Molecular Cell
Article

Optimizing Protein Stability In Vivo

Linda Foit,[1,6,7] Gareth J. Morgan,[3,4,7] Maximilian J. Kern,[1] Lenz R. Steimer,[1] Annekathrin A. von Hacht,[1] James Titchmarsh,[3,5] Stuart L. Warriner,[3,5] Sheena E. Radford,[3,4,*] and James C.A. Bardwell[1,2,*]
[1]Howard Hughes Medical Institute
[2]Department of Molecular, Cellular, and Developmental Biology
University of Michigan, Ann Arbor, MI 48109, USA
[3]Astbury Centre for Structural and Molecular Biology
[4]Institute for Molecular and Cellular Biology
[5]School of Chemistry
University of Leeds, Leeds LS2 9JT, UK
[6]Institute for Chemistry and Pharmacy, University of Münster, Münster 48149, Germany
[7]These authors contributed equally to this work
*Correspondence: s.e.radford@leeds.ac.uk (S.E.R.), jbardwel@umich.edu (J.C.A.B.)
DOI 10.1016/j.molcel.2009.11.022

SUMMARY

Identifying mutations that stabilize proteins is challenging because most substitutions are destabilizing. In addition to being of immense practical utility, the ability to evolve protein stability in vivo may indicate how evolution has formed today's protein sequences. Here we describe a genetic selection that directly links the in vivo stability of proteins to antibiotic resistance. It allows the identification of stabilizing mutations within proteins. The large majority of mutants selected for improved antibiotic resistance are stabilized both thermodynamically and kinetically, indicating that similar principles govern stability in vivo and in vitro. The approach requires no prior structural or functional knowledge and allows selection for stability without a need to maintain function. Mutations that enhance thermodynamic stability of the protein Im7 map overwhelmingly to surface residues involved in binding to colicin E7, showing how the evolutionary pressures that drive Im7-E7 complex formation have compromised the stability of the isolated Im7 protein.

INTRODUCTION

The factors that govern protein folding and stability have been well studied, but how evolution has generated sequences with the appropriate balance of functional, kinetic, and thermodynamic properties remains unresolved. Key to addressing this question is the need to understand why the protein products of evolution are usually only marginally stable and how different evolutionary pressures, such as functional requirements, the need to maintain solubility, or the need to fold rapidly, have been balanced to result in the current repertoire of amino acid sequences (Espargaro et al., 2008; Wurth et al., 2002; Gosavi et al., 2008). Since most random mutations destabilize folded proteins (DePristo et al., 2005), identifying stabilizing mutations is challenging but of great practical and intellectual value. It is generally accepted that proteins require a stability that exceeds a minimal threshold in order to fold and function (Bloom et al., 2007; DePristo et al., 2005). The thermodynamic stabilities of most water-soluble globular proteins are low, generally between 12 and 42 kJ/mol (DePristo et al., 2005). Despite a wealth of experiments to address the question of why proteins are only marginally stable, there is currently no consensus (Sanchez et al., 2006). Understanding why proteins are marginally stable is also important from a practical viewpoint, since this metastability makes proteins difficult to handle experimentally (Mehlin et al., 2006). The vast array of protein structure and sequence data now available has fueled theoretical and experimental studies to enhance protein stability (Eijsink et al., 2005; Roodveldt et al., 2005; Wunderlich et al., 2005). Although these studies have resulted in a number of helpful designs and algorithms that have had some success in predicting the effect of amino acid substitutions on protein stability, our understanding of the factors that govern protein stability remains far from complete. Furthermore, the question of how protein stability in the crowded cellular environment relates to measurements of dilute, pure proteins in vitro remains unresolved, despite recent innovative experiments to address this question (Ghaemmaghami and Oas, 2001; Ignatova and Gierasch, 2004; Mayer et al., 2007).

Here we present a powerful new method of evolving protein stability in vivo using a genetic selection that directly links the stability of a protein to antibiotic resistance and does not require any knowledge of the protein's structure or function. Uniquely, it offers a new route for improving our understanding of the fundamental basis of protein stability and expression in a cellular context. Unlike existing screening techniques that require laborious testing of individual variants (Magliery and Regan, 2004; Roodveldt et al., 2005), our approach enables direct selection for stable variants of a given test protein simply by demanding bacterial growth on an antibiotic. We demonstrate that both thermodynamic and kinetic stability are critical determinants of the expression and maintenance of proteins in vivo. Using the protein Im7, we find that stabilizing mutations predominately map to the surface used to bind its cognate toxin colicin E7, indicating how the need to maintain function has resulted in a

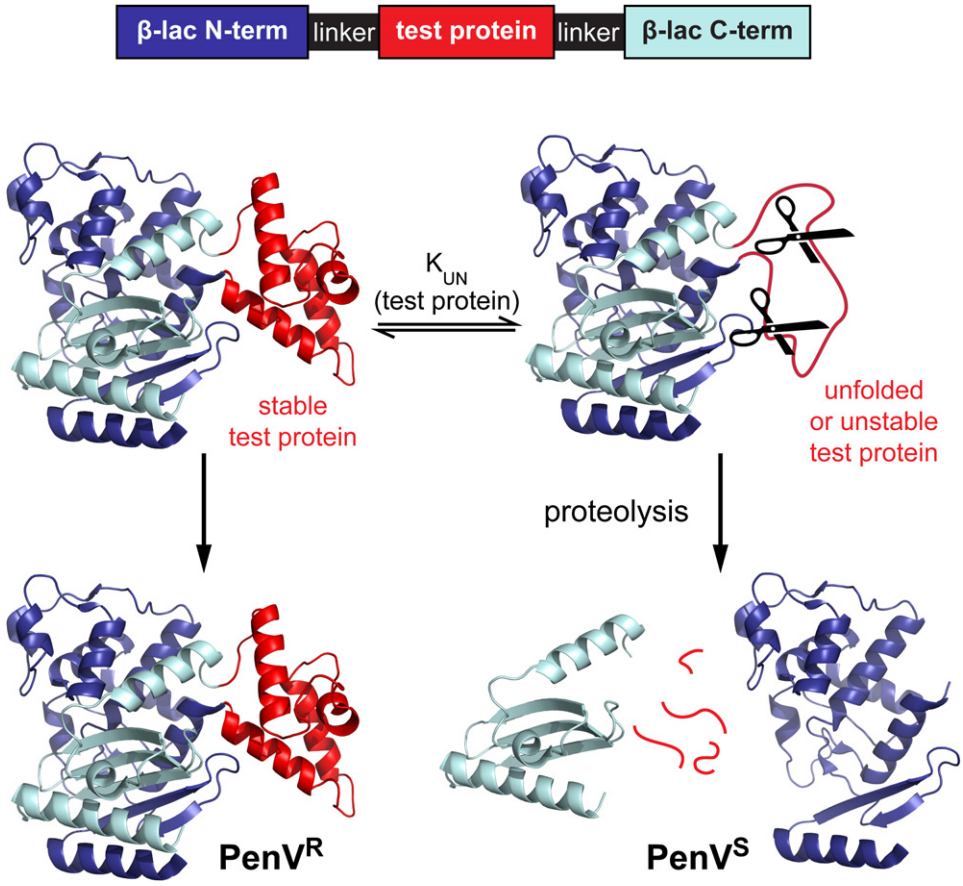


Figure 1. Schematic Diagram of the Tripartite System for the Evolution of Improved Protein Stability

The test protein (red) was inserted into TEM1-β-lactamase as part of a tripartite fusion. Poor folding and/or low stability of the test protein should result in an increased sensitivity of the tripartite fusion to cellular proteases and in turn to low levels of resistance to β-lactam antibiotics such as penicillin V. In the case of proper folding and/or increased stability of the test protein, the fusion protein should remain intact and confer high levels of antibiotic resistance. The level of in vivo penicillin resistance may thus be related to the folding/unfolding equilibrium constant of the inserted protein, designated K_{UN}. $PenV^S$ = penicillin V sensitive; $PenV^R$ = penicillin V resistant; linker = glycine-serine rich linkers of the following lengths: 27 residues for G-CSF, 33 residues for Im7 and cyt b_{562}, and 68 residues for MBP.

decrease in stability of the isolated Im7 protein that can be reversed upon mutation.

RESULTS

A Tripartite Fusion System for Linking Protein Stability to Antibiotic Resistance

The design principle for a system able to monitor the stability of proteins in vivo and to obtain stabilized protein variants by selection for antibiotic resistance without consideration of function is shown in Figure 1. The design is based on a tripartite fusion whereby the test protein is inserted into the sequence of an antibiotic resistance protein, TEM1-β-lactamase. This enzyme is known to tolerate insertions, deletions, and substitutions in the surface-exposed loop between residues 196 and 199 (Galarneau et al., 2002). Dissection of TEM1-β-lactamase between residues 196 and 197 results in two fragments that are catalytically inactive when separated but, if fused to proteins that interact, can reassemble to form the active enzyme (Galarneau et al., 2002). The idea behind our tripartite fusion approach is that upon the correct folding of the test protein, the two halves of β-lactamase will be brought close enough together to associate and thereby confer the expected resistance to β-lactam antibiotics. Unstable proteins, which are targeted for degradation by the cell's protein quality control machinery, should separate the two halves of β-lactamase and thereby result in reduced resistance of host cells to β-lactam antibiotics. We reasoned that β-lactamase activity—and therefore the resistance of the host cell to β-lactam antibiotics—will be directly linked to the in vivo stability of the test protein. Our approach should not only provide a convenient readout of the ability of the inserted protein to fold into a stable state in vivo but, by selecting for increased antibiotic resistance, should also provide a powerful route to evolve stabilizing mutations in proteins. TEM1-β-lactamase was chosen because it is a sensitive, robust, and selectable reporter that can be used in both prokaryotic and eukaryotic cells and is ideally suited for ultraminiaturization and high-throughput studies (Qureshi, 2007).

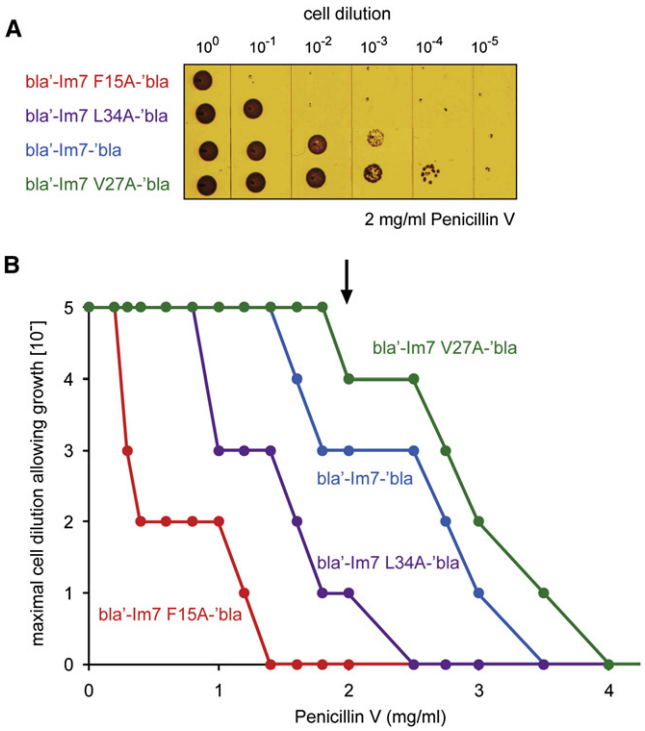


Figure 2. Protein Stability Correlates with Antibiotic Resistance

Mid-log phase cells of *E. coli* NEB10-β expressing TEM1-β-lactamase fused with wild-type Im7, the destabilized variants Im7 L34A or F15A, or the stabilized variant Im7 V27A were normalized to $A_{600\ nm}$ = 1. Serial dilutions of cultures from 10^0 to 10^{-5} were spotted on LB plates containing different concentrations of penicillin V. After 18 hr incubation at 37°C, growth or no growth for each dilution at each penicillin V concentration was scored and used to calculate MIC values as described in Figure S1.

(A) Spot titer from a single LB plate supplemented with 2 mg/ml penicillin V.

(B) Maximal cell dilution that allowed growth plotted against penicillin V concentration. The arrow indicates the penicillin V concentration shown in (A) (2 mg/ml). The published ΔG°_{UN} values for wild-type Im7, F15A, V27A, and L34A are −24.9, −9.7, −27.0, and −17.2 kJ/mol, respectively (Capaldi et al., 2002), which correspond to K_{UN} (the ratio of folded to unfolded protein) values of 40,000, 62, 96,000, and 1,500, respectively.

Antibiotic Resistance Correlates with Thermodynamic Stability for a Variety of Proteins

To determine whether the tripartite fusion system described above allows antibiotic resistance to be used to screen for protein stability, four different proteins were inserted into β-lactamase, along with 31 variants of these proteins that contain mutations known to alter their thermodynamic stability. Immunity protein 7 (Im7) (Capaldi et al., 2002), cytochrome b_{562} (cyt b_{562}) (Kamiya et al., 2001), human granulocyte colony-stimulating factor (G-CSF) (Bishop et al., 2001), and maltose-binding protein (MBP) (Chun et al., 1993) were chosen. Thus, our test set included proteins that vary in terms of size, sequence, structural characteristics, source, and presence of cofactors or disulfide bonds, allowing stringent testing of the general utility of the tripartite fusion system.

