



ELSEVIER

Protein stability by number: high-throughput and statistical approaches to one of protein science's most difficult problems

Thomas J Magliery^{1,2}, Jason J Lavinder^{3,a} and Brandon J Sullivan³

Most proteins are only barely stable, which impedes research, complicates therapeutic applications, and makes proteins susceptible to pathologically destabilizing mutations. Our ability to predict the thermodynamic consequences of even single point mutations is still surprisingly limited, and established methods of measuring stability are slow. Recent advances are bringing protein stability studies into the high-throughput realm. Some methods are based on inferential read-outs such as activity, proteolytic resistance or split-protein fragment reassembly. Other methods use miniaturization of direct measurements, such as intrinsic fluorescence, H/D exchange, cysteine reactivity, aggregation and hydrophobic dye binding (DSF). Protein engineering based on statistical analysis (consensus and correlated occurrences of amino acids) is promising, but much work remains to understand and implement these methods.

Addresses

¹ Department of Chemistry, The Ohio State University, 100 West 18th Avenue, Columbus, OH 43210, USA

² Department of Biochemistry, The Ohio State University, 100 West 18th Avenue, Columbus, OH 43210, USA

³ Ohio State Biochemistry Program, The Ohio State University, 100 West 18th Avenue, Columbus, OH 43210, USA

Corresponding author: Magliery, Thomas J
(magliery@chemistry.ohio-state.edu)

^a Present address: Department of Chemical Engineering, The University of Texas at Austin, 1 University Station CO400, Austin, TX 78712, USA.

Current Opinion in Chemical Biology 2011, 15:443–451

This review comes from a themed issue on
Molecular Diversity
Edited by T. Ashton Cropp and Dewey McCafferty

Available online 15 April 2011

1367-5931/\$ – see front matter
© 2011 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.cbpa.2011.03.015

Introduction

Site-directed mutagenesis, still the core technology of protein engineering, will turn 30 next year. The past three decades have seen well in excess of 100,000 mutations made (many more if we count combinatorial approaches) to probe and alter the structure, activity, folding and stability of a vast array of proteins with different folds and functions. A huge number of stability measurements have been amassed, in addition to a massive body of hypothesis-driven experiments designed to tease out the basis of protein stability. But predicting the stability of

protein mutants remains one of the great unsolved problems of protein science, proving itself more difficult than the prediction of protein structure or even the design of fairly efficient enzymes.

This difficulty is in spite of our actually knowing a great deal about the forces that dominate in protein folding [1,2] and perhaps even more about the atomic-resolution structures of folded proteins. So what's the problem? One problem is that despite large forces being at work in the structure of the folded state, such as the enthalpies associated with all the hydrogen bonds that form, the net stabilities of proteins are small—5–15 kcal mol⁻¹. This is because the forces acting on the unfolded state, such as all the hydrogen bond donors and acceptors that are satisfied by solvent, are also large. This marginal differential means that exquisite accuracy is required from fairly crude potential functions, and the problem is exacerbated by our inability to meaningfully model the unfolded state. Furthermore, it is difficult or impossible to model key aspects of protein folding, such as backbone motion or solvent entropy.

Even empirical approaches that attempt to extrapolate from training sets of thermodynamic data do not capture sufficient information to solve the problem, but it is less clear if the reasons for this are fundamental. On one hand, the standard methods of characterization—calorimetry or spectroscopically observed chemical or thermal denaturation—are slow and laborious. On the other hand, even 'large' databases are easily dwarfed by the size of sequence space, and it is certainly clear that the effect of a mutation is only meaningful in context. Mutating alanine to serine is vastly different in different scaffolds, in different secondary structures, with different packing densities or solvent exposures, or with different amino acids nearby. So while insight may not follow from numbers alone, there is a degree to which having large numbers of well-characterized and highly related mutants will shed light on the problem of protein stability. And even if it does not, the technology to enable those measurements will also enable brute-force approaches for engineering stability.

In recent years, the problem of protein stability has intersected with problems of large numbers in two interesting ways, which each is proving useful for engineering proteins for improved stability and elucidating the underlying reasons. The first is the development of fairly general high-throughput methods for measuring protein stability. The second is the use of statistics to predict

Table 1

High-throughput methods of measuring protein stability				
Method	Format	Scale	Notes	Refs
Thermal inactivation	Cell lysate	10^3 – 10^5	Requires an activity to screen in HT	[7]
β -lactamase loop insertion	Selection in bacteria	$\sim 10^9$	Tunable selection by varying antibiotic	[8*]
Pulse proteolysis	<i>In vitro</i> , minimal purification	~ 10	Detection by gel electrophoresis	[9]
Yeast surface reassembly	Screen in yeast, FACS	10^5 – 10^6	Requires a protein that can reassemble after dissection	[17*]
Intrinsic fluorescence	<i>In vitro</i> , purified protein	10 – 10^3	Direction comparison to LT methods, multiple points per protein, much larger scale possible with microfluidics	[22–24,26]
H/D exchange/MALDI MS (SUPREX)	<i>In vitro</i> , minimal purification	10 – 10^2	Requires multiple points per protein and MS automation	[27,28]
Cysteine reactivity	<i>In vitro</i> , purified protein	10 – 10^3	Requires multiple points per protein but scales well with fluorescence readout	[32,33*,46]
Differential static light scattering	<i>In vitro</i> , purified protein	10^2 – 10^4	No probe is required, but affected by solubility and kinetics	[34**]
Isothermal denaturation/GFP-Basta	<i>In vitro</i> , purified protein or lysate with GFP fusion	10^2 – 10^5	Uses unfolding/aggregation kinetics as a measure of stability, can be read out with multiple signals	[35,36]
Differential scanning fluorescence/HTTS	<i>In vitro</i> , purified protein	10^2 – 10^4	Usually done with a real-time PCR machine and SYPRO orange dye, dye might perturb the measurement	[38–40]
GFP fusion	Screen in cells	10^4 – 10^6	Reports both expression level and solubility, but easily scaled	[19–21]

stabilizing mutations from the very large number of sequences that have resulted from 15 years of genome sequencing. Here we will highlight some of the most important recent advances in these two areas.

