Directed evolution designed to optimize the *in vivo* protein folding environment

by

Shu Quan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in the University of Michigan
2010

Doctoral Committee:

Professor James C. A. Bardwell, Chair
Professor David P. Ballou
Professor Robert A. Bender
Associate Professor Matthew R. Chapman
Associate Professor Bruce A. Palfey

Dedication

To my parents

Acknowledgements

I would like to express my appreciation to my professors, labmates, friends and family for their constant support during my doctoral studies.

First and foremost I offer my sincerest gratitude to my supervisor, **Prof. James Bardwell**, who has supported, encouraged, and guided me throughout my graduate studies with his knowledge and patience. From the first day I joined Jim's lab, I realized every experiment needs critical controls. I have always been inspired by his smart ideas and his persistence in solving difficult problems. Jim has been more than a great scientific mentor to me. He realized my difficulties as an international student, and he cheered me up whenever I was upset by my poor English. After years of Jim's training and encouragement I gradually came to feel myself to be a confident, critical and independent researcher.

I would also like to thank **Prof. Ursula Jakob** for her support. She always inspired me with her solid advice, and particularly with the help she offered in developing my capacity to give effective scientific presentations.

I also would like to thank members of my thesis committee, **Prof. David Ballou**, **Prof. Bob Bender**, **Prof. Bruce Palfey**, and **Prof. Matt Chapman**. Without their constant support and interest in my projects, this thesis would not have been written.

In my daily work I have been blessed with a friendly and cheerful group of fellow students and postdocs. First, I must thank **Dr. Jonathan Pan**, for being my first supervisor in the lab and for giving many thoughtful suggestions in managing my personal life. I have been aided for many years in running equipment and doing assays by **Dr. Guoping Ren**, **Dr. Tim Tapley**, and **Dr. Ajamaluddin Malik**. They have deep knowledge of bacterial genetics or biochemistry and they are always willing to share their ideas with me. Especially, I would like to thank **Dr. Linda Foit** for setting up the β-lactamase system and sharing her valuable plasmid constructs with me. I would like to

thank **Dr. Tim Tapley**, **Philipp Koldewey**, **Nadine Kirsch**, **Jennifer Pfizenmaier**, and **Stephan Hofmann** for their hard work in making such rapid progress in the Spy project which I initiated. Particularly, I would like to thank **Prof. Miroslaw Cygler** and his research group for solving the crystal structure of Spy and **Prof. Zhaohui Xu** for his valuable input to the analysis of Spy's structure. I want to thank **Irmhild Schneider**, **Anne-Katrin von Hacht**, **Dr. Annie Hiniker** and **Dr. Luis Masip** for assisting my other projects in the past. I also want to thank **Bharath Mamathambika**, **Antje Mueller**, **Thomas Meier**, and **Tsinatkeab Hailu** for their daily advice. It is a pleasure to thank all other members in the Bardwell lab and the Jakob lab for providing me with such a welcoming working environment.

I am grateful to have nice friends whom I can hang out with after work: **Dr. Hui Li**, **Dr. Yang Cao**, **Daniel Smith** and **Mario Blanco**.

Finally, I would like to thank my parents and my husband for their love and support during these last six years.

Preface

The aims of the study

Proteins are synthesized on the ribosomes in the cytoplasm. Upon the completion of synthesis, most proteins must fold into precise three-dimensional structures to be biologically active. Cytoplasmic proteins fold and stay in the cytoplasm; however, those proteins destined for the bacterial inner membrane, periplasm, cell envelope, and extracellular space, must generally maintain an at least partially unfolded conformation under most circumstances before they reach their ultimate compartment where they can fold properly. During these processes, many proteins require the assistance of folding modulators, in order to adopt native structures.

When a polypeptide is released into the periplasm, it needs to fold into its native conformation, and may also need to acquire post-translational modifications. Though the periplasmic folding environment is thought to be relatively chaperone-deficient compared to the cytoplasm, there are periplasmic chaperones, foldases, and proteases that contribute to the folding and modification of periplasmic proteins. They include 1) chaperones: Skp, FkpA, SurA, LolA, Spy, PapD and FimC; 2) peptidyl-prolyl cis-trans isomerases (PPIases): SurA, FkpA, PpiA and PpiD; 3) thio-disulfide oxidoreductases: DsbA, DsbB, DsbC, DsbD, DsbG, DsbE and CcmH. 4) proteases: DegP, Prc, Pet, and OmpT. Cells have also developed sophisticated quality-control systems to monitor protein folding. They respond to protein unfolding or misfolding, usually by the upregulation of genes encoding folding modulators and proteases.

Chaperones interact with, stabilize, or promote the net folding/refolding process of protein substrates. Foldases including PPIases and thiol-disulfide oxidoreductases are real protein folding catalysts which act to accelerate the rate limiting steps during substrate folding. Both molecular chaperones and foldases are actively involved in the protein folding/refolding processes, and overexpression of them is a very promising approach to

the optimization of the folding environment. Proteases, on the other hand, degrade abnormal extracytoplasmic proteins and maintain the homeostasis of the cell envelope. The coordination of these folding modulators ensures the proper function of the cell under physiological conditions and cell survival under stress conditions. Understanding the function, mechanism, regulation, and cooperation of these folding modulators would enlarge the scope of optimizing the extracytoplasmic folding environment, and improve heterologous protein production in the periplasm.

Through powerful chromosomal engineering and directed evolution technologies, new or improved functions have been gained for known periplasmic folding modulators. This raises at least four important questions: 1) what factors determine the function of a periplasmic folding modulator? 2) How can one evaluate the folding status of a protein *in vivo*? 3) How complete is the list of known periplasmic chaperones and folding catalysts? 4) Is there any possibility of optimizing the folding environment for specific heterologous proteins?

To gain some insight into the first question, I studied the functional evolution of members of the thiol-disulfide oxidoreductases. I mutated the characteristic active site, CXXC, in a strongly reducing oxidoreductase, thioredoxin, and forced the bacteria to select for mutants that act as oxidizing oxidoreductases. I found that the CXXC motif is important in its role of governing many of the properties of members of the thioredoxin superfamily. This is a good example of the way that nature can use a single particular scaffold to evolve several different functions of different enzymes.

The success in investigating the functional relationship of thiol-disulfide oxidoreductases inspired me to further explore the folding environment of bacterial periplasm. Apparently, nature has been proven to be very powerful in evolving mutations for novel functions. Therefore, there is great potential to further optimize heterologous protein production in the periplasm and to look for novel folding modulators. To explore the second and the third question, I developed a dual selection system that links protein folding to antimicrobial resistance. Cells survive in the presence of higher concentration of two antimicrobials only when they express stable and well folded test proteins. By applying this "fold or die" principle regarding bacteria together with chromosomal

mutagenesis techniques, we selected for bacteria that enhanced their ability to fold the test protein. With this selection system, I was able to identify a new periplasmic chaperone called Spy. Our approach opened up a new route towards chaperone discovery and in custom tailoring the folding of heterologous proteins.

Chapter one of my thesis summarizes the extracytoplasmic protein quality control system. Firstly, I describe important periplasmic folding modulators, their functions, and their discovering process. I then introduce important members of the thiol-disulfide oxidoreductases family, especially pointing out features that might be important for their functions. Next, the periplamic stress responses to protein unfolding or misfolding conditions are discussed. Finally, the previously established β-lactamase selection system is introduced, which provides a means of measuring protein stability *in vivo*.

Chapter two is modified from my previous publication "The CXXC motif is more than a redox rheostat"[1] on the work of complementing DsbA with active site-altered thioredoxin. The detailed results, experiment procedures, and the significance of this work are discussed.

Chapter three is modified from the manuscript in review "Genetic selection designed to stabilize proteins uncovers a chaperone called Spy". It describes a dual selection system for finding new periplasmic folding modulators, and the identification of a novel chaperone Spy using this system. Functional characterization of Spy as a highly effective chaperone is also presented.

Finally, in Chapter four I discuss these important results, and the future directions, such as identifying the *in vivo* substrates of Spy, exploring its working mechanism based on its crystal structure, and custom tailoring the folding of other proteins, such as the aggregation prone MalE31, using our selection system.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D: two dimensional

AMS: 4-acetoamido-4'-maleimidylstilbene-2,2'-disulfonate

AP: alkaline phosphatase

Bla: β-lactamase

BPTI: bovine pancreatic trypsin inhibitor

Cd^R: cadmium resistance

EMS: ethyl methanesulfonate

GAPDH: glyceraldehyde-3-phosphate dehydrogenase

GCSF: granulocyte colony-stimulating factor

GFP: green fluorescent protein

GS linker: glycine-serine linker

HAMP: histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases

HPLC: high performance liquid chromatography

Im7: immunity protein 7

LDH: lactate dehydrogenase

MBP: maltose binding protein

MDH: malate dehydrogenase

MIC: minimal inhibitory concentration

OMP: outer membrane proteins

PenV: penicillin V

$PenV^R$: penicillin V resistance

PMSF: phenylmethylsulfonyl fluoride

PPIases: peptidyl-prolyl cis-trans isomerases

qRT-PCR: quantitative real-time reverse transcription polymerase chain reaction

SRP: signal recognition particle

uPA: urokinase (also called urokinase-type plasminogen activator)

vtPA: a truncated version of tissue plasminogen activator

Abstract

Proteins are often unstable. Although much is known about protein folding *in vitro*, how proteins fold and function in the cell is a fundamental important question. The Protein folding process is assisted by molecular chaperones and folding catalysts *in vivo*. Understanding how chaperones are regulated and how they function may provide new avenues for developing protein folding modulators. We used directed evolution which combines DNA manipulation and powerful selection procedures for beneficial mutations in proteins to specifically address these questions.

My work focused on two distinct, though closely related problems, both of which have to do with the directed evolution of the periplasmic folding environment of bacteria. My first set of experiments concerned the relationship between the CXXC active site and the functional properties of thiol-disulfide oxidoreductases. Thiol-disulfide oxidoreductases are involved in catalyzing disulfide bond formation, isomerization and reduction during protein folding. We selected for mutants in the CXXC motif of a reducing oxidoreductase, thioredoxin, that complement null mutants in the very oxidizing oxidoreductase, DsbA. We found that altering the CXXC motif affects not only the reduction potential of thioredoxin, but also the ability of the protein to interact with folding protein substrates and reoxidants. Furthermore, the CXXC motif also impacts the ability of thioredoxin to function as a disulfide isomerase. Our results indicate that the CXXC motif has the remarkable ability to confer a large number of very specific properties on thioredoxin related proteins, in addition to their usual roles of regulating redox potentials.

The second phase of my work sought to optimize the *in vivo* folding of proteins by linking folding to antibiotic resistance, thereby forcing bacteria to either effectively fold the selected proteins or perish. Here we were able to show that when *Escherichia coli* is challenged to fold a very unstable protein, it responds by overproducing a protein

called Spy, which increases the steady state level of unstable proteins up to nearly 700 fold. *In vitro* studies demonstrate that Spy functions as a very effective ATP-independent chaperone that suppresses protein aggregation and aids protein refolding. Our strategy opens up new routes for chaperone discovery and the custom tailoring of the *in vivo* folding environment. Spy forms thin flexible cradle-shape dimers, with an apolar concave surface. Spy is unlike the structure of any previously solved chaperone making Spy the prototypical member of a new class of small chaperones that facilitate protein refolding in the absence of energy cofactors. We propose a model for Spy action where it protects proteins *in vivo* by binding to the surface of unstable proteins coating them with a thin layer that inhibits proteolysis and/or aggregation.

Chapter 1

The protein folding environment in the periplasm of *Escherichia coli*

1.1 Protein folding *in vivo*

In 1973, Anfinsen raised the concept that all the information to govern the folding of a protein is contained in its amino acid sequence[2]. There is much evidence for this, as numerous proteins have been shown to fold spontaneously *in vitro*[3-5]. However, the folding process is more complicated in a living cell than in a test tube due to the complex nature of the cellular folding environment.

The cytoplasm is extremely crowded, where macromolecular concentrations can reach up to 400 grams per liter[6]. The bacterial periplasm has been shown to be three times more viscous than the cytoplasm[7], implying an even greater protein concentration. Environmental conditions such as increased temperature create problems in protein folding and solubility. Therefore, proper folding of proteins inside the protease rich cell in a timely fashion is a tremendous challenge. Indeed, the folding process is a multi-scale dynamic process, if one considers the synergies between protein synthesis, targeting, folding, modification, and degradation.

It was previously thought that small (less than 100 residues), single-domain host proteins may achieve their native conformation spontaneously owing to their fast folding kinetics[8], and that host proteins with large, multidomains or overexpressed recombinant proteins generally need the help of folding modulators[9-11]. More recent research suggested that all proteins may fold via one or more intermediate states that act as stepping stones to their native states (reviewed in Ref. [12]). Those folding intermediates, especially when kinetically stable, have the tendency to rapidly collapse into non-native conformations. The fact that partially folded proteins tend to aggregate more readily in a highly crowded cellular environment[13] may explain the need for folding modulators. Based on functionality and mode of action, those folding modulators can be divided into

several categories including 1) classical molecular chaperones, 2) foldases, 3) proteases, and 4) a group of highly specialized chaperones that provide the steric conformation necessary for folding[14, 15].

1.1.1 Protein folding modulators participate in protein quality control

Molecular chaperones are defined as "any protein which interacts, stabilizes, or helps a non-native protein to acquire its native conformation, but is not present in the final functional structure"[11]. They are not truly folding catalysts since they do not accelerate the rate-limiting steps in the folding pathway. Instead, molecular chaperones facilitate protein folding mainly by binding to folding intermediates or stress partially unfolded proteins to maintain these non-native species in soluble or translocation-competent states. Such binding is generally through the non-covalent hydrophobic interaction of regions on the chaperone and hydrophobic amino acid side chains on the substrate proteins. These side chains are normally buried in native proteins, but exposed in unfolded or partially folded polypeptides. Binding of chaperones to substrates should help to shield the hydrophobic regions on substrates from solvents or from each other. Therefore, premature aggregation or inappropriate arrangement of protein regions is reduced by the action of chaperones.

If a chaperone only does what is described above, it is called "a holding chaperone"[16]. In contrast, some "folding chaperones" promote the net folding/refolding of substrate proteins using conformational changes coupled with ATP hydrolysis, such as members from the Hsp60, Hsp70, and Hsp90 families (reviewed in Refs. [11, 17]). ATP-regulated substrate release allows folding to proceed; therefore, these "folding chaperones" optimize the efficiency of folding, in addition to preventing substrate aggregation along the folding pathway[11]. Due to the lack of ATP in the periplasm, periplasmic folding chaperones must use other mechanisms to couple substrate binding and release. A third category of chaperones is known as the "disaggregation chaperone". These chaperones are not involved in the folding process, but instead, use energy provided by ATP hydrolysis to actively solubilize protein aggregates so that the proteins can be refolded (reviewed in Ref. [18]). In most cases, they work in collaboration with folding chaperones.

On the other hand, the foldases are true folding catalysts since they accelerate the rate-limiting steps during substrate folding, such as disulfide bond formation and peptidyl-prolyl cis-trans isomerization. Both molecular chaperones and foldases are actively involved in the protein folding/refolding process, and their overexpression is very promising as an approach of optimizing the folding environment.

More recently, a class of chaperones that provide highly specific steric (structural) information to their target proteins has been identified (as reviewed in Ref [19]). These target proteins are usually impossible to fold into their native structure on a biologically-relevant time scale due to large kinetic barriers that slow the folding process. The folded states of these target proteins are usually less stable, or only marginally more stable, than their unfolded states. The steric chaperones seem to overcome high energy barriers and therefore accelerate folding. After removal of the steric chaperones, the folded target proteins are trapped in the folded states by high energy barriers. Thus it seems that these chaperones direct proteins to violate Anfinsen's rule, which states that the folded state of a protein contains the minimal free energy. Other possible exceptions to Anfinsen's rule are the intrinsically disordered proteins. Although these proteins do not have regular three dimensional structures, cells in fact tolerate and select for such proteins since they are usually associated with important cellular functions[20-22]. Examples for steric chaperones include bacterial protease-prodomains, lipase specific foldase (Lif), and fimbrial periplasmic chaperones such as PapD[19].

Finally, as another branch of the quality control system, cells have developed ways to proteolytically remove irreversibly damaged proteins. This maintains cellular homeostasis and allows the recycling of amino acids. Proteases are involved in these processes. Under certain circumstances, misfolded proteins form protease-resistant aggregates known as inclusion bodies, thus escaping the protein folding quality-control system.

It is noteworthy that the lines separating the various folding modulators are not obvious[23]. A foldase might have chaperone activity (such as protein disulfide isomerase DsbC[24]), a protease might switch to a chaperone under specific conditions (such as

DegP[25]), and a chaperone may perform the peptidyl-prolyl cis-trans isomerization function (such as SurA[26]).

1.1.2 Periplasmic folding modulators*

The periplasm is estimated to make up 5-16% of the volume of the entire cell[27-29]. The periplasmic environment is very different from the cytoplasmic environment. Due to the permeability of the outer membrane, which allows free passage of hydrophilic molecules smaller than about ~700Da[30]), the properties of the periplasm are largely dependent on the surrounding medium. As a consequence, the pH and ionic strength reflect the extracellular environment and are thus subject to fluctuation. In addition, the periplasm contains the peptidoglycan layer, which is involved in maintaining cell shape[31]. In contrast to the reducing, highly buffered, and energy-rich environment in the cytoplasm, the periplasm is an oxidizing, relatively unbuffered compartment that lacks any obviously energy source, such as ATP that is required for many cytoplasmic folding chaperones to work. Therefore, the periplasmic folding modulators must use different mechanisms to couple substrate binding and release.

Only a few periplasmic chaperones and folding catalysts have been identified so far. Based on the substrate specificity, periplasmic chaperones can be divided into specific and general chaperones. Specific chaperones include SurA and LolA, which help the folding of outer membrane proteins or lipoproteins, respectively. PapD and FimC and their related proteins are also specialized chaperones that function in pili biogenesis. Skp, FkpA, and HdeA are the only known periplasmic general chaperones. Skp probably has the broadest range of substrates, including both outer membrane proteins and periplasmic proteins[32], or recombinant proteins expressed in the periplasm[33-35]. FkpA was shown only to protect periplasmic proteins or recombinant proteins expressed in the periplasm but not the outer membrane proteins[36, 37]. In contrast to FkpA and Skp, which fulfill chaperone function under normal physiological conditions, HdeA mediates acid resistance in

* More detailed descriptions, as well as descriptions of cytoplasmic folding modulators can be found in Appendix.

pathogenic enteric bacteria[38, 39]. It prevents acid denatured proteins from aggregation and aids in their refolding when shifted to neutral pH[40].

Peptidyl prolyl cis/trans isomerases (PPIases) catalyze the cis-trans isomerization of peptide bonds with proline (Xaa-pro bond; Xaa, any amino acid before proline), which is a rate-limiting step in the folding process of many proteins[41]. SurA, PpiD, FkpA and PpiA are known periplasmic PPIases. It is noted that SurA and FkpA also perform chaperone functions, while PpiA and PpiD are solely peptidyl prolyl cis/trans isomerases.

Thiol-disulfide oxidoreductases catalyze disulfide bond formation and isomerization reactions. They include DsbA and DsbB for disulfide oxidation, and DsbC, DsbD, and DsbG for disulfide isomerization. They also include DsbE (CcmG) and CcmH involved in cytochrome biogenesis.

In the periplasm, the accumulation of misfolded protein is alleviated by more than 20 proteases/peptidases (a list of periplasmic proteases/peptidases is available at http://www.cf.ac.uk/biosi/staffinfo/ehrmann/tools/proteases/allproteases.html). Among these, DegP (HtrA, Do) plays a leading role in removing the misfolded proteins (reviewed in Refs. [42, 43]). DegP has been shown to mediate degradation of a number of misfolded cell envelope proteins *in vivo*[25, 44-46]. However, DegP does not cleave well-folded proteins[25]. Interestingly, DegP also has temperature-regulated chaperone activity[25].

1.1.3 Identification of periplasmic folding modulators

Surprisingly, only a few chaperones have been identified directly through their abilities to stabilize proteins or promote protein folding processes. These chaperones, such as PapD, are usually encoded in the same operon as the proteins they stabilize[47]. Therefore, identification of these specialized chaperones was relatively straightforward. The genes encoding chaperones have also been isolated because of mutational phenotypes. These phenotypes might not be obviously related to chaperone functions. For example, *surA* (for survival) was originally identified as a gene essential for survival in the stationary phase[48]. The gene *hdeA* was identified as a gene important for survival at low pH. It is generally accepted that low pH causes protein denaturation; however, the

chaperone role of HdeA remained undiscovered, even after the crystal structure of this protein had been solved. Instead, the identification of many chaperones was because of their increased expression at elevated growth temperatures[49]. Alternatively, proteins were suggested as having a chaperone-like activity when they were found to bind to or interact with specific proteins. For instance, Skp was found to bind to outer membrane proteins in a pull-down assay using the outer membrane porin, OmpF[50]. LolA was identified as a periplasmic carrier protein that facilitates the release of an outer membrane lipoprotein, Lpp, from the inner membrane[51].

Identification of periplasmic folding catalysts was historically much more straightforward than the identification of periplasmic chaperones. For one reason, they catalyze specific reactions, and deletion of important folding catalysts often leads to easily distinguishable phenotypes. Therefore, folding catalysts were identified through complementation experiments. For instance, DsbG was identified as its overexpression reduced the dithiothreitol sensitivity of strains lacking a functional disulfide bond formation system[52]. On the other hand, these folding catalysts have characteristic domains (PPIase domain or thioredoxin domain) that might be used to identify new members based on sequence homology or structural similarity. Examples include the identification of FkpA based on its sequence homology to the PPIase domain of macrophage infectivity potentiator (Mip) proteins of *Legionella pneumophila* and *Chlamydia trachomatis*[53].

Due to the difficulty of identifying chaperones directly based on their chaperone activities, people may wonder how complete the list of known chaperones is. One aim of my graduate study was to try to explore the possibility of identifying new periplasmic chaperones and folding catalysts. Previously, the lack of an easily detectable phenotype due to a chaperone deficiency has limited the direct uncovering of new chaperones. Our lab has set up a powerful selection system that links the folding status of an unstable test protein to a distinguishable phenotype: antibiotic resistance. With this system, we reasoned that new chaperones might be uncovered based on their ability to stabilize the test protein and therefore confer antibiotic resistance. A detailed description of this system is in section 1.3.

1.1.4 Extracytoplasmic stress responses

Any unfavorable growth condition that would lead to protein denaturation and accumulation of misfolded cell envelope proteins may activate extracytoplasmic stress responses. As a result, cells upregulate or downregulate the level of a series of proteins to combat stress. Actually, many of the proteins involved in protein quality control have been characterized from studies on the stress responses. *E. coli* has at least five (σ^E, Cpx, Bae, Psp and Rcs) extracytoplasmic stress response pathways that allow it to monitor extracytoplasmic homeostasis (reviewed in Ref. [54]). Each of these five stress response pathways monitors different stresses and regulates different sets of genes. There is, however, a degree of overlap between different response pathways. One feature that distinguishes the extracytoplasmic stress response systems from cytoplasmic responses is the individual localization of the sensor protein and the effector. When the stress is sensed, the signal must cross the cytoplasmic membrane to reach the transcription regulator.

1.1.4.1 Two-component signaling transduction systems: Cpx, Bae and Rcs

The basic two-component system involves a sensor kinase and response-regulator protein[55]. The sensor kinase directly phosphorylates its cognate regulator protein in response to a specific signal. Phosphorylation of the regulator modulates its activity.

The Cpx (conjugative pilus expression) pathway mainly responds to stress in the periplasm and is induced by elevated pH, altered inner membrane composition, and misfolding or overproduction of cell envelope proteins[56]. Its sensor protein, CpxA, is a transmembrane protein located in the inner membrane. CpxA senses the stress signal with its periplasmic domain, autophosphorylates at a conserved histidine, and then transfers this phosphate to a conserved aspartate of the cytoplasmic regulator CpxR (Figure 1.1). CpxR then activates genes responsible for envelope physiology, such as *degP*, *dsbA*, *ppiA* and *ppiD*. One small periplasmic protein, CpxP, is also upregulated. It serves as a negative modulator of the Cpx pathway by binding to the periplasmic sensing domain of CpxA. Under normal growth conditions, binding of CpxP to CpxA keeps CpxA

unphosphorylated. Under stress conditions, CpxP is cleaved by DegP, thus relieving inhibition of CpxA.

The Bae (bacteria adaptative response) pathway is involved in the detoxification in the periplasm of harmful compounds such as antibiotics, bile salts, and detergents. It only regulates a few genes; most of them encoding drug exporters, such as *mdtABCD* and *acrD*. One interesting gene of this regulon, *spy*, is very strongly upregulated at the mRNA level when Bae is active[57]. Spy has been previously known to massively overexpress during spheroplasting (a process that allows periplasmic components to leak out) and is also regulated by the Cpx pathway[57]. Similar to the Cpx pathway, the Bae system uses a histidine kinase (BaeS), and a transcription factor (BaeR) to mediate the stress response (Figure 1.1). In Chapter 3 I describe how I discovered Spy to be a novel periplasmic chaperone.

The Rcs (regulator of capsular synthesis) pathway senses envelope composition. It is activated by any perturbation to the peptidoglycan layer, and contributes to intrinsic antibiotic resistance[58]. The signal transduction cascade includes an outer membrane protein, RcsF, inner membrane proteins RcsC and RcsD, and cytoplasmic proteins RcsB and RcsA.

1.1.4.2 Other extracytoplasmic stress response: σ^E and Psp

The σ^E pathway is induced by heat shock, and mainly monitors outer membrane protein misfolding. Its activation is a proteolytic cascade that results in the destruction of the membrane-bound anti-sigma factor RseA and the sequential release of the transcription factor σ^E (reviewed in Ref. [59]). Under normal growth conditions, RseA binds to and inhibits σ^E with RseB stabilizing the interaction between RseA and σ^E. In the presence of stress, RseA is cleaved in a two-step process: first by DegS from the periplasmic side and then by RseP from the cytoplasmic side (Figure 1.1). DegS was proposed to sense protein misfolding signals in the periplasm. Cleavage of RseA leads to the release of its cytoplasmic domain with σ^E still bound. Then the cytoplasmic proteases ClpXP digest the cytoplasmic domain of RseA and releases σ^E. The transcription factor σ^E regulates a large number of genes encoding folding modulators, including *degP*, *skp*,

fkpA, *surA*, *dsbC* etc. σ^E regulon overlaps substantially with that of Cpx, indicating the partial functional redundancy of the two systems.

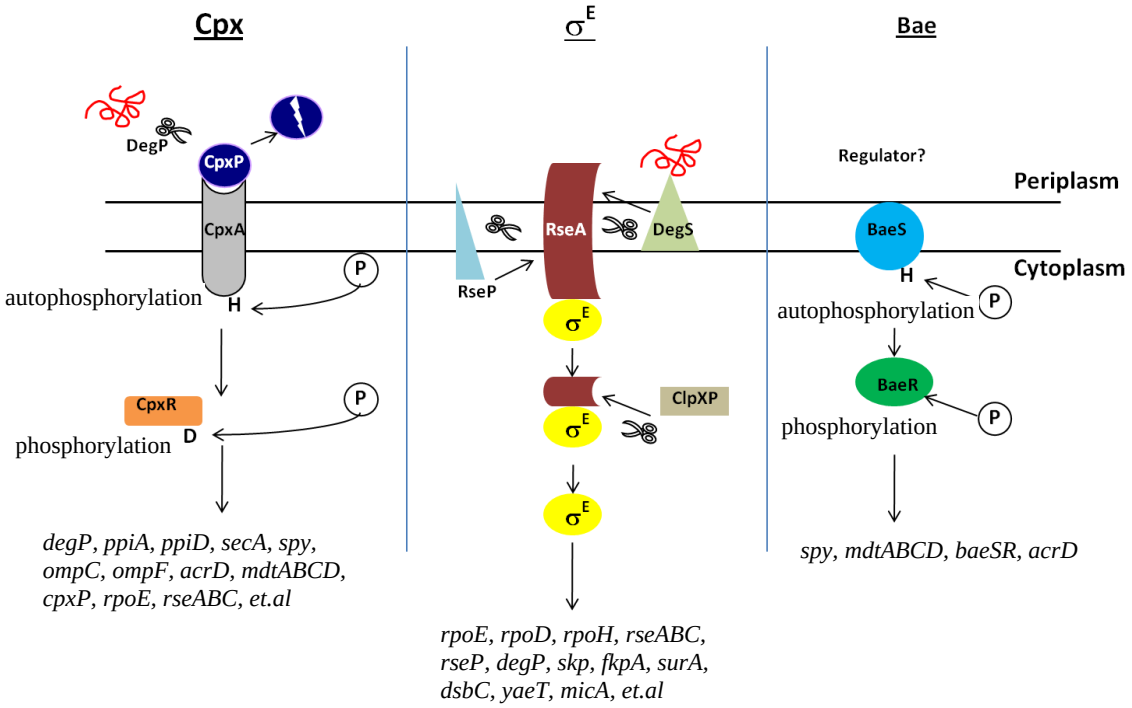


Figure 1.1 The Cpx, σ^E, and Bae stress response pathways
The activation process of each pathway is described in the text. Representative genes belonging to each regulon are listed at the bottom.

The Psp (phage shock protein) response monitors the integrity of the inner membrane and maintains the proton-motive force[60]. PspF is an activator of σ^{54}-containing RNA polymerase, and is normally inhibited by PspA. In the presence of stress signals, inner membrane protein PspB and PspC titrate away PspA, releasing PspF and initiating the Psp response.

1.2 Thiol-disulfide oxidoreductase: unique enzymes with a similar fold

One of the most important posttranslational modifications for exported proteins is the formation of disulfide bonds, which when present are important for the oxidative folding, activity, and stability of the proteins that contain them. Exported proteins show a strong bias for containing an even numbers of cysteines[61], implying that the majority of

cysteines are involved in disulfide bonds. Disulfide bond formation and isomerization result from thiol-disulfide exchange reactions between the substrate proteins and the thiol-disulfide oxidoreductases that are active in the periplasmic space (reviewed in Refs.[62, 63]). Most thiol-disulfide oxidoreductases belong to the thioredoxin superfamily. They contain a protein fold related to thioredoxin and usually have an active site motif that contains the sequence CXXC[64, 65]. These enzymes play an important role in protein folding and stability and are also responsible for bacterial pathogenicity[66].

Thiol-disulfide oxidoreductases are found in all living organisms. Thiol-disulfide oxidoreductases present in the periplasmic space act as thiol-disulfide oxidases or isomerases, while those found in the cytoplasm perform mainly reductive steps.

1.2.1 The thioredoxin fold and the CXXC active site

The basic tertiary structure of the thioredoxin superfamily that include a large number of thiol-disulfide oxidoreductases was first observed in the protein thioredoxin in 1975, and is called the thioredoxin fold[67]. The thioredoxin fold is a distinct structural motif consisting of four β strands and three flanking α helices[67] (Figure 1.2). Thiol-disulfide oxidoreductases having the thioredoxin fold belong to the thioredoxin superfamily. Although a basic scaffold of the thioredoxin superfamily, the thioredoxin fold is not restricted to the thioredoxin superfamily. Other *E. coli* proteins containing the thioredoxin fold include glutathione S-transferases, glutathione peroxidases, and alkyl hydroperoxidases (peroxiredoxins) [65, 68].

Members of the thioredoxin superfamily in *E. coli* include thioredoxin 1&2, glutaredoxin 1,2,3&4, DsbA, DsbC, DsbD, DsbG, DsbE, and a glutaredoxin-like protein, NrdH[69, 70]. These proteins have low sequence similarity and perform distinct functions (Figure 1.3 on page 14).

Another important feature that distinguishes members of the thioredoxin superfamily from other thioredoxin fold-containing proteins is the presence of an active site motif with a CXXC sequence, where X is any amino acid. The CXXC active site is located at the N-terminus of the α1 helix of the thioredoxin fold[68] (Figure 1.2). The active site contains the redox cysteine pairs and the XX dipeptide that is very important in

controlling the redox properties of the protein in which it is found[71, 72]. Several research groups have shown that it is possible to alter the redox properties of these oxidoreductases by changing the XX dipeptide in the CXXC motif. When changing the CXXC motif of one oxidoreductase to the sequence of another, the function of the original enzyme may be substantially altered. In some cases, the modified protein is even able to perform some of the activities of the protein from which the XX sequence was derived [73-75].

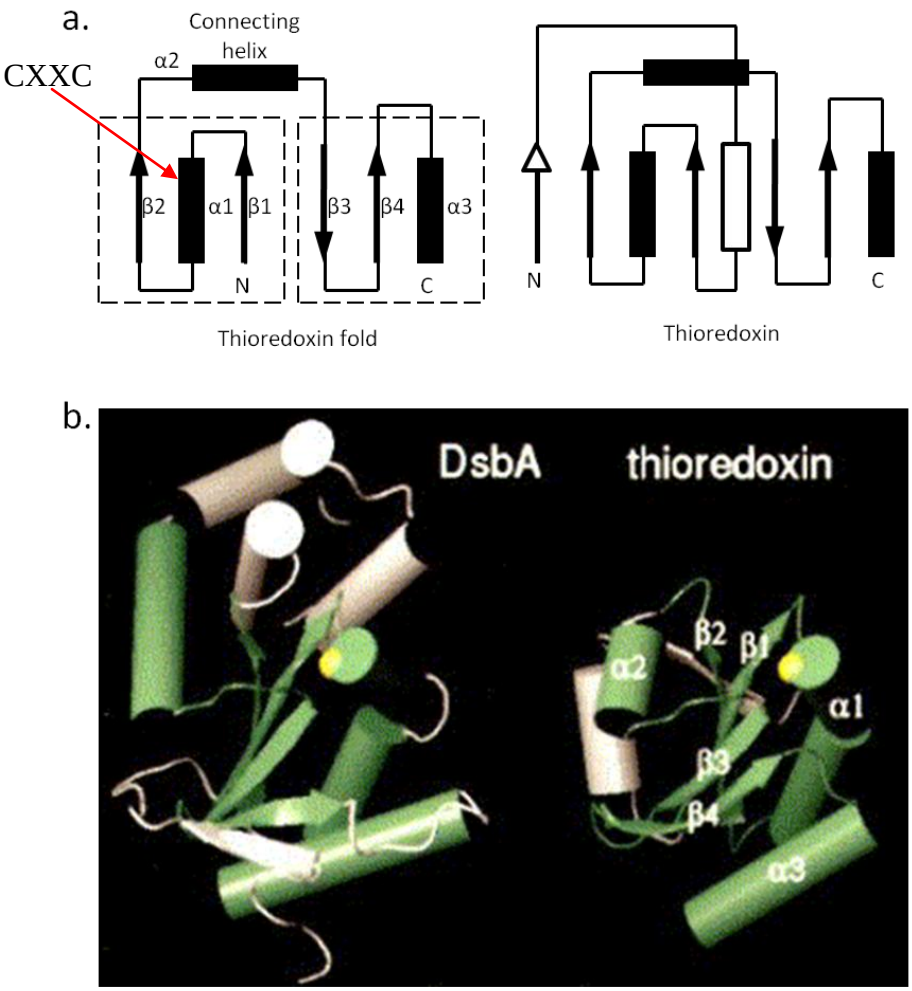


Figure 1.2 Architecture of the thioredoxin fold
The figure is modified from Fig 1& 3 in Ref. [68]. **a**, The β-sheet and α-helices of the thioredoxin fold can be subdivided into an N-terminal βαβ motif (boxed) and a C-terminal ββα motif (boxed), being connected by a loop of residues that incorporates a third helix. Thioredoxin contains one additional β-strand and one additional α-helix (white arrow and rectangle). **b**, structures of thioredoxin and DsbA. In DsbA, a second domain consisting of α-helices (white) is inserted into the thioredoxin motif (green). The first cysteine in the active site CXXC is colored yellow.

The effect of the XX dipeptide in the active site might be attributed to steric influences of their side chains. DsbA has a CPHC active site. The first residue of the XX dipeptide is proline, a residue that due to its ring structure comformationally restricts peptides to a greater extent than any of the other 18 amino acids with side chains. When the disulfide bond is formed in the active site of DsbA, the conformational restriction effect of proline might destabilize the disulfide bond, making it easy to break. If this is the case, it would contribute to the strong oxidizing activity of DsbA. Proline is also present at the active site of thioredoxin (CGPC), but is in the second position. Considering the large conformation space provided by the glycine residue, proline in the active site of thioredoxin might have only a marginal conformational effect[76]. In addition to these purely mechanical considerations, the XX dipeptide might also affect the electrostatic environment of the active site. For instance, the histidine residue in the active site of DsbA is known to stabilize the thiolate anion of Cys30[77]. The XX peptide has also been shown to influence the redox potential of the enzyme and the pK_a of the N-terminal cysteine in the CXXC motif. More extensive studies on the XX dipeptide suggested more of its functions, as described in Chapter 2.

1.2.2 Redox potential and pK_a

The redox potential and pK_a of the N-terminal active site cysteine are two important properties that influence the reactivity of thiol-disulfide oxidoreductases.

Redox potential measures the tendency of a chemical species to acquire electrons and thereby be reduced. It has long been proposed that redox potential affects the functions of proteins in the thioredoxin superfamily[71, 73, 75]. The intrinsic redox potentials of individual thiol-disulfide oxidoreductases can be quite different (Table 1.1). The oxidizing activity of a thiol-disulfide oxidoreductase is usually correlated with a high redox potential; while a reducing oxidoreductase often has a low redox potential. DsbA, with one of the highest redox potentials, is one of the most oxidizing oxidoreductases. It has been proposed that the XX dipeptide in the CXXC active site strongly influences the redox potential of the oxidoreductase[73, 75, 78].

Table 1.1 Redox potential of different oxidoreductases

Protein	Disulfide	Redox potential (mV)
DsbA	30-33	-122
DsbB	41-44	-69
	104-130	-186
DsbC	98-101	-140
DsbD α subunit	104-110	-229
DsbD γ subunit	461-464	-241
DsbG	126-129	-126
Thioredoxin	33-36	-270
DsbL	29-32	-95

Table is modified from Table. 1 in Ref. [76]. DsbL was recently identified to have an even higher redox potential than DsbA[79].

The pK_a value of a compound or a chemical group refers to the pH at which the compound/ chemical group is half ionized. The pK_a value of a cysteine determines the extent of ionization of the thiol group at any pH, and consequently, indicates the reactivity of the thiol group in thiol-disulfide exchange reactions. It also indicates the intrinsic chemical reactivity of the sulfur atom, including its reactivity when it is part of a disulfide bond[80]. Cys30 (the N-terminal cysteine in the CXXC active site) in DsbA has an extremely low pK_a value of 3.5, compared to the pK_a value of a normal cysteine of about 8.7[80]. It had been suggested that all of the following properties contribute to the exceptional oxidizing power of DsbA: the low pK_a value of Cys30, the high redox potential, and the energetic difference between the reduced and oxidized form of DsbA, with the oxidized form being less stable[81]. Importantly, the PH dipeptide in the active site of DsbA has been found to be critical in determining all of these important properties of DsbA[81]. Substitution of the PH dipeptide with other amino acid residues increased the pK_a of Cys30, decreased the energetic difference between the oxidized and reduced form of DsbA, changed the redox potential of DsbA, and thus made DsbA less oxidizing[81].

1.2.3 Representative thiol-disulfide oxidoreductases and their functions

Thiol-disulfide oxidoreductases are located in the periplasm, the inner membrane, and the cytoplasm of *E. coli.* The pathways for thiol-disulfide oxidation, reduction, and isomerization reactions are summarized in Figure 1.3.

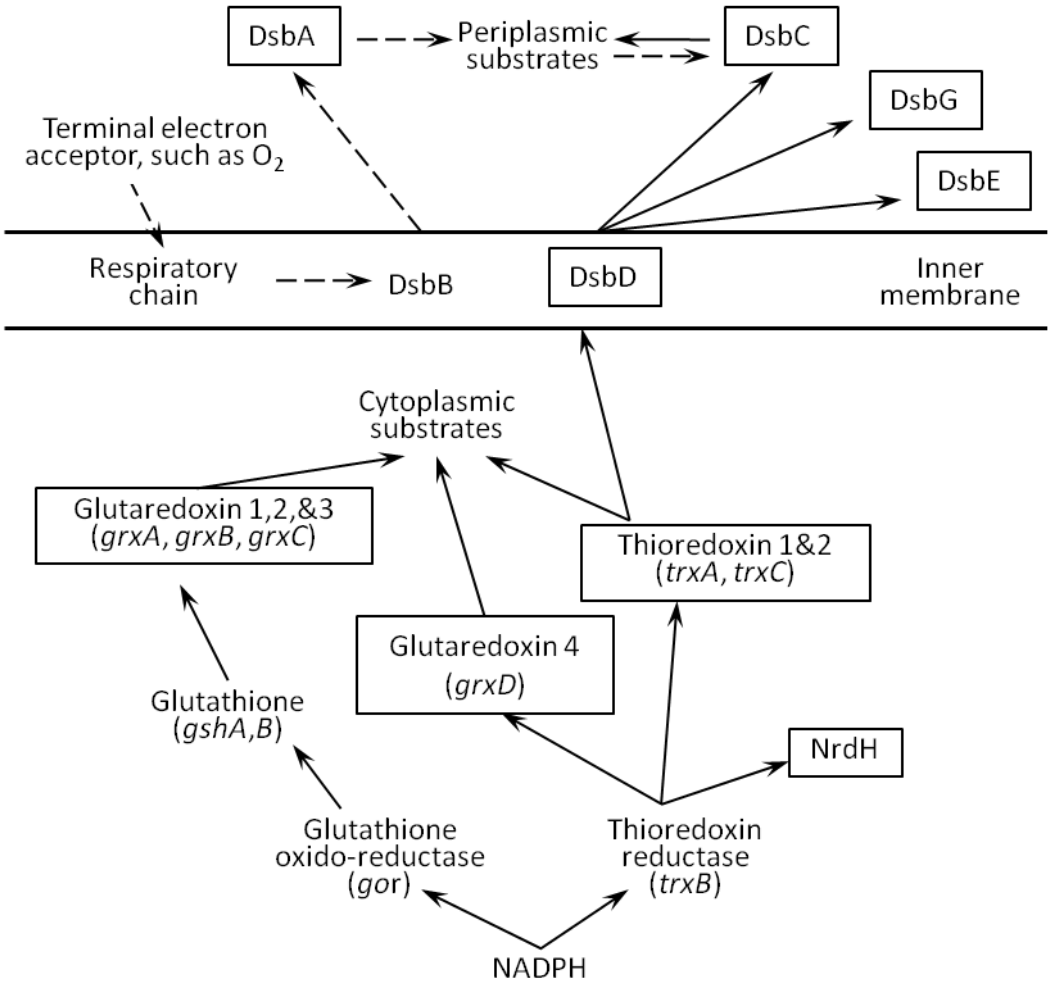


Figure 1.3 Major thiol-disulfide oxidoreductases in *E. coli*.
The figure was adapted from Fig.1 of Ref. [69] and information from Ref. [82]. Members of the thioredoxin superfamily are boxed. Members perform different functions: reduction (black arrow), oxidation (broken arrow), and disulfide bond isomerization (two arrows). The functions of DsbG, DsbE and NrdH are less well characterized. The thiol-disulfide oxidoreductase DsbB does not belong to the thioredoxin superfamily due to the lack of the thioredoxin fold.

1.2.3.1 Disulfide bond formation-DsbA and DsbB

The formation of disulfide bonds can occur spontaneously in the presence of oxygen, but this generally occurs relatively slowly. In the periplasm, disulfide bond formation is almost exclusively catalyzed by the thiol-disulfide oxidoreductase DsbA[83]. This 21 kDa protein donates its disulfide bond to substrate proteins via an unstable mixed disulfide intermediate. This mixed disulfide is resolved when another cysteine from the substrate attacks the mixed disulfide bond, resulting in the breakage of the intermediate and the net transfer of a disulfide bond from DsbA to the substrate. The oxidized form of DsbA is less stable than the reduced form. Therefore, it has the tendency to transfer its disulfide bond to substrate proteins[84]. In addition to its strong oxidizing thiol-disulfide oxidoreductase activity, DsbA has been reported to have chaperone-like activity[85].

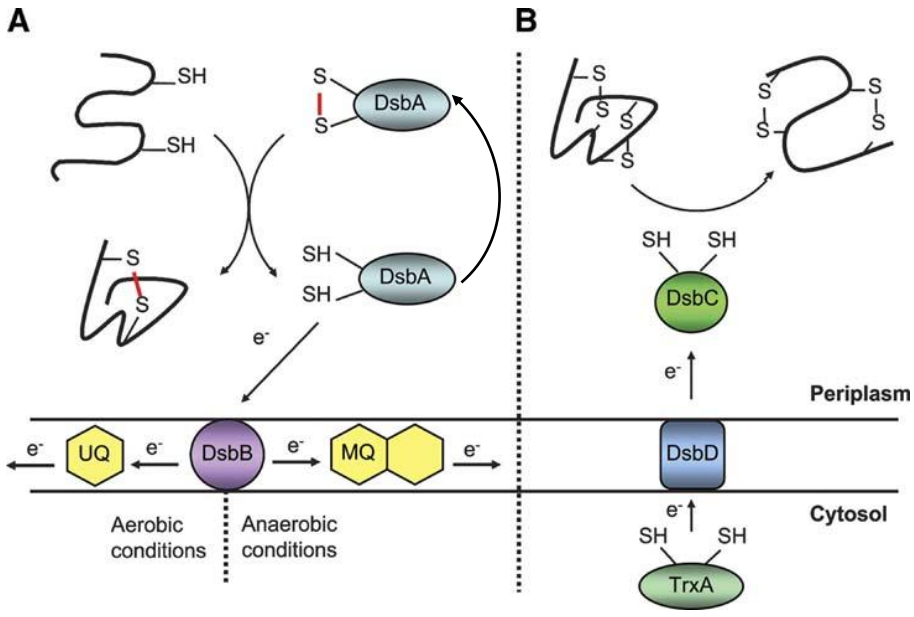


Figure 1.4 Mechanism of disulfide formation and isomerization in the bacterial periplasm.
The figure is modified from Fig.6 in Ref. [65]. **a**, DsbA resides in the periplasm of *E. coli* and introduces disulfide bonds into newly translocated polypeptides. For the system to be catalytic, reduced DsbA is reoxidized by the inner membrane protein DsbB. DsbB transfers electrons to the respiration chain with the final electron acceptor molecular oxygen. **b**, incorrect disulfide bonds are reshuffled by the periplasmic protein DsbC. DsbC is kept reduced by inner membrane protein DsbD. DsbD gains its reducing equivalents from the cytosolic protein thioredoxin (TrxA), which ultimately gains its reduced equivalents from NADPH (Figure 1.3).

Reduced DsbA is reoxidized by the inner membrane protein DsbB which transfers electrons to quinones in the respiratory chain (Figure 1.4 a). DsbB possesses four transmembrane helices and two periplasmic loops, which both contain catalytic cysteines[86, 87]. Cys41 and Cys44 are located in the smaller periplasmic loop within a CXXC motif. Cys104 and Cys133 reside on the bigger periplasmic loop, which contains regions that interact with DsbA. The reoxidation of DsbA occurs via a mixed disulfide bond between Cys104 of DsbB and Cys30 of DsbA. Cys33 of DsbA then attacks the mixed disulfide bond between DsbA and DsbB and resolves it[87]. Reoxidation of DsbA involves the relocation of disulfide bonds between the four catalytic cysteines of DsbB. Cys44 of DsbB is transiently reduced during this process, leading to the formation of Cys44-uniquinone charge-transfer complex[88, 89]. This results in a strong absorbance peak at 510 nm that can be used to monitor the interaction of DsbA and DsbB as described in Chapter 2.

1.2.3.2 Disulfide bond isomerization-DsbC, DsbD and DsbG

Since DsbA has a strongly oxidizing redox potential, it tends to form non-native disulfide bonds quite frequently, especially when the correct disulfide bond requires linking cysteines that are not consecutive in sequence. The misconnected disulfide bond is corrected by the disulfide isomerase DsbC (Figure 1.4 b). Isomerization is the process of reshuffling non-native disulfide bonds. This process involves sequential disulfide bond reduction and reoxidation. Reduced DsbC attacks the incorrect disulfide bond forming a mixed disulfide intermediate. This either leads to complete reduction of the incorrect disulfide bonds, or formation of a correct disulfide bond if the right pairing cysteine attacks and resolves the mixed disulfide[62]. DsbC is a V-shaped dimer with an uncharged cleft at the dimer interface that serves as a potential substrate binding site. Because of this structural feature, DsbC possibly has chaperone activity[24].

Organization of DsbC domains also showed striking similarity with the periplasmic chaperone FkpA[56]. Both enzymes are dimeric, with a C-terminal domain carrying the active site for disulfide isomerization or peptidyl-prolyl isomerization, a helix linker, and an N-terminal domain for dimerization and chaperone activity. Fusion of the dimerization domain and helix linker of FkpA with two DsbA proteins results in a V-shaped enzyme.

This new enzyme performs similarly to DsbC, suggesting that FkpA and DsbC might use similar mechanisms to couple chaperone activity and isomerase activity[90]. It also suggests that the functional evolution of proteins is the process of adding appendix structures to a more conserved structural scaffold, such as the thioredoxin domain or the PPIase domain.

To fulfill its function as an isomerase, DsbC is always kept in the reduced form by the inner membrane protein DsbD. DsbD gains electrons from the cytoplasmic protein thioredoxin, which ultimately acquires electrons from NADPH. DsbD also reduces DsbG, a protein that is probably also a disulfide isomerase. DsbG shares 24% sequence identity with DsbC, and is thought to be less effective in disulfide isomerization and have a narrower substrate range[91]. However, through laboratory evolution, the isomerase activity of DsbG could be substantially improved to resemble that of DsbC[92]. Recent studies have suggested that DsbG, together with DsbC, controls the global sulfenic acid content of the periplasm and protects thiols of single cysteine residues from oxidation to sulfenic acids[93].

1.2.3.3 Disulfide bond reduction: thioredoxins and glutaredoxins

Cytoplasmic proteins do not normally form disulfide bonds, although some proteins such as Hsp33 form transient disulfide bonds to regulate their functions under specific conditions[94]. Thiol groups in cytoplasmic proteins are kept reduced by two pathways: the thioredoxin pathway and the glutaredoxin/glutathione pathway. In total, *E. coli* has six thioredoxin or glutaredoxin proteins, namely thioredoxin 1, thioredoxin 2, glutaredoxin 1, glutaredoxin 2, glutaredoxin 3, and glutaredoxin 4[65]. Throughout my thesis, "thioredoxin" refers to thioredoxin 1, which is encoded by the *trxA* gene.

In the thioredoxin pathway, thioredoxin reductase uses the reducing equivalents from NADPH to maintain the reducing state of thioredoxin (Figure 1.3). Thioredoxin in turn can reduce disulfide bonds in various proteins, such as DsbD, peroxiredoxins, methionine sulfoxides, ribonucleotide reductase, and Hsp33 (reviewed in Ref. [95]). Thioredoxin therefore is involved in many important cellular reactions such as ribonucleotide reduction, sulfate assimilation, antioxidation, and so on[95, 96].

The glutathione/glutaredoxin system also uses NADPH in this case to reduce glutathione via glutathione oxidoreductase. Glutathione is then able to reduce the glutaredoxins (Figure 1.3). Like thioredoxins, glutaredoxins are involved in many cellular functions, and play an important role in redox signaling by regulating protein glutathionylation[97, 98].

1.2.4 Directed evolution of thiol-disulfide oxidoreductase

The thioredoxin fold in the example of thioredoxin itself is stable to up to 80˚C. Therefore, it is not surprising that during evolution many different enzymes have used this fold as a scaffold. Directed evolution that combines strong mutagenesis conditions and specific selection techniques allows us to mimic millions of years of evolution of the functions of enzymes within days. This technique, applied to members of the thioredoxin superfamily, has generated a large amount of valuable information for thiol-disulfide oxidoreductases[99-101]. The information gained from these experiments not only confirms the suitability of the thioredoxin scaffold as a basic structural and functional block, but also suggests the possibility of adding new functions to the existing functions of an enzyme.

For example, the substitution of the CGPC active site in thioredoxin by the CPHC active site as in DsbA makes thioredoxin a thiol-disulfide oxidase similar to DsbA[99]. Thioredoxin with its active site changed to CACC partially complements DsbA by the acquisition of a 2Fe–2S iron–sulfur cluster[100]. DsbC also partially complements DsbA if the regions in DsbC important for dimerization have been modified [101]. DsbG with a single mutation in its surface-exposed regions causes DsbG to resemble DsbC[92]. These examples suggest that the functions and substrate specificities of members from the thioredoxin superfamily can evolve through only a few amino acid changes.

In Chapter 2, we used the systematic approach of mutating the CXXC motif of thioredoxin to make it resemble DsbA. With this approach, we gained insight into the key features of this widely conserved motif.

1.3 Monitoring protein folding *in vivo*

Periplasmic expression in *E. coli* has long been used to produce biologically active recombinant proteins of pharmaceutical importance. However, the production of large and complex recombinant proteins, especially eukaryotic proteins, is often limited by poor folding or low yields. These problems might be attributed to different folding conditions or redox environment in the prokaryotic periplasm, compared with the endoplasmic reticulum (ER) of eukaryotic organisms. We wondered if it is possible to improve the folding environment of the periplasm by employing techniques of chromosomal mutagenesis. As a prerequisite we needed to set up a system that conveniently monitors protein folding *in vivo*. Most proteins, upon expression, do not confer a phenotype that can be adapted to a selection or screen method. Protein activity measurements, on the other hand, are often limited by strict requirement of protein purity and therefore may not be applicable for high-throughput assays using crude cell lysates.

1.3.1 The β-lactamase system

To circumvent the problems mentioned in the above section, our laboratory used a strategy to make a fusion protein of the test protein with a reporter protein, which conveys a selectable phenotype if it is well folded. The test protein is inserted into a permissive site of the reporter so that the folding of the test protein would greatly affect the folding status of the reporter protein. As a result, the phenotype conferred by the reporter would reflect the folding of the test protein.

One example of such a fusion protein previously developed in our laboratory is the β-lactamase system[102] (Figure 1.5). It is a powerful genetic system that directly links increased protein folding kinetics and thermodynamic stability to increased antibiotic resistance, establishing a quantitative *in vivo* measure of protein stability[102]. In this system, a guest protein is inserted via a glycine-serine (GS) linker into a permissive site in the TEM-1 β-lactamase. In the absence of efficient folding modulators, the unstable/ poorly folded guest protein is prone to degradation by periplasmic proteases. Proteolysis of the guest protein separates β-lactamase into two parts. The two parts of β-lactamase are ultimately degraded, leading to a decrease of the host's resistance towards β-lactam

antibiotics, such as penicillin V (PenV). On the other hand, well-folded guest proteins able to maintain a stable strucuture are protease resistant, leading to the assembly of a functional fusion protein that gives resistance to PenV. This system was therefore proposed to distinguish well-folded/stable proteins from poorly folded/unstable ones, by the means of conferring different levels of antibiotic resistance.

