Annual Review of Biochemistry

Principles of Protein Stability and Their Application in Computational Design

Adi Goldenzweig and Sarel Fleishman

Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel; email: sarel@weizmann.ac.il

Annu. Rev. Biochem. 2018. 87:8.1–8.25

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-062917-012102

Keywords

computational protein design, Rosetta, proteostasis, chaperones, bioinformatics, PROSS

Abstract

Proteins are increasingly used in basic and applied biomedical research. Many proteins, however, are only marginally stable and can be expressed in limited amounts, thus hampering research and applications. Research has revealed the thermodynamic, cellular, and evolutionary principles and mechanisms that underlie marginal stability. With this growing understanding, computational stability design methods have advanced over the past two decades starting from methods that selectively addressed only some aspects of marginal stability. Current methods are more general and, by combining phylogenetic analysis with atomistic design, have shown drastic improvements in solubility, thermal stability, and aggregation resistance while maintaining the protein's primary molecular activity. Stability design is opening the way to rational engineering of improved enzymes, therapeutics, and vaccines and to the application of protein design methodology to large proteins and molecular activities that have proven challenging in the past.

Contents

1. INTRODUCTION

Proteins play essential roles in all biological processes, functioning as exquisite catalysts, inhibitors, and sensors. To perform their function, most proteins need to fold into a single well-defined structure, the native state, and remain stably folded for an extended period of time. To achieve this, the native-state free energy must be lower than for unfolded or misfolded states. The extent to which the native state is more favorable than all competing states is broadly defined as protein stability. Protein stability has a wide range of practical manifestations in nature as well as in laboratory settings, including resistance to high temperature, denaturant, proteases, and nonphysiological pH; furthermore, in vivo yields of functional protein, solubility, and proper cellular or extracellular localization are often correlated with protein stability.

Natural proteins are often only marginally stable, with free-energy differences between the native state and the unfolded or misfolded states as low as 5 kcal/mol (1, 2). Since marginal stability often results in sensitivity to mutations or changes in the environment, it has broad implications for understanding constraints on protein evolution (3–5), the underlying causes of disease, and for engineering improved variants for research, biotechnology, and medicine. For instance, mutations that cause human disease often do not alter molecular function but rather decrease protein stability to the point that too little of the protein is available to perform its function (6–8); in an illuminating example, cancer-associated mutations in p53 often destabilize the protein, such that less than half of the expressed protein is properly folded (7). In other cases, disease-causing mutations may cause misfolding or lead to potentially cytotoxic aggregation and fibril formation (9, 10).

All organisms have evolved complex mechanisms that address the potential problems that arise from marginal protein stability; collectively, these mechanisms are known as the proteostasis network (PN). The PN is essential for many proteins to fold and function efficiently under both normal and stress conditions. Furthermore, cellular aging has been linked to a decrease in the functions of the PN that leads to the gradual formation of cytotoxic amyloids (10–17).

These tremendous implications for every aspect of biomedical research have led to an ongoing effort to decode the underlying principles that govern protein stability and its many manifestations.

In broad terms, protein stability is probed in two complementary directions: first, protein kinetic or thermodynamic stability as measured in vitro in the face of stress, such as denaturant, protease, or elevated temperature; and second, in vivo stability, or the efficiency of protein folding to the native state and maintenance in that state in living cells. Insights from both fields were gradually incorporated into methods for stability design over the last two decades. Early in the evolution of stability design, the intricacies involved in predicting mutational effects on stability suggested that the ultimate goal of stability design methods that are universally applicable might be unattainable (18); nevertheless, deeper understanding of the molecular and cellular determinants of stability together with substantial improvement in prediction accuracy have led to steady improvement in design methodology to the point of fully automated methods applicable to a wide range of protein stability problems. This review focuses on the major insights into the question of protein stability in vitro (Section 2) and in vivo (Section 3) and how these insights have laid the groundwork for the development of practical and general methods for computational stability design (Section 4).

From our perspective, perhaps the most exciting application of computational stability design is for design of new molecular activities. Notably, de novo designed proteins, which are small, rich in secondary structure, and devoid of molecular activity, can be very stable (19–22), yet newly designed enzymes or binders have typically shown low stability and expression levels (23, 24). Indeed, our own interest in stability design emerged from our observations that antibodies designed using standard computational strategies showed unacceptably low stability and in vivo expression levels (25). In the past, such bottlenecks were overcome by iterative experimental application of genetic randomization and selection for improved variants (23); with recent methodological improvements, these critical bottlenecks can now be addressed algorithmically. Therefore, in addition to enhancing stability in natural proteins, stability design may open the way to rational design of activities not seen in nature.

2. THERMODYNAMICS OF PROTEIN STABILITY

The free energy of unfolding of many proteins is strikingly small ($\Delta G_{\text{unfolding}} = 5$–$10$ kcal/mol) (1, 2), comparable to the contribution from only a few hydrogen bonds. This marginal thermodynamic stability has wide ramifications, because changes in the protein's environment or elimination by mutation of only a few interactions out of the thousands observed in the native state could tip the balance and turn an active protein into a nonfunctional, misfolded, or aggregated form. We start by asking why many natural proteins are only marginally stable and analyze the forces that stabilize the native state, as well as the design principles that disfavor non-native, misfolded, or aggregated forms. For historical perspectives on protein folding and stability, see References 1, 26, 27.

2.1. Why Are Many Proteins Marginally Stable?

Four complementary considerations can explain why many natural proteins are marginally stable:

1. Some molecular activities, such as binding and catalysis, necessitate compromising structural features. Enzyme active sites, for instance, comprise multiple uncompensated charges that are necessary for stabilizing the transition state (28). Accordingly, mutations in catalytic residues can stabilize the protein though they lower its activity (29, 30); similarly, protein-binding sites often contain hydrophobic, solvent-exposed surfaces that are essential to forming tight interactions with their targets (31). In fact, a possible explanation for why proteins are so large compared with the number of amino acids directly participating in

molecular activity is that large gains in folding free energy from other parts of the protein are necessary for the active site to form stably (32). In addition, some molecular activities require proteins to reversibly fold and unfold—for instance, for efficient transport across cell membranes (33) or to sense mechanical stress (34)—and high native-state stability may lock such proteins into one conformation. Similarly, conformational change, as seen for instance in allosteric communication between effector binding sites and active sites, may demand flexibility, while excessive stability may rigidify proteins (35, 36).

2. The unfolded and misfolded states of a protein outnumber the folded state by many orders of magnitude. Folding therefore carries a tremendous entropic penalty, estimated to be on the order of 70 kcal/mol for a 100-residue protein (37–39). The molecular contacts observed in the native state must compensate for this large penalty, but as discussed in Section 2.2, the net contribution of individual contacts may be quite small, thus requiring many stabilizing contacts to overcome the entropic penalty.
3. Marginal stability may regulate protein abundance by increasing cellular turnover rates (40, 41); excessively stable proteins may resist degradation and clearance by proteases. Marginal stability, according to this view, is adaptive, allowing the cell to recycle proteins rapidly.
4. As proteins mutate through evolution, selection pressures optimize stability but only up to the point at which no more gains in organism fitness are made. A protein that expresses to sufficient levels and does not form toxic misfolded or aggregated species is therefore unlikely to experience positive selection pressures to increase stability (**Figure 1**). Marginal stability may therefore reflect a balance between the accumulation of mutations through genetic drift and the necessity to maintain a high enough fraction of folded protein (42, 43).

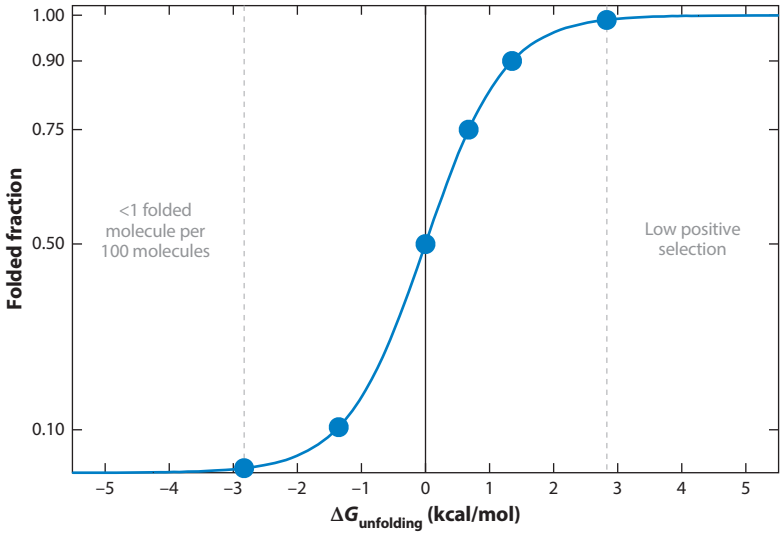

Figure 1

The thermodynamic basis for marginal stability in proteins. Many natural proteins are marginally stable owing to low selection pressures for higher stability. The x-axis shows the free-energy difference between misfolded or unfolded states and the folded state. The fraction of folded protein as a function of this free-energy difference shows a steep sigmoidal relationship, where at $\Delta G \sim 3$ kcal/mol approximately 99% of the protein is folded (*dashed horizontal line* on the right-hand side). Above this threshold, selection pressures for increasing the free-energy difference may be quite low.

