et al.*, 2007). Kopito, (2000), also suggests that, while inclusion body formation might be initiated at single nucleation points, the resulting aggregate also combines with other aggregates; this would form a large disordered structure of heterogeneous proteins as observed in bacterial

inclusion bodies (de Groot *et al.*, 2009). In eukaryotes, especially in mammalian cells, aggregates have been observed to form rigid, amyloid fibrils, which are highly ordered chains with high β-sheet content (Dobson, 2003; Stefani, 2004). The conformation stability of inclusion bodies is so high that often, the free energy associated with such stability is much lower than that of the native, globular protein, as seen in Fig. 1.3; a fact which would explain their prevalence in cells and their relative insolubility (Hartl & Hayer-Hartl, 2009).

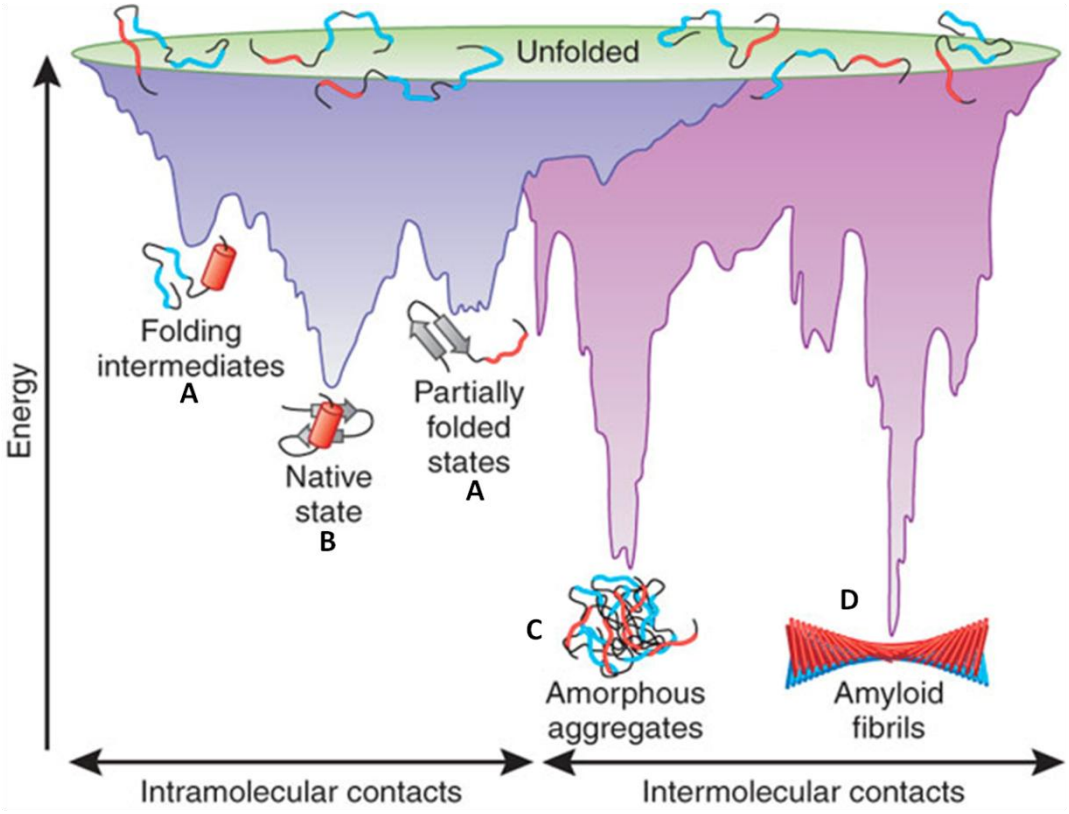

Figure 1.3. The path of least resistance. The energy landscape of various conformations during protein folding. Intramolecular contacts between residues of intermediates (A) promote native-like folding (B), shown by the blue region; while intermolecular associations, coloured in purple, most likely interdomain or subunit interactions, favour the formation of aggregates (C, D). In this case, IBs, especially the amyloid fibrils, have a lower conformational energy and therefore more stable than the native conformation. Source: Hartl & Hayer-Hartl, 2009.

Further analyses of inclusion body structure reveal that bacterial inclusion bodies have similar amyloid or amyloid-like precursors, suggesting a higher organisation in prokaryotic inclusion bodies (de Groot *et al.*, 2009; Díez-Gil *et al.*, 2010; Ventura & Villaverde, 2006) and an active role into physiological function rather than an inert one (Villaverde & Carrió, 2003).

1.3.2 Inclusion bodies have adverse or beneficial effects in organisms

The physiological role of inclusion bodies is not well understood and current literature tends to cite them as 'dead end' subjects, the waste products of poor folding machinery in the cell; the consensus opinion is that inclusion bodies are ubiquitous, inert bodies that play no chemical role in cellular metabolism.

While inclusion bodies may be chemically inert, they influence cellular activity significantly. They have been implicated in a wide range of genetic disorders in humans such as Alzheimer's, Parkinson's, Cystic Fibrosis (Dobson, 2001), Type II diabetes (González-Montalbán *et al.*, 2007), prion diseases such as bovine spongiform encephalopathy (BSE or Mad Cow's disease) and Creutzfeldt- Jacob disease (Stefani, 2004). In bacteria, a high concentration of inclusion bodies is toxic and often fatal (Saibil, 2008; Bösl *et al.*, 2006).

Regardless of inclusion body activity after they are formed, they are generally a response to cell stress, often brought on by extreme temperatures and heterologous protein expression (Díez-Gil *et al.*, 2010). Also, intrinsic faults encoded in the amino acid sequences due to DNA mutation, the fidelity of transcription and translational machinery (Kopito, 2000) may cause non-native association of protein domains and cause aggregation into inclusion bodies.

1.3.3 Mechanisms of preventing protein misfolding and aggregation

Cells have had to evolve adaptive mechanisms as a response to inclusion body formation: firstly, through proteolysis and secondly, via the use of accessory proteins known as molecular chaperones (Barnett *et al.*, 2000). In the former, misfolded proteins are targeted by ubiquitination to the cell's degrading machinery such as the 20S proteasome found in eukaryotes, homologs found in *E. coli*, such as Clp proteases (Liu *et al.*, 2002; Wickner *et al.*, 1999), as well as range of other intracellular proteases that target misfolded proteins by detecting exposed hydrophobic regions. The proteolytic pathway is itself a very large field of study with respect to protein misfolding and aggregation, however, it is not the focus of this review and is not treated in greater detail; instead, the focus now moves to molecular chaperones and the manner in which they bring about correct, native folding of proteins.

1.4 Molecular chaperones as tools for preventing misfolding and aggregation

Molecular chaperones are defined as proteins that enable folding of other proteins – nascent or misfolded – to reach their stable, native conformation but are not themselves part of the final product of the reaction (Vabulas *et al.*, 2010). They form transient and reversible associations with their substrate or client proteins, although they neither lower activation energy for the protein of interest nor confer steric information. As such, they are not enzymes or folding catalysts and are seen as 'facilitators' of folding, instead (Dobson, 2003; Hartl & Martin, 1995).

They constitute a highly diverse group of molecules with respect to size, structure and function. Chaperones are not limited to tertiary assembly of nascent polypeptides but also in the posttranslational assembly of proteins into their multimeric, quaternary structures (Makrides, 1996); in addition, they are involved in refolding of misfolded or aggregated proteins, disassembly of such aggregates for proteolysis, as well as directing the cell to apoptosis in case of severe damage (Saibil, 2008). A few molecular chaperones have also been linked to signal transduction pathways, as well as protein translocation to the periplasm (Baneyx & Mujacic, 2004; Bann *et al.*, 2004).

Molecular chaperones are ubiquitous in nature and are found in all eukaryotes, prokaryotes and archaea (Gething, 1996). They were initially observed during the heat shock response to thermal stress when they were designated 'heat shock' proteins (Hsps), but it is important to note that not all Hsps are chaperones and while most chaperones are induced to higher concentrations during heat shock, a number are expressed constitutively under normal, physiological temperatures (Hartl & Martin, 1995). During heat shock, a high number of proteins become misfolded or are translated incorrectly; molecular chaperones 'rescue' these proteins and prevent their aggregation or refold them so that the cell can maintain metabolic function under thermal stress, thereby improving the cell's thermotolerance (Thomas & Baneyx, 1998).

Chaperones are cytosolic proteins (Hartl & Hayer-Hartl, 2002) and despite their diversity, three loose classifications of molecular chaperones have been proposed: holdases, foldases and unfoldases or disaggregases (Hoffmann *et al.*, 2010). Holdases prevent aggregation by associating with the exposed hydrophobic regions commonly found in nascent polypeptides

or misfolded proteins. This classification only describes the chief function of each chaperone system; often, the properties of one chaperone may overlap with those of another and even encompass all three classes.

Chaperones either work as singular molecules or in sets to carry out their task, which increases their specificity as well as their substrate range (Kolaj *et al.*, 2009). During *de novo* synthesis, the nascent polypeptide might immediately associate with ribosome-associated chaperones such as trigger factor (TF) and DnaK, which act as holdases and prevent non-native association of domains, as well as exposed hydrophobic residues (Mayer & Bukau, 2005). The intermediate protein may then be picked up for further refolding by the GroEL/GroES foldases (Hartl & Hayer-Hartl, 2002) and if misfolding of the intermediate occurs, the ClpB disaggregase will disassemble it affording the protein a second chance at refolding (Doyle & Wickner, 2009). At this stage, any proteins that consistently misfold or aggregate despite chaperone-assisted folding may be targeted for proteolysis, as illustrated in Fig. 1.4 (Doyle & Wickner, 2009).

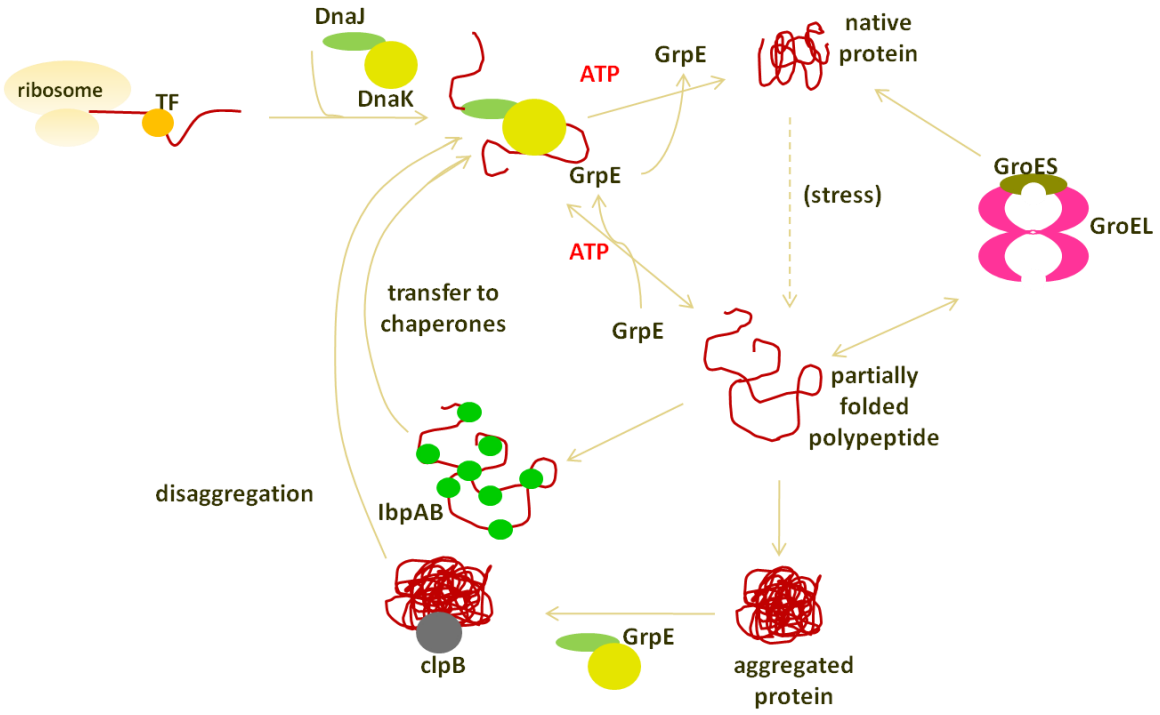

Figure 1.4. The network of molecular chaperones found in *E. coli*. While each chaperone can work individually to bring about native folding of proteins, they often work hand in hand with each other for optimal folding of client proteins. Adapted from Kolaj *et al.*, 2009.

The sequence of chaperone-mediated folding and unfolding can be summarised in three main steps: substrate recognition and binding, ATP-dependent folding/ unfolding and substrate release. However, the precise manner in which they actually bring about folding and unfolding and how molecular chaperones recognise client proteins is an undecided issue in research and is slowly unravelling (Boshoff *et al.*, 2004).

1.4.1 Molecular chaperones in *E. coli*

Molecular chaperones in *E. coli* are diverse but highly conserved proteins and are among the best characterised within the cell (Table 1.1) (Schlieker *et al.*, 2002). Increasing insight into their structure and mechanism simultaneously reveals possible functional mechanisms of other prokaryotic chaperones, as well as their eukaryotic homologs and vice versa (Haslberger *et al.*, 2010; Martin, 1997).

Table 1.1. Grouping of molecular chaperones found in living organisms and their *E. coli* homologs. Chaperones are highly conserved proteins and understanding the way they work in one organism often suggests how they function in other living systems. Source: Schlieker *et al.*, 2002.

chaperone family	structure	ATP	*E. coli* member	number of active species per cell	null mutant phenotype	function
HSP100	6-mer 	+	ClpA		No phenotype	protein degradation together with the ClpP Protease
			ClpB	500	Impaired thermotolerance	disaggregation of protein aggregates together with the DnaK system
HSP90	dimer 	+	HtpG	1050	Reduced growth rate at 44°C	unknown
HSP70	monomer 	+	DnaK co-chaperone DnaJ/GrpE	9900	Temperature-sensitive growth (39°C)	de novo protein folding prevention of protein aggregation at high temperatures regulation of the heat shock response disaggregation of protein aggregates together with ClpB
HSP60	14-mer 	+	GroEL co-chaperone GroES	1230	lethal	de novo protein folding prevention of protein aggregation at high temperatures
sHSP	8-24-mer 		IbpA IbpB	<600	No phenotype	prevention of protein aggregation at high temperatures
Trigger-factor	monomer 		Trigger-factor	20000	No phenotype	ribosome-associated chaperone de novo protein folding

Chaperones in this family include *E. coli* DnaK of the Hsp70 family (Bukau & Horwich, 1998). Foldases carry out complete refolding of misfolded polypeptide in sequestered

environment, enabling it to reach its native conformation without the crowding effect from the cytosol. The *E. coli* GroES/GroEL complex of the Hsp60 family is the main foldase within the cell (Walter & Buchner, 2002). Disaggregases or unfoldases, such as *E. coli* ClpB, participate in the disassembly of misfolded proteins and inclusion bodies; resulting proteins may then be targeted for proteolysis or refolding (Liu *et al.*, 2002).

While some functions can be assigned with confidence, it is important to note that slight alterations in selective and evolutionary pressure and may also cause some of these chaperones to be obsolete or become essential within certain organisms (Schlee & Reinstein, 2002); as such, there is a large body of literature available, as well as research that is devoted to the study chaperones and how they work in different organisms.

1.4.2 Trigger factor: a ribosome associated chaperone

1.4.2.1 Structural features of trigger factor

Trigger factor (TF) occurs as a ~ 50 kDa monomer or ~100 kDa dimeric protein; equilibrium shifts to one form or the other depending on a metabolic time scale (Genevaux *et al.*, 2004; Martinez-Hackert & Hendrickson, 2009). Tertiary structure is based primarily on α-helical coil although some domains are stabilised through β-sheets, (Fig. 1.5); the N-terminal domain contains a Phe-Arg-Lys motif that enables it to bind to the 50S ribosomal exit tunnel while the P domain houses a substrate binding cavity that binds peptides with regions of eight, consecutive basic or aromatic residues (Genevaux *et al.*, 2004; Patzelt *et al.*, 2001). A peptidyl-propyl isomerase domain (PPIase) catalyses the isomerisation of proline residues (Patzelt *et al.*, 2001; Scholz *et al.*, 1997). There is a C-domain whose function is unknown but has been implicated in contributing towards TF's chaperone activity (Kramer *et al.*, 2004); it forms 'arms' with two α-helical protrusions that extend from its surface and while tertiary arrangement of domains results in a 'cradle'-shaped protein with the C-domain in the middle, primary arrangement differs in that the 'arm' domain is at the C-terminal and the 'head', i.e. the PPIase domain is located at the centre of the molecule, from residues 149-250 (Maier *et al.*, 2005).

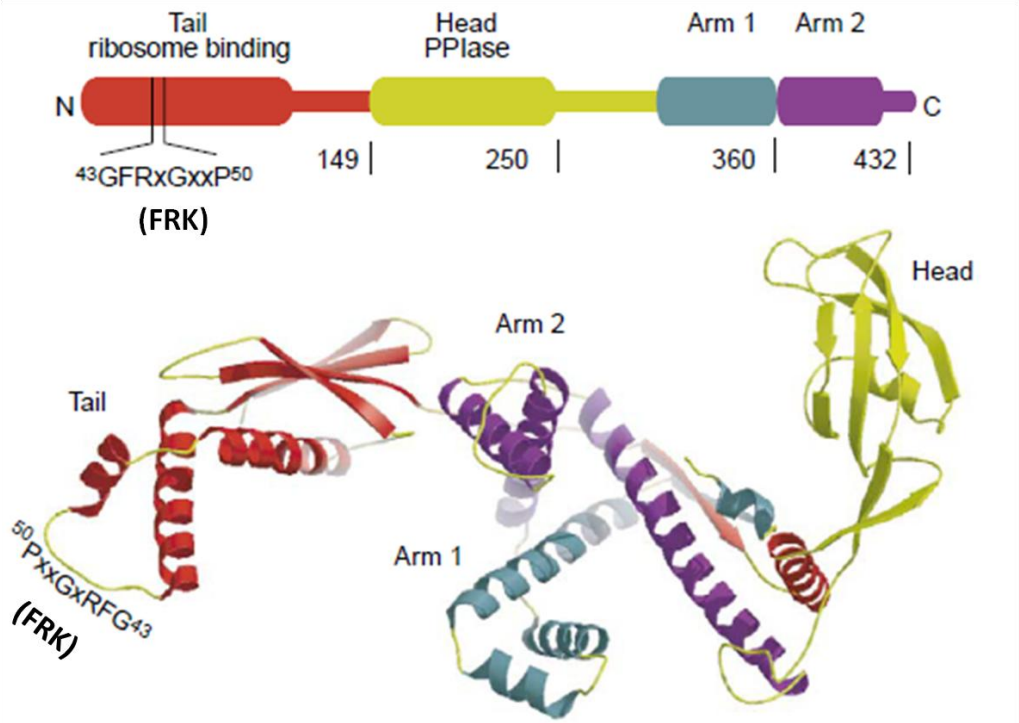

Figure 1.5. The three main domains of *E. coli*'s TF. Tertiary structure arrangement shows that the PPIase domain and N-terminal domain lie on the periphery of the molecule while the C-domain is sandwiched between them; however, the linear arrangement of domains shows the PPIase domain in the centre of the N- and C-terminal. Source: from Hartl & Hayer-Hartl, 2009; Maier *et al.*, 2005.

1.4.2.2 Mechanism of TF chaperone cycle

The TF mediated cycle is relatively simple and is unique in that it is the only molecular chaperone cycle that does not require ATP. Trigger factor binds to the 50S exit tunnel via the L23 region on the ribosome while the L29 region causes conformational changes in TF that expose it conserved motif into the tunnel, ready for the oncoming peptide (Baram *et al.*, 2005). As translation progresses, (Fig. 1.6), the exiting nascent strand increases ribosome/TF affinity for each other and they remain bound to each other. The exposed hydrophobic residues are immediately shielded by the substrate binding cavity where, the nascent chain will internalise its hydrophobic residues to achieve its native state. This equilibrium can be maintained for ~10s (Hartl & Hayer-Hartl, 2009), a sufficient time for short peptides to attain native state conformation; however, larger peptides – 15 kDa – are unable to bury their residues as they await translation of native domains further along the peptide chain and as these become available, initial TF/substrate complexes are destabilised and detach from the

ribosome, enabling uncomplexed TF from the monomer/dimer pool to bind again at the ribosome (Baram *et al.*, 2005; Maier *et al.*, 2003).

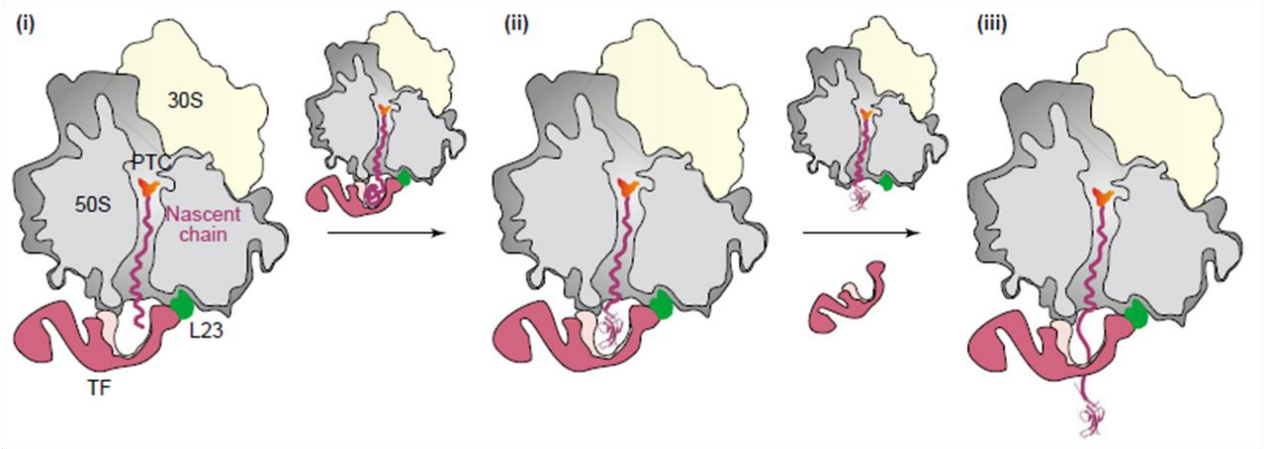

Figure 1.6. Cycle of TF-mediated folding at the ribosomal exit tunnel. Translated peptides are captured by TF in its 'cradle' structure (I) and prevented from interacting with the hydrophilic environment in the cytosol until they have buried their hydrophobic residues (II); if the peptide chain is longer and cannot be accommodated into the cradle at once, TF dissociated from the ribosome and another one takes its place to shield the elongating strand while the old molecule stays bound to the peptide until it has correctly folded (III). Source: Maier *et al.*, 2005.

Trigger factor is one of the most abundant molecular chaperones which is present in the cell as a monomer-dimer pool at equilibrium; it associates with the ribosomal exit tunnel in a 1:1 stoichiometric ratio and cytosolic abundance of ribosome/TF species is proposed to be ~90% of all ribosomes (Maier *et al.*, 2003; Maier *et al.*, 2005). In solution, the 'free', dimerised species of TF have a show decreased chaperone activity and soluble protein yield is much lower, in comparison to ribosome-bound TF (Scholz *et al.*, 1997). This might depend on the residence time of TF on the ribosome; in such a case, TF remains bound to the ribosome far longer and has a longer time to shield its substrate and is not required only for folding but for isomerisation of any propyl residues by the PPIase domain. As a free molecule in the cytosol, TF associations with substrate have a shorter half-life, in the order of milliseconds, usually observed for shorter peptides (Maier *et al.*, 2003).

1.4.2.3 Physiological role of Trigger factor

Trigger factor is the first chaperone that nascent polypeptides encounter upon exiting the ribosome (Maier *et al.,* 2005). It appears to be the only molecular chaperone that is specially adapted to bind to the ribosome, suggesting a role in the early protein biogenesis (Valent *et al.*, 1995); it also means that the range of functions it can carry out as a chaperone is limited to preventing protein aggregation (Deuerling *et al.,* 2003; Maier *et al.,* 2005) unlike the other chaperones, whose functions often extend to folding of proteins. In addition, TF is not known to act downstream of other chaperone systems (Hoffmann *et al.*, 2010) and so plays no part in rescuing misfolded proteins; nor is there any literature that suggests proteins folded by downstream chaperones are shuttled back to TF for refolding. According to Hartl & Hayer-Hartl, (2009), TF interacts with most of the 2400 proteins that exist in *E. coli*, ~ 70 % of which need no further folding by downstream chaperones like DnaK or the GroEL/GroES system (Hoffmann *et al.*, 2010).

Trigger factor is unique in that it is both an enzyme, through its PPIase domain, and a molecular chaperone through its C-terminal substrate domain. This is not to say it is a non-essential domain as mutants lacking it have a decreased viability (Kramer *et al.,* 2004). Another distinct property is the ATP-independent cycle that is absent in other chaperone families. The PPIase domain catalyses the *cis-trans* isomerisation of propyl residues, but studies show that it can also bind peptides with little or no proline (Maier *et al.,* 2003; Patzelt *et al.*, 2001).

It was suggested that TF generally folded short peptides exiting the ribosome, while the larger proteins were left for DnaK, yet, Maier *et al.,* (2003), confirm that larger proteins also associate with TF, and indeed shows a greater affinity, ~100-fold higher, for them. Its ability to assemble the S7 ribosomal protein, as shown by Martinez-Hackert & Hendrickson, (2009) indicates that it can act as an assembly factor for large proteins.

Expression of TF is constitutive (Hoffmann *et al.*, 2010) and is not increased upon cell stress. Mutants lacking TF show no change in phenotype, but double deletion of TF and DnaK are fatal to cells, especially above 30°C (Hartl & Hayer-Hartl, 2002; Hoffmann *et al.*, 2010) due to the fact that their holdase activities overlap functionally. As such, one or the other must be present, although, due to the lack of a heat shock activity, TF will not substitute DnaK at higher temperatures.

1.4.3 Hsp70 family: DnaK holdase

1.4.3.1 Structural organisation

DnaK in *E. coli* is a monomeric protein of the Hsp70 family which associates with an Hsp40 co-chaperone, DnaJ, and a nucleotide exchange factor, GrpE (Betiku, 2006). The DnaK chaperone is 70 kDa protein with an N-terminal ATPase domain of ~44 kDa and a C-terminal substrate binding domain (SBD) of ~27 kDa, which also has β-sheet domain that recognises extended regions of five to seven hydrophobic residues often exposed in client proteins (Hartl & Hayer-Hartl, 2009; Mayer & Bukau, 2005). An α-helical segment extends outwards from the N-terminal side of the SBD, which participates in ATP dependent opening and closing of the SBD, (Fig. 1.7).

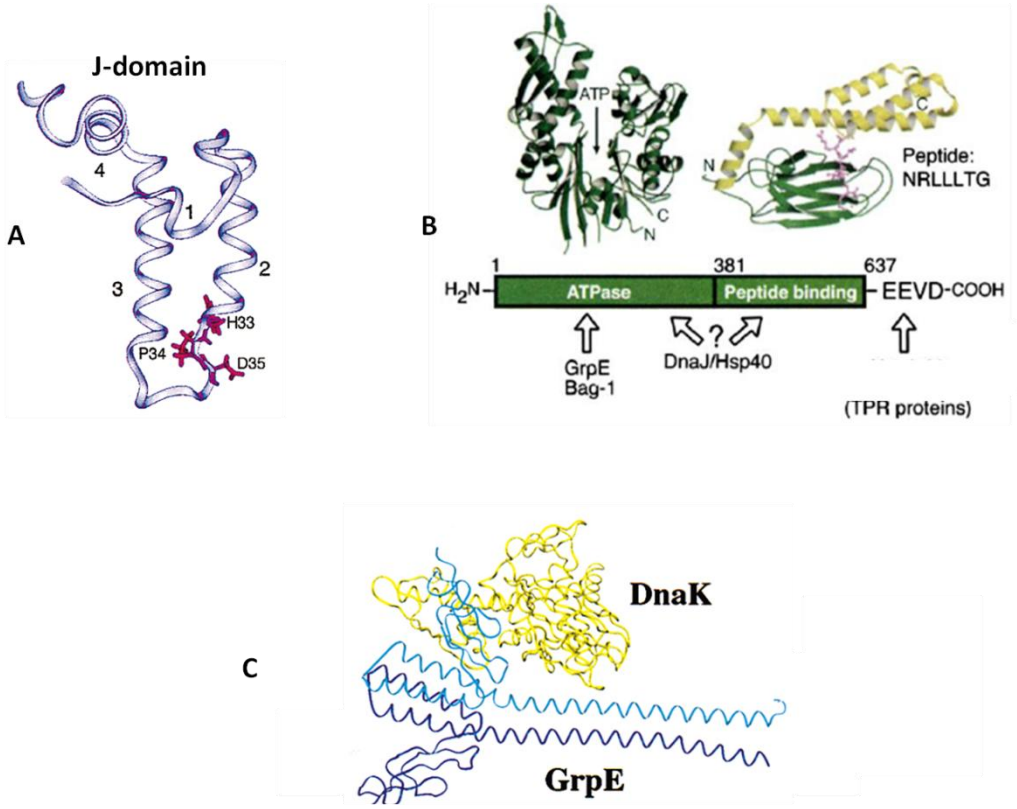

Figure 1.7. The arrangement of conserved structural domains in DnaK, DnaJ and GrpE. In A, the J domain of DnaJ is shown with the four helices labelld from 1-4. (B) Structure of DnaK showing 'lid' (in yellow) that covers the substrate binding domain (SBD) to trap the peptide (in purple) into the SBD cavity. (C) GrpE complexes with DnaK via an allosteric site in tis ATPase domain. Source: Bukau & Horwich 1998; Hartl & Hayer-Hartl 2002.

DnaJ is a 40 kDa chaperone with a highly conserved N-terminal domain of 73 - 78 residues known as the J domain (Fink 1999; Hennessy *et al.*, 2005); the J domain interacts with DnaK while the C-terminal domain is able to bind client proteins. DnaJ belongs to the Hsp40 family, with over 100 known homologs in different organisms; and while the J domain is highly conserved, no sequence homologs have been found for the C-terminal domain (Fink, 1999). GrpE is a 20 kDa homodimer and although it is unrelated to the chaperone family, it is always associated with the DnaK/DnaJ chaperone system where it acts allosterically as a nucleotide exchange factor (NEF) in the exchange of ATP, (Fig. 1.7) (Hartl & Hayer-Hartl, 2002; Hartl & Hayer-Hartl, 2009).

1.4.3.2 Functional cycle of DnaK chaperone

DnaK recognises unfolded peptides through the hydrophobic patch in the SBD. Client proteins include those with corresponding hydrophobic stretches of amino acids, particularly, leucine. These residues repeat every 40-100 amino acids and are often buried in properly folded proteins but exposed in misfolded one (Hartl & Hayer-Hartl, 2009; Hartl & Hayer-Hartl 2002).

In the ATP-dependent cycle, DnaJ binds to the exposed residues on a substrate protein with the C-terminal domain and acts to recruit proteins for DnaK (Hartl & Hayer-Hartl, 2009). At this stage, DnaK has low affinity for the substrate, ATP is bound to the ATPase domain and the 'lid' is in an open conformation which allows access to the SBD (Hartl & Martin, 1995). DnaJ has a high affinity for DnaK and binds to its SBD allosterically via the N-terminal J domain and this binding brings the substrate in close proximity with DnaK's SBD. Binding also effects hydrolysis of the bound ATP, which turns DnaK into a high-affinity molecule with its substrate (Mayer & Bukau, 2005). This causes conformational changes that draw the substrate further into the cavity of the SBD and the lid adopts a closed conformation that tethers the peptide in place, (Fig. 1.8), with subsequent release of DnaJ. This is the holdase activity of DnaK.

The release of the bound substrate is initiated by GrpE nucleotide exchange factor, which binds to an allosteric site within DnaK's ATPase domain. It triggers the release of ADP from DnaK, which reverts to its low substrate affinity state; the ADP binds to GrpE instead and unbinds from DnaK, which enables ATP to bind to DnaK again and start a new holdase cycle (Bukau & Horwich, 1998).

