Screening Libraries To Identify Proteins with Desired Binding Activities Using a Split-GFP Reassembly Assay

Meredith E. Jackrel[†], Aitziber L. Cortajarena[‡], Tina Y. Liu[‡], and Lynne Regan[†,‡,]*

[†]Departments of Chemistry and [‡]Molecular Biophysics & Biochemistry, Yale University, New Haven, Connecticut 06520

cbi9990825210008

Wе have previously described a split-green fluorescent protein (GFP) reassembly assay by which to detect protein−protein interactions (1−3). In this assay, green fluorescent protein is dissected into two fragments, which when expressed together in *Escherichia coli* do not reassemble into a fluorescent protein. If, however, the two fragments of GFP are each individually fused to two interacting proteins, then this interaction can mediate reassembly of the GFP, with consequent cellular fluorescence (Figure 1). Here, we describe the use of this assay to screen large libraries and to identify proteins with new binding specificities.

A protein that is to be used as a framework on which to introduce novel binding activities should ideally be small, stable, of known structure, and devoid of disulfide bonds. Here we describe the use of the versatile tetratricopeptide repeat (TPR) as such a framework. The TPR is 34 amino acids long and adopts a helix-turn-helix structure (4, 5). The natural function of TPR domains is ligand binding, and the minimal peptide-binding module comprises three tandem TPR repeats (6). The 3-TPR unit possesses all of the desirable characteristics described above and therefore represents an ideal framework on which to introduce novel binding activities. There are several additional properties of TPRs that make them an attractive framework. The peptide ligand is bound in an extended conformation, suggesting that virtually any peptide sequence could be recognized (7). In addition, there is little, if any, change in the TPR backbone upon ligand binding (8). It is therefore reasonable to consider the structure of an existing TPR−peptide complex as a guideline when designing TPRs with new binding specificities. Sequence alignment and analysis

ABSTRACT Designer protein modules, which bind specifically to a desired target, have numerous potential applications. One approach to creating such proteins is to construct and screen libraries. Here we present a detailed description of using a split-GFP reassembly assay to screen libraries and identify proteins with novel binding properties. Attractive features of the split-GFP based screen are the absence of false positives and the simplicity, robustness, and ease of automation of the screen. Here, we describe both the construction of a näive protein library, and screening of the library using the split-GFP assay to identify proteins that bind specifically to chosen peptide sequences.

*Corresponding author,
lynne.regan@yale.edu.
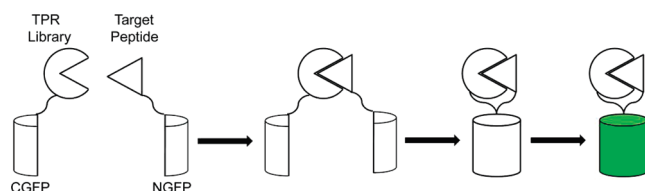
© 2010 American Chemical Society

Figure 1. Split-GFP reassembly assay as a useful tool for screening protein−protein interactions. Green fluorescent protein is dissected into two fragments. These fragments are fused to either the TPR library (CGFP) or the target sequence (NGFP). When the target peptide binds a representative TPR variant, they initiate the refolding of the GFP fragments. The two halves of GFP are then trapped and can be selected for upon fluorescence maturation.

of all TPRs suggest that they are naturally used to bind a wide variety of different ligands (*9, 10*). In addition, natural 3-TPR modules have demonstrated the ability to distinguish between two closely related sequences, indicating that they can potentially be engineered to differentiate between similar ligands with a high degree of specificity (*11−13*).

Here we describe the construction of a diverse library that is then screened for binding to a peptide of choice using a split-GFP reassembly assay coupled with fluorescence activated cell sorting (FACS). Characterization of variants selected from this library indicates that we have indeed successfully isolated TPR modules with new binding specificities.
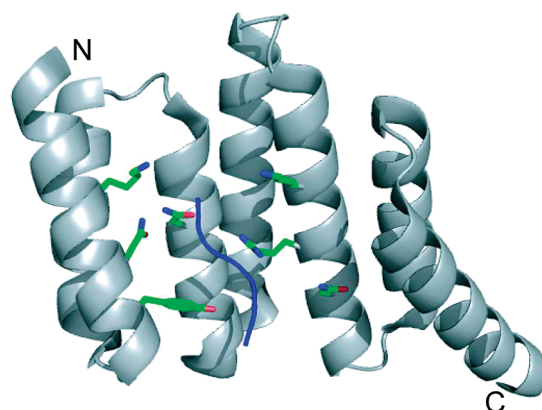

Figure 2. TPR scaffold and library design. The TPR2A structure (PDB ID 1ELR (*7*)) is shown with the residues randomized in this work (K229, N233, Y236, N264, K301, K305, and N308) depicted as green sticks. The Hsp90 peptide is shown as a blue ribbon.

RESULTS AND DISCUSSION

TPR Framework. Two well-characterized natural 3-TPR domains, TPR1 and TPR2A of the protein HOP, bind the C-terminal tails of Hsp70 and Hsp90, respectively. The C-terminal peptides of Hsp70 (PTIEEVD) and Hsp90 (MEEVD) are similar in sequence, yet both TPR1 and TPR2A discriminate against their noncognate ligands. The co-crystal structures of TPR2A and TPR1 with their peptide ligands reveal that the 3-TPR domains form a concave "cradle" in which the extended C-terminal peptides bind (Figure 2) (*7, 11, 13*). Here we use TPR2A as the framework for our redesigns.

Library Design. Our goal was to create a library that was not targeted toward any particular peptide ligand but instead highly randomized to encode a high degree of diversity. In choosing which residues to randomize, there is a balance between creating a library of maximum diversity and creating a library with a size that can be fully covered by the screening methods available.