The first three above considerations imply that marginal stability is an inevitable requirement of the chemistry of the polypeptide chain and molecular function, and of the need to regulate cellular protein levels. These are all important considerations in different contexts, but it is important to note that proteins from thermophiles can be much more stable than their mesophilic homologs, withstanding both high temperatures and denaturant concentrations (44–46). Furthermore, computer-designed proteins are often exceedingly stable (19, 21, 22). These observations suggest that marginal stability is often not a functional or physical necessity. It is therefore the fourth consideration—that evolutionary selection pressures weaken beyond a certain stability threshold—that dominates. This is an important conclusion in the context of stability design, as it suggests that in the majority of cases, the stability of natural proteins leaves substantial room for improvement without risk of impairing the protein's primary activity, as we describe in Section 4.

Energy function: a function that sums contributions to molecular energy, typically comprising van der Waals, electrostatics, hydrogen bonding, and solvation and may include statistical terms derived from experimental data

2.2. Thermodynamic Contributions to Protein Stability

Design of stable proteins often relies on quantitative modeling of the forces that stabilize biological molecules and of the effects of mutations on stability and function. Most design algorithms evaluate structure models using an energy function, which sums the dominant contributions to native-state energy. In Section 2.2.1, we briefly describe the four types of noncovalent interactions that are primarily responsible for the stability of the native state: the hydrophobic effect, van der Waals interactions, hydrogen bonding, and electrostatics (**Figure 2**) (47). A crucial complementary consideration, known as negative design, concerns design principles that do not necessarily

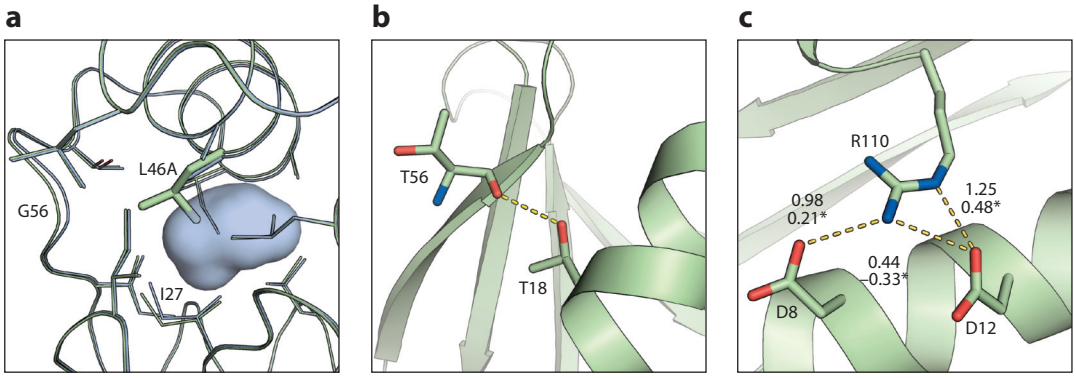

Figure 2

Proteins comprise thousands of individually weak interactions. (*a*) The tight packing of protein cores is driven by van der Waals interactions and the hydrophobic effect, and any cavity leads to a substantial penalty. Mutation Leu46Ala in the enzyme T4 lysozyme (PDB entries: 1L63 for wild type, *green*; 1L67 for Leu46Ala, *light blue*) induces cavity formation (*light blue*) and destabilizes the protein by −2.7 kcal/mol (50). The lysozyme backbone and some side-chain atoms adjust to minimize the cavity (backbone adjustment around Gly56 and the Ile27 Cδ1 displacement). (*b*) Thr18 in RNase T1 (PDB entry: 1RGG) forms a buried hydrogen bond (*dashed yellow line*) with the Thr56 backbone carbonyl. Mutating Thr18 to the isosteric amino acid Val penalizes stability by −1.4 kcal/mol (146), likely due to loss of the hydrogen bond, the burial of an unsatisfied donor (Thr56 carbonyl), and clashes or backbone changes to accommodate the methyl group in Val18, which is larger compared with the hydroxyl group in Thr18. (*c*) A charged triad on the surface of barnase demonstrates cooperativity and competition in protein interactions (PDB entry: 1BRN). Asp8 and Asp12 on the N-terminal helix form one and two charged hydrogen bonds, respectively, with the C-terminal Arg110 (*dashed yellow lines*). For every pair in the triad, two values are shown in kcal/mol units (55): (*i*) the pairwise interaction contribution measured against the wild-type background; and (*ii*) the interaction against a background in which the third residue in the triad is mutated to Ala (*). Positive values indicate stabilizing pairwise interactions. Owing to cooperativity, the pairwise contributions differ depending on whether the third position is mutated, and a stabilizing mutation in one context may be destabilizing in another. Abbreviation: PDB, Protein Data Bank.

contribute to the thermodynamic stability of the native state per se but contribute to the destabilization of unwanted, unfolded, misfolded, or aggregated states. Negative design is the subject of Section 2.2.2.

2.2.1. The dominant interactions contributing to native-state stability. The hydrophobic effect, that is, the preference of apolar amino acids to be sequestered from water, provides the dominant driving force for folding and for stability in all proteins (48). Furthermore, in well-packed protein cores, the hydrophobic amino acid residues form favorable van der Waals contacts (**Figure 2*a***), whereas polar and charged groups are primarily found on the protein surface. An obvious exception to this neat division is the burial in the protein core of backbone polar groups, which typically form regular hydrogen bonds within α helices and β sheets. Owing to their regularity and optimal hydrogen-bonding configuration, these interactions provide another important driving force for folding (26, 49).

Protein-engineering studies have yielded a quantitative understanding of hydrophobic core interactions and their effect on protein structure. The effects of mutations from large, well-packed residues in the protein core to alanine can be approximated by a combination of the losses of van der Waals packing and the hydrophobic effect upon folding; a single such mutation can result in a penalty of approximately 2 kcal/mol (**Figure 2*a***) (1, 26 , 50). These studies furthermore showed that the protein structure may adapt to such mutations by contracting around the cavity, provided that the surrounding backbone is flexible; rigid parts of the protein are therefore more sensitive to destabilizing mutations than flexible parts (50). Furthermore, to maximize the gains from van der Waals interactions, protein cores are almost as tightly packed as allowed by the excluded volume of the constituent atoms (51).

A hydrogen bond involves a hydrogen donor and acceptor of positive and negative partial charge, respectively (**Figure 2*b***). Since hydrogen bonds form at very close separation and are sensitive to the orientation of the donor and acceptor groups, the formation of hydrogen bonds leads to an entropic penalty and the loss of favorable contacts between the polar groups and water in the unfolded state. The net contribution of hydrogen bonds to stability is therefore 0.5–2 kcal/mol, depending on the bond orientation, whether it forms between charged or uncharged groups, and whether it is part of a network of polar interactions (1, 26, 52). Although the net contribution from hydrogen bonds is small, they are essential for specifying the native conformation, because they require precise geometry to maximize their energy contribution (53).

Salt bridges between two opposite charges (usually at a distance $\leq$5 Å) also contribute to thermodynamic stability but in most cases to a smaller extent (**Figure 2*c***) (54). Most salt bridges observed in proteins are surface exposed and typically contribute <1 kcal/mol to stability (52, 55, 56). Buried salt bridges may have large stabilizing effects of as much as 4–5 kcal/mol (57, 58), owing to the strength of Coulomb attraction in the low-dielectric protein core. Yet, buried salt bridges are rarely observed, as their formation is accompanied by a large penalty for transferring charged groups from water to a nonpolar environment that is not always fully compensated by stabilizing Coulomb interactions (26, 59). Indeed, where buried salt bridges occur, they are often part of a network of polar interactions that may provide additional stabilization (57). Similar to buried hydrogen bonds, buried salt bridges also have an important role in specifying the native conformation, since misfolded states, in which the charged residues are buried but the salt bridge is not optimized, may incur large penalties. Interestingly, salt bridges contribute more to stability at high temperatures, explaining why salt bridges are more common in proteins from thermophilic organisms (see sidebar titled Salt Bridges Contribute More to Native-State Stability at High Temperatures).

SALT BRIDGES CONTRIBUTE MORE TO NATIVE-STATE STABILITY AT HIGH TEMPERATURES

Proteins from hyperthermophilic organisms exhibit a variety of molecular features that explain their high thermal stability relative to their homologs from mesophilic organisms. While features vary among families, a higher propensity of salt bridges emerges as a general mechanism across families (44). This finding is intriguing, given that in mesophiles, salt bridges contribute relatively little to stability, since large desolvation penalties disfavor their formation. At elevated temperatures, however, favorable water–charge interactions (solvation) in the unfolded state weaken, and therefore water competes less with the formation of native-state salt bridges; by contrast, Coulomb attraction is relatively unaffected by temperature, leading to a larger net contribution to native-state stability (59). From a design standpoint, mutations that eliminate or create salt bridges may therefore have different effects on folding free energy depending on the temperature.

Coulomb interactions also play a role in interactions involving aromatic side chains. The aromatic ring core bears a partial negative charge, whereas the surrounding hydrogen atoms are partially positive. This charge distribution favors stacking configurations among aromatic residues (60) as well as interactions between positively charged Lys and Arg side chains and the ring core (61). These interactions make a small contribution to stability (~0.5 kcal/mol) (62).