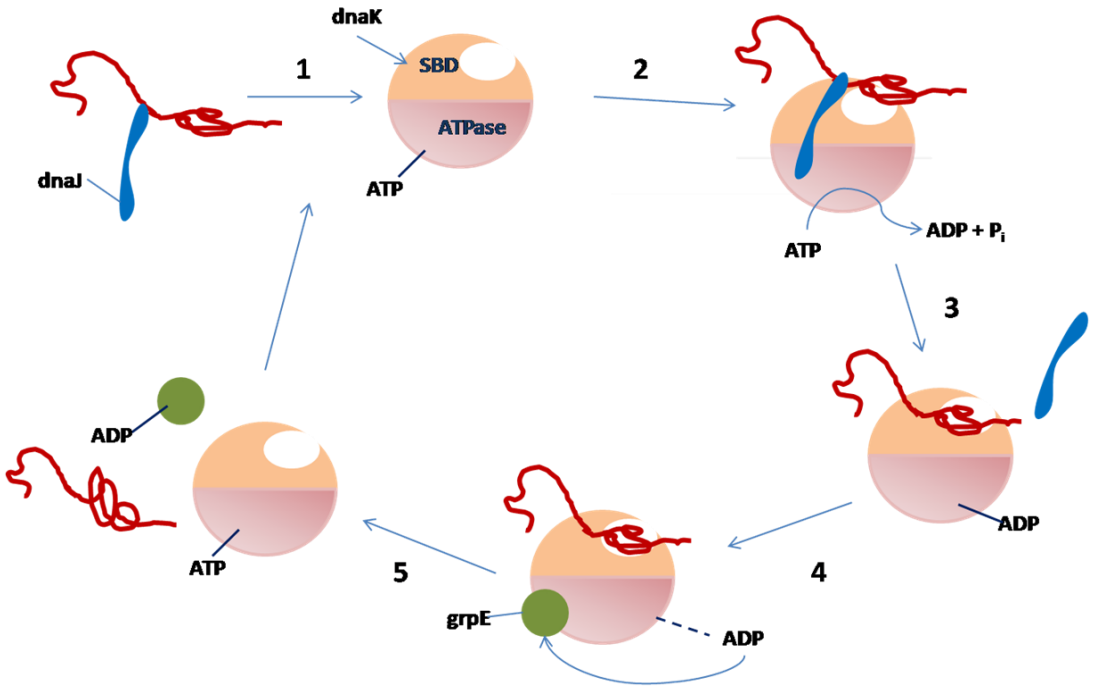

Figure 1.8. The functional cycle of DnaK with its co-chaperone, DnaJ. The 'holdase' activity is divided into a low affinity phase in which ATP is bound to DnaK and a high affinity phase in which ATP is hydrolysed to ADP allowing substrate retention at DnaK's substrate binding domain. Adapted from Mayer & Bukau, 2005.

The stable ADP-bound state of the DnaK/client protein complex lasts only as long as the time it takes for GrpE to be exchanged in the ATPase domain but this is sufficient time for native domains to associate and fold correctly. DnaK can bind substrate without the aid of DnaJ but it has been shown rate of binding is greatly enhanced by having DnaK as a co-chaperone (Hartl & Hayer-Hartl, 2002).

1.4.3.3 Physiological role of DnaK

The DnaK chaperone complex is another set of essential molecular chaperones inside the cell but unlike the chaperonins, it is more versatile with respect to their physiology. It can bind any exposed residues on any protein and is also involved at every step of a protein's life cycle: from a nascent polypeptide just out of the ribosome right its proteolysis by the cell's proteases (Hartl & Hayer-Hartl, 2009).

This is achieved by its ability to act in concert with other chaperone systems such as trigger factor at the ribosome; cycling substrates to and from the GroEL/GroEL system during

folding and refolding; and finally, it has been shown to associate with ClpB to disassemble aggregated proteins which are sometimes targeted for proteolysis.

Small proteins, ~57 residues, often do not require to pass through the DnaK chaperone as they fold spontaneously or are folded as they exit the ribosome by TF; however, about 20% of nascent polypeptides are known to associate with DnaK (Hartl & Hayer-Hartl 2009), slightly higher than the 15% proposed for GroEL (Hesterkamp & Bukau 1998). While most nascent proteins associate with trigger factor, initially, mutants lacking TF have shown to be viable as the workload is now transferred to DnaK and as a result, often show no phenotype to trigger factor deletion (Hartl & Martin, 1995; Hartl & Hayer-Hartl, 2002).

Short peptides may adequately fold by themselves and may not require DnaK assistance but Tomoyasu *et al.*, (2001) have also identified at least 93 *E. coli* proteins ranging from 21 – 167 kDa that are aggregation prone at the physiological temperature of 30°C, when DnaK is absent in the cell. DnaK works to shield exposed hydrophobic patches on polypeptides, to prevent possible aggregation of non-native domains, as well as to shield them from the aqueous cytosolic environment that could enhance their precipitation into inclusion bodies (Hesterkamp & Bukau, 1998; Hartl & Hayer-Hartl, 2002).

According to Carrió & Villaverde, (2005), DnaK interactions also extend to inclusion bodies, microscopy images show that in inclusion bodies, co-precipitated DnaK is localised at the perimeter of the aggregated molecules; suggesting that they also have an important role in the solubilising inclusion bodies. If the cell is to build thermotolerance, then it requires some kind of 'shock absorber' and that is provided by inclusion bodies. Therefore, degrading all the inclusion bodies would be physiologically detrimental; in this case, association of the DnaK system with inclusion bodies appears to delay the proteolytic activities of other Clp ATPases, which are discussed later (Haslberger *et al.*, 2010).

DnaK homologs and DnaJ-like proteins have been identified in *E. coli* and are also involved in preventing protein aggregation, although their activity is generally detected during cell stress, such as starvation. There appears to be no division of labour among these homologs and DnaK or DnaJ under physiological conditions but experimental data demonstrates their capacity to maintain cell viability in mutants lacking DnaK and DnaJ (Hartl & Hayer-Hartl, 2002; Hesterkamp & Bukau, 1998). Under normal, physiological conditions, DnaK binds unfolded proteins or σ^{32}, in a competitive manner; σ^{32} is an RNA polymerase recruiting factor that is associated with promoters of heat-shock genes (Guisbert *et al.*, 2004). Binding of σ^{32}

to DnaK renders the factor inactive and it cannot bind to heat-shock inducible promoters. However, when the concentration of unfolded proteins increases, due to a stress response, σ^{32} is competitively substituted by these proteins on DnaK's substrate cavity and enables recruitment of RNA polymerase to other heat-shock genes with subsequent increase in their translation and a decrease in protein aggregation (Guisbert *et al.*, 2004).

DnaK is also involved in a host of other non-folding functions that include activation of RepA/E and DnaA for chromosomal replication in *E. coli*; initiation of replication in λ phage DNA, the production of flagella and protein trafficking (Watanabe *et al.,* 2000), indicating why deleterious mutations of this gene are fatal to the cell.

1.4.4 Hsp60 family: The chaperonins

1.4.4.1 Structural organisation

Chaperonins are divided into group I chaperonins and occur in eubacteria, chloroplasts and mitochondria; group II chaperonins are found in eukaryotes and in archaea (Furutani *et al.*, 1998) . This section only treats group I chaperonins which share a number of homologous features with group II chaperonins.

The GroEL/GroES chaperone system is made up of the protein GroEL and its co-chaperone GroES, (Fig. 1.9). GroEL is a 60 kDa protein in the Hsp60 family known as chaperonins (Cheng *et al.*, 1989) while GroES is a 10 kDa protein belonging to the Hsp10 family. Chaperonins are made of two, barrel-like units of approximately 800 kDa (Vabulas *et al.*, 2010) and each unit consists of seven GroEL proteins arranged symmetrically around a central axis (Bukau & Horwich, 1998) forming apical, intermediate and equatorial domains (Hartl & Martin, 1995). The apical domain contains seven hydrophobic side chains – one per heptamer – and is the site of substrate and GroES, which binds via a small loop in it structure (Hartl & Martin, 1995; Saibil, 2008). The intermediate domain is flexible and enables some degree of conformational movement of the apical domain during GroES and substrate binding (Banach *et al.*, 2009). The equatorial domain provides residues to link the two heptameric rings back to back (Martin, 1997).

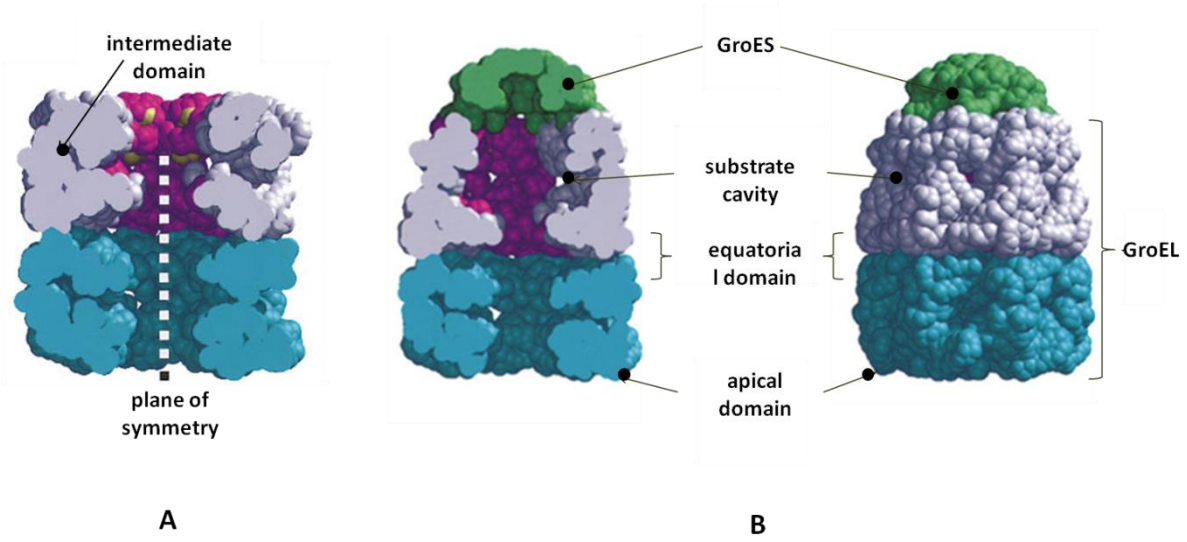

Figure 1.9. Vertical cross-sectional view of *E. coli* GroEL and GroES chaperones. The symmetrical double rings are joined at the equatorial plane while the intermediate domain provides hinge-like flexibility for the apical domain (A); the hydrophobic side chains in the apical domain are shown in yellow. GroES binds to the apical domain to form a sequestered, hydrophilic environment in which the client protein may fold. Adapted from Bukau & Horwich, 1998.

According to Bukau & Horwich, (1998) and Ranson *et al.*, (1998), GroES is an heptamer with the difference that it forms a lid-like structure as opposed to the open barrel formed by GroEL. Upon binding, GroES efectively caps the open-ended GroEL ring creating a cavity in the interior. This is the substrate cavity and traverses the entire apical, intermediate and equatorial domains; it is hydrophilic and provides an uncrowded environment, referred to as Anfinsen's cage, in which the substrate can fold (Ellis, 2003). The interior of the equatorial region also houses the ATPase domain; each heptamer has one which plays a role in conformational changes of the chaperonin cavity (Bukau & Horwich, 1998).

1.4.4.2 Mechanism of chaperonin-mediated folding

The folding cycle of GroEL can be summarised as binding, encapsulation and release of the client protein. It is initiated through substrate binding to the GroEL apical domain in a GroES/GroEL/ADP complex (Saibil, 2000). Client proteins with exposed hydrophobic residues interact with at least three hydrophobic side chains in the apical domain in a bid to bury their hydrophobic domains. This triggers binding of GroES to the apical domain via a flexible loop that forms a 'hinge' and ATP to the ATPase domains. Both events cause extensive torsional and conformational changes in the cavity: the substrate cavity's volume

increases approximately two fold while the hydrophobic residues in the apical domain are retracted further inwards into the GroEL wall. Without these apical residues, the exposed residues of the client protein are released and 'dropped' into the hydrophilic cavity. The protein is now in an uncharged, uncrowded environment, where it is allowed to fold. Hydrolysis of ATP to ADP in the equatorial domain takes approximately 20s to complete, which allows the encapsulated protein ample time to fold to its native state, without the risk of proteolysis (Betiku, 2006; Bigotti & Clarke, 2008; Ranson *et al.*, 1998).

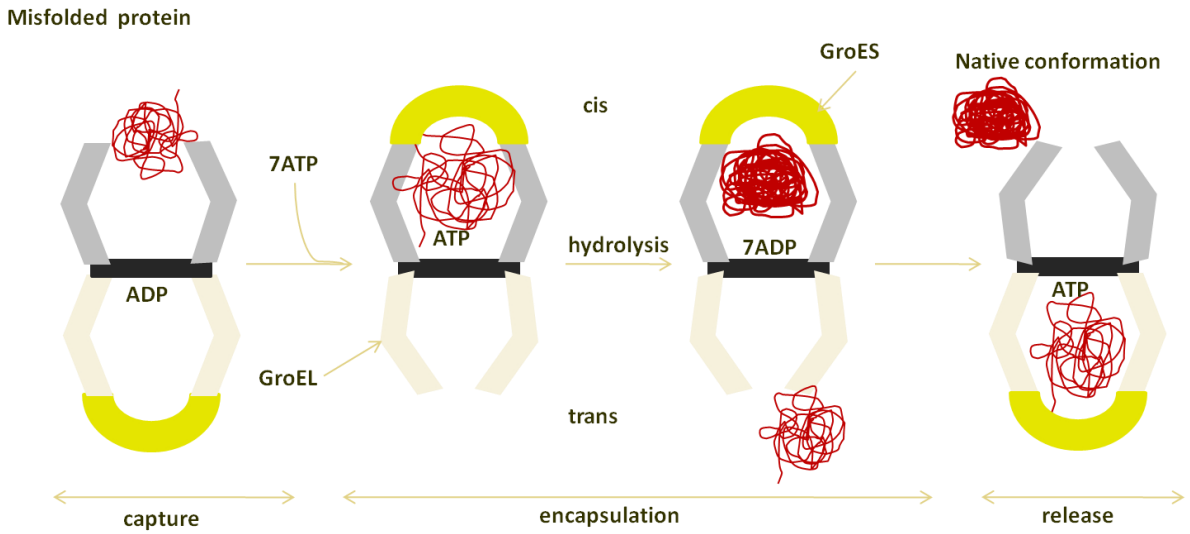

Figure 1.10. The functional cycle of GroEL and GroES chaperones in *E. coli*, illustrating the three steps of ATP dependent folding – capture or binding of the substrate protein, encapsulation into the GroEL by GroES and release of native protein. The diagram also shows the antagonistic relationship between the two GroEL rings with respect to substrate binding; both rings cannot be saturated with substrate at the same time and occurs out of phase, instead. Adapted from Bukau & Horwich, 1998 and Ranson *et al.*, 1998.

Substrate binding and folding occurs in both rings of GroEL but rather than doing in parallel, binding of the substrate to the GroEL rings of one complex appears to be mutually exclusive, also shown in Fig. 1.10. The substrate binds to one GroEL ring, termed the *cis* ring and causes conformational changes in the unbound ring, termed the *trans* ring. This results in narrowing of the opening so that substrates cannot bind to the trans hydrophobic residues (Saibil, 2000; Ranson *et al.*, 1998). Upon ATP hydrolysis in the *cis* ring, the *trans* ring is now able to bind substrate and ATP and this is what causes unhinging and subsequent release of the cis GroES and substrate protein (Ranson *et al.*, 1998). The former *trans* ring, now a *cis* ring, carries out folding events as described before, with folding being shuttled between the

two rings. The folding of nascent proteins and misfolded proteins is not a once-off interaction; it may be repeated as long as is required although little is known on how the cell determines this. It has been shown, however, that proteins that are not correctly folded are not often released completely into the cytosol but remain attached to the apical domain of GroEL via the hydrophobic residues, ready for another folding cycle. Folding by the GroEL/GroES complex seems to be, largely, a mechanical process based on hydrophobic interactions of related domains in the client protein (England & Pande, 2008), no data has been collected on biochemical interactions of its residues with the client molecule residues.

1.4.4.3 Physiological role of chaperonins

Chaperonins typically fold proteins in the 20-50 kDa range and the maximum size that can be accommodated in the substrate cavity is 60 kDa; when the chaperones encounter larger proteins, such as the 82 kDa *S. cerevisiae* mitochondrial aconitase, GroEL only associates with the exposed hydrophobic region, without the aid of GroES, thereby acting as a holdase rather than a foldase (Chaudhuri *et al.*, 2001). The holdase function of the chaperonins is, however, not efficiently modulated as its foldase activity when it encapsulates its substrate protein (Ellis, 2003).

Chaperonins are essential to the cell and are constitutively expressed, although they also play an important role in the rescue of misfolded proteins during heat shock (Becker & Craig, 1994; Guisbert *et al.*, 2004). They fold approximately 250 proteins, ~10%, within *E. coli* with 85 being stringently specific for GroEL and GroES, such as malate dehydrogenase or they are prone to aggregation and proteolysis (Hartl & Hayer-Hartl, 2009; Ranson *et al.*, 1998). They are the chief folding chaperones for large, multi subunit proteins, which tend to have complex α-helical and β-sheet arrangements that make them prone to aggregation (Saibil, 2000; Vabulas *et al.*, 2010).

Chaperonins and DnaK work in the same pathway, with DnaK working upstream and transferring its substrates to the chaperonin machinery. Likewise, chaperones that are able to leave the chaperonin cavity may still require DnaK assistance before achieving their final conformations (Becker & Craig, 1994; Braig 1998).

1.4.5 Hsp100 family: ClpB disaggregase machine

1.4.5.1 Structural features of ClpB

ClpB is a member of the Clp ATPases, large multi-subunit proteins, which include ClpA, ClpX and ClpY (Liu *et al.*, 2002). It belongs to a class of proteins known AAA+ – ATPase associated with a variety of cellular activities – which have conserved regions that participate in ATP binding, hydrolysis and subunit oligomerisation (Strub *et al.*, 2003). It is highly conserved in bacteria and as such, available crystallography data available for the thermophilic bacterium, *Thermus thermophilus*, is used as a structural and functional model for other prokaryotic ClpBs, including that of *E. coli* (Doyle & Wickner, 2009; Lee & Tsai, 2005).

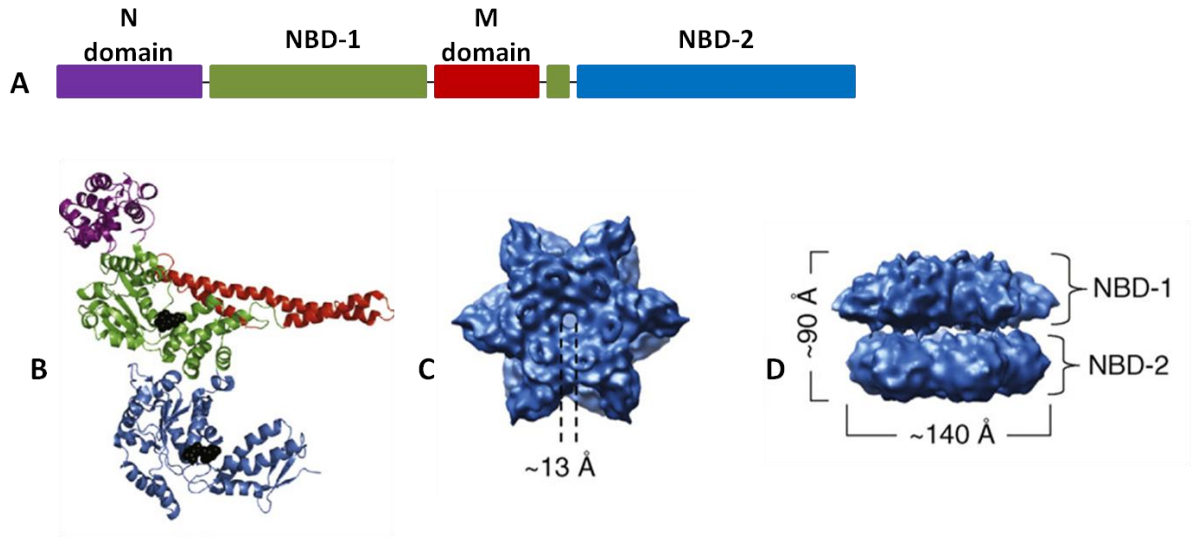

Figure 1.11. Tertiary structure of *T. thermophilus* ClpB chaperone. (A) Linear arrangement of residues for the ClpB protein. (B) Tertiary domain arrangement of one ClpB unit showing NBD-1 and NBD-2, flanking the M domain, which projects outwards. C and D show top and side views of the hexameric assembly of the protomer into a ring structure with a central pore. The characteristic 'star' shape is produced by the protruding M domain helices. Source: Doyle & Wickner, 2009.

ClpB is a hexameric ring assembly; each unit contains two ATPase or nucleotide binding domains (NBD 1 and 2) joined to each other via a middle (M) domain (Tek & Zolkiewski, 2002; Doyle & Wickner, 2009), (Fig. 1.11). The N domain influences substrate binding affinity while the C-terminal region, near NBD-2, regulates self association of protomers into

a multimeric protein (Barnett *et al.*, 2000). The hexamer forms a central, 13 Å pore that traverses the entire protein and which is important in protein binding, remodelling and disaggregation (Doyle & Wickner, 2009). This remodelling activity is due to the rich distribution of hydrophobic residues, especially tyrosine, which line the pore opening in NBD-1 and NBD-2 (Haslberger *et al.*, 2010).

Other prokaryotic members of the Hsp100 family are ClpA, ClpC, ClpX and ClpY which bind to ClpP or ClpQ peptidases to carry out proteolytic degradation (Tek & Zolkiewski, 2002; Maurizi & Xia, 2004). All have the two stacked NBD domains common to Clp proteins, with the exception of ClpX and ClpY which have a single NBD domain that carries out all the work (Lee & Tsai, 2005).

1.4.5.2 Mechanism of ClpB-mediated disaggregation

Disaggregation by ClpB is an ATP-dependent process that occurs via translocation through the central pore. ATP-/ADP-bound states of NBD-1 and NBD-2, respectively, occur out of phase with each other and in this state the ClpB hexamer is stable (Haslberger *et al.*, 2010; Maurizi & Xia, 2004). Equilibrium favours substrate binding through exposed hydrophobic domains on the misfolded or aggregated protein as well as the tyrosine residues around the pore lining, which are now displayed to the substrate. Due to the smallness of this pore, only one or two polypeptide strands can be accommodated so they are effectively 'threaded' through NBD-1. Hydrolysis of ATP in NBD-1 results in conformational changes whereby tyrosine residues in this domain are internalised and ATP binds NBD-2, the result of which is loss of substrate affinity by NBD-1 and increased substrate affinity by NBD-2. In this manner, polypeptides are unravelled strand by strand or domain by domain and pulled through the chaperone and subsequent hydrolysis of ATP in NBD-2, once more, reduces the domain's affinity for its substrate and effects the exit of the protein out of the chaperone, (Fig. 1.12a), (Doyle & Wickner, 2009; Haslberger *et al.*, 2010). Maurizi & Xia, 2004 suggest that the ATPase activity of NBD-1 is much weaker than that of NBD-2, one can speculate on whether it is this difference in ATPase activity that also directs the substrate to the second domain and subsequently leads to unidirectional ejection out of NBD-2 rather than NBD-1; however, no data exists for this view. The resulting polypeptide is available for refolding events that may lead to its native conformation; this may occur spontaneously or with the aid of the DnaK or the GroE chaperone systems (Maurizi & Xia, 2004), as previously described.

In addition to independent remodelling of proteins, ClpB also associates with the DnaK/DnaJ/GrpE system; in fact, ClpB is so often co-purified with the DnaK chaperones that the independent unfolding activity of ClpB has been identified only recently (Doyle & Wickner, 2009). The interaction between the DnaK complex and ClpB is unknown but in this case DnaK and its co-chaperones do not act holdases but rather as substrate recruiters for ClpB, (Fig. 1.12c) . In addition, other Clp ATPases such as ClpA, associate with the ClpP peptidase via their NBD-2 domains, in a manner that aligns their central pore to that of the protease; in this manner, protein aggregates are unfolded and threaded through the ATPases and accepted by ClpP which degrades them, (Fig. 1.12b) (Doyle & Wickner, 2009).

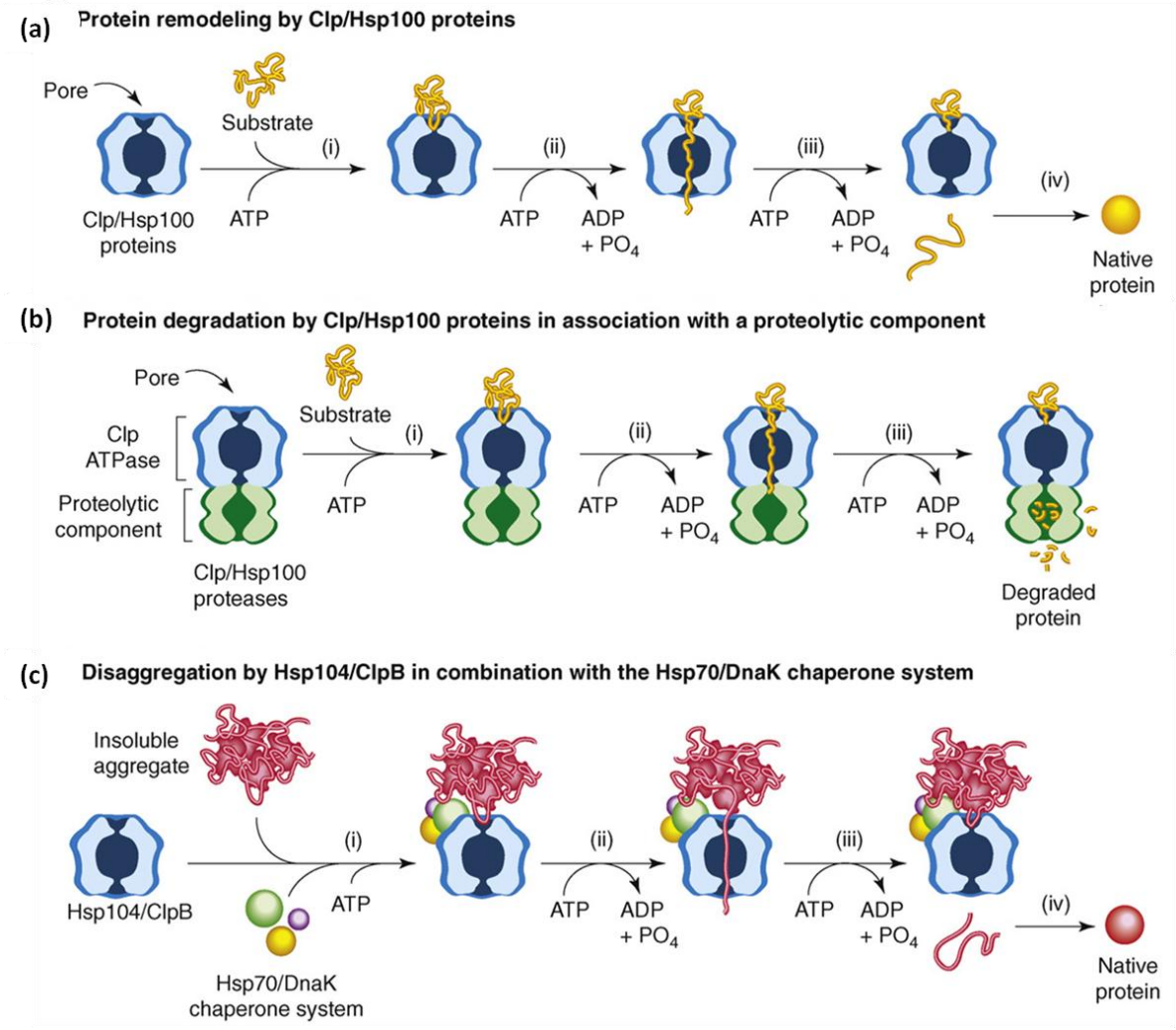

Figure 1.12. The ATP-dependent disaggregation of proteins by Clp ATPases. (A) Disaggregation of a polypeptide by the NBD domains as proposed for ClpB. (B) Association of a Clp ATPase with ClpP leads to protein unfolding and subsequent degradation. (C) DnaK/DnaJ/GrpE system interacts with ClpB to carry out large-scale unfolding of protein aggregates. Source: Doyle & Wickner, 2009.

1.4.5.3 Physiological role of Clp ATPases

Most of the Clp ATPases in *E. coli* are involved in protein degradation but ClpB is the only one that is known to disaggregate misfolded proteins; which in turn gives them a second chance ate refolding correctly (Tek & Zolkiewski, 2002). They are particularly important during heat shock and other forms of severe cell stress, where they are tasked with solubilising high concentrations of aggregated proteins (Kedzierska & Matuszewska, 2001); in this case, the rescued protein is transferred to holdases such as DnaK and the GroE (Watanabe *et al.*, 2000) foldase for refolding into the correct conformation, a process that restores their catalytic activity (Haslberger *et al.*, 2010). In addition, the initial response to heat-shock by ClpB is followed by a gradual increase in thermotolerance by cells (Haslberger *et al.*, 2010) and the ClpB/DnaK system is a key contributor to pathogenicity in some prokaryotes. Association of ClpB with DnaK/DnaJ/GrpE improves the ability to remodel heterogeneous aggregates at least 15 times more than if ClpB were to carry it out independently (Haslberger *et al.*, 2010). Previously, Doyle *et al.,* (2007) had presented evidence for independent remodelling of GFP and YFP, so while it does happen, this activity is particularly poor. DnaK and its co-chaperones participate in extracting single strands from aggregated protein masses and transfer them to ClpB (Haslberger *et al.*, 2010).

The ability of Hsp100 chaperones to associate with the cell's proteolytic machinery is important, especially if one considers the effect of cytotoxic concentrations of aggregated proteins in the cytosol. By targeting insoluble aggregates to proteases like ClpP, inclusion bodies are minimised and cell viability can be maintained.

1.4.6 HtpG and the small Hsps

In addition to the main families of molecular chaperones active within *E. coli*, several other molecular chaperones or chaperone-like proteins exist; they may be uniquely structured but their functions and mechanism are slight variations of the three described above. These chaperones may work independently; however, they are often observed working co-operatively with the three main chaperone systems in varying capacities. And while the Hsp60/Hsp70 chaperones are closely linked to early folding stages through to the mature, native phase of a protein, the other chaperones are restricted to the later stages of protein folding (Guisbert *et al.*, 2004; Shiau *et al.*, 2006). They are also important chaperones in *E. coli'*s heat-shock response, especially above temperatures of 45°C, or when the cell's main chaperone holdase, DnaK, has been deactivated (Thomas & Baneyx, 2000).

HtpG is the *E. coli* homolog for the Hsp90 family of chaperones; *in vitro*, it has ATP-dependent holdase activity but its *in vivo* physiological role, with respect to protein folding, is yet to be revealed (Csermely *et al.*, 1998); there is speculation that it is involved in protein secretion (Thomas & Baneyx, 2000). It forms a homodimer whose N-terminal domains (NTD) create a cleft in which the substrate protein can bind, as proposed by Shiau *et al.*, (2006), (Fig. 1.13A). Hydrophobic residues line the cleft in the middle (MD) domain, as well as the C-terminal domain (CTD) and are displayed to misfolded proteins; upon substrate binding, the resulting complex is stabilised by addition of ATP. When this is hydrolysed, the hydrophobic residues in the dimer are buried into the MD and CTD, the dimer adopts a closed conformation that brings the N-terminal domains in close proximity in a 'pincer' – like shape and the substrate is released from the cleft. As ADP dissociates from NTD, the conformational changes in the NTD relax the dimer and it adopts is open shape once more for another cycle of folding. HtpG is essential in preventing aggregation when cells are subjected to mild heat shock, ~42°C but it is unable to complement DnaK-deficient strains under the same conditions (Thomas & Baneyx, 1998).

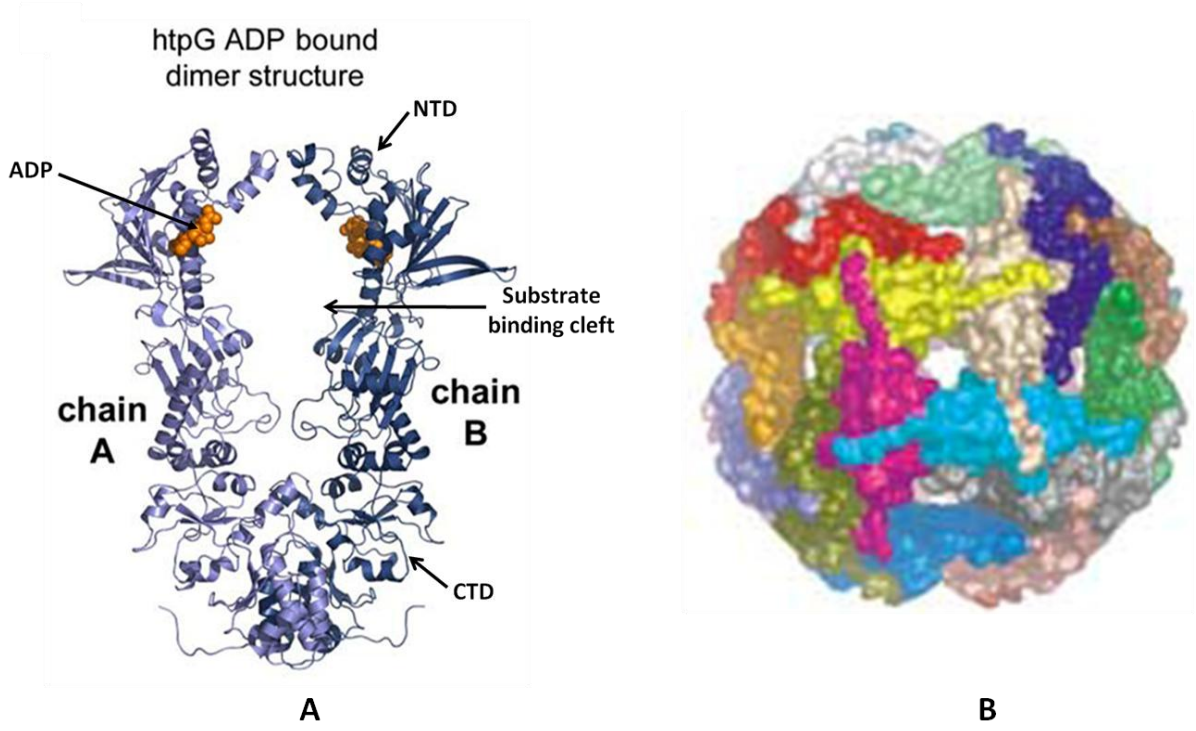

Figure 1.13. Structure of the HtpG chaperone dimer from *E. coli* (A) and the Hsp16.5, a member of the sHsp family, from *Methanococcus jannaschii* (B). In B, each coloured segment represents an sHsp protomer. Adapted from Shiau *et al.*, 2006; Stirling *et al.*, 2003.