We introduced long, flexible glycine/serine-rich linkers between residues 196 and 197 of TEM1-β-lactamase. The genes encoding the four test proteins were then cloned into the linker-encoding region to generate the desired tripartite fusions. Mutations known to alter the stability of the test proteins were introduced into these clones via site-directed mutagenesis. The antibiotic resistance of strains expressing each tripartite fusion was measured by spotting dilutions of cells onto plates containing increasing concentrations of the β-lactam antibiotic penicillin V. Figure 2A shows one such experiment: using plates containing 2 mg/ml penicillin V, we compared the growth of cells expressing the tripartite fusion containing wild-type Im7 (bla′-Im7-′bla) with that of two destabilized variants, F15A and L34A (bla′-F15A-′bla and bla′-L34A-′bla, respectively) and one stabilized variant, V27A (bla′-V27A-′bla) (Capaldi et al., 2002). Insertion of the destabilized variants resulted in clones with a significantly lower resistance toward penicillin V over a wide range of antibiotic concentrations, whereas insertion of the stabilized mutant resulted in higher resistance to penicillin V also over a wide antibiotic concentration range (Figure 2B). To obtain a single parameter that quantitatively describes the antibiotic sensitivity of each construct, the titration data were used to calculate the minimal inhibitory concentration (MIC) of antibiotic that inhibited cell growth (see the Experimental Procedures and see Figure S1 available online).

How MIC values relate to the in vitro thermodynamic stability of the inserted protein was assessed by determining the MIC values of cells expressing the tripartite fusions of the four test proteins (Im7, MBP, G-CSF, and cyt b_{562}) and their variants with published differences in thermodynamic stability (Bishop et al., 2001; Capaldi et al., 2002; Chun et al., 1993). We found a striking and significant correlation between thermodynamic stability of the inserted protein and MIC for all four test protein systems (Figure 3 and Figure S1). For Im7, all but one of the mutant proteins known to destabilize the protein in vitro conferred lower antibiotic resistance in vivo. The Im7 variants F41A and F84A, which are so destabilized that they mainly populate the unfolded state in vitro (precluding determination of their ΔG°_{UN} values) (Capaldi et al., 2002), showed dramatically reduced MIC values relative to the wild-type protein ($\ln[MIC_{mut}/MIC_{wt}] = -1.9$ and −1.1, respectively). Despite the presence of disulfide bonds in G-CSF, a heme cofactor in cyt b_{562}, and the large size of MBP, a significant correlation between $\Delta\Delta G^{\circ}_{UN}$ and MIC was also observed for these proteins (Figures 3B and 3C and Figure S1). The data indicate that the approach developed provides a robust and generic in vivo readout of protein stability for proteins with different sizes and structural properties, including both stabilizing and destabilizing mutations.

Selection for Im7 Mutants with Higher Levels of Antibiotic Resistance Yields Variants with Increased Thermodynamic Stability, Expression Levels, and Protease Resistance

Linking in vivo stability to antibiotic resistance provides a method for selecting stabilizing variants of proteins without requiring knowledge of the protein's structure or function. To explore this possibility, the Im7 portion of the fusion construct bla′-Im7-′bla was subjected to error-prone PCR. The resulting library of mutants was transformed into *E. coli* and plated onto media containing different concentrations of penicillin V. Colonies that displayed greater antibiotic resistance compared with cells

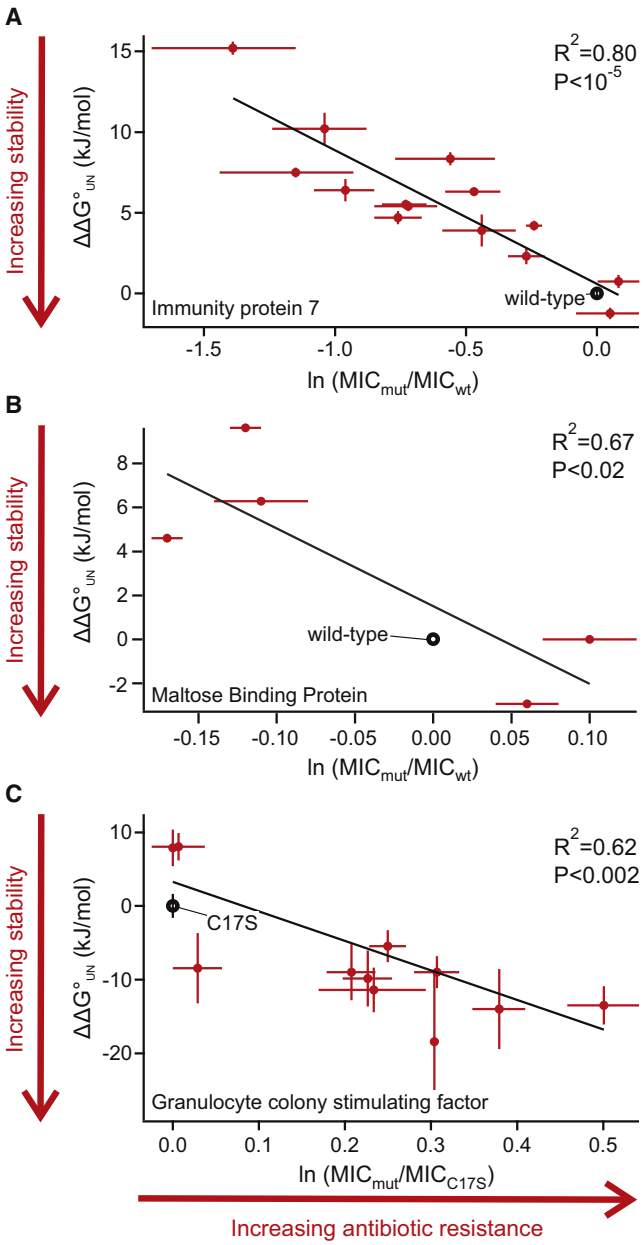


Figure 3. Antibiotic Resistance Correlates with Stability for a Variety of Proteins

Thermodynamically destabilized or stabilized mutants of Im7 (A), MBP (B), or G-CSF (C) 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 Figure S1. The average MIC for penicillin V, relative to the wild-type protein, is plotted against $\Delta\Delta G^{\circ}_{UN}$, where $\Delta\Delta G^{\circ}_{UN} = \Delta G^{\circ}_{UN}$ (mutant) – ΔG°_{UN} (wild-type). In the case of G-CSF, the pseudo-wild-type protein C17S was used because wild-type G-CSF has a high tendency to aggregate (Bishop et al., 2001). Error bars represent the standard deviation of the mean of the MIC or the standard deviation of the mean of the $\Delta\Delta G^{\circ}_{UN}$ measurements for Im7. Error bars for the $\Delta\Delta G^{\circ}_{UN}$ of G-CSF were taken from Bishop et al. (2001).

expressing bla′-Im7-′bla were selected, and their MIC values were determined by spot titration. In total, 17 single, 9 double, and 5 triple mutants were identified that resulted in MIC values up to 3-fold greater than that for wild-type Im7 (Tables S1 and S2).

To investigate whether the increased antibiotic resistance of the selected fusion constructs results in an increase in the in vitro thermodynamic stability of the mutant Im7 proteins, each variant was expressed and purified in the absence of the β-lactamase fusion, and its thermodynamic stability was determined by urea titration (see the Experimental Procedures and Supplemental Data). Of the 31 selected mutants, 26 showed an increased thermodynamic stability compared with wild-type Im7 (Figures 4A and 5A; Table S2), providing compelling evidence that antibiotic resistance provides a sensitive, quantitative measure of thermodynamic stability in vivo. Combining the data for the selected mutants with those for the destabilized Im7 variants (Figure 3A) resulted in a striking linear correlation between stability and antibiotic resistance that spanned over 20 kJ/mol in $\Delta\Delta G^{\circ}_{UN}$ and a 7-fold difference in MIC values (Figure 4A).

The in vivo stability and steady-state expression level of a protein is not just determined by the protein's thermodynamic stability but by its rate of synthesis or degradation by cellular proteases, as well as by its solubility or, for secreted proteins, efficiency of translocation. Often these properties are linked to each other: susceptibility to proteolysis, for instance, depends on the protein visiting a protease-labile conformation through global or local unfolding processes. Global unfolding is related to a protein's thermodynamic stability, whereas local unfolding depends on transient dynamics. Numerous findings indicate that the thermodynamic stability of a protein is an important determinant of its proteolytic susceptibility and therefore strongly influences its rate of degradation (Parsell and Sauer, 1989; Sieber et al., 1998; Park and Marqusee, 2005). Indeed, increased thermodynamic stability for Im7 variants was shown to correlate with a decrease in sensitivity toward proteinase K digestion in vitro (Figure S2). The increased thermodynamic stability of the selected Im7 variants also correlates with an increased steady-state expression level in vivo, as measured by quantitative western blots of lysates obtained from cells expressing the bla′-Im7-′bla constructs, analyzed using an anti-β-lactamase antibody. All fusion proteins were detected only in the soluble, periplasmic fractions, with no detectable material in the cytosol or insoluble fractions, indicating that neither the solubility nor the translocation efficiency of the tripartite fusions containing the different Im7 mutants was altered detectably with respect to wild-type Im7. Changes in expression level correlate well with levels of antibiotic resistance (Figure 4B) and with β-lactamase activity of cell extracts (Figure S2). This suggests that the level of resistance is determined by the relative amount of the fusion protein, rather than changes in the specific activity of β-lactamase in the different tripartite fusions. Selection for increased antibiotic resistance thus results in new protein sequences that display increases in thermodynamic stability and steady-state expression level. These results demonstrate that the tripartite fusion system can be used to improve the expression level of a chosen protein simply by selecting for

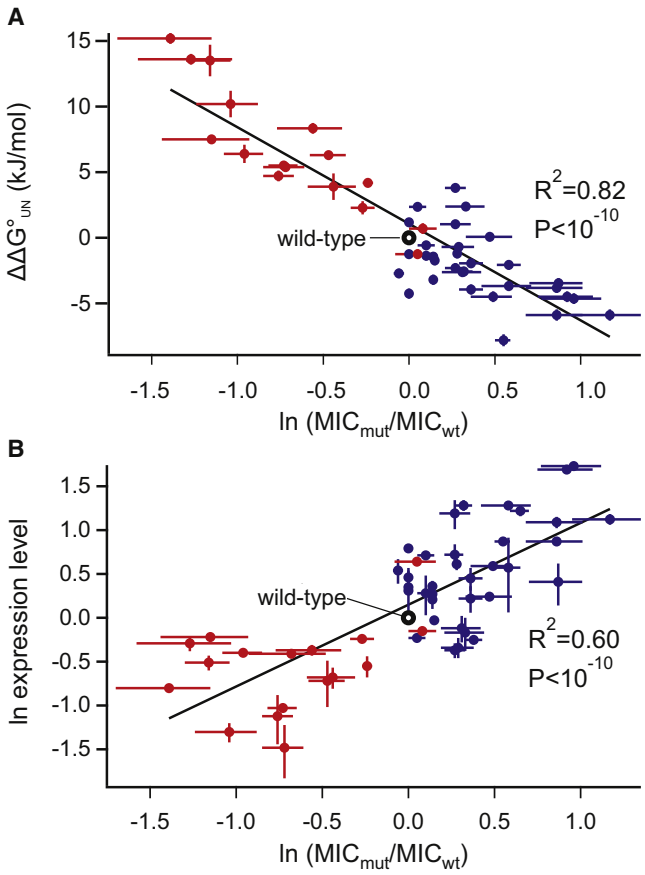


Figure 4. Antibiotic Resistance Correlates with Stability and Expression Levels of the Variant Proteins

Data points include published destabilized mutants (red) (Capaldi et al., 2002) and mutants selected on the basis of their increased level of antibiotic resistance (blue); see also Figure S2 and Table S1. (A) Thermodynamic stabilities of Im7 variants selected on the basis of an increased MIC. Stabilities were measured by denaturant titration (see the Experimental Procedures and the Supplemental Data). (B) Expression levels of tripartite fusion proteins that contained thermodynamically destabilized or stabilized mutants of Im7. Fusion proteins were detected by western blot using a β-lactamase antibody. Band intensities for tripartite fusions containing Im7 mutants were normalized to the expression level of β-lactamase containing wild-type Im7. Error bars represent the standard deviation of the mean of the MIC, the standard deviation of the mean of the expression levels measurements, or the standard deviation of the $\Delta\Delta G^{\circ}_{UN}$ measurements.

increased antibiotic resistance of fusions containing variants of the protein.

Antibiotic Resistance of the Fusion Protein Is Related to Kinetic Stability of the Inserted Im7 Sequence

The thermodynamic stability of a protein depends on the balance of its folding and unfolding rate constants (k_f and k_u, respectively) and is described by $\Delta G^{\circ}_{UN} = -RT\ln(k_f/k_u)$. To determine the effect of the introduced mutations on the kinetic stability of the Im7 variants, their unfolding rates were measured using stopped-flow fluorescence (Figure S3 and Table S2). The results showed that 27 of the 31 mutants selected for increased antibiotic resistance in vivo had increased kinetic stability (decreased k_u), with the most thermodynamically stable variants having a 5-fold decreased k_u compared with wild-type Im7 (Figure 5B and Table S2). Only three variants (V33E, V33E D63N, and N26K D49N S58R) that showed higher levels of antibiotic resistance as part of the tripartite fusion showed more rapid unfolding and lower thermodynamic stability than wild-type Im7 (Figure 5B, upper right quadrant). The in vitro proteolytic susceptibility of V33E was similar to that of wild-type Im7, whereas V33E D63N and N26K D49N S58R showed higher protease resistance, consistent with the enhanced MIC of at least the latter two mutants (Figure S2). Compared to wild-type Im7, the expression level of N26K D49N S58R is similar, that of V33E D63N is higher, and that of V33E is lower. For these variants, reduced local fluctuations may be mitigating the effect of global destabilization (Park and Marqusee, 2005), highlighting the complex interplay of global versus local dynamics in determining protein expression in vivo. None of the mutants isolated for higher MIC were both thermodynamically stabilized and kinetically destabilized (Figure 5B, upper left quadrant). Overall, therefore, the results reveal that slow unfolding from the native state is an important determinant of stability and expression levels.

