Genetic selection systems to study and optimize protein folding *in vivo*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Antje Schickert (geb. Müller)

aus Freudenstadt

Tübingen

2015

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	10. März 2015
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Volkmar Braun
2. Berichterstatter:	Prof. Dr. James C. A. Bardwell

To my family

Meiner Familie

Acknowledgments

I would like to express my gratitude and special thanks to my adviser and mentor Prof. Bardwell who has helped me to grow as a scientist. His encouragement, advice and support helped me tremendously during my work in the lab. His expertise in the field of microbiology and his enthusiasm were huge sources of innovation and motivation. Thank you for giving me the chance to work and learn in your lab and under your supervision.

I deeply appreciate all the help from my supervisor Prof. Braun. This work would not have been possible without his support and encouragement. His contagious enthusiasm for science and love for teaching has always fascinated and inspired me. Thank you for all our efforts and your tremendous help!

I want to thank Prof. Ursula Jakob for many ideas and critical input to improve my research. I loved to work with so many bright and exciting scientists at the University of Michigan. I want to thank Linda Foit who introduced me in the lab and taught me about the biosensors. Thank you to Michael Gray for the wonderful cooperation, great discussions and wonderful protocols.

I want to thank Scott Horowitz, Christopher Lennon and Shu Quan for their help and advice. Thanks for the great discussion, ideas and constructs.

Tsinatkeab Tadesse , Daniela Knoefler, Claudia Cremers, Bastian Groitl and Jan Dahl, special thanks go to you guys for all the advice, support and fun that made those years in Michigan very special and unforgettable.

Ein riesen Dankeschön auch an meine Eltern und meinen Bruder Andreas. Danke, dass ihr immer für mich da seid und mich unterstützt. Ihr seid die beste Familie die man sich wünschen kann und habt mir unglaublich geholfen.

Thank you to my husband Matthew Schickert who is the rock in my life. Thank you for your faith and patience and teaching me perseverance and conviction. You are my hero.

Table of Contents

1 Abbreviations

1,3 BPG	1,3-Bisphosphoglycerate
ACP	human muscle acylphosphatase
ANT	aminoglycoside adenylyltransferase
APH	Aminoglycoside-3-6 phosphotransferase-IIa
ATP	Adenosine triphosphate
BPTI	Bovine trypsin inhibitor
CAT	Chloramphenicol acetyltransferase
Cys	Cysteine
DHAP	Dihydroxyacetophosphate
DsbA	bacterial disulfide oxidoreductase; disulfide bond family
E. coli	*Escherichia coli*
ER	Endoplasmic reticulum
G3P	Glycerinaldehyde-3-phosphate
GAPDH	Glycerinaldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
hG-CSF	human granulocyte-colony stimulating factor
HOCL	Hypochlorous acid
Im7	Immunity protein 7
MgsA	Methylglyoxal synthase
MIC	Minimal inhibitory concentration
NAT	Nourseothricin acetyltransferase
PDI	Protein disulfide isomerase
PolyP	Polyphosphate
PolyQ	Polyglutamine
PPK	Polyphosphate kinase
PPX	Exopolyphosphatase
rpoH	RNA polymerase sigma-32 factor
S.cerevisiae	Saccharomyces cerevisiae
SDM	Site-directed mutagenesis
V. cholerae	*Vibrio cholerae*

2 Abstract

Most proteins have to fold into their native structure, a distinct three-dimensional fold, in order to perform their functions in the cell. Many proteins can reach their native state without the help of other macromolecules *in vitro*, but the crowded environment of the cell constitutes a more challenging folding environment. Biosensors can facilitate studying protein folding in the cell by linking protein stability of a guest protein to antibiotic resistance. Biosensors designed in our lab consist of an antibiotic resistance marker with a guest protein inserted into a permissive site. The folding status of the guest protein can be easily assessed by measuring the antibiotic resistance of cells expressing these biosensors.

This method allowed us to investigate the folding pathway of bovine pancreatic trypsin inhibitor (BPTI) directly in the complex environment of the cell. The biosensor proved especially powerful when we selected for protein variants that increased protein expression by demanding growth on high antibiotic concentrations. Most BPTI variants that achieved increased expression did so by eliminating or destabilizing the disulfide bond Cys14-Cys38 in BPTI. This had the effect of reducing the formation of kinetically trapped intermediates during folding, which are more susceptible to aggravation or proteolysis.

Previously the lab had developed biosensors that only worked in the periplasm of prokaryotes. I developed a biosensor that could function in the cytosolic compartment. While doing so I established a protocol that allows one to easily identify permissive sites in antibiotic resistance proteins. This approach gave rise to three new cytosolic biosensors, all of which exhibited a linear correlation between antibiotic resistance and protein stability *in vivo* for different guest proteins independently of the exact position of the permissive site in the resistance marker.

The expansion of our antibiotic resistant biosensors into the cytosolic compartment offers the opportunity to investigate the effect of chaperones on proteins directly in the cell. The stabilizing effect of polyphosphate was measured using the antibiotic resistance proteins, which provided *in vivo* evidence for polyphosphates chaperone activity.

Biosensors are efficient tools for investigating proteostasis *in vivo*. Using this method we can assess the influence of cellular components such as chaperones on protein folding and expression, and monitor the formation of amyloids. Furthermore, the biosensors allow us to evolve guest proteins in order to improve their folding properties simply by selecting for increased antibiotic resistance. This makes these biosensors effective tools for optimizing the expression of problematic biotechnological or pharmaceutically interesting proteins.

3 Zusammenfassung

Um ihre Funktion in der Zelle ausüben zu können, müssen viele Proteine sich in eine spezifische dreidimensionale Struktur, ihre native Form, falten. *In vitro* können viele Proteine ohne die Unterstützung andere Makromoleküle falten, aber im hoch konzentrierten Zellinneren sind die meisten Proteine auf Hilfe angewiesen.

Um das komplexe Zusammenspiel der *in vivo* Proteinfaltung zu erforschen, konstruierten wir Biosensoren, die aus einem Antibiotikaresistenzprotein bestehen, in die ein sogenanntes Gastprotein inseriert wurde. Da Proteinstabilität an Antibiotikaresistenz gekoppelt wurde, kann der *in vivo* Faltungsstatus des Gastproteins einfach ermittelt werden, indem die Antibiotikaresistenz der Biosensor-exprimierenden *E. coli* Zellen getestet wird.

Diese Methode ermöglicht uns den Faltungsweg des bovine pancreatic trypsin inhibitors (BPTI) direkt in der Komplexität des Periplasmas einer *E. coli* Zelle zu erforschen. Durch die Selektion für erhöhte Antibiotikaresistenz erlaubt der Biosensor die Identifikation von Proteinvarianten mit verbesserten Expressionseigenschaften. Die meisten der selektierten BPTI Varianten mit erhöhter Antibiotikaresistenz eliminierten die Disulfidbrücke zwischen Cysteinen 14 und 38, was die Bildung von kinetisch gefangenen Intermediaten im Faltungsprozess verhindert. Die erhöhte Expression der selektierten Varianten ist daher vermutlich auf das Umgehen dieser Zwischenstufen zurückzuführen, die aufgrund höherer Tendenz zur Aggregation oder Protolyse die Faltungseffizienz reduzieren.

Um die Anwendbarkeit unserer Methode zu erweitern, entwickelten wir Biosensoren für das Zytoplasma. Das Protokoll, um effizient geeignete Insertionsstellen (permissive Positionen) für Gastproteine in Antibiotikaresistenzproteinen zu identifizieren, konnte genutzt werden um insgesamt drei neue Biosensoren etablieren. Alle Biosensoren zeigen eine lineare Korrelation zwischen Antibiotikaresistenz und Proteinstabilität des Gastproteins, unabhängig von der genauen Position der permissiven Insertionsstelle.

Die Expansion der Biosensoren ins Zytosol erlaubt uns den Effekt von Chaperonen auf Proteine direkt in der Zelle zu erforschen. Die Resistenzproteine konnten als Sensor für Stabilisierung durch Polyphosphat genutzt werden und lieferte somit einen *in vivo* Indiz für dessen Chaperonaktivität.

Biosensoren sind effiziente Werkzeuge um Proteostasis in der Zelle zu untersuchen. Sie erlauben den Einfluss zellulärer Komponenten, wie Chaperone, auf Proteinfaltung und Proteinexpression zu ergründen und Proteinaggregation und Amyloidformation zu erforschen. Außerdem erlauben Biosensoren durch Selektion die Faltungseigenschaften von biotechnologisch oder pharmazeutischen Gastproteinen zu verbessern.

4 Publication list

4.1 Accepted papers

Genetic selection for enhanced folding *in vivo* targets the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor

Antioxidants & Redox Signaling 2011 Mar 23; 14(6):973-84

Linda Foit*, Antje Mueller-Schickert*, Bharath S. Mamathambika, Stefan Gleiter, Caitlyn L. Klaska, Guoping Ren, James C. A. Bardwell

*These two authors contributed equally to this work

Cytosolic Selection Systems to Study Protein Stability

Journal of Bacteriology 2014 Dec 15; 196 (24):4333-43.

Ajamaluddin Malik,* Antje Mueller-Schickert*, James C. A. Bardwell

*These two authors contributed equally to this work

Polyphosphate is a primordial chaperone

Molecular Cell 2014 Mar 20; 53(5):689-99

Michael J. Gray, Wei-Yun Wholey, Nico O. Wagner, Claudia M. Cremers, Antje Mueller-Schickert, Nathaniel T Hock, Adam G. Krieger, Erica M. Smith, Robert A. Bender, James C. A. Bardwell, Ursula Jakob

5 Personal Contribution

Genetic selection for enhanced folding *in vivo* targets the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor.

Determination of MICs of selected BPTI variants (Table 2) was performed by Antje Schickert and Bharath S. Mamathambika. Determination of MICs of folding intermediates generated by SDM was carried out by Antje Schickert (Table 2, Fig. 3 and Fig. 4A. Fig 4B). Fig. 4C was generated by Antje Schickert with Linda Foit. Determination of BPTI expression in cell extracts was performed by Antje Schickert (Fig. 5 and Fig. 6). Selection of BPTI variants was performed by Bharath S. Mamathambika (BLA system, Table 2, Fig. 2) and Stefan Gleiter (PROSIDE, Table 1). Determination of MICs of hG-CSF was performed by Bharath S. Mamathambika.

The resulting publication was written by Linda Foit, Antje Schickert and James C. A. Bardwell

Cytosolic Selection Systems to Study Protein Stability

The establishment of biosensor ANT was carried out by Antje Schickert. Antje Schickert performed the determination of permissive sites in ANT (Fig. 1 and Fig. 2 and Fig. S4) as well as the measurement of MICs of Im7 and AcP variants in ANT (Fig. 6 and Fig. S1 and Tables S2, S4 and S5), and the investigation of different permissive sites in ANT with Im7 variants (Fig. 7). Ajamaluddin Malik determined permissive sites in APH (Fig. 3, Fig. S2) and NAT (Fig. S7), measured MICs and soluble protein amounts of Im7 in APH (Fig. 4, Fig. 8 and Fig. S5 Table S6 Fig. S3, Table S2) and measured aggregation propensity in APH (Fig 5, Fig. 9 and Fig. S6).

The resulting publication was written by Antje Schickert, Ajamaluddin Malik and James C. A. Bardwell

Polyphosphate is a primordial chaperone.

Determination of biosensor stabilization by polyphosphate was performed by Antje Schickert (Fig. 4E, Fig. S2B and Fig. S2C).

Experiments shown in Fig. 1A 1B and 1E, 2A-D, 3E-H, 4A, 5A, 5B, 5D, 5E S1C, S1D S1H, S2A and Table S1 were performed by Michael J. Gray. Wei-Yun Wholey performed the experiments in Fig. 5C, S1B, S1B, S1E and S1G. Experiments in Fig 1D and S1F were performed by Claudia M. Cremers. Experiments shown in 3A, 4B, 4C and S4 were performed by Nathaniel T Hock. Experiments shown in Fig. 3B, 3C were performed by Adam G. Krieger. Nico O. Wagner performed experiments in Fig. 3D and S3

The resulting publication was written by Michael J. Grey and Ursula Jakob.

6 Introduction

6.1 Protein folding

Proteins, the most abundant and diverse macromolecules in living organisms, are the workforce of the cell. Thousands of different proteins are expressed in every cell and are involved in almost every cellular process including replication and transcription of DNA, translation of RNA, regulatory processes, catalysis, and signaling and transport (1,2).

Proteins are synthesized on ribosomes, where mRNAs are decoded into long linear amino acid chains. In order to be functional, most proteins have to fold into their native state, a three-dimensional structure which is specific for each protein. *In vitro*, most small proteins are able to refold very quickly after denaturation, independently of other molecules or energy sources. This demonstrates that the amino acid sequence itself contains all necessary information for the folding process (3).

During the folding process, an unfolded protein slides down an energy landscape in order to reach its native state, typically the most thermodynamically stable conformation (4) (Figure 1). Starting from the enormous conformational heterogeneity present in the unfolded protein, the number of native contacts in the protein increases during the folding process, while the number of accessible conformations declines, until the unique structure of the native state is reached. As the protein folds into its native state, the decline in free energy explains why most proteins fold quickly into their native structure in spite of the almost limitless number of interactions and conformations, (1,5). While the folding pathway of small proteins with fewer than 100 amino acids is frequently smooth and can often be described using two-state kinetics (6), most larger proteins fold via intermediates, represented by local minima in the energy landscape. Intermediates that can be kinetically trapped require reorganization of intramolecular interactions before folding can proceed. Due to the exposure of hydrophobic patches, these intermediates and other partially unfolded protein species are often aggregation prone. Chaperones inhibit these potentially toxic off-pathway reactions by binding to aggregation-prone species and reducing their free concentration and thus inhibiting the aggregation process. Chaperones can promote the folding of their substrate through multiple cycles of binding and release, which allow their substrates to fold.

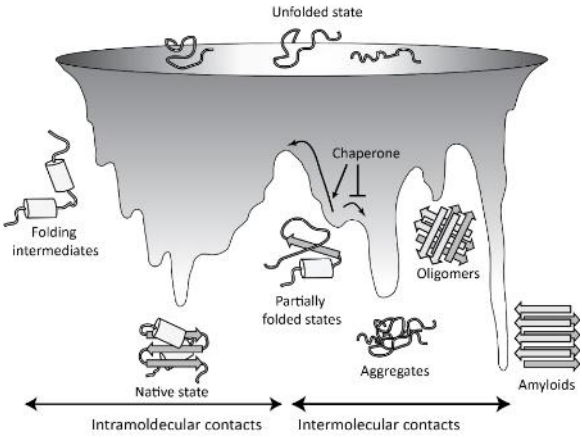


Figure 1: *protein folding in the folding funnel. The model describes protein folding on a funnel-shaped energy landscape. Unfolded proteins can occupy a multitude of different conformations. The formation of intramolecular contacts restricts the number of conformations a protein can form until finally the unique structure of the native state is reached. If an energy landscape is rugged this leads to the transient population of folding intermediates. For folding to proceed, these intermediates need to overcome free energy barriers. Intermediates and other partially unfolded states are prone to concentration-dependent aggregation. Aggregates can take on various forms including unstructured aggregates, oligomers or highly ordered amyloid fibrils. Chaperones partially prevent aggregation by binding to aggregation-sensitive folding intermediates and by maintaining a low local concentration. Chaperones can also help intermediate states pass over free energy barriers. (Figure modified from (1,5))*

6.1.1 Protein folding in the cell

The folding environment in the cell differs dramatically from the conditions present in a test tube. While folding *in vitro* is usually investigated in buffers with purified and highly diluted proteins, the environment in the cell is far more crowded and diverse. The cytosol is very viscous and protein concentrations there can reach 300-400 g/l (8). This molecular crowding leads to excluded volume effects favoring more compact folds over extended conformations and enhancing the aggregation propensity of proteins (9,10). Furthermore, protein folding *in vivo* is coupled to translation. This exposes the unfolded nascent chain to the cytosol while the rest of the protein is still being synthesized on the ribosome. Translation on polysomes further increases the local concentration of partially unfolded proteins (11) and therefore the probability of aggregation. Additionally, microorganisms are exposed to environmental stress conditions such as increased temperatures, chemicals or changing pH, all of which can induce protein unfolding (1).

Considering the complexity of the folding environment in the cell it is not surprising that most proteins *in vivo* depend on chaperones to fold into their native state (12). The cellular protein quality control system, a large network of proteins which includes chaperones and proteases, is responsible for maintaining proteostasis in the cell. Whereas chaperones assist in *de novo* folding and refolding of proteins, proteases degrade irreversibly misfolded proteins. This balance of protein synthesis, maintenance and degradation ensures that the cell functions efficiently.

Notwithstanding that the protein folding landscape *in vitro* is well understood and that many key proteins involved in the folding process have been identified, understanding how the complex cellular environment can reshape and influence those folding pathways remains a challenge (13). In recent years many promising techniques have been developed for investigating protein folding in the cell. In cell NMR (14,15), *in vivo* labeling of proteins with fluorescent dyes (16) and urea titration and temperature jump experiments (17,18) are just some examples of a steadily growing arsenal enabling one to measure protein stability in the cell. These methods are difficult to implement, however, and require not only expertise but also special equipment. Other, more generally accessible techniques, include the use of fusion proteins such as GFP as folding sensors. A test protein is fused with the sensor protein GFP and the readout of the sensor is directly linked to the stability and solubility of its fusion partner. The disadvantage of such an approach is that proteolytic digest can separate the test protein from GFP, leading to a false positive signal. This is especially problematic when the test protein is unfolded and therefore more proteolytic sensitive (19).

6.1.2 Consequences of protein misfolding in the cell

The inability of cells to maintain proper proteostasis is associated with several diseases. Cystic fibrosis, a so called loss-of-function disease, is caused by mutations that prevent a specific protein from folding correctly, which results in increased protein degradation.

Alzheimer's, Parkinson's, Huntington's and type 2 diabetes are categorized as gain-of-function diseases and are associated with the formation of amyloid fibrils. The increase in aggregation propensity of the precursor proteins can be linked to different causes, such as polyglutamine-expansions in the case of Huntington's disease, the-age related impaired function of the quality control system in the case of Alzheimer's (20) or infections with prions in the case of kuru or Creutzfeldt–Jakob disease.

Although the precursor proteins differ in sequence and structure, the amyloids which form in these diseases have common features (21). Most striking is the distinctive morphology, known as cross-β-conformation, which is present in most amyloids (22). Interestingly, fibril formation is not only

found in clinically relevant proteins, it also seems to be a common feature of many peptides and proteins (23,24).

6.2 Protein stability

6.2.1 Thermodynamic and kinetic protein stability

Considering the significant roles that they play in the living cell, it is somewhat surprising that proteins are only marginally stable. The thermodynamic stability of most proteins is only between -5 and -15 kcal/mol (25), which corresponds to approximately the energy contained in a single hydrogen bond (26). Thermodynamic stability (ΔG), is the free energy difference between the unfolded and the native state. High thermodynamic stability results in a higher concentration of folded molecules, lower thermodynamic stability shifts the equilibrium toward the unfolded conformation. Kinetic stability separates the folded state of a protein from intermediates or other partially folded species by a high free-energy barrier. This barrier ensures that the protein can remain in its functional (marginally stable) conformation.

Thermodynamic protein stability has a lower limit, ensuring that a significant proportion of a protein population is folded, functional and is able to evade degradation or aggregation (26). But the answer to the question of why proteins are so marginally stable is unclear. There are clear advantages to high stability, it has been proposed for example, that high stability confers proteins with the robustness to evolve by allowing them to accumulate beneficial mutations until their stability reaches a lower limit. (27). However there can be advantages to lower stability too, for instance, evidence suggests there may be a trade-off between protein function and stability. Enzymes need to be stable and well-folded in order to escape cellular proteases and stress conditions but they also need some flexibility in order to bind their substrate and catalyze reactions. Catalysis often requires energetically unfavorable environments, such as clusters of charged residues, which impedes stability (28,29). In such a scenario proteins with excessive stability may have their functionality affected. Yet another explanation postulates that the marginal stability of proteins is due in part to the high-dimensional space of possible sequences that a protein can form. Highly stable proteins are, according to this theory, only seldom observed because the sequences forming such proteins are extremely rare (30,31).

The stability of a protein can be altered by mutations, which can increase or decrease its stability compared to the wild-type protein. The effect a mutation has on the stability of a protein is expressed as the difference in the free energy of folding between the wild-type and the variant (ΔΔG).

Mutations can have very severe effects on protein stability, often destabilizing proteins by 0.5 – 5 kcal/mol, values that can be a considerable portion of the net free energy of folding (26). As discussed previously, many known diseases are directly caused by such destabilizing mutations that lead to protein unfolding, degradation or aggregation (32).

6.2.2 Stabilizing proteins

The insufficient stability and high aggregation propensity of many proteins is a commonly encountered practical problem that greatly impedes research and also interferes with many pharmaceutical and biotechnological applications. Frequently, medically or industrially interesting proteins are too unstable to produce and store in a cost-effective manner for successful usage as drugs. The expression of high levels of functional and stable proteins is often unsuccessful due to inefficient folding and aggregation. In order to increase protein expression, some proteins have been stabilized by rational design approaches (33). Many of the stabilizing mutations that have been identified in evolution experiments, however, are difficult to rationalize and therefore remain difficult to predict through rational design (34). Our understanding on how protein stability is determined by sequence and structure is still surprisingly limited. There is no reliable method to calculate protein stability or to accurately predict the influence of single point mutations (35). Additionally, although the increased stability and solubility of folding intermediates is often a critical step for successful protein expression, predictions only consider the thermodynamic stability of the folded protein and not the function of the protein (36). About 5% of mutations are found be stabilizing in experimental and computational approaches, but those mutations are mostly found in catalytic residues and interfere with protein function (35,36). Directed evolution, whereby an optimized protein variant is obtained via a Darwinian selection process, provides an alternative method for identifying stabilizing mutations. This method does not demand any prior knowledge about a protein's structure or function. The amount of protein variants that can be tested for specific attributes only depends on the size of the generated protein library.

6.2.3 Proteolytic protein stability

The degradation machinery of the cell recognizes partially or fully unfolded proteins. Therefore, degradation rates are mainly determined by thermodynamic protein stability (37). Misfolded proteins are also targeted by proteases and are selectively degraded (38). In eukaryotes, up to 30 % of newly synthesized proteins undergo proteolysis within several minutes, probably because of unsuccessful folding (39). Degradation is also used as a regulatory tool in processes such as cell cycle

control and signal transduction. The rapid turnover of regulatory proteins allows a fast adaption of the cell to changing environmental conditions (40).

E. coli has five ATP dependent proteolytic complexes in the cytosol (Lon, ClpAP, ClpXP, HslUV and FtsH) which differ in substrate specificity, expression pattern and localization. Each bacterial protease recognizes specific degradation tags or degrons in its target protein (41). These degrons are usually inaccessible when the protein is correctly folded. After Lon recognizes its specific degrons, a hydrophobic patch of 5-7 amino acids, it is able to unfold and degrade the whole protein in an ATP-dependent manner (42).

6.3 Biosensors for the *in vivo* folding environment

6.3.1 Protein insertion into permissive sites of a guest protein

The creation of fusion proteins is very common and is used for a great variety of applications such as protein purification and imaging, and in the development of new biopharmaceuticals (43). In most cases, fusions are made at the N- or C-terminus of a protein. Insertional fusions, in which one protein is inserted within the other protein, are much rarer and more difficult to obtain stable proteins from (44). In order to create an insertional fusion, the guest protein is inserted into a host protein, creating a tripartite fusion protein. Sites within a protein that tolerate the insertion of additional amino acids without the disruption of protein function are called permissive sites and they are quite rare (45). The majority of insertions disrupt function and or stability; indeed insertional mutations are generally considered to be null mutations. Permissive sites are mostly found on surface-exposed regions of proteins within flexible loop regions, where the interference of an inserted peptide with the fold, activity and the targeting of the host protein is minimal (46,47). Identification of permissive sites has been used to investigate membrane topology (46), structure-function relationships, and the mapping of active sites, binding sites and transmembrane domains (48,49). Insertion of short peptides into a carrier protein has been used to present foreign epitopes to the immune system, thus inducing the production of antibodies (50,51). Inserting GFP into a permissive site of a viral protein, produced a visual of the replication of Hepatitis C Virus (52)

6.3.2 Working principle of the biosensor for protein stability

The biosensor for protein stability is based on a tripartite fusion of an antibiotic resistance protein and a guest protein inserted into a permissive site (Figure 3). Linker sequences on both termini

of the guest protein the guest and the host protein enough flexibility to fold in their native conformations (53). The working principle of the biosensor is based on the observation that the proteolytic susceptibility of a protein is directly correlated to its thermodynamic stability (37). The biosensor utilizes this connection by linking proteolytic susceptibility and aggregation propensity of a guest protein to antibiotic resistance (Figure 2), thereby relating an easily observable and selectable property such as antibiotic resistance to *in vivo* protein stability.

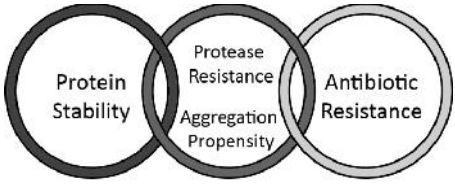


Figure 2: *The biosensors connect antibiotic resistance of cells to the guest proteins stability utilizing the guest proteins protease susceptibility or aggregation propensity*

Since the guest protein is better protected from proteolysis and aggregation (Figure 3 left side), *E. coli* expressing the biosensor with a well-folded and thermodynamically stable guest protein exhibit a high resistance towards the respective antibiotic. If the guest protein is thermodynamically unstable and unfolded it is more susceptible to proteolysis or aggregation (54), which results in decreased activity of the biosensor. *E. coli* expressing these constructs will have a lower antibiotic resistance compared to cells expressing a stable guest protein (Figure 3 right side).

The biosensors are a powerful tool that allow for us to select for conditions that improve the folding or stability of the guest protein without requiring further knowledge about the structure or function of the guest protein. Identifying stabilizing mutations is still a major challenge. Most mutations have deleterious effects on proteins, making screening processes too ineffective, and prediction programs are often too inaccurate to predict stabilizing mutations dependably (26). The ability to select for stabilized protein variants from a library of randomly mutated proteins is therefore of great value. Our biosensors allow us to select for stabilized proteins by simply demanding growth on increased antibiotic concentrations, rendering time-consuming screening processes obsolete (55,56). Enhanced folding properties can not only be achieved by mutating the guest protein itself but also by optimizing the folding environment for the guest protein. The latter can be obtained, for example, by changing the properties or availability of chaperones.

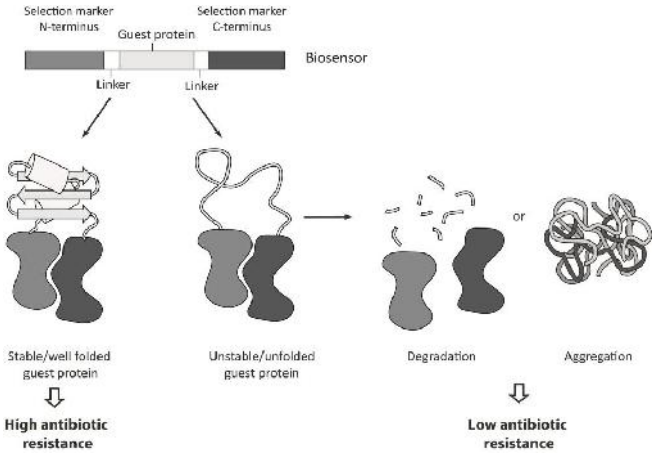


Figure 3: *Working principle of the biosensor for protein stability. A guest protein is inserted into the permissive site of an antibiotic resistance protein. The stability of this biosensor is directly dependent on the proteolytic susceptibility of the guest protein. If the guest protein is thermodynamically stable and well folded, cells expressing the biosensor will have a high antibiotic resistance. If the guest protein is unstable and unfolded it will be more vulnerable to cellular proteases (37) and the biosensor will be degraded. Cells expressing this construct will be more sensitive towards antibiotics.*

6.3.3 The β-lactamase biosensor system

β-lactam antibiotics like ampicillin and penicillin interfere with the synthesis of peptidoglycan, an essential component of the cell wall of gram-negative bacteria. Peptidoglycan, a meshwork made from peptide and glycan that can be covalently cross-linked, confers mechanic stability to the bacterial cell. By acting as pseudo-substrates and acetylating the active site of the enzyme catalyzing the crosslinking (transpeptidases), β-lactam antibiotics inhibit the crosslinking of the peptidoglycan, rendering the cell wall instable (57). The weakening of the cell wall leads to bulge formation of bacteria and can cause cell lysis. The β-lactamase biosensor is based on the antibiotic resistance protein TEM1 β-lactamase that confers resistance to β-lactam antibiotics and is commonly found in ampicillin-resistant *E. coli* strains (58,59). TEM1 β-lactamase inactivates β-lactam antibiotics by hydrolyzing the β-lactam of the antibiotic. The β-lactamase biosensor has been very successfully used to select for stabilized protein variants of Im7 and to identify a new periplasmic chaperone (55,60). The biosensor has also been shown to be able to measure the stabilizing effects of chemical chaperones on a guest protein (61).

6.3.4 The ANT biosensor system

The β-lactamase biosensor has been very useful in studying protein stability. In order to function, however, the system must be exported to the periplasm. Many cytosolic proteins contain cysteines that do not form disulfide bonds in their native conformation. Export to the periplasm would lead to the formation of non-native disulfide bonds accompanied by misfolding and very low antibiotic resistance, rendering the biosensor useless. In order to complement the β-lactamase biosensor we developed the aminoglycoside adenylyltransferase (ANT) selection system, which remains in the cytoplasm. The ANT system enables investigation of the cytoplasmic folding environment and examination of the influence of cytoplasmic chaperones, binding partners and other factors on the folding of the guest protein.

ANT is an aminoglycoside-modifying enzyme that confers resistance towards the antibiotics spectinomycin and streptomycin (62). Spectinomycin binds to the 30S subunit of the bacterial ribosomes and blocks protein biosynthesis by preventing elongation in the translocation step (63). Unlike other translation inhibitors such as streptomycin, spectinomycin does not induce misreading and is not bactericidal (64,65).

ANT adenylylates both spectinomycin and streptomycin at distinct hydroxyl groups (the 3" hydroxyl position of streptomycin and the 9 hydroxyl position of spectinomycin), thereby interfering with their binding to the ribosome (66). The protein is frequently found in isolates of gram-negative bacteria and *Staphylococcus aureus* (67–69) .

Aminoglycoside antibiotics are a large and diverse family of antibiotics that are produced by gram-positive soil actinomycetes. Spectinomycin, the first drug used to effectively treat tuberculosis (70), is still used to treat infections with *Neisseria gonorrhoeae* and in several veterinary applications (71,72). The recent and alarming increase of multi-drug-resistant bacteria has led to the reintroduction of old antibiotics such as spectinomycin, which are often administered in combination with other antibiotics in the treatment of multi-drug-resistant infections (73).

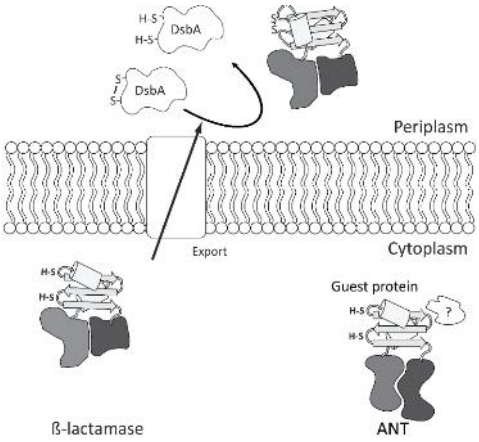


Figure 4: *Localization of the ß-lactamase and ANT based biosensors in the cell. The ANT system works in the cytosol of E. coli and allows investigation of the folding environment. Factors like chaperones, binding partners or mutations in the guest protein, that lead to a change in the guest protein's stability, can be monitored by changes in antibiotic resistance. The ß-lactamase tripartite system works in the periplasm of gram-negative bacteria. Export of the system can lead to disulfide bond formation in the oxidizing environment of the periplasm, mainly due to DsbA. Many cytosolic proteins have cysteines but do not form disulfide bonds in the cytosol. Export of these proteins could lead to non-native disulfide bond formation and misfolding of the guest protein. Therefore those proteins have to be investigated in their natural environment, the cytosol.*

7 Thesis objectives

The aim of this work was to develop and establish a cytosolic selection system for protein stability and to investigate protein folding in the cytosol and the periplasm using the ANT and β-lactamase biosensors.

The β-lactamase sensor was proposed to be used as a tool to investigate the *in vivo* folding pathway of BPTI. By selecting for BPTI variants capable of folding more efficiently than the wild-type protein in *E. coli* we hypothesized that we can identify and circumvent folding traps. The selection system can furthermore be used to compare the *in vivo* stability of different known BPTI intermediates and to assess their stability in *E. coli* by measuring their antibiotic resistance.

Fortunately, the experimental findings can be compared to *in vitro* folding, a far better understood process, thereby providing insight into the differences and similarities between *in vivo* and *in vitro* folding of this specific protein. By identifying similarities and differences between these diverse folding environments for BPTI, we might gain knowledge about the general aspects of *in vivo* and *in vitro* folding.

Another aim of this work was to establish a cytosolic biosensor based on aminoglycoside adenylyltransferase to widen the variety of test proteins that can be investigated. The β-lactamase sensor allows us to investigate the folding conditions in the periplasm but is unable to use cytosolic guest proteins because many of them contain cysteines which are improperly oxidized in the periplasmic compartment. In addition it obviously cannot be used to study the cytosolic folding environment. A cytosolic biosensor would complement the periplasmic biosensor. It would enable future investigation of the folding pathways of cytosolic proteins coupled with future examination of cytosolic folding helpers. This task involves establishing a protocol to easily identify permissive sites within antibiotic resistance markers, which is the first step in transforming them into biosensors. In order to be validated as biosensors I needed to test our constructs for their correlation between antibiotic resistance and stability of the inserted test protein. This process involves testing multiple permissive sites within the same biosensor as well as testing different test proteins to exclude the possibility that the biosensor was specific for only a single test protein.

8 Results and Discussion

8.1 Genetic selection for enhanced folding in vivo targets the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor.

Most of what we know about protein function and folding has been learned from *in vitro* experiments which are generally easier to perform and involve more precise tools than those available in the living cell. These biochemical studies are performed under readily characterized conditions involving purified components and buffers. Taking proteins out of the complex environment of the cell and studying them under simplified conditions has enabled researchers to gain valuable knowledge about biophysical and biochemical processes. Recognizing that conditions in the cell differ significantly from the conditions usually used for *in vitro* experiments, scientists have tried to mimic some of the *in vivo* conditions, such as molecular crowding, in the reaction tube, (74–76). While these *in vitro* experiments have increased our understanding of *in vivo* folding, they cannot recreate the complexity of the environment in the cell, with its multitude of interactions between chemically and structurally diverse macromolecules. Furthermore, cells contain a variety of proteins (including proteases and chaperones) known as the proteostasis network, that directly affects protein folding, stability and turnover (77,78).

The development of antibiotic resistance based biosensors has allowed us to monitor the folding status of guest proteins in the complexity of the living cell, reducing the previous need to mimic such conditions *in vitro*. The antibiotic resistance of cells expressing the biosensor is directly correlated with the *in vitro* stability of guest proteins (55,79).

E. coli is commonly used for the cost-effective production of heterologous proteins, including many of therapeutic and biotechnological importance. Posttranslational modifications, including disulfide bonds, which are essential for the structure, stability and function of proteins in which they are found (80,81), constitute one of the most critical steps in heterologous protein expression. The diverse features found in eukaryotes and prokaryotes are largely responsible for the efficiency difference in protein folding and disulfide bond formation. Eukaryotes have a specialized subcellular compartment for disulfide bond formation, the endoplasmic reticulum (ER), where protein disulfide isomerase (PDI) catalyzes the formation and rearrangement of disulfide bonds (82). *E. coli* relies on the export of proteins to the periplasm, an oxidizing environment where the Dsb system induces and isomerizes disulfide bonds (80,83). Most prokaryotic proteins that are exported to the periplasm have

fewer than four cysteines, making disulfide bond formation and isomerization relatively simple (84). However, the expression of heterologous proteins with multiple disulfides can pose a problem for *E. coli* (85).

One of the best studied model proteins for oxidative protein folding is bovine trypsin inhibitor (BPTI), a small protein which contains 58 amino acids. BPTI is stabilized by three non-consecutive disulfide bonds (86–88). The vast majority of folding studies with BPTI were performed *in vitro* where the refolding pathway from the fully reduced protein can be monitored, but much less is known about its folding inside the cell (88,89).

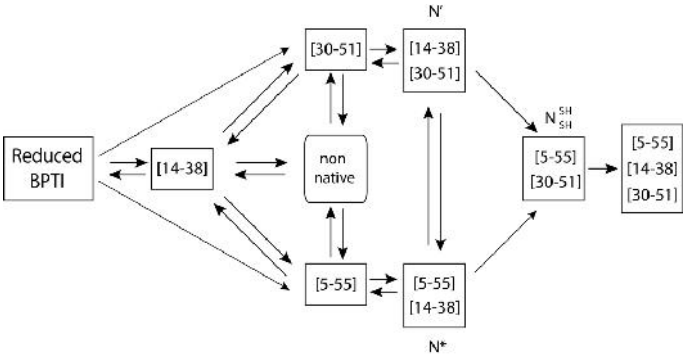


Figure 5: *In vitro folding Pathway of BPTI from the fully reduced form (left) to the native protein (right). BPTI forms one-disulfide intermediates, whereas [14-38] is more rapidly formed than [30-51] and [5-55]. Only [30-5] and [5-55] can form a second disulfide, creating the kinetically trapped intermediates N' and N*. While N* can be stable for weeks, N' slowly forms* N_{SH}^{SH} *which then rapidly forms disulfide [14-38] and folds to native BPTI. (modified from (56))*

BPTI's complex *in vitro* folding pathway involves multiple folding intermediates that have to pass through the stages of isomerization for the protein to fold into its native structure containing disulfides Cys5-55, Cys14-Cys38, and Cys30-Cys51 (**Figure 5**) (86,87). Starting from the fully reduced form, the protein rapidly forms one-disulfide intermediates BPTI [14-38], BPTI [5-55] and BPTI [30-51], whereas the formation of disulfide bond Cys14-Cys38 is the fastest (90). The disulfide Cys14-Cys38 has to be isomerized to either form disulfide bond Cys5-Cys55 or Cys30-Cys51 in order to proceed on the folding pathway, a step that might prove problematic for *E. coli's* disulfide bond machinery (91). Only the formation of the more thermodynamically stable BPTI [5-55] and BPTI [30-51] allows BPTI to form a second disulfide bond and proceed on its folding pathway (88,92). The two-disulfide intermediates BPTI [5-55; 14-38] (N*) and BPTI [14-38; 30-51] (N') are both kinetically trapped intermediates. The

isomerization of disulfide bonds to form BPTI [5-55; 30-51] (N_{SH}^{SH}) only occurs very slowly due to the inaccessibility of the free cysteines, which are buried within the structure of BPTI. N* does not undergo any significant changes and can be stable for weeks while N' slowly forms N_{SH}^{SH}, which then rapidly adds the third disulfide Cys14-Cys38 to form native BPTI (88,94).

The combination of a well-studied *in vitro* folding pathway, the readily characterized intermediates and the availability of numerous mutants makes BPTI an ideal candidate for studying and comparing *in vivo* protein folding pathways with data obtained in *in vitro* experiments. Two periplasmic biosensors were used to investigate the *in vivo* folding of BPTI. Both biosensors facilitate selection for BPTI variants with enhanced expression. Despite the high stability of BPTI, it is poorly expressed in *E. coli* (95). We reasoned then that it is therefore unlikely that expression would be improved by further stabilization of BPTI. Instead mutations might act by enhancing the folding pathway of BPTI in *E. coli,* and in doing so might help us to identify and understand problematic steps in the *in vivo* folding process. Insight obtained with this well-studied model protein could possibly be used to facilitate the folding of other proteins, even in cases where their folding pathways have not been fully investigated.

The first biosensor Stefan Gleiter, (a postdoc in our lab who proceeded me in this work) employed was a variation of the directed evolution method termed PROSIDE (96) that has previously been used to increase protein stability. In this approach, a test protein is inserted within the capsid protein g3p of the fd phage. The capsid protein is essential for phage infectivity and the rate of infectivity is directly related to the proteolytic sensitivity and stability of the inserted test protein. The second biosensor we used was the β-lactamase biosensor (BLA) developed in our lab (55). Similar to PROSIDE, this biosensor links proteolytic sensitivity of the test protein to a selectable marker, namely antibiotic resistance (Figure 3). BPTI variants with improved folding properties therefore increase plague size and phage titer (PROSIDE) or increase antibiotic resistance (BLA) compared to wild-type BPTI.

The insertion of wild-type BPTI into g3p significantly reduces the infectivity of the phage, in keeping with to the poor heterologous expression of BPTI in *E. coli* (89). In order to identify BPTI variants with improved folding properties, the gene encoding BPTI was randomly mutagenized by Stefan Gleiter. He tested the resulting BPTI library for increased phage infectivity in a growth competition selection. 42 independent g3p-BPTI variants that increased the infectivity of the phage could be identified. About half of those variants increased infectivity by deleting BPTI from the fusion.

Those variants therefore achieved infection levels comparable to wild-type g3p. The other 17 phage variants contained full-length BPTI exhibiting one or multiple mutations. In the case of single mutations (G12D and F33L) it is a straightforward matter of the mutation proving necessary and sufficient for the phenotype. Interpreting BPTI variants with multiple mutations is more complicated. 12 out of 33 single mutations came up at least twice independently. This makes it likely that the mutation improves folding and expression of BPTI rather than simply occurring in the background.

Interestingly, the selection process resulted in a vast number of BPTI variants that either disfavored the formation of the disulfide bond Cys14-Cys38 or even eliminated it completely. The four variants with the strongest phenotypes all eliminated the disulfide bond Cys14-Cys38 by mutating both cysteines involved. Seven other mutations not affecting the cysteines directly have been shown to destabilize Cys14-Cys38 (97).

The major disadvantage of the PROSIDE approach, is that the growth of the phage is coupled to the folding of the test protein, heightening the evolutionary pressure to delete the test protein. Additionally, the low growth rates of the phage make it difficult to obtain DNA in amounts necessary for sequencing and site-directed mutagenesis.

This problem was resolved by resorting to the β-lactamase biosensor developed in our laboratory (55). Growth of *E. coli* is only affected by the stability of the test protein when bacteria are under selection pressure, which reduces the risk of deletion of the test protein and simplifies cloning and sequencing of selected variants.

Bharath Mamathambika, a graduate student in our lab, inserted BPTI into a permissive site of β-lactamase, creating a tripartite fusion protein (BLA-BPTI). As with the PROSIDE approach, a randomly mutagenized BPTI library was generated. Bharat then selected for increased antibiotic resistance as compared to the BLA-BPTI wild-type. BLA-BPTI variants with increased MICs were isolated and sequenced. 13 BPTI variants with single point mutations and one variant with three mutations were subsequently identified. I determined Minimal inhibitory concentrations (MIC) and protein levels for every selected BLA-BPTI variant and then normalized to the resistance level and protein level of the BLA-BPTI wild-type protein.

I inserted four of the BPTI variants selected in PROSIDE that destabilize the Cys14-Cys38 disulfide bond along with a variant described in the literature in the β-lactamase system. I found that the degree of destabilization of disulfide bond Cys14-Cys38 correlates well with increased antibiotic resistance for the variants in the biosensor. I found that the complete elimination of disulfide bond Cys14-Cys38 consistently lead to the highest increases of MICs. I measured protein levels in western

blots of the BPTI variants that had the highest MIC. Both showed increased protein levels when expressed alone and in context of the β-lactamase fusion, confirming that the stabilization of the protein is not dependent on the presence of the fusion partner.

The biosensor proves to be a valuable tool for selection of variants with improved folding properties, and allows one to easily compare the stability of different BPTI intermediates. In order to further study the folding pathway of BPTI, I constructed the three single disulfide species of BLA-BPTI: BPTI [5-55] BPTI [14-38] and BPTI [30-51] that accumulate during the first stage of BPTIs folding pathway (Figure 5). I created the BPTI variants via site-directed mutagenesis in which the four cysteines not involved in the specific disulfide bond were replaced by serines. Each of the three variants exhibited a high correlation between antibiotic resistance and the similarity to BPTIs native structure as well as the contribution to overall protein stability of the single disulfide variants in wild-type BPTI. BPTI [14-38] is the least stable of these variants and had the lowest MIC in our experiments.