Screening for protein stability

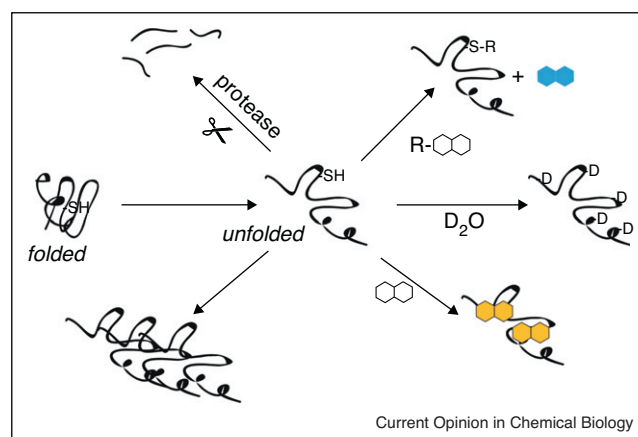
High-throughput approaches for measuring or improving protein stability generally fall into two categories; either they attempt to infer the stability from properties that are typically measured close to physiological conditions, or they perturb the conditions of the protein in some way and read out the stability (more or less) directly. For example, protein expression level, solubility, secretion, binding and enzymatic activity, and resistance to proteolysis may all be taken as indications of a stable protein [3]. In general, the Achilles' heel of these approaches is a lack of broad applicability (for example, many interesting proteins do not have an enzymatic function) and 'you-get-what-you-select-for' kinds of escape variants (for example, unstable but protease-resistant mutants). On the other hand, the problem with measuring the stability directly is the difficulty of miniaturization; circular dichroism and differential scanning calorimetry are not well suited to 96-well plates. But some creative ideas have been applied recently with both types of approaches, which we highlight here (Figure 1, Table 1). Several other recent reviews highlight other aspects of combinatorial approaches to protein biophysical properties [3–6].

Inferential screens for stability

A straightforward approach for selecting for thermostable proteins is to monitor protein activity at elevated temperatures or after heating. For example, thermal inacti-

vation was used to engineer an esterase with good tolerance of high temperatures but robust room-temperature activity, a feat that was not universally thought to be possible until it was demonstrated directly [7]. This approach is limited to proteins with an activity that can be assayed easily, and there is not a clear correlation

Figure 1



Principles of screening for protein stability. Most methods of screening for stability modify the unfolded state or observe some unique property of it—for example, by proteolysis, reaction with an exposed cysteine, amide proton exchange, hydrophobic dye binding, or aggregation. Typically, the protein solution is heated or challenged with increasing concentration of chemical denaturant to establish when the signal is observed. Because most HT stability screening methods involve observation of an irreversible reaction, care must be taken in interpreting the data as a change in equilibrium stability. While some methods enable measurements in the presence of other proteins, most require sufficient purification so that other proteins do not produce the unfolding signal.

between the degree of thermal inactivation and the stability since it is complicated by aggregation and folding rates. But this is a very practical approach to screening for proteins with improved stability and activity under various perturbing conditions. Screens for achieving a stability threshold based on binding or catalytic activity have formed the basis for several notable combinatorial experiments in protein design using λ suppressor, barnase, chorismate mutase, and Rop, to name a few [3].

Resistance to proteolysis has been used broadly to identify structured variants, particularly on phage particles. It has been difficult to rely on nonspecific proteolysis as a read-out of stability, because proteolysis rates are not just related to global stability but to local stability and substrate specificity. Recently, Bardwell and co-workers developed a system in which a protein of interest (POI) is inserted into a loop of TEM-1 β -lactamase, where it was hypothesized that lower-stability mutants of the POI would generally lead to greater degradation by cellular proteases [8[•]]. For several proteins, the log of the minimum inhibitory concentration (MIC) of antibiotic showed a striking correlation to the stability of mutants ($R^2 > 0.6$). The relationship was especially good for Immunity protein 7, where it was clear that expression level was correlated to stability. Often, expression level differences from varying rates of transcription and translation, solubility differences, or display differences (on phage or yeast) can be confounding factors to these types of inferential screens. But the authors convincingly showed that the system could be used to select for Im7 variants with improvements in both thermodynamic and kinetic stability. The selection is also tunable and demands that selected variants be soluble and expressible.

Park and Marqusee have recently employed pulse proteolysis in increasing concentrations of urea using thermolysin, which retains its activity in high urea, to measure folding ΔG values [9]. The method is read out by SDS-PAGE, but it can be applied to unpurified protein in crude lysate with sufficient overexpression or specific detection, making it suitable for fairly high-throughput quantitative determinations of stability. By adjusting the pulse time and using chemical denaturant, one can directly measure the fraction folded and avoid confounding differences in protease susceptibility under native conditions. This observation led Park *et al.* to challenge the entire *E. coli* proteome with a protease under native conditions to specifically identify resistant proteins [10[•]]. Maltose binding protein was a notable survivor of thermolysin treatment, and it achieves its resistance through kinetic stability to unfolding. While it may be a challenge to apply to libraries of mutants, this screening principle may be useful to shed light on determinants of kinetic stability.