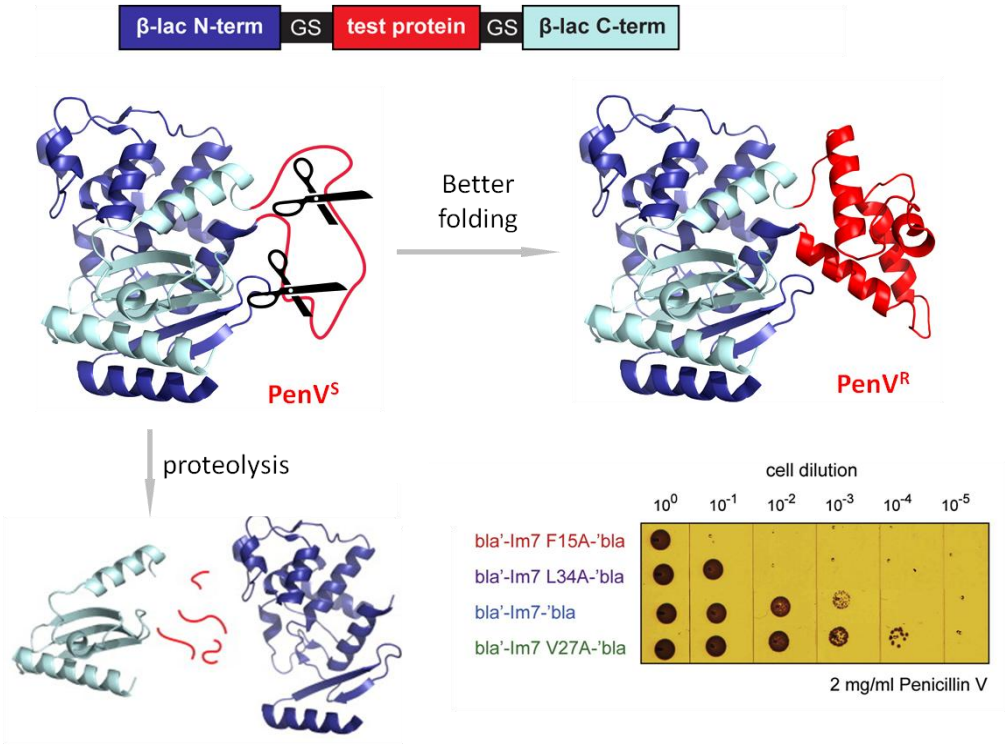


Figure 1.5 The β-lactamase system links the PenV resistance to the stability of the inserted protein

The figure is modified from Fig.1 and Fig.2 in ref. [102]. The level of penicillin resistance *in vivo* may be related to the folding status of the inserted protein. Poorly folded proteins are degraded by periplasmic proteases (represented by scissors), which leads to the irreversible separation of the two parts of β-lactamase and low level of PenV resistance.Well folded proteins are resistant to proteolysis, forming an active β-lactamase fusion and resulting high level of PenV resistance. $PenV^S$ = penicillin V sensitive. $PenV^R$ = penicillin V resistant. N-term: N-terminal part. C-term: C-terminal part. GS: the flexible glycine-serine linker. Right bottom corner shows a typical result from a spot titer experiment. Growth of the cells expressing different fusion proteins (such as bla'-Im7 F15A-bla', meaning Im7 F15A inserted into β-lactamase) was compared. F15A and L34A mutations destabilize Im7, making the variants fold worse than Im7 wild type; while V27A is a stabilizing mutation that improves the folding of Im7. Stabilization of Im7 leads to increased PenV resistance so that higher dilutions of the cells can grow on PenV plate, and vice versa.

This hypothesis was tested by insertion of three different guest proteins into β-lactamase: immunity protein 7 (Im7), a bacterial protein without cofactors or disulfides, granulocyte colony-stimulating factor (GCSF), a eukaryotic protein containing three disulfides, and maltose binding protein (MBP), a large bacterial periplasmic protein. Each of these three proteins has a well-studied *in vitro* folding pathway, and several destabilizing mutations are known[102]. It was clearly shown for each protein that the introduction of these destabilizing mutations led to an increased sensitivity to PenV (Figure 1.5 and Fig 3 in Ref. [102]). Moreover, for all of the three proteins including many of their mutants tested, a direct correlation between antibiotic resistance and the thermodynamic stabilities of the guest protein was found[102]. The highest antibiotic resistance always resulted from cells expressing β-lactamase being fused with the most stable test proteins. This suggests that the level of antibiotic resistance could be used as an indicator for the folding status of the guest protein. Therefore, one can theoretically select for cells containing a well-folded guest protein by simply plating cells on increasing concentrations of antibiotics. Indeed, by using the tripartite fusions between β-lactamase and unstable proteins, this system was successful in obtaining dramatically stabilized protein variants[102]. It is noted that the β-lactamase system provides a robust readout for protein folding/stability and a method for selecting stabilizing protein variants without requiring knowledge of the structure, function or activity of the test protein.

We hypothesized that this system might actually be able to select for host variants that alter the *in vivo* folding environment in ways that enhance protein folding or stability. If such variants could be obtained, we reasoned that their analysis might provide insight into the *in vivo* folding process. Chapter 3 describes the use of an enhanced version of this system to identify a new periplasmic chaperone called Spy. More applications of this system are discussed in Chapter 4.

1.3.2 Other systems generated to monitor protein folding *in vivo*

A few systems have been generated in other laboratories to qualitatively or quantitatively measure protein folding *in vivo* and to select for more stabilized protein variants. These systems include phage display methods and the green fluorescent protein (GFP) folding reporter system. One example of phage display methods is the Proside

(protein stability increased by directed evolution) system, which couples increased protease resistance of protein variants with the infectivity of a filamentous phage fd [103]. The folding reporter protein used in this system is the gene-3-protein (g3p) of the fd phage, which mediates the infectivity of the phage to *E. coli*. Test proteins or peptides are inserted between two of the three domains of the g3p protein[103]. The resulting chimeric protein is then subjected to proteolysis *in vitro* and selected for improved stability of the inserted part based on phage's infectivity. The method is, however, limited by the size of the inserted protein since it requires that the insertion does not abolish the assembly or infectivity of the phage. In the GFP folding reporter system, a test protein is fused to the N-terminus of GFP[104]. As a result, the folding and/or solubility of the upstream test protein affect the folding of GFP, and only correct folding of GFP leads to fluorescence. Through rounds of mutagenesis and screening for fluorescence, stabilized variants of test proteins can be identified. To solve the problems such as truncation artifacts (proteolytic destruction of the N-terminal test protein can relieve the interference of the test protein on the folding of GFP, and lead to false positive results), improved versions of GFP reporters have been developed[105]. These improved GFP reporters are circularly permuted so that insertions are now placed between two parts of GFP. The GFP reporter system and our β-lactamase system both provide robust measurement of protein stability *in vivo*. They differ in that the GFP system is a screening method which does not kill bacteria, while the β-lactamase system is a selection method which links to the survival of the bacteria.

1.4 Conclusion

E. coli has developed sophisticated systems both to assist protein folding in the periplasm and to respond to protein misfolding. Thiol-disulfide oxidoreductases are important enzymes mediating post-translational modification in proteins and aiding in protein folding. The thioredoxin fold and the active site CXXC are important for the functional properties of thiol-disulfide oxidoreductases. The β-lactamase selection system links the folding of unstable test proteins to antibiotic resistance, providing a means to conveniently monitor protein folding *in vivo*. This system together with mutagenesis techniques are useful in the chaperone discovery process.

Chapter 2
The CXXC motif is more than a redox rheostat[†]

Abstract: The CXXC active site motif of thiol-disulfide oxidoreductases is thought to act as a redox rheostat, whose sequence determines its reduction potential and with that its functional properties. We tested this idea by selecting for mutants in the CXXC motif of a reducing oxidoreductase, thioredoxin, that complement null mutants in a very oxidizing oxidoreductase, DsbA. We found that altering the CXXC motif affects not only the reduction potential of the protein, but also affects the ability of the protein to function as a disulfide isomerase, and also impacts its interaction with folding protein substrates and reoxidants. Surprisingly, nearly all our thioredoxin mutants had increased activity in disulfide isomerization *in vitro* and *in vivo*. Our results indicate that the CXXC motif has the remarkable ability to confer a large number of very specific properties on thioredoxin related proteins.

2.1 Introduction

Thiol-disulfide oxidoreductases are found in all living organisms. They catalyze the oxidation of protein thiols and the reduction and isomerization of disulfide bonds in proteins. They are thus critical for protein folding reactions. Thiol-disulfide oxidoreductases present in the periplasmic space act as oxidases or isomerases, while those in the cytoplasm perform mainly reductive steps. The CXXC active site of these

[†] This chapter is modified from the publication "The CXXC motif is more than a redox rheostat. Quan, S., Schneider, I., Pan, J., Hacht, A.V. & Bardwell, J.C. *The Journal of biological chemistry* (2007)". I performed all the experiments in the paper. JCAB designed and supvised the project. JP provided conceptual advice. IS set up the initial selection conditions for thioredoxin mutants. AH helped me with some spot titer experiments. I wrote the initial draft of the paper. JCAB edited it and wrote the final manuscript.

thioredoxin-related proteins is essential for their activity[65, 106, 107]. The sequence of the XX dipeptide located between the cysteines in the active site motif is very important in controlling the redox properties of the protein in which it is found[71, 72]; so much so that it has been termed a redox rheostat[108]. Several groups have shown that it is possible to alter the redox properties of these oxidoreductases by mutating the XX dipeptide in the CXXC motif. When changing the CXXC motif of one oxidoreductase to the sequence of another, the function of the original enzyme may be substantially altered. In some cases, the modified protein is even able to perform some of the activities of the protein from which the XX sequence was derived[73-75]. The effect these mutations have on the activity of the protein has been generally attributed to the effect they have on the redox potential of the protein[73, 75, 78]. We decided to test this hypothesis by selecting for mutants in the CXXC motif of the reducing protein thioredoxin that allow it to complement the very oxidizing protein DsbA.

The formation of disulfide bonds in *E. coli* requires DsbA[83]. DsbA oxidizes proteins secreted into the periplasm by rapidly exchanging its disulfide with reduced pairs of cysteines present in substrate proteins[84]. Reduced DsbA is reoxidized by the membrane protein DsbB[109]. DsbA has the highest redox potential known (E°'= –121 mV) among members of the thioredoxin related thiol-disulfide oxidoreductases. This is thought to be important for DsbA to be capable of rapidly oxidizing substrate proteins and keeping the periplasm in an oxidized state[81]. This high redox potential is understood in terms of the electrostatic forces in the vicinity of the active site disulfide, CPHC motif[81, 110]. Although the forces behind that control the redox potential of DsbA are well studied, the *in vivo* importance of this extremely high redox potential is not entirely clear. DsbA variants carrying mutations in the dipeptide in the CXXC active site can still function like wild-type DsbA in supporting normal cellular processes despite their less oxidizing redox potentials[81, 111]. In addition, no clear relationship exists between the redox potential of DsbA mutants and their ability to complement DsbA null strains[81].

Other members of the thioredoxin superfamily, when mutated, are capable of at least partially complementing *dsbA* when secreted into the periplasm. Jonda *et al.* tested four thioredoxin C<u>XX</u>C variants (harboring the XX dipeptides from the active site of DsbA, protein disulfide isomerase (PDI), glutaredoxin, and thioredoxin reductase) for

their ability to restore motility to a *dsbA*⁻ strain. They found that the DsbA, PDI and glutaredoxin-type sequence variants could partially complement *dsbA*⁻[99]. While thioredoxin is found nearly universally[112], DsbA is mainly restricted to the γ and β subgroup of proteobacteria, a clade that arose about 2.2 billion years ago[113], suggesting that thioredoxin and DsbA may have diverged at about this time. DsbA and thioredoxin overall only share 10% sequence identity[114]. Thus it is surprising that relatively minor alterations in the thioredoxin active site sequence can allow it to partially complement the very distantly related protein, DsbA.

The redox potential of thioredoxin is 160 mV lower than that of DsbA (E°' of thioredoxin = –270 mV[73]), and this is thought to contribute to its ability to keep the cytoplasmic protein cysteines reduced[115]. Previously, the ability of thioredoxin mutants to rescue *dsbA*⁻ mutants was mainly attributed to their increased redox potentials, which bring them closer to the extremely oxidizing redox potential of DsbA[72]. However, it should be noted that DsbC, which has an *in vivo* function as an isomerase, is nearly as oxidizing as DsbA. DsbC has a redox potential of –129 mV[116], while that of DsbA is –121 mV[84]. Yet DsbC in its function as an isomerase, reshuffle incorrect disulfide bonds, while DsbA oxidizes disulfide bonds. Thus, there are clearly more properties affecting the function of disulfide oxidoreductases than their redox potentials.

Previous studies on the role of the CXXC motif on the functional properties of thioredoxin-like proteins have typically limited their approach to changing the active site sequences of thioredoxin family members so that they precisely match those sequences of other family members. Then the properties of the mutated proteins were determined. We decided to take a much more comprehensive approach of randomly mutating the CXXC motif and then selecting active mutants in order to determine which features of the CXXC motif are involved in allowing thioredoxin to complement DsbA. With this approach, we hoped to gain insight into the key features of this widely conserved motif. We also hoped to further understand what makes DsbA such an effective disulfide catalyst and explore the functional similarities between thioredoxin and DsbA. The powerful selections available for DsbA activity and the wealth of information on thioredoxin provide an ideal situation to extensively investigate the relationships between

the active site sequence, the redox potential, and the *in vivo* function of DsbA as an oxidase.

In a process similar to natural selection, we made random alterations in the CXXC motif of *E. coli* thioredoxin and then applied selective pressure on cells, demanding a *dsbA*$^{+}$ phenotype. We found that the sequence of the CXXC motif has the remarkable ability not just in controlling the redox potential of thioredoxin related proteins, but also in controlling their ability to isomerize disulfides and their ability to interact with their reoxidants and folding proteins. Surprisingly, many of our complementing thioredoxin mutants were even more efficient than is wild-type DsbA in disulfide isomerization.

2.2 Results

2.2.1 Construction of thioredoxin mutant library

We decided to systematically explore the role played by the central "X" residues of the CXXC active site motif of thiol-disulfide oxidoreductases. To do this we randomly mutated this motif in the reducing protein, thioredoxin, and asked which mutants complement a null mutant in the oxidizing protein, DsbA. The plasmid pssTRX contains the coding sequences of thioredoxin fused to the export signal of DsbA^{99}. This plasmid was randomly mutated in the active site CXXC using multi site-directed mutagenesis. The resulting plasmid library was transformed into the *dsbA*$^{-}$ strain JP120 (see strain list on page 48). About 10,000 transformants were obtained on LB plates containing 200 μg/ml ampicillin. To judge the mutation frequency, 96 colonies were randomly chosen and plasmid DNA was prepared and sequenced. The sequencing results showed that 40% of the clones contained mutated CXXC sequences with an almost equal distribution of the four nucleotides A (36%) T (20%) G (24%) C (20%) in the mutated sites, indicating the mutagenesis was nearly random. Our aim in this mutagenesis was to obtain the vast majority of the 400 possible dipeptide combinations in the CXXC active site. Our library size of about 4000 and the almost equal distribution of nucleotides in the mutants made it likely that we achieved our aim.

2.2.2 Identification of thioredoxin mutants that complement *dsbA*

All the transformants generated from the random CXXC mutagenesis were replica-plated to plates containing 15 μM $CdCl_2$ to select for a *dsbA*$^+$ phenotype. Cadmium is a toxic divalent metal ion[117]. The high affinity of cadmium for protein thiol groups is believed to be the major biochemical basis of cadmium toxicity[118, 119]. In bacteria, *dsbA*$^-$ strains contain a much higher content of free thiol groups than *dsbA*$^+$ strains. Cd^{2+} binds to these free thiol groups and inhibits their proper folding[120, 121]. DsbA null strains are especially Cd^{2+} sensitive, failing to grow when exposed to 4 μM Cd^{2+}. In contrast, wild-type strains are resistant to up to 400 μM Cd^{2+}. JP120 transformed with plasmid encoding wild-type thioredoxin showed only residual growth at 15 μM Cd^{2+}. This provides a powerful selection for the *dsbA*$^+$ phenotype. By selecting for cadmium resistance in a *dsbA*$^-$ strain, we expected to recover thioredoxin mutants that could at least partially substitute for DsbA. We obtained 231 colonies resistant to 15 μM Cd^{2+} from the 4000 mutant colonies screened. Sequencing of the plasmids contained in these strains revealed 37 different combinations of amino acids present in the CXXC motif. We verified that the mutations present on the plasmids were sufficient for complementation of cadmium resistance by re-transforming the mutant plasmid DNA into the *dsbA*$^-$ strain JP120. We classified the cadmium resistance of the various mutants, by plating different dilutions of mid-log phase cultures onto LB ampicillin plates containing various $CdCl_2$ concentrations (Figure 2.1 and data not shown). The mutants were rank ordered from highest cadmium resistance to lowest. To verify that the cadmium resistance of these thioredoxin mutants reflects their ability to oxidize proteins, we tested Cd^R isolates for their ability to restore motility, a phenotype characteristic of *dsb*$^+$ strains. To be motile, *E. coli* must properly assemble its bacterial flagella, and to do this, a critical disulfide in the flagellar component FlgI must be introduced. Thus, *Δdsb* strains are nonmotile[122]. A good correlation exists (R=0.95) between the extent of motility and the degree of cadmium resistance (Figure 2.2). These results indicate that the thioredoxin mutants we had obtained were capable of restoring at least two of the phenotypes disrupted in *dsbA*$^-$ strains.

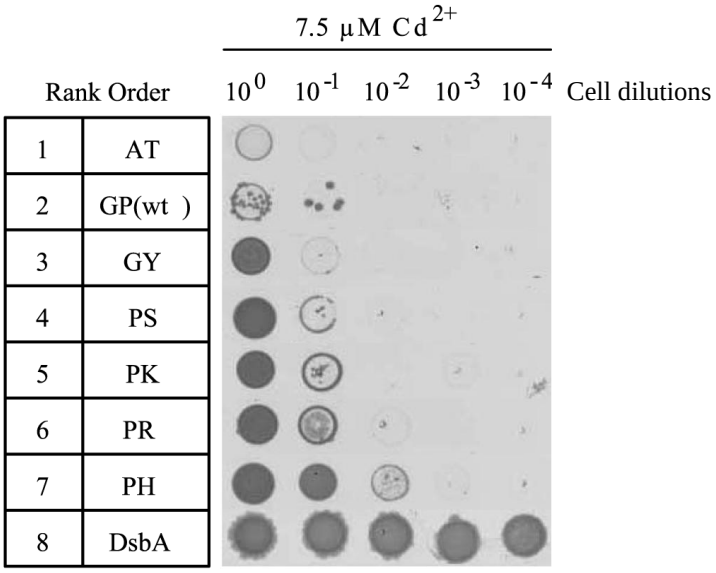


Figure 2.1 Thioredoxin active site mutants show various resistances to cadmium.
Mid-log phase cells were normalized to A_{600}=1. Serial dilutions of cultures from 10^0 to 10^{-4} were spotted on LB plates containing different concentration of cadmium. Cadmium resistance for each thioredoxin variant was ranked 1-7, from weak to strong, based on growth on LB ampicillin plates containing 7.5 μM, 10 μM and 15 μM Cd^{2+} after 18h incubation at 37ºC. Only the 7.5 μM Cd^{2+} plate is shown. The DsbA containing strain (ER1821, see strain list on page 48) showed the highest cadmium resistance and ranked 8. Variants are identified by their dipeptide sequence "XX" within the CXXC active site motif. AT is a *dsbA*-non-complementing thioredoxin mutant.

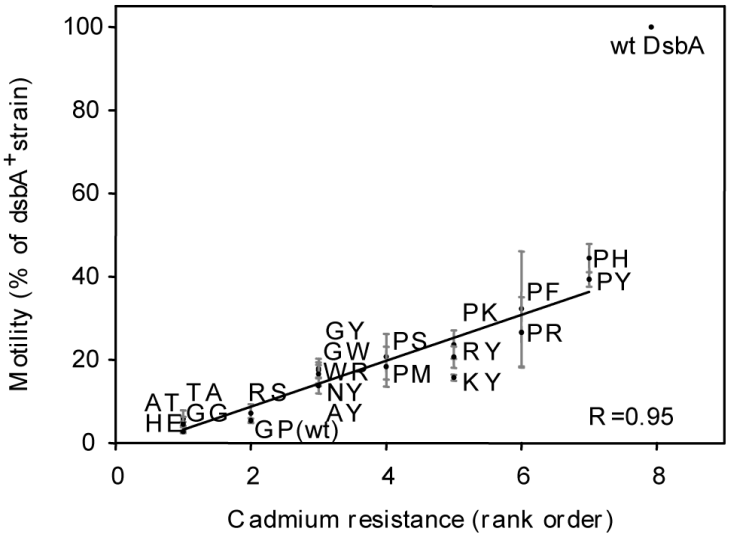


Figure 2.2 Correlation of motility and cadmium resistance for thioredoxin mutants.
Motility (% of dsbA strain) was plotted as a function of cadmium resistance (rank) for wild-type thioredoxin (wt) and 19 variants, including the non-complementing mutants RS, AT, HE, TA and GG. Data for wild-type DsbA (wt DsbA) strain was not used to develop the correlation.

2.2.3 Analysis of the dipeptide sequences in the CXXC motif of thioredoxin mutants

We then analyzed the dipeptide sequences in the CXXC motif of the plasmids that confer a *dsbA*$^+$ phenotype (Figure 2.3). We reasoned that the most cadmium resistant mutants were likely to survive the selection most often. We found a reasonable correlation (R=0.7) between the degree of cadmium resistance and the frequency at which mutants were obtained (Figure 2.4). Interestingly, for the more N-terminal position, the most frequently discovered amino acid is proline (65%), which was found about 16-fold more frequently than the 4% frequency expected by chance (Figure 2.3). It is worth noting that proline is found in the corresponding position of *E. coli* DsbA and in many DsbA homologs (76%) (Table 2.1). Thus, there seems to be a very strong bias towards proline at the N-terminal position. For the more C-terminal position in the CXXC motif, we obtained basic (His, Lys, Arg) or aromatic amino acids (Tyr, Phe) most frequently (51% and 32%, respectively). Histidine, tyrosine and phenylalanine were respectively found 9, 9 and 4-fold more frequently than expected by chance. The expected frequency for arginine is high because it is specified by 6 codons; however, the observed frequency was still 1.7-fold higher than expected by chance (Figure 2.3). These observations clearly indicate the preference for basic or aromatic amino acids at the more C-terminal position, and are consistent with the frequencies observed in naturally occurring DsbA homologs (79% for basic and 13% for aromatic amino acids at the more C-terminal position) (Table 2.1). Note that all these expected frequencies are what one calculates for the N- and C-terminal positions independently. The observed frequencies for the dipeptide sequences of the complementing mutants are on average 18-fold higher than the dipeptide frequencies expected by chance (Figure 2.4).

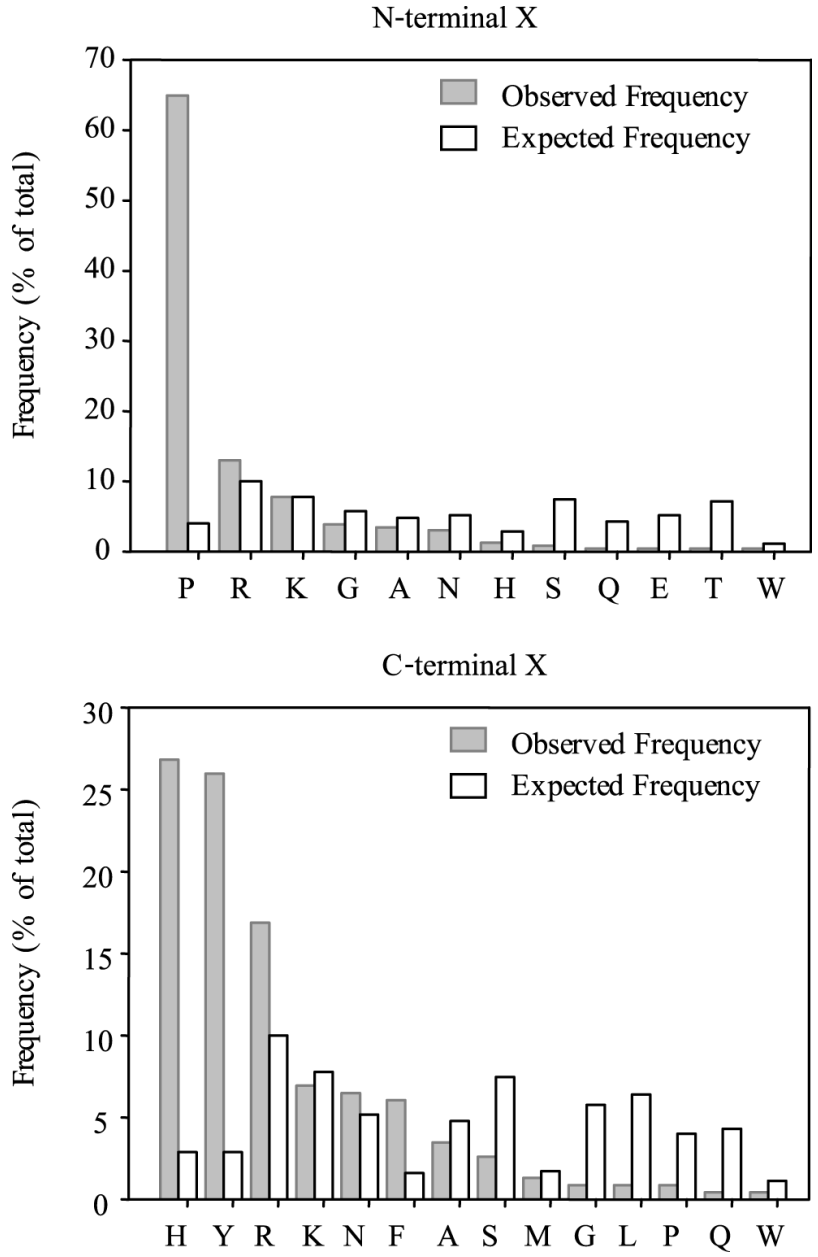


Figure 2.3 DsbA-complementing thioredoxin mutants CXXC active site motif with the sequence Cys-Pro-basic (or aromatic)-Cys.

Twelve amino acids were obtained at the more N-terminal position (top) and 14 were obtained at the more C-terminal position (bottom). Frequencies are given as percent of total. Expected frequency of each amino acid was calculated by adding together the frequencies of each of the natural codons of that amino acid. These codon frequencies were calculated by multiplying the abundance of each single nucleotide (A 36%, T 20%, G 24%, and C 20%) that was observed in unselected mutant sequences.

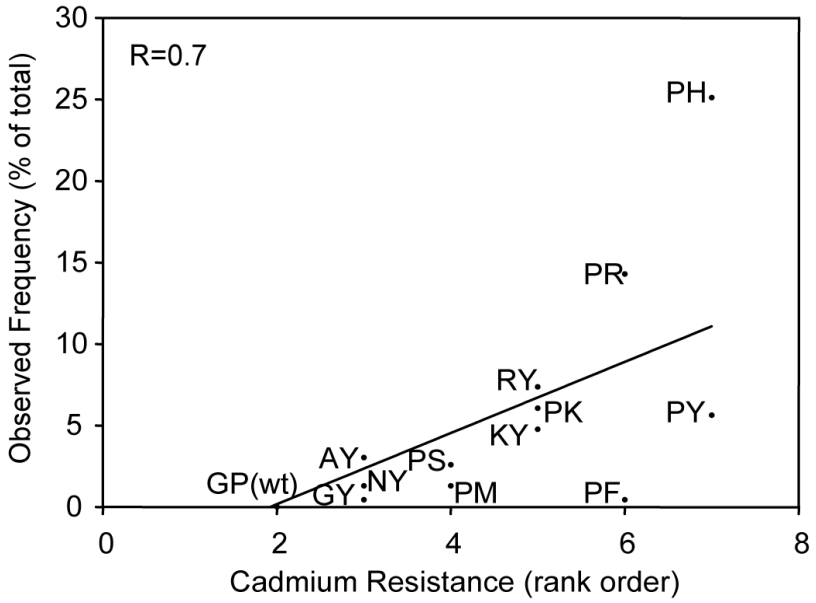


Figure 2.4 Correlation of cadmium resistance and observed and expected frequencies for the cadmium resistant thioredoxin mutants.
Note that the expected frequencies for all the cadmium resistant mutants range from only 0.064% to 0.746%, are much lower than the observed frequencies. For the PH mutant, the expected frequency is 218 fold lower than the observed frequency.

The mutant that most frequently survived the cadmium selection has the CPHC motif, observed in 58 out of 231 clones (25%), which is exactly the same CXXC sequence most frequently found in DsbA homologs in evolution. The expected frequency of this dipeptide in unselected clones is 218-fold lower. The short CPHC motif, which was optimized by evolution to work in DsbA, also seems to be optimized in conferring DsbA-like properties on thioredoxin, a protein that only shares 10% sequence identity with DsbA and is separated from DsbA by about 2 billion years of evolution[68]. We find the apparent portability of this CPHC motif remarkable.

Table 2.1 Analysis of dipeptide sequences in CXXC motif of DsbA and thioredoxin mutants

Amino acid in the more N-terminal positionin in CXXC (%)			Amino acid in the more C-terminal positionin in CXXC (%)		
	Nature evolved DsbA homologs	**Lab evolved thioredoxin mutants**		**Nature evolved DsbA homologs**	**Lab evolved thioredoxin mutants**
Proline	76	65	Histidine	79	27
Arginine	-	13	Tyrosine	8	26
Lysine	-	8	Arginine	-	17
Glycine	5	4	Lysine	-	7
Alanine	0.5	3	Phenylalanine	5	6
Asparagine	-	3	Asparagine	3	6
Serine	6	1	Alanine	4	3
Histidine	1	1	Serine	-	3
Glutamine	1	0.4	Proline	0.5	1
Glutamic acid	1	0.4	Methionine	-	1
Threonine	1	0.4	Glycine	-	1
Tryptophan	-	0.4	Leucine	-	1
Isoleucine	5	-	Tryptophan	0.5	0.4
Valine	3	-	Glutamine	-	0.4
Leucine	1	-			

The frequencies of occurrence for each of the amino acids in the N- and C-terminal positions in the CXXC motif of DsbA homologs from 181 species are compared with the frequencies for our laboratory evolved thioredoxin mutants showing resistance to 15 μM cadmium. The frequency for each amino acid is presented as a percentage, with the amino acids listed in the order with which they were found in the lab evolved thioredoxin mutants.

2.2.4 Kinetics of oxidation of thioredoxin mutants by DsbB *in vitro*

One trivial reason for the increased ability of our thioredoxin mutants to complement the *dsbA*$^+$ phenotype might be an increased expression level or an increased fraction of the oxidized form. Neither of these was the case, as quantitative western blots showed that the expression levels of the different thioredoxin mutants are almost identical to each other and to the level of expression observed for wild type (0.53-1.1 fold of wild type). AMS trapping showed that all the thioredoxin mutants are in the oxidized state.

Our mutants are DsbB dependent *in vivo* as judged by their inability to allow bacterial motility in a *dsbB*$^{-}$ strain (data not shown), suggesting that they are functioning by replacing DsbA in the periplasm. It is possible that our thioredoxin mutants are active in oxidizing proteins because they can be effectively oxidized by DsbB. To test this idea, we evaluated the oxidation efficiency of different thioredoxin mutants by DsbB using the enzyme monitored turnover method described by Gibson *et al.* [123]. This method utilizes the time course of the decay of an enzyme intermediate upon reaction with an excess of substrate in order to determine k_{cat} and apparent K_m values. A purple charge-transfer complex intermediate characterized by a strong absorbance peak at 510 nm is formed during the oxidation of DsbA by DsbB[88, 89]. Our experiments revealed that wild-type thioredoxin as well as its CXXC variants also induce an absorbance peak at 510 nm, which almost certainly corresponds to a very similar intermediate as is seen between DsbB and DsbA. We analyzed the enzyme monitored turnover absorbance curves at 510 nm and used these data to derive V_{max} and K_m values for both wild-type DsbA and wild-type thioredoxin and a large number of our thioredoxin mutants (Figure 2.5 and Table 2.2). A K_m of 14 ± 1 μM and a V_{max} of 0.74 ± 0.01 S^{-1} were obtained for the oxidation of wild-type thioredoxin by DsbB, similar to the previously published K_m of 20 ± 7 μM[99]. Surprisingly, these values are very similar to those of DsbA for DsbB (8 ± 1 μM). Thus, it appears that wild-type thioredoxin interacts with DsbB about as efficiently as DsbA does. This reduced the likelihood that our mutants can achieve their increased activity by increasing their catalytic efficiency with DsbB. Indeed, within the group of active thioredoxin mutants the correlation between *in vivo* motility values and their catalytic efficiency with DsbB is very poor. For simplicity, we refer to our thioredoxin mutants using only the dipeptide sequence of the active site; thus, the mutant with the CPHC sequence is referred to as "PH". We do note however, that the PH mutant, which has the best motility, does have a much better catalytic efficiency, i.e. a higher V_{max}/K_m value than almost all of the other mutants.

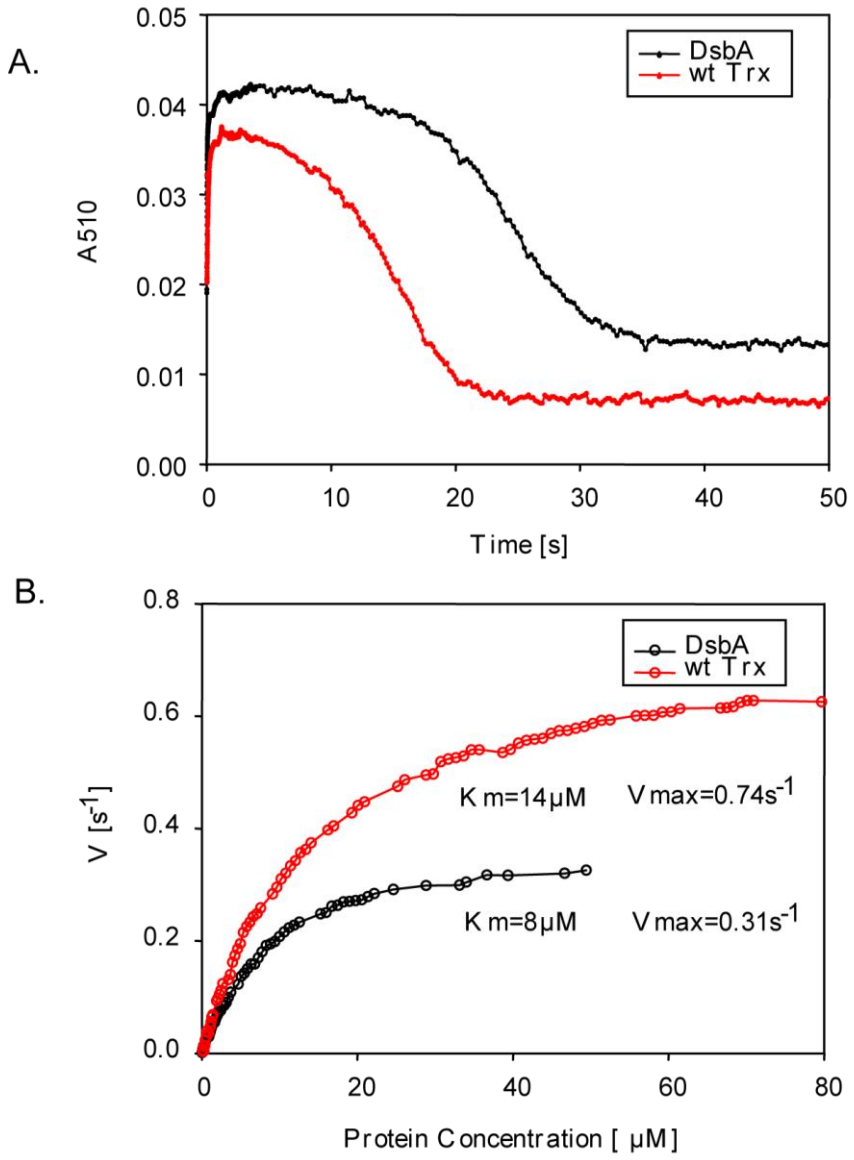


Figure 2.5 Enzyme monitored turnover by stopped flow.
A, Absorbance change at 510 nm upon mixing 80 µM DsbA or wild-type thioredoxin (wt Trx) with 10 µM DsbB and 200 µM Q1 in PND buffer at 10°C (see experimental procedures for details). **B**, V vs. [S] curves derived from traces in panel A by program A. The hyperbolic fits were made using SigmaPlot, giving a Km of 14 µM and a Vmax of $0.74s^{-1}$ for this reaction for wt Trx.

Table 2.2 Parameters for DsbA and the thioredoxin variants

CXXC variant	Observed frequency (%)	Motility (% of *dsbA*$^+$ strain)	Cadmium resistance (rank order)	Copper resistance (rank order)	E°′ (mV)	K_m with DsbB (μM)	V_{max} with DsbB (S^{-1})	V_{max}/K_m with DsbB ($S^{-1}\mu M^{-1}$)	k_2 with hirudin ($M^{-1}S^{-1}$)	Native hirudin at 2h (%)
DsbA	-	100	8	-	-121	8	0.31	0.041	2.0×10^6	3.8
DsbA complementing mutants										
PH	25	45	7	4	-210	22	1.03	0.049	7.6×10^5	50.5
PY	5.6	39	7	6	-197	7	0.09	0.014	ND	40.0
PF	0.4	32	6	6	-197	4	0.07	0.016	ND	49.3
PR	14	27	6	3	-203	121	2.19	0.018	2.4×10^5	12.1
PK	6	24	5	3	-207	121	2.09	0.018	ND	15.0
PS	2.6	21	4	6	-219	11	0.23	0.022	2.8×10^5	18.4
RY	7.4	21	5	7	-215	5	0.21	0.036	ND	98.3
PM	1.3	18	4	3	-216	21	0.41	0.020	ND	9.5
GY	0.4	18	3	6	-220	4	0.15	0.035	ND	57.0
GW	0.4	18	3	6	-220	7	0.21	0.033	ND	40.0
WR	0.4	17	3	3	-215	105	4.80	0.041	ND	10.7
KY	4.8	16	5	7	-214	10	0.51	0.051	1.6×10^6	79.1
NY	1.3	14	3	6	-214	6	0.21	0.033	ND	53.2
AY	3	14	3	7	-219	6	0.22	0.039	ND	47.0
Non-complementing mutants										
RS	0	7	2	3	-225	43	0.92	0.022	ND	18.4
TA	0	6	1	1	-236	-	-	-	ND	1.0
GP(wt)	0	5	2	6	-270	14	0.74	0.048	9.0×10^4	6.7
AT	0	5	1	3	-224	-	-	-	3.4×10^4	1.7
GG	0	5	1	3	-234	68	1.07	0.016	ND	6.1
HE	0	3	1	6	-229	28	0.35	0.013	ND	5.4

Results are shown for DsbA, wild-type thioredoxin and 19 thioredoxin active site variants studied by phenotypical assays and biochemical assays. From PH to AY are the DsbA-complementing thioredoxin mutants. A value of 0 of observed frequency indicated the mutants were non-complementing. In the pool of DsbA non-complementing mutants, AT and TA are not measurably oxidized by DsbB so that the K_m and V_{max} values could not be determined. ND: not determined.

Importantly, we tested the interaction of DsbB and five non-complementing control mutants (RS, TA, AT, GG, HE). No *in vitro* interaction can be detected between DsbB and AT or TA mutants. Thus, the inactivity of 40% of our mutants might be connected to their inability to interact with DsbB (Table 2.2). In contrast, all of the motile mutants

tested are able to interact with DsbB to some degree. Another trend we observed is that all the mutants with aromatic amino acids at the more C-terminal position had lower K_m values (Table 2.2), while some of the mutants with a basic amino acid in this position had large K_m values. This observation may indicate that the aromatic amino acid-containing mutants have a stronger tendency to form a substrate-enzyme complex with DsbB. Intriguingly, aromatic residues are also common in nature, found in 13% of DsbA homologs and are present in almost 100% of DsbC sequences. DsbC is *E. coli*'s principle disulfide isomerase[124]. We concluded from these experiments that the ability to be reoxidized by DsbB is necessary but not sufficient for thioredoxin mutants to be able to complement DsbA.

2.2.5 Measuring the redox potentials of thioredoxin mutants

One of the most prevalent hypotheses in the disulfide catalyst field is that the redox potential determines the function of thiodoxin-related thiol-disulfide oxidoreductases[71, 73, 75]. Thus, we determined the standard redox potentials of our thioredoxin mutants (Table 2.2). We found that the complementing mutants all showed an increased redox potential relative to wild type thioredoxin (average increase of 58 mV). On its own, this would help confirm that an increased redox potential is key in the ability of mutants to complement DsbA. However, in an important control experiment, we found that the five non-complementing control mutants also exhibited an increased redox potential relative to wild type thioredoxin. Although these were uniformly less oxidizing, the difference between the two groups was very small (Table 2.2). We observed a correlation between motility and redox potential for the complementing mutants (Figure 2.6). The more oxidizing the mutant, the better it is at complementing. The PH mutant is the most DsbA-like phenotypically, and not coincidentally, has the most DsbA-like sequence. However, with a redox potential of –210 mV, it is not the most oxidizing of the complementing mutants, which range from –220 to –197 mV. Importantly, it is only slightly more oxidizing than the non-complementing mutants, whose redox potentials range from –236 to –224 mV, or poorly complementing mutants like AY (–219mV) or NY (–214mV). Thus, while a more oxidizing redox potential may

be a prerequisite for allowing thioredoxin to functionally replace DsbA, it is clearly not the complete story.

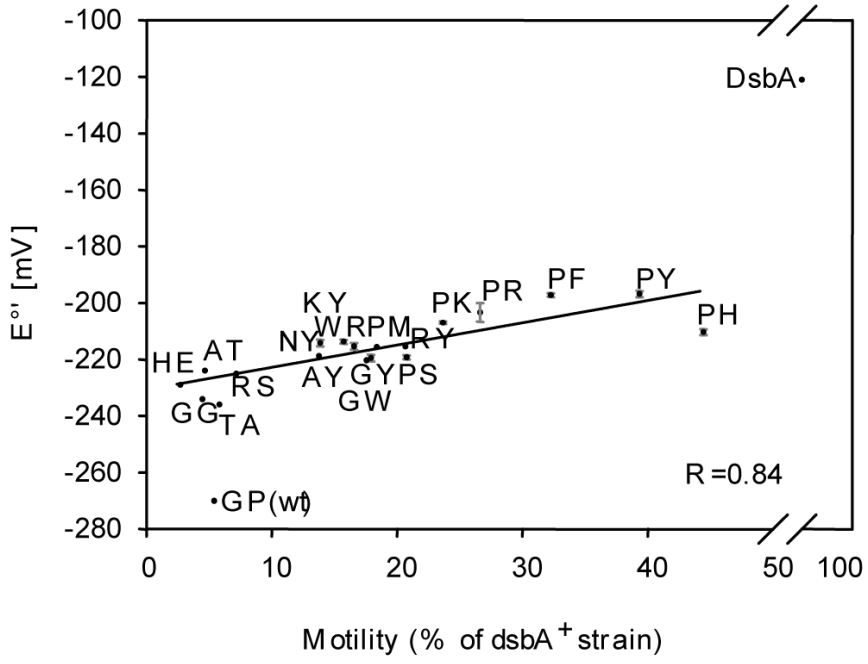


Figure 2.6 Correlation of standard redox potential (E°') and motility for thioredoxin mutants.

Curve was derived from 19 thioredoxin variants. Wild-type thioredoxin and DsbA were not used to calculate the correlation since their redox potentials are extremely low (for thioredoxin) or high (for DsbA). Note that the redox potential for the most motile thioredoxin mutant PH is not as high as other motile thioreodoxin mutants.

2.2.6 Thioredoxin mutants have enhanced protein oxidation and isomerase activity *in vitro* and *in vivo*

To obtain a more direct measure of the *in vivo* function of our thioredoxin mutants, we decided to examine how efficiently our thioredoxin mutants could oxidize substrate proteins *in vitro*. As a model substrate, we chose hirudin, a 65-residue protein with three disulfide bonds in the native state and a well-established model substrate for DsbA[125, 126]. This protein has a well studied disulfide linked *in vitro* folding pathway, it is commercially available, and the folding intermediates can be easily separated on HPLC. This is in contrast to the few known *E. coli* substrates for DsbC such as RNaseI, which

have none of these properties. To determine the rate at which our mutants oxidized the thiols in hirudin we followed their increase in the fluorescence that accompanies their incubation with reduced hirudin. A disulfide in the active site of thioredoxin acts to quench its tryptophan fluorescence[127]. The kinetics of the reaction are shown in Figure 2.8, with the second order rate constants for the oxidation of reduced hirudin by DsbA, thioredoxin and its variants shown in the figure legend. The complementing mutants were 3- to 11-fold faster at oxidizing hirudin than is wild type thioredoxin. A good correlation exists between these rate constants and their ability to restore motility *in vivo*, strongly suggesting that increased ability to oxidize proteins is crucial to their enhanced *in vivo* activity (Figure 2.7). One of the mutants (KY) is actually nearly as fast as DsbA's very rapid kinetics. DsbA is known for its extremely fast disulfide exchange reaction rates, which are about 3 orders of magnitude above known values for mono and dithiols[84]. DsbA is also much more reactive than thioredoxin; for example, DTT reduces DsbA about 1000 times faster than it does thioredoxin[128]. We find it noteworthy that some of our mutants approach the very rapid disulfide exchange kinetics of DsbA.

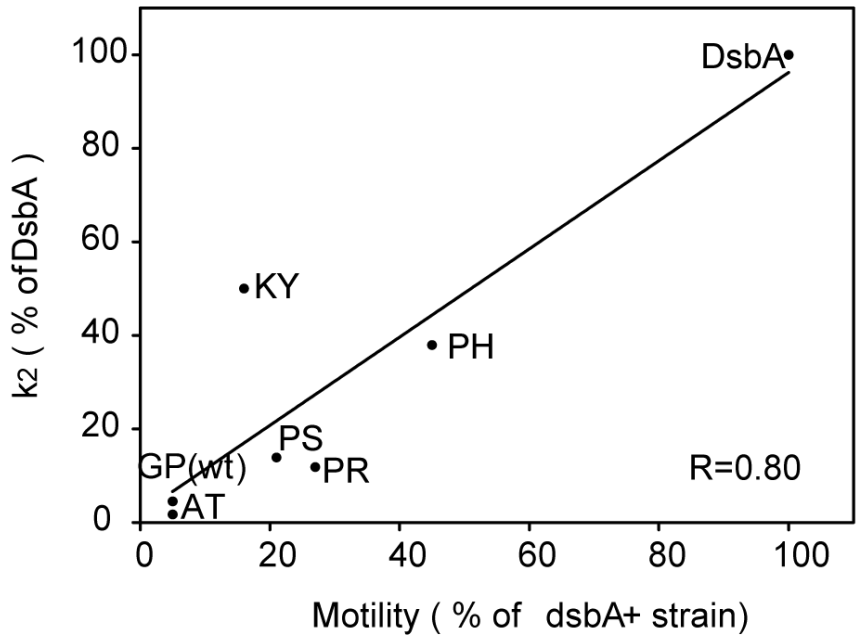


Figure 2.7 The second-order rate constants (k_2) of hirudin oxidative folding reaction and motility for thioredoxin variants are well correlated.

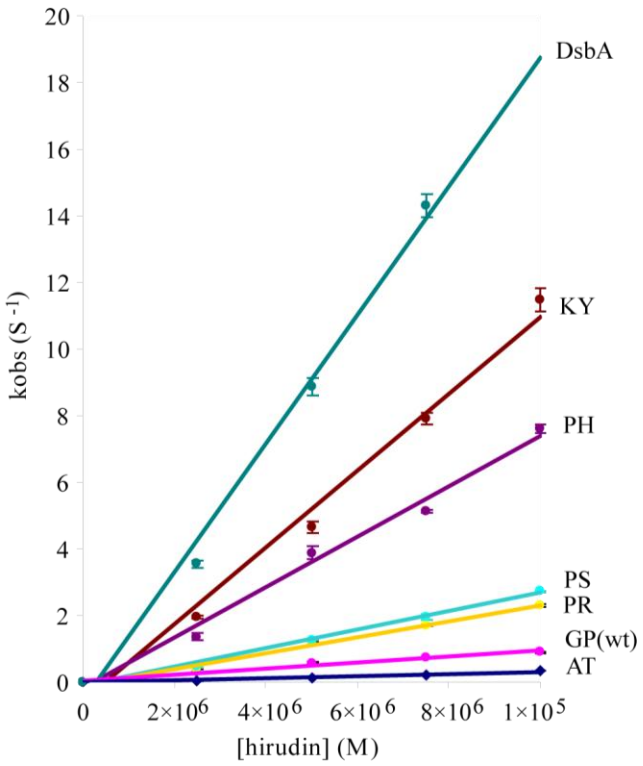


Figure 2.8 Kinetics of the oxidation of reduced hirudin by DsbA, thioredoxin, and its variants.

The second-order rate constants (k_2; M^{-1} S^{-1}) for the interaction of reduced hirudin and DsbA, thioredoxin, or its mutants were obtained by fitting the observed pseudo first-order rate constants (k_{obs}; S^{-1}) against hirudin concentrations (M). The slopes are the apparent second-order rate constants for the reaction, and the values are as follows: DsbA, 2.0×10^6 M^{-1} S^{-1}; KY mutant, 1.0×10^6 M^{-1} S^{-1}; PH mutant, 7.6×10^5 M^{-1} S^{-1}; PS mutant, 2.8×10^5 M^{-1} S^{-1}; PR mutant, 2.4×10^5 M^{-1} S^{-1}; thioredoxin (GP(wt), where wt is wild-type), 9.0×10^4 M^{-1} S^{-1}; and AT mutant, 3.4×10^4 M^{-1} S^{-1}.

We then wanted to test if our mutants were more efficient than wild type thioredoxin in the overall oxidative folding of hirudin, a process that involves both disulfide oxidation and isomerization steps. We were very pleased to find that all the DsbA-complementing mutants exhibited an enhanced ability to oxidatively refold reduced hirudin. They were not only more efficient compared to wild-type thioredoxin but in a real surprise, were also faster compared to wild type DsbA, showing complete refolding of hirudin at time points where wild type DsbA is just getting started (Figure 2.9 and Table 2.2). We find this surprising because DsbA has presumably been

optimized for the oxidative folding of proteins over millions of years of evolution, yet by changing just two residues in a related protein we appear to have been able to isolate a mutant protein that is superior to DsbA in the oxidative folding of at least one protein.

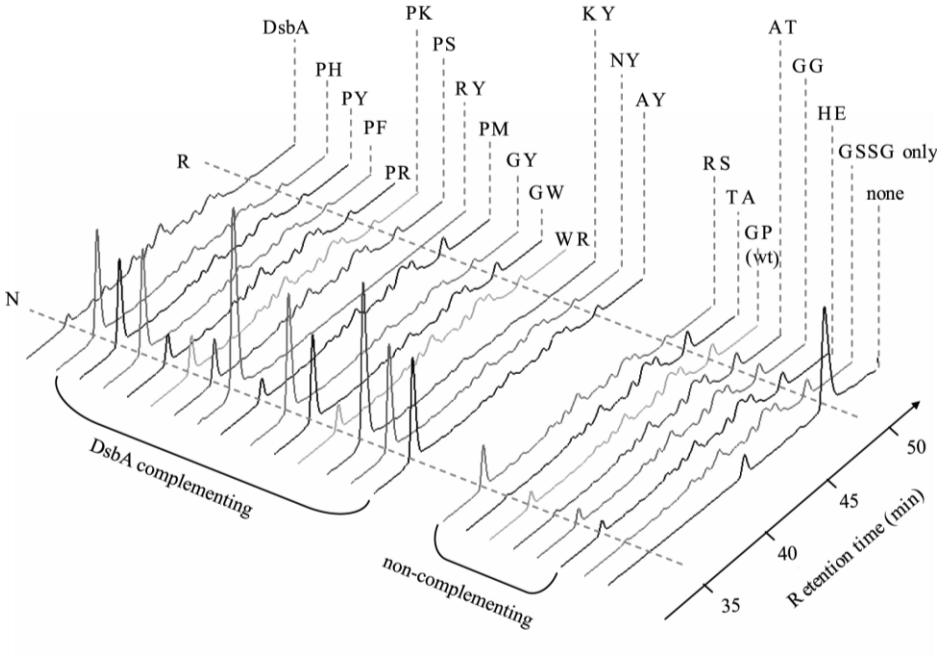


Figure 2.9 DsbA-complementing thioredoxin mutants fold reduced hirudin efficiently.

Reduced hirudin (24 μM) was incubated with oxidized glutathione (200 μM) in phosphate buffer. Then catalytic amount of oxidized thioredoxin variants or DsbA (1 μM) were added to initiate each folding reaction (see experimental procedures for details). Two hours after initiation, hirudin folding intermediates were separated by reverse-phase HPLC. The absorbance traces are aligned on the same time scale. Each trace is labeled by the catalyst added or by the name of the control. From left to right, the traces follow the order of decreased motility of the strains containing the corresponding thioredoxin variants or DsbA. The *dashed lines* indicate fully reduced (*R*) and native (*N*) hirudin.

The oxidative refolding of reduced hirudin involves two processes, an initial oxidation reaction, followed by isomerization of incorrect disulfides to generate native hirudin. Our thioredoxin mutants were better than DsbA in overall oxidative folding of hirudin and not quite as good as DsbA in the initial oxidation step, leading us to predict that they had isomerization ability that is superior to that of DsbA.

Using scrambled hirudin as a substrate we indeed found that our mutants can generate native hirudin from a disulfide scrambled hirudin substrate faster than can DsbA (Figure 2.10 A). Two of those tested (KY and PH) were very efficient, at least as effective as DsbC, the principle disulfide isomerase *E. coli* (Figure 2.10 A). This is surprising because the phenotype we selected for is the ability to complement DsbA (an oxidase) and not DsbC (an isomerase). We note that wild type thioredoxin does show some isomerase activity *in vivo* and *in vitro*, as has previously been reported [129, 130]. We note that scrambled hirudin is a very heterogenous substrate, consisting of at least 8 different species that are separable by HPLC (Figure 2.10 A). The thioredoxin mutants are apparently capable of accelerating the isomerization of these different species, strongly implying that these mutants are active in the isomerization of a variety of substrates.