Although it is more difficult to assess the individual effects of amino acid substitutions in variants with multiple mutations, the vast majority of the double and triple mutants are combinations of single mutants with known stabilizing effects, which are approximately additive for stability. To determine whether even more stable variants of Im7 could be created by combining selected mutations, the most stabilized single mutant (D35N) and the most stable double mutant (L18F D63N) were combined to create the triple mutant L18F D35N D63N. This variant showed the largest increase in thermodynamic and kinetic stability of all the selected mutants studied and showed high protease resistance (Figure S2). However, its MIC and expression level were lower than the best triple mutants we obtained. Systematic combination of mutations in Im7 variants selected for enhanced stability can thus be used to create variants with stabilities enhanced even further.

Various computational approaches have been used to predict changes in the stability of proteins upon mutation (Eijsink et al., 2005). The stability of the selected Im7 variants was determined using four such prediction algorithms (Figure 6 and Figure S4). Prediction of the effect of destabilizing core mutations was in general reasonably successful for the destabilized variants of Im7 created using systematic mutation of core residues (Capaldi et al., 2002). However, none of the four programs tested efficiently predicted $\Delta\Delta G^{\circ}_{UN}$ values for the stabilized mutants of Im7 obtained by selection (Figure 6 and Figure S4). This highlights the difficulty of predicting thermodynamically stabilizing mutations in proteins, especially when solvent-exposed residues are involved (Wunderlich et al., 2005), and thus reveals another potential benefit of the tripartite system in providing a new foundation for the generation of more reliable prediction programs.

Location and Function of the Stabilizing Mutations

Figure 7 shows the location of mutations in Im7 selected for increased antibiotic resistance displayed in the context of the

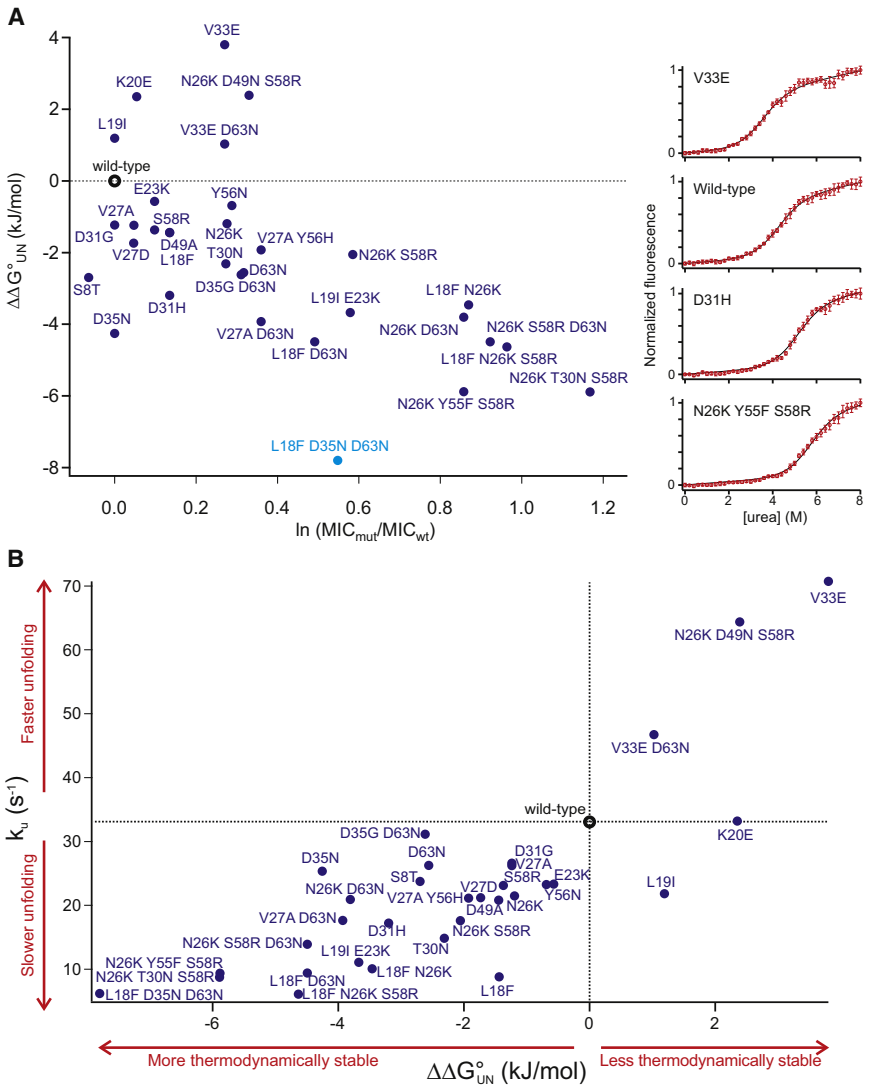


Figure 5. Im7 Mutants Selected for Increased Antibiotic Resistance Exhibit Enhanced Thermodynamic and Kinetic Stability

(A) Relative antibiotic resistance, expressed as $\ln(MIC_{mut}/MIC_{wt})$, is plotted against relative thermodynamic stability, $\Delta\Delta G^{\circ}_{UN}$, where $\Delta\Delta G^{\circ}_{UN} = \Delta G^{\circ}_{UN}$ (mutant) – ΔG°_{UN} wild-type). The designed variant L18F D35N D63N is highlighted in cyan. Equilibrium denaturant titrations for wild-type Im7 and different selected variants are shown on the right. Solid lines show the fit to a two-state equilibrium unfolding equation (Ferguson et al., 1999). Error bars represent the standard deviation of the fluorescence signal of each sample from the eight replicates.

(B) Relative contribution of thermodynamic stability ($\Delta\Delta G^{\circ}_{UN}$ measured by equilibrium urea titration) and kinetic stability (measured as the unfolding rate of the native protein) of selected (blue) and designed (cyan) Im7 variants, see also Figure S3 and Table S2.

protein's 87 residue sequence and 3D structure. The result is striking in that even though only 14 (16%) of Im7's residues are directly involved in binding to its cognate nuclease toxin, colicin E7 (Ko et al., 1999), 10 out of the 14 (71%) thermodynamically stabilizing single point mutations isolated (Figures 7A and 7B; Table S2) map to this binding interface, and a number of these have clear effects on binding (see Table S3). Three additional mutations, D35G, Y55F, and Y56H, which were found in double or triple mutants, but were not isolated as single mutants, also map to binding interface residues. The mutations Y55F and Y56H are stabilizing in the context of the other mutations present (Table S2). In addition, all evolved double and triple mutants exhibiting an increase in thermodynamic stability alter at least one residue that is directly involved in the interaction between Im7 and colicin E7. Despite this targeting of interface residues, 12 of the 23 variants that contain an interface residue mutation and demonstrate enhanced

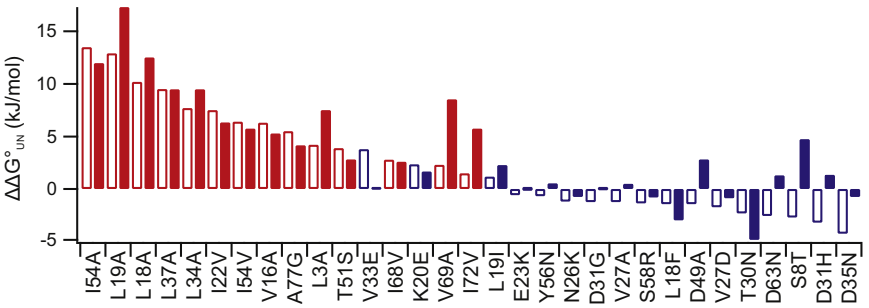


Figure 6. Stabilizing Mutations Are Poorly Predicted by Computational Methods

The effect of mutations on Im7 stability was determined experimentally (open bars) and calculated (solid bars) using the program FoldX; see also Figure S4. Only single mutations are considered here. Red bars indicate the data for hydrophobic truncation mutations; blue bars indicate mutations selected for antibiotic resistance. The variants are ordered in each plot according to the experimentally measured thermodynamic stability (least stable on the left and most stable on the right). FoldX predicted the stabilities of the published destabilized Im7 variants with some success (R^2 value was 0.62); it was unable to efficiently predict the effects of selected mutants (R^2 value was 0.02).

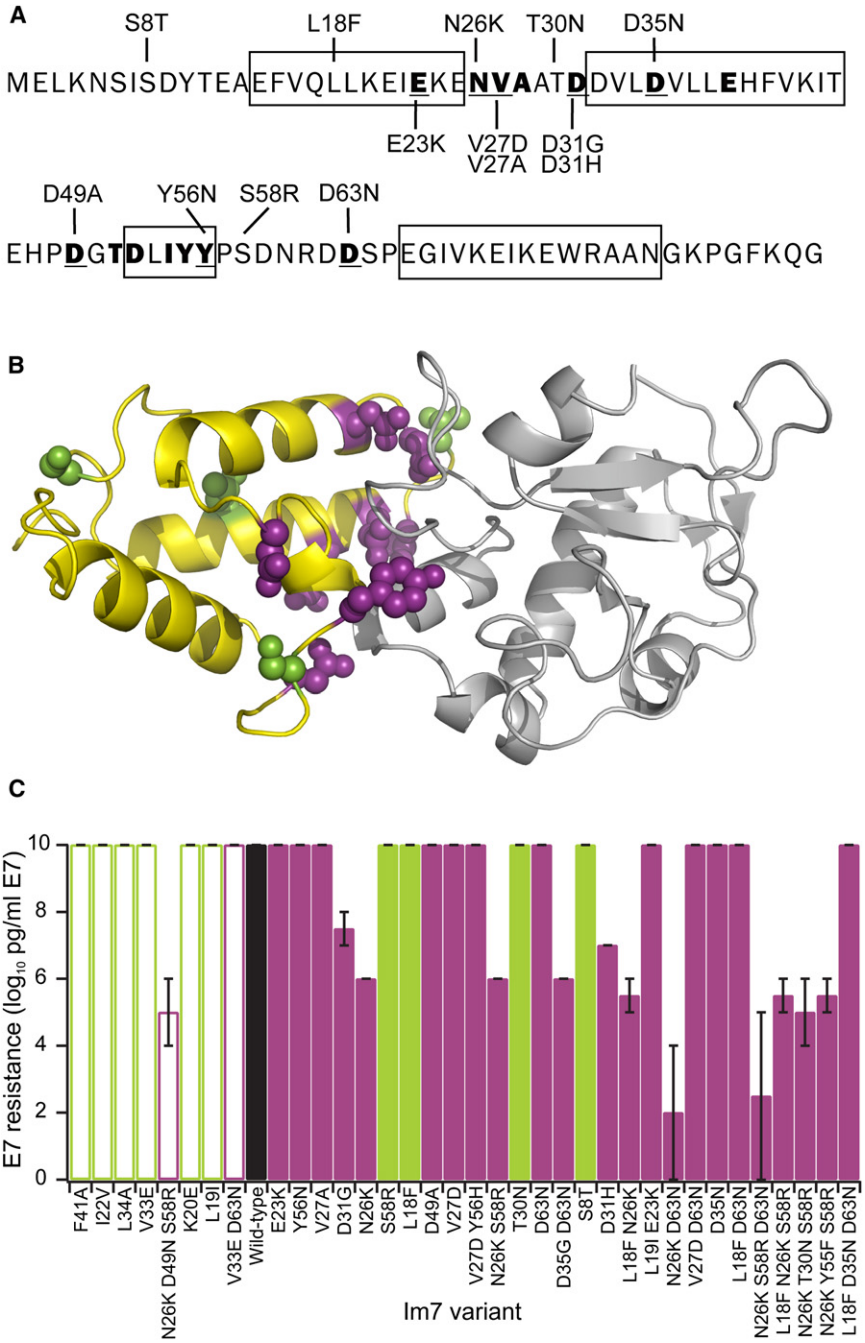


Figure 7. Stabilizing Mutants in bla′-Im7-′bla Are Concentrated in Functional Regions of the Inserted Protein

(A) Im7 sequence showing the 14 single mutations that show a thermodynamically stabilizing effect (Table S2). Binding interface residues identified in the Im7:colicin E7 cocrystal structure (Ko et al., 1999; PDB code 7CEI) are indicated by bold type. Binding residues that were found in mutants that are thermodynamically stabilized are underlined. The native α helices are shown by boxes; see also Figure S5 and Table S3.

(B) Cocrystal structure of Im7 (yellow) bound to the nuclease domain of colicin E7 (gray). Mutated residues selected for enhanced antibiotic resistance in variants containing single point mutations that increase thermodynamic stability and are also involved in the binding interface are highlighted in purple; selected residues that increase thermodynamic stability but are not directly involved in the binding interface are colored green.

