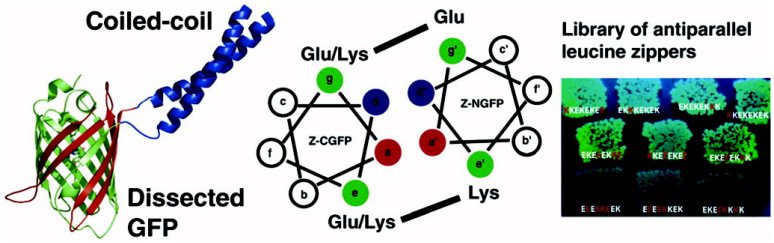
Article

Detecting Protein–Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism

Thomas J. Magliery, Christopher G. M. Wilson, Weilan Pan, Dennis Mishler, Indraneel Ghosh, Andrew D. Hamilton, and Lynne Regan

J. Am. Chem. Soc., **2005**, 127 (1), 146-157• DOI: 10.1021/ja046699g • Publication Date (Web): 02 December 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 20 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 12/02/2004

Detecting Protein−Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism

Thomas J. Magliery, Christopher G. M. Wilson, Weilan Pan, Dennis Mishler, Indraneel Ghosh,[†] Andrew D. Hamilton,[‡] and Lynne Regan*,[‡]

Contribution from the Department of Molecular Biophysics & Biochemistry, Yale University, P.O. Box 208114, New Haven, Connecticut 06520-8114

Received June 4, 2004; E-mail: lynne.regan@yale.edu

Abstract: Identification of protein binding partners is one of the key challenges of proteomics. We recently introduced a screen for detecting protein−protein interactions based on reassembly of dissected fragments of green fluorescent protein fused to interacting peptides. Here, we present a set of comaintained *Escherichia coli* plasmids for the facile subcloning of fusions to the green fluorescent protein fragments. Using a library of antiparallel leucine zippers, we have shown that the screen can detect very weak interactions ($K_D \approx 1$ mM). In vitro kinetics show that the reassembly reaction is essentially irreversible, suggesting that the screen may be useful for detecting transient interactions. Finally, we used the screen to discriminate cognate from noncognate protein−ligand interactions for tetratricopeptide repeat domains. These experiments demonstrate the general utility of the screen for larger proteins and elucidate mechanistic details to guide the further use of this screen in proteomic analysis. Additionally, this work gives insight into the positional inequivalence of stabilizing interactions in antiparallel coiled coils.

Introduction

Recent advances in genomics have rapidly expanded the number of putative proteins or open reading frames known in organisms from all three domains of life. Methods to identify the functions of those proteins, despite significant recent advancements, have not expanded apace.[1] One of the key clues to the function of an unknown protein is the identification of molecules with which it interacts. Computational and phylogenetic methods to this end have developed considerably, but these ultimately rely on databases of known interacting proteins.[2] The existing databases are relatively small and nonoverlapping, and therefore unreliable.

Binding partners for individual proteins can be identified through immunoprecipitation and related approaches such as tandem affinity purification,[3] but in vitro processing steps make it difficult to assess the in vivo significance of these results. Many known protein−protein interactions are not detected by these methods, in part because very tight binding is required. Purification methods of this type are essentially systematic, allowing examination of a single putative interaction at one time. The "classic" library-based method for identifying protein binding partners is yeast two-hybrid analysis,[4] but this method has considerable limitations: it must be done in yeast, it requires nuclear importation and function, it does not demand a direct interaction, and it can be confounded by proteins that activate transcription in the absence of a binding partner. Although two-hybrid methods allow detection of weaker interactions, they are hampered by abundant false positives. A number of assays have recently been used to circumvent some of these problems, including bacterial two-hybrid systems and functional interaction traps based on fusion to dissected fragments of dihydrofolate reductase or ubiquitin.[5−7] Also, in vitro protein microarray-based approaches have proved useful for identifying protein−protein interactions.[8,9]

Our group recently introduced a method for identifying and interrogating protein−protein interactions based on fusions to a dissected green fluorescent protein (GFP).[10] Briefly, it was shown that two fragments of GFP, split in a loop between residues 157 and 158, do not associate to give reassembled GFP when produced in *trans* in bacteria. However, by fusing strongly interacting antiparallel leucine zippers to the C- and N-termini of the N-terminal and C-terminal fragments of GFP, respectively, folding and fluorescence of the split GFP molecule are achieved (Figure 1 and Supporting Information). Thus, bacteria hosting these types of GFP fusions are only

[†] Present address: Department of Chemistry, University of Arizona, 1306 E. University Blvd., Tucson, AZ 85721-0041.

[‡] Also the Department of Chemistry, Yale University, P.O. Box 208107, New Haven, CT 06520-8107.

