Published in final edited form as:

J Mol Biol. 2008 October 10; 382(3): 721–733. doi:10.1016/j.jmb.2008.07.036.

High-throughput analysis of the protein sequence-stability landscape using a quantitative "yeast surface two-hybrid" system and fragment reconstitution

Sanjib Dutta¹, Akiko Koide, and Shohei Koide*

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637

Abstract

Stability evaluation of many mutants can lead to a better understanding of the sequence determinants of a structural motif and of factors governing protein stability and protein evolution. The traditional biophysical analysis of protein stability is low throughput, limiting our ability to widely explore the sequence space in a quantitative manner. In this study, we have developed a high-throughput library screening method for quantifying stability changes, which is based on protein fragment reconstitution and yeast surface display. Our method exploits the thermodynamic linkage between protein stability and fragment reconstitution and the ability of the yeast surface display technique to quantitatively evaluate protein-protein interactions. The method was applied to a fibronectin type III (FN3) domain. Characterization of fragment reconstitution was facilitated by the co-expression of two FN3 fragments, thus establishing a "yeast surface two-hybrid" method. Importantly, our method does not rely on competition between clones and thus eliminates a common limitation of high-throughput selection methods in which the most stable variants are predominantly recovered. Thus, it allows for the isolation of sequences that exhibits a desired level of stability. We identified over one hundred unique sequences for a β -bulge motif, which was significantly more informative than natural sequences of the FN3 family in revealing the sequence determinants for the β -bulge. Our method provides a powerful means to rapidly assess stability of many variants, to systematically assess contribution of different factors to protein stability and to enhance protein stability.

Keywords

protein stability; protein design; library screening; directed evolution; yeast surface display

Introduction

Quantifying the conformational stability of a large number of protein variants establishes the sequence-stability relationship for a specific structural motif or a fold. Comprehensive knowledge of the sequence-stability relationship is fundamentally important for protein structure prediction and protein design. A traditional strategy to establish this relationship is to produce a number of variants and measure their stability using biophysical techniques (i.e. "sequence first, stability second").¹ However, this strategy requires a number of time-

Corresponding author: E-mail, skoide@uchicago.edu.

¹Present address: Department of Biology, Massachusetts Institute of Technology.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

consuming steps including mutagenesis, expression, purification and biophysical measurements, and thus, although information rich, it is inherently low throughput. To circumvent the sampling limitation of biophysical approaches, naturally observed sequence diversity for a fold has been used as a source for the sequence-stability relationship. It is assumed that the occurrence of a given amino acid type at a position among aligned sequences is correlated with its importance for the stability or function of the protein. However, because protein stability determinants are usually distributed throughout a protein molecule and the thermodynamic contribution of a single position is not necessarily linked to the natural selection, residue conservation from natural sequences across species is not always a strong indicator of its thermodynamic importance as have been comprehensively demonstrated by mutational studies of a protein-protein interface.² Therefore, a high-throughput (HTP) method that quantifies protein conformational stability would have a large impact on our ability to generate a quantitative and comprehensive sequence-stability relationship.

HTP library selection strategies have been developed that identify variants with desired properties such as high stability from a large ensemble of variants.^{3; 4; 5; 6; 7; 8; 9} In this class of experiments, the linkage between the conformational stability and protein function is exploited. In a favorable case, protein stability is linked to the organism phenotype,¹⁰ and more commonly to the function of the protein of interest such as enzymatic activity and binding to its partner.^{11; 12; 13} Although variants with improved stability have been obtained from these function-based selections,⁷ this type of selection is effective within a narrow range of protein stability where a significant fraction of protein is denatured.^{12; 14} A widely used method to isolate stable protein variants in the absence of function-based selection is proteolytic selection coupled with phage display.⁶ Schmidt and coworkers have successfully isolated stable variants of several proteins using this approach,^{6; 15} although thermodynamic stability is not always correlated with proteolytic stability.⁵

A major limitation of library selection approaches is that these methods are by nature a competitive technique and more stable variants are always preferentially selected over less stable clones. As a result, it is difficult to identify a collection of variants with a wide range of stability in the presence of highly stable ones that dominate the selection. It is however possible to establish quantitative relationships among selected sequences from their relative population in the selected sequence pool.¹⁶ Using the "shotgun scanning" approach, Cochran and coworkers estimated the unfolding free energy differences ($\Delta\Delta G$) of amino acids at a given position with reasonable accuracy.¹³ This method, however, requires a large number (often in the hundreds) of sequences even for deriving a relatively small range of $\Delta\Delta G$ values.

Given the paucity of HTP methods for quantitative assessment of protein stability, in this work we aimed to establish a HTP screening technology for protein stability. We reasoned that, rather than a selection method that identifies variants with the highest level of stability in a population, a screening method that can yield the level of stability of each variant and isolate them according to an arbitrary level of stability (i.e. "stability first, sequence second") would be a more powerful tool for establishing the sequence-stability relationship. Such method requires the establishment of a parameter that can directly report the level of stability of a protein and a readout method for the parameter that does not subject each variant to competition with the rest of the population. We have fulfilled these requirements by a novel combination of protein fragment reconstitution and yeast surface display screening.

The regeneration of the function (or native fold) of a protein by its fragments is a well studied phenomenon and is referred to as fragment reconstitution (also as fragment complementation).^{17; 18; 19; 20; 21; 22; 23} The strength of interaction between the fragments has been demonstrated to be directly related to the stability of the uncut protein (i.e. $\Delta\Delta G_{\text{reconstitution}} \approx \Delta\Delta G_{\text{unfolding}}$ for the same mutation).^{24; 25} Therefore, fragment reconstitution converts the

task of quantifying protein stability into a task of quantifying protein-protein interaction affinity.

Yeast surface display is a HTP and quantitative screening technique.^{26; 27} In this methodology, a protein of interest is displayed on the yeast surface as a fusion protein to a cell surface protein Aga2p. Because yeast cells serve as large particles suitable for fluorescence activated cell sorting (FACS), the display level of the protein and ligand binding to it can be quantitatively determined for each cell expressing a unique variant. Therefore, protein variants can be sorted based on their own binding properties in the absence of competition with the rest of the population. A modern flow cytometer can sort millions of cells within a short period therefore facilitating an analysis of a large population of protein variants.

