JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Exploiting Elements of Transcriptional Machinery to Enhance Protein Stability

Nora H. Barakat, Nesreen H. Barakat, Lisa J. Carmody and John J. Love*

Department of Chemistry and Biochemistry, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-1030, USA

The correlation between protein structure and function is well established, yet the role stability/flexibility plays in protein function is being explored. Here, we describe an *in vivo* screen in which the thermal stability of a test protein is correlated directly to the transcriptional regulation of a reporter gene. The screen readout is independent of the function of the test protein, proteolytic resistance, solubility or propensity to aggregate indiscriminately, and is thus dependent solely on the overall stability of the test protein. The system entails the use of an engineered chimeric construct that consists of three covalently linked domains; a constant N-terminal DNA-binding domain, a variable central test protein, and a constant C-terminal transcriptional activation domain. The test proteins are mutant variants of the β1 domain of streptococcal protein G that span fairly evenly a thermal stability range from as low as 38 °C to above 100 °C. When the chimeric construct contains a test variant of low thermal stability, the reporter gene is up-regulated to a greater extent relative to that of more stable/less flexible variants. A panel of nine Gβ1 mutant variants was used to benchmark the screen, and spectroscopic methods were employed to characterize the thermal and structural properties of each variant accurately. The screen was combined with *in silico* methods to interrogate a library of randomized variants for selection of mutants of greater structural integrity.

Keywords: protein design; protein stability; conformational specificity; *in vivo* combinatorial screen; directed molecular evolution

*Corresponding author

Introduction

The field of protein design aspires to engineer proteins of specific function. Function is correlated directly to structure and, therefore, a specific target fold is normally chosen for a desired function. In addition to structure, protein dynamics correlate to function,[1] and in recent years increasing experimental evidence (driven primarily by high-field heteronuclear NMR) indicates that overall structural stability, or lack thereof, may play an important role in protein function.[2–4] The overall dynamic regime of protein backbone and side-chain atoms correlates to intrinsic stability and, thus, protein flexibility (or plasticity) may also correlate to function. Therefore,

the ability to tailor the intrinsic stability/flexibility of a designed protein to a desired function may ultimately enhance protein design endeavors. Furthermore, an *in vivo* screen for protein stability enables the rapid analysis of large portions of the sequence space of a particular fold.

To date, considerable efforts have been made to create *in vivo* screens designed to interrogate large combinatorial libraries of protein variants for increased stability,[5–8] or solubility.[9–12] (For comprehensive reviews, see Magliery.[13,14]) In addition to these screens, computational design methods have been used either alone,[15–19] or in combination with *in vivo* screens,[20,21] to design and select for protein variants with increased intrinsic stability and/or improved function. Of particular relevance to the work reported here are studies in which the compact, highly stable fold of the β1 domain of streptococcal protein G (Gβ1) has been the target of design efforts.[22–25] Phage display methods have been developed for the selection of folded Gβ1 variants,[26–29] and the close structural homologue protein L.[30] The general premise of these studies is

based on the valid assumption that destabilized variants do not fold properly, and thus do not bind the constant Fc region of immobilized antibodies as effectively as properly folded variants. As the relationship between higher stability/enhanced folding and stronger binding is not absolute (there are examples of induced fit or folding of partially disordered proteins upon binding or catalysis),[31,32] it is not unexpected that in some cases phage-based screens were able to select for folded variants but not for variants of greater thermal stability.[29,30]

Here, we report the development of an *in vivo* screen for protein stability that is completely independent of the function of the test protein. In addition, screen results correlate well to the measured stability of protein variants known to self-associate,[33] and thus reporter readout is independent of the solubility of the test protein or its propensity to aggregate indiscriminately. Structure-based rational design was used to create a panel of nine Gβ1 variants that span a thermal stability range from 38 °C to above than 100 °C. These variants were characterized in the context of the engineered *in vivo* screen and structural parameters were assessed with circular dichroism (CD), NMR and *in vitro* proteolysis. Preliminary screening of a randomized library, based on a destabilized variant, demonstrated the utility of the screen for selecting variants of greater thermal stability.
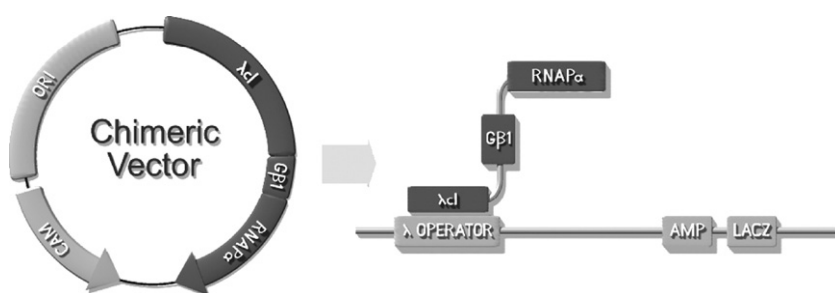
Results

Screen design

Creation of the chimeric *in vivo* screen entailed splicing the genes of different mutant variants of the Gβ1 domain between the genes for the proteins used as bait and prey in a bacterial two-hybrid screen (BacterioMatch® Two-Hybrid System, Stratagene).[34–36] The Gβ1 domains all start at amino acid position 2 (Thr) and end at position 56 (Glu) using the numbering scheme of the structure deposited in the RCSB Protein Data Bank (accession code 1PGA).[37] The resulting protein is a three-domain chimera in which the N-terminal DNA binding domain (full-length bacteriophage λcI) is linked to the N terminus of the Gβ1 variant and the C-

terminal transactivation domain (the N-terminal domain of the α-subunit of bacterial RNA polymerase (RNAPα)) is linked to the C terminus of the Gβ1 variant (Figure 1). The chimeric construct was created by subcloning the gene for RNAPα into the bait plasmid downstream of λcI. The Gβ1 genes were then subcloned between λcI and RNAPα in the newly created plasmid. There is a short linker between the C terminus of λcI and the N terminus of the Gβ1 variants (Ala$_3$-Ser) and a three amino acid linker (Gly-Asn-Ser) between the C terminus of the Gβ1 variant and the N terminus of RNAP-α. The three-domain chimera acts as a functional transcription factor upon binding the λ operator (located upstream of the reporter genes in the F′ episome) and results in the transcription of reporter genes (e.g. β-lactamase).

Design of Gβ1 mutants of variable stability

Nine Gβ1 mutant variants with different melting temperature (T_m) values were engineered for testing in the chimeric screen. Six of the nine variants originate from three "parent" variants; the wild-type domain (Gβ1-WT), and two variants termed monomer A (MonA) and monomer B (MonB). The MonA and MonB variants originated from an earlier design project in which a *de novo* protein interface was engineered by computationally docking the normally monomeric Gβ1 domain to itself, followed by the use of the ORBIT suite of protein design algorithms[17] to mutate specific interfacial side-chains with the goal of driving specific complex formation.[33,38] This design resulted in a pair of monomers that, upon generation in the laboratory, formed a heterodimer of modest binding affinity. Introduction of the mutations that resulted in the sequences for each monomer altered their thermodynamic properties relative to Gβ1-WT. The 12 mutations that resulted in the MonA sequence stabilized it to a hyperthermophile (i.e. $T_m > 100$ °C), while the eight for MonB were destabilizing, resulting in a T_m of 38 °C (Gβ1-WT ~85 °C).

Of the six additional variants, four were single-point mutants of Gβ1-WT: (1) Gβ1-W43A, (2) Gβ1-W43V, (3) Gβ1-W43Y, and (4) Gβ1-Y45A, and two variants were derived from MonB-WT: (1) MonB-A45V and (2) MonB-A45Y. The sequences for all nine variants are given in Table 1. The rationale for the

Figure 1. Chimeric vector and construct. The genes for the Gβ1 variants were subcloned into the chimeric vector between the genes for λcI and RNAPα, which resulted in the three domains linked together as follows - λcI-Gβ1-RNAPα. Binding of the protein chimera to the λ-operator up-regulates the reporter genes β-lactamase (AMP) and β-galactosidase (LACZ).