(1) Zhu, H.; Bilgin, M.; Snyder, M. *Annu. Rev. Biochem.* **2003**, *72*, 783.
(2) Salwinski, L.; Eisenberg, D. *Curr. Opin. Struct. Biol.* **2003**, *13*, 377.
(3) Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. *Methods* **2001**, *24*, 218.
(4) Fields, S.; Song, O. *Nature* **1989**, *340*, 245.
(5) Hays, L. B.; Chen, Y. S.; Hu, J. C. *Biotechniques* **2000**, *29*, 288.
(6) Pelletier, J. N.; Arndt, K. M.; Pluckthun, A.; Michnick, S. W. *Nat. Biotechnol.* **1999**, *17*, 683.
(7) Johnsson, N.; Varshavsky, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10340.
(8) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R.; Gerstein, M.; Snyder, M. *Science* **2001**, *293*, 2101.
(9) Newman, J. R.; Keating, A. E. *Science* **2003**, *300*, 2097.
(10) Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2000**, *122*, 5658.

10.1021/ja046699g CCC: $30.25

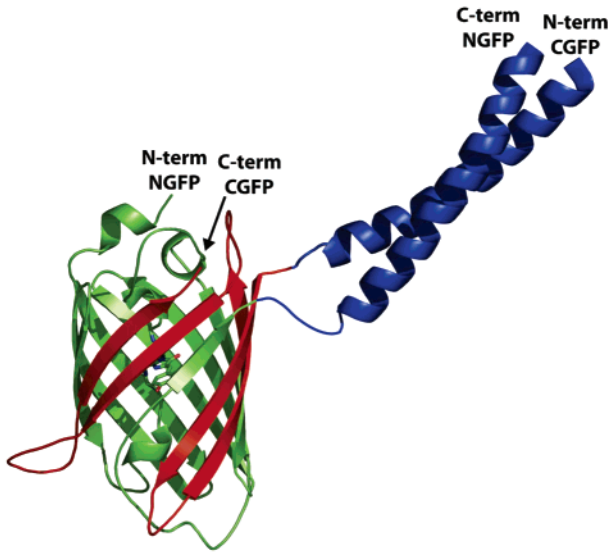


Figure 1. GFP reassembly by a protein−protein interaction. A schematic depiction of the reassembled GFP complex shows the NGFP fragment (residues 1−157, green), which contains the fluorophore, CGFP fragment (residues 158−238, red), and the antiparallel leucine zipper peptides (blue) fused at the point of dissection. The figure was rendered using PyMOL (http://www.pymol.org) from PDB file 1EMA (GFP) and a portion of the antiparallel leucine zipper from *Thermus thermophilus* seryl-tRNA synthetase (PDB file 1SER). The structures of the fusion junctions are unknown.

fluorescent if interacting proteins are fused to the GFP fragments.

GFP is an especially attractive molecule to use as an interaction trap, because it is known to express, fold, and fluoresce in virtually every cell type and subcellular structure in which it has been tested.[11] Moreover, there are likely topological limitations on the relative positions of the interacting proteins and the dissected GFP fragments, which presumably will strongly favor the detection of direct protein−protein interactions (as opposed to those that occur through complexes). Indeed, after our original report,[10] a similar approach was reported by Kerppola and colleagues.[12] Enhanced yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP) were split between residues 154 and 155, and the parallel leucine zipper domains of Jun and Fos were fused to the N-termini of the YFP or CFP fragments. The fusions were shown to reassemble in mammalian cells and properly report the subcellular localizations of bZIP transcription factors. Simultaneous observation of multiple protein−protein interactions was possible using both YFP and CFP fusion pairs.[13]

In our original implementation of this screen, both the N- and C-terminal fusions were expressed from essentially identical ampicillin-resistant pET11a plasmids under the control of the T7 promoter. Thus, only those cells that received both plasmids expressing the N- and C-terminal fusions were fluorescent upon cotransformation. Obviously, to be certain of the phenotype conferred by any particular interaction, and moreover to be able to interrogate interactions in library format, a vector system is required in which both the N- and C-terminal fusions are comaintained in cells. For the screen to be of wide applicability, it is also important to understand better the mechanism of the fusion-assisted GFP refolding reaction. What is the minimum affinity that is required for GFP reassembly? Does the affinity correspond to the degree of cellular fluorescence? Do other factors besides the protein−protein interaction affect the acquisition of fluorescence? Can transient interactions be trapped by this method? What types of polypeptides (peptides, large proteins, etc.) can be assayed with this screen?

Here, we present a set of compatible plasmids that permit the comaintenance and independent expression of N- and C-terminal GFP fragment fusions. These plasmids have been further engineered to facilitate more general use in screening applications by the introduction of linker sequences that permit rapid subcloning of "bait" and "prey" proteins of choice. Using a library of antiparallel leucine zippers that differ in their dissociation constants, we have estimated the minimum interaction strength (K_D) required for detection by this screen to be approximately 1 mM. Moreover, we have demonstrated that the protein−protein interaction is required for reassembly, and that the GFP refolding event is irreversible in vitro. Together, these data show that the method is extremely sensitive and suggest that it may be useful for detecting transient interactions. We have further used the screen to detect protein−peptide interactions in bacteria, which expands its scope beyond peptide−peptide interactions that were addressed previously.

Therefore, we provide the necessary data to guide the use of this screen for detecting protein−protein interactions of a researcher's choice. In addition, in the course of these studies, considerable insight has been gained into the interaction of antiparallel leucine zippers, whose basis of interaction is still incompletely understood.