When the sequences of all TPR repeats are aligned, certain positions in the binding pocket are "hypervariable" compared to the rest of the surface of the protein (*10*). Such hypervariability is analogous to that described for the complementarity-determining regions of antibodies by Wu and Kabat (*14*) and suggests that TPRs have the ability to bind a wide range of different ligands and that interactions with such ligands are mediated by the hypervariable residues. Considering all TPR sequences, there is a maximum of 7 hypervariable residues per repeat (*10*). Thus, a 3-TPR unit would potentially have 21 hypervariable positions.

Not all positions identified on the basis of their hypervariability when all TPR sequences are considered are necessarily involved in ligand recognition by a particular TPR module. We therefore examined the co-crystal structure of the natural 3-TPR unit, TPR2A of the protein HOP, in complex with its natural peptide ligand, the C-terminal pentapeptide of Hsp90 (MEEVD) (Figure 2) (*7*). Only about 11 of the 21 potential positions are on the ligand binding face and in proximity to the peptide ligand in the TPR2A−MEEVD complex, and therefore these were the only positions considered further.

Randomizing 6 positions with each of 19 different amino acids (we avoided the use of cysteine to prevent the potential introduction of disulfide bonds) would give a library size of 4.7×10^7, randomizing 7 positions would give a library size of 8.9×10^8, and randomizing 8 positions would give a library size of 1.6×10^{10}.

The maximum transformation efficiency using super-competent *E. coli* is approximately 10^9. We therefore made the compromise to randomize 7 positions, chosen on the basis of hypervariability and analysis of the TPR2A−MEEVD co-crystal structure (Figure 2). A library of this size can be virtually completely covered by our screening methods.

Constructing the Library. We sought to create a library with equal representation of each amino acid at each of the 7 positions. We therefore synthesized oligonucleotides using a trimer phosphoramidite mixture (corresponding to an equal mixture of 19 codons) at the positions to be varied (*15, 16*). The use of this mixture eliminates mutational bias resulting from degeneracy of the genetic code and also prevents the introduction of stop codons. Using a combination of Klenow fragment extensions and PCR amplifications, the set of six overlapping oligonucleotides were joined to create the gene library. The library was then cloned into the pMRBAD-link-CGFP vector, creating an in-frame fusion of the TPR and C-terminal fragment of GFP. We estimate that the library size is 2.7×10^8, which was determined by plating serial dilutions of the transformation mixture.

Screening the Library. Having created the library, the next step was to identify from its billions of members the variants that bind the target peptide. We employed a "split-GFP reassembly" screen, which was developed in our group (*1, 3*). When the two halves of GFP are co-expressed in *E. coli*, they do not assemble to create a fluorescent protein unless they are attached to two proteins or peptides that interact. Thus, if a target peptide is expressed on one-half of GFP and the TPR library on the other half, TPR variants that bind to the target can be identified by colony fluorescence. To screen a relatively small library, one can identify positive clones by screening for fluorescence on plates under inducing conditions. To screen larger libraries, it is more effec-
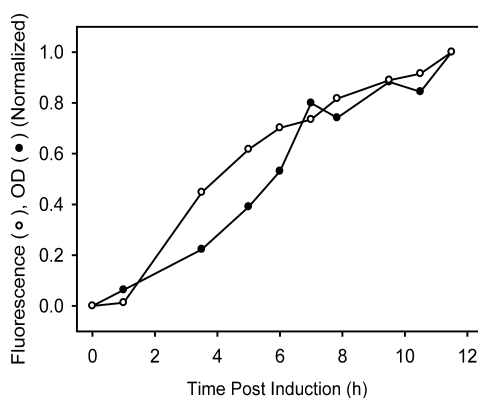

Figure 3. Fluorescence development and cell growth over time for TPR2A-Hsp90 mediated GFP reassembly in liquid culture. TPR2A-CGFP and NGFP-Hsp90 were co-transformed in BL21-Gold (DE3) and grown at 37 °C to an OD_{600} of 0.5, and then expression was induced. Growth was continued at 20 °C for 12 h. Fluorescence was measured by excitation at 397 nm, and emission values are reported for 505 nm.

tive to use automated fluorescence activated cell sorting (FACS).

Optimization of the FACS Sorting Protocol for Library Screening. FACS has seldom been used to sort *E. coli* and the split-GFP reassembly assay has not, to our knowledge, previously been used to screen a large library for protein−peptide binding interactions (*17*). We therefore optimized the sorting method before applying it to sort the library.

We first determined the optimal time after induction of GFP expression at which to sort the library (Figure 3). We monitored growth (OD 600) and fluorescence (emission at 505 nm) in liquid culture of *E. coli* BL21-Gold(DE3) cells expressing a positive control pair: TPR2A with the C-terminal peptide of Hsp90. Fluorescence was first observed 3 h following induction, and the

TABLE 1. FACS Sorting Results

Round	Control mixture[a]		c-Myc library		Dss1 library	
	Events screened	% GFP+	Events screened	% GFP+	Events screened	% GFP+
1	1×10^6	9.6	4.8×10^8	2.9	8.4×10^7	3.2
2	1×10^5	32.2	5.5×10^6	14.8	5.6×10^6	42.9
3	very few	~100	very few	>95	2.2×10^5	>95