In this study, we investigated the sequence-stability relationship of the tenth type III domain (FN3) of human fibronectin (FNfn10). FN3 has a β -sandwich fold with seven β -strands and ubiquitously found in animal proteins (Figure 1A).^{28; 29} FNfn10 is widely used as a molecular scaffold for engineering antibody mimics.^{30; 31; 32} Because FNfn10 is very stable ($T_m \sim 90^\circ$ C), function-based selection using aforementioned methods is unlikely to discriminate small stability changes among variants. Also FNfn10 contains flexible loops that may be sensitive to protease cleavage regardless of the overall stability,³³ rendering proteolysis-based selection difficult. These characteristics of FNfn10 call for the development of new HTP methods for stability measurement. We have previously demonstrated fragment reconstitution of FNfn10 and a qualitative correlation between the stability effects of mutations and their effects on *in vivo* affinity of fragment reconstitution as measured with yeast two-hybrid.²⁵ The full-length FNfn10 protein has been successfully displayed on the yeast surface.³⁴ Here we describe the establishment of a "yeast surface two-hybrid" method for FNfn10 fragment reconstitution and its capacity to provide a quantitative readout for the conformational stability of the uncut protein. We used this strategy to establish the sequence-stability relationship for a structurally conserved β -bulge motif in the FN3 fold. Importantly, our stability-based screening demonstrates distinct amino acid preferences for the β -bulge motif that is not evident from sequence variations of natural FN3 family members.

Results

Yeast surface display of FNfn10 fragments

In this work, we used FNfn10 fragments that are bisected between N52 and S53, which will be referred to as the N and C fragments, respectively. The N fragment includes β -strands A–C and the C fragment D–G (Figure 1A). These fragments reconstitute the native fold with high affinity.²⁵ We first displayed the N fragment on the yeast surface and separately prepared the C fragment as a ubiquitin fusion protein with a V5 tag (termed Ubi-C). We fused the less soluble N fragment²⁵ to Aga2p so that the binding of the more soluble C fragment to cells displaying the N fragment is monitored (Figure 1B). In addition, we introduced a stabilizing mutation, D7N,³⁵ in the N fragment to potentially improve the extent of reconstitution. The N_{D7N} fragment was robustly displayed on the yeast surface as detected with an antibody directed to the FLAG epitope tag attached to the C-terminus of the N_{D7N} fragment (Figure 1B and 1C). We have shown that the N fragment lacks a well-defined structure,²⁵ indicating that the quality control process that reduces the secretion/surface display of unfolded protein^{36; 37} is not stringent, at least for a short peptide like the N_{D7N} fragment (52 residues plus linkers).

We were unable to detect binding of the purified C fragment (Ubi-C) to the surface displayed N_{D7N} fragment even at micromolar concentrations of the C fragment (Figure 1D), although the dissociation constant (K_d) for these fragments as determined in a previous *in vitro* experiment is in the low nanomolar range.²⁵ We reasoned that the expressed N_{D7N} fragment may be misfolded in a conformation incapable of binding to the C fragment *in vitro*. Our display

system was based on a high copy number plasmid, which could lead to a high-level expression and aggregation. We tested a low copy number plasmid to lower the expression level of the N_{D7N} fragment and therefore reduce the chances of aggregation. Although this modification led to a higher fraction of cells displaying the N fragment, still no binding was observed (data not shown).

Yeast surface two-hybrid system rescued FNfn10 fragment reconstitution

We reasoned that coexpression of the N_{D7N} and C fragments might prevent the N_{D7N} fragment from entering non-productive pathways. Therefore we modified our display system so that both fragments were coexpressed from a single plasmid. The N_{D7N} fragment was displayed on the yeast surface, as in our first system described above, and the C fragment was secreted as a free protein into the media. The expression of both fragments was driven by the Gal 1/10 promoter (Supplementary Figure 1) that gives tight and inducible regulation and bidirectional control, which may lead to the production of stoichiometric amounts of the two fragments. With this coexpression system, successful reconstitution of the fragments on the yeast surface was observed, as evidenced by a high degree of correlation between N_{D7N} fragment expression and C fragment binding (Figure 1E). The surface display level of the N_{D7N} fragment was unaffected by the expression of the C fragment. When the surface displayed N fragment is removed by treating cells with dithiothreitol, no signal for the C fragment was detected (data not shown), indicating that only C-fragment bound to the N fragment gives rise to the signal. Because this system can detect the interaction of two proteins on yeast surface, we term it "yeast surface two-hybrid".

Quantitative evaluation of fragment reconstitution using yeast surface two-hybrid

Next, we examined whether our yeast surface two-hybrid system can quantitatively detect interactions of FN3 fragments. It should be noted that a lack of detectable interactions can arise either due to the absence of interactions between two functional fragments or the production of a non-functional fragment. In order to unambiguously distinguish these possibilities, we first assessed whether the N_{D7N} fragment coexpressed with a weakly associating mutant C fragment maintains tight binding to the wild-type C fragment added in trans after the removal of the weakly associating C fragment. Therefore, we introduced the I59A mutation in the C fragment, which is known to decrease the stability of uncut FNfn10 by $\sim 3.6 \text{ kcal mol}^{-1}$.³⁸ We have previously shown that this mutation greatly reduces the strength of fragment reconstitution in vivo.²⁵ The co-expression of the N_{D7N} fragment with the I59A C fragment did not significantly affect the surface display level of the N_{D7N} fragment but reduced the extent of fragment reconstitution to a level close to that observed in the absence of added C fragment (compare Figure 2A with Figure 1D). However, a subsequent addition of a purified, wild-type C fragment to these cells restored a high level of fragment reconstitution (Figure 2B). These results indicate that the coexpression of the weakly associating C fragment somehow protects the N fragment from entering a nonproductive pathway. Using this strategy, a titration curve of the C fragment was made (Figure 2C), which yielded a K_d value of $25.3 \pm 2 \text{ nM}$. Together, these results establish that when the N and C fragments are coexpressed a low level of binding readout most likely reports a low level of reconstitution, not non-functional display of the N fragment.