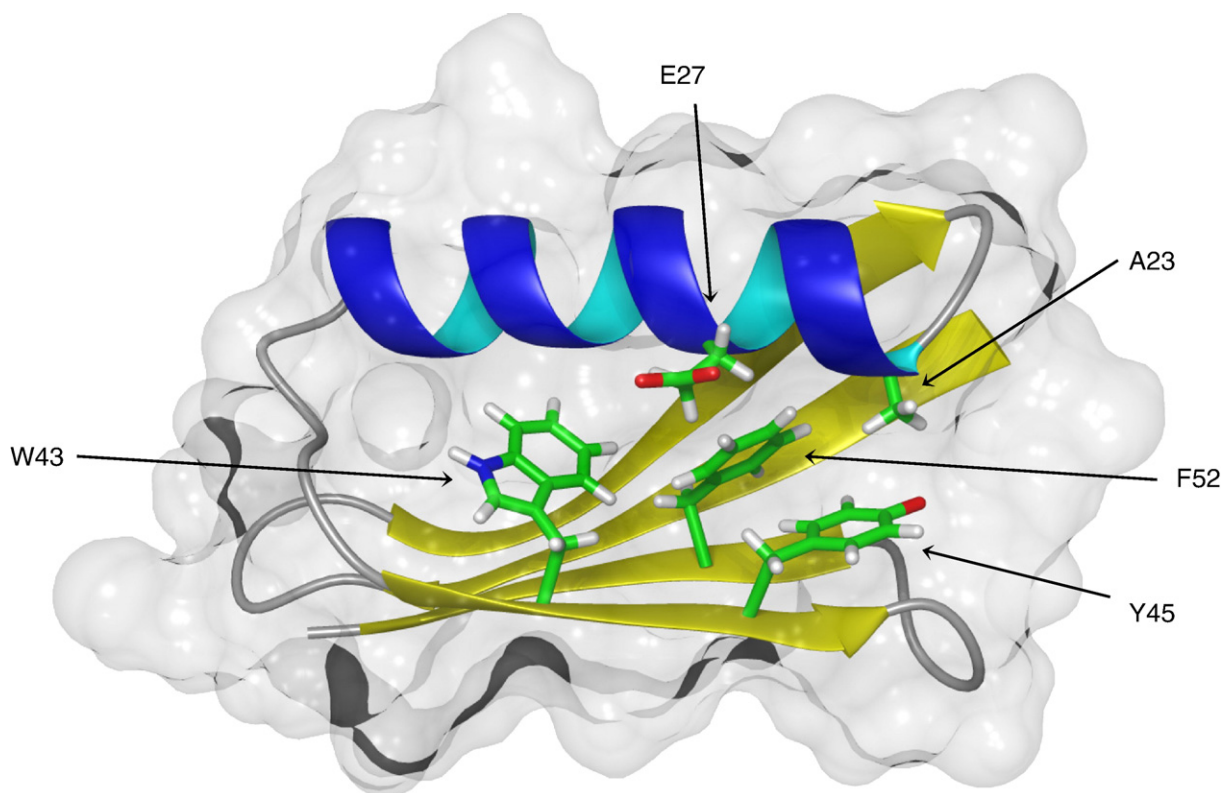**Table 1.** Amino acid sequences for Gβ1-WT and mutant variants

		1	10	20	30	40	50	T_m (°C)
Gβ1-WT	—	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE						85
Gβ1 W43A	—	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGE**A**TYDDATKTFTVTE						57
Gβ1 W43V	—	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGE**V**TYDDATKTFTVTE						66
Gβ1 W43Y	—	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGE**Y**TYDDATKTFTVTE						64
Gβ1 Y45A	—	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWT**A**DDATKTFTVTE						55
MonB-WT	—	MTYKLILNGKTLKGETTTEAVD**I**ATA**AD**VF**A**QYA**A**DNGV**K**GEWT**ADE**ATKTFTVTE						38
MonB A45V	—	MTYKLILNGKTLKGETTTEAVD<u>I</u>ATA<u>AD</u>VF<u>A</u>QYA<u>A</u>DNGV<u>K</u>GEWT**V**DEATKTFTVTE						54
MonB A45Y	—	MTYKLILNGKTLKGETTTEAVD<u>I</u>ATA<u>AD</u>VF<u>A</u>QYA<u>A</u>DNGV<u>K</u>GEWT**Y**DEATKTFTVTE						72
MonB ORDES	—	MTYKLILNGKTLKGETTTEAVD**K**ATA**K**DVF**A**QYA**A**DNGV**K**GEWT**F**DEATKTFTVTE						64
MonA	—	MTYKLILNGKTLKGE**FTA**EA**E**DAA**LA**E**YI**F**RA**LA**KAQ**GVDGEWTYDDATKTFTVTE						>100

Mutation(s) from the sequence just above each line are indicated in bold. The eight mutations for MonB-WT and the 12 for MonA are in bold. Mutation(s) of MonB-WT sequence are indicated in bold and the MonB-WT positions are underlined.

mutations that gave rise to the six mutants stems primarily from work performed by Kobayashi *et al.*,[39] and earlier work by Blanco *et al.*[40] Both groups analyzed a 16-residue peptide that encompassed the C-terminal β-hairpin of Gβ1 (residues 41–56). It was demonstrated initially that the isolated peptide significantly populates a native-like β-hairpin structure in aqueous solution, and that three aromatic residues (W43, Y45, and F52) cluster to form a small hydrophobic core (Figure 2).[40] Subsequent work demonstrated that upon mixing the C-terminal peptide with a peptide that comprised the first 40 Gβ1 residues, a complex was indeed formed, and exhibited spectroscopic and thermodynamic properties not unlike that of the intact domain (yet of expected lower stability). Alanine scanning of the C-terminal peptide further illustrated the critical structural role the cluster of aromatic residues plays, as mutation of any of these residues to alanine abrogated complex formation completely.[39]

Figure 2. Gβ1-WT structure and positions of specific residues. The backbone of Gβ1-WT is depicted as a blue ribbon for the α-helix and yellow ribbons for the β-sheet. Side-chains are shown for the residues that make up the C-terminal aromatic cluster (43, 45, 52), as well as two residues in close proximity (23 and 27). The transparent surface was generated with the program MSMS (using a probe sphere radius of 1.4 Å),[52] the backbone ribbons with the program MOLMOL,[53] and the final combined image was rendered with the program POVRay.

The rational design of the six additional variants was based on mutation of two of the aromatic residues that make up the C-terminal cluster (i.e. W43 and Y45). Mutation of W43 to alanine reduced the T_m from the wild-type value of 85 °C to 57 °C, thus confirming the critical role this residue plays in stabilizing the wild-type domain. We predicted that mutation of W43 to valine would have less of an effect on stability, as valine is larger and more hydrophobic than alanine, and thus would fill the space previously occupied by the tryptophan side-chain more efficiently. This prediction proved true, as the reduction of the measured T_m was not as drastic, and resulted in a T_m of 66 °C for the W43V mutation. Interestingly, and somewhat unexpectedly, the mutation of W43 to tyrosine was slightly more destabilizing in comparison to the valine mutant, as it resulted in a T_m of 64 °C. Finally, the critical role Y45 plays in stabilizing the overall fold was confirmed upon mutation to alanine, as it resulted in reduction of the T_m to 55 °C for the Gβ1-Y45A mutant.

The interfacial design that produced the MonB sequence resulted in the substitution of Y45 to alanine, which we surmised was the primary cause of the reduction in the thermal stability from 85 °C to 38 °C. Substitution back to tyrosine confirmed this assumption as the MonB-A45Y revision mutant resulted in an increase in the T_m from 38 °C to 72 °C. Substitution of the intermediate sized, non-polar residue valine at position 45 (MonB-A45V) resulted in a predictably more modest increase in thermal stability, from 38 °C to 54 °C.

An additional MonB variant, MonB-ORDES, was generated by performing computational mutagenesis on three select positions (23, 27, and 45) using the ORBIT design algorithms. These positions were chosen on the basis of the proven importance of position 45 for the structural integrity of the Gβ1 fold, and because positions 23 and 27 are in close spatial proximity to 45 (Figure 2). The Gβ1-WT sequence consists of A23, E27 and Y45, whereas the MonB-WT sequence has Ile, Ala and Ala at the respective positions. The ORBIT calculation performed on the free MonB-WT structure returned Lys, Lys and Phe at these positions, and resulted in an increase in the T_m from 38 °C to 64 °C.