To test if our mutants also have an improved ability to isomerize protein disulfides *in vivo* we tested if they could complement the copper sensitivity of null mutants in DsbC. Plasmids containing thioredoxin active site variants were transformed to the *dsbC* null strain BW25113 *dsbC::kan*. Copper can catalyze the formation of incorrect disulfides, including disulfide linked oligomers, which DsbC appears to be able to resolve[124]. Eight of the 14 of the active mutants we tested were either as copper resistant as *dsbC*$^{+}$ strains or nearly as copper resistant. This was in contrast to the 5 inactive mutants where all but one was no more copper resistant than the empty vector (which had a copper resistance rank of 3) (Figure 2.10 B, Table 2.2 and data not shown). Our mutants seem to act by both increasing their ability to oxidize and isomerize disulfides in proteins.

The mutants that seemed to show the best isomerase activity *in vivo* and *in vitro* are those with an aromatic or histidine amino acid in the C-terminal position of the CXXC motif (Table 2.2). (The side chain of histidine can also be considered aromatic; it meets the electron rule of aromatic amino acids in one of its protonation states.) Previous work on the *in vivo* folding of a multi-disulfide protein in *E. coli* by DsbC mutants with a randomized CXXC central dipeptide suggested a strong preference for hydrophobic and particularly aromatic amino acids at the C-terminal position[111]. Our study found a higher average rate of hirudin refolding for mutants with aromatic amino acids in the active site than for those with basic amino acids. This result is consisten with our hypothesis that the

aromatic amino acid-containing active site dipeptide mutants of thioredoxin may function as disulfide isomerases.

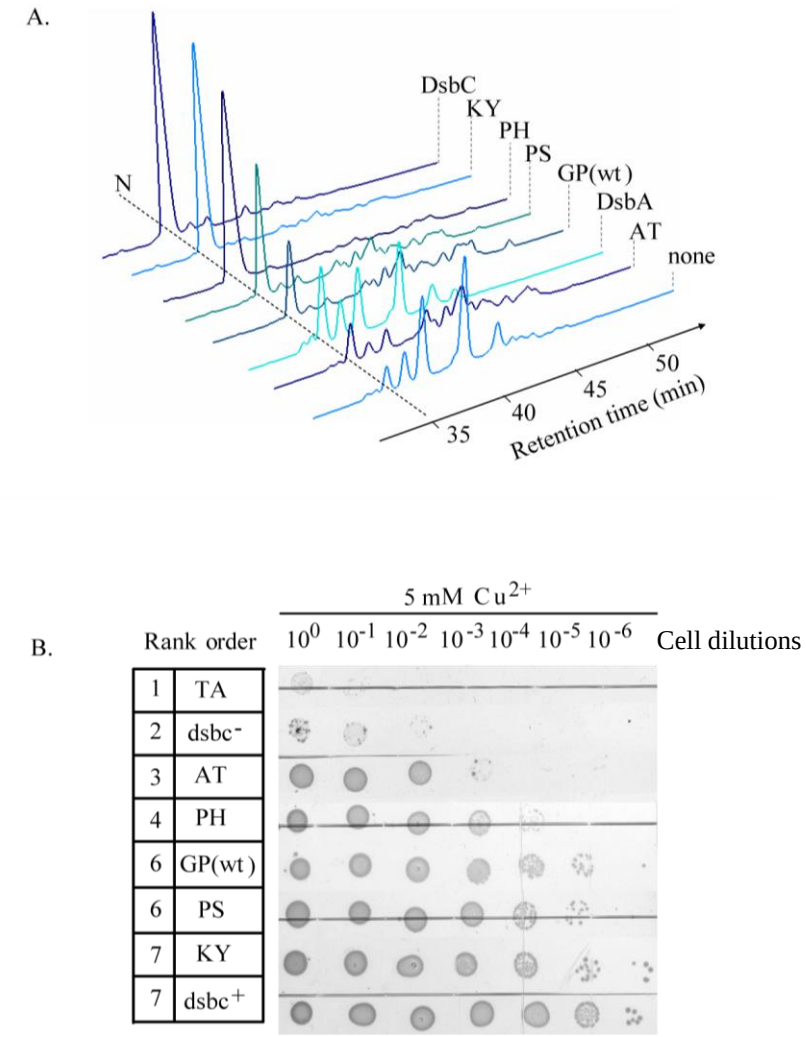


Figure 2.10 Thioredoxin active-site variants function as isomerases *in vitro* and *in vivo*.

A. scrambled hirudin (21 µM) was incubated with equal amount of freshly reduced thioredoxin variants, DsbA, and DsbC in the ismomerization buffer (see experimental procedures for details). Five minutes after initiation, hirudin folding intermediates were separated by reverse-phase HPLC. *N*, native hirudin. DsbA-complementing thioredoxin mutants folded scrambled hirudin more efficiently than did DsbA; some were even comparable with DsbC. The non-complementing AT mutant was worse than wild-type (*wt*) thioredoxin. **B.** thioredoxin mutants, as well as the wild type, conferred various copper resistances to a *dsbC*$^{-}$ strain. Mid-log phase cells were normalized to A_{600}=1. Serial dilutions of cultures from 10^{0} to 10^{-6} were spotted on BHI plate containing 5mM $CuCl_2$. Copper resistance for each thioredoxin C*XX*C variant was ranked 1-7, from weak to strong. Variants are identified by their *XX* dipeptide sequence within the C*XX*C active-site motif.

2.3 Discussion

In a process similar to natural selection, we identified amino acid combinations of the dipeptide in the CXXC motif of thioredoxin mutants that complement the oxidizing thiol-disulfide oxido-reductase DsbA. We show that this CXXC motif is a major factor in determining the function of the protein, in ways that go well beyond simple changes in the redox potential. Our results closely recapitulate evolution. A strong preference was found for proline at the more N-terminal position of the motif and a basic or aromatic amino acid at the more C-terminal position, very similar to what is observed for natural DsbA sequences. The CXXC sequence that best complemented DsbA deletion mutants was CPHC, exactly the same sequence that is most commonly found in natural DsbAs. This implies that the CPHC motif is optimal for conferring DsbA-like properties on both DsbA and thioredoxin, two proteins that share only 10% sequence identity.

The spectrum of the CXXC sequences found in our lab-evolved thioredoxins is slightly broader than what is found in evolution. In natural DsbAs, the only positively charged residue observed is histidine, while in our lab evolved DsbA-like thioredoxins the preference for positively charged residues is slightly more relaxed (His 27%, Arg 17%, Lys 7%). In our lab-evolved mutants, a positively charged residue was observed in the more N-terminal position 20% of the time when there was no positively charged residue in the C-terminal position, possibly indicating that an N-terminal positively charged residue may substitute for a C-terminal one.

In DsbA, the positively charged His32 residue is known to act to form a salt bridge with the thiolate anion of Cys30[77]. This stabilizes the reduced form of DsbA and thus contributes to some of the thermodynamic driving force behind DsbA's oxidizing power[81, 110]. Our observation that thioredoxin mutants frequently possess a positively charged residue at the C- or N-terminal positions is further evidence that this electrostatic interaction is important in determining the functional properties of the thioredoxin-related proteins. The aromatic amino acids we found may also act by stabilizing the N-terminal thiolate anion, in this case by a sulfur-aromatic ring interaction[131, 132]. The π-electron cloud on the faces of the aromatic ring is known to interact strongly with cations[133], so the ring edges are relatively positively charged and capable of interacting with negatively

charged thiolate ions. This type of reaction is common in proteins containing thiol groups[134, 135].

Our results suggest that proline is strongly favored as the N-terminal amino acid in the CXXC motif. This was unexpected as we know of no prior evidence implicating the importance of proline at this position. We reason that proline at the more N-terminal position may be important due to effects on the local conformation near the active site. The active site CPHC motif for DsbA is located at the N-terminus of the α1-helix, and proline tends to break helices. The positive dipole that occurs at the N-terminal end of the α1-helix is thought to contribute to the oxidizing power of DsbA[136, 137]. The interaction between the N-terminal thiolate anion in the CXXC motif and the α-helix dipole appears to be very sensitive to the microenvironment at the helix terminus[138, 139]. The precise positions of the proline may be important in breaking the helix and in generating the appropriate geometry for such an interaction.

The efficiencies of selected thioredoxin mutants were studied *in vitro*; the interaction with their oxidant DsbB, and with a substrate protein hirudin, as well as their redox potentials are also determined. All DsbA-complementing thioredoxin mutants have increased redox potentials as expected, since an increased redox potential ensures that the thioredoxin mutant will pass the threshold to be an effective oxidant. However, all five non-complementing mutants tested also exhibited increased redox potentials, strongly suggesting that an increased redox potential is not sufficient for making thioredoxin DsbA-like in function. That both complementing, and non-complementing mutants have a redox potential more oxidizing than thioredoxin is, upon reflection, not surprising; the function of thioredoxin in the cell is to reduce disulfides, so its sequence is likely to be optimized for this function. Random mutagenesis usually leads to a decline in a protein's structure and function and thus in the case of thioredoxin will be expected to move thioredoxin's redox potential in a less reducing direction, towards the reducing power present in a minimally structured peptide. Consistent with this, the redox potentials of our mutants are similar to the –230 mV value observed for a CAAC motif in an α-helical peptide free in solution[140]. Thus, a more oxidizing redox potential, though probably necessary for thioredoxin mutants to complement DsbA, cannot be the only requirement.

The interactions of DsbA with substrate proteins and the reoxidant DsbB are important components of DsbA's catalytic cycle. Surprisingly, our data did not establish a correlation between the efficiency of interaction with DsbB and the ability to complement DsbA. In fact, wild type thioredoxin interacts with DsbB about as efficient as does DsbA, indicating that interacting with DsbB may not be the rate limiting step in the catalytic cycle. However, we did observe that more rapid interaction with DsbB does help to improve the catalytic efficiency of the reoxidation reaction. This was seen for the CPHC mutant, whose relatively high efficiency of interaction with DsbB compensates for its only moderately oxidizing redox potential, making it a very strongly DsbA-complementing mutant. Although it is not a rate-determining step, abolishment of the interaction of thioredoxin variants with DsbB abolishes the whole catalytic cycle, resulting in the inability of several mutants to complement DsbA (AT and TA mutant).

We found that all of the DsbA complementing mutants exhibited an enhanced ability to oxidatively refold reduced hirudin compared to wild type thioredoxin. Surprisingly, we found that the majority of the DsbA complementing mutants are better at isomerizing scrambled hirudin than is DsbA and these isomerase activities can occur *in vivo*. The observation that our mutants selected on the basis of complementing DsbA function *in vivo* had increased isomerase activity raises the interesting possibility that DsbA has more isomerase function *in vivo* than previously thought. Although most of the *in vivo* and *in vitro* work on DsbA has emphasized its oxidase activity, it does exhibit measurable levels of isomerase activity *in vitro*[141]. Copper (II) apparently acts as a nonspecific oxidant *in vivo*, generating large numbers of nonspecific disulfides, including disulfide linked dimers and multimers[124]. Mutants in the gene for DsbC, which is thought to be *E. coli*'s principle isomerase, are copper sensitive, presumably because *dsbC*$^{-}$ strains are deficient in their ability to reshuffle these incorrect disulfides. It is interesting to note, however, that *dsbA*$^{-}$*dsbC*$^{-}$ double mutant strains are even more copper sensitive than *dsbC*$^{-}$ strains implying that DsbA and DsbC both participate in correcting the incorrect disulfides generated by copper[124]. Our finding that thioredoxin mutants that rescue DsbA null mutant phenotypes show enhanced isomerase activity is consistent with the notion that one of the properties that they are rescuing is isomerase activity. A number of the mutants selected on the basis of DsbA complementation, appear by our

tests to have thiol-disulfide oxidase activity *in vitro* as efficient as DsbA and isomerase activity *in vitro* as efficient as DsbC. If one thioredoxin related protein can "do it all" then why has evolution chosen to evolve both a DsbA based oxidation pathway and a DsbC based isomerization pathway? We do not have a clear answer to this question, but in general, multigene families are thought to have evolved to address questions of substrate specificity. We note that in the *in vivo* tests for oxidation and isomerization of our thioredoxin mutants, though they showed considerable activity, were not as efficient as is wild type *E. coli.* We also note that nature has chosen to give wild type DsbA an extremely oxidizing redox potential of E°' of –121mV. Although the CXXC motif is important in determining the redox potential of thioredoxin related proteins, it is unlikely to be the sole determinant. Thus, mutations in thioredoxin that are restricted to altering this motif, as ours were, are unlikely to be as effective as DsbA in oxidizing dithiols.

Despite almost 45 years of work on protein disulfide isomerases, surprisingly little is known about the *in vivo* requirements for disulfide isomerization. It is known that dimerization of two thioredoxin domains enhances disulfide isomerization activity[142, 143]. During the purification, different thioredoxin mutants eluted at approximately the same position as wild-type thioredoxin (monomer) on the gel filtration column, indicating it is unlikely that our thioredoxin mutants are dimers. It also seems likely that the redox potential of the isomerase is important for its function. The mechanism of isomerization requires that the isomerase functions as both an acceptor and a donor of disulfides; thus, a priori one would expect the redox potential of the isomerase to be finely balanced so that it can both accept and donate disulfides. Consistent with this notion, our thioredoxin mutants do have a redox potential intermediate between the strongly oxidizing redox potential of DsbA and the relatively reducing redox potential of thioredoxn, and the redox potential does seem to be important in their mechanism.

A selection for mutants in thioredoxin that complement null mutations in protein disulfide isomerase performed in yeast yielded mutations with a redox potential more oxidizing than that of the original thioredoxin[75]; however, it now appears that the sulfhydryl oxidation, not disulfide isomerization, is the principal function of protein disulfide isomerase in yeast[144]. This probably explains why more oxidizing mutants were obtained using this selection. It is interesting to note that in their selection, as well as in

ours, only a very limited range of mutants that allowed complementation were obtained. Interestingly, they also found that an aromatic residue (in their case tryptophan) worked well to cause thioredoxin to have isomerase activity *in vivo*. These results suggest that there are very specific sequences for good isomerase function that go beyond having the appropriate redox potential.

Overall, we conclude that at least three properties are required for a thioredoxin mutant to complement DsbA: an increased redox potential, the ability to be efficiently oxidized by DsbB, and the ability to oxidize substrate proteins. An increased redox potential seems to be the prerequisite. The other two properties, on the other hand, if missing, will abolish the ability to complement DsbA. The two properties sometimes can even compensate for relatively "low" redox potential and make the mutant protein very active *in vivo*. The total efficiency as an oxidant is a good combination of the three aspects.

Our work further establishes the vital importance of the CXXC residues in determining the functional properties of thiol-disulfide oxidoreductases of the thioredoxin superfamily. Remarkably, the short motif, which was optimized by evolution to work in DsbA; also seems to be optimized in conferring DsbA-like properties on thioredoxin, a protein separated from DsbA by more than 2 billion years of evolution. The ability of two amino acids to confer properties of one enzyme or another shows the great importance of these residues. In contrast to current thinking, this importance is not limited to the effect these changes have on the redox potential of the protein, but our work shows that these residues have important effects on interactions both with folding proteins and the reoxidant DsbB and on the ability of the protein to participate in disulfide isomerization reactions. We were surprised to find that thioredoxin mutants selected on their ability to complement the oxidizing thiol-disulfide oxidoreductase DsbA, turned out to be much more efficient in isomerizing disulfides than is DsbA, some were even as efficient as is DsbC, *E. coli*'s principle disulfide isomerase.

2.4 Experimental procedures

2.4.1 Strains and plasmids

The strains and plasmids used in the present study are listed in Table 2.3.

Table 2.3 Strains and plasmids used

Strain or plasmid	Genotype or relevant characteristics	Source
Strains		
ER1821	F⁻ *glnV44 e14*⁻ (*McrA*⁻) *rfbD1 relA1 endA1 spoT1 thi-1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
JP120	ER 1821 *dsbA::kan1, zih12::Tn10*	Lab collection
IS13	JP120, pssTRX	This study
JP373	ER1821, pKK233-2	Lab collection
WP591	DHB4, *ΔtrxA*	Ref. [115]
BW25113, *dsbC::kan*	*rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1 dsbC:: kan*	From Keio collection
Plasmids		
pssTRX	pASK40-ss-trxA, thioredoxin exported via the DsbA signal sequence	Ref. [99]
pKK233-2	Cloning vector, pBR322 origin, Amp^R	Pharmacia

All strains used were derivatives of *E. coli* K-12 strains.

2.4.2 Alignment of DsbA sequences

To identify homologs of DsbA in different bacteria, the nonredundant (nr) database at NCBI was searched with the *E. coli* K12 DsbA protein sequence (accession no. CAA56736 [GenBank]) using the PSI-BLAST algorithm with a threshold of 0.005. We used four rounds of iterations. DsbA family members can be distinguished from other thioredoxin family members because they contain an α-helical structural motif absent from other thioreodoxin related proteins, such as DsbC, DsbD, DsbG and thioredoxin [68]. All sequences were checked to verify that they contain sequences homologous to this α-helical domain. When sequences were available from multiple strains, one set of sequence data was chosen to represent that species. For paralogs existing in the same organism, we chose the sequence most homologous to the *E. coli* DsbA query sequence based on blast E values. In this way, nearly all of the sequences chosen are likely to be sequences of DsbA orthologs. Final alignment of the active site dipeptide sequences of DsbA homologs was done based on the sequences from 101 genera and 181 species.

2.4.3 Cloning and phenotypic assays

Construction of thioredoxin mutant library: The thioredoxin plasmid pssTRX with the DsbA signal sequence for periplasmic expression was a gift from R. Glockshuber. It was constructed by replacing the DsbA sequence in plasmid pDsbA3 with the thioredoxin sequence at NheI and BamHI restriction sites[99]. This plasmid was randomly mutated with the Quickchange Multi Site-Directed Mutagenesis Kit (Stratagene) to obtain mutations in the dipeptide in the CXXC motif of thioredoxin. The primer used to construct the mutant collection has six degenerate nucleotides for the dipeptide (Primer sequence as 5' TCGATTTCTGGGCAGAGTGGTGCNNNNNNTGCAAAATGATCGCCCCGATT 3'). Thermal cycling reaction was performed as directed; the product digested with *Dpn*I, and precipitated using the DNA Pellet Paint Co-precipitant (Novagen). The precipitated DNA was electroporated into 50 μl *E. coli* JP120 electrocompetent cells. Transformants were plated on Luria Bertrani (LB) plates supplemented with 200 μg/ml ampicillin and incubated at 37 °C for 14 h. Colonies obtained after electroporation were replicated using sterile filter papers (Whatman) onto LB plates supplemented with 15 μM $CdCl_2$ and 200 μg/ml ampicillin at 37 °C for 18 h. Strains were classified by their resistance to cadmium. The plasmids that conferred cadmium resistance in *dsbA*$^{-}$ strains were isolated via a mini prep kit (Promega) and the *trxA* genes encoded by those plasmids were sequenced.

Spot titers for cadmium resistance and copper resistance: Spot titers were performed to quantify the relative cadmium or copper resistance caused by the dipeptide change at the active site. Briefly mid-log phase cells ($OD_{600} \approx 1$) were serially diluted in sterile 170 mM NaCl solution. 2 μl of each dilution was plated onto LB plates supplemented with 200 μg/ml ampicillin and $CdCl_2$ in a concentration of 7.5 μM, 10 μM or 15 μM or onto BHI (Brain Heart Infusion) plates containing 5 mM $CuCl_2$. After 18 h at 37 °C, the growth of each thioredoxin mutant was compared and the cadmium or copper resistance of each of the thioredoxin mutants was ranked 1-7, with higher values indicating increased resistance.

Motility assay: The motility assay was performed in M9 minimal soft agar plates (0.2%) supplemented with 18 amino acids (excluding cysteine and methionine), 0.4%

glycerol, 2 μg/ml nicotinamide, 0.2 μg/ml riboflavin B2, 2 μg/ml thiamine B1, 2 μg/ml biotin, 1 mM $MgSO_4$ and with the appropriate antibiotic. Mid-log phase liquid cultures were diluted based on the optical density (OD) at 600 nm to normalized cell density for 1 OD. 2 μl of cells were then inoculated into the center of the motility plate. After 20 h incubation at 37 °C, the diameter of the swarm was measured.

2.4.4 Biochemical and biophysical assays

Protein purification: Thioredoxin variants were expressed and purified in the periplasm of a *trxA*$^{-}$ strain WP591 as previously described[72, 99]. Protein concentrations were determined by absorption at 280 nm using an extinction coefficient of 13980 cm^{-1} M^{-1} for reduced wild-type thioredoxin and 14105 cm^{-1} M^{-1} for oxidized wild-type thioredoxin. The extinction coefficients for thioredoxin mutants were calculated using the online program ProtParam (http://ca.expasy.org/tools/protparam.html)[145].

AMS trapping: To examine the redox states of the thioredoxin mutants, acid-precipitated proteins were solubilized in buffered sodium dodecyl sulfate (SDS) solution containing 4-acetoamido-4'-maleimidylstilbene-2,2'-disulfonate (AMS) as described previously[146]. The samples were incubated in the dark at 37°C for 1 hour. The alkylation was stopped by adding SDS-nonreducing loading buffer and analyzed by electrophoresis and western blotting.

Kinetics study for the interaction with DsbB by stopped flow: Stopped flow absorbance measurements were performed on a Hi-Tech Scientific SF61 instrument (1.0 cm path length) in single-mixing mode. The typical reaction contains around 100 μM of freshly reduced thioredoxin variants, 10 μM DsbB and 200 μM ubiquinone-1 (coenzyme Q1; Sigma). Thioredoxin variants and a DsbB/Q1 mix were incubated in PND buffer (50 mM Sodium Phosphate, 300 mM NaCl, 0.04% dodecyl maltoside, pH 8.0) at 10 °C before mixing. The absorbance after mixing was recorded at 510 nm. One data set contains three to four successive shots. Each trace from the enzyme monitored turnover experiment were simulated separately according to ref. [123] using Program A (developed by Chung-Yen Chiu, Rong Chang, and Joel Dinverno under the direction of David P.

Ballou, University of Michigan[147]). The resulting data was further fit by SigmaPlot using the Michaelis-Menten equation to give the value of K_m and V_{max}.

Redox potential measurement: Proteins whose redox potentials were to be measured were incubated with the gamma domain of DsbD, a protein of known redox potential, –235 mV [148] . At equilibrium, different protein species were separated by reverse-phase HPLC and the equilibrium constant (K_{eq}) of the reaction and the standard redox potential ($E^{o'}$) of the protein were determined as described [149]. For the mutants whose reduced and oxidized peaks could not be fully separated from either reduced or oxidized peak of γDsbD, the equilibrium constants with glutathione were measured as described[72]. A value of –240 mV was used for the standard redox potential of glutathione at pH 7.0 to calculate the standard redox potentials of the thioredoxin variants[150].

Oxidative folding of hirudin: 24 μM reduced hirudin and 200 μM oxidized glutathione were incubated in 100 mM sodium phosphate buffer containing 1 mM EDTA, pH 7.0, at 25 °C. Oxidized thioredoxin variants were added at a catalytic concentration of 1 μM to initiate each folding reaction. Aliquots of 120 μl were removed after different reaction times and quenched with 15 μl formic acid and 15 μl acetonitrile. Hirudin folding intermediates were separated by reverse-phase HPLC on a C18 column at 55 °C in a 19-25% acetonitrile gradient in 0.1% (v/v) trifluoroacetic acid. The absorbance was recorded at 220 nm.

Stopped flow fluorescence measurements were performed on a KinTek SF-2004 instrument in single-mixing mode. The typical reaction contains 0.5 μM of oxidized protein (DsbA, thioredoxin and its variants) and 2.5, 5, 7.5 and 10 μM reduced hirudin, respectively. The oxidized proteins and reduced hirudin were incubated in 100 mM sodium phosphate buffer containing 1 mM EDTA, pH 7.0 at 25 °C before mixing. An excitation wavelength of 295 nm and a Bandpass filter were used to monitor the fluorescence change of the proteins. Each stopped-flow trace was fit to a single exponential according to the equation: $F = F_o + \Delta F(1 - e^{-kobs \cdot t})$, where F is the fluorescence emission at known times; F_o is the fluorescence emission of completely oxidized protein and ΔF is the difference between the fluorescence emission of completely reduced and oxidized protein. k_{obs} is the peudo first-order rate constant.

Different k_{obs}'s are plotted against hirudin concentrations. The slope is the observed second-order rate constant of the reaction.

Isomerization of scrambled hirudin: 21 μM scrambled hirudin and 21 μM freshly reduced protein (DsbC, DsbA, thioredoxin and its variants) were incubated in the isomerization buffer (20 mM sodium phosphate, 130 mM sodium chloride, 0.13% polyethylene glycol 8000) at 25 °C. Aliquots of 120 μl were removed after different reaction times and quenched with 15 μl formic acid and 15 μl acetonitrile. Hirudin folding intermediates were separated by the same gradient as in the oxidative folding assay.

Chapter 3

Genetic selection designed to custom tailor the *in vivo* folding environment uncovers a chaperone called Spy[‡]

Abstract: Proteins are often unstable. Although much is known about protein folding *in vitro*[151], how proteins fold and function in the cell remains an unanswered question of fundamental importance[152]. To optimize the *in vivo* folding of proteins, we linked protein stability to antibiotic resistance[102], thereby forcing bacteria to effectively fold and stabilize proteins. When we challenged *Escherichia coli* to stabilize a very unstable protein, it responds by massively overproducing a small protein called Spy, which increases the level of unstable proteins up to nearly 700-fold. *In vitro* studies demonstrate that Spy functions as a very effective ATP-independent chaperone that suppresses protein aggregation and aids protein refolding. Our strategy opens up new routes for chaperone discovery and the custom tailoring of the *in vivo* folding environment. The crystal structure of Spy shows it to be an alpha helical protein that

[‡] This chapter is adapted from a manuscript "Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Quan, S., Tapley, T., Koldewey, P., Kirsch, N., Ruane, K., Pfizenmaier, J., Shi, R., Hofmann, S., Foit, L., Jakob, U., Xu, Z., Cygler, M. & Bardwell, J.C.A. Submitted". JCAB designed and supervised the project. I performed a large number of the experiments in this manuscript, including setting up the dual selection system, mutagenesis and selection, identification of Spy, quantification of Im7 and Spy levels, identification of Δ*spy* phenotype, and some of the *in vitro* chaperone activity assays. I contribute exclusively to Figs 3.1-3.9, Fig 3.11 and Tables 3.1-7 in this chapter and partially to Figs 3.10, 12, 13, and Fig 3.15. Other figures were prepared by TT, PK, MC, ZX and JCAB. TT and PK performed most of the *in vitro* activity assays (Figs 3.12, 13). TT and T.Franzmann performed analytical gel filtration and ultracentrifugation experiments (Fig 3.14). PK performed the casein binding and fluorescence anisotropy experiments (Fig 3.15). MC, KR and RS crystallized Spy. JCAB, MC and ZX analyzed the structure (Fig 3.16). JCAB analyzed Spy homologs (Fig 3.17). NK helped to analyze Im7 expression and made some strains and tested initial condition for chaperone assays. JP helped to search for phenotype and made some strains. SH helped with some spot titer experiments. UJ and LF gave technical support and conceptual advice. JCAB wrote the manuscript and I contributed.

forms cradle-shape dimers with highly apolar concave surface unlike the structure of any previously solved chaperone. We propose a model for Spy action where it binds the exposed hydrophobic segments of unfolded proteins on the concave side of the cradle, with the flexible N terminus providing additional contacts. Spy appears to represent the prototypical member of a new class of small chaperones that facilitate protein refolding in the absence of energy cofactors.

3.1 Introduction

The folding of many proteins is assisted by molecular chaperones and other folding helpers in the cell[153]. Many chaperones act by inhibiting off pathway events, such as aggregation and serve a broad range of substrates in a stoichiometric manner. *In vitro* assays for chaperone activity are thus almost by necessity, relatively insensitive, making these assays more useful in studying previously identified chaperones than in identifying new chaperones in crude cell lysates. Instead of being discovered directly because of their capacity to assist in *in vivo* protein folding, many chaperones were initially characterized because of their induction by stress conditions[154]. The lack of a sensitive and general *in vivo* assay for chaperone activity led us to wonder how complete the list of known chaperones is.

Previously, our laboratory developed a genetic system that directly links increased protein stability to increased antibiotic resistance, which provides a selectable and quantitative *in vivo* measure of protein stability[102]. We reasoned that this system might be used to directly select for bacterial variants with improved protein folding characteristics, and in doing so, uncover new chaperones. The selection system makes use of a tripartite fusion between β-lactamase and unstable proteins, which effectively links the stability of the inserted proteins to the penicillin resistance of the host strain. Previously it has been shown by our laboratory that β-lactamase tolerates the insertion of a well-folded protein in a surface loop of β-lactamase at amino acid 197, and still retains enzymatic activity[102]. However, insertion of unstable proteins in this site negatively affects β-lactamase activity and decreases penicillin resistance ($PenV^R$) *in vivo*[3]. This strategy allowed easy selection

for stabilized protein variants of a well characterized protein, immunity protein 7 (Im7)[102].

Now we hypothesized that a similar strategy might actually be used to select for host variants that alter the *in vivo* folding environment of individual proteins by enhancing specifically or generally protein folding or stability. We report here the development of an enhanced dual selection strategy based on the simultaneous use of two selectable markers, β-lactamase and DsbA. Both of the two markers provide readouts for *in vivo* protein folding/stability. We thought that this strategy might provide a direct route toward chaperone discovery, since historically, chaperone coexpression has often been used to enhance the expression of unstable proteins[155]. Considering our system functions in the periplasm, it also may provide an opportunity to explore protein folding in what has generally been considered to be a relatively chaperone poor environment.

Upon selection for host variants that stabilized a very unstable Im7 mutant, we obtained host variants that massively overproduce a previously not well characterized protein called Spy, which increases the expression level of unstable proteins up to nearly 700-fold. *In vitro* studies demonstrate that Spy functions as a very effective ATP-independent chaperone that suppresses protein aggregation and aids protein refolding. We obtained the crystal structure of Spy and found it to be the founding member of a new class of chaperone possessing a novel cradle-shaped structure.

Our overall experimental scheme is illustrated in Figure 3.1.

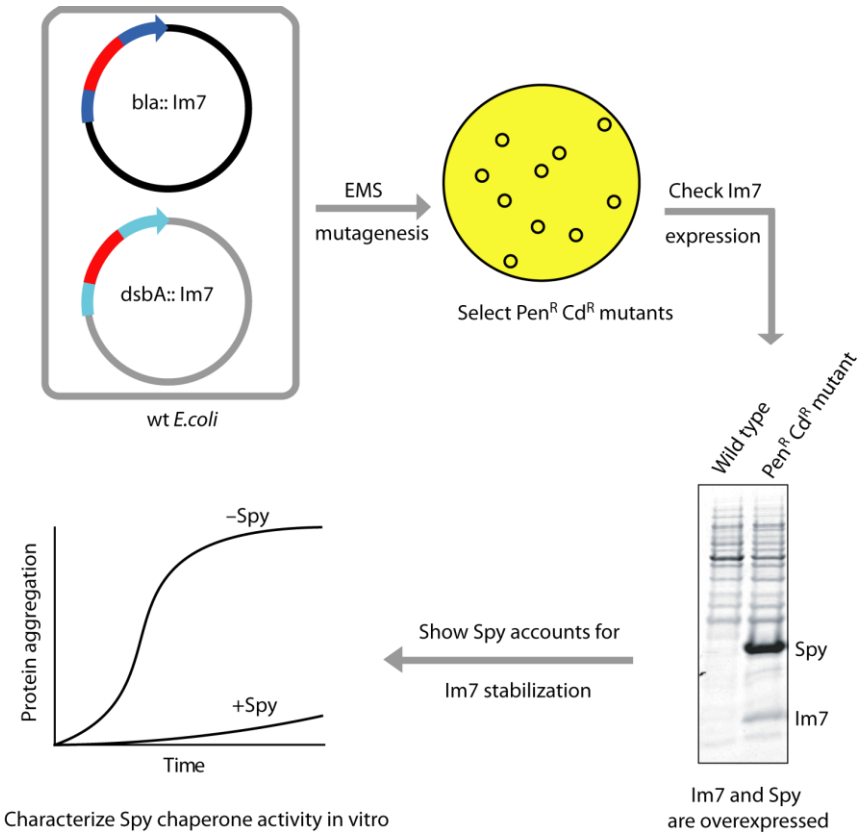


Figure 3.1 Overall experimental scheme
Fusion constructs that link the stability of the poorly folded Im7 protein to penicillin resistance (*bla::Im7*) and to cadmium resistance (*dsbA::Im7*) were introduced into the same *E. coli* strain. Following host chromosomal mutagenesis by ethyl methanesulfonate (EMS) treatment, mutants that simultaneously enhance both penicillin and cadmium resistance were selected. The goal is to find host variants that stabilize Im7. Levels of Im7 (with or without a fusion context) present in these strains were measured. The mutants were found to massively increase levels of Im7 and a host protein called Spy. Genetic experiments showed that increased Spy levels are necessary and sufficient to increase Im7 levels. The Spy protein was examined for molecular chaperone activity *in vitro* and found to be highly effective as a chaperone in preventing protein aggregation and aiding in refolding.

3.2 Results

3.2.1 Design of a dual tripartite fusion system to select for optimized protein folding *in vivo*

We designed a dual selection system whereby the same unstable test protein is inserted into the middle of two resistance markers. As the first selectable marker, we used our previously established β-lactamase system (Figure 3.2 a). As a second selectable marker, we constructed tripartite fusions with DsbA, a periplasmic protein that catalyzes disulfide bond formation and thereby encodes cadmium resistance ($CdCl_2^R$)[121] (Figure 3.2 b). Similar to the β-lactamase system, the DsbA system only confers resistance of the bacteria towards cadmium when the test protein inserted into DsbA folds properly.

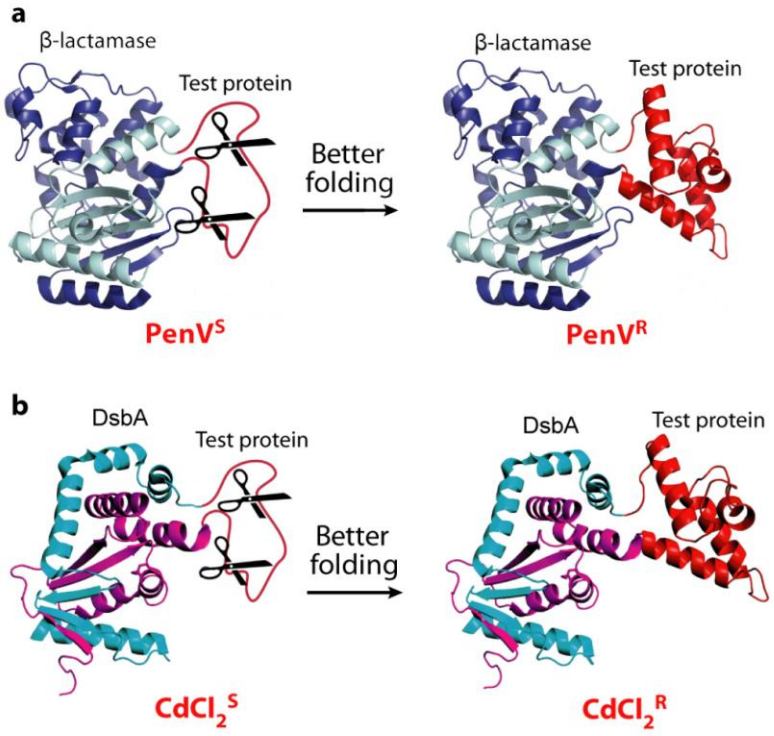


Figure 3.2 A dual fusion selection for enhancing *in vivo* protein stability
a, Unstable/poorly folded test proteins inserted into β-lactamase are degraded by cellular proteases producing penicillin sensitive ($PenV^S$) strains. Improving the folding of the test proteins so that they fold into stable strucutures increases penicillin resistance ($PenV^R$)[102]. This part of the figure is modified from Fig.1 in ref. [102]. **b**, Insertion of unstable test proteins into DsbA renders the strains expressing the fusion proteins sensitive to cadmium ($CdCl_2^S$). Improving the folding of the test proteins increases the strains' resistance towards cadmium ($CdCl_2^R$).

These two markers provide an independent measure of the folding/stability of the inserted protein and should avoid intrinsic antibiotic resistant mutations that are irrelevant to protein folding (Figure 3.3). In strains that co-express both tripartite fusions, we reasoned that host mutations that simultaneously increased penicillin V and cadmium resistance would likely be positively affecting the only property that these markers appear to have in common; namely, the stability of the inserted protein.

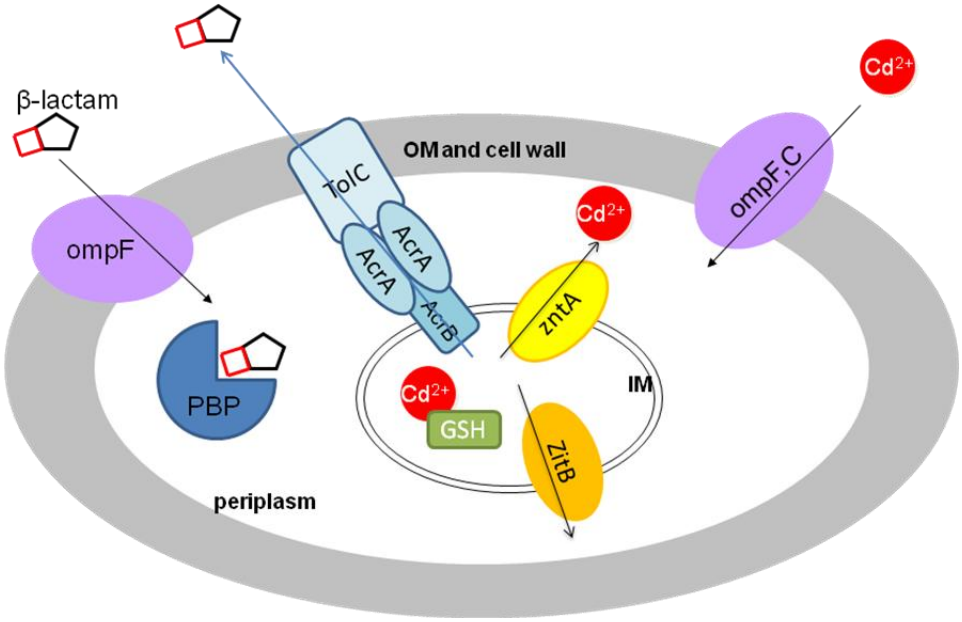


Figure 3.3 Mutations contributing to intrinsic PenV^R and Cd^R
Bacteria develop intrinsic drug resistance by changing the membrane permeability to decrease the drug uptake, by enzymatic inactivation of drugs, by target modification, and by active export of drugs. Chromosomal mutations contributing to penicillin resistance include mutations in the penicillin target, cell wall transpeptidases (PBPs), and in proteins constituting the efflux system for β-lactam antibiotics, such as AcrAB-TolC[156]. Mutations involved in conferring cadmium resistance are less characterized. Such mutations probably include mutations leading to overexpression of cytoplasmic GSH which quenches the toxicity of cadmium by binding to it, and the heavy metal transporters, ZntA and ZitB[157], which pump cadmium out from the cytoplasm. There are few overlapping mechanisms for bacteria to acquire intrinsic PenV and cadmium resistance. However, precautions must be taken when analyzing the phenotypes because it was hypothesized that both PenV and cadmium enter the periplasm through the outer membrane porins, OmpF and OmpC[158, 159]. One can imagine that mutations decreasing the levels of these porins might increase bacterial resistance to PenV or cadmium. Accordingly, I confirmed the absence of such mutations when I analyzed the mutant strains as I describe in the following chapters. (qRT-PCR experiment showed that the mRNA levels of OmpC and OmpF in mutant strains were not significantly different from the levels in wild type strains (Table 3.4 on page 74).) PenV^R: penicillin V resistance; Cd^R: cadmium resistance.

We first needed to identify a site within DsbA that could tolerate insertions and where the stability of the inserted protein affected the activity of DsbA. To identify a site within DsbA that might tolerate protein insertions, we examined positions in DsbA that can tolerate circular permutation[160], as the two events impose similar requirements on the structure of DsbA. In circular permutations, the N- and C-termini of a protein are fused and new termini are created. A variant of DsbA that is circularly permuted at amino acid T99 (Q100-T99) has very similar enzymatic properties and structure compared to wild-type DsbA[160]. We found that amino acid T99 of DsbA tolerated insertion of a 30-amino acid serine glycine linker. Thus, we inserted our test proteins into the linker we had inserted at amino acid T99 (Figure 3.4).

Immunity protein 7 (Im7) is a ~10 kDa helical protein with a well characterized folding pathway[161]. It binds endonuclease colicin E7 and inhibits the bactericidal activity of E7, allowing bacteria producing E7 to survive[162]. Insertion of variants of Im7 into amino acid T99 of DsbA revealed an excellent correlation between cadmium resistance and Im7 variant stability (Figure 3.5 b). A similar correlation between PenV resistance and protein stability was previously shown when variants of Im7 were inserted into β-lactamase[102] (Figure 3.5 a). Cadmium resistance also correlated well with penicillin resistance when these same Im7 variants were inserted into β-lactamase (Figure 3.5 c). The highest resistance to both antimicrobials came from the most s Im7 variants. These results suggested that host variants that act to improve the folding of Im7 and stabilize it should increase both penicillin and cadmium resistance and thus should be easily selected for on this basis.

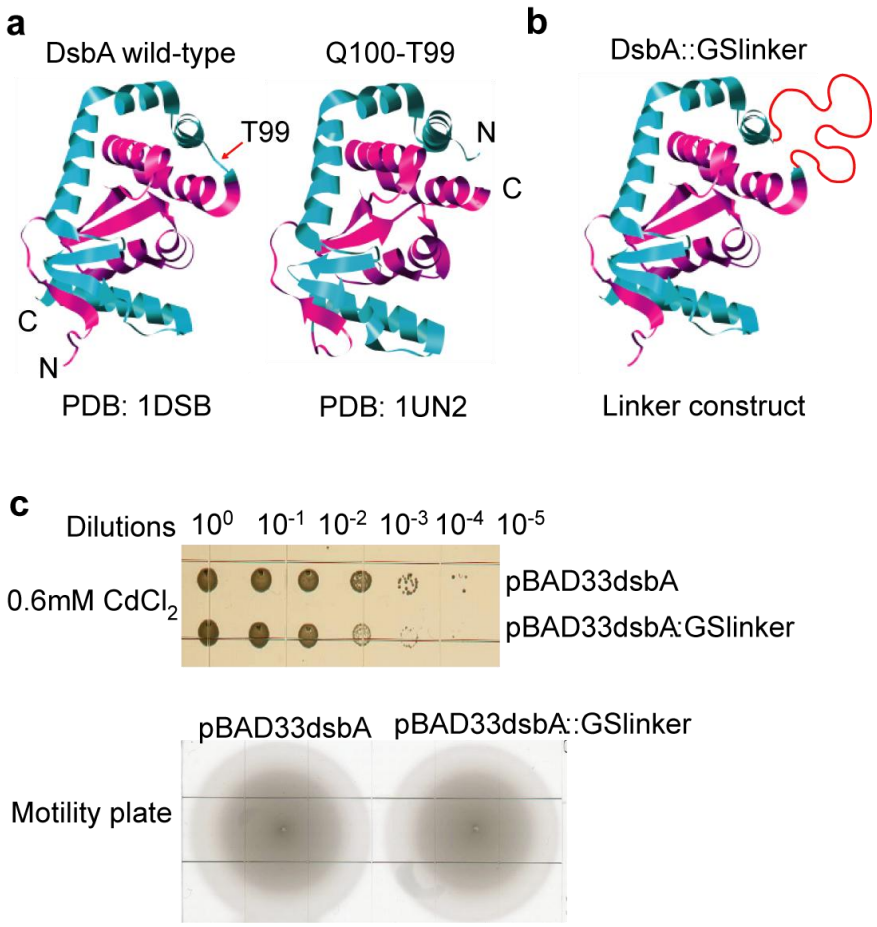


Figure 3.4 DsbA tolerates insertions at amino acid T99

In circular permutations, the original N- and C-termini of a protein are fused and new termini are created. We reasoned that if DsbA can tolerate circular permutation, it may also tolerate protein insertion at the same position as where it is circularly permuted, as the two events impose similar requirements on the structure of DsbA. **a**, Variants of DsbA that are circularly permuted at amino acid T99 (Q100-T99) are very similar in enzymatic properties[160] and in structure to wild-type DsbA, as seen by comparing the right and left structures in panel (**a**). **b**, Diagram of the protein construct (named DsbA::GSlinker) obtained by inserting a 30-amino acid glycine-serine linker $(GGGS)_2SS(GS)_4(GGGS)_2$ at position T99 in a pBAD33 plasmid-carrying wild type DsbA. **c**, The phenotypes of the pBAD33dsbA::GSlinker construct were almost indistinguishable from a wild-type pBAD33dsbA clone in two tests of DsbA activity: cadmium resistance (top), and bacterial motility (bottom). The spot titer experiment for cadmium resistance was performed as described in methods in Chapter 3. The motility assay was performed as described in experimental procedures in Chapter 2, except for the addition of 0.2% arabinose into the media to induce the expressing of DsbA and DsbA::GSlinker.

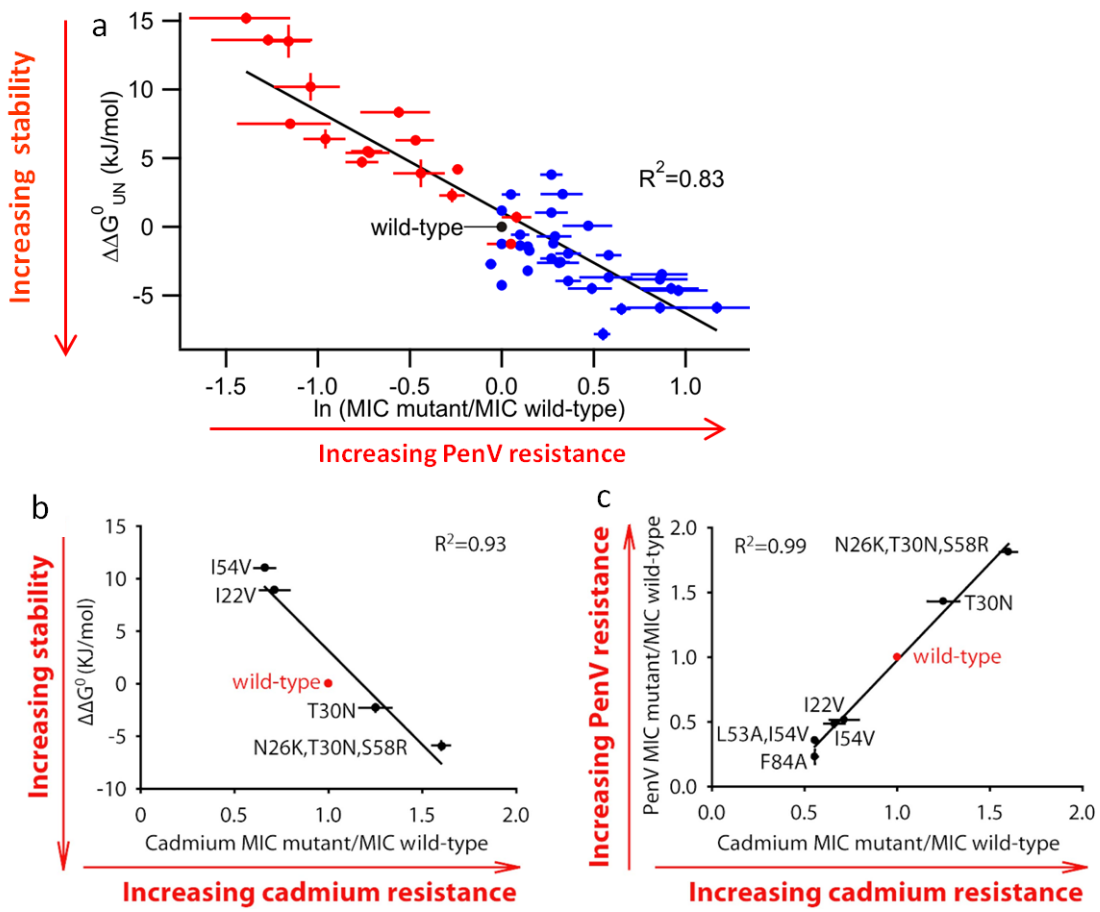


Figure 3.5 Antimicrobial resistance correlates with stability of Im7 variants.
a, Thermodynamic stability (given as $\Delta\Delta G^0$ relative to wild type Im7 and was determined by urea titration as described in Ref. [102]) of Im7 protein variants is highly correlated with the *in vivo* Penicillin resistance of tripartite β-lactamase fusions containing these variants [given as minimal inhibitory concentration (MIC) of PenV for variants relative to MIC for wild-type]. This part of the figure is adapted from Fig.4 in Ref. [102]. Red dots represent the published destabilizing Im7 variants; blue dots represent the selected Im7 variants on the basis of their increased PenV resistace [102]. **b**, Thermodynamic stability of Im7 protein variants[102] is correlated with the MIC of cadmium of cells expressing the tripartite DsbA fusions containing these variants. **c**, Penicillin resistance of cells expressing the β-lactamase-Im7 fusion proteins correlates well with their cadmium resistance when they also express the DsbA-Im7 fusion proteins with the same Im7 variants. Error bars indicate the standard deviation of 3 independent measurements.

3.2.2 Selection to identify mutants that simultaneously increase penicillin and cadmium resistance

With the dual selection system, we decided to search for host mutations that enable the proper folding of Im7-L53AI54A, a very unstable Im7 variant[163]. We randomly mutagenized strain SQ1306, which contains the DsbA-Im7-L53AI54A fusion and the β-lactamase-Im7-L53AI54A fusion, with the mutagen ethyl methanesulfonate (EMS). We selected first for variants with enhanced penicillin resistance, and then screened these variants for enhanced cadmium resistance. Several precautions were taken to reduce the probability that any increased resistances observed were due to factors such as simple reversion of the destabilizing mutations, transcriptional changes, plasmid copy number alterations, toxin resistance, or fusion specific effects. These included cloning the Im7 L53AI54A mutant under the control of two different promoters (β-lactamase promoter and pBAD), on plasmids with different copy number control mechanisms (pBR322 and p15A origin of replication), and using two different resistance markers in two different fusion contexts.

Plate screens revealed that about 12% of our PenV resistant strains had simultaneously acquired resistance to 0.5 mM $CdCl_2$. Ten independently isolated mutant strains (EMS1-EMS10) that were resistant to both PenV and $CdCl_2$ were selected for further analysis. The strains generated in this study are listed in Table 3.6 on page 92. Eight of these ten host variants contained elevated levels of the tripartite β-lactamase Im7 fusion protein (Figure 3.6) measured by western blot of total cell lysates. EMS4 and EMS9, containing the highest amount of the tripartite fusion protein and being resistant to high concentrations of both PenV and cadmium, were chosen for further analysis. Since preparation of the total cell extracts involved solubilization by boiling in a solution containing 2% SDS, this procedure is designed to measure the total amount of β-lactamase fusion protein present in the cells, including both soluble and insoluble protein. This is in contrast to the periplasmic extracts prepared for the quantifications of Im7 protein expressed in the absence of the fusion shown in Table 3.1 and Table 3.2 (on page 65-66) in which only soluble protein is measured; insoluble material will be left behind in the cellular fraction.

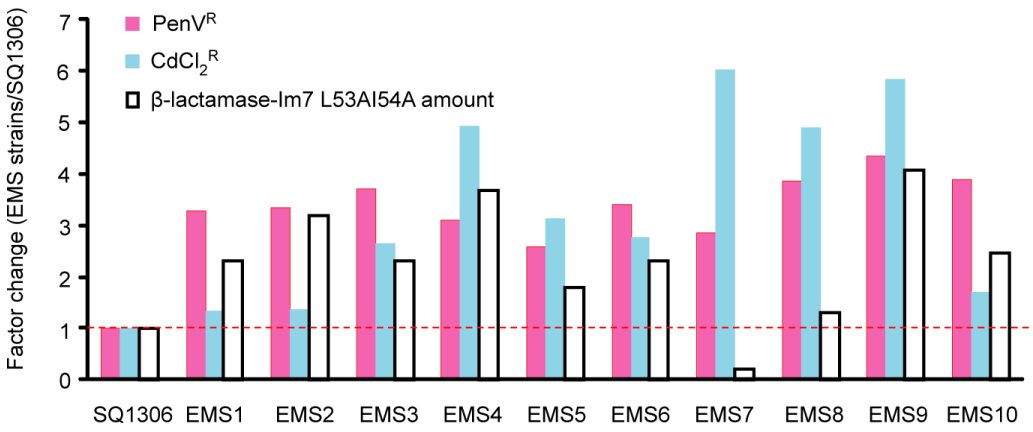


Figure 3.6 Penicillin V and $CdCl_2$ resistant strains ($PenV^R$ and $CdCl2^R$) show increased levels of β-lactamase fusion protein

The penicillin V and cadmium chloride resistance of the starting strain SQ1306 and strains EMS1 through EMS10 were measured by spot titer experiment as described in Materials and Methods. The levels of β-lactamase fusion protein in these strains were determined by Western blot analysis of total cell extracts. The bar graphs show levels relative to the starting strain (SQ1306) and represent the average of two measurements.

One variant, EMS7, was the only one that showed a decreased level of β-lactamase-Im7 L53AI54A. However, EMS7 was extremely resistant to cadmium, suggesting a different mechanism for this variant to acquire resistance. This raised an interesting question whether the dual selection system is necessary and sufficient to exclude mutations that are unrelated to protein folding. I will discuss this question in details in Chapter 4. However, the combination of the dual selection system with an additional screening technique (such as detection of the levels of the fusion proteins) clearly demonstrated the applicability of our approach to identifying possible host folding modulators.

To confirm that we had indeed obtained host variants that were affecting the folding and consequently protein levels of the Im7 protein itself, we expressed Im7 L53AI54A, Im7 wild type and two other destabilized Im7 mutant proteins on their own in the absence of the fusion using a third promoter (pTrc) carried by a third origin of replication (CDF). In contrast to the periplasmic extracts prepared from the un-mutated $PenV^S/CdCl_2^S$ strain SQ1306, which contained only low amounts of the Im7 proteins, both EMS4 and EMS9 strain backgrounds accumulated large amounts of Im7 proteins

(Figure 3.7). Quantification of the amounts of Im7 proteins present in periplasmic extracts using an Agent bioanalyzer showed that Im7 proteins made up 7-10% of the periplasmic contents in EMS4 and EMS9, a 34- to 92-fold increase over the amount present in wild type strains (Figure 3.7, Table 3.1 and Table 3.2). The Im7 proteins were released by periplasmic extraction in the absence of detergent, indicating that they accumulate in a soluble form.