Small heat shock proteins (sHsps), known as IbpA and IbpB in *E. coli*, are 10 – 30 kDa monomers (Thomas & Baneyx, 2000) that assemble into a wide variety of oligomeric conformations, seen in Fig. 1.13B (Stirling *et al.*, 2003). Thomas & Baneyx, (1998) observed that IbpA and IbpB work together with the DnaK chaperone system to counteract heat shock, although null mutants of these two chaperones produces a negligible decrease in metabolism under physiological conditions. Despite this, Ario de Marco *et al.*, (2007), identified 12 proteins that expressly require the sHsps. They participate in preventing aggregation of hydrophobic patches independently of ATP (Kolaj *et al.*, 2009) and rather than having a central cavity in which hydrophobic residues are shielded as observed in other chaperones, aggregated proteins are held on the surface of the oligomeric structures; however, the mechanism of interaction has not been elucidated (Kolaj *et al.*, 2009; Stirling *et al.*, 2003). IbpB chaperones also co-operate with other non-chaperone proteins involved in heterologous expression stress response to maintain cell membrane integrity (Ami *et al.*, 2009).

1.4.7 Molecular chaperones from extremophiles

As mentioned before, molecular chaperones are highly conserved across species and domains (Kolaj *et al.*, 2009). A number of the same chaperones found in thermophilic prokaryotes have been shown to carry out similar protein folding or disaggregating functions as their mesophilic counterparts (Sugimoto *et al.*, 2003; Watanabe *et al.*, 2000). These include DnaK and GrpE from *T. thermophilus* (Nakamura *et al.*, 2010), ClpB from the same organism (Schlee *et al.*, 2001), and GroEL homologs in the thermophilic cyanobacterium *Thermosynechococcus elongatus* (Sato *et al.*, 2008). However, while the main functional mechanisms are preserved, minor differences have been observed in these chaperone systems (Klostermeier *et al.*, 1999).

One of these differences is that organisms have only a few copies of molecular chaperones and seem adapted to folding their proteins using fewer chaperone systems; for example, ClpB is only unique to *T. thermophilus* and has yet to be identified in other thermophiles while Hsp70 is present in a few mesophilic and thermophilic archeons (Hofman-Bang *et al.*, 1999) but altogether absent in the hyperthermophilic species (Schlee & Reinstein, 2002).

Unlike the *E. coli* (*Ec*) DnaK system, ATPase activity of DnaK in *T. thermophilus* is stimulated by GrpE binding allosterically to it, rather than DnaJ (Groemping *et al.*, 2001). Due to this reversal, the rate of ATP hydrolysis is slower but the rate of nucleotide exchange

is faster and results in slow release of the substrate; this would make its holdase activity more efficient than that of *Ec*DnaK. In addition, an 8 kDa protein, identified as DafA, is essential as an assembly factor for the DnaK-DnaJ complex in *T. thermophilus* (Motohashi *et al.*, 1996).

Chaperonin homologs in archaea, known as thermosomes (Horwich *et al.*, 1999), have a double ring structure with 8 to 9 protomers (Furutani *et al.*, 1998). This is a common feature of Group II chaperonins which are also found in eukaryotes.

These studies have extended into characterising cold-adapted molecular chaperones as well; Robin *et al.*, (2009) have identified TF in a psychrophilic bacterium, *Psychrobacter frigidicola*, which is homologous to *E. coli* TF. It that has a higher capacity to fold proteins efficiently, compared to the *E. coli* homolog. Tosco *et al.*, (2003), also identified a 55 kDa GroEL homolog of *Pseudoalteromonus haloplanktis* that shares more than 80% amino acid identity with *Ec*GroEL and can actually use *Ec*GroES as a co-chaperone.

1.5 Heterologous expression in *E. coli* increases the formation of inclusion bodies

Heterologous expression, the production of proteins in a non-native host, has provided a means to study proteins from prokaryotes and eukaryotes, as well as exploit them commercially in the production of medicines and industrial biocatalysts (Littlechild *et al.*, 2007; Terpe, 2006).

Cells synthesise proteins in low concentrations, yet in order to make use of them, they are required in larger volumes than the cell usually makes, a problem that is solved by overexpression of the protein of interest in a suitable host (Villaverde & Carrió, 2003; Widersten, 1998). This is achieved by engineering multiple copies of a gene on chromosomal DNA or having a gene on a plasmid with a relatively high copy number to produce multiple copies of it (Georgiou & Valax, 1996). In addition, not all proteins can be overexpressed in their native hosts, due to reasons including poor expression ability, host pathogenicity, poorly understood metabolism, difficulty in culturing the host organism and cost effectiveness (Villaverde & Carrió, 2003); still, a few bacterial species, yeasts, filamentous fungi and animal cell lines have been found to be suitable expression hosts (Terpe, 2006).

Bacterial hosts, especially *E. coli*, are viewed as the most convenient expression vectors for a number of reasons; they are inexpensive, their biology is well characterised and several plasmid vectors and promoters have been developed to optimise protein expression. They are also easily cultured and their quick growth rates mean that large biomass may be harvested in a short time with relatively high yields of protein for laboratory and large-scale biotechnology applications (Georgiou & Valax, 1996; Terpe, 2006).

The biggest challenge to recombinant protein expression in *E. coli* and other bacterial hosts is aggregation of the expressed protein into inclusion bodies (Carrió & Villaverde, 2001); a problem that is magnified when overexpression of the recombinant protein is carried out (de Marco *et al.*, 2007). Expressed proteins readily accumulate in the cytosol and periplasm due to *E. coli*'s poor ability to translocate proteins to the extra-cellular environment. This increases cytosolic solute concentration and viscosity due to macromolecular crowding (de Marco *et al.*, 2007).

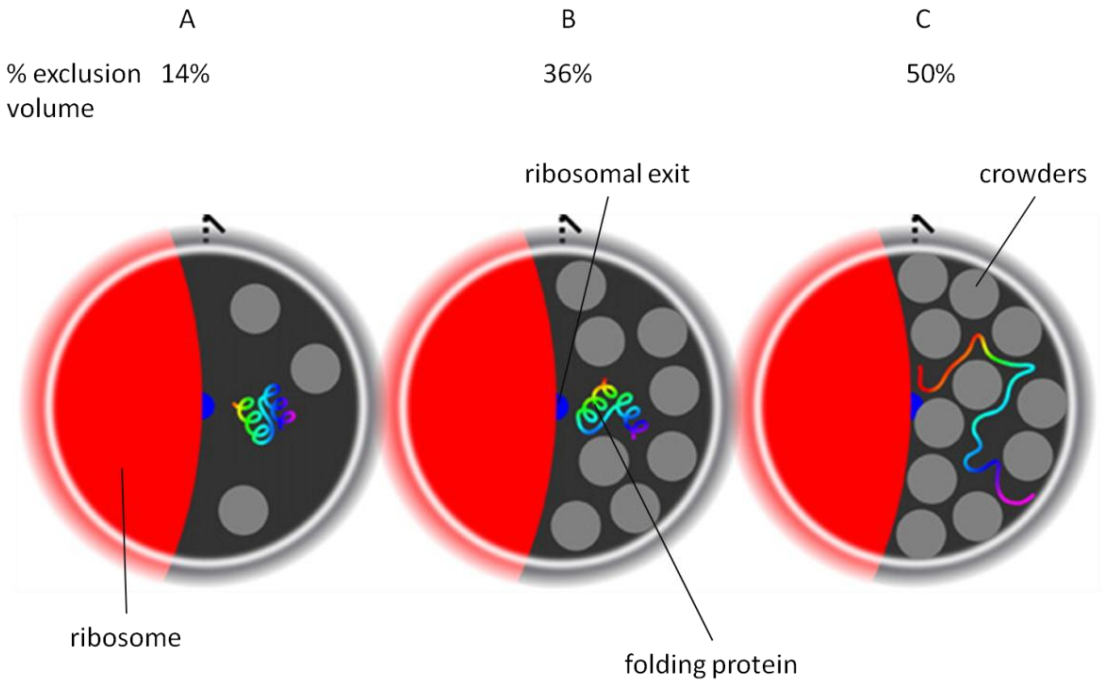

Figure 1.14. The effect of increased macromolecular crowding on the folding of a nascent polypeptide chain, as seen during heterologous overexpression. As the excluded volume increases, the volume available for the new protein to fold is reduced. Initially, this promotes compaction into native conformation (A); however, on increasing crowder concentration, compaction promotes non-native intermediate conformations (B) until it is so high that nascent polypeptides simple collapse into misfolded conformations that aggregate IBs. Adapted from Jefferys *et al.*, 2010.

In the case of heterologous overexpression, most foreign proteins are unable to fold to their native state in bacteria; this is especially true for eukaryotic proteins, several of which require a different environmental set-up – redox potential and pH – to fold correctly (Schrödel *et al.*, 2005). They also need slower rates of co-translational folding (Zhang & Ignatova, 2010), as well as alternative modes of post-translational modification processes that are absent in bacteria, such as intron processing (Jana & Deb, 2005; de Marco *et al.*, 2005).

Under these conditions, the likelihood of proteins exposing their hydrophobic regions and forming non-native intermediates is higher. As protein expression progresses, so the volume of protein increases, the cytosolic environment becomes highly reducing and the pH increases. Macromolecular crowding, (Fig.1.14), increases to a point where nascent chains have little room to fold correctly. All these factors have the cumulative effect of precipitating the protein out of solution to form inclusion bodies (Fink, 1998).

Unlike the inclusion bodies found under normal physiological conditions, however, these are more homogenous with respect to protein composition due to the large concentration of the expressed protein. In most cases, the aggregated proteins are not only made of misfolded proteins; a greater part are intact proteins with native conformation, an important factor when considering how to increase soluble protein recovery in biotechnology (de Marco *et al.*, 2005). Homologous overexpression of native *E. coli* proteins also results in inclusion bodies and de Marco *et al.*, (2005) have related this to the 'unnatural' condition in which they are expressed, i.e. above physiological concentrations.

Overexpression does not only affect the protein of interest but also the host; *E. coli* undergoes structural and physiological changes that include reduced fluidity and permeability of the membrane, modification of intracellular lipid components (Ami *et al.*, 2009) and a decrease in cell biomass as the cells metabolism is shifted towards protein synthesis (Tolia & Joshua-Tor, 2006).

1.5.1 Improving protein recovery in *E. coli*

Despite the high prevalence of inclusion bodies during protein expression, there are several techniques of improving the yield of active, soluble proteins in biotechnology. *In vitro* methods are based on Anfinsen's principle of self-assembly and involve resolubilising inclusion bodies, complete denaturation of proteins and *in vitro* refolding in a controlled environment (de Marco *et al.*, 2005; Schrödel *et al.*, 2005). However, yields of soluble

protein recovered from this are often low in comparison to the time-intensive procedure and protein activity is not guaranteed (de Marco *et al.*, 2005).

In vitro protein folding is limited in terms of how it may be optimised to improve protein solubility; however, *in vivo* expression show a greater range of properties that can be manipulated to increase not only the concentrations of the protein of interest, but also its solubility and activity upon extraction and purification (Georgiou & Valax, 1996; Terpe, 2006). Strategies may involve adjusting fermentation temperature below the optimum for slow expression of the protein of interest (Georgiou & Valax, 1996) and using strong promoters to increase expression levels, or employ weak promoters for tight, regulated expression of toxic proteins that are normally aggregated to reduce cytotoxicity (Terpe, 2006; Kim & Lee, 2008). The use of thioredoxin, NusA (de Marco *et al.*, 2004), His-tags and other affinity tags (Terpe, 2003) or green fluorescent proteins (GFPs) (Sevastsyanovich *et al.*, 2009) as fusion partners is widely used since research has shown how a highly soluble protein, fused to a less soluble partner may drive equilibrium of both towards solubility rather than aggregation and improve subsequent purification steps using affinity chromatography (Makrides, 1996; Schrödel *et al.*, 2005). In the case of using large proteins such as GFPs as fusion partners, the main drawback is that once purified, the protein couple has to be cleaved by a site-specific protease such as tobacco etch virus (TEV) protease (Fang *et al.*, 2007) although this is time consuming and retaining activity of the cleaved product may not be successful (Sørensen & Mortensen, 2005; Terpe, 2003).

Other strategies for optimising correct folding of proteins during heterologous expression are codon optimisation to improve translation efficiency of rare codons and improving translation initiation. Also, targeting expressed proteins to the periplasm or extracellular environment alleviates the build-up of proteins to cytotoxic concentrations that trigger protein aggregation (Makrides, 1996).

1.5.2 Molecular chaperones in heterologous expression of proteins

The ability to fold proteins correctly by molecular chaperones has been exploited extensively to improve yields of soluble, active protein expressed in *E. coli* and *S. cerevisiae* (Makrides, 1996).

Practical approaches to co-expressing chaperones are carried out on chaperone encoding plasmids (de Marco & de Marco, 2004) although stimulation of heat-shock genes with benzyl

alcohol also results in high concentrations of endogenous chaperones (de Marco *et al.*, 2005). Chaperones are expressed as single proteins or as an entire chaperone system (Kolaj *et al.*, 2009); however, Makrides, (1996), proposes the latter to be a better strategy, as chaperones seldom act independently and overexpression of one chaperone may result in plasmid instability and its subsequent loss. Nishihara *et al.*, (1998), developed several vectors for the co-expression of TF, DnaK-DnaJ-GrpE and GroEL/GroES in different chaperone sets, (Fig. 1.15); they are now available commercially through Takara Bio Incorporated, a bioengineering company.

Overexpression of native *E. coli* chaperones is known to improve protein expression of several heterologous proteins with varying degrees of success; production of scFv antibody fragment with DnaK/DnaJ/GrpE increased 100-fold; interleukin-1 solubility was increased only with the overexpression of DnaK and DnaJ (Kolaj *et al.*, 2009). Some heterologous proteins such as human thromboxane, human lysozyme or *Agrobacterium tumefaciens* β-glucosidase are often found in inclusion bodies, and co-expression with GroEL/GroES improves their solubility; even solubility of aconitase which cannot fit into the chaperonin cavity is increased by 40% while it's activity only by 1.5-fold (Kolaj *et al.*, 2009). This is further evidence that chaperonin chamber is required for the correct folding of proteins, while the exterior only participates in preventing aggregation (Ellis, 2003). At least 21 proteins from eukaryotes, eubacteria and archaea are substrates for GroE chaperones (Kolaj *et al.*, 2009) which also points towards a promiscuous affinity for proteins.

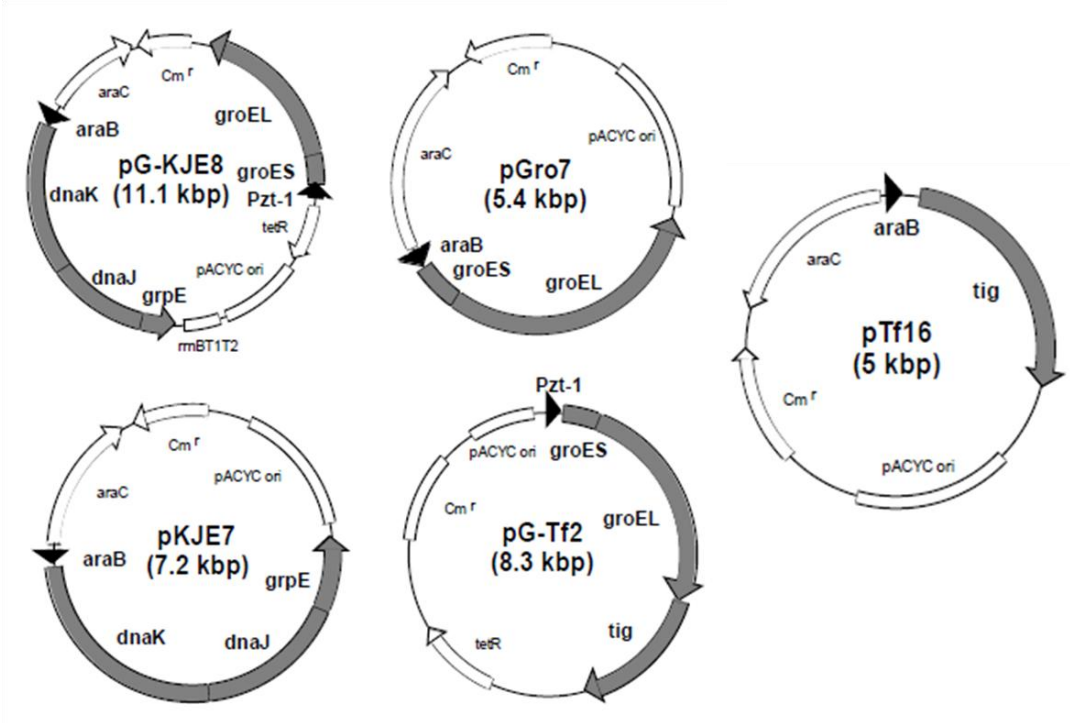

Figure 1.15. Commercial plasmids for the co-expression of *E. coli* molecular chaperones with target proteins from Takara Bio Inc. Adapted from Takara, 2009.

Despite widespread success, this system is not a foolproof solution to increasing solubility; a few proteins fail to be refolded by GroEL/GroES, have poor enzyme activities and cause the overall decrease in viability of the expression host (Kolaj *et al.*, 2009). Expression of the sHsps improved solubility of enhanced GFP (Han *et al.*, 2004), while combined expression of DnaK and ClpB improved solubility in of two aggregation prone species (de Marco *et al.*, 2005); this could be due to the disaggregase mechanism of ClpB (Schlieker *et al.*, 2002) rather than interaction with early nascent strands. Overexpression of TF alone increased solubility of human endostatin, combined expression of TF and GroEL/GroEL increased solubility endostatin, human lysozyme and human oxygen-regulated protein ORP150 (Nishihara *et al.*, 2000).

Solubility and activity are two distinct features of a biologically active protein and increased solubility does not mean a corresponding or equal improvement in a protein's activity (de Marco *et al.*, 2005). Some groups have reported poor or no activity values despite producing highly soluble proteins (Butz *et al.*, 2003). Combinatorial expression of different chaperone systems may improve solubility and protein activity (Nishihara *et al.*, 1998; Schlieker *et al.*, 2002) but evidence as to why this is more effective is speculative and has not been

demonstrated experimentally, as most co-expression studies are aimed at high production of proteins rather than the biochemical properties of the co-expression system (de Marco *et al.*, 2007).

Chaperones are not exclusively used to overproduce soluble proteins *in vivo*; Jhamb *et al.*, (2008) have described the use of *in vitro* refolding of proteins by immobilised chaperones while Bergeron *et al.*, (2009) successfully created a protein-thermosome chimera that could be used as an immobilised biocatalyst, with a longer half-life than that of uncoupled enzymes.

Since molecular chaperones are highly conserved, the effect of co-expressing heterologous chaperones with other proteins in *E. coli* has demonstrated comparable results, with respect to yields of active, soluble protein, to those of endogenous *E. coli* chaperones. The protein-chaperone syatem described earlier used a thermosome from the hyperthermophilic *Methanocaldococcus jannschii*. A small heat-shock protein, CsHsp17.5, from *Castanea sativa* or chestnut, was expressed in *E. coli* and improved thermotolerance at 50°C (Soto *et al.*, 1999). DnaK from gram-positive *Tetragenococcus halophilus* was able to refold the lactate dehydrogenase protein from *Lactococcus lactis*, *in vitro* and increased halotolerance of *E. coli* cells during fermentation (Sugimoto *et al.*, 2003).

However, above 43°C, *T. halophilus* DnaK is unable to complement *E. coli* DnaK due to a 23 residue region that is absent in its ATPase domain; this region is required for DnaJ and GrpE binding in gram-negative bacteria but not in gram-positive species (Sugimoto *et al.*, 2007). Expression of *Plasmodium falciparum* (*Pf*) Hsp70, *Pf*Hsp70, in *E. coli* improved the host's thermotolerance while co-expression of *Pf*Hsp70 with another *P. falciparum* protein, *Pf*GTP cyclohydrolase, showed significant improvement in solubility and activity (Shonhai *et al.*, 2005; Stephens *et al.*, 2011). One unique study carried out by Mahin *et al.*, (2010), demonstrated the ability of heterologously expressed *E. coli* DnaK to suppress thermosensitivity in *Lactococcus bacillus*, so far no other literature describes heterologous expression of *E. coli* chaperones in other organisms.

Chapter 2 . Cloning of a chaperone co-expression vector for heterologous expression in *E. coli* using thermophilic chaperones from *T. thermophilus* and *T. scotoductus*

2.1 Abstract

A set of vectors that were used for chaperone co-expression were constructed. The promoter, P_{BAD}, was amplified alongside its regulator-encoding gene, AraC, by PCR. The KJEA operons, encoding thermophilic DnaK, DnaJ, GrpE and DafA genes, from *Thermus scotoductus* SA-01 and *Thermus thermophilus* HB8 were similarly amplified from genomic DNA extracted from these organisms. Fragments were then subcloned into pGEM®-T easy to give a range of recombinant vectors. The P_{BAD} promoter was later fused to *Ts*KJEA and *Tt*KJEA. The resulting fusion genes, alongside AraC were finally cloned into both pET22 and pET28 to yield four recombinant vectors: p22*Ts*K, p22*Tk*K, p28*Ts*K and p28*Tt*K. The KJEA operon is under the control of the arabinose-inducible P_{BAD} promoter and its regulator protein, encoded by AraC. Transformed *E. coli* BL21 (DE3) carrying the chaperone plasmids were induced with L-arabinose and it was discovered that 5 mg ml^{-1} of L-arabinose is the concentration at which P_{BAD} produces maximum yields of the DnaK chaperone proteins. Expression of *Tt*KJEA from p22*Tt*K was unsuccessful, probably due to a fault in orientation of P_{BAD} relative to AraC; however, successful expression of the *Ts*KJEA was achieved and is comparable to the expression of DnaK proteins encoded by pKJE7, a commercial, DnaK co-expression vector developed by Takara Bio Incorporated. For both *Ts*KJEA and pKJE7, expression is high for DnaK but extremely low for the other operon proteins, DnaJ, GrpE and DafA, as quantified by SDS-PAGE; this does not indicated non-expression but differential expression levels of polycistronic mRNA, which is regulated by transcription terminators and mRNA stability, a common phenomenon in prokaryotic operons.

2.2 Introduction

Molecular chaperones are accessory folding modulators that enable a protein to fold to its native conformation by shielding its hydrophobic residues during the folding process, thereby, preventing non-native association of protein domains that could lead to a misfolded product (Bukau *et al.,* 2006; Dobson, 2003). In prokaryotes, such as *Escherichia coli*, misfolded proteins obtained by overexpression aggregate into insoluble aggregates known as inclusion bodies that are not only toxic to the cell but also amount to losses of soluble, active proteins recovered in research and biotechnological applications (de Groot *et al.*, 2009; Roodveldt *et al.*, 2005).

Although there are methods to resolubilise inclusion bodies into active proteins, the process is time consuming and cost intensive, with limited success at retaining protein activity (de Marco *et al.,* 2005; Singh & Panda, 2005). Molecular chaperones have been exploited extensively in heterologous expression systems as a way of improving the yield and activity of these recombinant proteins *in vivo* (Kolaj *et al.*, 2009).

Common protocols for co-expressing proteins alongside molecular chaperones employ the use of plasmid vectors; in this case, chaperone-encoded plasmids are co-transformed with a plasmid encoding the protein of interest into a host cell and expressed simultaneously (de Marco & de Marco, 2004). In addition to finding compatible plasmids, tight selective pressure must be maintained to ensure their stability within the host cell (Tolia & Joshua-Tor, 2006). Routine use of *E. coli* for the overexpression of a wide variety of recombinant proteins means that several plasmids have been developed commercially to meet this demand and finding compatible plasmids is no longer a great obstacle (Makrides, 1996; Terpe, 2006); however, multiple transformations of a single culture yields few or no transformants (Goldsmith *et al.*, 2007; Tolia & Joshua-Tor, 2006).

A novel approach by Bergeron *et al.,* (2009) and Kyratsous *et al.,* (2009) uses molecular chaperones fused to the protein of interest as a way of increasing the protein solubility and half-life. While this approach might be useful in generating robust enzymes for large scale fermentations, cleavage of the fused chaperone is necessary when a pure protein is required and this extends the purification procedure (Terpe, 2003) with partial success, as the cleavage process might be incomplete (Stevens, 2000).

Most of the co-expression vectors described make use of chaperones derived from *E. coli* (de Marco *et al.*, 2007; Nishihara *et al.*, 1998) although a recent vector developed by Stephens *et al.*, (2011) employs a heterologous Hsp70 chaperone from *Plasmodium falciparum* and has been demonstrated to fold *P. falciparum* related proteins expressed in *E. coli* relatively well. Despite the number of studies regarding heterologous expression in *E. coli* of chaperones from diverse organisms, such as psychrophiles (Robin *et al.*, 2009), thermophilic archaea (Bergeron *et al.*, 2009) and halophiles (Sugimoto *et al.*, 2003), none of them have effectively studied the effect of overexpression of heterologous molecular chaperones on recombinant protein overproduction. In this study, two novel approaches to co-expressing prokaryotic chaperones to improve heterologous expression in *E. coli* will be explored.

Firstly, rather than using *E. coli*-derived chaperones, thermophilic chaperones will be used; this project will study the effect of the DnaK operon from *Thermus thermophilus* and *Thermus scotoductus* on folding recombinant proteins in *E. coli*. These bacteria belong to a group of thermophilic organisms that live in environments of 65°C to 80°C (Beffa *et al.*, 2007) and while *T. thermophilus* is commonly found in compost heaps, *T. scotoductus* has been isolated in deep surfaces such as thermal springs and deep mines (Balkwill *et al.*, 2004). The DnaK chaperone system in both organisms is arranged as an operon in the order DnaK, GrpE, DnaJ and the assembly factor DafA (Motohashi *et al.*, 1996). This operon – KJEA – has a low ATPase activity compared to that of *E. coli* (Groemping *et al.*, 2001; Osipiuk & Joachimiak, 1997). In addition to being adapted to fold proteins at such extreme conditions, the study proposes that the increased substrate affinity makes it an efficient holdase in comparison to the mesophilic DnaK, making it a more efficient chaperone to prevent protein aggregation.

Secondly, the study looks to adapt the double co-expression system described previously. Rather than constructing independent chaperone- and protein-encoded plasmids, the entire expression system will be modified and adapted for a single vector by placing the DnaK operon and protein of interest under two different promoters. The pETDuet-1 vectors by Novagen have used a similar system of putting genes under the same promoter on different loci of one plasmid (Lan *et al.*, 2006; Tolia & Joshua-Tor, 2006) but this project proposes that using two strong, tightly-regulated but unrelated promoters will allow for overproduction and independent fine-tuning of both the chaperone proteins and the protein interest.

2.3 Methods and materials

2.3.1 Bacterial strains, plasmids and oligonucleotide primers

The following tables list all the bacterial strains, plasmids and oligonucleotide primers used for the construction of the chaperone co-expression vectors.

Table 2.1. List of bacterial strains used in this study.

Organism	Strain	Reference
E. coli	*E. coli* XL10 Gold	Stratagene
	E. coli BL21(DE3)	
	E. coli BL21(DE3)/pRARE[a]	
	E. coli BL21(DE3)/pKJE7[b]	
T. thermophilus	*T. thermophilus* HB8	ATCC 27634
T. scotoductus	*T. scotoductus* SA-01	ATCC 700910

[a, b] Strains were developed by Dr. J. Van Marvijk, University of the Free State.

Table 2.2. List of plasmid vectors used in this study.

Plasmid	Description	Source
pGEM®-T easy	Linear vector with 'T' overhangs for subcloning of *Taq* polymerase-amplified DNA.	Promega
pET22	High-copy number plasmid with T7/lac promoter, for high level protein expression and C-terminal His-tag for purification. Confers ampicillin resistance.	Novagen
pET28	High-copy number plasmid with T7/lac promoter, for high level protein expression, as well as N- and C-terminal His-tags for purification. Confers kanamycin resistance.	Novagen
pBAD	High copy plasmid with *E. coli*-derived arabinose promoter and C-terminal His-tag for high level protein expression.	Invitrogen
pRARE	Plasmid for expression of rare tRNAs from the Novagen Rosetta-gami™ B strain. It was isolated and transformed into *E. coli* BL21(DE3)[a].	Novagen

pGEMT-P	pGEM®-T easy plasmid ligated to P_{BAD} amplicon during subcloning after PCR amplification.	This study
pGEMT-A	pGEM®-T easy plasmid ligated to araC amplicon during subcloning after PCR amplification.	This study
pGEMT-*Ts*K	pGEM®-T easy plasmid ligated to *Ts*KJEA amplicon from *T. scotoductus* SA-01.	This study
pGEMT-*Tt*K	pGEM®-T easy plasmid ligated to *Tt*KJEA amplicon from *T. thermophilus* HB8.	This study
pGEMT-P*Ts*K	pGEM®-T easy plasmid ligated to a P_{BAD}-*Ts*KJEA fusion PCR amplicon.	This study
pGEMT-P*Tt*K	pGEM®-T easy plasmid ligated to a P_{BAD}-*Tt*KJEA fusion PCR amplicon.	This study
pET22A	An amplicon of AraC gene is ligated into pET22.	This study
pET28A	An amplicon of AraC gene is ligated into pET28.	This study
p22*Ts*K	Final chaperone plasmid construct containing the full KJEA operon from *T. scotoductus* SA-01, under the control of the P_{BAD} promoter and AraC regulator and a pET22 backbone.	This study
p22TkK	Final chaperone plasmid construct containing the full KJEA operon from *T. thermophilus* HB8, under the control of the P_{BAD} promoter and AraC regulator and a pET22 backbone.	This study
p28*Ts*K	Final chaperone plasmid construct containing the full KJEA operon from *T. scotoductus* SA-01, under the control of the P_{BAD} promoter and AraC regulator and a pET28 backbone.	This study
p28*Tt*K	Final chaperone plasmid construct containing the full KJEA operon from *T. thermophilus* HB8, under the control of the P_{BAD} promoter and AraC regulator and a pET28 backbone.	This study

[a] Strains were developed by Dr. J. Van Marvijk, University of the Free State.

Table 2.3. List of primers used in this study. All primers were obtained from Bioneer or Integrated DNA Technologies (IDT).

Primer	5'→ 3' Sequence	Comments and restriction sites added
araBAD-1F	CGG CAT GCA AAC CAA TTG TCC ATA TTG C	*Sph*I
araBAD-1R	CGT CTA GAT AAT TCC TCC TGT TAG CCC	*Xba*I
araC-1F	CGA GAT CTT TAT GAC AAC TTG ACG GCT ACA TC	*Bgl*II
araC-1R	CGA GAT CTA TGG CTG AAG CGC AAA ATG AT	*Bgl*II
*Ts*KJE-1F	CCT CTA GAA TGC TTG AAA TGA GAG GTG TAG CT	*Xba*I
*Ts*KJE-1R	CAG CAT GCT TAG CGG CGC TCC AGG ATC	*Sph*I
*Tt*KJE-1F	CGT CTA GAA TGG CCA AGG CAG TGG GCA TTG AC	*Xba*I
*Tt*KJE-1R	CAG CAT GCC TAA GTG CGC TCC AGG ATC TC	*Sph*I
Ts/*Tt*KJE-1F	GCC TAC GGC CTG GAC AAG AAG GG	Internal sequencing primer for KJEA.
Ts/*Tt*KJE-2F	GAG ACC AAG GGC GGG GTG AT	Internal sequencing primer for KJEA.
Ts/*Tt*KJE-3F	GAC GCC GAC TAC AAG CCC GC	Internal sequencing primer for KJEA.
Ts/*Tt*KJE-4F	TTC CAG CGG GGC TTC CGC	Internal sequencing primer for KJEA.
Ts/*Tt*KJE-5F	GTC TTC CGC CTC GAG GGC	Internal sequencing primer for KJEA.
T7	GTA ATA CGA CTC ACT ATA	5' – forward sequencing primer in pGEM® -T easy.
Sp6	TAC GAT TTA GGT GAC ACT ATA G	3' – reverse sequencing primer in pGEM® -T easy.

2.3.2 General experimental procedures

All chemicals and reagents used to growth media were analytical grade and supplied by Sigma-Aldrich or Merck.

2.3.2.1 Molecular cloning techniques

Unless otherwise stated, all recombinant DNA techniques used are described by Sambrook *et al.*, (1989).

All PCR reactions were prepared using the Expand Long Template PCR system from Roche Applied Sciences by adding together 1 µl of double stranded DNA template, 1 µl each of 10 mM of the relevant forward and reverse primers, 1 µl of 10mM dNTP mix, 5 µl of 10 X Expand Long Template buffer system 2, and 0.5 µl (1 unit) of Expand Long Template polymerase; the reaction was made up to 50 µl with PCR-grade distilled water. The reactions were carried out in the Applied Biosystems Thermocycler 2720.