To determine the level of quantitateness of our yeast surface two-hybrid method, we compared interactions of additional fragment pairs: (i) N_{WT} and C_{WT} ; (ii) N_{WT} and C_{I59V} ; and (iii) N_{WT} and C_{V50A} . The I59V and V50A mutations in the C fragment are known to destabilize FNfn10 to different degrees.³⁸ We observed that the display levels of the N fragments in these contexts were similar, unaffected by the stability of the complex (Figure 3A). The levels of fragment reconstitution were clearly influenced by these mutations (Figure 3A). To assess the degree of fragment reconstitution, we defined a parameter "B/E ratio" as the ratio of the PE

fluorescence intensity (i.e. binding) over the FITC fluorescence intensity (i.e. expression). The B/E ratio for the cells ranked within the top 10% of surface display showed a remarkable correlation ($R=0.99$) to the stability of the uncut, parent protein (Figure 3B). We used only high-displaying cells to eliminate larger errors in the B/E ratio of low displaying cells. As the stabilities of the parent proteins varied over a range of $2.4 \text{ kcal mol}^{-1}$, the yeast surface two-hybrid can quantitatively monitor protein-protein interactions over this dynamic range and therefore provide quantitative measures of FNfn10 stability.

Mapping the sequence space using yeast surface two-hybrid

To investigate whether our approach can rapidly evaluate the sequence space for a given structural motif, we first analyzed two single positions. We chose these positions within the N fragment so that potential effects of mutations on the surface display level can be detected and compensated. Also the identical wild-type C fragment was used to minimize differences in expression level across all variants in a library. First, we randomized W22, a buried residue in the core using the NNK codon (where N is a mixture of A, T, G and C and K is a mixture of G and T; it encodes all amino acids) and analyzed the mixture of yeast cells by FACS (Figure 4B). Only ~3.7 % of high-expression cells exhibited strong binding, indicating most of the library members were incapable of reconstitution and thus severely destabilizing. This ratio of 3.7% is very close to 3.8 % that corresponds to 1 member out of a total of 31 members of the NNK library (32 codon combinations minus one stop codon). Sequencing of individual clones revealed that only Trp was present at the position in the clones exhibiting strong binding, supporting the conclusion from the population analysis. The yeast surface two-hybrid data suggest that FNfn10 cannot accommodate any other amino acid at position 22 without severe destabilization. This is not surprising because W22 is highly buried in the core and also highly conserved throughout the FN3 family³⁹. The W22F mutation has been reported to destabilize FNfn10 by $1.94 \text{ kcal mol}^{-1}$, which would exhibit very weak signal in the yeast surface two-hybrid assay.³⁸

In the second case we chose Leu19, a surface exposed amino acid. (Figure 4A). A similar analysis of a library in which Leu19 is randomized with the NNK codon indicated that ~71% showed binding comparable to or higher than the wild type population. Sequencing of 24 clones from this population showed ten different amino acids with no clear bias for the wild type residue (F (1), I (1), K (1), L (2), M (2), Q (1), R (2), S (7), T (5), V (1); the numbers in parenthesis indicate the number of occurrence). Thus both the population and sequence analyses indicate that position 19 can tolerate many amino acid types with minimal effect on the stability. This result is in agreement with the fact that position 19 is not highly conserved among natural FN3 domains.³⁹ Therefore, using our method the degree of amino acid substitution tolerance at a single position can be immediately determined from the population analysis of a randomized library.

Next, we examined the sequence space for a three-residue cluster that forms a structurally conserved β -bulge among FN3 domains. Residues Val11, Ala12 and Leu19 in the N fragment constitute the β bulge (Figure 1A). In the conventional β bulge nomenclature, the two residues on the bulged side are termed “1” and “2” (Val11 and Ala12, respectively) and the residue on the opposite side is termed “X” (Leu19).⁴⁰ To identify sequence variants of the β bulge that cause different degrees of stability change, we constructed a yeast surface two-hybrid library in which these three residues were randomized with all possible amino acids ($20^3 = 8,000$ amino acid combinations from $32^3 = 32,768$ DNA combinations; referred to as “bulge library”). As in the single-position libraries, the C fragment in this library was the wild type. As expected, the naïve library showed diverse degrees of binding (Figure 4C). Cells were sorted based on the level of binding into four classes; E (excellent), G (good), M (moderate) and P (poor) (Figure 4C, D). Forty eight clones from each of these four regions were recovered and

individually analyzed. Individual clones showed the same extent of binding as observed for the class from which they were selected, indicating stringent discrimination during sorting (Supplementary Figure 2). Amino acid sequences of individual clones revealed that most clones from class E have the consensus sequence β -Xaa- β where β is a β -branched amino acid and Xaa is any amino acid (Figure 5A). The remaining clones from this class had a β -branched amino acid in at least one of positions "1" or "X" (i.e. β -Xaa-Xaa or Xaa-Xaa- β) or mutations at positions 7 or 9 (see below). Clones from class G revealed more variability in their sequences (Figure 5B). Only, 3 out of a total of 39 clones had the β -Xaa- β sequence though a substantial number (23 of 39) had the β -Xaa-Xaa or Xaa-Xaa- β sequence. It is worth mentioning that we were able to retrieve the wild type sequence from this bin. In a stark contrast, none of the analyzed clones from classes M and P (i.e. weak binding bins) had the β -Xaa- β motif, and only a small number (6 of 40 in class M, and 12 of 41 in class P) had the β -Xaa-Xaa or Xaa-Xaa- β sequences (Figure 5C, D).

It should be noted that only ~30% of clones from the L19 library contained a β -branched amino acid (see above), although L19 is enriched with a β -branched amino acid in the β -bulge library screening. This difference is probably due to different levels of screening stringency. In the L19 library screening, only the "wild type-like or better" population was recovered (Fig. 4 A and B), while clones were finely binned in the β -bulge library screening (Fig. 4C).