Circular dichroism

Far ultraviolet (UV) circular dichroism (CD) spectra were collected for ten variants and the spectra collected on the nine mutant variants are highly similar to that of Gβ1-WT (Supplementary Data Figure 1(a)). These results indicate that the mutant variants most likely maintain the α/β fold topology of the wild-type domain. The melting temperatures for all ten variants were measured using standard thermal denaturation monitored by CD at 218 nm (Table 1; and Supplementary Data Figure 1(b)). The slopes of the melting curves are fairly similar and indicate cooperative

unfolding, and thus provide further evidence that the mutant variants maintain the general wild-type fold.
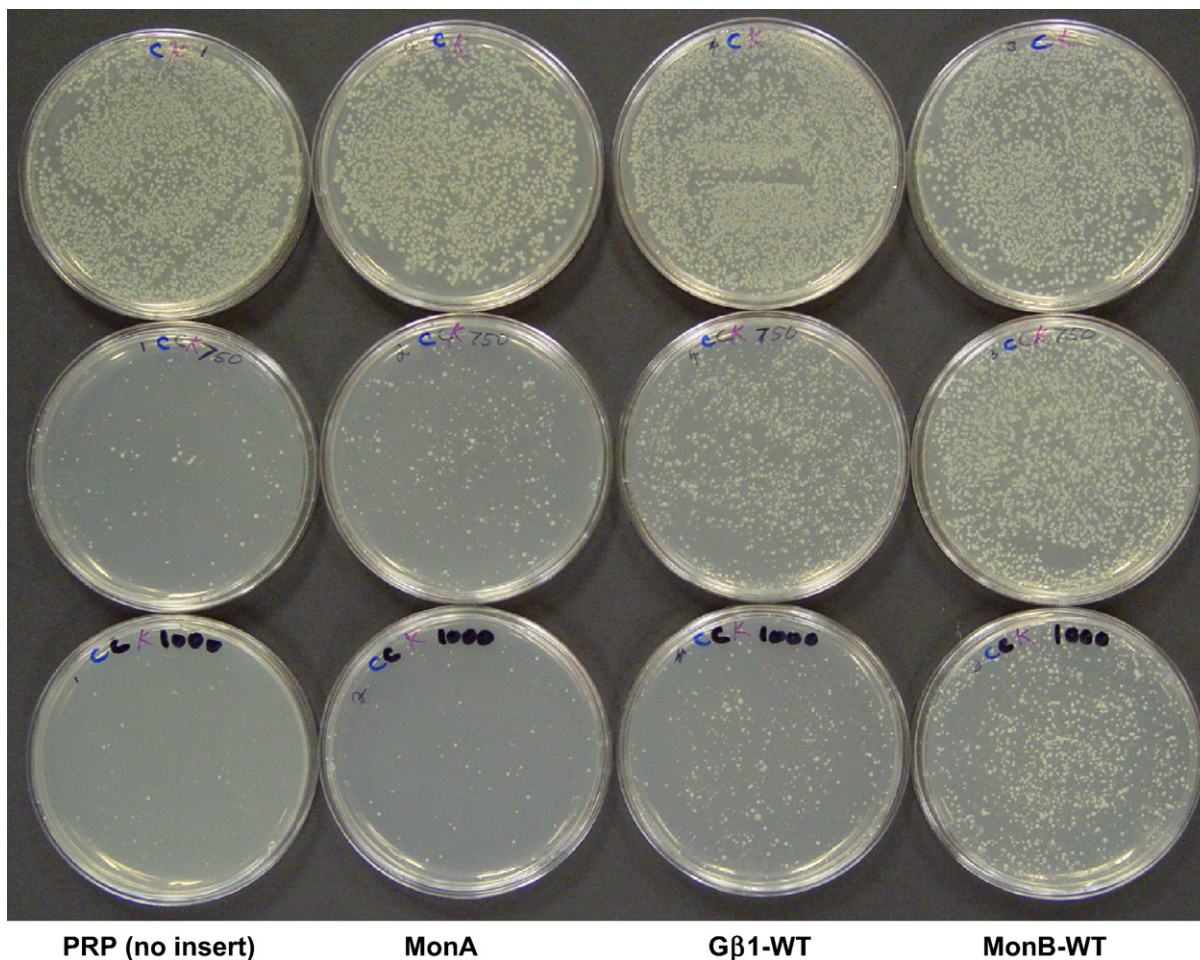
Screening mutants on increasing amounts of reporter antibiotic

The genes for Gβ1-WT, MonA, and MonB-WT were cloned into the chimeric construct and initially tested for growth on agar plates that contained increasing amounts of the reporter antibiotic carbenicillin (Figure 3). It is evident from the relatively large disparity in growth rates and colony numbers for these variants that the chimera that contains the least stable variant, MonB-WT, clearly functions as a most effective transcription factor. These results indicate an inverse relationship between transformation efficiency and thermal stability, as the most stable variant, MonA, gave rise to the fewest colonies, Gβ1-WT slightly more, and the least stable variant, MonB-WT, the greatest number of colonies.

Tests of bacterial growth-rates for the additional six mutant variants (Table 1) were conducted in liquid medium. Cells were grown for a set amount of time (e.g. 4 h) in medium that contained a fixed amount of reporter antibiotic (e.g. 3000 μg/ml of carbenicillin). In general, the results (Figure 4) agree with the trend observed for the plate screening i.e. the chimeras that contain variants of lower thermal stability (and thus potentially higher intrinsic flexibility) up-regulate the reporter genes to a greater extent as compared to those of higher thermal stability. Although a general trend is evident, there are a number of variants that up-regulate the reporter gene to a amount that is comparable to the least stable variant, MonB-WT, yet these variants (MonB-A45V, Gβ1-W43A, Gβ1-Y45A) have T_m values that are higher by at least 15 deg. C. There are a number of potential reasons for this finding; for example, the fact that screen readout may be more highly sensitive to differences in T_m for variants at the higher end of the thermal stability range (greater than 60 °C). Furthermore, in addition to the screen being sensitive to thermal stability, it may be sensitive to other physical characteristics of the variants, such as intrinsic flexibility. To explore the potential role flexibility plays in this screen, 2D heteronuclear NMR spectra were collected on all variants.
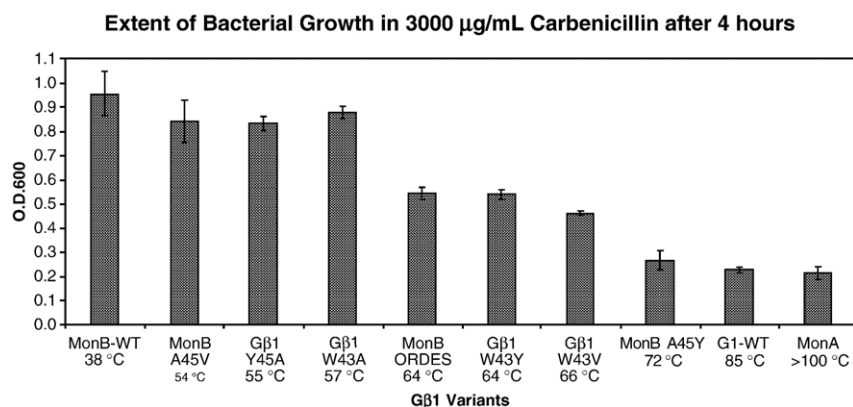
NMR spectroscopy

[^{1}H,^{15}N] heteronuclear single quantum coherence (HSQC) spectra were collected for ten variants (four representative 2D HSQC spectra are shown in Figure 5). The significant signal dispersion and high resolution exhibited in the spectra recorded for Gβ1-WT is nearly unparalleled, and reflects both its compact overall structure and, as noted in the original report of the Gβ1 structure, the fact that approximately 95% of the residues are involved in regular, well-
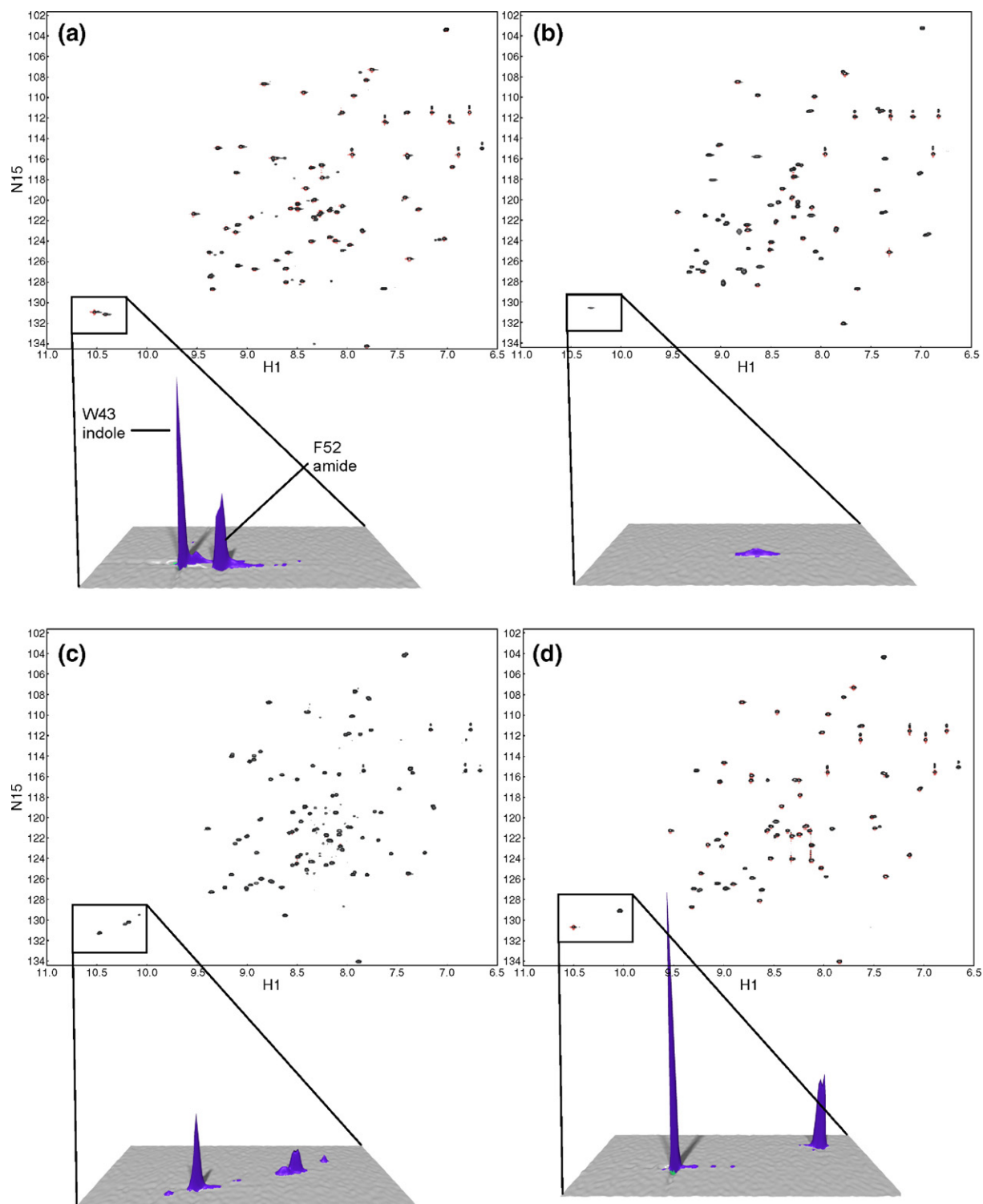
PRP (no insert)   **MonA**   **Gβ1-WT**   **MonB-WT**