I further investigated the expression levels of these variants and found a good correlation between protein levels and antibiotic resistance. An exception was the tripartite fusion with BPTI [14-38], for which I measured protein levels comparable to BLA-BPTI wild-type even though antibiotic resistance was significantly reduced. My analysis showed that BLA-BPTI [14-38] is predominantly found in the insoluble fraction of the periplasm and that only a small fraction of the protein is periplasmic, soluble and therefore expected to be functional. The amount of functional periplasmic protein amount correlates well to the antibiotic resistance measured for this variant.

Even though the elimination of disulfide Cys14-Cys38 thermodynamically destabilizes BPTI, its destabilization increases expression levels of BPTI in *E. coli* (95). *In vitro* as well as in *E. coli* the major step limiting the rate of the folding pathway is the formation and isomerization of the Cys14-Cys38 disulfide (88).

Elimination of disulfide Cys14-Cys38 could allow BPTI to rapidly form either [5-55] or [30-51], both of which can progress rapidly on the folding pathway (Figure 6). The formation of [14-38] does not substantially stabilize BPTI (92), possibly leaving the protein in an unstructured and therefore protease- and aggregation-sensitive conformation. The early formation of the kinetically trapped intermediate [14-38] might therefore be responsible for the observed low antibiotic resistance and aggregation in the periplasm. The lack of *in vivo* enzymatic support in isomerizing [14-38] might also

help explain why its elimination increases expression significantly, as the uncatalyzed isomerization rate is expected to be slow.

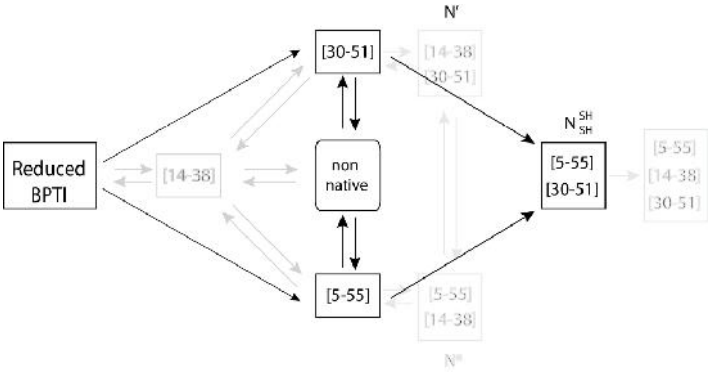


Figure 6: *Simplification of BPTI folding by elimination of Cys14-Cys38. Mutating Cys14 and Cys38 in BPTI can potentially accelerate BPTI folding. Eliminating the aggregation prone one disulfide intermediate [14-38] could accelerate the formation of the productive folding intermediates [30-52] and [5-55] which can be converted to N_{SH}^{SH} and thereby circumvent the formation of the kinetically trapped intermediates N' and N* which are potentially prone to aggregation or proteolysis.*

BPTI [5-55] and [30-51] are more stable, forming a structure more similar to the native conformation of BPTI. This might save both variants from aggregation or digestion by proteases, thereby increasing antibiotic resistance and expression levels. A larger population of the protein can efficiently fold into the native conformation and successfully prevent degradation. If this explanation is correct, the increased antibiotic resistance and elevated protein levels we observed with these constructs are a consequence of simplifying the BPTI folding pathway.

The elimination of Cys14-Cys38 could also play an essential role in speeding the later stages of the BPTI folding process (Figure 6). BPTI N* and BPTI N' are kinetically trapped intermediates that need to reduce or isomerize Cys14-Cys38 in order to proceed to the native fold via the native-like N_{SH}^{SH} (88). PDI and its less effective prokaryotic counterpart DsbC predominantly catalyzes the isomerization of Cys14-Cys38 in N' and N* (99,100) I measured MICs and protein levels for the three two-disulfide intermediates N', N* and N_{SH}^{SH}. N' showed the lowest MIC and protein levels of the three intermediates: this finding is consonant with the observation from the secondary literature that N' remains largely unfolded in *E. coli* (98). N_{SH}^{SH} on the other hand, the most stable of these two-disulfide intermediates, also has the highest antibiotic resistance and protein levels in the β-lactamase system (101).

Elimination of Cys14-Cys38 prevents the formation of the kinetically trapped intermediates N' and N* and could therefore facilitate the folding process by circumventing possibly protease sensitive or aggregation prone species.

Protein levels and antibiotic resistance levels of isolated and generated BPTI variants I measured demonstrate a high correlation, proving that biosensors a good read-out method for steady-state protein levels in the cell. The results are also prompting us to conclude that proteolysis rather than aggregation leads to the reduced antibiotic resistance of destabilized variants. The exception is BPTI [14-38] where aggregation in the periplasm and not proteolysis seems to be the cause for reduced antibiotic resistance.

The similarity between known *in vitro* kinetic traps and selected mutants circumventing these same traps *in vivo,* suggests that the *in vitro* and *in vivo* folding pathway of BPTI share significant similarities. Interestingly, although the selection did produce BPTI variants with enhanced thermodynamic stability, most of the variants identified for increased expression displayed a lower stability than wild-type BPTI. The variant seemed to predominantly circumvent kinetically trapped intermediates that are prone to proteolysis.

8.2 Cytosolic Selection Systems to Study Protein Stability

Our establishment of new biosensors expands our ability to monitor the complex folding environment in the cytosol of prokaryotes and potentially in eukaryotes. While the function of β-lactamase is dependent on export to the periplasm of gram negative prokaryotes (where it prevents destruction of the cell wall caused by β-lactam antibiotics), the new biosensors are based on antibiotic resistance proteins that work in the cytosol. We developed three biosensors based on the resistance markers nourseothricin acetyltransferase (NAT), Aminoglycoside-3-6 phosphotransferase-IIa (APH) and aminoglycoside adenylyltransferase (ANT), which confer resistance to the antibiotics nourseothricin, kanamycin/G418 and spectinomycin respectively.

The development of new biosensors based on ANT, APH and NAT creates further possibilities for investigating cytosolic proteins and growing our understanding of the folding environment in the cytoplasm. The cytosol with its multitude of chaperones is a good environment to express proteins that do not require disulfide bond formation. The three new biosensors developed for the cytoplasm not only increase the repertoire of guest proteins available, they also allow facilitate the investigation of amyloid formation and can possibly be used in eukaryotic cells.

The amino acid sequence of many proteins of biomedical or biotechnological importance cannot be altered. In order to optimize the expression of those proteins, it is necessary to improve and adapt the folding environment to the specific protein. In the future, those newly established biosensors might be used to select for optimized folding conditions for various proteins whose sequence needs to remain unaltered.

The initial step in transforming those antibiotic resistance markers into biosensors involves the identification of permissive sites. Such sites are distinguished by allowing the insertion of a guest protein without disrupting the function of the host protein. Permissive sites are mostly located within loop regions on the surface of the three-dimensional protein structure, where the interference of the guest protein with protein function and folding of the host is minimal (102). Because proteins only allow the insertion of additional amino acids in loop regions in their surface area, not in buried areas or active sites, linker-scanning methods were traditionally used to investigate protein structure and function (103,104). In the pentapetide scanning mutagenesis, a transposon is used that integrates at random positions into the sequence of a host protein, creating a library with insertions distributed within the host protein sequence (105). The removal of the transposon leaves a pentapeptide insertion behind and the library containing the pentapeptide insertions can be screened for successful

folding. Analysis of the pentapeptide position in successfully folded proteins might provide valuable insight into the structure and function of the protein.

Ajamaluddin Malik and I used this same method to identity permissive sites within all three cytosolic proteins, allowing us to create a library covering a vast amount of pentapeptide insertion site that can subsequently be screened for tolerance of protein insertions. While I developed the biosensor based on ANT, Ajamaluddin Malik developed the biosensors based on APH and NAT.

I mixed the transposon donor plasmid pGS4 with a recipient plasmid containing our antibiotic resistance protein. Transposase TnsABC integrates the transposon with very little sequence specificity into the recipient plasmid. I transformed the transposition reaction into *E. coli*, and identified successful transpositions in the antibiotic resistance marker by loss of the specific antibiotic resistance encoded by the resistance marker and gain of chloramphenicol resistance encoded by the transposon. Insertions that lead to loss of antibiotic resistance are expected to be within the promoter region or the open reading frame of the resistance marker. I evaluated the obtained library of transposon integrations by sequencing to ensure random transposon integrations throughout the open reading frame of the host protein.

The transposon is flanked by two restriction sites, either PmeI or NotI, which can be used to remove the transposon from the host protein, leaving behind a 15 bp scar sequence at the original insertion site. This scar sequence encodes a pentapeptide and contains a unique restriction site (PmeI or NotI) that allows the insertion of proteins in later steps. This method permits the generation of random pentapeptide insertion libraries, which can be used to identify permissive sites simply by testing for functional proteins by selecting for regained antibiotic resistance rendered by the host protein. Functional host proteins were subsequently sequenced in order to identify the exact position of the pentapeptide insertion. Proteins that allow the insertion of a pentapeptide in their permissive site without loss of functionality were then tested for tolerance of a full-length protein. I chose the small prokaryotic protein immunity protein 7 (Im7) to further test the pentapetide permissive sites. This helical 10 kDa protein has been shown to work very well in the β-lactamase system. Furthermore, the availability of multiple variants with known *in vitro* stabilities simplifies the verification process of the biosensors in the steps that follow (55). After insertion of Im7 with flanking glycine-serine (GS) linkers, which allow for flexibility between the fusion partners, I generated Im7 variants covering a wide range of *in vitro* stabilities.

Ajamaluddin Malik found permissive sites within APH to be located on the surface area of the protein, in same of the loop regions. Interestingly, he could not identify permissive sites in linker regions of the protein. This result underscores the difficulty of predicting the permissive sites, even if

the structure of the host protein is known. In many instances in which the structure of the protein is unknown, as is the case for ANT and NAT, adopting an experimental approach for discover permissive sites becomes inevitable.

After identification of permissive sites in APH Ajamaluddin Malik tested the correlation between *in vitro* stability of the guest proteins with the antibiotic resistance of cells expressing the biosensor. Destabilized Im7 variants in the biosensor display a firm correlation to antibiotic resistance, but variants with stabilities higher than the wild-type have similar MICs as wild-type Im7. This could be due to a saturation effect whereby the linear correlation is lost and further stabilization of the host protein does not translate into increased antibiotic resistance. The reduced antibiotic resistance in destabilized Im7 variants is apparently due to increased aggregation rather than proteolysis of the biosensor, since the total protein amount of biosensor is constant for all tested Im7 variants. Ajamaluddin Malik could show that antibiotic resistance as well as soluble protein are inversely correlated with increasing polyQ chain length in the biosensor, suggesting that the biosensor can be used as a sensor for protein aggregation and amyloid formation.

The low sensitivity for stabilized protein variants in the APH sensor prompted us to develop a second biosensor. I chose ANT, which confers resistance to spectinomycin and streptomycin as host protein. I identified permissive sites following the same method previously developed for APH. I was able to identify pentapeptide permissive sites in three hotspots (at the N terminus, the middle and the C terminus of ANT) and tested those pentapeptide permissive sites for tolerance of a full length protein Im7. I found the distribution of Im7 permissive sites followed a pattern similar to that seen in the pentapeptide permissive sites. This suggests that many (but not all) sites that tolerate the insertion of short fragments also permit the insertion of a longer protein. I tested a library of Im7 insertion in ANT for spectinomycin resistance and identified two subgroups. One group exhibited a very high spectinomycin resistance level, comparable to ANT without Im7 insertion. Sequencing revealed that all clones had in-frame insertions of Im7 in the N terminus of ANT. I found the second subgroup exhibited lower resistance levels. Insertions were in-frame and were located in the middle and the C terminus of ANT. Additionally I found Im7 insertions in the N terminus of ANT, which lead to frame sifts. After concluding that the N terminus of ANT is likely not essential for function, I concentrated instead on permissive sites in the middle and in the C terminus of ANT.

The optimal biosensor yields a linear correlation over a wide range of antibiotic concentrations and stabilities of the guest protein. I measured the antibiotic resistance of 13 Im7 variants that cover a wide range of stabilities in the biosensor (insertion behind amino acid 155), and found that they

exhibit a linear correlation with their *in vitro* stabilities. Although stabilized variants cluster more than destabilized variants, the ANT biosensor yields a much broader range than the APH-based biosensor.

In order to prove the versatility of the ANT biosensor, I switched the guest protein and tested the sensitivity range with variants of the small eukaryotic enzyme human muscle acylphosphatase (AcP). I found the biosensor shows a high correlation between *in vitro* stability and antibiotic resistance as seen before for Im7, demonstrating that the system is not limited to the Im7 with which it was established.

To further evaluate the influence of the exact location of the insertion, I tested three additional permissive sites in ANT. Two insertion sites in the middle of the protein (behind amino acid 153 and 157) and one in the C terminus (behind amino acid 289) were evaluated following the procedure previously used for permissive site 155 before. I inserted five Im7 variants that cover a broad range of *in vitro* stabilities. Although the sensitivity varies slightly between different permissive sites, I found all four permissive sites show a linear correlation. This result suggests that the basic functional mechanism is mostly independent of the exact location of the permissive site.

APH not only confers resistance to kanamycin in *E. coli* but also to G418 in *S. cerevisiae*. In an attempt to further broaden the biosensor approach to eukaryotes, Ajamaluddin Malik tested the antibiotic resistance and protein levels of six Im7 variants. As seen in *E. coli* we observed a correlation between Im7 stability and both antibiotic resistance and soluble protein amounts, while the total protein amount of the biosensor remained constant. This result suggests that low antibiotic resistance is a consequence of aggregation and not proteolysis. Ajamaluddin Malik found the same inverse correlation between polyQ chain lengths and antibiotic resistance as well as soluble biosensor protein as seen in *E. coli*.

Adopting the linker scanning approach used for APH and ANT, Ajamaluddin Malik was able to identify two permissive sites in NAT. But a stabilized Im7 variant failed to further increase antibiotic resistance, leading to the assumption that wild-type Im7 might represent the upper limit of the assay in NAT.

8.3 Polyphosphate is a primordial chaperone.

It is generally accepted that life on earth has developed from an RNA world, where RNA molecules stored genetic information and catalyzed enzyme reactions at the same time. Those tasks were eventually assumed by DNA and proteins, leading to the formation of life as we know it today. But once proteins assumed a workhorse role within the cell a proteostasis network also became necessary. Chaperones and proteases had to help proteins to fold in their three-dimensional structure and withstand stress and recycle unwanted proteins. Inorganic polyphosphate (polyP) could be a member of this ancient proteostasis network. PolyP is a simple molecule formed by phosphor anhydride bonded inorganic polyphosphate which is highly conserved among organisms.

Hypochlorous acid (HOCL) stress induces the formation and accumulation of polyP in *E. coli* and *V. cholearae* and polyP increases resistance of *E. coli* and *V. cholerae* towards HOCL stress. Cells that lack polyphosphate kinase (Δppk), the enzyme that generates polyP from inorganic phosphate, are much more sensitive to HOCL treatment than wild-type strains. Strains lacking exopolyphosphatase (Δppx), the enzyme that catalyzes the breakdown of polyP to inorganic phosphate, accumulate more polyP after HOCL treatment and are more resistant than wild-type cells.

ATP levels in HOCL stressed organisms decline rapidly but a strain that is unable to synthesize polyP (Δppk) retains most of its ATP following HOCL stress. These results suggest that declining ATP levels are indeed a consequence of polyP synthesis, and that they do not result involuntarily from HOCL stress and the blocking of glycolysis, but rather act as a protection mechanism against HOCL.

The protective effect of polyP in heat shock conditions was tested in an *rpoH*-deficient *E. coli* strain. This strain lacks most protein chaperones and is incapable of inducting a heat shock response, thus rendering the strain very sensitive to heat shock. PolyP reduced the level of aggregated proteins and increased heat resistance in the *rpoH* strain in a similar manner to the protein chaperones DnaK/J. This demonstrates that polyP chaperone activity against aggregation prone proteins is not HOCL specific.

In order to test the chaperone function of polyP *in vivo*, wild-type *E. coli* and a Δppk mutant were exposed to HOCL stress. The expression of heat shock genes, induced by unfolded proteins, was determined. Interestingly, the expression of heat shock genes including Hsp33 and DnaK was much higher in cells lacking PPK. The amount of aggregation inversely correlated with the amount of polyP in the strains, suggesting that polyP is able to prevent aggregation of unfolded proteins. This could

indicate that polyP works as a general chaperone preventing proteins from unfolding and shielding them from aggregation.

The biosensors developed in our lab was used to investigate the ability of polyP to stabilize proteins even under non-stress conditions. Wild-type *E. coli*, *Δppk*, and Δppx p*ppk* were used to create a variety of different *in vivo* polyP levels. I expressed the biosensors with ANT, β-lactamase and chloramphenicol acetyltransferase (CAT) in these three strains and determined antibiotic resistance to all three antibiotics. PPK over expressing strains that have the largest amount of intrinsic polyP also yielded the highest resistance against spectinomycin. Wild-type *E. coli* demonstrated intermediate resistance against spectinomycin and strains lacking polyP (Δppk) yielded the lowest resistance.

The polyP-deficient strain only showed a slight decrease in spectinomycin resistance in the absence of ANT. The observed increase in antibiotic resistance in cells containing polyP is therefore likely due to protein stabilization caused by polyP. I found the same correlation between antibiotic resistance and polyP levels for CAT, but not for β-lactamase which is exported to the periplasm of *E. coli* and hence is not influenced by the cytoplasmic polyP.

Michael Grey showed further in *in vitro* experiments with purified polyP that the molecule is able to inhibit the aggregation of chemically and thermally denatured luciferase, a standard assay for chaperone activity. This demonstrations that the chaperone activity can indeed be attributed to polyP.

If polyP is incubated with luciferase during the inactivation process, luciferase can be successfully refolded by DnaK/DnaJ/GrpE, suggesting that polyP is capable of keeping the unfolded protein in a conformation that can be refolded. The unfolding of luciferase is irreversible if luciferase is inactivated and aggregates in the absence of polyP. PolyP binds to unfolded proteins in an ATP independent manner, rescuing proteins from aggregation and supporting their refolding once stress conditions have passed.

Michael Grey showed that the protective abilities of polyP are correlated with chain length. Although even shorter polyP chains in the range of 14-60 mers were able to protect proteins from aggregating, longer polyP chains with 200-1300 mers were clearly more effective. This might answer the question as to why *E. coli* prefers to accumulate longer polyP chains in the face of stressful conditions.

Experiments show that PPX is a redox-regulated enzyme which is reversibly oxidized by HOCL and thereby rendered transiently inactive. This transient inactivation supports the fast accumulation

of polyP during stress conditions. *In vivo* experiments demonstrate that the polyP accumulation in wild-type cell levels off 30 min after HOCL treatment, while the accumulation in a Δppx strain continues. The reactivated PPX in wild-type *E. coli* can disassemble polyP and replenish cellular ATP and P_i levels. PolyP can be formed rapidly following HOCL stress, when most protein chaperones are inactivated. Therefore polyP could serve as an ATP storage molecule until more favorable growth conditions are reestablished. The depletion of ATP during HOCL stress could prevent the unnecessary use of energy for protein synthesis and chaperone cycles until oxidative stress conditions have ceased and those processes can be performed more efficiently. When stress conditions have passed, PPX becomes reactivated and degrades polyP. The generated ATP is needed by protein chaperones to refold proteins bound by polyP and for *de novo* protein synthesis.

9 Summary

A large number of diseases directly result from protein misfolding, degradation, aggregation or amyloid formation (106). Understanding protein folding and proteostasis is imperative if we want to prevent or reverse such diseases. Additionally, effective heterologous protein expression plays an important role in biotechnological and biomedical approaches. Studying and optimizing protein folding in the cell remains a challenging task, however, due to the complex environment of the living cell (77).

The biosensors constitute a valuable tool for investigating the *in vivo* folding environment of *E. coli*. The sensors not only allow us to assess and compare *in vivo* stability of different protein variants; they also permit the selection of protein variants with enhanced expression. Prior knowledge of the structure or function of the guest proteins is unnecessary. The selection process uniquely allows us to test whole libraries of protein variants and to allow *E. coli* to evolve the protein to improve its expression by increasing the thermodynamic stability of the protein, reducing its aggregation propensity or its proteolytic susceptibility.

The system permits easy use and is operational in commonly used *E. coli* strains and plasmids. The screen for permissive sites is very straightforward and allows the development of new biosensors based on new antibiotic resistance proteins or other markers. The cytosolic biosensors we developed greatly increase the arsenal of proteins that can be investigated. The new sensors have been shown to not only link antibiotic resistance to protein stability but also serve as a readout for protein aggregation and amyloid formation. Additionally, the developed biosensor offers us the chance of further broadening the approach by investigating the complex folding environment in eukaryotes.

The biosensors can also be used to change and optimize the folding environment of the cell. Especially when it is essential not to change the sequence of guest protein itself, the cell itself can be tuned to express a desired protein more efficiently. This could be a novel option for pharmaceutical proteins in particular. The strong correlation between *in vitro* stability and *in vivo* antibiotic resistance found in all biosensors developed in our lab could indicate that *in vitro* stability is the dominant factor for *in vivo* stability. Low thermodynamic stability of a protein correlates to the time the protein spends in a more unfolded confirmation, which in general is also more proteolytic sensitive. But factors such as aggregation propensity also play a major role in determining the antibiotic resistance, as was

observed when aggregation-sensitive polyQ stretches were inserted in the APH biosensor (79). The biosensors can be used as a readout system for aggregation and possibly also for amyloid formation.

Interestingly, while selecting for stabilized variants of human muscle acylphosphatase with the ANT system (data not shown), we identified multiple variants that had increased MICs compared to the wild-type protein. The increased MIC could be confirmed with the β-lactamase system, but when those variants were purified and their thermodynamic stabilities were measured, they did not show any stabilization compared to the wild-type protein (unpublished data). It is very likely that those proteins could be stabilized *in vivo* but not under *in vitro* conditions. It is conceivable that the mutation increased the ability of the protein to interact with chaperones or decreased the protein's susceptibility for aggregation or proteolysis in the cell without altering its stability (107,108). The lack of adequate antibodies made it very difficult to follow up on those results.

These results nevertheless demonstrate that the *in vivo* environment could influence protein stability and proteostasis in ways that are not easily accessible in the reaction tube. One of the advantages of a selection system is to permit *E. coli* the opportunity to develop a way of improving protein stability. By investigating those solutions we may learn a lot about the cellular folding processes.

10 References

1. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011 Jul 21;475(7356):324–32.
2. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. The Molecules of Life. 2000 [cited 2014 Oct 24]; Available from: http://www.ncbi.nlm.nih.gov/books/NBK21473/
3. Anfinsen CB. Principles that Govern the Folding of Protein Chains. Science. 1973 Jul 20;181(4096):223–30.
4. Dobson CM, Šali A, Karplus M. Protein Folding: A Perspective from Theory and Experiment. Angew Chem Int Ed. 1998;37(7):868–93.
5. Clark PL. Protein folding in the cell: reshaping the folding funnel. Trends Biochem Sci. 2004 Oct;29(10):527–34.
6. Jackson SE. How do small single-domain proteins fold? Fold Des. 1998 Aug;3(4):R81–91.
7. Jahn TR, Radford SE. The Yin and Yang of protein folding. FEBS J. 2005;272(23):5962–70.
8. Zimmerman SB, Trach SO. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J Mol Biol. 1991 Dec 5;222(3):599–620.
9. Van den Berg B, Ellis RJ, Dobson CM. Effects of macromolecular crowding on protein folding and aggregation. EMBO J. 1999 Dec 15;18(24):6927–33.
10. Zhou H-X. Polymer crowders and protein crowders act similarly on protein folding stability. FEBS Lett. 2013 Mar 1;587(5):394–7.
11. Gershenson A, Gierasch LM. Protein folding in the cell: challenges and progress. Curr Opin Struct Biol. 2011 Feb;21(1):32–41.
12. Ellis RJ, Minton AP. Protein aggregation in crowded environments. Biol Chem. 2006 May 1;387(5):485–97.
13. Bartlett AI, Radford SE. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. Nat Struct Mol Biol. 2009 Jun;16(6):582–8.
14. Burz DS, Dutta K, Cowburn D, Shekhtman A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). Nat Methods. 2006 Feb;3(2):91–3.
15. Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dötsch V. High-Resolution Macromolecular NMR Spectroscopy Inside Living Cells. J Am Chem Soc. 2001 Mar 1;123(10):2446–7.
16. Ignatova Z, Gierasch LM. A method for direct measurement of protein stability in vivo. Methods Mol Biol Clifton NJ. 2009;490:165–78.
17. Ebbinghaus S, Dhar A, McDonald JD, Gruebele M. Protein folding stability and dynamics imaged in a living cell. Nat Methods. 2010 Apr;7(4):319–23.
18. Ignatova Z, Krishnan B, Bombardier JP, Marcelino AMC, Hong J, Gierasch LM. From the test tube to the cell: Exploring the folding and aggregation of a β-clam protein. Pept Sci. 2007;88(2):157–63.
19. Mayer S, Rüdiger S, Ang HC, Joerger AC, Fersht AR. Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro. J Mol Biol. 2007 Sep 7;372(1):268–76.
20. Ben-Zvi A, Miller EA, Morimoto RI. Collapse of proteostasis represents an early molecular event in Caenorhabditis elegans aging. Proc Natl Acad Sci. 2009 Sep 1;106(35):14914–9.
21. Dobson CM. The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci. 2001 Feb 28;356(1406):133–45.
22. Sunde M, Blake C. The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. In: Frederic M. Richards DSE and PSK, editor. Advances in Protein Chemistry [Internet]. Academic Press; 1997 [cited 2013 Aug 14]. p. 123–59. Available from: http://www.sciencedirect.com/science/article/pii/S0065323308603204

23. Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, et al. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci. 1999 Mar 30;96(7):3590–4.
24. Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM. Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci. 1998 Apr 14;95(8):4224–8.
25. Magliery TJ, Lavinder JJ, Sullivan BJ. Protein stability by number: high-throughput and statistical approaches to one of protein science's most difficult problems. Curr Opin Chem Biol. 2011 Jun;15(3):443–51.
26. DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet. 2005 Sep;6(9):678–87.
27. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci. 2006 Apr 11;103(15):5869–74.
28. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science. 1995 Jan 20;267(5196):383–6.
29. Zhu ZY, Karlin S. Clusters of charged residues in protein three-dimensional structures. Proc Natl Acad Sci. 1996 Aug 6;93(16):8350–5.
30. Goldstein RA. The evolution and evolutionary consequences of marginal thermostability in proteins. Proteins Struct Funct Bioinforma. 2011 May 1;79(5):1396–407.
31. Taverna DM, Goldstein RA. Why are proteins marginally stable? Proteins Struct Funct Bioinforma. 2002 Jan 1;46(1):105–9.
32. Yue P, Li Z, Moult J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. J Mol Biol. 2005 Oct 21;353(2):459–73.
33. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of Protein Pharmaceuticals: An Update. Pharm Res. 2010 Apr 1;27(4):544–75.
34. Eijsink VGH, Bjørk A, Gåseidnes S, Sirevåg R, Synstad B, Burg B van den, et al. Rational engineering of enzyme stability. J Biotechnol. 2004 Sep 30;113(1–3):105–20.
35. Roodveldt C, Aharoni A, Tawfik DS. Directed evolution of proteins for heterologous expression and stability. Curr Opin Struct Biol. 2005 Feb;15(1):50–6.
36. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol. 2009 Oct;19(5):596–604.
37. Parsell DA, Sauer RT. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. J Biol Chem. 1989 May 5;264(13):7590–5.
38. Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. Nature. 2003 Dec 18;426(6968):895–9.
39. Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature. 2000 Apr 13;404(6779):770–4.
40. Ciechanover A, Orian A, Schwartz AL. Ubiquitin-mediated proteolysis: biological regulation via destruction. BioEssays News Rev Mol Cell Dev Biol. 2000 May;22(5):442–51.
41. Baker TA, Sauer RT. ATP-dependent proteases of bacteria: recognition logic and operating principles. Trends Biochem Sci. 2006 Dec;31(12):647–53.
42. Baker TA, Sauer RT. ATP-dependent proteases of bacteria: recognition logic and operating principles. Trends Biochem Sci. 2006 Dec;31(12):647–53.
43. Schmidt SR. Fusion-proteins as biopharmaceuticals--applications and challenges. Curr Opin Drug Discov Devel. 2009 Mar;12(2):284–95.
44. Doi Y, Arakawa Y. 16S Ribosomal RNA Methylation: Emerging Resistance Mechanism against Aminoglycosides. Clin Infect Dis. 2007 Jul 1;45(1):88–94.
45. Hofnung M, Bedouelle H, Boulain JC, Clement JM, Charbit A, Duplay P, et al. Genetic approaches to the study and use of proteins: Random point mutations and random linker insertions. Bull Inst Pasteur. 1988;8(1):95–101.

46. Charbit A, Ronco J, Michel V, Werts C, Hofnung M. Permissive sites and topology of an outer membrane protein with a reporter epitope. J Bacteriol. 1991 Jan 1;173(1):262–75.
47. Manoil C, Bailey J. A simple screen for permissive sites in proteins: analysis of Escherichia coli lac permease. J Mol Biol. 1997 Mar 28;267(2):250–63.
48. Anton BP, Raleigh EA. Transposon-Mediated Linker Insertion Scanning Mutagenesis of the Escherichia coli McrA Endonuclease. J Bacteriol. 2004 Sep;186(17):5699–707.
49. Kilburn L, Poole K, Meyer J-M, Neshat S. Insertion Mutagenesis of the Ferric Pyoverdine Receptor FpvA ofPseudomonas aeruginosa: Identification of Permissive Sites and a Region Important for Ligand Binding. J Bacteriol. 1998 Dec 15;180(24):6753–6.
50. Coëffier E, Clément J-M, Cussac V, Khodaei-Boorane N, Jehanno M, Rojas M, et al. Antigenicity and immunogenicity of the HIV-1 gp41 epitope ELDKWA inserted into permissive sites of the MalE protein. Vaccine. 2000 Nov 22;19(7–8):684–93.
51. Martineau P, Guillet J-G, Leclerc C, Hofnung M. Expression of heterologous peptides at two permissive sites of the MalE protein: antigenicity and immunogenicity of foreign B-cell and T-cell epitopes. Gene. 1992 Apr 1;113(1):35–46.
52. Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, Goff SP, et al. Insertion of Green Fluorescent Protein into Nonstructural Protein 5A Allows Direct Visualization of Functional Hepatitis C Virus Replication Complexes. J Virol. 2004 Jul 15;78(14):7400–9.
53. Chen X, Zaro JL, Shen W-C. Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev. 2013 Oct 15;65(10):1357–69.
54. Tomoyasu T, Mogk A, Langen H, Goloubinoff P, Bukau B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. Mol Microbiol. 2001;40(2):397–413.
55. Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, et al. Optimizing Protein Stability In Vivo. Mol Cell. 2009 Dec 11;36(5):861–71.
56. Foit L, Mueller-Schickert A, Mamathambika BS, Gleiter S, Klaska CL, Ren G, et al. Genetic Selection for Enhanced Folding In Vivo Targets the Cys14-Cys38 Disulfide Bond in Bovine Pancreatic Trypsin Inhibitor. Antioxid Redox Signal. 2010 Nov 29;14(6):973–84.
57. Walsh C. Molecular mechanisms that confer antibacterial drug resistance. Nature. 2000 Aug 17;406(6797):775–81.
58. Cooksey R, Swenson J, Clark N, Gay E, Thornsberry C. Patterns and mechanisms of beta-lactam resistance among isolates of Escherichia coli from hospitals in the United States. Antimicrob Agents Chemother. 1990 May;34(5):739–45.
59. Salverda MLM, De Visser JAGM, Barlow M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiol Rev. 2010 Nov;34(6):1015–36.
60. Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, Pfizenmaier J, et al. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nat Struct Mol Biol. 2011 Mar;18(3):262–9.
61. Hailu TT, Foit L, Bardwell JCA. In vivo detection and quantification of chemicals that enhance protein stability. Anal Biochem. 2013 Mar 1;434(1):181–6.
62. Shaw KJ, Rather PN, Hare RS, Miller GH. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. Microbiol Rev. 1993 Mar;57(1):138.
63. Bilgin N, Richter AA, Ehrenberg M, Dahlberg AE, Kurland CG. Ribosomal RNA and protein mutants resistant to spectinomycin. EMBO J. 1990 Mar;9(3):735.
64. Anderson P, Davies J, Davis BD. Effect of spectinomycin on polypeptide synthesis in extracts of Escherichia coli. J Mol Biol. 1967 Oct 14;29(1):203–15.
65. Davies J, Anderson P, Davis BD. Inhibition of Protein Synthesis by Spectinomycin. Science. 1965 Sep 3;149(3688):1096–8.

66. Hollingshead S, Vapnek D. Nucleotide sequence analysis of a gene encoding a streptomycin/spectinomycin adenyltransferase. Plasmid. 1985 Jan;13(1):17–30.
67. Clark NC, Olsvik Ø, Swenson JM, Spiegel CA, Tenover FC. Detection of a Streptomycin/Spectinomycin Adenylyltransferase Gene (aadA) in Enterococcus faecalis. Antimicrob Agents Chemother. 1999 Jan 1;43(1):157–60.
68. Courvalin P, Fiandt M. Aminoglycoside-modifying enzymes of Staphylococcus aureus; Experssion in Escherichia coli. Gene. 1990 May;9(3–4):247–69.
69. Shaw KJ, Hare RS, Sabatelli FJ, Rizzo M, Cramer CA, Naples L, et al. Correlation between aminoglycoside resistance profiles and DNA hybridization of clinical isolates. Antimicrob Agents Chemother. 1991 Nov 1;35(11):2253–61.
70. Schatz A, Bugle E, Waksman SA. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*†. Exp Biol Med. 1944 Jan 1;55(1):66–9.
71. Schwarz S, Chaslus-Dancla E. Use of antimicrobials in veterinary medicine and mechanisms of resistance. Vet Res. 2001 May;32(3/4):201–25.
72. Teuber M. Veterinary use and antibiotic resistance. Curr Opin Microbiol. 2001 Oct 1;4(5):493–9.
73. Falagas ME, Grammatikos AP, Michalopoulos A. Potential of old-generation antibiotics to address current need for new antibiotics. Expert Rev Anti Infect Ther. 2008 Oct;6(5):593–600.
74. Álamo M del, Rivas G, Mateu MG. Effect of Macromolecular Crowding Agents on Human Immunodeficiency Virus Type 1 Capsid Protein Assembly In Vitro. J Virol. 2005 Nov 15;79(22):14271–81.
75. Chebotareva NA, Kurganov BI, Livanova NB. Biochemical effects of molecular crowding. Biochem Mosc. 2004 Nov 1;69(11):1239–51.
76. Zhou H-X, Rivas G, Minton AP. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences *. Annu Rev Biophys. 2008 Jun;37(1):375–97.
77. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol. 2009 Jun;16(6):574–81.
78. Phillip Y, Schreiber G. Formation of protein complexes in crowded environments – From in vitro to in vivo. FEBS Lett. 2013 Apr 17;587(8):1046–52.
79. Malik A, Mueller-Schickert A, Bardwell JCA. Cytosolic Selection Systems to Study Protein Stability. J Bacteriol. 2014 Sep 29;JB.02215–14.
80. Bardwell JCA, McGovern K, Beckwith J. Identification of a protein required for disulfide bond formation in vivo. Cell. 1991 Nov 1;67(3):581–9.
81. Darby N, Creighton TE. Probing protein folding and stability using disulfide bonds. Mol Biotechnol. 1997 Feb 1;7(1):57–77.
82. Laboissière MCA, Sturley SL, Raines RT. The Essential Function of Protein-disulfide Isomerase Is to Unscramble Non-native Disulfide Bonds. J Biol Chem. 1995 Nov 24;270(47):28006–9.
83. Marco A de. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb Cell Factories. 2009 May 14;8(1):26.
84. Dutton RJ, 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. 2008 Aug 19;105(33):11933–8.
85. Herrmann JM, Riemer J. Three Approaches to One Problem: Protein Folding in the Periplasm, the Endoplasmic Reticulum, and the Intermembrane Space. Antioxid Redox Signal. 2014 Feb 2;21(3):438–56.
86. Darby NJ, Morin PE, Talbo G, Creighton TE. Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. J Mol Biol. 1995;249(2):463–77.
87. Goldenberg DP. Native and non-native intermediates in the BPTI folding pathway. Trends Biochem Sci. 1992 Jul;17(7):257–61.

88. Weissman JS, Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. Science. 1991 Sep 20;253(5026):1386–93.
89. Ostermeier M, Georgiou G. The folding of bovine pancreatic trypsin inhibitor in the Escherichia coli periplasm. J Biol Chem. 1994 Aug 19;269(33):21072–7.
90. Dadlez M, Kim PS. Rapid formation of the native 14-38 disulfide bond in the early stages of BPTI folding. Biochemistry (Mosc). 1996 Dec 17;35(50):16153–64.
91. Hiniker A, Collet J-F, Bardwell JCA. Copper stress causes an in vivo requirement for the Escherichia coli disulfide isomerase DsbC. J Biol Chem. 2005 Oct 7;280(40):33785–91.
92. Bulaj G, Goldenberg DP. Early events in the disulfide-coupled folding of BPTI. Protein Sci Publ Protein Soc. 1999 Sep;8(9):1825–42.
93. Creighton TE, Goldenberg DP. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. J Mol Biol. 1984 Nov 5;179(3):497–526.
94. Mamathambika BS, Bardwell JC. Disulfide-Linked Protein Folding Pathways. Annu Rev Cell Dev Biol. 2008;24(1):211–35.
95. Nilsson B, Berman-Marks C, Kuntz ID, Anderson S. Secretion incompetence of bovine pancreatic trypsin inhibitor expressed in Escherichia coli. J Biol Chem. 1991 Feb 15;266(5):2970–7.
96. Martin A, Schmid FX, Sieber V. Proside: a phage-based method for selecting thermostable proteins. Methods Mol Biol Clifton NJ. 2003;230:57–70.
97. Goldenberg DP, Bekeart LS, Laheru DA, Zhou JD. Probing the determinants of disulfide stability in native pancreatic trypsin inhibitor. Biochemistry (Mosc). 1993 Mar 1;32(11):2835–44.
98. Creighton TE. Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. J Mol Biol. 1977 Jun 25;113(2):275–93.
99. Weissman JS, Kimt PS. Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. Nature. 1993 Sep 9;365(6442):185–8.
100. Zapun A, Missiakas D, Raina S, Creighton TE. Structural and functional characterization of DsbC, a protein involved in disulfide bond formation in Escherichia coli. Biochemistry (Mosc). 1995 Apr 18;34(15):5075–89.
101. Wagner G, Kalb (gilboa) AJ, Wüthrich K. Conformational Studies by 1H Nuclear Magnetic Resonance of the Basic Pancreatic Trypsin Inhibitor after Reduction of the Disulfide Bond between Cys-14 and Cys-38. Eur J Biochem. 1979 Apr 1;95(2):249–53.
102. Collinet B, Hervé M, Pecorari F, Minard P, Eder O, Desmadril M. Functionally Accepted Insertions of Proteins within Protein Domains. J Biol Chem. 2000 Jun 9;275(23):17428–33.
103. Haller AA, Semler BL. Linker scanning mutagenesis of the internal ribosome entry site of poliovirus RNA. J Virol. 1992 Aug 1;66(8):5075–86.
104. Wang S, Deng L, Brown ML, Agellon LB, Tall AR. Structure-function studies of human cholesteryl ester transfer protein by linker insertion scanning mutagenesis. Biochemistry (Mosc). 1991 Apr 1;30(14):3484–90.
105. Hallet B, Sherratt DJ, Hayes F. Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. Nucleic Acids Res. 1997 May 1;25(9):1866–7.
106. Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. Annu Rev Biochem. 2006;75(1):333–66.
107. Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, et al. Kinetic partitioning of protein folding and aggregation. Nat Struct Mol Biol. 2002 Feb;9(2):137–43.
108. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide andprotein aggregation rates. Nature. 2003 Aug 14;424(6950):805–8.

11 Appendix

11.1 Attached accepted papers

Genetic selection for enhanced folding *in vivo* targets the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor

Antioxidants & Redox Signaling 2011 Mar 23; 14(6):973-84

Linda Foit*, Antje Mueller-Schickert*, Bharath S. Mamathambika, Stefan Gleiter, Caitlyn L. Klaska, Guoping Ren, James C. A. Bardwell

*These two authors contributed equally to this work

Cytosolic Selection Systems to Study Protein Stability

Journal of Bacteriology 2014 Dec 15; 196 (24):4333-43.

Ajamaluddin Malik,* Antje Mueller-Schickert*, James C. A. Bardwell

*These two authors contributed equally to this work

Polyphosphate is a primordial chaperone

Molecular Cell 2014 Mar 20; 53(5):689-99

Michael J. Gray, Wei-Yun Wholey, Nico O. Wagner, Claudia M. Cremers, Antje Mueller-Schickert, Nathaniel T Hock, Adam G. Krieger, Erica M. Smith, Robert A. Bender, James C. A. Bardwell, Ursula Jakob

ANTIOXIDANTS & REDOX SIGNALING
Volume 14, Number 6, 2011
© Mary Ann Liebert, Inc.
DOI: 10.1089/ars.2010.3712

ORIGINAL RESEARCH COMMUNICATION

Genetic Selection for Enhanced Folding *In Vivo* Targets the Cys14-Cys38 Disulfide Bond in Bovine Pancreatic Trypsin Inhibitor

Linda Foit,* Antje Mueller-Schickert,* Bharath S. Mamathambika, Stefan Gleiter, Caitlyn L. Klaska, Guoping Ren, and James C.A. Bardwell

Abstract

The periplasm provides a strongly oxidizing environment; however, periplasmic expression of proteins with disulfide bonds is often inefficient. Here, we used two different tripartite fusion systems to perform *in vivo* selections for mutants of the model protein bovine pancreatic trypsin inhibitor (BPTI) with the aim of enhancing its expression in *Escherichia coli*. This trypsin inhibitor contains three disulfides that contribute to its extreme stability and protease resistance. The mutants we isolated for increased expression appear to act by eliminating or destabilizing the Cys14-Cys38 disulfide in BPTI. In doing so, they are expected to reduce or eliminate kinetic traps that exist within the well characterized *in vitro* folding pathway of BPTI. These results suggest that elimination or destabilization of a disulfide bond whose formation is problematic *in vitro* can enhance *in vivo* protein folding. The use of these *in vivo* selections may prove a valuable way to identify and eliminate disulfides and other rate-limiting steps in the folding of proteins, including those proteins whose *in vitro* folding pathways are unknown. *Antioxid. Redox Signal.* 14, 973–984.

Introduction

DISULFIDE BONDS are critical posttranslational modifications that regulate the folding and stability of proteins. The activities of many extracellular and pharmaceutically relevant proteins such as insulin and the tissue plasminogen activator rely on the formation of their correct disulfide bond pattern (22). *Escherichia coli* is a preferred host for heterologous protein expression due to its simple and inexpensive cultivation, ease of genetic manipulation, and the comprehensive understanding of its genetics (40). Because of its ability to form disulfide bonds, the oxidizing environment of the bacterial periplasm in particular is often the compartment of choice for the expression of disulfide bond-containing proteins (3).

Frustratingly, despite the extremely high native-state stability of proteins with multiple disulfide bonds (34), heterologous proteins with complex disulfide bond patterns often express very poorly in *E. coli* (32). One example is the bovine pancreatic trypsin inhibitor (BPTI) (31), which in its mature form contains three nonconsecutive disulfide bonds involving all of its six cysteines. Fully oxidized, native BPTI has a melting temperature of over 100°C (36). However, the protein folds poorly in bacteria, resulting in very low expression yields (31). The *in vitro* folding pathway of BPTI is probably the most intensively studied of all proteins and is dominated by the formation and isomerization of its disulfide bonds (1, 16, 45). Conversely, much less is known about the *in vivo* folding of BPTI, especially in heterologous systems (31–33). It is still not clear whether the folding challenges BPTI faces *in vitro* are also responsible for its relatively low expression in heterologous systems compared to that of other proteins (31).

In this study, we utilized two independent genetic systems to select for BPTI variants that exhibit improved expression and folding properties in the *E. coli* periplasm. The nature of the BPTI variants obtained supports the theory that destabilization or elimination of the native disulfide bond Cys14-Cys38 allows for a more productive route of folding toward the native state *in vivo*. The use of *in vivo* selections may therefore prove to be a valuable way to identify and circumvent critical rate-limiting steps in the *in vivo* folding of proteins, including those proteins whose *in vitro* folding pathways are unknown.

Department of Molecular, Cellular and Developmental Biology, Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan.
*These two authors contributed equally to this work.