All agarose electrophoresis analyses were run on 1% w/v agarose gel (Seakem) dissolved in TAE buffer [0.1 M Tris, 0.05 M EDTA (pH 8.0) and 0.1 mM glacial acetic acid] at a constant voltage of 90V for 30 to 45 min. Gels were analysed using the Biorad gel documentation system. When the agarose gels were used in DNA isolations, the DarkReader Transilluminator from Clare Chemicals was used. For reference, the molecular ladder, GeneRuler^{TM} DNA Ladder Mix, from Fermentas was used.

Ligation of amplicons and plasmid vectors was carried out in a 10 µl reaction mix consisting of 4 µl of insert to be ligated, 3 µl of the linearised, dephosphorylated vector, 1 µl of 10 X T4 ligase buffer, 1 µl of 10 u µl^{-1} T4 DNA ligase and made up to volume with distilled water. The exception was cloning into pGEM$^{®}$-T easy, in which case 3 µl of insert were used with 0.25 µl of the cloning vector.

Dephosphorylation of plasmid was performed in a 30 µl reaction containing 24 µl of linearised, purified plasmid, 3 µl of 5 u µl^{-1} Antarctic Phosphatase (New England Biolabs) and 3 µl of 10 X Antarctic Phosphatase buffer at 37°C, overnight. The enzyme was inactivated at 75°C for 10 min and the reaction was cooled to room temperature before use.

All restriction endonucleases, T4 DNA ligase and their corresponding buffers described here were supplied by Fermentas. Restriction digests were performed in 10 µl reactions containing 3 µl of the plasmid DNA template, 0.5 µl of 10 u µl^{-1} restriction endonuclease, 1 µl of 10 X

endonuclease buffer and made up to volume with distilled water. In the case of double digests, the enzyme units were adjusted as specified by Fermentas and when digesting a template for downstream experiments such as ligations, 6 µl of DNA template was used instead. Results were analysed by agarose gel electrophoresis. Digests were incubated for 3 hours to allow complete digestion at 37°C.

2.3.2.2 Transformation, selection of colonies and preparation of plasmid DNA

Transformation of *E. coli* strains was performed using the protocol described by Chen *et al.*, (1992), as follows: 100 µl of competent cells, according to the method described by Inoue *et al.,* (1990) were incubated with the appropriate volume of transforming DNA on ice for 15 min, followed by heat shocked at 42°C for 1 min. The cells were cold-shocked on ice-water slurry for 2 min. To this, 500 µl of Luria-Bertani (LB) broth (5 g yeast extract, 10 g tryptone, 5 g NaCl per litre of distilled water) was added, followed by incubation of the transformation mix at 37°C, with shaking for 45 min. The culture was centrifuged at 6000 x *g* for 1 min, 450 µl of the broth was discarded while the remaining 150 µl was used to resuspend the cells and plate them on LB agar plates (LB broth with 15g per litre of bacteriological agar) supplemented with the appropriate selective antibiotic.

Enumeration of single colonies was carried out by growing them in 5 ml LB broth containing the appropriate selective antibiotic at 37°C, overnight, in a shaking incubator. For plasmid DNA isolations, 1ml of this culture was used as described below.

Plasmid DNA isolations (minipreps) were carried out to screen for positive clones using the lysis by boiling method adapted from Sambrook *et al.,* (1989), as follows: 1 ml of *E. coli* overnight culture was harvested and centrifuged at 6000 x *g* to pellet the cells; the supernatant was discarded while the pellet was resuspended in 350 µl of STET buffer [8% w/v Sucrose, 5% v/v Triton X-100, 50 mM EDTA, 50 mM Tris (pH 8.0)]. For lysis, 25 µl of lysozyme [10 mg ml^{-1}] was added and the suspension was boiled at 100°C for 44 s. The reaction mix was centrifuged at 13 000 x *g* for 10 min; the resulting pellet was removed with a sterile toothpick and discarded. The supernatant was precipitated with 40 µl of 2.5 M sodium acetate (pH 5.2) and 420 µl isopropanol at -20°C for 20 min. The mixture was centrifuged at 4°C, 13 000 x *g*, for 20 min, the supernatant was aspirated and the pellet was washed in 70% ethanol. This mix was centrifuged for 10 min at 4°C, 13 000 x *g*, followed by aspiration of the supernatant. The pellet was dried and resuspended in 40 µl of TE and RNAse [10 µg ml^{-1}]. For the purposes of enumerating plasmids from positive clones,

however, minipreps were performed using the Bioflux Plasmid DNA Extraction Kit. All DNA concentrations were measured using the NanoDrop ND-1000 Spectrophotometer from NanoDrop Technologies.

2.3.2.3 Sequencing reactions

Sequencing of cloned inserts was performed while ligated into pGEM®-T easy and again when they were cloned into the pET vector backbones. Reactions were set up using the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) as follows: 1 µl DNA template, 1 µl of 3.2 pM primer, 2 µl of 5 X sequencing buffer, 0.5 µl premix and 6.5 µl PCR-grade distilled water. Thermal cycling was carried out for 25 cycles at 94°C for 10 s, 50°C for 5 s and 60°C for 4 min.

The EDTA-ethanol precipitation method was used to prepare samples for sequencing as follows: samples were mixed with 10 µl of distilled water, 5 µl of 125 mM EDTA (pH 8.0) and 60 µl of 100% ethanol. Samples were vortexed and incubated at room temperature for 15 min, then centrifuged at 4°C, 13 000 x *g*, for 20 min. The supernatant was aspirated, the pellet was washed in 120 µl of 70% ethanol and centrifuged and 4°C, 13 000 x *g*, for 10 min. Ethanol was aspirated and the samples were dried and then sequenced using the Hitachi 3130x*I* Genetic Analyser, also from Applied Biosystems. Sequence chromatograms were analysed with Geneious v4.8.2 (Biomatters) software and aligned to the original sequences using ClustalW software hosted by the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

2.3.2.4 Protein expression analyses

Samples were prepared for SDS-PAGE by adding 5 µl of the sample to an equal volume of 2 X Laemmli sample buffer (Biorad). Gels for SDS-PAGE were cast with 10% bis/acrylamide according to the procedure described by Laemmli *et al.*, (1970), using the 0.75 mm Mini-PROTEAN® Tetra system (Biorad). Samples were then electrophoresed alongside PageRuler™ Prestained Protein ladder (Fermentas) in the Mini-PROTEAN® Tetra Cell at 100V until the ladder had resolved. Gels were developed by coomassie staining according to the method described by Fairbanks *et al.*, (1971) and evaluated.

2.3.4 Extraction of genomic DNA from *T. scotoductus* and *T. thermophilus*

T. scotoductus SA-01 and *T. thermophilus* HB8 were grown in TYG [5 g Tryptone, 3 g Yeast extract and 1 g D-Glucose per litre of dH_2O] overnight at 65°C. Genomic DNA (gDNA) was isolated from these cultures according to the method described by Labuschagne & Albertyn, (2007), as follows: an overnight culture of either *T. scotoductus* SA-01 or *T. thermophilus* HB8 (1ml) was aliquoted into a microcentrifuge tube and centrifuged at 13 000 x *g* for 1min. The supernatant was aspirated and 500 µl of DNA isolation buffer [100 mM Tris-HCl (pH 8.0); 50 mM EDTA; 1% SDS], as well as 200 µl of glass beads were added to the pelleted cells and vortexed for 4 min. The suspension was cooled on ice, after which 250 µl of 7 M ammonium acetate was added, followed by vortexing. The suspension was incubated at 65°C for 5 min and cooled on ice for 5 min. After adding 500 µl of chloroform, vortexing and centrifuging at 4°C, 13 000 x *g*, for 5 min, the supernatant was retained and precipitated with 1 X isopropanol for 5 min at room temperature, followed by centrifuging at 4°C, 13 000 x *g*, for 5 min. The resulting pellet was washed in 70% ethanol and centrifuged at 4°C, 13 000 x *g*. The supernatant was aspirated and the pellet was dried at 37°C for 2 hrs, followed with resuspension in TE [10 mM Tris-HCl (pH 7.5), 1 mM EDTA] and RNAase [10 µg ml^{-1}].

2.3.5 PCR amplification of P_{BAD}, AraC, *Ts*KJEA and *Tt*KJEA

The arabinose promoter, P_{BAD}, was amplified from the pBAD plasmid (Invitrogen) for 35 cycles [94°C for 30 s, 55°C for 30 s, 72°C for 45 s] with the Expand Long Template PCR system using primers araBAD-1F and araBAD-1R. The gene which codes for a regulator protein for this promoter, AraC, was amplified from the same plasmid [94°C for 30 s, 60°C for 30 s, 72°C for 1 min] using araC-1F and araC-1R.

TsKJEA and TtKJEA were amplified from gDNA extracted from *T. scotoductus* SA-01 and *T. thermophilus* HB8, respectively, [94°C for 30 s, 60°C for 30 s, 72°C for 2 min] for 35 cycles using the corresponding primer pairs *Ts*KJE-1F/ *Ts*KJE-1R and *Tt*KJE-1F/ *Tt*KJE-1R.

2.3.6 Subcloning of PCR fragments into pGEM®-T easy

Amplicons for subcloning were electrophoresed, excised from the agarose gel and purified with the Bioflux Gel Extraction Kit according to the manufacturer's instructions. Purified fragments were ligated into pGEM®-T easy using T4 DNA ligase overnight at 4°C. The

reaction mix was used to transform *E. coli* XL10 Gold and plated onto LB_{AIX} agar plates (60 μg ml^{-1} ampicillin, 1 mM IPTG and 40 μg ml^{-1} X-gal).

White colonies were picked and inoculated into LB_{Amp} [60μg ml^{-1}], grown overnight at 37°C and harvested. Plasmid DNA minipreps were carried out using lysis by boiling method described by Sambrook *et al.*, (1989) and resulting DNA was used to screen for positive clones using restriction digests. To screen for clones possessing pGEMT-P, a double digest was performed with *Sph*I and *Xba*I to release the P_{BAD} fragment, while pGEMT-A was screened by digesting the plasmid with *Bgl*II to release the AraC gene. The two KJEA vectors, pGEMT-*Ts*K and pGEMT-*Tt*K were screened by digesting the extracted plasmids with *Xba*I and *Sph*I.

2.3.6.1 Constructing a P_{BAD}-KJEA fusion

In order to fuse the arabinose promoter to the KJEA operon, pGEMT-P, pGEMT-*Ts*KJEA and pGEMT-*Tt*KJEA were digested with *Sph*I and *Xba*I to release the P_{BAD}, *Ts*KJEA and *Tt*KJEA fragments, respectively. The digests were electrophoresed on agarose, excised and purified from the gel. Ligation of P_{BAD} to *Ts*KJEA and P_{BAD} to *Tt*KJEA was performed using T4 DNA ligase at 4°C, overnight. The ligation mix served as the template for PCR amplification to obtain a P_{BAD}-KJEA fusion amplicon by using the primer-pairs araBAD-1F/*Ts*KJEA-1R and araBAD-1F/*Tt*KJEA-1R. Successful amplification would only occur with the correct orientation of the promoter fused to the KJEA operon via the *Xba*I site. Thermal cycling was carried out [94°C for 30 s, 57°C for 30 s, 72°C for 2.5 min] for 35 cycles using Expand Long Template polymerase.

Amplicons of P_{BAD}-*Ts*KJEA and P_{BAD}-*Tt*KJEA were ligated into pGEM®-T easy at 4°C with T4 DNA ligase, overnight. The ligation mix was used to transform competent *E. coli* XL10 Gold cell, which were the plated on LB_{AIX} agar plates. White colonies were inoculated into LB_{Amp} broth and grown at 37°C, overnight. Plasmid DNA minipreps were carried out and screening for positive clones containing the recombinant pGEMT-P*Ts*K and pGEMT-P*Tt*K was performed by digesting the minipreps with *Sph*I.

2.3.7 Construction of pET22A and pET28A

The AraC gene was digested from pGEMT-A by *Bgl*II, separated on agarose gel and purified. The pET22 and pET28 plasmids were also digested with *Bgl*II to linearise them,

dephosphorylated and run on agarose. Linearised vector was cut from the gel and purified. The AraC gene was ligated into the *Bgl*II sites of the linearised vectors with T4 DNA ligase at 4°C, overnight and the resulting mix was used to transform competent *E. coli* XL10 Gold cells that were plated onto LB_{Amp} or LB_{Kan} [30μg ml^{-1} kanamycin] agar.

Colonies were inoculated into LB_{Amp} or LB_{Kan} broth and grown overnight at 37°C. Cells were harvested for plasmid DNA minipreps which were digested with *Bgl*II to screen for positive clones bearing the AraC insert.

2.3.8 Construction of p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K

To make the final plasmid constructs, pET22A and pET28A were linearised with *Sph*I and dephosphorylated as described for previous plasmids. The promoter-fusion fragments, P_{BAD}-*Ts*KJEA and P_{BAD}-*Tt*KJEA, were digested from pGEMT-P*Ts*K and pGEMT-P*Tt*K by with *Sph*I. The fusion genes were purified from agarose gels; each fusion was then ligated into the *Sph*I sites of linearised pET22A and pET28A. The ligation mix was used to transform competent *E. coli* XL10 Gold cells. Transformants were plated onto LB_{Amp} or LB_{Kan} agar plates. Colonies that grew were used to inoculate LB_{Amp} or LB_{Kan} broth and grown overnight at 37°C. Cells were harvested for plasmid DNA minipreps and screened for positive clones by performing restriction digests with *Sph*I.

2.3.9 Induction of the KJEA operon with L-arabinose

Induction of P_{BAD} was carried out using L-arabinose to test for the expression of *Ts*DnaK and *Tt*DnaK chaperone proteins in *E. coli* and to determine the optimum concentration of L-arabinose required to produce maximum expression of the KJEA operon.

Competent *E. coli* BL2 (DE3) was transformed with p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K and plated on either LB_{Amp} or LB_{Kan}. Single colonies from these plates were inoculated into LB_{Amp} or LB_{Kan} and grown overnight. Of this, 100 μl of each transformant was used to inoculate six Erlenmeyer flasks containing 30ml of LB_{Amp} or LB_{Kan}. The cells were grown to an OD_{600} of ~0.8 – 1 at 37°C, after which induction was initiated by adding the following concentrations of L-arabinose: [0.0005 mg, 0.005 mg, 0.05 mg, 0.5 mg and 5 mg] ml^{-1}. Expression was conducted for 6hrs at 30°C and 2ml samples were collected from each flask, at the end of the run. As a negative control, *E. coli* BL21 (DE3) was transformed with empty

pET22 and pET28 plasmids, while a commercial plasmid from Takara, pKJE7 that contains *E. coli*-derived (*Ec*) DnaK, DnaJ and GrpE was used to compare chaperone expression levels.

The harvested cells were centrifuged at 6000 x *g* to pellet them; the supernatant was discarded and the cells were resuspended in 500 µl 50 mM EDTA supplemented with cOmplete, Mini, EDTA-free Protease Inhibitor tablets (Roche Applied Science). Cells were lysed using the Bandelin Sonoplus sonicator at 55% power for 30 s and placed on ice. The lysate was centrifuged at 4°C, 13 000 x *g*, for 10 min to pellet cell debris and insoluble protein fractions. The supernatant, or soluble fraction, was transferred to clean microcentrifuge tubes and analysed by SDS-PAGE.

The recombinant KJEA plasmids were also transformed into *E. coli* BL21 (DE3)/pRARE strains to improve expression of DnaK proteins by supplementing *E. coli* with rare codons. The cells were plated onto $LB_{Cm/\ Amp}$ or $LB_{Cm/\ Kan}$, with chloramphenicol (Cm) at a concentration of 34 µg ml^{-1}. Expression was carried out as described above.

2.4 Results and discussion

2.4.1 Extraction of genomic DNA from *T. scotoductus* and *T. thermophilus*

Genomic DNA was extracted from *T. scotoductus* SA-01 and *T. thermophilus* HB8 (Fig. 2.1). This served as the source for KJEA-encoding genes, DnaK chaperone system in these organisms, which would be used to construct the chaperone co-expression plasmids proposed in this study.

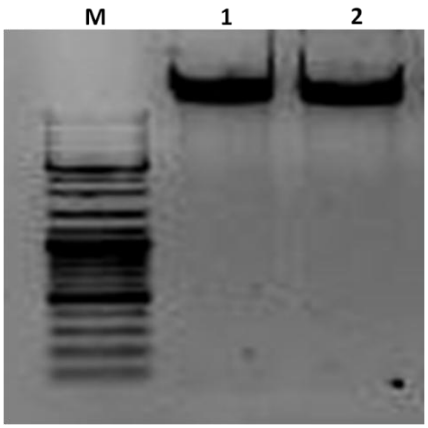

Figure 2.1. Extraction of total gDNA from *T. scotoductus* (lane 1) and *T. thermophilus* (lane 2). Lane M represents the 10kB molecular weight marker, GeneRuler ™ DNA Ladder Mix, which was used throughout this study for agarose gels.

2.4.2 PCR amplification of P_{BAD}, AraC, *Ts*KJEA and *Tt*KJEA

The *E. coli*-derived arabinose promoter, P_{BAD} and the gene encoding the regulator protein, AraC, were obtained by PCR from pBAD. The expected sizes for the promoter, 324bp, and AraC, 891bp, are shown in Fig. 2.2. Gene clusters encoding the four KJEA operon proteins, DnaK, DnaJ, GrpE and DafA, were similarly isolated from gDNA extracted from *T. scotoductus* SA-01 and *T. thermophilus* HB8 to give products of 3576bp and 3518bp, respectively (Fig. 2.2). All amplicons were subsequently excised from the agarose gel and purified for further use in subcloning of the fragments into pGEM®-T easy.

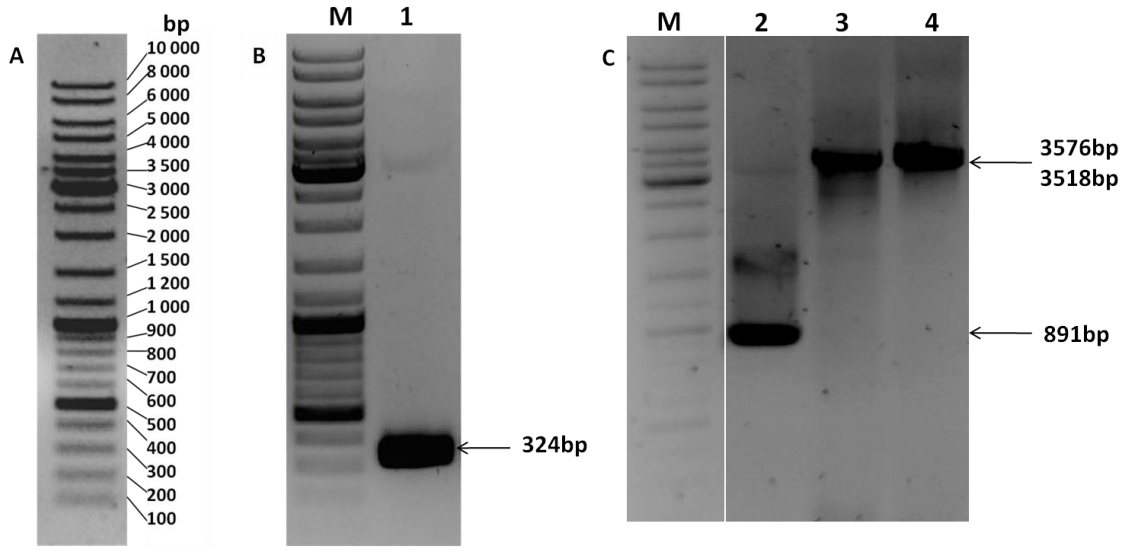

Figure 2.2. Amplification of P_{BAD} (B), AraC, *Ts*KJEA and *Tt*KJEA (C) by PCR. The promoter and its regulator protein-encoding gene (lanes 1 and 2, respectively) were amplified from the same plasmid, pBAD. The KJEA operon genes (lanes 3 and 4) were amplified from *T. scotoductus* and *T. thermophilus* gDNA, respectively. The molecular weight marker, M, is shown in A.

2.4.3 Subcloning of PCR fragments into pGEM®-T easy

PCR fragments of P_{BAD}, AraC, *Ts*KJEA and *Tt*KJEA were gel-purified and ligated into pGEM®-T easy vector. All clones were verified for correct insertion using restriction digests. Screening for positive clones of pGEMT-P, pGEMT-*Ts*K and pGEMT-*Tt*K was carried out by restriction analysis using *Sph*I and *Xba*I (Fig. 2.3). *Bgl*II restriction analysis was used to screen for recombinant pGEMT-A, which bears the AraC gene (Fig. 2.3). Successful clones were digested with the same enzymes to liberate these fragments from their plasmid backbone, for further downstream cloning of chaperone co-expression vectors.

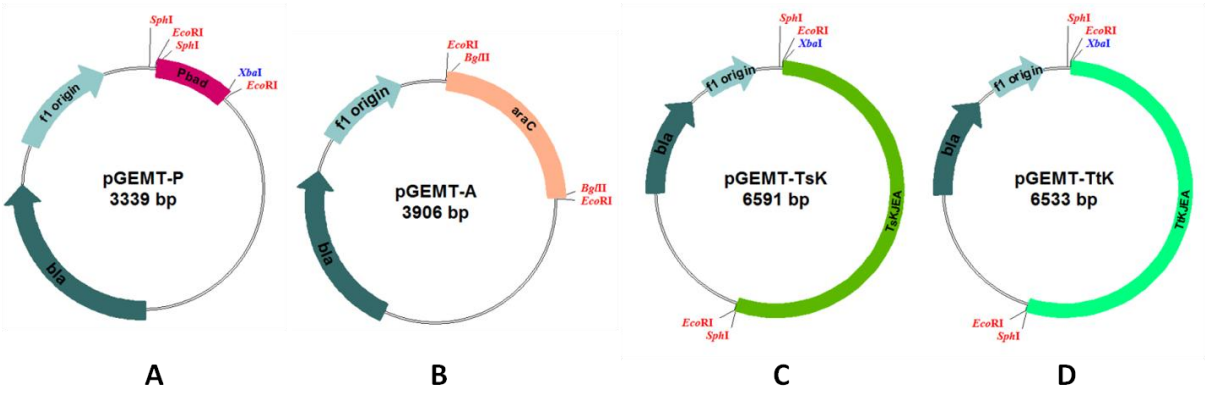

Figure 2.3. Diagrammatic representation of recombinant vectors constructed by ligating P_{BAD}, AraC, *Ts*KJEA and *Tt*KJEA into the 3015bp GEM®-T easy vector. P_{BAD} and the two DnaK operons are shown with the *Sph*I/ *Xba*I restriction sites flanking their 5' and 3' ends (A, C and D) while AraC is inserted with *Bgl*II.

2.4.3.1 Constructing a P_{BAD}-KJEA fusion

The gel-purified fragment of P_{BAD} was ligated at the 3' end to the 5' ends of either *Ts*KJEA or *Tt*KJEA via the *Xba*I restriction site to produce a fusion gene of the promoter and chaperone. To screen for the correct fusion gene possessing this link, PCR was carried out using the promoter's forward primer, araBAD-1F and the operons' reverse primers, *Ts*KJE-1R or *Tt*KJE-1R. The use of this primer pair ensures that only fusions linked in the following orientation: 5'-*Sph*I – P_{BAD} – *Xba*I –KJEA – *Sph*I-3' are amplified (Fig 2.4).

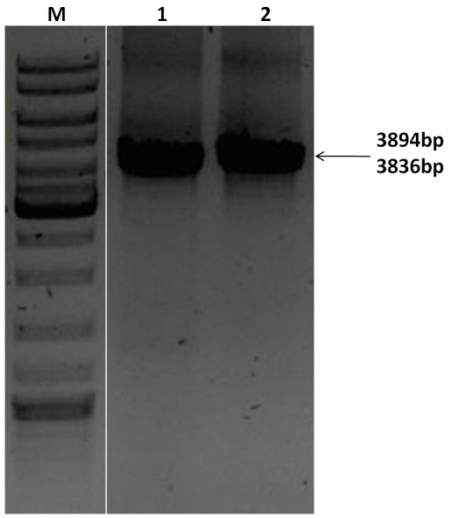

Figure 2.4. P_{BAD}-KJEA fusion genes obtained by PCR. Lane 1 shows P_{BAD}-*Ts*KJEA fusion gene (3894bp), while lane 2 shows the P_{BAD}-*Tt*KJEA fusion gene (3836bp). M is the molecular weight marker.

Amplicons of the two fusion genes were gel-purified and ligated into pGEM®-T easy. To screen for positive clones, possessing recombinant plasmids pGEMT-PTsK and pGEMT-PTtK, restriction digests with *Sph*I, were performed (Fig 2.5). The resulting fusion fragments were later purified from the gel and used in the final construction of the chaperone vectors.

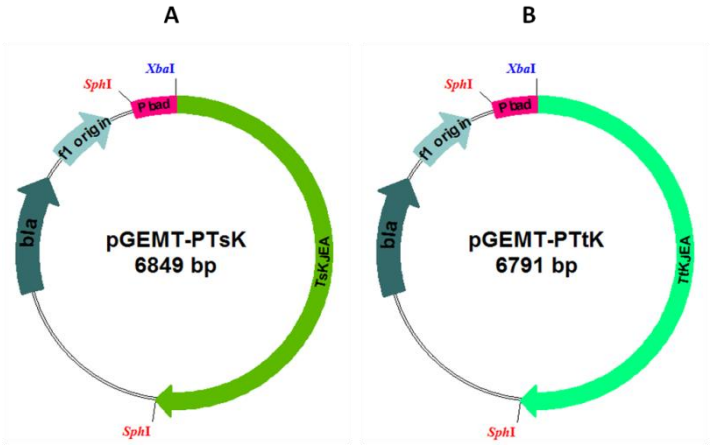

Figure 2.5. Diagrammatic representation of P_{BAD} fused to *Ts*KJEA operon (A) and *Tt*KJEA operon (B) in pGEM®-T easy. The vectors shown here were confirmed by digesting plasmid DNA with *Sph*I.

2.4.4 Construction of pET22A and pET28A

The regulator-encoding gene, AraC, was digested from pGEMT-A with *Bgl*II and gel-purified. The pET22 and pET28 vectors were similarly linearised with *Bgl*II and dephosphorylated. The purified AraC gene was cloned into the complementary *Bgl*II sites of pET22 and pET28 to yield recombinant pET22A and pET28A (Fig. 2.6). Screening for positive clones was performed y restriction analysis with *Bgl*II (Fig 2.7)

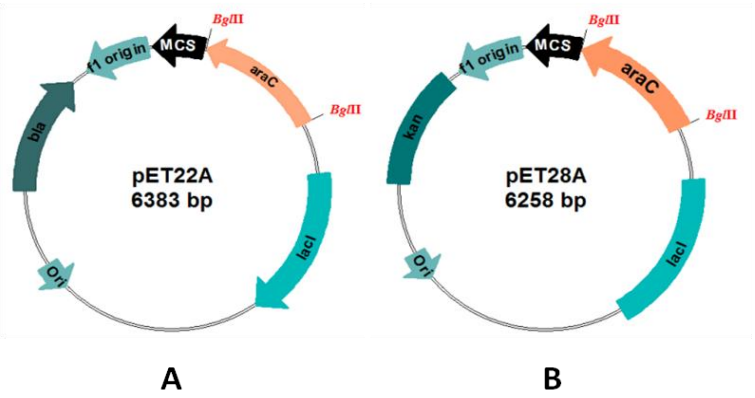

Figure 2.6. Diagrammatic representation of the recombinant pET22A and pET28A vectors, cloned by ligating *Bgl*II-digested AraC into pET22 and pET28.

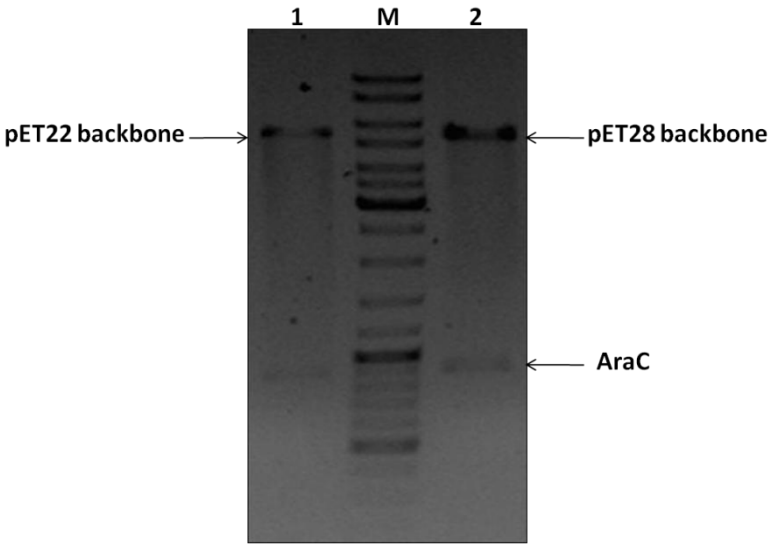

Figure 2.7. Restriction digest of pET22A (lane 1) and pET28A (lane 2) with *Bgl*II.

2.4.5 Construction of p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K

In the final step of designing the chaperone co-expression vectors, the promoter-KJE fusion genes, P_{BAD}-*Ts*KJEA and P_{BAD}-*Tt*KJEA, were digested from pGEMT-P*Ts*K and pGEMT-P*Tt*K with *Sph*I and ligated into *Sph*I-linearised pET22A and pET28A. The resulting recombinant vectors are the DnaK-encoding chaperones p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K, shown in Fig 2.8, which will be used to express heterologous proteins in *E. coli*. Screening for insertion into the *Sph*I sites was confirmed by restriction digests (Fig 2.9A) and orientation was checked by restriction analysis with *Bgl*II (Fig 2.9B).

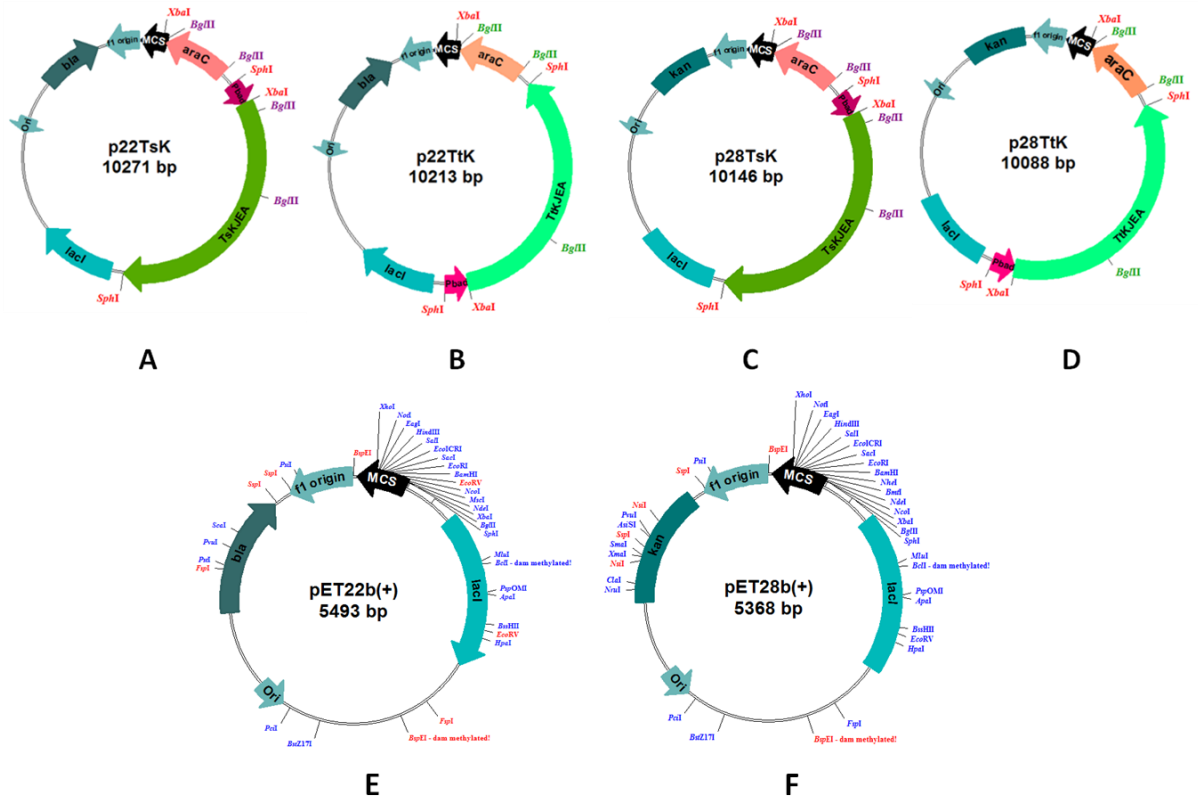

Figure 2.8. DnaK chaperone co-expression vectors. The recombinant vectors, shown from A to D, are derived from pET22 and pET28 plasmids. E and F will be used as the negative controls for chaperone expression.