Figure 3. Reporter gene expression for Gβ1 variants. The uppermost row corresponds to control plates that contain no reporter antibiotic. The middle row contains plates with 750 µg/ml of carbenicillin and the lowest row contains plates with 1000 µg/ml of carbenicillin. The first column corresponds to the vector without a Gβ1 variant insert (i.e. λ*cI* fused to RNAPα with a short intervening insert; referred to as PrP) and the other variants are listed below the columns.

ordered secondary structure elements.[41] Even though there are differences in apparent conformational flexibility based on the *in vivo* screen results, and different T_m values, all variants exhibit signal dispersion comparable to that of Gβ1-WT. This finding is in agreement with the CD results, and further indicates that the overall Gβ1 fold topology is likely maintained for all variants.

Figure 4. Correlation between bacterial growth and protein thermal stability. The extent of growth for the panel of mutant variants of the Gβ1 domain is shown. The height of each bar corresponds to the bacterial cell density for each mutant after growth for 4 h in liquid LB medium that contained 3000 µg/ml of carbenicillin. The A_{600} value corresponds to the ability of each variant to up-regulate the β-lactamase reporter gene. The parent name, the mutation(s) that gave rise to each variant, and the T_m values are given below each bar.

Figure 5. Representative [^{1}H, ^{15}N] HSQC spectra. (a) Gβ1-WT, (b) Gβ1-W43A, (c) MonB-WT, and (d) Gβ1-Y45A. The insets illustrate either the peak for the backbone amide of F52 and/or the peak that corresponds to W43 indole.

Although signal dispersion is maintained for all variants there is a trend in signal resolution that generally parallels the screen results, i.e. variants of lower stability exhibit lower resolution (broader line-widths) in comparison to those of higher stability. This trend is most evident in the spectra for the least stable variant, MonB-WT, and for the

next to least stable variant, MonB-A45V (Figure 6(e) and (d), respectively). In comparison to the Gβ1-WT spectrum (Figure 6(j)), there are sections in these spectra in which overlapping peaks obscure the baseline (indicated by bars below in Figure 6). In addition, MonB-WT exhibits evidence of multiple conformations that exchange slowly on
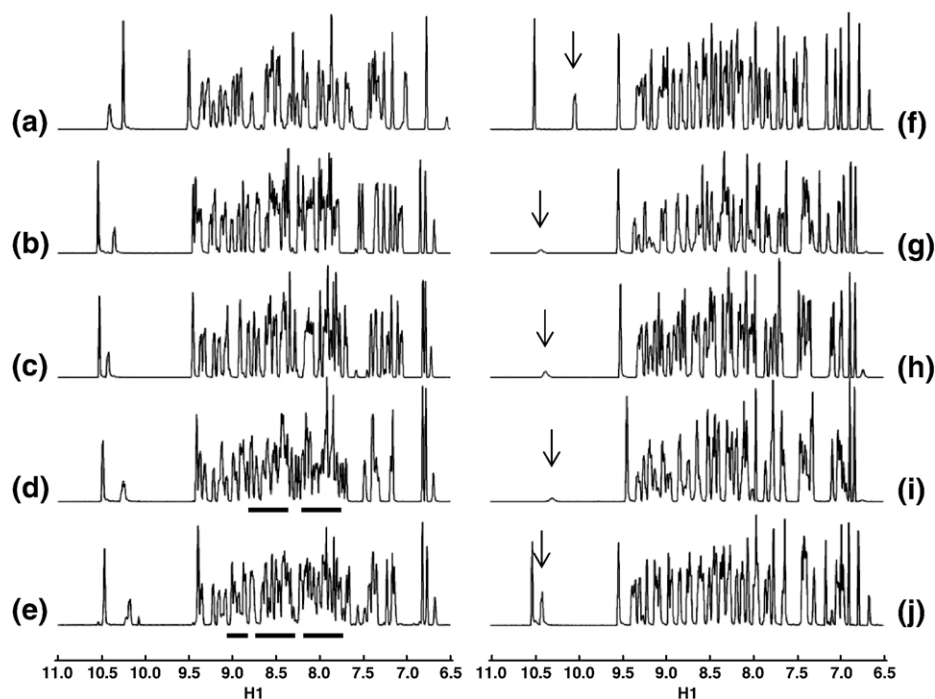
the NMR time-scale as there are multiple sets of peaks in the 2D HSQC and at least four peaks in the downfield region for the tryptophan indole proton of W43 and the amide proton of F52 (the peaks between ~10 ppm and ~10.5 ppm in Figure 5(c)). This finding was not unexpected, as MonB-WT is known to self-associate and form amyloid-like fibers when incubated and agitated near its melting temperature.[33]

There is a series of three single-point mutants of Gβ1-WT for which the *in vivo* screen results nicely reflect the peak shape of the associated NMR spectra (Gβ1-W43A, Gβ1-W43Y, and Gβ1-W43V). The peak corresponding to the backbone amide of F52 is considerably broadened for these mutants relative to Gβ1-WT (Figures 5(b) and 6(g)–(i)). The relatively extreme broadening of this peak indicates greater internal motion of backbone and side-chain atoms in this region, and further illustrates the important role the centrally located W43 side-chain plays in stabilizing the structure of the Gβ1 domain. In light of these mutations, the increased dynamics are not unexpected, i.e. mutation of tryptophan to alanine results in an internal cavity that corresponds to the difference between the volume of the tryptophan side-chain *versus* that for alanine (237.6 $\mathring{A}^3$ and 91.5 $\mathring{A}^3$, respectively).[42] The broadened peak for F52 likely reflects internal motion of backbone and side-chain atoms as they rearrange, and/or collapse, to compensate for the resulting cavity. The increased dynamics are due to fluctuations between structural states that are similar in energy and thus not resolved with a single structural solution. The degree of broadening for the F52 amide nicely reflects differences in the melting temperatures as well as the *in vivo* screen results, i.e. the W43A mutant has the lowest T_m (57 °C), the highest *in vivo* screen readout and the broadest F52 NMR peak; the intermediate variant, W43Y, has a higher T_m (64 °C), exhibits an intermediate *in vivo* screen readout and has a broad F52 NMR peak; the most stable variant in this series, W43V, has a slightly higher T_m (66 °C), reduced *in vivo* screen readout and an F52 NMR peak that, although still broad, is sharper than that for the other two variants. The result that the F52 peak shape is considerably broader for these three mutants in comparison to the less stable variants MonB-WT and MonB-A45V is an excellent example of how local flexibility can often be decoupled from overall thermal stability, and demonstrates that higher thermal stability does not always correlate to greater conformational specificity (i.e. decreased number of possible conformations). The higher T_m values for the W43 mutants likely reflect increased hydrophobic contacts relative to the less stable variants, yet the NMR results, in conjunction with the *in vivo* screen readout, reflect the fact that the structural integrities of these variants are fairly similar.