In class E, there were a few clones that did not have the consensus motif but instead had an additional mutation at either position 7 or 9 (Figure 5A), which were probably introduced due to PCR errors during library construction. The presence of these mutations in class E can be rationalized by our group's previous finding that the electrostatic repulsion between the carboxyl groups of Asp7 and Glu9 destabilizes FNfn10 and the removal of the negative charge of Asp7 leads to stabilization.³⁵ While these variants needed to be excluded from the dataset for sequence space mapping, they demonstrate that our strategy can be employed to identify variants that are more stable than the wild type from a large library constructed by a technique such as error-prone PCR.

Three clones from each of the four classes were reformatted into full-length (uncut) FNfn10 and their stability was determined *in vitro* (Table 1). The mutations from classes E and G resulted in the level of stability that is similar to or slightly higher than the stability of the wild type protein. Those from classes M and P were destabilizing by varying amounts. The degree of fragment reconstitution for the different pairs as measured by FACS showed a good correlation with the stability of the mutant proteins ($R=0.89$; Figure 5E), indicating that the yeast surface two-hybrid readout is indeed capable of identifying protein sequences with a predefined level of stability.

A comparison of the experimentally defined sequence space with that derived from natural sequence diversity

To examine whether we can obtain similar information about the sequence determinants of the β -bulge from sequence analysis of natural FN3 domains, we analyzed the amino acid variation in the three positions of the β -bulge for all 821 natural FN3 sequences available in the SMART database⁴¹ (Version 5.1). Unlike the E class clones from our library these FN3 family members did not exhibit clear sequence bias toward β -branched amino acids at positions "1" or "X" (data not shown). Since our library selection revealed a clear sequence bias for the β -Xaa- β motif, we also examined the amino acid distribution at these two positions in the natural FN3 sequences (Figure 6A). Interestingly, we observed that residue pairs overrepresented in this natural sequence dataset populate similar regions of sequence space to those occupied by sequences from the E and M classes (Figure 6B). The preference by the natural sequences implies an evolutionary pressure for these positions that is probably linked to stability. While

this dataset clearly revealed the presence of a sequence bias at these positions, the small number of sequences precluded an estimation of relative stability using Boltzmann analysis.

In contrast, the data from the yeast surface two-hybrid analysis allowed for a quantitative "sequence landscape" for the two positions (Figure 6B). Importantly, only a few residue pairs appeared in different stability bins (half circles with different colors in Figure 6B), suggesting the capability of our method to finely discriminate stability differences. Among these overlapping pairs all but one are from adjacent stability bins, implying that the contribution to stability for the most part is accounted from these two positions "1" and "X" and not significantly influenced by residue identity at position "2". The occurrence of the pair "L11/V19" in both the E (excellent) and P (poor) class is an exception to this rule, probably due to the presence of a highly destabilizing proline at position "2" in the clone from the P class (clone P5; Figure 5D).

Discussion

We have developed a variation of yeast surface display, "yeast surface two-hybrid". While co-expression of components has been used for yeast surface display of heterodimeric and oligomeric proteins such as Fab and streptavidin,^{42,43} to our knowledge our system is the first application of yeast surface display to the assembly of complexes from short unstructured peptides and a quantification of the affinity of such interactions. It is likely that the yeast surface two-hybrid can be readily employed to study similar types of peptide assemblies such as coiled coils.

Although the N fragment of FNfn10 is disordered in isolation,²⁵ it is robustly displayed on the yeast surface. This is in contrast to previous observations that secretion and surface-display efficiency in yeast is strongly correlated to the thermodynamic stability.^{36,37} Our results, however, is in agreement with a recent study showing indistinguishable display levels for the molten globule and stably folded forms of a small protein.⁴⁴ Together, these results suggest that, at least for small proteins, the quality control mechanism involving yeast surface display is not as stringent as originally thought, and that it cannot be used to eliminate unstructured proteins.

The cosecreted C fragment serves as a chaperone for maintaining the N fragment in a functional form, suggesting that the formation of the complex early in the secretory pathway prevented the N fragment from being displayed in a form not suited for binding. Interestingly, even the presence of a mutant C fragment with low affinity was sufficient. We speculate that because of spatial constraints in the yeast cells the effective concentrations of these fragments during biosynthesis is much higher than those after surface display and secretion, and thus the low affinity is sufficient for the transient formation of the complex. Coexpression of single chain antibodies (scFv) with chaperones has been shown to increase the secretion efficiency of scFv in yeast.⁴⁵ Thus, coexpression of a binding partner may expand the range of proteins that can be functionally displayed on the yeast surface.

By combining yeast surface two-hybrid and fragment reconstitution, we have established a HTP method for quantitatively evaluating protein stability and identifying variants within a predefined range of stability. These unique characteristics are made possible by the fact that yeast surface display is a screening method that evaluates individual members rather than a selection method that isolates the best members from a population and that the screening readout (the B/E ratio) given for each library member directly indicates the level of stability. Selection-based methods do not have such readout, and thus it is necessary to set up a competition among library members from which one could obtain population differences among them that is then converted into free energy differences using the Boltzmann analysis.

Because the stability difference is proportional to the logarithm of population difference in the Boltzmann analysis, a large dataset would be necessary if one wishes to obtain a statistically significant value for a large stability difference. The requirement of the population ratio between variants to derive free energy differences makes it difficult to analyze higher-order correlations involving multiple positions. For example, analyzing 1200 clones was not sufficient for obtaining statistically significant free energy values for many pair-wise contributions to protein G stability.¹³ This difficulty is compounded by the natural consequence of selection in which stable clones dominate the selected population and limit its sequence diversity. In contrast, in our method stability of each sequence can be readily assigned based on its B/E ratio, and thus sparse coverage of all possible sequences in a dataset is not a limiting factor. The dynamic range of our method corresponds to ~ 2.4 kcal mol⁻¹ of stability change, which equates to a 1000-fold difference in the equilibrium constant. This is similar to the range obtained earlier from shotgun scanning data.¹³ We envision that the dynamic range of our method can be further expanded by including a denaturant or a stabilizing reagent in the reaction.

The FACS profiles of the two single position libraries (Fig. 4) showed stark contrast and they immediately revealed the sequence space allowed for these positions. The sequence space information from our experiments approximately recapitulated natural sequence variations seen for the FN3 family. While these studies were performed for the proof of concept and thus did not offer new insights into the thermodynamic roles of the FN3 residues studied, such experiments can be easily refined using finer stability bins. Also, this type of analysis would allow for rapid and economical mapping of the sequence-stability relationship for a protein with few homologues.