Materials and Methods

Cloning of phages for the protein stability increased by directed evolution system

fd phage derivatives were grown on XL1-Blue *E. coli* cells (Stratagene, La Jolla, CA). For cloning and sequencing of phage DNA, the replicative form was prepared using a Qiagen Miniprep Kit (Qiagen, La Jolla, CA). To construct tripartite fusions within the g3p protein of fd, a derivative of fd phage called fdP213 was used. This variant carries the g3p protein stabilizing mutations T13I, T101I, Q129H, D209Y, and P213G (29). The gene for BPTI was amplified from pTI103 plasmid (15) using the primers 5′-TAATTAGGGCCCGG CCTGACTTCTGC-3′ and 5′-GCTAATAGGGCCCCACCAC AGGTCCTCATGCA-3′. These primers introduced *Apa*I sites at the 5′ and 3′ ends of the gene for mature portion of BPTI, enabling its insertion into the *Apa*I site that had been previously engineered into the *g3p* gene of fd phage by Krebber *et al.* (25). This allowed the in-frame cloning of *BPTI* into the *g3p* gene with no change in the amino acid sequence of BPTI, resulting in the phage PHSG7. It also placed BPTI between the N- and the C-terminal domains of the g3p protein. Inverted inserts were selected against because they created stop codons in *g3p*, preventing phage infectivity. The resulting fusion protein is referred to as "g3p′-BPTI WT-′g3p."

Directed evolution of BPTI in the phage system

Mutagenesis of the *BPTI* gene was performed according to the GeneMorph II EZClone domain mutagenesis kit (Stratagene) with some minor modifications. Briefly, the *BPTI* sequence was amplified from pHSG7 DNA in an error-prone polymerase chain reaction (PCR) using the Mutazyme II enzyme mix and primers 5′-CTGTCAATGCTCCGTCCGG GGCC-3′ and 5′-GTACCAGAAGCCATGGCCGGCTG-3′, which annealed up- and downstream of the *BPTI* gene. The PCR was performed as recommended by Stratagene, except that the annealing temperature was 55°C and the elongation time was 1 min. The resulting pool of mutagenized *BPTI* genes was then used to perform a second, nonmutagenic PCR on pHSG7 DNA, according to the GeneMorph II protocol, to complete the synthesis of the fd phage. After digestion with *Dpn*I (New England Biolabs, Beverly, MA), the remaining PCR product was precipitated. The resuspended DNA pellet was transformed into XL1-Blue cells, and the cultures were grown in LB for 14 h at 37°C. This simple outgrowth served as a growth competition selection that allowed the most infectious phage to outgrow less infectious phages in the culture. This protocol was designed to enable the selection of mutations within the *BPTI* gene that conferred a growth advantage to the g3p-BPTI tripartite fusion containing fd phage. Subsequently, the cultures were pelleted and the phage containing supernatant was plated using 0.7% top agar LB in different dilutions. Plaques larger than those of the unmutagenized pHSG7 phage (which generated tiny, almost invisible plaques) were picked and used to re-infect 10 ml of an overnight culture of XL1-Blue cells. These cells were again grown for 14 h to grow up the large plaque phage clones. After harvesting, the phage titer [plaque forming units (pfu)/ml] was determined, and the sequence of the *BPTI* gene determined from the isolated replicative form of the phage.

Directed evolution of BPTI in the β-lactamase system

To construct a tripartite fusion between BPTI WT and β-lactamase, the gene for BPTI was first PCR-amplified from pT1103 using primers 5′-CTGTCAATGCTCCGTCCGGG GCC-3′ and 5′-GTACCAGAAGCCATGGCCGGCTG-3′ and then cloned into the *Pfo*I site within the β-lactamase gene in pBR322*link to obtain pBR322*link-BPTI. The resulting fusion protein is referred to as bla′-BPTI WT-′bla. The *BPTI* gene was then randomly mutagenized, and clones that showed an enhanced resistance to penicillin V (Pen V) or ampicillin were isolated, and the minimal inhibitory concentration (MIC) of antibiotic necessary to inhibit their growth was determined as previously described (13). Additional specific mutations in BPTI were generated using the QuikChange™ Site-Directed Mutagenesis Kit (Qiagen).

Determination of expression levels

Whole cell extracts were prepared as previously described (13).

For the separation of the soluble and insoluble fractions, mid-log-phase cells were pelleted (13,000 *g*, 4°C, 5 min) and adjusted to $OD_{600\ nm} = 7$ in lysis buffer (10 m*M* Tris, 2 m*M* EDTA, 1 mg/ml lysozyme, and protease inhibitor). One hundred microliters of this mixture was subjected to 10 freeze–thaw cycles and the soluble and the insoluble fractions were separated by centrifugation (13,000 *g*, 4°C, 10 min). The supernatant after centrifugation was the soluble fraction. The pellet (insoluble fraction) was resuspended in a total volume of 100 μl resuspension buffer (3.2 m*M* Na_2HPO_4, 0.5 m*M* KH_2PO_4, 1.3 m*M* KCl, 135 m*M* NaCl, 2 m*M* EDTA, and protease inhibitor, pH 7.4).

For the preparation of periplasmic extracts, a volume of mid-log-phase cells corresponding to $OD_{600\ nm} = 7$ was pelleted (13,000 *g*, 4°C, 10 min) and resuspended in 70 μl osmotic shock buffer (0.2 M Tris pH 8, 0.5 M sucrose, 1 m*M* EDTA). After incubation on ice for 30 min, the supernatant (periplasmic extract) and the pellet (cytoplasmatic fraction) were separated by centrifugation (13,000 *g*, 4°C, 10 min).

To obtain samples in which bla′-link-BPTI-′bla would still contain its signal sequence, mid-log phase cells expressing bla′-link-BPTI-′bla under arabinose control were induced with 2% arabinose. After 2 min, the cells were treated with 0.08% sodium azide, incubated for another 18 min, and whole cell extracts were prepared as described above. For the expression of BPTI in the absence of β-lactamase, BPTI was expressed from pET11a (New England Biolabs) under a T7 promoter. Protein production was induced with 5 μM IPTG.

Protein samples were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis and Western blot analysis. Proteins were detected using a monoclonal rabbit antibody raised against TEM1-β-lactamase (Millipore, Billerica, MA) or reduced BPTI (generous gift from George Georgiou) as a primary and HRP-goat anti-rabbit IgG (Thermo Scientific, Waltham, MA) as a secondary antibody. Various dilutions of the protein samples for constructs containing WT BPTI and the BPTI variants were loaded on the gel. For each dilution, the ratio of band intensities for the mutant BPTI and WT BPTI was determined and averaged.

Results

To select for variants of BPTI that show enhanced expression and folding in the bacterial periplasm, we employed two independent tripartite fusion approaches that link the *in vivo* folding of BPTI to the activity of a reporter protein. The first approach was a modification of the protein stability increased by directed evolution (PROSIDE) technique (29); the second approach was the split β-lactamase system developed in our lab (13). In the PROSIDE approach, a guest protein is inserted between the N- and C-terminal domains of the g3p capsid protein of the fd phage (Fig. 1A). Both domains must remain covalently linked for g3p to be functional and confer phage infectivity. If the inserted protein is poorly folded or unstable, it becomes proteolytically sensitive. Degradation of the inserted protein will lead to a separation of the N- and C-terminal domains and a loss of infectivity. However, if the inserted guest protein folds well, the two domains of g3p remain covalently linked, and the phage carrying the fusion protein is infectious. This method therefore directly links the *in vitro* stability of the guest protein with phage infectivity. Improved folding efficiency of the guest protein in the periplasm should result in a higher infectivity of the phage, resulting in larger plaque sizes and increased phage titers.

In the β-lactamase system, the guest protein is inserted between the N- and C-terminal halves of TEM1-β-lactamase, a periplasmic enzyme that confers resistance to β-lactam antibiotics (13) (Fig. 1B). We used this technique previously to monitor the *in vivo* stability of a variety of different proteins and to identify variants with improved thermodynamic and kinetic stability by selecting for increased levels of antibiotic resistance (13). As in the PROSIDE system, poor folding of the guest protein makes it susceptible to proteolytic degradation, leading to a separation of the two halves of the reporter protein. Bacteria expressing such a construct are therefore sensitive to antibiotic stress. A well-folded guest protein, however, ensures that the fusion remains intact, thereby leading to high levels of antibiotic resistance.

Selecting for BPTI variants that lead to increased infectivity in the PROSIDE system

The phage system links the proteolytic stability of a guest protein to phage infectivity and therefore to the rate of phage amplification. Insertion of WT BPTI into the protein g3p of the fd phage (referred to as "g3p′-BPTI WT-′g3p″) resulted in a dramatic decrease in phage titer (from $\sim 10^{11}$ pfu/ml for WT fd phage to $\sim 10^{6}$ pfu/ml for g3p′-BPTI WT-′g3p, Table 1). This result is consistent with the poor *in vivo* folding and expression of BPTI noted previously (33). We reasoned that by mutagenizing the BPTI portion of the fusion and selecting for variants that show improved phage growth, we should be able to identify mutants with improved *in vivo* expression and folding properties. About 22 independent mutagenesis reactions yielded a total of 42 phage variants that resulted in increased plaque sizes and higher phage titers. Sequence analysis revealed two classes of fd variants that improved phage growth. The first class consisted of 25 independent deletion variants in which all or nearly all of the BPTI encoding sequence was selectively removed from the *g3p* gene. These deletion variants all showed titers and plaque sizes comparable to the original WT phage (data not shown). The remaining 17 phage variants had acquired single or multiple mutations in the BPTI sequence and showed increased plaque size and titer compared to g3p′-BPTI WT-′g3p (listed in Table 1).

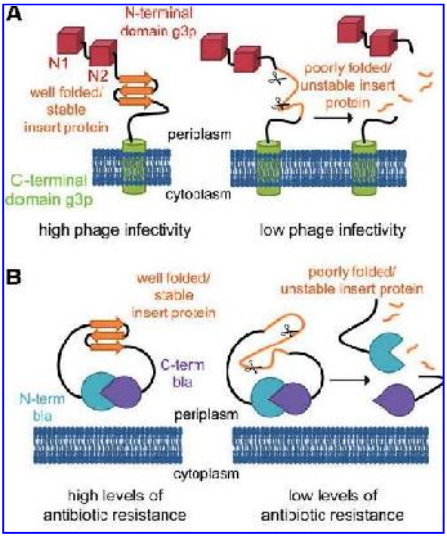


FIG. 1. Directed evolution of proteins using the modified protein stability increased by directed evolution and the β-lactamase tripartite fusion systems. The test protein BPTI (shown in orange) is inserted into either the minor coat protein of the fd phage, g3p, **(A)** or β-lactamase **(B)**, creating a tripartite fusion. **(A)** During phage replication in *Escherichia coli*, g3p is first synthesized as an integral membrane protein (shown) before it is assembled into the phage coat (not shown). The N-terminal domain of g3p carries the recognition site for the bacterial receptors the phage binds to during infection, and the C-terminal domain anchors g3p into the phage coat. Therefore, both domains must remain covalently linked for g3p to be functional and maintain phage infectivity. **(B)** β-lactamase (bla) is a periplasmic enzyme that inactivates β-lactam antibiotics. The ability of the tripartite fusions to confer phage infectivity **(A)** or resistance to antibiotics **(B)** is strongly dependent on the *in vivo* stability of the guest protein that is inserted into the reporter protein. If the test protein is stable and folded properly, the two halves of the reporter protein remain covalently linked and confer activity (left). If the inserted protein is not properly folded or unstable, it is prone to degradation by periplasmic proteases (depicted as scissors) and is likely to be degraded (right). This leads to a separation of the N- and C-terminal fragments of g3p/β-lactamase, resulting in a decreased phage titer/antibiotic resistance (13). BPTI, bovine pancreatic trypsin inhibitor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article at www.liebertonline.com/ars).

For the two BPTI variants with single mutations (F33L and G12D), it is clear that the single mutation is sufficient for the enhanced growth phenotype. The vast majority of BPTI variants isolated, however, contained multiple amino acid substitutions, complicating interpretation. Many of the residues

Table 1. Phage Titers for Bovine Pancreatic Trypsin Inhibitor Variants Obtained with the Phage System

g3p'-BPTI-'g3p variant	*Relative phage growth*
WT fd phage	+++++
BPTI WT	+/−
G12D	+++
F33L	+++
C14S, **C38S**	+++++
C14R, **C38**Y, P2S, D3F	++++
C14Y, **C38**W, **A40**D	++++
C14Y, **C38**Y, **R42**S, K46M	++++
Y35N, **A40**S, K41N	+++
R17G, G37N	+++
A16T, **I18**N, **S47**N	+++
A16N, **V34**E	+++
I18S, **A40**S	+++
L6Q, **R17**S	+++
V34E, **S47**N	+++
F4L*, **A16T**, **R39**N	+++
F4L*, **K26N**, **R39**S	+++
P13S, **Y35G**	+++
Y35N, G36D, R42S, S47N	+++

Insertion of WT BPTI into the g3p encoding sequence of the phage decreased the phage titer from $\sim 10^{11}$ pfu/ml (+++++) to 10^6 pfu/ml (+/−). Titers varied significantly from experiment to experiment; values shown are the average of 3 experiments.++++corresponds roughly to a phage titer of 10^{10} pfu/ml,+++roughly to 10^9 pfu/ml. Residues in bold were found to be substituted in multiple independent mutagenesis reactions. In most of these cases, the chemical nature of the amino acid substitution was the same. Underlined are the BPTI exact amino acid substitutions that were previously found to selectively reduce the Cys14-Cys38 disulfide upon adding DTT *in vitro* (2, 18) and/or are known to destabilize this particular disulfide bond *in vitro* (17). Dotted underlined variants eliminate the Cys14-Cys38 disulfide bond completely. Variants that have been shown to decrease the accumulation of kinetic traps in the folding pathway of BPTI *in vitro* are marked with * (50).

BPTI, bovine pancreatic trypsin inhibitor; DTT, dithiothreitol; pfu, plaque forming unit; WT, wild-type.

substituted were isolated on multiple occasions; 12 of the 33 point mutations identified affected residues that were found to be altered in multiple, independent mutagenesis reactions (shown bold in Table 1). We reasoned that the more often a residue was independently replaced in different BPTI variants, the more likely that alteration of this residue is important for robust growth of the phage expressing the g3p-BPTI variant. Among the isolated BPTI variants, six residues (F4, R17, I18, V34, R39, and R42) were independently found to be mutated twice, four (A16, Y35, A40, and S47) were mutated three times, and two (C14 and C38) were mutated four times. It is striking that of the 15 BPTI variants with multiple amino acid substitutions, the four variants exhibiting the largest increase in phage titer all eradicate the Cys14-Cys38 disulfide by eliminating both of the involved cysteines. Remarkably, seven of the substitutions that do not affect the cysteines are exactly the same as substitutions that have previously been found to destabilize Cys14-Cys38 *in vitro* (underlined in Table 1) (17).

To confirm that these amino acid substitutions are, on their own, sufficient to increase phage titers compared to phage containing WT BPTI, we attempted to construct the individual mutations in the phage system by site-specific mutagenesis. However, the very strong growth disadvantage conferred upon fd by g3p fusions containing WT BPTI and the resulting low titer greatly decreased the amount of DNA that could be isolated from the WT BPTI construct, making site-specific mutagensis very difficult. When high yields of DNA were obtained, we observed that it was often the result of the spontaneous evolution of phage containing deletions or other mutations within BPTI. The underlying basis of this problem is unfortunately exactly the same as the basis of the genetic selection; namely, the coupling of the growth of the phage to the folding of the test protein.

Selecting for BPTI variants that lead to increased antibiotic resistance in the β-lactamase system

To circumvent the problem of low DNA levels and frequent excisions of the BPTI gene from the phage genome, we decided to establish growth conditions in which the folding of the inserted protein could be selectively uncoupled from the growth of strains containing the selectable marker. This was not possible for the g3p tripartite fusions because g3p is required for phage propagation and consequently phage DNA amplification. We therefore employed a second genetic selection system, the β-lactamase system. Here, the tripartite fusion is expressed from a bacterial plasmid that also encodes for an additional selectable marker, tetracycline resistance. Growing cells containing this plasmid in tetracycline allows for an easy amplification and manipulation of plasmid DNA without any requirement for efficient folding of the tripartite β-lactamase protein. The β-lactamase system further provides a very reproducible readout for the relative level of antibiotic resistance conferred by different tripartite fusions (13).

The construction of a plasmid expressing a tripartite fusion between BPTI WT and β-lactamase (referred to as "bla'-BPTI WT-'bla") and the determination of the resulting level of antibiotic resistance is described in Materials and Methods. Antibiotic resistance is given as the MIC that prevents cell growth. The level of resistance caused by a given BPTI mutant is normalized to the level of resistance conferred by bla'-BPTI WT-'bla. The insertion of BPTI into β-lactamase resulted in a massive drop in antibiotic resistance compared to cells expressing WT β-lactamase, similar to the drop in phage infectivity that accompanied the insertion of BPTI into g3p (Fig. 2). As with the phage selection, we randomly mutagenized the BPTI portion of the tripartite fusion gene; however, in this case, we selected for increased antibiotic resistance to identify BPTI variants that show enhanced protein levels and folding in the periplasm. Using this approach, we isolated 13 point variants and 1 triple variant (P13S C14Y G28R) that showed increased MIC values compared to cells expressing bla'-BPTI WT-'bla (Table 2). We also constructed point variants P13S and G28R in the BPTI WT background by site-specific mutagenesis, and they also showed increased levels of antibiotic resistance. Variants isolated with the β-lactamase system that eliminate or destabilize the Cys14-Cys38 disulfide bond are underlined in Table 2.

For the two point mutations that led to the largest increase in MIC, BPTI G28W and K26M, we also tested the steady-state protein levels in the periplasm. These measurements were made both for BPTI when it was part of the tripartite fusion, as well as the isolated BPTI protein expressed on its own in the absence of β-lactamase. In both these contexts the protein levels of BPTI G28W and K26M in whole cell extracts were

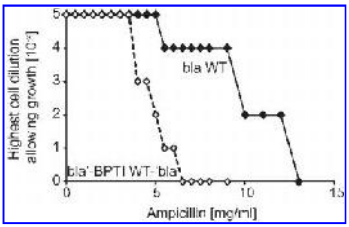


FIG. 2. Insertion of BPTI into β-lactamase leads to a massive drop in antibiotic resistance compared to cells expressing the unmodified enzyme. Mid-log phase cells of *E. coli* MG1655 Δ*ampC* expressing TEM1-β-lactamase (bla WT) or TEM1-β-lactamase fused with WT BPTI (bla'-BPTI WT-'bla), respectively, were normalized to $OD_{600\,nm} = 1$. Serial dilutions of cultures from 10^0 to 10^{-5} were spotted on LB plates containing different concentrations of ampicillin. After 18 h of incubation at 37°C, growth or no growth for each dilution for each ampicillin concentration was determined. WT, wild-type.

significantly increased compared to the corresponding WT constructs (2.5 ± 0.2-fold for BPTI G28W and 2.8 ± 0.3-fold for K26M in the fusion context, compared to the protein level of bla-BPTI WT-bla; 1.5 ± 0.1-fold for BPTI G28W and 1.7 ± 0.2-fold for K26M in the isolated protein in the absence of the fusion, compared to the protein level of WT BPTI itself. This suggests that the β-lactamase system works well for the selection of protein variants with increased expression in *E. coli*.

Destabilization of Cys14-Cys38 in vitro *correlates with increased levels of antibiotic resistance in the β-lactamase system* in vivo

Because disulfide bonds act to stabilize proteins, we were initially surprised that the elimination or destablization of Cys14-Cys38 in BPTI was actually reflected in improved phenotypes in the phage system. We were therefore interested in whether the degree of destabilization of the Cys14-Cys38 disulfide as determined *in vitro* would correlate with increased protein stability *in vivo* as reflected by improved phenotypes in the tripartite fusion systems. In contrast to the phage system, which was characterized by a rather high propensity for recombination events that lead to the elimination of the protein inserted into g3p and phage titers that varied significantly from experiment to experiment, the β-lactamase system is genetically very stable and has been shown to provide a very reproducible phenotype for a variety of insert proteins (13). We introduced four of the amino acid substitutions that were obtained in the phage selection as point mutations into the bla'-BPTI WT-'bla background, and one mutant, Y35L, that was derived from the literature (17). All five BPTI variants are known to destabilize this disulfide bond *in vitro* (17). Remarkably, the extent of thermodynamic destabilization of Cys14-C38 caused by a given point mutation *in vitro* correlated well with the level of antibiotic resistance in the β-lactamase system (Fig. 3). Consistent with these

TABLE 2. MINIMAL INHIBITORY CONCENTRATION VALUES FOR BOVINE PANCREATIC TRYPSIN INHIBITOR VARIANTS OBTAINED WITH THE β-LACTAMASE SYSTEM

BPTI variants obtained in the selection		*BPTI variants constructed via site-directed mutagenesis*	
bla'-BPTI-'bla	*MIC ± SD*	*bla'-BPTI-'bla*	*MIC ± SD*
bla WT	**2.07 ± 0.07**	A16T	0.98 ± 0.02
BPTI WT	**1.00 ± 0.00**	F33L	0.94 ± 0.06
N24H	1.07 ± 0.06	Y35N	1.01 ± 0.08
P13S[a]	1.10 ± 0.02	Y35L	1.07 ± 0.01
I18L	1.10 ± 0.02	N24A	1.08 ± 0.08
I19F	1.17 ± 0.01	Y35G	1.12 ± 0.04
Y10N	1.23 ± 0.02	Y23L	1.18 ± 0.01
Q31K	1.23 ± 0.01		
A25V	1.23 ± 0.05	[14–38]	0.13 ± 0.02
Q31R	1.25 ± 0.01	[30–51]	0.90 ± 0.03
P13L	1.36 ± 0.04	[5–55]	1.46 ± 0.03
G28R[a]	1.37 ± 0.01		
C14W	1.43 ± 0.06	[5–55] [14–38][b]	0.74 ± 0.02
N24K	1.50 ± 0.09	[14–38] [30–51][b]	1.10 ± 0.03
K26M	1.88 ± 0.09	[5–55] [30–51][b]	1.70 ± 0.06
G28W	1.91 ± 0.10		
P13S C14Y **G28**R[a]	1.35 ± 0.08		

The relative level of antibiotic resistance of cells expressing different bla'-BPTI-'bla variants is given as the minimal inhibitory concentration (MIC) that prevents cell growth (see the Materials and Methods section). Residues in bold were substituted more than once in independent mutagenesis reactions. In most of these cases, the chemical nature of the amino acid substitution was the same. Residues underlined were identified in a screen for DTT-sensitive mutants as single point mutations or are known to destabilize the Cys14-Cys38 disulfide *in vitro* (2, 17, 50). The dotted underlined variants prevent formation of Cys14-Cys38.

[a]P13S and G28R were selected as part of a triple mutant (P13S C14Y G28R) and subsequently constructed as single point mutations *via* site-directed mutagenesis.

[b]Relative protein levels of [5–55] [14–38], [14–38] [30–51], and [5–55] [30–51] in the β-lactamase system were 0.9 ± 0.1, 1.2 ± 0.2, and 1.6 ± 0.1, respectively.

SD, standard deviation.

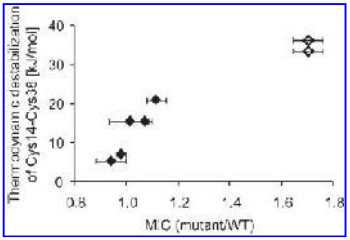


FIG. 3. Increased destabilization of the Cys14-Cys38 disulfide bond *in vitro* correlates with increased antibiotic resistance in the β-lactamase system *in vivo*. Different variants of BPTI were inserted into β-lactamase *via* flexible linkers. The level of antibiotic resistance for cells expressing the corresponding fusion constructs was determined as described in the Materials and Methods section. BPTI variants containing single amino acid substitutions that were either identified in the phage selection (A16T, F33L, Y35N, and Y35G) or obtained from the literature (Y35L) are indicated with filled diamonds (17). The relative level of antibiotic resistance (indicated by MIC) compared to a tripartite fusion containing BPTI WT *in vivo* is plotted against the thermodynamic destabilization $\Delta\Delta G$ of the Cys14-Cys38 disulfide bond in the mutant proteins as determined *in vitro* ($\Delta\Delta G = R\,T\,\ln[C_{eff}(WT)/C_{eff}(mut)]$, where C_{eff} (WT) and C_{eff} (mut) are the effective concentrations of the Cys14-Cys38 disulfide bond in WT BPTI or the mutant proteins, respectively) (17). Stability values for a BPTI species in which the formation of Cys14-Cys38 was prevented by either carboxymethylation or carboxyamidomethylation were taken from Schwarz *et al.* (36) and plotted against the level of resistance conferred by bla′-BPTI C14S C38S-′bla. These latter variants are proteins with chemical side chain modifications rather than mutations, and their thermodynamic destabilization values were determined using a different methodology. Therefore, their absolute values may not be precisely comparable to the single amino acid substitutions; they are shown as open diamonds. The R^2 value for the single point variants (*filled diamonds*) is 0.89, and the R^2 for the complete data set is 0.91. Data are expressed as mean values ± SD. MIC, minimal inhibitory concentration; SD, standard deviation.

findings, complete removal of the Cys14-Cys38 disulfide bond leads to an even higher MIC (Fig. 3, open diamonds). Importantly, the elimination of the Cys14-Cys38 disulfide bond increases the protein level of BPTI in *E. coli* both when it is part of the tripartite fusion as well as when BPTI is expressed on its own (1.6 ± 0.1-fold increase in protein level relative to bla′-BPTI WT-′bla in presence of β-lactamase and 2.3 ± 0.2-fold increase in protein level relative to BPTI WT in absence of β-lactamase, respectively). These results fit nicely with a previous observation that the elimination of this disulfide by serine substitutions leads to a 7-fold increase in the expression level of BPTI in the *E. coli* periplasm (31).

The stabilities of single disulfide-bonded BPTI species in vitro *correlate with the level of antibiotic resistance in the β-lactamase system* in vivo

Because the β-lactamase system has been shown to be a good readout for the *in vivo* stability of a variety of proteins (13), we decided to use it to assess which of the different intermediates that accumulate in BPTI's folding pathway *in vitro* might be especially problematic for BPTI's folding *in vivo*. Traditionally, much information about the folding pathway of BPTI has been obtained by studying the properties of analogs of folding intermediates (11) in which particular pairs of cysteines were chemically blocked or replaced with other amino acids, especially serine (6). To create three BPTI variants that can only form a single native disulfide bond, we used site-directed mutagenesis to replace all but one cysteine pair with serines. We list the disulfide mutants we constructed and define our nomenclature in Table 3. Among the single disulfide-bonded species containing native disulfides, BPTI [5–55] has the most native-like conformation *in vitro*, whereas [14–38] has the least (26, 41). Introduction of [5–55] into the β-lactamase system resulted in the highest MIC, whereas introduction of the least structured and least stable species ([14–38]) resulted in the lowest MIC, suggesting that the latter intermediate does not form a stable conformation in the bacterial periplasm (Fig. 4A). These results indicate that, for single disulfide-bonded species of BPTI, the extent of native structure *in vitro* is correlated to the MIC *in vivo*. Consistently, the levels of antibiotic resistance for BPTI [5–55], [14–38], and [30–51] correlate very well with the contribution of the particular disulfide bond to the stability of the folded conformation, as reflected by the effective concentration of the disulfide bond in the native state (7, 23) (Fig. 4B). Interestingly, there is an inverse correlation between the MIC and the tendency of a disulfide to form in reduced BPTI: the higher the conformational preference for formation of a particular native disulfide in the reduced state, the lower the MIC of the BPTI variant that can only form this particular disulfide (Fig. 4C). An inverse correlation was also found between the MIC and the reactivity of the involved cysteines, with higher reactivites

TABLE 3. NOMENCLATURE FOR BOVINE PANCREATIC TRYPSIN INHIBITOR VARIANTS USED IN THIS STUDY

BPTI species	*Alternative name*	*Contains disulfide bond(s)*	*Corresponds to BPTI variant in the β-lactamase system*
[5–55]	–	Cys5-Cys55	C14S C30S C38S C51S
[14–38]	–	Cys14-Cys38	C5S C30S C51S C55S
[30–51]	–	Cys30-Cys51	C5S C14S C38S C55S
[5–55; 14–38]	–	Cys5-Cys55 and Cys14-Cys38	C30S C51S
[5–55; 30–51]	–	Cys5-Cys55 and Cys30-Cys51	C14S C38S
[14–38; 30–51]	–	Cys14-Cys38 and Cys30-Cys51	C5S C55S
[5–55; 14–38; 30–51]	Wild-type	Cys5-Cys55, Cys14-Cys38 and Cys30-Cys51	

CysX-CysY refers to a disulfide bond formed between cysteine residue X and cysteine residue Y. BPTI variants representing folding intermediates found *in vitro* were obtained by substituting cysteines not involved in disulfide bond formation with serines.

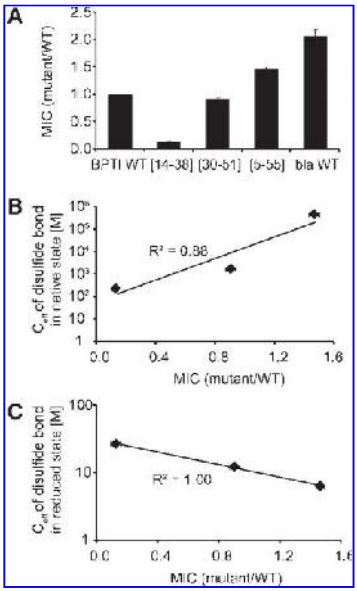


FIG. 4. In BPTI variants that can only form a single disulfide bond, the stability of the disulfide bond *in vitro* correlates with the level of antibiotic resistance in the *β*-lactamase system *in vivo*. (A) Three variants of BPTI representing single disulfide-bonded intermediates ([14–38], [30–51], or [5–55], respectively) were inserted into *β*-lactamase *via* flexible linkers. In all variants, the cysteines not involved in formation of the disulfide bond indicated were substituted with serines. Bla WT indicates WT *β*-lactamase without any insertions, and BPTI WT refers to the construct bla-′BPTI WT-′bla. The level of antibiotic resistance (indicated by MIC) for cells expressing the corresponding fusion constructs was determined as described in the Materials and Methods section and is normalized to cells expressing bla′-BPTI WT-′bla. **(B)** The "effective concentration" of two interacting groups in the native-like conformation can be used as an expression of the free energy of the intermolecular interaction (23). The higher the effective concentration, the larger is the contribution of the particular interaction to the stability of the folded formation. Values for the effective concentrations were taken from Creighton and Goldenberg (7). **(C)** In the reduced state, the kinetic effective concentration can be used to assess the extent to which two cysteines are brought into proximity to each other (9). Values for the effective concentration for the three disulfide bonds indicated were taken from Dadlez and Kim (9). Data are expressed as mean values ± SD.

resulting in lower MICs (data not shown) (9). Further, we observed that a single disulfide-bonded variant of BPTI, [5–55], which is substantially less stable than the native protein (43), exhibits a higher MIC than WT BPTI in the *β*-lactamase system (Table 2), and this is testimony to the severe folding problems WT BPTI faces in the periplasm of *E. coli*.

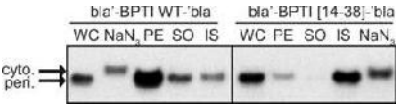


FIG. 5. The majority of BPTI [14–38] detected in whole cell extracts is periplasmic, but insoluble. Cell fractions were prepared from *E. coli* MG1655b Δ*amp*C cells expressing bla′-BPTI WT-′bla or bla′-BPTI [14–38]-′bla, respectively, as described in the Material and Methods section. The tripartite fusions were detected by Western blot using an anti-*β*-lactamase antibody. In cells treated with sodium azide, the Sec pathway is blocked and the export of the tripartite fusions prevented (NaN_3). Here, tripartite fusion proteins remain in the cytosol and retain their signal sequence, which is usually cleaved upon export in the periplasm. Tripartite fusions that were exported (as detected in the periplasmic extracts [PE], indicated with peri.) can therefore be distinguished from fusion proteins that were retained in the cytosol (as detected in the sample treated with NaN_3, indicated with cyto.) by their molecular weight. The blot shows that for both bla′-BPTI WT-′bla and bla′-BPTI [14-38]-′bla, all fusion protein detected in whole cell extracts (WC) is located in the periplasm. However, for bla′-BPTI [14-38]-′bla the amount of soluble protein is much lower than for bla′-BPTI WT-′bla (compare soluble [SO] and insoluble fractions [IS]).

Because the MIC is a measure of how many molecules are successfully exported, we wondered if the low MIC caused by bla′-BPTI [14–38]-′bla was the result of a preferred formation of this disulfide in the cytosol, preventing export. bla′-BPTI [14–38]-′bla showed a protein level in whole cell extracts that was substantially higher than expected from its MIC value (protein level was 1.1 ± 0.1, MIC was only 0.13 ± 0.0 compared to bla′-BPTI WT-′bla). Subsequent analysis of soluble/insoluble and periplasmic/cytosolic fractions of cells expressing bla′-BPTI [14–38]-′bla indicated that the majority of the protein was periplasmic, but insoluble (Fig. 5). This result is in stark contrast to that obtained for the rest of the bla′-BPTI-′bla variants, which showed a very good correlation between level of antibiotic resistance and steady-state levels of the fusion proteins in whole cell extracts (Fig. 6). The latter observation is consistent with previous experiments using the *β*-lactamase system, which showed that the level of antibiotic resistance is a good measure for the relative level of the fusion protein in the periplasm (13). However, the actual amount of soluble, periplasmic bla′-BPTI [14–38]′-bla reflected the observed level of antibiotic resistance for bla′-BPTI [14–38]-′bla very well (periplasmic level compared to bla′-BPTI WT-′bla was 0.11 ± 0.05).

The β-lactamase system is also suitable to assess the effect of free cysteines on the in vivo *stability of a protein*

During heterologous protein expression in *E. coli*, the presence of unpaired cysteines in recombinant, disulfide bond containing proteins can promote the formation of incorrect intra- and intermolecular disulfide bonds, leaving the protein in inactive and aggregation-prone conformations (27). Elimination of free cysteines *via* site-directed mutagenesis has successfully been used to increase the stability, specific activity, and expression of these proteins and reduce their aggregation propensity (14, 30). However, since the

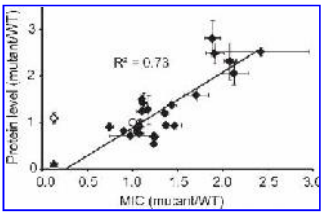


FIG. 6. Steady-state protein levels of different bla′-BPTI-′bla variants *in vivo* correlate well with the corresponding levels of antibiotic resistance. The relative steady-state protein levels in whole cell extracts were detected by Western blot using an anti-β-lactamase antibody. Band intensities of β-lactamase (*circle*) or tripartite fusions containing different BPTI variants (*filled diamonds*) were normalized to the protein levels of cells expressing bla′-BPTI WT-′bla (*open circle*). In the case of bla′-BPTI [14–38]′-bla, the protein level was determined in whole cell extracts (*open diamond*) and in soluble, periplasmic fractions (*triangle*). In both cases, the protein level was adjusted to the protein level of bla′-BPTI WT′-bla in whole cell extracts or the soluble, periplasmic cell fraction, respectively. The trend line includes the value for the soluble periplasmic fraction of bla′-BPTI [14–38]′-bla and excludes the value for the whole cell extract of bla′-BPTI [14–38]′-bla. Data are expressed as mean values ± SD.

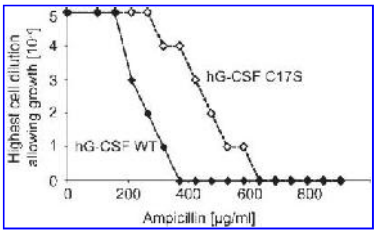


FIG. 7. Substitution of Cys17 in hG-CSF with serine leads to an increase in antibiotic resistance as part of the β-lactamase system. WT hG-CSF, which contains two disulfide bonds (Cys36-Cys42 and Cys64-Cys74) and one unpaired cysteine (Cys17), is prone to aggregation, even under native-like conditions (35). Substitution of the free cysteine with serine significantly decreases the proteins propensity to aggregate (35). Mid-log phase cells of *E. coli* MG1655 Δ*ampC* expressing bla′-hG-CSF WT-′bla or bla′-hG-CSF C17S-′bla, respectively, were normalized to $OD_{600\ nm} = 1$. Serial dilutions of cultures from 10^0 to 10^{-5} were spotted on LB plates containing different concentrations of ampicillin. After 18 h of incubation at 37°C, growth or no growth for each dilution for each ampicillin concentration was determined.

substitution of cysteines has also been shown to decrease protein stability by reducing hydrophobic interactions or generating cavities in the protein core (30), the effect of such mutations is hard to predict and can largely depend on the amino acid the cysteine is replaced with (21).

WT human granulocyte-colony stimulating factor (hG-CSF), which forms two natural disulfide bonds and contains one unpaired cysteine (Cys17), has a high propensity to aggregate, even under native-like conditions (35). This aggregation propensity is significantly decreased when the free cysteine is eliminated (35). In agreement with this observation, introduction of hG-CSF C17S into the β-lactamase system leads to an substantial increase in antibiotic resistance over a range of PenV concentrations compared to WT hG-CSF (Fig. 7) (13). This suggests that the β-lactamase technique can be used to identify cysteine mutations that are beneficial for the expression and *in vivo* stability of the corresponding protein by randomly mutating the proteins cysteines and simply selecting for increased levels of antibiotic resistance.

Discussion

Elimination or destabilization of the Cys14-Cys38 disulfide of BPTI leads to an increased expression in the bacterial periplasm

Despite the increasing number of studies investigating the folding and stability of proteins in the living cell (20), the fundamental question as to what extent *in vitro* folding experiments reflect the actual folding pathway *in vivo* has not been sufficiently answered for many proteins. In this study, we chose a multi-disulfide-bonded protein that is particularly poorly expressed in heterologous systems with the aim of selecting for variants that show improved folding and expression in the periplasmic space of *E. coli*. The test protein BPTI and many of its mutants have already been extensively characterized *in vitro*, allowing a direct comparison of mutants described in the literature with variants obtained in our *in vivo* selection.

Both genetic selection systems employed in this study were designed to exclusively identify variants that are successfully exported to the oxidizing environment of the periplasm, ensuring that we isolated BPTI variants showing increased protein levels in the bacterial compartment that would favor the correct formation of the protein's disulfide bonds. Strikingly, BPTI variants that resulted in the strongest phenotypes indicative of improved folding and expression in both selection systems completely eliminated one of BPTI's three native disulfide bonds, Cys14-Cys38. When analyzing our isolated BPTI variants in which no cysteines were mutated, we further noticed a high degree of overlap with BPTI variants sensitive to the reducing agent dithiothreitol (DTT) that were previously isolated in a screen by Coplen (2). This screen identified BPTI variants that are able to fold into an active conformation but, unlike WT BPTI, rapidly lose their protease inhibitory activity upon addition of DTT and are consequently inactivated in the presence of trypsin (2, 18). In total, our selections for increased folding and expression of BPTI in the bacterial periplasm yielded 8 variants (G12D, A16T, Y35N, Y35G, F33L, G28W, P13S, and G36D) that are known to be DTT-sensitive as single amino acid substitutions *in vitro* and/or were shown reduce the effective concentration of the Cys14-Cys38 disulfide, increase the rate constant for its reduction, and reduce its stability *in vitro* by up to 21 kJ/mol (BPTI Y35G) (2, 17). Five of the DTT-sensitive mutants (G12D, A16T, Y35N, Y35G, and F33L) have further been shown to selectively make the Cys14-Cys38 disulfide susceptible to reduction by DTT (17). For these variants, it was suggested that the resulting loss of BPTI activity was due to an increased flexibility of the BPTI variants and that the mutations might

favor the intramolecular rearrangement of intermediates (2, 17, 18). Altogether, 65% of the BPTI variants obtained in the *in vivo* phage selection contain amino acid substitutions that either destabilize or eliminate the Cys14-Cys38 disulfide bond *in vitro*. This is a remarkable overlap, considering that there are 1102 single amino acid substitutions possible for the BPTI sequence. The idea that destabilization or removal of Cys14-Cys38 is key to improving the folding of BPTI in the bacterial periplasm is further supported by our findings that the level of antibiotic resistance in the β-lactamase system measured *in vivo* directly correlates with the extent of thermodynamic destabilization of the Cys14-Cys38 disulfide caused by a given single point mutation *in vitro* (Fig. 3) and that the relative protein level of BPTI C14S C38S expressed on its own is significantly enhanced compared to WT BPTI.

Elimination of the Cys14-Cys38 disulfide might prevent its unfavorable, early formation of in vivo

That mutations reducing a protein's *in vitro* stability lead to an increase in expression *in vivo* might appear counterintuitive. However, the formation of Cys14-Cys38 is particularly problematic in BPTI's *in vitro* folding pathway due to its early but nonproductive formation in BPTI folding and its involvement in the rate-limiting step of the overall BPTI folding

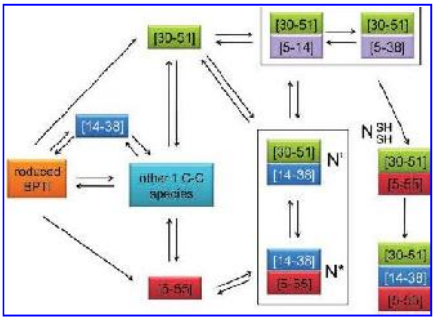


FIG. 8. Consensus folding pathway of WT BPTI *in vitro*, modified from Creighton and Goldenberg (7) and Weissman and Kim (8, 45, 46). Starting from fully reduced BPTI (far left), the [14–38] intermediate is formed much more rapidly than any other species containing one disulfide-bond (9). However, the vast majority of [14–38] molecules are either reduced or isomerized before they form the thermodynamically more stable intermediates [30–51] or [5–55] (1). The most productive folding pathway for BPTI proceeding from the single disulfide-bonded species occurs *via* the two intermediates [5–14; 30–51] and [5–38; 30–51], which are in rapid equilibrium with each other and contain non-native disulfides (5, 10). At neutral pH, a large fraction of the BPTI molecules accumulates as kinetic traps (N′ and N*), which only slowly isomerize toward the very native-like species (45, 46). Once NH^{SH}_{SH} is formed, the Cys14-Cys38 disulfide is introduced rapidly to obtain native BPTI (4). Note that the relative distribution of the double disulfide-bonded species is dependent on the pH (10, 45). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article at www.liebertonline.com/ars).

reaction (Fig. 8) (6, 46, 49). Due to the high reactivity of Cys14 and Cys38 and the significant conformational preference for introducing the Cys14-Cys38 disulfide during BPTI refolding *in vitro* at neutral pH, the [14–38] intermediate is formed much more rapidly from reduced BPTI than any other species containing a single disulfide bond, native or non-native (1, 8, 9). Despite this, [14–38] is not a very productive folding intermediate. The vast majority of [14–38] molecules are either reduced or isomerized before they accumulate in the form of the more stable single disulfide-bonded species [5–55] and [30–51] *in vitro* (41, 45, 46). However, the reduction or isomerization of the Cys14-Cys38 disulfide that would allow for this efficient folding pathway might be more difficult in the very oxidizing environment of the periplasm (38) than in *in vitro* refolding reactions, which tend to be done under less oxidizing conditions. Importantly, in the absence of other stabilizing effects such as an additional disulfide bond, the sole formation of Cys14-Cys38 in reduced BPTI does not substantially stabilize the native conformation *in vitro* (12). In *E. coli*, BPTI [14–38] does not appear to adopt a soluble, stable conformation, as reflected by its low MIC and its accumulation in the insoluble fraction in the β-lactamase system. Therefore, the early and rapid formation of Cys14-Cys38 in the periplasm might favor the accumulation of BPTI in an aggregation prone state, reducing the number of molecules that can proceed folding toward [5–55] and [30–51], which are thermodynamically more stable *in vitro* (1). Consistently, BPTI variants that mimic single disulfide-bonded intermediates [5–55] and [30–51] also show an increased protein level compared to [14–38] *in vivo* as reflected by their increased MICs in the β-lactamase system. In fact, the stabilities of BPTI's native disulfide bonds as determined *in vitro* correlate well with the observed antibiotic resistance of the corresponding tripartite fusions *in vivo*, suggesting that the β-lactamase system is well suited to assess the stabilities of different BPTI variants in the living cell.

