Limitations of yeast surface display in engineering proteins of high thermostability

Sheldon Park1, Yao Xu1, Xiaoran Fu Stowell1,4, Feng Gai1,4, Jeffery G. Saven1,2 and Eric T. Boder2,3,5

1Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, 2Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104 and 3Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, PA 19104, USA
4Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
5To whom correspondence should be addressed.
E-mail: boder@seas.upenn.edu

Engineering proteins that can fold to unique structures remains a challenge. Protein stability has previously been engineered via the observed correlation between thermal stability and eukaryotic secretion level. To explore the limits of an expression-based approach, variants of the highly thermostable three-helix bundle protein α3D were studied using yeast surface display. A library of α3D mutants was created to explore the possible correlation of protein stability and fold with expression level. Five efficiently expressed mutants were then purified and further studied biochemically. Despite their differences in stability, most mutants expressed at levels comparable with that of wild-type α3D. Two other related sequences (α2A and α3B) that form collapsed, stable molten globules but lack a uniquely folded structure were similarly expressed at high levels by yeast display. Together these observations suggest that the quality control system in yeast is unable to discriminate between well-folded proteins of high stability and molten globules. The present study, therefore, suggests that an optimization of the surface display efficiency on yeast may yield proteins that are thermally and chemically stable yet are poorly folded.

Keywords: α3D/helical protein/protein library/yeast surface display

Introduction

Making detailed biophysical measurements on a large number of protein variants is highly laborious and adds to the difficulty of studying the sequence–structure relationship in proteins. This problem is particularly poignant when studying proteins for which a functional assay, such as ligand binding or catalysis, cannot be easily devised. In such cases, one necessarily resorts to often slow biochemical and biophysical techniques for characterization of structure and stability to deduce the underlying physical principles. Since most available biochemical/biophysical techniques are best suited for low throughput analysis, a high throughput screen to correlate sequence with structure would vastly accelerate the engineering of stably folded proteins. In the present study, we used yeast surface display to explore if an efficient readout method may be developed to relate protein sequence with structure. Less stable proteins spend more time in partially unfolded conformations. Since unfolding usually leads to loss of function and may cause aggregation, the relative stability between native and denatured states is a crucial parameter to optimize when considering protein-based applications. The sequence–structure relationship may be studied independently of protein function using a proteolysis-based phage library screen (Sieber et al., 1998; Finucane et al., 1999). In one implementation, the protein of interest is expressed between a coat protein of the phage (e.g. pIII) and an anchoring tag (e.g. the hexahistidine sequence). During the subsequent adsorption–proteolysis–release–amplification cycle, unstable mutants are depleted from the pool due to their increased proteolytic susceptibility, thus enriching the population of stability-enhanced mutants in the library. Additionally, an added infectivity domain may be used to allow selective amplification of stable mutants following a protease treatment (Kristensen and Winter, 1998; Sieber et al., 1998). Protease sensitivity, a hallmark of unstable proteins, thus allows indirect discrimination between unstable and stable proteins.

As an alternative to phage, yeast display has a number of attractive features as a protein engineering platform (Boder and Wittrup, 1997; 1998; 2000). Yeast libraries permit the use of the eukaryotic transcriptional and translational machinery for protein expression. Yeast libraries may be screened with flow cytometry, thus providing quantitative analysis of individual clones in real time. Additionally, studies of single-chain antibodies, single-chain T-cell receptors, major histocompatibility complexes, and bovine pancreatic trypsin inhibitor have suggested that the quality control (QC) system in yeast modulates the display or secretion efficiency of a protein based on its thermal stability (Kowalski et al., 1998a, b; Kieke et al., 1999; Shusta et al., 1999; Shusta et al., 2000; Hagihara and Kim, 2002; Starwalt et al., 2003; Esteban and Zhao, 2004). The reported degradation of unstable or misfolded proteins in the endoplasmic reticulum (ER) is attributed to the QC apparatus consisting of ER-resident chaperones such as BiP, GRP94, calnexin and protein disulfide isomerase, and provides a built-in screening system against poorly folded peptide variants to augment the overall quality of the peptide library (Ellgaard and Helenius, 2003). The possibility of preferentially expressing stability-enhanced mutants, therefore, creates a unique opportunity to develop a general strategy based on the expression level for engineering de novo proteins that are not only stable but well-folded as well.