Though we observed near perfect correlation between stability and affinity of fragment reconstitution for our reference set of mutants ($R=0.99$; Figure 3B), the correlation for the entire set of mutants was not as good ($R=0.89$; Figure 5E). Several possible reasons for the lower correlation can be considered. First, our screening involves several washing steps and thus it is not solely dictated by the equilibrium affinity of the fragments. Second, for the perfect correlation, a mutation needs to equally affect the native state of the uncut protein and the reconstituted fragments (and equivalently the unfolded state of the uncut protein and the dissociated state of the fragments). It is possible that a mutation has differential long-range effects on the region near the cleavage site between the uncut and cut proteins. Likewise, a mutation can differentially affect the unfolded/dissociated states, e.g. by forming different residual structures. Although it may be difficult to completely eliminate such undesired effects due to the fundamental plasticity of proteins,^{46; 47} these complications can be minimized by choosing a cleavage site that is distantly located from a region of interest.

The imperfect correlation between stability and affinity for the entire set of mutants may preclude the application of our method as a complete substitute for biophysical measurements of individual mutants to obtain accurate $\Delta\Delta G$ values for stability changes. Also one needs a set of mutants with known stability for the calibration of the B/E ratio. While these can be viewed as limitations, a clear strength of our method lies in identifying pools of different stability ranges that can subsequently be analyzed rigorously.

Among HTP methods, our method is unique in that it can identify sequences based on their *absolute* stability rather than *relative* stability within a particular sequence repertoire. This is made possible by the fact that the B/E ratio can be calibrated a priori with a set of clones of known stability. This is a clear advantage when attempting to identify variants that are more stable than the starting protein because the FACS profile of a library can immediately tell whether such variants exist in the library.

For the β -bulge motif, we identified 149 unique sequences with assigned stability. While the estimated stability values do contain errors of ~ 0.5 kcal/mol, the ability to rapidly assign stability to many sequences gives a unique opportunity to produce a high-quality database for mapping a sequence landscape. It is interesting that the distribution of natural sequences generally agrees with that of our library analysis but the former is significantly broader than the latter, suggesting that the stability of this β -bulge motif is an important but not dominant determinant of the natural FN3 sequences. Because the "synthetic" sequence landscape from our library is devoid of any other mutations outside the region of interest and the screening is solely based on stability, it was straightforward to interpret the importance of the dominant β -Xaa- β motif to the stability of this β -bulge. It is conceivable that one can expand this type of analysis to a structural element involving a larger number of amino acid residues.

In summary, the combination of yeast surface two-hybrid and fragment reconstitution provides a powerful means to rapidly map the sequence landscape for a structural motif and to isolate stable variants. The highly controlled sequence diversity in the synthetic sequence landscape is superior to the natural sequence landscape for extracting purely energetic determinants of a structural motif. Many proteins with diverse folds have been reconstituted from fragments (e.g. β -galactosidase,⁴⁸ ribonuclease,¹⁷ green fluorescence protein,^{49; 50} dihydrofolate reductase²³ and ubiquitin²²), and a method to identify a successful bisection point has been developed.⁵¹ At the same time, modern flow-cytometers are commonly available and high-throughput DNA sequencing is becoming increasingly affordable. Therefore, this method should be broadly applicable to many systems, which would significantly expand the sequence information available to a variety of structural motifs.

Materials and Methods

Strains and Media

Yeast strain EBY 100 was purchased from Invitrogen. Plasmid pRS424 was a generous gift from Dr. Gela Tevzadze (University of Chicago). Yeast cells were grown in selective glucose (SDCAA) or galactose media (SGCAA) following previously published protocols.⁵² Antibodies for fluorescent labeling were purchased from Sigma, St. Louis, MO.

Vector construction

The yeast display vector for FNfn10 fragment reconstitution is based on the pRS424 plasmid⁵³ (Supplementary Figure 1). A 676 nucleotide *Bam*HI-*Eco*RI fragment containing the *GAL1-10* upstream activating sequence including the transcriptional start sites for the *GAL1* and *GAL10* genes from pBM150⁵⁴ was inserted into the polylinker site for pRS424, resulting in vector pRS424-GAL(*TRP1*). A PCR fragment containing the Aga2 open reading frame, FNfn10 N-fragment followed by a FLAG tag with flanking *Xho*I and *Eco*RI sites was inserted into pRS424-GAL (*TRP1*) to generate pRS-FN-ABC. Next, a PCR fragment containing the Aga2 signal sequence, FNfn10 C-fragment followed by a V5 tag was cloned into the *Spe*I and *Not*I sites of pRS-FN-ABC to generate the final vector pRS-FN-ABC-DEFG. The presence of two KpnI sites between the ends of the N-fragment of pRS-FN-ABC-DEFG facilitated the construction of an acceptor vector for yeast homologous recombination. To prepare the acceptor vector pRS-FN-ABC-DEFG was digested with KpnI to remove the N-fragment and religated to generate pRS-FN-DEFG. Mutations in the N and C fragment were created by subcloning appropriate PCR fragments in the plasmid pRS-FN-ABC-DEFG.

Library Construction

All mutagenesis experiments were done using pRS-FN-ABC-DEFG as the template using PCR with a mutagenic primer containing a degenerate codon (MNN, N denotes a mixture of A, T, G and C, and M a mixture of A and C). To prepare a library the yeast strain EBY100 was

transformed with the cut acceptor vector and a PCR fragment following the procedure of Gietz et al.⁵⁵

Induction and fluorescence labeling of yeast cells

Cells from freshly grown culture in SDCAA were transferred to 2 ml of SGCAA with a starting O.D. of ~0.2 and then grown at 30 degrees for 24–32 hours before harvesting.