(C) In vivo activity of all selected variants, measured by the resistance of *E. coli* cells expressing each Im7 variant to grow in the presence of colicin E7. The variants are ordered by thermodynamic stability, with the most stable on the right. Thermodynamically stabilized variants are shown in solid bars; thermodynamically destabilized variants are shown in open bars; colors are as in (B). Wild-type Im7 is shown in black. Three mutants created by deletion of hydrophobic core residues I22V, L34A, and F41A (Capaldi et al., 2002) (far left) were assayed for comparison with the selected variants. These show no measurable reduction in activity. Error bars represent the standard error from a minimum of two replicates.

thermodynamic stability show no measurable decrease in ability to protect cells against E7 induced toxicity in vivo (Figure 7C). These results are consistent with the known ability of immunity proteins to protect against colicin toxicity even when binding affinity is reduced 10^7-fold (from 10 pM to 1 μM) (Li et al., 2004). Indeed, every variant that showed decreased survival in this assay contained a mutation of N26, D31, or D35, consistent with the known crucial involvement of these side chains in determining binding affinity (Li et al., 2004; Ko et al., 1999). The effects of mutations in the binding interface were further investigated by isothermal titration calorimetry (ITC) and modeling (see Figure S5 and Table S3). The variants N26K and D31H, which were identified by the in vivo assay as reducing the resistance of cells to applied E7, show binding affinities (K_d) for E7 in vitro that are decreased by 60- and 50-fold, respectively. Variants containing mutations of other known binding site residues (Ko et al., 1999) show smaller effects of up to 10-fold reduction in affinity, while the double and triple mutants N26K S58R and N26K S58R D63N decreased the K_d by 350- and 480-fold, respectively (Figure S5 and Table S3). The effects of the selected stabilizing mutations on the association between Im7 and E7 were also modeled (Table S3), and the results are consistent with the functional effects observed. The ability to maintain function in the stabilized mutants, even though residues in the binding interface have been targeted by selection, shows that the tripartite system can be used to evolve proteins with enhanced stability and expression level without ablating their function entirely.

DISCUSSION

Antibiotic Resistance Can Be Used to Predict and Evolve the Stability of Proteins

Understanding how a protein's stability and level of expression are determined in vivo is one of the key unsolved problems in basic biology. The vast majority of studies aimed at addressing the origins of protein stability have been performed in vitro, where techniques capable of determining ΔG°_{UN} can be applied with accuracy. Although some reports suggest that in vitro and in vivo stability may be correlated (Ghaemmaghami and Oas, 2001; Ignatova and Gierasch, 2004), the generality of this relationship remains unclear because measurement of stability in vivo is complex and difficult to quantify in situ with confidence. To address this problem, we developed a generally applicable technique that directly links the stability of a given protein to an easily detectable phenotype: antibiotic resistance. A significant correlation was found between the level of antibiotic resistance observed in vivo and the thermodynamic stability measured in vitro for the 62 mutants of four proteins tested. The tripartite fusion system thus provides a selectable in vivo readout for the thermodynamic stability of proteins of various sizes, sequences, and structures; from both bacterial and eukaryotic sources; and independent of the presence of cofactors or disulfide bonds. The excellent correlation observed between in vivo antibiotic resistance and in vitro thermodynamic stability provides good evidence that thermodynamic stability is one of the key determinants of in vivo protein stability. Unfolding rates also appear to be important in determining in vivo stability and expression level, presumably because a slow unfolding rate provides enhanced resistance to protease digestion. Local structural fluctuations have also been shown to be important in determining the susceptibility of proteins to proteolysis (Park and Marqusee, 2005), and such fluctuations could represent an upper limit on the in vivo stability of proteins once global unfolding is no longer limiting. Im7 undergoes local unfolding at equilibrium involving unfolding of its third helix (Gorski et al., 2004). Modulating such fluctuations could alter the protein's sensitivity to proteolysis. The observation that the expression levels of the fusion proteins correlate well with the measured levels of antibiotic resistance further demonstrates the feasibility of using directed evolution with antibiotic selection to select for proteins with enhanced expression and stability in vivo. The tripartite system developed here could also be used to enhance protein solubility in aggregation-prone systems. Wild-type G-CSF is poorly soluble and shows MIC values in our system 1.8-fold lower than the more soluble variant C17S. Similarly, the aggregation-prone MBP mutant G32 I33P has a MIC value 5.6-fold lower than wild-type MBP (data not shown).

Why proteins are marginally stable and what role marginal stability plays in determining a protein's functional properties is not clear. Despite a wealth of experiments and theoretical approaches to address the question of why proteins are only marginally stable, there is currently no consensus (Sanchez et al., 2006; DePristo et al., 2005; Tokuriki et al., 2008) It has been suggested that proteins may be forced into marginal stability by the destabilizing effects of random mutation, which may act to decrease a protein's stability until it reaches a threshold value below which function begins to be affected (Bloom et al., 2006, 2007; Taverna and Goldstein, 2002). Alternatively, it has been discussed that marginal stability may be functionally beneficial or that stability and function trade off (Bahar et al., 2007; DePristo et al., 2005; Ferreiro et al., 2007; Tsou, 1998; Beadle and Shoichet, 2002; Schreiber et al., 1994: Tokuriki et al., 2008). Since the biological function of most proteins depends on their ability to fold into thermodynamically stable structures, the common occurrence of mutations that reduce both function and stability does not indicate whether a protein's stability is optimal or merely above a particular minimal threshold. (Beadle and Shoichet, 2002; Dessailly et al., 2007). On the other hand, in vitro evolution experiments, which have shown that it is possible to dramatically stabilize enzymes without sacrificing their biological functions (Bloom et al., 2006), usually involve screening for enzyme activity following a challenge to stability (denaturant, temperature, or pH) (Eijsink et al., 2005). Function and stability are thus under simultaneous selection in both natural and experimental evolution experiments, making it difficult to establish how each has influenced the other. Our approach, by contrast, selects stabilizing mutations in vivo without any selection for function. Therefore, it provides an opportunity to separate the effects of these different evolutionary forces. We utilized our selection system for stabilizing mutations in proteins to directly access a stability range beyond the marginal stability of a given protein and simulate an evolutionary scenario in which function is not under evolutionary pressure. We reasoned that if the marginal stability of globular proteins is simply the result of random destabilizing mutations, then stabilizing mutations should occur at functional and nonfunctional residues with the same likelihood. If, on the other hand, protein stability has been limited by the requirement of the protein to be functional, stabilizing mutations would be expected to map mainly to functional residues. It has previously been shown that some functional residues including residues involved in binding sites can be suboptimal for stability (Beadle and Shoichet, 2002; Chen and Lim, 2008). Rather than altering a limited number of known functional residues and evaluating the effect on the marginal stability of a protein, our inverse approach allows us to examine a much larger number of stabilizing mutants to investigate why proteins are only marginally stable. Our results provide an intriguing example of the evolutionary compromise between function and stability. For Im7, we found a striking overrepresentation of stabilizing mutations at residues involved in binding of its partner colicin E7. It would appear, then, that functional requirements at these sites have led to a compromise in the stability of Im7. The function of Im7 is to bind tightly to colicin E7, inhibiting its nuclease activity (Wallis et al., 1995). In order to be functional, Im7 needs to bind E7 tightly while maintaining sufficient in vivo stability to escape cellular proteolytic processes. Thermodynamic principles dictate that binding will act to enhance the stability of proteins when they form a complex to an extent determined by the strength of the interaction. Proteins that display limited thermodynamic stability in their unliganded form thus possess the opportunity for an enhanced binding affinity to a partner protein by exploiting binding-induced stabilization. These considerations, and our data, indicate that the stability of the

Im7 monomer is optimized rather than maximized in evolution, as is likely to be generally the case for proteins that are involved in binding reactions, and which may be true for any protein that requires flexibility for its function. Stabilization of the Im7 monomer is hence most easily achieved by mutations at the binding interface.

Using this binding protein proved fortuitous, since Im7 has a substantial number of solvent-exposed residues that are directly involved in binding and hence do not exhibit the structural tension created by the organization of catalytic residues in active sites. Active site configurations have been shown to be inherently unfavorable for protein stability for a number of examples (Tokuriki et al., 2008). The remarkably strong affinity between Im7 and E7 ($K_d = 10^{-15}$ M) also has the advantage that a complete loss of function due to a single point mutation is unlikely. Indeed, immunity proteins confer resistance to noncognate colicin toxins despite wide variation in their binding regions (Wallis et al., 1995). It is therefore striking that 11 of the 23 mutants that thermodynamically stabilized Im7 decreased the activity of Im7 in vivo.

Advantages of In Vivo Selection for Protein Stability Using Antibiotic Resistance

Because the marginal stability of proteins poses such important scientific and practical problems, there have been numerous attempts to develop techniques to screen or select for proteins with increased stability or solubility (Chautard et al., 2007; Magliery and Regan, 2004; Muller and Johnsson, 2008; Philipps et al., 2003; Wunderlich et al., 2005). A method with a quantitative, selectable output, however, has proven elusive. Two methodologies that perhaps come closest to this goal are the GFP-based approaches pioneered by Waldo and coworkers and the PROSIDE method developed by Schmid and coworkers (Sieber et al., 1998; Waldo, 2003). Waldo's method and other screens for protein solubility are typically based on fusions to reporter constructs such as GFP where folding and activity of the reporter is dependent on the folding and/or solubility of the test protein. However, simple proteolytic destruction of the test protein, as might be expected for poorly folded proteins, can relieve the interference and lead to false positive results. Since a protein's solubility is not necessarily related to its activity or thermodynamic stability (Famm et al., 2008; Garcia-Fruitos et al., 2007), it is not possible to separate which of these phenomena causes the enhanced antibiotic resistance in our examples; however, the results highlight the potential of the tripartite system as a means to select variants of proteins that are expressed in soluble form and are thus easier to produce, handle, and characterize.

Generally, selections in which only the desired variants survive are preferable to screens in which the specific properties of each variant must be determined individually. Previous studies have described selection strategies to stabilize proteins (Magliery and Regan, 2004; Roodveldt et al., 2005). The PROSIDE method (Sieber et al., 1998) links the infectivity of a filamentous phage to the in vitro protease and denaturant resistance of protein variants fused into the phage protein g3p. This approach has been used to enhance the stability of proteins and to select for structure, folding, and enzymatic activity (reviewed by Magliery and Regan, 2004). In contrast to the tripartite system described here, phage display-based methods are in vitro selections that appear to be most suitable for further enhancing the stability of already stable, small, single-domain proteins that can be readily displayed on the surface of phage.

The selection system described here provides a strategy with the unique capability of directly quantifying protein stability in vivo. Our system combines a number of advantages: it can be applied to proteins irrespective of the stability of the wild-type protein; is based on a selection rather than a screen; is easy to set up using commonly available *E. coli* strains, plasmids, and reagents; can be applied to a variety of proteins; requires no knowledge of the protein's structure, folding, or activity; has a very low false positive rate (only 4 of 31 Im7 variants were not stabilized either thermodynamically or kinetically); and gives a quantitative output of the thermodynamic stability of a protein. Systematic studies show that up to 94% of eukaryotic proteins fail to express in *E. coli* or form insoluble inclusion bodies (Mehlin et al., 2006), and other proteins that can be expressed in *E. coli* often lack sufficient in vitro stability to allow for detailed biochemical or structural studies, highlighting the need for systems able to enhance the biochemical properties of proteins of choice.

The data generated from such in vivo selections could also be used to enhance the power of stability prediction algorithms to include variants that are more adventurous than those obtained to date by rational targeted mutagenesis approaches—an important objective, given the paucity of current algorithms to predict stability when solvent-exposed residues are altered (Figure S5). Further, it presents a new approach to investigate the interaction of folding proteins with their in vivo environment, providing a route by which organisms tailored for the expression of a desired protein may be evolved by isolating *E. coli* strains with mutations in folding factors, chaperones, or other proteins that alter folding. While the system developed here has focused on the *E. coli* periplasm as the targeted folding environment, the strategy of using a tripartite fusion as a suitable selection marker is readily transferable to other folding compartments. Finally, the ability to evolve stability separately from function provides an exciting new framework to dissect the relative importance of the different evolutionary pressures that determined today's protein sequences.

EXPERIMENTAL PROCEDURES

Detailed descriptions of all experimental procedures are provided in the Supplemental Data.

Testing Antibiotic Resistance

Spot titer tests were performed to determine the level of antibiotic resistance of cells expressing β-lactamase tripartite fusions. Serial dilutions of mid-log phase cells of *E. coli* NEB10β were spotted onto LB plates supplemented with increasing concentrations of penicillin V. After incubation at 37°C for 18 hr, the minimal concentration of penicillin V preventing cell growth, defined as the MIC, was determined as described in the Supplemental Data.

Selection of Im7 Mutants with Increased Antibiotic Resistance

Mutagenesis of the Im7 gene was performed by the MEGAWHOP technique as detailed in the Supplemental Data. The mutated DNA was transformed into chemically competent NEB10β cells and subjected to selection on LB penicillin V plates. The template plasmid pBR322*link-Im7 WT was used as

a control. After incubation overnight at 37°C, single colonies of the library were picked from plates containing penicillin V concentrations that did not allow any cell growth of pBR322*link-Im7 WT. The mutated plasmid DNA was isolated, sequenced, and retransformed into fresh NEB10β cells for determination of MIC values.

β-Lactamase Expression Levels and In Vivo Colicin Assays

Expression levels were measured by western blot, using antibodies against wild-type TEM1 β-lactamase (Millipore). Endogenous MBP was used as an internal control. The ability of different Im7 variants to inhibit the toxicity of *E. coli* colicin E7 was determined by assessing the survival of cells expressing those variants when colicin E7 was titrated onto cells growing on agar (Wallis et al., 1995).