Although the relation between protein stability and secretion efficiency in yeast has been examined, it is not clear whether the proofreading machinery can discriminate well-structured proteins from collapsed, molten-globule structures independent of their stabilities. The structural fidelity of the QC machinery may be compromised when applied to variants of a protein having extreme stability,
e.g. proteins from thermophiles, where the relative significance of a perturbation imparted by a mutation may be less than in a protein of marginal stability. Herein, ‘high thermal stability’ includes proteins having no native structure but that, upon thermal or chemical denaturation, exhibit a transition from a collapsed, molten-globule state to a largely unfolded state. Furthermore, most proteins that have been engineered in yeast contain one or more disulfide bonds in the native structure, leaving the question open whether the lessons from these studies extend to proteins such as the three-helix bundle α3D, that are stabilized through noncovalent tertiary interactions. Since intracellular proteins often lack disulfide bonds, an examination of the expression profile of α3D and its mutants in yeast may prove a model system for engineering other disulfide-free proteins in the future. In addition to serving a broad range of biological functions (Schneider et al., 1998), helical proteins also provide a useful test case since the fold is based on well-characterized secondary structural motifs, α helix and interhelical turns, making biophysical experiments easier to interpret than those involving proteins of greater structural complexity.

We expressed α3D and a library of α3D mutants on yeast with the goals of: (i) testing the utility of the eukaryotic folding QC mechanism in selecting well-structured, stable helical proteins; and (ii) engineering proteins of very high thermal stability by yeast display. Our studies suggest that, within the context of high melting temperature parent phenotypes, yeast expression level has insufficient resolution to distinguish structural variations. To that end we show that engineered three-helix bundle proteins of high stability, α3A, α3B and α3D, are all expressed at comparable levels on yeast, although only α3D has a well-formed tertiary structure. Structural heterogeneity and the absence of well-defined structures have previously been shown for α3A and α3B (Bryson et al., 1998). We therefore conclude that the intrinsic stability of a protein as well as the nature of stabilizing interactions may limit the effectiveness of yeast display in selecting and identifying well-structured proteins.

### Methods

**Yeast display and flow cytometry**

A synthetic gene containing residues 2–73 of de novo designed α3D was constructed from four overlapping primers, which were PCR extended to yield the full-length sequence. A library of α3D mutants was constructed using degenerate oligonucleotides with mutations at positions that encode the core residues of the third helix. The composition of the bases at each randomized position was adjusted to substitute four aliphatic side chains (V, I, L and M) at V53, L56, A60, I63 and L67 and to replace Y70 with an equal mix of Y and F (Table I). After PCR amplification, the full-length gene product was digested with NheI and XhoI restriction enzymes and ligated in frame to the C-terminus of Aga2. The library was transformed into EBY100 by the lithium acetate method (Gietz and Woods, 2002), which yielded ∼10^5 independent clones. The culture was grown in a glucose-containing medium to an optical density of 600 nm (OD_{600}) of 2.0 and was switched to a galactose-containing medium for 16 h at 30°C to induce protein expression. The epitopes were detected using either monoclonal c-Myc antibody (9E10) or monoclonal biotinylated HA antibody (12CA5), and visualized using either phycoerythrin-conjugated anti-mouse rabbit antibody (c-Myc) or quantum red-conjugated streptavidin (HA). Fluorescence-assisted cell sorting was performed on Vantage Diva (Beckton Dickerson).

### Computational sequence probabilities and energies

The sequence probabilities and conformational energies of the α3D mutants were computed as described previously (Kono and Saven, 2001; Calhoun et al., 2003; Fu et al., 2003). The template structures used in the calculations were the 13 NMR structures of wild-type α3D [where ‘wild type’ refers to the originally designed α3D sequence as published in Walsh et al. (1999); NMR coordinates kindly provided by W. F. DeGrado, University of Pennsylvania].