Elimination of the Cys14-Cys38 disulfide might further prevent the accumulation of protease-sensitive intermediates in later stages of BPTI folding

Preventing or decreasing the rapid formation of Cys14-Cys38 in the early folding of BPTI *in vivo* might be sufficient to explain the superior phenotypes of our selected BPTI variants that eliminated or destabilized this disulfide. However, formation and subsequent reduction of Cys14-Cys38 is also critical in later folding stages *in vitro*, which might also help explain the enhanced antibiotic resistance of tripartite fusions in which both Cys14 and Cys38 are replaced. At neutral pH, a large fraction of BPTI molecules accumulates in the form of the kinetic traps [5–55; 14–38] (N*) or [14–38; 30–51] (N′) (7, 46). For folding to proceed, N* and N′ must reduce Cys14-Cys38 and unfold substantially to isomerize into the native-like NH^{SH}_{SH} [5–55; 30–51] (46). The reduction and isomerization of Cys14-Cys38 appears to be the rate-limiting step in the productive folding of BPTI *in vitro* at neutral pH (5, 7). Reduction and isomerization of Cys14-Cys38, however, might be especially problematic in the periplasm, which is high in oxidase but low in isomerase activity (19).

All the double disulfide-bonded intermediates N*, N′, and NH^{SH}_{SH} show substantial native-like structure *in vitro* (39, 42). However, the amount of native structure in a BPTI intermediate and its lifetime under optimized conditions *in vitro* do

not necessarily indicate the proteolytic sensitivity of that species *in vivo*. Although the kinetic traps N′ and N* persist for long periods of time under appropriate conditions *in vitro* (6, 46, 49), the majority of N′ molecules are actually unfolded at physiological temperatures (7), suggesting that this species might be prone to proteolysis at 37°C *in vivo*. Of the three intermediates, NH^{SH}_{SH} is the most stable and native-like species (4, 44) and is known to fold into its active, native conformation both *in vitro* under physiological temperatures and when expressed in the periplasm of *E. coli* (28). Although its melting temperature is about 30°C lower than that of WT BPTI, it is still almost 40°C higher than the cultivation temperature of *E. coli*, suggesting that NH^{SH}_{SH} is relatively stable *in vivo* (44). In good agreement with these findings, the BPTI variant [5–55; 30–51] that represents the NH^{SH}_{SH} species leads to the highest MIC and protein level among the three double disulfide-bonded variants tested (Table 2). On the other hand, the two variants that imitate the kinetically trapped species N* and N′ ([5–55; 14–38] and [14–38; 30–51], respectively) exhibit substantially lower *in vivo* stabilities, as reflected by decreased levels of antibiotic resistance and steady-state protein levels in the β-lactamase system, suggesting that they are proteolytically more sensitive *in vivo*.

Because antibiotic resistance is determined by the steady-state levels of soluble and periplasmic protein, it is possible that the low MIC values of N* and N′ are caused in part by a problematic early formation of Cys14-Cys38, inducing protein aggregation. However, the very good correlation between antibiotic resistance and protein level in whole cell extracts for N*, N′, and NH^{SH}_{SH} suggests that aggregation is not the dominating factor for the decreased MIC of N* and N′ (Table 2). In contrast to [14–38], N* and N′ both possess an additional, stabilizing disulfide bond, allowing the protein to reach other, more folded and less aggregation prone conformations. Remarkably, the phenotype conferred by bla′-BPTI WT-′bla is intermediate (Fig. 4A). This may result from a mixed population in which some molecules populate the very stable, native state (leading to high levels of antibiotic resistance) and other molecules populate the kinetically trapped intermediate states N* or N′ (leading to low levels of antibiotic resistance).

In addition to its impact on the early folding events of BPTI, the elimination or destabilization of Cys14-Cys38 might therefore also be associated with a simplification of BPTI's folding pathway *in vivo* in later folding stages, in which the accumulation of the proteolytically sensitive intermediates N* and N′ is prevented or decreased. Variants that destabilize Cys14-Cys38 might favor the formation of double disulfide-bonded species with non-native disulfides, which are more likely to undergo intramolecular rearrangements and are part of the most productive folding route of BPTI *in vitro*, as it has been already shown for several of the BPTI variants isolated in our selection (F33L, Y23L, F4L, and Y35L) (50). This might lead to a preferred formation of the stable and very native-like species NH^{SH}_{SH}.

Interestingly, in variants of BPTI in which the formation of the Cys14-Cys38 disulfide bond was prevented, the rate of formation of native-like BPTI decreased from $9\,s^{-1}$ to $1.6\times10^{-3}\,s^{-1}$ at 25°C and pH 8.7 *in vitro* (15). However, this effect was diminished when experiments were performed at 37°C and at a less basic pH (8.3); under these conditions, only a 3-fold decrease in the overall folding rate for native BPTI was observed (28). Our finding that BPTI C14S C38S shows increased protein levels compared to BPTI WT when expressed in the bacterial periplasm in absence of β-lactamase, an observation also made by (31), suggests that this slightly higher rate of formation of the native or native-like protein for BPTI WT compared to [5–55; 30–51] *in vitro* cannot, if even relevant *in vivo*, compensate for the potentially negative impact of the presence of Cys14-Cys38 in the early and later folding stages of BPTI *in vivo*.

The observation that the elimination of a stabilizing disulfide bond enhances the expression of BPTI might provoke the question why it was not eliminated by nature in the course of evolution. However, BPTI variants in which the Cys14-Cys38 disulfide is eliminated by substituting one or both of the involved cysteines are known to be cleaved by trypsin up to 10,000-fold more rapidly (47) and show decreased secretion efficiencies in yeast by up to 30% (24), suggesting that Cys14-Cys38 is important for the *in vivo* function and protease resistance of the protein.

Dependence of BPTI folding on the presence of folding factors

Our data suggest that the elimination or reduction of the kinetic traps N* and N′ *in vivo* might increase the level of BPTI in the bacterial periplasm. In BPTI's natural environment, the endoplasmatic reticulum, these species are likely to be isomerized by the eukaryotic disulfide bond isomerase (protein disulfide isomerase [PDI]), which has been shown to promote isomerization of these kinetic traps and enhance the formation of native BPTI by 27-fold *in vitro* (48). In fact, co-expression of rat PDI has been successfully used to increase the expression of BPTI in *E. coli* (32). Compared to its eukaryotic counterpart PDI, the isomerase activity of the bacterial enzyme responsible for disulfide bond isomerization in the periplasm, DsbC, is relatively low (37). Consistent with that, we did not find the level of antibiotic resistance conferred by various BPTI tripartite fusions to be strongly dependent on the presence of native levels of DsbC (data not shown). BPTI folding in *E. coli*, however, severely depends on the activity of the disulfide bond oxidase DsbA (33). The importance of DsbA for BPTI folding is supported by our finding that *dsbA*$^-$ strains expressing different bla-BPTI fusions showed decreased levels of antibiotic resistance compared to a *dsbA*$^+$ strain (data not shown).

Conclusion

Presumably due to the differences in the redox-environment of the endoplasmatic reticulum and the bacterial periplasm, the heterologous expression of disulfide-rich, heterologous proteins in *E. coli* is often very challenging (32). Our results for BPTI and hG-CSF indicate that our tripartite fusion systems are well suited to enhance expression and folding of difficult to express heterologous proteins in the bacterial periplasm by random mutation of the target gene and a simple selection for improved phenotypes in our systems. The feasibility of our approach is supported by the good correlation between the relative level of antibiotic resistance and the relative level of the tripartite fusion in the β-lactamase system (Fig. 6) and the enhanced protein levels of selected BPTI variants compared to BPTI WT (determined for double mutant C14S C38S and single mutants K26M and G28W) in absence of β-lactamase

we observe. Further, the majority of variants selected for increased expression of BPTI eliminates or destabilizes the Cys14-Cys38 disulfide bond, suggesting that circumvention of Cys14-Cys38-related folding problems in early or later folding stages may improve the expression of BPTI *in vivo* and compensate for the apparent loss in stability *in vitro*. For proteins with unpaired cysteines like hG-CSF, our systems can be used to easily identify if and how free cysteines can be substituted to increase expression and *in vivo* stability of the protein. Our results therefore emphasize the power of genetic selection to identify and circumvent critical steps and residues problematic for in the *in vivo* folding pathway of proteins without the need to perform time-consuming site-directed mutagenesis experiments. Although we used an extremely well-studied model protein here, this technique should be applicable to a variety of target proteins with potentially unknown folding pathways. In cases where an unmutagenized target protein is preferred, our selection systems also provide the convenient possibility of fine-tuning the redox-environment of the periplasm for the enhanced expression of a target protein by mutating co-expressed folding factors such as oxidoreductases or proteins involved in redox-maintenance.

Acknowledgments

We thank George Georgiou for the generous gift of BPTI antibody. We are further grateful to F.X. Schmid and B. Eckert for providing us with fd phage.

Author Disclosure Statement

No competing financial interests exist.

References

1. Bulaj G and Goldenberg DP. Early events in the disulfide-coupled folding of BPTI. *Protein Sci* 8: 1825–1842, 1999.
2. Coplen LJ, Frieden RW, and Goldenberg DP. A genetic screen to identify variants of bovine pancreatic trypsin inhibitor with altered folding energetics. *Proteins* 7: 16–31, 1990.
3. Cornelis P. Expressing genes in different *Escherichia coli* compartments. *Curr Opin Biotechnol* 11: 450–454, 2000.
4. Creighton T. Disulphide-coupled protein folding pathways. *Philos Trans R Soc Lond B Biol Sci* 348: 5–10, 1995.
5. Creighton TE. Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J Mol Biol* 113: 275–293, 1977.
6. Creighton TE. Protein folding pathways determined using disulphide bonds. *Bioessays* 14: 195–199, 1992.
7. Creighton TE and Goldenberg DP. Kinetic role of a metastable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J Mol Biol* 179: 497–526, 1984.
8. Dadlez M and Kim PS. A third native one-disulphide intermediate in the folding of bovine pancreatic trypsin inhibitor. *Nat Struct Biol* 2: 674–679, 1995.
9. Dadlez M and Kim PS. Rapid formation of the native 14–38 disulfide bond in the early stages of BPTI folding. *Biochemistry* 35: 16153–16164, 1996.
10. Darby NJ, Morin PE, Talbo G, and Creighton TE. Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. *J Mol Biol* 249: 463–477, 1995.
11. Darby NJ, van Mierlo CP, and Creighton TE. The 5–55 single-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor. *FEBS Lett* 279: 61–64, 1991.
12. Ferrer M, Barany G, and Woodward C. Partially folded, molten globule and molten coil states of bovine pancreatic trypsin inhibitor. *Nat Struct Biol* 2: 211–217, 1995.
13. Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, Warriner SL, Radford SE, and Bardwell JC. Optimizing protein stability *in vivo*. *Mol Cell* 36: 861–871, 2009.
14. Fremaux I, Mazeres S, Brisson-Lougarre A, Arnaud M, Ladurantie C, and Fournier D. Improvement of Drosophila acetylcholinesterase stability by elimination of a free cysteine. *BMC Biochem* 3: 21, 2002.
15. Goldenberg DP. Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and – 38 thiols. *Biochemistry* 27: 2481–2489, 1988.
16. Goldenberg DP. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 17: 257–261, 1992.
17. Goldenberg DP, Bekeart LS, Laheru DA, and Zhou JD. Probing the determinants of disulfide stability in native pancreatic trypsin inhibitor. *Biochemistry* 32: 2835–2844, 1993.
18. Goldenberg DP, Berger JM, Laheru DA, Wooden S, and Zhang JX. Genetic dissection of pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 89: 5083–5087, 1992.
19. Hiniker A, Collet JF, and Bardwell JC. Copper stress causes an *in vivo* requirement for the *Escherichia coli* disulfide isomerase DsbC. *J Biol Chem* 280: 33785–33791, 2005.
20. Ignatova Z. Monitoring protein stability *in vivo*. *Microb Cell Fact* 4: 23, 2005.
21. Ishikawa M, Iijima H, Satake-Ishikawa R, Tsumura H, Iwamatsu A, Kadoya T, Shimada Y, Fukamachi H, Kobayashi K, Matsuki S, *et al.* The substitution of cysteine 17 of recombinant human G-CSF with alanine greatly enhanced its stability. *Cell Struct Funct* 17: 61–65, 1992.
22. Kibria FM and Lees WJ. Balancing conformational and oxidative kinetic traps during the folding of bovine pancreatic trypsin inhibitor (BPTI) with glutathione and glutathione disulfide. *J Am Chem Soc* 130: 796–797, 2008.
23. Kirby AJ. Effective molarities for intramolecular reactions. *Adv Phy Org Chem* 17: 183–278, 1980.
24. Kowalski JM, Parekh RN, and Wittrup KD. Secretion efficiency in *Saccharomyces cerevisiae* of bovine pancreatic trypsin inhibitor mutants lacking disulfide bonds is correlated with thermodynamic stability. *Biochemistry* 37: 1264–1273, 1998.
25. Krebber C, Spada S, Desplancq D, Krebber A, Ge L, and Pluckthun A. Selectively-infective phage (SIP): a mechanistic dissection of a novel *in vivo* selection for protein-ligand interactions. *J Mol Biol* 268: 607–618, 1997.
26. Krokoszynska I, Dadlez M, and Otlewski J. Structure of single-disulfide variants of bovine pancreatic trypsin inhibitor (BPTI) as probed by their binding to bovine beta-trypsin. *J Mol Biol* 275: 503–513, 1998.
27. Leatherbarrow RJ and Fersht AR. Protein engineering. *Protein Eng* 1: 7–16, 1986.
28. Marks CB, Naderi H, Kosen PA, Kuntz ID, and Anderson S. Mutants of bovine pancreatic trypsin inhibitor lacking cysteines 14 and 38 can fold properly. *Science* 235: 1370–1373, 1987.
29. Martin A, Schmid FX, and Sieber V. Proside: a phage-based method for selecting thermostable proteins. *Methods Mol Biol* 230: 57–70, 2003.

30. McRee DE, Redford SM, Getzoff ED, Lepock JR, Hallewell RA, and Tainer JA. Changes in crystallographic structure and thermostability of a Cu, Zn superoxide dismutase mutant resulting from the removal of a buried cysteine. *J Biol Chem* 265: 14234–14241, 1990.
31. Nilsson B, Berman-Marks C, Kuntz ID, and Anderson S. Secretion incompetence of bovine pancreatic trypsin inhibitor expressed in *Escherichia coli*. *J Biol Chem* 266: 2970–2977, 1991.
32. Ostermeier M, De Sutter K, and Georgiou G. Eukaryotic protein disulfide isomerase complements *Escherichia coli* dsbA mutants and increases the yield of a heterologous secreted protein with disulfide bonds. *J Biol Chem* 271: 10616–10622, 1996.
33. Ostermeier M and Georgiou G. The folding of bovine pancreatic trypsin inhibitor in the *Escherichia coli* periplasm. *J Biol Chem* 269: 21072–21077, 1994.
34. Raines RT. Ribonuclease A. *Chem Rev* 98: 1045–1066, 1998.
35. Raso SW, Abel J, Barnes JM, Maloney KM, Pipes G, Treuheit MJ, King J, and Brems DN. Aggregation of granulocyte-colony stimulating factor *in vitro* involves a conformationally altered monomeric state. *Protein Sci* 14: 2246–2257, 2005.
36. Schwarz H, Hinz HJ, Mehlich A, Tschesche H, and Wenzel HR. Stability studies on derivatives of the bovine pancreatic trypsin inhibitor. *Biochemistry* 26: 3544–3551, 1987.
37. Segatori L, Paukstelis PJ, Gilbert HF, and Georgiou G. Engineered DsbC chimeras catalyze both protein oxidation and disulfide-bond isomerization in *Escherichia coli*: reconciling two competing pathways. *Proc Natl Acad Sci USA* 101: 10018–10023, 2004.
38. Sevier CS and Kaiser CA. Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* 3: 836–847, 2002.
39. Stassinopoulou CI, Wagner G, and Wuthrich K. Two-dimensional 1H NMR of two chemically modified analogs of the basic pancreatic trypsin inhibitor. Sequence-specific resonance assignments and sequence location of conformation changes relative to the native protein. *Eur J Biochem* 145: 423–430, 1984.
40. Swartz JR. Advances in *Escherichia coli* production of therapeutic proteins. *Curr Opin Biotechnol* 12: 195–201, 2001.
41. van Mierlo CP, Darby NJ, Keeler J, Neuhaus D, and Creighton TE. Partially folded conformation of the (30–51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. 1H and 15N resonance assignments and determination of backbone dynamics from 15N relaxation measurements. *J Mol Biol* 229: 1125–1146, 1993.
42. van Mierlo CP, Darby NJ, Neuhaus D, and Creighton TE. (14–38, 30–51) double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: a two-dimensional 1H nuclear magnetic resonance study. *J Mol Biol* 222: 353–371, 1991.
43. van Mierlo CP, Darby NJ, Neuhaus D, and Creighton TE. Two-dimensional 1H nuclear magnetic resonance study of the (5–55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol* 222: 373–390, 1991.
44. Wagner G, Kalb AJ, and Wuthrich K. Conformational studies by 1H nuclear magnetic resonance of the basic pancreatic trypsin inhibitor after reduction of the disulfide bond between Cys-14 and Cys-38. Influence of charged protecting groups on the stability of the protein. *Eur J Biochem* 95: 249–253, 1979.
45. Weissman JS and Kim PS. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat Struct Biol* 2: 1123–1130, 1995.
46. Weissman JS and Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 253: 1386–1393, 1991.
47. Zakharova E, Horvath MP, and Goldenberg DP. Functional and structural roles of the Cys14-Cys38 disulfide of bovine pancreatic trypsin inhibitor. *J Mol Biol* 382: 998–1013, 2008.
48. Zapun A and Creighton TE. Effects of DsbA on the disulfide folding of bovine pancreatic trypsin inhibitor and alpha-lactalbumin. *Biochemistry* 33: 5202–5211, 1994.
49. Zhang JX and Goldenberg DP. Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. *Biochemistry* 32: 14075–14081, 1993.
50. Zhang JX and Goldenberg DP. Mutational analysis of the BPTI folding pathway: I. Effects of aromatic—>leucine substitutions on the distribution of folding intermediates. *Protein Sci* 6: 1549–1562, 1997.

Address correspondence to:
Dr. James C.A. Bardwell
Department of Molecular, Cellular and Developmental Biology
Howard Hughes Medical Institute
University of Michigan
Ann Arbor, MI 48109

E-mail: jbardwel@umich.edu

Date of first submission to ARS Central, October 18, 2010; date of final revised submission, November 28, 2010; date of acceptance, November 28, 2010.

Abbreviations Used

BPTI = bovine pancreatic trypsin inhibitor
DTT = dithiothreitol
hG-CSF = human granulocyte-colony stimulating factor
LB = lysogeny broth, also known as Luria-Bertani broth
MIC = minimal inhibitory concentration
PCR = polymerase chain reaction
PDI = protein disulfide isomerase
pfu = plaque forming unit
PROSIDE = protein stability increased by directed evolution
SD = standard deviation
WT = wild-type

Cytosolic Selection Systems To Study Protein Stability

Ajamaluddin Malik,* Antje Mueller-Schickert, James C. A. Bardwell

Howard Hughes Medical Institute, Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA

Here we describe biosensors that provide readouts for protein stability in the cytosolic compartment of prokaryotes. These biosensors consist of tripartite sandwich fusions that link the *in vitro* stability or aggregation susceptibility of guest proteins to the *in vivo* resistance of host cells to the antibiotics kanamycin, spectinomycin, and nourseothricin. These selectable markers confer antibiotic resistance in a wide range of hosts and are easily quantifiable. We show that mutations within guest proteins that affect their stability alter the antibiotic resistances of the cells expressing the biosensors in a manner that is related to the *in vitro* stabilities of the mutant guest proteins. In addition, we find that polyglutamine tracts of increasing length are associated with an increased tendency to form amyloids *in vivo* and, in our sandwich fusion system, with decreased resistance to aminoglycoside antibiotics. We demonstrate that our approach allows the *in vivo* analysis of protein stability in the cytosolic compartment without the need for prior structural and functional knowledge.

Protein folding has been intensely studied *in vitro*, providing us with a detailed understanding of this process. However, the simplified conditions typically used *in vitro* (i.e., the use of single purified proteins at very low concentrations) differ substantially from the conditions present in the crowded environment in living cells (1). Although many proteins can fold *in vitro* without the assistance of other proteins, most proteins *in vivo* appear to depend on the help of molecular chaperones to fold into their native conformation (2). Protein misfolding *in vivo* has been linked to various disease states, including Alzheimer's disease (3), Parkinson's disease (4), and cystic fibrosis (5), emphasizing the urgent need to better understand the folding process that occurs inside the cell.

In an attempt to get a better understanding of the *in vivo* folding process, several *in vitro* methods have been developed to mimic intracellular conditions. However, replicating the complexity of the cytoplasm in a test tube is a challenging task (6, 7). The difficulty in reproducing the complex environment of living cells has led to the development of methods that allow the measurement of protein stability and folding kinetics inside the cell (8–10). Our lab recently developed biosensors that link protein stability to antimicrobial resistance, allowing the investigation of the periplasmic folding environment (11, 12). Here, we describe the establishment of new protein folding sensors based on proteins involved with resistance to spectinomycin, kanamycin, and nourseothricin (ClonNAT) that expand our approach beyond the periplasm into the more complex folding environment of the cytosol. Using the three guest proteins immunity protein 7 (Im7), human muscle acylphosphatase (AcP), and polyglutamine (polyQ) tracts, we show that the antibiotic resistance conferred by the biosensors is highly correlated with the *in vitro* stability and aggregation susceptibility of the guest proteins. These new biosensors thus provide additional tools to investigate *in vivo* protein folding in the cytosol of a variety of hosts.

MATERIALS AND METHODS

Bacterial strains and expression vectors. *Escherichia coli* strain NEB10β (New England BioLabs) was used for cloning and protein expression. To develop a sandwich fusion system (Fig. 1) based on kanamycin resistance, we used a vector (pAM18) derived from pCR-Blunt II-TOPO (Invitrogen), which carries the gene (*aphA-2*) for the kanamycin resistance protein aminoglycoside-3′-phosphotransferase IIa (APH; GenBank accession number WP_004614937). To eliminate expression of the toxic CcdB product present on this plasmid, we introduced a stop codon (TAA) by mutating the tyrosine at amino acid position 5 of the *ccdB* gene. This was done via site-directed mutagenesis using primers P1 and P2 (see Table S1 in the supplemental material). To increase the proportion of the plasmid taken up by the *aphA-2* gene, which encodes kanamycin resistance, the size of the plasmid was decreased by NsiI digestion followed by religation, which removed a 3.4-kb NsiI fragment. The resulting 3.4-kbp plasmid vector was named pAM15.

To develop a sandwich fusion system based on streptomycin/spectinomycin resistance, the *aadA* gene, which encodes resistance to streptomycin and spectinomycin via an adenylyltransferase [ANT(3″)(9); GenBank accession number ABF67771], was cloned from pBAD43 into pBR322 using the restriction enzyme HindIII, creating pAMS1.

To develop a sandwich fusion system based on nourseothricin resistance, the *nat1* gene encoding nourseothricin acetyltransferase (NAT; GenBank accession number CAA51674), present on plasmid p4339 (obtained from Amy Chang, University of Michigan), was PCR amplified using primers P134 and P135 to introduce EcoRI sites and then ligated into EcoRI-cleaved pBR322 to generate the plasmid pAM103. The ability of the AM101 strain harboring pAM103 to confer nourseothricin resistance was maximal on Terrific broth plates, so this medium was used in experiments involving nourseothricin.

Linker scanning mutagenesis of resistance genes. The overall mutagenesis screen strategy is illustrated in Fig. 2. To identify sites within antibiotic resistance genes that would permit the insertion of pentapeptides, we used a GPS-LS linker scanning system with the conditions recommended by the manufacturer (New England BioLabs). The following describes the procedure used for the *aphA-2* gene (GenBank accession number WP_004614937), which encodes kanamycin resistance. A similar but not identical protocol was followed to identify sites permissive for the

Received 13 August 2014 Accepted 23 September 2014

Published ahead of print 29 September 2014

Address correspondence to James C. A. Bardwell, jbardwel@umich.edu.

* Present address: Ajamaluddin Malik, Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia.

A.M. and A.M.-S. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.02215-14.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.02215-14

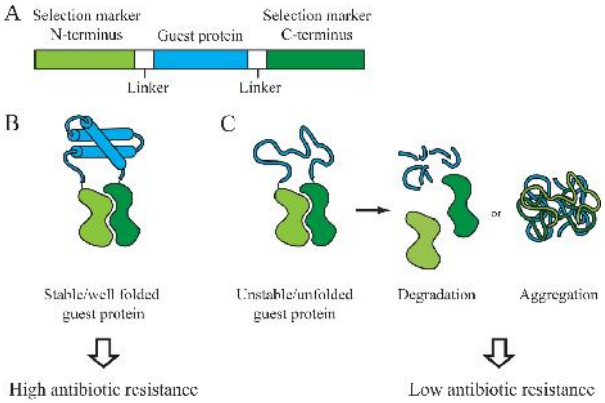


FIG 1 Working principle of the tripartite fusion biosensor for the cytosolic folding environment. (A) The guest protein (blue) is inserted into a permissive site in the antibiotic resistance protein (green), dividing the host protein into an N terminus and a C terminus. The guest protein and the host protein are connected through GS linkers (white), which increase flexibility and allow both proteins to fold into their native conformation. The antibiotic resistance of cells expressing the biosensor is directly correlated to the stability of the guest protein. (B) If the guest protein is stable and well folded, the tripartite system will confer high levels of antibiotic resistance to cells expressing the construct. (C) If the guest protein is unfolded or unstable, the biosensor is rendered nonfunctional by either degradation by cellular proteases or aggregation. Cells expressing this construct have significantly lower antibiotic resistance.

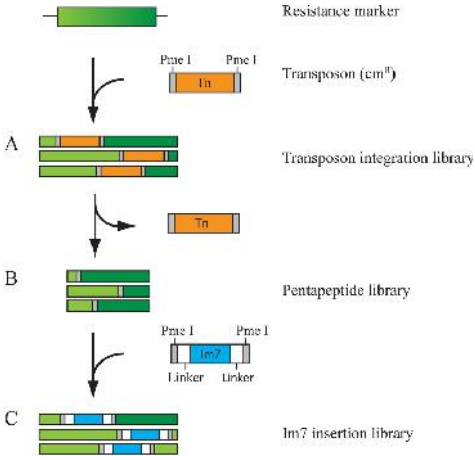


FIG 2 Identification of permissive sites in the resistance marker using the GPS-LS linker scanning system. (A) The transposon (orange) integrates randomly into the coding sequence of the resistance marker (green), creating a transposon integration library. Clones containing the desired transposon insertions in the resistance marker are identified by resistance to chloramphenicol and sensitivity to spectinomycin, kanamycin, or nourseothricin. (B) Digestion of the library with the restriction enzyme PmeI removes the transposon and after religation leaves behind a pentapeptide insertion at the former insertion site. *E. coli* is transformed with the pentapeptide library and selected for regained spectinomycin, kanamycin, or nourseothricin resistance, allowing the growth of cells that express the resistance marker with pentapeptide insertions in permissive sites. (C) The guest protein Im7 is cloned into the pentapeptide library using PmeI, creating the Im7 insertion library. The Im7 library is transformed into *E. coli* and selected for spectinomycin, kanamycin, or nourseothricin resistance.

aadA gene (GenBank accession number ABF67771), which encodes spectinomycin resistance, and the *nat1* gene, which encodes nourseothricin resistance (see the information in the supplemental material).

Twenty nanograms of the transposon donor plasmid pGPS4, which contains transprimer-4, which carries the *cat* gene encoding chloramphenicol resistance (Cm^r) (GenBank accession number WP_002361567), was mixed with 80 ng of the recipient plasmid pAM15, which carries the *aphA-2* gene encoding kanamycin resistance, and the TnsABC transposase enzyme to initiate *in vitro* transposition. The transposase acts to integrate the transposon randomly in the recipient plasmid sequence. NEB10β cells were then transformed with the transposition reaction by electroporation. Cells were incubated for 1 h at 37°C and plated onto plates containing chloramphenicol (15 μg/ml) to select for plasmids that were successful recipients in the *in vitro* transposition reaction. The donor plasmid contains the R6K origin, whose replication requires the π protein to be supplied in *cis*. This plasmid is thus not able to replicate in NEB10β cells. Therefore, only those events that result in successful transposition into the recipient plasmid result in chloramphenicol-resistant colonies.

To screen for those reactions that resulted in insertion of transprimer-4 into *aphA-2* and had thus abolished kanamycin resistance, we simply replica plated the transformants onto Luria-Bertani (LB) plates containing either chloramphenicol alone (15 μg/ml; LB Cm) or both chloramphenicol (15 μg/ml) and kanamycin (100 μg/ml).

Because subsequent steps required the identification of sites that could tolerate the insertion of an open reading frame (ORF) into *aphA-2*, it was very important to eliminate all kanamycin-resistant bacteria at this stage to prevent contamination in subsequent steps. Thus, the kanamycin-sensitive (Kan^s) clones were purified to single colonies and subjected to four rounds of replica plating.

The transposons inserted by the GPS-LS linker scanning system were flanked by PmeI restriction enzyme sites. This design allows the facile removal of nearly all of the inserted *cat* gene, which encodes chloramphenicol resistance, by PmeI digestion followed by religation. The remaining 15-bp transposition scar contains a PmeI restriction site and encodes a tiny 5-amino-acid ORF. This approach allowed us to identify sites in *aphA-2* that would tolerate a 5-amino-acid insertion and to easily introduce test protein ORFs via the PmeI site. The Kan^s clones were grown

overnight with shaking at 37°C in 100 μl LB Cm contained in 96-well microtiter plates that had been covered with AirPore tape sheets (Qiagen). Seventy-microliter aliquots from each of these Kan^s cultures were pooled, and plasmid DNA was prepared and digested with PmeI. The linearized DNA from the Kan^s library was gel purified and religated. This library contained the linker scanners and was therefore named the Kan^{LS} library. It was transformed into NEB10β cells and selected on LB plates containing 200 μg/ml kanamycin, followed by sequencing using primer P3 in order to determine which sites within *aphA-2* could sustain 5-amino-acid insertions and still cause kanamycin resistance.

Cloning of Im7 into the Kan^{LS} library. The *ceiE7* gene (GenBank accession number Q03708) (11), altered so that it would encode the Im7 F84A mutant flanked by glycine-serine (GS) linkers, was amplified from pAM120 using primers P4 and P5. The GS linkers consisted of 17 amino acids each (upstream sequence, GLNGSGSGSGSGSGSGSGS; downstream sequence, GSSSGSSGSGSGGGGSGLNG). The primers were designed to introduce PmeI restriction sites at both ends of the gene encoding Im7, *ceiE7*. The PCR product was ligated into the pCR-Blunt II-TOPO vector (Invitrogen). This clone was then digested with PmeI to isolate the Im7 F84A ORF and subsequently ligated into the Kan^{LS} library, which had been PmeI digested and dephosphorylated. Those sites within *aphA-2* that allowed phenotypic Kan^r even after the insertion of the Im7 ORF were isolated by transformation into electrocompetent NEB10β cells, followed by plating on LB plates that contained 100 μg/ml kanamycin. The sites within *aphA-2* that were permissive for the Im7 F84A insertion were identified by sequencing using primer P3. The Im7 F84A mutation inserted into amino acid position 55 of *aphA-2* (*aphA-2aa55*; present on pAM108) was reversed to generate the wild-type (WT) Im7 gene *ceiE7* using primers P6 and P7, generating pAM107. To further optimize APH as a biosensor, we evaluated strains containing either Im7 F84A (strain AM236) or wild-type Im7 (strain AM235) for differences in antibiotic resistance after growth on various media (LB, Terrific broth, MacConkey agar, nutrient broth, and M63), at various temperatures (37°C and 42°C), in various strain backgrounds (strains BL21 and MG1655), and with 4 different kanamycin-related antibiotics (kanamycin, neomycin, paramomycin, and gentamicin). The conditions that showed the largest difference in antibiotic resistance between tripartite fusions containing the unstable Im7 F84A mutant and wild-type Im7 were those in which the BL21(DE3)-RIPL strain was grown in LB medium at 37°C, where a very satisfying 5-log-unit difference in cell survival was seen over a nearly 2,000-μg/ml range (see Fig. S2 in the supplemental material). Subsequently, individual point mutants of known thermodynamic stability (F15A, V33E, L34A, L53A, I54V, N26K-S58R, V69A, N26K-T30N-S58R, L3A, D35N-D63N, V27A-D63N, L18F-D35N-D63N, L18A) were introduced into the Im7 WT tripartite fusion by site-directed mutagenesis using primers P8-P9, P10-P11, P12-P13, P14-P15, P16-P17, P46-P47-P48-P49, P52-P53, P44-P45-P48-P49, P30-P31, P25-P26-P36-P37, P25-P26-P40-P41, P38-P39-P25-P26-P36-P37, and P42-P43, respectively. Various variants of Im7 were introduced into the pentapeptide permissive sites in the APH and NAT proteins in a similar manner. The derivation of these plasmids is described in Table S5 in the supplemental material.

Replacing Im7 with another guest protein (AcP) in the spectinomycin resistance marker. Digestion of the pBR322 plasmid that contains the *aadA* gene with the *ceiE7* gene encoding the guest protein Im7 inserted after amino acid 155 of ANT (pAMS2) with XhoI and SacI removes only *ceiE7* from the plasmid and leaves the majority of the GS linker in the plasmid, creating pAMS3 (see Fig. S1 in the supplemental material). This simplifies the swapping of different guest proteins in the biosensor. Note that the digest shortens the original upstream GS linker in pAMS3 from 17 to 11 amino acids, followed by the XhoI restriction site (which encodes Leu and Glu); the downstream linker is similarly shortened from 17 to 13 amino acids and is preceded by a SacI restriction site (which encodes Glu and Leu). The *acpY2* gene, encoding the enzyme AcP (GenBank accession number P14621), was thus cloned into pAMS3 using XhoI and SacI, creating pAMS34. This construct was further modified using primers P52 and P53 in order to create a cysteine-less variant of AcP (C21S) called pAMS35. The use of this cysteine-less variant, which was present in all subsequent AcP constructs, facilitates comparison with previously published stability data (13). This pseudo wild-type variant is referred to as the WT throughout the study. pAMS35 therefore has the biosensor with *acpY2* inserted after amino acid position 155 of *aadA*. AcP variants Y11F, V20A, M61A, L65V, and E83D all contain mutation C21S and were created using site-directed mutagenesis with primers P56-P57, P58-P59, P60-P61, P62-P63, and P64-P65, respectively (plasmids pAMS36 to pAMS40; see Table S5 in the supplemental material).

Cloning of polyglutamines at position 55 in the *aphA-2* gene, which encodes APH. To facilitate the insertion of additional peptides at permissive position 55 within the APH protein, a unique BamHI site was introduced at the corresponding position of the *aphA-2* gene in pAM15 by site-directed mutagenesis using primers P18 and P19. Different lengths of polyQ-encoding DNA flanked by BamHI ends was amplified from p416/PQ103 (14) using primers P20 and P21. PCR products ranging in size from 50 to 500 bp were isolated from a 1% preparative agarose gel. Subclones of polyQ-encoding DNA of various lengths were made in the pCR-Blunt II-TOPO vector. Three different lengths of polyQ inserts (20, 45, and 87 Q residues), which were isolated by BamHI digestion and ligated into a BamHI-linearized dephosphorylated vector followed by transformation into NEB10β cells, were identified by sequencing. The corresponding plasmids were called pAM79, pAM80, and pAM81. We noticed that the polyQ87 tract in pAM81 was somewhat unstable, readily generating N-terminal 40-nucleotide deletions in APH. To prevent these specific recombinations, several silent mutations were generated at positions 41 to 43 of APH in pAM81 using primers P22 and P23 to generate pAM180. This plasmid was stable, so the changes appear to have reduced homologous recombinations between the ORF and its promoter.

MIC determination in *E. coli*. Antibiotic resistance was measured in terms of the MIC, i.e., the lowest concentration of an antibiotic that inhibits the growth of a microorganism. MIC experiments were performed as described by Foit et al. (11) with the following modifications. Prior to spot titer experiments for kanamycin or nourseothricin resistance, cultures were made by inoculating a single colony into 5 ml LB containing either 25 μg/ml zeocin or 200 μg/ml ampicillin, followed by incubation overnight at 37°C without shaking to obtain an optical density (OD) at 600 nm (OD_{600}) of ~0.1 to 0.2. Note that this overnight selection was not for the partially crippled antibiotic resistance marker but, rather, was for the alternative antibiotic resistance marker present on the plasmid. Following this standing overnight growth, tubes were transferred to a shaking incubator at 37°C, where cells rapidly entered log-phase growth and after 2 to 3 h had reached late log phase (OD_{600} = 1 to 1.5). Spot titer experiments were done on prewarmed LB agar plates containing increasing concentrations of kanamycin or nourseothricin. The growth of cultures for determining MICs for spectinomycin resistance was done in a slightly different fashion. Single colonies were inoculated from plates and grown overnight in culture tubes containing 5 ml LB and ampicillin (100 μg/ml) in a rotary drum incubator at 37°C. Overnight cultures were diluted 1:100 and grown to log phase under the same conditions. Cells were pelleted and adjusted to an OD_{600} of 1 with phosphate-buffered saline (PBS). Cells were serially diluted, and dilutions of 10^0 to 10^{-5} were spotted on LB plates containing increasing concentrations of spectinomycin.

After incubation overnight at 37°C, MICs for every variant were determined by evaluating the growth of each dilution on antibiotic at each concentration. The average of the MICs of dilutions of 10^{-1} to 10^{-5} was calculated. Experiments were repeated at least three times, and the results are shown as means ± standard errors of the means.

Fractionation of *E. coli* extracts. Cultures expressing *aphA-2aa55*::*ceiE7* variants were grown overnight in LB containing 25 μg/ml zeocin. The culture density was then adjusted to an OD_{600} of 3.5 by adding 10 mM Tris, pH 8.0. For soluble protein extraction, 0.5 ml of these cultures was pelleted at 15,700 × *g* for 10 min, and the pellets were resuspended in 500 μl lysis buffer (1 mg/ml lysozyme, 5 mM EDTA in Tris-buffered saline).

One hundred microliters of 0.1-mm-diameter glass beads (BioSpec) was added to each tube, and the tubes were incubated on ice for 30 min, during which time they were vortexed four times at top speed for 30 s each time to lyse the cells and break up big aggregates. Aggregates were separated by centrifugation at 15,700 × *g* for 10 min. Aggregated proteins in the pellet were extracted by boiling the pellet for 5 min in 600 μl of 1× SDS loading buffer. Two hundred microliters of the supernatant was mixed with 40 μl of 5× SDS loading buffer and denatured by boiling for 3 min at 95°C. This fraction was designated the soluble fraction. Samples from the soluble and insoluble fractions (15 μl each) were loaded on precast 4 to 12% bis-Tris polyacrylamide gels. Proteins were blotted on an iBlot system (Invitrogen) according to the manufacturer's instructions. Protein fusions containing the kanamycin resistance protein were detected with Western blots using anti-neomycin phosphotransferase II (Sigma) antibody at a 1:10,000 dilution. Horseradish peroxidase-conjugated goat anti-rabbit IgG at a 1:10,000 dilution was used as the secondary antibody. SuperSignal West Pico chemiluminescent substrate (Pierce) was used to visualize antibody binding.

For fractionation of the polyQ sandwich fusion proteins, 0.2-ml samples from the overnight cultures (OD_{600} = 5) were incubated with lysis buffer on ice. To break the cells, freezing and thawing were performed for 3 rounds by cycling between a dry ice ethanol bath and a water bath held at 37°C. The supernatant obtained after centrifugation at 15,700 × *g* for 10 min was considered the soluble fraction. The insoluble pellet was dissolved in 240 μl 1× SDS loading buffer, and 40 μl 5× SDS loading buffer was added to the 200-μl supernatant. Samples were denatured by boiling at 95°C for 3 min, and 20 μl from both fractions was loaded onto precast 4 to 12% bis-Tris gels. Western blotting and development were done as described above.

Cloning of *aphA-2aa55*::*ceiE7* variants and polyQ sandwich fusions in *Saccharomyces cerevisiae*. Primers P24 and P25 were designed to amplify sandwich fusions from the pTOPO vector, and these fusions were cloned into the pYES2.1 TOPO TA expression shuttle vector. Clones were sequenced with primers P26 and P27.

MIC determination in *S. cerevisiae*. Sandwich fusions were transformed into the INVSc1 strain of *S. cerevisiae* expressed in synthetic complete (SC) minimal broth medium minus uracil (SC−U) supplemented with 2% galactose under the control of the *GAL* promoter for 48 h. All cultures were adjusted to an OD_{600} of 0.2. Fully induced cultures (3.2 μl) were inoculated in 100 μl yeast extract-peptone dextrose (YPD) medium containing various concentrations of G418 in 96-well plate formats. Microtiter plates were sealed with tape pads (Qiagen) and incubated at 30°C for 20 h. The OD_{600} was measured with a Synergy HT plate reader (BioTek Instruments, Inc.).

Analysis of sandwich fusions in *S. cerevisiae*. Sandwich fusions with Im7 variants or polyQ were expressed in SC−U supplemented with 2% galactose for 48 h. Duplicate 0.2-ml samples from the cultures were adjusted to an OD_{600} of 5. Cells were pelleted by centrifuging at 6,500 × *g* for 5 min. For soluble protein extraction, pellets were treated with 200 μl yeast lysis solution (Amersham) according to the manufacturer's instructions. For total protein analysis, the pellet was dissolved in 240 μl 1× SDS loading buffer. Glass beads (~100 μl; diameter, 0.5 mm) were added to each of the tubes. The tubes were incubated on ice and vortexed at high speed six times for 30 s each time. For soluble protein extraction, centrifugation was done at 15,700 × *g* for 10 min. One hundred microliters supernatant was aspirated, and 20 μl 5× SDS loading buffer was added. Samples were boiled at 95°C for 3 min. For the loading control, 15-μl samples were loaded on 7% Tris-acetate gels, and for Western blot analysis, 20-μl samples were loaded on 4 to 12% bis-Tris gels. Blotting and development were done as described above.

RESULTS

Development of a protein folding biosensor. The principle of a sandwich folding biosensor is illustrated in Fig. 1. It involves the insertion of a guest protein (shown in blue) into the middle of a selectable marker. The concept is that if the guest protein is folded properly, it should bring the N- and C-terminal halves of the selectable marker close together, enabling the marker to fold properly and thus function. However, if the guest protein is unstable, it will be cleaved by the plethora of proteases present *in vivo* or it will be prone to aggregation. Both proteolysis and aggregation should, in theory, decrease the activity of the biosensor. In principle, if one is able to couple *in vivo* proteolysis or aggregation to an appropriate selectable marker, one should be able to study protein stability and aggregation propensity *in vivo*. If the selectable marker is efficient enough, one should also be able to select for more stable or less aggregation-prone variants and to identify host mutants that show improved folding environments. We previously developed biosensors based on two periplasmic proteins: the β-lactamase protein, which confers ampicillin resistance, and the DsbA protein, which catalyzes disulfide bond formation and confers cadmium resistance. The selective power of this approach allowed the facile isolation of stabilized protein variants (11) and the discovery of a new periplasmic chaperone called Spy (11, 15). Although powerful, these systems cannot be utilized to investigate the folding environment of the cytoplasm because export to the periplasm is essential for β-lactamase function and for the cadmium resistance encoded by the DsbA gene. We therefore decided to search for selectable markers that could expand our approach into the cytosolic compartment and potentially into other organisms as well.

Antibiotic resistance markers are ideal for our purposes because antibiotics very effectively kill cells or inhibit their growth, their resistance markers often provide very high levels of protection, and they often work on a wide variety of species. Antimicrobial resistance can easily be determined in spot titer experiments, and the effects of different conditions or mutations on protein stability or aggregation propensity can be compared. We decided to investigate markers for kanamycin, spectinomycin, and nourseothricin resistance. The aminoglycoside-3′-phosphotransferase IIa (APH) gene, which encodes kanamycin resistance in bacteria and G418 resistance in yeast, was chosen because its structure is known and it is commonly used as a selection marker (16, 17). The streptomycin/spectinomycin adenylyltransferase ANT(3″)(9) protein (18) was similarly chosen because spectinomycin and streptomycin are also commonly used, inexpensive, and readily available (19). Resistance to the antibiotic nourseothricin (20) is conferred by nourseothricin acetyltransferase (NAT). This antibiotic, though much more expensive and less available than kanamycin or spectinomycin, was chosen because it is effective not just against Gram-negative bacteria but also against Gram-positive bacteria, mycobacteria, mycoplasmas, protista, yeasts, and plants (21–24).

For the sandwich fusion approach to be successful, we needed to find sites within these antibiotic resistance markers that would tolerate the insertion of folded proteins and be responsive to the stability or aggregation susceptibility of the inserted protein, allowing us to link antibiotic resistance to protein stability or aggregation susceptibility. These types of sites are presumably quite rare, as most insertions are disruptive to protein function (25). We decided to develop a general protocol to find sites within selectable markers that would tolerate insertions and then to find within this set those in which the stability of the inserted protein determined the selectable marker's activity. To accomplish this, we first constructed a library of insertions within the antibiotic resistance

genes, screened them for sites that would tolerate short amino acid insertions, and then screened these for sites that would tolerate the insertion of a complete protein. Finally, we screened these sites for those in which the stability of the inserted protein determined the level of antibiotic resistance. This procedure is diagramed in Fig. 2 and described in detail in the Materials and Methods.