Equilibrium and Kinetic Stability Measurements, Limited Proteolysis, and Binding Affinity Measurements

His-tagged Im7 variants were expressed and purified essentially as previously described (Capaldi et al., 2002). Biophysical measurements were carried out at 37°C in 50 mM Tris-HCl containing 0.4 M Na_2SO_4 at pH 7.5. Equilibrium denaturant titrations were carried out in a 384-well microplate format using a Hamilton Microlab Star robot to set up samples and a PerkinElmer EnVision plate reader to measure fluorescence, based on the previously published method (Ferguson et al., 1999). Unfolding kinetics were measured by stopped-flow fluorescence (Capaldi et al., 2002). Proteolytic susceptibility of Im7 variants was measured by comparing band intensities of each variant on an SDS-PAGE gel before and after digestion with proteinase K for 5 min at 37°C. The affinity of Im7 variants for the colicin E7 DNase domain was determined by measuring the ability of the mutants to displace bound Im8 from E7 using ITC as described in the Supplemental Data.

SUPPLEMENTAL DATA

Supplemental Data include five figures, three tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00863-6.

ACKNOWLEDGMENTS

We thank Colin Kleanthous for the generous gifts of colicin E7 and Im8, as well as many insightful discussions; Nicola Meenan, Nick Housden, and Wei Feng Xue for help with ITC measurements and analysis; Tsinatkeab Tadesse for help with measuring MBP stabilities; and Dana Reichmann and Vladimir Potapov for modeling binding affinities. We also thank Ursula Jakob for critical reading of the manuscript. This work was funded by the Howard Hughes Medical Institute (HHMI), the National Institutes of Health (NIH), The Wellcome Trust (reference 062164//z/00/Z), and the University of Leeds. The data presented here are part of a Ph.D. dissertation conducted by L.F. for the University of Münster, Institute for Chemistry and Pharmacy.

Received: March 27, 2009
Revised: August 4, 2009
Accepted: October 24, 2009
Published: December 10, 2009

REFERENCES

Bahar, I., Chennubhotla, C., and Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Curr. Opin. Struct. Biol. *17*, 633–640.

Beadle, B.M., and Shoichet, B.K. (2002). Structural bases of stability-function tradeoffs in enzymes. J. Mol. Biol. *32*, 285–296.

Bishop, B., Koay, D.C., Sartorelli, A.C., and Regan, L. (2001). Reengineering granulocyte colony-stimulating factor for enhanced stability. J. Biol. Chem. *276*, 33465–33470.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA *103*, 5869–5874.

Bloom, J.D., Raval, A., and Wilke, C.O. (2007). Thermodynamics of neutral protein evolution. Genetics *175*, 255–266.

Capaldi, A.P., Kleanthous, C., and Radford, S.E. (2002). Im7 folding mechanism: misfolding on a path to the native state. Nat. Struct. Biol. *9*, 209–216.

Chautard, H., Blas-Galindo, E., Menguy, T., Grand'Moursel, L., Cava, F., Berenguer, J., and Delcourt, M. (2007). An activity-independent selection system of thermostable protein variants. Nat. Methods *4*, 919–921.

Chen, Y.-C., and Lim, C. (2008). Common physical basis of macromolecule-binding sites in proteins. Nucleic Acids Res. *36*, 7078–7087.

Chun, S.Y., Strobel, S., Bassford, P., Jr., and Randall, L.L. (1993). Folding of maltose-binding protein. J. Biol. Chem. *268*, 20855–20862.

DePristo, M.A., Weinreich, D.M., and Hartl, D.L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. Nat. Rev. Genet. *6*, 678–687.

Dessailly, B.H., Lensink, M.F., and Wodak, S.J. (2007). Relating destabilizing regions to known functional sites in proteins. BMC Bioinformatics *8*, 141.

Eijsink, V.G.H., Gåseidnes, S., Borchert, T.V., and van den Burg, B. (2005). Directed evolution of enzyme stability. Biomol. Eng. *22*, 21–30.

Espargaro, A., Castillo, V., de Groot, N.S., and Ventura, S. (2008). The *in vivo* and *in vitro* aggregation properties of globular proteins correlate with their conformational stability. J. Mol. Biol. *378*, 1116–1131.

Famm, K., Hansen, L., Christ, D., and Winter, G. (2008). Thermodynamically stable aggregation-resistant antibody domains through directed evolution. J. Mol. Biol. *376*, 926–931.

Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C., and Radford, S.E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J. Mol. Biol. *286*, 1597–1608.

Ferreiro, D.U., Hegler, J.A., Komives, E.A., and Wolynes, P.G. (2007). Localizing frustration in native proteins and protein assemblies. Proc. Natl. Acad. Sci. USA *104*, 19819–19824.

Galarneau, A., Primeau, M., Trudeau, L.-E., and Michnick, S.W. (2002). β-Lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein-protein interactions. Nat. Biotechnol. *20*, 619–622.

Garcia-Fruitos, E., Aris, A., and Villaverde, A. (2007). Localization of functional polypeptides in bacterial inclusion bodies. Appl. Environ. Microbiol. *73*, 289–294.

Ghaemmaghami, S., and Oas, T.G. (2001). Quantitative protein stability measurement *in vivo*. Nat. Struct. Biol. *8*, 879–882.

Gorski, S.A., Le Duff, C.S., Capaldi, A.P., Kalverda, A.P., Beddard, G.S., Moore, G.R., and Radford, S.E. (2004). Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. J. Mol. Biol. *337*, 183–193.

Gosavi, S., Whitford, P.C., Jennings, P.A., and Onuchic, J.N. (2008). Extracting function from a β-trefoil folding motif. Proc. Natl. Acad. Sci. USA *105*, 10384–10389.

Ignatova, Z., and Gierasch, L.M. (2004). Monitoring protein stability and aggregation *in vivo* by real-time fluorescent labeling. Proc. Natl. Acad. Sci. USA *101*, 523–528.

Kamiya, N., Okimoto, Y., Ding, Z., Ohtomo, H., Shimizu, M., Kitayama, A., Morii, H., and Nagamune, T. (2001). How does heme axial ligand deletion affect the structure and the function of cytochrome b(562). Protein Eng. *14*, 415–419.

Ko, T.P., Liao, C.C., Ku, W.Y., Chak, K.F., and Yuan, H.S. (1999). The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure *7*, 91–102.

Li, W., Keeble, A.H., Giffard, C., James, R., Moore, G.R., and Kleanthous, C. (2004). Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. J. Mol. Biol. *337*, 743–759.

Magliery, T.J., and Regan, L. (2004). Combinatorial approaches to protein stability and structure. Eur. J. Biochem. *271*, 1595–1608.

Mayer, S., Rudiger, S., Ang, H.C., Joerger, A.C., and Fersht, A.R. (2007). Correlation of levels of folded recombinant p53 in *Escherichia coli* with thermodynamic stability *in vitro*. J. Mol. Biol. *372*, 268–276.

Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N., et al. (2006). Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. Mol. Biochem. Parasitol. *148*, 144–160.

Muller, J., and Johnsson, N. (2008). Split-ubiquitin and the split-protein sensors: chessman for the endgame. ChemBioChem *9*, 2029–2038.

Park, C., and Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nat. Methods *2*, 207–212.

Parsell, D.A., and Sauer, R.T. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. J. Biol. Chem. *264*, 7590–7595.

Philipps, B., Hennecke, J., and Glockshuber, R. (2003). FRET-based *in vivo* screening for protein folding and increased protein stability. J. Mol. Biol. *327*, 239–249.

Qureshi, S.A. (2007). β-lactamase: an ideal reporter system for monitoring gene expression in live eukaryotic cells. Biotechniques *42*, 91–96.

Roodveldt, C., Aharoni, A., and Tawfik, D.S. (2005). Directed evolution of proteins for heterologous expression and stability. Curr. Opin. Struct. Biol. *15*, 50–56.

Sanchez, I.E., Tejero, J., Gomez-Moreno, C., Medina, M., and Serrano, L. (2006). Point mutations in protein globular domains: contributions from function, stability and misfolding. J. Mol. Biol. *363*, 422–432.

Schreiber, G., Buckle, A.M., and Fersht, A.R. (1994). Stability and function: two constraints in the evolution of barstar and other proteins. Structure *2*, 945–951.

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

Taverna, D.M., and Goldstein, R.A. (2002). Why are proteins marginally stable? Proteins *46*, 105–109.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS Comput. Biol. *4*, e1000002. 10.1371/journal.pcbi.1000002.

Tsou, C.L. (1998). Active site flexibility in enzyme catalysis. Ann. N Y Acad. Sci. *864*, 1–8.

Waldo, G.S. (2003). Genetic screens and directed evolution for protein solubility. Curr. Opin. Chem. Biol. *7*, 33–38.

Wallis, R., Leung, K.-Y., Pommer, A.J., Videler, H., Moore, G.R., James, R., and Kleanthous, C. (1995). Protein-protein interactions in colicin E9 DNase-immunity protein complexes. Biochemistry *34*, 13751–13759.

Wunderlich, M., Martin, A., Staab, C.A., and Schmid, F.X. (2005). Evolutionary protein stabilization in comparison with computational design. J. Mol. Biol. *351*, 1160–1168.

Wurth, C., Guimard, N.K., and Hecht, M.H. (2002). Mutations that reduce aggregation of the Alzheimer's Aβ42 peptide. J. Mol. Biol. *319*, 1279–1290.

Exploring the interplay of stability and function in protein evolution

New methods further elucidate why protein stability is necessarily so tenuous and stability-increasing mutations compromise biological function.

Gustavo Caetano-Anollés[1)2)*] *and Jay Mittenthal*[2)3)]

A new split β-lactamase assay promises experimental testing of the interplay of protein stability and function. Proteins are sufficiently stable to act effectively within cells. However, mutations generally destabilize structure, with effects on free energy that are comparable to the free energy of folding. Assays of protein functionality and stability *in vivo* enable a quick study of factors that influence these properties in response to targeted mutations. These assays can help molecular engineering but can also be used to target important questions, including why most proteins are marginally stable, how mutations alter structural makeup, and how thermodynamics, function, and environment shape molecular change. Processes of self-organization and natural selection are determinants of stability and function. Non-equilibrium thermodynamics provides crucial concepts, *e.g.*, cells as emergent energy-dissipating entities that do work and build their own parts, and a framework to study the sculpting role of evolution at different scales.

Keywords:
fitness; mutation; protein folding; split β-lactamase assay; stability-function tradeoff

DOI 10.1002/bies.201000038

[1)] Department of Crop Sciences, University of Illinois, Urbana, IL, USA
[2)] Institute for Genomic Biology, University of Illinois, Urbana, IL, USA
[3)] Department of Cell and Developmental Biology, University of Illinois, Urbana, IL, USA

***Corresponding author:**
Professor Gustavo Caetano-Anollés
E-mail: gca@uiuc.edu

Introduction

Proteins change gradually, usually by accumulation of amino acid substitutions. An average protein with ~300 amino acid residues, if mutated, defines a permutational space of possible sequences that is far more than astronomical ($\sim 10^{321}$–10^{469}) [1]. As expected, only a minute fraction of possible sequences have materialized in the ~4 billion-year-long history of life ($\sim 10^{32}$). This exploration of sequence space has resulted in the discovery of the ~1200 protein folds that are known. These folds are highly ordered three-dimensional structures that define the shape of the protein molecules and embed all known biological functions. While natural proteins generally fold into well-defined structures and are associated with clearly defined functions, they are only marginally stable, with free energies of folding in the range 20–60 kJ mol^{-1} [2]. In fact, mutations generally destabilize structure with effects on free energy that are comparable to the free energy of folding. The interplay between the search for stable fold structures in sequence space and the discovery and maintenance of useful functions seems of crucial evolutionary significance. Its exploration by experimental means is a fundamental endeavor.

A recent and ingenious split β-lactamase method that selects for antibiotic resistance and stable proteins now provides a way to explore this interplay without requiring that we know the structure of a protein [3]. The method exploits the tolerance of the TEM-1-β-lactamase enzyme of *Escherichia coli* to host the insertion of a ''test protein'' fragment in a loop on the surface of the enzyme [4]. If the test protein is stable enough to bring the halves of a β-lactamase enzyme together so that they can associate effectively, the *E. coli* cells will be resistant to β-lactam antibiotics (Fig. 1). If the stability of the test protein is compromised by mutation, the two β-lactamase halves will not associate and the bacterial cells will be less resistant to the antibiotic, unless mutations induce structural change that still keeps the ends of the β-lactamase halves together. Foit *et al.* [3] discovered that there was a striking correlation between antibiotic resistance and thermodynamic stability of mutated test