Individually, the noncovalent interactions described above make small contributions to folding free energy, and the aggregate of thousands of interactions is needed to offset the large loss in entropy upon folding (47). It should, however, be noted that the relative contribution of a given interaction may strongly depend on structural context, and interactions that contribute little to stability as a pair may become more dominant within a cooperative network of interacting groups. For instance, hydrogen bonding carries an entropic penalty due to its precise geometric requirements. Within an organized network of polar interactions, however, the entropic cost of freezing each group is paid once, though each such group may participate in multiple hydrogen bonds within a network (**Figure 2*c***) (55). Thus, protein cooperativity is another important principle for stabilizing the native state and also explains why mutational effects on stability strongly depend on the molecular context.

2.2.2. Negative-design principles that destabilize misfolded and aggregated states. The features that stabilize the native state are collectively known as positive-design elements. Conversely, features that destabilize misfolded, aggregated, or unfolded states are known as negative-design elements and play an essential role in specifying the native state (63). In considering the many conformational degrees of freedom of any protein, it becomes obvious that non-native or misfolded states may outnumber the native state by orders of magnitude. If the native state is energetically similar to the misfolded states, the latter might outcompete the former, leading to what is known as frustrated folding, in which the protein might be trapped in stable folding intermediates instead of the native state (35, 64, 65). Accordingly, native-state stability is a function of the free-energy difference not only between the native and unfolded states but also between the native and misfolded states. To counter misfolding, proteins have evolved negative-design elements—features that do not necessarily contribute to the energy of the folded state but rather destabilize undesired states.

On first consideration, it may appear that negative-design elements would be impossible to predict and therefore to model, since misfolded, unfolded, and aggregated states are numerous and their structures are largely unknown (65). Nevertheless, some of the negative-design principles

are surprisingly simple and are related to hydrophobicity, charge, and secondary structure (39). For instance, the first description of negative design came from studying the molecular causes of sickle cell anemia (9). In this severe genetic disease, a point mutation replaces a negatively charged Glu with a hydrophobic Val on hemoglobin. Although this mutation does not appreciably destabilize hemoglobin or reduce its affinity for oxygen, the replacement of a charged surface with a hydrophobic one induces fibril formation and deformation of the red blood cell.

A second example of negative design highlights the role of secondary structure. β-strand–containing proteins are especially prone to aggregation, because this backbone conformation can form nonspecific interactions with other strands within the molecule or with other molecules. To counter this problem, β-strands at the edge of sheets often contain Pro residues or charges that would disrupt nonspecific pairing (66). Furthermore, alternating polar/hydrophobic sequence stretches are likely to form β-strands, which may drive amyloid formation and are therefore depleted in natural sequences (67). Finally, mutational effects on aggregation can be approximated by a simple combination of change in charge, hydrophobicity, and the propensity to convert from an α-helical to a β-strand conformation (68). Thus, despite the complexities of misfolding and aggregation pathways (17), simple principles that do not require detailed atomistic modeling can be used to analyze and potentially mitigate these undesired outcomes.

2.3. Biomolecular Energy Functions Model Energy Contributions to Native-State Stability

All atomistic design software packages rely on energy functions to compute the relative energies of biomolecular states and rank the effects of mutations on stability (69–73). The complexities of accurate energy calculations are partly due to the fact that the noncovalent bonds that stabilize the native state are individually weak and their strengths depend on the molecular context, including whether they are exposed to solvent and whether they participate in cooperative networks. Nevertheless, some contributions to native-state stability, including van der Waals packing and hydrogen bonds, can be computed accurately and have allowed the design of new proteins to atomic accuracy (19, 21, 22, 74). Contributions to folding free energy from conformational entropy, however, require intensive sampling of non-native conformations and typically cannot be taken into account in energy functions. Furthermore, many-body interactions, such as between polar groups and water (75) and cooperative polar interaction networks, are difficult to accurately model, and some of these may require the use of specialized search heuristics to design correctly (20, 24). Therefore, although the rules of protein stability are broadly known, their accurate representation is the subject of ongoing research (76). In the context of design of functional proteins, the requirements of molecular activity, including the formation of preorganized cavities for catalysis or hydrophobic surface patches for binding, may exacerbate the problem of misfolding and aggregation (39). As we see in Section 4, stabilizing natural proteins while maintaining their primary activity therefore requires consideration of both the positive-design elements that improve the native-state energy and the negative-design elements that block misfolding and aggregation.

In conclusion, from a thermodynamic perspective, the native state is determined by many weak atomic contacts that together only marginally overcome the large entropic loss of conformational degrees of freedom in the unfolded state. It is furthermore clear that any strain in the native state, such as core cavities, unsatisfied hydrogen bonds, or salt bridges, can carry a penalty in native-state stability. Biomolecular energy functions capture the dominant contributions to native-state stability, but stability is additionally determined by the free-energy difference between the native and misfolded states. In Section 4, we see that by eliminating strain in the native state, replacing it with favorable molecular interactions, and considering the negative-design principles described

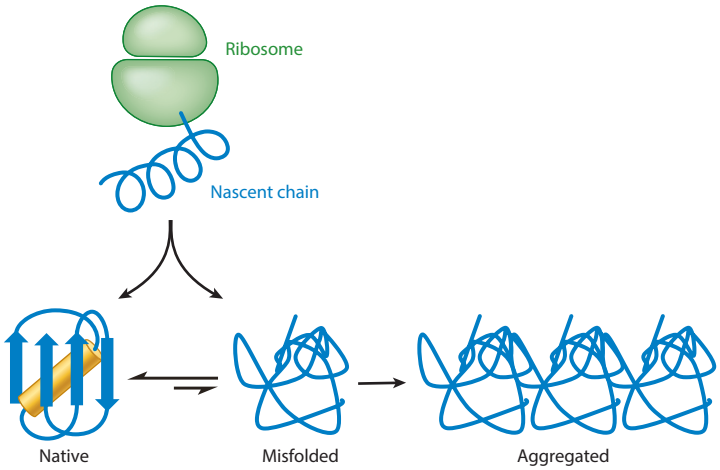

Figure 3

The possible fates of the nascent protein chain in vivo. Protein synthesis by the ribosome is typically much slower than the folding rate. Therefore, the nascent chain may adopt secondary structure and some tertiary contacts as it emerges from the ribosome exit tunnel. Hydrophobic and uncharged surfaces may form transient non-native contacts leading to misfolding and terminal aggregation. The protein may also fold into the native state, although the native state too may misfold and aggregate, a process that may be accelerated under stress—for instance, owing to high temperature, denaturant, or altered pH. In the cell, chaperones may interact with any of these states, blocking misfolding and aggregation and maintaining the protein in the native state.

above, large gains in stability can be made. However, since proteins are typically produced by dedicated and elaborate cell machinery, we shall first discuss how this machinery blocks misfolding and aggregation, thereby contributing to another important determinant of protein stability—in vivo expression of natively folded protein.

3. PROTEIN EXPRESSION IN LIVING CELLS

Although many proteins are only marginally stable and prone to misfolding, within their natural hosts, the majority fold efficiently and are stable in the folded state. Efficient folding of marginally stable proteins is enabled by dedicated machinery, the PN, that limits protein misfolding and aggregation (**Figure 3**). In non-natural contexts, however, such as in the overexpression of foreign proteins, the host PN may be overwhelmed, limiting expressibility—that is, the yields of natively folded and functional protein. One of the principal goals of stability design is therefore to design expressible variants even in the absence of the protein's native PN. In this section, we therefore briefly review how the PN limits misfolding and aggregation, what molecular features trigger PN involvement in folding, and whether some of these features could be eliminated by design to lower the protein's dependence on the PN for efficient folding. For recent reviews of the PN's involvement in health and disease, we refer the reader to References 77–80.

Anfinsen's (81) landmark experiments in the 1950s and 1960s showed that following complete chemical denaturation, the small model protein ribonuclease (124 amino acids) spontaneously refolds to its native functional state without additional cellular components. This demonstration led to the thermodynamic hypothesis—perhaps the most fundamental and influential concept in protein chemistry—which argues that all the information for folding is encoded in the primary protein sequence. Similarly, many other small proteins were shown to refold spontaneously from

the unfolded state to the folded state with no stable intermediates between these two states (82); therefore, these proteins were termed two-state folders.

Two-state folders: proteins for which only two states, folded and unfolded, are observed and that have no stable intermediates

Although many general insights that have withstood the test of time emerged from studying small two-state folders (64, 83), it must be kept in mind that large proteins comprise a major fraction of every proteome, with the median size of proteins being 300–400 amino acids (84). The large size of most proteins immensely complicates the theoretical and practical considerations related to folding compared with the situation with small two-state folders. Specifically, proteins with >100 amino acids typically fold via stable intermediates; thus, large proteins are typically not two-state folders (85). Stable folding intermediates have local structures that may lead to misfolded or aggregated states, contributing to marginal stability (**Figure 3**) (83, 86).