Split-protein reassembly, also called protein-fragment complementation, has proven useful for identifying

protein interactions in living cells, wherein reassembly of fragments of DHFR [11], GFP [12,13], luciferase [14] or other proteins is driven by the interaction between POIs fused to the fragments. Split fluorescent proteins reassemble irreversibly, making them useful for detecting weak interactions but generally unsuitable for measuring binding constants [15]. By contrast, split luciferase reassembles reversibly, which has been exploited to look at interaction dynamics in cells but so far not to look at stability directly [16]. Koide and co-workers recently combined yeast surface display with protein reassembly, and they demonstrated that FACS-detected reassembly could be used to measure stabilities and enrich in mutants with a defined range of stabilities [17[•]]. A human fibronectin type III domain (FN3) was split, with one fragment displayed on the cell surface and the other secreted into the medium. The fragments were fused to two epitopes for fluorescently labeled antibodies, such that FACS could resolve the display and reassembly levels on each cell. The log of this ratio correlated well with the change in binding energy ($R^2 = 0.8$) for a series of mutants that form a β -bulge in FN3. While not every protein will reassemble, and the binding energies may not perfectly match the stabilities of the full-length proteins, the ability to rapidly determine and sort for absolute protein stabilities is especially useful.

Despite the complicating irreversibility of split GFP reassembly, Linse and colleagues demonstrated that the split fragments of the B1 domain of protein G (GB1) could drive the reassembly of the known interaction-detection fragments of GFP, and that fluorescence was related to the thermal stability of the corresponding GB1 full-length protein with the same mutations [18]. This result is somewhat surprising, considering that in general cellular fluorescence from split GFP reassembly does not quantitatively correspond to the binding affinity [15]. This screen is also limited to proteins that can be dissected to reassemble, and while it lacks inherent controls for expression level differences, it is simpler in its implementation than yeast display complementation screening.

Waldo *et al.* introduced a screen for soluble proteins based on fusion of a 'folding reporter' GFP to the C-terminus of a POI [19]. The GFP only folds and becomes fluorescent if the fused POI folds and is soluble. One substantial improvement to the screen was the dissection of a 'superfolder' GFP into a tagging fragment of 15 amino acids from the C-terminus and a 215 aa 'detector' fragment [20^{••}]. These GFP fragments spontaneously reassemble, and the tag influences the solubility of the fusion less than the original folding reporter. Procedures for HT implementation have been described recently [21]. Solubility and stability are not generally directly related, but solubility is another key biophysical property that cannot be predicted and requires HT screening methods.

Direct, small-scale screens for stability

Of all the traditional methods of measuring protein stability, thermal and chemical denaturation monitored by intrinsic fluorescence of aromatic amino acids is the most straightforward to miniaturize. Stites *et al.* developed an early home-built autotitrator for semi-automated denaturation measurements [22]. Edgell, Pielak and colleagues carried out pioneering work in this area using autotitration methods and robotics with a standard fluorimeter [23]. Dalby and colleagues extended the method considerably by adapting it to microtiter plates with autotitration, which dramatically increased the throughput [24[•]]. Mayo and co-workers recently coupled this system with computational design of proteins libraries, enabling exhaustive characterization of computational predictions in a reasonable experimental timeframe [25^{••}]. Dalby and co-workers have gone on to further miniaturize their method to nanoliter scale using microfluidics, which enables screening with very small amounts of proteins (perhaps only 10⁸ molecules) [26]. This approaches a scale where concomitant miniaturization of protein production is a challenge for libraries, but it has great promise immediately for protein–ligand interactions. Of course, these methods rely on the presence of an intrinsic fluorophore, which not all proteins possess, and they require a fair amount of specialized equipment even in their simplest implementation. However, they are likely to produce measurements that directly compare to those taken by standard methods.

Hydrogen–deuterium exchange has been used extensively to measure the stability, dynamics and folding of proteins by NMR and mass spectrometry. Oas and co-workers developed a HT screen called stability of unpurified proteins from rates of H/D exchange, or SUPREX [27]. Cell lysate from 200 μ L of cell culture is exposed to a pulse of D₂O in varying concentrations of chemical denaturant, and the sample is dried with MALDI matrix for rapid acquisition. The method is complicated by aggregation or low expression. Also, EX2 conditions (wherein folding is faster than the intrinsic exchange rates of the protons) are required to extract thermodynamic parameters. But Oas has successfully used this method to measure protein stabilities in living cells [28]. (Ignatova and Gierasch have made similar measurements recently using biarsenical dyes as the readout instead of MALDI mass spectrometry [29].) Fitzgerald and colleagues recently described a variant of SUPREX based on oxidation rates (SPROX) that addresses some of the complications of H/D exchange for these sorts of experiments, such as resolving power, ion suppression, chromatographic separation and reversibility of modification [30].

Other reactions can also be used to monitor protein stability. For example, Silverman and Harbury developed a method called misincorporation proton–alkyl exchange

(MPAX), which uses weak missense suppressors to make random, residue specific Cys mutations throughout a protein of interest [31]. The burial of these Cys residues is interrogated by alkylation, which can be read out through mass spectrometry or chemical scission and PAGE. The method is especially useful for measuring the stabilities of proteins that do not refold reversibly, since the measurement is made under native conditions. Hellinga, Oas and colleagues have developed a related HT method called quantitative cysteine reactivity (QCR), which uses gel-shift as a readout, as well as a fast (fQCR) variant with a fluorescence readout [32,33[•]]. Like H/D exchange, thermodynamic parameters can only be extracted in the EX2 regime. Also, an appropriate buried Cys residue (ideally only one) is required, or the protein must be engineered with some peril of changing the protein thermodynamics. This method was demonstrated at picomole (nanogram) scale using HT gene fabrication and cell-free transcription-translation, which is a very exciting frontier in HT stability measurements.

A somewhat different measurement that is applicable to proteins that unfold irreversibly and aggregate—which represents a large fraction of interesting proteins—has been called differential static light scattering (DSLS). Senisterra *et al.* reported the use of a home-built instrument (which is now commercially available) that is capable of light scattering measurements of protein aggregation in 384-well format [34^{••}]. It is worth noting that 600 nm absorbance in a standard plate reader is also a reasonable way to measure aggregation. The chief advantage of this method is its simplicity, as no intrinsic or extrinsic probes are required. Besides the limitation to aggregating proteins, this non-equilibrium method could be confounded by dramatic changes in the kinetics of unfolding or aggregation for different mutants. But these effects appear to be small in proof-of-principle experiments.