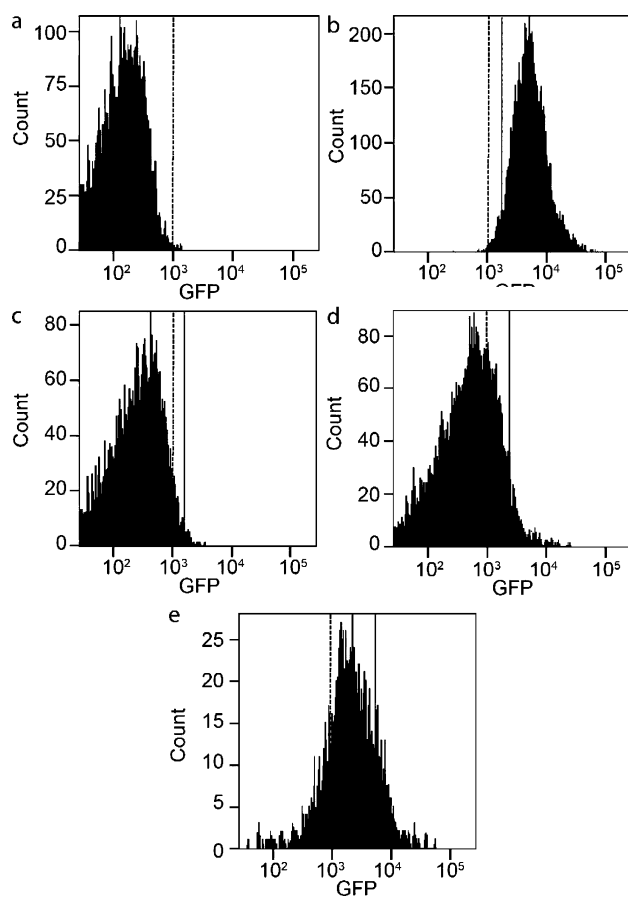[a]The control mixture contained 10% GFP + *E. coli*

Figure 4. FACS sorting of *E. coli* expressing reassembled GFP fragments. A negative control of TPR2A with a non-interacting leucine zipper peptide was sorted (a) as well as a positive control mixture of TPR2A with Hsp90 (b) to set the threshold for the GFP+ population. To sort the library, BL21-Gold(DE3) *E. coli* was co-transformed with NGFP-Dss1 and the TPR-CGFP library. Three sequential rounds of sorting are shown (c−e). The collected populations are indicated by solid lines. In the first round of sorting (c), 3.2% of the library registered as GFP+. The top 2−4% of the population was collected, outgrown, and resorted (d). By the third round of sorting (e), over 95% of the population registered as GFP+. The dashed line is drawn at the same fluorescence level in each panel and is meant as a guide to the reader to denote the separation of the positive and negative peaks. Note the shifting of the population toward a higher level of fluorescence in later rounds of sorting.

Figure 5. Specificity of the anti-c-Myc TPR variants. *E. coli* comaintaining the cognate and non-cognate plasmid pairs are plated on inducing solid media to visualize fluorescence. a) Representative anti-c-Myc TPRs (A4 and A5) as well as the TPR2A control are shown with their cognate and noncognate peptides. The anti-c-Myc TPR variants fluoresce at similar levels with c-Myc and Hsp90, which have similar sequences, but do not fluoresce with the unrelated Dss1 peptide. b) TPR24 and TPRE2 paired with c-Myc. TPR24 is a weakly fluorescent anti-c-Myc variant obtained from FACS sorting the library and was used as a template for an error prone PCR library, from which TPRE2 was selected.

fluorescence level per cell reached a maximum 6 h post-induction. In addition, we chose this time point because the cells are in the log phase of growth, which is optimal for cell sorting.

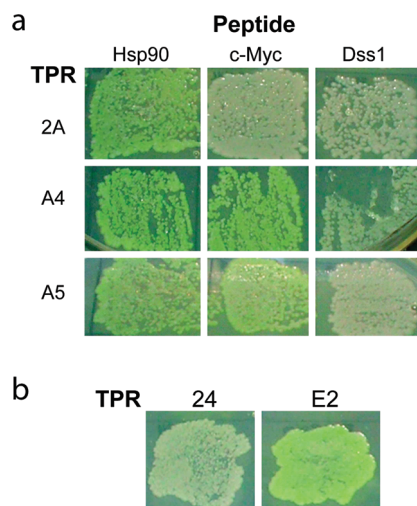To determine an optimum sorting protocol, we used trial experiments sorting control mixtures with known quantities of fluorescent and nonfluorescent cells. We determined that it was preferable to perform iterative rounds of sorting followed by outgrowth with a lower fluorescence threshold for positive cells rather than to perform fewer rounds of sorting with a more stringent fluorescence threshold.

Initially we sorted the cells at a rate of 40,000 events per second, which would allow the entire library to be sorted in just over 6 h. While ideally a population 10 times the library size would be sorted to ensure each representative is properly screened at least once, 60 h of cell sorting is impractical. However, when sorting control mixtures at 40,000 events per second, the enrichment only rose from 11.1% GFP+ to 15.6% GFP+ in the second round. Lowering the throughput rate from 40,000 to 20,000 events per second dramatically improved the enrichment of the collected population. Sort-
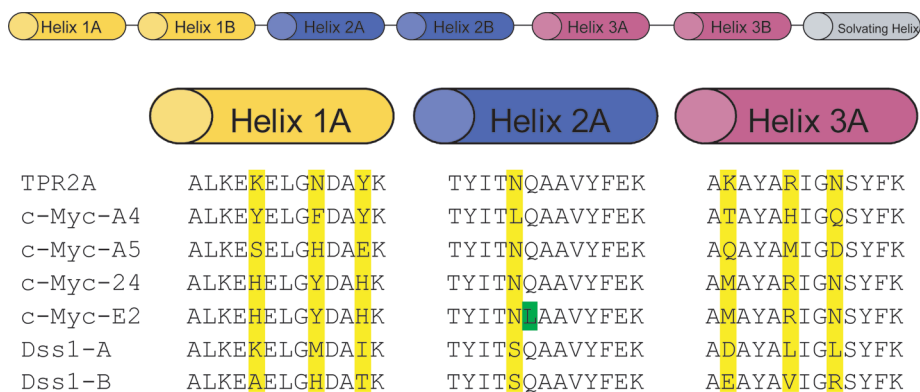
Figure 6. Sequences of selected TPR Variants. The sequences of the helices with randomized positions, relative to the full-length sequence depicted above, are shown for several anti-c-Myc and anti-Dss1 TPR variants. The randomized positions are highlighted in yellow. For TPRE2, a variant selected from error-prone PCR, the Q265L mutation is shown in green, while the R331C mutation does not fall on a binding helix.

ing the control mixture at this speed, we observed an enrichment from 9.6% GFP+ in the first round to 32.2% GFP+ in the second round. After a third round of sorting, virtually 100% of the population was GFP+ (Table 1), and this was confirmed by plating. A speed of 20,000 events per second was therefore used to sort the library.