3.2.3 Spy overexpression is necessary and sufficient to enhance Im7 expression

In examining the periplasmic extracts of the Im7 overexpressing strains, we noted the strong induction of a 15.9 kD protein, which we identified by mass spectrometry as Spy (Spheroplast protein Y)[164]; this protein accounted for up to 43% of total periplasmic content in eight of the ten tested $PenV^R/CdCl_2^R$ EMS mutant strains (Figure 3.7, Figure 3.8 and Table 3.2). Spy expression is mainly under the control of the Bae and the Cpx extracytoplasmic stress response systems[57, 164, 165], which are induced by a variety of stress conditions known to cause protein unfolding and aggregation[166, 167].

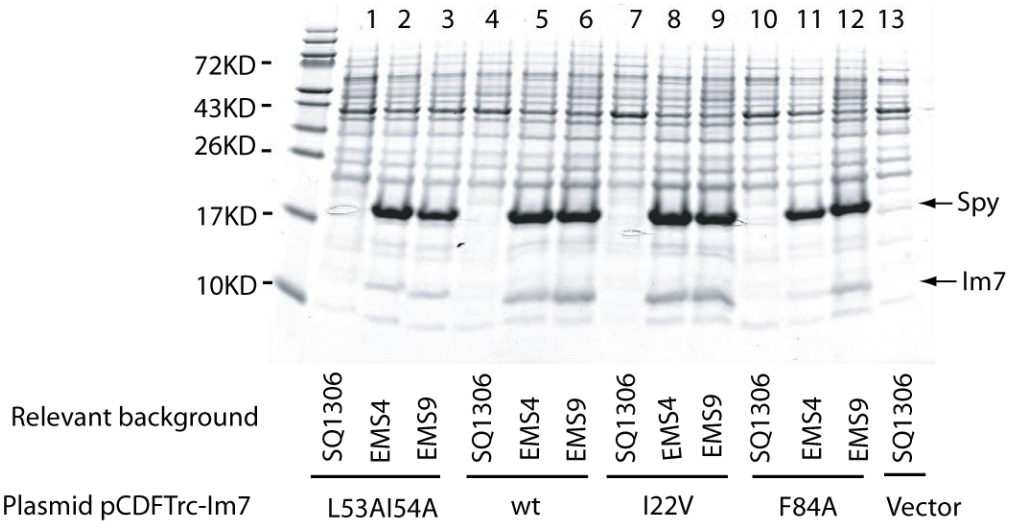


Figure 3.7 Im7 and Spy are abundant in the periplasm of EMS strains
The $PenV^R$ and $CdCl_2^R$ resistant mutants EMS4 and EMS9 and the unmutagenized SQ1306 strain were transformed with plasmids encoding wild-type Im7 (wt) or the destabilized variants: L53AI54A, I22V, and F84A[102, 163]. Periplasmic extracts were prepared and analyzed by SDS-PAGE. Im7 became 7-10% of total periplasmic proteins in EMS4 and EMS9. Strikingly, Spy made up to 43% of the periplasmic contents in EMS4 and EMS9 but almost invisible in SQ1306.

Table 3.1 Spy induction increases Im7 levels substantially

Strains compared	Relevant host genotypes* compared	Im7 variant	Fold increase in Im7 level
SQ1406/SQ1414	EMS4/wt	L53AI54A	66±28
SQ1410/SQ1414	EMS9/wt	L53AI54A	92±44
SQ1805/SQ1809	*spy*$^{++}$/wt	L53AI54A	280±145
SQ1826/SQ1830	*ΔbaeSR, spy*$^{++}$/ *ΔbaeSR*	L53AI54A	684±162
SQ1405/SQ1413	EMS4/wt	wt	34±12
SQ1409/SQ1413	EMS9/wt	wt	43±24
SQ1804/SQ1808	*spy*$^{++}$/wt	wt	165±93
SQ1825/SQ1829	*ΔbaeSR, spy*$^{++}$/ *ΔbaeSR*	wt	211±83
SQ1407/SQ1415	EMS4/wt	I22V	51±9
SQ1411/SQ1415	EMS9/wt	I22V	48±23
SQ1806/SQ1810	*spy*$^{++}$/wt	I22V	103±14
SQ1827/SQ1831	*ΔbaeSR, spy*$^{++}$/ *ΔbaeSR*	I22V	470±176
SQ1408/SQ1416	EMS4/wt	F84A	43±18
SQ1412/SQ1416	EMS9/wt	F84A	67±16
SQ1807/SQ1811	*spy*$^{++}$/wt	F84A	694±243
SQ1828/SQ1832	*ΔbaeSR, spy*$^{++}$/ *ΔbaeSR*	F84A	589±102

Data indicate extent of Im7 overproduction in various strains (see Table 3.6 on page 92 for details of strains). Im7 L53AI54A, wt, and two other destabilizing variants, I22V and F84A were expressed in EMS4, EMS9, and the unmutagenezied strain. In the EMS4 and EMS9 backgrounds, Spy was overexpressed due to the constitutive mutations in the *baeS* gene. To directly assess the overexpression effect of Spy and to determine whether other BaeSR regulated gene products also contribute to the increased level of Im7, Spy was expressed under a Trc promoter in a wild type strain and a *ΔbaeSR* strain, which are transformed with the various Im7 plasmids, respectively. Fold increases in Im7 levels were quantified with an Agilent Bioanalyzer 2100 by direct comparison between periplasmic extracts prepared from various pairs of strains (see methods in Chapter 3). Values are the average of measurements from 3 biological samples ± standard deviation.

*spy^{++} designates cells that overexpress Spy from pTrc-spy.

Table 3.2 Im7 accumulates in the presence of Spy overexpression

Relative genotype*	Actual strain	Im7 variant	Spy%	Im7%
SQ1306	SQ1414	L53AI54A	N.D.	0.1±0.07
EMS4	SQ1406	L53AI54A	48±7	7±2
EMS9	SQ1410	L53AI54A	35±13	8±2
wt	SQ1809	L53AI54A	N.D.	0.02±0.01
spy++	SQ1805	L53AI54A	43±2	7±3
ΔbaeSR	SQ1830	L53AI54A	N.D.	0.01±0.004
ΔbaeSR, spy++	SQ1826	L53AI54A	40±4	7±1
SQ1306	SQ1413	wt	N.D.	0.3±0.1
EMS4	SQ1405	wt	42±5	10±1
EMS9	SQ1409	wt	35±11	10±1
wt	SQ1808	wt	N.D.	0.1±0.02
spy++	SQ1804	wt	40±1	7±1
ΔbaeSR	SQ1829	wt	N.D.	0.1±0.04
ΔbaeSR, spy++	SQ1825	wt	41±2	10±4
SQ1306	SQ1415	I22V	N.D.	0.2±0.1
EMS4	SQ1407	I22V	43±5	9±2
EMS9	SQ1411	I22V	36±12	8±1
wt	SQ1810	I22V	N.D.	0.1±0.01
spy++	SQ1806	I22V	35±9	6±1
ΔbaeSR	SQ1831	I22V	N.D.	0.01±0.006
ΔbaeSR, spy++	SQ1827	I22V	39±3	6±1
SQ1306	SQ1416	F84A	N.D.	0.2±0.1
EMS4	SQ1408	F84A	44±5	8±2
EMS9	SQ1412	F84A	31±13	7±1
wt	SQ1811	F84A	N.D.	0.01±0.004
spy++	SQ1807	F84A	40±8	6±1
ΔbaeSR	SQ1832	F84A	N.D.	0.01±0.003
ΔbaeSR, spy++	SQ1828	F84A	40±5	6±1

Percentage of Im7 and Spy in the total periplasm in different strains strains (see Table 3.6 on page 92 for details of strains) was analyzed with an Agilent Bioanalyzer 2100 as described in Materials and Methods. Values are the average of measurements from 3 biological samples ± standard deviation.

*spy^{++} designates cells that overexpress Spy from pTrc-spy.

N.D., not detectable.

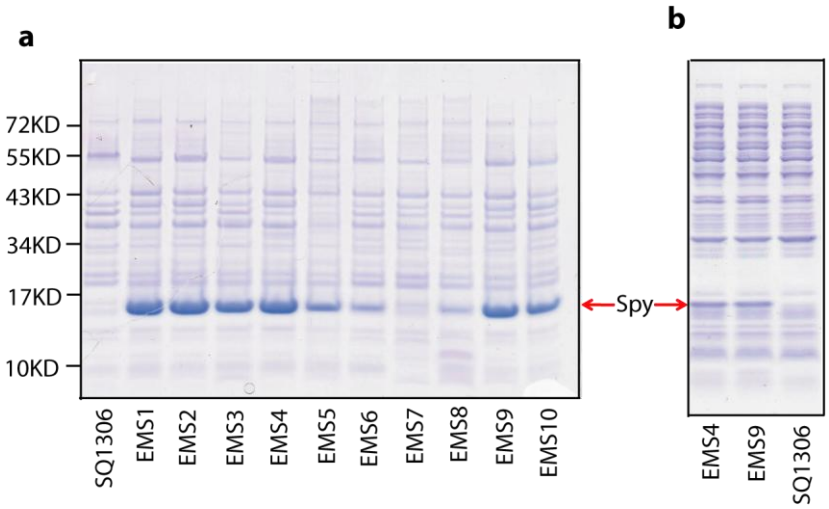


Figure 3.8 The periplasmic protein Spy is massively induced in the majority of $PenV^R/CdCl_2^R$ strains

a, Periplasmic extracts were prepared from the unmutated strain SQ1306 and the $PenV^R/CdCl_2^R$ strains EMS1-EMS10. The periplasmic proteins were analyzed by SDS-PAGE followed by Coomassie blue staining. All $PenV^R/CdCl_2^R$ strains showed increased levels of Spy except EMS7, which was the only host variant to show lower levels of β-lactamase fusion in our original screen (Figure 3.6). Spy was present in only moderate levels in EMS 8, the strain that had only slightly increased levels of the β-lactamase Im7 fusion protein (Figure 3.6). **b**, Spy is visible in whole cell extracts from EMS4 and EMS9.

3.2.3.1 Single mutation in *baeS* caused overexpression of Spy

To identify the mutation(s) that caused the massive up-regulation of Spy expression in EMS4 and EMS9, we sequenced the *spy* gene, its regulators *cpxARP* and *baeRS*, and a number of other candidate genes, *surA*, *skp*, *fkpA*, *prc*, *degP*, *ptr*, *ompP*, *ompT*, *rcsABCDF*, and *mdtABCD* and their upstream regulator sequences. We discovered that both EMS4 and EMS9 contained mutations in *baeS* but not in any of the other sequenced genes or regulatory sequences. We then sequenced the *baeS* gene in all other independently isolated $PenV^R/CdCl_2^R$ strains (EMS1-10) and determined that it was mutated in all strains except EMS7, the mutant that failed to overproduce Spy or Im7 (Table 3.3). BaeS is a putative histidine kinase that together with the proposed response regulator BaeR makes up the two-component BaeSR envelope stress response regulation system[57, 164-166]; this system regulates *spy* and a few other periplasmic stress genes.

Table 3.3 Mutations in baeS are found in all but one independently isolated $PenV^R/CdCl_2^R$ strain

Strain	Mutation in *baeS*
EMS1	R150W
EMS2	R150W
EMS3	P192L
EMS4	R416C
EMS5	R150W
EMS6	D268N
EMS7	None
EMS8	D268N
EMS9	E264K
EMS10	D268N

The *baeS* gene had in total 5 different mutations identified in 9 EMS strains. EMS7 was the only mutated strain that did not contain a mutation in *baeS*. It was also the only mutated strain which did not increase the amount of bla-Im7 L53AI54A tripartite fusion protein (Figure 3.6).

BaeS has a periplasmic domain, two transmembrane helices, and a cytoplasmic domain, which can be further divided into a highly conserved histidine kinase domain and the HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) linker. HAMP is a common structural element between the sensor domain and the kinase domain of a histidine kinase, presumably mediating the signal transduction between the two domains[168]. Mutations in BaeS are mapped to various regions of BaeS (Figure 3.9). R150W, isolated independently three times, is in the periplasmic domain of BaeS. P192L is in the HAMP linker of BaeS. The rest of the mutations are located in the histidine kinase domain of BaeS. Therefore, BaeS might be altered in the ability to sense periplasmic stress, to transmit the signal from its sensor domain to its kinase domain, or to modulate its kinase activity. P192, E264, D268, and R416 are highly conserved amino acids among BaeS homologs from different bacterial species. Transcriptional analysis of EMS4 revealed not only a massive up-regulation of *spy* mRNA but also a significant induction of the two other clearly defined downstream targets of BaeSR: *mdtA* and *acrD*. Moving the *baeS-R416C* mutant allele from EMS4 into a fresh unmutagenized strain background by P1 transduction gave very similar inductions for these three genes, suggesting that *baeS-R41C* was sufficient to cause the constitutive activation of the

BaeSR envelope stress response (Table 3.4 on page 74). Other mutations might have a similar effect but we have not yet tested them.

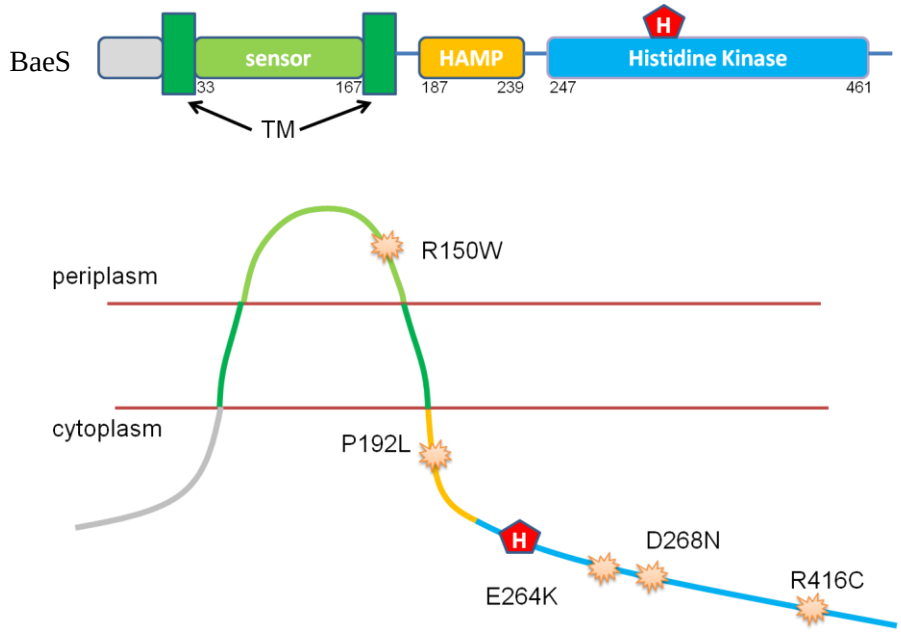


Figure 3.9 Mutations mapped to various regions of BaeS
Predicted domain division of BaeS. BaeS has a periplasmic sensor domain and a cytoplasmic histidine kinase domain, connected by a HAMP linker. TM: transmembrane helices. Mutations in BaeS are located on various regions of BaeS. Information on the predicted domains of BaeS was acquired from http://www.uniprot.org/uniprot/P30847.

3.2.3.2 Mutations in *baeS* are necessary and sufficient for increased PenV^R and Im7 levels

To establish whether the mutation present in EMS4 (*baeS-R416C*) is sufficient for enhanced Im7 expression, we conducted a number of genetic experiments. We moved the *baeS-R416C* mutant allele from EMS4 into a fresh, un-mutagenized strain and found that the levels of Im7 and PenV^R were indeed very similar to the levels observed in EMS4. These results demonstrate that the *baeS-R416C* mutation is sufficient to enhance Im7 expression. In the reciprocal experiment we replaced the *baeS-R416C* mutation from EMS4 with a wild type sequence, resulting in the high PenV sensitivity originally observed in the unmutated parental strain, demonstrating that the *baeS-R416C* mutation is also necessary for increased PenV^R and Im7 level.

The experiments were performed as follows. The *baeS-R416C* mutant allele was moved into the wild-type SQ1306 background by P1 co-transduction with a linked marker *yegL::kan* (~10 kb upstream of *baeS*) from the Keio knockout collection[169]. Penicillin resistance co-transduced at a rate of 100% with the *baeS-R416C* mutation. (All kan^R colonies tested (including SQ1640) that received the *baeS-R416C* mutation as assayed by PCR amplification and sequencing the *baeS* gene were penicillin resistant.) All kan^R transductants tested that received only the *yegL::kan* marker but remained wild-type for *baeS* (including SQ1594) were penicillin sensitive (Figure 3.10 a). We were able to show that *baeS* mutations were sufficient to enhance the level of Im7 once transferred to a wild type strain (Figure 3.10 b).

In the reciprocal experiment, when the *baeS* wild-type gene was moved by P1 co-transduction with the *yegL::kan* marker into the EMS4 background, each time the *baeS* wild-type gene was acquired (*e.g.*, SQ1608), the EMS4 background became penicillin sensitive. Each time the *baeS-R416C* mutation was not replaced by wild type sequence (*e.g.*, SQ1606), the strain remained penicillin resistant. Identical results were obtained upon moving the *baeS-E264K* mutant allele in and out of the wild-type and EMS9 strain backgrounds. These results clearly demonstrated that the mutations in *baeS* are necessary for the phenotypes we observed (Figure 3.10 a).

To test if alterations in *spy* expression or other genes under *baeS* control were responsible for $PenV^R$, we disrupted the *spy*, *baeS, baeSR*, *mdtABCD,* and *acrD* genes in EMS4. The *spy* gene*, baeS* alone, and *baeSR* were necessary for PenV resistance, but *mdtABCD* and *acrD* were not (Figure 3.10 c).

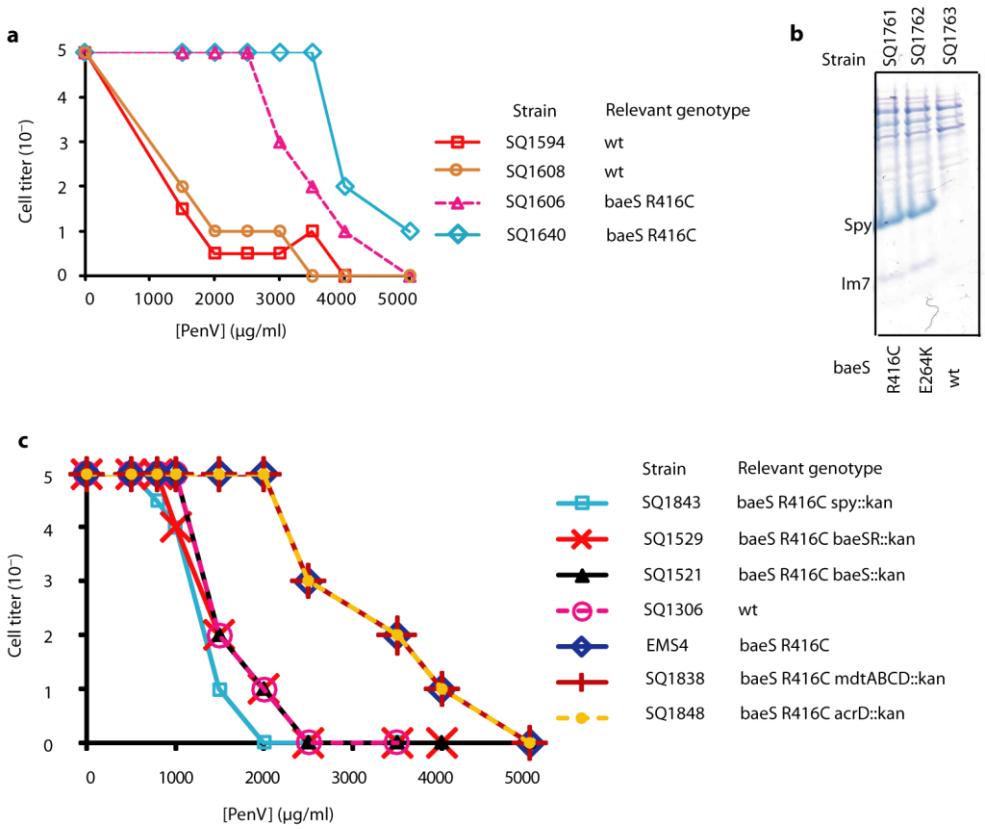


Figure 3.10 baeS mutations are necessary and sufficient to cause Spy and Im7 overproduction

a, Log-phase cells of different strains were normalized to A_{600}=1, serially diluted and plated onto plates containing different concentrations of penicillin V. The highest dilutions of cells at which cells still grew at each PenV concentration were plotted against PenV concentrations to generate a titration curve. Penicillin resistance of EMS4 is due to the presence of the *baeS-R416C* mutation. Strains that contain the *baeS-R416C* mutation (SQ1606 and SQ1640) are penicillin resistant, whereas those that are wild-type for *baeS* (SQ1594 and SQ1608) are penicillin sensitive. See text for details. **b**, Im7 L53AI54A was overproduced in strains with wild-type backgrounds that have received (via P1 co-transduction) the *baeS-R416C* mutation (SQ1761) or the *baeS-E264K* mutation (SQ1762), but not in strain that remains wild-type for *baeS* (SQ1763). **c**, Similar spot titer experiments were performed as described in panel a. Deletion of the *spy* gene, *baeS* alone, and *baeSR* in the *baeS R416C* background dramatically decreased the PenV resistance to the levels similar to that of a wild type strain (SQ1306), indicating that these genes were necessary for PenV resistance. The *mdtABCD* gene and *acrD* were not necessary for PenV resistance since their deletions had no effect on PenV resistance. For ease in comparison, the order of the traces left to right corresponds to the strain order given top to bottom.

3.2.3.3 Spy overexpression is necessary and sufficient to lead to the phenotypes, even in the absence of a functional Bae pathway

To determine if Spy overproduction alone is responsible for the enhanced levels of the unstable Im7 protein, or whether other BaeSR regulated proteins are involved as well, we overproduced Spy from the pTrc promoter to levels similar to those seen in the EMS4 mutant, both in wild-type strains and in *baeSR* null strains. Overexpression of Spy, even in the absence of a functional BaeSR system, led to similar levels of soluble Im7 in the various destabilized Im7 mutants as were observed in EMS4 by SDS-PAGE analysis (Figure 3.11). More precise quantification of Im7 and Spy levels (obtained using an Agilent Bioanalyzer 2100) showed that upon Spy overexpression, periplasmic Im7 levels increased 100- to 700-fold (Table 3.1). We conclude that Spy overproduction is necessary and sufficient to increase the levels of soluble periplasmic Im7.

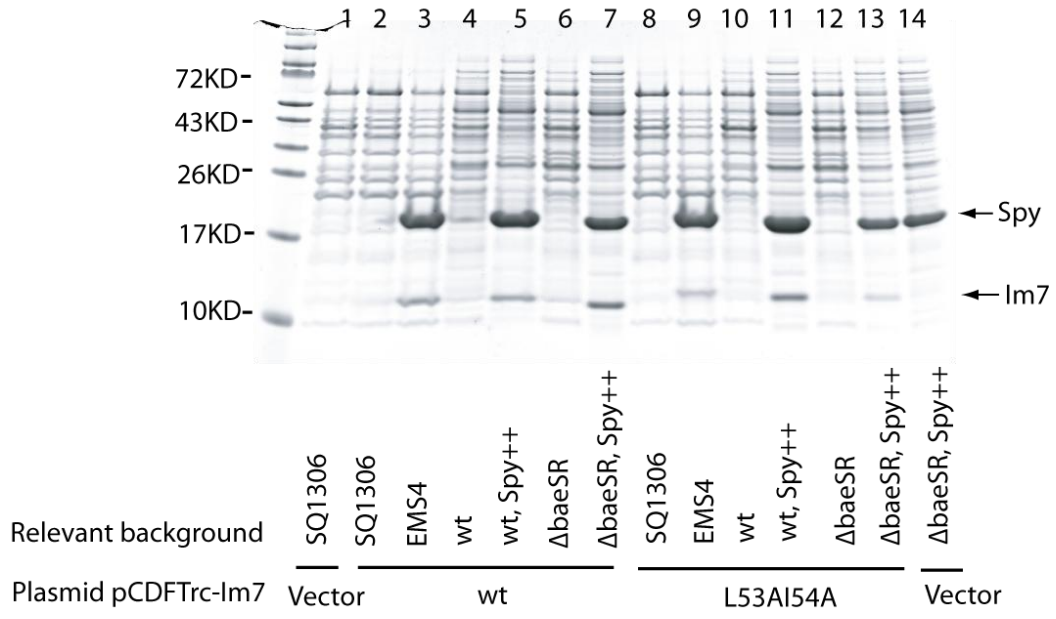


Figure 3.11 Overexpression of Spy increased the amount of soluble Im7 in the periplasm

Spy overexpression is sufficient to enhance Im7 levels in both wild-type and *ΔbaeSR* backgrounds to the levels seen in EMS4. Plasmid-encoded Im7 and its destabilized variants were expressed in SQ765 wild-type (wt) or *ΔbaeSR* backgrounds, with or without the co-expression of plasmid-encoded Spy (designated as Spy++).

3.2.4 Spy has chaperone activity *in vitro*

Though sequences homologous to Spy are present in a wide variety of bacterial species, very little was previously known about Spy function. Deletion of *spy* was reported to cause slight induction of *degP* and *rpoH*, two genes that are under the control of *rpoE*, a stress response system involved in outer membrane protein biogenesis[170]. This observation led to the suggestion that Spy may also be involved in this process[170]. However, we were unable to detect significant induction of these or other periplasmic stress regulated genes upon deletion of *spy* in our strain background with quantitative reverse transcription polymerase chain reaction (qRT-PCR) (Table 3.4), suggesting that *spy* deletion does not cause significant defects in membrane integrity.

Our results that overexpression of Spy leads to the accumulation of an otherwise highly unstable protein, however, suggested that Spy might function as a chaperone that facilitates protein folding in the bacterial periplasm. To assess chaperone activity, we purified Spy and tested its influence on the aggregation of a number of substrate proteins *in vitro*. We first tested the effect of Spy on the aggregation of thermally denatured malate dehydrogenase (MDH) and found that addition of increasing amounts of Spy led to a significant reduction of protein aggregation (Figure 3.12 a). Even sub-stoichiometric quantities of Spy effectively inhibited the aggregation process, suggesting that Spy is a highly efficient chaperone. Analysis of the effects of Spy on urea-denatured MDH (Figure 3.12 b) revealed a similar result and showed that Spy effectively prevents MDH from aggregating. We also found that Spy, which is strongly induced by ethanol[57], protects MDH from ethanol-mediated aggregation (Figure 3.12 e). Ethanol is a long known protein denaturant and one of the most potent inducers of the *E. coli* heat shock response[170]. Given that heat, urea and ethanol differ dramatically in their mode of protein unfolding, we concluded that Spy must have a general affinity for unfoled proteins.

To assess whether other proteins are protected against stress-induced protein unfolding by Spy as well, we compared the aggregation of chemically denatured aldolase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in the absence and presence of Spy (Figure 3.12 c,d). Again, we observed a strong suppression of protein aggregation when stochiometric amounts of Spy were present. We have shown that Spy

overexpression leads to the accumulation of higher levels of normally unstable proteins (Figure 3.7 and Figure 3.11). These results strongly suggested that Spy is a newly discovered general chaperone in the periplasm of *E. coli*.

Table 3.4 Effect of Spy overproduction and deletion of the *spy* gene on the message levels of various envelope stress regulated genes

Gene	Regulated by	Factor change (EMS4/ wt[1])	Range	Factor change (*baeS R416C*[2]/ wt[3])	Range	Factor change (*Δspy*[4]/ wt[5])	Range	Factor change (*spy++*[6]/ wt[5])	Range
spy	*bae,cpx*	**279**	204-342	**253**	201-295	NA	NA	**1.5×10^4**	1.2-2.0 $\times 10^4$
mdtA	*bae,cpx*	**74**	53-101	**61**	45-80	1.4	1.2-1.7	*4.5*	3.1-6.5
acrD	*bae*	**13**	9.7-18.7	**16**	13-19	1.0	0.8-1.2	1.7	1.1-2.3
rseP	*rpoE*	1.1	1.0-1.4	0.7	0.6-0.8	0.9	0.7-1.1	1.1	0.9-1.4
rpoH	*rpoE*	1.1	0.9-1.4	1.2	1.0-1.4	1.1	1.0-1.3	0.9	0.6-1.1
yebE	*cpx,rpoE*	1.4	1.0-1.8	1.6	1.3-1.8	1.3	1.1-1.6	0.9	0.7-1.3
cpxP	*cpx*	1.3	1.0-1.5	1.3	1.1-1.5	1.2	1.0-1.4	0.8	0.7-0.9
dsbA	*cpx*	NA	NA	1.5	1.1-2.2	1.1	0.7-1.6	1.1	1.0-1.3
osmB	*rcs*	1.6	1.3-2.0	0.7	0.5-0.9	2.3	1.4-3.6	0.4	0.2-0.9
pspA	*psp*	1.6	1.3-2.0	2.3	2.0-2.6	1.0	0.9-1.2	*3.1*	2.0-4.8
degP	*rpoE,cpx*	1.1	0.9-1.4	1	0.7-1.3	1.2	0.9-1.6	0.7	0.6-0.8
ompF	*rpoE,cpx*	0.3	0.3-0.4	1.3	1.5-2.1	0.9	1.0-1.4	2.1	0.4-0.7
ompC	*rpoE,cpx*	0.7	0.6-0.7	1.6	1.1-1.4	1.0	0.8-1.1	0.8	1.4-2.8
ibpA	*rpoH*	0.7	0.5-1.0	1.4	1.4-1.9	0.9	0.7-1.4	0.9	0.5-1.0
tolC		1.8	1.5-2.3	1.8	1.3-1.5	1.2	0.8-1.0	0.5	0.9-1.0
marR		1	0.8-1.1	0.9	0.8-1.0	1.0	0.8-1.4	1.4	1.1-1.8

Genes exhibiting more than a factor of 5 change in mRNA levels are shown in bold. Genes exhibiting a factor of 2.5 to 5 change in mRNA levels are shown in italics. The listed genes were chosen because they: (1) were the most induced in the various regulatory systems [as described previously[57], the factor change under an active regulatory system was: *mdtA* (5.3/*baeR*); *degP* (56.3/r*poE*); *rseP* (7.4/*rpoE*); *yebE* (9.8/*cpx*); *osmB* (11.4/*rcs*), and *pspA* (11.8/*psp*)], (2) contributed to intrinsic drug resistance (*marR*, *tolC*, *mdtA*, *acrD*, *ompC*, and *ompF*), or (3) function in protein folding (*degP*, *dsbA*, and *ibpA*). DegP is a protease, DsbA is a disulfide catalyst, and IbpA is a cytoplasmic chaperone.

NA, not applicable.

Strains used: 1. SQ1306, 2. SQ1698, 3. SQ1712, 4. SQ1731, 5. SQ765, 6. SQ1796. See Table 3.6 on page 92 for details.

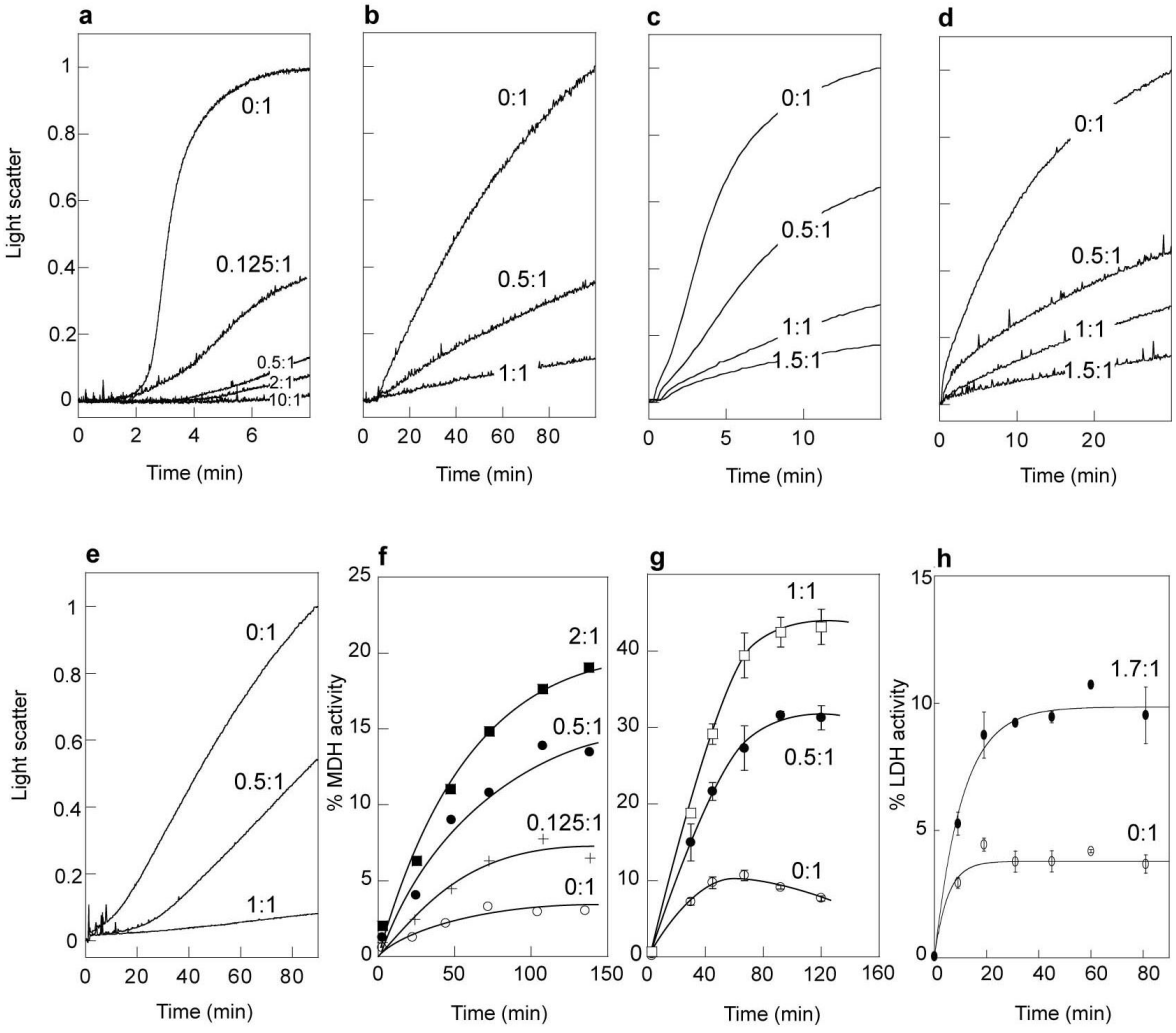


Figure 3.12 Spy has chaperone activity[§]

a–e, Spy suppresses protein aggregation as monitored by light scattering. Aggregate formation in the absence (0:1) and presence of increasing amounts of Spy (ratios given are Spy:substrate) was monitored for thermally-denatured MDH (**a**), urea-denatured MDH (**b**), guanidine-denatured aldolase (**c**), guanidine-denatured GAPDH (**d**) and ethanol-denatured MDH (**e**). **f-h**, Spy enhances protein refolding as assessed by recovery of enzymatic activity. Refolding was monitored in the absence (0:1) and presence of Spy for thermally-denatured MDH (**f**), guanidine-denatured MDH (**g**) and guanidine-denatured LDH (**h**). Error bars indicate the standard deviation of 3 independent measurements. Ultracentrifugation and gel filtration studies (Figure 3.15 on page 82) indicate that Spy is dimeric in solution so Spy concentrations are given as a dimer.

[§] I performed experiments as presented in panel a & f. The rest experiments were performed by Tim Tapley and Philipp Koldewey. Tim Tapley prepared this figure.

Most known ATP-dependent chaperones, such as the DnaK and GroEL systems, function as folding chaperones. They use cycles of ATP-binding and hydrolysis to regulate substrate binding and release, thus facilitating protein folding. In contrast, most known ATP-independent chaperones function as holding chaperones. They prevent protein aggregation under stress conditions, and transfer their substrates to ATP-dependent folding chaperones upon return to non-stress conditions. Although Spy is localized to the ATP-devoid environment of the bacterial periplasm, it appeared to support the accumulation and thus folding of Im7 variants under non-stress conditions. These results raised the intriguing possibility that Spy might function as an ATP-independent folding chaperone. Alternatively, Spy might function as a very effective holding chaperone and transfer substrates to other ATP-independent folding chaperone for folding.

To investigate the possibility that Spy functions as a folding chaperone, we tested whether Spy aids in folding of denatured protein substrates *in vitro*. Our results suggested that Spy not only prevented protein aggregation but also significantly increased the refolding yield of both thermally and chemically denatured MDH and chemically denatured lactate dehydrogenase (LDH) (Figure 3.12 f-h). These assays clearly showed that Spy indeed functions as a highly efficient folding chaperone.

Since the assays were performed in the absence of any cofactors, these results strongly suggested that Spy has intrinsic protein folding capacity, providing an excellent explanation how Spy overexpression is sufficient to significantly increases the amount of folded Im7 protein *in vivo*.

3.2.5 Conditions induce *spy*

Transcription of *spy* is regulated by the Cpx and the Bae stress response pathways[57]. The known induction signals of *spy* include spheroplasting[171], overexpression of misfolded periplasmic or outermembrane proteins, such as PapG and NlpE[164], indole[164], condensed tannins[172], ethanol[57], metals[173], low pH, deletion of *dsbA*[174], as well as overexpression of transcription activators. Conditions inducing *spy* are summarized in Table 3.5.

Table 3.5 Conditions that induce *spy*

Conditions that induce *spy*	Fold increase in *spy* message levels	Method
spheroplasting[171]	>15	*spy* promoter-lacZ
indole[164]	14 (1.5 ΔbaeSR)	qRT-PCR
ethanol[57,]*	~4	microarray
butanol*	~6	microarray
condensed tannins (1%)[172]	494/18	qRT-PCR/microarray
model apple juice (pH 3.7)[175]	7	Microarray
HCl (pH 5.5)*	~3	Microarray
ΔdsbA[174]	5	Microarray
overexpression of misfolded PapG[164]	>8	*spy* promoter-lacZ
overexpression of NlpE[164]	>9	*spy* promoter-lacZ
mecillinam & cefsulodin*	>3	microarray
$ZnCl_2$[173]	~3	*spy* promoter-lacZ
copper-glycine[173]	>5	microarray
BaeR overexpression (from pUCbaeR)[176]	640/140 >20	qRT-PCR/microarray *spy* promoter-lacZ
cpxA24 (mutation constitutively activates CpxA)[170]	>10	*spy* promoter-lacZ
ΔbaeR[176]	2.2 >20	qRT-PCR *spy* promoter-lacZ
ΔphoBR[176]	650	qRT-PCR
ΔcreBC[176]	240	qRT-PCR

Spy was induced by various conditions that activate the Cpx or Bae response. Indole treatment causes membrane derangement, resulting in the reduction of oxygen by semiquinones and the peroxidation of lipids[177]. lacZ: gene encoding β-galactosidase, a reporter to monitor transcription from the *spy* promoter. *: Unless specified, the data for microarray experiments were obtained from http://www.prfect.org/index.php?option=com_wrapper&Itemid=419

The induction of Spy by these conditions led us to hypothesize that deletion of spy might result in growth defects in bacteria under these conditions. We created a *Δspy* mutation in four different strain backgrounds: MG1655, MC4100, BW25113, and BL21. We then grew these strains under normal and spy-inducing conditions and compared their growth. In addition, we tested whether *spy* null strains exhibit deficient growth when the culture medium (or agar plate) contained chemicals known either as environmental pollutants[178], or as those which induce heat shock responses[179, 180]. These chemicals

include ether, benzene, chlorpyrivos, 2,4-dichloroaniline, dioctylphthalate, hexachlorobenzene, pentachlorophenol, trichloreoethylene, isopropanol, propanol, methanol, sodium proprionate, formamide, 4-nitrophenol, phenol, sodium acetate, $CdCl_2$, NaCl, H_2O_2, hypochloride, naladixic acid, menadione, 2,4-dinitrophenol, dibucaine, paraquat, and bile salts. Unexpectedly we did not observe any obvious growth defects of *spy* null strains resulting from contact with these chemicals including the ones induce *spy*. The only exception was tannic acid, a polyphenol, which I discuss in great details in the following section. We further tested various antibiotics and detergents, and chemicals known to inhibit the growth of strains which are devoid of functional Cpx or Bae pathways[181]. They include puromycin, ampicillin, mecillinam, novobiocin, mitomycin C, cefsuolodin, nafcillin, cloxacillin, kanamycin, chloramphenicol, spectinomycin, geneticin, EDTA, Triton, SDS, gallic acid, nickel chloride and sodium tungstate. Here also, however, we did not observe any phenotypes. In summary, the only phenotype observed caused by the *spy* deletion is the sensitivity to tannic acid.

3.2.6 Spy mediates resistance to tannic acid and protects activities of proteins from tannic acid inactivation

Tannic acid is a specific commercial form of tannin. Tannins are synthesized by plants to protect against bacterial and fungal infections, and only small quantities are required to aggregate proteins, a feature that may be responsible for their antimicrobial activity[182]. Tannins have long been used in tanning leather and are present in red wine, tea, and unripe fruit. It is thought that the astringent taste of food substances that contain tannins is due to the crosslinking of mouth proteins[183]. The astringent taste of strong tea is often reduced by addition of milk, which drives coprecipitation of the tannins with the disordered protein casein that is present in milk[183]. RT-PCR-based measurements indicate that *spy* mRNA is induced nearly 500-fold in response to tannin treatment[172]. The gene for *ibpB*, an *E. coli* small heat shock protein homologue also involved in inhibiting protein aggregation, is induced 48-fold by tannins, making it second only to *spy* in induction by tannins[172].

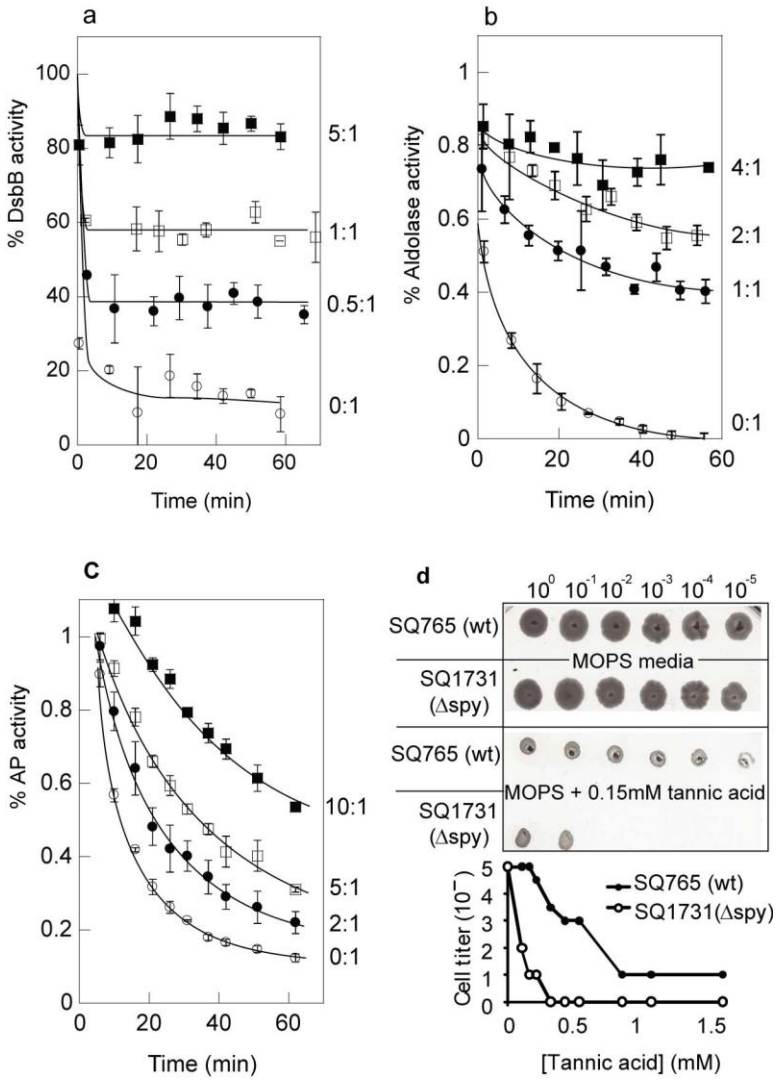


Figure 3.13 Spy protects DsbB, aldolase, and alkaline phosphatase from tannic acid-induced activity loss[**]

Spy protects DsbB, aldolase, and alkaline phosphatase (AP) from tannin inactivation. Spy concentrations are given as a dimer. Error bars indicate the standard deviation from 3 independent measurements. **a**, Enzymatic activity of *E. coli* DsbB (0.5 μM) over time incubated in 100 μM tannic acid in the absence (0:1) or presence of increasing amounts of Spy (ratios given are Spy:substrate). **b**, Enzymatic activity of rabbit muscle aldolase (0.5 μM) over time incubated in 16 μM tannic acid in the absence or presence of increasing amounts of Spy. **c**, The enzymatic activity of *E. coli* alkaline phosphatase (1 μM) over time incubated in 500 μM tannic acid in the absence or presence of increasing amounts of Spy. **d**, Spy deletion strains are tannin sensitive.

[**] I performed the tannin sensitivity experiments as presented in panel d. Stephan Hofmann assisted with the spot titer. The activity measurements of enzymes were performed by Tim Tapley and Philipp Koldewey. Tim Tapley prepared panel a-c of this figure.

We tested the effects of Spy on tannin-mediated inactivation of the *E. coli* membrane protein DsbB (Figure 3.13 a), aldolase (Figure 3.13 b), and the *E. coli* periplasmic protein alkaline phosphatase (Figure 3.13 c). Spy protected all three proteins from tannic acid-induced activity loss. Thus, Spy induction may be involved in protecting cells from tannin-induced protein aggregation and inactivation. We found that *spy* null mutants are sensitive to tannins (Figure 3.13 d), consistent with the previously reported tannin sensitivity of *baeSR* null mutants[172].

3.2.7 Spy may not have a regulatory role like its homolog CpxP

The Spy protein is 29% identical to CpxP, an inhibitory component of the CpxRA regulatory system. Both Spy and CpxP appear to be mostly α-helical in structure (Figure 3.14). CpxP binds to the periplasmic domain of CpxA, inhibiting its autokinase activity. Presence of unfolded proteins causes the release of CpxP, thereby activating the Cpx response[184, 185]. These results suggested that CpxP might be acting as one of the very few known periplasmic chaperones, targeting itself and its unfolded protein cargo to the protease DegP for degradation[186, 187]. To assess whether Spy might be similarly involved in regulating BaeSR or the other periplasmic stress response systems, we carried out RT-PCR (Table 3.4 on page 74) in strains either lacking or overexpressing *spy*. We found no notable influence on the expression of *baeS*, *cpx*, *rpoE*, *rpoH*, *rcs* or *psp* regulated genes, suggesting that Spy, unlike CpxP, may not be playing a major regulatory role. Instead, our results strongly suggest that Spy is directly functional as a molecular chaperone, supporting the folding of soluble proteins in the *E. coli* periplasm.

3.2.8 Structure of Spy reveals an alpha helical cradle-shaped chaperone

To gain insights into the mechanism of Spy's chaperone action, our collaborators crystallized and determined the three-dimensional structure of Spy[††]. Spy contains 161 residues of which 23 constitute the signal peptide that directs Spy to the periplasm. Our residue numbers refer to the mature, periplasmic form of Spy, which contains 138 residues.

[††] The crystal structure of Spy was solved by Miroslaw Cygler, Karen M. Ruane, and Rong Shi from McGill University, Canada. Jim Bardwell, Miroslaw Cygler, and Zhaohui Xu analyzed the strucuture.

The crystal structure shows that Spy molecules associate into tightly bound dimers. Size exclusion chromatography, analytical ultracentrifugation (Figure 3.15) and dynamic light scattering analysis (data not shown) confirmed this oligomerization state and revealed the presence of Spy dimers also in solution. Each Spy monomer is highly α-helical, in agreement with the CD spectrum (Figure 3.14).

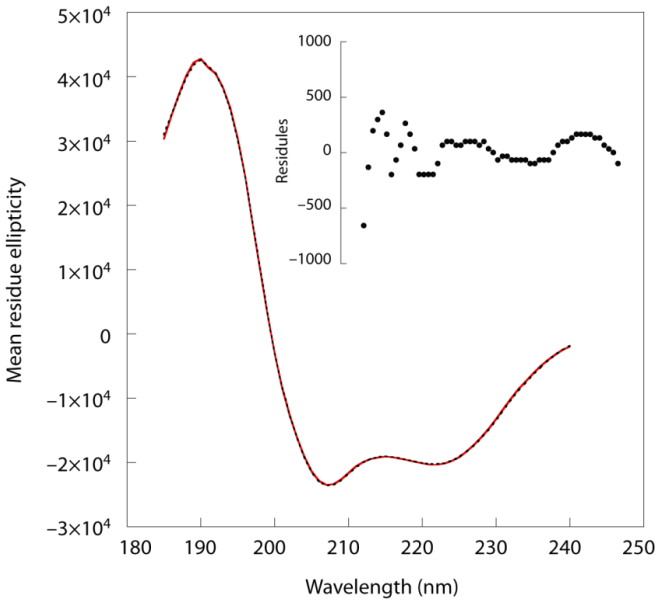


Figure 3.14 Circular dichroism indicates that Spy is a predominantly α-helical protein[‡‡]

Far UV circular dichroism (CD) spectra of Spy were collected and then analyzed using the CDSSTR algorithm on the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) to assess secondary structure content. The raw data are shown as a solid red line; the dashed black line (which almost precisely overlays the raw data trace) represents the reconstructed spectrum computed based on the following secondary structure content: 77% helix, 4% strand, 5% turn and 14% disordered. Residuals (the difference between the experimental spectrum and reconstructed spectrum) are shown in the inset. The related protein CpxP is also predicted by Jpred3 http://www.compbio.dundee.ac.uk/www-jpred/ to be mostly α-helical in secondary structure.

The Spy dimer is formed through the antiparallel coiled-coiled interaction of helices α3 with α3' (prime indicates the second monomer) and antiparallel interactions between α4 and α2' (Figure 3.16). The shape of the dimer is rather unusual, with one surface highly concave and the other convex, highly reminiscent of a cradle. The bottom of this cradle is formed by helices α3, the sides by the connection between helices α1 and

[‡‡] I performed the experiment. Tim Tapley analyzed the data.

α2 and the tips by helices α1 and α4 (Figure 3.16 a). The disordered N- and C-termini potentially extend over the concave surface of the cradle (Figure 3.16 a.b). The thickness of the dimer is in most places equivalent to just the width of a single α helix. The contacts between the two monomers are extensive, burying a surface of ~1850 $Å^2$ per monomer upon dimerization and suggest high stability of dimers. The most intimate contacts between the two monomers are at the ends of the cradle. The concave surface is also lined with many conserved apolar side chains including Leu34, Ile42, Met46, Pro56, Met64, Met85, Met97, Ile103, and Phe115, localized in two patches in the interior of the cradle giving it an overall apolar/hydrophobic character (Figure 3.16 b). The convex surface has on the other hand a hydrophilic character typical for water-soluble proteins.

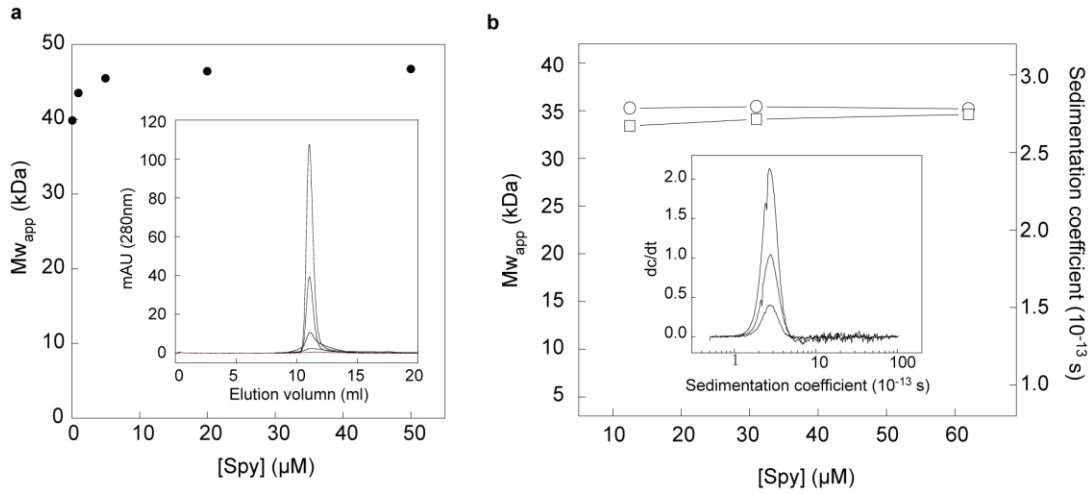


Figure 3.15 Size exclusion chromatography and analytical ultracentrifugation indicate that Spy is an elongated dimer[§§]

a, Size exclusion chromatography performed over a range of Spy concentrations (0.1–50 μM monomer concentration) yield an apparent molecular weight (MW_{app}) of ~45 kDa, which is 2.8 times the expected size for monomeric Spy. The apparent molecular weight decreased only very slightly at very low Spy concentrations (100 nM). The inset shows the raw chromatograms. **b**, Concentration-dependent sedimentation velocity analytical ultracentrifugation analysis of Spy at indicated concentrations. Molecular weight (□) and Sedimentation coefficient (○) were derived from 2-dimensinal spectrum analysis and plotted as the function of the tested Spy concentration. The molecular weight determined was ~30 kDa, suggesting that Spy forms dimers (the calculated mass of the Spy monomer is 15.9 kDa). Inset: dc/dt boundary analysis of Spy at 62, 31 and 12.5 μM (concentration given as a monomer). Samples were spun at 42,000 rpm in 10 mM HEPES, 100 mM NaCl, pH 7.5.

[§§] Tim Tapley and Titus Fransmann performed these experiments and prepared this figure.