The concurrent behavior in the trends observed for the NMR peak shape and the *in vivo* screen does not hold for one particular variant, Gβ1-Y45A. This variant exhibits fairly high *in vivo* screen readout (and has a relatively low T_m of 55 °C) yet has relatively sharp NMR lines (Figure 6(f)). For this variant, the F52 amide peak is shifted furthest

Figure 6. This Figure displays the 1D projections of the 2D HSQC spectra for ten variants: (a) MonA; (b) MonB-ORDES; (c) MonB-A45Y; (d) MonB-A45V; (e) MonB-WT; (f) Gβ1-Y45A; (g) Gβ1-W43Y; (h) Gβ1-W43V; (i) Gβ1-W43A; (j) Gβ1-WT. Regions of relatively poor resolution are indicated by a black bar below, and the resonances for backbone amide proton for F52 are indicated by an arrow.

Table 2. Proteolysis results

Protein variant	Time to achieve 50% reduction of the parent peak (min.)
MonB[a]	5.0
MonB-A45V	15.5
Gβ1-Y45A	17.5
Gβ1-W43Y	28.0
Gβ1-W43A	32.0
Gβ1-W43V	46.0
MonB-ORDES[a]	360.0
MonB-A45Y	364.0
Gβ1-WT	3224.0
Mon A[b]	>10,000.0

[a] Estimated by SDS-PAGE; for all other variants the 50% degradation time was obtained by HPLC.
[b] Higher enzyme concentration, 20 mg/ml *versus* 0.33 mg/ml for all other variants.

upfield in both the proton and nitrogen dimensions (Figure 5(d)). This likely reflects a repositioning of the side-chain of F52 (as it is in close proximity to Y45 in the wild-type structure), yet may be due to additional backbone and side-chain rearrangement to compensate for the loss of the tyrosine side-chain at position 45. To confirm the malleability implied by the relatively high *in vivo* screen readout for this mutant, and to verify the structural integrity of the other variants, *in vitro* proteolysis experiments were performed on all variants.

In vitro proteolysis analysis

Cleavage of the peptide bond by proteases is dependent on the ability of the protease to bind the protein substrate and induce the proper stereospecific structure that is amenable to enzymatic degradation in the active site of the protease. Proteins of lower overall stability tend to be more malleable and, thus, more prone to proteolytic cleavage relative to stable variants and, therefore, an inverse correlation exists between protein stability (or plasticity) and proteolysis. To further characterize the intrinsic stabilities of the panel of Gβ1 mutant variants, we subjected all variants to *in vitro* proteolysis analysis using equal quantities of the serine proteases trypsin and chymotrypsin.

The results of the proteolysis analysis are shown in Table 2, where the amount of time required to degrade 50% of the protein is listed for each variant. In general, the results correlate closely with the *in vivo* screen results, and the measured T_m values for each variant. In fact, the most stable variant, the hyperthermophile variant MonA, needed a 60 times greater amount of both enzymes to achieve 50% degradation in a reasonable time-frame (29 h), whereas the 50% degradation times for the least stable variants (MonB-WT, MonB-A45V) were approximately 5 min and 15 min, respectively. As opposed to the stable structure implied by the relatively narrow NMR line-widths for the Gβ1-Y45A variant, the proteolysis analysis proved this variant is indeed quite malleable, as the time to reach 50% degradation was quite short

(~17.5 min), confirming what was observed in the *in vivo* screen.

Library generation and preliminary screening

To explore the utility of the *in vivo* screen for selection of Gβ1 variants of greater stability, we created a library of mutants using the sequence of the least stable variant, MonB-WT as a starting point. Position 45 is crucial for stability, as mutation from Tyr to Ala proved to be a dominant reason for the low level of stability observed for MonB-WT (mutation back to Tyr increased T_m from 38 °C to 72 °C). Therefore, this region of the MonB-WT sequence was targeted for improvements in stability, using both computational redesign and physical library screening. Positions 23 and 27 were targeted due to their close proximity to position 45 (Figure 2).

In addition to screening an actual library of mutant genes, we used computational methods to virtually screen the three select positions for amino acid residues that would stabilize the MonB variant. To this end, the ORBIT design algorithms were used to mutate and energetically assess rotameric descriptions of all amino acids except for glycine, proline, and cysteine at the three respective positions. The resulting amino acid residues (K23, K27, and F45) were introduced into the MonB sequence, cloned into the chimeric construct and displayed an intermediate *in vivo* screen readout (Table 3).

A library of genes was engineered that contained random bases at the codons for positions 23, 27 and 45, while the remainder of the genes contained codons from the MonB-WT sequence. From this preliminary screen, five variants were selected that have melting temperatures greater than MonB-WT (Table 3). Three variants have melting temperatures that are 6–9 deg. C higher than that of MonB-WT (44 °C, 47 °C, and 47 °C) and screen-based readouts that are comparable to that of the parent sequence. Each of these variants has small to medium sized amino acid residues at position 45 (A, V, T) and two had serine at position 27. On the other hand, two of the selected variants are considerably more stable than MonB-WT (59 °C and 69 °C) and exhibit higher conformational stability, as judged by the screen

Table 3. Amino acid sequences at positions 23, 27, and 45 for Gβ1-WT, MonB-WT, MonB-ORDES and the five mutants selected from the chimeric screen

Protein variants	Position 23	Position 27	Position 45	T_m (°C)	OD_{600}[a]
Gβ1-WT	A	E	Y	85°	0.23 ± 0.01
MonB-WT	I	A	A	38°	0.95 ± 0.09
MonB-ORDES	K	K	F	64°	0.55 ± 0.03
1	I	S	A	44°	0.75 ± 0.06
2	Y	S	V	47°	0.92 ± 0.08
3	Y	Y	T	47°	0.81 ± 0.05
4	P	L	L	59°	0.63 ± 0.07
5	L	Y	W	69°	0.41 ± 0.05

[a] The A_{600} corresponds to the absorbance taken after 4 h of growth in medium that contained 3000 μg/ml of carbenicillin.

results. These variants have larger and more hydrophobic residues at the respective positions and, interestingly, one has proline at position 23, which is near the N terminus of the helix and thus may provide additional stability to this region of the protein.

Discussion

The motivation for the screen design was driven primarily by the significant disparity in the thermal stability observed between the least and most stable variants (MonA and MonB-WT). We assumed that the large difference in the measured T_m values corresponded to differences in structural integrity and, thus, inherent malleability. This assumption was confirmed by *in vitro* proteolysis experiments performed on the three parent Gβ1 variants (i.e. MonA, MonB-WT, and Gβ1-WT). It was assumed initially that, in the context of the chimeric construct, the measured differences in stability might either manifest in variable proteolytic resistance *in vivo* (less stable variants would be cleaved at a higher rate and thus up-regulate the reporter genes to a lesser extent) or to differences in the overall structural integrity (stability) of the entire chimeric construct. The results reported herein clearly indicate the latter to be the case. In addition, there were no detectable differences in proteolytic resistance observed *in vivo* based on Western blot analysis of the chimeric constructs that contained the MonA, Gβ1-WT, and MonB-WT sequences (data not shown).

The spectroscopic and *in vitro* analysis, combined with the *in vivo* screen, provided considerable insights into the pivotal roles W43 and Y45 play in maintaining the structural integrity of the Gβ1 fold. The Y45A mutation in the MonB-WT variant is the probable cause for its low stability and high *in vivo* screen readout and, therefore, we chose to randomize position 45 as well as two other positions in close proximity. As a benchmark to compare our selection results against, we performed computational mutagenesis on the three select positions using the ORBIT design algorithms. The design was successful, as the resulting sequence has an increased T_m of 64 °C (from 38 °C) and exhibited intermediate *in vivo* screen readout.