Induced yeast cells from 2 ml of culture were harvested by centrifugation, washed with TBS (50mM Tris, 100mM sodium chloride, pH 8) and used immediately or after overnight storage at 4 degrees for labeling. Approximately 10^6 yeasts were suspended in BSS (50mM Tris, 100mM sodium chloride, pH 8, 1mg/ml BSA) containing 100-fold diluted Anti-FLAG rabbit antibody (Sigma) and 500-fold diluted Anti-V5 monoclonal antibody and incubated on ice for 30–60 minutes. Cells were then pelleted and washed once with BSS before incubating with secondary antibodies, 100-fold-diluted fluorescein-conjugated (FITC) goat anti-rabbit antibody and R-Phycoerythrin (PE)-conjugated goat antimouse IgG (Sigma) for 30–60 minutes. Finally, cells were washed once before suspending in 300 μ l of BSS for flow cytometry analysis.

Flow cytometric analysis and sorting

Labeled yeast cells were analyzed on a BD FACScan flow cytometer at the University of Chicago Flow Cytometry Facility. The yeast cell population was gated by forward light scatter to reject clumped cells. Typically, data for 10,000 events were collected for analysis. Sorting was carried out on a BD FACS Aria. Approximately $1-2 \times 10^7$ cells were sorted with sorting windows set to collect the top 80% of expression positive cells. The sorting windows were calibrated using cells expressing four standard FN3 fragment pairs (wild type, D7N, I59V and V50A). After 3 rounds, sorted populations were plated on SDCAA plates to obtain single clones. Individual clones were further characterized by flow cytometry and their DNA sequences were determined.

Quantitative equilibrium binding experiments of yeast displayed N fragment with the purified C fragment were performed as described.⁵² To quantify the extent of binding a PERL script was written to calculate the PE/FITC ratio of the top 1000 high-expressing cells. This binding/expression (B/E) ratio was used to determine the equilibrium dissociation constant (K_d) according to the equation, $B/E = m_1 + (m_2 - m_1) * x / (K_d + x)$ where x is the concentration of the C fragment, m_1 is B/E in the absence of the C fragment, and m_2 is B/E with the saturating concentration of the fragment.

Protein expression and purification

The *E. coli* expression vector for wild type FNfn10 has been described.³⁵ The gene for the ubiquitin fusion of the FNfn10 C fragment with the V5 tag ("Ubi-C") was subcloned into the expression vector pET24a (Novagen) so that the ubiquitin gene is fused N terminal to the C fragment.⁵⁶ All mutations were introduced using standard PCR methods and the genes of all the mutants were verified by DNA sequencing. All the mutant proteins were expressed as soluble proteins and subsequently purified using metal affinity chromatography, as previously described.³⁰

Protein stability measurement

Chemical denaturation reactions were monitored using Trp fluorescence. All fluorescence experiments were performed at 30 °C using a Spectronics AB-2 spectrofluorometer equipped with an automated titrator as described previously.^{30; 57} For chemical denaturation measurements, proteins were dissolved at a final concentration of 1 μ M in 20mM sodium citrate

buffer at pH 6 containing 100mM sodium chloride. Increasing concentrations of guanidine thiocyanate (GuSCN) were added and the change in fluorescence from the single tryptophan residue in FNfn10 was monitored as described.³⁰ The GuSCN concentrations before and after a titration experiment were determined using an Abbe refractometer (Spectronics Instruments) as described.⁵⁸ Data were fitted with the two-state unfolding model³⁰ using the program Igor Pro 3.0 (Wavemetrics, Lake Oswego, Oregon).

Analysis of natural sequences of the FN3 family

Natural sequences for FN3 were retrieved from the SMART database (Version 5.1) along with the associated ClustalW multiple sequence alignment (MSA). The frequency of occurrences for single amino acid or residue pairs was calculated from the MSA data using a Perl script.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ryan Gilbreth and the members of the University of Chicago Flow Cytometry Facility for assistance. This work was supported in part by National Institutes of Health grants R01-GM72688 and R21-CA132700 and by the University of Chicago Cancer Research Center.

References

1. Shortle D. Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem* 1989;264:5315–5318. [PubMed: 2647715]
2. Pal G, Kouadio JL, Artis DR, Kossiakoff AA, Sidhu SS. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J Biol Chem* 2006;281:22378–22385. [PubMed: 16762925]
3. Magliery TJ, Regan L. Combinatorial approaches to protein stability and structure. *Eur J Biochem* 2004;271:1595–1608. [PubMed: 15096198]
4. Watters AL, Baker D. Searching for folded proteins in vitro and in silico. *Eur J Biochem* 2004;271:1615–1622. [PubMed: 15096200]
5. Bai Y, Feng H. Selection of stably folded proteins by phage-display with proteolysis. *Eur J Biochem* 2004;271:1609–1614. [PubMed: 15096199]
6. Sieber V, Pluckthun A, Schmid FX. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* 1998;16:955–960. [PubMed: 9788353]
7. Jung S, Honegger A, Pluckthun A. Selection for improved protein stability by phage display. *J Mol Biol* 1999;294:163–180. [PubMed: 10556036]
8. Lim WA, Sauer RT. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 1989;339:31–36. [PubMed: 2524006]
9. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. Protein design by binary patterning of polar and nonpolar amino acids. *Science* 1993;262:1680–1685. [PubMed: 8259512]
10. Das G, Hickey DR, McLendon D, McLendon G, Sherman F. Dramatic thermostabilization of yeast iso-1-cytochrome c by an asparagine----isoleucine replacement at position 57. *Proc Natl Acad Sci U S A* 1989;86:496–499. [PubMed: 2536164]
11. Kim DE, Gu H, Baker D. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc Natl Acad Sci U S A* 1998;95:4982–4986. [PubMed: 9560214]
12. Kotz JD, Bond CJ, Cochran AG. Phage-display as a tool for quantifying protein stability determinants. *Eur J Biochem* 2004;271:1623–1629. [PubMed: 15096201]
13. Distefano MD, Zhong A, Cochran AG. Quantifying beta-sheet stability by phage display. *J Mol Biol* 2002;322:179–188. [PubMed: 12215423]