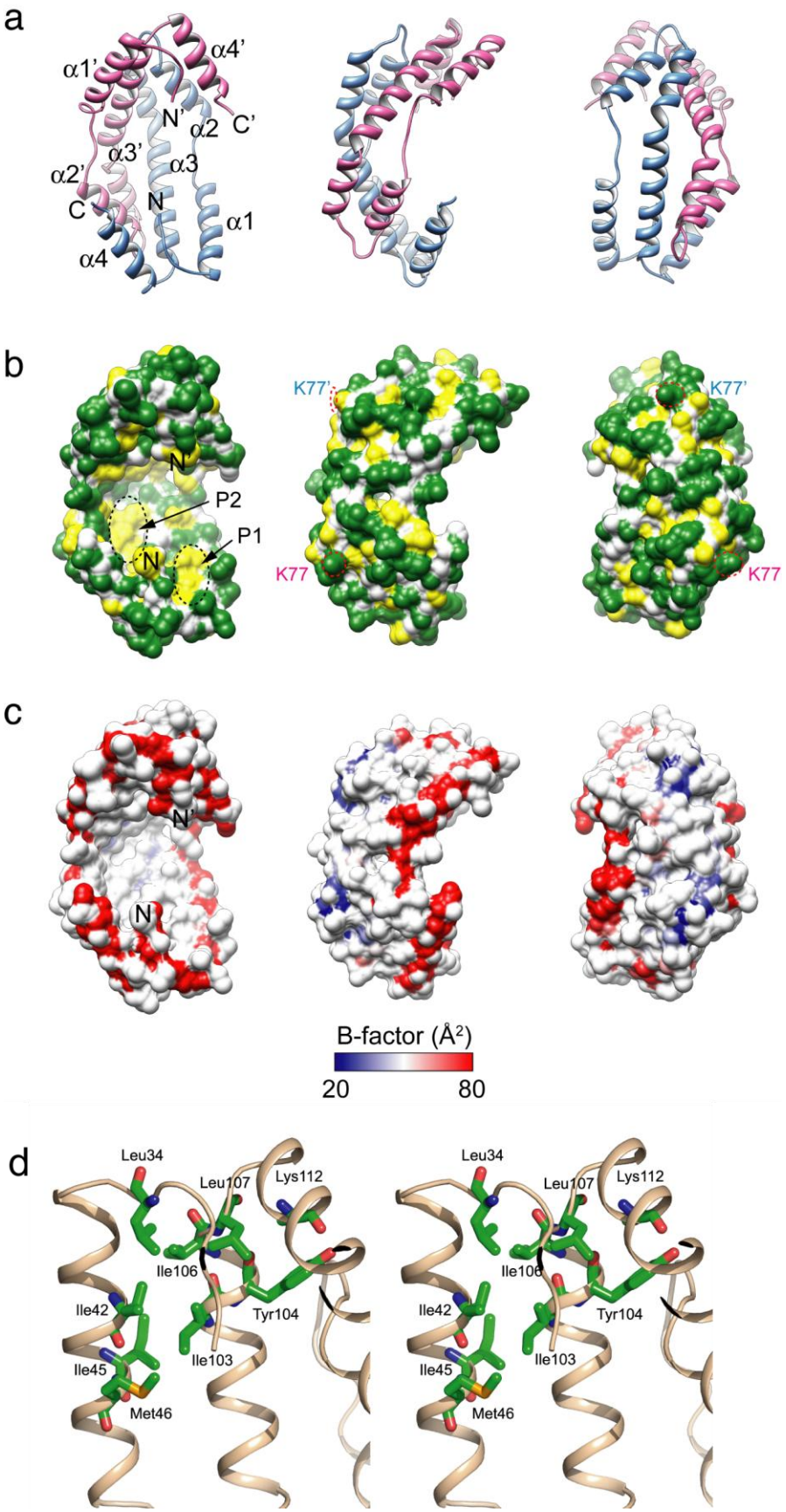

a
α1'
α4'
N'
α2
C'
α3'
α3
α2'
C
N
α1
α4
b
N'
P2
P1
N
K77'
K77
K77'
K77
c
N'
N
B-factor (Å²)
20
80
d
Leu34
Leu107
Lys112
Ile106
Ile42
Tyr104
Ile45
Ile103
Met46
Leu34
Leu107
Lys112
Ile106
Ile42
Tyr104
Ile45
Ile103
Met46

Figure 3.16 The crystal structure of the Spy dimer shown in three orientations; rotated by 90° along the vertical axis[***].

a. Ribbon drawing shows an all-alpha helical structure. One subunit is colored light blue and the other is colored magenta. The N- and C-termini as well as the secondary structural elements of the molecule are labeled. **b.** The surface properties of Spy. The molecular surface of Spy is colored based on the underlying atoms: all backbone atoms, white; all polar and charged side-chain atoms, green; all hydrophobic side-chain atoms, yellow. Note the concave surface of the molecule has two significant yellow hydrophobic patches. The lower right patch (P1) is composed of Leu34, Ile42, Met46 and Ile103. The upper left patch (P2) is composed of Pro56, Met64, Ile68, Met85, Met93 and Met97. Lysine 77, which was labeled with bimane for some of the experiments shown in Figure 3.18 on page 88 is circled in red. **c**. Structural flexibility of Spy dimer. The molecular surface representing the Spy backbone atoms is colored based on the average backbone B-factors for each residue. Note the rim lining the concave surface has higher B-factors indicating greater structural flexibility. In particular, the N- and C-termini are highly mobile. **d**. The cluster of hydrophobic residues at the tip of the cradle. These residues are well conserved among homologous sequences (Figure 3.17).

CLUSTAL W (1.83) multiple sequence alignment[†††]

```
gi|194438564      ----ADTTTAAPADAKP---------------------MMHHKGKFGP----HQDMMFK 30
gi|238894281      ----AETTT-APSDSKPM--------------------MMHHKGGPG-----QHDMMFK 29
gi|258635767      ---AADNLTPPPAGSEKP--------------------MHKPPMRHG-----GMDMMFK 31
gi|253686940      -----NEASKGPGPEHKM--------------------MMKEGREHRGM--MGPDAMFK 32
gi|259908592      ---AADNLTPPPAGAEKP--------------------GMQKPPRHG-----GPHEMFK 31
gi|251791069      ------ADTSMPDQGARM--------------------MKHHDGGRGERGMMEENMMFK 33
gi270264557       ----ADTTAAPATAADAA--------------------PMKMMHHKGE----GKGGPFA 31
gi|85059846       -----ADTATAPAASESA--------------------KMYHHPKHGP----MMEQMFK 30
gi|49176443       ----AEVGSGDNWHPG---------------------------EELTQ----RSTQSHMFD 26
gi|152972721      ----AEVVTSVNWLPG---------------------------DEGGQ----RGSQSHMFD 26
gi|270265193      ----ADTTSETAPTPG---------------------------NDAITR-VPG--QHHMFD 27
gi|238764463      --------ADKAGATD---------------------GWCHGDGAMMNKKDGRGHHNMFD 31
gi|258637607      ----AETTTIDEMHQN---------------------------GGLTTGSMTQNPQSHMFD 30
gi|271502506      IEHHSSVDSDASAESNT-----------------PQVGNCYRDDSATKREES--QQGMFD 41
gi|85060157       -------ADNNTLES---------------------------WQQDAITHRIQES-HYSMFD 27
gi|238921654      -------DEGEPVCS---------------------------WYQGNDMMQQSMD--HNYMFD 27
gi|300714680      -----AVTTTDEMHQD---------------------------NAATR-SMTQIPQSHMFD 28
gi|253686545      ------AESGDAPASGW----------------------HIDDSATKGVSG--QQGMFD 29
gi|253991780      ----ETTNTTHAATTPGICAS---CGY-------KSSEGHHYHNN-EAGDN----DHLFN 41
gi|227354740      ----ETSNTNNQPAGQQQYNDNYHCGYGYGMHDGRDNRGHRGQRGHMMGNNNFGESRMFN 56
gi|284008856      ----ETADVNLAATQSAKQIQAVKHDSPISK---RNITRDRLCNPYDIFVP------LIK 47
gi|139522416      ----DTADTDDTPEAANPSPYCLSYEH-------KRDSGYYRSDEHNYNYS-----YVFG 44
gi|290477254      ----GTAGTDN-VIADDFSQYCMPYGN-------KNNVSYCLNDERNYNYS-----CIFA 43
gi|188533989      ---AADNLTPPPSGAEKT--------------------MMPERPHHG-----GPHEMFK 31
gi|259906788      -------ATTGEMHQD---------------------------DGTNR-TLRQVPQSNMFD 26
gi|37528606       ----GTASTMHAATVPGNCAP---CEY-------KNVGSYHYRNNNDIGDN----DHMFN 42
                                                                            :
```

[***] Jim Bardwell, Miroslaw Cygler, and Zhaohui Xu prepared this figure.

[†††] Jim Bardwell performed and analyzed the Spy sequence alignment and prepared this figure.

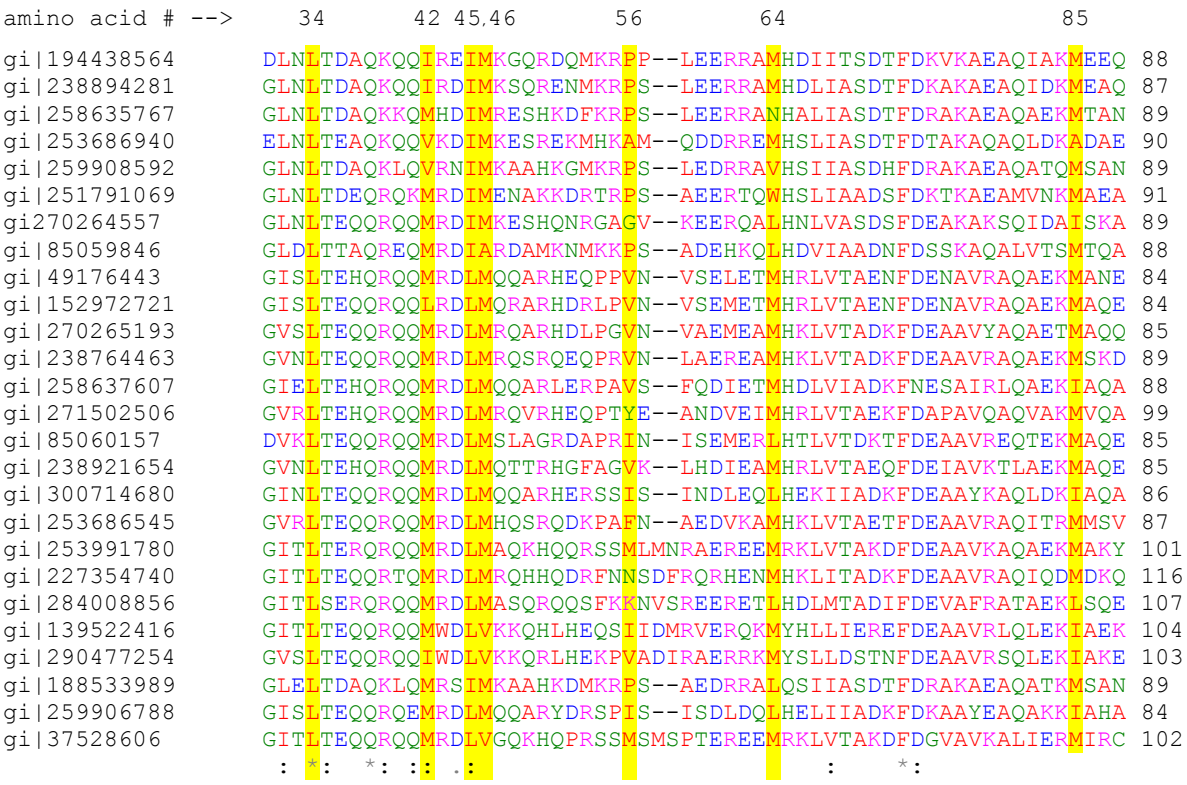

amino acid # --> 34 42 45,46 56 64 85
gi|194438564 DLNLTDAQKQQIREIMKGQRDQMKRPP--LEERRAMHDIITSDTFDKVKAEAQIAKMEEQ 88
gi|238894281 GLNLTDAQKQQIRDIMKSQRENMKRPS--LEERRAMHDLIASDTFDKAKAEAQIDKMEAQ 87
gi|258635767 GLNLTDAQKKQMHDIMRESHKDFKRPS--LEERRANHALIASDTFDRAKAEAQAEKMTAN 89
gi|253686940 ELNLTEAQKQQVKDIMKESREKMHKAM--QDDRREMHSLIASDTFDTAKAQAQLDKADAE 90
gi|259908592 GLNLTDAQKLQVRNIMKAAHKGMKRPS--LEDRRAVHSIIASDHFDRAKAEAQATQMSAN 89
gi|251791069 GLNLTDEQRQKMRDIMENAKKDRTRPS--AEERTQWHSLIAADSFDKTKAEAMVNKMAEA 91
gi270264557 GLNLTEQQRQQMRDIMKESHQNRGAGV--KEERQALHNLVASDSFDEAKAKSQIDAISKA 89
gi|85059846 GLDLTTAQREQMRDIARDAMKNMKKPS--ADEHKQLHDVIAADNFDSSKAQALVTSMTQA 88
gi|49176443 GISLTEHQRQQMRDLMQQARHEQPPVN--VSELETMHRLVTAENFDENAVRAQAEKMANE 84
gi|152972721 GISLTEQQRQQLRDLMQRARHDRLPVN--VSEMETMHRLVTAENFDENAVRAQAEKMAQE 84
gi|270265193 GVSLTEQQRQQMRDLMRQARHDLPGVN--VAEMEAMHKLVTADKFDEAAVYAQAETMAQQ 85
gi|238764463 GVNLTEQQRQQMRDLMRQSRQEQPRVN--LAEREAMHKLVTADKFDEAAVRAQAEKMSKD 89
gi|258637607 GIELTEHQRQQMRDLMQQARLERPAVS--FQDIETMHDLVIADKFNESAIRLQAEKIAQA 88
gi|271502506 GVRLTEHQRQQMRDLMRQVRHEQPTYE--ANDVEIMHRLVTAEKFDAPAVQAQVAKMVQA 99
gi|85060157 DVKLTEQQRQQMRDLMSLAGRDAPRIN--ISEMERLHTLVTDKTFDEAAVREQTEKMAQE 85
gi|238921654 GVNLTEHQRQQMRDLMQTTRHGFAGVK--LHDIEAMHRLVTAEQFDEIAVKTLAEKMAQE 85
gi|300714680 GINLTEQQRQQMRDLMQQARHERSSIS--INDLEQLHEKIIADKFDEAAYKAQLDKIAQA 86
gi|253686545 GVRLTEQQRQQMRDLMHQSRQDKPAFN--AEDVKAMHKLVTAETFDEAAVRAQITRMMSV 87
gi|253991780 GITLTERQRQQMRDLMAQKHQQRSSMLMNRAEREEMRKLVTAKDFDEAAVKAQAEKMAKY 101
gi|227354740 GITLTEQQRTQMRDLMRQHHQDRFNNSDFRQRHENMHKLITADKFDEAAVRAQIQDMDKQ 116
gi|284008856 GITLSERQRQQMRDLMASQRQQSFKKNVSREERETLHDLMTADIFDEVAFRATAEKLSQE 107
gi|139522416 GITLTEQQRQQMWDLVKKQHLHEQSIIDMRVERQKMYHLLIEREFDEAAVRLQLEKIAEK 104
gi|290477254 GVSLTEQQRQQIWDLVKKQRLHEKPVADIRAERRKMYSLLDSTNFDEAAVRSQLEKIAKE 103
gi|188533989 GLELTDAQKLQMRSIMKAAHKDMKRPS--AEDRRALQSIIASDTFDRAKAEAQATKMSAN 89
gi|259906788 GISLTEQQRQEMRDLMQQARYDRSPIS--ISDLDQLHELIIADKFDKAAYEAQAKKIAHA 84
gi|37528606 GITLTEQQRQQMRDLVGQKHQPRSSMSMSPTEREEMRKLVTAKDFDGVAVKALIERMIRC 102
amino acid # --> 97 103-4-6-7 112 115
gi|194438564 RKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKGKMPAT-AE--------- 138
gi|238894281 HKAMALSRLETQNKIYNILTPEQKKQFNANFEKHLTERNAPAGKMPAP-AE--------- 137
gi|258635767 AKERAVAMMETQNKLYNVLTPEQKKQFNANFEKRLTEKRPHDGKMPPPPAEGE------- 142
gi|253686940 HKARMLNGLETKNKIYNVLTPEQKKQYNDNFEKRLTQPPRPDGKPVP--AE--------- 139
gi|259908592 SQQRTMLMLETQNKMYNVLTPEQKKQFNQNFEKRLAEKPQHEDRMPP--ADD-------- 139
gi|251791069 NKAHMLKRLELHNQMYNVLTPEQKKQFNDNFAKRQQEPVAP--KAP-------------- 135
gi270264557 QSEHMLERAKAENKMYNLLTPEQKKQYNENYQKREQKMMDHMNKMKSKMTTEQ------- 142
gi|85059846 QNERMLAHLEMQNKMYNVLTPTQKQEFNKKFEEHQTKIMEHGNN---------------- 132
gi|49176443 QIARQVEMAKVRNQMYRLLTPEQQAVLNEKHQQRMEQLRDVTQWQKSSSLKLLSSSNSRS 144
gi|152972721 QVARQVEMAKVRNQMYHLLTPEQQAVLNAKHQQRMDQLREVARMQKGSAMMLSSSSNTLQ 144
gi|270265193 QVKRQVEMARVRNQMYNLLTPQQKSVLDQKHQQRMQQMEQQISGLQQTSAQKLSMTE--- 142
gi|238764463 QIDRQVEMAKVRNQMFNLLTPEQKAVLNQKHQQRIEQMQQAPAAQ-PSSAQK-------- 140
gi|258637607 QVEQNIVMARVRNQMYHLLTPAQQATVQKNYERRLDEMRRLSELQPSSPLQAVSSTSSNQ 148
gi|271502506 QIARQVEIARVSNQMYNMLTPEQKAVLNQKHEQAMQAVRQQMPAS-DSSSQNLQLSRQGQ 158
gi|85060157 QVARQVEMARVRNQMYNLLTPEQQQILEQKHLQRVSDLKLQMADKISTSAQKPAVNE--- 142
gi|238921654 SVIRQVEMARIRNMVFNLLTSDQKAQLNARYQQKIAGWQQQIAYMQNSSAQK-------- 137
gi|300714680 EVSRQVEMARIRNQMYHLLTPAQQDVLNEKHQQRMSEMRKLTNMQQASSLQAVSSTGSNQ 146
gi|253686545 QIERQIQMTRVRNQMYNLLTPAQKEILDVKHKQRMKEMQQQISMFNQMAAPAPGTTNQTE 147
gi|253991780 DIERQVEMARIHHQMYQLLTPEQQVQLEQQHRQYMSQLSN---------------------- 141
gi|227354740 AIERRVEMAKVHNQMYQLLTPEQKAQLEKNYQQRVSSVSQQNQQTQEKAN---------- 166
gi|284008856 IVEQQIEMARIYHQFYKLLTHDQKMILEKQHQKQLSRSKY---------------------- 147
gi|139522416 NIDLGVEIARIRNQMYQLLTPEQKERLYKRYEGQTAQEMH---------------------- 144
gi|290477254 NVDLSVEIARIRNQIYQLLTTEQKVQLQRRYEVRVNREMN---------------------- 143
gi|188533989 GQQRALSMMETQNKLYNVLTAEQKKQFNQNFEKRLAEKPHHQDGMPP--ADD-------- 139
gi|259906788 EVARQVEMGRVRNQMYHLLTPQQQSILQQKHQQRLGELRRLTNMQLSSPLQAASSTDSTP 144
gi|37528606 NVEHQVEMARIHHQMYQLLTPEQQAQLEQHYRQRMSQLPN---------------------- 142

Figure 3.17 Alignment of sequences homologous to Spy.
The NCBI nr database was searched using blastp using the *E. coli* MG1655 Spy protein sequence gi|194438564. All sequences that showed a E value of ≤0.001 were selected (225 sequences in total) and a distance tree of these sequences with treeview default NCBI settings was displayed. This tree showed that Spy homologues were present in a wide variety of enterobacteria, proteobacteria, and some cyanobacteria. Then one sequence (leaf) from each branch of the tree was selected to provide a good representation of the diversity present. The selected sequences were aligned with ClustalW (www.ebi.ac.uk/clustalw) using the default parameters. The proteins are identified by their gi numbers, first sequence shown gi|194438564 is *E. coli* Spy. For this sequence the signal sequence was omitted, so that the numbering is for the mature portion of the Spy protein. The residues highlighted in yellow and numbered at the top of the alignment are those mentioned in the main text. The colors and consensus symbols shown are the ClustalW at EBI default values namely: (the following information is taken directly from the website www.ebi.ac.uk/clustalw).

AVFPMILW	RED	Small (small+ hydrophobic)
DE	BLUE	Acidic
RK	MAGENTA	Basic
STYHCNGQ	GREEN	Hydroxyl + Amine + Basic - Q

"Consensus symbols: an alignment will display by default the following symbols denoting the degree of conservation observed in each column: "*" means that the residues or nucleotides in that column are identical in all sequences in the alignment. ":" means that conserved substitutions have been observed, according to the color table above. "." means that semi-conserved substitutions are observed."

3.2.9 Substrate binding to the concave side of the cradle‡‡‡

Based on the crystal structure and the fact that many previously identified substrate binding sites of chaperones are hydrophobic in nature, we hypothesized that the concave side of the cradle may serve as substrate-binding site of Spy. In addition, based on our recent findings that the acid-activated periplasmic chaperone HdeA functions as a largely disordered protein when it supports the refolding of acid-denatured proteins[188], we considered whether the disordered regions of Spy had a similar role.

In order to test these hypotheses, two cysteine variants of Spy were generated by Philipp Koldewey; one with a unique cysteine at position 24, which is in the disordered N-terminal region of Spy on the concave side of the cradle (Figure 3.16 b). Another cysteine was engineered at position 77, which is on the opposite (convex) side of Spy

‡‡‡ Philipp Koldewey performed all the experiments in this section.

(Figure 3.16 b). These unique cysteine residues were used to fluorescently label Spy with monobromobimane, a small thiol specific fluorophore. We decided to use casein as substrate protein since it is natively unfolded and soluble. Thus, neither aggregation nor refolding of the protein should interfere with our measurements. Fluorescence anisotropy was used to investigate whether the introduced mutations affected substrate-binding. As shown in Figure 3.18 a, both labeled Spy variants exhibited an increase in anisotropy upon casein binding, suggesting that they retain the ability to bind substrates. This result agreed with studies using analytical gel filtration, which revealed the presence of an apparently stable complex between Spy and casein (Figure 3.18 b). These results suggested that the two substituted amino acids are either not directly involved in substrate binding or that their substitutions did not strongly affect substrate binding.

We then measured fluorescence emission spectra to monitor potential environmental changes in the vicinity of the two cysteines, which would serve as direct read-out for substrate interaction at or near those specific sites. As shown in Figure 3.18 c-d, the fluorescence of bimane attached via Cys24 is partially quenched and slightly blue-shifted upon binding of casein. In contrast, the fluorescence of bimane attached via Cys77 is slightly increased by casein binding and no shift in the emission maximum was detected. These results suggested that substrate binding occurs on the concave side of the cradle, possibly within the cradle itself. The small change of bimane fluoresence observed at Cys77 could be due to an overall conformational change in the small Spy protein that occurs upon substrate binding.

In order to confirm that casein was a suitable substrate protein, i.e., that it is bound in the bona fide substrate binding site in our fluorescence experiments, a competition assay between MDH and casein was carried out. Because we know that Spy enhances the refolding of chemically denatured MDH, we reasoned that addition of casein should decrease the MDH refolding rate if indeed casein and MDH bind to the same apolar surface of Spy. As shown in Figure 3.18 e, an equimolar addition of casein significantly decreases the refolding rate of MDH. We interpret this as being due to a direct competition for the same site on Spy, namely the conserved apolar cradle we had originally hypothesized as the substrate binding site.

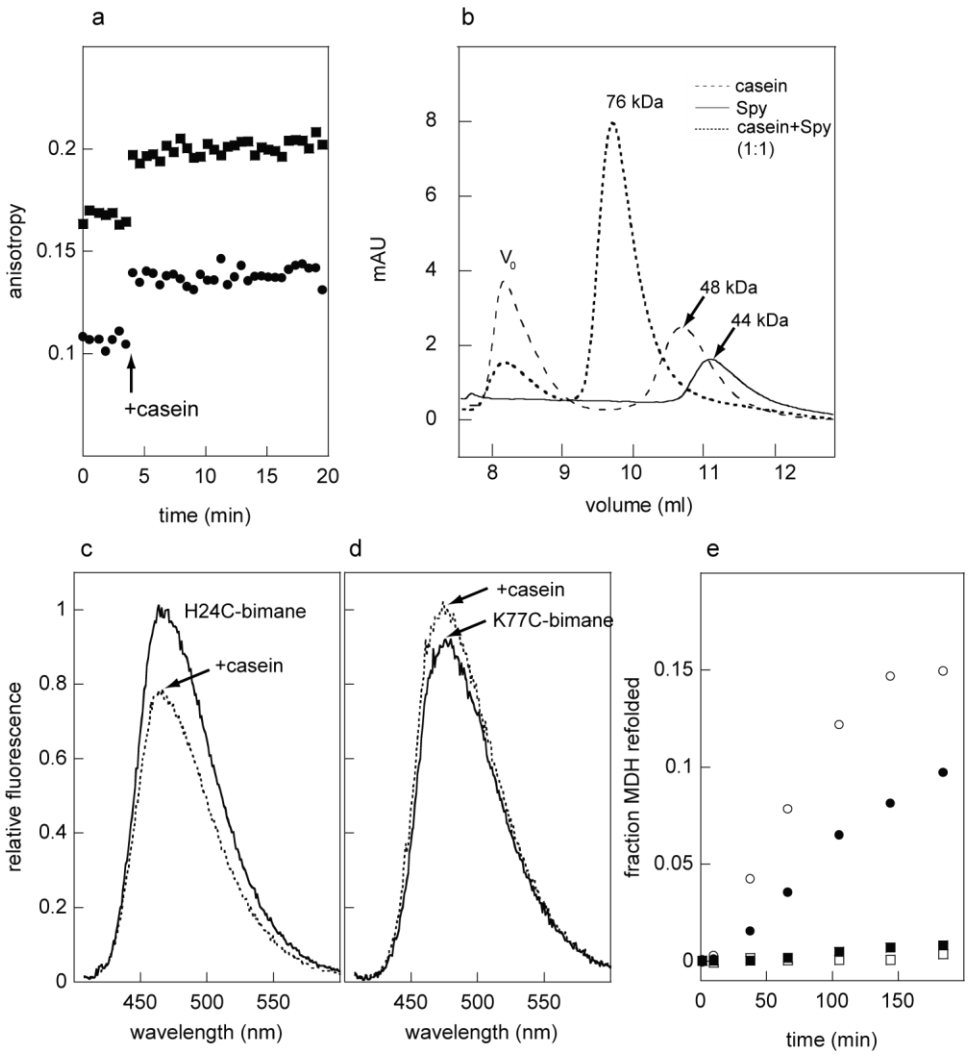


Figure 3.18 Spy binds the disordered model substrate protein casein on the concave side of the cradle[§§§].

a, Fluorescence anisotropy of bimane-labeled Spy H24C (squares) and Spy K77C (circles) before and after addition of an equimolar amount of casein. **b,** Analytical gel filtration of Spy alone, casein alone, or a 1:1 mixture of Spy and casein. The MW of a dimer of Spy is 31kD, but it elutes as a 44 kDa molecule, which is consistent with the elongated form of the dimer as seen in the crystal structure. The actual MW of casein is 23-26 kDa. **c,** Normalized fluorescence emission spectra of bimane-labeled Spy H24C in the absence (solid line) or presence (dashed line) of an equimolar amount of casein. **d**, Normalized fluorescence emission spectra of bimane-labeled Spy K77C in the absence (solid line) or presence (dashed line) of an equimolar amount of casein. **e,** Competition assay between casein and MDH for Spy binding. Refolding of chemically denatured MDH alone (open squares), MDH+casein (1:1; closed squares), MDH+Spy (1:1; open circles) or MDH+casein+Spy (1:1:1; closed circles) was monitored by recovery of enzymatic activity. Recovered activity is plotted as a fraction of the activity of native (i.e. non-denatured) MDH.

[§§§] Philipp Koldewey and Tim Tapley prepared this figure.

3.3 Discussion

It may seem surprising that a chaperone as effective as Spy has remained unstudied in an organism as well characterized as *E. coli*. However, because chaperones are usually effective only in stoichiometric quantities, chaperone assays are generally not sensitive enough to enable their purification from crude lysates by activity. Instead, chaperones have often been first identified because their genes are induced in protein unfolding conditions. Our approach of linking folding to selectable markers opens up the possibility of directing evolution to alter the *in vivo* folding environment to specifically enhance the folding of a given unstable protein and provides a new route for chaperone discovery. In our experiments, the vast majority of mutations that resulted in improved folding of the unstable Im7 protein induced the Spy protein, suggesting that the overproduction of this chaperone is the most genetically accessible way of enhancing the levels of unstable Im7 protein. Although Spy chaperones the folding of a wide variety of substrate proteins *in vitro*, Spy overproduction *in vivo* is clearly not a panacea that solves all *in vivo* protein folding problems. The *baeS-R416C* mutation, for instance, did not increase PenV resistance of tripartite fusions between β-lactamase and RNaseI or the highly aggregation-sensitive mutant of maltose binding protein (MBP-G32DI33P). Instead, our approach offers a method to force bacteria to tailor their *in vivo* folding environment to stabilize specific test proteins. Applying our selection to other poorly folded test proteins may prove useful in the customization of bacteria to stabilize specific poorly folded proteins, in facilitating chaperone discovery efforts, and in enhancing our understanding of the *in vivo* protein folding environment.

Spy has a unique cradle shape that is unlike any other chaperone with known structure. It lacks a significant globular core and would therefore be expected to display higher flexibility than the average globular protein. Its thinness places a disproportionate number of side chains on the protein surface. The Spy protein cradle is so thin that a number of residues including Leu34, Ile42, Met46 and Ile103 form both part of the structural "core" of the Spy structure and also part of a surface exposed patch. The highest backbone temperature factors (B-factors) are observed in surfaces and bumps extending from the concave side of the cradle, particularly in the connectors between

helices α1 and α2, which form the sides of the cradle and suggest potential for bending and twisting of the molecule (Figure 3.16 c). The shape combined with observed flexibility is suggestive of a mechanism for Spy action that involves the shielding of aggregation-sensitive regions on substrate proteins, which are revealed upon partial unfolding, by interaction with the relatively apolar surface present on the interior (concave) side of the cradle structure. This overall mechanism is consistent with the proposed mechanisms for a wide variety of chaperones that shield aggregation regions of proteins. A common structural feature in chaperones that enable them to specifically recognize folding intermediates are exposed hydrophobic surfaces. These surfaces serve to bind to hydrophobic segments existed in folding intermediates and release them to bulk solvent in a controlled fashion. The surfaces can be quite extensive, as seen in GroEL, or rather limited as seen in Hsp70[189]. While the near million-dalton GroEL with its ring structure can bind an entire polypeptide chain in an enclosed manner, the cradle-shaped small chaperone Spy will likely bind a segment of chain in a more open fashion. However, one could also imagine a scenario where two or more Spy dimers form a surface coat to enclose the substrate polypeptide. The flexible termini might provide additional, extended non-specific contacts with the substrate protein as seen with the close approach of His24 to the substrate. This histidine lies in a disordered region, 5 residues beyond the first residue resolved in the crystal structure, which is located at the interior corner of the cradle (Figure 3.16 b). This binding mode is consistant with the lack of fluorescence quench of derivatized Lys77 upon substrate binding. This residue lies on the outside (convex) side of the cradle.

The highly flexible nature of Spy may allow the chaperone to accommodate a variety of partially unfolded protein substrates. It has been proposed that structural disorder could play a fundamental role in the function of many chaperones[190]. The basis of their high flexibility, which is usually a high ratio of charged to hydrophobic residues, allows chaperones to keep their bound substrate proteins soluble and to prevent their aggregation. The extremely high flexibility of the periplasmic chaperone HdeA, for example, appears to allow this chaperone to bind numerous substrate proteins and prevent aggregation[188]. Spy may thus act as a cradle shaped wrapper that thinly coats proteins, protecting them from aggregation and possibly proteolysis as well. The levels of unstable

Im7 mutants go from barely detectable levels to nearly 10% of the periplasm upon Spy overproduction, likely because it is very effective in preventing *in vivo* proteolysis by promoting well folded proteins. This wrapper has two surfaces, a concave interior hydrophobic surface designed to interact with substrates, and a convex hydrophilic outer surface designed to maintain solubility of the chaperone-substrate complex. This mechanism is consistent with the high abundance of Spy under stress conditions and may explain how Spy is able to increase the levels of unstable proteins in the cell by up to nearly 700 fold.

Although suppression of protein aggregation does not require energy, release and refolding of bound substrate proteins usually does[40]. ATP is absent in the periplasm and in our assays was not required for the refolding function of Spy, indicating that the chaperone activity of Spy is ATP-independent. Thus, Spy appears to be one of the very few chaperones that actively support protein refolding in the absence of any obvious energy source, suggesting that Spy uses a different mechanism to control substrate binding and release from previously characterized chaperones. How Spy enables refolding in the absence of energy cofactors is a provocative question for future research. At this point we can only speculate that the apparently highly dynamic nature of the Spy structure may allow structural fluctuations that not only allow it to mold to various proteins but also enables it to release them.

3.4 Methods

3.4.1 Strains and plasmids used in this study

Table 3.6 Strains and plasmids

Strain	Genotype or relevant characteristics	Source
EMS1	SQ1306, *baeS R150W*	This study
EMS2	SQ1306, *baeS R150W*	This study
EMS3	SQ1306, *baeS P192L*	This study
EMS4	SQ1306, *baeS R416C*	This study
EMS5	SQ1306, *baeS R150W*	This study
EMS6	SQ1306, *baeS D268N*	This study
EMS7	SQ1306, $PenV^R$, $CdCl_2^R$	This study
EMS8	SQ1306, *baeS D268N*	This study
EMS9	SQ1306, *baeS E264K*	This study
EMS10	SQ1306, *baeS R150W*	This study
SQ765	MG1655 (*$F^-\lambda^-$ ilvG$^-$ rfb-50 rph-1*), *ΔhsdR*	This study
SQ1306	SQ765, *ΔampC, ΔdsbA*, pBR322bla::GSlinkerIm7L53AI54A pBAD33dsbA::GSlinkerIm7L53AI54A	This study
SQ1405	EMS4, pCDFTrc-ssIm7wt	This study
SQ1406	EMS4, pCDFTrc-ssIm7L53AI54A	This study
SQ1407	EMS4, pCDFTrc-ssIm7I22V	This study
SQ1408	EMS4, pCDFTrc-ssIm7F84A	This study
SQ1409	EMS9, pCDFTrc-ssIm7wt	This study
SQ1410	EMS9, pCDFTrc-ssIm7L53AI54A	This study
SQ1411	EMS9, pCDFTrc-ssIm7I22V	This study
SQ1412	EMS9, pCDFTrc-ssIm7F84A	This study
SQ1413	SQ1306, pCDFTrc-ssIm7wt	This study
SQ1414	SQ1306, pCDFTrc-ssIm7L53AI54A	This study
SQ1415	SQ1306, pCDFTrc-ssIm7I22V	This study
SQ1416	SQ1306, pCDFTrc-ssIm7F84A	This study
SQ1436	SQ1306, pCDFTrc	This study
SQ1594	SQ1306, *yegL::kan*	This study
SQ1640	SQ1306, *baeS R416C, yegL::kan*	This study

SQ1606	EMS4, *yegL::kan*	This study
SQ1608	EMS4, *baeS wt, yegL::kan*	This study
SQ1843	EMS4, *spy::kan*	This study
SQ1529	EMS4, *baeSR::kan*	This study
SQ1521	EMS4, *baeS::kan*	This study
SQ1838	EMS4, *mdtABCD::kan*	This study
SQ1848	EMS4, *acrD::kan*	This study
SQ1698	SQ765, *ΔyegL, baeS R416C*	This study
SQ1700	SQ765, *ΔyegL, baeS E264K*	This study
SQ1712	SQ765, *ΔyegL*	This study
SQ1761	SQ1698, pCDFTrc-ssIm7L53AI54A	This study
SQ1762	SQ1700, pCDFTrc-ssIm7L53AI54A	This study
SQ1763	SQ1712, pCDFTrc-ssIm7L53AI54A	This study
SQ1804	SQ765, pCDFTrc-ssIm7wt, pTrc-spy	This study
SQ1805	SQ765, pCDFTrc-ssIm7L53AI54A, pTrc-spy	This study
SQ1806	SQ765, pCDFTrc-ssIm7I22V, pTrc-spy	This study
SQ1807	SQ765, pCDFTrc-ssIm7F84A, pTrc-spy	This study
SQ1808	SQ765, pCDFTrc-ssIm7wt	This study
SQ1809	SQ765, pCDFTrc-ssIm7L53AI54A	This study
SQ1810	SQ765, pCDFTrc-ssIm7 I22V	This study
SQ1811	SQ765, pCDFTrc-ssIm7 F84A	This study
SQ1731	SQ765, *Δspy*	This study
SQ1777	SQ765, *ΔbaeSR*	This study
SQ1825	SQ1777, pCDFTrc-ssIm7wt, pTrc-spy	This study
SQ1826	SQ1777, pCDFTrc-ssIm7L53AI54A, pTrc-spy	This study
SQ1827	SQ1777, pCDFTrc-ssIm7I22V, pTrc-spy	This study
SQ1828	SQ1777, pCDFTrc-ssIm7F84A, pTrc-spy	This study
SQ1829	SQ1777, pCDFTrc-ssIm7wt, pTrc99a	This study
SQ1830	SQ1777, pCDFTrc-ssIm7L53AI54A, pTrc99a	This study
SQ1831	SQ1777, pCDFTrc-ssIm7I22V, pTrc99a	This study
SQ1832	SQ1777, pCDFTrc-ssIm7F84A, pTrc99a	This study
SQ1833	SQ1777, pCDFTrc, pTrc-spy	This study
SQ1796	SQ765, pTrc-spy	This study

All strains were derivatives of *E. coli* K-12 strains.

Plasmid	Relevant characteristics	Source
pBR322bla::GS linker Im7 L53AI54A	Im7 L53AI54A inserted into *bla* at aa197 via GS linker	Foit *et al.*[102]
pBAD33dsbA::GS linker Im7 L53AI54A	Im7 L53AI54A inserted into *dsbA* at aa99 via GS linker	This study
pCDFTrc	Expression vector with the CloDF13-derived CDF replicon[191], Trc promoter and kanamycine resistant marker	This study
pCDFTrc-ssIm7 wt	Im7 wt with signal sequence of *dsbA* cloned into pCDFTrc vector	This study
pCDFTrc-ssIm7 L53AI54A	Im7 L53AI54A with signal sequence of *dsbA* cloned into pCDFTrc vector	This study
pCDFTrc-ssIm7 I22V	Im7 I22V with signal sequence of *dsbA* cloned into pCDFTrc vector	This study
pCDFTrc-ssIm7 F84A	Im7 F84A with signal sequence of *dsbA* cloned into pCDFTrc vector	This study
pET28bsumo-spy	Spy without signal sequence cloned into pET28bsumo vector, for Spy purification	This study
pTrc99a	Expression vector	Pharmacia
pTrc-spy	Spy cloned into pTrc99a derived vector pssTrx[24]	This study

3.4.2 Cloning and strains construction

Plasmid construction. The wild-type *dsbA* gene including the ribosome binding site was amplified from the MG1655 genome with primer 1 (5'CGC GGG TAC CAG GAG GAA TTC ATG AAA AAG ATT TGG CTG GCG3') and primer 2 (5'CGC GTC TAG ATT ATT TTT TCT CGG ACA GAT ATT TCA C3'), and was cloned into pBAD33a via KpnI and XbaI sites. A 30-amino acid GS linker[102] was introduced into *dsbA* with a procedure similar to that described previously[102], using primer 3 (5'CGC GTC CGG GAG CGG TTC CGG AAG CGG AGG AGG TGG TTC AGG CGG AGG TGG AAG CCA GAC CAT TCG TTC TGC TTC TGA TAT CCG CGA TGT ATT TAT C3') and primer 4 (5'CGC GTC CCG GAT CCT GAG CTC GAG CCA CCA CCA CCA GAA CCA CCA CCA CCG GTT TTC TGT ACG CCT TCA AAC AGC GGA ACA GTC ACT TTG TC3'). Im7 L53AI54A, amplified using primer 5 (5'CGC GGG ATC CGA ACT GAA AAA TAG TAT TAG TG3') and primer 6 (5'CGC GGG ATC CGC

CCT GTT TAA ATC CTG GC3') with pBR322bla::GS linker Im7 L53AI54A[102] as template, was inserted into the linker region via the BamHI site.

Vector pCDFTrc for Im7 expression was constructed by ligation of the 0.9 kb NdeI- ApaI fragment from the pTrc99a derived vector pssTrx[99] and the 2.8 kb NdeI-ApaI fragment from pCDFDuet-1 (Novagen). The resulting plasmid was digested by AgeI and NheI, and the larger fragment was ligated through the same restrictions sites with a linear fragment containing the kanamycin resistant marker and its promoter. The NdeI site was then changed to the BamHI site by QuickChange mutagenesis (Stratagene) to create the pCDFTrc vector. C-terminal 6-His tagged Im7 (wild-type and 3 variants) was inserted into the pCDFTrc vector via the BamHI site using primer 7 (5′GCG CGG ATC CAT GGA ACT GAA AAA TAG TAT TAG3′) and primer 8 (5′GCG CGG ATC CTT AAT GGT GAT GGT GAT GAT GGC CCT GTT TAG CTC CTG GCT TAC C3′). A two-step polymerase chain reaction (PCR) was performed to add the signal sequence of *dsbA* in front of the Im7 gene. Step 1: primer 9 (5′GCG CGG ATC CAT GAA AAA GAT TTG GCT GGC3′) and primer 10 (5′GTG TAA TCA CTA ATA CTA TTT TTC AGT TCC GCC GAT GCG CTA AAC GC3′) were used to amplify the signal sequence of *dsbA*. In a different reaction, primer 11 (5′GCG TTT AGC GCA TCG GCG GAA CTG AAA AAT AGT ATT AGT GAT TAC AC3′) and primer 12 (5′GCG CGG ATC CTT AAT GGT GAT G3′) were used to amplify the Im7 gene. Primers 10 and 11 reverse complement each other and correspond to 15 base pairs (bp) of the *dsbA* signal sequence and the first 32 bp of the coding sequence of Im7. Step 2: full length amplification products that were generated from the above two reactions were used as templates to amplify the Im7 gene that now contained the *dsbA* signal sequence, using primers 9 and 12. The resulting fragment was ligated into pCDFTrc via the BamHI site to obtain pCDFTrc-ssIm7. Although the *dsbA* signal sequence was added, Im7 remains under the control of the trc promoter, which is different than those promoters that drive either bla::Im7 expression (bla) or DsbA:Im7 expression (pBAD). These Im7 expression plasmids are under the CDF copy control mechanism, which is independent from the copy control mechanism that governs either the *bla::Im7* fusion (the pMB1 ori) or the *dsbA::Im7* fusion (the p15A ori).

To obtain pTrc-spy, the vector pssTrx was digested with NdeI and BamHI and then ligated with the PCR-amplified *spy* gene that had been cleaved with the same restriction enzymes. *Spy* was amplified from the MG1655 genome with primer 13 (5′GCG CCA TAT GCG TAA ATT AAC TGC ACT G3′) and primer 14 (5′GCG CGG ATC CTT ATT CAG CAG TTG CAG GCA3′).

Gene disruption. The *spy, baeS, baeSR* and *mdtABCD* genes were disrupted through recombination between short homologous DNA regions according to the method of Wanner[192]. This method involves amplifying the kanamycin resistant gene and flanking by Flp recognition sites using primers that also contain homology to the gene to be disrupted. Primers for gene disruption were as follows. *spy*: 5′GAA AGC CGT AAT AAA TAA CTG AAA GGA AGG ATA TAG AAT ATT CCG GGG ATC CGT CGA CC3′ and 5′GTG GAC AAG ACC GGC GGT CTT AAG TTT TTT GGC TGA AAG ATG TAG GCT GGA GCT GCT TCG3′; *baeS*: 5′CAA AAT GTA GCT ATT TCG CGG CGA AAA AGG AGC GCG CAA TTC CGG GGA TCC GTC GAC C3′ and 5′CAA AAT ACG CGG TGT GTT TTC GTC GAT TGG TAA CTC GGT GTA GGC TGG AGC TGC TTC G3′; *baeSR*: same forward primer as for *baeS* disruption and reverse primer 5′GAA CGC CAT CTC CGG CTA ACA AAA TAA TGT CGC TAA AAT GTA GGC TGG AGC TGC TTC G3′; *mdtABCD*: 5′TTC CGC GAA ACG TTT CAG GAA GAG AAA CTC TTA ACG ATG ATT CCG GGG ATC CGT CGA CC3′ and 5′CAG AAC TTC ATT GCG CGC TCC TTT TTC GCC GCG AAA TAG CTG TAG GCT GGA GCT GCT TC3′. Following amplification and linear recombination into the chromosome, the kanamycin resistance gene was removed by transformation with a plasmid that encodes the flip recombinase. Since this procedure requires a step of introducing the cells with an Amp^R plasmid pKD46 (that encodes the λ red recombinase), the Amp^R pBR322bla::GS linker Im7 L53AI54A plasmid in SQ1306 and EMS4 was cured by a selection on fusaric acid agar plates as described previously[193]. After the target gene was disrupted in the resulting strain, pKD46 was cured by growth at high temperature as previously described and pBR322bla::GS linker Im7 L53AI54A was retransformed. The gene *acrD* was disrupted by P1 transduction with donor phage

prepared from the *acrD::kan* strain from the Keio collection[169]. Kanamycin (30 μg/ml) was used for the selection of gene disruption.

3.4.3 Mutagenesis and selection

Mid-log phase culture (A_{600nm}≈ 0.5, 40 ml) of SQ1306 was harvested and washed twice with 40 ml buffer A (10.5 g/L K_2HPO_4, 4.5 g/L KH_2PO_4, 1 g/L $(NH_4)_2SO_4$, and 0.5 g/L sodium citrate•$2H_2O$). Cells were pelleted and resuspended in buffer A to a normalized cell density of A_{600nm}= 1.0. Then, 2 ml aliquots were supplemented with 30 μl of the mutagen ethyl methanesulfonate (EMS) (Sigma, 99%) and incubated at 37 °C with 200 rpm shaking for 15 to 180 min. Cells of each aliquot were pelleted and washed twice with 5 ml buffer A. Cells were finally resuspended in 2 ml buffer A and the survival percentage following mutagenesis was determined by titering the cells on LB agar plates. The remaining cells were supplemented with 15% v/v glycerol and stored at –80 °C. Cells treated with EMS for less than 60 min showed a 100% survival rate, whereas survival decreased to 25% after treatment for 180 min. For selection of clones with increased PenV resistance, each pool of EMS treated cells was grown at 37 °C for 4 hours. Cell density was normalized to A_{600nm}=1.0, and a serial dilution was made in a sterile 170 mM NaCl solution. A100 μl aliquot of the 10^{-3} dilution (approximately 10^5 cells) was spread on LB agar plates supplemented with 1500 μg/ml PenV and 0.2% w/v arabinose. After 18 hours incubation at 37 °C, the penicillin resistant colonies were picked, inoculated, and screened for cadmium resistance via a spot titer method (described below). A total of 263 penicillin resistant colonies were obtained from 18 plates. No cells resistant to 1500 μg/ml PenV were observed in the absence of mutagenesis.

3.4.4 Phenotypic assays

Spot titers for PenV and cadmium resistance. Spot titers were measured to quantify the relative PenV and cadmium resistance conferred by the β-lactamase-Im7 or DsbA-Im7 tripartite fusions. The cells were grown, serially diluted in 10-fold increments and plated as described in the experimental procedures of Chapter 2 with the following modifications: 2 μl of each dilution was plated onto LB plates supplemented with 0.2%

arabinose and 1,500 to 6,000 μg/ml PenV or 0.075 to 0.6 mM $CdCl_2$. For each strain, the maximal cell dilutions allowing cell growth at each PenV or cadmium concentration were plotted against PenV or cadmium concentrations, in order to generate a titration curve. Ratios of minimal inhibitory concentration (MIC) values (MIC mutant/ MIC wt) were calculated from the titration curve as described in Ref[102]. Firstly, the maximal penicillin or cadmium concentrations that cells can survive to at each dilution were obtained from the curve. The ratios of MIC mutant/ MIC wt at each cell dilution were then calculated. The overall MIC mutant/ MIC wt value was an average of the ratios obtained from the 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} dilutions.

Quantitative RT-PCR. Total RNA was extracted from mid-log phase cultures using the RNeasy Mini Kit (Qiagen). For SQ1796, cells were induced to overexpress Spy by culturing with 1 mM IPTG for 3 hours at 37 °C. Genomic DNA contamination was removed with a TURBO DNA-free kit (Ambion). Extracted RNA (1 μg) was used for the reverse transcription polymerase chain reaction (RT-PCR) using SuperScript III reverse transcriptase kit (Invitrogen) with RNase inhibitor (Invitrogen) in a 30 μl reaction. Fifty nanograms of cDNA was used per 10 μl q-PCR reaction as described previously[57] with a Mastercycler EP Realplex real-time PCR system (Eppendorf). Primers were designed by D-lux™ Designer (Invitrogen: http://escience.invitrogen.com/lux/). Cycle threshold (Ct) values were calculated with the Realplex software (Eppendorf) and then analyzed with REST 2009 (Qiagen).

3.4.5 Protein purification and biochemical assays

Spy purification. The gene for *spy* minus its signal sequence was amplified from the MG1655 genome using primer 15 (5′CGC GGG ATC CGC AGA CAC CAC TAC CGC AG3′) and primer 16 (5′CGC GCT CGA GTT ATT CAG CAG TTG CAG GCA3′) and ligated into a pET28b-based vector[194] (generously provided by the Ming Lei lab at the University of Michigan) with sequences encoding a 6-His-SUMO tag fused at the N terminus. The expressed His-SUMO tag can be specifically removed by ULP1 protease, leaving a single serine residue at the N terminus of the protein. *E. coli* BL21(DE3) bearing the pET28bSUMO-spy plasmid was grown at 37 °C to early log phase in LB media containing 100 μg/ml of kanamycin and then shifted to 20 °C; Spy expression was

induced by addition of IPTG to 0.1 mM. After 16 hours of expression, cells were pelleted by centrifugation and resuspended in lysis buffer containing 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol, 3 mM 2-mercaptoethanol, 0.5 mg/ml lysozyme, 0.05 mg/ml DNase, 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1 tablet of protease inhibitor cocktail (complete mini EDTA-free, Roche) per 10 ml. Cells were lysed using a French press for 3 cycles at 1300 psi. After centrifugation for 1.5 hours at 14,000 rpm at 4 °C, the supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 2 hours at 4 °C to allow binding of the His-tagged Spy protein. The beads were washed with 40 bead volumes of lysis buffer, allowed to settle, and resuspended in lysis buffer containing 75 μg/ml ULP1 protease (provided by the Ming Lei lab at the University of Michigan). After overnight digestion at 4 °C, the mixture was applied to a mini-chromatography column (Novagen Cat. No.69673-3), and the cleaved Spy protein was collected in the flow-through. Spy was further purified by ion exchange chromotography. The cleaved Spy protein solution, which contained 400 mM NaCl, was diluted 6-fold with 20 mM HEPES, 0.5 mM EDTA, 5 mM 2-mercaptoethanol, at pH 6.5 before loading onto a 5 ml SP-FF cation exchange column (GE healthcare). Spy was then eluted with a linear gradient from 0-0.6 M NaCl in 20 mM HEPES, 0.5 mM EDTA, 5 mM 2-mercaptoethanol, pH 6.5. The resulting Spy protein was estimated to be >95% pure as analyzed by SDS-PAGE, and was stored in small aliquots at −80 °C. Prior to each experiment, Spy was then exchanged into the appropriate matching buffer (typically 40 mM HEPES, 100 mM NaCl, pH 7.5) with a PD-10 desalting column (GE Healthcare).

Quantification of proteins. The amount of the tripartite fusion protein β-lactamase-Im7L53AI54A in SQ1306 and EMS1-10 was quantified by Western blot with anti-β-lactamase antibody using whole cell lysate as described previously[102] with minor modifications. To quantify plasmid-encoded Im7 (wild-type and variants) expressed in the absence of the fusion, cells were grown to mid-log phase in LB medium at 37 °C. Im7 protein expression was induced with 2 mM IPTG for 2 hours. Addition of IPTG also induces the expression of Spy when the pTrc-spy plasmid is present. Periplasmic extractions were prepared as described previously[195], and the proteins were separated using a Bioanalyzer 2100 (Agilent) with the high sensitivity protein 250 kit and the conditions specified by the manufacturer. To visualize the tiny signal corresponding to

Im7 present in the absence of Spy expression, a pico labeling protocol was applied (Agilent technical note, publication number: 5990-3703EN*)*. Prior to labeling, periplasmic extractions were diluted to different extents to ensure the linear relationship between Im7 amount and signal. Protein ratios were determined by integrating the trough-to-trough peak area. The high sensitivity 250 kit and the pico labeling protocol allowed the visualization and quantification of the very small amount of Im7 present in the absence of Spy, an amount that was not detectable on either Coomassie blue stained SDS-PAGE gels or on the Agilent bioanalyzer using either the 80 or 230 protein kits. Although excellent for determining the ratio of a single defined protein present in different sample preparations, we reasoned that differences in labeling efficiencies between different proteins would make this protocol suboptimal for quantifying percentages of the total protein. Instead, the percentage of Spy and Im7 was quantified in overproducing strains using the protein 80 kit (Agilent) without labeling. Areas corresponding to Spy or Im7 peaks were compared to the total area of all proteins present in the periplasmic extract. In the absence of Spy, the Im7 peak was not detectable; therefore, the percentage of Im7 was calculated by dividing the Im7 percentage in the presence of Spy overexpression with the fold increase of Im7 level. For example, Im7 is 10% of total periplasm in SQ1405. Im7 expression increased 34-fold in SQ1405 compared to SQ1413. The percentage of Im7 in SQ1413 was therefore calculated to be 0.3%. The percentage of Spy present in periplasmic extracts in different EMS strains was quantified similarly with the protein 80 kit.

CD spectra and analysis[****]. Far UV CD spectra (240-185 nm) of Spy were obtained using a Jasco J-810 spectropolarimeter. Spectra of 5 μM Spy in 10 mM sodium phosphate buffer, pH 7.5 were collected at 20 °C using a 0.1 cm path length quartz cuvette. Analysis to predict secondary structure content used the CDSSTR algorithm and reference set #3 on the DichroWeb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml)[196, 197].

[****] Tim Tapley wrote the method of CD analysis , analytical gel filtration, and fluorescence experiments.