A variation on this theme is isothermal denaturation (ITD), in which the rate of irreversible denaturation is observed, typically at a temperature just below that of melting. In principle, this denaturation can be observed by loss of a signal such as CD or shift of a signal such as UV absorbance or fluorescence. For proteins that aggregate, light scattering is also possible. ITD measurements are highly reproducible and have been reported to be more sensitive to small changes in stability, which is especially useful for ligand binding studies. Senisterra *et al.* adapted the method to HT using their 384-well scattering apparatus, with the additional advantage that ITD required less protein than comparable methods [35]. ITD measurements require *a priori* knowledge of the protein's approximate melting temperature, which could be problematic for protein libraries, and presumably could be very sensitive to changes in kinetics that may not be directly linked to equilibrium stability.

Schaeffer and colleagues have introduced an *in vitro* hybrid of the GFP fusion method for solubility and ITD, which can measure stability without purification [36[•]]. Here, the POI is fused to the N-terminus of GFP. The protein, purified or in lysate, is then subjected to ITD in HT format. The method, called GFP-Basta, is only applicable to proteins that aggregate upon unfolding and is limited by the GFP aggregation and photophysics, but practically these are not very significant limitations for most POIs.

One especially promising method called differential scanning fluorimetry (DSF) is simple, broadly applicable and requires little specialized equipment. A method called ThermoFluor was developed by 3D Pharmaceuticals, now owned by Johnson & Johnson, which reports on the perturbation of the melting temperature of a receptor by a potential ligand through addition of an extrinsic fluorophore [37[•]]. Hydrophobic dyes such as ANS are quenched in aqueous solvent but become fluorescent in organic solvent or when bound to molten globules or protein unfolding intermediates.

Most laboratory implementations of DSF, which is used extensively to optimize buffer conditions for crystallography [38,39], use real-time PCR machines that typically lack filter sets in the blue. Consequently, dyes such as SYPRO Orange have been widely used instead of ANS. RT-PCR machines enable DSF in 96 and 384-well formats with $\sim 20 \mu\text{L}$ of solution, where $\sim 1 \mu\text{g} \mu\text{L}^{-1}$ solutions are required. Nordlund and colleagues have shown that DSF is applicable to a broad range of proteins but that some proteins bind to SYPRO Orange in the folded state [38]. Magliery and co-workers demonstrated that for a series of related mutants of a protein, the correspondence between T_m values determined from CD thermal denaturation and DSF is excellent [40^{••}]. The reverse-format protein-engineering implementation of ThermoFluor, in which the conditions and ligands are held constant and the protein varied, was called High-Throughput Thermal Scanning (HTTS). It has been applied to core and loop libraries of four-helix bundle proteins to elucidate determinants of stability (Lavinder, J.J., Hari, S.B., Sen, S. and TJM, in preparation). DSF is surprisingly reversible through the melting point [41], although dye-protein aggregates appear upon extended heating of the denatured state. It is likely that the dye itself will perturb the apparent melting point, but the ΔT_m values correspond quite well to calorimetric and spectroscopic measurements.

All of these methods require miniaturization and high-throughput handling of protein expression, purification, and conceivably library construction. Growth of bacteria in 1–2 mL of culture in 96 deep-well plates is the technology of choice for most of these methods, where some amount of robotic liquid handling for plate or bead-based

affinity purifications (particularly IMAC) is helpful. For the most part, these are achieved with considerable home optimization at present. Platforms for HT oligonucleotide and gene synthesis [42,43] and *in vitro* protein expression [44] stand to expand the screening front-end further, but these are still far from straightforward implementation in most labs.

Membrane proteins and antibodies

Two targets of great interest in the pharmaceutical industry are particularly challenging for adapting to stability screens. Membrane proteins make up a large fraction of all drug targets, but they are difficult to work with *in vitro*, particularly for structural studies. Many recent successes in membrane protein crystallography have been born of strategies to stabilize the POI [45]. The most rapidly expanding area in pharmaceuticals is that of biologics, which antibodies and antibody-like molecules dominate at present. But their generally poor biophysical properties make them difficult to engineer and formulate as drugs.

Membrane proteins are often difficult to express or purify and are only stable in detergent or lipid formulations. Stevens and colleagues adapted cysteine reactivity reported by fluorescence, similar to the fQCR method, for membrane proteins in detergent [46]. They demonstrated its use on a lipophilic model protein, a monotopic membrane protein, and an integral membrane protein from the GPCR family. More recently, Cherezov, Stevens and colleagues have expanded this method by examining protein unfolding in the lipidic cubic phase with intrinsic fluorescence or fluorescence upon cysteine modification as a read-out (LCP- T_m) [47]. The method was additionally used for ITD over long time frames for membrane proteins. Baldwin and co-workers also developed an ITD screen for membrane proteins in detergent [48]. DSF has been applied to membrane proteins, but the high fluorescence background of the dye in detergent is a complicating factor [49]. Dyes that are more specific for proteins over lipids may improve this.

Even many full-length monoclonal antibodies are limited in their use as therapeutics by marginal stability and aggregation. Formats that are more straightforward to engineer and express, such as Fabs and scFvs, often suffer from decreased stability, and more significant engineering for humanization and generation of bispecific species can compromise stability further. DSF and DSLS have both been applied successfully to formulation studies of monoclonal antibodies [50,51]. Thermal inactivation screening has also been used as a means of establishing sufficient stability for scFv variants [52]. Cysteine reactivity has been applied to mAb stability, which seems particularly apt given the importance of disulfide bonds for antibody stability [53]. Little has been published on the application of these sorts of screens to

engineering antibody stability, but much of this work is behind industry doors at present.