In addition to these considerations, in the context of any living cell, protein folding occurs in a crowded environment (17). Indeed, the cytosol can be thought of as a highly concentrated solution of roughly 300–400 g of protein per liter cytosol (87), far exceeding concentrations in typical in vitro experiments and posing a further risk of nonspecific interactions, misfolding, and aggregation. Misfolding and aggregation are problems not only owing to loss of the target protein; in fact, owing to their persistence, protein aggregates may have gain-of-function pathological effects, especially in long-living cells, such as neurons (80, 86, 88).

3.1. A Protein Quality Control Network Ensures Robust Folding in All Organisms

Given the complex considerations related to the folding of most cellular proteins, it is not surprising that in all organisms proteome integrity is maintained by a variety of mechanisms, collectively known as the PN (77, 80, 89). It has also been noted that more complex proteomes, such as those in mammals, have more elaborate PNs with more layers of control than simpler proteomes (78, 79). The PN includes components that regulate all steps in the protein life cycle, starting from just outside the ribosomal peptide exit tunnel, as the nascent chain is synthesized. Regulation continues through assistance to proper folding, maintenance of the folded state, extraction of protein chains from aggregates, and finally, proteolysis of unneeded, terminally misfolded, or aggregated proteins (**Figure 3**).

An elaborate PN, as seen in multicellular organisms, also provides an essential buffer for proteome maintenance in the face of environmental stress (90). The PN also provides a necessary buffer for mutations that may destabilize the protein, either for adaptive reasons, since sequence features that are necessary for activity are often destabilizing (4, 29), or accumulated through genetic drift. The PN may therefore be partly responsible for the marginal stability of many natural proteins, as it weakens evolutionary selection pressures for independent folding (see sidebar titled Variability in Chaperone Dependence Among Homologous Proteins). Dependence on the source organism's PN is seen most clearly in the heterologous overexpression of proteins. For instance, the overexpression of mammalian proteins in microbial hosts often results in low or even no detectable yields of functional protein, a situation that can be partly alleviated by simultaneous overexpression of bacterial chaperones (91), or alternatively, by production in specialized cell cultures derived from multicellular organisms with elaborate PNs.

3.1.1. Secondary structure elements form already during protein synthesis. Even prior to engagement with any PN component, protein folding in vivo is subject to different constraints than in vitro refolding. Protein synthesis by the ribosome occurs at a rate of ∼20 amino acids per second in prokaryotes and at a somewhat slower rate in eukaryotes (5–9 amino acids per second) (92, 93). Both rates are strikingly slow compared with the typical folding rates of many

VARIABILITY IN CHAPERONE DEPENDENCE AMONG HOMOLOGOUS PROTEINS

The molecular determinants of chaperone dependence are still the subject of intense investigation. In several cases, chaperone-dependent proteins have chaperone-independent homologs in other organisms (142, 143). One example is *S*-adenosylmethionine synthase (MetK). While *Escherichia coli* MetK (*Ec*MetK) is a GroEL/GroES obligate substrate, its *Ureaplasma urealyticum* ortholog (*Uu*MetK), sharing 45% sequence identity with *Ec*MetK, is not (143). Moreover, GroEL-dependent *Uu*MetK mutants were obtained by random mutagenesis, including by introducing only one or two mutations, and those were seen to result in higher aggregation (144). Similarly, when mouse dihydrofolate reductase (DHFR) is expressed recombinantly in *E. coli*, it interacts strongly with GroEL, while *E. coli* DHFR does not (145). The two DHFRs differ primarily in three surface loops, and grafting two of these loops from mouse into *E. coli* DHFR turns the latter into a GroEL-dependent variant. Many folds, including, for instance, TIM barrels, have chaperone-dependent representatives and chaperone-independent ones within the same organism, although they share the same general folding constraints. These and other findings imply that chaperone dependence is not an intrinsic requirement of a fold or a molecular function; instead, it is likely acquired by random drift under weak selection pressure for chaperone independence.

protein domains, which only require micro- to milliseconds to fold (82). Given that peptide-chain synthesis is the rate-limiting step in the formation of the folded state, it is clear that in vivo protein folding may start cotranslationally (77, 94–96). Thus, in cotranslational folding, transient interactions involving the nascent chain, the ribosome, and proteins surrounding the ribosomal peptide exit tunnel impact a protein's folding trajectory—a situation altogether different from in vitro refolding experiments, in which the entire protein chain is simultaneously available for forming native interactions.

The ribosome architecture also modulates the folding trajectory. The ribosomal peptide exit tunnel, spanning from the peptidyltransferase center to the peptide exit pore, is ∼100 Å long, suitable for an extended protein chain of ∼30 amino acids (94, 97). The tunnel diameter varies and reaches 20 Å in parts, sufficiently wide to accommodate the formation of α-helices and even small tertiary structures already within the tunnel (98, 99). Furthermore, the exit tunnel lining is negatively charged, thereby inducing partial hydrophobic collapse and promoting formation of secondary structural elements in this secluded environment (100–104). Finally, as the nascent chain emerges from the exit tunnel, the ribosome surface and its associated chaperones maintain the nascent chain in a partly unfolded state, comprising mostly secondary structure elements and almost lacking tertiary structure, thereby disfavoring local non-native contacts that may lead to misfolding or aggregation (77, 80). It is clear from these considerations that the ribosome is an active player in the initial steps of secondary and local tertiary structure formation of the nascent chain.

From the standpoint of stability design, given the considerations above, one way in which proteins could be designed to reduce their dependence on the PN is by improving the protein sequence's conformity to the native-state backbone. Increased sequence–backbone conformity would accelerate the formation of native secondary structure and reduce the formation of misfolded states. As we see in Section 4, this is an important mechanism by which protein stability and expressibility can be increased.

3.1.2. Protein-folding chaperones prevent misfolding and aggregation. A protein chain may interact with chaperones at every phase of its existence—during synthesis, including as the nascent chain emerges from the ribosome; through all folding stages; and in the folded state (**Figure 3**). The

protein chain may also be rescued by chaperones from misfolded or aggregated states and brought back into the folding pathway. A chaperone is defined as any protein that interacts transiently with non-native conformations of other proteins to promote their folding (or refolding) into their native state (105). By their interactions with client proteins, chaperones may block some of the folding trajectories leading to undesired intermediates that are prone to misfolding and aggregation (106, 107). Estimates put the fraction of the proteome that interacts with at least one chaperone in any organism at more than 70% (77, 108); however, determining what fraction of these interactions is obligatory chaperone interactions is difficult, owing to overlap and redundancy in the functions of many chaperones.

Obligatory chaperone interactions: proteins that must interact with a specific chaperone family to reach their folded state under physiological conditions

The molecular mechanisms by which chaperones assist folding for a wide range of distinct client proteins remained elusive until the past decade. Recent advances in X-ray crystallography, single-molecule microscopy, and NMR have provided detailed information about client–chaperone interactions at the molecular level. A feature common to many chaperones is that they bind large, solvent-exposed hydrophobic surfaces (77, 80, 94, 109–111)—precisely the molecular features that lead to misfolding and aggregation. Affinity to the client proteins is typically low, enabling fast cycles of binding, release, and rebinding of the client protein until the native state is stably formed and the hydrophobic surfaces are sequestered from the cytoplasm. Since the PN recognizes exposed hydrophobic patches on proteins, one common approach to reduce dependence on the PN for efficient folding is to eliminate by design surface hydrophobic patches (see Section 4).

To conclude, efficient folding is a prerequisite for obtaining high yields of functional protein in vivo. Folding efficiency is encoded, at least to some extent, in the protein primary sequence, but for many large, multidomain proteins, which are often the subject of stability design, it is also determined by interactions with the PN. This consideration can become important in overexpression and particularly in heterologous hosts, which may lack specific chaperones that are essential to folding. Protein design for intrinsic expressibility, that is, independent of specific chaperones, is therefore highly desired. Nevertheless, designing highly expressible proteins is challenging given that folding intermediates, misfolded states, and interaction sites with the PN are numerous, and molecular information about them is rare. Even without knowledge of alternative states, however, some of the features that trigger the involvement of the PN in protein folding can be eliminated by design to improve intrinsic expressibility. In the next section, we review how the principles of native-state stability and in vivo folding efficiency are used to design more stable and expressible proteins.

4. PROTEIN STABILITY DESIGN

Proteins are increasingly being used as research reagents, as catalysts for biochemical transformations, and in biomedicine. The considerations outlined above regarding thermal stability, misfolding, aggregation, and heterologous overexpression are substantial and sometimes represent insurmountable constraints on the path to application. Furthermore, engineering enhanced protein activity, such as binding affinity or catalytic rate, is often constrained by the marginal stability of the target protein (4). Protein engineering therefore often comprises laborious and iterative steps to enhance activity and then to improve or regain stability (112–114). More broadly, from the perspective of fundamental research, rational design of stable protein variants provides an ultimate test of our understanding of the rules governing protein structure, stability, function, and expressibility.