Analytical gel filtration. The oligomeric state of Spy was investigated by analytical gel filtration using a Superdex75 column (GE Healthcare). Due to the very low intrinsic extinction coefficient of Spy, fluorescein-labeled protein was used for experiments performed at low protein concentrations. Labeling with fluorescein isothiocyanate (FITC) (Invitrogen) was performed according to instructions from the manufacturer. Various concentrations of Spy (0.1–50 μM, concentration given as a monomer) were injected onto the column using a 100 μL loop, and the protein was eluted with HN buffer (40 mM HEPES-KOH, 100 mM NaCl, pH 7.5). The apparent molecular weight of Spy was determined by comparison to a calibration curve generated with globular proteins of known molecular weights. The low molecular weight standard kit was purchased from GE healthcare and used as suggested by the manufacturer.

Fluorescence Labeling. Labeling of Spy variants (H24C and K77C) was performed by first reducing both mutant proteins in 50 mM sodium phosphate buffer (pH 8) with 10 mM DTT, on ice for 1 h. Excess DTT was removed with a NAP5 desalting column (GE Healthcare). Spy mutants were then incubated with a 3-fold molar excess of monobromobimane in 50 mM sodium phosphate buffer (pH 8) for 1 h at RT. Excess dye was removed with NAP5 desalting columns. Labeling efficiency was determined spectrophotometrically. A saturated stock solution of monobromobimane (Invitrogen) was prepared and the concentration was determined as suggested by the manufacturer, using $\varepsilon_{380} = 5000\ M^{-1}cm^{-1}$. A defined dilution of the free dye in sodium phosphate buffer was prepared and the extinction coefficient ε at 398 nm ($\varepsilon\ (bimane)_{398\ nm} = 4643.7\ M^{-1}cm^{-1}$) as well as a correction factor (1.15 = A280/A398) for the absorbance at 280 nm were calculated. The absorbance at 398 nm was then used to determine the concentration of the dye in the labeled sample and the concentration of the Spy protein ($\varepsilon\ (Spy)_{280\ nm} \approx 1490\ M^{-1}cm^{-1}$) was finally calculated by subtracting the proportional absorbance of the dye at 280 nm from the total absorbance of the sample using the correction factor. The label efficiency appeared to be 1.00 for H24C and 0.98 for K77C.

Fluorescence Anisotropy. Bimane-labeled Spy mutants (0.5 μM final concentration) were incubated in 40 mM Hepes buffer containing 100 mM NaCl (pH 7.5)

at 20°C for 5 min in the absence of casein, then an equimolar amount of casein was added. Anisotropy (r) was calculated according to the following equations:

$$G = I_{hv}/I_{hh} \quad [1]$$

$$r = (I_{vv} - G*I_{vh})/(I_{vv} + 2GI_{vh}) \quad [2]$$

where G is the instrument correction factor, r is anisotropy, and I is the fluorescence intensity measured with polarizers in the orientations indicated by the subscripts[198]. Anisotropy was recorded with a Cary Eclipse Spectrofluorimeter using λ_{ex} = 398 nm (10 nm bandpass) and λ_{em} = 466 nm (10 nm bandpass).

Fluorescence Emission Spectra. Bimane-labeled Spy mutants (0.5 μM) were incubated in 40 mM Hepes, 100 mM NaCl (pH 7.5) at 20°C for 5 min. Emission spectra were then recorded before and after addition of a 3-fold molar excess of casein. Emission spectra were recorded from 408-650 nm, using an excitation wavelength of 398 nm (5 nm spectral bandpass).

3.4.6 Chaperone activity assays††††

Aggregation assays. Thermal aggregation of malate dehydrogenase (MDH) was monitored by light scattering at 43 °C. MDH from pig heart mitochondria (Roche) in 50 mM potassium phosphate, 0.5 mM EDTA, pH 7.5, was diluted to a final concentration of 154 nM into pre-warmed 40 mM HEPES-KOH (pH 7.5) in the absence or presence of various concentrations of Spy. Aggregation of MDH at 43 °C was monitored by light scattering at excitation/emission wavelength of 360 nm using a F-4500 fluorescence spectrophotometer (Hitachi).

Guanidine-denatured MDH, aldolase, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were prepared by incubating 50 μM of the respective protein in 5.4 M guanidine HCl, 50 mM Tris, 20 mM DTT, pH 7.5 for ≥ 2 hours at room temperature. Urea-denatured MDH was prepared by incubating 50 μM MDH in 6.6 M urea containing 10 mM DTT for >1 hr at room temperature. MDH, Aldolase, and

†††† I wrote the methods for MDH aggregation and refolding assays. The rest methods in this section were written by Tim Tapley.

GAPDH were then diluted to final concentrations of 500, 300, and 500 nM, respectively, into HN buffer (40 mM HEPES-KOH, 100 mM NaCl, pH 7.5) pre-equilibrated at 20 °C in the absence or presence of various concentrations of Spy. Aggregation of ethanol-denatured MDH was monitored by incubating 1 μM MDH in HN buffer containing 10% v/v ethanol at 30 °C in the absence or presence of various concentrations of Spy. Light scattering was monitored using a Varian Eclipse spectrofluorimeter equipped with a Peltier temperature control unit. Excitation (2.5 nm spectral band-pass) and emission (5 nm spectral band-pass) were set to 500 nm. The light scattering signal of proteins in the absence of Spy was used for normalization. Aldolase from rabbit heart muscle, alkaline phosphatase from *E. coli*, and GAPDH from baker's yeast were purchased from Sigma. MDH from pig heart mitochondria was purchased from Roche.

MDH refolding assay. MDH activity was assayed as described previously[199]. Briefly, 5–50 μl of the reaction mixture (inactivation or refolding) was mixed with 950 μl substrate (0.15 mM NADH and 1 mM oxaloacetate in 50 mM sodium phosphate buffer, pH 8.0). The absorbance change at 340 nm was monitored and the initial slope was calculated and divided by the value corresponding to 100% activity (*i.e.*, that of an equivalent concentration of native MDH) to obtain the percentage of native MDH activity. To measure the thermal inactivation of MDH, 5-50 μl aliquots were removed at various times and assayed for MDH activity. To initiate refolding of thermally unfolded MDH, samples were shifted to 25 °C at a defined time after the start of thermal inactivation, and aliquots were taken to monitor reactivation.

LDH refolding assay. LDH (from rabbit muscle, Roche) was exchanged into HN buffer (40 mM HEPES, 100 mM NaCl, pH 7.5) using a NAP5 desalting column (GE Healthcare). LDH was then denatured by addition of guanidine HCl to a final concentration of 3.8 M and incubated at room temperature for >2 hours. The influence of Spy on LDH refolding in HN buffer at 20 °C was then tested by diluting denatured LDH 166-fold into HN buffer, to a final concentration of 0.3 μM in the absence or presence of 0.5 μM Spy. Aliquots were tested at various times for LDH activity according to a protocol provided by Sigma. The activity assay buffer contained 1.1 mM pyruvate and 0.12 mM NADH in 100 mM sodium phosphate, pH 7.5. LDH activity was measured by

monitoring the conversion of NADH to NAD^+, which results in a decrease in absorbance at 340 nm.

MDH refolding competition assay. Competition assay with denatured Malate dehydrogenase (MDH) and α-casein was performed as described for MDH alone except that different amounts of α-casein were added to the Spy-WT containing buffer before denatured MDH was added. Ratios of casein: MDH of 1:1, 4:1 and 8:1 were used.

Tannic acid inactivation of enzyme activity. All samples were equilibrated for 5 min at 20 °C before addition of tannic acid (Sigma). Spy was added just prior to temperature equilibration. Aliquots were taken from the assays at different time points (0–60 min), added to the activity assay buffer, and assayed in triplicate.

Bacterial alkaline phosphatase (Sigma) was incubated in HN buffer (40 mM HEPES-KOH, 100 mM NaCl, pH 7.5) at a final concentration of 1 μM in the absence or presence of increasing amounts of Spy and supplemented with 500 μM tannic acid. Then 5 μl aliquots were removed at various times and assayed for alkaline phosphatase activity by mixing with 995 μl alkaline phosphatase activity buffer (100 mM glycine, 1 mM $MgCl_2$, 1 mM $ZnCl_2$, 6 mM p-nitrophenyl phosphate, pH 10.4). Alkaline phosphatase activity was determined by monitoring absorbance changes at 405 nm due to the formation of p-nitrophenol resulting from cleavage of inorganic phosphate from the p-nitrophenyl phosphate substrate (according to the protocol provided by Sigma).

DsbB was purified as previously described[200]. DsbB (0.5 μM) was incubated with 100 μM tannic acid in a buffer containing 50 mM Na phosphate, pH 6.0, 300 mM NaCl, 0.1% n-dodecyl-β-D-maltopyranoside at 20 °C in the presence or absence of Spy. At the time points indicated, 4 μl of the sample was mixed with 196 μl DsbB activity buffer (50 mM Na phosphate, pH 6.0, 300 mM NaCl, 25 μM DsbA, and 25 μM ubiquinone Q1), and DsbB activity was determined as previously described[200].

Aldolase (0.5 μM) was incubated with 16 μM tannic acid in a buffer containing 100 mM K phosphate, pH 7.5, 0.5 mM EDTA at 20 °C in the presence or absence of Spy. At the indicated time points, 5 μl aliquots were withdrawn and mixed with 196 μl aldolase activity buffer (0.1 M Tris, 1.9 mM fructose 1,6-diphosphate, 0.13 mM β-

NADH, 5 units α-glycerophosphate dehydrogenase/triosephosphate isomerse, pH 7.5), and aldolase activity was determined by monitoring the decrease in absorbance at 340 nm due to conversion of β-NADH to β-NAD^+, according to the protocol provided by Sigma. Spy sediments upon ultracentrifugation as a 30 kDa species, indicating that Spy forms a dimer in solution. Thus, except where specifically noted, all concentrations of Spy given in this manuscript are of the dimer. The ratios of Spy:substrate were also calculated based on the assumption that Spy is active as a dimer.

3.4.7 Structural analysis[‡‡‡‡]

Crystallization. Initial crystallization conditions of the His_6-Spy were identified using the $AmSO_4$ suite (Qiagen Inc, Canada).The SeMet substituted protein was crystallized under the same conditions. The best diffracting crystals were obtained by mixing 1 μl protein in the final buffer with 1 μl of reservoir solution containing 0.3 M $CdCl_2$ and 2.4 M $AmSO_4$, under the vapor diffusion hanging drop method. Prior to data collection the crystals were cryoprotected in paratone and flash frozen in liquid nitrogen.

Data collection and processing. The diffraction data were collected from a single crystal at the CMCF-1 beamline at the Canadian Light Source (CLS), Saskatoon, Saskatchewan at the Se absorption edge (wavelength of 0.9792 Å) to 2.6 Å resolution. Data were processed and scaled using DENZO and SCALEPACK in the HKL2000suite[201].

Structure determination and refinement. The structure was solved by the single anomalous diffraction method. Of the 22 expected selenium atoms in the asymmetric unit, 8 were located and refined using autoSHARP[202]. These sites were used to obtain preliminary phases. The model was build manually in Coot[203] using the Se sites as reference points. Of the 146 residues (138 protein + 8 tag), 92 showed clear backbone density in each monomer. Several cycles of refinement using REFMAC5[204] followed by model rebuilding were carried out. The final refinement was performed with PHENIX[205] and included the TLS model for thermal motions. The pertinent data are shown in Table

[‡‡‡‡] Miroslaw Cygler wrote this section.

3. The coordinates and structure factors have been deposited in the Protein Databank with accession code 3O39.

Table 3.7 Data collection and refinement statistics for Spy

Data collection	
Space group	P62
Cell dimensions	
a, b, c (Å)	68.97, 68.97, 124.28
Resolution (Å)*	50-2.6 (2.66-2.6)
R merge*	0.116 (0.575)
Mean $I/\sigma I$*	37.1 (2.9)
Completeness (%)*	97.8 (5.2)
Multiplicity*	9.3 (5.2)
No. of unique reflections	10139
Refinement	
Resolution (Å)	2.6
R_{work}	0.243
R_{free}	0.281
No. of atoms	
Protein	1512
Solvent	21
Ions (Cd)	9
B factors (Å^2)	
Protein	81
Solvent	63
Ions (Cd)	83
Rms Derivations	
Bond length (Å)	0.014
Bond angle (°)	1.32
PDB code	3O39

Numbers in parenthesis correspond to the high-resolution outer shell.

Chapter 4
Discussion and future direction

4.1 Further characterization of Spy

4.1.1 Identification of the *in vivo* substrates of Spy

Spy has been shown to be an efficient ATP-independent chaperone *in vitro*. *In vivo*, its massive expression greatly enhanced periplasmic Im7 expression. The mechanism by which Spy binds substrate proteins is not entirely understood. It is particulary interesting how Spy can aid in substrate refolding in an energy independent manner. It is also unclear what the physiological substrates of Spy are. I propose several experiments to address these questions. To identify the physiological substrates, I plan to use two-dimensional (2D) gel electrophoresis to compare the periplasmic contents from the wild type, *spy* null, and Spy overexpression strains. This technique has proven to be effective in the identification of substrates for DsbA and DsbC[206].

Preliminary experiments were carried out to compare the periplamic contents from a wild type strain (SQ1712) and a strain with the *baeS 416C* mutation (SQ1698, see the strain list in Chapter 3 for details) that led to the massive production of Spy. Pseudo-2D gel experiments performed with an Agilent 2100 bioanalyzer shows that the Spy overexpression strain SQ1698 contains more periplasmic proteins than the wild type strain SQ1712 (Figure 4.1). This observation suggests that Spy stabilizes a number of periplasmic proteins, and that these proteins might be potential substrates of Spy. I plan to use traditional 2D gel technique and mass spectrometry to identify these potential Spy substrates.

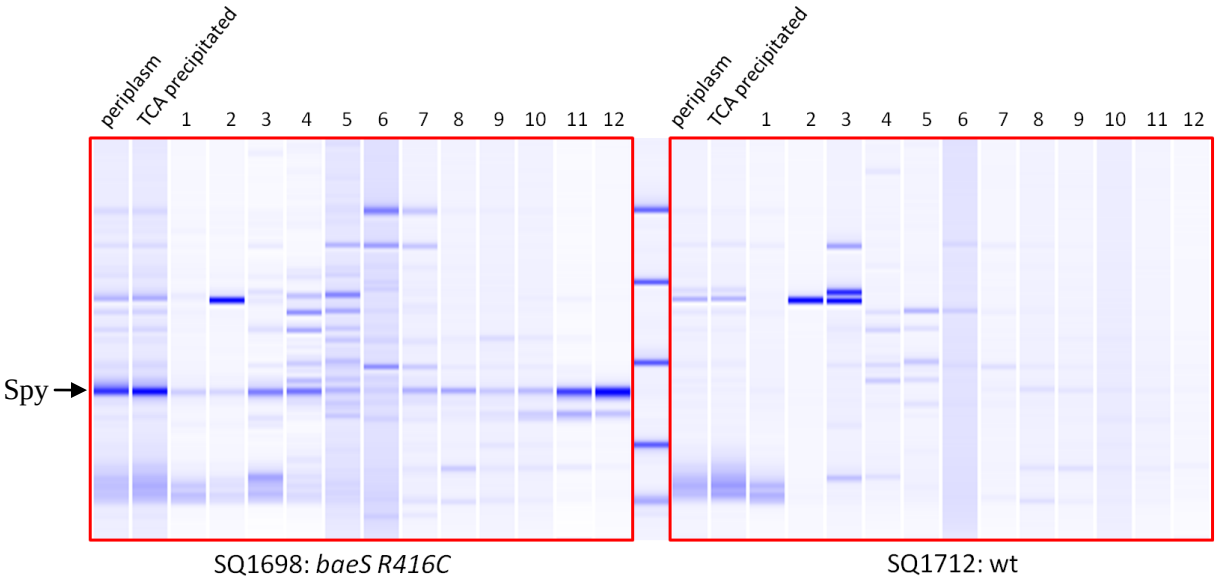


Figure 4.1 Pseudo 2D gel of the periplasmic samples from a wild type and a Spy-overexpression strain

Approximately 300 ml of mid-log phase cells (A_{600}=0.5) grown in LB medium were harvested for periplasmic extraction (methods as described in Ref.[206]). The periplasmic contents were precipitated by incubating the periplasmic samples with 15% TCA on ice for 1h, followed by centrifugation at 13,000 rpm for 20 min. The pellets were washed twice with ice-cold acetone, air dried, and resuspended in 450 µl of reducing 2D buffer. The first dimension of the pseudo 2D gel was separated by isoelectric focusing with an Agilent 3100 off-gel machine. Samples (150µl) were loaded into each well on top of the IPG strip and were separated into 12 fractions along a pH gradient, ranging from pH 3-10. After separating the first dimension, fractionated proteins were recovered from each well in the sample buffer. Each fraction was then subjected to analysis using an Agilent 2100 bioanalyzer with protein 80 chips (as described in Chapter 3). Periplasm: periplasmic extraction from SQ1698 or SQ1712 without any further treatment; TCA precipitated: periplasmic extraction from SQ1698 or SQ1712 precipitated by 15% TCA and resuspended in the 2D sample buffer. TCA precipitation was efficient to precipitate almost all the proteins in the periplasmic extracts. These experiments show a number of differences in proteins expressed in the *baeS R416C* mutant compared to the wild type strain, in addition to Spy overproduction. These proteins might be *in vivo* substrates for Spy that were stabilized by Spy overproduction.

4.1.2 Identification of the substrate binding sites of Spy

The crystal structure suggests that Spy forms a cradle-shaped dimer with unstructured regions at both ends and a cluster of hydrophobic residues inside of the cradle. It has been proposed that structural disorder might play a fundamental role in the

function of many chaperones[190]. It is also well accepted that exposed hydrophobic surfaces in chaperones enable them to specifically recognize protein folding intermediates.We propose that both of the unstructured regions and the hydrophobic regions of Spy are potentially involved in substrate recognition, substrate binding, and substrate protection of Spy's chaperone action. We therefore would like to identify the substrate binding site(s) of Spy through mutagenesis study of these regions.

The structured fragment of Spy contains residues from Phe29 to Thr124. The 28 N-terminal residues and 14 C-terminal residues could not be resolved in the crystal structure, indicating that they are highly disordered. Trypsin digestion of the mature Spy protein reveals a protease-stable fragment composed of residues 21-130[§§§§]. This fragment includes the structured fragment of Spy with additional eight N-terminal and six C-terminal residues. These additional residues are resistant to trypsin, although they are disordered, probably due to the lack of protease cleavage sites in them. We wonder whether the protease-resistant fragment of Spy containing residues 21-130 retains basal chaperone activity and whether the unstructured regions at the N- and C-termini aid in the chaperone activity. To identify whether the unstructured regions at the N or the C terminus contribute to the chaperone activity of Spy, I am going to make various truncation mutants of Spy and measure their activities both *in vivo* and *in vitro*. I will first make the Δ1-28 and the Δ125-138 mutants of Spy separately and in combination. If these truncated regions turn out to be important for the chaperone activity of Spy, I will shorten the truncated part to identify the minimal region at the termini required for Spy's chaperone activity.

We suggest that substrates bind to the concave side of Spy and hypothesize that mutations affecting the hydrophobic environment of the concave side of Spy would diminish its chaperone activity. I will focus on the residues that are implicated in the hydrophobic exposed-surface and are involved in the hydrophobic core, namely Pro56, Met64, Ile68, Met85, Met93, Met97, Leu34, Ile42, Met46 and Ile103 (Figure 3.16). I plan to make serine substitutions to these residues and test chaperone activity of the mutants. I will start with making the L34S, L42S, I68S and I103S single or

[§§§§] Tim Tapley and Philipp Koldewey performed the experiment.

combinational mutants since these mutated residues are conserved or semi conserved among Spy homologs in different species (Figure 3.17). This strategy will hopefully help to identify substrate binding residues in Spy.

4.2 Use of the dual selection system to customize the folding of other test proteins

4.2.1 Problems of protein solubility

The dual selection system works in the periplasm of *E. coli*. As discussed in Chapter 1, this particular compartment facilitates the formation of disulfide bonds, which is important for the stability and activity of many recombinant proteins. Therefore, we propose to test if the dual selection system could be adapted to customize the folding environment of proteins with disulfide bonds.

Initially, I planned to clone six disulfide bond-containing proteins into *bla* (the gene encoding TEM-1 β-lactamase) on the pBR322 vector. These proteins include bovine pancreatic trypsin inhibitor (BPTI), RNaseI, RNaseA, a truncated version of tissue plasminogen activator (vtPA), urokinase (uPA), and proinsulin (Figure 4.2). However, I have only been successful in cloning of BPTI and RNaseI. The insertion of the larger proteins with more complex disulfide patterns exclusively led to frame shifts in the cloned genes. The constitutive expression of these proteins from the bla promoter might be toxic to the cell; as a result, the cell responded by only allowing variants with disrupted reading frames to grow. To circumvent this problem, I cloned these genes into the β-lactamase gene, *bla*, on a low copy vector under the control of a pBAD promoter (pBAD43). The tight regulation of the pBAD promoter, in the absence of the inducer, finally enabled me to obtain the constructs. When arabinose is present (0.2%), cells containing such a construct still exhibit normal growth rate, indicating there is no major effect of the fusion proteins when expressed at lower levels. Therefore, fusion proteins made in the pBAD43 vectors were used for further analysis. Similarly, genes encoding these proteins were inserted into *dsbA* on the pBAD33 vector. The plasmids pBAD33 and pBAD43 are both under arabinose control, but have different origin of replication and are, therefore, compatible with each other.

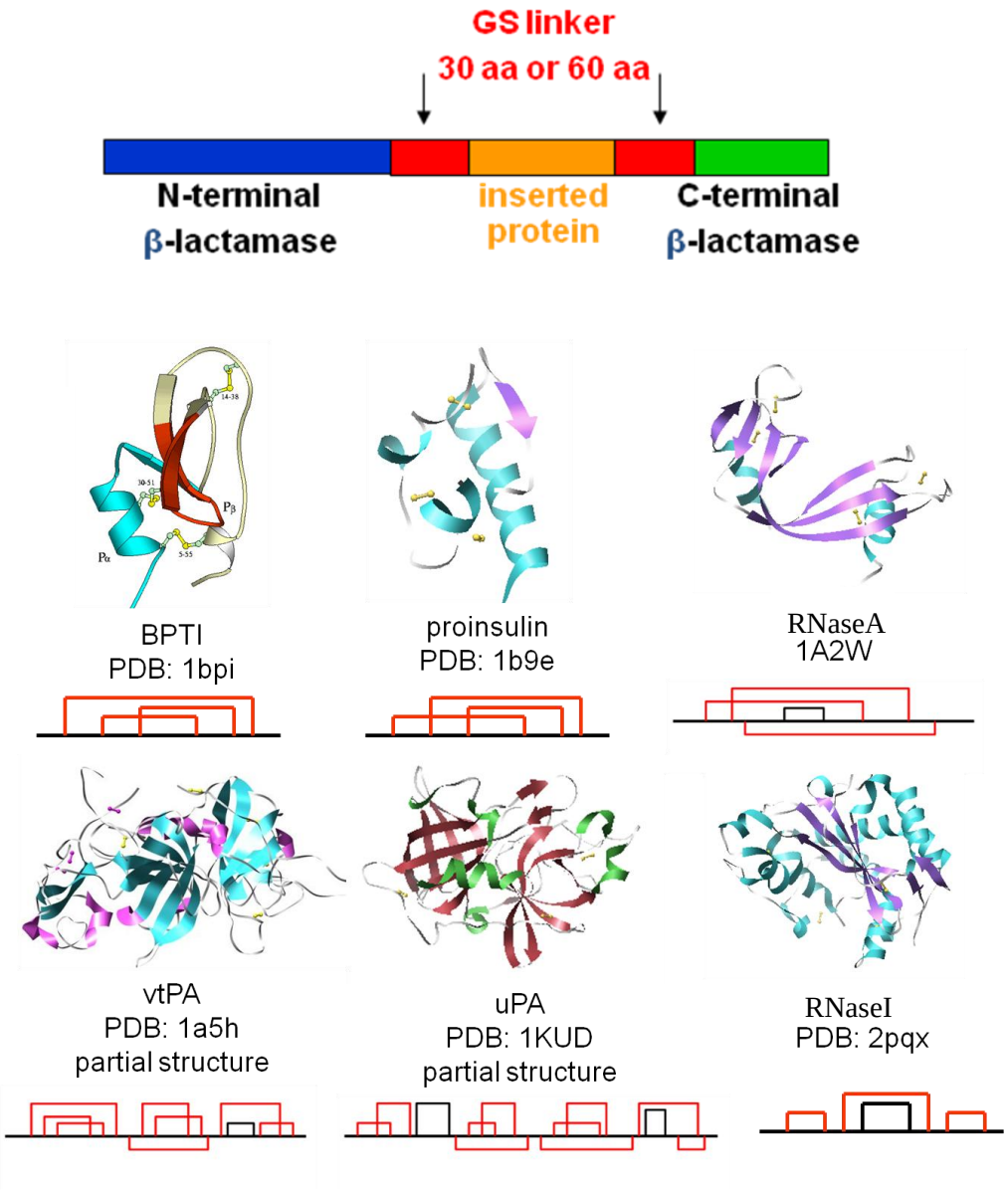


Figure 4.2 Proteins with multiple disulfide bonds

Six test proteins with multiple disulfide bonds are inserted into *bla* on the pBAD43 vector via a 30aa or a 60aa glycine-serine (GS) linker depending on the size of the protein. The disulfide bond patterns of these proteins are shown below their crystal structures. RNaseI is an *E. coli* protein located in the periplasm, so it is a physiological relevant substrate to study periplasmic folding modulators. The other proteins are from eukaryotic origins. BPTI and proinsulin contain 3 nonconsecutive disulfide bonds (red); RnaseA contains 3 nonconsecutive and one consecutive disulfide bonds (black); vtPA, a truncated form of tissue plasminogen activator, contains 8 nonconsecutive and one consecutive disulfide bonds; uPA, the urokinase-type plasminogen activator, contains 10 nonconsecutive and two consecutive disulfide bonds; RnaseI, contains one nonconsecutive and three consecutive disulfide bonds.

In studying the PenV resistance of these fusion constructs, I found an interesting correlation of the size of the inserted proteins, the number of nonconsecutive cysteines, and their PenV or Cadmium resistance. Larger proteins usually require sophisticated machinery to help efficient folding and quite often, these proteins have more disulfide bonds than the smaller ones. Proteins such as vtPA and uPA are therefore the most difficult targets to fold, leading to the lowest PenV or cadmium resistance (Figure 4.3). Intriguingly, the size of the protein seems to be a more important factor than the number of disulfide bonds (compare the R^2 values of the upper left and lower left panels in Figure 4.3). For example, the big protein, pepsinogen, confers very low PenV resistance to the cell when made as a fusion protein with β-lactamase, even though this protein contains no disulfide bonds (Figure 4.3).

The periplasmic disulfide isomerase DsbC has been shown to increase the expression and disulfide bond formation in *E. coli* and eukaryotic proteins having non-consecutive disulfide bonds. For example, overexpression of DsbC results in 16-fold higher yields of active vtPA[207] and its absence causes ~ 10-fold decrease of RnaseI activity[206] and a nearly 100-fold decrease of uPA activity[208] compared to expression in wild type strains. Therefore, my initial test was focused on whether DsbC overexpression increases the PenV resistance of cells containing β-lactamase tripartite fusions with the six test proteins shown in Figure 4.2. To my surprise, neither the deletion of dsbC nor its overexpression changes the PenV resistance (data not shown and Figure 4.4 a) of strains except for the one expressing bla-RNaseI fusion, which showed slight dependence on DsbC (data not shown). On the other hand, a strain expressing vtPA alone in the absence of fusion partner still showed dependence on DsbC (Figure 4.4 b). Therefore, the fusion partner might slightly inhibit the efficient interaction of DsbC with the test proteins, or create additional folding problems that DsbC was not able to solve.

To figure out the reason for DsbC-independence, I first determined whether the fusion proteins were secreted into the periplasm as expected. NaN_3 (0.02%) was used to block the Sec-dependent translocation and the size of the fusion proteins were checked by SDS-page followed by western blot with antibody against β-lactamase. I discovered that all of the six fusion proteins are secreted (Figure 4.5).

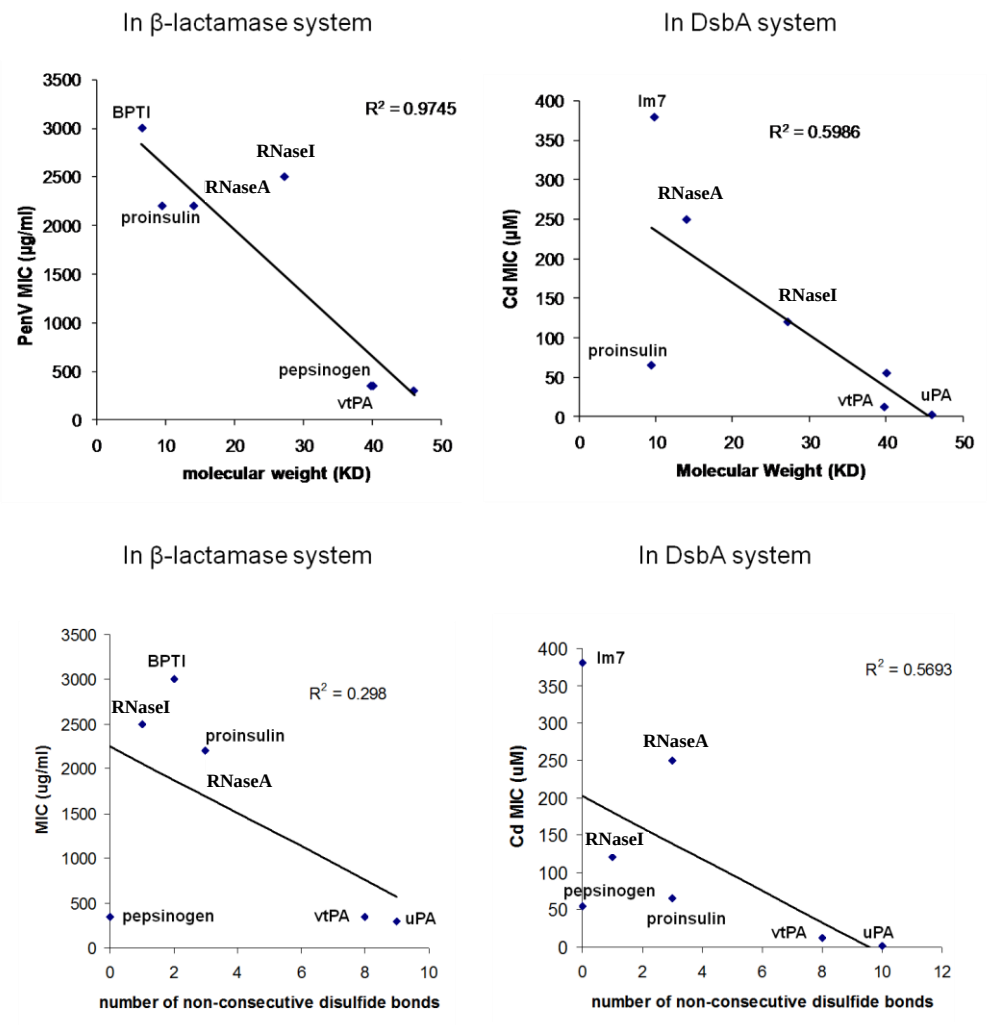


Figure 4.3 Bigger protein or protein with more nonconsecutive disulfide bonds results in lower PenV^R or Cd^R

The minimal inhibitory concentration (MIC) for PenV or cadmium of strains expressing the tripartite fusion proteins was determined by spot titer as described in methods in Chapter 3. The MICs for PenV or cadmium are plotted against the molecular weight or number of non-consecutive disulfide bonds of different inserted proteins. Linear regression was used to evaluate the correlations. Larger proteins are often more difficult to fold, resulting in lower antimicrobial resistance (upper panels). Disulfide bonds, especially the nonconsecutive ones, require effective oxidation and isomerization machinery to form correctly. This created an extra burden on the folding of the protein and therefore lowered the resistance (lower panels).

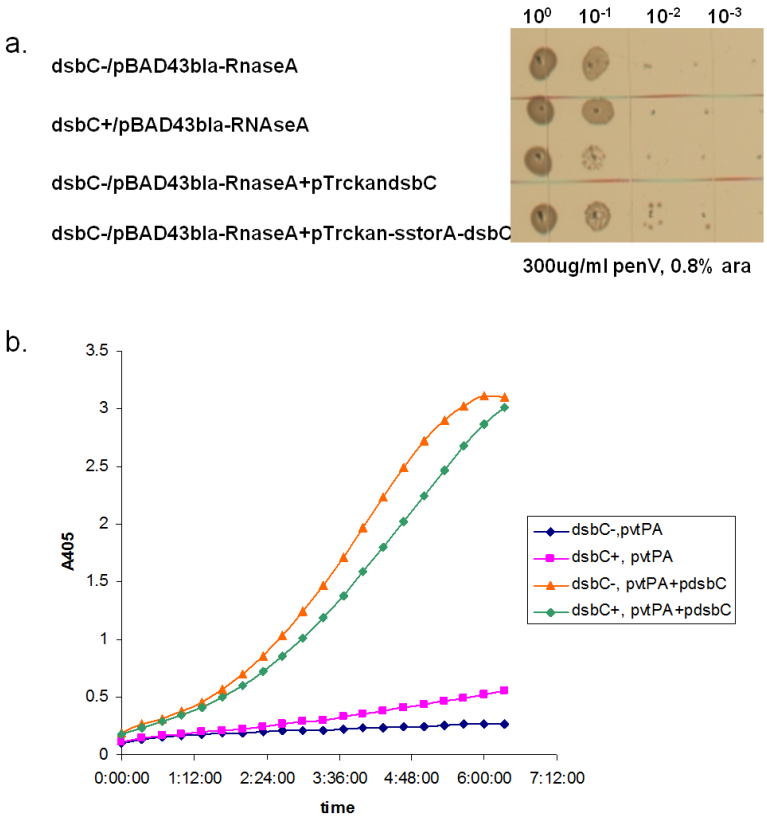


Figure 4.4 The fusion proteins with β-lactamase showed independence of DsbC for folding

a, RNaseA fused with β-lactamase was expressed in a wild type or a *dsbC*¯ strain following induction with 0.8% arabinose. DsbC was overexpressed from a Trc promoter with DsbC's signal sequence or with the Tat signal sequence in the *dsbC* strain. The penicillin V resistance of different strains was assayed by spot titer experiment as described in methods in Chapter 3. Neither the deletion of *dsbC* nor overexpression of DsbC changes the PenV resistance of cells expressing pBAD43bla-RnaseA. Using a Tat-dependent signal sequence seems to increase the amount of DsbC even further but still failed to increase $PenV^R$. Similar results were obtained for β-lactamase fusion constructs with vtPA, uPA, BPTI, and pepsinogen. **b**, vtPA expressed alone in the absence of any fusion partner clearly showed dependence on DsbC. Periplasmic extracts were prepared from each strain at mid-log phase and vtPA activity was assayed for its ability to hydrolyze a chromogenic substrate (spectrozyme PL) resulting in the absorbance at 405nm. Little vtPA activity was detected in the *dsbC* null strain (blue curve). Weak activity of vtPA was detected in the wild type stain (magenta curve), and overexpression of DsbC from a plasmid substantially increased vtPA activity (orange and green curves). pvtPA: plasmid encoding vtPA. pdsbC: DsbC overexpression plasmid.

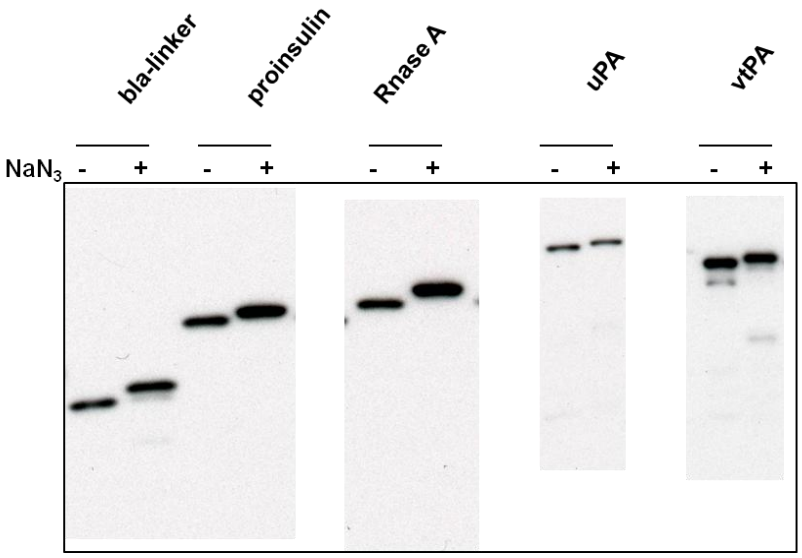


Figure 4.5 The fusion proteins are secreted into the periplasm
Cells were grown to mid-log phase without arabinose induction. Then, 0.02% of NaN_3 was added to one group of cultures to block the Sec-dependent translocation. After incubation at 37°C for 30min, arabinose was added to a final concentration of 0.2% to induce the fusion proteins. After additional 3h incubation, cells were spun down and the whole cell lysates were prepared by boiling the cell pellets in reducing sample buffer containing 2% of SDS. The fusion proteins were detected with western blot using the anti-β-lactamase antibody, which was previously shown to effectively detect bla-Im7 fusion constructs. For each fusion protein compared (with or without NaN_3 treatment), the lower band represents a smaller molecular whose signal sequence was cleaved off upon successfully translocation. The higher band represents the species that was trapped in the cytoplasm with an intact signal sequence. When the Sec pathway functioned normally, there was only one band for all the fusion proteins as well as the Bla-linker, indicating a complete translocation. Results were similar for bla-linker-BPTI/RNaseI proteins (not shown).

Another reason why overexpression of DsbC failed to increase penicillin resistance of cells expressing the fusion proteins might be that these fusion proteins are insoluble and DsbC is not able to help insoluble proteins. To check the solubility of the secreted proteins, I prepared the soluble and insoluble fractions of each test strain. After lysozyme treatment and three cycles of freezing and thawing, the soluble and insoluble materials were separated by centrifugation. For bla-BPTI (wild type or its variant C14SC38S), only half of the protein was soluble. For proinsulin and RNaseA, the majority of the tripartite fusion proteins were associated with the insoluble fraction, though some soluble proteins were detectable. For uPA, vtPA, and RnaseI, unfortunately almost all the fusion proteins

are insoluble. On the other hand, at least 90% of bla-linker was soluble and more than half of bla-Im7 (wt and F84A variant) was soluble (Figure 4.6). I concluded, therefore, that the insolubility of the fusion proteins might largely mask any DsbC-dependence of those fusion proteins.

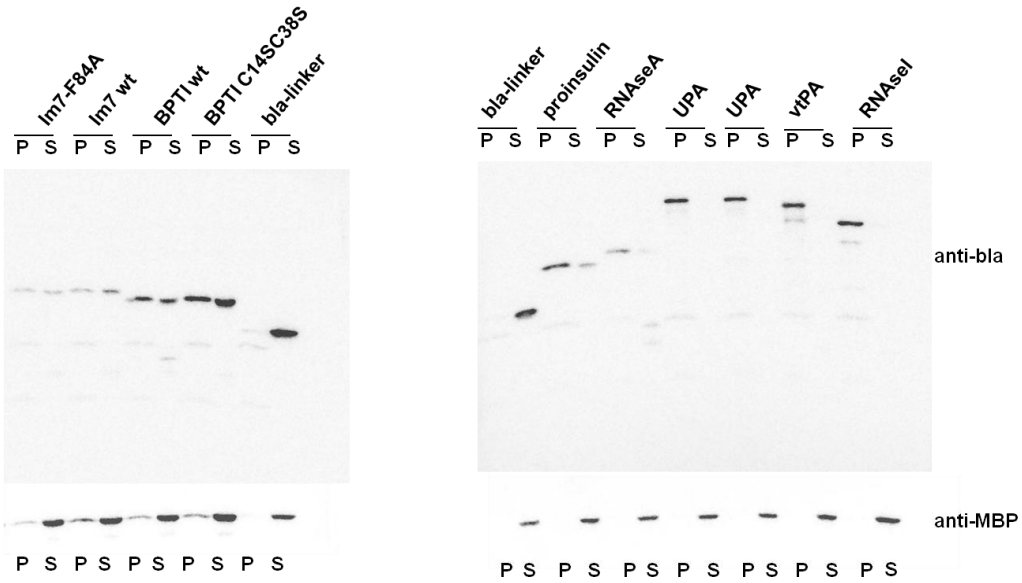


Figure 4.6 Solubility test of the fusion constructs
Cells for protein solubility test were grown to mid-log phase, then induced with 0.2% arabinose and 0.3% maltose for 3h before harvest. The cell pellets were resuspended in PBS buffer containing 1 mg/ml lysozyme, and then subjected to freezing and thawing in a dry ice/ethanol bath and an ice cold water bath. After three cycles, the cell suspension was nearly clear. 1 unit of DNase was added to the cell suspension to remove DNA and facilitate downstream application. Insoluble materials were pelleted by centrifugation at 13,000 rpm for 20 min at 4°C. The supernatant was carefully transferred to another tube and the pellet was resuspended in PBS to the original volume. Samples were then separated by SDS-PAGE and blotted with antibody against β-lactamase or MBP. The internal control, MBP, was shown to associate with the soluble fraction for most samples. For Bla-Im7 wt/F84A and Bla-BPTI wt/C14SC38S samples, there was some insoluble MBP detected, which might be an artifact because of incomplete cell lysis. Therefore, the 50% solubility of those proteins was underestimated. P: insoluble fraction (pellet). S: soluble fraction.

Until now, in contrast to ClpB, DsbC has not been reported to actively solublize protein aggregates. A number of events must occur to link protein folding to antibiotic

resistance in this situation: firstly, the fusion proteins have to be rescued from aggregation, and then the folding of the substrate proteins must be promoted to enable the assembly of active β-lactamase to confer antibiotic resistance. Alternatively, the cells must be efficient at folding these eukaryotic proteins to prevent aggregation in the first place, which might be more difficult compared to rescuing an unstable yet soluble prokaryotic protein, Im7. Spy overproduction, unfortunately, was not able to solve these problems, at least for the case of RNaseI. The selection for chromosomal mutations that solubilize difficult targets such as RNaseI and aid in their folding would definitely enlarge our understanding of periplasmic chaperones and possibly identify novel periplasmic disaggregation chaperones.

4.2.2 Is the dual selection system really necessary?

The dual selection system was used to identify potential periplasmic folding modulators. In the PenV selection and cadmium screening process, only colonies that showed a strong increase of the resistance to both antimicrobials were chosen to be further analyzed. Nine of the ten resistant EMS strains turned out to have a mutation in *baeS*. EMS7, the only strain that does not contain a mutation in *baeS*, however, showed an extremely high level of cadmium resistance and a slight PenV resistance. Moreover, the same mutation of *baeS* in EMS6, EMS8, and EMS10 conferred similar levels of PenV resistance but different levels of cadmium resistance (Figure 3.6 in Chapter 3). Furthermore, the reconstituted strain SQ1640 (SQ1306, *baeS R416C, yegL::kan*) showed almost the same level of PenV resistance as EMS4 (Figure 3.10 in Chapter 3), but apparently lower cadmium resistance than EMS4 (data not shown). Therefore, these EMS strains probably contain other mutations that contribute to intrinsic cadmium resistance.

For these reasons, we wondered whether the cadmium selection system is really necessary. To test this, the *baeS* genes in other EMS strains (EMS11-22) were amplified and sequenced[*****]. These 12 EMS strains were PenV resistant but were only slightly more cadmium resistant than the starting strain, SQ1306, or as cadmium sensitive as SQ1306. Surprisingly, 8 of the 12 EMS strains were found to contain mutations in *baeS*.

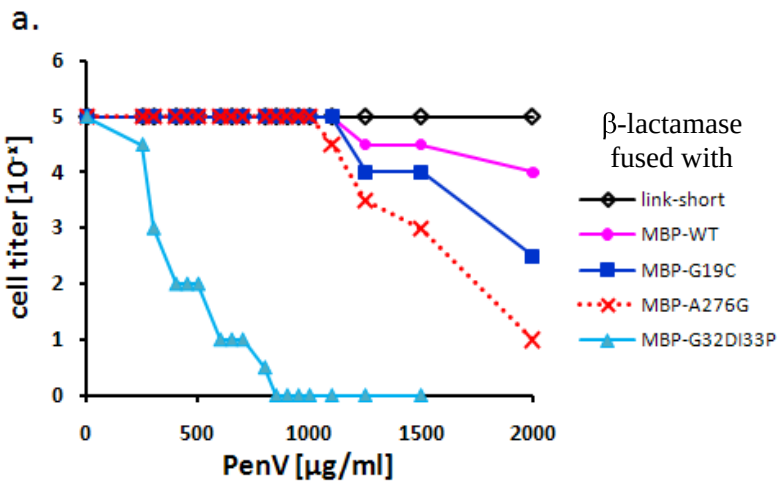
[*****] The experiment was performed by Jennifer Pfizenmaier.

EMS14, EMS15, and EMS21 contain the P255S mutation in *baeS*. EMS17 has the G421D mutation, EMS19 has the P255L mutation, EMS20 has the G261D mutation, and EMS18 has the G324DR416H mutation in *baeS*. EMS13 contains the same *baeS E264K* mutation as in EMS9, with a similar PenV resistance as EMS9. Western blot showed that the total amount of bla-Im7 L53AI54A was similar in EMS9 and EMS13 (data not shown), and all of the EMS strains with a *baeS* mutation increased the Bla::Im7 L53AI54A total amount (data not shown). Thus, it seems that the screening for cadmium resistance is not absolutely necessary, as long as the level of the fusion proteins is measured. Though cadmium screening provides a quick and robust identification of interesting mutants, considering the high toxicity of the cadmium, we suggest using the measurement of protein level as a second screening method as described in section 3.2.2 in Chapter 3. This strategy was used to select for folding factors of MalE31, an aggregation prone mutant of the periplasmic maltose binding protein (MBP).

4.2.3 Example on β-lactamase-MalE31[†††††]

As previously discussed, the presence of disulfide bonds makes the folding process more complicated for large, aggregation-prone proteins. To test the folding functions alone, we studied the folding of a disulfide bond-free but aggregation prone protein, MalE31 (MBP G32DI33P), in the β-lactamase system. MBP captures maltose in the periplasm and transports the bound maltose to the inner membrane ABC transporter MalFGK, which then undergoes conformational changes to release maltose into the cytoplasm. MBP is among the proteins whose folding pathways have been extensively studied. It has variants exhibiting different folding abilities with measured thermodynamic stability values[209]. One of its variants, MalE31, forms aggregates in the periplasm. Upon insertion of the different variants into the β-lactamase system, we established a qualitative correlation of PenV resistance and the thermodynamic stability of MBP variants (Figure 4.7). It is noteworthy that the MalE31 variant showed dramatically decreased PenV resistance compared to other MBP variants, which are not aggregation prone (Figure 4.7).

[†††††] These experiments were performed by Jennifer Pfizenmaier. I supervised the cloning procedure and provided suggestions and feedback.



b.

MBP variant	ΔΔG⁰
wt	0
G19C	2.3
A276G	1.5
G32DI33P	N.A

Figure 4.7 Destabilized MBP variants confer lower PenV resistance
Penicillin V resistance of cells expressing different MBP variants fused to β-lactamse were determined by spot titer experiment as described in methods in Chapter 3. The maximal dilutions of cells which still allowed the cell growth at each penicillin V concentration were plotted against penicillin V concentrations. **a**, The titration curves show that the highest PenV resistance comes from the relatively most stable variant of MBP, namely the wild type. The aggregation prone variant G32DI33P (MalE31) results in the lowest level of PenV resistance. **b**, Measured thermodynamic stabilities of MBP variants as presented by $\Delta\Delta G^0$ values[209]. A positive value in $\Delta\Delta G^0$ corresponds to a decrease in stability. Accordingly, the G19C variant is less stable than the A276G variant. This variant, however, survives to higher PenV concentration than the A276G variant in the spot titer experiment. Therefore, the PenV resistance provided a qualitative measure of the stability of β-lactamase-MBP fusion proteins.

We thought with the Bla-MBP G32DI33P system, we could select for chromosomal mutations that solubilize and probably stabilize the highly aggregation-prone proteins. We used western blot with anti-β-lactamase or anti-MBP antibody to screen for increased solubility. Initially we obtained mutants that contain more soluble protein; however, these mutants all turned out to be revertants to wild type MBP after we sequenced the bla-MalE31 plasmid. We attributed this problem to the homologous recombination of the *bla-malE31* gene with the chromosomal wild type *malE*, and we deleted the endogenous *malE* to prevent this problem. We also constructed a plasmid to overexpress MalE31 in the absence of any fusion context, which would provide additional information on whether the chromosomal mutations really helped to solublize

this aggregation-prone variant. With the anti-MBP antibody, both the levels of MBP itself or in the fusion context with β-lactamase could be simultaneously detected. We are now working on the mutagenesis and screening of a *malE* null strain containing both plasmids.

4.2.4 PolyQ-containing proteins are not aggregating in the periplasm, suggesting the existence of disaggregation factors

Neurodegenerative diseases are usually linked to the abnormal expansion of a polyglutamine (polyQ) track in disease-associated genes[210]. The expanded polyQ domain causes the affected protein to become aberrantly folded and aggregation-prone, which is toxic for the organism. Morimoto *et.al* proposed a mechanism for the toxic effects of the polyQ aggregates based on studies of the temperature-sensitive/polyQ *C. elegans* strains (reviewed in Ref. [211]). In their model, the polyQ proteins exert toxicity by putting stress on the general cellular folding machinery. Under normal growth conditions, misfolded proteins are adequately dealt with by the cellular folding surveillance system and the chaperone network. However, the expression of polyQ proteins overwhelmed the system, leading to the imbalanced homeostasis and toxicity. The toxic effect of polyQ, therefore, is not restricted to just causing one particular protein to aggregate and lose its function. Consistent with this model, the insertion of a polyQ track into proteins that are completely unrelated to the known neurodegenerative diseases produced aggregates and neurodegeneration in mice[212].

To study the effect of polyQ on the conformational changes and folding of proteins, peptides containing different lengths of polyQ were synthesized and characterized[213]. PolyQ monomers are disordered, and associate into aggregates when the length is greater than 16 residues[213]. Although the aggregates formed by Q16 are still soluble, those formed by Q20 or longer are sedimentable over time[213]. With their intrinsic aggregation-prone properties, these longer polyQ tracts destabilize and cause partial unfolding of the proteins into which they are inserted, and mediate the aggregation of the unfolded intermediates by interglutamine interactions[214]. The length threshold of a polyQ track to mediate aggregation of affected proteins is thought to be 35-45 consecutive glutamine residues[215].

With the dual selection system, I wondered whether the insertion of polyQ would cause aggregation of the tripartite fusion proteins and result in PenV or cadmium sensitivity. I inserted 20Q, 45Q, 51Q, and 87Q into both β-lactamase and DsbA via a 30 amino acid serine-glycine linker, and tested the resistance of the corresponding strains. Although insertion of 20Q had almost no effect, insertion of 45Q, 51Q, and 87Q dramatically decreased resistance to PenV and cadmium, respectively (Figure 4.8).

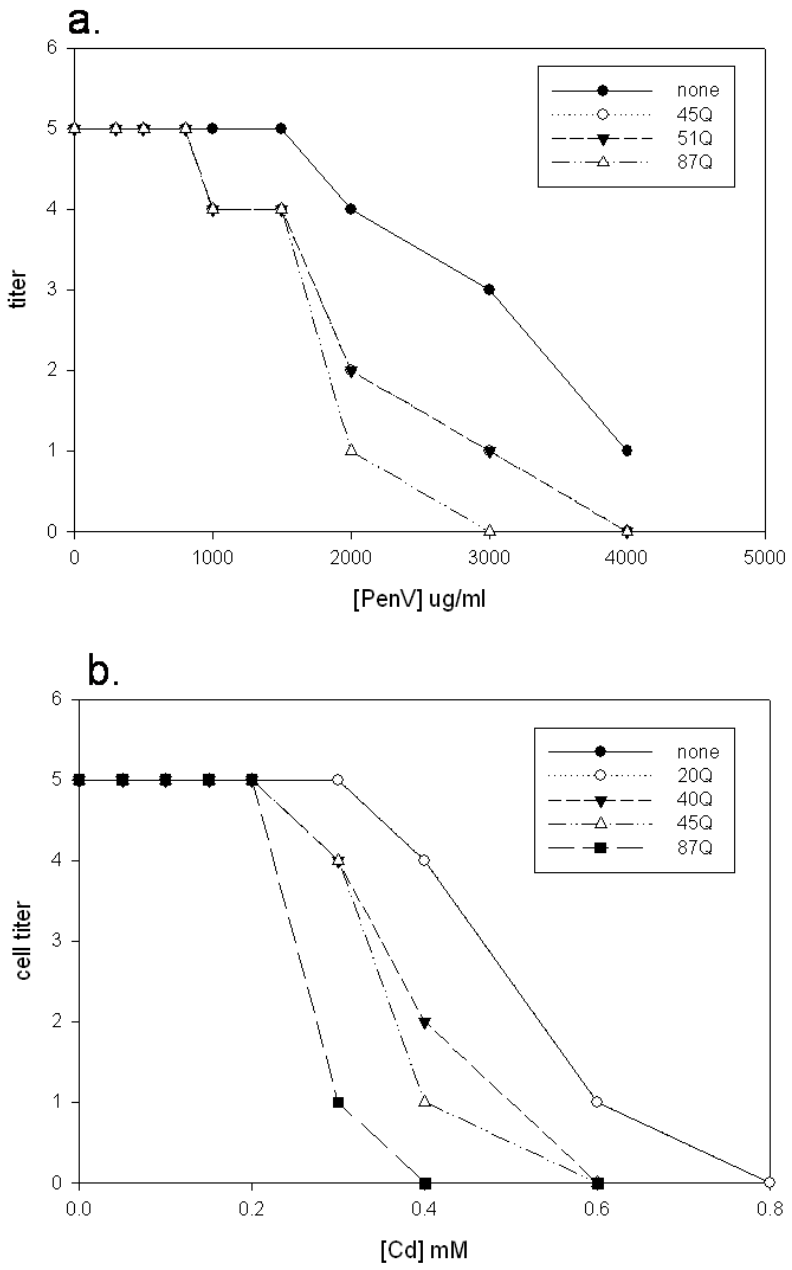


Figure 4.8 Insertion of polyQ into the dual selection system

a, PolyQ tracks of different lengths were inserted into pBAD33bla-linker (β-lactamase cloned into pBAD33). The expression of the tripartite fusion proteins was induced by the addition of 0.8% arabinose in the LB plates. The spot titer curves showed that resistance of the cells towards PenV decreased when the length of polyQ increased. The 45Q curve overlapped the 51Q curve. None: no polyQ insertion in β-lactamase. **b**, Similar trend of cadmium resistance was shown: the insertion of longer polyQ resulted in lower cadmium resistance. PolyQ was inserted into pBAD33DsbAlinker and the fusion proteins were induced by 0.1% arabinose. Insertion of 20Q into DsbA had almost no effect on cadmium resistance as was reflected by the overlap of the 20Q curve and no insertion curve. None: no polyQ insertion into DsbA.