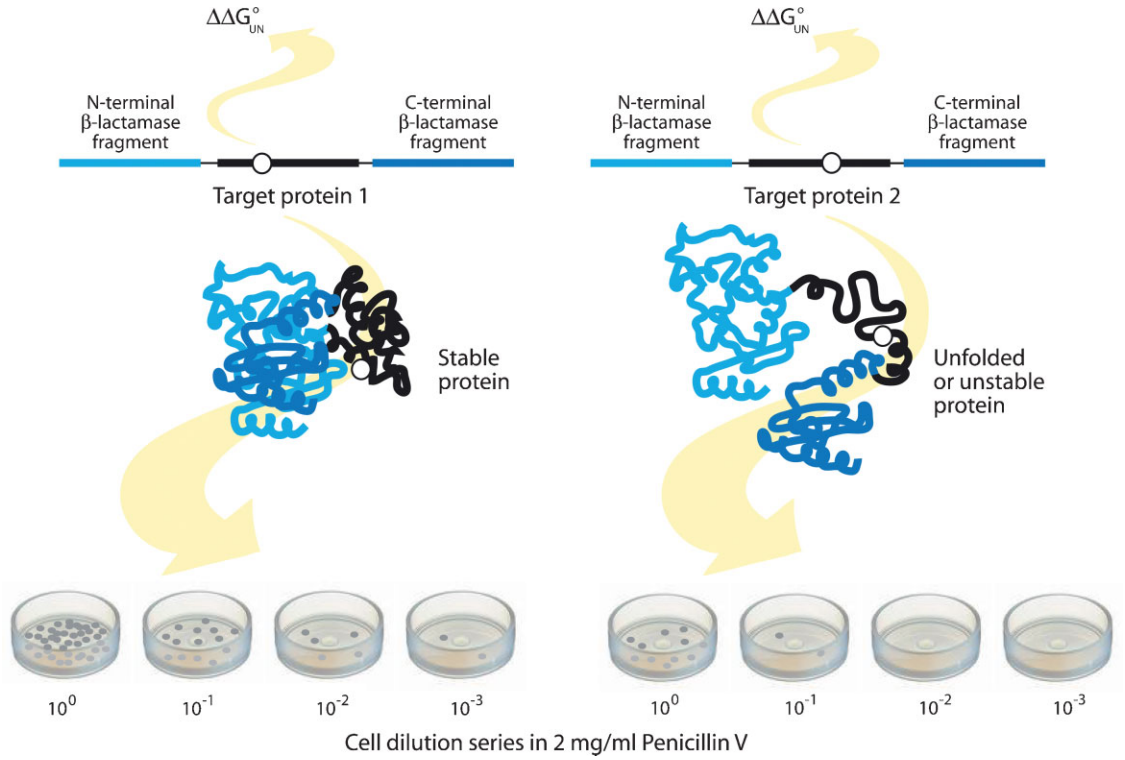


Figure 1. The split β-lactamase method. Test proteins with different mutations are inserted into TEM1-β-lactamase. Mutations on the test protein can either decrease its stability or make the reporter non-functional. Cellular proteases that are part of the quality control machinery of the cell will then degrade the poorly folded or unstable complexes, decreasing the resistance of the bacterial host to antibiotics. Plating transformed bacterial cells on an antibiotic dilution series and calculating the maximal cell dilution that allows growth can assay antibiotic resistance levels. Protein stability measured as difference in free energy ($\Delta\Delta G^{\circ}_{UN}$) between mutant and wild-type test proteins is assayed with biophysical methods such as equilibrium denaturant titrations or unfolding kinetics measured by stop-flow fluorescence [3].

proteins over a sevenfold range of penicillin V concentrations and over a 10–30 kJ mol^{-1} change in free energy of protein stability. This correlation was consistently recovered when studying mutant variants of toxin-binding immunity protein 7 (Im7), cytochrome b_{562}, human granulocyte colony-stimulating factor, and maltose-binding protein, which were generated by error-prone PCR. Remarkably, in these studies, mutations that enhance thermodynamic stability of Im7 map to surface residues involved in the binding of the protein to its cognate toxin (colicin E7). This observation reveals how evolutionary constraints imposed by the interaction of Im7 and E7 compromise the stability of Im7.

The method identifies stabilizing mutations in target proteins and complements a host of other strategies. However, its generic and quantitative properties are rivaled by two other approaches, the Proside *in vitro* selection method [5] and the green fluorescent protein (GFP) folding interference assay [6, 7]. The Proside method inserts a repertoire of protein variants between the N-terminal and C-terminal domains of the gene-3-protein (g3p) of the filamentous phage fd [8]. Since infectivity of the phage is lost if the insert is unstable and is degraded, rounds of *in vitro* proteolysis, infection, and phage propagation can be used to enrich for those protein variants that are most stable. The GFP reporter assay uses a C-terminal fusion to evaluate the success of the folding of an upstream-fused polypeptide. Fluorescence of the GFP fusion is positively correlated with folding of the target polypeptide. Cycles of mutagenesis and selection for directed evolution can make use of this feature to engineer protein variants with increased solubility, finding mutations that for example reduce disease-linked peptide aggregation. However, truncation artifacts can affect measurements of folding and solubility. Mitigating these effects require the use of novel GFP insertion reporters with circular permutations [9]. Both approaches have been successfully used in directed evolution applications. Interestingly, Proside was used to select for mutants of TEM-1-β-lactamase that have increased folding stability and optimize the packing of helical segments in the protein [10].

The assay of Foit *et al.* [3] has potential to analyze the effects of mutation not only on functionality and thermodynamic and kinetic stability, but also on general structural makeup of the fold and on biophysical properties such as solubility and hydrophobicity. Many folds tolerate a number of secondary-structure embellishments that are sometimes associated with changes in function [11]. It would be valuable to study how embellishments affect both stability and function. Since protein expression appears to be linked to stability, the method can also be used to enhance expression levels in protein engineering [12]. This link suggests that the quality control machinery of the cell and the proteostasis network that guarantee the correct folding of proteins [13] are needed for

protein turnover and are important determinants of biological function. Genetic selection approaches such as the split β-lactamase assay can be used to study how the molecular components of quality control and proteostasis impact protein evolution in microbial or eukaryotic organisms. Similarly, the effects of biotic and abiotic stress, or even lifestyle, on protein stability can be assayed. This could clarify for example the molecular adaptations in microbes and other organisms that make them capable of living in extreme environments [14].

Function, stability and design

Since the interplay between stability and functionality is crucial for the effective design of proteins, one might try to increase both the stability of an enzyme and its catalytic activity. However, Foit *et al.* [3] found that most mutations that increase stability occurred at the active site of one of the proteins they tested, Im7. Thus, functionality compromises stability, at least in this case. This trade-off helps to explain why most proteins are marginally stable. Recent protein ''resurrection'' experiments also manifest the interplay of function and stability [15]. The sequence of an ancient vertebrate glucocorticoid receptor, an important steroid hormone-regulated transcription factor, was inferred by phylogenetic reconstruction. This protein was then biochemically synthesized, expressed in cultured cells, and its structure analyzed by X-ray crystallography. Remarkably, the history of mutations recapitulating the evolution of the receptor's hormone specificity showed that some mutations repositioned crucial residues for receptor-ligand and intraprotein interactions but were destabilizing. These initial changes of large phenotypic effect were then followed by mutations that stabilized the new function [15]. A provocative suggestion to increase the potential of these kinds of experiments is to test resurrected proteins using the split β-lactamase strategy, avoiding the limiting crystallographic step. The assay could also improve our ability to engineer proteins. It might be possible to evolve a test protein within cells and, in response to an extracellular signal, have the cell insert the test protein into the antibiotic resistance protein, so the assay could be performed. Investigators could then test the function and stability of the protein during its evolution *in situ*. This procedure might be used to modify an enzyme so it performs a novel but related function, or to see how changes in the intracellular milieu affect protein stability.

To appreciate the interplay between stability and functionality more deeply, note that functionality influences the fitness of an organism. Fitness is the net rate of reproduction, birth rate minus death rate. The stability of a system's state measures the tendency to return to that state after a perturbation. A protein is mutationally stable if its minimum free energy (mfe) shape is maintained despite diverse neutral mutations. It is thermodynamically stable if its mfe shape is maintained over a wide range of environmental conditions. It is kinetically stable if it has few ways to misfold, *i.e.* it strongly tends to fold into the mfe shape [16]. Stability does not depend on functionality. Thus, the assay for stability of Foit *et al.* [3] measures the tendency of a wild-type or mutated protein, active or inactive, to retain its native shape in an intracellular environment.

To design macromolecules, for a specified sequence of monomers, it helps to have models to compute the low-lying free energy levels, the corresponding repertoire of shapes, and the energy landscape of transitions between shapes that characterizes the kinetics of folding. Such models are available for predicting RNA secondary structure (here shape) – the pattern of local base pairing in an RNA molecule [17]. Using these models, Ancel and Fontana [18] examined the interplay between fitness and stability, by modeling the evolution of RNA shape. The fittest phenotype was a target shape. Because the Boltzmann distribution determines the probability of occurrence of a shape, RNA molecules spend more time in shapes with lower free energy. So, selection for the target shape favors mutants having that shape more often because they have lower energy. These mutants also have lower plasticity – fewer thermally accessible shapes, most of which resemble the target shape. During selection, the variability of shapes under point mutation also decreases; most mutants with one or two point mutations fold to nearly the target shape. Thus, evolved RNAs lock into the target shape, with relatively low plasticity and variability.

It is likely that this scenario also describes the evolution of proteins. Natural proteins can accumulate mutations and still fold into stable structures. This supports Kimura's theory that much of protein evolution is neutral [19]. Models of protein folding show that a funnel in the energy landscape directs folding into a protein's native shape, which is robust to environmental change [20, 21]. Since protein stability promotes mutational robustness [22], a feature that has been experimentally verified with variants of subtilisin and TEM-1-β-lactamase, the link between stability and tolerance to mutation can be used to study neutral evolution of proteins and can be derived directly from mutational changes in free energy [23]. In this regard, the methods here described provide tools to test mathematical theories of how thermodynamics shape neutral protein evolution [23]. For example, the split β-lactamase strategy could be used to test if indeed neutral evolution leads to marginally stable proteins.

Models to compute low-lying free energy levels, shapes, and energy landscapes are not available for many real proteins. Computational methods for inferring the tertiary structure of a protein from its sequence give accurate results for about 60% of sequenced proteins. Molecular dynamics (MD) methods allow simulation of protein folding, ligand binding and catalysis, and stability of a protein. However, available algorithms limit the long MD simulations needed for these purposes. Attard [24] has offered a new non-equilibrium statistical mechanics and thermodynamics, based on transition entropy, ''the number of molecular configurations associated with a transition between macrostates in a specified time'' ([24] p. 74). This formalism offers new, more efficient methods for MD simulation and shows that commonly used equations of motion are unphysical.

Conclusions

Experiments and models of protein evolution can be used to explore the trade-off between protein functionality and stability. While the assay of Foit *et al.* [3] provides an

additional tool to close the gap between theory and experimentation within a cellular framework, the work of Attard [24] shows that non-equilibrium statistical mechanics may provide new methods for designing proteins. Non-equilibrium thermodynamics (NET) may help to answer questions about the trade-off: why does it occur? What determines the balance between functionality and stability? Are such tradeoffs more widespread?

To address these questions, let us look at an organism from the perspective of NET. An organism is an engine that converts energy from a gradient of free energy into work. The work is used to build and run the organism's structure, which performs tasks sufficient for its survival. Energy-dissipating macromolecules, such as template-directed synthesizers of nucleic acids and proteins, perform work and build more parts. Such an engine is called a self-organizing or emergent dissipative system – an EDS. As the engine operates, it promotes the flow of energy down the free energy gradient, dissipating it, and producing entropy [25].

In an organism, macromolecules self-organize, as in the folding of a protein to a stable shape. However, to perform an energy-dissipating function such as catalysis, a macromolecule must be able to change shape; it must not be too stable [14]. This balance between functionality and stability may maximize the conversion of free energy to work within organisms, thereby maximizing their fitness. Organisms compete with each other and with other EDS for energy from gradients. Organisms are self-reproducing EDSs; fitter organisms, with a greater net rate of proliferation, garner a greater share of the available energy, and dissipate gradients faster. The rate of entropy production depends on the fitness and the number of organisms, and on their size. It has been demonstrated empirically and justified theoretically that diverse EDSs have a maximal entropy production rate [26].

The preceding argument suggests that at levels of organization higher than individual macromolecules, from multimolecular assemblies to cells and multicellular organisms, marginal stability of the functional state may balance self-organization and flexibility. Similarly, Kauffman [27] proposed that gene regulatory networks operate at a boundary balancing order and chaos. NET may provide ways for calculating such balances, and to understand evolution more generally [28].

Acknowledgments

We acknowledge helpful discussions with Elbert Branscomb and the National Science Foundation (MCB-0749836 to GCA) for ongoing support. Any opinions, findings, and conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