Protein native-state stability and expressibility are, in principle, related, even though the former is a thermodynamic property and the latter is primarily governed by the folding trajectory. This

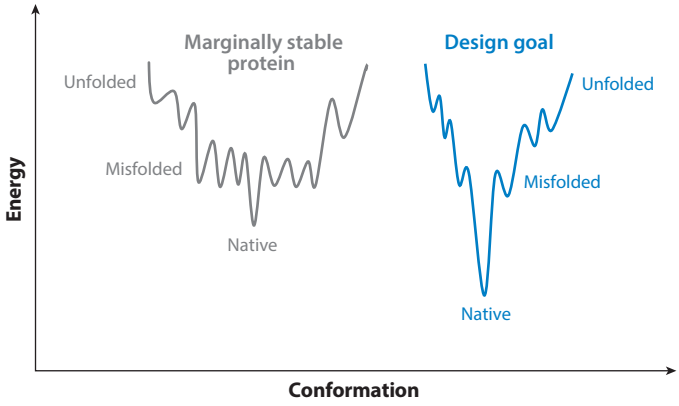

Figure 4

Schematic representation of the folding landscape of a marginally stable protein versus the goal of the design process. In this scheme, the marginally stable protein has many competing misfolded states of only slightly higher energy than the native state, and its folding is therefore frustrated. In the folding landscape of a successfully designed protein, by contrast, some misfolded states are eliminated and the difference in energy between the native and unfolded and remaining misfolded states is greater. Accordingly, the designed protein may preferentially fold into the native state even without the assistance of chaperones from its source organism.

relation is due to the fact that protein stability is determined by the free-energy difference between the folded and misfolded or unfolded states. Misfolded states may be thought of as traps that distract the nascent chain from folding into the native state, thereby frustrating the folding trajectory and lowering the yields of natively folded protein (35, 63, 64); misfolded states may also lead to terminal aggregation (**Figure 3**). Therefore, the free-energy difference between the folded state and misfolded states is also one of the determinants of expressibility. We may use the schematic in **Figure 4** to consider the goal of stability design: The energy landscape of a marginally stable target protein comprises numerous misfolded states, which are close in energy to the native state, and may therefore be highly populated. In the native host, the PN blocks these states, but the misfolded states may limit expressibility in heterologous hosts and limit the protein's lifetime in vitro. The stability design goal is to increase the energy gap between the natively folded state and misfolded or unfolded states. The energy gap can be increased by lowering the native-state energy, but it is also beneficial to eliminate as many misfolded states as possible and maintain the remaining ones at high energy relative to the new native-state energy, thus obtaining unfrustrated folding (83, 115). Thus, by increasing the gap between the natively folded state and misfolded or unfolded states, both thermal stability and expressibility are improved.

In principle, lowering the native-state energy requires a molecular structure of the native state and a design algorithm that captures the positive-design elements considered in Section 2. The schematic in **Figure 4** reveals, however, that this alone would not be enough, since such a design algorithm may inadvertently lower the energy of misfolded states or even introduce new misfolded states that were not populated in the original protein. Misfolded states may appear, on first consideration, to complicate the design goal tremendously, because we have almost no structural information on misfolded states and therefore no way to model them. Nevertheless, we saw that some of the structural features that affect misfolding and aggregation (Section 2.2.2) and trigger PN involvement in folding (Section 3.1.2) may obey simple negative-design principles related to charge, hydrophobicity, and secondary structure propensity (68); these principles do not require detailed atomistic information on interactions in the misfolded states. To achieve the

design goal of **Figure 4**, one must therefore lower the native-state energy, subject to the negative-design principles that limit misfolding and aggregation and also subject to the requirements of maintaining the protein's desired activity.

4.1. Phylogeny-Based Stability Design

Since phylogenetic analysis does not require a structure or a model of protein energetics, it can potentially circumvent the challenge of accurate modeling of native-state energetics. At the core of phylogeny-based stability design methods lies the notion that any extant protein may have accumulated destabilizing mutations during the course of its evolution, so long as those do not lower the natively folded fraction below physiological requirements (**Figure 1**). Viewed as a family of homologs, however, the most prevalent (consensus) identity at each amino acid position is likely to also be the most stabilizing; this most prevalent identity is also likely to conform with protein folding and with the protein's primary function, since it was selected through evolution in a majority of homologs evolving under different physiological contexts and demands. Conversely, identities that are rare or absent in a family of homologs are likely to be detrimental to function, stability, or expressibility, and their depletion may indicate some negative-design requirements. Thus, the history of mutations in the protein family may reveal specific solutions used in nature to stabilize the target protein's native state relative to misfolded and unfolded states without harming its primary activity.

On the basis of this idea, a method termed consensus design has been in use for more than two decades (116). In this method, amino acid identities in the target protein that differ from the family consensus are selectively changed to the consensus identity. Where sequence alignments are unambiguous and a large number of sequence homologs is available—as is the case for immunoglobulins, repeat proteins (117), and certain enzymes—approximately half of the consensus mutations were seen to be individually stabilizing (116, 118), and some multipoint consensus mutants were seen to have improved thermal resistance of more than 20°C. Although clearly successful and influential (119, 120), consensus design relies on high-confidence sequence alignments, which are not always available. Furthermore, by not accounting for the atomic details of the target protein, consensus design is prone to false-positive predictions, whereby mutations predicted to stabilize the protein instead destabilize it or harm its activity, particularly mutations in positions that are spatially close to one another. Consensus design therefore often requires laborious rounds of sequence design and experimental testing (121).

4.2. Structure-Based Stability Design

To eliminate some of the uncertainty inherent in conformational sampling and in evaluating the energetics of biomolecules (see Sections 2.2.2 and 2.3), early work in structure-based stability design focused on one or another aspect of protein stability. For instance, difficulties in modeling protein-surface electrostatics and solvation led to solutions that exclusively dealt with filling cavities in the protein core, where computational modeling is more accurate (122, 123). Conversely, some methods focused on increasing the number of charged amino acid residues on the protein surface, raising solubility and thermal resistance without requiring a detailed atomistic model (124, 125). Related methods scanned for sequence and structure patterns that are common in aggregation-prone regions, such as hydrophobic stretches, and introduced mutations to charged residues, thereby reducing aggregation and increasing solubility (126). Other studies focused on rigidifying the backbone by replacing the flexible amino acid Gly with Ala or introducing the rigid amino acid Pro where the native-state backbone conformation allowed such changes; such

rigidifying mutations may lower the entropy of the unfolded state, thereby improving the native-state free energy in relation to the unfolded state (127–129). These studies therefore address some of the principles of protein stability, protein expressibility, and the involvement of PN in protein folding. Furthermore, they require only limited modeling or none at all, thereby reducing risk from inaccuracies in biomolecular energetics. Although these methods have been successful in particular cases, each is appropriate to a subset of protein stability problems and therefore is not general.

Protein expressibility: yields of natively folded, active protein due to factors encoded in the primary sequence, rather than in the messenger RNA (mRNA) (e.g., codon usage or mRNA structure)

4.3. Stability Design Using Hybrid Approaches Comprising Phylogeny and Atomistic Design

The structure-based stability design methods described above address individual aspects of protein marginal stability, such as core cavities, surface hydrophobicity, and the flexibility of the unfolded state. In principle, energy functions used in modeling and design encode all of the thermodynamic terms described in Section 2, and design algorithms should therefore address all aspects of positive design simultaneously. Indeed, de novo designed folds have shown remarkable thermal resistance (19, 21, 74), but these folds were all based on ideal, strain-free backbones of high secondary structure content and were devoid of activity. It therefore remained unclear whether stability could be reliably improved by design without harming protein activity (119), especially as stability and molecular activity may trade off (4, 29, 30). It also remained unclear whether stability design methods could substantially improve protein expressibility in the absence of the native host's PN, particularly in large mammalian proteins that are prone to misfolding.

To address the problem of designing stable variants of natural proteins without harming their primary activity, two recent independent studies combined structure modeling and evolutionary information during design. In one method termed FireProt, two atomistic design algorithms were combined with evolutionary conservation and covariation analyses (130). Specifically, evolutionarily conserved positions and positions for which covariance analysis indicated dependence between positions were restricted to wild-type identities. Furthermore, FireProt focused design calculations on regions, which are likely to be the source of stability problems, and away from regions that are involved in activity. The method was applied to design two model enzymes, and in both cases, large gains in melting temperature ($>20°C$) were observed.