To test whether the decreased resistance was caused by decreased protein solubility, or by the conformational constraints from the larger inserts, I examined the solubility of the tripartite fusion proteins with β-lactamase. Surprisingly, the insertion of 45Q, 51Q, and 87Q did not result in aggregation of the tripartite fusion proteins; instead, the majority of the fusion proteins were recovered from the soluble fraction (Figure 4.9). Examination of the stacking gel showed no insoluble materials. The soluble fraction of the polyQ-containing proteins showed two adjacent bands, probably indicating the degradation of the fusion protein. The two bands may not be the precursor and the mature form of the fusion protein, since the smaller band was absent in the whole cell samples. This phenomenon was only restricted to proteins with the polyQ insert but not the Bla-linker protein, suggesting that the degradation was caused by the insertion of polyQ. However, the degradation did not separate β-lactamase into two parts completely; otherwise this would lead to bands of even lower molecular weights. Therefore, the polyQ insertion might still stick together after partial protease digestion, leading to a slightly smaller fusion protein.

A similar experiment was carried out to examine the solubility of the cytoplasmic aminoglycoside phosphotransferase protein with polyQ insertions[‡‡‡‡‡]. Insertion of the same polyQ tracts into a cytosolic protein clearly diminished the solubility of the fusion proteins (Figure 4.10). Even the insertion of only 20 consecutive glutamines made more than half of the fusion protein insoluble. The insertion of 87Q made more than 90% of the fusion protein insoluble. These results clearly demonstrated the ability of polyQ to induce the aggregation of the affected proteins in the cytoplasm. As I discussed in

[‡‡‡‡‡] The experiment was performed by Ajmal Malik.

Chapter one, the environment of the periplasm differs greatly from that of the cytoplasm. Therefore, the inability of polyQ-containing proteins to aggregate in the periplasm might be the consequence of residing in a different environment. On the other hand, no aggregation might reflect a more powerful periplasmic chaperone net work and a folding surveillance system which better buffers the homeostasis of the periplasm.

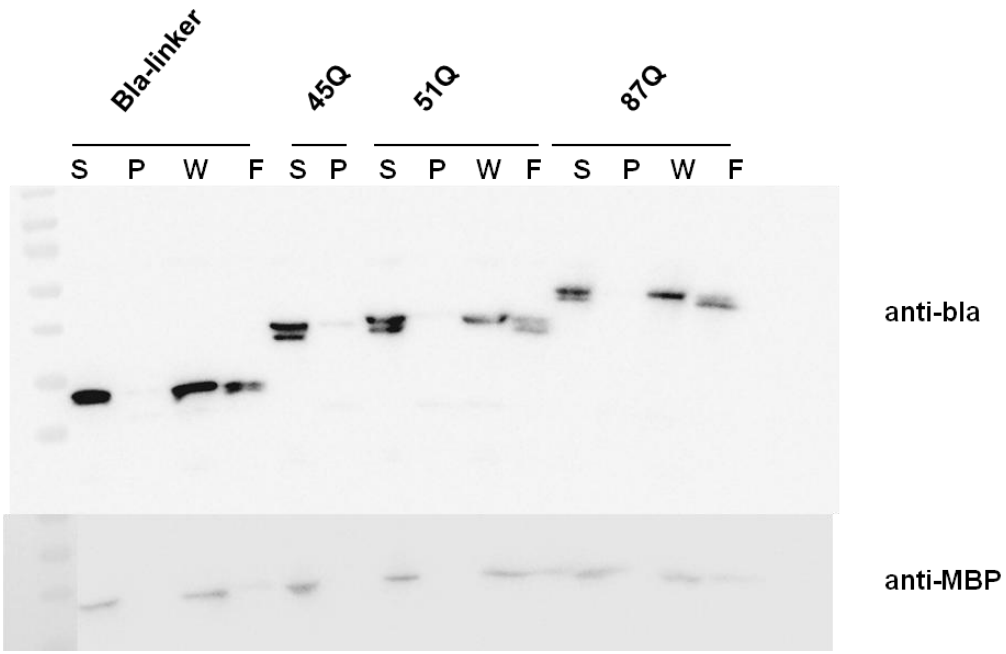


Figure 4.9 Bla-polyQ tripartite fusion proteins are soluble in the periplasm
Cells expressing different tripartite fusion proteins were grown to mid-log phase and induced with 0.2% arabinose and 0.3% maltose (for MBP induction) for 3 hours before harvest. To prepare whole cell lysates, cells were normalized to A_{600}=1.0, spun down and resuspended in reducing loading buffer containing 2% SDS. Some samples were treated with 88% formic acid to solubilize possible amyloid-like aggregates. After the addition of formic acid, samples were spun in a speed vacuum for 1h until all the formic acid was evaporated. Then the dried samples was resuspended in reducing loading buffer and sonicated until all the materials were solubilized. To prepare soluble and insoluble fractions, cells were lysed by the freeze-thaw method as previously described. The soluble fraction was the supernatant after spinning the samples at 13,000 rpm for 30min. The pellet was resuspended in an equal amount of reducing SDS buffer to make the insoluble fraction. After boiling for 5min, the whole cell lysates or the soluble/insoluble fractions were separated by SDS-PAGE and were blotted by anti-β-lactamase antibody or anti-MBP antibody. S: soluble fraction; P: pellet; W: whole cell; F: formic acid treated. Bla-linker: β-lactamase with the GS linker but no polyQ insertion.

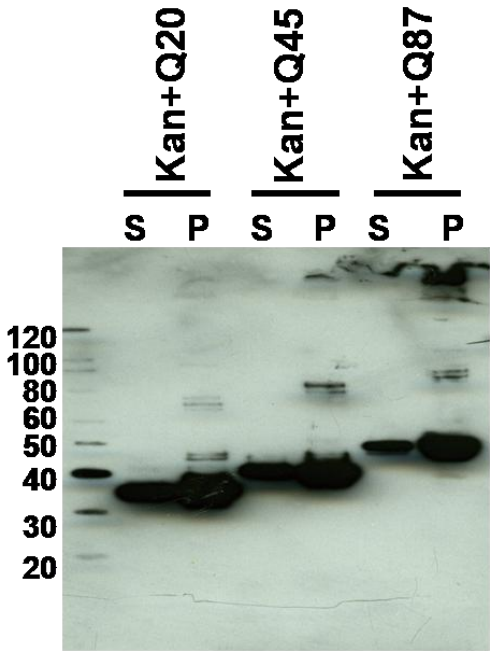


Figure 4.10 PolyQ induces aggregation when inserted into a cytosolic protein[§§§§§]
Q20, Q45, and Q87 were inserted into the cytoplasmic protein aminoglycoside phosphotransferase. Cells were fractionated into soluble and insoluble fractions in a similar procedure as described in Figure 4.9. The majority of the fusion proteins were insoluble and the insolubility increased upon the length increase of polyQ. S: soluble fraction; P: pellet.

The formation of periplasmic amyloid fibers has not previously been reported. CsgA and CsgB form curli, the functional bacterial amyloids, on the surface of *E. coli*[216]. If the outer membrane channel for the secretion of the curli subunits is blocked, CsgA and CsgB are confined in the periplasm[217]. Instead of polymerizing into fibers, these curli subunits are degraded by proteases. This observation, together with the inability of the polyQ-containing proteins to aggregate in the periplasm, suggested the possible existence of a powerful disaggregation system for the amyloidogenic proteins. Searching for chromosomal mutations that promote periplasmic amyloid formation would help to elucidate the mechanisms of amyloidosis, identify factors to prevent its occurance, and

[§§§§§] The figure was prepared by Ajmal Malik.

contribute to futher understanding of the folding and unfolding environment of the periplasm.

To summarize my thesis work, I have used directed evolution techniques to optimize the folding environment of the bacterial periplasm. I investigated the important role of the CXXC active site in thiol-disulfide oxidoreductases from the thioredoxin superfamily. Our results and previous results in the literature indicated that this active site has the remarkable ability to confer a large number of very specific properties on thioredoxin related proteins, including redox potential, enzymatic reactivity, interaction with substrate proteins and reoxidants, as well as the ability to function as a disulfide isomerase. This is a good example of the way that nature can use a particular scaffold to evolve functional properties of enzymes.

For my second project, I developed a dual selection system based on the previously established β-lactamase system. This dual selection can conveniently monitor protein folding *in vivo* by linking protein folding to the bacterial resistances towards two different antimicrobials. I used this system and directed evolution techniques to identify periplasmic folding modulators that can improve the folding of a very unstable test protein. I uncovered Spy as a novel chaperone and performed experiments (with contributions from others) to show that Spy is acting independently of any energy cofactor as an effective folding chaperone. In collaboration with Prof. Cygler we obtained the crystal structure of Spy. The structure reveals a flexible cradle-shaped dimer, with an apolar concave surface. We plan to carry out experiments to identify the physiological substrates of Spy and to investigate its chaperone mechanism. Our strategy opens up new routes for chaperone discovery and the custom tailoring of the *in vivo* folding environment for different proteins.

Appendix

Major cytoplasmic and periplasmic folding modulators

Cytoplasmic folding modulators

Chaperones involved in the *de novo* folding of cytosolic proteins

The folding of a cytoplasmic protein may start after the polypeptide is released from the ribosome (posttranslationally), or it may occur cotranslationally. Chaperones can participate in early stages, when the polypeptide is still being synthesized. This is functionally significant, since the N-terminus of a nascent polypeptide chain is exposed in a partially folded, aggregation-sensitive state during translation. Chaperones are often required to protect these nascent polypeptide chains, and to block the nonnative inter-chain interactions[218]. In bacteria, trigger factor binds directly to the ribosomal subunit in proximity to the polypeptide exit site (reviewed in Ref. [219]). It binds linear polypeptides and thus maintains elongated polypeptides in a non-aggregation state until sufficient structural information for proper folding becomes available via continued chain elongation. Trigger factor has a peptidyl-prolyl cis-trans isomerase (PPIase) domain and therefore has the ability to catalyze prolyl cis-trans isomerization. Evidence showed that trigger factor occupies almost 90% of ribosomes *in vivo*[220]. Around 70% of trigger factor client proteins may fold spontaneously upon releasing from the ribosome without further assistance[11].

Cytosolic polypeptides that need further assistance are passed on to the ATP dependent DnaK-DnaJ-GrpE chaperone system for folding (reviewed in Ref. [11]). DnaK is the prokaryotic homolog of eukaryotic Hsp70 proteins[221]. It consists of an N-terminal nucleotide binding domain, and a C-terminal substrate binding domain. The latter can be further divided to a β sandwich subdomain and an α-helical lid[222]. Substrates are delivered to ATP-bound DnaK by its cochaperone DnaJ (an Hsp40 homolog) (Figure 5.1). DnaJ accelerates ATP hydrolysis leading to closure of the α-helical lid and tight

substrate binding by DnaK. DnaJ then dissociates from DnaK, and the nucleotide exchange factor GrpE binds to DnaK. GrpE induces ADP dissociation and ATP binding to DnaK, resulting in the lid opening and substrate releasing. Released substrates may then directly acquire their native conformations, or they may rebind to DnaK for another cycle of chaperone-assisted folding. Alternatively they can be transferred to a downstream chaperonin system for additional refolding. Binding to DnaK protects the hydrophobic regions of unfolded substrates, but there is also evidence suggesting that DnaK unfolds kinetically trapped folding intermediates, to convert them into productive folding intermediates (reviewed in Ref. [223]). Besides DnaK-DnaJ, *E. coli* K12 strains encode two other proteins with homology to Hsp70 and five additional partner J-domain proteins (reviewed in Ref. [224]). It should be noted that there is partial overlap between trigger factor clients and DnaK-DnaJ clients, and DnaK-DnaJ can work in conjunction with trigger factor while the polypeptide is still being synthesized on the ribosome. However, DnaK does not bind to the ribosome directly as trigger factor does. Around 20% of proteins adopt their native structures with the help of the DnaK-DnaJ-GrpE chaperone system.

The remaining 10% of the proteins are captured by the GroEL-GroES chaperonin system for effective folding (reviewed in Refs.[225-227]). Chaperonins are large, double ringed protein complexes that provide a cage-like cylindrical structure for a polypeptide to fold inside. GroEL (an Hsp60 homolog) forms a double ring structure of two stacked heptamers. The co-chaperone GroES (an Hsp10 homolog) binds to one end of the ring and seals the chamber. Substrate proteins bind GroEL initially through interaction with its hydrophobic patches at the end of the open ring (Figure 5.2). The subsquential binding of ATP and GroES to the substrate binding end of GroEL causes a conformational change of the chaperonin leading to the enlargement of the chamber and the displacement of the substrate into the chamber. The conformational change of GroEL may cause the unfolding of the substrates, so that misfolded proteins have another chance to fold correctly. The chamber of GroEL is thought to act as a micro-folding environment where the substrates are shielded and aggregation is no longer a problem. The inner wall of the chamber is highly hydrophilic and net negatively charged, properties thought to promote folding of substrate proteins by accumulating ordered water molecules, thereby

generating the force to bury exposed hydrophobic residues in the substrates[228]. The releasing of substrate proteins is coupled to ATP-ADP exchange. Once the bound ATP is hydrolyzed on one ring, another substrate, together with an ATP molecule and a second GroES, can bind to the other ring of GroEL, triggering the release of both the folded substrate and the GroES that was initially bound on the opposite side (Figure 5.). *In vivo*, GroEL-GroES clients are usually between 20-50 kDa with α/β or α+β domain topologies. However, GroEL is capable of binding almost half of the proteins of the *E. coli* proteome *in vitro*, if those proteins are unfolded or partially folded[229].

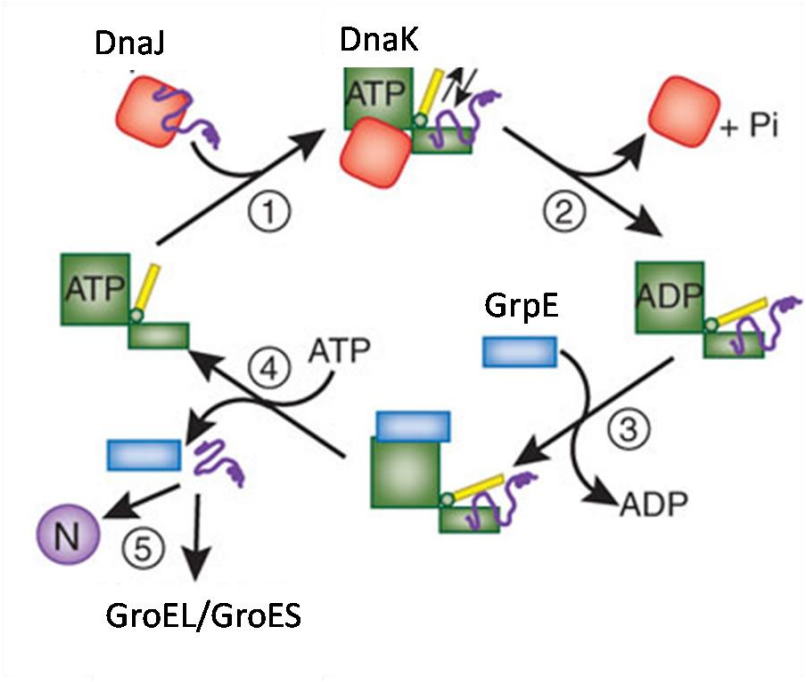


Figure 5. 1 DnaK-DnaJ-GrpE reaction cycle
Both the figure and figure legend were adapted from Fig 5 in Ref.[230] (1) DnaJ delivers substrate to ATP-bound DnaK. The ATP-bound state of DnaK has high association and dissociation rates for substrates. (2) Hydrolysis of ATP to ADP, accelerated by DnaJ, results in closing of the α-helical lid (yellow) and tight binding of substrate by DnaK. DnaJ dissociates from DnaK. (3) Dissociation of ADP catalyzed by GrpE. (4) Opening of the α-helical lid, induced by ATP binding, results in substrate release. (5) Released substrate either folds to native state (N), is transferred to downstream chaperones or rebinds to DnaK.

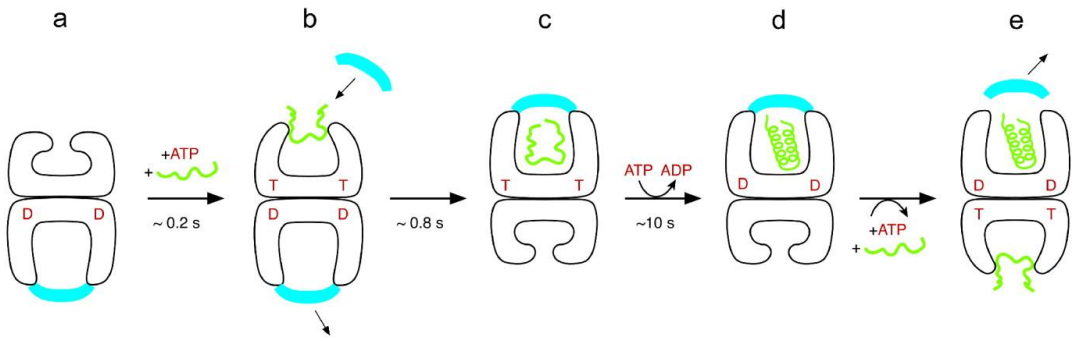


Figure 5. 2 GroEL-GroES reaction cycle
The figure was adapted from Fig. 2 in Ref. [226]. **a**, GroEL forms an asymmetric double ring structure with GroES (cyan) to seal one of the rings. While a substrate protein is folding in the closed ring (not shown on the figure), another non-native substrate (green) may be delivered to the open ring of GroEL. **b**, ATP binding to the apical ring of GroEL triggers a conformational change that results in substrate capture into the chamber. ATP binding also sends a signal to the opposite ring to eject its ligands: GroES, folded substrate (not shown), and ADP. **c,** substrate encapsulation is completed upon binding of GroES. Substrate folds in the hydrophilic chamber. ATP is then hydrolyzed to ADP, which primes the substrate release by weakening the affinity of GroEL and GroES. **d**, substrate release is triggered by the binding of ATP, another substrate, and another GroES (not shown) to the ring on the opposite side. **e**, a new round of substrate refolding cycle begins.

To summarize, trigger factor, DnaK-DnaJ-GrpE, and GroEL-GroES are the main chaperones that mediate *de novo* protein folding in the cytoplasm. Trigger factor does not rely on an energy source to perform its chaperone function. The other two systems use ATP hydrolysis coupled to conformational changes to provide energy.

Cytosolic chaperones involved in stress response

Under stress-induced denaturation conditions, DnaK-DnaJ-GrpE and GroEL-GroES systems refold denatured proteins. Interestingly, those chaperones, along with many others, are themselves heat shock proteins whose synthesis is strongly induced by high growth temperatures[49]. Their increased expression under heat or other stress conditions that increase cellular protein misfolding indicates an important role for them in combating stress. The DnaK-DnaJ-GrpE and GroEL-GroES systems often work in cooperation with other heat shock cytoplasmic chaperones that usually function as holding chaperones. Under severe stress conditions, these holding chaperones prevent

unfolded proteins from irreversible aggregation and from overloading the DnaK-DnaJ-GrpE and GroEL-GroES systems. When the stress conditions are relieved, the holding chaperones pass the unfolded proteins to DnaK-DnaJ or GroEL-GroES for active refolding. Holding chaperones include Hsp33, Hsp31, and small heat shock proteins, IbpA and IbpB.

Hsp33 was originally identified as a heat shock protein, but its chaperone function is redox regulated on the posttranscriptional level, which distinguishes it from other temperature-regulated cytosolic chaperones[231]. It has four conserved redox sensitive cysteines, which coordinate a zinc atom under normal growth conditions [232]. The reduced form of Hsp33 has no apparent chaperone activity; however, under oxidative stress conditions, such as treatment with hypochlorite, the cysteines in Hsp33 become oxidized, leading to intramolecular disulfide bond formation and release of zinc. This also causes partial unfolding, dimerization of the protein, and exposure of a high affinity substrate binding site for unfolded proteins (reviewed in Ref. [94]). It is noted that under severe oxidative stress conditions that act to decrease cellular ATP levels, the efficiency of DnaK-DnaJ-GrpE can be compromised. Therefore, under these circumstances, cells rely mainly on Hsp33 to manage oxidative protein misfolding[233]. Oxidized Hsp33 is reduced by the glutaredoxin and thioredoxin systems, which do not dissociate substrates from Hsp33 but rather prime Hsp33 for fast inactivation[234]. When conditions return to normal, interaction of Hsp33-substrate complexes with the functional DnaK-DnaJ-GrpE system, releases substrate proteins from Hsp33 and transfers the substrate to refolding either by the DnaK-DnaJ-GrpE system alone or this system acting together with GroEL-GroES[234].

Hsp31(YedU) forms homodimers and is thought to capture early unfolded intermediates under stress conditions[235]. Its mode of action relies on temperature-driven conformational changes to expose hydrophobic regions to capture substrates. Hsp31 does not require energy; however, binding of ATP inhibits its chaperone activity at a high temperature, probably by restricting the substrate binding site[236]. Hsp31 has also been shown to have peptidase activity[237] and contributes to acid resistance in stationary-phase *E. coli*[238].

IbpA and IbpB are small heat shock proteins (MW=16 kDa). They were identified by their ability to associate tightly with inclusion bodies formed either during heterologous protein production[239] or from endogeneous *E. coli* proteins that aggregated due to heat shock[240, 241]. Evidence shows that IbpA and IbpB are not only involved in heat defense, but might also be involved in resistance towards oxidative stress[242-245]. The binding of IbpA and IbpB to aggregated proteins protects the enzymatic activity of the aggregated proteins[246], slows down the proteolytic removal of aggregates[247], maintains proteins in a disaggregation-competent state[248, 249], and facilitates subsequent refolding of proteins by the ATP-dependent chaperones DnaK-DnaJ-GrpE[250].

Cytosolic chaperones involved in solubilizing protein aggregates

If both holding chaperones and folding chaperones fail to rescue the unfolded substrates from aggregation, *E. coli* possesses a third line of defense to combat the deleterious effects of protein aggregation. Through destructive (ClpAP, ClpXP, ClpCP[251, 252]) or non-destructive protein disaggregation pathways, aggregates are digested or solubilized. An enzyme called ClpB (an Hsp104 homolog), which functions to actively solubilize aggregates, is central to the non-destructive disaggregation pathway[18, 253]. ClpB belongs to the AAA+ (ATPases associated with various cellular activities) superfamily of ATPases and works in conjunction with the DnaK chaperone system. DnaK assists the disaggregation process during the initial steps probably by extracting polypeptides from aggregates. It does this by presenting the unstructured regions of those polypeptides to ClpB and modulating the ATPase cycles of ClpB[254]. DnaK might also assist in the refolding process of the solubilized polypeptides. ClpB unfolds substrates by forcibly threading polypeptide chains through its central cavity and it is highly selective for unstructured regions within proteins[255]. The energy for substrate threading and release are provided by coupling ATP hydrolysis to these processes. Interestingly, ClpB/DnaK seem to be preferred to facilitate the extraction of substrates from the small heat shock proteins[256], indicating a greater dis-aggregation role of DnaK than of GroEL. Due to the size limitation of its encapsulated folding cavity, GroEL may not be as powerful in disaggregation as the ClpB/DnaK system. In fact, interaction with GroEL might drive larger proteins towards the formation of inclusion bodies[257]. However, once

the aggregates are solubilized, the addition of GroEL-GroES increases the refolding yield[258].

Other chaperones involved in stress response include the Hsp90 homolog HtpG[259], which may assist *de novo* protein folding in slightly stressed *E. coli* cells[260].

Cytosolic chaperones involved in protein translocation

After achieving their native conformation, cytosolic proteins are retained in the cytoplasm. However, those proteins that are destined to have a periplasmic location, a cell envelope location, or to be exported to the extracellular space, need to be transported to their appropriate locations. Secreted proteins are made as precursors with N-terminal signal sequences. These sequences direct the precursors to the translocon and are cleaved off when the precursors reach their final destinations. The majority of secreted proteins are transported posttranslationally in an extended "export-competent" conformation by the Sec-dependent pathway [261]. A subset of proteins are exported via the signal recognition particle (SRP)-dependent pathway in a cotranslational manner. A third secretion pathway, Tat (twin-arginine dependent pathway), exclusively deals with folded proteins or partially folded proteins[262]. This pathway allows the secretion of proteins that assemble subunits or bind cofactors in the cytoplasm prior to their secretion[262].

Although not thoroughly studied, generalized cytoplasmic folding factors might interact with precursor proteins, helping to regulate their folding and in some cases, guiding the precursors from the ribosome to the translocon (reviewed in Ref. [263]). Indeed, the chaperonin GroEL has been shown to form stable complexes with purified precursors from the Sec dependent secretion pathway[264] and may also interact with the peripheral membrane protein SecA, a protein known as a central component of the translocation machinery[265]. Similarly, trigger factor functions in both the Sec- and SRP- dependent secretion pathways. Trigger factor functions by binding nascent polypeptides, holding them in an unfolded state for a relatively long time, allowing them to be efficiently engaged by the translocation-specific chaperone, SecB[266]. Trigger factor also binds the ribosome-nascent polypeptide complex simultaneously with SRP until the complex reaches the membrane-bound SRP receptor FtsY[267]. In the case of substrates destined for Sec-dependent transport, binding of trigger factor to ribosome-nascent polypeptides

prevents the binding of SRP, thus ensuring the specificity of SRP for binding rather hydrophobic signal sequences[266]. It is noteworthy that the effects of these chaperones in translocation *in vivo* are still ambiguous, and they may not be absolutely required for efficient translocation. For example, the involvement of trigger factor increases and stabilizes the precursor of leech carboxypeptidase inhibitor (pre-LCI) in the cytoplasm, and binding to trigger factor can lead to precursor accumulation and decreased efficiency for translocation across the inner membrane[268].

The Sec translocation-specific chaperone, SecB, has anti-folding properties and a broad substrate specificity, similar to other general holding chaperones. Importantly, it is capable of targeting its substrates to SecA for subsequent translocation through the SecYEG translocon[269, 270]. SecB is a tetramer consisting of a dimer of dimers[269, 270]. It is a highly acidic protein with two hydrophobic grooves on the tetramer surface that serve as the polypeptide binding sites. Substrate proteins are thought to wrap into the surface grooves of SecB in an extended conformation. In such a conformation they are protected from aggregation. The chaperone role of SecB is not restricted to the secretion pathway, since overexpression of SecB has been shown to be able to suppress the growth defect of strains lacking both TF and DnaK[271]. *In vitro*, SecB has been shown to disaggregate the insulin B-chain[272]. Another Sec-specific factor, SecA, performs multiple functions. In addition to its well established function of guiding the translocation process[273], SecA also possesses chaperone activity of promoting the folding of nonsecretory proteins[274]. Thus, SecA acts in the quality control to ensure that only unfolded precursors are translocated.

Cytosolic chaperone systems might also help in the maturation process of some proteins secreted via the Tat pathway. DnaK was found to be important for stabilizing such proteins[275] and together with another chaperone, SlyD, serve as general Tat-signal binding proteins[276]. DmsD, a Tat-specific chaperone, was shown to interact with general chaperones to escort its substrate through a cascade of chaperone-assisted folding events[277]. Additional studies will hopefully further clarify the influence of cytosolic chaperones on the folding process of extracytoplasmic proteins.

Periplasmic folding modulators

The periplasmic environment is very different from the cytoplasmic environment. Due to the permeability of the outer membrane, the properties of the periplasm are largely dependent on the surrounding medium. In contrast to the reducing, highly buffered, and energy-rich environment in the cytoplasm, the periplasm is an oxidizing, relatively unbuffered compartment that lacks any obvious energy source, such as ATP that is required for many cytoplasmic folding chaperones to work. Therefore, the periplasmic folding modulators must use different mechanisms to couple substrate binding and release.

Specialized chaperones

SurA and LolA- chaperones for outer membrane proteins or lipoproteins

SurA has been suggested to have a role in the biogenesis of outer membrane proteins. The gene *surA* was originally identified as essential for survival in stationary phase[48]. Strains devoid of *surA* are slightly mucoid in rich medium and are sensitive to bacitracin, vancomycin, and bile salts, suggesting that SurA plays a role in maintaining the integrity of the outer membrane[26]. Mutants null with respect to *surA* also have reduced levels of properly folded outer membrane proteins, such as LamB[278], OmpA, OmpF[26], and OmpC. SurA also possesses PPIase activity but it's chaperone activity is independent of its PPIase activity, since a mutant of *surA* lacking its two PPIase domains almost completely complements *surA in vivo* and this SurA variant has undiminished chaperone-like activity *in vitro*[279]. SurA is a dumbbell shaped molecule with an extension in the core structure module, which might be a potential substrate binding site[280]. Surprisingly, the crystal structure of the N-terminal chaperone domain of SurA closely resembles the C-termimal substrate binding domain of trigger factor, despite the lack of sequence homology between SurA and trigger factor[219].

Selection for short peptides that bind SurA using the phage display technique suggested that SurA binds preferentially to an aromatic-polar-aromatic-nonpolar-proline (Ar-polar-Ar-nonpolar-Pro) motif. The first three residues of this motif occurs quite frequently in outer membrane proteins[281]. Other techniques less stringent for detecting

binding affinity confirmed the preference of SurA to bind aromatic residues[282]. All of the above evidence leads to the role of SurA in outer membrane protein biogenesis. However, there is currently no clear evidence for the involvement of SurA in periplasmic protein folding[26].

LolA is a specialized chaperone involved in outer membrane lipoprotein biogenesis (reviewed in Ref. [283]). It was originally identified as a protein essential for releasing the major outer membrane lipoprotein (Lpp) from the inner membrane[284]. LolA was later shown to facilitate the release of other outer membrane lipoproteins, such as Pal, NlpB, Slp, and RlpA[285]. The LolA-dependent release was thought to be crucial for the efficient and specific incorporation of lipoproteins into the outer membrane[285]. The released lipoprotein forms a complex with LolA and is targeted to the outer membrane lipoprotein receptor LolB[286]. The lipoprotein is then released from LolA and transferred to LolB[287]. Lipoproteins are highly hydrophobic due to their N-terminal lipids; therefore, LolA and LolB are thought to shield the hydrophobic regions of lipoproteins during their transportation[283].

PapD, FimC-chaperones involved in pili biosynthesis

Pili are adhesive organelles occuring on the bacterial surface. Like other secreted proteins, the translocation of pili subunits across the inner membrane, through the periplasm, and their secretion to the outside of the cell, followed by assembly on cell surface, occur with the help of chaperone proteins (reviewed in Refs.[288-291]). The PapD-like superfamily of periplasmic chaperones is specialized for pili biogenesis. There are more than 30 papD-like chaperones that facilitate the assembly of pili and non-pili organelles (reviewed in Ref.[292]). PapD itself and FimC are the two most studied members of the PapD superfamily. PapD assembles P pili[293] and FimC assemble the Type 1 pili[294], respectively. These chaperones interact with newly translocated pili subunits in the periplasm and facilitate their release from the inner membrane[295, 296]. Binding to these chaperones stabilizes the pili subunits, caps their interactive surfaces, and prevents the aggregation of these subunits in the periplasm[296, 297]. In the absence of PapD-related chaperones, pili subunits are subject to aggregation in the periplasm. This aggregation is sensed by the Cpx and σ^E stress response pathways, leading to an increased level of the

degP protease, which facilitates the removal of pili aggregates. The activation of these stress response pathways also induces DsbA, an enzyme responsible for the catalysis of disulfide bond formation within pili subunits[295, 298, 299].

Following secretion and binding in the periplasm, the chaperone-pili subunit complexes are targeted to the outer membrane usher[300, 301], which forms a donut-shaped channel big enough for the pili subunits to pass through[302]. Binding of the chaperone-subunit complexes to usher drives the release of pili subunits from the chaperone, and stabilizes usher in an assembly-competent conformation, which in turn, allows the initiation of pili assembly[303]. In contrast to many other chaperone systems, PapD-like chaperones work as monomers. They act to allow the folding of pili subunits to occur during pili assembly[296]. Folded subunits remain bound to the chaperones until they reach the outer membrane usher, where upon they are released[296]. The PapD chaperone donates a β strand to the pili subunits. This enables the pili subunits to fold[304] in a process termed donor strand complementation[305]. When the pili subunits polymerize, the β strand, which was donated by the PapD chaperone to one pili subunit, is displaced by the N-terminal extension of another pili subunit in a process termed donor strand exchange[304]. As a result, PapD is absent from the final assembly of pili subunits. PapD and FimC are thus highly specialized chaperones that provide steric information for pili subunit folding.

Generic chaperones: FkpA, Skp and HdeA

A global screening for genes that suppress one of the extracytoplasmic stress responses, the σ^E response, identified *fkpA* and two other chaperone genes, *surA* and *skp*[306]. Overexpression of FkpA prevents the formation of inclusion bodies by a defective folding variant of the *E . coli* maltose-binding protein, MalE31[36]. The overproduction of FkpA alone or as a fusion partner also increases the yield and/or solubility of heterologous proteins, such as antibody single-chain Fv (scFv) fragments[37], recombinant penicillin acylase (PAC)[307], and retroviral envelope proteins[308]. The chaperone-like activity of FkpA is hypothesized to be due to its interaction with early folding intermediates that prevents their aggregation. Additionally, the interaction of FkpA with partially unfolded protein species may drive a productive protein refolding pathway. Similar to SurA, the chaperone-like activity of FkpA is independent of its PPIase

activity[36]. Its crystal structure reveals a V-shaped homodimer. The N-terminal domain of FkpA is responsible for its chaperone activity and dimerization, and the C-terminal domain is responsible for its PPIase activity[309]. Consistent with the substrate diversity of FkpA, the whole molecule exhibits structural flexibility[310]. FkpA protects periplasmic proteins or recombinant proteins expressed in the periplasm; there is no evidence for its participation in outer membrane biogenesis.

Skp was initially suggested to assist in outer membrane protein folding because of its affinity for outer membrane proteins such as OmpF[311], OmpA[311, 312], OmpC, and LamB[311]. It is thought that Skp assists outer membrane proteins (OMPs) both to fold and to insert them into the outer membrane. Skp may function cooperatively with lipopolysaccharide (LPS), a major component of the outer membrane of Gram-negative bacteria, to maintain the unfolded OMPs in a state ready for membrane insertion[313] . Strains depleted of *skp* are viable[311]. However, these strains show a decreased steady state level of outer membrane proteins[311], and have an altered outer membrane structure and physiology[312]. Proteomics and *in vivo* expression studies suggested that Skp has a broad substrate specificity that includes both outer membrane proteins and native periplasmic proteins[32], as well as recombinant proteins expressed in the periplasm[33-35], supporting its role as a general chaperone in the periplasm.

The 17 kDa protein Skp forms a homo-trimer in solution, and the structure resembles a jellyfish[314]. The main body of this chaperone consists of a tightly packed β-barrel, which is surrounded by the C-terminal helices of the three subunits[315, 316]. The jellyfish-shaped cavity formed by these tentacles can almost accommodate a folded-like structure of proteins of ~ 25 kDa[315]. However, if its unfolded substrates exist as molten globules, which are more loosely packed, the maximal capacity of Skp would be smaller[315]. A working model has been proposed[315] where one Skp trimer captures a single OMP[317] as soon as it translocates into the periplasm. Skp sequesters the hydrophobic regions of OMP inside of its cavity. This seems to not only prevent aggregation of OMPs or premature folding of these regions, but also allow independent folding of the other domains outside of the cavity[318]. Skp then binds lipopolysaccharide (LPS) to facilitate the delivery and insertion of OMP into the outer membrane. Consistent with this model, Skp was shown to exhibit two different states *in vivo*, one of which is associated with the

outer membrane components[319]. These two different states might represent different stages in the association/dissociation cycle of Skp from the outer membrane.

In contrast to FkpA and Skp, which fulfill chaperone functions under normal physiological conditions, HdeA mediates acid resistance in pathogenic enteric bacteria[38, 39]. The gastrointestinal tract of mammals has a pH of 1 to 3. In this pH range proteins generally denature[320, 321]. The *hdeA* gene is among those induced under acidic conditions[322]. HdeA is proposed to use a novel strategy to modulate its chaperone function. It exists in an ordered conformation as a dimer under normal growth conditions and is inactive in this conformation and does not bind proteins. Acidic pH induces a rapid conformation change in HdeA, allowing it to adopt a disordered monomer form as a chaperone that is able to bind substrates[188, 323]. The hydrophobic regions of HdeA that are exposed under acidic conditions play a major role in substrate protection. The exposure of its highly charged terminal segment allows the HdeA-substrate complex to be soluble[324]. HdeA is one of the smallest chaperones known (MW~ 9.7kDa) yet its active form is able to adopt different conformations that allow the binding of a broad range of substrates[188]. The *hdeB* gene is in the same operon as *hdeA*, and has been reported to have a similar phenotype as HdeA does, which confers protection at acidic pH[325]. However, our lab has been unable to show any chaperone activity for HdeB using the protein substrates that HdeA protects from aggregation[******]. During the renaturation process, substrates are slowly released from HdeA in such a way that the concentration of aggregation-prone intermediates is kept below the threshold where they could aggregate; therefore, the substrates are devoid of aggregation, and could be effectively refolded[326].

Compared to the essential role of generic cytoplasmic chaperones such as GroEL, the effects of periplasmic folding modulators seem to be less important, since deletion mutations in the genes for periplasmic chaperones generally cause no or little phenotypic response. However, certain combinations of two or more knockout mutations in periplasmic chaperones (e.g., *skp* and *degP*[312], *surA* and *ppiD*[327]) cause severe phenotypes, indicating that they may play very important though overlapping roles.

[******] Tim Tapley's results.

Peptidyl prolyl isomerase

Peptidyl prolyl cis/trans isomerase (PPIases) catalyzes the cis-trans isomerization of peptide bonds with proline (Xaa-pro bond; Xaa, any amino acid before proline). This isomerization step is a rate limiting step in the folding process of many proteins[41]. SurA, PpiD, FkpA and PpiA are the known periplasmic PPIases.

SurA and FkpA, as previously discussed, are two PPIases that also have chaperone activity. Interestingly, studies show that the prolyl isomerase domains alone have a relatively narrow substrate specificity. However, the presence of the chaperone domains enables the efficient transfer of protein substrates to the prolyl isomerase site. This confers these enzymes with a broad substrate specificity[328]. The other two PPIases, PpiD and PpiA, have not yet been demonstrated to have chaperone activity.

The gene *ppiD* was isolated as a multicopy suppressor of *surA* in a genetic screen[329]. Deletion of the gene results in reduced levels of OMP, hypersensitivity to hydrophobic antibiotics and detergents, and the induction of the σ^E response[329]. PpiD is anchored to the inner membrane with a catalytic domain facing the periplasm[329]. Similar to Skp, PpiD has been demonstrated to influence the release of newly synthesized outer membrane proteins from the cytoplasmic membrane into the periplasm[330]. Since PpiD is anchored in the inner membrane near the SecYEG translocon, it was proposed that PpiD interacts with substrate polypeptides as soon as they emerge from the translocon. Skp could then bind to substrates following their release into the periplasm[330]. Though the substrate specificities of PpiD and SurA are thought to overlap, SurA appears to be more specific than ppiD[331].

PpiA is not particularly well characterized as an enzyme. Strains that are *ppiA*(*rotA*) null are viable, and show no detectable change in the steady-state expression level of periplasmic or outer membrane proteins[332]. Coexpression of PpiA showed no effect on recombinant protein expression[37] and no influence on the kinetics of periplasmic protein folding[332]. The exact function of PpiA remains to be established.

Proteases

Prematurely terminated polypeptide, protein folding intermediates that are kinetically trapped off-pathway, and partially folded proteins that failed to be rescued by folding modulators are generally degraded. This prevents the accumulation of protein aggregates and allows the recycling of amino acids.

Lon[333], ClpYQ[334], ClpAP, ClpXP and membrane associated FtsH[335] are ATP-dependent proteases that mediate protein degradation in the cytoplasm. Aberrantly folded substrates are firstly recognized and captured by the ATPase domains in the proteases. These substrates are then unfolded and transferred to the protease domains of the proteases (reviewed in Ref. [336]). Protein substrates are cleaved into peptides, which are further degraded by peptidases.

In the periplasm, misfolded proteins are degraded by more than 20 proteases/peptidases. The list of periplasmic proteases/peptidases is available at http://www.cf.ac.uk/biosi/staffinfo/ehrmann/tools/proteases/allproteases.html. Due to the lack of energy source in the periplasm, periplasmic proteases must act differently from ATP-dependent cytoplasmic proteases.

The DegP (HtrA, Do) protease plays a leading role in removing the misfolded proteins in the periplasm (reviewed in Ref.[42, 43]). The *degP* gene was initially identified through mutational analysis of strains that were unable to grow at elevated temperature[337] or failed to degrade abnormal periplasmic proteins[338]. DegP mediates the degradation of a number of misfolded periplasmic or cell envelope proteins, such as MBP[44], PhoA[45] , MalS[25] and OmpF[46], and the cytoplasmic protein TreF that is mislocated to the periplasm[339], as well as recombinant proteins expressed in the periplasm[340]. However, DegP does not cleave well folded proteins, for example, native MalS[25]. It seems that DegP has a broad substrate specificity towards misfolded proteins. *In vitro*, DegP recognizes heat-denatured proteins in general rather than a particular set of proteins[341]. DegP cleaves misfolded protein substrates at discrete Val/Xaa or Ile/Xaa sites (Xaa meaning any amino acid), which are normally buried in the hydrophobic core of native proteins[342]. DegP is a serine protease and its activity can be inhibited by

diisopropylfluorophosphate (DFP), but not by phenylmethylsulfonyl fluoride (PMSF) that inhibits most serine proteases[343].

The mature DegP monomer contains an N-terminal protease domain and two C-terminal PDZ domains. PDZ domains are structural modules that mediate protein-protein interactions. DegP's PDZ1 domain is important for its substrate recognition and protease activity, while PDZ2 domain is involved in maintaining the hexameric conformation[344]. The crystal structure of DegP suggests it to be a hexameric "cage" composed of two trimeric rings. Each trimer of DegP forms a funnel-like structure in such a way that the three protease domains from each subunit form the top of the funnel and the six PDZ domains extend away. The hexamer of DegP is formed by the staggered association of two trimers. In the hexameric cage of DegP, the six protease domains constitute the top and bottom, whereas the 12 PDZ domains form the mobile side wall of the cage. The movement of the PDZ domains controls the accessibility of the proteolytic sites, which are located inside of the cage[345].

Interestingly, DegP has not only protease activity, but also chaperone activity[25]. Therefore, understanding the underlying mechanism of the functional switch of DegP is important. In particular, how does DegP distinguish a substrate that can be refolded from ones that are severely damaged and have to be degraded? It is demonstrated that temperature is an essential factor to govern the functional switch of DegP[25]. At low temperatures (28°C and below), DegP shows more chaperone activity than protease activity. At higher temperatures (37°C and above), DegP predominantly shows protease activity. Studies on a mutant form of DegP ($DegP_{S210A}$) that is devoid of protease activity showed that this mutant is active as a chaperone at any temperature, suggesting that the functional switch of being a protease or a chaperone could be mediated by minor changes in the structure of DegP which might be induced by the changing temperatures.

Other periplasmic proteases that are potentially involved in protein quality control are Prc, Pet, and OmpT. Their absence was shown to result in reduced proteolysis of recombinant proteins produced in the periplasm[346].

References

1. Quan, S., Schneider, I., Pan, J., Hacht, A.V. & Bardwell, J.C. The CXXC motif is more than a redox rheostat. *J. Biol. Chem.* (2007).
2. Anfinsen, C.B. Principles That Govern Folding of Protein Chains. *Science* **181**, 223-230 (1973).
3. Jaenicke, R. Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117-237 (1987).
4. Crooke, E., Brundage, L., Rice, M. & Wickner, W. ProOmpA spontaneously folds in a membrane assembly competent state which trigger factor stabilizes. *EMBO J.* **7**, 1831-1835 (1988).
5. Baldwin, R.L. How Does Protein Folding Get Started. *Trends Biochem. Sci.* **14**, 291-294 (1989).
6. Ellis, R.J. & Minton, A.P. Cell biology: join the crowd. *Nature* **425**, 27-28 (2003).
7. Mullineaux, C.W., Nenninger, A., Ray, N. & Robinson, C. Diffusion of green fluorescent protein in three cell environments in Escherichia coli. *J. Bacteriol.* **188**, 3442-3448 (2006).
8. Jackson, S.E. How do small single-domain proteins fold? *Fold. Des.* **3**, R81-R91 (1998).
9. Walter, S. & Buchner, J. Molecular chaperones--cellular machines for protein folding. *Angew Chem Int Ed Engl* **41**, 1098-1113 (2002).
10. Hartl, F.U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571-579 (1996).
11. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**, 574-581 (2009).
12. Brockwell, D.J. & Radford, S.E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17**, 30-37 (2007).
13. Ellis, R.J. & Minton, A.P. Protein aggregation in crowded environments. *Biol Chem* **387**, 485-497 (2006).
14. Baneyx, F. Keeping up with protein folding. *Microb Cell Fact* **3**, 6 (2004).
15. Thomas, J.G., Ayling, A. & Baneyx, F. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli - To fold or to refold. *Appl Biochem Biotech* **66**, 197-238 (1997).
16. Beissinger, M. & Buchner, J. How chaperones fold proteins. *Biological Chemistry* **379**, 245-259 (1998).
17. Saibil, H.R. Chaperone machines in action. *Curr Opin Struc Biol* **18**, 35-42 (2008).
18. Doyle, S.M. & Wickner, S. Hsp104 and ClpB: protein disaggregating machines. *Trends Biochem Sci* **34**, 40-48 (2009).
19. Pauwels, K., Van Molle, I., Tommassen, J. & Van Gelder, P. Chaperoning Anfinsen: the steric foldases. *Mol Microbiol* **64**, 917-922 (2007).

20. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.R., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C.H., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, M., Garner, E.C. & Obradovic, Z. Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59 (2001).
21. Serdyuk, I.N. Structured proteins and proteins with intrinsic disorder. *Mol Biol* **41**, 262-277 (2007).
22. Chen, X., Solomon, W.C., Kang, Y., Cerda-Maira, F., Darwin, K.H. & Walters, K.J. Prokaryotic Ubiquitin-Like Protein Pup Is Intrinsically Disordered. *J Mol Biol* **392**, 208-217 (2009).
23. Wang, C.C. & Tsou, C.L. Enzymes as chaperones and chaperones as enzymes. *Febs Lett* **425**, 382-384 (1998).
24. Chen, J., Song, J.L., Zhang, S., Wang, Y., Cui, D.F. & Wang, C.C. Chaperone activity of DsbC. *J Biol Chem* **274**, 19601-19605 (1999).
25. Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339-347 (1999).
26. Lazar, S.W. & Kolter, R. SurA assists the folding of Escherichia coli outer membrane proteins. *Journal of Bacteriology* **178**, 1770-1773 (1996).
27. J.-M. Chuysen, R.Hakenbeck. New comprehensive biochemistry: Bacterial Cell Wall, Vol. 27. (ed. R.Hakenbeck. J.-M. Chuysen) (Elsevier Science B.V., 1994).
28. Winkler, H.H. Rickettsial cell water and membrane permeability determined by a micro space technique. *Appl Environ Microbiol* **31**, 146-149 (1976).
29. Graham, L.L., Beveridge, T.J. & Nanninga, N. Periplasmic space and the concept of the periplasm. *Trends Biochem Sci* **16**, 328-329 (1991).
30. Nakae, T. & Nikaido, H. Outer Membrane as a Diffusion Barrier in Salmonella-Typhimurium - Penetration of Oligosaccharides and Polysaccharides into Isolated Outer Membrane-Vesicles and Cells with Degraded Peptidoglycan Layer .4. *J Biol Chem* **250**, 7359-7365 (1975).
31. Silhavy, T.J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**, a000414.
32. Jarchow, S., Luck, C., Gorg, A. & Skerra, A. Identification of potential substrate proteins for the periplasmic Escherichia coli chaperone Skp. *Proteomics* **8**, 4987-4994 (2008).
33. Bothmann, H. & Pluckthun, A. Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat Biotechnol* **16**, 376-380 (1998).
34. Chatterjee, D.K. & Esposito, D. Enhanced soluble protein expression using two new fusion tags. *Protein Expres Purif* **46**, 122-129 (2006).
35. Ha, S.C., Pereira, J.H., Jeong, J.H., Huh, J.H. & Kim, S.H. Purification of human transcription factors Nanog and Sox2, each in complex with Skp, an Escherichia coli periplasmic chaperone. *Protein Expres Purif* **67**, 164-168 (2009).
36. Arie, J.P., Sassoon, N. & Betton, J.M. Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of Escherichia coli. *Mol Microbiol* **39**, 199-210 (2001).
37. Ramm, K. & Pluckthun, A. The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA - II. Isomerase-independent chaperone activity in vitro. *J Biol Chem* **275**, 17106-17113 (2000).

38. Zhao, B.Y. & Houry, W.A. Acid stress response in enteropathogenic gammaproteobacteria: an aptitude for survival. *Biochem Cell Biol* **88**, 301-314 (2010).

39. Gajiwala, K.S. & Burley, S.K. HdeA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *J Mol Biol* **295**, 605-612 (2000).

40. Tapley, T.L., Franzmann, T.M., Chakraborty, S., Jakob, U. & Bardwell, J.C.A. Protein refolding by pH-triggered chaperone binding and release. *Proc Natl Acad Sci U S A.* **107**, 1071-1076 (2010).

41. Schmid, F.X. Prolyl Isomerase - Enzymatic Catalysis of Slow Protein-Folding Reactions. *Annu Rev Bioph Biom* **22**, 123-143 (1993).

42. Clausen, T., Southan, C. & Ehrmann, M. The HtrA family of proteases: Implications for protein composition and cell fate. *Mol Cell* **10**, 443-455 (2002).

43. Meltzer, M., Hasenbein, S., Mamant, N., Merdanovic, M., Poepsel, S., Hauske, P., Kaiser, M., Huber, R., Krojer, T., Clausen, T. & Ehrmann, M. Structure, function and regulation of the conserved serine proteases DegP and DegS of Escherichia coli. *Res Microbiol* **160**, 660-666 (2009).

44. Betton, J.M., Sassoon, N., Hofnung, M. & Laurent, M. Degradation versus aggregation of misfolded maltose-binding protein in the periplasm of Escherichia coli. *J Biol Chem* **273**, 8897-8902 (1998).

45. Sone, M., Kishigami, S., Yoshihisa, T. & Ito, K. Roles of disulfide bonds in bacterial alkaline phosphatase. *J Biol Chem* **272**, 6174-6178 (1997).

46. Misra, R., Castillokeller, M. & Deng, M. Overexpression of protease-deficient DegP (S210A) rescues the lethal phenotype of Escherichia coli OmpF assembly mutants in a degP background. *J Bacteriol* **182**, 4882-4888 (2000).

47. Lindberg, F., Tennent, J.M., Hultgren, S.J., Lund, B. & Normark, S. PapD, a periplasmic transport protein in P-pilus biogenesis. *J. Bacteriol.* **171**, 6052-6058 (1989).

48. Tormo, A., Almiron, M. & Kolter, R. surA, an Escherichia coli gene essential for survival in stationary phase. *J. Bacteriol.* **172**, 4339-4347 (1990).

49. Georgopoulos, C. & Welch, W.J. Role of the Major Heat-Shock Proteins as Molecular Chaperones. *Annu. Rev. Cell Biol.* **9**, 601-634 (1993).

50. Chen, R. & Henning, U. A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. *Mol. Microbiol.* **19**, 1287-1294 (1996).

51. Matsuyama, S., Tajima, T. & Tokuda, H. A novel periplasmic carrier protein involved in the sorting and transport of Escherichia coli lipoproteins destined for the outer membrane. *EMBO J.* **14**, 3365-3372 (1995).

52. Andersen, C.L., Matthey-Dupraz, A., Missiakas, D. & Raina, S. A new Escherichia coli gene, dsbG, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins. *Mol. Microbiol.* **26**, 121-132 (1997).

53. Horne, S.M. & Young, K.D. Escherichia coli and other species of the Enterobacteriaceae encode a protein similar to the family of Mip-like FK506-binding proteins. *Arch. Microbiol.* **163**, 357-365 (1995).

54. MacRitchie, D.M., Buelow, D.R., Price, N.L. & Raivio, T.L. Two-component signaling and gram negative envelope stress response systems. *Adv Exp Med Biol* **631**, 80-110 (2008).

55. Chang, C. & Stewart, R.C. The two-component system - Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol* **117**, 723-731 (1998).

56. Miot, M. & Betton, J.M. Protein quality control in the bacterial periplasm. *Microb Cell Fact* **3**, 4 (2004).

57. Bury-Mone, S., Nomane, Y., Reymond, N., Barbet, R., Jacquet, E., Imbeaud, S., Jacq, A. & Bouloc, P. Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet.* **5**, e1000651 (2009).

58. Laubacher, M.E. & Ades, S.E. The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. *J. Bacteriol.* **190**, 2065-2074 (2008).

59. Ades, S.E. Regulation by destruction: design of the sigmaE envelope stress response. *Curr Opin Microbiol* **11**, 535-540 (2008).

60. Darwin, A.J. The phage-shock-protein response. *Mol Microbiol* **57**, 621-628 (2005).

61. Dutton, R.J., Boyd, D., Berkmen, M. & Beckwith, J. Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11933-11938 (2008).

62. Mainathambika, B.S. & Bardwell, J.C. Disulfide-Linked Protein Folding Pathways. *Annu Rev Cell Dev Bi* **24**, 211-235 (2008).

63. Kadokura, H. & Beckwith, J. Mechanisms of Oxidative Protein Folding in the Bacterial Cell Envelope. *Antioxid Redox Signal* (2010).

64. Atkinson, H.J. & Babbitt, P.C. An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. *PLoS Comput. Biol.* **5**, e1000541 (2009).

65. Pan, J.L. & Bardwell, J.C. The origami of thioredoxin-like folds. *Protein Sci* **15**, 2217-2227 (2006).

66. Heras, B., Shouldice, S.R., Totsika, M., Scanlon, M.J., Schembri, M.A. & Martin, J.L. DSB proteins and bacterial pathogenicity. *Nat Rev Microbiol* **7**, 215-225 (2009).

67. Holmgren, A., Soderberg, B.O., Eklund, H. & Branden, C.I. Three-dimensional structure of Escherichia coli thioredoxin-S2 to 2.8 A resolution. *Proc Natl Acad Sci U S A* **72**, 2305-2309 (1975).

68. Martin, J.L. Thioredoxin--a fold for all reasons. *Structure* **3**, 245-250 (1995).

69. Aslund, F. & Beckwith, J. The thioredoxin superfamily: redundancy, specificity, and gray-area genomics. *J Bacteriol* **181**, 1375-1379 (1999).