Interrogation of the randomized library resulted in the selection of five variants that exhibit greater structural integrity than the parent MonB-WT sequence and one variant that has a higher T_m compared to the MonB-ORDES sequence (Table 3). These findings illustrate the viability for using this screen to select stable mutants from a pool of randomized variants. A limitation of the current rendition of the screen is that the less stable variants are more efficient at up-regulating the reporter gene and, thus, have a selective advantage over variants of greater stability. The screen will likely become more effective for interrogating libraries of greater complexity upon switching the reporter gene from one that confers antibiotic resistance to one that is toxic and, thus, kills bacteria that harbor less stable variants. The point that less stable variants are more efficient at recruiting the endogenous transcriptional machinery raises an important question: why do more stable variants function as less efficient transcription factors in the context of this chimeric screen? Insights into what might be happening on a molecular level are derived from two sources.
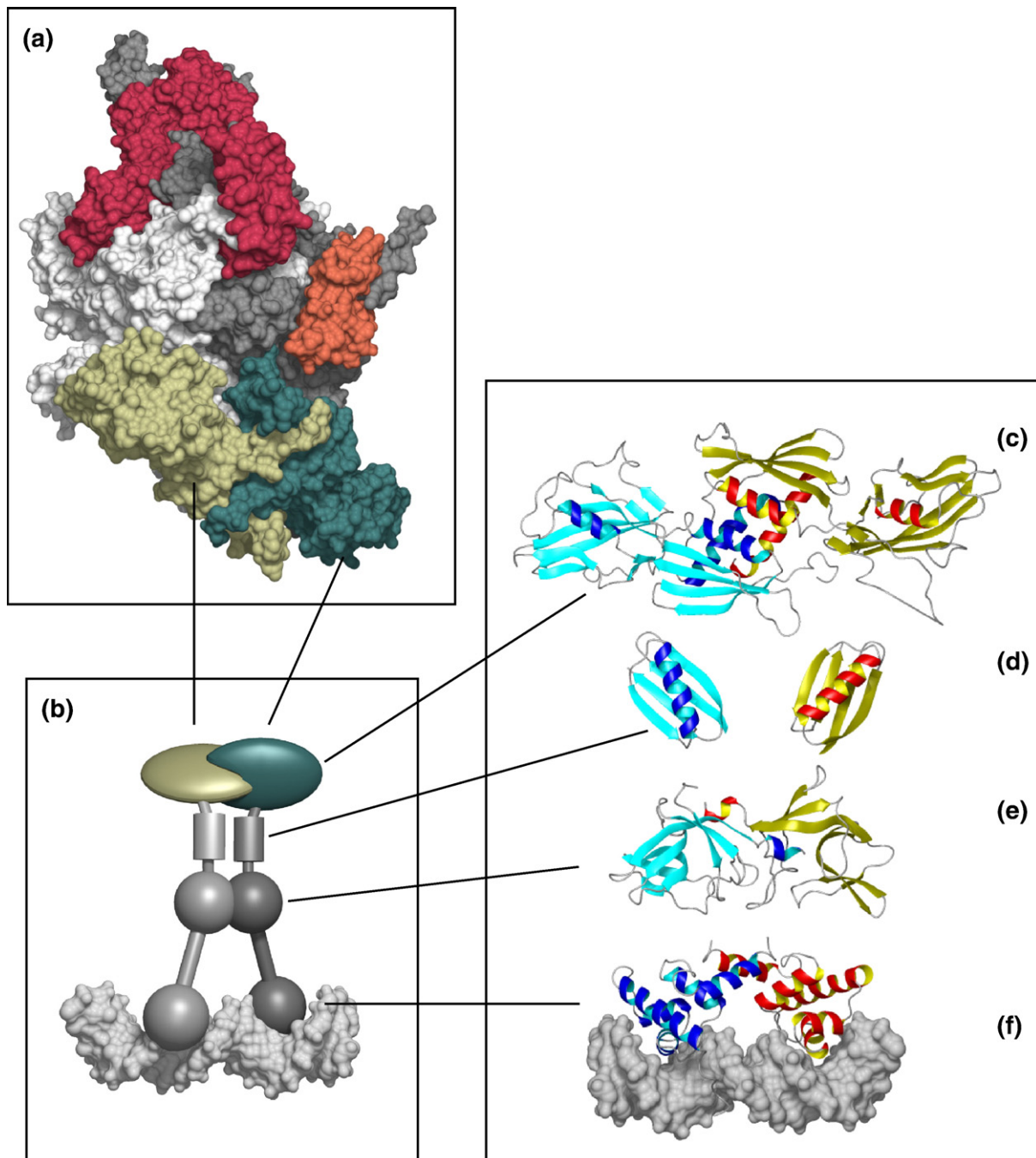
The first stems from work in which Gβ1 was displayed on the surface of filamentous bacteriophage with the goal of exploring the structural integrity of a library of mutant variants.[29] The screen proved to be successful at selecting folded variants, yet did not result in the selection of variants with significantly increased thermal stability. The inability to select for more stable variants may be due to inherent structural properties of the tight Gβ1 fold, as an unexpected discovery demonstrated that phage displaying stable Gβ1 variants (wild type) gave rise to unnaturally small plaques, yet unstable mutants gave rise to the normal larger plaque phenotype.[29] The final step in phage assembly entails transport of phage coat proteins to the bacterial membrane, where they are assembled on viral DNA and extruded through the bacterial membrane. Since the Gβ1 variants were fused to the gpIII coat protein for display purposes, it was hypothesized that the presence of the Gβ1 domain may inhibit transport of gpIII across the bacterial membrane. Mutations that disrupt the structural integrity of the tightly folded Gβ1 domain may allow assembly to occur more efficiently, imparting a growth advantage for phage harboring less stable variants. In light of the inverse relationship, we observe between intrinsic stability and transcriptional activation in the *in vivo* chimeric screen, we consider this explanation to be highly plausible.

In addition to the Gβ1 domain, other highly stable proteins have proved refractory to display on filamentous phage.[43] For example, designed ankyrin repeat proteins that are expressed in soluble form with high yields in *Escherichia coli* show high thermodynamic stability and fast cooperative folding, and are resistant to proteolysis but not displayed efficiently on filamentous phage.[44] This problem was alleviated successfully upon altering the secretion pathway from one in which translocation across the bacterial membrane occurs post-translationally to one in which it occurs cotranslationally, thus providing direct empirical evidence that highly stable protein folds can inhibit, or partially block, certain cellular processes.

A second source of insight as to why the more stable Gβ1 variants ultimately function as less efficient transcription factors in the chimeric screen is provided upon analyzing the crystal structures of all the components that comprise the chimeric construct (Figure 7(c)–(f)). Transcriptional activation begins at the level of DNA where interactions occur between the N-terminal DNA-binding domain of λcI and operator DNA (Figure 7(f)).

The N-terminal DNA-binding domain of λ*cI* is known to dimerize weakly, and the crystal structure reveals that there are regions where the domains contact one another and thus may form favorable intermolecular contacts.[45] The N-terminal domain of λ*cI* is separated from the C-terminal "dimerization" domain (Figure 7(e)) by a relatively unstructured ~40 residue linker (not shown).

During the phage life-cycle the C-terminal domain of λ*cI* mediates dimerization as well as the interactions responsible for the cooperative binding of two repressor dimers to pairs of operator sites.[46] The C terminus of λ*cI* is connected to the N terminus of the Gβ1 variants through a four residue linker (Ala₃-Ser) and is located at the end of the small symmetrically related helical segments

Figure 7. Model of the chimeric construct. (a) Surface rendering of the crystal structure of the bacterial RNA polymerase holoenzyme from *T. thermophilus* (1IW7). The coloring scheme for the subunits is as follows: two α subunits, olive green and dark green; β, light gray; β′, dark gray; ω, orange; σ, red. (b) A static schematic illustrating the connectivity of the domains that comprise the chimeric construct. (c)–(f) Backbone ribbon depictions of the crystal structures of the actual domains of the chimeric construct. (c) N-terminal domain of the α-subunit of the *E. coli* RNA polymerase (2DF); (d) the β1-domain of protein G (1PGA); (e) the C-terminal dimerization domain of the λ*cI* repressor (1F39); (f) the N-terminal domain of the λ*cI* repressor in complex with the λ operator DNA (1LMB).

found in close proximity to the dimer interface (illustrated as blue and red ribbons in Figure 7(e)). Finally, the C terminus of the Gβ1 variants is connected to the N terminus of RNAPα (Figure 7 (c)) by a three residue linker (Gly-Asn-Ser). Here again, it is quite apparent from the crystal structure[47] that the N-terminal domain of RNAPα must dimerize properly to form the correct orientation that provides the specific geometry necessary for recruitment of the other highly intertwined components of the RNA polymerase (Figure 7(a)).[48]

We believe that the intervening Gβ1 test variants act in a sense as "molecular rheostats", providing variable resistance in allowing the attached domains to achieve the correct dimer orientations. Stable Gβ1 variants likely maintain the tight α/β fold that possibly inhibits, or partially blocks, optimal dimerization of the other components of the chimeric construct. On the other hand, the less stable variants are more likely to exist in a dynamic equilibrium between partially unfolded states, which more readily allow the optimal intermolecular interactions necessary to achieve the proper shape, and chemical complementarity for enhanced recruitment of the endogenous RNA polymerase. In addition, lack of structural stability may enhance reporter gene expression by enabling the chimeric construct, bound at the λcI operator site, to position RNAPα in proximity to promoter elements more effectively, thus stabilizing the binding of RNA polymerase and activating transcription from the test promoter.

The exploitation of readily available, well-characterized elements of transcriptional machinery from different organisms has provided a novel method to explore the structural determinants of a particular protein fold (Gβ1), a fold that continues to provide insights into factors important for protein structure as well as folding. The unique application of these transcriptional elements represents new possibilities for the creation of novel combinatorial screens that should provide yet more opportunities to explore large regions of protein sequence space rapidly and accurately.

Materials and Methods

Materials

All chemicals and reagents were of the highest quality and obtained from either Sigma-Aldrich or Fisher Scientific International Inc. unless stated otherwise. Oligonucleotides were obtained from Integrated DNA Technologies. Pfu Turbo® DNA polymerase was obtained from Stratagene. Restriction enzymes and buffers were obtained from New England Biolabs. The BacterioMatch® Two-Hybrid System was obtained from Stratagene. All cloning was performed with either *E. coli* XL1-Blue from Stratagene or Top 10 from Invitrogen. Protein expression was performed in BL21(DE3) from Novagen. Point mutations were produced using the QuikChange® method (Stratagene).