Use of the Split-GFP Reassembly Assay in Screening Libraries. The plasmid encoding the NGFP-target peptide was first transformed into BL21-Gold(DE3), and we can routinely make these cells electrocompetent with transformation efficiencies of approximately 10^9 cfu μg^{-1} plasmid. We then transformed the library into this strain achieving efficiencies above 8×10^8.

As the first targets to screen the TPR library against, we chose the c-Myc epitope tag and the small protein Dss1. The c-Myc tag's sequence, EQKLISEEDL, is similar to the C-terminal sequence of TPR2A's cognate ligand, Hsp90 (MEEVD). Dss1 is a 70 amino acid protein that binds the tumor suppressor protein BRCA2, with no sequence similarity to the MEEVD sequence (18). We fused the c-Myc epitope tag, the full-length Dss1 sequence, the 19 residue N-terminal epitope of Dss1, and the 19 residue C-terminal epitope of Dss1 to NGFP for screening purposes.

Using our optimized protocol, we first sorted the library conservatively, with fairly permissive gating for fluorescence. Using the same gates as were set for the control, 2.9% of the library sorted for c-Myc binding was GFP+ (Table 1). We collected the top 2−4% of the population and resorted a second time, and the GFP+

population increased to 14.8%. A small fraction of these cells were sorted a third time, "analytically" to check the success of the second round of sorting. Over 95% of the cells were recorded as fluorescent. Therefore, the cells collected in the second sort were allowed to recover in liquid media and then plated under inducing conditions to select colonies for further analysis. Following streaking to single colonies to confirm the phenotype, the fluorescent clones were sequenced. The same procedures were used for sorting the library for Dss1 binding (Table 1 and Figure 4).

Specificity of the Selected TPR Variants. It is a truism that in selections you get what you select for. We selected for fluorescence, with the hope that fluorescence reflected TPR-peptide binding. One can envisage other pathways to fluorescence, so the first test we performed was to analyze the peptide binding specificity of the selected TPRs in the split-GFP reassembly assay. With this goal, we co-transformed the TPRs with peptides they were *not* selected against, *i.e.*, anti-c-Myc TPRs with the Dss1 and Hsp90 peptides (Figure 5, panel a). None of the anti-c-Myc TPRs coexpressed with Dss1 exhibited the bright fluorescence associated with cognate pairing. Interestingly, there was some fluorescence when the anti-c-Myc TPRs were co-expressed with the Hsp90 peptide, but as discussed above, there are similarities between the Hsp90 and c-Myc sequences, and therefore this observation was not completely unexpected.
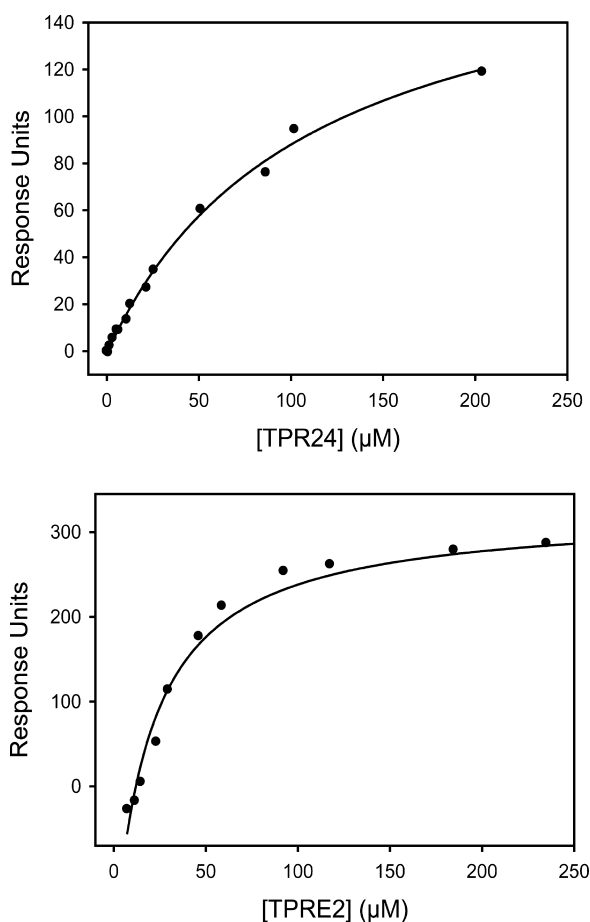
Figure 7. Binding analysis of TPR-peptide interactions. Surface plasmon resonance (SPR) binding assays were used to quantify binding. Biotinylated c-Myc peptide was immobilized on the chip, and increasing concentrations of TPR were passed over the chip. The response units at equilibrium are plotted versus their corresponding concentrations, and the data were fit to a 1:1 binding model to determine the dissociation constants (K_d). The K_d for TPR24-c-Myc = 108 μM, and that for TPRE2-c-Myc = 20.1 μM.

Results of Screening. After streaking to single colonies, the TPRs from several fluorescent clones were sequenced, and the sequences were aligned and analyzed. The simplest result of such a selection is that a clear consensus sequence emerges. This result would suggest that the screen has selected for TPRs that bind a single "epitope" on the target peptide. In the anti-Dss1 screen, using the C-terminal 19 amino acids, 96 clones were sequenced and 84 gave readable sequences, of which 55 were unique, and showed a clear consensus.

On Helix A3, at position 301, 80% of the variants had a negatively charged residue, and at positions 305 and 308, small hydrophobic residues were found in 67% and 80% of the variants, respectively. The other 4 randomized residues did not show a notable consensus. We chose 2 of these proteins for detailed analysis and demonstrated that they both bind specifically and with a K_d of approximately 10 μM to their target peptide (see the preceding Letter).