Protein stability from sequence statistics

Most of the methods described above are useful for sorting out which members of a library are folded and stable. This enables the researcher to make mutations according to some hypothesis of design or even entirely at random, and locate variants with suitable physical properties. But it is often not simple to find rare stabilizing mutations, and such mutations may compromise other features of the protein, such as enzymatic function or expression level. An alternative approach to identifying sites of stabilizing mutations is to turn to statistical analysis of the natural repertoire of the motif, domain, protein or fold of interest. An attractive idea is that making mutations to the most common amino acid in some position of a protein is likely to be beneficial, and indeed these so-called ‘consensus’ mutations are tolerated and stabilizing far more frequently than at random or even from the best predictions today. But implementation is harder than it sounds. Multiple-sequence alignment (MSA) is often challenging, especially in poorly conserved regions or loops, leading to high noise. Most sites in proteins are not well conserved, and taking the most common amino acid in these positions is often little better than picking one at random. Moreover, some positions, especially weakly conserved ones, can be seen to vary together—that is, to be ‘correlated’—although these correlations are only sometimes close in space and are of uncertain significance in most cases (Box 1). The very large number of protein sequences available today makes these kinds of approaches worthy of greater attention in the years ahead.

Consensus

Steinbacher, Pluckthun and colleagues found that about half of the mutations made to an antibody V_{κ} domain were

Box 1 Consensus and correlation

A ‘consensus’ residue is simply the most common amino acid in one position of a family of proteins—that is, in a column of a multiple sequence alignment. It is not always easy to determine the consensus sequence of a protein, because it is not easy to align stretches of sequence that are poorly conserved or have insertions or deletions (such as loops). Also, many positions are only weakly conserved and may use nearly all 20 amino acids with some frequency. A correlation is, fundamentally, when a pair of residues in two positions is observed more or less frequently than expected by chance. For example, if Ala is seen in position A in 20% of sequences, and if it is in position B in 20% of sequences, then we would expect Ala–Ala pairs in A–B in 4% of sequences. If we observed it in all 20% of the sequences, that might represent a strong correlation. Compared to consensus, many more sequences are necessary to be confident of the significance of correlations since there are 400 possible pairs in any two positions. Information theory (e.g. relative entropy and mutual information) can be used to quantify how biased or conserved a position is, and how interconnected the distributions of two positions are.

stabilizing [54**]. Wirtz and Steipe went on to use this concept to generate hyperstable V_H domains for intracellular expression of Fvs in *E. coli* [55]. Wyss and colleagues made a number of variants of fungal phytases based on the consensus of a very small number of closely related sequences (less than 20), and they found that even the full consensus sequences were active and significantly thermostabilized [56,57]. (‘Full consensus’ means the most common amino acid from the MSA was used in every position of the protein.) More recent efforts have focused on the design of ubiquitous motifs, such as ankyrin repeats [58,59] and tetratricopeptide repeats [60]. Consensus variants of both of these repeats have been assembled into very stable domains and engineered using library and rational methods for novel binding properties.

The origin of consensus stabilization is not entirely clear. One possibility is that individual proteins in the MSA only avail themselves of as many stabilizing mutations as necessary for function, but that consensus amalgamates these mostly additive mutations. It is also not yet clear why only half of the mutations are stabilizing. Some light was shed on this recently by Arnold, Hilvert and co-workers, who showed that consensus mutations from library selections were also stabilizing [61*]. The authors suggested that this method benefited from the removal of phylogenetic artifacts. Other factors stemming from correlation and poorly conserved residues also probably play a role. A related approach to making consensus mutations is to make ‘ancestral’ mutations by tracing mutations back to early sequences along the phylogenetic path. These mutations also turn out to be stabilizing about half the time [62]. Tawfik and colleagues have incorporated ancestral mutations into the family shuffling of paraoxonase-3 to successfully yield stable, active chimeric enzymes [63].

Correlation

An additional layer of complication in the statistical analysis of MSAs is that not all positions are statistically independent. Ranganathan and colleagues developed a method called statistical coupling analysis (SCA), which is a perturbation-based approach to identifying overrepresented pairs of amino acids [64]. They showed that inclusion of both consensus and correlation information was necessary for the design of folded WW domains—meaning that variants that were plausible from positional distributions alone were often unfolded if they did not capture enough correlation information [65**]. The meaning of these types of correlations is even less well understood than the etiology of consensus stabilization. Many correlated residues are not close in space, but many can also be assembled into networks of interacting residues that connect distant regions of a protein fold. A number of studies have identified roles for correlated residues in allosteric regulation [66–68]. Ranganathan has recently

devised a new kind of SCA calculation and has used it to identify independent clusters of co-evolving residues in protein families (sectors) [69]. Mutations to different sectors in trypsin demonstrated that one had structural and the other had functional consequences. There is a great deal of work that lies ahead to understand the meaning of correlations in a general way.

Magliery and Regan applied consensus analysis and SCA to TPR motifs, which uncovered two subfamilies with distinct alternative networks of interacting residues and resulted in an algorithm for identifying active-site residues [70,71]. Among the most striking of the results of that study was an explanation for the unusually high charge of a consensus TPR motif (-7) despite the average TPR having a zero net charge. Charge neutralization occurred by correlations in weakly conserved positions on the surface of the motif. This effect was not always local (for example, two residues close in space forming a salt bridge), suggesting at least one mechanism for important non-local correlations. Magliery and colleagues recently engineered two closely related consensus variants of triosephosphate isomerase (TIM) from slightly different sequence databases (BJS, Durani, V. and TJM, submitted). The two variants differed dramatically in their physical properties and activity, with one of them having wild-type like kinetics and the other being weakly active and poorly folded. Both variants were much more substantially different from any natural TIM than the two variants were from each other. The only apparent difference between these two variants is the extent to which they complete networks of correlated residues. This type of host-guest approach will hopefully shed light on the physical meaning of correlated positions.