Construction of the chimeric construct plasmid

The genes for the chimeric construct (RNAPα and λcI) were sub-cloned from plasmids pTRG, and pBT obtained from a commercially available bacterial two-hybrid system (BacterioMatch® Two-Hybrid System, Stratagene). The gene for RNAP-α was PCR amplified from the pTRG plasmid and cloned into the pBT vector downstream of λcI *via* EcoRI/BamHI (all oligonucleotide sequences are listed in Supplementary Data). The genes for the Gβ1 variants were sub-cloned into the chimeric vector with the engineered restriction sites NotI and EcoRI. The λ-operator DNA and the reporter genes are located on the F′ episome, which is supplied in the bacterial strain XLl-Blue F′ (Stratagene).

Creation of the Gβ1 mutant variants

Point mutants of Gβ1-WT and MonB were produced using the QuikChange® method (Stratagene) with oligonucleotide primers (Integrated DNA Technologies) that contained the appropriate DNA base changes. Synthetic DNA oligonucleotides were used for recursive PCR synthesis of the genes for the MonB-ORDES variant and the genes in the randomized library (oligonucleotide sequences are listed in Supplementary Data).

Assessing bacterial growth on plates containing reporter antibiotic

Chimeric construct plasmids containing the genes for five Gβ1 variants (MonA, Gβ1-WT, MonB-WT, MonB A45V, and MonB A45Y) were transformed into 50 μl of BacterioMatch two-hybrid system reporter strain competent cells (Stratagene). Cells (80 μl) from each transformation reaction were plated on LB control plates (CK) that contained chloramphenicol (12.5 μg/ml) and kanamycin (50 μg/ml). The function of the control plates was to verify colony numbers in the absence of reporter antibiotic. The same volume of cells (80 μl) was plated on reporter antibiotic plates (CCK) that, in addition to chloramphenicol (12.5 μg/ml) and kanamycin (50 μg/ml), contained carbenicillin (both 750 μg/ml and 1000 μg/ml). All plates were incubated at 37 °C for approximately 20 h.

Assessing bacterial growth in liquid medium containing reporter antibiotic

Plasmids that contained the chimeric construct and the genes for all Gβ1 variants tested were transformed into BacterioMatch two-hybrid system reporter strain competent cells (Stratagene) and plated on control CK plates that contained kanamycin (50 μg/ml), and chloramphenicol (12.5 μg/ml). Colonies were picked from these CK plates and grown in 20 ml of LB liquid medium with kanamycin (50 μg/ml), and chloramphenicol (12.5 μg/ml) overnight at 37 °C. From the overnight cultures, cells were diluted to A_{600} of 0.1 in 10 ml of fresh LB containing kanamycin (50 μg/ml), and chloramphenicol (12.5 μg/ml). The cultures were then incubated for 15 min at 37 °C and the A_{600} was measured again to verify that all cultures were the same density. Carbenicillin (3000 μg/ml) was then added, and the cultures were incubated for a total of 4 h, at which time the A_{600} was measured to ascertain the extent of growth.

Protein expression and purification

For protein expression, the genes for the Gβ1 variants were sub-cloned into pET-21a (Novagen) and transformed into BL21(DE3). After growth to an A_{600} of approximately 1, the cells were induced with 1 mM IPTG. The expressed proteins were isolated using a freeze/thaw method,[49] and purification was accomplished with reverse-phase HPLC using a linear 1% min^{-1} acetonitrile/water gradient containing 0.1% (v/v) trifluoroacetic acid. Concentrations of all variants were determined in 6 M guanidine hydrochloride using standard extinction coefficients for the tryptophan and tyrosine residues. Proteins that were ^{15}N-labeled for NMR studies were prepared with standard M9 minimal medium using [^{15}N]ammonium sulfate (2 g/l). Protein purity was verified with SDS-PAGE and reverse phase HPLC, and the correct molecular mass was confirmed by mass spectrometry.

Circular dichroism

The CD data were collected on Jasco-810 spectrometer equipped with a thermoelectric unit and using a 0.1 mm path-length cell. Protein samples were 50 μM in 50 mM sodium phosphate at pH 6.5. Thermal melts were monitored at 218 nm. Data were collected every 1 deg.C with an equilibration time of 2 min. Far-UV spectra were acquired in the continuous mode at 25 °C with 1 nm bandwidth and a 4 s response time. For the thermal denaturation curves, the data were normalized by first shifting all points linearly, such that the $[\theta]_{218}$ value at 5 °C was zero. Then a scaling factor was obtained for each set by dividing the maximum $[\theta]_{218}$ value for all sets at 95 °C (i.e. 53.0) by the $[\theta]_{218}$ value at 95 °C for each set. All data points for each set were then scaled by the unique scaling factor calculated for each set.

NMR spectroscopy

NMR spectra were collected at 293 K on a Varian UnityPlus 600 MHz spectrometer equipped with an HCN triple-resonance probe with triple-axis pulse field gradients. Protein concentrations were ~1.25 mM in 50 mM sodium phosphate at pH ~6.5. Standard 2D [^{1}H, ^{15}N] HSQC spectra were collected for all variants except those selected from the screen. The programs NMRDraw and NMRPipe[50] were used to process the NMR data, and the program NMRView (One Moon Scientific, Inc.) was used to generate and analyze the spectra.

In vitro proteolysis analysis

The enzymes used in this assay were bovine α-chymotrypsin and trypsin obtained from Calbiochem® (catalog# 230832) and the trypsin from ICN Biomedicals, Inc™ (catalog# 101171). Each enzyme was reconstituted to a stock concentration of 20 mg/ml in 50 mM sodium phosphate buffer (pH 6.8) and used to make a 1:60 (v/v) dilution of the 20 mg/ml stock solution to a final concentration of 0.33 mg/ml.

The reactions were carried out at 20 °C, with a constant concentration of enzyme (approximately 1% of the weight of the protein being hydrolyzed), and a constant concentration of each variant (0.625 mM). After fixed amounts of time, each reaction was stopped by adding 500 μl of cold 50 mM sodium phosphate buffer (pH 6.8), and immediately injected onto an HPLC C-18 analytical column. Reactions times were adjusted until approximately 50% degradation of the parent (uncut) peak was achieved.

For two variants (MonB-WT and MonB-ORDES) it was necessary to resolve the digested fragments by SDS-PAGE as opposed to HPLC, as these variants precipitate readily in the acetonitrile/water/trifluoroacetic acid buffer used for HPLC. The reactions for these variants were terminated upon the addition of 10 μl of 2 × SDS loading buffer and boiled for 10 min. The samples were then loaded onto standard SDS/15% (w/v) polyacrylamide gels, run and stained with standard procedures. The 50% degradation time-point was estimated from the multiple time-points resolved on the resulting polyacrylamide gel.

Computational design

Protein design calculations were performed with the ORBIT design algorithms and incorporated backbone-dependent side-chain rotamers derived from the library of Dunbrack and Karplus.[51] Calculations were carried out on positions 23, 27, and 45 using the MonB-WT sequence and previously calculated structure. The residue class assigned to these positions was boundary with solvation effects included, and all amino acids were considered in the calculation except for glycine, proline, and cysteine. The following positions, classified as core residues, were floated and calculated with solvation energy: Y3, L5, A20, A26, F30, A34, and F52. Positions M1, T18, T25, V29, A31, Y33, W43, and K50 were classified as boundary residues and floated with solvation energy included. Solvation energy was not included for the following floated positions, which were classified as surface: K4, V21, D22, A24, D28, Q32, T44, D46, E47, A48, T49, T51, and T53.

Library generation and screening

Recursive PCR was used to construct a library of genes using oligonucleotides that contained random bases at the codons for positions 23, 27 and 45, while the remainder of the genes contained codons from the MonB-WT sequence. Screening of the resulting library entailed primary and secondary steps due to the fact that, in the context of current rendition of the screen, the difference in the growth-rate for the most stable variant (MonA) and the least stable variant (MonB) is similar to natural variations in bacterial growth rates.

For the primary screen, 50 μl of BacterioMatch two-hybrid system reporter strain competent cells (Stratagene) were incubated initially for 15 min with β-mercaptoethanol to increase transformation efficiency. The plasmids that contained the library of MonB mutant variants (randomized at positions 23, 27, and 45) were transformed into the β-mercaptoethanol-treated competent cells and plated onto both control CK plates (kanamycin 50 μg/ml and chloramphenicol 12.5 μg/ml) and reporter CCK plates (kanamycin 50 μg/ml, chloramphenicol 12.5 μg/ml and carbenicillin 1000 μg/ml). Plates were incubated overnight and the next day slow-growing (small) colonies were selected and grown in liquid medium overnight for the purpose of plasmid amplification and isolation, using standard plasmid isolation protocols.