In contrast, our screening of the library against the c-Myc and full-length Dss1 peptides produced many potential hits but no clear consensus sequence (Figure 6). In the screen for c-Myc binding TPRs, 48 green colonies were sequenced, of which 31 contained unique sequences. We speculate that for both c-Myc and full-length Dss1, we may not be enriching for TPR variants against a single epitope. The selected TPR populations may contain TPRs that recognize several different epitopes; therefore no single consensus is evident.

The 31 TPR variants from the library screened against c-Myc were visually scored for fluorescence on plates, and the 12 brightest clones were selected for further characterization. These variants were subcloned from the CGFP vector into the pProExHTA expression vector as His-tagged fusions, overexpressed, and purified. Using circular dichroism (CD), we consistently observed that the CD spectra of the selected TPRs were essentially identical to that of the parent, TPR2A. In addition, all proteins were folded and exhibited cooperative unfolding transitions, with melting temperatures (T_m) ranging from 24 to 56 °C. The T_m of TPR2A is 56 °C. We observed no obvious relationship between protein stability and binding affinity. These results indicated that by altering the hypervariable residues on the binding interface we did not disrupt the structure of the proteins.

Binding Assays of TPRs. To quantify binding, we used surface plasmon resonance, immobilizing N-terminally biotinylated peptides and monitoring the binding of TPRs (Figure 7). From these data, dissociation constants for the TPR—peptide interactions could be calculated.

Of the 12 TPR variants, 2 were too prone to precipitation to make further characterization possible. The remaining 10 variants bound the peptide with affinities in the micromolar range. All of the anti-c-Myc TPRs bind to both the c-Myc and Hsp90 peptides with similar affinity. In contrast, TPR2A binds the Hsp90 peptide but shows no detectable binding to the c-Myc peptide. These results mimic the *in vivo* fluorescence results in which the

anti-c-Myc-TPR-CGFP fusions give fluorescent colonies when paired with either NGFP-c-Myc or NGFP-Hsp90. However, TPR2A-CGFP only gives fluorescent colonies when paired with NGFP-Hsp90 but not NGFP-c-Myc.

Enhancing the Affinity of the Variants. Following several rounds of FACS sorting, the affinity and specificity of the selected proteins may be sufficient for the desired applications (see the preceding Letter). However, when greater affinity is necessary, random mutagenesis or rational design can be used. We applied a random mutagenesis strategy to TPR24, a variant with modest affinity for the c-Myc peptide and relatively weak fluorescence.

We used error-prone PCR with TPR24 as the template to construct a series of small libraries with varying mutational frequencies. We aimed for an average of $2-3$ amino acid mutations per 3-TPR domain. These libraries were fused to CGFP, and the ligation mixtures were transformed into BL21-Gold(DE3)-NGFP-c-Myc. Because these libraries were relatively small, FACS sorting was not necessary. Instead the cells were plated directly on inducing media, and approximately 10^4 colonies were screened by eye under UV illumination. Most colonies appeared to fluoresce similarly to colonies containing the parent TPR24-CGFP, NGFP-c-Myc pair. One colony, TPRE2, fluoresced very brightly, far greater than TPR24 (Figure 5, panel b) and was selected for further analysis. TPRE2 has two point mutations, Q265L and R331C, relative to the parent TPR24. The Q265L mutation is directly adjacent to the fourth position we randomized, whereas the R331C mutation is distant from the binding site. We did not analyze these variants as single mutants. Binding analysis of TPRE2 revealed an improved dissociation constant of 20 μM as compared to a dissociation constant of 108 μM for TPR24 (Figure 7).

Conclusions. A major challenge of protein design is to create useful new proteins that interact specifically with desired biological targets. In this work we have shown that the split-GFP reassembly assay is a gener-

ally applicable library screening strategy to identify novel binding modules.

Alternative screening methods have also been developed. Phage display is perhaps the most common screening method, though ribosome display has also been championed by some researchers. There are advantages and disadvantages associated with any screening method. Ribosome display is performed entirely *in vitro*, allowing for large libraries to be constructed without the constraints of the transformation efficiency of *E. coli* (*19*). However, the system has many practical limitations that have hampered the expansion of its use, such as the difficulty in maintaining the physical linkage between a phenotype and its genotype (mRNA). Although phage display has been particularly successful for selections involving short peptides, for various reasons many proteins do not display in a folded, functional form on phage. In addition, fusion of a peptide or protein to the phage coat protein can compromise infectivity, leading to false positives (*20–23*).

A major advantage of the split-GFP screening method is that, in our experience, we have never observed false positives. We may well have false negatives, though this is difficult to determine. The specificity of the TPR−peptide interactions has been demonstrated to be faithfully reproduced in the assay, and weak binding is sufficient to give a positive result. In addition, a positive interaction produces an easily detectable fluorescent signal that can be analyzed by high-throughput methods using FACS to quickly screen a large library and isolate positive hits. Proteins are expressed inside the cell, so there is virtually no risk of losing the linkage between genotype and phenotype. We consider the fact that the screen gives a strong signal from a relatively weak interaction to be a positive attribute, as it increases the chances of identifying positive hits in the first round of screening. One could consider it a limitation if extremely high affinity binders are desired. However, more stringent rounds of screening or design could easily be applied to reach higher affinities, once an initial hit has been identified.

METHODS

All enzymes were purchased from New England Biolabs, except where noted. All oligonucleotide and peptide synthesis, as well as DNA sequencing, was performed by the W. M. Keck Foundation Biotechnology Resource Laboratory (Yale University). Selective media contained 100 μg mL^{-1} ampicillin and/or 35 μg mL^{-1} kanamycin. Inducing media for the split-GFP reassembly assay was supplemented with 100 μM IPTG and 0.2% arabi-

nose. To screen for split-GFP reassembly on plates, the plates were incubated for $8-16$ h at 30 °C followed by 2 days at 20 °C. Fluorescence was typically noted after 24 h. Reducing the incubation temperature is critical for the development of fluorescence.