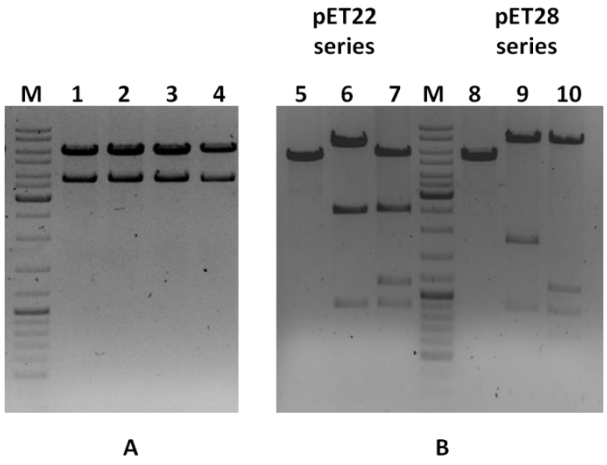

Figure 2.9. Digestion of the recombinant chaperone plasmids with *Sph*I, A. Lanes 1 to 4 show p22TtK, p22TsK, p28TtK and p28TsK, respectively. Digestion of these vectors with *Bgl*II gives the profile shown in B, where lanes 6 and 7 are p22TtK and p22tsK, respectively while lanes 9 and 10 are p28TtK and p28TsK, respectively. The negative control vectors pET22 and pET28, which contain no chaperones, are shown in lanes 5 and 8, respectively.

Successful construction of plasmid vectors encoding the *T. scotoductus* and *T. thermophilus* DnaK system was achieved through a series of molecular cloning procedures. An important feature of these vectors is the presence of an entire operon of the chaperone and its co-chaperone proteins, as opposed to cloning only one component of the chaperone system. While it has been possible to improve heterologous expression by co-expressing individual chaperones, the lack of accessory chaperone proteins appears to cause instability in the plasmid and it is quickly lost from the host (Makrides, 1996; Sugimoto *et al.*, 2008). This would mean that proteins in the chaperone system are adapted to work closely with each other to coordinate folding of substrate proteins, and are unsuited to an independent metabolic state.

The use of a strong promoter such as P_{BAD} will ensure high level of transcription in the DnaK operon which should increase the concentration of proteins translated from it (Guzman *et al.*, 1995). This promoter is also tightly regulated by its regulator, AraC, which acts as an activator upon induction with L-arabinose. Having an inducible promoter enables controlled expression of the DnaK genes and fine-tuning of inducer concentrations to optimise expression levels of the operon (Guzman *et al.*, 1995; Smith & Schleif, 1978).

Cloning of KJEA genes into a non-coding region of the plasmid two pET vectors leaves the multiple cloning sites (MCS) available to clone a range of genes that might be of interest, under the control of the T7/lac promoter. This is also a strong promoter that will ensure high level of expression of the gene under its control (Makrides, 1996; Stevens, 2000). The presence of two tightly regulated promoters allows expression of either protein to be tightly controlled (de Marco & de Marco, 2004; Nishihara *et al.*, 1998). Other than the addition of molecular chaperones, no components were removed from the pET vectors and they retain the properties that make them popular vectors for protein expression in *E. coli*, such as their high copy number, His-tags for purification – two, in the case of the pET28 series – and a peptide leader sequence for extra-cellular transport of the expressed protein in the pET22 series.

2.4.6 Evaluation of chaperone coexpression plasmids by induction with arabinose

The four vectors, p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K, were transformed into *E. coli* BL21 (DE3) cells. To test for the expression of the KJEA operon, varying concentrations of L-arabinose were used to induce the P_{BAD} promoter and expression profiles were analysed by SDS-PAGE, (Fig. 2.10). The pKJE7 vector from Takara Bio Incorporated, encoding *E. coli* DnaK, DnaJ and GrpE were also tested for chaperone expression levels using L-arabinose.

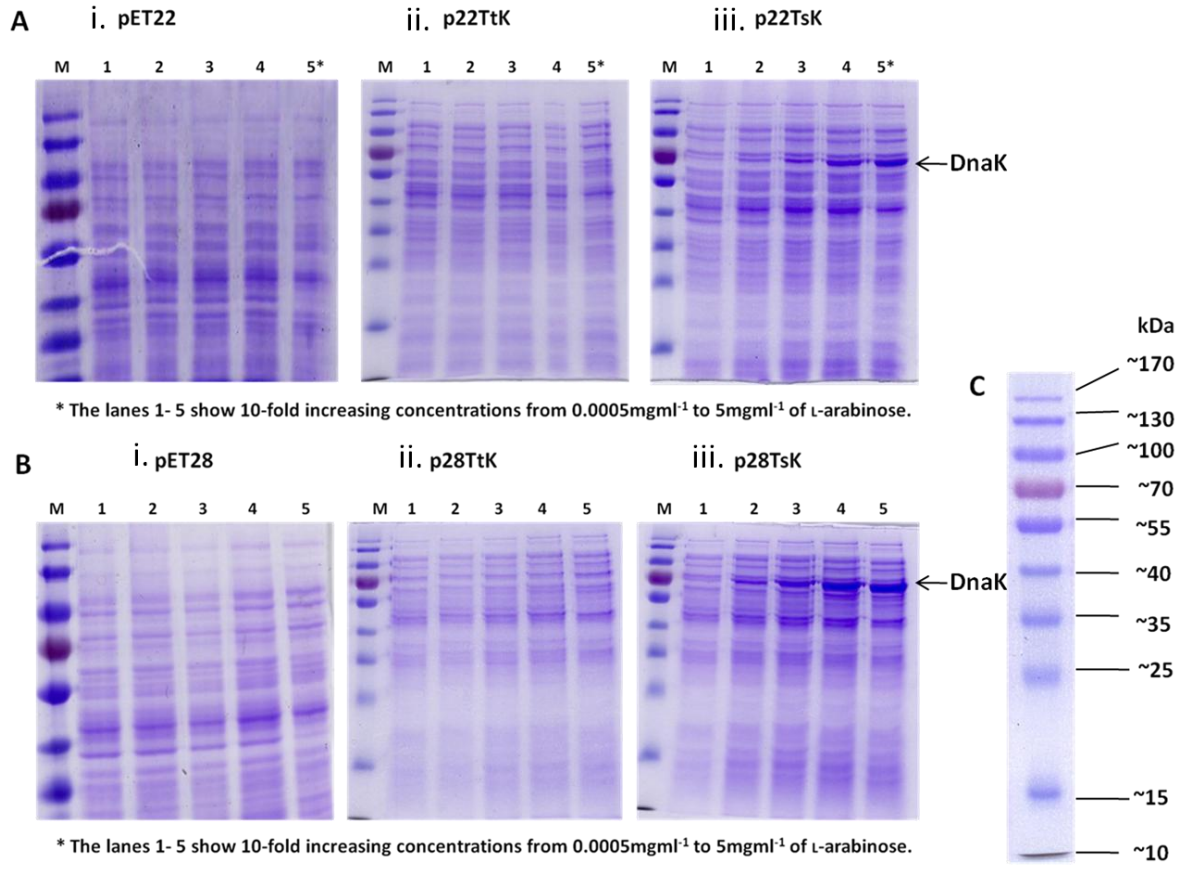

Figure 2.10. Induction of the P_{BAD} promoter in chaperone co-expression vectors by different concentrations of L-arabinose after 6 hours. The pET22-derived series of plasmids is shown in A, while B shows the pET28-derived vectors. Expression of the ~66kDa DnaK chaperone is visible in both p22*Ts*K and p28*Ts*K but not in p22*Tt*K or p28*Tt*K plasmids. The negative controls, pET22 and pET28 show no DnaK expression. M, the molecular marker used throughout the study in all SDS-PAGE analyses is shown in C. The co-chaperone proteins DnaJ, GrpE and DafA could not be visualised on SDS-PAGE gel as they may be present in low concentrations. Cells grown for 24 hours gave the same profiles as those grown for 6 hours.

Cells expressing no heterologous chaperones, i.e. the negative controls bearing pET22 and pET28 vectors, show no apparent increase in DnaK or co-chaperone concentrations. The expression profile of cells transformed by p22*Tt*K and p28*Tt*K also shows that no chaperones could be expressed from this vector. The reason for this is not entirely clear; all inserts used to clone the chaperone plasmids were verified by sequencing and found to contain no mismatches compared to the original sequences. Initially, it was thought to be linked to differences in codon usage between *E. coli* and thermophilic bacteria (Kim & Lee, 2008) despite the fact that none of the current literature states that as an obstacle to expressing heterologous chaperones. To eliminate this as a cause for non-expression, *E. coli* Bl21 (DE3) bearing the pRARE plasmid were transformed with chaperone coexpression plasmids. The pRARE plasmid expresses rare codons in *E. coli*, constitutively. However, Fig 2.11 shows that after inducing the promoter with 5 mg ml^{-1} of L-arabinose, the expression profile remains unchanged.

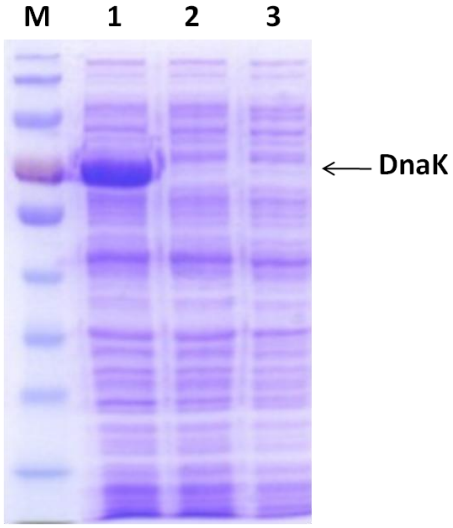

Figure 2.11. Expression of *T. scotoductus* and *T. thermophilus* DnaK operon proteins with the aid of rare codons. In this case, only cells transformed with p22*Ts*K (lane 1), p22*Tt*K (lane 2) and pET22 (lane 3) are shown.

It is likely that the problem arises from the promoter-regulator interactions of P_{BAD} and AraC. Digestion of the plasmid with *Bgl*II reveals that P_{BAD}-*Tt*KJEA is inverted, i.e. its orientation in relation to AraC is reversed, relocating P_{BAD} ~3.5kb away from its regulator (Fig 2.8B and D). Transcription of AraC is initiated at the AraC promoter (P_C), which lies in the 324bp region encoding P_{BAD}; in addition, AraC acts as an activator in the presence of L-arabinose and binds to P_{BAD} to initiate transcription of genes downstream of this promoter (Smith &

Schleif, 1978). With the promoter so far from the AraC gene in p22*Tt*K and p28*Tt*K, it is possible that RNA polymerase was unable to even begin transcription of AraC, although it is still able to bind to P_C (Cass & Wilcox, 1988). As a result, AraC protein was not translated, or it was present at physiologically insignificant levels such that, it was unable to initiate transcription from P_{BAD} and produce significant concentrations of *Tt*KJEA. This can be tested by screening for a clone with the normal orientation and carrying out the same induction studies to see whether this causes a change in expression of *Tt*KJEA.

In contrast, L-arabinose was able to induce expression of the operon from P_{BAD} (Fig.2.10). Maximum expression of the operon was achieved with 5 mg ml^{-1} of L-arabinose, which is consistent with literature (Nishihara *et al.*, 1998; de Marco & de Marco, 2004; Haacke *et al.*, 2009), but already, it is evident that P_{BAD} can be induced using arabinose at concentrations as low as 0.005 mg ml^{-1} (Fig 2.10B.III), further evidence that this is a strong promoter.

Only the ~ 66kDa *Ts*DnaK protein can be seen distinctly on the SDS-PAGE gel (Fig. 2.10A.III and B.III). The other three proteins, *Ts*DnaJ (~ 31kDa), *Ts*GrpE (~20kDa) and *Ts*DafA (~8kDa) cannot be detected on these gels. The main reason for this might be related to the manner in which polycistronic mRNA is processed within the prokaryotic cell. During transcription, the entire operon is transcribed but mRNA is progressively degraded from the 3' end; this causes an increase in the concentration of mRNA for the gene closest to the promoter while decreasing that of genes further downstream (Byrne *et al.*, 2008; Klug, 1993). The stability of mRNA is often influenced by the presence of transcription terminator sequences that form loops to delay degradation; consequently, the concentration of translated protein from polycistronic mRNA is, usually, asymmetrical (Jana & Deb, 2005).

The regulation of relative gene expression by mRNA stability appears to be of physiological significance. Genes whose products are required at high concentrations are closest to the operon (Shimamoto *et al.*, 1994). Sugimoto *et al.*, (2008) demonstrated that overexpression of GrpE alone promotes abnormal cell division and decreases activity of the DnaK system in *E. coli*; it is possible to infer that this is the case in *T. scotoductus* and *T. thermophilus* since their chaperones have conserved roles within these organisms. Furthermore, Osipiuk & Joachimiak, (1997) described a terminator sequence found between *T. thermophilus* DnaK and GrpE which accounts for higher concentrations of DnaK in relation to the other operon proteins and related this to literature which states that *E. coli* DnaK is often expressed in 10-fold higher concentrations than DnaJ. All this explains why the three co-chaperone proteins

cannot be visualised on SDS-PAGE gels; however, a sensitive method such as real-time PCR may be used to provide the relative concentrations of mRNA transcripts from these proteins (Livak & Schmittgen, 2001).

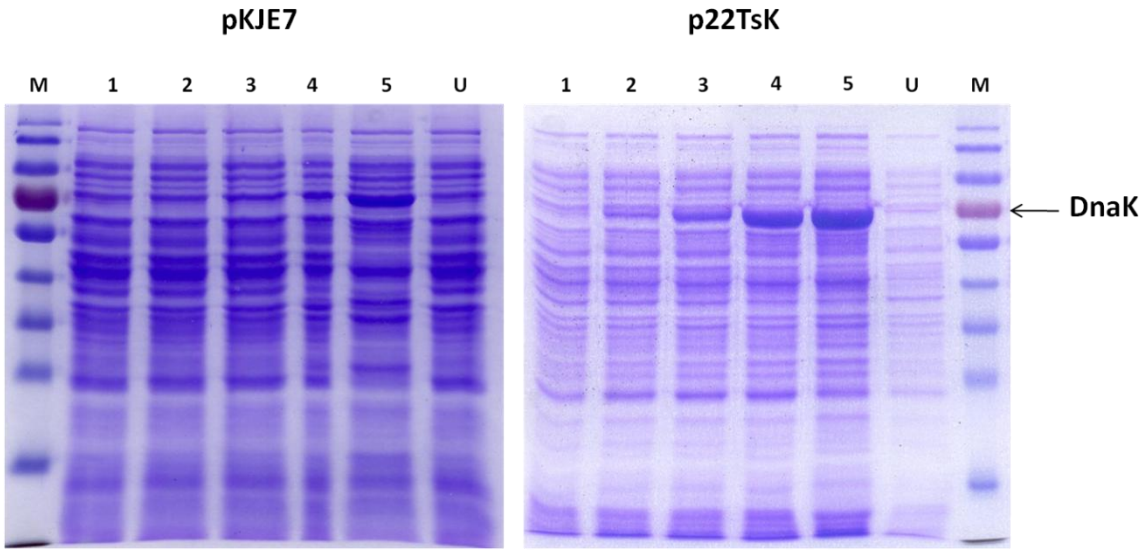

Figure 2.12. Comparison of DnaK expression from pKJE7 and the chaperone vector p22TsK. The lanes 1- 5 show 10-fold increasing concentrations from 0.0005 mg ml^{-1} to 5 mg ml^{-1} of L-arabinose. The pKJE7 plasmid encodes *E. coli* DnaK chaperone proteins while p22*Ts*K encodes *T. scotoductus* DnaK chaperone proteins. In both vectors, only the DnaK chaperone is visible. U is the uninduced control, with no added L-arabinose and M is the protein ladder.

In comparison, expression profiles of cells expressing pKJE7 chaperone proteins appear to be relatively similar to that of *Ts*KJEA (Fig 2.12). And just as with the chaperone thermophilic system, *Ec*DnaJ and *Ec*GrpE cannot be seen on SDS-PAGE gel. A similar regulatory mechanism of the *E. coli* DnaK operon could also be occurring; the *E. coli* DnaK operon is arranged as DnaK, DnaJ and GrpE and a terminator sequence similar to the one found between *Tt*DnaK and *Tt*GrpE has been found between *Ec*DnaK and *Ec*DnaJ (Osipiuk & Joachimiak, 1997; Yochem *et al.*, 1978).

2.4.7 Concluding remarks

Chaperone co-expression plasmids were constructed using *T. scotoductus* and *T. thermophilus* DnaK operon genes. The ability to express the corresponding proteins at moderate temperatures was demonstrated by placing them under the control of the arabinose promoter, P_{BAD}. However, this was only examined with SDS-PAGE. Quantitative analyses to

determine the concentration of individual cannot be determined as chaperone proteins lack His-tags that would enable their purification from the cell extract for subsequent determination of protein content. For this study, this is no setback as downstream expression studies will be carried out under similar conditions such that concentrations of chaperones remain relatively similar for all vectors.

Chapter 3 . Evaluation of heterologously expressed *Thermus thermophilus* DNA polymerase in *Escherichia coli* using the thermophilic DnaK chaperone system from *Thermus scotoductus* and *Thermus thermophilus*.

3.1 Abstract

Heterologous expression of *T. thermophilus* DNA polymerase I, *Tth* pol and sGFP –coupled *Tth* pol was carried out in *E. coli* using *T. scotoductus* DnaK chaperones to aid protein folding. Expression was carried out from a single vector encoding both the molecular chaperones and the polymerases. A commercial plasmid, pKJE7, encoding the *E. coli* DnaK system, was also used to compare the effect of mesophilic chaperones on protein folding. Expression was successfully accomplished for all vectors and purification of uncoupled *Tth* pol was achieved even from the negative controls, which expressed no heterologous chaperones. However, sGFP-coupled *Tth* pol was only purified from strains co-expressing *Ts*DnaK chaperone. Activity assays to test processivity, thermostability and fidelity of recombinant *Tth* pol were carried out with PCR and commercial *Taq* polymerase was used for comparative purposes. Results show that recombinant *Tth* pol was able to amplify 771bp sGFP DNA with high fidelity and was comparable to the commercial *Taq* pol. Recombinant *Tth* pol expressed with the commercial chaperone vector yielded less product than that of the thermophilic chaperone although, it had the same fidelity in amplifying sGFP. Extended incubation at 95°C rendered the polymerase ineffective to amplify sGFP, although *Taq* pol remained stable enough to amplify it. On increasing template size, results show that recombinant *Tth* pol only amplified a 3518bp DNA fragment at low yields compared to *Taq* pol while neither polymerase could amplify a 6.6 kb plasmid template; however, processivity and thermostability are inherent features of DNA polymerases and cannot be improved by molecular chaperones.

3.2 Introduction

The mesophilic eubacterium *Escherichia coli* is one of the most widely exploited hosts used to express a number of biotechnologically relevant proteins (Georgiou & Valax, 1996). One of the advantages of using *E. coli* is the ability to express high quantities of heterologous proteins that are often produced in the native host in low concentrations. This is facilitated by the number of expression vectors that have been developed for *E. coli* but are unavailable for other organisms (Jana & Deb, 2005; Terpe, 2006). The availability of cloning and expression vectors also allows for the production of recombinant proteins developed through *in vitro* processes such as mutagenesis. Other factors that make *E. coli* the preferred expression host include its low cost, rapid growth rate and simple fermentation conditions compared to other expression systems (Makrides, 1996).

The main bottleneck to heterologous expression in *E. coli*, however, is the tendency for a number of these proteins to form homogenous, insoluble aggregates known as inclusion bodies (de Groot *et al.*, 2009; Villaverde & Carrió, 2003). A fraction of these aggregates may be easily resolubilised *in vitro* but the greater part require very strong denaturing conditions followed by *in vitro* refolding that often diminish the protein's activity or its concentration (de Marco *et al.*, 2005).

Several methods have been developed not only to increase the yield of heterologous proteins but to improve solubility and therefore, activity of the expressed proteins. Of these, the use of fusion partners such as glutathione S-transferase (GST), maltose binding protein (MBP), NusA and polyhistidine tags to enhance *in vivo* solubility is common practice (Terpe, 2003). Increasingly, novel methods for improving protein solubility look to exploiting the cells built-in folding modulators: molecular chaperones (Bergeron *et al.*, 2009; Nishihara *et al.*, 1998) .

Molecular chaperones interact transiently with their client proteins to prevent aggregation of hydrophobic patches on the client proteins or prevent non native association of protein domains during folding that may result in protein misfolding (Hartl & Hayer-Hartl, 2009). By expressing individual chaperones or a system of chaperones simultaneously with the protein of interest has demonstrated the chaperones' ability to counter aggregation and effectively improve heterologous protein solubility and activity (Kolaj *et al.*, 2009).

Co-expression of molecular chaperones with other proteins employs chaperone plasmids which are co-transformed with protein-encoding vectors into the expression host. While this

has been a successful approach (de Marco *et al.*, 2007; Kondo *et al.*, 2000), the process requires relatively high selective pressure to maintain multiple plasmids stably within the host and double transformations may be achieved with some difficulty (Tolia & Joshua-Tor, 2006).

In this project, it has been proposed that a single vector with the ability to express both the chaperone proteins and protein of interest under two separate promoters may circumvent this problem. The vectors constructed contain the thermophilic KJEA operon from *Thermus thermophilus* or *Thermus scotoductus*.

The effect of chaperone co-expression was evaluated using *Tth* DNA polymerase I, a 94 kDa, thermostable polymerase isolated from *T. thermophilus*. The polymerase has Mg^{2+}-dependent DNA polymerase activity and Mn^{2+}-dependent reverse transcriptase activity; as such it is extensively used in standard PCR and real time PCR (Rüttimann *et al.*, 1985). It lacks a 3'→ 5' exonuclease activity which lowers its fidelity, compared to commercial, high fidelity polymerases such as *Pfu* pol, a polymerase from *Pyrococcus furiosus* and KOD from *Thermococcus kodakaraensis* (Arezi, 2003; Benson *et al.*, 2003). As it is thermostable and lacks proof reading activity, it has been suggested as a possible candidate for sequencing (Arakawa *et al.*, 1996). Its processivity – the number of nucleotides incorporated in a single association/dissociation cycle – is around 60nt per second, a low processivity considering that some polymerases, such as KOD have a processivity of more than 300nt per second (Arezi, 2003; Foord & Rose, 1994).

The polymerase has also been fused to superfolder GFP – sGFP – a 28 kDa fluorescent protein developed by Pédelacq *et al.*, (2006), which is finding wide application as a thermostable fluorescent tag. Often, engineered GFPs are prone to aggregation from misfolding when expressed in *E. coli*; however, sGFP was demonstrated to fold stably and even promote folding of the fusion partner (Cava *et al.*, 2008; Pédelacq *et al.*, 2006). This is a useful property for the folding of large proteins as short protein tags, such His-tags, could be buried within the protein and not only render it difficult to detect but also difficult to purify (Niiranen *et al.*, 2007).

3.3 Methods and materials

3.3.1 List of strains, plasmids and oligonucleotide primers

The following is a list of all bacterial strains, plasmids and oligonucleotide primers used for the cloning and expression of *Tth* DNA polymerase in *E. coli*.

Table 3.1. List of bacterial strains used in this study.

Organism	Strain	Reference
E. coli	*E. coli* XL10 Gold	Stratagene
	E. coli BL21(DE3)	
	E. coli BL21(DE3)/pRARE[a]	
	E. coli BL21(DE3)/pKJE7[b]	

[a, b] Strains were developed by Dr. J. Van Marvijk, University of the Free State.

Table 3.2. List of plasmids used in this study.

Plasmid	Description	Source
pGEM®-T easy	Linear vector with 'T' overhangs for subcloning of *Taq* polymerase-amplified DNA.	Promega
pET22	High-copy number plasmid with T7/lac promoter, for high level protein expression and C-terminal His-tag for purification. Confers ampicillin resistance.	Novagen
pET28	High-copy number plasmid with T7/lac promoter, for high level protein expression, as well as N- and C-terminal His-tags for purification. Confers kanamycin resistance.	Novagen
pRARE	Plasmid for expression of rare tRNAs from the Novagen Rosetta gami™ B strain. It was isolated and transformed into *E. coli* BL21(DE3)[c].	Novagen
p22*Ts*K	Final chaperone construct containing the full KJEA operon from *T. scotoductus* SA-01, under the control of the P_{BAD} promoter and AraC regulator and a pET22 backbone.	Chapter 2
p22TkK	Final chaperone plasmid construct containing the full	Chapter 2

	KJEA operon from *T. thermophilus* HB8, under the control of the P_{BAD} promoter and AraC regulator and a pET22 backbone.	
p28*Ts*K	Final chaperone plasmid construct containing the full KJEA operon from *T. scotoductus* SA-01, under the control of the P_{BAD} promoter and AraC regulator and a pET28 backbone.	Chapter 2
p28*Tt*K	Final chaperone plasmid construct containing the full KJEA operon from *T. thermophilus* HB8, under the control of the P_{BAD} promoter and AraC regulator and a pET28 backbone.	Chapter 2
pGEMT-Pol	Recombinant pGEM®-T easy vector encoding *Tth* DNA polymerase.	This study
pGEMT-sGFP	Recombinant pGEM®-T easy vector encoding sGFP.	This study
p22tK-P	Chaperone plasmid encoding *Tt*KJEA and *Tth* pol I gene.	This study
pMHPnqo-sGFP	Source plasmid from which sGFP was amplified.	Prof. J. Berenguer, Madrid, Spain
p22sK-P	Chaperone plasmid encoding *Ts*KJEA and *Tth* pol I gene.	This study
p22-P	pET22 plasmid encoding *Tth* pol I gene; will serve as a negative control.	This study
p22sK-PF	Chaperone plasmid encoding *Ts*KJEA, *Tth* pol I gene and sGFP fusion gene.	This study
p22tK-PF	Chaperone plasmid encoding *Tt*KJEA, *Tth* pol I and sGFP fusion gene.	This study
p22-PF	pET22 plasmid encoding *Tth* pol I and sGFP gene; will serve as the negative control.	This study
pKJE7	Plasmid chaperone encoding *E. coli* DnaK, DnaJ and GrpE.	Takara Bio Incorporated
p22CYP153A6	6.6kb plasmid which served as template for recombinant *Tth* polymerases	Dr D. Opperman, Bloemfontein, South Africa
P22*Tt*K	3.5kb fragment of *Tt*KJEA operon will serve as template for recombinant *Tth* polymerases	This study

Table 3.3. List of oligonucleotide primers used for this study. All primers were obtained from Bioneer or Integrated DNA Technologies (IDT).

Primer	**5'→ 3' Sequence**	**Restriction sites added/comments**
CYP153A6THIS_F	CAC CAC CAC CAC CAC CAC TGA GAT AAG CTT	Amplifies 6.6 kb plasmid.
CYP153A6THIS_R	GGC GTT GAT GCG CAC GGG CAG	Amplifies a 6.6kb plasmid.
sGFP-1F	CCA TAT GCC ATG GAC TAG TAT CGA TGA ATT	*Nde*I
sGFP-1R	CAA GCT TTT TGT AGA GCT CAT CCA TGC CAT	*Hind*III
sGFP-2R	CAA GCT TTT ATT TGT AGA GCT CAT CCA TGC CAT GTG T	*Hind*III
Sp6 upstream	ATT TAG GTG ACA CTA TAG	Sequencing primer for genes cloned in pET22 MCS
T7 terminator	GCT AGT TAT TGC TCA GCG G	Sequencing primer for genes cloned in pET22 MCS
*Tth*DNAPolI-1F	CCA TAT GGA GGC GAT GCT TCC GCT CTT TGA A	*Nde*I
*Tth*DNAPolI-1R	CCA TAT GGC CCT TGG CGG AAA GCC AGT	*Nde*I
*Tt*KJE-1F	CGT CTA GAA TGG CCA AGG CAG TGG GCA TTG AC	*Xba*I
*Tt*KJE-1R	CAG CAT GCC TAA GTG CGC TCC AGG ATC TC	*Sph*I

3.3.2 General experimental procedures

All chemicals and reagents used to growth media were analytical grade and supplied by Sigma-Aldrich or Merck.

3.3.2.1 Molecular cloning techniques

Unless otherwise stated, all recombinant DNA techniques used are described by Sambrook *et al.*, (1989).

All PCR reactions were prepared using the Expand Long Template PCR system from Roche Applied Sciences by adding together 1 µl of double stranded DNA template, 1 µl each of 10 mM of the relevant forward and reverse primers, 1 µl of 10mM dNTP mix, 5 µl of 10 X Expand Long Template buffer system 2, and 0.5 µl (1 unit) of Expand Long Template polymerase; the reaction was made up to 50 µl with PCR-grade distilled water. The reactions were carried out in the Applied Biosystems Thermocycler 2720.

All agarose electrophoresis analyses were run on 1% agarose gel (Seakem) dissolved in TAE buffer [0.1 M Tris, 0.05 M EDTA (pH 8.0) and 0.1 mM glacial acetic acid] at a constant voltage of 90V for 30 to 45 min. Gels were analysed using the Biorad gel documentation system. When the agarose gels were used in DNA isolations, the DarkReader Transilluminator from Clare Chemicals was used. For reference, the molecular ladder, GeneRuler^{TM} DNA Ladder Mix, from Fermentas was used.

Ligation of amplicons and plasmid vectors was carried out in a 10 µl reaction mix consisting of 4 µl of insert to be ligated, 3 µl of the linearised, dephosphorylated vector, 1 µl of 10 X T4 ligase buffer, 1 µl of 10 u µl^{-1} T4 DNA ligase and made up to volume with distilled water. The exception was cloning into pGEM$^{®}$-T easy, in which case 3 µl of insert were used with 0.25 µl of the cloning vector.

Dephosphorylation of plasmid was performed in a 30 µl reaction containing 24 µl of linearised, purified plasmid, 3 µl of 5 u µl^{-1} Antarctic Phosphatase (New England Biolabs) and 3 µl of 10 X Antarctic Phosphatase buffer at 37°C, overnight. The enzyme was inactivated at 75°C for 10 min and the reaction was cooled to room temperature before use.

All restriction endonucleases, T4 DNA ligase and their corresponding buffers described here were supplied by Fermentas. Restriction digests were performed in 10 µl reactions containing 3 µl of the plasmid DNA template, 0.5 µl of 10 u µl^{-1} restriction endonuclease, 1 µl of 10 X

endonuclease buffer and made up to volume with distilled water. In the case of double digests, the enzyme units were adjusted as specified by Fermentas and when digesting a template for downstream experiments such as ligations, 6 µl of DNA template was used instead. Results were analysed by agarose gel electrophoresis. Digests were incubated for 3 hours to allow complete digestion at 37°C.

3.3.2.2 Transformation, selection of colonies and preparation of plasmid DNA

Transformation of *E. coli* strains was performed using the protocol described by Chen *et al.*, (1992), as follows: 100 µl of competent cells, according to the method described by Inoue *et al.,* (1990) were incubated with the appropriate volume of transforming DNA on ice for 15 min, followed by heat shocked at 42°C for 1 min. The cells were cold-shocked on ice-water slurry for 2 min. To this, 500 µl of Luria-Bertani (LB) broth (5 g yeast extract, 10 g tryptone, 5 g NaCl per litre of distilled water) was added, followed by incubation of the transformation mix at 37°C, with shaking for 45 min. The culture was centrifuged at 6000 x *g* for 1 min, 450 µl of the broth was discarded while the remaining 150 µl was used to resuspend the cells and plate them on LB agar plates (LB broth with 15g per litre of bacteriological agar) supplemented with the appropriate selective antibiotic.

Enumeration of single colonies was carried out by growing them in 5 ml LB broth containing the appropriate selective antibiotic at 37°C, overnight, in a shaking incubator. For plasmid DNA isolations, 1ml of this culture was used as described below.

Plasmid DNA isolations (minipreps) were carried out to screen for positive clones using the lysis by boiling method adapted from Sambrook *et al.,* (1989), as follows: 1 ml of *E. coli* overnight culture was harvested and centrifuged at 6000 x *g* to pellet the cells; the supernatant was discarded while the pellet was resuspended in 350 µl of STET buffer [8% w/v Sucrose, 5% v/v Triton X-100, 50 mM EDTA, 50 mM Tris (pH 8.0)]. For lysis, 25 µl of lysozyme [10 mg ml^{-1}] was added and the suspension was boiled at 100°C for 44 s. The reaction mix was centrifuged at 13 000 x *g* for 10 min; the resulting pellet was removed with a sterile toothpick and discarded. The supernatant was precipitated with 40 µl of 2.5 M sodium acetate (pH 5.2) and 420 µl isopropanol at -20°C for 20 min. The mixture was centrifuged at 4°C, 13 000 x *g*, for 20 min, the supernatant was aspirated and the pellet was washed in 70% ethanol. This mix was centrifuged for 10 min at 4°C, 13 000 x *g*, followed by aspiration of the supernatant. The pellet was dried and resuspended in 40 µl of TE and RNAse [10 µg ml^{-1}]. For the purposes of enumerating plasmids from positive clones,

minipreps were performed using the Bioflux Plasmid DNA Extraction Kit. All DNA concentrations were measured using the NanoDrop ND-1000 Spectrophotometer from NanoDrop Technologies.

3.3.2.3 Sequencing reactions

Sequencing of cloned inserts was performed while ligated into pGEM®-T easy and again when they were cloned into the pET vector backbones. Reactions were set up using the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) as follows: 1 µl DNA template, 1 µl of 3.2 pM primer, 2 µl of 5 X sequencing buffer, 0.5 µl premix and 6.5 µl PCR-grade distilled water. Thermal cycling was carried out for 25 cycles at 94°C for 10 s, 50°C for 5 s and 60°C for 4 min.