Conclusions and outlook

First-principles computational methods are likely to remain far from a comprehensive predictive model for some time, until better potential functions and better treatments of the unfolded state can be incorporated. Even empirical parameterization is very difficult given our sparse coverage of sequence space in thermodynamics studies. But there has been a dramatic increase in efforts to bridge that gap in the past 5–10 years in the form of new HT screens for foldedness, thermodynamic stability, solubility and kinetic stability. In the next decade, with improved methods of HT gene construction, handling, expression and purification, these new stability screening methods will give us a vastly richer and more detailed view of the effects of mutations on protein physical properties. And in the meantime, these methods are immediately adaptable for screening random libraries to improve the physical properties of proteins for easier handling, crystallization and structural studies, and superior biotherapeutics. Application of these methods to biophysically ‘unfriendly’ proteins that are larger, more complex and do not refold spontaneously is likely to

change our view of protein folding for the majority of proteins.

Random screening in the absence of any information is often a slog, and any information to narrow down libraries to find stabilizing mutations is welcome. Protein sequence statistics can be a useful tool for guiding combinatorial experiments and limiting possibilities in difficult engineering experiments. The molecular etiology of the effects of consensus and correlated mutations remains a difficult problem, but the combination of screening methods with these kinds of calculations in the next decade will accelerate research towards an understanding of those effects.

Protein stability remains one of the most difficult problems in protein science, but its illumination by experiments that take advantage of large numbers, both experimentally and statistically, offers new hope for a solution in the years ahead.

Acknowledgements

The authors thank the NIH (R01 GM083114 and U54 NS058183 to TJM) and The Ohio State University for support. JJJ was an NIH CBIP fellow and a fellow of the Great Rivers affiliate of the AHA. BJS was an NIH CBIP fellow and is a Presidential fellow of The Ohio State University.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Dill KA: **Dominant forces in protein folding.** *Biochemistry* 1990, **29**:7133-7155.
 2. Rose GD, Wolfenden R: **Hydrogen bonding, hydrophobicity, packing, and protein folding.** *Annu Rev Biophys Biomol Struct* 1993, **22**:381-415.
 3. Magliery TJ, Regan L: **Combinatorial approaches to protein stability and structure.** *Eur J Biochem* 2004, **271**:1595-1608.
 4. Bommaris AS, Broering JM, Chaparro-Riggers JF, Polizzi KM: **High-throughput screening for enhanced protein stability.** *Curr Opin Biotechnol* 2006, **17**:606-610.
 5. Roodveldt C, Aharoni A, Tawfik DS: **Directed evolution of proteins for heterologous expression and stability.** *Curr Opin Struct Biol* 2005, **15**:50-56.
 6. Senisterra GA, Finerty PJ Jr: **High throughput methods of assessing protein stability and aggregation.** *Mol Biosyst* 2009, **5**:217-223.
 7. Giver L, Gershenson A, Freskgard PO, Arnold FH: **Directed evolution of a thermostable esterase.** *Proc Natl Acad Sci USA* 1998, **95**:12809-12813.
 8. Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA,
 - Titchmarsh J, Warriner SL, Radford SE, Bardwell JC: **Optimizing protein stability in vivo.** *Mol Cell* 2009, **36**:861-871.
 Since this method is based on a tunable selection, it can be used to very rapidly identify stable, expressible variants from large libraries.
 9. Park C, Marqusee S: **Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding.** *Nat Methods* 2005, **2**:207-212.
 10. Park C, Zhou S, Gilmore J, Marqusee S: **Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis.** *J Mol Biol* 2007, **368**:1426-1437.

Interestingly, protease-resistant proteins were found to lack unstructured regions and have high kinetic stability, rather than be sequences that avoid proteolytic specificity. The sequence basis of kinetic stability is very poorly understood, but kinetic stability is especially advantageous for practical applications such as therapeutics.

11. Pelletier JN, Campbell-Valois FX, Michnick SW: **Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments.** *Proc Natl Acad Sci USA* 1998, **95**:12141-12146.
12. Ghosh I, Hamilton AD, Regan L: **Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein.** *J Am Chem Soc* 2000, **122**:5658-5659.
13. Hu CD, Chinenov Y, Kerppola TK: **Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation.** *Mol Cell* 2002, **9**:789-798.
14. Paulmurugan R, Umezawa Y, Gambhir SS: **Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies.** *Proc Natl Acad Sci USA* 2002, **99**:15608-15613.

15. Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L: **Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism.** *J Am Chem Soc* 2005, **127**:146-157.
16. Stefan E, Aquin S, Berger N, Landry CR, Nyfeler B, Bouvier M, Michnick SW: **Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo.** *Proc Natl Acad Sci USA* 2007, **104**:16916-16921.

17. Dutta S, Koide A, Koide S: **High-throughput analysis of the protein sequence-stability landscape using a quantitative yeast surface two-hybrid system and fragment reconstitution.** *J Mol Biol* 2008, **382**:721-733.

While not all proteins are likely to reassemble on dissection, this method does not require any intrinsic probe or property of the analyte protein. It also has excellent potential for engineering stability into protein complexes, where no engineered dissection point is required.

18. Lindman S, Hernandez-Garcia A, Szczepankiewicz O, Frohm B, Linse S: **In vivo protein stabilization based on fragment complementation and a split GFP system.** *Proc Natl Acad Sci USA* 2010, **107**:19826-19831.
19. Waldo GS, Standish BM, Berendzen J, Terwilliger TC: **Rapid protein-folding assay using green fluorescent protein.** *Nat Biotechnol* 1999, **17**:691-695.
20. Cabantous S, Terwilliger TC, Waldo GS: **Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.** *Nat Biotechnol* 2005, **23**:102-107.

Like kinetic stability, the sequence basis of solubility is also very poorly understood, and yet it is crucial for most practical applications of proteins. The tagging peptide approach is minimally perturbing compared to the original folding reporter GFP.