### Soluble protein production

Five α3D mutants selected from the screen were PCR amplified and cloned into the GST expression vector with the Tobacco Etch Virus (TEV) cleavage site between GST and α3D (the plasmid was a gift from G. Van Duyne, University of Pennsylvania). The fusion protein was expressed in *Escherichia coli* BL21 (DE3) (Novagen) by adding 100 μM isopropyl-β-d-thiogalactoside to 250 ml Luria–Bertani culture at OD_{600} = 0.4–0.6 for 3 h. The bacterial pellet was obtained by centrifugation at 4°C, lysed with 10 ml of B-Per (Pierce), and mixed with 2 ml of Glutathione resin (Invitrogen) pre-equilibrated in 50 mM Tris, pH 8.0 and 100 mM NaCl (Buffer A). Phenylmethylsulfonyl fluoride (1 mM) was added to prevent proteolytic degradation during binding (2 h at 4°C), after which the resin was washed with at least 100 ml of Buffer A. α3D mutants were recovered by adding 10 μl of TEV protease (Invitrogen) directly to the resin-bound GST fusion at 22°C. Additional 5 μl of TEV was added 18 h later and the proteolysis was allowed to proceed for additional 48 h. The α3D mutants released into the solution were filtered through a Centricon concentrator (Amicon) with 50 kDa cutoff to remove high molecular weight species and finally concentrated using a Centricon with 3 kDa cutoff. The purity of the protein was checked by running an SDS–PAGE gel, which showed a single band when stained with Coomassie blue.

### Fourier transform infrared spectroscopy

The thermal denaturation of each mutant was monitored using Fourier transform infrared (FTIR) spectroscopy. Temperature-dependent FTIR spectra were collected on a

<table>
<thead>
<tr>
<th>Codon position</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliphatic</td>
<td>1</td>
<td>57</td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>19</td>
<td>15</td>
<td>37</td>
</tr>
<tr>
<td>Aromatic</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>42</td>
<td>0</td>
</tr>
</tbody>
</table>

A degenerate DNA primer was synthesized with the corresponding mix of DNA bases and used to assemble the full α3D gene for the library.

---

**Table I.** The percentage of the bases in the codon used to encode residue positions 53, 56, 60, 63, 67 (aliphatic) and 70 (aromatic), which were replaced with other hydrophobic residues in the library.
Magn-I-IR 860 spectrometer (Nicolet, WI) using 2 cm⁻¹ resolution. A CaF₂ sample cell that is divided into two compartments using a Teflon spacer of 52 µm was employed to allow the separate measurements of the sample and buffer under identical conditions. Temperature control was achieved by a thermostatted copper block. To correct for slow instrument drift, a computer controlled translation stage was used to move both the sample and buffer side of the sample cell in and out of the infrared beam alternately after each spectrum of eight scans. The final results consist of an average of 32 such spectra, both for sample and buffer. Protein solutions used in the FTIR measurements were prepared by dissolving lyophilized protein samples into 50 mM Tris D₂O solution (pH = 7.6) to the final concentration of ~3 mg/ml. Using a global fitting procedure (Zhu et al., 2004), the temperature-dependent FTIR spectra (in the amide I' region) were analyzed to yield information regarding the thermal stability of the protein of interest. Specifically, the reported thermal melting behavior of the α3D mutants was determined from the band centered at ~1660 cm⁻¹ that is associated with the non-helical amides (Zhu et al., 2003, 2004).

**Fluorescence spectroscopy**

α3D and its mutants studied here each contain a single Trp residue. Guanidine hydrochloride (GdnHCl) induced unfolding of these proteins was monitored by measuring the Trp fluorescence at 22°C as a function of the molar concentration of the denaturant [GdnHCl]. The fluorescence spectra were collected using a quartz cuvette with a 1 cm path length on a Fluorolog 3.10 spectrofluorometer (Jobin Yvon Horiba, NJ) with λex = 290 nm and 1 nm spectral resolution. The protein concentration, determined by Trp absorbance at 280 nm, ranged from 3 to 4 µM (at pH 7.6). Both wild type and the mutants showed a single Trp emission band near 335 nm. Addition of GdnHCl shifted the emission peak to a higher wavelength of 350 nm, and simultaneously reduced the peak intensity. The integrated Trp emission peak area as a function of [GdnHCl] (0–6 M) is well described by a two-state unfolding model (Walsh et al., 2001b), and the free energy of unfolding was treated as a linear function of [GdnHCl], i.e. ΔGobs = ΔGU + m*[GdnHCl], where ΔGU is the free energy for unfolding in water (0 M GdnHCl) and m measures the degree of folding/unfolding cooperativity. Cm, the GdnHCl concentration corresponding to ΔGobs = 0, is reported in Table II.