The EDTA-ethanol precipitation method was used to prepare samples for sequencing as follows: samples were mixed with 10 µl of distilled water, 5 µl of 125 mM EDTA (pH 8.0) and 60 µl of 100% ethanol. Samples were vortexed and incubated at room temperature for 15 min, then centrifuged at 4°C, 13 000 x *g*, for 20 min. The supernatant was aspirated, the pellet was washed in 120 µl of 70% ethanol and centrifuged and 4°C, 13 000 x *g*, for 10 min. Ethanol was aspirated and the samples were dried and then sequenced using the Hitachi 3130x*I* Genetic Analyser, also from Applied Biosystems. Sequence chromatograms were analysed with Geneious v4.8.2 (Biomatters) software and aligned to the original sequences using ClustalW software hosted by the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

3.3.3 Cloning of *Tth*Pol and sGFP into chaperone vectors

3.3.3.1 PCR amplification of *Tth*Pol I and sGFP

The gene encoding *T. thermophilus* DNA polymerase I, *Tth*Pol I, was amplified from gDNA extracted from *T. thermophilus* HB8. Thermal cycling was carried out with an initial denaturation of 94°C for 3 min and the program of [94°C for 10 s, 59°C for 30 s and 68°C for 3 min] repeated for 35 cycles, as well as a final extension of 4 min, using the primer pair *Tth*DNA PolI-1F/*Tth*DNA PolI and Expand Long Template Polymerase.

The plasmid, pMHPnqo-sGFP, supplied by Prof. J. Berenguer, was the template used to amplify sGFP. The PCR reaction was cycled [94°C for 30 s, 55°C for 30 s and 68°C for 1

min] for 35 cycles using the sGFP-1F/sGFP-1R or sGFP-1f/sGFP-2R primer pairs. The sample was initially denatured at 94°C for 2 min and had a final extension of 3 min at 68°C.

3.3.3.2 Subcloning of *Tth*PolI and sGFP into pGEM®-T easy

The *Tth*Pol I amplicon and sGFP amplicons were ligated into pGEM®-T easy, using T4 DNA ligase at 4°C, overnight. The ligation mix was used to transform competent *E. coli* XL10Gold cells, which were plated onto LB_{AIX} plates.

White colonies were inoculated into LB_{Amp} broth and grown overnight at 37°C. Plasmid minipreps were carried out on colonies and screening was performed by restriction analysis to identify recombinant pGEMT-Pol and pGEMT-sGFP. Insertion of *Tth*PolI was confirmed by restriction digestion by *Nde*I, while insertion of sGFP was confirmed by digestion with *Nde*I/ *Hind*III double digest.

3.3.3.3 Cloning *Tth*PolI and sGFP into the chaperone coexpression plasmids

Restriction digests were carried out using *Nde*I and *Nde*I/*Hind*III digest to release *Tth*PolI and sGFP from pGEMT-Pol and pGEMT-GFP. The fragments were separated on agarose and purified with the Bioflux Gel Extraction Kit. The chaperone plasmids (p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K) constructed previously were linearised, first, with *Nde*I and *Hind*III, as was pET22 and pET28. The linearised vectors were gel-purified. Purified sGFP was then ligated into these vectors via *Nde*I and *Hind*III, using T4 DNA ligase. The ligation reaction was used to transform competent *E. coli* XL10 Gold and plated onto LB_{Amp} (pET22) or LB_{Kan} (pET28). Colonies were inoculated into LB_{Amp} or LB_{Kan} broth and grown overnight at 37°C. Cells were harvested and plasmid DNA minipreps were performed to isolate plasmid. Positive clones, bearing the recombinant plasmids, p22-F, p22sK-F, p22tK-F, p28-F, p28sK-F or p28tK-F, were identified by digesting the plasmid with *Nde*I and *Hind*III.

Plasmids containing sGFP were linearised further with *Nde*I and gel-purified. The plasmids were dephosphorylated and ligated to the *Tth*PolI gene via *Nde*I by T4 DNA ligase. The ligation reaction was used to transform competent *E. coli* XL10 Gold and plated onto LB_{Amp}. Colonies were inoculated into LB_{Amp} broth and grown overnight at 37°C. Plasmid DNA minipreps obtained from these cultures were screened for positive clones, bearing recombinant p22-PF, p22sK-PF and p22tK-PF, by *Nde*I restriction analysis. Correct orientation of *Tth*PolI was confirmed by digesting plasmids with *Bgl*II.

The same process was repeated to construct chaperone co-expression vectors containing only *Tth*PolI; with the exception that chaperone co-expression vectors were digested only with *Nde*I to yield recombinant p22-P, p22sK-P and p22tK-P. Correct orientation of *Tth*PolI in the MCS was confirmed by digesting the plasmids with *Bgl*II.

3.3.4 Heterologous overexpression and purification of *Tth*PolI and *Tth*PolI-sGFP

3.3.4.1 Expression of *Tth*PolI and *Tth*PolI-sGFP

The six recombinant plasmids, possessing either *Tth*PolI or *Tth*PolI-sGFP fusion, were used to transform competent *E. coli* BL21 (DE3). In addition p22-P and p22-PF was also transformed into *E. coli* BL21 (DE3)/pKJE7, a strain that possesses the commercial plasmid pKJE7, which encodes *E. coli* derived (*Ec*) DnaK operon. Transformants were grown on LB_{Amp} or $LB_{Amp/Cm}$ plates, the latter in case of strains bearing the commercial vector. Single colonies were inoculated into 5 ml of LB_{Amp} or $LB_{Amp/Cm}$ and grown overnight at 37°C. The overnight culture was used to inoculate 500 ml of LB supplemented with the appropriate antibiotic. Cultures were divided into 2 X 250 ml LB broth in 1000ml shake flasks. Cells were grown to an OD_{600} of ~0.8 – 1 at 37°C. Chaperone expression was induced by a final concentration of 0.5 mg ml^{-1} of L-arabinose. Expression was carried out at 30°C for 2 hrs, followed by induction with IPTG to a final concentration of 1mM. Expression was carried out for a further 6 hrs at 30°C, after which cells were harvested by centrifugation at 6000 x *g*, 4°C, for 3 min in 50 ml falcon tubes. The cells were the resuspended in 30 ml of 20 mM sodium phosphate buffer (pH 7.4).

3.3.4.2 Purification of *Tth*PolI and *Tth*PolI-sGFP

The resuspended cells were washed three times in 30 ml Binding buffer (20 mM sodium phosphate, 0.5 M NaCl and 20 mM imidazole), pH 7.4, and centrifuged at 6000 x *g* at each wash to pellet the cells. The washed cells were resuspended in 30 ml Binding buffer supplemented with cOmplete, Mini, EDTA-free Protease Inhibitor tablets. Lysis was carried out by single passage of the cell suspension through the Constant Cell Disruptor: One Shot Model (Constant Systems) at 32.5 kpsi. The suspension was cleared of the insoluble fraction by centrifuging at 6000 x *g*, 4°C, for 20 min. The lysate was then ultracentrifuged under vacuum at 30 000 rpm, 4°C for 1 h 30 min using the Optima™ L-100XP ultracentrifuge

(Beckman-Coulter). Purification of *Tth*PolI protein and sGFP-coupled *Tth*PolI from the cell-free extract was carried out by using 5 ml His Trap™ FF columns and the Akta Prime Plus purification system (GE Healthcare); the proteins, already in Binding buffer, were loaded onto the column and eluted using a 20 – 500 mM imidazole gradient. The Elution buffer contained 20 mM sodium phosphate, 0.5 M NaCl and 0.5 M imidazole. Fractions containing protein were pooled, dialysed against 20 mM sodium phosphate buffer (pH 7.4) using SnakeSkin® Pleated Dialysis tubing, 10 000 MWCO (Thermo Scientific) and concentrated with 10 kDa spin columns (Amicon Bioseparations).

From this point, recombinant polymerases will be designated as *Tth*PolI_N, *Tth*PolI_T, *Tth*PolI_S and *Tth*PolI_E to denote polymerases purified from cells possessing p22-P, p22tK-P, p22sK-P and pKJE7/p22-P, respectively. The sGFP-coupled polymerases are designated *Tth*PolI-sGFP_N, *Tth*PolI-sGFP_T, *Tth*PolI-sGFP_S and *Tth*PolI-sGFP_E and were purified from strains containing recombinant vectors p22-PF, p22tK-PF, p22sK-PF and pKJE7/p22-PF. The commercial polymerases, from Supplier A and Supplier B are designated Pol_A and Pol_B, respectively.

Total protein concentration was calculated using the Micro BCA™ Protein Assay Kit (Pierce Biotechnology). To standardise protein units of enzyme to be used in downstream assays, the protein concentration for 1 unit of activity from two commercial DNA polymerases was also calculated using the same kit. The same protein content in all the other purified polymerases was used as a unit measure for these proteins.

3.3.5 Evaluation of processivity, fidelity and thermotolerance in *Tth*PolI and *Tth*PolI-sGFP

3.3.5.1 Evaluating the processivity of purified DNA polymerases

To measure the size range of DNA templates for the purified *Tth*PolI and sGFP-coupled *Tth*PolI, PCR reactions were set up for each polymerase as 50 µl reactions containing ~ 150 ng DNA template, 2.5 µl each of 10 mM forward and reverse primers, 1 µl of 10 mM dNTP mix, 5 µl of 25 mM $MgCl_2$ (Supplier B), 5 µl of 5 X DNA polymerase buffer (Supplier B), 2.5 units of DNA polymerase enzyme and a volume of PCR-grade distilled water to make up the 50 µl volume. The templates supplied were the sGFP gene (771bp) in recombinant pGEMT-GFP, the *Tt*KJEA operon (3518bp) in p22*Tt*K and the plasmid p22Cyp153A6

(6.6kbp); the primer pairs used to amplify each template, respectively, were sGFP-1F/sGFP-1R, *Tt*KJE-1F/*Tt*KJE-1R and CYP153A6THIS_F/ CYP153A6THIS_R. Thermal cycling was performed for 25 cycles as follows: [95°C for 30 s, 55°C for 30 s and 74°C for 1 min] for the sGFP template; [95°C for 30 s, 60°C for 30 s and 74°C for 4 min] for the *Tt*KJE template and [95°C for 10 s, 60°C for 30 s and 74°C for 7 min] for the p22Cyp153A6 template.

3.3.5.2 Evaluating the fidelity of purified DNA polymerases

Amplicons obtained from the PCR amplification of sGFP by heterologously expressed DNA polymerases were sequenced using the primer pair sGFP-1F/sGFP-1R. Reactions were set up using the BigDye® Terminator v3.1Cycle sequencing kit as 1 µl sGFP PCR template, 1 µl of 3.2 pM forward and reverse primer, 2 µl of 5 X sequencing buffer, 0.5 µl Premix and 5.5 µl of PCR-grade water. Thermal cycling was performed for 25 cycles at 94°C for 10 s, 50°C for 5 s and 60°C for 4 min. Preparation of reactions for sequencing, analysis and alignment was carried out as described in General Procedures.

3.3.5.3 Evaluation of thermostability of purified polymerases

The pGEMT-GFP plasmid was used as template to amplify sGFP by PCR using heat treated DNA polymerase. The reactions were prepared as 50 µl mixes containing ~ 150 ng pGEMT-GFP template, 2.5 µl each of 10 mM sGFP-1F and sGFP-1R primers, 1 µl of 10 mM dNTP mix, 5 µl of 10 X DNA polymerase buffer, 5 µl $MgCl_2$, 2.5 units of DNA polymerase and PCR-grade water. Heat treatment was carried out by setting up the following PCR program: the initial heat treatment was carried out for 1 hr at 95°C and immediately after, thermal cycling was performed for 25 cycles [94°C for 10 s, 50°C for 5 s and 60°C for 4 min].

3.4 Results and discussion

3.4.1 Cloning of sGFP and *Tth*Pol into the chaperone plasmids

The sGFP gene, encoding a thermostable green fluorescent protein, and *Tth*PolI, encoding thermostable DNA polymerase from *T. thermophilus* were amplified by PCR (Fig 3.1), subcloned into pGEM®-T easy and finally cloned into the MCS of the chaperone co-expression vectors constructed previously, to give a recombinant DNA polymerase fused to sGFP. An uncoupled DNA polymerase was also cloned into the chaperone vectors to compare expression capacity and activity.

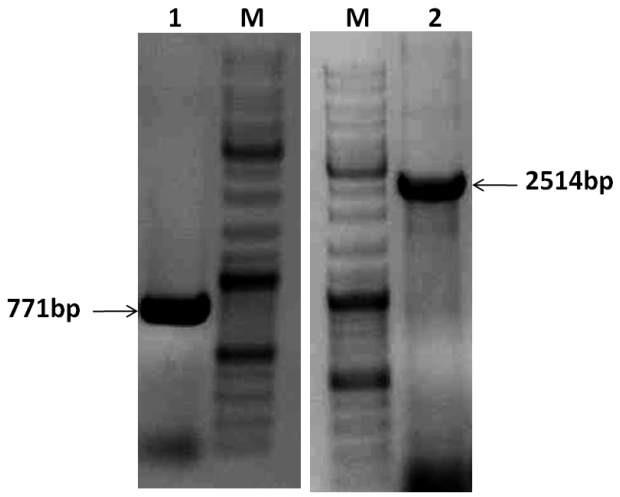

Figure 3.1. PCR amplification of the 771 bp fragment encoding sGFP (lane 1) and the 2514 bp fragment encoding *Tth*PolI (lane 2) from *T. thermophilus* HB8. M is the 10kb molecular weight marker used for all agarose gels in this study.

Construction of plasmid vectors to express *Tth*PolI and *Tth*PolI fused to sGFP were relatively successful. From pET22, the following series of vectors were obtained: p22-P, p22tK-P and p22sK-P to express *Tth*PolI while p22-P, p22tK-P, and p22sK-P vectors were developed to express sGFP-coupled *Tth*PolI.

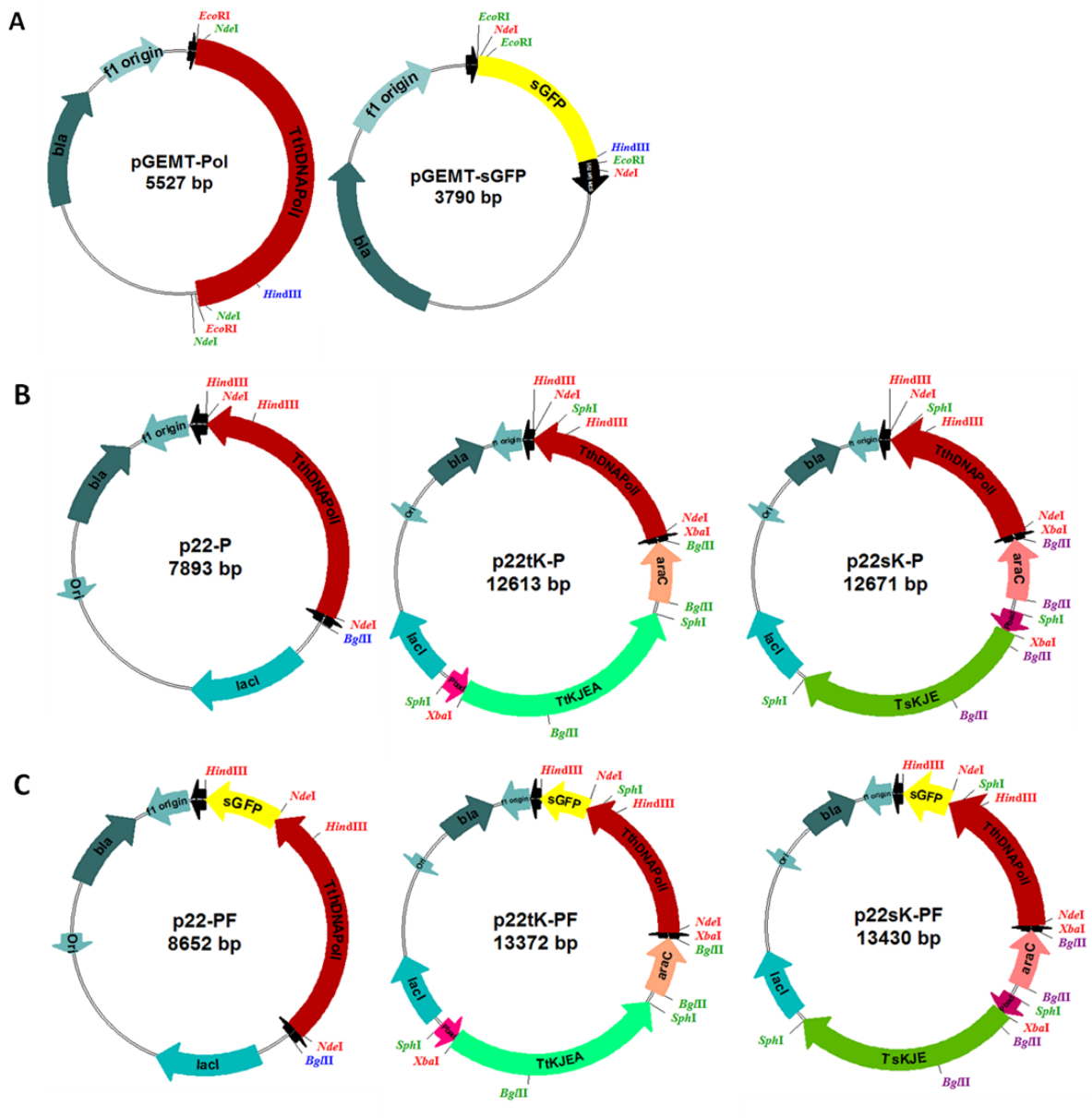

Figure 3.2. Diagrammatic representation of chaperone plasmids with the cloned fragments of sGFP and *Tth*PolI. In A, the fragments are ligated into pGEM® T easy vectors during the subcloning steps. In B, the series developed for the uncoupled *Tth*PolI gene using the pET22 backbone is shown while C shows the *Tth*PolI-sGFP gene-fusion.

Restriction analysis of the plasmids shown in Fig. 3.2 was performed using *Nde*I and *Bgl*II to confirm correct orientation of the *Tth*PolI gene with respect to the T7/lac promoter (Fig. 3.3).

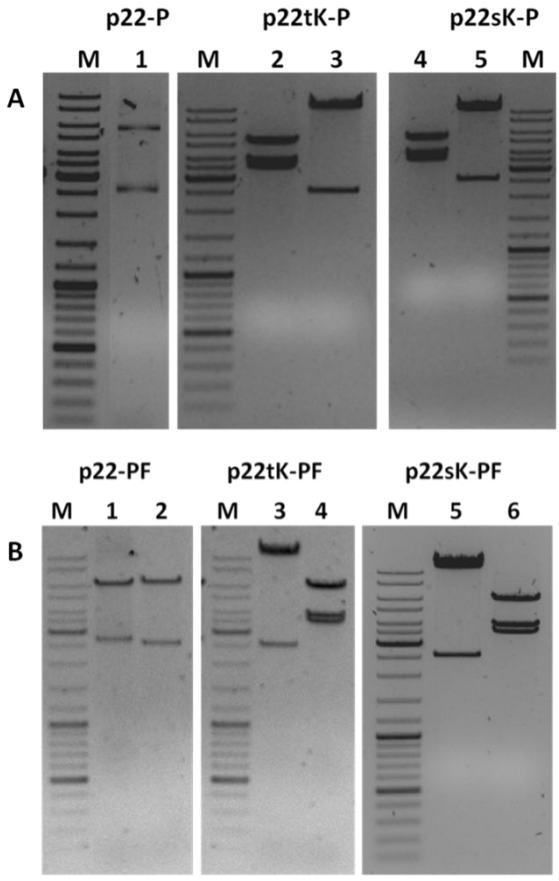

Figure 3.3. Restriction digest of plasmid vectors expressing *Tth*Pol and *Tth*PolI-sGFP. In A, lanes 1, 2 and 4 show the series developed to express *Tth*PolI digested with *NdeI* while lanes 2 and 5 show the plasmids digested with *BglII*. In B, plasmids developed to express *Tth*PolI-sGFP fusion gene are shown digested by *NdeI* (lanes 1, 3 and 5) and *BglII* (lanes 2, 4 and 6).

Developing a series of DNA polymerase expression vectors based on the pET28 vectors was not successful and the obstacle preventing this is yet to be identified. Difficulties were encountered in ligating the Tth PolI and sGFP fragments in the MCS. This is not problem of experimental procedure but rather the limited time available to complete all constructs. The main advantage of having the pET28-derived chaperone vectors was the presence of N-terminal and C-terminal His-tags as opposed the single C-terminal His-tag of the pET22-based vectors. The presence of two fusion tags not only increases versatility of the expression vector, it may also improve purification of the protein to be expressed from it by increasing the protein's affinity for the column with which it is purified (Khan *et al.*, 2006). However, the single His-tag in the pET22-derived series is still sufficient to enable purification of the protein expressed from them.

3.4.2 Heterologous expression and purification of *Tth*PolI and sGFP-couple *Tth*PolI using thermophilic DnaK chaperone system

3.4.2.1 Effect of heterologous expression on protein content and biomass

Expression of *Tth*PolI and *Tth*PolI-sGFP was initiated from the chaperone plasmids at the T7/lac promoter after 2 hours of inducing DnaK expression with L-arabinose. The resulting proteins, *Tth*PolI$_N$, *Tth*PolI$_T$, *Tth*PolI$_S$, *Tth*PolI-sGFP$_N$, *Tth*PolI-sGFP$_T$ and *Tth*PolI-sGFP$_S$ were purified by Ni-affinity purification. For comparison, the commercial pKJE7 plasmids were also used for expression, after which *Tth*PolI$_E$ and *Tth*PolI-sGFP$_E$ were similarly purified.

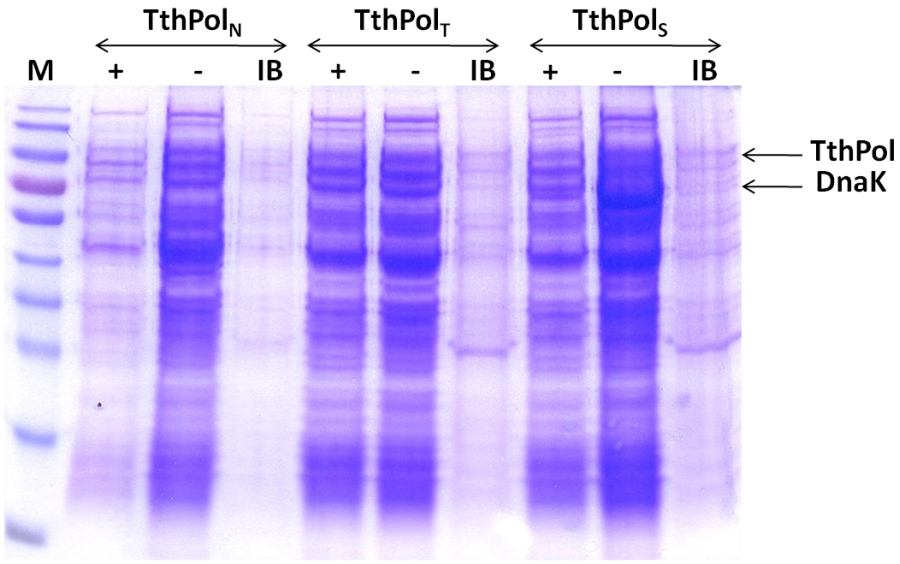

Figure 3.4. Expression profiles of recombinant *Tth*Pol in *E. coli*, as shown by SDS-PAGE. Lanes showing IPTG induction (+) and no IPTG induction (-) are shown along the insoluble fraction (IB) after 6 hours. M is the molecular weight marker used for all protein gels. In all profiles, the decrease in protein content can be observed after induction with IPTG as a result of increasing metabolic load.

Heterologous expression of ~94kDa *Tth*PolI and ~132kDa sGFP-coupled *Tth*PolI proteins appears to take place but with relatively low amounts of protein produced, such that they are hardly visible on SDS-PAGE (Fig 3.4). Induction by IPTG shows no marked change in the expression profiles except between those of p22sK-P, p22sK-PF. While the negative control does not express any thermophilic DnaK chaperones because it lacks them, p22tK-P does not express its *Tt*KJEA proteins due to lack of promoter activation by AraC regulator protein.

The same is observed for p22-PF and p22tK-PF. However, in the *Ts*KJEA expression vectors, p22sK-P, the level of DnaK decreases when IPTG is added to the expression media, indicating that the drop in concentration of DnaK must be due to successful expression of an additional protein, *Tth*PolI$_S$, taking place (Fig. 3.4). Kolaj *et al.*, (2009), report this to be a common feature of expression systems, where the total cytosolic protein content remains constant but the relative concentrations of each protein differ as expression proceeds. This, in turn, arises from competition between the two promoters, P_{BAD} and T7/lac, where RNA polymerase units may be divided between the two promoters. To allow the production of sufficient chaperone proteins, induction of P_{BAD} was performed two hours before that of T7/lac; the extended time should immediately provide DnaK chaperones ready to prevent aggregation of newly synthesised *Tth*PolI (de Marco & de Marco, 2004).

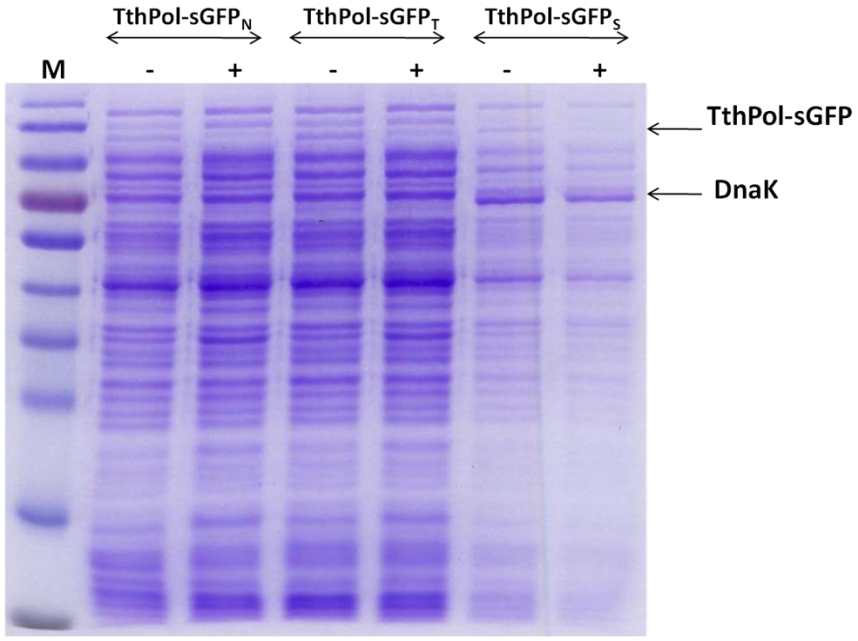

Figure 3.5. Expression profiles of recombinant *Tth*Pol-sGFP in *E. coli*, as shown by SDS-PAGE. Lanes showing IPTG induction (+) and no IPTG induction (-) are shown. M is the molecular weight marker used for all protein gels. The decrease in protein content can be observed in p22sK-PF fractions (expressing *Ts*DnaK chaperone system and *Tth*Pol-sGFP$_S$) after induction with IPTG as a result of increasing metabolic load.

For the expression of *Tth*PolI-sGFP$_N$, *Tth*PolI-sGFP$_T$ and *Tth*PolI-sGFP$_S$, the change in expression is marked by a decrease in cell biomass, which is seen on SDS-PAGE gel as a decrease in the overall intensity of protein content (Fig 3.5). This change is observed for the IPTG-induced and IPTG-uninduced cells containing p22sK-PF. This is because the metabolic

load in these cells has been increased by the four heterologous proteins, *Ts*DnaK, *Ts*DnaJ, *Ts*GrpE and *Ts*DafA; upon induction with IPTG, the fifth protein, *Tth*PolI-sGFP$_S$, also adds to this load. For the vectors expressing no chaperones, the negative control (p22-PF) and p22tK-PF, there is no such reduction in biomass as only one protein, *Tth*PolI-sGFP$_N$ or *Tth*PolI-sGFP$_T$ is being expressed in each. Change in growth rate is another feature of heterologous expression systems and is especially marked for plasmid-encoded proteins due to the strong promoters used to produce them (Bentley *et al.*, 1990; Scott *et al.*, 2010).

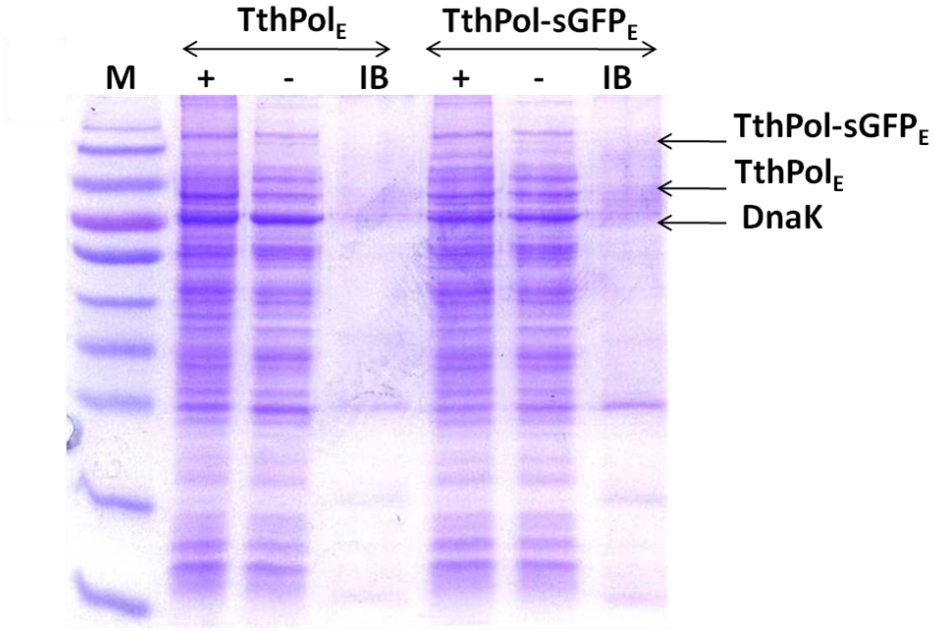

Figure 3.6. Expression profiles of recombinant *Tth*Pol$_E$ and *Tth*Pol-sGFP$_E$ in *E. coli*. Lanes showing IPTG induction (+) and no IPTG induction (-) are shown along the insoluble fraction (IB). M is the molecular weight marker. For these plasmids, the decrease in protein content as a result of increasing metabolic is not evident and has been attributed to plasmid segregation, whereby one plasmid is selectively expressed due to its higher abundance in the cell culture.

On comparing these results with the SDS-PAGE profiles of the commercial chaperone vectors, the first difference to note is that DnaK concentration does not decrease upon induction with IPTG and there is no detectable decrease in biomass (Fig. 3.6). This could be because the chaperone expression system is more efficient at producing native chaperones, compared to heterologous chaperones. Osipiuk & Joachimiak, (1997), reported that *Tt*DnaK, *Tt*DnaJ and *Tt*GrpE have a relatively high proline content and 55.5% amino acid similarity to *Ec*DnaK, 43.1% to *Ec*DnaJ and only 28.1% to *Ec*GrpE; these differences might cause slower expression of the thermophilic proteins. In addition, the use of multiple plasmids in a single

vector may cause plasmid segregation so that they are asymmetrically divided amongst replicating cells. This imbalance might favour one plasmid species over the other and the end result is that a high concentration of proteins on the abundant plasmid accumulates while that of the other plasmid remains low (Popov *et al.*, 2011). It is possible that pKJE7, being transformed first into *E. coli* BL21 (DE3), has a higher abundance that p22-P and p22-PF, which were transformed second, causing an uneven distribution of recombinant plasmids in the culture. As such, expression from pKJE7 is generally guaranteed while expression from the other two plasmids occurs at extremely low levels. This provides another reason as to why heavy selective pressure must be maintained and why a single expression vector would be desirable, as it eliminates the problem of plasmid segregation; all proteins are expressed from one vector and differential expression arises only from the two promoters.

Table 3.4. The relationship between biomass and protein concentration in the expression of *Tth* DNA polymerases using molecular chaperones.The two commercial polymerases were used to standardise protein concentrations to be used for activity assays. Using Pol_A, 5 units of enzyme = 452.7ng protein. Total protein concentration was calculated after purifying *Tth* polymerase from the cell-free extract with BCA protein assay kit.

Vector used	**Protein purified**	**[a]Biomass (g)**	**[Protein] (µg ml^{-1})**
p22-P	*Tth*PolI$_N$	3.76	262.9
p22tK-P	*Tth*PolI$_T$	3.42	201.8
p22sK-P	*Tth*PolI$_S$	3.18	207.7
p22-PF	*Tth*PolI-sGFP$_N$	2.78	149.9
p22tK-PF	*Tth*PolI-sGFP$_T$	4.21	218.3
p22sK-PF	*Tth*PolI-sGFP$_S$	4.07	168.3
p22-P/pKJE7	*Tth*PolI$_E$	3.57	206.0
p22-PF/pKJE7	*Tth*PolI-sGFP$_E$	2.69	166.4
	[b]Pol$_A$		452.7
	[b]Pol$_B$		630.9

[a] The figures quoted for mass are those of wet biomass.
[b] The two commercial polymerases were used as supplied.