21. Listwan P, Terwilliger TC, Waldo GS: **Automated, high-throughput platform for protein solubility screening using a split-GFP system.** *J Struct Funct Genomics* 2009, **10**:47-55.
22. Stites WE, Byrne MP, Aviv J, Kaplan M, Curtis PM: **Instrumentation for automated determination of protein stability.** *Anal Biochem* 1995, **227**:112-122.
23. Edgell MH, Sims DA, Pielak GJ, Yi F: **High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer.** *Biochemistry* 2003, **42**:7587-7593.
24. Aucamp JP, Cosme AM, Lye GJ, Dalby PA: **High-throughput measurement of protein stability in microtiter plates.** *Biotechnol Bioeng* 2005, **89**:599-607.

This microtiter plate implementation of an intrinsic fluorescence assay for chemical stability works with very small amounts of protein and makes measurements that are directly comparable to reliable low-throughput approaches. It also is ideally suited to extract ligand binding constants.

25. Allen BD, Nisthal A, Mayo SL: **Experimental library screening demonstrates the successful application of computational**

protein design to large structural ensembles. *Proc Natl Acad Sci USA* 2010, **107**:19838-19843.

An exhaustive examination of computationally guided library designs was enabled by the microtiter intrinsic fluorescence approach (ref. [24*]).

26. Gaudet M, Remtulla N, Jackson SE, Main ER, Bracewell DG, Aeppli G, Dalby PA: **Protein denaturation and protein:drugs interactions from intrinsic protein fluorescence measurements at the nanolitre scale.** *Protein Sci* 2010, **19**:1544-1554.
27. Ghaemmghami S, Fitzgerald MC, Oas TG: **A quantitative, high-throughput screen for protein stability.** *Proc Natl Acad Sci USA* 2000, **97**:8296-8301.
28. Ghaemmghami S, Oas TG: **Quantitative protein stability measurement in vivo.** *Nat Struct Biol* 2001, **8**:879-882.
29. Ignatova Z, Gierasch LM: **Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling.** *Proc Natl Acad Sci USA* 2004, **101**:523-528.
30. West GM, Tang L, Fitzgerald MC: **Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy.** *Anal Chem* 2008, **80**:4175-4185.
31. Silverman JA, Harbury PB: **Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange.** *J Biol Chem* 2002, **277**:30968-30975.
32. Isom DG, Marguet PR, Oas TG, Hellinga HW: **A miniaturized technique for assessing protein thermodynamics and function using fast determination of quantitative cysteine reactivity.** *Proteins* 2010.
33. Isom DG, Vardy E, Oas TG, Hellinga HW: **Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity.** *Proc Natl Acad Sci USA* 2010, **107**:4908-4913.

Borrowing from the mathematics of H/D exchange, chemical reactivity of cysteine residues requires less specialized equipment and is effective at a very small scale for measuring stability and binding affinity. This demonstration is interfaced with HT gene synthesis and cell-free expression, setting the stage for a very HT approach to libraries of distantly related proteins. In the paper, the readout was electrophoretic mobility, but an even faster approach (ref. [32]) is to use a fluorogenic reagent.

34. Senisterra GA, Markin E, Yamazaki K, Hui R, Vedadi M, Awrey DE: **Screening for ligands using a generic and high-throughput light-scattering-based assay.** *J Biomol Screen* 2006, **11**:940-948.

A large number of proteins of practical interest unfold irreversibly and aggregate when they do so. This can be observed without any intrinsic or extrinsic probe, making it suitable for a very wide range of proteins and conditions.

35. Senisterra GA, Soo Hong B, Park HW, Vedadi M: **Application of high-throughput isothermal denaturation to assess protein stability and screen for ligands.** *J Biomol Screen* 2008, **13**:337-342.
36. Moreau MJ, Morin I, Schaeffer PM: **Quantitative determination of protein stability and ligand binding using a green fluorescent protein reporter system.** *Mol Biosyst* 2010, **6**:1285-1292.

By fusing an irreversibly unfolding protein to GFP, isothermal denaturation can be observed for virtually any protein in a simple UV-VIS plate reader. And no purification is required as long as the expression levels are sufficient.

37. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P *et al.*: **High-density miniaturized thermal shift assays as a general strategy for drug discovery.** *J Biomol Screen* 2001, **6**:429-440.

The ThermoFluor assay was originally designed to screen for ligand binding in a pseudo-calorimetric way that could easily be scaled and did not depend on any particular property of the receptor. The extremely simple implementation of dye addition and heating in a multiwell fluorimeter makes this screen widely accessible.

38. Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P: **ThermoFluor-based high-throughput stability optimization of proteins for structural studies.** *Anal Biochem* 2006, **357**:289-298.

39. Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H *et al.*: **Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination.** *Proc Natl Acad Sci USA* 2006, **103**:15835-15840.
40. Lavinder JJ, Hari SB, Sullivan BJ, Magliery TJ: **High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering.** *J Am Chem Soc* 2009, **131**:3794-3795.
- Unrelated proteins bind differently to the reporter dyes in DSF, so it was not clear if it could be used for protein engineering. Here the melting temperatures of a series of related core mutants of a model protein were measured by CD and DSF, and the changes in T_m s were well-correlated.
41. Layton CJ, Hellinga HW: **Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes.** *Biochemistry* 2010, **49**:10831-10841.
42. Cox JC, Lape J, Sayed MA, Hellinga HW: **Protein fabrication automation.** *Protein Sci* 2007, **16**:379-390.
43. Kosuri S, Eroshenko N, Leproust EM, Super M, Way J, Li JB, Church GM: **Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips.** *Nat Biotechnol* 2010, **28**:1295-1299.
44. Jewett MC, Swartz JR: **Rapid expression and purification of 100 nmol quantities of active protein using cell-free protein synthesis.** *Biotechnol Prog* 2004, **20**:102-109.
45. Rosenbaum DM, Rasmussen SG, Kobilka BK: **The structure and function of G-protein-coupled receptors.** *Nature* 2009, **459**:356-363.
46. Alexandrov AI, Mileni M, Chien EY, Hanson MA, Stevens RC: **Microscale fluorescent thermal stability assay for membrane proteins.** *Structure* 2008, **16**:351-359.
47. Liu W, Hanson MA, Stevens RC, Cherezov V: **LCP-Tm: an assay to measure and understand stability of membrane proteins in a membrane environment.** *Biophys J* 2010, **98**:1539-1548.
48. Postis VL, Deacon SE, Roach PC, Wright GS, Xia X, Ingram JC, Hadden JM, Henderson PJ, Phillips SE, McPherson MJ *et al.*: **A high-throughput assay of membrane protein stability.** *Mol Membr Biol* 2008, **25**:617-624.
49. Yeh AP, McMillan A, Stowell MH: **Rapid and simple protein-stability screens: application to membrane proteins.** *Acta Crystallogr D Biol Crystallogr* 2006, **62**:451-457.
50. He F, Hogan S, Latypov RF, Narhi LO, Razinkov VI: **High throughput thermostability screening of monoclonal antibody formulations.** *J Pharm Sci* 2010, **99**:1707-1720.
51. Goldberg DS, Bishop SM, Shah AU, Sathish HA: **Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: role of conformational and colloidal stability.** *J Pharm Sci* 2010.
52. Miller BR, Demarest SJ, Lugovskoy A, Huang F, Wu X, Snyder WB, Croner LJ, Wang N, Amatucci A, Michaelson JS *et al.*: **Stability engineering of scFvs for the development of bispecific and multivalent antibodies.** *Protein Eng Des Sel* 2010, **23**:549-557.
53. Lacy ER, Baker M, Brigham-Burke M: **Free sulfhydryl measurement as an indicator of antibody stability.** *Anal Biochem* 2008, **382**:66-68.
54. Steipe B, Schiller B, Pluckthun A, Steinbacher S: **Sequence statistics reliably predict stabilizing mutations in a protein domain.** *J Mol Biol* 1994, **240**:188-192.
- Although it makes intuitive sense that consensus mutations might be tolerated or even beneficial, it certainly is not obvious that so many consensus mutations would be stabilizing or that the mutations would have largely additive effects. Surprisingly little more has been uncovered about how to choose consensus mutations or how they work than the authors found in this work.
55. Wirtz P, Steipe B: **Intrabody construction and expression III: engineering hyperstable V(H) domains.** *Protein Sci* 1999, **8**:2245-2250.
56. Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon AP, Wyss M: **The consensus concept for thermostability engineering of proteins: further proof of concept.** *Protein Eng* 2002, **15**:403-411.
57. Lehmann M, Pasamontes L, Lassen SF, Wyss M: **The consensus concept for thermostability engineering of proteins.** *Biochim Biophys Acta* 2000, **1543**:408-415.
58. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, Grutter MG: **Designed to be stable: crystal structure of a consensus ankyrin repeat protein.** *Proc Natl Acad Sci USA* 2003, **100**:1700-1705.
59. Mosavi LK, Minor DL Jr, Peng ZY: **Consensus-derived structural determinants of the ankyrin repeat motif.** *Proc Natl Acad Sci USA* 2002, **99**:16029-16034.
60. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L: **Design of stable alpha-helical arrays from an idealized TPR motif.** *Structure* 2003, **11**:497-508.
61. Jackel C, Bloom JD, Kast P, Arnold FH, Hilvert D: **Consensus protein design without phylogenetic bias.** *J Mol Biol* 2010, **399**:541-546.
- The consensus sequence of mutants from selection from a combinatorial library was also found to be stabilized. Interestingly, while these consensus mutants had good activity, there was not a correlation between how many consensus residues a mutant had and its activity. Experiments such as this one to tease out the effects of different selective pressures in evolution will help us understand the origin of consensus stabilization.
62. Watanabe K, Ohkuri T, Yokobori S, Yamagishi A: **Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree.** *J Mol Biol* 2006, **355**:664-674.
63. Khersonsky O, Rosenblat M, Toker L, Yacobson S, Hugenmatter A, Silman I, Sussman JL, Aviram M, Tawfik DS: **Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization.** *Biochemistry* 2009, **48**:6644-6654.
64. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286**:295-299.
65. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R: **Evolutionary information for specifying a protein fold.** *Nature* 2005, **437**:512-518.
- This experiment unambiguously shows that design merely from the distribution of amino acids at each position of an MSA is not as effective as also considering sequence correlation in engineering folded, soluble proteins. While the exact molecular explanation for this effect is not entirely clear, this work strongly motivates studies to understand the effects of correlated mutations, which generally occur in weakly conserved positions.
66. Hatley ME, Lockless SW, Gibson SK, Gilman AG, Ranganathan R: **Allosteric determinants in guanine nucleotide-binding proteins.** *Proc Natl Acad Sci USA* 2003, **100**:14445-14450.
67. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R: **Surface sites for engineering allosteric control in proteins.** *Science* 2008, **322**:438-442.
68. Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R, Gierasch LM: **An interdomain sector mediating allostery in Hsp70 molecular chaperones.** *Mol Syst Biol* 2010, **6**:414.
69. Halabi N, Rivoire O, Leibler S, Ranganathan R: **Protein sectors: evolutionary units of three-dimensional structure.** *Cell* 2009, **138**:774-786.
70. Magliery TJ, Regan L: **Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif.** *J Mol Biol* 2004, **343**:731-745.
- A detailed study of correlation in a small repeat motif demonstrated both that alternative networks of interacting residues could be found in the same alignment, and that some correlation effects, such as charge compensation, do not reflect the interaction of two residues close in space.
71. Magliery TJ, Regan L: **Sequence variation in ligand binding sites in proteins.** *BMC Bioinformatics* 2005, **6**:240.