14. Zhou HX, Hoess RH, DeGrado WF. In vitro evolution of thermodynamically stable turns. *Nat Struct Biol* 1996;3:446–451. [PubMed: 8612075]
15. Wunderlich M, Martin A, Staab CA, Schmid FX. Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* 2005;351:1160–1168. [PubMed: 16051264]
16. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A* 2000;97:8950–8954. [PubMed: 10908667]
17. Kato I, Anfinsen CB. On the stabilization of ribonuclease S-protein by ribonuclease S-peptide. *J Biol Chem* 1969;244:1004–1007. [PubMed: 5769175]
18. Ullmann A, Jacob F, Monod J. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *J Mol Biol* 1967;24:339–343. [PubMed: 5339877]
19. Sancho J, Fersht AR. Dissection of an enzyme by protein engineering. The N and C-terminal fragments of barnase form a native-like complex with restored enzymic activity. *J Mol Biol* 1992;224:741–747. [PubMed: 1569553]
20. de Prat Gay G, Fersht AR. Generation of a family of protein fragments for structure-folding studies. 1. Folding complementation of two fragments of chymotrypsin inhibitor-2 formed by cleavage at its unique methionine residue. *Biochemistry* 1994;33:7957–7963. [PubMed: 8011658]
21. Shiba K, Schimmel P. Functional assembly of a randomly cleaved protein. *Proc Natl Acad Sci U S A* 1992;89:1880–1884. [PubMed: 1542687]
22. Johnsson N, Varshavsky A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci U S A* 1994;91:10340–10344. [PubMed: 7937952]
23. Pelletier JN, Campbell-Valois FX, Michnick SW. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. USA* 1998;95:12141–12146. [PubMed: 9770453]
24. Berggard T, Julenius K, Ogard A, Drakenberg T, Linse S. Fragment complementation studies of protein stabilization by hydrophobic core residues. *Biochemistry* 2001;40:1257–1264. [PubMed: 11170451]
25. Dutta S, Batori V, Koide A, Koide S. High-affinity fragment complementation of a fibronectin type III domain and its application to stability enhancement. *Protein Sci* 2005;14:2838–2848. [PubMed: 16199661]
26. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15:553–557. [PubMed: 9181578]
27. Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* 2000;97:10701–10705. [PubMed: 10984501]
28. Main AL, Harvey TS, Baron M, Boyd J, Campbell ID. The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* 1992;71:671–678. [PubMed: 1423622]
29. Leahy DJ, Aukhil I, Erickson HP. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* 1996;84:155–164. [PubMed: 8548820]
30. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol* 1998;284:1141–1151. [PubMed: 9837732]
31. Xu L, Aha P, Gu K, Kuimelis R, Kurz M, Lam T, Lim A, Liu H, Lohse P, Sun L, Weng S, Wagner R, Lipovsek D. Directed evolution of high-affinity antibody mimics using mRNA display. *Chem Biol* 2002;9:933–942. [PubMed: 12204693]
32. Lipovsek D, Lippow SM, Hackel BJ, Gregson MW, Cheng P, Kapila A, Wittrup KD. Evolution of an Interloop Disulfide Bond in High-Affinity Antibody Mimics Based on Fibronectin Type III Domain and Selected by Yeast Surface Display: Molecular Convergence with Single-Domain Camelid and Shark Antibodies. *J Mol Biol* 2007;368:1024–1041. [PubMed: 17382960]
33. Carr PA, Erickson HP, Palmer AG 3rd. Backbone dynamics of homologous fibronectin type III cell adhesion domains from fibronectin and tenascin. *Structure* 1997;5:949–959. [PubMed: 9261088]

34. Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci U S A* 2007;104:6632–6637. [PubMed: 17420456]
35. Koide A, Jordan MR, Horner SR, Batori V, Koide S. Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. *Biochemistry* 2001;40:10326–10333. [PubMed: 11513611]
36. Kowalski JM, Parekh RN, Mao J, Wittrup KD. Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. *J Biol Chem* 1998;273:19453–19458. [PubMed: 9677365]
37. Shusta EV, Kieke MC, Parke E, Kranz DM, Wittrup KD. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J Mol Biol* 1999;292:949–956. [PubMed: 10512694]
38. Cota E, Hamill SJ, Fowler SB, Clarke J. Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *J Mol Biol* 2000;302:713–725. [PubMed: 10986129]
39. Jee JG, Ikegami T, Hashimoto M, Kawabata T, Ikeguchi M, Watanabe T, Shirakawa M. Solution structure of the fibronectin type III domain from *Bacillus circulans* WL-12 chitinase A1. *J Biol Chem* 2002;277:1388–1397. [PubMed: 11600504]
40. Richardson JS, Getzoff ED, Richardson DC. The beta bulge: a common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci U S A* 1978;75:2574–2578. [PubMed: 275827]
41. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 1998;95:5857–5864. [PubMed: 9600884]
42. van den Beucken T, Pieters H, Steukers M, van der Vaart M, Ladner RC, Hoogenboom HR, Hufton SE. Affinity maturation of Fab antibody fragments by fluorescent-activated cell sorting of yeast-displayed libraries. *FEBS Lett* 2003;546:288–294. [PubMed: 12832056]
43. Furukawa H, Tanino T, Fukuda H, Kondo A. Development of novel yeast cell surface display system for homo-oligomeric protein by coexpression of native and anchored subunits. *Biotechnol Prog* 2006;22:994–997. [PubMed: 16889375]
44. Park S, Xu Y, Stowell XF, Gai F, Saven JG, Boder ET. Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng Des Sel* 2006;19:211–217. [PubMed: 16537642]
45. Shusta EV, Raines RT, Pluckthun A, Wittrup KD. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat Biotechnol* 1998;16:773–777. [PubMed: 9702778]
46. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286:295–299. [PubMed: 10514373]
47. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10:59–69. [PubMed: 12483203]
48. Rossi F, Charlton CA, Blau HM. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci U S A* 1997;94:8405–8410. [PubMed: 9237989]
49. Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* 2005;127:146–157. [PubMed: 15631464]
50. Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 2005;23:102–107. [PubMed: 15580262]
51. Ostermeier M, Nixon AE, Shim JH, Benkovic SJ. Combinatorial protein engineering by incremental truncation. *Proc Natl Acad Sci U S A* 1999;96:3562–3567. [PubMed: 10097076]
52. Boder ET, Wittrup KD. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol* 2000;328:430–444. [PubMed: 11075358]
53. Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* 1992;110:119–122. [PubMed: 1544568]