According to Table 3.4, there is a decrease in biomass from cells transformed with p22sK-P (3.18 g) compared to those transformed with p22-P (3.76 g) and p22tK-P (3.42 g). There is

also a corresponding decrease in protein concentration, from 262.9 μg ml^{-1} in p22-P to 207.7 μg ml^{-1}, in p22sK-P and p22sK-PF. As described, this has been attributed to increase in metabolic load during heterologous expression. A similar trend is observed in total protein concentration from p22-PF (149.9 μg ml^{-1} with 2.78 g of biomass) to p22sK-PF (4.07 g of cells produce only 168.3 μg ml^{-1} protein). Overall decrease in mass is observed in the negative control stains when expression changes from ~94 kDa *Tth*PolI to ~122 kDa *Tth*PolI-sGFP.

The only exceptions are the p22tK-P and p22tK-PF strains from which the concentration of purified protein remains relatively unchanged in relation to biomass. As *Tth*PolI-sGFP$_T$ was not purified (Fig. 3.7 and Fig. 3.8), it is assumed that the protein measured from p22tK-PF strain is of contaminant proteins that were co-purified on the Ni-affinity column. After expression, *Tth*PolI-sGFP$_E$ was purified without success and remained in the unbound fraction.

No significant change is observed from the inclusion body fraction. Generally, the polymerase would be seen as a dense band in the insoluble fraction if it had been misfolded; as such, it is not possible to determine how much of the protein aggregated, if at all, or was rescued from aggregation by the *Ts*DnaK chaperone system. It could also indicate that *Tth* polymerase is not prone to aggregation when heterologously expressed in *E. coli*. That the polymerase misfolds and is not easily purified when fused to sGFP is not an indication that it is also an aggregation-prone species, it appears to retain high solubility despite the misfolding (Kopito, 2000).

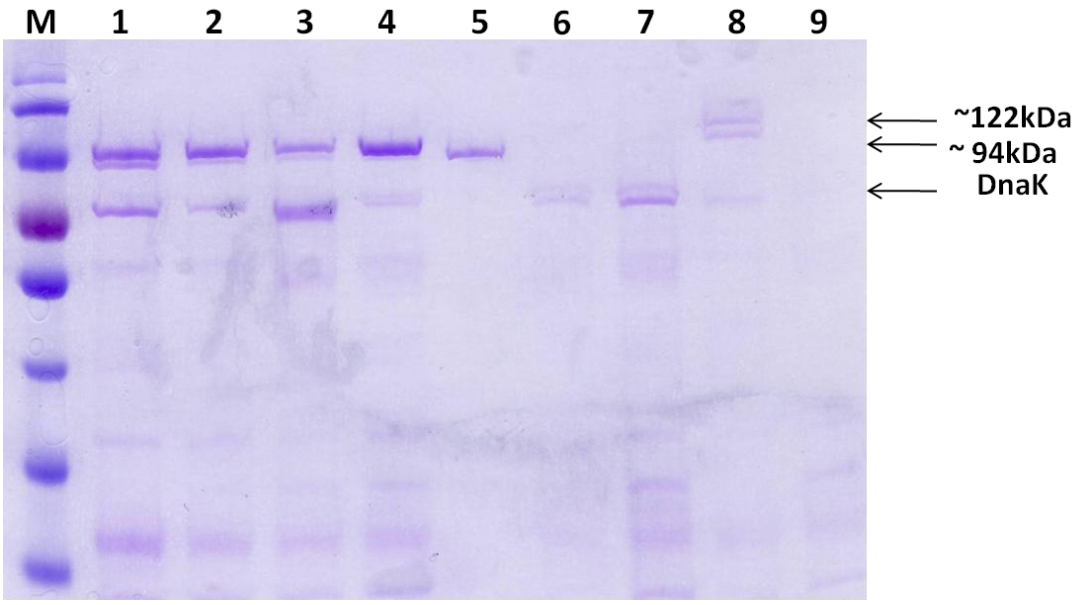

Figure 3.7. The recombinant *Tth* polymerases purified from *E. coli* by Ni-affinity chromatography. Lanes 1 to 4 show *Tth*Pol_N, *Tth*Pol_T and *Tth*Pol_S and *Tth*Pol_E, respectively. Lane 5 shows Pol_A, a commercial *Taq* Polymerase that was used to standardise protein units. Lanes 6, 7 and 9 are the purified fractions for *Tth*Pol-sGFP$_N$, *Tth*Pol-sGFP$_T$ and *Tth*Pol-sGFP$_E$ from which no DNA polymerase was recovered. Lane 8 is the purified fraction for *Tth*Pol-sGFP$_S$ showing the ~ 122kDa fusion protein. The equivalent of 50 units were loaded for each polymerase but band intensities differ as the total amount of protein purified also contained a fraction of protein contaminants, which decrease the actual concentration of *Tth* polymerase.

The presence of co-purified proteins is also observed by SDS-PAGE gel in the other protein fractions, which indicates that this particular method of protein purification is not overly efficient in excluding untagged proteins from co-eluting with the protein of interest (Fig. 3.7) (Smith, 2005). It also means that the protein concentration values that are attributed to the yield of *Tth* polymerases are not accurate and a fraction of those values are contributed by the contaminants. Their contribution to protein yield can only be calculated by purifying the polymerases further by a technique such as size-exclusion chromatography and determining total protein concentration again (Voedisch & Thie, 2010). The differences in protein concentration reported do not take into consideration protein lost during the purification process, nonetheless, it is important to be aware that it does occur and might have a significant impact on the final results.

In Fig. 3.7, a ~ 70kDa band was observed in the purified fractions. This band corresponds to DnaK, which has been identified as a common protein contaminant in purification procedures

(Ratelade *et al.,* 2009). Rial & Ceccarelli, 2002, report that the peptide region connecting fusion proteins to each other are hotspots for DnaK binding. Considering that this protein was overexpressed and that it associates with heterologous proteins to prevent their aggregation, it is expected that it would be present in high enough concentrations to be co-purified with the *Tth* polymerase despite it not being fused to an affinity tag. Perhaps if the Ni-affinity column had been more efficient in selecting only Hi-tagged proteins, DnaK could have been eluted with all the other cellular proteins.

3.4.2.2 Effect of thermophilic DnaK chaperone proteins on folding ability

The effect of DnaK chaperones on folding is not evident in the uncoupled *Tth* polymerases, as protein was efficiently purified for all three plasmids. However, there is a difference in the purification profile of the sGFP-coupled *Tth* Polymerases, which is observed more easily as the fused sGFP fluoresces under UV light. Fig. 3.8 shows that *Tth* polymerase was purified only from cells expressing chaperone proteins, which is shown by an increase in fluorescence of the purified fractions and a corresponding decrease in fluorescence in the fraction of unbound cellular proteins. For *Tth*PolI-sGFP$_N$ and *Tth*PolI-sGFP$_T$, the proteins remained in the unbound fraction and were unable to bind to the Ni-affinity column. By comparing these results to the uncoupled polymerases, it may be concluded that the larger size, ~ 28kDa of sGFP, in the sGFP-coupled polymerases creates a dependence on folding modulators to attain the correct native conformation, which is absent in the smaller ~94kDa proteins. A number of proteins have been reported to have obligate dependence on the presence of certain chaperones such as DnaK and chaperonins due to a large size or requiring multi-subunit assembly and from these profiles, it is evident that the larger the protein, the greater its reliance on molecular chaperones to fold it correctly (Hartl & Hayer-Hartl, 2009; Ranson *et al.*, 1998).

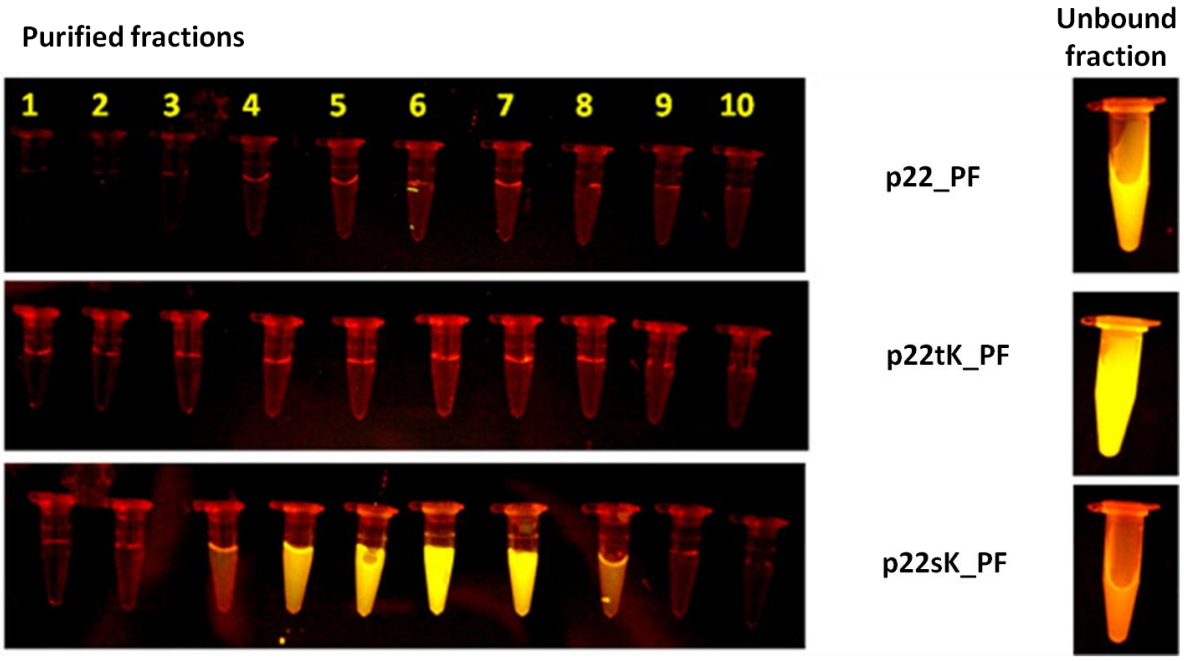

Figure 3.8. Purification of sGFP-coupled *Tth*PolI from p22-PF, p22tK-PF and p22sK-PF by Ni-affinity chromatography. The sGFP fluoresces under UV and enables localisation of sGFP-tagged proteins, in this case, the *Tth* DNA polymerase. The effect of the DnaK chaperone system on enhancing folding is easily seen in p22sK-PF fraction; *Tth*Pol-sGFP$_S$ was the only polymerase to be successfully purified of the sGFP-coupled *Tth* polymerases. The other two proteins, *Tth*PolI-sGFP$_N$ and *Tth*PolI-sGFP$_T$ come from plasmids that do not express DnaK chaperone proteins (p22-PF and p22tK-PF) and were eluted with all the other cellular proteins that could not bind to the purification column, causing the corresponding unbound fraction of eluents to fluoresce brighter that the purified fractions.

After expression, *Tth*PolI-sGFP$_E$ was purified without success and remained in the unbound fraction. It could not be detected on SDS-PAGE gel and reinforces the idea that as a protein increases in size, so does its dependence on molecular chaperones to assist it in protein folding. It also reveals the superior holdase function in the thermophilic *Ts*DnaK proteins, as it was possible to purify a *Tth* DNA polymerase fused to a sGFP partner. *Tt*DnaKJEA was able to fold a fused, ~122kDa protein at mesophilic temperature (30°C) and enable its purification while, the *Ec*DnaKJE chaperones, are unable to fold a heterologous protein at their native physiological temperature.

3.4.3 Evaluation of activity in *Tth*PolI and *Tth*PolI-sGFP expressed using thermophilic DnaK chaperone proteins

3.4.3.1 *Tth* polymerase processivity and thermostability

Elongation of sGFP was accomplished by all polymerases except *Tth*PolI-sGFP$_N$, *Tth*PolI-sGFP$_T$ and *Tth*PolI-sGFP$_E$. Reactions using these protein fractions were primarily carried out to test whether there may be small amounts of recombinant polymerase that were purified but could not be detected by SDS-PAGE and as they are negative for polymerase activity, it is final proof that they were not purified due to poor folding in the protein (Fig. 3.9).

The efficiency can only be measured by observing the intensity of bands created by the 771 bp sGFP amplicons (Fig. 3.9A); the bands show that *Tth*PolI$_N$, *Tth*PolI$_T$ and *Tth*PolI$_S$, are as efficient as the commercial Pol$_A$ or Pol$_B$. All three *Tth* polymerases show the same intensity, therefore, the effect of *Ts*DnaK chaperone system on *Tth*PolI$_S$ cannot be evaluated. There is a visible difference between the activity of *Tth*PolI$_S$, which is higher, and *Tth*PolI$_E$, which appears to be much lower. These differences could arise from the different efficiencies of thermophilic and mesophilic DnaK contributing to native folding of the polymerase which, in turn, translates to an increase in activity. It could also be related to the plasmid segregation phenomenon described earlier, in which uneven distribution of compatible plasmids occurs within a cell; it might be that the plasmid encoding *Tth*PolI$_E$, p22-P, was in low abundance in the culture despite maintaining selective pressure on both plasmids. Both these shortcomings are effectively solved by the single expression vector that was constructed using thermophilic DnaK chaperone proteins.

Although a product was obtained for *Tth*PolI-sGFP$_S$, it is much less compared to that of the commercial Pol$_A$ and Pol$_B$, as well as the uncoupled *Tth* polymerases. Possible causes might be due to the low concentration of actual *Tth*PolI-sGFP$_S$ present in the purified fraction. The effect of the coupled sGFP on polymerase activity is also unknown. Pédelacq *et al.*, (2006), report it to have no effect on the activity of its fusion partners but in this case, the interaction between sGFP and *Tth*PolI at the tertiary level is unknown; it might be that sGFP does not inhibit the enzyme allosterically but rather slows the enzyme by acting as ‘dead weight’, with respect to catalytic activity.

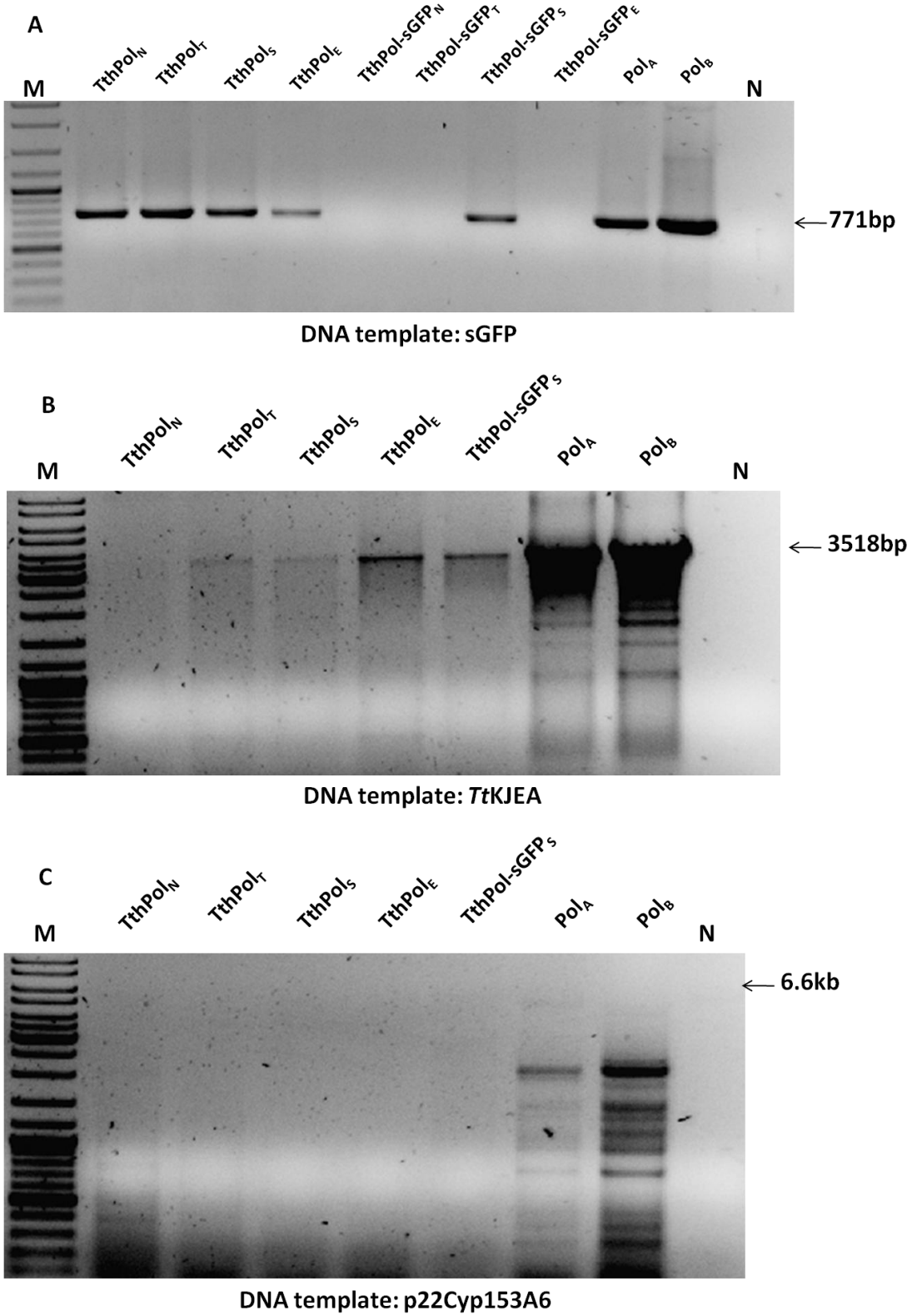

Figure 3.9. PCR amplification of sGFP, *Tt*KJEA and p22Cyp153A6 by recombinant *Tth* DNA polymerase. Each lane represents a DNA template amplified by the specific polymerase indicated at the top of the image. N represents the negative control for the PCR reactions in which DNA polymerase has been omitted. M is the molecular weight marker.

Considering that the unit concentration for recombinant *Tth* polymerases is lower than that of the commercial *Taq* polymerases, these polymerases still amplified a short fragment of DNA with an efficacy that is comparable to that of commercial polymerases. If the precise units of enzyme for both *Tth* polymerase and *Taq* polymerase were to be matched, it is possible *Tth* polymerase would outperform the *Taq* polymerases.

In the reaction performed to amplify *Tt*KJEA, the *T. thermophilus* DnaK operon, the overall activity of *Tth* polymerase decreased (Fig. 3.9B). Hardly any band is observed for *Tth*PolI$_N$ and although band intensity increased steadily from *Tth*PolI$_T$ to *Tth*PolI-sGFP$_S$ -amplified *Tt*KJEA, those of Pol$_A$ and Pol$_B$ is much brighter, indicating that *Taq* polymerase was more efficient in long-distance PCR. Among the five *Tth* polymerases, only those purified from chaperone-coexpressing strains – *Tth*PolI$_S$, *Tth*PolI$_E$ and *Tth*PolI-sGFP$_S$ – have a higher product yield. It is possible to conclude that *Ts*DnaK- mediated protein folding improved the enzymes' activities. *Ts*DnaK's efficiency as a molecular chaperone is comparable to that of *E. coli*-derived DnaK chaperone system available commercially. In the long-distance PCR performed to amplify the 6.6 kb p22Cyp153A6 plasmid, no product was obtained from recombinant *Tth* polymerase proteins and only a mixture of DNA fragments was obtained by using Pol$_A$ and Pol$_B$ (Fig. 3.9C).

Poor amplification of the longer DNA fragments by the *Tth p*olymerases could be related to their intrinsic processivity and thermostability, which are not conferred nor enhanced by molecular chaperones. The processivity of *Tth* polymerase and has been cited as 60 nucleotide/s (Foord & Rose, 1994) at 72°C; in addition, the half-life of recombinant *Tth*PolI is 20 min at 95°C while the half life of recombinant *Taq* pol is 40 min at the same temperature and has a processivity of 50-60 nucleotide/s at 72°C (Corless *et al.*, 2000; Foord & Rose, 1994). *Taq* polymerase is, therefore, more thermotolerant. For the PCR reactions performed, extension cycles were relatively long to accommodate the low processivity. Enzymes were subjected to extended thermal conditions and possible denaturing, loss of activity or both occurred in both *Taq* and *Tth* DNA polymerases. Either, recombinant *Tth* Polymerases have a low processivity due to the fact that these enzymes were unable to amplify p22Cyp153A6, or their shorter half life means that they are as processive as *Taq* polymerase but they lose activity before elongating a significant amount of DNA, in comparison with *Taq* polymerase. This might account for the decrease in band intensity observed for *Tt*KJE amplicons. In addition, Pol$_A$ and Pol$_B$ also seem to have poor affinity for templates longer than 3.5 kb, resulting in DNA fragments of variable size but the

recombinant *Tth* polymerase only yielded single, discreet bands of the template which shows a higher affinity for longer templates than for *Taq* polymerase.

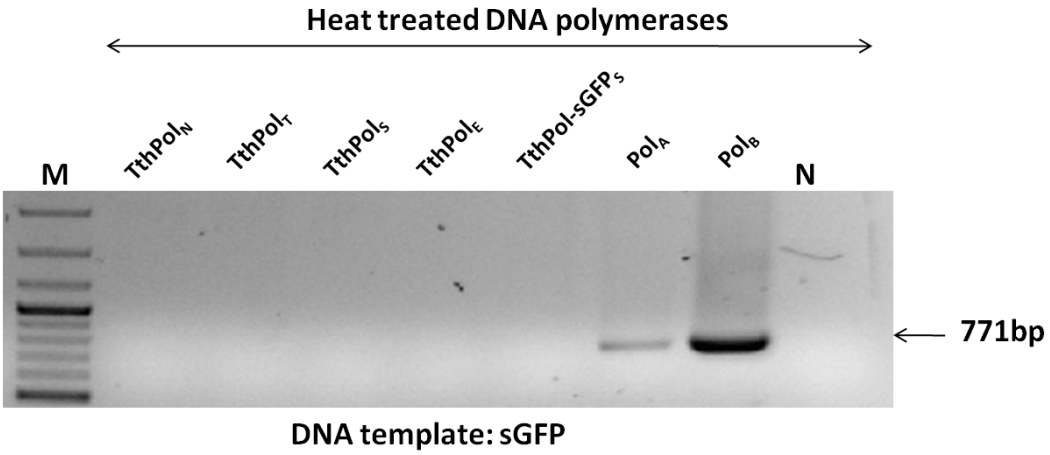

Figure 3.10. PCR amplification of sGFP by heat treated recombinant *Tth* polymerase. Polymerases were subjected to extended incubation at 95°C for 1 hour and immediately used to amplify sGFP. N represents the negative control for the PCR reaction, in which DNA polymerase has been omitted and M is the molecular weight marker.

Further evidence of differences in thermostability between recombinant *Tth* polymerase and *Taq* polymerase is shown in Fig 3.10. Here, prolonged incubation at 95°C renders all the previously active *Tth* polymerases inactive and sGFP is not amplified. However, Pol_A and Pol_B are still active enough to produce detectable quantities of sGFP by PCR. At the same time, the actual PCR cycle is shorter than that of elongating *Tt*KJEA or p22Cyp153A6 and as a result, does not go over the time in which *Taq* polymerase can resist thermal inactivation. In this case, *Ts*DnaK chaperone proteins would be unable to improve the activity of recombinant *Tth* polymerase beyond the protein's own intrinsic properties. Commercial variants of recombinant *Tth* polymerase exist and would have made a more reliable comparison of the chaperone effect; however, they were not available for this study.

3.4.3.2 Effect of thermophilic DnaK chaperone system on fidelity of recombinant *Tth* polymerase in PCR amplification

Sequencing was performed on *Tth*Pol-amplified sGFP to evaluate the enzyme's fidelity when amplifying short fragments.

*Tth*Pol has no 3'→5' exonuclease activity; such, is unable to correct misincorporated nucleotides; its error rate of 7.7 x 10^{-5} per base is comparable to that of *Taq* polymerase, which has been calculated to be around 1 x 10^{-5} per base (Rittié & Perbal, 2008). Therefore, short templates may be replicated with relatively high fidelity while longer templates will have misincorporated nucleotides. Fig 3.11 shows that the 771bp sGFP template was successfully amplified by all the DNA polymerases with no mismatched nucleotides, in comparison to the original sequence obtained from the plasmid.

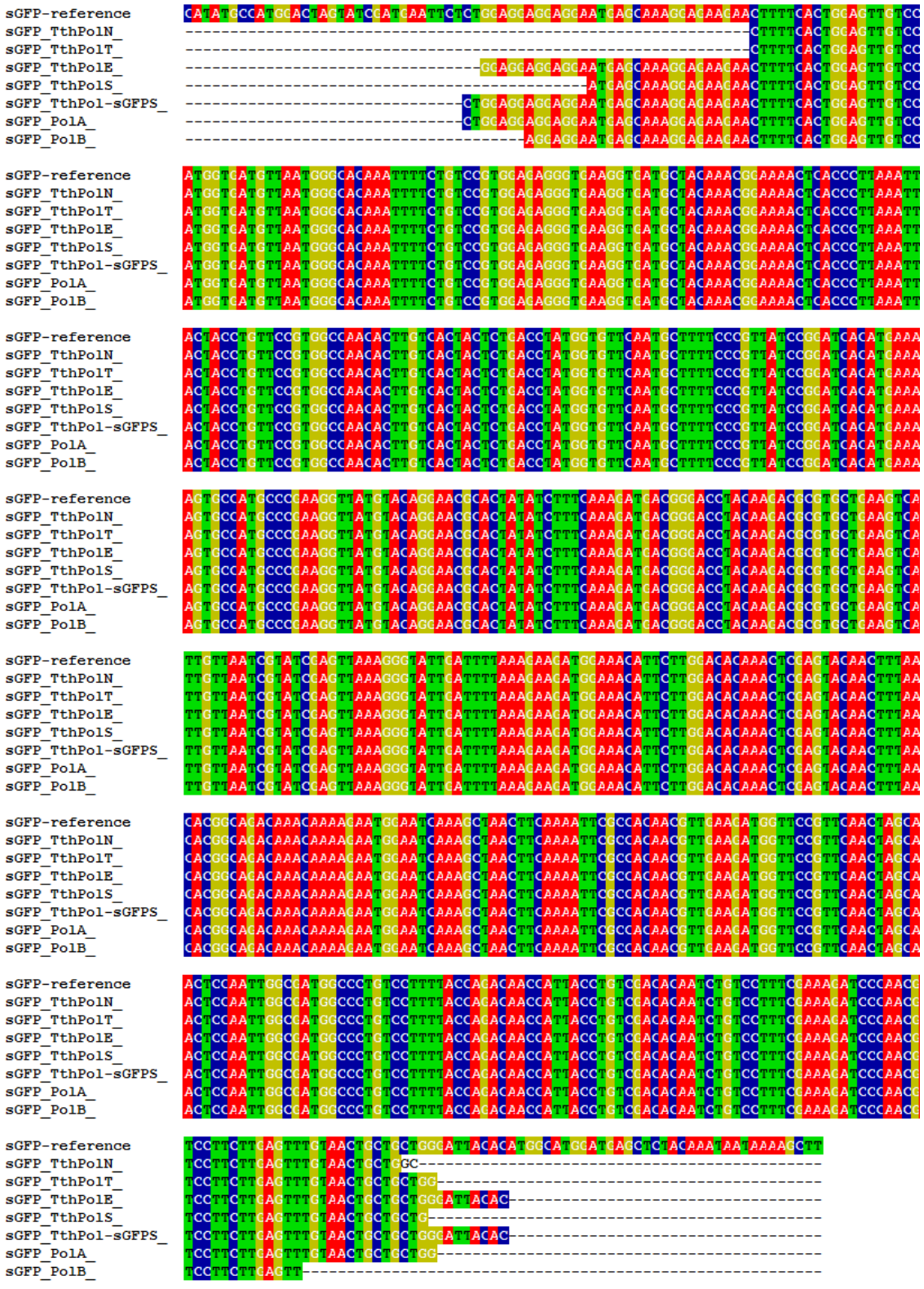

Figure 3.11. Multiple alignment showing sequencing results of sGFP amplicons amplified by recombinant *Tth* polymerases. The full reading frame of the original sGFP gene was used as the reference and commercial *Taq* polymerases, Pol_A and Pol_B were used for comparisons of the fidelity in nucleotide incorporation.

3.5 Concluding remarks

Expression of *T. scotoductus* DnaK chaperone system was shown to improve the folding of sGFP-coupled *Tth* DNA polymerase and thus its solubility. Consequently, this enabled purification of the protein fusion from cells. For the expression of uncoupled *Tth* polymerase, folding and solubility do not appear to be dependent on molecular chaperones and is expressed as a highly soluble and active protein in strains with no additional chaperones.

Recombinant *Tth* polymerase has a processivity that is comparable to commercial *Taq* polymerase and amplifies short double-stranded templates with high fidelity; however, it is not as thermostable and since all these properties are intrinsic features of the polymerase, they cannot be improved by molecular chaperones. Comparison of *Ts*DnaK chaperones to commercially available *Ec*DnaK chaperones showed no significant differences in expression but the *Ts*DnaK system had the superior folding effect as it was able to enhance folding of sGFP-coupled *Tth* Polymerase and allow its purification whereas the same protein fusion could not be purified from cells expressing the commercial *Ec*DnaK system. This demonstrates that heterologously expressed molecular chaperones could provide novel folding modulators to improve expression in *E. coli*; thermophilic organisms could be the source of such chaperones. Their chaperones are adapted to fold proteins at extreme temperatures but, as shown by *T. scotoductus*, they can also fold heterologous proteins at mesophilic temperatures.

The single chaperone co-expression vectors developed in this study are a novel approach to expressing heterologous proteins alongside molecular chaperones. With this system, double transformation is avoided, the selection pressure to maintain multiple plasmids within the cell is reduced and the problem of plasmid segregation, which results in competitive expression between multiple plasmids, is circumvented. Instead, the single co-expression vector maintains high expression of multiple proteins by using two strong but tightly-regulated promoters that can be individually controlled and fine-tuned by the researcher.

3.5.1 Future Work

Molecular chaperones play diverse roles in cellular metabolism and as a result, standardising the cell environment to measure one aspect of their physiology can be difficult. For this project, optimising research conditions such as standardising the volume of cells loaded for SDS-PAGE samples would show clear differences in protein expression profiles for the

different vectors. However, expression from these vectors occurs at different rates or not at all; therefore, different growth intervals are required for each to produce optimum cell volumes. Current results only show that cell mass, and thus, protein concentration, decreases with increasing protein expression. This is misleading in that it might be mistaken for a decrease in the efficiency of molecular chaperones to improve protein production.

Western blotting would be useful in detecting expression of *Thermus* DnaK chaperone proteins as it is highle sensitive, if antibodies could be raised for them. Though DnaK is visible on SDS-PAGE gels, DnaJ, GrpE and DafA cannot be seen and evidence of their contribution to DnaK in protein folding remains speculative if SDS-PAGE is relied on. Anti-Dnak antibodies were not available at the time this project could be developed in future. His-tags may also be added after the individual proteins in the operon to enable purification and detection; however, the protein-encoding genes overlap in the operon. Sequence manipulation during cloning is required to put the His-tags in frame with the genes even then, their purification may be unsuccessful. In this case, the use of anti-chaperone antibodies is a better option as extensive modification of the operon sequence is avoided.

Molecular chaperones tend to have high similarity, therefore, anti-chaperone antibodies from other organisms, such as *E. coli*, if they are readily available. Anti-chaperone antibodies from *E. coli* could be used only if the *E. coli* strains were chaperone deficient mutants, to prevent the antibodies from reacting with their native chaperones.

Other chaperone systems from *T. thermophilus* and *T. scotoductus* such as GroEL/GroES and ClpB may be substituted on the single co-expression vector; their ability to improve recombinant protein production in this system may then be evaluated.

Bibliography

Ami, D, Natalello, A, Schultz, T, Gatti-Lafranconi, P, Lotti, M, Doglia, SM, & de Marco, A (2009) Effects of recombinant protein misfolding and aggregation on bacterial membranes. *Biochim Biophys Acta* **1794**: 263-9.

Anfinsen, CB (1973) Principles that govern the folding of protein chains. *Science* **181**: 223-30.

Anfinsen, CB, Haber, E, Sela, M, & White, F. H (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U. S. A.* **47**: 1309-14.

Arakawa, T, Jongsareejlt, B, Tatsumi, Y *et al.*, (1996) Application of N-terminally truncated DNA polymerase from *Thermus thermophilus* (Δ*Tth* polymerase) to DNA sequencing and polymerase chain reactions : Comparative study of Δ*Tth* and wild-type *Tth* polymerases. *DNA Res* **3**: 87-92.

Arezi, B (2003) Amplification efficiency of thermostable DNA polymerases. *Anal Biochem* **321**: 226-235.

Baker, D (2000) A surprising simplicity to protein folding. *Nature* **405**: 39-42.

Balkwill, DL, Kieft, TL, Tsukuda, T, Kostandarithes, HM, Onstott, TC, Macnaughton, S, Bownas, J & Fredrickson, JK (2004) Identification of iron-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles* **8**: 37-44.

Banach, M, Stapor, K & Roterman, I (2009) Chaperonin structure - the large multi-subunit protein complex. *Int J Mol Sci* **10**: 844-61.

Baneyx, F & Mujacic, M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol* **22**: 1399-408.

Bann, JG, Pinkner, JS, Frieden, C & Hultgren, SJ (2004) Catalysis of protein folding by chaperones in pathogenic bacteria. *Proc Natl Acad Sci U. S. A.* **101**: 17389-93.