70. Jordan, A., Aslund, F., Pontis, E., Reichard, P. & Holmgren, A. Characterization of Escherichia coli NrdH. A glutaredoxin-like protein with a thioredoxin-like activity profile. *J Biol Chem* **272**, 18044-18050 (1997).

71. Huber-Wunderlich, M. & Glockshuber, R. A single dipeptide sequence modulates the redox properties of a whole enzyme family. *Folding & design* **3**, 161-171 (1998).

72. Mossner, E., Huber-Wunderlich, M. & Glockshuber, R. Characterization of Escherichia coli thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. *Protein Sci* **7**, 1233-1244 (1998).

73. Krause, G., Lundstrom, J., Barea, J.L., Pueyo de la Cuesta, C. & Holmgren, A. Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in Escherichia coli thioredoxin. *The Journal of biological chemistry* **266**, 9494-9500 (1991).

74. Lundstrom, J., Krause, G. & Holmgren, A. A Pro to His mutation in active site of thioredoxin increases its disulfide-isomerase activity 10-fold. New refolding systems for

reduced or randomly oxidized ribonuclease. *The Journal of biological chemistry* **267**, 9047-9052 (1992).

75. Chivers, P.T., Laboissiere, M.C. & Raines, R.T. The CXXC motif: imperatives for the formation of native disulfide bonds in the cell. *The EMBO journal* **15**, 2659-2667 (1996).
76. Wouters, M.A., Fan, S.W. & Haworth, N.L. Disulfides as Redox Switches: From Molecular Mechanisms to Functional Significance. *Antioxid Redox Sign* **12**, 53-91 (2010).
77. Guddat, L.W., Bardwell, J.C., Glockshuber, R., Huber-Wunderlich, M., Zander, T. & Martin, J.L. Structural analysis of three His32 mutants of DsbA: support for an electrostatic role of His32 in DsbA stability. *Protein Sci* **6**, 1893-1900 (1997).
78. Schultz, L.W., Chivers, P.T. & Raines, R.T. The CXXC motif: crystal structure of an active-site variant of Escherichia coli thioredoxin. *Acta crystallographica* **55**, 1533-1538 (1999).
79. Grimshaw, J.P.A., Stirnimann, C.U., Brozzo, M.S., Malojcic, G., Gruetter, M.G., Capitani, G. & Glockshuber, R. DsbL and DsbI form a specific dithiol oxidase system for periplasmic arylsulfate sulfotransferase in uropathogenic Escherichia coli. *J Mol Biol* **380**, 667-680 (2008).
80. Nelson, J.W. & Creighton, T.E. Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation in vivo. *Biochemistry (Mosc.)* **33**, 5974-5983 (1994).
81. Grauschopf, U., Winther, J.R., Korber, P., Zander, T., Dallinger, P. & Bardwell, J.C. Why is DsbA such an oxidizing disulfide catalyst? *Cell* **83**, 947-955 (1995).
82. Fernandes, A.P., Fladvad, M., Berndt, C., Andresen, C., Lillig, C.H., Neubauer, P., Sunnerhagen, M., Holmgren, A. & Vlamis-Gardikas, A. A novel monothiol glutaredoxin (Grx4) from Escherichia coli can serve as a substrate for thioredoxin reductase. *J Biol Chem* **280**, 24544-24552 (2005).
83. Bardwell, J.C., McGovern, K. & Beckwith, J. Identification of a protein required for disulfide bond formation in vivo. *Cell* **67**, 581-589 (1991).
84. Zapun, A., Bardwell, J.C. & Creighton, T.E. The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. *Biochemistry* **32**, 5083-5092 (1993).
85. Zheng, W.D., Quan, H., Song, J.L., Yang, S.L. & Wang, C.C. Does DsbA have chaperone-like activity? *Arch Biochem Biophys* **337**, 326-331 (1997).
86. Jander, G., Martin, N.L. & Beckwith, J. 2 Cysteines in Each Periplasmic Domain of the Membrane-Protein Dsbb Are Required for Its Function in Protein Disulfide Bond Formation. *EMBO J.* **13**, 5121-5127 (1994).
87. Inaba, K., Murakami, S., Suzuki, M., Nakagawa, A., Yamashita, E., Okada, K. & Ito, K. Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell* **127**, 789-801 (2006).
88. Inaba, K., Takahashi, Y.H., Fujieda, N., Kano, K., Miyoshi, H. & Ito, K. DsbB elicits a red-shift of bound ubiquinone during the catalysis of DsbA oxidation. *The Journal of biological chemistry* **279**, 6761-6768 (2004).
89. Tapley, T.L., Eichner, T., Gleiter, S., Ballou, D.P. & Bardwell, J.C. Kinetic Characterization of the Disulfide Bond-forming Enzyme DsbB. *The Journal of biological chemistry* **282**, 10263-10271 (2007).

90. Arredondo, S., Segatori, L., Gilbert, H.F. & Georgiou, G. De novo design and evolution of artificial disulfide isomerase enzymes analogous to the bacterial DsbC. *J Biol Chem* **283**, 31469-31476 (2008).

91. Bessette, P.H., Cotto, J.J., Gilbert, H.F. & Georgiou, G. In vivo and in vitro function of the Escherichia coli periplasmic cysteine oxidoreductase DsbG. *J Biol Chem* **274**, 7784-7792 (1999).

92. Hiniker, A., Ren, G.P., Heras, B., Zheng, Y., Laurinec, S., Jobson, R.W., Stuckey, J.A., Martin, J.L. & Bardwell, J.C.A. Laboratory evolution of one disulfide isomerase to resemble another. *P Natl Acad Sci USA* **104**, 11670-11675 (2007).

93. Depuydt, M., Leonard, S.E., Vertommen, D., Denoncin, K., Morsomme, P., Wahni, K., Messens, J., Carroll, K.S. & Collet, J.F. A periplasmic reducing system protects single cysteine residues from oxidation. *Science* **326**, 1109-1111 (2009).

94. Kumsta, C. & Jakob, U. Redox-Regulated Chaperones. *Biochemistry-Us* **48**, 4666-4676 (2009).

95. Meyer, Y., Buchanan, B.B., Vignols, F. & Reichheld, J.P. Thioredoxins and Glutaredoxins: Unifying Elements in Redox Biology. *Annu. Rev. Genet.* **43**, 335-367 (2009).

96. Vlamis-Gardikas, A. The multiple functions of the thiol-based electron flow pathways of Escherichia coli: Eternal concepts revisited. *Biochim. Biophys. Acta* **1780**, 1170-1200 (2008).

97. Fernandes, A.P. & Holmgren, A. Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antioxid. Redox. Signal.* **6**, 63-74 (2004).

98. Michelet, L., Zaffagnini, M., Massot, V., Keryer, E., Vanacker, H., Miginiac-Maslow, M., Issakidis-Bourguet, E. & Lemaire, S.D. Thioredoxins, glutaredoxins, and glutathionylation: new crosstalks to explore. *Photosynth Res* **89**, 225-245 (2006).

99. Jonda, S., Huber-Wunderlich, M., Glockshuber, R. & Mossner, E. Complementation of DsbA deficiency with secreted thioredoxin variants reveals the crucial role of an efficient dithiol oxidant for catalyzed protein folding in the bacterial periplasm. *The EMBO journal* **18**, 3271-3281 (1999).

100. Masip, L., Pan, J.L., Haldar, S., Penner-Hahn, J.E., DeLisa, M.P., Georgiou, G., Bardwell, J.C. & Collet, J.F. An engineered pathway for the formation of protein disulfide bonds. *Science* **303**, 1185-1189 (2004).

101. Bader, M.W., Hiniker, A., Regeimbal, J., Goldstone, D., Haebel, P.W., Riemer, J., Metcalf, P. & Bardwell, J.C.A. Turning a disulfide isomerase into an oxidase: DsbC mutants that imitate DsbA. *Embo J* **20**, 1555-1562 (2001).

102. Foit, L., Morgan, G.J., Kern, M.J., Steimer, L.R., von Hacht, A.A., Titchmarsh, J., Warriner, S.L., Radford, S.E. & Bardwell, J.C. Optimizing protein stability in vivo. *Mol Cell.* **36**, 861-871 (2009).

103. Sieber, V., Pluckthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-960 (1998).

104. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691-695 (1999).

105. Cabantous, S., Rogers, Y., Terwilliger, T.C. & Waldo, G.S. New Molecular Reporters for Rapid Protein Folding Assays. *PLoS ONE* **3**, - (2008).

106. Kadokura, H., Katzen, F. & Beckwith, J. Protein disulfide bond formation in prokaryotes. *Annual review of biochemistry* **72**, 111-135 (2003).

107. Carvalho, A.P., Fernandes, P.A. & Ramos, M.J. Similarities and differences in the thioredoxin superfamily. *Progress in biophysics and molecular biology* **91**, 229-248 (2006).

108. Chivers, P.T., Prehoda, K.E. & Raines, R.T. The CXXC motif: a rheostat in the active site. *Biochemistry* **36**, 4061-4066 (1997).

109. Bardwell, J.C., Lee, J.O., Jander, G., Martin, N., Belin, D. & Beckwith, J. A pathway for disulfide bond formation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 1038-1042 (1993).

110. Gane, P.J., Freedman, R.B. & Warwicker, J. A molecular model for the redox potential difference between thioredoxin and DsbA, based on electrostatics calculations. *Journal of molecular biology* **249**, 376-387 (1995).

111. Bessette, P.H., Qiu, J., Bardwell, J.C., Swartz, J.R. & Georgiou, G. Effect of sequences of the active-site dipeptides of DsbA and DsbC on in vivo folding of multidisulfide proteins in Escherichia coli. *Journal of bacteriology* **183**, 980-988 (2001).

112. Eklund, H., Gleason, F.K. & Holmgren, A. Structural and functional relations among thioredoxins of different species. *Proteins* **11**, 13-28 (1991).

113. Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E. & Stanhope, M.J. Universal trees based on large combined protein sequence data sets. *Nature genetics* **28**, 281-285 (2001).

114. Martin, J.L., Bardwell, J.C. & Kuriyan, J. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature* **365**, 464-468 (1993).

115. Prinz, W.A., Aslund, F., Holmgren, A. & Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *The Journal of biological chemistry* **272**, 15661-15667 (1997).

116. Zapun, A., Missiakas, D., Raina, S. & Creighton, T.E. Structural and functional characterization of DsbC, a protein involved in disulfide bond formation in Escherichia coli. *Biochemistry* **34**, 5075-5089 (1995).

117. Vallee, B.L. & Ulmer, D.D. Biochemical effects of mercury, cadmium, and lead. *Annual review of biochemistry* **41**, 91-128 (1972).

118. Chrestensen, C.A., Starke, D.W. & Mieyal, J.J. Acute cadmium exposure inactivates thioltransferase (Glutaredoxin), inhibits intracellular reduction of protein-glutathionyl-mixed disulfides, and initiates apoptosis. *The Journal of biological chemistry* **275**, 26556-26565 (2000).

119. Figueiredo-Pereira, M.E., Yakushin, S. & Cohen, G. Disruption of the intracellular sulfhydryl homeostasis by cadmium-induced oxidative stress leads to protein thiolation and ubiquitination in neuronal cells. *The Journal of biological chemistry* **273**, 12703-12709 (1998).

120. Rensing, C., Mitra, B. & Rosen, B.P. Insertional inactivation of dsbA produces sensitivity to cadmium and zinc in Escherichia coli. *Journal of bacteriology* **179**, 2769-2771 (1997).

121. Stafford, S.J., Humphreys, D.P. & Lund, P.A. Mutations in dsbA and dsbB, but not dsbC, lead to an enhanced sensitivity of Escherichia coli to Hg2+ and Cd2+. *FEMS microbiology letters* **174**, 179-184 (1999).

122. Dailey, F.E. & Berg, H.C. Mutants in disulfide bond formation that disrupt flagellar assembly in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 1043-1047 (1993).

123. Gibson, Q.H., Swoboda, B.E. & Massey, V. Kinetics and Mechanism of Action of Glucose Oxidase. *The Journal of biological chemistry* **239**, 3927-3934 (1964).

124. Hiniker, A., Collet, J.F. & Bardwell, J.C. Copper stress causes an in vivo requirement for the Escherichia coli disulfide isomerase DsbC. *The Journal of biological chemistry* **280**, 33785-33791 (2005).

125. Wunderlich, M., Otto, A., Maskos, K., Mucke, M., Seckler, R. & Glockshuber, R. Efficient catalysis of disulfide formation during protein folding with a single active-site cysteine. *Journal of molecular biology* **247**, 28-33 (1995).

126. Chatrenet, B. & Chang, J.Y. The folding of hirudin adopts a mechanism of trial and error. *The Journal of biological chemistry* **267**, 3038-3043 (1992).

127. Holmgren, A. Tryptophan fluorescence study of conformational transitions of the oxidized and reduced form of thioredoxin. *The Journal of biological chemistry* **247**, 1992-1998 (1972).

128. Wunderlich, M., Otto, A., Seckler, R. & Glockshuber, R. Bacterial protein disulfide isomerase: efficient catalysis of oxidative protein folding at acidic pH. *Biochemistry* **32**, 12251-12256 (1993).

129. Hawkins, H.C., Blackburn, E.C. & Freedman, R.B. Comparison of the activities of protein disulphide-isomerase and thioredoxin in catalysing disulphide isomerization in a protein substrate. *The Biochemical journal* **275 (Pt 2)**, 349-353 (1991).

130. Pigiet, V.P. & Schuster, B.J. Thioredoxin-catalyzed refolding of disulfide-containing proteins. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 7643-7647 (1986).

131. Morgan, R.S., Tatsch, C.E., Gushard, R.H., McAdon, J. & Warme, P.K. Chains of alternating sulfur and pi-bonded atoms in eight small proteins. *International journal of peptide and protein research* **11**, 209-217 (1978).

132. Zauhar, R.J., Colbert, C.L., Morgan, R.S. & Welsh, W.J. Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **53**, 233-248 (2000).

133. Dougherty, D.A. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163-168 (1996).

134. Pal, D. & Chakrabarti, P. Different types of interactions involving cysteine sulfhydryl group in proteins. *Journal of biomolecular structure & dynamics* **15**, 1059-1072 (1998).

135. Britto, P.J., Knipling, L. & Wolff, J. The local electrostatic environment determines cysteine reactivity of tubulin. *The Journal of biological chemistry* **277**, 29018-29027 (2002).

136. Guddat, L.W., Bardwell, J.C. & Martin, J.L. Crystal structures of reduced and oxidized DsbA: investigation of domain motion and thiolate stabilization. *Structure* **6**, 757-767 (1998).

137. Aqvist, J., Luecke, H., Quiocho, F.A. & Warshel, A. Dipoles localized at helix termini of proteins stabilize charges. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 2026-2030 (1991).

138. Kortemme, T. & Creighton, T.E. Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. *Journal of molecular biology* **253**, 799-812 (1995).

139. Kortemme, T., Darby, N.J. & Creighton, T.E. Electrostatic interactions in the active site of the N-terminal thioredoxin-like domain of protein disulfide isomerase. *Biochemistry* **35**, 14503-14511 (1996).

140. Iqbalsyah, T.M., Moutevelis, E., Warwicker, J., Errington, N. & Doig, A.J. The CXXC motif at the N terminus of an alpha-helical peptide. *Protein Sci* **15**, 1945-1950 (2006).

141. Joly, J.C. & Swartz, J.R. Protein folding activities of Escherichia coli protein disulfide isomerase. *Biochemistry* **33**, 4231-4236 (1994).

142. Sun, X.X. & Wang, C.C. The N-terminal sequence (residues 1-65) is essential for dimerization, activities, and peptide binding of Escherichia coli DsbC. *The Journal of biological chemistry* **275**, 22743-22749 (2000).

143. Zhao, Z., Peng, Y., Hao, S.F., Zeng, Z.H. & Wang, C.C. Dimerization by domain hybridization bestows chaperone and isomerase activities. *The Journal of biological chemistry* **278**, 43292-43298 (2003).

144. Solovyov, A., Xiao, R. & Gilbert, H.F. Sulfhydryl oxidation, not disulfide isomerization, is the principal function of protein disulfide isomerase in yeast Saccharomyces cerevisiae. *The Journal of biological chemistry* **279**, 34095-34100 (2004).

145. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. in The Proteomics Protocols Handbook. (ed. J.M. Walker) (Humana Press, 2005).

146. Ritz, D. & Beckwith, J. Redox state of cytoplasmic thioredoxin. *Methods in enzymology* **347**, 360-370 (2002).

147. Grzyska, P.K., Ryle, M.J., Monterosso, G.R., Liu, J., Ballou, D.P. & Hausinger, R.P. Steady-state and transient kinetic analyses of taurine/alpha-ketoglutarate dioxygenase: effects of oxygen concentration, alternative sulfonates, and active-site variants on the FeIV-oxo intermediate. *Biochemistry* **44**, 3845-3855 (2005).

148. Rozhkova, A., Stirnimann, C.U., Frei, P., Grauschopf, U., Brunisholz, R., Grutter, M.G., Capitani, G. & Glockshuber, R. Structural basis and kinetics of inter- and intramolecular disulfide exchange in the redox catalyst DsbD. *The EMBO journal* **23**, 1709-1719 (2004).

149. Collet, J.F., Riemer, J., Bader, M.W. & Bardwell, J.C. Reconstitution of a disulfide isomerization system. *The Journal of biological chemistry* **277**, 26886-26892 (2002).

150. Aslund, F., Berndt, K.D. & Holmgren, A. Redox potentials of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *The Journal of biological chemistry* **272**, 30780-30786 (1997).

151. Jaenicke, R. Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155-241 (1999).

152. Gierasch, L.M. & Gershenson, A. Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nat. Chem. Biol.* **5**, 774-777 (2009).

153. Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. & Hartl, F.U. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122**, 209-220 (2005).

154. Duguay, A.R. & Silhavy, T.J. Quality control in the bacterial periplasm. *Biochim Biophys Acta* **1694**, 121-134 (2004).

155. Kolaj, O., Spada, S., Robin, S. & Wall, J.G. Use of folding modulators to improve heterologous protein production in Escherichia coli. *Microbial Cell Factories* **8**, - (2009).

156. Nikaido, H., Basina, M., Nguyen, V. & Rosenberg, E.Y. Multidrug efflux pump AcrAB of Salmonella typhimurium excretes only those beta-lactam antibiotics containing lipophilic side chains. *J Bacteriol* **180**, 4686-4692 (1998).

157. Helbig, K., Bleuel, C., Krauss, G.J. & Nies, D.H. Glutathione and transition-metal homeostasis in Escherichia coli. *J Bacteriol* **190**, 5431-5438 (2008).

158. Jaffe, A., Chabbert, Y.A. & Semonin, O. Role of Porin Proteins Ompf and Ompc in the Permeation of Beta-Lactams. *Antimicrob Agents Ch* **22**, 942-948 (1982).

159. Brocklehurst, K.R. & Morby, A.P. Metal-ion tolerance in Escherichia coli: analysis of transcriptional profiles by gene-array technology. *Microbiol-Uk* **146**, 2277-2282 (2000).

160. Hennecke, J., Sebbel, P. & Glockshuber, R. Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J Mol Biol.* **286**, 1197-1215 (1999).

161. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nature Structural & Molecular Biology* **16**, 318-324 (2009).

162. Kleanthous, C., Hemmings, A.M., Moore, G.R. & James, R. Immunity proteins and their specificity for endonuclease colicins: telling right from wrong in protein-protein recognition. *Mol Microbiol* **28**, 227-233 (1998).

163. Spence, G.R., Capaldi, A.P. & Radford, S.E. Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J. Mol. Biol.* **341**, 215-226 (2004).

164. Raffa, R.G. & Raivio, T.L. A third envelope stress signal transduction pathway in Escherichia coli. *Mol Microbiol* **45**, 1599-1611 (2002).

165. MacRitchie, D.M., Buelow, D.R., Price, N.L. & Raivio, T.L. Two-component signaling and gram negative envelope stress response systems. *Adv. Exp. Med. Biol.* **631**, 80-110 (2008).

166. Duguay, A.R. & Silhavy, T.J. Quality control in the bacterial periplasm. *Biochim. Biophys. Acta* **1694**, 121-134 (2004).

167. Price, N.L. & Raivio, T.L. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J. Bacteriol.* **191**, 1798-1815 (2009).

168. Khorchid, A. & Ikura, M. Bacterial histidine kinase as signal sensor and transducer. *Int J Biochem Cell B* **38**, 307-312 (2006).

169. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. & Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2** (2006).

170. Raivio, T.L., Laird, M.W., Joly, J.C. & Silhavy, T.J. Tethering of CpxP to the inner membrane prevents spheroplast induction of the cpx envelope stress response. *Mol Microbiol* **37**, 1186-1197 (2000).

171. Hagenmaier, S., Stierhof, Y.D. & Henning, U. A new periplasmic protein of *Escherichia coli* which is synthesized in spheroplasts but not in intact cells. *J. Bacteriol.* **179**, 2073-2076 (1997).

172. Zoetendal, E.G., Smith, A.H., Sundset, M.A. & Mackie, R.I. The BaeSR two-component regulatory system mediates resistance to condensed tannins in Escherichia coli. *Appl Environ Microbiol* **74**, 535-539 (2008).

173. Yamamoto, K., Ogasawara, H. & Ishihama, A. Involvement of multiple transcription factors for metal-induced spy gene expression in Escherichia coli. *J Biotechnol* **133**, 196-200 (2008).

174. Vertommen, D., Depuydt, M., Pan, J., Leverrier, P., Knoops, L., Szikora, J.P., Messens, J., Bardwell, J.C. & Collet, J.F. The disulphide isomerase DsbC cooperates with the oxidase DsbA in a DsbD-independent manner. *Mol Microbiol* **67**, 336-349 (2008).

175. Bergholz, T.M., Vanaja, S.K. & Whittam, T.S. Gene expression induced in Escherichia coli O157:H7 upon exposure to model apple juice. *Appl. Environ. Microbiol.* **75**, 3542-3553 (2009).

176. Nishino, K., Honda, T. & Yamaguchi, A. Genome-wide analyses of Escherichia coli gene expression responsive to the BaeSR two-component regulatory system. *Journal of bacteriology* **187**, 1763-1772 (2005).

177. Garbe, T.R., Kobayashi, M. & Yukawa, H. Indole-inducible proteins in bacteria suggest membrane and oxidant toxicity. *Arch Microbiol* **173**, 78-82 (2000).

178. Blom, A., Harder, W. & Matin, A. Unique and Overlapping Pollutant Stress Proteins of Escherichia-Coli. *Appl. Environ. Microbiol.* **58**, 331-334 (1992).

179. Vandyk, T.K., Smulski, D.R., Reed, T.R., Belkin, S., Vollmer, A.C. & Larossa, R.A. Responses to Toxicants of an Escherichia-Coli Strain Carrying a Uspa'--Lux Genetic Fusion and an Escherichia-Coli Strain Carrying a Grpe'--Lux Fusion Are Similar. *Appl. Environ. Microbiol.* **61**, 4124-4127 (1995).

180. Vanbogelen, R.A., Acton, M.A. & Neidhardt, F.C. Induction of the Heat-Shock Regulon Does Not Produce Thermotolerance in Escherichia-Coli. *Genes Dev.* **1**, 525-531 (1987).

181. Zhou, L., Lei, X.H., Bochner, B.R. & Wanner, B.L. Phenotype microarray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems. *J. Bacteriol.* **185**, 4956-4972 (2003).

182. Zanchi, D., Narayanan, T., Hagenmuller, D., Baron, A., Guyot, S., Cabane, B. & Bouhallab, S. Tannin-assisted aggregation of natively unfolded proteins. *Europhys. Lett.* **82**, 58001 (2008).

183. Hofmann, T., Glabasnia, A., Schwarz, B., Wisman, K.N., Gangwer, K.A. & Hagerman, A.E. Protein binding and astringent taste of a polymeric procyanidin, 1,2,3,4,6-penta-O-galloyl-beta-D-glucopyranose, castalagin, and grandinin. *J Agr Food Chem* **54**, 9503-9509 (2006).

184. Danese, P.N. & Silhavy, T.J. CpxP, a stress-combative member of the Cpx regulon. *J. Bacteriol.* **180**, 831-839 (1998).

185. Raivio, T.L., Popkin, D.L. & Silhavy, T.J. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *J. Bacteriol.* **181**, 5263-5272 (1999).

186. DiGiuseppe, P.A. & Silhavy, T.J. Signal detection and target gene induction by the CpxRA two-component system. *J. Bacteriol.* **185**, 2432-2440 (2003).

187. Isaac, D.D., Pinkner, J.S., Hultgren, S.J. & Silhavy, T.J. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc Natl Acad Sci U S A.* **102**, 17775-17779 (2005).

188. Tapley, T.L., Korner, J.L., Barge, M.T., Hupfeld, J., Schauerte, J.A., Gafni, A., Jakob, U. & Bardwell, J.C.A. Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5557-5562 (2009).

189. Bukau, B. & Horwich, A.L. The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366 (1998).

190. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169-1175 (2004).

191. Nijkamp, H.J.J., Delang, R., Stuitje, A.R., Vandenelzen, P.J.M., Veltkamp, E. & Vanputten, A.J. The complete nucleotide-sequence of the bacteriocinogenic plasmid Clodf13. *Plasmid* **16**, 135-160 (1986).

192. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A.* **97**, 6640-6645 (2000).

193. Bochner, B.R., Huang, H.C., Schieven, G.L. & Ames, B.N. Positive selection for loss of tetracycline resistance. *J. Bacteriol.* **143**, 926-933 (1980).

194. Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y. & Lei, M. Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc Natl Acad Sci U S A.* **103**, 13956-13961 (2006).

195. Hiniker, A. & Bardwell, J.C.A. In vivo substrate specificity of periplasmic disulfide oxidoreductases. *J Biol Chem.* **279**, 12967-12973 (2004).

196. Whitmore, L. & Wallace, B.A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392-400 (2008).

197. Whitmore, L. & Wallace, B.A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668-673 (2004).

198. Lakowicz, J. Principles of fluorescence spectroscopy, Vol. XXVI. (Springer, New York; 2006).

199. Hayer-Hartl, M. Assay of malate dehydrogenase. A substrate for the *E. coli* chaperonins GroEL and GroES. *Methods Mol. Biol.* **140**, 127-132 (2000).

200. Bader, M.W., Xie, T., Yu, C.A. & Bardwell, J.C. Disulfide bonds are generated by quinone reduction. *J. Biol. Chem.* **275**, 26082-26088 (2000).

201. Otwinowski, Z. & Minor, W. in Macromolecular Crystallography, Pt A, Vol. 276 307-3261997).

202. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215-230 (2007).

203. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).

204. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255 (1997).

205. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**, 213-221 (2010).

206. Hiniker, A. & Bardwell, J.C. In vivo substrate specificity of periplasmic disulfide oxidoreductases. *J. Biol. Chem.* **279**, 12967-12973 (2004).

207. Kim, J.Y., Fogarty, E.A., Lu, F.J., Zhu, H., Wheelock, G.D., Henderson, L.A. & DeLisa, M.P. Twin-arginine translocation of active human tissue plasminogen activator in Escherichia coli. *Appl Environ Microb* **71**, 8451-8459 (2005).

208. Rietsch, A., Bessette, P., Georgiou, G. & Beckwith, J. Reduction of the periplasmic disulfide bond isomerase, DsbC, occurs by passage of electrons from cytoplasmic thioredoxin. *J Bacteriol* **179**, 6602-6608 (1997).

209. Chun, S.Y., Strobel, S., Bassford, P., Jr. & Randall, L.L. Folding of maltose-binding protein. Evidence for the identity of the rate-determining step in vivo and in vitro. *J. Biol. Chem.* **268**, 20855-20862 (1993).

210. Cummings, C.J. & Zoghbi, H.Y. Trinucleotide repeats: mechanisms and pathophysiology. *Annu. Rev. Genomics Hum. Genet.* **1**, 281-328 (2000).

211. Prahlad, V. & Morimoto, R.I. Integrating the stress response: lessons for neurodegenerative diseases from C. elegans. *Trends Cell Biol.* **19**, 52-61 (2009).

212. Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.t., Lindsey, R., Hersch, S.M., Jope, R.S., Albin, R.L. & Detloff, P.J. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* **91**, 753-763 (1997).

213. Walters, R.H. & Murphy, R.M. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* **393**, 978-992 (2009).

214. Barton, S., Jacak, R., Khare, S.D., Ding, F. & Dokholyan, N.V. The length dependence of the polyQ-mediated protein aggregation. *J. Biol. Chem.* **282**, 25487-25492 (2007).

215. Ross, C.A., Poirier, M.A., Wanker, E.E. & Amzel, M. Polyglutamine fibrillogenesis: the pathway unfolds. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1-3 (2003).

216. Wang, X. & Chapman, M.R. Curli provide the template for understanding controlled amyloid propagation. *Prion* **2**, 57-60 (2008).

217. Robinson, L.S., Ashman, E.M., Hultgren, S.J. & Chapman, M.R. Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol. Microbiol.* **59**, 870-881 (2006).

218. Hartl, F.U. & Hayer-Hartl, M. Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858 (2002).

219. Maier, T., Ferbitz, L., Deuerling, E. & Ban, N. A cradle for new proteins: trigger factor at the ribosome. *Curr Opin Struct Biol* **15**, 204-212 (2005).

220. Patzelt, H., Kramer, G., Rauch, T., Schonfeld, H.J., Bukau, B. & Deuerling, E. Three-state equilibrium of Escherichia coli trigger factor. *Biol Chem* **383**, 1611-1619 (2002).

221. Bardwell, J.C.A. & Craig, E.A. Major Heat-Shock Gene of Drosophila and the Escherichia-Coli Heat-Inducible Dnak Gene Are Homologous. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**, 848-852 (1984).

222. Zhu, X.T., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E. & Hendrickson, W.A. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-1614 (1996).

223. Slepenkov, S.V. & Witt, S.N. The unfolding story of the Escherichia coli Hsp70 DnaK: is DnaK a holdase or an unfoldase? *Mol Microbiol* **45**, 1197-1206 (2002).

224. Genevaux, P., Georgopoulos, C. & Kelley, W.L. The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions. *Mol Microbiol* **66**, 840-857 (2007).

225. Horwich, A.L., Apetri, A.C. & Fenton, W.A. The GroEL/GroES cis cavity as a passive anti-aggregation device. *Febs Lett* **583**, 2654-2662 (2009).

226. Horwich, A.L. & Fenton, W.A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys* **42**, 83-116 (2009).

227. Lund, P.A. Multiple chaperonins in bacteria - why so many? *Fems Microbiol Rev* **33**, 785-800 (2009).

228. England, J.L., Lucent, D. & Pande, V.S. A role for confined water in chaperonin function. *Journal of the American Chemical Society* **130**, 11838-11839 (2008).

229. Viitanen, P.V., Gatenby, A.A. & Lorimer, G.H. Purified Chaperonin 60 (Groel) Interacts with the Nonnative States of a Multitude of Escherichia-Coli Proteins. *Protein Sci* **1**, 363-369 (1992).

230. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574-581 (2009).

231. Jakob, U., Muse, W., Eser, M. & Bardwell, J.C. Chaperone activity with a redox switch. *Cell* **96**, 341-352 (1999).

232. Jakob, U., Eser, M. & Bardwell, J.C.A. Redox switch of Hsp33 has a novel zinc-binding motif. *J Biol Chem* **275**, 38302-38310 (2000).

233. Winter, J., Linke, K., Jatzek, A. & Jakob, U. Severe oxidative stress causes inactivation of DnaK and activation of the redox-regulated chaperone Hsp33. *Mol Cell* **17**, 381-392 (2005).

234. Hoffmann, J.H., Linke, K., Graf, P.C.F., Lilie, H. & Jakob, U. Identification of a redox-regulated chaperone network. *Embo J* **23**, 160-168 (2004).

235. Mujacic, M., Bader, M.W. & Baneyx, F. Escherichia coli Hsp31 functions as a holding chaperone that cooperates with the DnaK-DnaJ-GrpE system in the management of protein misfolding under severe stress conditions. *Mol Microbiol* **51**, 849-859 (2004).

236. Sastry, M.S.R., Korotkov, K., Brodsky, Y. & Baneyx, F. Hsp31, the Escherichia coli yedU gene product, is a molecular chaperone whose activity is inhibited by ATP at high temperatures. *J Biol Chem* **277**, 46026-46034 (2002).

237. Malki, A., Caldas, T., Abdallah, J., Kern, R., Eckey, V., Kim, S.J., Cha, S.S., Mori, H. & Richarme, G. Peptidase activity of the Escherichia coli Hsp31 chaperone. *J Biol Chem* **280**, 14420-14426 (2005).

238. Mujacic, M. & Baneyx, F. Chaperone Hsp31 contributes to acid resistance in stationary-phase Escherichia coli. *Appl Environ Microbiol* **73**, 1014-1018 (2007).

239. Allen, S.P., Polazzi, J.O., Gierse, J.K. & Easton, A.M. Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli. *J. Bacteriol.* **174**, 6938-6947 (1992).

240. Laskowska, E., Wawrzynow, A. & Taylor, A. IbpA and IbpB, the new heat-shock proteins, bind to endogenous Escherichia coli proteins aggregated intracellularly by heat shock. *Biochimie* **78**, 117-122 (1996).

241. Kuczynska-Wisnik, D., Kedzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipinska, B. & Laskowska, E. The Escherichia coli small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. *Microbiology-Sgm* **148**, 1757-1765 (2002).

242. Kitagawa, M., Matsumura, Y. & Tsuchido, T. Small heat shock proteins, IbpA and IbpB, are involved in resistances to heat and superoxide stresses in Escherichia coli. *FEMS Microbiol. Lett.* **184**, 165-171 (2000).

243. Kitagawa, M., Miyakawa, M., Matsumura, Y. & Tsuchido, T. Escherichia coli small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. *Eur. J. Biochem.* **269**, 2907-2917 (2002).

244. Matuszewska, E., Kwiatkowska, J., Kuczynska-Wisnik, D. & Laskowska, E. Escherichia coli heat-shock proteins IbpA/B are involved in resistance to oxidative stress induced by copper. *Microbiology-Sgm* **154**, 1739-1747 (2008).

245. Matuszewska, E., Kwiatkowska, J., Ratajczak, E., Kuczynska-Wisnik, D. & Laskowska, E. Role of Escherichia coli heat shock proteins IbpA and IbpB in protection of alcohol dehydrogenase AdhE against heat inactivation in the presence of oxygen. *Acta Biochim. Pol.* **56**, 55-61 (2009).

246. Kuczynska-Wisnik, D., Zurawa-Janicka, D., Narkiewicz, J., Kwiatkowska, J., Lipinska, B. & Laskowska, E. Escherichia coli small heat shock proteins IbpA/B enhance activity of enzymes sequestered in inclusion bodies. *Acta Biochim. Pol.* **51**, 925-931 (2004).

247. LeThanh, H., Neubauer, P. & Hoffmann, F. The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant Escherichia coli and delay degradation of inclusion bodies. *Microbial Cell Factories* **4**, - (2005).

248. Matuszewska, M., Kuczynska-Wisnik, D., Laskowska, E. & Liberek, K. The small heat shock protein IbpA of Escherichia coli cooperates with IbpB in stabilization of thermally aggregated proteins in a disaggregation competent state. *J. Biol. Chem.* **280**, 12292-12298 (2005).

249. Ratajczak, E., Zietkiewicz, S. & Liberek, K. Distinct Activities of Escherichia coli Small Heat Shock Proteins IbpA and IbpB Promote Efficient Protein Disaggregation. *J. Mol. Biol.* **386**, 178-189 (2009).

250. Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E. & Bukau, B. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol. Microbiol.* **50**, 585-595 (2003).

251. Baker, T.A. & Sauer, R.T. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**, 647-653 (2006).

252. Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., Neher, S.B., Oakes, E.S.C., Siddiqui, S.M., Wah, D.A. & Baker, T.A. Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* **119**, 9-18 (2004).

253. Barends, T.R.M., Werbeck, N.D. & Reinstein, J. Disaggregases in 4 dimensions. *Curr Opin Struc Biol* **20**, 46-53 (2010).

254. Doyle, S.M., Hoskins, J.R. & Wickner, S. Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system. *P Natl Acad Sci USA* **104**, 11138-11144 (2007).

255. Haslberger, T., Zdanowicz, A., Brand, I., Kirstein, J., Turgay, K., Mogk, A. & Bukau, B. Protein disaggregation by the AAA plus chaperone ClpB involves partial threading of looped polypeptide segments. *Nature Structural & Molecular Biology* **15**, 641-650 (2008).

256. Mogk, A., Schlieker, C., Friedrich, K.L., Schofeld, H.J., Vierling, E. & Bukau, B. Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. *J Biol Chem* **278**, 31033-31042 (2003).

257. Carrio, M.M. & Villaverde, A. Role of molecular chaperones in inclusion body formation. *Febs Lett* **537**, 215-221 (2003).

258. Veinger, L., Diamant, S., Buchner, J. & Goloubinoff, P. The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J Biol Chem* **273**, 11032-11037 (1998).

259. Bardwell, J.C. & Craig, E.A. Eukaryotic Mr 83,000 heat shock protein has a homologue in Escherichia coli. *Proc Natl Acad Sci U S A* **84**, 5177-5181 (1987).

260. Thomas, J.G. & Baneyx, F. ClpB and HtpG facilitate de novo protein folding in stressed Escherichia coli cells. *Mol Microbiol* **36**, 1360-1370 (2000).

261. Muller, M., Koch, H.G., Beck, K. & Schaefer, U. Protein traffic in bacteria: Multiple routes from the ribosome to and across the membrane. *Prog Nucleic Acid Re* **66**, 107-157 (2001).

262. Sargent, F. The twin-arginine transport system: moving folded proteins across membranes. *Biochem Soc T* **35**, 835-847 (2007).

263. Fisher, A.C. & DeLisa, M.P. A little help from my friends: Quality control of presecretory proteins in bacteria. *J Bacteriol* **186**, 7467-7473 (2004).

264. Lecker, S., Lill, R., Ziegelhoffer, T., Georgopoulos, C., Bassford, P.J., Jr., Kumamoto, C.A. & Wickner, W. Three pure chaperone proteins of Escherichia coli--SecB, trigger factor and GroEL--form soluble complexes with precursor proteins in vitro. *Embo J* **8**, 2703-2709 (1989).

265. Bochkareva, E.S., Solovieva, M.E. & Girshovich, A.S. Targeting of GroEL to SecA on the cytoplasmic membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **95**, 478-483 (1998).

266. Beck, K., Wu, L.F., Brunner, J. & Muller, M. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *Embo J* **19**, 134-143 (2000).

267. Buskiewicz, I., Deuerling, E., Gu, S.Q., Jockel, J., Rodnina, M.V., Bukau, B. & Wintermeyer, W. Trigger factor binds to ribosome-signal-recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. *P Natl Acad Sci USA* **101**, 7902-7906 (2004).

268. Puertas, J.M., Nannenga, B.L., Dornfeld, K.T., Betton, J.M. & Baneyx, F. Enhancing the secretory yields of leech carboxypeptidase inhibitor in Escherichia coli: influence of trigger factor and signal recognition particle. *Protein Expr Purif* (2010).

269. Bechtluft, P., Nouwen, N., Tans, S.J. & Driessen, A.J.M. SecB-A chaperone dedicated to protein translocation. *Mol Biosyst* **6**, 620-627 (2010).

270. Xu, Z.H., Knafels, J.D. & Yoshino, K. Crystal structure of the bacterial protein export chaperone SecB. *Nat Struct Biol* **7**, 1172-1177 (2000).

271. Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C. & Genevaux, P. SecB is a bona fide generalized chaperone in Escherichia coli. *P Natl Acad Sci USA* **101**, 7583-7588 (2004).

272. Panse, V.G., Vogel, P., Trommer, W.E. & Varadarajan, R. A thermodynamic coupling mechanism for the disaggregation of a model peptide substrate by chaperone SecB. *J Biol Chem* **275**, 18698-18703 (2000).

273. Economou, A. & Wickner, W. Seca Promotes Preprotein Translocation by Undergoing Atp-Driven Cycles of Membrane Insertion and Deinsertion. *Cell* **78**, 835-843 (1994).

274. Eser, M. & Ehrmann, M. SecA-dependent quality control of intracellular protein localization. *P Natl Acad Sci USA* **100**, 13231-13234 (2003).

275. Perez-Rodriguez, R., Fisher, A.C., Perlmutter, J.D., Hicks, M.G., Chanal, A., Santini, C.L., Wu, L.F., Palmer, T. & DeLisa, M.P. An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *J Mol Biol* **367**, 715-730 (2007).

276. Graubner, W., Schierhorn, A. & Bruser, T. DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone. *J Biol Chem* **282**, 7116-7124 (2007).

277. Li, H., Chang, L., Howell, J.M. & Turner, R.J. DmsD, a Tat system specific chaperone, interacts with other general chaperones and proteins involved in the molybdenum cofactor biosynthesis. *Biochim Biophys Acta* **1804**, 1301-1309 (2010).

278. Rouviere, P.E. & Gross, C.A. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev.* **10**, 3170-3182 (1996).

279. Behrens, S., Maier, R., de Cock, H., Schmid, F.X. & Gross, C.A. The SurA periplasmic PPlase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J.* **20**, 285-294 (2001).

280. Bitto, E. & McKay, D.B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489-1498 (2002).

281. Bitto, E. & McKay, D.B. The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J. Biol. Chem.* **278**, 49316-49322 (2003).

282. Xu, X.H., Wang, S.Y., Hu, Y.X. & McKay, D.B. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J. Mol. Biol.* **373**, 367-381 (2007).

283. Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in E-coli. *Bba-Mol Cell Res* **1693**, 5-13 (2004).

284. Matsuyama, S., Tajima, T. & Tokuda, H. A Novel Periplasmic Carrier Protein Involved in the Sorting and Transport of Escherichia-Coli Lipoproteins Destined for the Outer-Membrane. *Embo J* **14**, 3365-3372 (1995).

285. Yokota, N., Kuroda, T., Matsuyama, S. & Tokuda, H. Characterization of the LolA-LolB system as the general lipoprotein localization mechanism of Escherichia coli. *J Biol Chem* **274**, 30995-30999 (1999).

286. Matsuyama, S., Yokota, N. & Tokuda, H. A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of Escherichia coli. *Embo J* **16**, 6947-6955 (1997).

287. Okuda, S. & Tokuda, H. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *P Natl Acad Sci USA* **106**, 5877-5882 (2009).

288. Knight, S.D., Berglund, J. & Choudhury, D. Bacterial adhesins: structural studies reveal chaperone function and pilus biogenesis. *Curr Opin Chem Biol* **4**, 653-660 (2000).

289. Piatek, R., Zalewska, B., Bury, K. & Kur, J. The chaperone-usher pathway of bacterial adhesin biogenesis - from molecular mechanism to strategies of anti-bacterial prevention and modem vaccine design. *Acta Biochim Pol* **52**, 639-646 (2005).

290. Proft, T. & Baker, E.N. Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* **66**, 613-635 (2009).

291. Waksman, G. & Hultgren, S.J. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* **7**, 765-774 (2009).

292. Sauer, F.G., Knight, S.D., Waksman, G. & Hultgren, S.J. PapD-like chaperones and pilus biogenesis. *Semin Cell Dev Biol* **11**, 27-34 (2000).

293. Lindberg, F., Tennent, J.M., Hultgren, S.J., Lund, B. & Normark, S. Papd, a Periplasmic Transport Protein in P-Pilus Biogenesis. *J Bacteriol* **171**, 6052-6058 (1989).

294. Jones, C.H., Pinkner, J.S., Nicholes, A.V., Slonim, L.N., Abraham, S.N. & Hultgren, S.J. Fimc Is a Periplasmic Papd-Like Chaperone That Directs Assembly of Type-1 Pili in Bacteria. *P Natl Acad Sci USA* **90**, 8397-8401 (1993).

295. Jones, C.H., Danese, P.N., Pinkner, J.S., Silhavy, T.J. & Hultgren, S.J. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *Embo J* **16**, 6394-6406 (1997).

296. Soto, G.E., Dodson, K.W., Ogg, D., Liu, C., Heuser, J., Knight, S., Kihlberg, J., Jones, C.H. & Hultgren, S.J. Periplasmic chaperone recognition motif of subunits mediates quaternary interactions in the pilus. *Embo J* **17**, 6155-6167 (1998).

297. Kuehn, M.J., Normark, S. & Hultgren, S.J. Immunoglobulin-Like Papd Chaperone Caps and Uncaps Interactive Surfaces of Nascently Translocated Pilus Subunits. *P Natl Acad Sci USA* **88**, 10586-10590 (1991).

298. Hung, D.L., Raivio, T.L., Jones, C.H., Silhavy, T.J. & Hultgren, S.J. Cpx signaling pathway monitors biogenesis and affects assembly and expression of P pili. *Embo J* **20**, 1508-1518 (2001).

299. Vogt, S.L., Nevesinjac, A.Z., Humphries, R.M., Donnenberg, M.S., Armstrong, G.D. & Raivio, T.L. The Cpx envelope stress response both facilitates and inhibits elaboration of the enteropathogenic Escherichia coli bundle-forming pilus. *Mol Microbiol* **76**, 1095-1110 (2010).

300. Thanassi, D.G., Saulino, E.T. & Hultgren, S.J. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr Opin Microbiol* **1**, 223-231 (1998).

301. Sauer, F.G., Remaut, H., Hultgren, S.J. & Waksman, G. Fiber assembly by the chaperone-usher pathway. *Bba-Mol Cell Res* **1694**, 259-267 (2004).

302. Thanassi, D.G., Saulino, E.T., Lombardo, M.J., Roth, R., Heuser, J. & Hultgren, S.J. The PapC usher forms an oligomeric channel: Implications for pilus biogenesis across the outer membrane. *P Natl Acad Sci USA* **95**, 3146-3151 (1998).

303. Saulino, E.T., Thanassi, D.G., Pinkner, J.S. & Hultgren, S.J. Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *Embo J* **17**, 2177-2185 (1998).

304. Barnhart, M.M., Pinkner, J.S., Soto, G.E., Sauer, F.G., Langermann, S., Waksman, G., Frieden, C. & Hultgren, S.J. PapD-like chaperones provide the missing information for folding of pilin proteins. *P Natl Acad Sci USA* **97**, 7709-7714 (2000).

305. Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J. & Waksman, G. Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058-1061 (1999).

306. Missiakas, D., Betton, J.M. & Raina, S. New components of protein folding in extracytoplasmic compartments of Escherichia coli SurA, FkpA and Skp/OmpH. *Mol Microbiol* **21**, 871-884 (1996).

307. Wu, M.S., Pan, K.L. & Chou, C.P. Effect of heat-shock proteins for relieving physiological stress and enhancing the production of penicillin acylase in Escherichia coli. *Biotechnol Bioeng* **96**, 956-966 (2007).

308. Scholz, C., Schaarschmidt, P., Engel, A.M., Andres, H., Schmitt, U., Faatz, E., Balbach, J. & Schmid, F.X. Functional solubilization of aggregation-prone HIV envelope proteins by covalent fusion with chaperone modules. *J Mol Biol* **345**, 1229-1241 (2005).

309. Saul, F.A., Arie, J.P., Vulliez-le Normand, B., Kahn, R., Betton, J.M. & Bentley, G.A. Structural and functional studies of FkpA from Escherichia coli, a cis/trans peptidyl-prolyl isomerase with chaperone activity. *J Mol Biol* **335**, 595-608 (2004).

310. Hu, K.F., Galius, V. & Pervushin, K. Structural plasticity of peptidyl-prolyl isomerase sFkpA is a key to its chaperone function as revealed by solution NMR. *Biochemistry-Us* **45**, 11983-11991 (2006).

311. Chen, R. & Henning, U. A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. *Mol Microbiol* **19**, 1287-1294 (1996).

312. Schafer, U., Beck, K. & Muller, M. Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J Biol Chem* **274**, 24567-24574 (1999).

313. Bulieris, P.V., Behrens, S., Holst, O. & Kleinschmidt, J.H. Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide. *J Biol Chem* **278**, 9092-9099 (2003).

314. Schlapschy, M., Dommel, M.K., Hadian, K., Fogarasi, M., Korndorfer, I.P. & Skerra, A. The periplasmic E-coli chaperone Skp is a trimer in solution: biophysical and preliminary crystallographic characterization. *Biological Chemistry* **385**, 137-143 (2004).

315. Walton, T.A. & Sousa, M.C. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol Cell* **15**, 367-374 (2004).

316. Korndorfer, I.P., Dommel, M.K. & Skerra, A. Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. *Nature Structural & Molecular Biology* **11**, 1015-1020 (2004).

317. Qu, J., Mayer, C., Behrens, S., Holst, O. & Kleinschmidt, J.H. The trimeric periplasmic chaperone skp of Escherichia coli forms 1 : 1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions. *J Mol Biol* **374**, 91-105 (2007).

318. Walton, T.A., Sandoval, C.M., Fowler, C.A., Pardi, A. & Sousa, M.C. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *P Natl Acad Sci USA* **106**, 1772-1777 (2009).

319. De Cock, H., Schafer, U., Potgeter, M., Demel, R., Muller, M. & Tommassen, J. Affinity of the periplasmic chaperone Skp of Escherichia coli for phospholipids, lipopolysaccharides and non-native outer membrane proteins. Role of Skp in the biogenesis of outer membrane protein. *Eur J Biochem* **259**, 96-103 (1999).

320. Fink, A.L., Calciano, L.J., Goto, Y., Kurotsu, T. & Palleros, D.R. Classification of Acid Denaturation of Proteins - Intermediates and Unfolded States. *Biochemistry-Us* **33**, 12504-12511 (1994).

321. Goto, Y., Takahashi, N. & Fink, A.L. Mechanism of Acid-Induced Folding of Proteins. *Biochemistry-Us* **29**, 3480-3488 (1990).

322. Tucker, D.L., Tucker, N. & Conway, T. Gene expression profiling of the pH response in Escherichia coli. *J. Bacteriol.* **184**, 6551-6558 (2002).

323. Hong, W.Z., Jiao, W.W., Hu, J.C., Zhang, J.R., Liu, C., Fu, X.M., Shen, D., Xia, B. & Chang, Z.Y. Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. *J Biol Chem* **280**, 27029-27034 (2005).

324. Wu, Y.E., Hong, W.Z., Liu, C., Zhang, L.Q. & Chang, Z.Y. Conserved amphiphilic feature is essential for periplasmic chaperone HdeA to support acid resistance in enteric bacteria. *Biochem J* **412**, 389-397 (2008).

325. Kern, R., Malki, A., Abdallah, J., Tagourti, J. & Richarme, G. Escherichia coli HdeB is an acid stress chaperone. *J Bacteriol* **189**, 603-610 (2007).

326. Tapley, T.L., Franzmann, T.M., Chakraborty, S., Jakob, U. & Bardwell, J.C.A. Protein refolding by pH-triggered chaperone binding and release. *P Natl Acad Sci USA* **107**, 1071-1076 (2010).

327. Rizzitello, A.E., Harper, J.R. & Silhavy, T.J. Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. *J Bacteriol* **183**, 6794-6800 (2001).

328. Jakob, R.P., Zoldak, G., Aumuller, T. & Schmid, F.X. Chaperone domains convert prolyl isomerases into generic catalysts of protein folding. *P Natl Acad Sci USA* **106**, 20282-20287 (2009).

329. Dartigalongue, C. & Raina, S. A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in Escherichia coli. *Embo J* **17**, 3968-3980 (1998).

330. Antonoaea, R., Furst, M., Nishiyama, K. & Muller, M. The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry-Us* **47**, 5649-5656 (2008).

331. Stymest, K.H. & Klappa, P. The periplasmic peptidyl prolyl cis-trans isomerases PpiD and SurA have partially overlapping substrate specificities. *Febs J* **275**, 3470-3479 (2008).

332. Kleerebezem, M., Heutink, M. & Tommassen, J. Characterization of an Escherichia coli rotA mutant, affected in periplasmic peptidyl-prolyl cis/trans isomerase. *Mol Microbiol* **18**, 313-320 (1995).

333. Tsilibaris, V., Maenhaut-Michel, G. & Van Melderen, L. Biological roles of the Lon ATP-dependent protease. *Res Microbiol* **157**, 701-713 (2006).

334. Zolkiewski, M. A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Mol Microbiol* **61**, 1094-1100 (2006).

335. Narberhaus, F., Obrist, M., Fuhrer, F. & Langklotz, S. Degradation of cytoplasmic substrates by FtsH, a membrane-anchored protease with many talents. *Res Microbiol* **160**, 652-659 (2009).

336. Gottesman, S. Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Bi* **19**, 565-587 (2003).

337. Lipinska, B., Sharma, S. & Georgopoulos, C. Sequence-Analysis and Regulation of the Htra Gene of Escherichia-Coli - a Sigma-32-Independent Mechanism of Heat-Inducible Transcription. *Nucleic Acids Res* **16**, 10053-10067 (1988).

338. Strauch, K.L. & Beckwith, J. An Escherichia-Coli Mutation Preventing Degradation of Abnormal Periplasmic Proteins. *P Natl Acad Sci USA* **85**, 1576-1580 (1988).

339. Uhland, K., Mondigler, M., Spiess, C., Prinz, W. & Ehrmann, M. Determinants of translocation and folding of TreF, a trehalase of Escherichia coli. *J Biol Chem* **275**, 23439-23445 (2000).

340. Arslan, E., Schulz, H., Zufferey, R., Kunzler, P. & Thony-Meyer, L. Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb3 oxidase in Escherichia coli. *Biochem Biophys Res Commun* **251**, 744-747 (1998).

341. Laskowska, E., KuczynskaWisnik, D., SkorkoGlonek, J. & Taylor, A. Degradation by proteases Lon, Clp and HtrA, of Escherichia coli proteins aggregated in vivo by heat shock; HtrA protease action in vivo and in vitro. *Mol Microbiol* **22**, 555-571 (1996).

342. Kolmar, H., Waller, P.R.H. & Sauer, R.T. The DegP and DegQ periplasmic endoproteases of Escherichia coli: Specificity for cleavage sites and substrate conformation. *J Bacteriol* **178**, 5925-5929 (1996).

343. Lipinska, B., Zylicz, M. & Georgopoulos, C. The Htra (Degp) Protein, Essential for Escherichia-Coli Survival at High-Temperatures, Is an Endopeptidase. *J Bacteriol* **172**, 1791-1797 (1990).

344. Iwanczyk, J., Damjanovic, D., Kooistra, J., Leong, V., Jomaa, A., Ghirlando, R. & Ortega, J. Role of the PDZ domains in Escherichia coli DegP protein. *J Bacteriol* **189**, 3176-3186 (2007).

345. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455-459 (2002).

346. Meerman, H.J. & Georgiou, G. Construction and Characterization of a Set of Escherichia-Coli Strains Deficient in All Known Loci Affecting the Proteolytic Stability of Secreted Recombinant Proteins. *Bio-Technol* **12**, 1107-1110 (1994).