Construction of the TPR Library. The TPR library was synthesized using a series of Klenow extensions and PCR reactions using 6 overlapping oligonucleotides (*24*) based on the sequence

of TPR2A. We selected 7 positions for randomization: K229, N233, Y236, N264, K301, K305, and N308 (Figure 2). The six oligonucleotide sequences are (1) 5′-atggctaagcaggcactgaaa-gaaxxxgagctggggxxxgatgccxxxaagaagaaagactttgacacagcctt-gaagcattacgacaaag-3′, (2) 5′-cttttcaaagtataccgctgcttgxxxggtaatgt-aagtcatgttagtggggtccagctccttggctttgtcgtaatgcttcaag-3′, (3) 5′-cagcggtatactttgaaaagggcgactacaataagtgccgggagctttg-tgagaaggccattgaagtggggagagaaaaccg-3′, (4) 5′-ttttcttctttgaagtaggaxxxcccaatxxxagcatatgcxxxggcaatct-gtcgatagtcctctcggttttctctccccacttc-3′, (5) 5′-tcctacttcaaagaa-gaaaagtacaaggatgccatccatttctataacaagtctctggc-agagcaccgaacc-3′, (6) 5′-cccttgctccttcaggattttctccgcctgttggcacttttttagcacgtctggggttcggtg-ctctgccagag-3′. Here xxx denotes positions where a single codon from a mixture of 19 trimer phosphoramidites (Cys is excluded) (Glen Research) was added during synthesis.

Pairs of overlapping oligonucleotides (1-2, 3-4, and 5-6) were joined by Klenow extensions followed by a series of two PCR amplifications. The first PCR amplification fused the Klenow products 3-4 and 5-6 together using the primers 5′-taataagacgtcccttgctccttcaggattttc-3′ and 5′-cagcggtatactttgaaaag-3′. A final PCR amplification joined the remaining Klenow fragment 1-2 with the product of the first PCR reaction, 3-6, using the primers 5′-taataaccatggctaagcaggcactgaaag-3′ and 5′-taataagacgtcccttgctccttcaggattttc-3′. The full-length library of TPRs along with the pMRBAD-link-CGFP vector were digested with NcoI and AatII and ligated using T4 DNA Ligase. At each step, the DNA was quantified to ensure the amount of DNA was maintained at least 2 orders of magnitude above 10^9 molecules to preserve the library's diversity. The ligation mixture was cleaned by phenol extraction followed by ethanol precipitation. The total ligation product (1.5 μg of DNA) was transformed in 10 batches by electroporation into 500 μL of ElectroMAX DH10B cells (Invitrogen), with transformation efficiencies greater than 10^{10} transformants μg^{-1} of DNA. Transformations were performed at 1.7 keV using 1 mm electroporation cuvettes (Biorad). The cells were recovered in 1 L of prewarmed SOC media for 1 h. To calculate library size, serial dilutions were plated and the colonies were counted. After 1 h recovery in SOC media, kanamycin (35 μg mL^{-1}) was added and the cells were grown overnight at 37 °C. The overnight culture was used for a large-scale DNA plasmid purification using a maxiprep kit (Qiagen) followed by phenol extraction and ethanol precipitation before use in the split-GFP assay.

Construction of NGFP—Peptide Fusions. The sequence corresponding to the c-Myc epitope tag, EQKLISEEDL, was prepared by Klenow extension of the oligonucleotides 5′-taataactcga-gcgaacagaaactgattagcgaagaa-3′ and 5′-taataaggatccttacagat cttcttcgctaatcagttt-3′. The insert was then digested with BamHI and XhoI and ligated into pET11a-link-NGFP.

The sequence corresponding to full length human Dss1, MSEKKQPVDLGLLEEDDEFEEFPAEDWA-GLDEDEDAHVWEDNWDDDNVEDDFSNQLRAELEKHGYKMETS, was prepared from 4 overlapping oligonucleotides. Two Klenow extensions were performed, the first using the oligonucleotides 5′-taataactcgagcatgtcagagaaaaagcagccggtagacctgggtctgtt-agaggaagacgacgagtttgaa-3′ and 5′-gacatgtgcatcttcatcttcatctaatccagtccagccttcggcagggaact-cttcttcaaactcgtcgtc-3′, and a second using the oligonucleotides 5′-gatgaagatgcacatgtctgggaggataattgggatgatgacaatgtaga-ggatgacttctctaatcagttacga-3′ and 5′-taataaggatccttactatgaagtct-ccatcttataaccatgtttctctagttcagctcgtaactgattagagaa-3′. These two fragments were joined by PCR amplification, digested with BamHI and XhoI, and ligated into pET11a-link-NGFP.

The sequence corresponding to the 19 C-terminal residues of Dss1, FSNQLRAELEKHGYKMETS, was synthesized by Klenow extension of two overlapping oligonucleotides: 5′-taataactcgagcttttctaaccagttacgtgctgaattagagaaacatgg-3′ and 5′-taataaggatccttatcaagaagtctccatcttataaccatgtttctctaattcag-3′. The product of the Klenow extension was digested with BamHI and XhoI and ligated into pET11a-link-NGFP.

The construction of NGFP-Hsp90 has been reported previously (1).

Each NGFP-peptide fusion was chemically transformed in BL21-Gold(DE3) (Stratagene), and selected on ampicillin-supplemented media. The cells were then made electrocompetent using standard protocols. Transformation efficiency was determined by transforming the kanamycin-resistant plasmid encoding TPR2A-CGFP and plating serial dilutions. Transformation efficiency typically exceeded 10^8 cfu μg^{-1}.

Monitoring Fluorescence Development over Time. The protocol of Merkel and Regan was adapted with modifications (25). In duplicate, 250 μL of BL21-Gold(DE3)-NGFP-Hsp90 was transformed by electroporation with 0.5 μg of TPR2A-CGFP in 5 batches. The entire transformation was quenched in 1 L of 2xYT media. The cells were allowed to recover for 1 h at 37 °C, with shaking at 250 rpm, prior to addition of antibiotics. Growth was continued until an OD$_{600}$ of 0.5 was reached. The cells were then cooled and induced. Growth was continued at 20 °C, and at each time point 500 μL of cells was pelleted and resuspended in 1 mL of PBS. The resuspensions were stored at −20 °C. Fluorescence was quantified by excitation at 397 nm with 5 nm slit width, and emission was monitored from 410−600 nm (values are reported for 505 nm). Cell density of the resuspension was determined by OD$_{600}$.