Baram, D, Pyetan, E, Sittner, A, Auerbach-Nevo, T, Bashan, A & Yonath, A (2005) Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc Natl Acad Sci U. S. A.* **102**: 12017-22.

Barnett, ME, Zolkiewska, A & Zolkiewski, M (2000) Structure and activity of ClpB from *Escherichia coli*. Role of the amino-and -carboxyl-terminal domains. *J Biol Chem* **275**: 37565-71.

Becker, J & Craig, E (1994) Heat-shock proteins as molecular chaperones. *Eur J Biochem* **219**: 11-23.

Beffa, T, Blanc, M, Lyon, P-F, Vogt, G, Marchiani, M, Fischer, JL, Aragno, M (1996) Isolation of *Thermus* strains from hot composts (60 to 80°C). *Appl Environ Microbiol* **62**: 1723-27.

Benson, LM, Null, AP & Muddiman, DC (2003) Advantages of *Thermococcus kodakaraenis* (KOD) DNA Polymerase for PCR-mass spectrometry based analyses. *J Am Soc Mass Spectrom* **14**: 601-4.

Bentley, WE, Mirjalili, N, Andersen, DC, Havis, RH & Kompala, DS (1990) Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng* **35**: 668-81.

Bergeron, LM, Gomez, L, Whitehead, TA & Clark, DS (2009) Self-renaturing enzymes : Design of an enzyme-chaperone chimera as a new approach to enzyme stabilization. *Biotechnol Bioeng* **102**: 1316-22.

Betiku, E (2006) Molecular chaperones involved in heterologous protein folding in *Escherichia coli*. *Biotechnol Mol Biol Rev* **1**: 66-75.

Bigotti, MG & Clarke, AR (2008) Chaperonins: The hunt for the Group II mechanism. *Arch Biochem Biophys* **474**: 331-9.

Boshoff, A, Nicoll, WS, Hennessey *et al.*, (2004) Molecular chaperones in biology , medicine and protein biotechnology. *S Afr J Sci* **100**: 665-77.

Bösl, B, Grimminger, V, & Walter, S (2006) The molecular chaperone Hsp104 − A molecular machine for protein disaggregation. *J Struct Biol* **156**: 139-48.

Braig, K (1998) Chaperonins. *Curr Opin Struct Biol* **8**: 159-65.

Bukau, B & Horwich, AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**: 351-66.

Bukau, B, Weissman, J & Horwich, A (2006) Molecular chaperones and protein quality control. *Cell* **125**: 443-51.

Butz, JA, Niebauer, RT & Robinson, AS (2003) Co-expression of molecular chaperones does not improve the heterologous expression of mammalian G-protein coupled receptor expression in yeast. *Biotechnol. Bioeng.* **84**: 292-304.

Byrne, GA, Boland, CA, O' Connell, EP, & Meijer, WG (2008) Differential mRNA stability of the vapAICD operon of the facultative intracellular pathogen Rhodococcus equi. *FEMS Microbiology Letters* **280**: 89-94.

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

Carrió, MM & Villaverde, A (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Letters* **489**: 29-33.

Cass, LG, & Wilcox, G (1988) Novel activation of araC expression and a DNA site required for araC autoregulation in *Escherichia coli* B/r. *J Bacteriol* **170**: 4174-80.

Cava, F, de Pedro, MA, Blas-Galindo, E, Waldo, GS, Westblade, LF, & Berenguer, J (2008) Expression and use of superfolder green fluorescent protein at high temperatures *in vivo*: a tool to study extreme thermophile biology. *Environ Microbiol* **10**: 605-13.

Chaudhuri, TK, Farr, GW, Fenton, WA, Rospert, S, & Horwich, AL (2001) GroEL/GroES-mediated folding of a protein too large to be encapsulated. *Cell* **107**: 235-46.

Chen, Di, Yang, B, & Kuo, T (1992) One-step transformation of yeast in stationary phase. *Curr Genet* **21**: 83-4.

Cheng, MY, Hartl, FU, Martin, J, Pollock, RA, Kalousek, F, Neupert, W, Hallberg, EM, Hallberg, RL, & Horwich, AL (1989) Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* **337**: 620-5.

Corless, CE, Guiver, M, Borrow, R, Edward-Jones, V, Kaczmarski, EB, & Fox, AJ (2000) Contamination and Sensitivity Issues with a Real-Time Universal 16S rRNA PCR. *J Clin Microbiol* **38**: 1747-52.

Crick, F (1970) Central Dogma of Molecular Biology. *Nature* **227**: 561-63.

Csermely, P, Schnaider, T, Soti, C, Prohászka, Z, & Nardai, G (1998) The 90-kDa molecular chaperone family : structure , function , and clinical applications. A comprehensive review. *Pharmacol Ther* **79**: 129-68.

de Groot, NS, Sabate, R, & Ventura, S (2009) Amyloids in bacterial inclusion bodies. *Trends Biochem Sci* **34**: 408-16.

de Marco, A, & de Marco, V (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. *J Biotechnol* **109**: 45-52.

de Marco, A, Deuerling, E, Mogk, A, Tomoyasu, T, & Bukau, B (2007) Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol* **7**: 32-40.

de Marco, V, Stier, G, Blandin, S, & de Marco, A (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem Biophys Res Commun* **322**: 766-71.

de Marco, A, Vigh, L, Diamant, S, & Goloubinoff, P (2005) Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperon* **10**: 329-39.

Deuerling, Elke *et al.*, (2003) Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Mol Microbiol* **47**: 1317-28.

Dhar, A, Samiotakis, A, Ebbinghaus, S, Nienhaus, L, Homouz, D, Gruebele, M & Cheung, MS (2010) Structure , function , and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci U. S. A.* **107**: 17586-91.

Díez-Gil, C, Krabbenborg, S, García-Fruitós, E *et al.*, (2010) The nanoscale properties of bacterial inclusion bodies and their effect on mammalian cell proliferation. *Biomaterials* **31**: 5805-12.

Dinner, AR, Sali, A, Smith, LJ, Dobson, CM, & Karplus, M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **4**: 331-39.

Dobson, CM (2001) The structural basis of protein folding and its links with human disease. *Phil Trans R Soc Lond Ser B* **356**: 133-45.

Dobson, CM (2003) Protein folding and misfolding. *Nature* **426**: 884-90.

Dobson, CM (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**: 3-16.

Doyle, SM, Shorter, J, Zolkiewski, M, Hoskins, JR, Lindquist, S, & Wickner, S (2007) Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. *Nat Struct Mol Biol* **14**: 114-22.

Doyle, SM, & Wickner, S (2009) Hsp104 and ClpB: protein disaggregating machines. *Trends Biochem Sci* **34**: 40-8.

Ellis, RJ (2003) Protein Folding: Importance of the Anfinsen Cage. *Curr Biol* **13**: R881-83.

Engel, R, Westphal, AH, Huberts, DHEW, Nabuurs, SM, Lindhoud, S, Visser, AJWG, & van Mierlo, CPM (2008) Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem* **283**: 27383-94.

England, JL, & Pande, VS (2008) Potential for modulation of the hydrophobic effect inside chaperonins. *Biophys J* **95**: 3391-9.

Evans, MS, Sander, IM, & Clark, PL (2008) Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol* **383**: 683-92.

Fairbanks, G, Steck, TL, & Wallach, DF (1971) Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* **10**: 2606-17.

Fang, L, Jia, K-Z, Tang, Y-L, Ma, D-Y, Yu, M, & Hua, Z-C (2007) An improved strategy for high-level production of TEV protease in *Escherichia coli* and its purification and characterization. *Protein Expression Purif* **51**: 102-9.

Fink, AL (1998) Protein aggregation: folding aggregates , inclusion bodies and amyloid. *Fold Des* **3**: 9-23.

Fink, AL (1999) Chaperone-mediated protein folding. *Physiol Rev* **79**: 425-49.

Foord, OS, & Rose, EA (1994) Long-distance PCR. *Genome Res* **3**: S149-61.

Furutani, M, Iida, T, Yoshida, T, & Maruyama, T (1998) Group II chaperonin in a thermophilic methanogen, *Methanococcus thermolithotrophicus*. Chaperone activity and filament-forming ability. *J Biol Chem* **273**: 28399-407.

Genevaux, P, Keppel, F, Schwager, F, Langendijk-Genevaux, PS, Hartl, FU, & Georgopoulos, C (2004) *In vivo* analysis of the overlapping functions of DnaK and trigger factor. *EMBO rep* **5**: 195-200.

Georgiou, G, & Valax, P (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr Opin Biotechnol* **7**: 190-97.

Gershenson, A, & Gierasch, LM (2011) Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21**: 32-41.

Gething, MJ (1996) Molecular chaperones: clasping the prize. *Curr Biol* **6**: 1573-6.

Goldsmith, M, Kiss, C, Bradbury, ARM, & Tawfik, DS (2007) Avoiding and controlling double transformation artifacts. *Prot Eng Des Sel* **20**: 315-8.

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

Groemping, Y, Klostermeier, D, Herrmann, C, Veit, T, Seidel, R, & Reinstein, J (2001) Regulation of ATPase and chaperone cycle of DnaK from *Thermus thermophilus* by the nucleotide exchange factor GrpE. *J Mol Biol* **305**: 1173-83.

Guisbert, E, Herman, C, Lu, CZ, & Gross, CA (2004) A chaperone network controls the heat shock response in *E. coli*. *Gene Dev* **18**: 2812-21.

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

Haacke, A, Fendrich, G, Ramage, P, & Geiser, M (2009) Chaperone over-expression in *Escherichia coli*: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. *Protein Expression Purif* **64**: 185-93.

Han, M-J, Park, SJ, Park, TJ, & Lee, SY (2004) Roles and applications of small heat shock proteins in the production of recombinant proteins in *Escherichia coli*. *Biotechnol Bioeng* **88**: 426-36.

Hardesty, B, Tsalkova, T, & Kramer, G (1999) Co-translational folding. *Curr Opin Struct Biol* **9**: 111-14.

Hartl, FU, & Martin, J (1995) Molecular chaperones in cellular protein folding. *Curr Opin Struct Biol* **5**: 92-102.

Hartl, FU, & Hayer-Hartl, M (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nat Struct Mol Biol* **16**: 574-81.

Hartl, FU, & Hayer-Hartl, M (2002) Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **295**: 1852-58.

Haslberger, T, Bukau, B, & Mogk, A (2010) Towards a unifying mechanism for ClpB / Hsp104- mediated protein disaggregation and prion propagation. *Biochem Cell Biol* **88**: 63-75.

Hennessy, F, Boshoff, A, & Blatch, GL (2005) Rational mutagenesis of a 40 kDa heat shock protein from *Agrobacterium tumefaciens* identifies amino acid residues critical to its *in vivo* function. *Int J Biochem Cell Biol* **37**: 177-91.

Hennessy, F, Nicoll, WS, Zimmermann, R, Cheetham, ME, & Blatch, GL (2005) Not all J domains are created equal : Implications for the specificity of Hsp40 – Hsp70 interactions. *Protein Sci* **14**: 1697-1709.

Hesterkamp, T, & Bukau, B (1998) Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in *E . coli*. *EMBO J* **17**: 4818-28.

Hoffmann, A, Bukau, B, & Kramer, G (2010) Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta* **1803**: 650-61.

Hofman-Bang, J, Lange, M, de Macario, EC, Macario, AJL, & Ahring, BK (1999) The genes coding for the hsp70(dnaK) molecular chaperone machine occur in the moderate thermophilic archaeon *Methanosarcina thermophila* TM-1. *Gene* **238**: 387-95.

Horwich, AL, Weber-Ban, EU, & Finley, D (1999) Chaperone rings in protein folding and degradation. *Proc Natl Acad Sci U. S. A.* **96**: 11033-40.

Inoue, H, Nojima, H, & Okayama, H, (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-38.

Itzhaki, L, & Wolynes, P (2010) Nature and nurture in protein folding and binding. *Curr Opin Struct Biol* **20**: 1-2.

Jana, S, & Deb, JK (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl Microbiol Biotechnol* **67**: 289-98.

Jefferys, BR, Kelley, LA, & Sternberg, MJE (2010) Protein folding requires crowd control in a simulated cell. *J Mol Biol* **397**: 1329-38.

Jhamb, K, Jawed, A, & Sahoo, D (2008) Immobilized chaperones: A productive alternative to refolding of bacterial inclusion body proteins. *Process Biochem* **43**: 587-97.

Kedzierska, S, & Matuszewska, E (2001) The effect of co-overproduction of DnaK/DnaJ/GrpE and ClpB proteins on the removal of heat-aggregated proteins from *Escherichia coli* ΔclpB mutant cells - new insight into the role of Hsp70 in a functional cooperation with Hsp100. *FEMS Microbiol Lett* **204**: 355-60.

Khan, F, He, M, & Taussig, MJ, (2006) Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on Ni-nitrilotriacetic acid surfaces. *Anal Chem* **78**: 3072-9.

Kim, S, & Lee, SB (2008) Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. *Protein Expression Purif* **62**: 116-9.

Klostermeier, D, Seidel, R, & Reinstein, J (1999) The functional cycle and regulation of the *Thermus thermophilus* DnaK chaperone system. *J Mol Biol* **287**: 511-25.

Klug, G (1993) The role of mRNA degradation in the reguiated expression of bacteriai photosynthesis genes. *Mol Microbiol* **9**: 1-7.

Kolaj, O, Spada, S, Robin, S, & Wall, JG (2009) Use of folding modulators to improve heterologous protein production in *Escherichia coli*. *Microb Cell Fact* **8**: 9-26.

Kondo, A, Kohda, J, Endo, Y, Shiromizu, T, Kurokawa, Y, Nishihara, K, Yanagi, H, Yura, T, & Fukuda, H (2000) Improvement of productivity of active horseradish peroxidase in *Escherichia coli* by coexpression of Dsb proteins. *J Biosci Bioeng* **90**: 600-6.

Kopito, RR (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol* **10**: 524-530.

Kramer, G, Rutkowska, A, Wegrzyn, RD, *et al.*, (2004) Functional Dissection of *Escherichia coli* Trigger Factor : Unraveling the function of individual domains. *J Bacteriol* **186**: 3777-84.

Kramer, G, Ramachandiran, V, & Hardesty, B (2001) Cotranslational folding — omnia mea mecum porto ? *Int J Biochem Cell Biol* **33**: 541-53.

Kramer, G, Boehringer, D, Ban, N, & Bukau, B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16**: 589-97.

Kyratsous, CA, Silverstein, SJ, DeLong, CR, & Panagiotidis, CA (2009) Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*. *Gene* **440**: 9-15.

Labuschagne, M, & Albertyn, J (2007) Cloning of an epoxide hydrolase-encoding gene from *Rhodotorula mucilaginosa* and functional expression in *Yarrowia lipolytica*. *Yeast* **24**: 69-78.

Laemmli, UK, (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-85.

Lan, WS, Gu, JD, Zhang, JL, Shen, BC, Jiang, H, Mulchandani, A, Chen, W, & Qiao, CL (2006) Coexpression of two detoxifying pesticide-degrading enzymes in a genetically engineered bacterium. *Int Biodeterior Biodegrad* **58**: 70-6.

Lee, S, & Tsai, FTF (2005) Molecular chaperones in protein quality control. *J Biochem Mol Biol* **38**: 259-65.

Littlechild, JA, Guy, J, Connelly, S, Mallett, L, Waddell, S, Rye, CA, Line, K, & Isupov, M (2007) Natural methods of protein stabilization: thermostable biocatalysts. *Biochem Soc Trans* **35**: 1558-63.

Liu, Z, Tek, V, Akoev, V, & Zolkiewski, M (2002) Conserved amino acid residues within the amino-terminal domain of ClpB are essential for the chaperone activity. *J Mol Biol* **321**: 111-20.

Livak, KJ, & Schmittgen, TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* 25: 402-8. http://www.ncbi.nlm.nih.gov/pubmed/11846609 (Accessed July 12, 2010).

Mahin, A-A, Sugimoto, S, Higashi, C, Matsumoto, S, & Sonomoto, K (2010) Improvement of multiple-stress tolerance and lactic acid production in *Lactococcus lactis* NZ9000 under conditions of thermal stress by heterologous expression of *Escherichia coli* dnaK. *Appl Environ Microbiol* **76**: 4277-85.

Maier, R, Eckert, B, Scholz, C, Lilie, H, & Schmid, F-X (2003) Interaction of Trigger Factor with the ribosome. *J Mol Biol* **326**: 585-92.

Maier, T, Ferbitz, L, Deuerling, E, & Ban, N (2005) A cradle for new proteins : trigger factor at the ribosome. *Curr Opin Struct Biol* **15**: 204-12.

Makrides, SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* **60**: 512-38.

Marquardt, T, & Helenius, A (1992) Misfolding and aggregation of newly synthesized proteins in the endoplasmic reticulum. *J Cell Biology* **117**: 505-13.

Martin, J (1997) Molecular chaperones and mitochondrial protein folding. *J Bioenerg Biomembr* **29**: 35-43.

Martinez-Hackert, E, & Hendrickson, WA (2009) Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* **138**: 923-34.

Maurizi, MR, & Xia, D (2004) Protein binding and disruption by Clp/Hsp100 chaperones. *Structure* **12**: 175-83.

Mayer, MP, & Bukau, B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* **62**: 670-84.

Minton, AP (1992) Confinement as a determinant of macromolecular structure and reactivity. *Biophys J* **63**: 1090-100.

Minton, AP (2006) Quick guide macromolecular crowding. *Curr Biol* 16: 269-71.

Mittal, J, & Best, RB (2010) Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J* **98**: 315-20.

Motohashi, K, Yohda, M, Endo, I, & Yoshida, M (1996) A novel factor required for the assembly of the DnaK and DnaJ chaperones of *Thermus thermophilus*. *J Biol Chem* **271**: 17343-48.

Nakamura, A, Takumi, K, & Miki, K (2010) Crystal Structure of a thermophilic GrpE Protein : Insight into thermosensing function for the DnaK chaperone system. *J Mol Biol* **396**: 1000-11.

Niiranen, L, Espelid, S, Karlsen, CR, Mustonen, M, Paulsen, SM, Heikinheimo, P, & Willassen, NP (2007) Comparative expression study to increase the solubility of cold adapted *Vibrio* proteins in *Escherichia coli*. *Protein Expression Purif* **52**: 210-8.

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

Nishihara, K, Kanemori, M, Yanagi, H, & Yura, T (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl Environ Microbiol* **66**: 884-9.

Osipiuk, J, & Joachimiak, A (1997) Cloning , sequencing , and expression of dnaK-operon proteins from the thermophilic bacterium *Thermus thermophilus*. *Biochim Biophys Acta* **1353**: 253-65.

Patzelt, H, Rüdiger, S, Brehmer, D, *et al.*, (2001) Binding specificity of *Escherichia coli* trigger factor. *Proc Natl Acad Sci U. S. A.* **98**: 14244-9.

Popov, M, Petrov, S, Nacheva, G, Ivanov, I, & Reichl, U (2011) Effects of a recombinant gene expression on ColE1-like plasmid segregation in *Escherichia coli*. *BMC Biotechnol* **11**: 18-29.

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

Ranson, NA, White, HE, & Saibil, HR (1998) Chaperonins. *Biochem J* **333**: 233-42.

Ratelade, J, Miot, M-C, Johnson, E, Betton, J-M, Mazodier, P, & Benaroudj, N (2009) Production of recombinant proteins in the lon-deficient BL21 (DE3) strain of *Escherichia coli* in the absence of the DnaK chaperone. *Appl Environ Microbiol* **75**: 3803-7.

Rial, DV, & Ceccarelli, EA (2002) Removal of DnaK contamination during fusion protein purifications. *Protein Expression Purif* **25**: 503-7.

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

Rittié, L, & Perbal, B (2008) Enzymes used in molecular biology: a useful guide. *J Cell Commun Signal* **2**: 25-45.

Robin, S, Togashi, DM, Ryder, AG, & Wall, JG (2009) Trigger factor from the psychrophilic bacterium *Psychrobacter frigidicola* is a monomeric chaperone. *J Bacteriol* **191**: 1162-8.

Roodveldt, C, Aharoni, A, & Tawfik, DS (2005) Directed evolution of proteins for heterologous expression and stability. *Curr Opin Struct Biol* **15**: 50-6.

Rüttimann, C, Cotorás, M, Zaldívar, J, & Vicuña, R (1985) DNA polymerases from the extremely thermophilic bacterium *Thermus thermophilus* HB-8. *Eur J Biochem* **149**: 41-6.

Saibil, H (2000) Molecular chaperones: containers and surfaces for folding, stabilising or unfolding proteins. *Curr Opin Struct Biol* **10**: 251-8.

Saibil, HR (2008) Chaperone machines in action. *Curr Opin Struct Biol* **18**: 35-42.

Samanta, D, Das, A, Bhattacharya, A, Basu, A, Das, D, & DasGupta, C (2009) Mechanism of ribosome assisted protein folding: a new insight into rRNA functions. *Biochem Biophys Res Commun* **384**: 137-40.

Sambrook, J, Fritsch, EF, & Maniatis, T, (1989) Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, U. S. A.

Sato, S, Ikeuchi, M, & Nakamoto, H (2008) Expression and function of a groEL paralog in the thermophilic cyanobacterium *Thermosynechococcus elongatus* under heat and cold stress. *FEBS Lett* **582**: 3389-95.

Schlee, S, Groemping, Y, Herde, P, Seidel, R, & Reinstein, J (2001) The chaperone function of ClpB from *Thermus thermophilus* depends on allosteric interactions of its two ATP-binding sites. *J Mol Biol* **306**: 889-99.

Schlee, S, & Reinstein, J (2002) The DnaK / ClpB chaperone system from *Thermus thermophilus*. *Cell Mol Life Sci* **59**: 1598-1606.

Schlieker, C, Bukau, B, & Mogk, A (2002) Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. *J Biotechnol* **96**: 13-21.

Scholz, C, Stoller, G, Zarnt, T, Fischer, G, & Schmid, FX (1997) Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO J* **16**: 54-8.

Schrödel, A, Volz, J, & de Marco, A (2005) Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. *J Biotechnol* **120**: 2-10.

Scott, M, Gunderson, CW, Mateescu, EM, Zhang, Z, & Hwa, T (2010) Inter-dependence of cell growth and gene expression : origins and consequences. *Science* **330**: 1099-1102.

Sevastsyanovich, Y, Alfasi, S, Overton, T, Hall, R, Jones, J, Hewitt, C, & Cole, J (2009) Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. *FEMS Microbiol Lett* **1**: 86-94.

Shiau, AK, Harris, SF, Southworth, DR, & Agard, DA (2006) Structural Analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**: 329-40.

Shimamoto, T, Noguchi, K, Kuroda, M, Tsuda, M, & Tsuchiya, T (1994) Transcriptional attenuation and differential mRNA stability in the regulation of the *Escherichia coli* melibiose operon. *J Biochem* **115**: 1185-9.

Shonhai, A, Boshoff, A, & Blatch, GL (2005) *Plasmodium falciparum* heat shock protein 70 is able to suppress the thermosensitivity of an *Escherichia coli* DnaK mutant strain. *Mol Genet Genomics* **274**: 70-8.

Singh, SM, & Panda, AK (2005) Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* **99**: 303-10.

Smith, BR, & Schleif, R (1978) Nucleotide sequence of the L-arabinose regulatory region of *Escherichia coli* K12. *J Biol Chem* **253**: 6931-33.

Smith, C (2005) Striving for purity : advances in protein purification. *Nat Methods* **2**: 71-7.

Soto, A, Allona, I, Collada, C, Guevara, M-A, Casado, R, Rodriguez-Cerezo, E, Aragoncillo, C, & Gomez, L (1999) Heterologous expression of a plant small heat-shock protein enhances *Escherichia coli* viability under heat and cold stress. *Plant Physiol* **120**: 521-8.

Stefani, M (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta* **1739**: 5-25.

Stephens, LL, Shonhai, A, & Blatch, GL (2011) Co-expression of the *Plasmodium* falciparum molecular chaperone, PfHsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, PfGCHI. *Protein Expression Purif* **77**: 159-65.

Stevens, RC (2000) Design of high-throughput methods of protein production for structural biology. *Structure* **8**: R177-85.

Stirling, PC, Lundin, VF, & Leroux, MR (2003) Getting a grip on non-native proteins. *EMBO Rep* **4**: 565-70.

Strub, C, Schlieker, C, Bukau, B, & Mogk, A (2003) Poly-L-lysine enhances the protein disaggregation activity of ClpB. *FEBS Lett* **553**: 125-30.

Sugimoto, S, Higashi, C, Saruwatari, K, Nakayama, J, & Sonomoto, K (2007) A gram-negative characteristic segment in *Escherichia coli* DnaK is essential for the ATP-dependent cooperative function with the co-chaperones DnaJ and GrpE. *FEBS Lett* **581**: 2993-9.

Sugimoto, S, Nakayama, J, Fukuda, D, Sonezaki, S, Watanabe, M, Tosukhowong, A, & Sonomoto, K (2003) Effect of heterologous expression of molecular chaperone DnaK from *Tetragenococcus halophilus* on salinity adaptation of *Escherichia coli*. *J Biosc Bioeng* **9**6: 129-33.

Sugimoto, S, Saruwatari, K, Higashi, C, & Sonomoto, K (2008) The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK – DnaJ – GrpE chaperone system and for cell division. *Microbiology* **154**: 1876-85.

Sørensen, HP, & Mortensen, KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* **115**: 113-28.

Takara (2009) Chaperone Plasmid Set Manual. **1-8**.

Tek, V, & Zolkiewski, M (2002) Stability and interactions of the amino-terminal domain of ClpB from *Escherichia coli*. *Protein Sci* **11**: 1192-8.

Terpe, K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60**: 523-33.

Terpe, K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72**: 211-22.

Thomas, JG, & Baneyx, F (2000) ClpB and HtpG facilitate *de novo* protein folding in stressed EsIcherichia cIoli cells. *Mol Microbiol* **36**: 1360-70.

Thomas, JG, & Baneyx, F (1998) Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *in vivo*. *J Bacteriol* **180**: 5165-172.

Tolia, NH, & Joshua-Tor, L (2006) Strategies for protein coexpression in *Escherichia coli*. *Nat Methods* **3**: 55-64.

Tomoyasu, T, Mogk, A, Langen, H, Goloubinoff, P, & Bukau, B (2001) Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol Microbiol* **40**: 397-413.

Tosco, A, Birolo, L, Madonna, S, Lolli, G, Sannia, G, & Marino, G, (2003) GroEL from the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC 125 : molecular characterization and gene cloning. *Extremophiles* **7**: 17-28.

Vabulas, RM, Raychaudhuri, S, Hayer-Hartl, M, & Hartl, FU (2010) Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harb Perspect Biol* **2**: 1-18.

Valent, QA, Kendall, DA, High, S, Kusters, R, Oudega, B, & Luirink, J (1995) Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *EMBO J* **14**: 5494-505.

Vendruscolo, M, & Paci, E (2003) Protein folding: bringing theory and experiment closer together. *Curr Opin Struct Biol* **13**: 82-7.

Ventura, S, & Villaverde, A (2006) Protein quality in bacterial inclusion bodies. *Trends Biotechnol* **24**: 179-185.

Villaverde, A, & Carrió, MM (2003) Protein aggregation in recombinant bacteria : biological role of inclusion bodies. *Biotechnol Lett* **25**: 1385-95.

Voedisch, B, & Thie, H, (2010) Size exclusion chromatography. From: Antibody engineering. Kontermann, R, & Dübel, S (Editors). Chapter **38**: 607-12. Springer Berlin Heidelberg, Germany.

Walter, S, & Buchner, J (2002) Molecular chaperones - cellular machines for protein folding. *Angew Chem* **41**: 1098-113.

Wang, J, Zhang, K, Lu, H, & Wang, E (2005) Quantifying kinetic paths of protein folding. *Biophys J* **89**: 1612-20.

Watanabe, K, Iwashiro, T, & Suzuki, Y (2000) Features of DnaK operon genes of the obligate thermophile *Bacillus thermoglucosidasius* KP1006. *Anton Leeuw Int J G* **77**: 241-50.

Watanabe, YH, Motohashi, K, Taguchi, H, & Yoshida, M (2000) Heat-inactivated proteins managed by DnaKJ-GrpE-ClpB chaperones are released as a chaperonin-recognizable non-native form. *J Biol Chem* **275**: 12388-92.

Wickner, S, Maurizi, MR, & Gottesman, S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**: 1888-93.

Widersten, M (1998) Heterologous expression in *Escherichia coli* of soluble active-site random mutants of haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 by coexpression of molecular chaperonins GroEL/ES. *Protein Expression Purif* **13**: 389-95.

Yochem, J, Uchida, H, Sunshine, M, Saito, H, Georgopoulos, CP, & Feiss, M (1978) Genetic analysis of two genes, DnaJ and DnaK, necessary for *Escherichia coli* and bacteriophage lambda DNA replication. *Mol Gen Genet* 164: 9-14.

Zhang, G, & Ignatova, Z (2010) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr Opin Struct Biol* **21**: 25-31.

Zhou, H-X (2008) Protein folding in confined and crowded environments. *Arch Biochem Biophys* **469**: 76-82.

Zimmerman, SB, & Harrison, B (1987) Macromolecular crowding increases binding of DNA polymerase to DNA: an adaptive effect. *Proc Natl Acad Sci U. S. A.* **84**: 1871-5.

Summary

Molecular chaperones are proteins which enable other protein molecules to fold to their native conformation and this property has been widely used to improve the solubility of proteins expressed in *Escherichia coli*. The effect of co-expressing of heterologous, thermophilic DnaK chaperones alongside with *Thermus thermophilus* DNA polymerase in *E. coli* was investigated in this study. A novel approach of co-expressing these proteins was also attempted.

To construct the plasmid vectors, high-copy expression vectors pET22 and pET28 commercial plasmid were used to provide the backbone for the new vectors. The KJEA operon, encoding DnaK chaperone, DnaJ co-chaperone, GrpE nucleotide exchange factor and the DnaK/DnaJ assembly factor, DafA, was amplified from *Thermus scotoductus* SA-01 and *T. thermophilus* HB8 by PCR. Similarly, the arabinose-inducible promoter P_{BAD} and its regulator protein-encoding gene, AraC, were amplified from the pBAD commercial vector. All fragments were subcloned into pGEM®-T easy before cloning them into the pET vectors. P_{BAD} was first ligated to either *Ts*KJEA or *Tt*KJEA to make a promoter-DnaK fusion gene that was then subcloned into pGEM®-T easy and later cloned into pET22 and pET28, along with AraC to yield four recombinant vectors, p22*Ts*K, p22*Tt*K, p28*Ts*K and p28*Tt*K.

Induction of P_{BAD} in these vectors with 5 mg ml^{-1} of L-arabinose resulted in expression of DnaK chaperone proteins from only p22*Ts*K and p28tsK, which express *T. scotoductus* DnaK proteins. The problem in p22*Tt*K and p28*Tt*K has been attributed to non-expression of AraC protein due to the long distance between AraC and its promoter region which lies in the P_{BAD} region, and is inverted in these vectors; however, this has yet to be investigated.

The *T. thermophilus* DNA polymerase, *Tth*PolI was cloned into the MCS of the chaperone expression vectors. The same polymerase was also fused to a superfolder green fluorescent protein, sGFP, and cloned into the chaperone vectors. Only the pET22-series of chaperone vectors were used as cloning into the pET28 line was unsuccessful. Expression of these two proteins was initiated by induction of the T7/lac promoter with 1mM IPTG. A commercial plasmid, pKJE7, encoding *E. coli* DnaK, DnaJ and GrpE was also used to co-express both proteins, for comparison. Expression was achieved in all DnaK-expressing vectors as well as non-expressing negative controls. Purification of uncoupled *Tth*PolI by affinity

chromatography was possible from cell expressing DnaK and the negative controls; however, only *Tth*Pol-sGFP protein expressed from p22*Ts*K was purified successfully, demonstrating the need for molecular chaperones when folding large proteins and the superiority, in folding activity, of the thermophilic DnaK chaperone system, in comparison to the *E. coli* system.

Activity assays were carried out to test the processivity, thermostability and fidelity of *Tth* polymerases purified from these vectors. Results show that *Tth* Polymerase amplifies short fragments, such as 771 bp sGFP, with high fidelity and is comparable to commercial *Taq* polymerase. However, *Tth* polymerase purified from strains co-expressing the commercial DnaK proteins had a poorer activity and yielded lower product than the other polymerases.

It was able to amplify 3518bp-*Tt*KJEA operon but the yield of product was lower than that obtained from the two commercial *Taq* polymerases. It was also unable to amplify a 6.6 kb plasmid, p22Cyp153A6, although even the commercial *Taq* polymerases only produced a mixture of DNA fragments, none of which were the correct size. This problem has been linked to the thermostability of *Tth* polymerase, which has a half-life of 20 min while that of *Taq* polymerase is 40 min. This means that while, *Tth* might be able to amplify large fragments, as a result of its low processivity, it is soon denatured from the high temperature cycles in long-distance PCR and substantial amplicons are not generated in time but *Taq* polymerase, though stable, has poor affinity for longer templates and dissociates before it can complete elongation of the template. Extended incubation of *Tth* polymerase at 95°C inactivates it and it is unable to amplify even the relatively short sGFP template. According to literature, the poor thermostability is a property of *Tth* polymerase and cannot be altered or improved by molecular chaperones.