**Results and discussion**

**Yeast expression of α3D**

The stable three-helix bundle α3D has previously been engineered *de novo* using a combination of structural motif design and computational sequence calculations (Bryson et al., 1998). Its sequence includes structural elements known to stabilize the formation of secondary structures, such as optimized turn sequences and capping residues. Hairpin loop sequences were chosen to bias the directionality of interhelical turns, while the formation of alternate competing structures was prevented through strategic placement of charged residues. Its hydrophobic interior residues were identified through iterative application of a genetic algorithm. The protein has well-separated ¹H NMR peaks and its binding of the hydrophobic dye ANS is comparable with that of the native protein lysozyme (Bryson et al., 1998). Although a high resolution structure of the protein obtained by NMR validates a three-helix structure (Walsh et al., 1999), the hydrophobic core of α3D is still more fluctuational (Walsh et al., 2001b) and its side chains are more dynamic than in natural proteins (Walsh et al., 2001a). Therefore, the designed protein lacks some of the structural and dynamic characteristics of a natural protein, and the quality of its core packing may be further improved.

Studies have shown that yeast modulates protein expression based on thermal stability as long as the folding rate is not limiting (Parekh and Wittrup, 1997; Kowalski et al., 1998a; Kieke et al., 1999; Shusta et al., 1999). Both ER-associated degradation (Kostova and Wolf, 2003) and lysosomal proteolysis (Coughlan et al., 2004) have been suggested as the basis of yeast proofreading activity. Nevertheless, the generality of this proofreading activity in regulating the quality of secreted proteins has not been established. In particular, it is not clear whether the activity is applicable to properties other than melting temperature. Based on the large number of enzymes and chaperones implicated in the processing of extracellular proteins in the yeast ER (Wang and Chang, 1999; Jarosch et al., 2002; Elkabetz et al., 2004; Smith et al., 2004), it is possible that proteins destined for secretion undergo a rigorous structural screening before they exit from the ER. We tested the response of the yeast proofreading machinery to a protein with a fluctuating core by expressing α3D using yeast surface display (Figure 1) (Boder and Wittrup, 2000). The high melting temperature (Tm ~93°C) of α3D ensures the protein is folded at the induction temperature of 30°C (Walsh et al., 1999) and allows us to examine the effectiveness of the yeast screen in discriminating well-ordered proteins.

The protein is tagged with the c-Myc epitope (amino acid sequence: EQKLISEEDL) at the C-terminus and with the HA epitope (amino acid sequence: YPYDVPDYA) between Aag2p and α3D. These epitopes are used for immunodetection of the complex with monoclonal antibodies and fluorophore-conjugated secondary antibodies. Intracellular proteolysis of unstable or misfolded protein would lead to loss of c-Myc and result in reduced fluorescence. The observed fluorescence level of α3D-expressing yeast cells by flow cytometry shows that the protein is expressed intact at a high level (Figure 1C).

**Yeast display of a three-helix bundle library**

We investigated whether the stability of expressed α3D depends on the precise sequence of core residues by constructing a patterned α3D library, in which the interior residues from the third helix were randomized to other hydrophobic residues. The conservative nature of the substitutions are likely to maintain the three-helix bundle structure of α3D, yet such changes in the hydrophobic core may result in α3D variants of differing stability. We compared the expression pattern of the library with that of wild-type α3D. The majority of the mutants were labeled efficiently for both HA and c-Myc and thus appeared along the diagonal of the cytogram, where the c-Myc to HA ratio (R) is close to unity (Figure 2). A small set of clones well separated from the main group demonstrated values of the ratio R substantially smaller than unity (R ~0.02). However,

213
sequencing ten clones from this region revealed that they each contained one or more premature stop codons in the gene introduced during PCR, which resulted in loss of the c-Myc tag. For the majority of the mutants, comprising 2000 different sequences, conservative mutations in the hydrophobic core do not appear to significantly alter the expression level.

Previously, differences in melting temperature of 10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.

Previously, differences in melting temperature of <10°C have led to significantly altered display levels of mutant single-chain T cell receptors (Shusta et al., 1999); thus mutants in the α3D library may have similar denaturation temperatures. The limited dispersion seen in the expression profile of the α3D library do not appear to significantly alter the expression level.
function of temperature (Figure 3A). Similarly, the absorbance at ~1630 cm⁻¹, which arises from hydrated amides (Huang et al., 2001), decreases monotonically with temperature (data not shown). These data are consistent with the selected mutants maintaining the general three-helix bundle structure of α3D. Additionally, the data presented in Figure 3A indicate that all five mutants show only partial unfolding up to 80°C (the highest temperature used in the present study), suggesting that the thermal melting temperature (T_m) of these proteins is higher than 80°C. However, an analysis based on the available denaturation curves shows that mutants M1 and M3 are less stable than M2, M4, and M5, with M1 being the least stable.