To construct libraries of insertion sites within the antibiotic resistance genes, the GPS-LS linker scanning system, a transposon-based mutagenesis system commercially available from New England BioLabs, was used. This system enables the rapid construction of an extensive set of random insertions of a mini-transposon consisting of the chloramphenicol resistance marker flanked by PmeI sites into plasmid DNA.

The development of a kanamycin resistance-based stability biosensor is described here in detail. Stability biosensors based on other antibiotic resistance markers were generated in a similar fashion and are described in the Materials and Methods and in the information in the supplemental material.

We first transposed a chloramphenicol resistance (Cm^r) marker from the transposon donor plasmid pGPS4 into a recipient plasmid, pAM15, which contains the kanamycin resistance marker. We screened the ~30,000 chloramphenicol-resistant insertions that we obtained for those that were kanamycin sensitive in order to obtain a library of clones containing insertion sites within the *aphA-2* gene, which normally encodes kanamycin resistance. After four rounds of purification, 1,800 insertions verified to be kanamycin sensitive were obtained. The insertion sites were determined for 50 of these insertions. All were within *aphA-2* and its promoter region, and all were in different sites, consistent with previous observations that the Tn*9* transposase, on which the GPS-LS linker scanning system is based, inserts with very little sequence specificity. Because the number of insertions that we obtained ($n = 1,800$) exceeded the number of codons ($n = 264$) in the *aphA-2* gene by a factor of ~6, we reasoned that our insertion library was likely reasonably complete.

The transposons were excised from the pooled library, leaving behind a pentapeptide insertion carrying a PmeI site. Those sites within APH that could tolerate insertions of 5 amino acids and still confer kanamycin resistance were readily obtained simply by plating the religated library pool on medium containing 200 μg/ml kanamycin (26). Determination of the insertion site within 46 randomly picked Kan^r clones (see Fig. S2 in the supplemental material) indicates that the majority of the pentapeptide permissible insertion sites are present in surface loops in the N-terminal domain of the APH protein. We then discovered which of these sites could tolerate larger insertions by simply recleaving the Cm^r Kan^s library of all 1,800 clones with PmeI, followed by the kanamycin-resistant insertional cloning of a larger open reading frame. For a guest protein, we initially picked immunity protein 7 (Im7), a 10-kDa, helical *E. coli* protein. We chose Im7 as a model protein because of the ready availability of Im7 variants with a wide spectrum of known thermodynamic stabilities (11, 27). This facilitated the following step in our procedure: namely, to test different permissive sites for their correlation between antibiotic resistance and protein stability. For initial experiments, we chose to investigate a very destabilized variant, Im7 F84A. Im7 F84A was chosen as a guest protein because it is unstable *in vivo* and *in vitro* but is still capable of expressing significant resistance to penicillin when tested in the β-lactamase stability biosensors (11, 27). We ligated the plasmids with pAM108, a construct consisting of Im7 F84A and flanking glycine-serine (GS) linkers at its C and N termini (see Fig. S1 in the supplemental material). The linkers were designed to provide flexibility and reduce the steric interference of the fusion proteins, allowing them to fold independently into their native conformations.

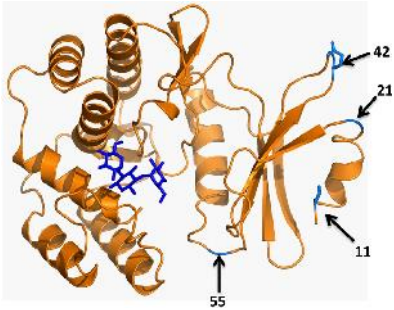


FIG 3 (A) Sites in the structure of aminoglycoside-3′-phosphotransferase IIa (PDB accession number 1ND4), which confers resistance to kanamycin, that we found to be suitable for the insertion of proteins.

Four sites at positions 11, 21, 42, and 55 within the APH protein that tolerate the insertion of Im7 F84A and still exhibit resistance to at least 100 μg/ml kanamycin were found (Fig. 3; see also Fig. S2 in the supplemental material). The APH protein, present in plasmid pAM108, with insertions of Im7 at position 55 was selected for further analysis.

Optimization of insertion sites. For a stability biosensor to be effective, it should give a linear readout over a wide range of stabilities of the guest protein. For biosensors derived from the APH protein, this means that there should be a large difference in kanamycin resistance for sandwich fusions containing inserted guest proteins of various stabilities, and their kanamycin resistance should be directly related to the stabilities of the guest proteins. Thus, we restored the wild-type Im7 sequence in our Im7 F84A construct and then introduced into Im7 a wide range of mutations that are known to affect Im7 protein stability (see Tables S5 and S2 in supplemental material). The relative MICs conferred by different sandwich fusions showed a very nice relationship with the stabilities of Im7 variants for all the destabilized variants and the wild-type protein (Fig. 4A). Variants of Im7 with stabilities greater than the stability of the wild type, however, showed MIC values that scattered around the wild-type MIC, indicating that the biosensors' capacity to measure stabilities may saturate for proteins with stabilities greater than that of wild-type Im7 (−24.9 kJ/mol) (27).

The amount and solubility of the sandwich fusions were evaluated via Western blotting using both whole-cell lysates and soluble fractions (Fig. 4B and C; see also Fig. S3A and B in supplemental material). We observed increased aggregation of the sandwich fusion as the stability of the guest protein inserted into the tripartite fusion decreased. It has previously been observed that when polyglutamine tracts longer than ~35 residues are present in proteins, they have a strong tendency to form amyloids *in vitro* and *in vivo*, with an increasing length of polyglutamine being associated with more severe amyloidogenesis (28). We inserted

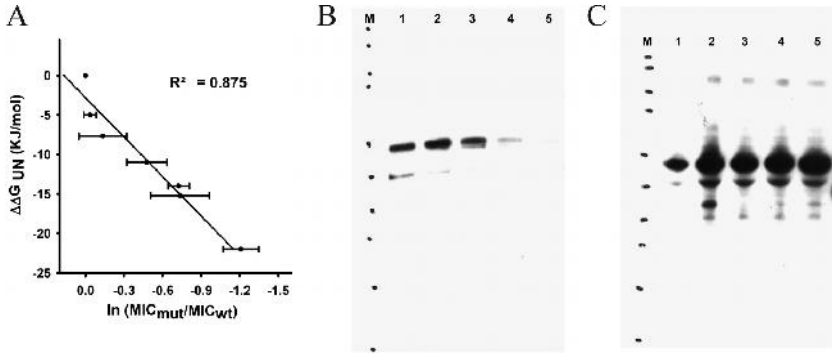


FIG 4 Relationship between stabilities of Im7 guest proteins and kanamycin resistance in *E. coli* BL21(DE3)-RIPL. (A) Position 55 in APH (which causes kanamycin resistance) was evaluated by inserting different thermodynamically destabilized mutants of Im7. The kanamycin resistances of cells containing the fusions with the destabilized Im7 mutants relative to those of cells containing WT Im7 correlated well with the stabilities of the guest proteins. From left to right, the strains used were AM235, AM240, AM258, AM243, AM242, AM247, and AM236, respectively. UN, unfolding; MIC_{mut}, MIC for the mutant; MIC_{wt}, MIC for the WT. (B and C) Western blots showing the amount and solubility of the sandwich fusions in the soluble fraction (B) and in whole-cell lysates (C). Lanes: M, molecular size marker; 1, WT Im7 fused at position 55 (AM235); 2, Im7 V33E fused at position 55 (AM240); 3, Im7 L34A fused at position 55 (AM258); 4, Im7 F15A fused at position 55 (AM247); and 5, Im7 F84A fused at position 55 (AM236).

polyglutamine tracts of three different lengths (20, 45, and 87 Q residues) at position 55 in APH and evaluated their effect on the solubility of the resulting sandwich fusions and on their ability to cause kanamycin resistance. As shown in Fig. 5, as the length of the polyglutamine tracts within sandwich fusions increased, the kanamycin resistance of cells expressing these fusions decreased (Fig. 5A); the amount of soluble material also decreased (Fig. 5B). Conversely, the proportion of SDS-insoluble material expressed increased with increasing polyglutamine tract length.

This APH biosensor thus appears to function as a readout for the aggregation propensity of polyglutamine tracts in the range of 20 to 87 amino acids and as a readout for Im7 stability in the range of 0 to −24.9 kJ/mol. The periplasmic β-lactamase-based stability biosensor that we previously developed was sensitive over a broader range. We wondered if biosensors based on resistance to another antibiotic might be responsive to a broader range of protein stabilities, so we repeated the biosensor development process using the streptomycin/spectinomycin adenylyltransferase ANT(3″) (9) protein, which confers spectinomycin/streptomycin resistance.

Identification of pentapeptide permissive sites in ANT. Pentapeptide permissive sites could be identified in a fashion similar to that described for APH by selecting colonies for regained spectinomycin resistance and identifying permissive sites by sequencing. While the transposon insertions appeared to be randomly distributed throughout *aadA*, pentapeptide permissive sites were

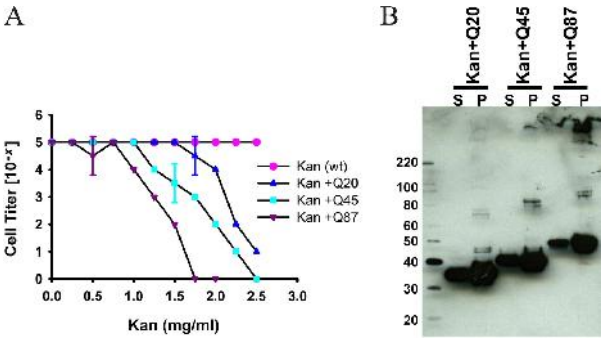


FIG 5 Correlation between amyloidogenicity and kanamycin resistance in *E. coli*. (A) Polyglutamine tracts of three different lengths (20, 45, and 87 Q residues, designated by +Q20, +Q45, and +Q87) were directly fused at position 55 in APH, and the resulting sandwich fusions were expressed in NEB10β. Cell survival at maximum dilutions was scored for different concentrations of kanamycin. The Kan WT is strain AM144, Q20 is strain AM200, Q45 is strain AM201, and Q87 is strain AM308. (B) Western blots showing the distribution of polyQ sandwich fusion proteins in the soluble fraction (lanes S) and in the corresponding cell pellets of the strains (lanes P). Numbers on the left are molecular size markers (in kilodaltons).

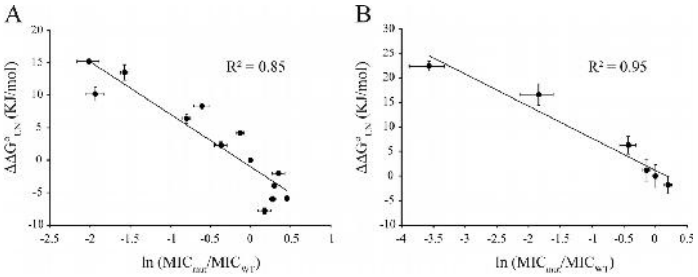


FIG 6 Correlation between antibiotic resistance and protein stabilities of Im7 and AcP variants in ANT. The *in vitro* stabilities of Im7 variants (A) and AcP variants (B) were plotted against the MICs for bacteria expressing the respective variants in the tripartite fusion Im7, which were (from left to right) strains AMS6, AMS15, AMS11, AMS5, AMS4, AMS8, AMS10, AMS2, AMS14, AMS12, AMS13, AMS7, and AMS9, and AcP, which were (from left to right) strains AMS37, AMS40, AMS38, AMS39, AMS35, and AMS36. The *in vitro* protein stabilities were taken from the literature (13, 27). MICs for cells expressing the constructs were measured in spot titer experiments. For both guest proteins, we see a significant correlation between stability and spectinomycin resistance.

found in only three distinct hot spots in the protein (see Fig. S4 in the supplemental material).

Im7 permissive sites in ANT show a distribution pattern similar to that of pentapeptide permissive sites. We again reasoned that only sites that tolerate a pentapeptide insertion would be likely to accept the insertion of a model protein. Im7 with linker sequences was thus cloned in the PmeI site carried by the pentapeptide insertion. The library containing Im7 in ANT pentapeptide permissive sites was transformed and selected on plates containing 100 μg/ml spectinomycin. Plasmids from resistant colonies were sequenced in order to identify Im7 permissive sites. Sequencing of resistant colonies showed that sites tolerating Im7 insertions also accumulated in three hot spots in ANT in a pattern very similar to that found for the pentapeptide library (see Fig. S4C and D in the supplemental material), allowing us to conclude that many sites that allow the insertion of a pentapeptide also allow the insertion of a full-length protein. The three hot spots in ANT that permitted Im7 insertions are positions 75 to 81, 152 to 157, and 289 to 317 (see Fig. S4C and D in the supplemental material). The last two groups of insertions were, as expected, in frame. Surprisingly, many of the insertions located in the N terminus (positions 75 to 81) caused frameshifts.

To further investigate this result, we screened 500 clones for spectinomycin resistance by letting them grow overnight in increasing concentrations of spectinomycin. By measuring the final ODs of the overnight cultures, we separated the insertions into two subgroups: (i) those with high spectinomycin resistance (for which MICs were indistinguishable from the MIC obtained for wild-type ANT) and (ii) those with moderate spectinomycin resistance. All the clones with high spectinomycin resistance had in-frame insertions and were located in the N terminus of ANT (see Fig. S4C in the supplemental material). The subgroup with significantly lower resistance had insertions in the N terminus, the middle, or the C terminus of the protein (see Fig. S4D in the supplemental material). We reasoned that clones with frameshifts in the N terminus might express truncated but still fully functional versions of ANT that use an alternate initiation codon that lies downstream of the Im7 insertion site. If this were the case, it seemed unlikely that the insertion of proteins with various stabilities at these positions would alter spectinomycin sensitivity. Therefore, we focused on insertion sites in the middle and the C terminus of the protein.

A spectinomycin-based biosensor links antibiotic resistance and *in vitro* stability for different guest proteins. Initially, we chose an insertion site in the middle of ANT (at position 155) present on plasmid pAMS2, to test the relationship between the *in vitro* stability of the Im7 guest protein and the antibiotic resistance of cells expressing the fusion protein. Using site-directed mutagenesis and pAMS2 as a template, we generated Im7 variants (pAMS4 to pAM15) covering a wide stability range (see Table S2 in the supplemental material). We then measured the MICs of cells expressing those fusion proteins (AMS2 and AMS4 to AMS15). Figure 6A shows that there is a very good correlation between the *in vitro* stability of the tested Im7 variants and the antibiotic resistance of cells expressing the corresponding fusion proteins. Variants that are destabilized compared to the stability of the WT *in vitro* show low antibiotic resistance, and in contrast to the kanamycin resistance-based protein stability biosensor, stabilized variants have resistances higher than the resistance of the wild type. Thus, it appears that the spectinomycin-based stability biosensors have a significantly broader range than the kanamycin-based ones.

The spectinomycin-based system contains XhoI and SacI restriction sites, placed close to Im7 within the GS linker sequence, to facilitate the easy replacement of Im7 with different guest proteins (Fig. S1 in the supplemental material). To determine if this spectinomycin-based system functions as a stability biosensor with a different guest protein (in addition to Im7), we tested its functionality with human muscle acylphosphate (AcP). This 98-amino-acid enzyme was chosen because the *in vitro* stabilities of many of its variants are known (13, 29). We switched the Im7 protein in our ANT tripartite fusion (pAMS2) with wild-type AcP (creating pAMS34) (see Fig. S1 in the supplemental material) and then used site-directed mutagenesis to generate a series of AcP variants (pAMS35 to pAMS40; see Table S5 in the supplemental material) with a wide range of *in vitro* stabilities (see Table S4 in the supplemental material). The spectinomycin MICs of these constructs measured in spot titer experiments were very well correlated to the *in vitro* stabilities of the inserted AcP protein variants (13) (Fig. 6B), showing that this system is not limited to Im7.

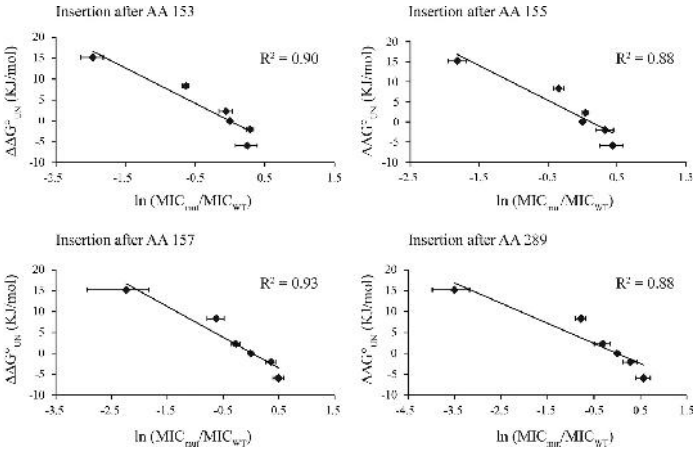


FIG 7 Four different permissive sites in ANT were evaluated with six Im7 variants spanning a wide range of *in vitro* stabilities. The MIC values were determined for the following strains, from left to right: for the insertion site at amino acid (AA) 153, strains AMS18, AMS19, AMS17, AMS16, AMS20, and AMS21; for the insertion site at amino acid 155, AMS6, AMS4, AMS8, AMS2, AMS7, and AMS9; for the insertion site at amino acid 157, AMS24, AMS25, AMS23, AMS22, AMS26, and AMS27; and for the insertion site at amino acid 289, AMS30, AMS31, AMS29, AMS28, AMS32, and AMS33. Independently of the exact location of the permissive site in ANT, we found that all biosensors showed a clear correlation between the *in vitro* stabilities of the Im7 inserts and the antibiotic resistance of the corresponding fusions.

To test whether our spectinomycin-based biosensor results depend on the position of the insertion site, we repeated the Im7 experiments for three additional permissive sites: two in the middle (positions 153 and 157) and one at the C terminus (position 289) of ANT (Fig. 7). MICs were measured in strains AMS2, AMS4, and AMS6 to AMS9 for insertion site 155, AMS16 to AMS21 for insertion site 153, AMS22 to AMS27 for insertion site 157, and AMS28 to AMS33 for insertion site 289 (see Table S5 in the supplemental material). For all tested sites, a very clear correlation between the *in vitro* stability of the insert and the *in vivo* antibiotic resistance of the corresponding construct was shown. For all four sites tested, there was a drop-off in apparent sensitivity for the most stable Im7 variant (Y11F; $\Delta\Delta G^0$, 1.8 kJ/mol) (13). This suggests that the spectinomycin-based stability biosensors may, like the kanamycin-based sensors, also work over a broader but still limited range of corresponding *in vitro* stabilities.

Toward the development of a stability biosensor that functions in the eukaryote *S. cerevisiae*. Finally, we attempted to develop a stability biosensor that functions in yeast. The APH protein confers resistance to the aminoglycoside antibiotic G418 in *S. cerevisiae* (30). Six different APH sandwich fusions containing Im7 guest proteins with various stabilities were cloned behind the *GAL* promoter and grown with induction using 2% galactose for 48 h. Following induction, ~2×10^4 yeast cells were inoculated into 100 μl YPD medium containing increasing concentrations of G418 (0 to 53 mg/ml) and grown at 30°C for 20 h; growth was monitored by measuring the OD_{600} (see Fig. S5 in the supplemental material). Growth inhibition by G418 of *S. cerevisiae* cells expressing these different sandwich fusions was linearly correlated with the stabilities of the Im7 guest proteins in the fusion; i.e., fusions containing proteins with greater stability showed greater antibiotic resistance (Fig. 8A).

The amount and solubility of the sandwich fusions were quantified via Western blotting using both whole-cell lysates and soluble fractions (Fig. 8B and C; see also Fig. S6 in the supplemental material). The amount of soluble material varied directly with the stability of the inserted Im7 protein, whereas the total amount of sandwich fusion remained constant. We next tested the solubility and activity of the APH sandwich fusions containing polyglutamine tracts. *S. cerevisiae* cells expressing polyQ inserted at amino acid 20 (polyQ20) or polyQ87 under galactose control were inoculated into 100 μl YPD containing 0 to 53 mg/ml of G418. After 20 h of incubation at 30°C, the growth of *S. cerevisiae* cells expressing fusions containing polyQ87 was significantly lower in the presence of G418 than that of yeast cells expressing fusions containing polyQ20 (Fig. 9A). Western blots (Fig. 9B and C) revealed that the total concentration of material was similar for both sandwich fusions, but their solubility differed, with the fusion containing polyQ87 being considerably less soluble (Fig. 9B). We do not have any specific structural information about how the polyQ tracts behave in the context of the fusion; we do know, however, that as the polyQ tracts get longer, the antibiotic resistances caused by the corresponding fusions go down and the solubility of the corresponding fusions also goes down.

These results suggested that APH/NAT-based biosensors could be used to monitor protein stability and the aggregation tendency in yeast. The differences in OD exhibited in Fig. 9A, however, are substantially less than the >5-log-unit difference in the numbers of CFU seen with the kanamycin selection in *E. coli* (Fig. 3B). We therefore attempted to develop a stability biosensor based on nourseothricin acetyltransferase, which causes resistance to the antibiotic nourseothricin in a wide range of prokaryotic as well as eukaryotic hosts. Using a procedure similar to that used for the kanamycin and spectinomycin resistance markers, we identified

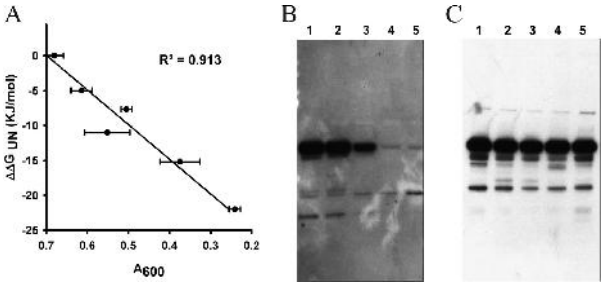


FIG 8 Relationship between stabilities of guest proteins and growth inhibition of *S. cerevisiae*. (A) APH sandwich fusions containing Im7 guest proteins with different stabilities were cloned behind the *GAL* promoter, and cells were grown using 2% galactose. Twenty thousand fully induced yeast cells were inoculated into 100 μl YPD containing increasing concentrations of the antibiotic G418 and incubated at 30°C for 20 h. The growth of *S. cerevisiae* at 400 μg/ml G418 was plotted against the *in vitro* stabilities of the Im7 variants. From top to bottom, the strains were AYC29, AYC34, AYC35, AYC39, AYC30, and AYC42, carrying the WT and Im7 variants V33E, L34A, I54V, F15A, and F84A, respectively. (B and C) Western blots quantifying sandwich fusion material obtained from soluble fractions (B) and whole-cell lysates (C). Lanes 1 to 5, strains AYC29 (lanes 1), AYC34 (lanes 2), AYC35 (lanes 3), AYC30 (lanes 4), and AYC42 (lanes 5), with WT Im7 and Im7 variants V33E, L34A, F15A, and F84A at position 55 in APH.

two permissive sites in the NAT protein that resulted in a very significant drop in resistance to nourseothricin in *E. coli* compared to that obtained with the protein with wild-type Im7 when they contained the unstable Im7 F15A insert (see Fig. S7 in the supplemental material). In both cases, cells containing wild-type and hyperstable variants of Im7 had similar nourseothricin resistances, suggesting that, similar to kanamycin resistance in *E. coli*, the *in vitro* stability exhibited by wild-type Im7 (15.2 kJ/mol) may represent the limit of the assay. Unfortunately, yeast cells containing our various tripartite fusions were equally and extremely nourseothricin resistant. Concentrations of nourseothricin as high as 5,000 μg/ml on plates showed no effect on yeast growth, independently of the NAT Im7 fusion that it contained. This was true when these fusions were cloned on both high-copy-number (pYES-derived) and low-copy-number (pYC-derived) plasmids. In liquid culture, growth could not be completely inhibited even at 33,000 μg/ml nourseothricin, and tripartite fusions showed very similar growth curves under nourseothricin selection pressure, irrespective of the *in vitro* stability of the Im7 that they contained.

DISCUSSION

Levels of protein expression are influenced by the folding capacity of the host and the intrinsic stability or aggregation propensity of the protein of interest (11, 31, 32). To study protein stability *in vivo*, we previously developed a selection system based on sandwich fusions to the TEM-1 β-lactamase protein, which causes re-

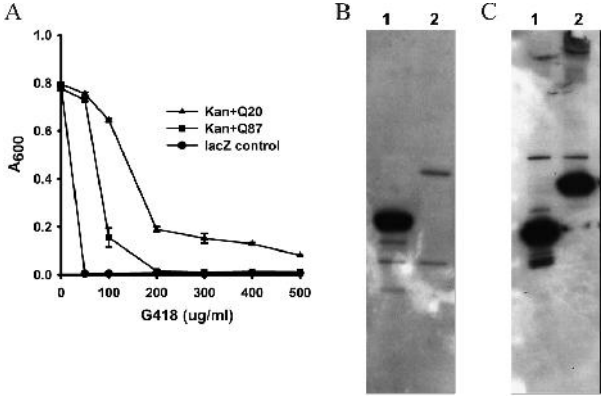


FIG 9 Relationship between the length of polyQ tracts in the fusion and growth inhibition of *S. cerevisiae*. (A) APH sandwich fusions containing polyQ20 or polyQ87 inserts were grown in yeast cells under galactose control. Twenty thousand fully induced cells were inoculated into 100 μl YPD medium containing increasing concentrations of G418. Growth was monitored after 20 h of incubation at 30°C. The strains used were AYC44 for Kan::Q20 and AYC46 for Kan::Q87. (B and C) Western blots quantifying sandwich fusion material obtained from soluble fractions (B) and whole-cell lysates (C). Lanes 1, APH with the polyQ20 insert at position 55 (AYC44); lanes 2, APH with the polyQ87 insert at position 55 (AYC46).

sistance to β-lactam antibiotics (11). This system proved to be effective in analyzing the *in vivo* stability of periplasmic proteins. This paper describes the development of a similar system based on several cytosolically expressed genes that encode resistance to aminoglycoside antibiotics. These sandwich fusion-based systems directly link the stability and solubility of model proteins to a phenotype that is easily selectable in *E. coli* and, in the case of G418, are screenable in the yeast *S. cerevisiae*. Insertions are usually deleterious, frequently disrupting the protein structure at the insertion site in catastrophic ways (25, 33). It is difficult to predict those rare sites within a protein that can accommodate insertions without severe functional disruption (34).

We therefore developed an approach to experimentally determine insertion-permissive sites within selectable markers. This protocol is very straightforward. It functioned well with all three antibiotic resistance markers that we tested and is potentially generalizable to any selectable marker carried by a plasmid. Sites within a host protein's structure that permit the insertion of a guest protein have historically been most commonly found in linker regions on the surface of the host protein. This is presumably because at these locations, the insertion of a guest protein does not interfere with the overall fold of the host protein (35, 36). This localization also exposes the guest protein to cellular proteases that recognize unstable and unfolded protein variants and degrade the fusion protein in a rate dependent on the guest protein's stability (37). The structure of the kanamycin resistance protein is known, and our permissive insertion sites do generally occur on surface regions in some (but not all) loops, but not within linker regions. Thus, even for proteins whose structure is known, like the APH protein, it would have been difficult to predict *a priori* where permissive insertion sites might lie, emphasizing the need to take an experimental approach. This is even more important for selectable markers like the ANT and NAT proteins, whose structures have not yet been solved.

The sandwich biosensor assays described here were developed to monitor *in vivo* protein stability and aggregation propensity in the cytosolic folding environment. Using these methods, we demonstrate a striking correlation between *in vivo* antibiotic resistances (the selectable traits) and *in vitro* stability for two different proteins, immunity protein 7 from *E. coli* and human muscle phosphatase. We also show that polyQ tracts of increasing length are associated with an increased tendency to form amyloids *in vivo* and are associated with a decreased resistance to aminoglycoside antibiotics in our sandwich fusion system. We demonstrate that our approach allows the quantitative analysis of protein stability in the cytosolic compartment without the need for prior structural and functional knowledge. These various selection systems will likely be useful in identifying folding modulators that enhance the stability of very unstable proteins (15).

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute, for which J.C.A.B. is an investigator.

We thank Linda Foit and Ursula Jakob for useful discussions.

REFERENCES

1. **Hartl FU, Hayer-Hartl M.** 2009. Converging concepts of protein folding in vitro and in vivo. Nat. Struct. Mol. Biol. **16:**574–581. http://dx.doi.org/10.1038/nsmb.1591.
2. **Ellis RJ, Hartl FU.** 1999. Principles of protein folding in the cellular environment. Curr. Opin. Struct. Biol. **9:**102–110. http://dx.doi.org/10.1016/S0959-440X(99)80013-X.
3. **Selkoe DJ.** 2002. Alzheimer's disease is a synaptic failure. Science **298:**789–791. http://dx.doi.org/10.1126/science.1074069.
4. **Narhi L, Wood SJ, Steavenson S, Jiang Y, Wu GM, Anafi D, Kaufman SA, Martin F, Sitney K, Denis P, Louis J-C, Wypych J, Biere AL, Citron M.** 1999. Both familial Parkinson's disease mutations accelerate α-synuclein aggregation. J. Biol. Chem. **274:**9843–9846. http://dx.doi.org/10.1074/jbc.274.14.9843.
5. **Powell K, Zeitlin PL.** 2002. Therapeutic approaches to repair defects in ΔF508 CFTR folding and cellular targeting. Adv. Drug Deliv. Rev. **54:**1395–1408. http://dx.doi.org/10.1016/S0169-409X(02)00148-5.
6. **Cheung MS, Klimov D, Thirumalai D.** 2005. Molecular crowding enhances native state stability and refolding rates of globular proteins. Proc. Natl. Acad. Sci. U. S. A. **102:**4753–4758. http://dx.doi.org/10.1073/pnas.0409630102.
7. **Stagg L, Christiansen A, Wittung-Stafshede P.** 2011. Macromolecular crowding tunes folding landscape of parallel α/β protein, apoflavodoxin. J. Am. Chem. Soc. **133:**646–648. http://dx.doi.org/10.1021/ja107638e.
8. **Ebbinghaus S, Dhar A, McDonald JD, Gruebele M.** 2010. Protein folding stability and dynamics imaged in a living cell. Nat. Methods **7:**319–323. http://dx.doi.org/10.1038/nmeth.1435.
9. **Ignatova Z, Krishnan B, Bombardier JP, Marcelino AMC, Hong J, Gierasch LM.** 2007. From the test tube to the cell: exploring the folding and aggregation of a β-clam protein. Pept. Sci. **88:**157–163. http://dx.doi.org/10.1002/bip.20665.
10. **Ghaemmaghami S, Oas TG.** 2001. Quantitative protein stability measurement in vivo. Nat. Struct. Mol. Biol. **8:**879–882. http://dx.doi.org/10.1038/nsb1001-879.
11. **Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, Warriner SL, Radford SE, Bardwell JCA.** 2009. Optimizing protein stability in vivo. Mol. Cell **36:**861–871. http://dx.doi.org/10.1016/j.molcel.2009.11.022.
12. **Hailu TT, Foit L, Bardwell JCA.** 2013. In vivo detection and quantification of chemicals that enhance protein stability. Anal. Biochem. **434:**181–186. http://dx.doi.org/10.1016/j.ab.2012.11.022.
13. **Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM.** 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat. Struct. Mol. Biol. **6:**1005–1009. http://dx.doi.org/10.1038/14890.
14. **Krobitsch S, Lindquist S.** 2000. Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. Proc. Natl. Acad. Sci. U. S. A. **97:**1589–1594. http://dx.doi.org/10.1073/pnas.97.4.1589.
15. **Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, Pfizenmaier J, Shi R, Hofmann S, Foit L, Ren G, Jakob U, Xu Z, Cygler M, Bardwell JCA.** 2011. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nat. Struct. Mol. Biol. **18:**262–269. http://dx.doi.org/10.1038/nsmb.2016.
16. **Karimi R, Ehrenberg M.** 1996. Dissociation rates of peptidyl-tRNA from the P-site of E. coli ribosomes. EMBO J. **15:**1149.
17. **Nurizzo D, Shewry SC, Perlin MH, Brown SA, Dholakia JN, Fuchs RL, Deva T, Baker EN, Smith CA.** 2003. The crystal structure of aminoglycoside-3′-phosphotransferase-IIa, an enzyme responsible for antibiotic resistance. J. Mol. Biol. **327:**491–506. http://dx.doi.org/10.1016/S0022-2836(03)00121-9.
18. **Clark NC, Olsvik Ø, Swenson JM, Spiegel CA, Tenover FC.** 1999. Detection of a streptomycin/spectinomycin adenylyltransferase gene (aadA) in Enterococcus faecalis. Antimicrob. Agents Chemother. **43:**157–160.
19. **Hollingshead S, Vapnek D.** 1985. Nucleotide sequence analysis of a gene encoding a streptomycin/spectinomycin adenyltransferase. Plasmid **13:**17–30. http://dx.doi.org/10.1016/0147-619X(85)90052-6.
20. **Krügel H, Fiedler G, Smith C, Baumberg S.** 1993. Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene nat1 from Streptomyces noursei. Gene **127:**127–131. http://dx.doi.org/10.1016/0378-1119(93)90627-F.
21. **Alshahni MM, Makimura K, Yamada T, Takatori K, Sawada T.** 2010. Nourseothricin acetyltransferase: a new dominant selectable marker for the dermatophyte Trichophyton mentagrophytes. Med. Mycol. **48:**665–668. http://dx.doi.org/10.3109/13693780903330555.
22. **Kochupurakkal BS, Iglehart JD.** 2013. Nourseothricin N-acetyl trans-

ferase: a positive selection marker for mammalian cells. PLoS One **8:**e68509. http://dx.doi.org/10.1371/journal.pone.0068509.
23. **Van Driessche B, Tafforeau L, Hentges P, Carr AM, Vandenhaute J.** 2005. Additional vectors for PCR-based gene tagging in Saccharomyces cerevisiae and Schizosaccharomyces pombe using nourseothricin resistance. Yeast **22:**1061–1068. http://dx.doi.org/10.1002/yea.1293.
24. **Van TT, Rooney PJ, Knoll LJ.** 2006. Nourseothricin acetyltransferase: a positive selectable marker for Toxoplasma gondii. J. Parasitol. **92:**668–670. http://dx.doi.org/10.1645/GE-706R.1.
25. **Cutler TA, Mills BM, Lubin DJ, Chong LT, Loh SN.** 2009. Effect of interdomain linker length on an antagonistic folding-unfolding equilibrium between two protein domains. J. Mol. Biol. **386:**854–868. http://dx.doi.org/10.1016/j.jmb.2008.10.090.
26. **Betton J-M, Jacob JP, Hofnung M, Broome-Smith JK.** 1997. Creating a bifunctional protein by insertion of β-lactamase into the maltodextrin-binding protein. Nat. Biotechnol. **15:**1276–1279. http://dx.doi.org/10.1038/nbt1197-1276.
27. **Capaldi AP, Kleanthous C, Radford SE.** 2002. Im7 folding mechanism: misfolding on a path to the native state. Nat. Struct. Biol. **9:**209–216. http://dx.doi.org/10.1038/nsb757.
28. **Reddy PH, Williams M, Tagle DA.** 1999. Recent advances in understanding the pathogenesis of Huntington's disease. Trends Neurosci. **22:**248–255. http://dx.doi.org/10.1016/S0166-2236(99)01415-0.
29. **Van Nuland NA, Chiti F, Taddei N, Raugei G, Ramponi G, Dobson CM.** 1998. Slow folding of muscle acylphosphatase in the absence of intermediates. J. Mol. Biol. **283:**883–891. http://dx.doi.org/10.1006/jmbi.1998.2009.
30. **Lang-Hinrichs C, Berndorff D, Seefeldt C, Stahl U.** 1989. G418 resistance in the yeast Saccharomyces cerevisiae: comparison of the neomycin resistance genes from Tn5 and Tn903. Appl. Microbiol. Biotechnol. **30:**388–394. http://dx.doi.org/10.1007/BF00296629.
31. **Espargaró A, Castillo V, de Groot NS, Ventura S.** 2008. The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. J. Mol. Biol. **378:**1116–1131. http://dx.doi.org/10.1016/j.jmb.2008.03.020.
32. **Mayer S, Rüdiger S, Ang HC, Joerger AC, Fersht AR.** 2007. Correlation of levels of folded recombinant p53 in Escherichia coli with thermodynamic stability in vitro. J. Mol. Biol. **372:**268–276. http://dx.doi.org/10.1016/j.jmb.2007.06.044.
33. **Collinet B, Hervé M, Pecorari F, Minard P, Eder O, Desmadril M.** 2000. Functionally accepted insertions of proteins within protein domains. J. Biol. Chem. **275:**17428–17433. http://dx.doi.org/10.1074/jbc.M000666200.
34. **Ruth N, Quinting B, Mainil J, Hallet B, Frère J-M, Huygen K, Galleni M.** 2008. Creating hybrid proteins by insertion of exogenous peptides into permissive sites of a class A β-lactamase. FEBS J. **275:**5150–5160. http://dx.doi.org/10.1111/j.1742-4658.2008.06646.x.
35. **Charbit A, Ronco J, Michel V, Werts C, Hofnung M.** 1991. Permissive sites and topology of an outer membrane protein with a reporter epitope. J. Bacteriol. **173:**262–275.
36. **Manoil C, Bailey J.** 1997. A simple screen for permissive sites in proteins: analysis of Escherichia coli lac permease. J. Mol. Biol. **267:**250–263. http://dx.doi.org/10.1006/jmbi.1996.0881.
37. **Parsell DA, Sauer RT.** 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. J. Biol. Chem. **264:**7590–7595.

SUPPLEMENT

Identification of permissive sites

The identification of permissive sites in *aadA* and *nat1* was performed in similar manner as for APH described in Material and Methods.

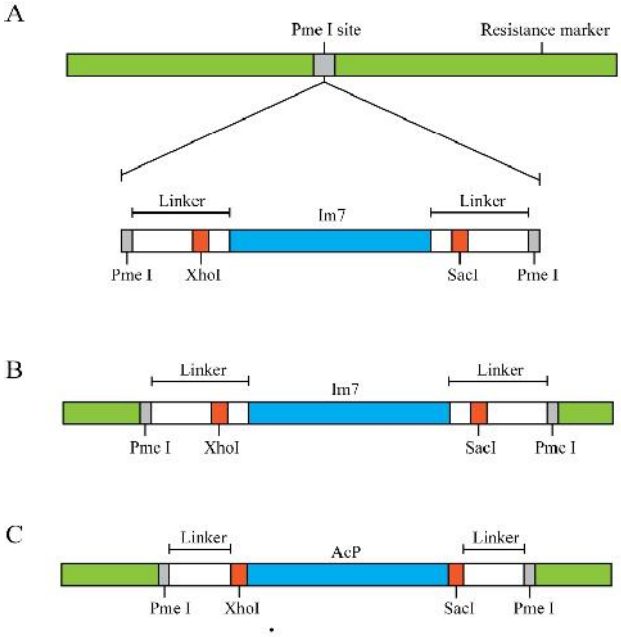
Successful transposon insertions in aadA could be detected loss of spectinomycin resistance. We identified 5000 clones that had lost their spectinomycin resistance, indicating transposon integration into aadA or its promoter region. To confirm the variability of our library, we sequenced a subset of clones and found that transposon insertions appeared to be distributed randomly throughout the ANT coding sequence (Fig. 6A). If these insertions are assumed to be completely random, we expect near complete coverage of the possible insertion sites within *aadA*, allowing us to screen through many possible sites for those permissive for insertion.

The plasmids of spectinomycin sensitive clones were pooled and the transposon was removed from *aadA* by digesting with restriction endonuclease Pme I. Again, after religation of the plasmids, a pentapeptide insertion scar that includes a Pme I restriction site remains at the former transposon integration site.

Transformation of the library and selection for regained resistance to 100 μg/ml spectinomycin only allows for the growth of colonies whose pentapetide insertion is in an ANT site that is at least partially permissive. Clones with insertions that significantly disturb ANT folding or function will not grow on 100 μg/ml spectinomycin. The wild-type pAMS1 plasmid confers resistance to more than 10,000 μg/ml spectinomycin. Sequencing of 40 resistant clones revealed three distinct hotspots in the protein that are permissive for pentapeptide insertion (Supplementary Fig. S4B)

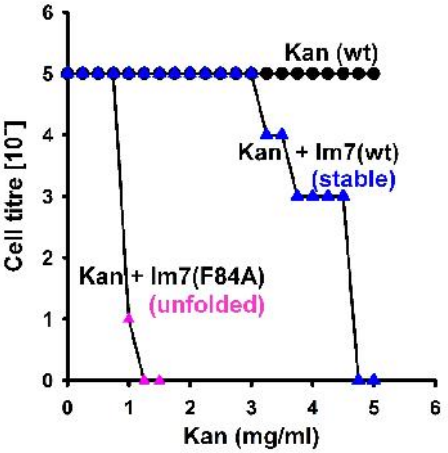
To clone Im7 into the pentapetide insertion library, plasmids from the library were pooled, purified, and digested with Pme I (which cuts inside the pentapetide insertion), then ligated with wild-type Im7 flanked by 17 amino acid GS linkers. The resulting plasmids were transformed into *E. coli* and selected on plates containing 100 μg/ml spectinomycin, only allowing growth of clones that had Im7 insertions in a permissive site in ANT.

Nat1 linker scanning was done with the Mutation Generation System (MGS) Kit (Thermoscientific), which inserts Not I-containing linkers instead of Pme I-containing linkers.



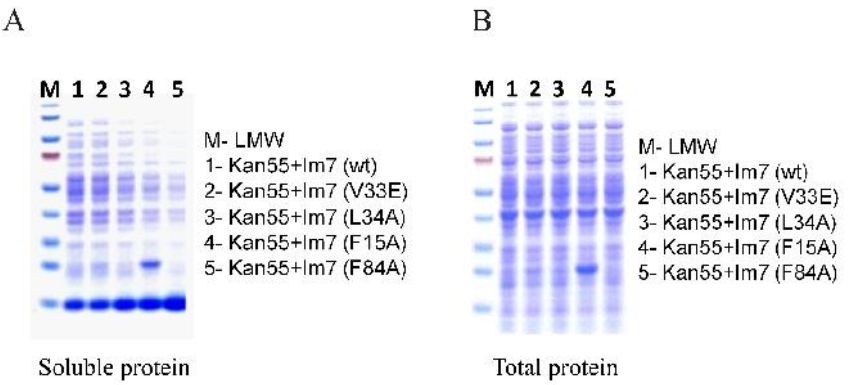
Supplementary Fig. S1:

A The pentapetide library of the resistance markers ANT and Kan^R (green) contains a Pme I site (grey) that is left behind after transposon excision. The Im7 construct we inserted into the pentapeptide library has Pme I sites (grey) on both sides of the protein and flexible GS-linkers (white). **B** The ANT biosensor after the insertion of the Im7 construct in the pentapeptide library (pAMS2 has the insertion at amino acid 155 of ANT). The Im7 construct is set up in such a way that the guest protein can be easily switched. The GS linkers contain Xho I and Sac I restriction sites (red), allowing directional cloning of new guest proteins. By digest with the two enzymes the majority of the linker is left behind in the resistance marker and allows simple substitution of new guest proteins. **C** ANT biosensor after Im7 was exchanged with AcP (pAMS34).