54. Johnston M, Davis RW. Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1984;4:1440–1448. [PubMed: 6092912]
55. Gietz RD, Woods RA. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 2002;350:87–96. [PubMed: 12073338]
56. Ohnishi S, Koide A, Koide S. Solution Conformation and Amyloid-like Fibril Formation of a Polar Peptide Derived from a β -Hairpin in the OspA Single-Layer β -Sheet. *J. Mol. Biol* 2000;301:477–489. [PubMed: 10926522]
57. Koide S, Bu Z, Risal D, Pham T-N, Nakagawa T, Tamura A, Engelman DM. Multi-step denaturation of *Borrelia burgdorferi* OspA, a protein containing a single-layer β -sheet. *Biochemistry* 1999;38:4757–4767. [PubMed: 10200164]
58. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH--a hierarchic classification of protein domain structures. *Structure* 1997;5:1093–1108. [PubMed: 9309224]

Figure 1.

(A) The amino acid sequence of FNfn10 shown according to the FN3 secondary structure. Residues in a β -strand are shown as white circles. Loop residues are shown shaded in light grey. β -Strand residues whose side chain constitutes the hydrophobic core are enclosed with a thicker line. The residues forming the conserved β bulge (V11, A12 and L19) are shown shaded in dark grey. The arrow marks the position where FNfn10 was cleaved to generate the N and C fragments. Residue numbers for the residues mutated in this study are shown. (B) Schematic drawing of the yeast surface two-hybrid system. The N fragment is expressed as a fusion to the yeast cell surface protein, Aga2p, and the C fragment is secreted. Expression of the N fragment and binding to C fragment is monitored by immunofluorescent detection of FLAG tag and of V5 tags located at the carboxyl termini of the N and C fragments, respectively. (C) A histogram of FITC fluorescence of uninduced yeast cells (filled) or yeast cells induced

for the expression of the FNfn10 N_{D7N} fragment (line) as monitored by immunofluorescent detection of the FLAG tag. (D, E) FITC fluorescence intensity (horizontal axis) and PE fluorescence intensity (vertical axis) of individual yeast cells displaying only the N_{D7N} fragment in the presence of purified C fragment (1 μM) (D) or the cosecreted C fragment using the yeast surface two-hybrid (E).

Figure 2. Binding profiles of cells displaying the N_{D7N} fragment with the coexpressed I59A C fragment. Flow cytometric bivariate plots of N_{D7N} fragment display (x-axis) and C-fragment binding (y-axis) in the absence (A) or presence (B) of 500nM purified C fragment. (C) The B/E ratio (calculated as the PE/FITC ratio for the top 1000 expressing clones) plotted against the C-fragment concentration. The curve is the best fit of the 1:1 binding model.

Figure 3.

(A) Yeast surface two-hybrid analysis of four FNfn10 fragment pairs: D7N (red), wild type (green), I59V (blue) and V50A (pink). The data are plotted in the same manner as in Figure 2. Four separate datasets are shown superimposed. (B) Correlation of the yeast surface two-hybrid readout (B/E ratio) with the conformational stability of the uncut protein. The line shows the linear regression ($R=0.99$). $\Delta\Delta G$ is defined as $\Delta G(\text{Mutant}) - \Delta G(\text{Wild Type})$.

Figure 4.

Yeast surface two-hybrid evaluation of FN3 sequence space. (A and B) Screening of single-position libraries, L19X (A) and W22X (B). The numbers in A and B show the percentage of cells in the gated window among the top 10% expression-positive population. (C) The profile of the “bulge library” in which three residues were randomized. The sorting windows for the E, G, M and P classes are shown. Note that only a small number of cells are displayed for clarity. (D) The profiles of the four populations of cells after three rounds of sorting. Data for the E (red), G (green), M (blue) and P (pink) classes are overlaid.

Figure 5. Sequences of clones from different classes of the “bulge library”. (A) class E, (B) class G, (C) class M and (D) class P. Amino acids with a β -branched residue as well as sequences with the consensus β -X- β motif where β refers to a β -branched residue are shaded in gray. The wild type sequence retrieved from class G is shown shaded in green. Clones containing additional mutations at position 7 or 9 are shaded in yellow, which were excluded from the sequence logo analysis. (E) Correlation between the yeast surface two-hybrid readout with stability change for the entire set of FNfn10 mutants used in this study ($R=0.89$). The reference clones (those

in Figure 3) are represented as closed circles and mutants are represented as follows: class E ($\circ$), G ($\diamond$), M (Δ), P (∇).

Figure 6.

(A) The amino acid distribution at positions 1 and X of the β -bulge motif found for natural FN3 sequences in the SMART database.⁴¹ The sequence pairs are shaded according to the frequency of occurrences. (B) Distribution of residue pairs from yeast surface two-hybrid screening. Residue pairs from separate bins are colored differently as indicated. Pairs occurring in two separate bins are shown as half circles with the corresponding colors for each bin. Note that each pair identified in yeast surface two-hybrid sorting already has an assigned stability value.

Table 1
Sequences of selected clones from the FNfn10 bulge library and their effects on the stability of the uncut protein. Mutated residues are shown in bold.

Clone	Sorting Class	Position			$\Delta\Delta G$ (kcal mol ⁻¹)	log(B/E ratio)
		11	12	19		
WT		V	A	L	0	1.3
E24	E	M	S	T	0.3±0.06	1.5
E3		V	A	T	0.2±0.05	1.7
E18		I	A	V	0.8±0.1	1.6
G1	G	V	V	T	0.6±0.06	1.5
G14		L	V	T	0.7±0.05	1.5
G18		R	L	V	-0.2±0.02	1.3
M3	M	S	R	Y	-0.73±0.1	0.75
M4		S	K	S	-1.4±0.1	1.0
M6		C	A	H	-1.1±0.08	1.0
P1	P	P	Q	V	-1.1±0.08	0.47
P2		D	S	S	-0.9±0.07	0.72
P3		K	T	C	-1.2±0.1	0.75