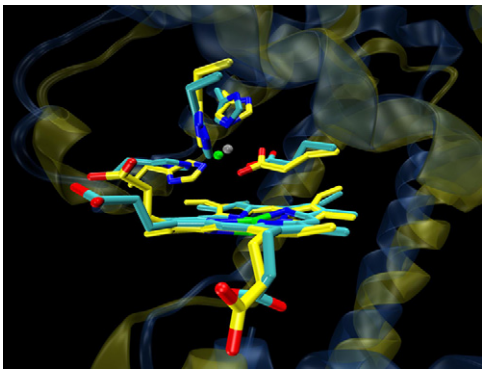
References

1. **Caetano-Anollés G**, **Wang M**, **Caetano-Anollés D**, *et al*. 2009. The origin, evolution and structure of the protein world. *Biochem J* **417**: 621–37.
2. **Fersht AR.** 1999. *Structure and Mechanism in Protein Science*. New York: W. H. Freeman.
3. **Foit L**, **Morgan GJ**, **Kern MJ**, *et al*. 2009. Optimizing protein stability *in vivo*. *Mol Cell* **36**: 861–71.
4. **Galarneau A**, **Primeau M**, **Trudeau L-E**, *et al*. 2002. β-Lactamase protein fragment complementation assays *in vivo* and *in vitro* sensors of protein-protein interactions. *Nat Biotechnol* **20**: 619–22.
5. **Martin A**, **Schmid FX**, **Sieber V.** 2003. Proside: a phage-based method for selecting thermostable proteins. *Methods Mol Biol* **230**: 57–70.
6. **Waldo GS**, **Standish BM**, **Berendzen J**, *et al.* 1999. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17**: 691–5.
7. **Waldo GS.** 2003. Genetic screens and directed evolution for protein solubility. *Curr Opin Chem Biol* **7**: 33–8.
8. **Sieber V**, **Pluckthun A**, **Schmid FX.** 1998. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**: 955–60.
9. **Cabantous S**, **Rogers Y**, **Terwilliger TC**, *et al.* 2008. New molecular reporters for rapid protein folding assays. *PLoS ONE* **3**: e2387.
10. **Kather I**, **Jakob RP**, **Dobbek H**, *et al.* 2008. Increased folding stability of TEM1 beta-lactamase by *in vitro* selection. *J Mol Biol* **383**: 238–51.
11. **Reeves GA**, **Dallman TJ**, **Redfern OC**, *et al.* 2006. Structural diversity of domain superfamilies in the CATH database. *J Mol Biol* **360**: 725–41.
12. **Wiseman RL**, **Kelly JW.** 2009. Evolving protein stability through genetic selection. *Mol Cell* **36**: 730–1.
13. **Balch WE**, **Morimoto RI**, **Dillin A**, *et al.* 2008. Adapting proteostasis for disease intervention. *Science* **319**: 916–9.
14. **Fields PA.** 2001. Protein function and thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol A* **129**: 417–31.
15. **Ortlund EA**, **Bridgham JT**, **Redinbo MR**, *et al.* 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**: 1544–8.
16. **Fontana W.** 2002. Modeling 'evo-devo' with RNA. *Bioessays* **24**: 1164–77.
17. **Schuster P.** 2010. Genotypes and phenotypes in the evolution of molecules. In Caetano-Anollés G, ed; *Evolutionary Genomics and Systems Biology*. Hoboken: Wiley-Blackwell. pp. 123–52.
18. **Ancel LW**, **Fontana W.** 2000. Plasticity, evolvability, and modularity in RNA. *J Exp Zool* **288**: 242–83.
19. **Kimura M.** 1987. Molecular evolutionary clock and the neutral theory. *J Mol Evol* **26**: 24–33.
20. **Taverna DM**, **Goldstein RA.** 2002. Why are proteins so robust to site mutations? *J Mol Biol* **315**: 479–84.
21. **Wroe R**, **Bornberg-Bauer E**, **Chan HS.** 2005. Comparing folding codes in simple heteropolymer models of protein evolutionary landscapes: robustness of the superfunnel paradigm. *Biophys J* **88**: 118–31.
22. **Bloom JD**, **Silberg JJ**, **Wilke CO**, *et al.* 2005. Thermodynamic prediction of protein neutrality. *Proc Natl Acad Sci USA* **102**: 606–11.
23. **Bloom JD**, **Raval A**, **Wilke CO.** 2008. Thermodynamic of neutral protein evolution. *Genetics* **175**: 255–66.
24. **Attard P.** 2009. The second entropy: a general theory for non-equilibrium thermodynamics and statistical mechanics. *Annu Rep Prog Chem Sect C* **105**: 63–173.
25. **Cottrell A.** 1979. The natural philosophy of engines. *Contemp Phys* **20**: 1–10.
26. **Niven RK.** 2009. Steady state of a dissipative flow-controlled system and the maximum entropy production principle. *Phys Rev E Stat Nonlin Soft Matter Phys* **80**: 021113.
27. **Kauffman SA.** 1993. *The Origins of Order*. Oxford: Oxford University Press.
28. **Hoelzer GA**, **Smith E**, **Pepper JW.** 2006. On the logical relationship between natural selection and self-organization. *Evol Biol* **19**: 1785–94.

Synthetic Biology Select

Cell

In this issue, *Leading Edge* highlights synthetic biology, a field that is deepening our understanding of the design principles of life and increasing our mastery of biological systems for practical applications in biotechnology and medicine. Reports from the recent literature, discussed in this Synthetic Biology Select, include the rational design of a functional enzyme, an approach for creating more stable proteins, and systems-level analyses that reveal unexpected sophistication in the regulatory networks of microorganisms.



Overlay of the crystal structure of the rationally designed protein (cyan) with the structure predicted by computational modeling (yellow). Image courtesy of Y. Lu.

A How to Guide for Building an Enzyme

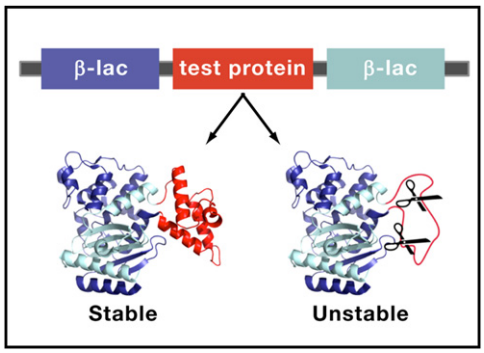
In a forceful demonstration of rational protein design, Yeung et al. (2009) report the transformation of sperm whale myoglobin into a functional nitric oxide reductase. Adding to the difficulty of this feat, crystal structures for native nitric oxide reductases are currently lacking. Hence, the proposed mechanism of catalysis had to be inferred from homology to subunits of heme copper oxidases (such as cytochrome *c* oxidase). The authors' efforts ultimately entailed the introduction of three histidines and one glutamine to create a non-heme binding site for iron, an essential cofactor for reductase activity. This synthetic site is then shown to confer iron binding, and remarkably, the engineered protein acquires the ability to reduce nitric oxide (NO) to nitrous oxide (N_2O). When the designed protein is crystallized, the structure of the rationally designed iron binding site closely matches the structure predicted by computational modeling. These findings show how synthetic enzymes can provide insight into the mechanism of action of enzymes and illustrate the power of in silico modeling for the design of synthetic proteins.

N. Yeung et al. (2009). Nature. Published online November 25, 2009. 10.1038/nature08620.

Selecting for Stability

Most amino acid substitutions decrease a protein's stability. Given this tendency, what tools do protein engineers have at their disposal when enhanced stability is their goal? Recent work by Foit et al. (2009) provides a clever means to select for improved stability through an in vivo screen in the bacterium *Escherichia coli*. To do this, they insert a test protein between the N- and C-terminal halves of TEM1-β-lactamase, an antibiotic resistance protein. These tripartite fusion proteins are then expressed in *E. coli* and grown in the presence of the antibiotic penicillin V, which TEM1-β-lactamase counteracts. Using this system, the authors show for a collection of test proteins that antibiotic resistance correlates with their stability. Taking this a step further into the realm of protein engineering, the authors then test a library of mutants for the protein Im7 (immunity protein 7) and succeed in identifying mutations that improve Im7's stability. However, these amino acid substitutions come with an interesting side effect, in that many disrupt the interaction of Im7 with the toxin colicin, its native interacting partner. Thus, this approach offers a straightforward means to engineer a protein with a desired stability and also reveals evidence of evolutionary tradeoffs between function and stability. For in-depth discussion of the tradeoffs involved in designing a different type of biological system—signal transduction pathways—see the Review in this issue by Christina Kiel, Eva Yus, and Luis Serrano (page 33).

L. Foit et al. (2009). Mol. Cell ***35****, 861–871.*



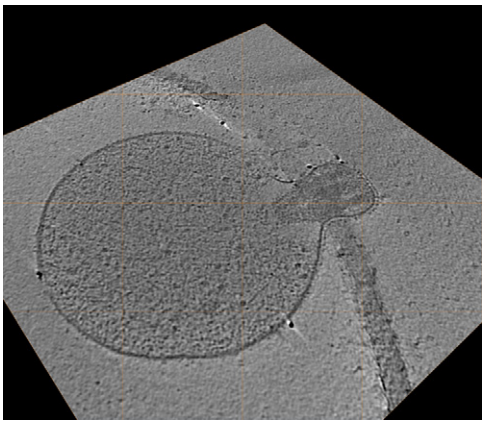
Tripartite fusion proteins in which a test protein is inserted into the antibiotic resistance gene *TEM1-β-lactamase* (*β-lac*) can be used to select for amino acid substitutions that improve protein stability. Figure courtesy of J. Bardwell.

How Yeast Know When It's Time to Grow

It makes intuitive sense that the availability of energy, such as glucose, would be a primary determinant of a cell's growth rate. However, according to a recent examination by Youk and van Oudenaarden (2009), the growth landscape for yeast is substantially more complex than anticipated. In yeast, glucose uptake is modulated by glucose sensors, which promote the expression of six primary hexose transporters (HXT). Youk and van Oudenaarden decouple the

sensing of extracellular glucose from its transport by deleting the endogenous HXT genes. They then replace the genes one at a time and put them under the control of a promoter that is induced by doxycycline but not glucose. Surprisingly, this decoupling reveals that an increase in the concentration of glucose does not invariably stimulate the rate of yeast growth. In some cases, increasing the glucose concentration actually decreases the growth rate. Moreover, the HXT replacement strains all behave differently with respect to the effect of glucose concentration on growth rate. Yet, this is not to say that growth rate is unpredictable. By plotting all of the data together, the authors show that growth rate is determined by two independent variables, the rate of glucose uptake and the extracellular concentration of glucose, and that these variables can be fit into a relatively simple equation that defines a growth landscape. This analysis reveals the functional importance of the interaction between the modules that control glucose import and sensing. By uncovering the underlying design principles controlling growth rate in yeast, this study may inform ongoing efforts to engineer biological systems that respond dynamically to changing environmental conditions. The theme of engineering dynamic control into heterologous pathways is explored in an Essay in this issue by William Holtz and Jay Keasling (page 19).

*H. Youk and A. van Oudenaarden (2009). Nature **462**, 875–879.*



A cryo-electron tomogram of *Mycoplasma pneumoniae*. Image courtesy of A.-C. Gavin.

Dissecting and Reconstructing *Mycoplasma*

Three recent studies (Kühner et al., 2009; Yus et al., 2009; Güell et al., 2009) suggest that the gene, protein, and metabolic networks of *Mycoplasma pneumoniae*, a bacterium with one of the smallest known genomes, display regulatory control previously thought to be limited to eukaryotes. In the first paper, Kühner et al. use mass spectrometry as part of a genome-wide screen to define the protein complexes of *M. pneumoniae*. This effort provides evidence that the organism has ~200 protein complexes that form from among the organism's 689 open reading frames. Building on existing structural data, the authors then map some of the most prevalent proteins and complexes onto cellular electron tomograms to produce a dramatic representation of cellular organization. Yus et al. undertake the reconstruction of the metabolic network of *M. pneumoniae* and use it to design a minimal culture medium containing only 19 ingredients to support growth. Their study further shows that *M. pneumoniae* has a greater capacity to adjust gene expression in response to specific metabolic perturbations than might be anticipated given that it is adapted to a specialized environment and has very few transcription factors compared to more complex bacteria. The transcriptional dynamics of *M. pneumoniae* are examined in-depth by Güell et al. Their study reveals the existence of 341 operons, many of which display production of alternate transcripts under different growth conditions. In addition, antisense transcripts are found in unanticipated large numbers. The sophisticated way in which gene expression is regulated in *M. pneumoniae* indicates the presence of underappreciated mechanisms that control gene expression, which when more completely understood might be used in the engineering of microbial gene networks. (The use of transcriptional feedback mechanisms in the design of cellular memory networks is discussed in this issue in an Essay on page 13 by Devin Burrill and Pamela Silver.) Cumulatively, these ambitious studies add new layers of insight into biological design and will undoubtedly provide food for thought for those undertaking the construction of synthetic biological systems or contemplating the challenges of creating synthetic organisms.

*S. Kühner et al. (2009). Science **326**, 1235–1240.*
*E. Yus et al. (2009). Science **326**, 1263–1268.*
*M. Güell et al. (2009). Science **326**, 1268–1271.*

Robert P. Kruger

Molecular Cell
Previews

Evolving Protein Stability through Genetic Selection

R. Luke Wiseman[2,3,*] and Jeffery W. Kelly[1,3,*]
[1]Department of Chemistry
[2]Department of Chemical Physiology
[3]Department of Molecular and Experimental Medicine
The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
*Correspondence: wiseman@scripps.edu (R.L.W.), jkelly@scripps.edu (J.W.K.)
DOI 10.1016/j.molcel.2009.11.023

In this issue of *Molecular Cell*, Foit et al. (2009) probe cellular protein folding using a split β-lactamase approach for evolving protein stability in the absence of any requirement for function.

Since the discovery that protein folding in vitro is primarily determined by the chemical information within the polypeptide sequence, much has been learned about the details of the physical chemical forces that mediate protein folding in dilute solutions at moderate temperatures. Unfortunately, we know much less about how the kinetic competition between folding, misfolding, and aggregation is managed in the crowded environment of the cell, where the protein concentration is very high, facilitating aggregation. It is now becoming apparent that protein folding in vivo is mediated by interactions between the folding polypeptide chain and the components of the cellular proteostasis network (Balch et al., 2008)—including chaperones, folding enzymes, trafficking factors, and degradation machinery—that together influence the stability, concentration, and localization of proteins in the cell. Due to these dynamic interactions, successful protein folding in a cell is dependent on both the kinetic and thermodynamic stability of the protein of interest and the relative capacity of the proteostasis network under a given set of cellular conditions (Wiseman et al., 2007).

Very little is known about why evolution has selected the protein sequences it has from a vast number of possibilities. The interactions between folding polypeptide chains and the proteostasis network (dependent on these sequences) have complicated our understanding of how protein sequences have evolved to optimize their function for cell survival. Protein function in vivo is dictated both by the inherent activity of the folded protein and its intracellular concentration (influenced by folding energetics). The involvement of the proteostasis network (Balch et al., 2008) in defining the population of folded proteins in vivo has required evolution to efficiently optimize the amino acid sequences to take advantage of the proteostasis network, while maintaining sufficient protein activity for function in vivo.