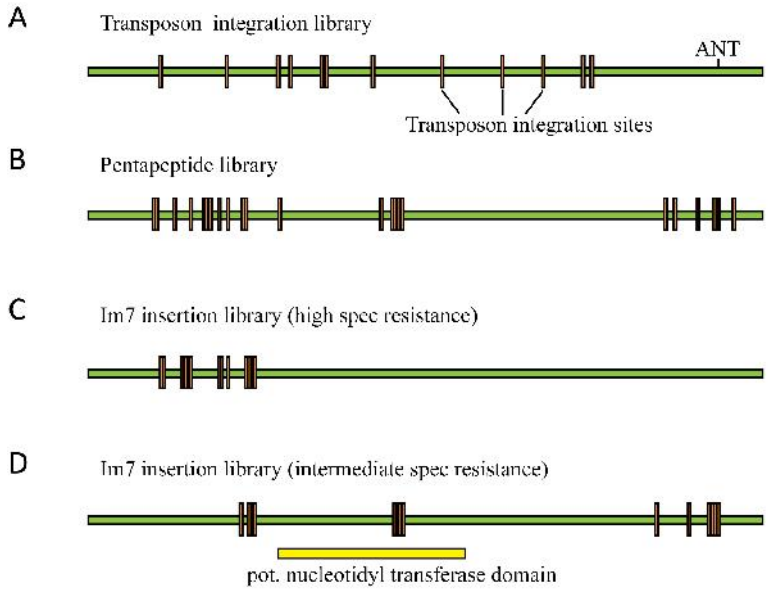
Supplementary Fig. S2

Comparison of antibiotic resistance in the kanamycin based biosensor. Cells expressing The unfolded Im7 variant F84A have a much lower antibiotic resistance than cells expressing the biosensor with Im7 wt. Maximal cell dilution of the cultures expressing Im7 (WT) and (F84A)-sandwich fusion at position 55 that allowed growth on different concentration of kanamycin shown with blue triangle and purple triangle, respectively. Bl21 (DE3)-RIPL strain expressing wt Im7 (stable) (AM235) and its unfolded variant F84A (AM236) fused at position 55 of the APH protein . Cultures were grown to mid-log phase and normalized to A_{600}=1. Cultures were 10 fold serially diluted from 10^0 to 10^5 were spotted on various concentration of kanamycin containing LB plates. Incubation was made 24 hrs at 37°C. At each dilution, growth or no growth was counted on each kanamycin plates to calculate MIC values.



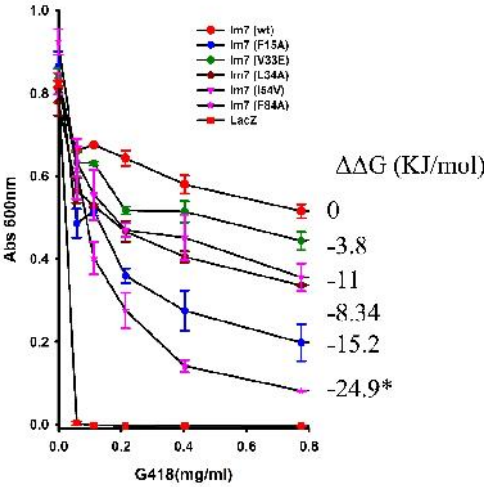
Supplement Fig. S3:

Loading control of soluble and total cell extract. (A) Soluble protein from cultures expressing Kan::Im7 variants was extracted by enzymatic treatment and glass bead grilling. Equal amount of soluble protein was analyzed on SDS-PAGE by Coomassie Blue staining (B) Whole cell expressing Kan::Im7 variants were boiled with SDS-loading dye and 15 ul was analyzed on SDS-PAGE by Coomassie Blue staining. For both A and B lane 1, AM235; lane 2, AM240; lane 3, AM 258; lane 4, AM247 and lane 5, AM236



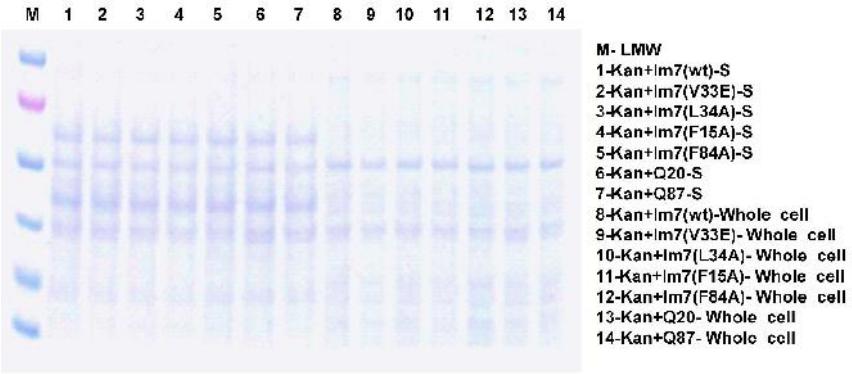
Supplementary Fig. S4

Transposon insertion sites during the identification of permissive sites in ANT. (A) Transposon integration sites (orange) seem to be randomly distributed throughout the ANT coding sequence (green). (B) Permissive sites in the pentapeptide library (orange) are found in three hotspots in ANT. (C) Im7 insertion sites (orange) and high spectinomycin resistance have insertions exclusively in the N terminus of ANT. (D) Im7 insertions with intermediate spectinomycin resistance have insertions in three hotspots of ANT. While insertions in the middle and the C terminus were all in-frame, all insertions in the N terminus led to frame shifts in ANT. It is possible that these clones use a downstream start codon to express a truncated version of ANT that confers lower levels of spectinomycin resistance.



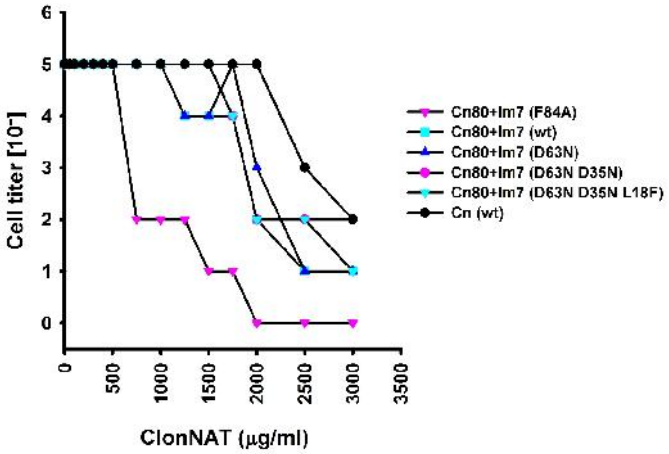
Supplementary Fig. S5:

Growth inhibition in S. cerevisiae expressing tripartite fusion under G418 selection pressure. From top to bottom in the graph, Im7 variants were wt, V33E, I54V, L34A, F15A and F84A, corresponding to strains A–YC 29, 34, 39,35, 30 and 42.Negative control expressing lacZ (A–YC43) showed with hexagon



Supplementary Fig. S6:

Loading control of soluble and total cell extract of S. cerevisiae. Lanes 1-7, soluble protein from S. cerevisiae expressing Kan::Im7 variants were isolated by treatment with yeast lysis solution; lanes 8-14, S. cerevisiae pellets boiled with SDS loading dye. From each sample, equal volume was loaded on SDS-PAGE and stained with Coomassie Blue. List all strains by number(Stains analysed in lanes 1-7 and 8-14 were A–YC29,34,35,30,42,44 and 46)



Supplementary Fig. S7:

Antibiotic resistance of Cn::Im7 variants. The antibiotic resistance of wild type Cn (A YC53) showed with circle. Im7 wt fused at Cn80 (A–YC48) showed with cyan square, Im7 F15A (A–YC47) with inverted blue triangle, Im7 D63N (A–YC50) with blue triangle, Im7 D63N D35N (A YC51) with pink circle and Im7 D63N D35N L18F (A–YC52) with inverted cyan triangle.

Supplement Table S1: Primers

P1	CAGTTTAAGGTTTAATCCTATAAAAGAG
P2	CTCTTTTATAGGATTAAACCTTAAACTG
P9	TTCAATTTCCTTAAGAAGTTGAACAGCCTCAGCCTCTGTGTAATCACTAATA
P10	TTCGAGTAACACATCTAACTCATCATCAGTTGCAGCAAC
P11	GCTGCAACTGATGATGTGTTAAATGTGTTACTCGAACACTTT
P12	GTTGCTGCAACTGATGATGTGGCAGATGTGTTACTCGAAC
P13	GTTCGAGTAACACATCTGCCACATCATCAGTTGCAGCAAC
P14	CTGAGCATCCAGATGGAACGGATGCGATTTATTATCCTAGTGATAATAG
P15	CTATTATCACTAGGATAATAAATCGCATCCGTTCCATCTGGATGCTCAG
P16	GCATCCAGATGGAACGGATCTGGTTTATTATCCTAGTGATAATAGA
P17	TCTATTATCACTAGGATAATAAACCAGATCCGTTCCATCTGGATGC
P18	GTCAAGACCGACCTGTCCGGATCCGCCCTGAATGAACTGC
P19	GCAGTTCATTCAGGGCGGATCCGGACAGGTCGGTCTTGAC
P20	GGATCCCAACAGCAGCAACAGCAACAACAG
P21	GGATCCCTGCTGTTGTTGCTGTTGCTGCTG
P22	GTCATGGTTGAACAAGATGGATTGCACGCAGGTTCTC
P23	TCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATG
P24	ATTGTTAATATACCTCTATACTTTAACGTC
P25	AGAATCGAGACCGAGGAGAGGGTTAGGGAT
P26	CCTAGTGATAATAGAGACAATAGCCCCGAAGGGA
P27	TCCCTTCGGGGCTATTGTCTCTATTATCACTAGG
P28	GTTGCTGCAACTGATGATGTGGCAGATGTGTTACTCGAACACTT
P29	AAGTGTTCGAGTAACACATCTGCCACATCATCAGTTGCAGCAAC
P30	CGGTTCCGGGAGCAGGGAAGCGAAAAATAGTATTAGTGATT
P31	AATCACTAATACTATTTTTCGCTTCCCTGCTCCCGGAACCG
P32	CATCCAGATGGAACGGATCTGGTCTATTATCCTAGTGATAATA
P33	TATTATCACTAGGATAATAGACCAGATCCGTTCCATCTGGATG
P34	GAGCATCCAGATGGAACGGATGCGATCTATTATCCTAGTGATAAT
P35	ATTATCACTAGGATAATAGATCGCATCCGTTCCATCTGGATGCTC
P36	GCTGCAACTGATGATGTGTTAAATGTGTTACTCGAACACTTT
P37	AAAGTGTTCGAGTAACACATTTAACACATCATCAGTTGCAGC
P38	ATTACACAGAGGCTGAGTTTGTTCAATTTCTTAAGGAAATTGAAAAAG
P39	CTTTTTCAATTTCCTTAAGAAATTGAACAAACTCAGCCTCTGTGTAAT
P40	TAAGGAAATTGAAAAAGAGAATGCTGCTGCAACTGATGATGTGTTAG
P41	CTAACACATCATCAGTTGCAGCAGCATTCTCTTTTTCAATTTCCTTA
P42	TGATTACACAGAGGCTGAGTTTGTTCAAGCTCTTAAGGAAATTGAAAAAGAG
P43	CTCTTTTTCAATTTCCTTAAGAGCTTGAACAAACTCAGCCTCTGTGTAATCA
P44	CTTAAGGAAATTGAAAAAGAGAAGGTTGCTGCAAATGATGATGTGTTAGATGTGT
P45	ACACATCTAACACATCATCATTTGCAGCAACCTTCTCTTTTTCAATTTCCTTAAG
P46	CTTAAGGAAATTGAAAAAGAGAAGGTTGCTGCAACTGATGATGTG
P47	CACATCATCAGTTGCAGCAACCTTCTCTTTTTCAATTTCCTTAAG

P48	GATCTGATCTATTATCCTAGGGATAATAGAGACGATAGCCC
P49	GGGCTATCGTCTCTATTATCCCTAGGATAATAGATCAGATC
P50	TATTAGTGATTACACAGAGGCTGAGGCTGTTCAACTTCTTAAGGAAATTGAA
P51	TTCAATTTCCTTAAGAAGTTGAACAGCCTCAGCCTCTGTGTAATCACTAATA
P52	GATAGCCCCGAAGGGATTGCCAAGGAAATTAAAGAATGG
P53	CCATTCTTTAATTTCCTTGGCAATCCCTTCGGGGCTATC
P54	GTCGCGTTCAAGGCGTCAGTTTTCGCATGTATACC
P55	GGTATACATGCGAAAACTGACGCCTTGAACGCGAC
P56	GCACAGTCCCTGAAATCGGTTGATTTTGAAGTCTTTGGT
P57	ACCAAAGACTTCAAAATCAACCGATTTCAGGGACTGTGC
P58	GGTCGCGTTCAAGGCGCCTGTTTTCGCATGTAT
P59	ATACATGCGAAAACAGGCGCCTTGAACGCGACC
P60	GCCCGAAGATAAAGTTAATTCCGCGAAATCTTGGTTGAGCAAAGTG
P61	CACTTTGCTCAACCAAGATTTCGCGGAATTAACTTTATCTTCGGGC
P62	AATTCCATGAAATCTTGGGTGAGCAAAGTGGGTTCGC
P63	GCGAACCCACTTTGCTCACCCAAGATTTCATGGAATT
P64	GATCGCACGAACTTTAGTAACGATAAGACCATTTCAAAATTGGAATA
P65	TATTCCAATTTTGAAATGGTCTTATCGTTACTAAAGTTCGTGCGATC
P134	GAATTCACATGGAGGCCCAGAATACCC
P135	CAGTATAGCGACCAGCATTCACGAATTC

Supplement Table S2: Thermodynamic stabilities of Im7 variants used in the biosensors

Im7 variant	Stability $\Delta\Delta G°_{UN}$ (KJ/mol)	ln (MIC_{mut}/MIC_{WT}) ANT system	SEM	ln (MIC_{mut}/MIC_{WT}) kan system	SEM
WT	0.00	0.00	0.00	0.00	0.00
V33E	-3.8			-0.04	0.03
L34A	-8.3	-0.60	0.05	-0.14	0.11
I54V	-6.4	-0.80	0.02	-0.48	0.09
L53A	-13.6			-0.73	0.05
F15A	-15.2	-2.02	0.02	-0.74	0.13
F84A	-24.9 *			-1.21	0.08
N26K S58R	2.05	0.35	0.11		
V69A	-2.30	-0.37	0.05		
N26K T30N S58R	5.89	0.45	0.02		
L3A	-4.2	-0.13	0.04		
I54A	-13.50	-1.57	0.01		
D35N D63N	5.97	0.28	0.05		
V27A D63N	3.93	0.30	0.04		
L18F D35N D63N	7.80	0.18	0.09		
L18A	-10.20	-1.94	0.02		

Supplement Table S3: Antibiotic resistance of Im7 variants inserted in different permissive sites of ANT

	Permissive site 153		Permissive site 157		Permissive site 289	
	ln (MIC_{mut}/MIC_{WT})	SEM	ln (MIC_{mut}/MIC_{WT})	SEM	ln (MIC_{mut}/MIC_{WT})	SEM
WT	0.00	0.00	0.00	0.00	0.00	0.00
V69A	-0.06	0.09	-0.27	0.06	-0.31	0.12
F15A	-1.97	0.02	-2.24	0.05	-3.50	0.01
L34A	-0.63	0.02	-0.62	0.08	-0.78	0.05
N26K S58R	0.29	0.06	0.36	0.12	0.28	0.20
N26K T30N S58R	0.25	0.20	0.49	0.16	0.57	0.27

* Im7 F84A is so destabilized that it mainly populates the unfolded state *in vitro*. This precludes the accurate determination of its ΔG°UN values (27) For the purposes of this paper it is assumed to be fully unfolded and thus have a ΔG°UN value = 0 kJ/mol or 24.9 kJ/mol greater than wild type Im7 which has a ΔG°UN value = -24.9kJ/mol list all strains by number

Supplement Table S4: Thermodynamic stabilities and MICs of AcP variants tested in the ANT system

AcP variant	Stability $\Delta\Delta G^{\circ}_{UN}$ (KJ/mol)	ln (MIC_{mut}/MIC_{WT})	SEM
C21S	0.00	0.00	0.00
Y11F C21S	1.80	0.20	0.07
C21S L65V	-22.40	-3.57	0.01
C21S E83D	-6.30	-0.43	0.08
C21S V20A	-1.20	-0.14	0.05
C21S M61A	-16.60	-1.84	0.04

Supplement Table S5 strains and plasmids used in this work

Strain or plasmid	Genotype	Relevant description or derivation
NEB10β	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80*dlacZΔ*M15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	
AMS1	NEB10β pAMS1	
AMS2	NEB10β pAMS2	
AMS3	NEB10β pAMS3	
AMS4	NEB10β pAMS4	
AMS5	NEB10β pAMS5	
AMS6	NEB10β pAMS6	
AMS7	NEB10β pAMS7	
AMS8	NEB10β pAMS8	
AMS9	NEB10β pAMS9	
AMS10	NEB10β pAMS10	
AMS11	NEB10β pAMS11	
AMS12	NEB10β pAMS12	
AMS13	NEB10β pAMS13	
AMS14	NEB10β pAMS14	
AMS15	NEB10β pAMS15	
AMS16	NEB10β pAMS16	
AMS17	NEB10β pAMS17	
AMS18	NEB10β pAMS18	
AMS19	NEB10β pAMS19	
AMS20	NEB10β pAMS20	
AMS21	NEB10β pAMS21	
AMS22	NEB10β pAMS22	
AMS23	NEB10β pAMS23	
AMS24	NEB10β pAMS24	
AMS25	NEB10β pAMS25	
AMS26	NEB10β pAMS26	
AMS27	NEB10β pAMS27	
AMS28	NEB10β pAMS28	
AMS29	NEB10β pAMS29	
AMS30	NEB10β pAMS30	
AMS31	NEB10β pAMS31	
AMS32	NEB10β pAMS32	
AMS33	NEB10β pAMS33	
AMS34	NEB10β pAMS34	
AMS35	NEB10β pAMS35	
AMS36	NEB10β pAMS36	
AMS37	NEB10β pAMS37	
AMS38	NEB10β pAMS38	
AMS39	NEB10β pAMS39	

Strain or plasmid	Genotype	Relevant description or derivation
AMS40	NEB10β pAMS40	
BL21-codon plus (DE3)-RIPL	E. coli B F $^{-}$ ompT hsdS ($r_B^- m_B^-$) dcm + Tet ^{r} gal λ (DE3) endA Hte [argU proL Cam ^{r}] [argU ileY leuW Strep/Spec ^{r}]	New England Biolabs
INVSc1	*MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52*	Dr. Anuj Kumar (University of Michigan)
AM101	NEB10b pAM103	
AM144	NEB10b pAM15	
AM152	NEB10b p*4339*	
AM200	NEB10b pAM79	
AM201	NEB10b pAM80	
AM235	BL21-codon plus (DE3)-RIPL pAM107	
AM236	BL21-codon plus (DE3)-RIPL pAM108	
AM240	BL21-codon plus (DE3)-RIPL pAM112	
AM242	BL21-codon plus (DE3)-RIPL pAM114	
AM243	BL21-codon plus (DE3)-RIPL pAM115	
AM246	BL21-codon plus (DE3)-RIPL pAM118	
AM247	BL21-codon plus (DE3)-RIPL pAM119	
AM248	NEB10b pAM120	
AM249	BL21-codon plus (DE3)-RIPL pAM121	
AM258	BL21-codon plus (DE3)-RIPL pAM130	
AM282	NEB10b pAM154	
AM283	NEB10b pAM155	
AM284	NEB10b pAM156	
AM285	NEB10b pAM157	
AM286	NEB10b pAM158	
AM288	NEB10b pAM160	
AM290	NEB10b pAM162	
AM291	NEB10b pAM163	
AM308	NEB10b pAM180	
AM323	NEB10b p416-103Q	
aYC-28	INVSc1 pAM154	
aYC-29	INVSc1 pAM154::ceiE7	
aYC-30	INVSc1 pAM154::ceiE7F15A	
aYC-34	INVSc1 pAM154::ceiE7V33E	
aYC-35	INVSc1 pAM154::ceiE7L34A	
aYC-39	INVSc1 pAM154::ceiE7I54V	
aYC-42	INVSc1 pAM154::ceiE7F84A	
aYC-43	INVSc1 pYES2.1/V5-His-TOPO::LacZ	
aYC-44	INVSc1 pAM154::Q20	
aYC-46	INVSc1 pAM154::Q87	
aYC-47	INVSc1 pYES2.1/V5-His-TOPO::nat1aa80::ceiE7 F15A	Im7 F15A inserted after amino acid 80 of nat1

Strain or plasmid	Genotype	Relevant description or derivation
aYC-48	INVSc1 pYES2.1/V5-His-TOPO::nat1aa80::ceiE7	Im7 wt inserted after amino acid 80 of nat1
aYC-50	INVSc1 pYES2.1/V5-His-TOPO::nat1aa80::ceiE7D63N	Im7 D63N inserted after amino acid 80 of nat1
aYC-51	INVSc1 pYES2.1/V5-His-TOPO::nat1aa80::ceiE7D63N,D35N	Im7 D63N, D35N inserted after amino acid 80 of nat1
aYC-52	INVSc1 pYES2.1/V5-His-TOPO::nat1aa80::ceiE7D63N,D35N,L18F	Im7 D63N, D35N, L18F inserted after amino acid 80 of nat1
aYC-53	INVSc1 pYES2.1/V5-His-TOPO::nat1	nat1 cloned on pYES2.1 vector
pAMS1	pBR322::aadA	aadA (encoding StrepR) from pBAD43 cloned into pBR322 using Hind III
pAMS2	pAMS1 aadAaa155::ceiE7	ceiE7 (gene encoding Im7) inserted after amino acid 155 of aadA
pAMS3	pAMS1 aadAaa155::linker	Linker inserted after amino acid 155 of aadA (encoding StrepR)
pAMS4	pAMS 2 aadAaa155::ceiE7L34A	Im7 L34A inserted after amino acid 155 of aadA
pAMS5	pAMS2 ceiE7I54V	Im7 I54V inserted after amino acid 155 of aadA
pAMS6	pAMS2 ceiE7F15A	Im7 F15A inserted after amino acid 155 of aadA
pAMS7	pAMS2 ceiE7N26K S58R	Im7 N26K S58R inserted after amino acid 155 of aadA
pAMS8	pAMS2 ceiE7V69A	Im7 V69A inserted after amino acid 155 of aadA
pAMS9	pAMS2 N26KT30NS58R	Im7 N26K T30N S58R inserted after amino acid 155 of aadA
pAMS10	pAMS2 ceiE7L3A	Im7 L3A inserted after amino acid 155 of aadA
pAMS11	pAMS2 ceiE7I54A	Im7 I54A inserted after amino acid 155 of aadA
pAMS12	pAMS2 ceiE7D35ND63N	Im7 D35N D63N inserted after amino acid 155 of aadA
pAMS13	pAMS2 ceiE7V27AD63N	Im7 V27A D63N inserted after amino acid 155 of aadA
pAMS14	pAMS2 ceiE7L18FD35ND63N	Im7 L18F D35N D63N inserted after amino acid 155 of aadA
pAMS15	pAMS2 ceiEL18A	Im7 L18A inserted after amino acid 155 of aadA
pAMS16	pBR322::aadAaa153::ceiE7	ceiE7 (gene encoding Im7) inserted after amino acid 153 of aadA
pAMS17	pAMS16 ceiE7V69A	Im7 V69A inserted after amino acid 153 of aadA
pAMS18	pAMS16ceiE7F15A	Im7 F15A inserted after amino acid 153 of aadA
pAMS19	pAMS16ceiE7L34A	Im7 L34A inserted after amino acid 153 of aadA
pAMS20	pAMS16 ceiE7N26KS58R	Im7 N26K S58R inserted after amino acid 153 of aadA
pAMS21	pAMS16 ceiE7N26KT30NS58R	Im7 N26K T30N S58R inserted after amino acid 153 of aadA
pAMS22	pBR322::aadAaa157::ceiE7	ceiE7 (gene encoding Im7) inserted after amino acid 157 of aadA
pAMS23	pAMS22 ceiE7V69A	Im7 V69A inserted after amino acid 157 of aadA
pAMS24	pAMS22 ceiE7F15A	Im7 F15A inserted after amino acid 157 of aadA

Strain or plasmid	Genotype	Relevant description or derivation
pAMS25	pAMS22 ceiE7L34A	Im7 L34A inserted after amino acid 157 of aadA
pAMS26	pAMS22 ceiE7N26KS58R	Im7 N26K S58R Im7 inserted after amino acid 157 of aadA
pAMS27	pAMS22ceiE7N26KT30NS58R	Im7 N26K T30N S58R Im7 inserted after amino acid 157 of aadA
pAMS28	pBR322::aadAaa289::ceiE7	ceiE7 (gene encoding Im7) inserted after amino acid 289 of aadA
pAMS29	pAMS28 ceiE7V69A	Im7 V69A inserted after amino acid 289 of aadA
pAMS30	pAMS28 ceiE7F15A	Im7 F15A inserted after amino acid 289 of aadA
pAMS31	pAMS28 ceiE7L34A	Im7 L34A inserted after amino acid 289 of aadA
pAMS32	pAMS28 ceiE7N26KS58R	Im7 N26K S58R inserted after amino acid 289 of aadA
pAMS33	pAMS28 ceiE7N26KT30NS58R	Im7 N26K T30N S58R inserted after amino acid 289 of aadA
pAMS34	pAMS1 aadAaa155::acpy2	Acyp2 encoding AcP inserted after amino acid 155 of aadA
pAMS35	pAMS34 acpy2C21S	AcP C21S AcP inserted after amino acid 155 of aadA
pAMS36	pAMS34 acpy2Y11FC21S	AcP Y11F C21S AcP inserted after amino acid 155 of aadA
pAMS37	pAMS34 acpy2C21SL65V	AcP C21S L65V AcP inserted after amino acid 155 of aadA
pAMS38	pAMS34 acpy2C21SE83D	AcP C21S E83D AcP inserted after amino acid 155 of aadA
pAMS39	pAMS34 acpy2V20AC21S	AcP C21S V20A AcP inserted after amino acid 155 of aadA
pAMS40	pAMS34 acpy2C21SM61A	AcP C21S M61A AcP inserted after amino acid 155 of aadA
pAM 15	pCR-Blunt II-TOPO- ccdB5	Point mutaion in *ccdB* gene at amino acid 5 to intoduce stop codon and NsiI digested and religated vector
pAM 18	pCR-Blunt II-TOPO- subclone	Subclone on pCR-Blunt II-TOPO- vector
pAM 79	pAM15aphA-2aa55::Q20	polyQ(20) inserted after amino acid 55 of aphA-2
pAM 80	pAM15aphA-2aa55::Q45	polyQ(45) inserted after amino acid 55 of aphA-2
pAM 81	pAM15aphA-2aa55::Q81	polyQ(87) inserted after amino acid 55 of aphA-2
pAM 103	pBR322::nat1	nat1 (encoding ClonNAT resistance) from plasmid p4339 (or pAM31) cloned into pBR322 using EcoRI
pAM 107	pAM15aphA-2aa55::ceiE7	ceiE7 (gene encoding Im7) inserted after amino acid 55 of aphA-2
pAM 108	pAM15aphA-2aa55::ceiE7 F84A	Im7 F84A inserted after amino acid 55 of aphA-2
pAM 112	pAM15aphA-2aa55::ceiE7 V33E	Im7 V33E inserted after amino acid 55 of aphA-2
pAM 114	pAM15aphA-2aa55::ceiE7 L53A	Im7 L53A inserted after amino acid 55 of aphA-2
pAM 115	pAM15aphA-2aa55::ceiE7 I54V	Im7 I54V inserted after amino acid 55 of aphA-2
pAM 118	pAM15aphA-2aa42::ceiE7 F84A	Im7 F84A inserted after amino acid 42 of aphA-2
pAM 119	pAM15aphA-2aa55::ceiE7 F15A	Im7 F15A inserted after amino acid 55 of aphA-2

Strain or plasmid	Genotype	Relevant description or derivation
pAM 120	pBR322 bla196::ceiE7 F84A	Im7 F84A inserted after amino acid 196 of bla in pBR322
pAM 121	pAM15aphA-2aa21::ceiE7 F84A	Im7 F84A inserted after amino acid 21 of aphA-2
pAM 130	pAM15aphA-2aa55::ceiE7 L34A	Im7 L34A inserted after amino acid 55 of aphA-2
pAM 154	pYES2.1/V5-His-TOPO::aphA-2	aphA-2 (encoding Kan resistance) cloned in pYES2.1/V5-His-TOPO
pAM 155	pYES2.1/V5-His-TOPO::aphA-2aa55::ceiE7	Im7 inserted after amino acid 55 of aphA-2 in pYES2.1/V5-His-TOPO
pAM 156	pYES2.1/V5-His-TOPO::aphA-2aa55::ceiE7 F84A	Im7 F84A inserted after amino acid 55 of aphA-2 in pYES2.1/V5-His-TOPO
pAM 157	pYES2.1/V5-His-TOPO::aphA-2aa55::ceiE7 F15A	Im7 F15A inserted after amino acid 55 of aphA-2 in pYES2.1/V5-His-TOPO
pAM 158	pYES2.1/V5-His-TOPO::aphA-2aa55::Q20	polyQ(20) inserted after amino acid 55 of aphA-2 in pYES2.1/V5-His-TOPO
pAM 160	pYES2.1/V5-His-TOPO::aphA-2aa55::Q87	polyQ(87) inserted after amino acid 55 of aphA-2 in pYES2.1/V5-His-TOPO
pAM 162	pAM15aphA-2aa55::ceiE7 V33E	Im7 V33E inserted after amino acid 55 of aphA-2 pYES2.1/V5-His-TOPO
pAM 163	pYES2.1/V5-His-TOPO::lacZ	lacZ gene coding for beta-galactosidase cloned in
pAM 180	pAM15aphA-2 (SDM41-43)aa55::Q87	polyQ(87) inserted in silent mutated aphA-2 variant after amino acid 55
p416-103Q		(14)
pGPS4	New England Biolab	
p4339	pCR-TOPO::natR-MX4	Dr. Charles Boone lab (U Toronto),

The source of all strains and plasmids is this study with the exception of NEB10β and BL21 which were obtained from New England Biolabs.

Molecular Cell
Article

Polyphosphate Is a Primordial Chaperone

Michael J. Gray,[1] Wei-Yun Wholey,[1,2] Nico O. Wagner,[1] Claudia M. Cremers,[1] Antje Mueller-Schickert,[1,3] Nathaniel T. Hock,[1] Adam G. Krieger,[1] Erica M. Smith,[1] Robert A. Bender,[1] James C.A. Bardwell,[1,2,3] and Ursula Jakob[1,2,*]
[1]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA
[2]Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI 48109, USA
[3]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI 48109, USA
*Correspondence: ujakob@umich.edu
http://dx.doi.org/10.1016/j.molcel.2014.01.012

SUMMARY

Composed of up to 1,000 phospho-anhydride bond-linked phosphate monomers, inorganic polyphosphate (polyP) is one of the most ancient, conserved, and enigmatic molecules in biology. Here we demonstrate that polyP functions as a hitherto unrecognized chaperone. We show that polyP stabilizes proteins in vivo, diminishes the need for other chaperone systems to survive proteotoxic stress conditions, and protects a wide variety of proteins against stress-induced unfolding and aggregation. In vitro studies reveal that polyP has protein-like chaperone qualities, binds to unfolding proteins with high affinity in an ATP-independent manner, and supports their productive refolding once nonstress conditions are restored. Our results uncover a universally important function for polyP and suggest that these long chains of inorganic phosphate may have served as one of nature's first chaperones, a role that continues to the present day.

INTRODUCTION

It is generally agreed that protein synthesis evolved from a world in which RNA served both genetic and catalytic roles in biology, although the driving forces and requirements for the transition to the protein world still remain unclear (Noller, 2004). A major question that has puzzled researchers for a long time is, how do proteins, which are born as linear chains of amino acids, achieve the intricate three-dimensional structures necessary for proper function? Anfinsen's classic experiments (Anfinsen, 1973), which showed that the specific structure of a protein is solely determined by its amino acid sequence, seemed to provide the long-sought answer as to how proteins could have evolved to play such central roles in biology. However, it has become increasingly clear that within the crowded environment of the cell, many proteins require a cohort of molecular chaperones, proteases, and regulatory signaling pathways, collectively called the proteostasis network, to fold, function, and withstand stress conditions (Powers and Balch, 2013). This realization has raised new questions, particularly regarding the potential coevolution of proteins and the proteostasis mechanisms necessary to keep them stable and soluble. We have now identified a primordial member of the proteostasis network, the prebiotic molecule inorganic polyphosphate (polyP). Synthesized in vivo from ATP and consisting entirely of high-energy phospho-anhydride-bonded inorganic phosphate (Achbergerová and Nahálka, 2011; Rao et al., 2009), these universally conserved molecules exhibit all of the characteristics of an efficient protein chaperone, making polyP one of the most ancient chaperones known.

Chaperone discovery is difficult. Chaperone-deficient cells exhibit many different, seemingly unrelated, and often overlapping phenotypes. These pleiotropic phenotypes are the result of the involvement of molecular chaperones in the folding, assembly and disassembly, transport, and degradation of a large number of different proteins. Therefore, the loss of a chaperone can often lead to unpredictable functional effects in the cell (Kim et al., 2013; Powers and Balch, 2013). Compounding this problem, the in vitro assays for chaperones are neither specific nor sensitive enough to enable their purification from crude lysates by activity. It is not surprising, therefore, that new chaperones continue to be discovered even in very well characterized organisms such as *Escherichia coli* (Quan et al., 2011).

Cells deficient in polyP show a multitude of different phenotypic traits, similar to the pleiotropic phenotypes exhibited by chaperone-deficient cells. Bacteria or unicellular eukaryotes lacking polyP are sensitive to a number of different stress conditions, including heat shock and heavy-metal exposure, and are defective in virulence, biofilm formation, and motility (Docampo et al., 2010; Rao et al., 2009). In higher eukaryotes, polyP is known to play a central role in blood clotting, and is involved in apoptosis, mTOR activation, and neuronal signaling (Holmstrom et al., 2013; Kulakovskaya et al., 2012; Moreno and Docampo, 2013; Smith et al., 2010). The underlying physiological role of polyP has been attributed to diverse functions of the molecule, including phosphate and energy storage (polyP is isoenergetic to ATP), metal chelation, pH buffering, and regulatory interactions (Kornberg et al., 1999; Kulakovskaya et al., 2012; Rao et al., 2009). However, there is no satisfactory explanation for a general mechanism by which polyP affects these seemingly unrelated processes in the cell.

Here, we show that bacteria, in response to protein-unfolding oxidative stress (i.e., hypochlorous acid [HOCl]), redirect cellular ATP to polyP, resulting in a more than 10,000-fold increase in stress resistance. We demonstrate that polyP functions as a global, highly effective, and wholly inorganic

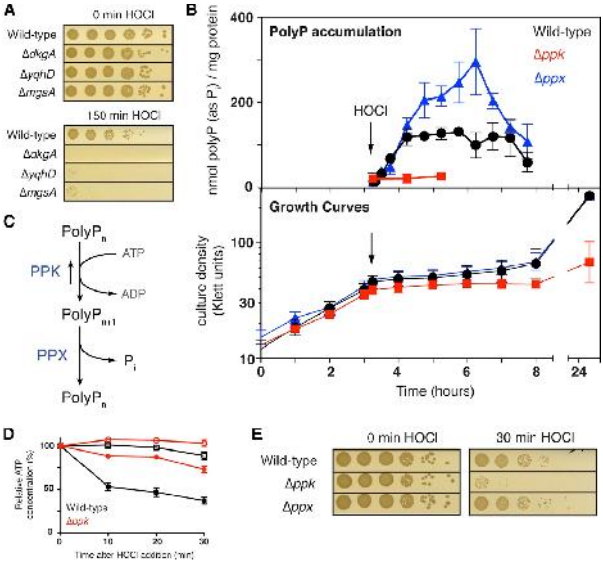


Figure 1. ATP-Derived PolyP Protects against HOCl

(A) Exponentially growing *E. coli* wild-type and mutant strains were incubated with 2 mM HOCl for the indicated time, serially diluted, and spot titered. See also Figures S1A and S1B.

(B) Growth (lower panel) and intracellular polyP concentration (upper panel) of *E. coli* wild-type (black circles), Δ*ppk* (red squares), and Δ*ppx* (blue triangles) (mean ± SD). Arrow indicates addition of 1 mM HOCl. See also Figures S1C–S1E.

(C) PolyP synthesis and degradation pathway in *E. coli*.

(D) ATP content of log phase *E. coli* wild-type (black squares) and Δ*ppk* (red circles) with (closed symbols) or without (open symbols) treatment with 1 mM HOCl, as a percentage of the initial value for each sample (mean ± SD). See also Figure S1F.

(E) Exponentially growing *E. coli* wild-type and mutant strains were incubated with 2.5 mM HOCl and then diluted and spot titered. See also Figures S1G and S1H.

chaperone that stabilizes unfolding proteins, prevents protein aggregation both in vitro and in vivo, and maintains proteins in a refolding-competent form. These results help to explain the long known but largely unexplained role of polyP in protecting organisms against stress conditions, and suggest that polyP may have served as one of nature's first chaperones.

RESULTS

Phosphate Starvation Is an Immediate Response to Oxidative Protein Unfolding Stress

A reexamination of a recent microarray analysis of *E. coli* gene-expression changes in response to the proteotoxic oxidant HOCl, a potent physiological antimicrobial, showed that the expression of at least 12 phosphate starvation-induced genes is highly upregulated (Gray et al., 2013b). This result was consistent with our earlier studies, which revealed that HOCl-treated cells substantially increase their toxic methylglyoxal production (Gray et al., 2013b), a reaction that is driven by low phosphate and high triose phosphate concentrations, and used to restore inorganic phosphate pools (Figure S1A available online; Booth et al., 2003). As expected, mutant bacteria carrying deletions in the enzymes DkgA or YqhD, which detoxify the accumulating electrophile methylglyoxal (Figure S1A), were found to be very sensitive to HOCl treatment (Figure 1A). Surprisingly, however, *E. coli* and *Vibrio cholera* mutants lacking the enzyme that makes the toxic electrophile (i.e., methylglyoxal synthase [MgsA]) were also highly HOCl sensitive (Figures 1A and S1B). These results strongly imply that replenishment of cellular phosphate pools plays a crucial role in oxidative stress defense in bacteria. This was very intriguing because phosphate is known to be entirely nonreactive with reactive chlorine species, and therefore its cellular levels should not be affected by these oxidants (Deborde and von Gunten, 2008).

Severe Oxidative Stress Leads to PolyP Accumulation

Because phosphate constitutes the building block of polyP, we considered whether polyP accumulation was triggered by HOCl stress, thereby resulting in phosphate starvation. As shown in Figure S1C, microscopic examination of HOCl-treated *E. coli* cells stained with DAPI revealed a significant accumulation of yellow fluorescent foci characteristic of DAPI bound to polyP (Aschar-Sobbi et al., 2008). Similar polyP-containing bodies (i.e., metachromatic granules, volutin granules, etc.) have been observed in many stress-exposed organisms, but the purpose of such bodies still remains largely unclear (Docampo et al., 2010; Rao et al., 2009). Very few yellow foci were detected in untreated cells or, as previously observed (Ault-Riché et al., 1998; Winter et al., 2005), in cells treated with the nonproteotoxic oxidant hydrogen peroxide (Figure S1C). Gel analysis (Figure S1D) and quantitative polyP measurements (Figures 1B and S1E) confirmed these results and showed that *E. coli* and *V. cholerae* cells accumulate substantial and comparable amounts of polyP within minutes after HOCl treatment. These results likely provide the missing link between HOCl stress and phosphate starvation.

PolyP is synthesized from ATP by polyphosphate kinase (PPK). Although this is a reversible process, the majority of polyP is degraded to inorganic phosphate by exopolyphosphatase (PPX) under physiological conditions (Figure 1C). Significant polyP accumulation upon HOCl stress should therefore result

in a noticeable decline in cellular ATP levels, and would explain why HOCl-stressed organisms appear phosphate starved. Intriguingly, all organisms investigated so far have been shown to experience a very rapid decline in cellular ATP levels upon severe oxidative stress treatment (Barrette et al., 1987; Hyslop et al., 1988; Winter et al., 2005). This ATP decline was previously attributed to the observed cessation of glycolysis, triggered by the oxidative inactivation of glyceraldehyde-3-phosphate dehydrogenase (GapDH), and leading to the stress-mediated accumulation of dihydroxyacetophosphate (DHAP), the phosphate donor and substrate of methylglyoxal synthase (Figure S1A). In light of our results, however, we now wondered whether accumulation of polyP might contribute to the observed ATP decline. To test this idea, we treated wild-type *E. coli* and the Δ*ppk* mutant strain (which is no longer able to respond to HOCl treatment with the accumulation of polyP; Figures 1B and S1D) with HOCl and measured the intracellular ATP levels. Upon exposure to HOCl, wild-type *E. coli* showed the previously observed rapid decrease in ATP (Winter et al., 2008), whereas Δ*ppk* mutant cells maintained most of their ATP levels (Figure 1D). A comparison of the absolute ATP levels measured in both strain backgrounds (Figure S1F) revealed that within the first 20 min of HOCl treatment, ATP levels decreased from 1.7 μM ATP/OD_{600} to 0.8 μM ATP/OD_{600} in wild-type cells, whereas Δ*ppk* strains, which lost significant viability during this time frame (Figure 1E), only showed a decrease in ATP levels from 2.4 μM ATP/OD_{600} to 2.1 μM ATP/OD_{600} (Figure S1F). These results suggest that wild-type bacteria actively redirect a substantial proportion of their cellular ATP pool to form polyP upon exposure to severe oxidative stress, and raise the intriguing possibility that the loss of ATP, long assumed to be a symptom of oxidative stress, may actually be part of an adaptive oxidative stress response aimed at rapidly accumulating large quantities of polyP. It is worth mentioning that the Δ*ppk* mutant cells contained substantially higher ATP concentrations than wild-type cells, even under nonstress conditions (Figure S1F). Although the implications of this result remain unclear, it is possible that polyP synthesis and degradation act as an ATP-draining futile cycle in wild-type cells, that the presence of polyP stimulates ATP consumption by other enzymes, or that the absence of polyP slows cellular metabolism and/or ATP consumption.

PolyP Accumulation Plays a Critical Role in HOCl Defense

To directly test whether accumulation of polyP contributes to the HOCl resistance of bacteria, we compared the HOCl resistance of *E. coli* and *V. cholerae* wild-type cells with that of strains that lack polyP (i.e., Δ*ppk* strains) or overaccumulate polyP (i.e., Δ*ppx* strains). As shown in Figures 1E, S1G, and S1H, Δ*ppk* cells are exquisitely sensitive to treatment with either HOCl or *N*-chlorotaurine, a physiologically relevant in vivo secondary oxidation product of HOCl that is present at high concentration in neutrophils (Nagl et al., 2000). In contrast, strains lacking the polyP-degrading exopolyphosphatase PPX (i.e., Δ*ppx* strains) accumulate higher levels of polyP upon HOCl treatment and are slightly more resistant to reactive chlorine species than wild-type cells (Figures 1E and S1H). These results are in agreement with prior studies from Arthur Kornberg's laboratory and others (Crooke et al., 1994; Kornberg et al., 1999; Rao et al., 2009), which showed that polyP-deficient organisms exhibit increased sensitivity to a variety of environmental stress conditions, including amino acid starvation, osmotic stress, and heat shock.

PolyP Functions as a General Molecular Chaperone In Vivo

Many stress conditions, including those shown to necessitate polyP accumulation in bacteria for improved survival, cause protein unfolding and aggregation (Kornberg et al., 1999; Winter et al., 2008). As a first test to determine whether polyP might be involved in maintaining proteostasis, we monitored the expression of heat-shock genes in HOCl-treated wild-type and Δ*ppk* strains. Stress-induced protein unfolding is the primary trigger of the *E. coli* heat-shock response, making upregulation of heat-shock gene expression a bellwether for the collapse of proteostasis (Guisbert et al., 2008). In the absence of stress, there was no discernible difference in heat-shock gene expression in the wild-type and Δ*ppk* strains (Figure S2A). However, upon exposure of both strains to HOCl treatment, we found that most of the tested heat-shock genes, including the heat-shock-regulated chaperones Hsp33 and DnaK, were more highly upregulated in Δ*ppk* as compared with the wild-type (Figure 2A). Expression of *rpoH*, which encodes the heat-shock sigma factor, was also slightly increased. In contrast, marker genes for DNA damage (e.g., *sulA*), which is often cited as a major threat during oxidative stress conditions (Imlay, 2013), did not show any additional upregulation in *E. coli* Δ*ppk* strains (Figure 2B). These results suggest that the presence of polyP reduces the need for other chaperones to combat oxidative protein unfolding, implying that polyP might function as a physiologically relevant chaperone.

To directly test this idea, we treated *E. coli* wild-type, Δ*ppk*, and Δ*ppx* strains with HOCl and analyzed the extent of protein aggregation in these cells. As previously observed (Winter et al., 2008), HOCl treatment led to substantial accumulation of insoluble proteins in wild-type *E. coli* within 30 min of incubation (Figure 2C). However, cells lacking *ppk* accumulated slightly more insoluble protein after HOCl treatment than the wild-type, whereas cells lacking *ppx*, and therefore the ability to degrade accumulating polyP, showed substantially less protein aggregation (Figure 2C). These results suggest that polyP plays an important role in protecting cells against toxic protein aggregation, and that levels of polyP accumulation might be inversely correlated with intracellular protein damage.