Fluorescence-Activated Cell Sorting. The libraries were prepared for FACS by transforming 1 μg of DNA of the TPR library in 10 batches into 500 μL of BL21-Gold(DE3)-NGFP-target peptide, and outgrowth and expression was performed using the procedures discussed above. For screening by iterative FACS, the cells were harvested 6 h following induction. The cells were washed three times in PBS and resuspended to give a FACS throughput rate of 17−24,000 events per second. All sorting was performed on a FACSAria (BD Biosciences) using doublet discrimination. The GFP+ gate was set by sorting cells co-expressing the TPR2A-CGFP, NGFP-Hsp90 pair. For sorting the library against c-Myc, the top 2−4% of the population was collected, though only 0.3% registered as GFP+ on the basis of the placement of the gates from sorting the positive control. The cells were sorted for 5.5 h into 2xYT media, and 9.2 million GFP+ cells were collected. The GFP+ cells were collected and sorted again, with a throughput rate of 1400 events per second. The collected cells were sorted a final time, and following a 30 min recovery at 37 °C with shaking, dilutions were plated under inducing conditions. Screening of the Dss1 library was conducted following the same protocols, with the exception that following each round of sorting, the cells were allowed to recover overnight in liquid media followed by induction and the next round of sorting.

Construction of Second Generation Library by Random Mutagenesis. In constructing this library, we aimed for an average of 2−3 amino acid mutations per 3-TPR domain, which we estimated would require 15 DNA mutations per kb. As this mutation rate is difficult to achieve using traditional error-prone PCR methods, we utilized the GeneMorphII Random Mutagenesis Kit, which has a higher mutational frequency and less bias than techniques that employ unbalanced dNTP concentrations. To construct the library, PCR was performed using TPR24 as a template according to the manufacturer's protocols. In order to alter the mutational frequency, the amount of starting template was varied. Three libraries were constructed with 100, 7, and 1

pg of template DNA, and a fourth library was constructed from a 1,000-fold dilution of the 7 pg mutagenized DNA. The amplification products were purified using a PCR cleanup kit (Qiagen) followed by digestion with NcoI and AatII. The digestion products were purified to ensure removal of the starting template and then ligated into pMRBAD-link-CGFP.

Each ligation mixture was transformed into BL21-Gold(DE3)-NGFP-c-Myc and plated on inducing media. Approximately 10⁴ colonies were screened, and 3 bright and 2 dim colonies were sequenced from each of the 100 and 7 pg libraries; 10 bright and 7 dim colonies were sequenced from the 1 pg library; and 8 bright and 6 dim colonies were sequenced from the library mutagenized twice.

Cellular Fluorescence Measurements. Each selected colony was restreaked 3−5 times to obtain single colonies for sequencing, ensure homogeneity, and reconfirm phenotype. To reconfirm phenotype, following each round of restreaking, single colonies were grown overnight in liquid culture in the absence of IPTG and arabinose to allow the cells to recover, as otherwise fluorescence was found to diminish in successive generations. To compare fluorescence qualitatively, 10 μL of 1:10⁴ dilutions were plated under inducing conditions in parallel. To screen the TPRs with the different NGFP-peptide fusions, the TPR-CGFP plasmids were isolated from the NGFP-target peptide plasmid. Plasmid DNA purified from E. coli co-maintaining the two plasmids was diluted 1:10³, 1 μL was transformed in DH10β, and colonies were selected on kanamycin-supplemented media. Removal of the NGFP plasmid was confirmed by ensuring there was no cellular growth after overnight incubation on ampicillin-supplemented media. The isolated plasmids were then purified and co-transformed in BL21-Gold (DE3) with the appropriate NGFP-peptide fusion.

Protein Subcloning, Expression, and Purification. The TPRs were cloned from the pMRBAD-CGFP vector by PCR amplification with the primers 5′-taataaggatccaagcaggcactgaaag-3′ and 5′-taataaaagctttcattgctccttcaggattttc-3′. The PCR products were digested with BamH I and HindIII and ligated into a modified pProEx-HTA vector (Gibco), and the sequences were verified.

The TPR proteins were expressed and purified as described elsewhere, with modifications (26). Briefly, the plasmids were transformed into BL21(DE3). Overnight cultures were used to inoculate 1 L of LB media, followed by growth at 37 °C to an OD_{600} of 0.5. Expression was induced with 0.6 mM IPTG, and the cells were grown for 5 h at 30 °C. The cells were harvested, frozen, and resuspended in lysis buffer (50 mM Tris, 300 mM NaCl, 5 mM β-mercaptoethanol, pH 8.0) with one tablet of complete EDTA-free protease inhibitor cocktail (Roche). The suspension was sonicated, and the lysate cleared by centrifugation for 1 h at 35,000g. The proteins were purified using Ni-NTA resin (Qiagen) according to manufacturer's protocols and dialyzed (Pierce) into 50 mM Tris, 150 mM NaCl, 5 mM β-mercaptoethanol, pH 8.0. Protein concentrations were determined by measuring UV absorption at 280 nm, using extinction coefficients calculated from amino acid composition (27).

Circular Dichroism Measurements. Circular dichroism spectra were acquired with 6 μM protein samples in PBS using an AVIV model 215 CD spectrophotometer (AVIV Instruments).

Far-UV CD (190−260 nm) spectra were recorded at 25 °C to confirm proper folding. Thermal denaturation curves were recorded by monitoring ellipticity at 222 nm while heating from 4 °C to 98 °C in 1 °C increments with an equilibration time of 1 min at each temperature. Melting temperatures (T_m) were estimated as the temperature at which half of the sample was unfolded.