In this issue of *Molecular Cell*, Foit et al. (2009) have created a very interesting genetic selection-based approach to understand protein stability in the cell. This approach is based on the observation that cleavage of the TEM-1 β-lactamase protein between residues 196 and 197 affords two fragments that are catalytically inactive, unless fused to proteins that interact and thus bring the fragments in proximity, allowing an active β-lactamase enzyme to form. By incorporating a "guest" protein between residues 196 and 197 of β-lactamase (Figure 1A), the extent of cellular guest folding can be quantified by β-lactamase activity, easily discerned by measuring the concentration of a cytotoxic β-lactamase substrate that permits growth. They explored the hypothesis that guest proteins exhibiting stable structures in the periplasmic space of *Escherichia coli* will exhibit maximal β-lactam resistance, while largely unstable proteins would display reduced antibiotic resistance due to degradation of the guest protein by the cellular quality control machinery (Figure 1B). Using over 60 mutants of four structurally distinct guest proteins, they demonstrated a striking correlation between antibiotic resistance and the kinetic and thermodynamic stability of the guest proteins, establishing a robust, generic and quantitative readout of protein stability in vivo. This system complements existing technology to monitor protein folding and stability in vivo, including protein-GFP folding reporters developed by Waldo (Waldo, 2003) and the PROSIDE method developed by Schmid (Sieber et al., 1998), by establishing a quantitative, selectable output to evolve proteins based on their stability in the absence of any requirement for function.

Applying their split β-lactamase approach to explore the potential for evolving protein stability in vivo, Foit et al. (2009) performed error-prone PCR on the gene encoding immunity protein 7 (Im7) and incorporated these constructs into the β-lactamase reporter as the guest protein. Using the resistance to an increasing concentration of the β-lactam penicillin V for selection, they demonstrated the ability to stabilize Im7 through evolving specific mutations in the gene product. The quantitative output (antibiotic resistance) varied over a 7-fold range for a 20 kJ/mol change in stability. Interestingly, the vast majority of mutants that stabilize the protein map to the functional surface of Im7 required for binding its cognate toxin colicin, suggesting that protein evolution has maximized protein stability within the constraint of not disrupting activity. Despite mutations being primarily targeted to the functional face of Im7, a subset of the mutations selected for maintained the function of the protein (colicin inhibition) and demonstrated a significantly higher thermodynamic and kinetic stability, confirming the hypothesis that protein stability can be evolved

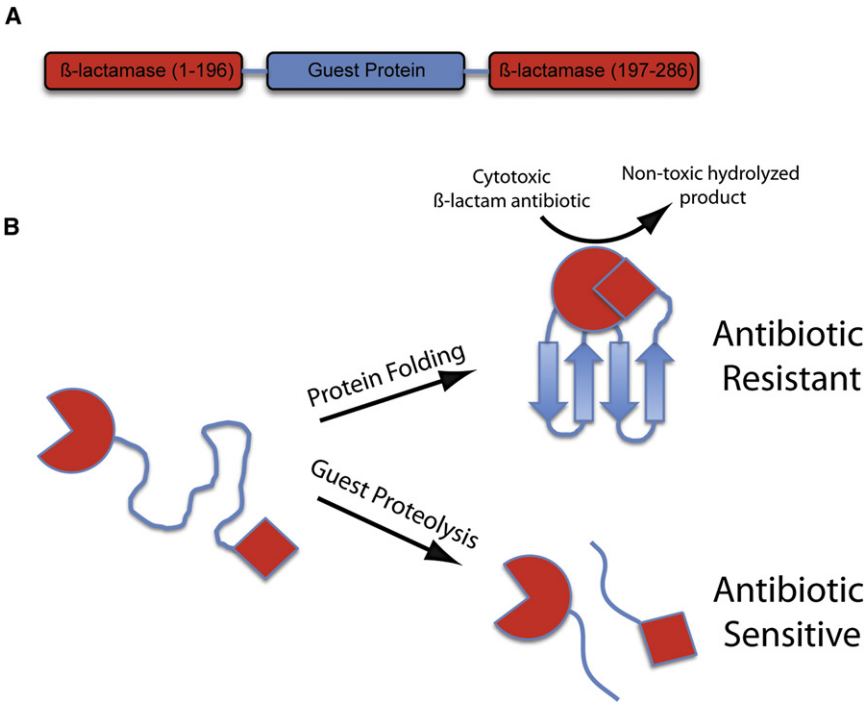


Figure 1. Split β-Lactamase Assay for Monitoring Protein Folding In Vivo

(A) Tripartite construct of a β-lactamase reporter with a guest protein incorporated between residues 196 and 197 of TEM1 β-lactamase.

(B) Mechanism of the β-lactamase folding reporter. Guest proteins able to fold to a stable structure in the periplasm confer antibiotic resistance, while unstable or unfolded sequences are degraded and hence antibiotic resistance is lost by dissociation of β-lactamase into two fragments. The level of antibiotic resistance correlates with the kinetics and thermodynamics of the protein fold, providing a sensitive and quantitative measure of protein folding in vivo.

independent of activity. A variety of control experiments revealed that selection for antibiotic resistance resulting in increased thermodynamic stability also increased the expression level of soluble periplasmic proteins in *E. coli*. This demonstrates that the split β-lactamase approach can be used as a practical tool to enhance protein expression levels, implying that the resulting sequences must harbor the ability to interact with the proteostasis network to maximize cellular protein levels. Many proteins apparently exhibit less than maximal thermodynamic and kinetic stability in vivo, suggesting that dynamic protein folding and unfolding may be required for trafficking, function, and turnover.

The development of genetic selection approaches, such as the Bardwell and Radford split β-lactamase approach described in this issue, has opened the door to exploring many aspects of the role of the proteostasis network in the maintenance and evolution of the proteome. The majority of eukaroytic proteins fail to express in *E. coli*, and we argue that this is partly because prokaryotes lack the sophisticated proteostasis network with which these proteins coevolved in eukaryotes. Using a genetic selection approach, the role of specific components of the mammalian proteostasis network on the folding of a given guest protein can now be evaluated, potentially providing significant insights into their roles in maintaining and/or regulating the cellular proteome. Similarly, the genetic approach to monitor protein folding also provides a sensitive measure of the impact of stress responses and transcriptional programs on the stability and integrity of the proteome, revealing the biological mechanisms for protein maintenance and evolution in response to a given cellular stimulus. By taking advantage of known fluorescent substrates of β-lactamase, this split β-lactamase approach should exhibit utility in mammalian cells and animal models to accurately and sensitively report on the extent of guest protein folding in the complex eukaryotic environment.

The split β-lactamase approach, and related folding sensors under development, should provide the necessary quantification to understand how protein sequence and the proteostasis network coevolved to define the cellular proteome. The ability to monitor protein conformational equilibria in vivo, previously a significant obstacle for protein folding studies in vivo, is now possible thanks to the efforts of the Bardwell and Radford groups, and this approach has the potential to answer many of the important biological questions related to the mechanisms by which proteins fold and evolve to optimize their function in a given cellular environment.

REFERENCES

Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Science *319*, 916–919.

Foit, L., Morgan, G.J., Kern, M.J., Steimer, L.R., von Hacht, A.A., Titchmarsh, J., Warriner, S.L., Radford, S.E., and Bardwell, J.C.A. (2009). Mol. Cell *36*, this issue, 861–871.

Sieber, V., Pluckthun, A., and Schmid, F.X. (1998). Nat. Biotechnol. *16*, 955–960.

Waldo, G.S. (2003). Curr. Opin. Chem. Biol. *7*, 33–38.

Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., and Balch, W.E. (2007). Cell *131*, 809–821.

PROTEIN FOLDING

Thermodynamic compromise

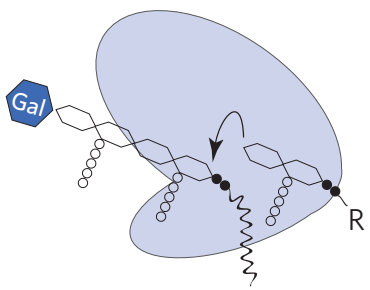
Mol. Cell **36**, 861–871 (2009)

A full understanding of the factors that govern protein stability *in vivo* is incomplete since studies of the kinetics and thermodynamics of protein folding have almost exclusively come from *in vitro* studies of dilute, pure proteins. Foit *et al.* have now developed a system to evolve protein stability *in vivo* by fusing a model protein between the two halves of the antibiotic resistance enzyme β-lactamase. Among known variants of the Im7 test protein that alter its stability *in vitro*, there was a strong correlation between thermodynamic stability and the minimal inhibitory concentration when expressed *in vivo*. By selecting for enhanced antibiotic resistance in a library of mutagenized Im7, the authors identified 13 residues whose mutation caused increased stability when tested *in vitro* (shown as spheres on the figure). Increased stability for Im7 variants correlated with decreased sensitivity towards protease digestion, an increase in steady state expression levels and increased kinetic stability, a measure related to protein unfolding rates. Functionally, a majority of the stabilized Im7 mutations (shown as blue spheres) map to the interface of Im7 binding with its cognate toxin, colicin E7. Some mutations reduced binding, but none abolished it. These results highlight a balance between functional, kinetic and thermodynamic properties and help to explain why protein evolution does not necessarily favor the most stable fold. *MB*

GLYCOBIOLOGY

Probing polymerization

J. Am. Chem. Soc., published online 17 December 2009, doi:10.1021/ja909325m



DEBORAH PERLSTEIN

Peptidoglycan glycosyltransferases create the bacterial cell wall in a processive reaction using glycolipids in the membrane. In particular, these enzymes catalyze the transfer of a disaccharide unit from a glycolipid in the donor site to a growing polymer chain in the acceptor site, followed by translocation of the chain to the donor site and a return to the beginning of the reaction. Previous work using chemically synthesized substrate analogs had shown that variations in the conformation, saturation and length of the lipid chain prevent homopolymerization, indicating that one or both of the donor and acceptor sites have strict substrate specificity. To investigate the unique properties of the two sites, Perlstein *et al.* used a second modified substrate, terminating in a galactose residue, which could serve as a glycosyl donor but not an acceptor. By testing the original analogs in the presence of the galactose construct, the authors were able to identify whether the acceptor site could accommodate the lipid modifications. Surprisingly, all of the analogs tested were converted to product, demonstrating that the donor site is responsible for the strict specificity observed. Though it remains to be seen how exactly the lipid tail impacts enzyme function—affecting initial binding versus translocation, for example—these results demonstrate an elegant strategy for studying a repetitive reaction and provide new insights into peptidoglycan biosynthesis. *CG*

EPIGENETICS

Methylation coordination

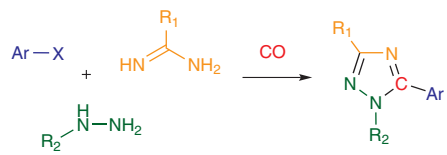
Nat. Struct. Mol. Biol., published online 20 December 2009, doi:10.1038/nsmb.1753

The reversible methylation of lysine residues within histones is a central regulatory mechanism for eukaryotic gene expression. Because histones often contain activating marks, such as Lys4-trimethylated histone 3 (H3K4me3), along with repressive marks such as H3K9me2 or H3K27me2, it remains unclear how histone-modifying enzymes manage these conflicting signals. Horton *et al.* now suggest that certain histone lysine demethylases target different dimethylated lysine residues by virtue of their structural architectures. PHF8 and KIAA1718 are two related demethylases that contain an N-terminal plant homeodomain (PHD) that binds to H3K4me3 sites and a jumonji demethylase domain that removes downstream dimethyllysine marks. H3K9me2 is a poor substrate for PHF8, but the addition of an upstream H3K4me3 modification enhances the ability of PHF8 to remove K3K9me2 marks by 12-fold. In contrast, KIAA1718 is active on H3K9me2 alone, but addition of a H3K4me3 mark abrogates demethylase activity. Structural analyses of PHF8 and KIAA1718 showed that although their PHD and jumonji domains adopt virtually identical structures, the relative three-dimensional relationship of these two domains was strikingly different: PHF8 bends to bring the PHD and jumonji sites in proximity, while KIAA1718 adopts a more extended conformation that enables it to target the more distant H3K27me2 modification. In addition to providing new insights into the 'histone code', the study opens the possibility that histone-modifying enzymes could be rationally engineered. *TLS*

CHEMICAL SYNTHESIS

Triazole in four parts

Angew. Chem. Int. Ed., published online 8 December 2009, doi:10.1002/anie.200905897



Triazoles, which are five-membered aromatic heterocycles containing two carbon and three nitrogen atoms, are commonly found in bioactive molecules. Access to highly substituted 1,2,4-triazoles from aryl halides typically requires complex multistep synthesis involving transition metal–catalyzed C-C coupling that proceeds via unstable metallo-containing intermediates. Staben and Blaquiere now report an alternative synthetic approach to trisubstituted 1,2,4-triazoles via Pd-catalyzed C-C-N coupling to form carbonyl precursors that cyclize to yield the triazole products. The modular synthetic approach involves the initial carbonylative coupling of amidines to aryl or heteroaryl halides, followed by *in situ* reaction with monosubstituted hydrazines. This four-component, one-pot reaction proceeded in good yield with a wide range of commercially available reactants. A variety of aryl and heteroaryl halides as well as alkyl, aryl and heteroaryl amidines and hydrazines were well tolerated. The pharmaceutical relevance of this synthetic approach was demonstrated by the synthesis of deferasirox, a metal chelator that is approved for the treatment of iron-overload disease. This method provides a new, facile approach for accessing a broad range of analogs of this important scaffold. *JK*

Written by Mirella Bucci, Catherine Goodman, Joanne Kotz & Terry L. Sheppard