To test whether this protein-protective effect was specific to HOCl stress conditions or was of general significance, we took advantage of the temperature-sensitive, *rpoH*-deficient *E. coli* strain BB7224. This strain is unable to induce the heat-shock response, is largely deficient in protein chaperones, and is extremely sensitive to the protein unfolding effects of heat treatment (Guisbert et al., 2008; Tomoyasu et al., 2001). Overexpression of select chaperones, such as the DnaK/DnaJ system, has been shown to mitigate heat-induced protein aggregation and rescue the temperature-sensitive phenotype of this strain (Tomoyasu et al., 2001). We therefore reasoned that if polyP exerts general chaperone-like protein protection, overproduction of

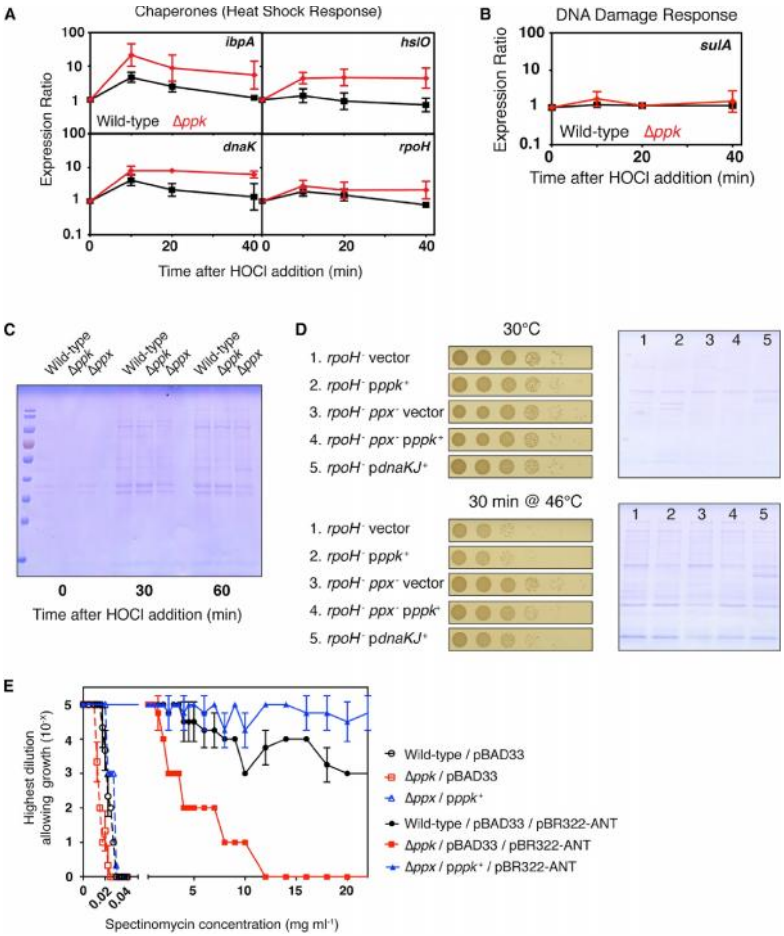


Figure 2. PolyP Is a Protein-Protective Chaperone In Vivo

(A and B) Wild-type (black) and Δ*ppk* (red) *E. coli* were grown to log phase and then treated with 0.4 mM HOCl. Expression of (A) the heat-shock genes *ibpA*, *hslO*, *dnaK*, and *rpoH* or (B) the DNA damage indicator gene *sulA* was measured by qRT-PCR (mean ± SD). See also Figure S2A.

(C) Insoluble protein fractions from exponentially growing *E. coli* wild-type and mutant strains before and after addition of 1 mM HOCl.

(D) Survival of (left panels) and insoluble protein fractions from (right panels) exponentially growing *E. coli rpoH* strains before and after a shift from 30° to 46°C.

(E) *E. coli* strains containing no polyP (red squares), wild-type (black circles), or higher than normal levels of polyP (blue triangles) were grown to log phase, serially diluted, spot titered on agar containing different concentrations of spectinomycin, and scored for growth (mean ± SD). See also Figures S2B and S2C.

polyP should protect this strain against heat shock in a way comparable to that observed for protein chaperones. Since heat stress does not induce significant accumulation of polyP in vivo (Ault-Riché et al., 1998), presumably because the RpoH-dependent proteostasis machinery is fully functional under such stress conditions, we constructed *E. coli* strains that were previously shown to contain different levels of polyP (Crooke et al., 1994). We generated BB7224 derivatives with and without functional

PPX, containing plasmids expressing *ppk* (p*ppk*$^+$) from arabinose-inducible promoters. All BB7224 derivatives that were engineered to overproduce polyP showed substantial improvements in heat-shock survival (Figure 2D, left panels) and decreased levels of insoluble protein aggregates (Figure 2D, right panels). The results were comparable to those seen in cells overproducing DnaK/DnaJ (Figure 2D; Tomoyasu et al., 2001), strongly indicating that polyP plays a direct role in maintaining protein homeostasis in vivo.

To test the ability of polyP to stabilize proteins even under non-stress conditions, we exploited the recent observation that the level of a strain's antibiotic resistance can be used as readout for the in vivo stability of the antibiotic resistance protein (Foit and Bardwell, 2013). We therefore coexpressed various antibiotic resistance proteins in wild-type, Δ*ppk*, or Δ*ppx* p*ppk*$^+$ strains and tested for antibiotic resistance. We observed by far the largest effect of polyP on the spectinomycin resistance-conferring enzyme aminoglycoside 3′-adenyltransferase (ANT), which showed significantly higher antibiotic resistance in PPK-overexpressing strains than in the wild-type, and much lower antibiotic resistance in strains lacking polyP (Figure 2E). Only a very small decrease in spectinomycin resistance was observed for the polyP-deficient strain in the absence of the ANT gene, indicating that the observed effect is largely due to the effect of polyP on ANT. Some apparent stabilization was also seen for chloramphenicol acetyltransferase, but not for the ampicillin resistance protein β-lactamase, which is expressed in the periplasm of *E. coli* (Figures S2B and S2C). These results demonstrate that polyP acts to stabilize cytoplasmic proteins in vivo and effectively protects bacteria against stress conditions that cause protein unfolding and aggregation.

PolyP Functions as a General Molecular Chaperone In Vitro

Although our studies demonstrated that polyP works well in a strain background that lacks most chaperones and proteases, the possibility still remained that the observed in vivo effects of polyP are indirect and mediated by a potential influence of polyP on components of the proteostasis machinery. We therefore decided to directly test, using standard in vitro chaperone assays, whether polyP, like a true protein chaperone, recognizes and binds in vitro unfolding proteins and prevents their irreversible aggregation. We first analyzed the influence of polyP on the aggregation of a variety of previously established chaperone substrate proteins. One such substrate, luciferase, aggregated rapidly when diluted from the urea-denatured form into buffer (Figure 3A, black trace). Astonishingly, micromolar concentrations of polyP (expressed in terms of total phosphate concentration due to the heterogeneous nature of commercially available polyP [Ault-Riché et al., 1998]) inhibited the aggregation of chemically denatured luciferase, with 500 μM polyP completely abolishing luciferase aggregation (Figure 3A, blue trace). PolyP was even more effective in protecting luciferase against thermal aggregation, which typically occurs when luciferase is incubated at temperatures above 40°C (Figure 3B, black trace). The presence of 1 μM polyP was sufficient to significantly reduce thermal aggregation, and 100 μM polyP completely prevented aggregate formation (Figure 3B, compare red and blue lines). To determine whether the observed chaperone activity is indeed polyP specific and not mediated by additional components in our polyP preparations, we added the highly active, polyP-degrading exopolyphosphatase from *Saccharomyces cerevisiae* (ScPPX) (Wurst and Kornberg, 1994) to preformed polyP-luciferase complexes at elevated temperatures. We reasoned that any luciferase that was subsequently released because of polyP degradation should rapidly aggregate under these conditions. Indeed, addition of ScPPX, which has no discernible chaperone activity or aggregation tendency itself (Figure S3A), resulted in immediate, dose-dependent aggregation of luciferase (Figure 3C, blue and green lines). These results showed that polyP was indeed the chaperone-active component in this assay.

The remarkable ability of polyP to stabilize proteins became clearly apparent when we compared the thermal stability of luciferase in the absence and presence of polyP. Although luciferase was completely insoluble upon incubation at 85°C in the absence of polyP (Figure 3D, inset) and had no discernible secondary structure (Figure 3D, black dotted trace), the presence of polyP maintained luciferase in a fully soluble and highly structured form for at least 20 min at these near-boiling temperatures (Figure 3D, red trace). These results suggested that polyP keeps luciferase soluble by stabilizing its secondary structure elements. Consistent with this protein-stabilizing effect, incubation of luciferase with increasing amounts of polyP also increasingly delayed the thermal inactivation of luciferase (Figure 3E, left). However, a quite unexpected finding was that, like a true protein chaperone, polyP maintained thermally inactivated luciferase in a state that was competent for refolding by the DnaK/DnaJ/GrpE system. This result became evident when we diluted luciferase, which had been thermally inactivated either in the absence or presence of polyP, into 25°C buffer containing the DnaK/DnaJ/GrpE ATP-dependent chaperone system. Whereas no significant refolding of luciferase was detected in the sample that lacked polyP during the inactivation (Figure 3E, right inset), significant refolding was achieved when luciferase was heat treated in the presence of polyP (Figure 3E, right). These effects of polyP are very comparable to the effects observed with general protein chaperones such as Hsp33 or the small heat-shock proteins, which bind unfolding proteins during heat inactivation and transfer their clients to the DnaK/DnaJ/GrpE chaperone system upon a temperature shift for refolding (Haslbeck et al., 2005; Hoffmann et al., 2004; Mogk et al., 1999).

Importantly, polyP's protein-protective effects were not restricted to thermally or chemically unfolded luciferase, but extended also to heat- or HOCl-induced protein aggregation of citrate synthase, another commonly used chaperone substrate. For both thermally and HOCl-unfolded citrate synthase, increasing amounts of polyP in the incubation reaction increasingly prevented protein aggregation (Figures 3F and 3G). To obtain a general overview of the proteins that are protected by polyP, and to detect any potential client specificity, we added increasing amounts of polyP to crude extracts of *E. coli* *ppk*::*kan*$^+$ (which expresses neither PPK nor PPX) or wild-type and incubated the cell lysates at heat-shock temperatures. This strategy has been extensively used to detect clients of protein chaperones. It is based on the observation that most of the proteins that aggregate upon stress treatment in intact cells also

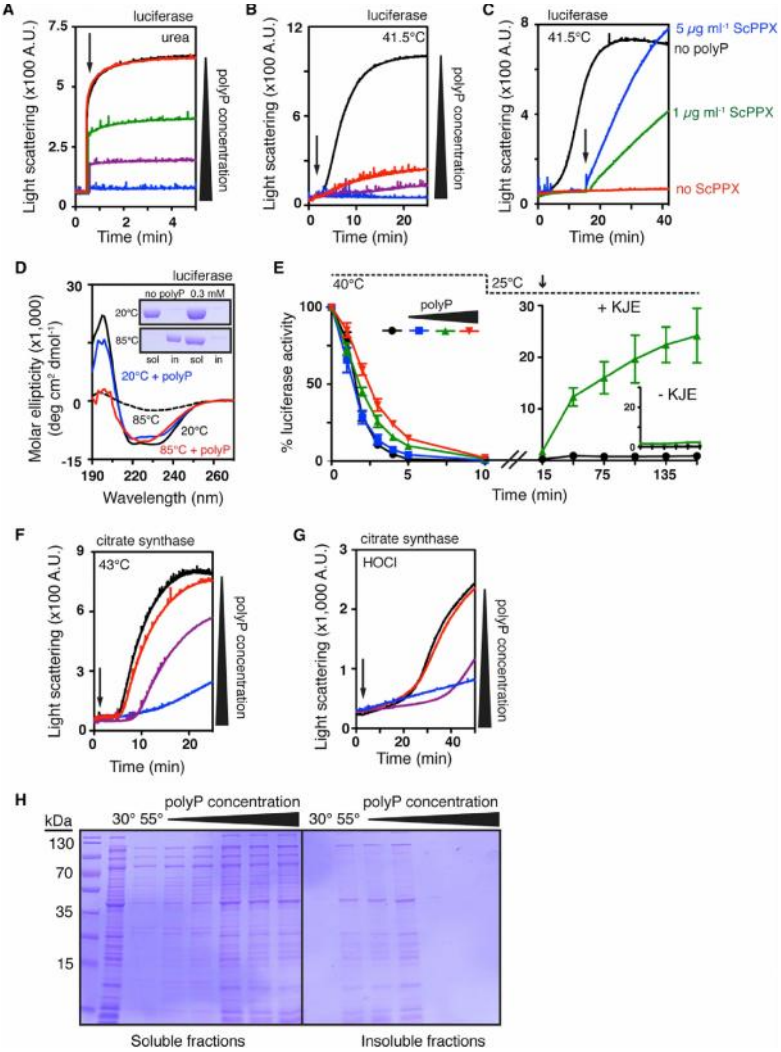


Figure 3. PolyP Is a Protein-Protective Chaperone In Vitro

(A) Aggregation of urea-denatured luciferase upon its dilution (arrow) into buffer containing no (black), 0.5 μM (red), 5 μM (green), 50 μM (purple), or 500 μM (blue) polyP. PolyP concentrations are expressed in terms of concentration of inorganic phosphate equivalents (P_i).

(B) Thermal aggregation of luciferase upon its dilution into prewarmed buffer (arrow) containing no (black), 1 μM (red), 10 μM (purple), or 100 μM (blue) polyP.

(C) Thermal aggregation of luciferase upon its dilution into prewarmed buffer containing no (black) or 0.5 mM polyP (blue, green, red) and 50 μM $MgCl_2$. Arrow indicates addition of 5 μg/ml (blue) or 1 μg/ml (green) ScPPX. See Figure S3A for additional controls.

(legend continued on next page)

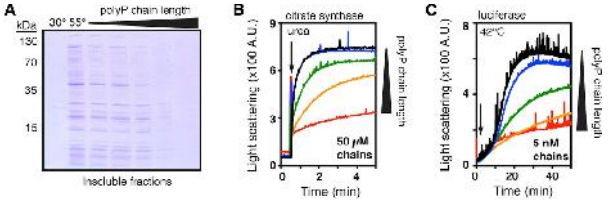


Figure 4. PolyP Chain Length Influences Chaperone Activity

(A) Crude lysates of *E. coli ppk*::*kan*$^+$ were incubated for 30 min at 30°C or 55°C, with 0 or 5 mM of different-length polyP; from left to right: heterogeneous short-chain (average 45-mer), 14-mer, 60-mer, 130-mer, and long-chain (200–1,300-mer). Insoluble fractions were separated and examined by SDS-PAGE.

(B and C) Aggregation of urea-denatured citrate synthase (B) or thermally denatured luciferase (C) upon their dilution into buffer (arrow) containing no polyP (black) or the indicated concentrations of different-length polyP: 14-mer (blue), 60-mer (green), 130-mer (orange), or 300-mer (red). Concentrations were determined according to the length of the polyP chains. Corresponding experiments using total P_i concentration are shown in Figures S4A and S4B.

aggregate in stress-treated cell lysates (Tomoyasu et al., 2001). We found that addition of polyP broadly protected a large range of different proteins against thermal aggregation and maintained them in a soluble form (Figures 3H and S3B). Importantly, the polyP concentrations necessary to protect bacterial proteins ex vivo (2–20 mM) were very similar to both the concentrations required to protect citrate synthase from aggregation in vitro (Figures 3F and 3G) and the ~50 mM polyP concentrations that have been measured in stressed *E. coli* cells (Ault-Riché et al., 1998). These results strongly suggest that polyP is a promiscuous and general protein chaperone.

PolyP Chain Length Determines Its Chaperone Efficacy

The length of polyP chains varies dramatically in nature and depends on both the organism and the cell type (Rao et al., 2009). Although functional differences between long- and short-chain polyPs have been noted, the reasons for these differences are unclear (Smith et al., 2010). To assess whether chain length affects the ability of polyP to protect against protein aggregation, we analyzed the influence of the same concentration (5 mM, based on P_i units) of different defined-length polyPs on the aggregation of thermally unfolding proteins in bacterial cell lysates. We tested the effects of homogeneous preparations of 14-mer, 60-mer, and 130-mer, as well as commercially available heterogeneous mixtures of short-chain polyPs (which were used for all other experiments except where indicated; average: 45 P_i units) and long-chain polyP polymers (range: 200–1,300 P_i units). At this concentration, all of the tested polyP preparations exerted some degree of protein protection (Figure 4A). However, by far the most effective chaperones were the long polyP chains, consisting of either 130-mer or a mixture of long-chain polyP (Figure 4A, right-most two lanes). Both preparations almost completely prevented protein aggregation in *E. coli* cell lysates heated to 55°C. We obtained similar results when we tested the effect of different-chain-length polyPs on the aggregation of urea-denatured citrate synthase or thermally denatured luciferase in vitro. At the minimal concentration of defined-length polyPs that effectively protected against aggregation, calculated either in terms of P_i units (Figures S4A and S4B) or in terms of polyP chains (Figures 4B and 4C), we observed a clear dependence of polyP chaperone efficacy on chain length: the shorter the chain length, the less effective was the protection. These results are consistent with the finding that bacteria preferentially accumulate long-chain polyP (up to 800 P_i units) upon stress conditions (Ault-Riché et al., 1998; Kornberg et al., 1999). Remarkably, 5 nM of 300-mer polyP chains was sufficient to nearly completely protect 130 nM luciferase against aggregation (Figure 4C, red trace). Higher concentrations (50 μM of chains) were required to protect 80 nM citrate synthase. This result may indicate that polyP exerts its chaperone activity differently with different substrates.

Accumulation of PolyP Is Redox Regulated

Our in vivo polyP measurements revealed that for the first 60 min of HOCl treatment, Δ*ppx* cells accumulated polyP with the same kinetics and to the same extent as HOCl-treated wild-type cells (Figure 1B, top panel, compare blue and black lines). This result was very unexpected, since deletion of PPX was predicted to abolish polyP hydrolysis to inorganic phosphate and therefore should lead to higher levels of polyP in these strains. We therefore considered the possibility that HOCl treatment might transiently inactivate PPX. This would allow for the HOCl-mediated accumulation of polyP and explain the inorganic phosphate starvation phenotype during HOCl stress. A similar posttranslational regulation has been proposed to trigger polyP accumulation during the *E. coli* amino acid starvation response. Under these stress conditions, the small signal molecule guanidine 5′,3′-bis-diphosphate (ppGpp) appears to be responsible for the transient

(D) CD spectra of luciferase incubated with or without 0.3 mM polyP at 20°C or 85°C. Inset shows SDS-PAGE of soluble and insoluble luciferase fractions after 20 min incubation at the indicated temperature and spectrum determination.

(E) Left panel: thermal inactivation of luciferase with no (black), 0.1 mM (blue), 1 mM (green), or 10 mM polyP (red) at 40°C. Right panel: reactivation of luciferase thermally inactivated in the absence (black) or presence (green) of 1 mM polyP upon a shift to permissive temperatures and addition of DnaK, DnaJ, GrpE, and MgATP (KJE). Inset shows reactivation in the absence of KJE. Error bars indicate mean ± SD.

(F and G) Thermal- (F) or HOCl- (G) induced aggregation of citrate synthase with no (black), 1 mM (red), 10 mM (purple), or 100 mM (blue) polyP. Arrow indicates time of citrate synthase addition (in F) or HOCl addition (in G).

(H) Crude lysates of *E. coli ppk*::*kan*$^-$ were incubated for 30 min at 30°C or 55°C, with 0, 0.2, 1, 2, 10, or 20 mM polyP. Soluble and insoluble fractions were separated and examined by SDS-PAGE. See Figure S3B for results with wild-type lysates.

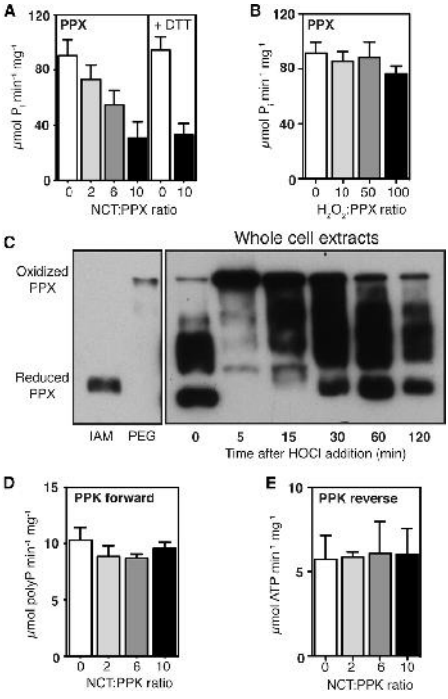


Figure 5. PPX Is a Redox-Regulated Enzyme

(A) Specific activity of PPX after incubation with different concentrations of *N*-chlorotaurine (NCT) (mean ± SD). To test for reversibility, oxidatively inactivated PPX (NCT:PPX 10:1) was incubated for 1 hr with 5 mM DTT and assayed again. See also Figure S4A.

(B) Specific activity of PPX after treatment with H_2O_2 (mean ± SD).

(C) *E. coli* overexpressing PPX was grown to log phase and then treated with 1 mM HOCl. Reduced cysteine thiols were alkylated with iodoacetamide (IAM). Oxidized cysteine thiols were reduced and alkylated with PEG-maleimide (PEG), with 2 kDa molecular mass added per modified cysteine. PPX was visualized by western blot. Fully IAM-labeled ("Reduced") or PEG-maleimide-labeled ("Oxidized") PPX standards are from the same blot, with longer exposure times.

(D and E) Specific activity of PPK's forward and reverse reactions (see Figure 1C) before and after treatment with NCT (mean ± SD). See also Figure S4B.

inhibition of PPX (Kuroda et al., 1997). However, since HOCl treatment does not induce gene-expression changes consistent with ppGpp accumulation (Durfee et al., 2008; Gray et al., 2013b), that mechanism is unlikely to account for HOCl-mediated polyP accumulation. Instead, HOCl is a highly thiol-reactive agent (Gray et al., 2013a), and there is large precedence for organisms using redox-regulated proteins to rapidly mount a stress defense against oxidative protein unfolding stress (Antelmann and Helmann, 2011). We therefore tested the redox sensitivity of PPX in vitro and in vivo. Purified PPX proved to be highly sensitive to inactivation by *N*-chlorotaurine (which causes many fewer nonspecific oxidation artifacts than HOCl [Chapman et al., 2003]) but not by H_2O_2 (compare Figures 5A and 5B). Mass-spectrometric analysis of *N*-chlorotaurine-treated PPX (Table S1) revealed the formation of sulfonic acid, an irreversible thiol modification, on Cys169, which is located in the predicted binding site for polyP (Alvarado et al., 2006), and, to a lesser extent, on the surface-exposed Cys85 (Figure S5A). This result helps to explain not only why oxidation of PPX leads to its inactivation but also why this inactivation is irreversible in vitro (Figure 5A). To monitor PPX oxidation and test for its reversibility in vivo, we conducted differential thiol-trapping experiments in *E. coli* overexpressing PPX (endogenous PPX levels were undetectable with our antibodies) at different time points after HOCl treatment. We alkylated all in vivo reduced cysteines and then labeled all in vivo oxidized cysteines, upon their ex vivo reduction, with the 2 kDa thiol-specific alkylating agent PEG-maleimide. This modification, which indicates the presence of in vivo reversible thiol modifications, leads to significant mass shifts that can be visualized by SDS-PAGE and western blotting. Within the first 5 min of HOCl treatment, the majority of endogenous PPX shifted to a slower-migrating species, indicating reversible thiol oxidation in PPX (Figure 5C). After about 60 min of HOCl treatment, at least 50% of PPX molecules were again in their reduced state, correlating well with the time at which polyP accumulation began to level off in wild-type cells while it continued to increase in Δ*ppx* cells (Figure 1B). These results provide evidence that PPX is a redox-regulated enzyme whose HOCl-mediated oxidation and concomitant transient inactivation contribute to the rapid polyP accumulation and P_i depletion in HOCl-stressed bacterial cells. Purified PPK, which also contains a cysteine residue in close proximity to its active site (Zhu et al., 2005; Figure S5B), was not affected by treatment with *N*-chlorotaurine (Figures 5D and 5E), indicating that the ability of PPK to synthesize polyP and potentially convert it back to ATP is not affected by HOCl treatment.

DISCUSSION

Here, we provide evidence that polyP is an ancient, universally conserved, highly effective, and wholly inorganic protein-protective chaperone, which may go some way toward explaining the complex pleiotropic phenotypes associated with polyP deficiencies in both prokaryotes and eukaryotes (Docampo et al., 2010; Rao et al., 2009). Our studies in bacteria identified polyP as a key component of a powerful, redox-regulated system for dealing with the proteotoxic effects of fast-acting oxidants such as HOCl. PolyP appears to counteract these proteotoxic effects by stabilizing proteins, preventing irreversible aggregation and maintaining them in a refolding-competent conformation. These are all typical features of protein chaperones, such as the small heat-shock proteins and Hsp33. Like protein chaperones, polyP does not appear to have any significant substrate specificity and stabilizes a wide variety of different proteins. Although we cannot exclude the possibility that polyP also has additional indirect or regulatory effects on proteostasis in vivo,

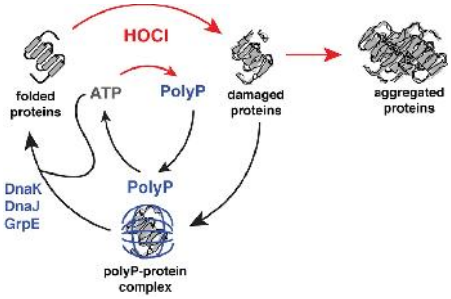


Figure 6. Model of the PolyP Chaperone Cycle
The antimicrobial oxidant HOCl damages proteins, causing the formation of cytotoxic protein aggregates. HOCl stimulates rapid conversion of cellular ATP to polyP through the oxidative inactivation of PPX. This conversion conserves high-energy phospho-anhydride bonds while downregulating cellular processes that require ATP, including ATP-dependent chaperones such as DnaK. PolyP functionally replaces these chaperones by forming stable complexes with unfolding proteins, keeping them soluble and refolding competent. Upon relief of stress conditions, polyP may be either degraded to free phosphate by PPX or reconverted to ATP by PPK. Restoration of cellular ATP pools reactivates ATP-dependent chaperones and allows for the effective refolding of polyP-protected proteins by the DnaK/DnaJ/GrpE complex.

it is clear from our results that polyP is able to directly stabilize a wide variety of proteins against multiple forms of unfolding stresses.

These results raise the obvious question of how polyP works as a chaperone. Most known chemical chaperones work in a protective osmolyte-like fashion, requiring very high (often molar) concentrations and stabilizing proteins via their strong interactions with the solvent (Canchi and García, 2013). In contrast, protein chaperones work in stoichiometric fashion and contain either defined binding sites, which consist of a mixture of hydrophobic and charged residues (Kim et al., 2013), or intrinsically disordered protein regions, which form upon client binding in a scaffold-like fashion (Kim et al., 2013; Reichmann et al., 2012). PolyP chains are effective at low micromolar concentrations, and their ability to protect proteins against protein aggregation increases with the length of their chain. It is therefore possible that polyP functions as a chemical scaffold, keeping proteins soluble by stabilizing secondary motifs. Alternatively, or in combination with the above functions, ionic interactions between the negatively charged polyP and positive side chains in proteins, as well as the high concentration of cations associated with polyP (Kulaev et al., 2004), might contribute to the stabilization effect.

Synthesis of polyP does not require transcription or translation. This makes polyP an excellent chaperone during stress conditions, such as HOCl stress, that not only cause protein unfolding but also inhibit new protein translation and inactivate ATP-dependent chaperones, such as the DnaJ/DnaK/GrpE system (Ling and Söll, 2010; Winter et al., 2008). In fact, it is tempting to speculate that synthesis of polyP is part of a larger scheme in which the high-energy phosphate bonds of ATP are fully preserved during the period in which ATP-dependent processes, such as protein translation and ATP-dependent molecular chaperone function, are stalled. This mechanism avoids the costly de novo synthesis of oxidation-prone polypeptide chains and prevents futile cycles between chaperones and unfolding clients under conditions that are nonpermissive for folding (Kim et al., 2013). Converting ATP directly into an oxidation-resistant chemical chaperone that binds tightly to and stabilizes unfolding proteins provides immediate compensation for the lack of chaperones. Upon relief of stress, polyP can then be rapidly reconverted to ATP by PPK, restoring cellular energy pools and allowing ATP-dependent chaperones to refold polyP-stabilized proteins (Figure 6). Our findings not only expand the complex redox-regulated network that bacteria use to resist the protein-damaging effects of HOCl (Drazic et al., 2013; Gray et al., 2013b; Winter et al., 2008) but also demonstrate a fundamentally important function for polyP, one of the most conserved molecules in biology. PolyP's protein-protective chaperone activity may be key to understanding its fundamental roles and diverse phenotypes in growth, development, virulence, and stress response in both prokaryotes and eukaryotes.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Growth Conditions, and Molecular Methods

All strains and plasmids used in this study are listed in the Table of Strains and Plasmids in Supplemental Experimental Procedures. Bacterial manipulations, protein purifications, enzyme assays, thiol trapping, and mass spectrometry were carried out according to standard methods; details are available in Supplemental Experimental Procedures.

Phenotypic and Transcriptional Analyses

HOCl survival assays and quantitative RT-PCR (qRT-PCR) of *E. coli* were performed as previously described (Gray et al., 2013b). The survival assays used for *V. cholerae* were the same as those employed for *E. coli* except that 550 μM HOCl was used instead of 2 mM HOCl, and 2 mM methionine instead of 10 mM sodium thiosulfate ($Na_2S_2O_3$) was used to quench HOCl. For *N*-chlorotaurine tolerance, *E. coli* was grown in 3-(N-morpholino)propanesulfonic acid (MOPS) medium at 37°C to an OD_{600} ~0.4, and then concentrated by centrifugation and resuspended to OD_{600} = 0.1 in the same medium with and without 120 μM *N*-chlorotaurine. Growth curves were collected using a Bio-Tek Synergy HT plate reader, and cultures were incubated at 37°C with constant shaking.

Quantitative High-Throughput PolyP Assay

Intracellular polyP levels were measured according to Ault-Riché et al. (1998) with slight modifications. A Biomek FX fluid-handling robot (Beckman Coulter) was used to automate polyP extraction, digestion, and measurement; full details are available in Supplemental Experimental Procedures.

In Vivo Protein Aggregation Assays

Membrane protein-free cellular insoluble protein fractions were prepared by a modification of the method of Tomoyasu et al. (2001). Briefly, cells equivalent to 4 ml of OD_{600} = 1 were harvested by centrifugation and resuspended in 40 μl Buffer 1 (10 mM potassium phosphate [pH 6.5], 1 mM EDTA, 20% [w/v] sucrose, 1 mg/ml lysozyme, 50 U/ml Benzonase nuclease [Merck]), and then incubated 30 min on ice and frozen at −80°C. After thawing on ice and addition of 360 μl Buffer 2 (10 mM potassium phosphate [pH 6.5], 1 mM EDTA), the cells were transferred to 2 ml microfuge tubes containing ~200 μl 0.5 mm glass beads (BioSpec Products) and shaken for 30 min at 1,400 rpm, 8°C, to lyse the cells completely. Then 200 μl aliquots were taken and insoluble fractions were separated by centrifugation (20 min at 16,100 *g*, 4°C), rinsed once with

Buffer 2, once with Buffer 3 (Buffer 2 plus 2% Nonidet P-40 [ICN Biomedicals]), and again with Buffer 2, and then visualized by reducing SDS-PAGE. For HOCl stress, *E. coli* strains were grown in MOPS medium at 37°C to OD_{600} ~1 and then diluted to OD_{600} = 0.35 with fresh medium. HOCl was added to 1 mM and incubation was continued. Samples were taken at the indicated time points, with HOCl quenched by immediate addition of $Na_2S_2O_3$ to 10 mM. For heat stress, BB7224-derived *E. coli* strains were grown at 30°C in lysogeny broth (LB) medium containing ampicillin and 1% arabinose to OD_{600} ~1, and then diluted to OD_{600} = 0.6 with fresh medium and incubated at 46°C. Samples were taken at the indicated time points and cooled rapidly, and cell survival was assessed at 30°C by serial dilution on LB agar plates containing 1% arabinose.

In Vivo Effect of PolyP on Antibiotic Resistance Protein Stability

The *E. coli* strains MG1655/pBAD33, Δ*ppk*/pBAD33, and Δ*ppx*/pPPK1 were transformed with pBR322-ANT. Overnight cultures of these strains were diluted 1:100 in fresh LB and grown to log phase at 37°C. OD_{600} was normalized to one with PBS. Cells were serially diluted in PBS, and 2 μl of 10^0–10^{-5} dilutions were spotted on LB agar plates containing 1% arabinose and the indicated concentrations of antibiotics.

In Vitro Protein Aggregation Assays

For aggregation of denatured proteins, citrate synthase or luciferase (12 μM) was denatured in urea (6.5 M for luciferase, 7.5 M for citrate synthase) for 2 hr and then diluted to 60 nM (luciferase) or 80 nM (citrate synthase) in 40 mM potassium phosphate (pH 7.5) at 30°C containing polyP as indicated. For thermally induced protein aggregation, citrate synthase (0.30 μM) was incubated in 40 mM potassium phosphate (pH 7.5) at 43°C, or luciferase (0.13 μM) was incubated in 40 mM HEPES (pH 7.5) at 41.5°C–43°C, with the indicated amounts of polyP. For HOCl-induced protein aggregation, citrate synthase (3 μM) was incubated in 40 mM potassium phosphate (pH 7.5) at 30°C with 350 μM HOCl. Light scattering was measured at λ_{ex} and λ_{em} = 360 nm using a Hitachi F4500 fluorescence spectrophotometer with a thermostated cuvette holder under constant stirring. All experiments were performed at least in triplicate. Each panel shows representative results obtained on a single day with a single batch of protein. For ex vivo aggregation assays, 200 μg aliquots of crude cell lysates with polyP added as indicated were incubated for 30 min with shaking (650 rpm) at the indicated temperatures. Soluble and insoluble fractions were separated by centrifugation (20 min at 16,100 *g*, 4°C) and visualized by reducing SDS-PAGE.

Thermal Inactivation and Refolding of Luciferase

DnaK, DnaJ, and GrpE were purified as previously described (Hoffmann et al., 2004). Luciferase (4 μM) was incubated in 10 mM potassium phosphate (pH 7.5) at 40°C and samples were removed at the indicated time points. After 10 min, the samples were transferred to 25°C for 5 min and then diluted 1:40 into 40 mM HEPES (pH 7.5), 50 mM KCl, 0.1 mg/ml bovine serum albumin, 2 mM MgATP with or without the addition of 1 μM DnaK, 0.2 μM DnaJ, and 1 μM GrpE. Luciferase activity was determined by luminescence (Lundin, 2000) in reaction mixtures containing 20 nM luciferase, 25 mM Tricine (pH 7.8), 5 mM $MgSO_4$, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 35 μM luciferin, and 2 mM MgATP.

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra (190–260 nm) were measured for 0.4 mg/ml luciferase in 10 mM potassium phosphate (pH 7.5) with or without 0.3 mM polyP at 20°C and 85°C, using a J-810 CD spectrophotometer (Jasco). After measurement, soluble and insoluble protein fractions of each treatment were separated by centrifugation (20 min at 16,100 *g*, 4°C) and visualized by reducing SDS-PAGE.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.01.012.

ACKNOWLEDGMENTS

We thank Dr. Carol Gross for her extremely helpful advice on preparing this manuscript. We are grateful to Dr. Toshikazu Shiba (Regenetiss, Japan) for providing us with polyphosphates of defined chain length. This work was funded by National Institutes of Health grants GM065318 and AI097893 to U.J., F32-GM096613 to M.J.G., and T32-GM007315 to W.Y.W. J.C.A.B. is a Howard Hughes Medical Institute investigator.

Received: July 29, 2013
Revised: December 10, 2013
Accepted: January 16, 2014
Published: February 20, 2014

REFERENCES

Achbergerová, L., and Nahálka, J. (2011). Polyphosphate—an ancient energy source and active metabolic regulator. Microb. Cell Fact. *10*, 63.

Alvarado, J., Ghosh, A., Janovitz, T., Jauregui, A., Hasson, M.S., and Sanders, D.A. (2006). Origin of exopolyphosphatase processivity: fusion of an ASKHA phosphotransferase and a cyclic nucleotide phosphodiesterase homolog. Structure *14*, 1263–1272.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230.

Antelmann, H., and Helmann, J.D. (2011). Thiol-based redox switches and gene regulation. Antioxid. Redox Signal. *14*, 1049–1063.

Aschar-Sobbi, R., Abramov, A.Y., Diao, C., Kargacin, M.E., Kargacin, G.J., French, R.J., and Pavlov, E. (2008). High sensitivity, quantitative measurements of polyphosphate using a new DAPI-based approach. J. Fluoresc. *18*, 859–866.

Ault-Riché, D., Fraley, C.D., Tzeng, C.M., and Kornberg, A. (1998). Novel assay reveals multiple pathways regulating stress-induced accumulations of inorganic polyphosphate in Escherichia coli. J. Bacteriol. *180*, 1841–1847.

Barrette, W.C., Jr., Albrich, J.M., and Hurst, J.K. (1987). Hypochlorous acid-promoted loss of metabolic energy in Escherichia coli. Infect. Immun. *55*, 2518–2525.

Booth, I.R., Ferguson, G.P., Miller, S., Li, C., Gunasekera, B., and Kinghorn, S. (2003). Bacterial production of methylglyoxal: a survival strategy or death by misadventure? Biochem. Soc. Trans. *31*, 1406–1408.

Canchi, D.R., and García, A.E. (2013). Cosolvent effects on protein stability. Annu. Rev. Phys. Chem. *64*, 273–293.

Chapman, A.L., Winterbourn, C.C., Brennan, S.O., Jordan, T.W., and Kettle, A.J. (2003). Characterization of non-covalent oligomers of proteins treated with hypochlorous acid. Biochem. J. *375*, 33–40.

Crooke, E., Akiyama, M., Rao, N.N., and Kornberg, A. (1994). Genetically altered levels of inorganic polyphosphate in Escherichia coli. J. Biol. Chem. *269*, 6290–6295.

Deborde, M., and von Gunten, U. (2008). Reactions of chlorine with inorganic and organic compounds during water treatment-Kinetics and mechanisms: a critical review. Water Res. *42*, 13–51.

Docampo, R., Ulrich, P., and Moreno, S.N.J. (2010). Evolution of acidocalcisomes and their role in polyphosphate storage and osmoregulation in eukaryotic microbes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 775–784.

Drazic, A., Miura, H., Peschek, J., Le, Y., Bach, N.C., Kriehuber, T., and Winter, J. (2013). Methionine oxidation activates a transcription factor in response to oxidative stress. Proc. Natl. Acad. Sci. USA *110*, 9493–9498.

Durfee, T., Hansen, A.M., Zhi, H., Blattner, F.R., and Jin, D.J. (2008). Transcription profiling of the stringent response in Escherichia coli. J. Bacteriol. *190*, 1084–1096.

Foit, L., and Bardwell, J.C. (2013). A tripartite fusion system for the selection of protein variants with increased stability in vivo. Methods Mol. Biol. *978*, 1–20.

Gray, M.J., Wholey, W.Y., and Jakob, U. (2013a). Bacterial responses to reactive chlorine species. Annu. Rev. Microbiol. *67*, 141–160.

Gray, M.J., Wholey, W.Y., Parker, B.W., Kim, M., and Jakob, U. (2013b). NemR is a bleach-sensing transcription factor. J. Biol. Chem. *288*, 13789–13798.

Guisbert, E., Yura, T., Rhodius, V.A., and Gross, C.A. (2008). Convergence of molecular, modeling, and systems approaches for an understanding of the Escherichia coli heat shock response. Microbiol. Mol. Biol. Rev. *72*, 545–554.

Haslbeck, M., Franzmann, T., Weinfurtner, D., and Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. Nat. Struct. Mol. Biol. *12*, 842–846.

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

Holmstrom, K.M., Marina, N., Baev, A.Y., Wood, N.W., Gourine, A.V., and Abramov, A.Y. (2013). Signalling properties of inorganic polyphosphate in the mammalian brain. Nat. Commun. *4*, 1362.

Hyslop, P.A., Hinshaw, D.B., Halsey, W.A., Jr., Schraufstätter, I.U., Sauerheber, R.D., Spragg, R.G., Jackson, J.H., and Cochrane, C.G. (1988). Mechanisms of oxidant-mediated cell injury. The glycolytic and mitochondrial pathways of ADP phosphorylation are major intracellular targets inactivated by hydrogen peroxide. J. Biol. Chem. *263*, 1665–1675.

Imlay, J.A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat. Rev. Microbiol. *11*, 443–454.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. Annu. Rev. Biochem. *82*, 323–355.

Kornberg, A., Rao, N.N., and Ault-Riché, D. (1999). Inorganic polyphosphate: a molecule of many functions. Annu. Rev. Biochem. *68*, 89–125.

Kulaev, I.S., Vagabov, V.M., and Kulakovskaya, T.V. (2004). The Biochemistry of Inorganic Polyphosphates. (Chichester, England: John Wiley & Sons, Ltd.), pp. 3–13.

Kulakovskaya, T.V., Vagabov, V.M., and Kulaev, I.S. (2012). Inorganic polyphosphate in industry, agriculture and medicine: modern state and outlook. Process Biochem. *47*, 1–10.

Kuroda, A., Murphy, H., Cashel, M., and Kornberg, A. (1997). Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphosphate in Escherichia coli. J. Biol. Chem. *272*, 21240–21243.

Ling, J., and Söll, D. (2010). Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. Proc. Natl. Acad. Sci. USA *107*, 4028–4033.

Lundin, A. (2000). Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. Methods Enzymol. *305*, 346–370.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rüdiger, S., Röder, D., Langen, H., and Bukau, B. (1999). Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. EMBO J. *18*, 6934–6949.

Moreno, S.N., and Docampo, R. (2013). Polyphosphate and its diverse functions in host cells and pathogens. PLoS Pathog. *9*, e1003230.

Nagl, M., Hess, M.W., Pfaller, K., Hengster, P., and Gottardi, W. (2000). Bactericidal activity of micromolar N-chlorotaurine: evidence for its antimicrobial function in the human defense system. Antimicrob. Agents Chemother. *44*, 2507–2513.

Noller, H.F. (2004). The driving force for molecular evolution of translation. RNA *10*, 1833–1837.

Powers, E.T., and Balch, W.E. (2013). Diversity in the origins of proteostasis networks–a driver for protein function in evolution. Nat. Rev. Mol. Cell Biol. *14*, 237–248.

Quan, S., Koldewey, P., Tapley, T., Kirsch, N., Ruane, K.M., Pfizenmaier, J., Shi, R., Hofmann, S., Foit, L., Ren, G., et al. (2011). Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nat. Struct. Mol. Biol. *18*, 262–269.

Rao, N.N., Gómez-García, M.R., and Kornberg, A. (2009). Inorganic polyphosphate: essential for growth and survival. Annu. Rev. Biochem. *78*, 605–647.

Reichmann, D., Xu, Y., Cremers, C.M., Ilbert, M., Mittelman, R., Fitzgerald, M.C., and Jakob, U. (2012). Order out of disorder: working cycle of an intrinsically unfolded chaperone. Cell *148*, 947–957.

Smith, S.A., Choi, S.H., Davis-Harrison, R., Huyck, J., Boettcher, J., Rienstra, C.M., and Morrissey, J.H. (2010). Polyphosphate exerts differential effects on blood clotting, depending on polymer size. Blood *116*, 4353–4359.

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

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

Winter, J., Ilbert, M., Graf, P.C., Ozcelik, D., and Jakob, U. (2008). Bleach activates a redox-regulated chaperone by oxidative protein unfolding. Cell *135*, 691–701.

Wurst, H., and Kornberg, A. (1994). A soluble exopolyphosphatase of Saccharomyces cerevisiae. Purification and characterization. J. Biol. Chem. *269*, 10996–11001.

Zhu, Y., Huang, W., Lee, S.S., and Xu, W. (2005). Crystal structure of a polyphosphate kinase and its implications for polyphosphate synthesis. EMBO Rep. *6*, 681–687.