A second hybrid method, termed PROSS, which was developed by the authors, subjects the entire protein to Rosetta design, except for the active or binding site (131). PROSS starts by two-step filtering of all amino acid identities that are likely to be destabilizing as single-point mutations relative to the wild-type native state: First, phylogenetic analysis rules out amino acid identities that are rarely observed in homologs (**Figure 5**); and second, Rosetta modeling is used to scan all identities that passed the first filter and to eliminate single-point mutations that destabilize the native state (see **Supplemental Video 1**). These two steps result in a reduced sequence space, in which all point mutations are predicted to be stabilizing. At the last step, Rosetta designs optimal combinations of mutations from this reduced sequence space, taking into account all interactions between mutated and unmutated positions, as encoded in the energy function. In this last step, inspired by the notion of consensus design, the energy function used to mutate the protein is augmented with a biasing potential that favors amino acid identities according to their frequency in the multiple-sequence alignment. This phylogeny-based biasing potential allows amino acid identities that are considered by Rosetta to be neutral, or even slightly destabilizing, to be incorporated into the designed variants. PROSS therefore combines aspects of consensus design with atomistic modeling and may therefore introduce both positive and negative design elements, including elements that address thermal stability and expressibility. We tested this approach by

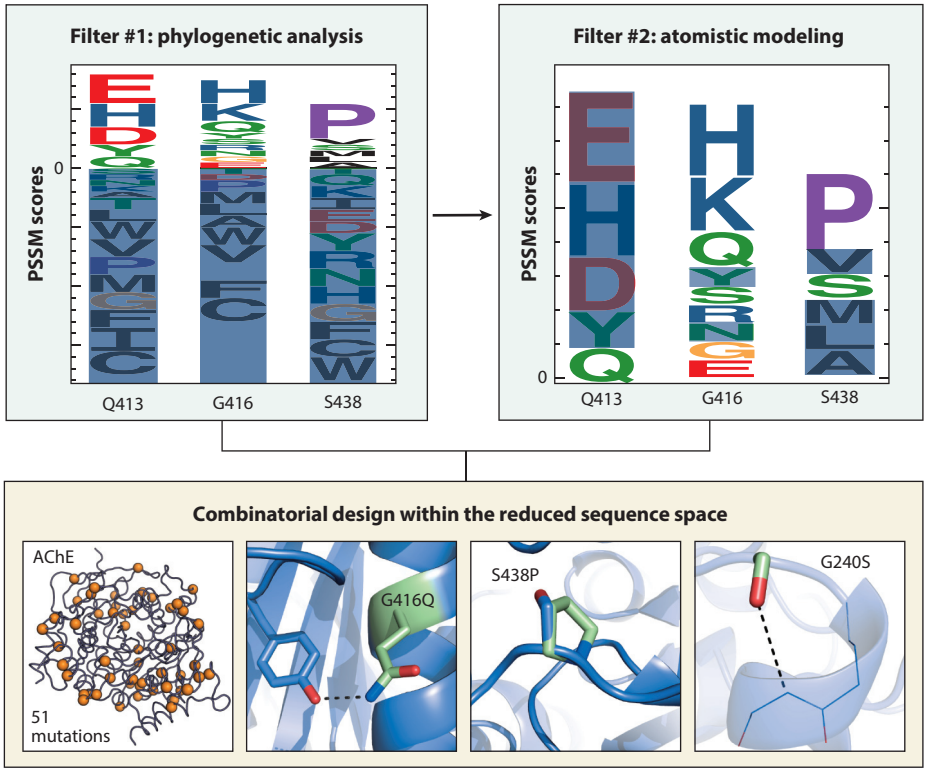

Figure 5

Illustration of the PROSS workflow applied to human acetylcholinesterase (AChE) (131). Filter #1: A multiple-sequence alignment (MSA) of homologous sequences is used to generate a position-specific scoring matrix (PSSM), which weights amino acid identities according to their likelihood of appearance in the MSA (high PSSM scores represent more likely amino acids). Shaded identities (PSSM <0) are eliminated from the sequence space available for design. Three representative positions on human AChE are noted. Filter #2: Each of the remaining identities is subjected to Rosetta atomistic modeling (**Supplemental Video 1**), and identities with unfavorable $\Delta\Delta G$ are eliminated (*shaded identities*). At the final step, all amino acid positions outside the active site are subjected to combinatorial sequence design within the reduced sequence space of the two filters. The AChE design shown here (PDB entry: 5HQ3) comprises 51 mutations relative to human AChE (PDB entry: 4EY7), yet has identical enzymatic properties, 20°C higher thermal resistance, and 1,800-fold higher expression levels in bacterial cells. Mutated positions are indicated in orange spheres, and several representative mutations are shown in green sticks, relative to the wild-type background in blue: (*left*) Gly416 on an exposed helical surface is mutated to Gln, increasing the amino acid conformity with the helical backbone and introducing an additional hydrogen bond with a neighboring tyrosine. (*center*) Ser438Pro rigidifies the loop backbone. (*right*) Gly240Ser forms helix-capping hydrogen bonds. Additional abbreviation: PDB, Protein Data Bank.

designing up to 5 variants, each comprising 10–70 mutations from wild type, for each of 5 unrelated enzymes and a vaccine immunogen (131, 132). In all cases, all or most variants showed large gains in thermal and aggregation resistance with no impact on the proteins' primary activity.

A notable effect of applying PROSS to proteins from eukaryotes was large gains in bacterial expression levels—gains of as many as three orders of magnitude (131, 132). Although expression levels and thermal stability are known to correlate (41, 133, 134), the large gains we observed were unusual and prompted us to analyze the molecular details of the designed variants. Structural

analysis suggested that for each protein tested, the method improved a variety of molecular properties, though the dominant improvement was often different for each target; in one case, PROSS introduced 17 core mutations, whereas in others, surface polarity was improved through the introduction of additional charges and polar amino acids. We also found that many mutations improved local sequence–backbone conformity, by introducing side chain–backbone hydrogen bonds in helix caps and rigidifying the backbone through mutation to Pro or away from Gly in appropriate places (**Figure 5**). This analysis suggested a potential molecular mechanism by which the mutations might improve intrinsic expressibility and lower dependence on the native PN. Specifically, the mutations address some of the negative-design elements described above, including improvement of sequence–backbone compatibility, and the removal of aggregation-prone hydrophobic surface patches. Whereas the wild-type protein may necessitate its native PN for efficient folding, the designed variants eliminate destabilizing or aggregation-prone features, achieving the design goal of **Figure 4**. Indeed, a recent study showed that a PROSS-stabilized green fluorescent protein (GFP) folded in *Escherichia coli* independent of the chaperone GroEL, whereas wild-type GFP formed tight interactions with GroEL and required it for in vivo folding (135). It therefore appears that combining information from evolutionary conservation analysis with atomistic design addresses both positive- and negative-design elements, promoting both thermal stability and intrinsic expressibility, including chaperone independence, while maintaining the protein's primary molecular activity.

5. CONCLUSIONS

Stability design methodology has improved through a deeper understanding of the underlying causes of marginal protein stability. Protein engineering studies through the 1980s and 1990s laid the groundwork for understanding the thermodynamic underpinnings of protein stability and folding pathways. Later work provided crucial information on the determinants of misfolding and aggregation, the involvement of molecular chaperones in the intricate process by which a large polypeptide chain folds into a unique native state without being trapped in misfolded or aggregated states, and how the native state persists. It became clear that since molecular chaperones are essential for folding native proteins, heterologous protein expression might suffer from the absence of required chaperones. Further work provided an important insight about the difficulties that lay ahead for stability design: First, mutations that improve stability may reduce molecular activity, a situation that is unacceptable in almost all cases; and second, stabilizing mutations individually make small contributions to stability, and only multiple mutations have an appreciable effect. Understanding these challenges led to stability design solutions that either eschewed structure modeling, as in consensus design, or used only structure modeling components that were deemed relatively safe, such as backbone rigidification, core packing design, or surface charge design. Parallel improvements in energy functions (76, 136), which were not reviewed here, increased the accuracy of modeling and design. Finally, in recent years, hybrid methods combining phylogenetic analysis with structure modeling and atomistic design have proven effective in significantly improving both thermodynamic stability and expressibility even in the absence of the native PN through the introduction of dozens of mutations. Thus, the goal of design methods applicable to a wide range of marginal stability problems may be within reach.

The ability to rationally stabilize proteins provides important opportunities for future research. Enzymes from eukaryotes or from organisms that defy culturing in the laboratory often require laborious cycles of protein engineering to achieve high expressibility in preferred heterologous hosts (137); these cycles may now be considerably shortened. Similarly, subunit vaccines, whereby surface proteins from pathogens are used in vaccination instead of the live or attenuated pathogen

itself, are a promising area for human and veterinary animal health; however, these surface proteins are often unstable (138), and stability design may be used to rapidly improve them (132). Additionally, stabilized and highly expressing protein variants may serve as better starting points for engineering enhanced function or altered specificity (114). And finally, basic research into the molecular determinants of low stability and expressibility may benefit from contrasting marginally stable natural proteins and their stabilized counterparts and may in turn provide important feedback for improving algorithms for stability design. An important area we did not cover in this review relates to membrane proteins, for which low stability and expressibility are often as problematic as for soluble proteins (139). Recent improvements in understanding the sequence determinants of membrane protein energetics and expression may extend stability design to this class of proteins (140, 141). These recent and future developments of stability design may therefore accelerate numerous fields of biomolecular research.

SUMMARY POINTS

1. Many natural proteins are only marginally stable, with energy gaps between the native and unfolded or misfolded states as low as 5 kcal/mol, equivalent to the net contribution from only a few hydrogen bonds.
2. Marginal stability sensitizes proteins to changes in environment, lowers their expressibility, increases costs of production, and therefore limits their usefulness in research and applications.
3. The noncovalent forces that stabilize the native state are individually weak, and only the aggregate of thousands of such interactions overcomes the entropy loss upon folding.
4. Proteins embody elements of negative design to prevent misfolding and aggregation, but these elements are potentially challenging to model as misfolded and aggregated states are numerous and structurally uncharacterized.
5. Marginal stability is overcome in living systems by an elaborate network of proteins known as the PN that limits misfolding and aggregation.
6. Phylogenetic analysis focuses stability design on mutations that have been preferred in the target protein's evolution, thereby incorporating both positive and negative design elements.
7. Current algorithms use combinations of phylogenetic and structure-modeling approaches and have resulted in large increases in thermal stability, aggregation resistance, and expressibility even in the absence of the host's PN, without loss in the protein's primary activity.