While the results of the thermal melting experiments suggest that M1–M5 are highly thermostable, the lack of data points at temperatures significantly above their individual T_m prevents an accurate estimate of their stability. The difficulty of measuring T_m in a non-denaturing solution was also evident for wild-type α3D, which undergoes partial unfolding at temperatures above 90°C in the absence of a denaturant (Bryson et al., 1998). To quantitatively evaluate the folding stability of these mutants, we performed a chemical denaturation study using GdnHCl as the denaturant and the lone tryptophan at position 4 (W4) as the fluorescent probe. The total tryptophan fluorescence intensity, a measure of solvent exposure of W4, undergoes a sigmoidal transition from high to low as the proteins unfold at high GdnHCl concentrations (Figure 3B). We obtained ΔG_U, m value, and C_m for each mutant by fitting the denaturation curves to a two-state model (Walsh et al., 2001b) (Table II). The fitted free energy of unfolding, ΔG_U is smaller for M1 and M3 than for wild type, whereas it is larger for M2, M4, and M5 than for wild type. While these observations are in agreement with the FTIR measurements, the measured differences in stability do not translate into differential expression level on yeast (Figure 2). These results thus suggest that an inability of the yeast QC system to discriminate among highly thermostable proteins may contribute to the apparent homogeneity in the expression level of α3D mutants.

Yeast expression of α3A and α3B
As a difference in stability alone does not appear to modulate the expression level, we examined whether the structural integrity of a protein may alter the efficiency of surface display. To that end, we expressed two progenitors of α3D (α3A and α3B), which have been shown to form structures that have structural and biophysical properties characteristic of nonnative-like proteins (Bryson et al., 1998). For example, α3A exists in a monomer/dimer/trimer equilibrium as measured by sedimentation equilibrium; while α3B does not form a well-determined structure and is suspected to exist in two competing topologies based on enhanced ANS binding and significant line broadening in 1H NMR spectra. As a result, these proteins would be expected to be degraded by the yeast QC system in the ER and appear at reduced levels on the cell surface. Surprisingly, the observed c-Myc levels of α3A and α3B are comparable with that of α3D, suggesting that they still pass the secretion QC checkpoint equally well (Figure 4). One possible explanation is that despite lacking a rigorously defined tertiary structure (Bryson et al., 1998), these proteins may form stable collapsed globules with secondary structure that protect them against degradation in the ER (at least when they are fused to Aga2p). A chemical denaturation experiment of α3B shows that the protein undergoes a two-state transition with ΔG_U = 7.3 kcal/mol, comparable with that of wild-type α3D (ΔG_U = 5.65 kcal/mol). Combined, both yeast display and in vitro biochemical studies suggest that the yeast QC machinery does not have the requisite resolution to distinguish extremely stable, structured helical proteins.

Implications for future protein engineering
One of the goals of protein engineering is to identify and produce proteins with targeted functionalities in large quantities to enable a variety of applications. For example, the demand for biologically active insulin, a few tons per year (Petrides et al., 1995), requires heterologous expression of stability-enhanced proinsulin in E.coli or in yeast. Similarly
Table II. Thermodynamic parameters obtained from the GdnHCl induced denaturation of α3D and its mutants. The errors are ±0.50 kcal/mol for $\Delta G_{0l}$, ±0.05 kcal/mol/M for $m$ and ±0.03 M for $C_m$

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$\Delta G_{0l}$ (kcal/mol)</th>
<th>$m$ (kcal/mol/M)</th>
<th>$C_m$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>5.65</td>
<td>2.18</td>
<td>2.59</td>
</tr>
<tr>
<td>M1</td>
<td>4.90</td>
<td>1.79</td>
<td>2.73</td>
</tr>
<tr>
<td>M2</td>
<td>6.84</td>
<td>2.51</td>
<td>2.73</td>
</tr>
<tr>
<td>M3</td>
<td>5.44</td>
<td>1.85</td>
<td>2.94</td>
</tr>
<tr>
<td>M4</td>
<td>6.91</td>
<td>2.56</td>
<td>2.70</td>
</tr>
<tr>
<td>M5</td>
<td>6.82</td>
<td>2.39</td>
<td>2.85</td>
</tr>
</tbody>
</table>