The secondary step of the screen entailed transformation of the isolated plasmids into BacterioMatch two-hybrid system reporter strain competent cells and subsequent plating onto control CK and reporter CCK plates (carbenicillin 1000 μg/ml). The growth rates of the

colonies on the CCK plates were compared to control plates that contained cells transformed with chimeric construct plasmids that contained the genes for MonA, MonB-WT and Gβ1-WT. Variants that grew more slowly than MonB-WT on CCK plates (carbenicillin 1000 μg/ml) were tested in the next step of the screening process.

To further verify the reduced growth-rates of the variants selected for in the previous step, additional measurements were made on cells grown in 96-well plates in LB medium. For each slow-growing library variant isolated from the previous step, six colonies were picked from the CK plates and inoculated separately into 1 ml of LB liquid medium that contained kanamycin (50 μg/ml) and chloramphenicol (12.5 μg/ml). Cultures were grown overnight at 37 °C and diluted the next day into 180 μl of LB in a 96-well plate that contained kanamycin (50 μg/ml), chloramphenicol (12.5 μg/ml) and carbenicillin (2000 μg/ml). The 96-well plate was incubated in a micro-titer plate reader for 24 h at 37 °C with continuous shaking and A_{600} values were measured every 15 min. The six growth curve values for each variant were averaged and compared to the averaged growth curves for the MonA, MonB-WT and Gβ1-WT variants. The genes for slow-growing library variants were sequenced and sub-cloned into a pET-21a expression vector for subsequent protein production purposes.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.10.091

References

1. Agarwal, P. (2006). Enzymes: an integrated view of structure, dynamics and function. *Microb. Cell Fact.* **5**, 2.
2. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
3. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell. Biol.* **6**, 197–208.
4. Mittermaier, A. & Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**, 224–228.
5. Bai, Y. & Feng, H. (2004). Selection of stably folded proteins by phage-display with proteolysis. *Eur. J. Biochem.* **271**, 1609–1614.
6. Hecht, M. H., Das, A., Go, A., Bradley, L. H. & Wei, Y. (2004). *De novo* proteins from designed combinatorial libraries. *Protein Sci.* **13**, 1711–1723.
7. MacBeath, G., Kast, P. & Hilvert, D. (1998). Probing enzyme quaternary structure by combinatorial mutagenesis and selection. *Protein Sci.* **7**, 1757–1767.
8. Magliery, T. J. & Regan, L. (2004). A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics. *Protein Eng. Des. Sel.* **17**, 77–83.
9. Philipps, B., Hennecke, J. & Glockshuber, R. (2003). FRET-based *in vivo* screening for protein folding and increased protein stability. *J Mol Biol*, **327**, 239–249.
10. Auf der Maur, A., Tissot, K. & Barberis, A. (2004). Antigen-independent selection of intracellular stable antibody frameworks. *Methods*, **34**, 215–224.
11. Cabantous, S., Pedelacq, J. D., Mark, B. L., Naranjo, C., Terwilleger, T. C. & Waldo, G. S. (2005). Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J. Struct. Funct. Genomics*, **6**, 113–119.
12. Waldo, G. S. (2003). Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33–38.
13. Magliery, T. J. & Regan, L. (2004). Library approaches to biophysical problems. *Eur. J. Biochem.* **271**, 1593–1594.
14. Magliery, T. J. & Regan, L. (2004). Combinatorial approaches to protein stability and structure. *Eur. J. Biochem.* **271**, 1595–1608.
15. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. (2005). Computational thermostabilization of an enzyme. *Science*, **308**, 857–860.
16. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Jr., Stoddard, B. L. *et al.* (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
17. Dahiyat, B. I. & Mayo, S. L. (1997). *De novo* protein design: fully automated sequence selection. *Science*, **278**, 82–87.
18. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470–475.
19. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
20. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* **351**, 1160–1168.
21. Hayes, R. J., Bentzien, J., Ary, M. L., Hwang, M. Y., Jacinto, J. M., Vielmetter, J. *et al.* (2002). Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl Acad. Sci. USA*, **99**, 15926–15931.
22. Byeon, I. J., Louis, J. M. & Gronenborn, A. M. (2003). A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J. Mol. Biol.* **333**, 141–152.
23. Byeon, I. J., Louis, J. M. & Gronenborn, A. M. (2004). A captured folding intermediate involved in dimer-

ization and domain-swapping of GB1. *J. Mol. Biol.* **340**, 615–625.

24. Goehlert, V. A., Krupinska, E., Regan, L. & Stone, M. J. (2004). Analysis of side-chain mobility among protein G B1 domain mutants with widely varying stabilities. *Protein Sci.* **13**, 3322–3330.

25. Louis, J. M., Byeon, I. J., Baxa, U. & Gronenborn, A. M. (2005). The GB1 amyloid fibril: recruitment of the peripheral beta-strands of the domain swapped dimer into the polymeric interface. *J. Mol. Biol.* **348**, 687–698.

26. Distefano, M. D., Zhong, A. & Cochran, A. G. (2002). Quantifying [beta]-sheet stability by phage display. *J. Mol. Biol.* **322**, 179–188.

27. Kotz, J. D., Bond, C. J. & Cochran, A. G. (2004). Phage-display as a tool for quantifying protein stability determinants. *E. J. Biochem.* **271**, 1623–1629.

28. Alexander, P. A., Rozak, D. A., Orban, J. & Bryan, P. N. (2005). Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code. *Biochemistry*, **44**, 14045–14054.

29. O'Neil, K. T., Hoess, R. H., Raleigh, D. P. & DeGrado, W. F. (1995). Thermodynamic genetics of the folding of the B1 immunoglobulin-binding domain from strepto-coccal protein G. *Proteins: Struct. Funct. Genet.* **21**, 11–21.

30. Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995). A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* **4**, 1108–1117.

31. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K. & Hilvert, D. (2004). An enzymatic molten globule: efficient coupling of folding and catalysis. *Proc. Natl Acad. Sci. USA*, **101**, 12860–12864.

32. Love, J. J., Li, X., Chung, J., Dyson, H. J. & Wright, P. E. (2004). The LEF-1 high-mobility group domain under-goes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry*, **43**, 8725–8734.

33. Shukla, U. J., Marino, H., Huang, P.-S., Mayo, S. L. & Love, J. J. (2004). A designed protein interface that blocks fibril formation. *J. Am. Chem. Soc.* **126**, 13914–13915.

34. Dove, S. L., Joung, J. K. & Hochschild, A. (1997). Activation of prokaryotic transcription through arbi-trary protein-protein contacts. *Nature*, **386**, 627–630.

35. Dove, S. L. & Hochschild, A. (1998). Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12**, 745–754.

36. Dove, S. L. & Hochschild, A. (2004). A bacterial two-hybrid system based on transcription activation. *Methods Mol. Biol.* **261**, 231–246.

37. Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. (1994). Two crystal structures of the B1 immunoglo-bulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry*, **33**, 4721–4729.

38. Huang, P.-S., Love, J. J. & Mayo, S. L. (2005).

Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J. Comput. Chem.* **26**, 1222–1232.

39. Kobayashi, N., Honda, S., Yoshii, H. & Munekata, E. (2000). Role of side-chains in the cooperative beta-hairpin folding of the short C-terminal fragment derived from streptococcal protein G. *Biochemistry*, **39**, 6564–6571.

40. Blanco, F. J., Rivas, G. & Serrano, L. (1994). A Short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Struct. Biol.* **1**, 584–590.

41. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. *et al.* (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, **253**, 657–661.

42. Chothia, C. (1975). Structural invariants in protein folding. *Nature*, **254**, 304–308.

43. Wilson, D. R. & Finlay, B. B. (1998). Phage display: applications, innovations, and issues in phage and host biology. *Canad. J. Microbiol.* **44**, 313–329.

44. Steiner, D., Forrer, P., Stumpp, M. T. & Pluckthun, A. (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nature Biotechnol.* **24**, 823–831.

45. Pabo, C. O. & Lewis, M. (1982). The operator-binding domain of [lambda] repressor: structure and DNA recognition. *Nature*, **298**, 443–447.

46. Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M. (2000). Crystal structure of the [lambda] repressor C-terminal domain provides a model for cooperative operator binding. *Cell*, **101**, 801–811.

47. Zhang, G. & Darst, S. A. (1998). Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science*, **281**, 262–266.

48. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. *et al.* (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6[thinsp]A resolution. *Nature*, **417**, 712–719.

49. Johnson, B. H. & Hecht, M. H. (1994). Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology (NY)*, **12**, 1357–1360.

50. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimen-sional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

51. Dunbrack, R. L., Jr & Karplus, M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574.

52. Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). Reduced surface: an efficient way to compute mole-cular surfaces. *Biopolymers*, **38**, 305–320.

53. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.