FUTURE ISSUES

1. Exciting areas for applying stability design include enzymes for high-temperature reactors and antibodies and vaccines for long-term storage and high production yields.
2. Stable, high-expression designs may be contrasted with their natural counterparts to understand the molecular determinants of low expression and particularly interactions with the PN.

3. Membrane proteins are often unstable and exhibit low expressibility; future methods should account for thermal stability and the determinants of membrane protein expression.
4. Evolutionary coupling between amino acid positions may enable better modeling and design of the cooperative interaction networks in protein cores and active sites.

DISCLOSURE STATEMENT

S.F. is a paid consultant to IgC Bio Ltd., which holds a license to the PROSS stability design algorithm.

ACKNOWLEDGMENTS

The authors thank Amnon Horovitz for discussions on the role of cooperativity and chaperones in protein folding and Ingémar Andre, Dan Tawfik, and Nicholas Schafer for critical reading of the manuscript. Research in the Fleishman laboratory is supported by a Starter's Grant from the European Research Council (335439), the Israel Science Foundation through its Center of Excellence in Structural Cell Biology (1775/12) and its India-Israel Joint Research Program (2281/15), by IgC Bio Ltd., and by a charitable donation from Sam Switzer and family.

LITERATURE CITED

1. Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29(31):7133–55
2. Fersht A. 1999. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: Freeman
3. Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10(12):866–76
4. Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19(5):596–604
5. Bershtein S, Goldin K, Tawfik DS. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* 379(5):1029–44
6. Yue P, Li Z, Moult J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* 353(2):459–73
7. Bullock AN, Fersht AR. 2001. Rescuing the function of mutant p53. *Nat. Rev. Cancer* 1(1):68–76
8. Protasevich I, Yang Z, Wang C, Atwell S, Zhao X, et al. 2010. Thermal unfolding studies show the disease causing F508del mutation in CFTR thermodynamically destabilizes nucleotide-binding domain 1. *Protein Sci.* 19(10):1917–31
9. Pauling L, Itano HA, Singer SJ, Wells IC. 1949. Sickle cell anemia, a molecular disease. *Science* 110(2865):543–48
10. Stefani M, Dobson CM. 2003. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* 81(11):678–99
11. Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, et al. 2015. Widespread proteome remodeling and aggregation in aging *C. elegans*. *Cell* 161(4):919–32
12. Taylor RC, Dillin A. 2011. Aging as an event of proteostasis collapse. *Cold Spring Harb. Perspect. Biol.* 3(5):aa004440
13. Gidalevitz T, Ben-Zvi A, Ho KH, Brignull HR, Morimoto RI. 2006. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* 311(5766):1471–74

14. Ben-Zvi A, Miller EA, Morimoto RI. 2009. Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *PNAS* 106(35):14914–19
15. Lin H, Bhatia R, Lal R. 2001. Amyloid β protein forms ion channels: implications for Alzheimer's disease pathophysiology. *FASEB J.* 15(13):2433–44
16. Poirier MA, Li H, Macosko J, Cai S, Amzel M, Ross CA. 2002. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. *J. Biol. Chem.* 277(43):41032–37
17. Chiti F, Dobson CM. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–66
18. Zhao H, Arnold FH. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* 12(1):47–53
19. Huang P-S, Boyken SE, Baker D. 2016. The coming of age of de novo protein design. *Nature* 537(7620):320–27
20. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, et al. 2016. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* 352(6286):680–87
21. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364–68
22. Dahiyat BI, Mayo SL. 1997. De novo protein design: fully automated sequence selection. *Science* 278(5335):82–87
23. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, et al. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–95
24. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, et al. 2011. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332(6031):816–21
25. Baran D, Pszolla MG, Lapidoth GD, Norn C, Dym O, et al. 2018. Principles for computational design of binding antibodies. *PNAS*. In press
26. Pace CN, Scholtz JM, Grimsley GR. 2014. Forces stabilizing proteins. *FEBS Lett.* 588(14):2177–84
27. Dill KA, MacCallum JL. 2012. The protein-folding problem, 50 years on. *Science* 338(6110):1042–46
28. Warshel A, Levitt M. 1976. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103(2):227–49
29. Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995. A relationship between protein stability and protein function. *PNAS* 92(2):452–56
30. Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008. How protein stability and new functions trade off. *PLOS Comput. Biol.* 4(2):e1000002
31. Halperin I, Wolfson H, Nussinov R. 2004. Protein-protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure* 12(6):1027–38
32. Srere PA. 1984. Why are enzymes so big? *Trends Biochem. Sci.* 9(9):387–90
33. Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. 2009. Importing mitochondrial proteins: machineries and mechanisms. *Cell* 138(4):628–44
34. Tskhovrebova L, Trinick J, Sleep JA, Simmons RM. 1997. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* 387(6630):308–12
35. Ferreiro DU, Komives EA, Wolynes PG. 2013. Frustration in biomolecules. *Q. Rev. Biophys.* 47(4):1–97
36. Tsai C-J, Nussinov R. 2014. A unified view of "how allostery works." *PLOS Comput. Biol.* 10(2):e1003394
37. Brady GP, Sharp KA. 1997. Entropy in protein folding and in protein–protein interactions. *Curr. Opin. Struct. Biol.* 7(2):215–21
38. Levinthal C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* 65:44–45
39. Fleishman SJ, Baker D. 2012. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 149(2):262–73
40. McLendon G, Radany E. 1978. Is protein turnover thermodynamically controlled? *J. Biol. Chem.* 253(18):6335–37
41. Parsell DA, Sauer RT. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J. Biol. Chem.* 264(13):7590–95

42. Warszawski S, Netzer R, Tawfik DS, Fleishman SJ. 2014. A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *J. Mol. Biol.* 426(24):4125–38
43. Goldstein RA. 2011. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins* 79(5):1396–1407
44. Szilágyi A, Závodszky P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* 8(5):493–504
45. Jaenicke R, Böhm G. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* 8(6):738–48
46. Steiner P, Sauer U. 2001. Proteins induced during adaptation of *Acetobacter aceti* to high acetate concentrations. *Appl. Environ. Microbiol.* 67(12):5474–81
47. Honig B, Yang AS. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* 46:27–58
48. Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11(4):281–96
49. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, et al. 2011. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* 408(3):514–28
50. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, et al. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255(5041):178–83
51. Chothia C. 1975. Structural invariants in protein folding. *Nature* 254(5498):304–8
52. Albeck S, Unger R, Schreiber G. 2000. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J. Mol. Biol.* 298(3):503–20
53. Yang AS, Honig B. 1995. Free energy determinants of secondary structure formation: I. α-Helices. *J. Mol. Biol.* 252(3):351–65
54. Honig B, Nicholls A. 1995. Classical electrostatics in biology and chemistry. *Science* 268(5214):1144–49
55. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR. 1990. Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216(4):1031–44
56. Makhatadze GI, Loladze VV, Ermolenko DN, Chen X, Thomas ST. 2003. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* 327(5):1135–48
57. Anderson DE, Becktel WJ, Dahlquist FW. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* 29(9):2403–8
58. Tissot AC, Vuilleumier S, Fersht AR. 1996. Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry* 35(21):6786–94
59. Elcock AH. 1998. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* 284(2):489–502
60. McGaughey GB, Gagné M, Rappé AK. 1998. π-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* 273(25):15458–63
61. Dougherty DA. 2007. Cation-π interactions involving aromatic amino acids. *J. Nutr.* 137(6, Suppl. 1):1504S–8S; discussion, 1516S–17S
62. Prajapati RS, Sirajuddin M, Durani V, Sreeramulu S, Varadarajan R. 2006. Contribution of cation-π interactions to protein stability. *Biochemistry* 45(50):15000–10
63. Abkevich VI, Gutin AM, Shakhnovich EI. 1994. Free energy landscape for protein folding kinetics: intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.* 101(7):6052–62
64. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21(3):167–95
65. Schafer NP, Kim BL, Zheng W, Wolynes PG. 2014. Learning to fold proteins using energy landscape theory. *Isr. J. Chem.* 54(8–9):1311–37
66. Richardson JS, Richardson DC. 2002. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *PNAS* 99(5):2754–59
67. Broome BM, Hecht MH. 2000. Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J. Mol. Biol.* 296(4):961–68

68. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. 2003. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424(6950):805–8
69. Das R, Baker D. 2008. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* 77:363–82
70. Dahiyat BI, Mayo SL. 1996. Protein design automation. *Protein Sci.* 5(5):895–903
71. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* 33(Suppl. 2):W382–88
72. Yin S, Ding F, Dokholyan NV. 2007. Eris: an automated estimator of protein stability. *Nat. Methods* 4(6):466–67
73. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, et al. 1995. Protein design: a hierarchic approach. *Science* 270(5238):935–41
74. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, et al. 2012. Principles for designing ideal protein structures. *Nature* 491(7423):222–27
75. Lazaridis T, Karplus M. 1999. Effective energy function for proteins in solution. *Proteins* 35(2):133–52
76. Park H, Bradley P, Greisen P Jr., Liu Y, Mulligan VK, et al. 2016. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* 12(12):6201–12
77. Balchin D, Hayer-Hartl M, Hartl FU. 2016. In vivo aspects of protein folding and quality control. *Science* 353(6294):aac4354
78. Labbadia J, Morimoto RI. 2015. The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* 84:435–64
79. Powers ET, Balch WE. 2013. Diversity in the origins of proteostasis networks—a driver for protein function in evolution. *Nat. Rev. Mol. Cell Biol.* 14(4):237–48
80. Saibil H. 2013. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* 14(10):630–42
81. Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181(4096):223–30
82. Jackson SE. 1998. How do small single-domain proteins fold? *Fold. Des.* 3(4):R81–91
83. Shakhnovich EI. 1997. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* 7(1):29–40
84. Brocchieri L, Karlin S. 2005. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33(10):3390–3400
85. Brockwell DJ, Radford SE. 2007. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17(1):30–37
86. Dobson CM. 2003. Protein folding and misfolding. *Nature* 426(6968):884–90
87. Ellis RJ, Minton AP. 2006. Protein aggregation in crowded environments. *Biol. Chem.* 387(5):485–97
88. Douglas PM, Dillin A. 2010. Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* 190(5):719–29
89. Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16(6):574–81
90. Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396(6709):336–42
91. Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459(7247):668–73
92. Chaney JL, Clark PL. 2015. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* 44:143–66
93. Young R, Bremer H. 1976. Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochem. J.* 160(2):185–94
94. Kramer G, Boehringer D, Ban N, Bukau B. 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 16(6):589–97
95. Trovato F, O'Brien EP. 2016. Insights into cotranslational nascent protein behavior from computer simulations. *Annu. Rev. Biophys.* 45:345–69
96. Evans MS, Sander IM, Clark PL. 2008. Cotranslational folding promotes β-helix formation and avoids aggregation in vivo. *J. Mol. Biol.* 383(3):683–92

97. Voss NR, Gerstein M, Steitz TA, Moore PB. 2006. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* 360(4):893–906
98. Kosolapov A, Deutsch C. 2009. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 16(4):405–11
99. Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, et al. 2015. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* 12(10):1533–40
100. Lu J, Deutsch C. 2005. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 12(12):1123–29
101. Woolhead CA, McCormick PJ, Johnson AE. 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* 116(5):725–36
102. Bhushan S, Gartmann M, Halic M, Armache J-P, Jarasch A, et al. 2010. α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 17(3):313–17
103. Lu J, Kobertz WR, Deutsch C. 2007. Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* 371(5):1378–91
104. Ziv G, Haran G, Thirumalai D. 2005. Ribosome exit tunnel can entropically stabilize α-helices. *PNAS* 102(52):18956–61
105. Hartl FU. 1996. Molecular chaperones in cellular protein folding. *Nature* 381(6583):571–79
106. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. 2013. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* 82:323–55
107. Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, et al. 2010. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* 142(1):112–22
108. Gong Y, Kakihara Y, Krogan N, Greenblatt J, Emili A, et al. 2009. An atlas of chaperone-protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. *Mol. Syst. Biol.* 5(1):275
109. Rüdiger S, Germeroth L, Schneider-Mergener J, Bukau B. 1997. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* 16(7):1501–7
110. Sekhar A, Rosenzweig R, Bouvignies G, Kay LE. 2015. Mapping the conformation of a client protein through the Hsp70 functional cycle. *PNAS* 112(33):10395–400
111. Sekhar A, Rosenzweig R, Bouvignies G, Kay LE. 2016. Hsp70 biases the folding pathways of client proteins. *PNAS* 113(20):2794–2801
112. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. *PNAS* 103(15):5869–74
113. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444(7121):929–32
114. Goldsmith M, Aggarwal N, Ashani Y, Jubran H, Greisen PJ, et al. 2017. Overcoming an optimization plateau in the directed evolution of highly efficient nerve agent bioscavengers. *Protein Eng. Des. Sel.* 30(4):333–45
115. Shakhnovich EI, Gutin AM. 1993. Engineering of stable and fast-folding sequences of model proteins. *PNAS* 90(15):7195–99
116. Steipe B, Schiller B, Plückthun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* 240(3):188–92
117. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, Grutter MG. 2003. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *PNAS* 100:1700–5
118. Lehmann M, Pasamontes L, Lassen SF, Wyss M. 2000. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* 1543(2):408–15
119. Magliery TJ. 2015. Protein stability: computation, sequence statistics, and new experimental methods. *Curr. Opin. Struct. Biol.* 33:161–68
120. Porebski BT, Buckle AM. 2016. Consensus protein design. *Protein Eng. Des. Sel.* 29(7):245–51
121. Sullivan BJ, Nguyen T, Durani V, Mathur D, Rojas S, et al. 2012. Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability. *J. Mol. Biol.* 420(4–5):384–99

122. Korkegian A, Black ME, Baker D, Stoddard BL. 2005. Computational thermostabilization of an enzyme. *Science* 308(5723):857–60
123. Borgo B, Havranek JJ. 2012. Automated selection of stabilizing mutations in designed and natural proteins. *PNAS* 109(5):1494–99
124. Lawrence MS, Phillips KJ, Liu DR. 2007. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* 129(33):10110–12
125. Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, et al. 2012. Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* 19(4):449–55
126. Ganesan A, Siekierska A, Beerten J, Brams M, Van Durme J, et al. 2016. Structural hot spots for the solubility of globular proteins. *Nat. Commun.* 7:10816
127. Dombkowski AA, Sultana KZ, Craig DB. 2014. Protein disulfide engineering. *FEBS Lett.* 588(2):206–12
128. Hardy F, Vriend G, Veltman OR, van der Vinne B, Venema G, Eijsink VG. 1993. Stabilization of *Bacillus stearothermophilus* neutral protease by introduction of prolines. *FEBS Lett.* 317(1–2):89–92
129. Matthews BW, Nicholson H, Becktel WJ. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *PNAS* 84(19):6663–67
130. Bednar D, Beerens K, Sebestova E, Bendl J, Khare S, et al. 2015. FireProt: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLOS Comput. Biol.* 11(11):e1004556
131. Goldenzweig A, Goldsmith M, Hill SE, Gertman O, Laurino P, et al. 2016. Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell.* 63(2):337–46
132. Campeotto I, Goldenzweig A, Davey J, Barfod L, Marshall JM, et al. 2017. One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *PNAS* 114(5):998–1002
133. Kieke MC, Shusta EV, Boder ET, Teyton L, Wittrup KD, Kranz DM. 1999. Selection of functional T cell receptor mutants from a yeast surface-display library. *PNAS* 96(10):5651–56
134. Kwon WS, Da Silva NA, Kellis JT Jr. 1996. Relationship between thermal stability, degradation rate and expression yield of barnase variants in the periplasm of *Escherichia coli*. *Protein Eng.* 9(12):1197–1202
135. Bandyopadhyay B, Goldenzweig A, Unger T, Adato O, Fleishman SJ, et al. 2017. Local energetic frustration affects the dependence of green fluorescent protein folding on the chaperonin GroEL. *J. Biol. Chem.* 292:20583–91
136. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, et al. 2010. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487:545–74
137. Joyce MG, Zhang B, Ou L, Chen M, Chuang G-Y, et al. 2016. Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat. Struct. Mol. Biol.* 23(9):811–20
138. Kotecha A, Seago J, Scott K, Burman A, Loureiro S, et al. 2015. Structure-based energetics of protein interfaces guides foot-and-mouth disease virus vaccine design. *Nat. Struct. Mol. Biol.* 22(10):788
139. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. 2008. Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *PNAS* 105(3):877–82
140. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, et al. 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433(7024):377–81
141. Elazar A, Weinstein J, Biran I, Fridman Y, Bibi E, Fleishman SJ. 2016. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *eLife* 5:12125
142. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang H-C, et al. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122(2):209–20
143. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. 2010. A systematic survey of in vivo obligate chaperonin- dependent substrates. *EMBO J.* 29(9):1552–64
144. Ishimoto T, Fujiwara K, Niwa T, Taguchi H. 2014. Conversion of a chaperonin GroEL-independent protein into an obligate substrate. *J. Biol. Chem.* 289(46):32073–80
145. Clark AC, Hugo E, Frieden C. 1996. Determination of regions in the dihydrofolate reductase structure that interact with the molecular chaperonin GroEL. *Biochemistry* 35(18):5893–901
146. Takano K, Scholtz JM, Sacchettini JC, Pace CN. 2003. The contribution of polar group burial to protein stability is strongly context-dependent. *J. Biol. Chem.* 278(34):31790–95

RELATED RESOURCES

1. The Protein Repair One Stop Shop (PROSS) webserver for stability design: **http://pross.weizmann.ac.il**
2. The FireProt webserver for stability design: **http://loschmidt.chemi.muni.cz/fireprot**
3. FoldX for prediction of the thermodynamic effects of single-point mutations: **http://foldxsuite.crg.eu/**
4. Eris for prediction of the thermodynamic effects of single-point mutations: **http://eris.dokhlab.org/**