for other model proteins that were engineered using yeast (e.g., BPTI, single-chain antibodies and T-cell receptors and MHC molecules), the secretion level of proinsulin from yeast was shown to correlate with thermal stability (Kjeldsen et al., 2002). The relationship between stability and expression level can be understood intuitively: Unstable proteins frequently adopt unfolded conformations that lead to either degradation or retention in the ER, whereas stable proteins predominantly adopt folded native conformations and escape degradation. Yet, there are temperature-sensitive T4 lysozyme mutants whose half-lives in E.coli are significantly longer than that of a thermostable mutant (Inoue and Rechsteiner, 1994a, b). Furthermore, it is unlikely that structural destabilization is the sole factor that led to a reduction in the expression level of BPTI lacking the disulfide bond between residues 14 and 38; wild-type BPTI has $T_m = 105$°C, and even with a change in the free energy of unfolding $\Delta G = 3.7$ kcal/mol, the mutant should be able to fold stably to the native structure at the expression temperature of 30°C (Kowalski et al., 1998b). The folding rate has been suggested as another potential determinant for protein expression. Although studies using wild-type and mutant BPTIs have shown that only their thermal stabilities correlate with expression (Kowalski et al., 1998b), a mutation in the cystic fibrosis transduction regulator missing phenylalanine at position 508, the most common form of cystic fibrosis genotype, is believed to be defective in folding rather than destabilized in the folded state (Qu et al., 1997). These counterexamples to the rule of stability versus expression level highlight the difficulty of capturing the efficiency of protein degradation and export in a single thermodynamic parameter.

Considering the common practice of stability optimization as a means to increase production and engineer novel proteins, it is important to examine the limitations of such an approach and to identify the regime in which the hypothesis fails. We applied yeast surface display to examine a series of helical bundles that have previously been characterized using biochemical and biophysical methods. Additionally, we have computed the conformational energy of α3B, α3D and sequences selected from the α3D library to investigate whether there are quantitative differences in the conformational energies due to modified interactions of side chains (data not shown). Both experimental and computational analyses identified characteristics that could have invoked the yeast QC system to discriminate among the tested sequences, including the heterogeneous tertiary structures of α3A and α3B. The inability of yeast to triage poorly folded structures thus raises a concern that some effects are too subtle to be detected in an in vivo high throughput screen. At the moment the most plausible explanation for this shortcoming is that the melting temperatures of these proteins are simply too high. The threshold at which stability becomes so high that differences in expression level can no longer be observed is difficult to estimate as it depends on numerous factors that are beyond the current understanding of the secretory pathway. Nonetheless, previous studies of engineered single-chain T-cell receptors indicate melting temperatures in the range ~40–60°C yield easily detectable expression level changes (Shusta et al., 1999, 2000). The results presented here strongly suggest that the expression level may be a misleading indicator of structural integrity at the high stability end, and orthogonal screens would be required to identify sequences that are truly native-like from those sequences that fold to structures with high $T_m$ yet lack a well-defined topology.

**Conclusion**

The systematic identification of stable sequences compatible with a given target structure has broad implications for understanding protein folding, development of therapeutic proteins, and discovery of novel future biomaterials. However, the lack of a high throughput assay to rapidly discern stable, well-structured proteins poses a significant technical challenge in the field. To better understand the potentials and limitations of yeast display as a protein engineering platform, we analyzed the yeast surface expression level of three-helix bundle proteins. Our studies, which combine yeast display with in vitro biochemical and
spectroscopic measurements, show that the yeast QC system is unable to differentiate well-folded variants from collapsed, molten globules among proteins of high stability. These observations corroborate the conclusions from an independent study that the secretion level alone can be a poor indicator of protein stability or structural integrity (Hagihara and Kim, 2002). Complementary screens or analyses that interrogate properties other than thermal stability [e.g. computational library analysis and design (Park, et al., 2004)] may help reduce false positives and engineer sequences with native-like structures.

Acknowledgements

This work was supported by the NSF under MRSEC grant DMR 00-79909 (E.T.B., F.G., J.G.S.) and DMR 05-20020 (J.G.S.) and by the NIH (GM61267; J.G.S.).

References


Received July 15, 2005; revised December 2, 2005; accepted January 30, 2006

Edited by James Marks