Surface Plasmon Resonance Binding Assays. Surface plasmon resonance (SPR) measurements were performed on a Biacore 3000 system (Biacore). All measurements were performed in HBS-EP buffer (150 mM NaCl, 5 mM EDTA, 0.005% polysorbate 20, 10 mM HEPES, pH 7.4). N-Terminally biotinylated 24-mer peptides corresponding to Hsp90 (biotin-SAAVTEEMPPLEGDDDTSRMEEVD-COOH) and linker-c-Myc (biotin-SGSGSSGGSGSGSGSSEQKLISEEDL-COOH) were used for these measurements. To ensure binding was specific to c-Myc rather than the linker, a peptide corresponding to the 10 C-terminal residues of c-Myc was also synthesized and purified for competition experiments.

Using standard amine coupling, 3500 RU of NeutrAvidin (Pierce) was immobilized on a CM5 chip. 300−350 RU of biotinylated peptide was captured on the chip, followed by blocking with free biotin. An identical channel was constructed with all reagents except peptide to enable background subtraction. For the binding assays, 120 μL of purified proteins were injected over the chip at a flow rate of 40 μL min⁻¹ using the KINJECT mode followed by a 300 s dissociation period. Regeneration of the surface was achieved using three 40-μL injections of 1 M NaCl.

To calculate equilibrium dissociation constants (K_d), the average response values at equilibrium (R_{eq}) were plotted versus concentration. The curves were fit to a one-site binding model using SigmaPlot (Systat Software, Point Richmond, CA) using the equation

$$R_{eq} = \frac{R_{max}[P]}{K_d + [P]}$$

where R_{eq} is the average equilibrium response, R_{max} is the equilibrium response at saturation, K_d is the dissociation constant, and [P] is protein concentration.

REFERENCES

1. Magliery, T. J., Wilson, C. G., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., and Regan, L. (2005) Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism, *J. Am. Chem. Soc. 127*, 146−157.

2. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein, *J. Am. Chem. Soc. 122*, 5658−5659.

3. Wilson, C. G., Magliery, T. J., and Regan, L. (2004) Detecting protein-protein interactions with GFP-fragment reassembly, *Nat. Methods 1*, 255−262.

4. Lamb, J. R., Tugendreich, S., and Hieter, P. (1995) Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci. 20*, 257−259.

5. Kajander, T., Cortajarena, A. L., Main, E. R., Mochrie, S. G., and Regan, L. (2005) A new folding paradigm for repeat proteins, *J. Am. Chem. Soc. 127*, 10188−10190.

6. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix, *Trends Biochem. Sci. 28*, 655−662.

7. Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine, *Cell 101*, 199−210.

8. Cortajarena, A. L., and Regan, L. (2006) Ligand binding by TPR domains, *Protein Sci. 15*, 1193−1198.

9. Magliery, T. J., and Regan, L. (2004) Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif, *J. Mol. Biol. 343*, 731−745.

10. Magliery, T. J., and Regan, L. (2005) Sequence variation in ligand binding sites in proteins, *BMC Bioinf. 6*, 240.

11. Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F. U. (2002) Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes, *J. Biol. Chem. 277*, 19265–19275.

12. Cortajarena, A. L., Wang, J., and Regan, L. (2010) Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand, *FEBS J. 277*, 1058–1066.

13. Kajander, T., Sachs, J. N., Goldman, A., and Regan, L. (2009) Electrostatic interactions of hop TPR domains with HSP70 and HSP90—computational analysis and protein engineering, *J. Biol. Chem. 284*, 25364–25374.

14. Wu, T. T., and Kabat, E. A. (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity, *J. Exp. Med. 132*, 211–250.

15. Virnekas, B., Ge, L., Pluckthun, A., Schneider, K. C., Wellnhofer, G., and Moroney, S. E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis, *Nucleic Acids Res. 22*, 5600–7.

16. Sondek, J., and Shortle, D. (1992) A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites, *Proc. Natl. Acad. Sci. U.S.A. 89*, 3581–3585.

17. Daugherty, P. S., Iverson, B. L., and Georgiou, G. (2000) Flow cytometric screening of cell-based libraries, *J. Immunol. Methods 243*, 211–227.

18. Marston, N. J., Richards, W. J., Hughes, D., Bertwistle, D., Marshall, C. J., and Ashworth, A. (1999) Interaction between the product of the breast cancer susceptibility gene BRCA2 and DSS1, a protein functionally conserved from yeast to mammals, *Mol. Cell. Biol. 19*, 4633–4642.

19. Takahashi, T. T., Austin, R. J., and Roberts, R. W. (2003) mRNA display: ligand discovery, interaction analysis and beyond, *Trends Biochem. Sci. 28*, 159–165.

20. Atwell, S., Ultsch, M., De Vos, A. M., and Wells, J. A. (1997) Structural plasticity in a remodeled protein-protein interface, *Science 278*, 1125–1128.

21. Atwell, S., and Wells, J. A. (1999) Selection for improved subtiligases by phage display, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9497–9502.

22. Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000) High copy display of large proteins on phage for functional selections, *J. Mol. Biol. 296*, 487–495.

23. Steiner, D., Forrer, P., and Pluckthun, A. (2008) Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display, *J. Mol. Biol. 382*, 1211–1227.

24. Kajander, T., Cortajarena, A. L., and Regan, L. (2006) Consensus design as a tool for engineering repeat proteins, *Methods Mol. Biol. 340*, 151–170.

25. Merkel, J. S., and Regan, L. (2000) Modulating protein folding rates *in vivo* and *in vitro* by side-chain interactions between the parallel beta strands of green fluorescent protein, *J. Biol. Chem. 275*, 29200–29206.

26. Jackrel, M. E., Valverde, R., and Regan, L. (2009) Redesign of a protein-peptide interaction: characterization and applications, *Protein Sci. 18*, 762–774.

27. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci. 4*, 2411–2423.