Molecular Cell, Volume *53*

Supplemental Information

Polyphosphate Is a Primordial Chaperone

Michael J. Gray, Wei-Yun Wholey, Nico O. Wagner, Claudia M. Cremers, Antje Mueller-Schickert, Nathaniel T. Hock, Adam G. Krieger, Erica M. Smith, Robert A. Bender, James C.A. Bardwell, and Ursula Jakob

SUPPLEMENTAL DATA:

FIGURES:

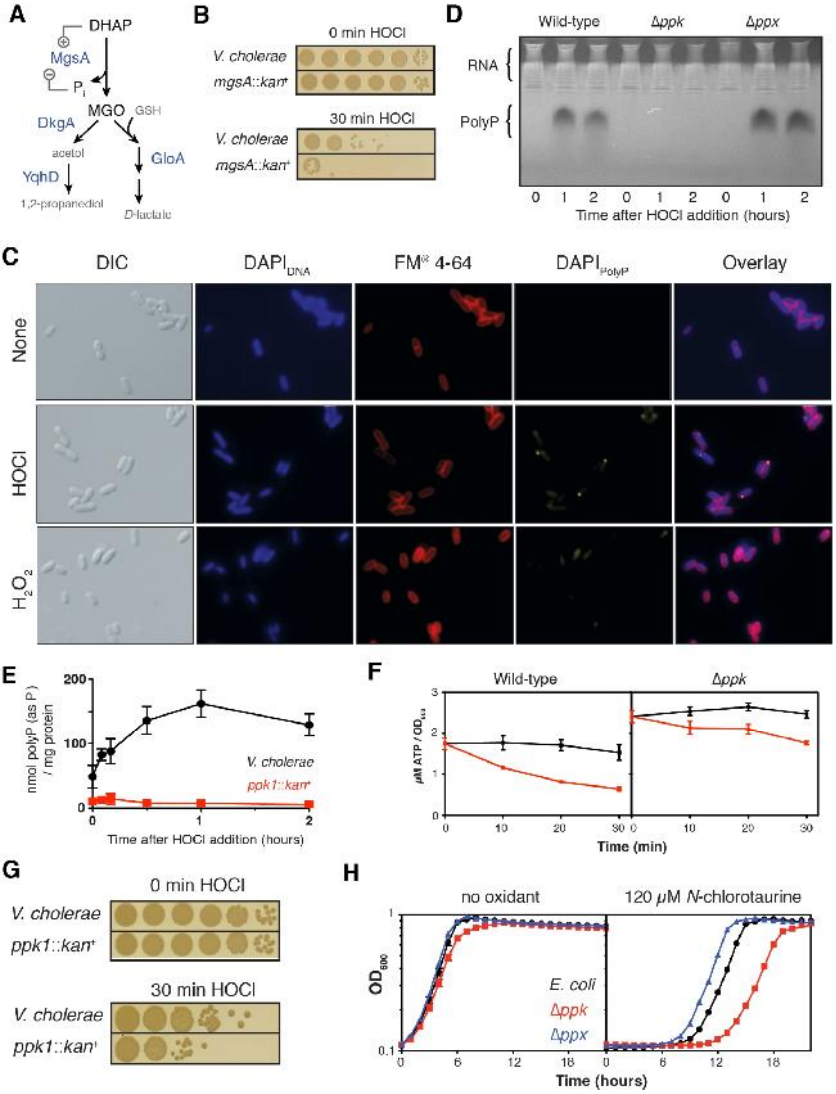


Figure S1, Related to Figure 1. ATP-derived PolyP Protects Against HOCl.

(A) Methylglyoxal (MGO) biosynthesis and detoxification pathways in *E. coli*. Abbreviations: DHAP, dihydroxyacetone phosphate; P_i, inorganic phosphate; GSH, reduced glutathione.

(B) Exponentially growing *V. cholerae* were incubated with 550 µM HOCl, serially diluted and spot-titered.

(C) *E. coli* MG1655 was grown to log phase, then incubated in MOPS medium containing no additive (None), 2 mM HOCl, or 2 mM H_2O_2 for 30 min. Cells were stained with DAPI and FM® 4-64 and visualized by differential interference contrast (DIC) microscopy. DNA ($DAPI_{DNA}$), cell membranes (FM® 4-64), and polyP granules ($DAPI_{polyP}$) were visualized by fluorescence microscopy. Overlay shows results for all strains.

(D) *E. coli* strains were grown to log phase in MOPS medium, then treated with 1 mM HOCl. PolyP was separated on TBE-Urea PAGE gels (Bio-Rad) and visualized by negative DAPI staining.

(E) PolyP content of *V. cholerae* after addition of 550 µM HOCl (mean ± SD).

(F) Cellular ATP levels in MG1655 (Wild-type) and the Δ*ppk* mutant grown to log phase, then treated with 0 (black) or 1 mM (red) HOCl; ATP concentration is expressed as µM ATP per OD_{600} of cell culture (mean ± SD).

(G) Exponentially growing *V. cholera* were incubated with 550 µM HOCl, then diluted and spot-titered.

(H) *E. coli* wild-type (black circles), Δ*ppk* (red squares), and Δ*ppx* (blue triangles) strains were grown at 37°C with aeration in MOPS medium with or without 120 µM *N*-chlorotaurine. Growth was monitored by A_{600} (mean ± SD).

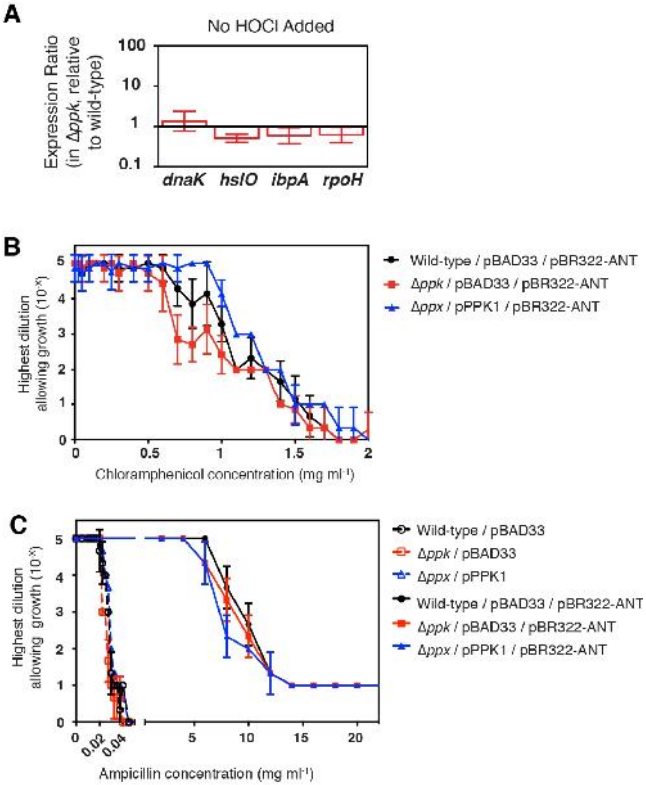


Figure S2, Related to Figure 2. PolyP is a Protein-Protective Chaperone *In Vivo*.

(A) Relative expression of heat shock genes in wild-type and Δ*ppk* strains in the absence of HOCl stress was measured by qRT-PCR (mean ± SD).

(B) and (C) Plasmid-bearing *E. coli* containing no polyP (red squares), wild-type (black circles), or higher than normal levels of polyP (blue triangles) were grown to log phase, serially diluted, spot-titered on agar containing different concentrations of chloramphenicol (B) or ampicillin (C) and scored for growth (mean ± SD).

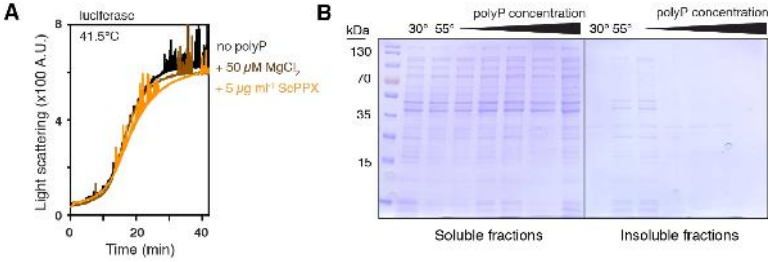


Figure S3, Related to Figure 3. PolyP is a Protein-Protective Chaperone *In Vitro*.

(A) Thermal aggregation of luciferase upon its dilution into pre-warmed buffer (black) or buffer containing 50 μM $MgCl_2$ (brown) or 5 μg ml^{-1} ScPPX (orange).

(B) Crude lysates of *E. coli* MG1655 were incubated 30 minutes at 30° or 55°C, with 0, 0.2, 1, 2, 6, or 10 mM polyP. Soluble and insoluble fractions were separated and examined by SDS-PAGE.

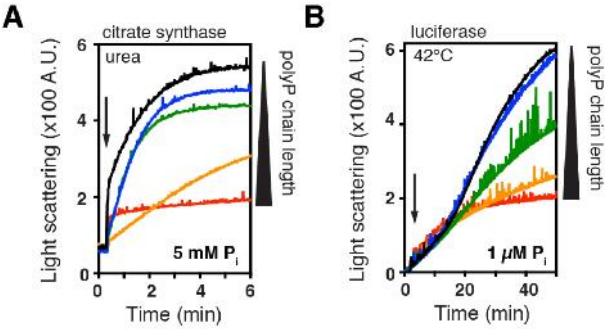


Figure S4, Related to Figure 4. PolyP Chain Length Influences Chaperone Activity

Aggregation of (A) urea-denatured citrate synthase or (B) thermally-denatured luciferase upon their dilution into buffer containing no polyP (black) or the indicated concentrations of different length polyP: 14-mer (blue), 60-mer (green), 130-mer (orange), or 300-mer (red). Concentrations were determined in terms of total phosphate concentration. Arrows indicate time of protein addition.

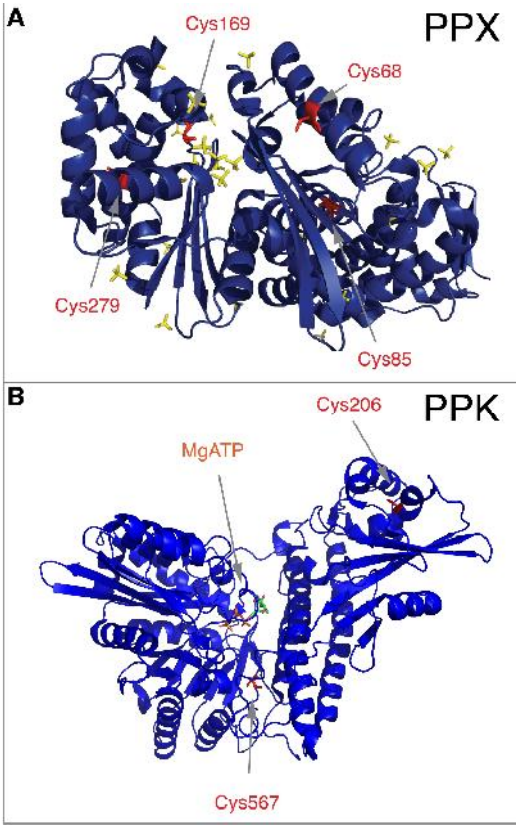


Figure S5; Related to Figure 5. PPX is a Redox-Regulated Enzyme.

(A) The structure of *E. coli* PPX monomer (PDB 1U67) (Alvarado et al., 2006) is shown. Cysteine residues are indicated in red, sulfate ions (thought to indicate the region of PPX to which polyP binds) are indicated in yellow. Cys169 is located directly within the predicted polyP binding site of PPX.

(B) The structure of a monomer of PPK (PDB 1XDO) (Zhu et al., 2005). Cysteine residues are indicated in red and MgATP in the active site is shown as sticks. The highly conserved Cys567 is located directly within the predicted active site of PPK.

SUPPLEMENTAL TABLES:

Supplemental Table S1, Related to Figure 5. Cysteine Oxidation State of PPX *In Vitro*.

no oxidant added[a]

Cysteine	IAM (-SH)	NEM (-SOH)[b]	$-SO_2H$	$-SO_3H$	no modification
68	22[c]	16	0	1	0
85	42	33	0	1	0
169	24	14	0	1	0
279	2	3	0	0	0

10:1 *N*-chlorotaurine

Cysteine	IAM (-SH)	NEM (-SOH)	$-SO_2H$	$-SO_3H$	no modification
68	25	17	0	0	0
85	25	32	0	12	0
169	7	7	2	24	2
279	1	3	0	1	0

[a] PPX incubated aerobically without reducing agent; resulting in the observed proportion of NEM-modified (reversibly oxidized) cysteine residues in the "no oxidant" PPX treatment.

[b] Includes all DTT-reversible cysteine oxidations.

[c] Spectral counts for PPX peptides containing cysteine residues with the indicated modifications identified by LC MS/MS (MS Bioworks).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES:

Bacterial Strains and Growth Conditions

All DNA manipulations were carried out in *E. coli* XL1-Blue (Stratagene). The identity of inserts in all plasmids constructed during this study was confirmed by sequencing (GENEWIZ, Inc. or GenScript, Inc.). Null mutations in *E. coli* MG1655 (Blattner et al., 1997) and BB7224 (Tomoyasu et al., 2001) were constructed as previously described (Baba et al., 2006; Datsenko and Wanner, 2000; Silhavy et al., 1984). *V. cholerae* El Tor C6706*lacZ* and its *ppk1*-TnFGL3::*kan*$^+$ and *mgsA*-TnFGL3::*kan*$^+$ derivatives were from the defined transposon insertion library of Cameron *et al.* (Cameron et al., 2008), and were provided by Victor DiRita (University of Michigan). *E. coli* and *V. cholerae* were grown at 37°C in lysogenic broth (LB; Fisher) or in MOPS minimal medium (Teknova) containing 0.2% glucose, 1.32 mM K_2HPO_4, and 10 μM thiamine. Unless indicated, chemicals were from Fisher or Sigma-Aldrich. Short-chain heterogeneous sodium polyP (Type 45 sodium phosphate glass) was from Sigma-Aldrich or Acros Organics. Long-chain heterogeneous sodium polyP was from KeraFAST. Defined-length polyphosphates were a gift from Dr. Toshikazu Shiba (Regenetiss, Japan). *N*-chlorotaurine was synthesized before each use (Peskin and Winterbourn, 2001).

Sequence Analysis and Primer Design

Gene and protein sequences were obtained from the Integrated Microbial Genomes database (http://img.jgi.doe.gov). PCR and sequencing primers were designed with Web Primer (http://www.yeastgenome.org/cgi-bin/web-primer). qRT-PCR primers were designed with Primer3 0.4.0 (http://frodo.wi.mit.edu/primer3/input.htm) for the following genes: *rrsD*, 5' AAG AAC TTA CCT GGT CTT GAC ATC 3' and 5' CAG TTT ATC ACT GGC AGT CTC CTT 3'; *ibpA*, 5' TGC TAT TGG ATT TGA CCG TTT G 3' and 5' CGG CAC GTT ATA CGG AGG GTA GCC G 3'; *hslO*, 5' TGC CGC AAC ATG ACC AAT TAC ATC GC 3' and 5' TCA CCA TCA AAC TTC AGC GTA GCG GT 3'; *dnaK*, 5' ACA GCA CCC GTA AGC AGG TTG AAG AA 3' and 5' TGG

GCG ATT TCC ATC AGT TTC TGG GA 3'; *rpoH*, 5' GAT TTG ATT CAG GAA GGT AAC ATC G 3' and 5' GTT TTA CGC AGG TTG AAG AAC AGT 3'; *sulA*, 5' CGG GCT TAT CAG TGA AGT TGT CTA T 3' and 5' CTG GCT AAT CTG CAT TAC TTT CGT T 3'.

Strain Construction

In-frame replacements of complete coding sequences with chloramphenicol resistance cassettes (Datsenko and Wanner, 2000) were constructed for the following genes in *E. coli* MG1655, using the indicated primers: *mgsA*: 5' ATA AGT GCT TAC AGT AAT CTG TAG GAA AGT TAA CTA CGG ATG TAC ATT ATG GTG TAG GCT GGA GCT GCT TC 3' and 5' TGG CGA GAA AAC CGT AAG AAA CAG GTG GCG TTT GCC ACC TGT GCA ATA TTA CAT ATG AAT ATC CTC CTT AG 3'; *yqhD*: 5' GGC AGA TCG TTC TCT GCC CTC ATA TTG GCC CAG CAA AGG GAG CAA GTA ATG GTG TAG GCT GGA GCT GCT TC 3' and 5' TAA GTC TGG ACG AAA TGC CCG AAA ACG AAA GTT TGA GGC GTA AAA AGC TTA CAT ATG AAT ATC CTC CTT AG 3'; *dkgA*: 5' ACA CCT CAC CGG AGC CTG CTC CGG TGA GTT CAT ATA AAG GAG GAA CGT ATG GTG TAG GCT GGA GCT GCT TC 3' and 5' TGA AAA GTC CGG TAG CGG AAC ATT ACC GCC ACC GGG AGA ATT TGC ATG TTA CAT ATG AAT ATC CTC CTT AG 3'; *ppx*: 5' TGG CGA TTT ATG ACT ACA TCA AAT CAC TCG AAC AAC CTG AAT AAC CCT ATG GTG TAG GCT GGA GCT GCT TC 3' and 5' AAG TGC CTG AAT AAT GCG GGC CGA CAT TTC TCG TCG GCC CGC AAA GTA TTA CAT ATG AAT ATC CTC CTT AG 3'. The presence of the desired alleles was confirmed by PCR amplification of each locus. P1*vir* transduction (Silhavy et al., 1984) was used to move the *ppk*::*kan*$^{+}$ allele from the Keio collection (Baba et al., 2006) into MG1655 and to move the *ppx*::*cat*$^{+}$ allele from MJG308 into BB7224. Chloramphenicol and kanamycin resistance cassettes were resolved where indicated to generate in-frame deletions (Datsenko and Wanner, 2000). The presence of the desired alleles was confirmed by PCR amplification and sequencing.

The *ppk* gene (2,067 bp) plus 20 bp of upstream sequence was amplified from *E. coli* MG1655 genomic DNA with primers 5' ACC GGT ACC AAA ACG GAG TAA AAG TGG TAA TGG G 3' and 5' CTT AAG CTT TTA TTC AGG TTG TTC GAG TGA TTT G 3' and cloned into the *Kpn*I and *Hin*dIII sites of plasmid pBAD33 to yield plasmid pPPK1. The *ppk* gene was subcloned from pPPK1 into the *Kpn*I and *Hin*dIII sites of plasmid pBAD18 to yield plasmid pPPK7.

The *dnaKJ* locus (3,136 bp) plus 19 bp of upstream sequence was amplified from *E. coli* MG1655 genomic DNA with primers 5' ACC GGT ACC ATA TAG TGG AGA CGT TTA GAT GGG 3' and 5' AGA TCT AGA TTA GCG GGT CAG GTC GTC AAA G 3' and cloned into the *Kpn*I and *Xba*I sites of plasmid pBAD18 to yield plasmid pDNAKJ2.

The spectinomycin resistance gene (*aadA*), encoding aminoglycoside 3' adenyltransferase (ANT), was amplified from pBAD43 (Stewart et al., 1999) using primers 5' AAG CTT ATG CGC TCA CGC AAC TGG TCC AGA ACC 3' and 5' AAG CTT ATT ATT TGC CGA CTA CCT TGG TGA TC 3' and cloned into the *Hin*dIII site of plasmid pBR322 (Bolivar et al., 1977) to yield plasmid pBR322-ANT. Strains MJG356 (MG1655 / pBAD33), MJG360 (Δ*ppk* / pBAD33), and MJG361 (Δ*ppx* / pPPK1) were transformed with pBR322-ANT to yield strains AMS346, AMS348, and AMS353, respectively.

In Vivo ATP Measurements

ATP measurements were carried out as described (Yang et al., 2002) with minor modifications. MG1655 wild-type and Δ*ppk* deletion cells were grown in MOPS medium at 37°C to an OD_{600} of approximately 0.5, at which point cells were treated with 1 mM HOCl. At the indicated time points, 100 µl bacterial culture was added to 900 µl boiling 40 mM HEPES (pH 7.8), 4 mM $MgSO_4$ and rapidly shaken for 4 min at 99°C. After boiling, the samples were transferred on ice, and the total ATP content was determined using a luciferase activity assay. For this, 50 µl samples were transferred in triplicate in a 96 well plate format, 150 µl assay buffer (140 µM

luciferin, 0.1 μM luciferase, 0.1 mg ml^{-1} BSA in 100 mM KH_2PO_4 pH 7.8, 25 mM glycylglycine, 0.2 mM EDTA) was added, and bioluminescence was recorded for 2 min.

PolyP Measurements

PolyP granules were visualized by fluorescent microscopy as described (Aschar-Sobbi et al., 2008), with modifications as follows. *E. coli* cultures were prepared as for HOCl survival assays, stained 10 min at room temperature with 50 μg ml^{-1} 4'-6-diamidino-2-phenylindole (DAPI) and 1 μg ml^{-1} N-(3-triethylammoniumpropyl)-4-(6-(4-(diethylamino)phenyl)hexatrienyl)pyridinium dibromide (FM® 4-64) (Molecular Probes), then fixed to glass slides using poly-L-lysine and Citifluor mountant media (Ted Pella, Inc.). Cells were visualized by differential interference contrast (DIC) microscopy, and DNA, polyP granules, and cell membranes were visualized by fluorescence microscopy using the 100X oil immersion objective of an Olympus BX61 upright microscope (Olympus America, Inc.) controlled by the Metamorph Basic software package (v. 7.7.2.0) (Molecular Devices, Inc.). DNA was visualized by DNA-DAPI fluorescence (λ_{ex} = 387 ± 11 nm, λ_{em} = 440 ± 40 nm), polyP granules were visualized by polyP-DAPI fluorescence (λ_{ex} = 420 ± 40 nm, λ_{em} = 535 ± 30 nm) (Aschar-Sobbi et al., 2008), and cell membranes were visualized by FM® 4-64 fluorescence (λ_{ex} = 560 ± 25 nm, λ_{em} 607 ± 34 nm).

Cellular polyP content was also visualized by gel electrophoresis and negative DAPI staining as described (Smith and Morrissey, 2007). Samples were prepared for gel analysis by resuspending 1 mL of *E. coli* culture in 50 μl 50 mM Tris-HCl (pH 8), boiling 10 min, centrifuging 1 min at 16,100 x g, and diluting supernatants 1:1 in 6X DNA loading dye (Promega).

Intracellular polyP levels were quantified as previously described (Ault-Riche et al., 1998), with modifications as follows. A Biomek® FX fluid handling robot (Beckman Coulter) was used to automate polyP extraction, digestion, and measurement. *E. coli* strains were grown at 37°C with aeration in 50 ml MOPS medium (in 300 ml Klett flasks) to an optical density of 40 - 50 Klett units (OD_{600} ~ 0.35), tracking growth with a Klett-Summerson colorimeter. HOCl was added to 1

mM final concentration and 2 ml samples were collected by centrifugation, resuspended in 0.5 ml GITC Lysis Buffer (4 M guanidinium isothiocyanate, 50 mM Tris-HCl [pH 7]), incubated at 95°C for 5 min, then stored at -80°C. Protein concentration was determined by Bradford assay (Bio-Rad). PolyP was precipitated by addition of 0.5 ml 95% ethanol and 30 µl 10% SDS, then bound to 10 µl glassmilk (0.1 g ml^{-1} silicon dioxide in 6 M guanidine-HCl) (Boyle and Lew, 1995). Samples were transferred to AcroPrep™ Advance 96 filter plates (Pall), rinsed with 5 mM Tris-HCl [pH 7.5], 50 mM NaCl, 5 mM EDTA, 50% ethanol, and then dried by vacuum to remove residual ethanol. PolyP was eluted with 50 mM Tris-HCl (pH 8), and degraded to P_i using the yeast exopolyphosphatase ScPPX (Wurst and Kornberg, 1994). P_i concentration was measured using a colorimetric Na_2MoO_4 / malachite green assay (Carter and Karl, 1982). PolyP concentrations were expressed in terms of individual P_i units, normalized to mg of protein in the original sample.

ScPPX Purification

The *PPX1* gene (1,194 bp) was amplified from *S. cereviseae* genomic DNA with primers 5' GTC TAG ACA TAT GTC GCC TTT GAG AAA GAC GG 3' and 5' GAA TTC GGA TCC TCA CTC TTC CAG GTT TGA GT 3' and cloned into the *NdeI* and *Bam*HI sites of plasmid pET-15b to yield the N-terminally His_6-tagged ScPPX expression plasmid pScPPX2. ScPPX was overproduced from pScPPX2 in *E. coli* BL21(DE3) (Novagen). A single colony of the overexpressing strain was resuspended in 5 ml of LB broth, and 1 ml of this suspension was inoculated into 1 liter of LB broth containing ampicillin. Cultures were grown overnight at 37°C without shaking. To induce expression, these cultures were incubated at 37°C with shaking (200 rpm) for 30 min, then IPTG was added to 1 mM, and cultures were incubated for an additional 4 h at 37°C with shaking. Cells were harvested by centrifugation (20 min at 4000 rpm) and resuspended in 50 mM sodium phosphate, 0.5 M NaCl, 5 mM imidazole (pH 7.4). Lysozyme (1 mg ml^{-1}) and Benzonase (50 U ml^{-1}, Merck) were added and cells were incubated 30 min on ice.

Cells were broken by sonication (2 cycles of 2 min at 5 sec on, 5 sec off on ice). Cell lysate was cleared by centrifugation (20 min at 15,000 g at 4°C), filtered through a 0.8 µm syringe filter (Gelman), and applied to a 5 ml nickel-charged HiTrap Chelating HP column (GE Biosciences) at 4 ml min^{-1} with a P1 peristaltic pump. The column was rinsed with 50 ml 50 mM sodium phosphate, 0.5 M NaCl, 5 mM imidazole (pH 8), then with 50 ml 50 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole (pH 8). ScPPX was eluted with 50 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole (pH 8), and fractions containing ScPPX were concentrated to 1 mg ml^{-1}, dialyzed against 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 10% (v/v) glycerol, and stored at 4°C. Specific activity, determined in 20 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 50 mM ammonium acetate, but otherwise as described below for *E. coli* PPX, was 535 ± 105 µmol P_i min^{-1} mg $ScPPX1^{-1}$.

PPK and PPX Purification

PPK was purified as described (Zhu et al., 2003), with minor modifications. The *ppk* gene (2,067 bp) without the stop codon was amplified from *Escherichia coli* MG1655 genomic DNA with primers 5' ATG CAT ATG GGT CAG GAA AAG CTA TAC ATC G 3' and 5' CTT AAG CTT TTC AGG TTG TTC GAG TGA TTT G 3' and cloned into the *Nde*I and *Hind*III sites of plasmid pET-21b to yield the C-terminally His_6-tagged PPK expression plasmid pPPK2. PPK was overproduced from pPPK2 in *E. coli* BL21(DE3) (Novagen); 5 ml of an overnight culture of the overexpressing strain was inoculated into 1 liter of protein expression medium (12 g l^{-1} tryptone, 24 g l^{-1} yeast extract, 4% glycerol [vol / vol], 178 mM potassium phosphate [pH 7]) and incubated at 37°C with shaking to OD_{600} = 1. The culture was cooled to 20°C, then induced with 150 µM IPTG and incubated overnight at 20°C with shaking. Cells were harvested by centrifugation (10 min @ 6,000 g @ 4°C), resuspended in 50 mM Tris-HCl (pH 7.5), 10% sucrose (w / vol), 300 µg ml^{-1} lysozyme, then incubated 45 min on ice followed by 10 min at 37°C. Insoluble material was removed by centrifugation (10 min @ 16,100 g @ 4°C), the

supernatant was removed, and the pellet was resuspended by sonication in 50 mM Tris-HCl (pH 7.5), 10% sucrose (w / vol), 5 mM $MgCl_2$, 30 U ml^{-1} Benzonase (Merck) with Complete Mini, EDTA-free protease inhibitor cocktail (Roche). PPK was extracted from membranes by adding solid KCl to 1 M, then adding 1/10 volume 1 M Na_2CO_3 and stirring 30 min at 4°C. The resulting solution was sonicated for 2 min (5 sec on, 5 sec off) and cell debris was removed by centrifugation (1 h @ 20,000 g @ 4°C). The supernatant was immediately diluted 1:1 with distilled water, then applied to a 5 ml nickel-charged HiTrap Chelating HP column (GE Biosciences) equilibrated with 50 mM HEPES, pH 7.5, 0.5 M NaCl, 15% glycerol, 50 mM imidazole. PPK was eluted with a 17 column-volume linear gradient to 0.5 M imidazole, using an ÄKTA FPLC (Amersham Pharmacia). Fractions containing pure PPK were pooled and dialyzed overnight against 20 mM HEPES, pH 8, 150 mM NaCl, 15% glycerol, 1 mM EDTA, 5 mM DTT and stored at -80°C.

The *E. coli* PPX overexpression vector pEcPPX was a gift from David Sanders (Purdue University) (Alvarado et al., 2006), and PPX was purified by anion exchange and hydroxyapatite affinity chromatography. PPX was overproduced from pEcPPX in *E. coli* BL21(DE3) (Novagen); a single colony of the overexpressing strain was resuspended in 5 ml of LB broth, and 1 ml of this suspension was inoculated into 1 liter of LB broth containing kanamycin. Cultures were grown overnight at 37°C without shaking. To induce expression, these cultures were incubated at 37°C with shaking (200 rpm) for 30 min, then IPTG was added to 1 mM, and cultures were incubated for an additional 2.5 h at 37°C with shaking. Cells were harvested by centrifugation (20 min @ 4,000 rpm), resuspended in 10 ml HEPES buffer (50 mM; pH 8) containing 50 mM NaCl, 1 mM EDTA, 2 mM DTT, and 10% (w/v) glycerol, then lysed by two passages through a French press (19,000 psi). The lysate was clarified by centrifugation (1 h @ 20,000 g @ 4°C) and passage through a 0.8 µm syringe filter, then applied to a 5 ml HiTrap Q anion exchange column (GE Biosciences). Proteins were eluted with a 30 column-volume linear gradient to 1 M NaCl using an ÄKTA fast pressure liquid chromatography (FPLC) system (Amersham

Pharmacia), and fractions containing PPX were pooled and concentrated to 5 ml using an Amicon Ultra 30-K MWCO centrifugal concentrator. They were then dialyzed against 50 mM HEPES (pH 8), 50 mM NaCl, 10% glycerol, 1 mM DTT, 5 mM K_2HPO_4 and applied to a 69 ml hydroxyapatite column (Bio-Rad). PPX was eluted with a 3 column-volume linear gradient to 400 mM K_2HPO_4, and fractions containing pure PPX were pooled, concentrated, dialyzed against 50 mM HEPES / KOH (pH 8), 50 mM NaCl, 10% (w/v) glycerol, 1 mM EDTA, 1 mM DTT, and stored at -80°C.

PPK and PPX Activity Assays

PPK and PPX activity were assayed as described (Akiyama et al., 1993; Tzeng and Kornberg, 2000), with minor modifications as follows. Oxidized PPK samples were prepared by incubating PPK for 15 min at 37°C in 20 mM sodium phosphate (pH 8), 150 mM NaCl, 15% (v/v) glycerol with or without the indicated molar ratios of *N*-chlorotaurine. Oxidants were removed with P-30 gel chromatography columns (Bio-Rad). Forward (polyP-generating) reaction mixtures contained 50 mM HEPES (pH 7.5), 50 mM ammonium sulfate, 5 mM $MgCl_2$, and 10 nM PPK. After equilibration (1 min) at 37°C, reactions were started by addition of ATP to 10 mM, and aliquots were taken at 1 min intervals, stopping the reaction by mixing with 4 volumes of GITC Lysis Buffer. PolyP content of reaction aliquots was assayed as described above for *in vivo* polyP measurements, and specific activities were calculated as µmol polyP formed (in individual P_i units) min^{-1} mg PPK^{-1}. Reverse (ATP-generating) reaction mixtures contained 50 mM HEPES (pH 7.5), 50 mM ammonium sulfate, 2 mM MgADP, and 50 nM PPK. After equilibration (1 min) at 37°C, reactions were started by addition of polyP to 500 µM, and aliquots were taken at 30 sec intervals, stopping the reaction by incubating 5 min at 99°C. The ATP content of PPK reaction aliquots was measured using luciferase (Lundin, 2000) and specific activities were calculated as µmol ATP formed min^{-1} mg PPK^{-1}.

Oxidized PPX samples were prepared by incubating PPX for 15 min at 37°C in 50 mM potassium phosphate (pH 8), 175 mM KCl containing the indicated molar ratios of *N*-chlorotaurine or H_2O_2. Oxidants were removed with P-30 gel chromatography columns (Bio-Rad). Reaction mixtures contained 50 mM HEPES (pH 8), 175 mM KCl, 1 mM $MgCl_2$, and 1 nM PPX. After equilibration (1 min) at 37°C, reactions were started by addition of polyP to 200 μM, and aliquots were taken at 1 min intervals, stopping the reaction by mixture with 0.7 volume of Buffer A (4 parts 2 N HCl, 3 parts 0.1 M Na_2MoO_4). Free phosphate was measured with the Na_2MoO_4 / malachite green colorimetric assay (Carter and Karl, 1982), and specific activities were calculated as μmol P_i released min^{-1} mg PPX^{-1}.

Mass Spectrometry for Cysteine Thiol Status.

PPX, with or without *N*-chlorotaurine oxidation (as above), was prepared for mass spectrometry by differential thiol trapping. PPX samples (20 μg) were precipitated with trichloroacetic acid (TCA), then incubated 1 hr at room temperature in pH 8.5 denaturing alkylating buffer (DAB) (200 mM Tris-HCl, 6 M urea, 10 mM EDTA, 0.5% sodium dodecylsulfate [SDS]) containing 100 mM iodoacetamide (IAM) to irreversibly alkylate reduced cysteine residues. After reduction in DAB (pH 8.5) containing 100 mM DTT, samples were further modified by incubation for 1 hr in DAB (pH 7) containing 100 mM *N*-ethylmaleimide to irreversibly modify any cysteine which had been reversibly oxidized in the original sample. Cysteine modification status was determined by LC-MS/MS of tryptically-digested samples (MS Bioworks, Ann Arbor, MI).

***In Vivo* Thiol Trapping of PPX.**

E. coli BL21 containing plasmid pEcPPX was grown at 37°C in MOPS medium to OD_{600} = 0.4 - 0.5. PPX expression was induced with 50 μM IPTG for 30 min, then cells were treated with 1 mM HOCl. Before and after treatment, 2 ml aliquots were lysed in the presence of 10% trichloroacetic acid (TCA) to prevent further thiol oxidation. After 30 min incubation on ice,

precipitated proteins were pelleted by centrifugation (13,000 rpm, 20 min, 4°C). The protein pellet was resuspended in denaturing alkylating buffer (DAB, 6 M Urea, 200 mM Tris-HCl [pH 8.5], 10 mM EDTA, 0.5% w/v SDS) supplemented with 100 mM IAM to irreversibly alkylate all reduced cysteines and incubated 1 hr at 25°C with shaking. Samples were again precipitated with TCA to remove unbound IAM and pelleted by centrifugation. For differential thiol trapping with PEG-maleimide, protein pellets were resuspended in DAB buffer supplemented with 10 mM DTT and incubated for 1 h at 25°C to reduce oxidized cysteines. Excess DTT was removed by TCA precipitation and centrifugation. All newly accessible cysteines were then modified with 10 mM of the thiol-specific alkylation reagent PEG-maleimide, which adds 2 kDa to every modified cysteine. Samples were incubated for 1 hr at 25°C, again precipitated with TCA to remove unbound PEG-maleimide, and then suspended in reducing loading buffer. As controls, purified PPX protein samples were reduced with DTT, then alkylated with either IAM or PEG-maleimide, as described above. All samples were separated by SDS-PAGE and visualized by western blot using polyclonal anti-PPX antibodies (Pacific Immunology).

Crude Cell Lysate Preparation

E. coli strains were incubated overnight in 30 mL LB at 37°C with shaking (200 rpm), then subcultured into 2 L LB and grown at 37°C with shaking to OD_{600}= 0.6-0.7. Cells were harvested by centrifugation and resuspended in 25 mL ice-cold 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. Cells were then spun down again (10 min @ 8,000 rpm @ 4°C) and resuspended in 10 mL ice-cold buffer. Phenylmethanesulfonylfluoride (PMSF) was added to 1 mM and the cells were broken with a French press (2 passages at 20,000 psi). Cell debris was spun out (5 min @ 3,000 g) and aliquots of the supernatant were stored at -80°C. Before use, lysates were thawed on ice, spun down (5 min @ 16,100 g), and exchanged into 10 mM potassium phosphate (pH 7.5) using 7K MWCO Zeba Spin Desalting Columns (Thermo Scientific). Lysates were spun again (20 min @ 16,100 g @ 4°C) to remove any remaining insoluble material before use.

Table of Strains and Plasmids.

Strain name	Marker(s)[a]	Relevant Genotype	Source[b]
E. coli strains:			
XL1-Blue	Tc^R Nx^R	*endA1 gyrA96*(Nx^R) *thi-1 recA1 relA1 lac glnV44* F'[::Tn10(*tet*$^+$) *proAB*$^+$ *lacI*$_q$ Δ(*lacZ*)M15] *hsdR17*(r_K^- m_K^+)	Stratagene
BL21(DE3)		F^-, *ompT gal dcm lon hsdSB*(r_B^- m_B^-) λ(DE3 [*lacI lacUV5*-T7 gene 1 *ind1 sam7 nin5*])	Novagen
BB7224	Sm^R Km^R	F^-, λ^-, $e14^-$, [*araD139*]$_{B/r}$ Δ(*argF-lac*)*169 flhD5301* Δ(*fruK-yeiR*)*725*(*fruA25*) *relA1 rpsL150*(Sm^R) *rbsR22* Δ(*fimB-fimE*)*632*(::IS1) *ptsF25 zhf*::Tn*10*(Tc^S) *suhX401 deoC1 araD*$^+$ *rpoH*::*kan*$^+$	(Tomoyasu et al., 2001)
MG1655		F^-, λ^-, *rph-1 ilvG*$^-$ *rfb-50*	(Blattner et al., 1997)
MJG154	Cm^R	MG1655 *mgsA*::*cat*$^+$	
MJG155	Cm^R	MG1655 *yqhD*::*cat*$^+$	
MJG176		MG1655 Δ*mgsA*	
MJG177		MG1655 Δ*yqhD*	
MJG184	Cm^R	MG1655 *dkgA*::*cat*$^+$	
MJG187		MG1655 Δ*dkgA*	
MJG208	Km^R	MG1655 *ppk*::*kan*$^+$	
MJG224		MG1655 Δ*ppk*	
MJG308	Cm^R	MG1655 *ppx*::*cat*$^+$	
MJG315		MG1655 Δ*ppx*	
MJG356	Cm^R	MG1655 / pBAD33 (*cat*$^+$)	
MJG360	Cm^R	MG1655 Δ*ppk* / pBAD33 (*cat*$^+$)	
MJG361	Cm^R	MG1655 Δ*ppx* / pPPK1 (*ppk*$^+$ *cat*$^+$)	
MJG491	Sm^R Km^R Ap^R	BB7224 / pBAD18 (*bla*$^+$)	
MJG492	Sm^R Km^R Ap^R	BB7224 / pPPK7 (*ppk*$^+$ *bla*$^+$)	
MJG513	Sm^R Km^R	BB7224 *ppx*::*cat*$^+$	

	Cm^R		
MJG514	Sm^R Km^R Cm^R Ap^R	BB7224 *ppx*::*cat*$^+$ / pBAD18 (*bla*$^+$)	
MJG515	Sm^R Km^R Cm^R Ap^R	BB7224 *ppx*::*cat*$^+$ / pPPK7 (*ppk*$^+$ *bla*$^+$)	
MJG531	Sm^R Km^R Ap^R	BB7224 / pDNAKJ2 (*dnaKJ*$^+$ *bla*$^+$)	
AMS346	Cm^R Sp^R Tc^R Ap^R	MG1655 / pBAD33 (*cat*$^+$) / pBR322-ANT (*aadA*$^+$ *tet*$^+$ *bla*$^+$)	
AMS348	Cm^R Sp^R Tc^R Ap^R	MG1655 Δ*ppk* / pBAD33 (*cat*$^+$) / pBR322-ANT (*aadA*$^+$ *tet*$^+$ *bla*$^+$)	
AMS353	Cm^R Sp^R Tc^R Ap^R	MG1655 Δ*ppx* / pPPK1 (*ppk*$^+$ *cat*$^+$) / pBR322-ANT (*aadA*$^+$ *tet*$^+$ *bla*$^+$)	
V. cholerae El Tor strains:			
C6706*lacZ*	Sm^R		(Cameron et al., 2008)
EC2798	Sm^R Km^R	*mgsA*-TnFGL3::*kan*$^+$	(Cameron et al., 2008)
EC21753	Sm^R Km^R	*ppk1*-TnFGL3::*kan*$^+$	(Cameron et al., 2008)
Plasmids:			
pET-15b	Ap^R	N-terminal His_6 tag overexpression vector	Novagen
pET-21b(+)	Ap^R	C-terminal His_6 tag overexpression vector	Novagen
pKD3	Cm^R Ap^R	*cat*$^+$ chloramphenicol resistance cassette donor	(Datsenko and Wanner, 2000)
pKD46	Ap^R	λ Red recombinase$^+$	(Datsenko and Wanner, 2000)
pCP20	Cm^R Ap^R	Flp recombinase$^+$	(Datsenko and Wanner, 2000)
pBAD18	Ap^R	cloning vector with P_{BAD} arabinose-inducible promoter	(Guzman et al., 1995)

pBAD33	Cm^R	cloning vector with P_{BAD} arabinose-inducible promoter	(Guzman et al., 1995)
pBAD43	Sp^R Ap^R	cloning vector with P_{BAD} arabinose-inducible promoter	(Stewart et al., 1999)
pBR322	Tc^R Ap^R	cloning vector	(Bolivar et al., 1977)
pBR322-ANT	Sp^R Tc^R Ap^R	*aadA*$^+$	
pPPK1	Cm^R	*ppk*$^+$	
pPPK2	Ap^R	*ppk*$^+$ translational fusion to C-terminal His_6 tag for protein purification	
pPPK7	Ap^R	*ppk*$^+$	
pDNAKJ2	Ap^R	*dnaKJ*$^+$	
pScPPX2	Ap^R	*S. cerevisiae PPX1*$^+$ translational fusion to N-terminal His_6 tag for protein purification	
pEcPPX	Km^R	*ppx*$^+$	(Alvarado et al., 2006)

[a] Abbreviations: Tc^R, tetracycline resistance; Nx^R, nalidixic acid resistance; Cm^R, chloramphenicol resistance; Ap^R, ampicillin resistance; Km^R, kanamycin resistance; Sp^R, spectinomycin resistance; Sm^R, streptomycin resistance; Tc^S, tetracycline sensitivity.

[b] Unless otherwise indicated, all strains and plasmids were generated in the course of this work.

SUPPLEMENTAL REFERENCES:

Akiyama, M., Crooke, E., and Kornberg, A. (1993). An exopolyphosphatase of Escherichia coli. The enzyme and its ppx gene in a polyphosphate operon. J Biol Chem *268*, 633-639.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.*, et al.* (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453-1462.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene *2*, 95-113.

Boyle, J.S., and Lew, A.M. (1995). An inexpensive alternative to glassmilk for DNA purification. Trends Genet *11*, 8.

Cameron, D.E., Urbach, J.M., and Mekalanos, J.J. (2008). A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proc Natl Acad Sci U S A *105*, 8736-8741.

Carter, S.G., and Karl, D.W. (1982). Inorganic phosphate assay with malachite green: an improvement and evaluation. J Biochem Biophys Methods *7*, 7-13.

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

Lundin, A. (2000). Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. Method Enzymol *305*, 346-370.

Peskin, A.V., and Winterbourn, C.C. (2001). Kinetics of the reactions of hypochlorous acid and amino acid chloramines with thiols, methionine, and ascorbate. Free Radic Biol Med *30*, 572-579.

Silhavy, T.J., Berman, M.L., and Enquist, L.W., eds. (1984). Experiments with gene fusions (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Smith, S.A., and Morrissey, J.H. (2007). Sensitive fluorescence detection of polyphosphate in polyacrylamide gels using 4',6-diamidino-2-phenylindol. Electrophoresis *28*, 3461-3465.

Stewart, E.J., Katzen, F., and Beckwith, J. (1999). Six conserved cysteines of the membrane protein DsbD are required for the transfer of electrons from the cytoplasm to the periplasm of Escherichia coli. EMBO J *18*, 5963-5971.

Tzeng, C.M., and Kornberg, A. (2000). The multiple activities of polyphosphate kinase of Escherichia coli and their subunit structure determined by radiation target analysis. J Biol Chem *275*, 3977-3983.

Yang, N.C., Ho, W.M., Chen, Y.H., and Hu, M.L. (2002). A convenient one-step extraction of cellular ATP using boiling water for the luciferin-luciferase assay of ATP. Anal Biochem *306*, 323-327.

Zhu, Y., Lee, S.S., and Xu, W. (2003). Crystallization and characterization of polyphosphate kinase from Escherichia coli. Biochem Biophys Res Commun *305*, 997-1001.