Materials and Methods

Detailed methods of plasmid construction, construction of the leucine zipper library, analysis of the library, SPR analysis of the peptide interactions, CD analysis of the peptide interactions, construction of the tetratricopeptide repeat (TPR) and ligand GFP fragment fusions, SPR analysis of the TPR−peptide interactions, analysis of the persistence of refolded GFP, and CD analysis of refolded GFP under denaturating conditions can be found in the Supporting Information.

Screening. Compatible plasmids (e.g., pMRBAD−Z−CGFP and pET11a−Z−NGFP) were either cotransformed or sequentially transformed into BL21(DE3) *Escherichia coli* by electroporation. (Sequential transformation means that cells transformed with one plasmid were made competent for electroporation by standard methods and then transformed with the second plasmid.) Cells were screened on LB agar supplemented with 35 μg mL^{-1} kanamycin, 100 μg mL^{-1} ampicillin, 10 μM IPTG, and 0.02−0.2% arabinose. Cells were either grown for 3 days at room temperature (22 °C) or grown for 8−16 h at 30 or 37 °C followed by 1−2 days of incubation at room temperature. Fluorescence was observed under a hand-held long-wave UV lamp (365 nm).

Fluorescence Quantitation. Overnight cultures (LB/Kan/Amp) of analyte fusion pairs (e.g., Hsp90−NGFP/TPR2A−CGFP) expressed in BL21(DE3) were diluted 1:10^4, and 10 μL was plated on screening medium in duplicate as above. After 3 days at room temperature, colonies were resuspended in 4 mL of 50 mM Tris−HCl (pH 8) and 300 mM NaCl. These were diluted 50-fold in the same buffer, and the OD_{600} was measured to compare cell densities. These were lysed with lysozyme as described for the TPRs (Supporting Information). After centrifugation, the cleared lysates were diluted 10-fold in the Tris−HCl buffer and observed by fluorescence spectroscopy as described for GFP kinetics (below). Fluorescence was normalized for cell density.

Persistence of Refolded GFP. To examine both the stability of the reassembled GFP complex and the dependence of the stability on the

(11) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509.
(12) Hu, C. D.; Chinenov, Y.; Kerppola, T. K. *Mol. Cell* **2002**, *9*, 789.
(13) Hu, C. D.; Kerppola, T. K. *Nat. Biotechnol.* **2003**, *21*, 539.

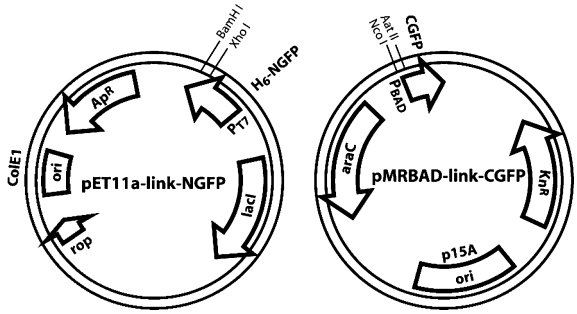


Figure 2. Maps of compatible fusion expression plasmids. The fusions used in our original report were both expressed from pET11a plasmids, similarly to pET11a−link−NGFP. Restriction sites for in-frame fusion cloning are indicated. The "link" vectors have a short DNA linker between the restriction sites. In the plasmids pET11a−Z−NGFP and pMRBAD−Z−GFP, the linker sequence has been replaced by DNA encoding a designed antiparallel leucine zipper (see Figure 4). The figure was created in part with pDRAW32 (http://www.acaclone.com).

leucine zipper interaction, a cleavable construct was created. Starting with the Z−CGFP fusion, the amino acid sequence GGENLYFQG was inserted between the leucine zipper peptide and the TSGGSG linker to the CGFP fragment. TEV protease cleaves the Gln−Gly peptide bond in that sequence.

Two 50 μL fractions of each of the purified Z−NGFP/Z−CGFP and Z−NGFP/Ztev−CGFP complexes were incubated overnight at room temperature, one each with addition of 10 U of rTEV protease (Invitrogen). Then, 30 μL of each sample was acetone precipitated for SDS−PAGE, and 10 μL was diluted with 90 μL of GFP buffer for fluorescence spectroscopy ($\lambda_{excitation}$ = 468 nm, $\lambda_{emission}$ = 505 nm).

Kinetics of Unfolding for Refolded GFP. Urea denaturation was carried out in 50 mM Tris−HCl (pH 8), 300 mM NaCl, and 0−7.2 M urea, as appropriate. Fluorescence was observed with $\lambda_{excitation}$ = 468 nm and $\lambda_{emission}$ = 505 nm. Data were collected in three time regimes: 0, 1.5, 3, 4.5, 5, 5.5, 6, 6.5, and 7.2 M urea every 12 h for 3 days, 0, 3, 3.5, 4, and 4.5 M urea every 24 h for 7 days, and 0, 5.75, 6, 6.25, and 6.5 M urea every 4 h for 2 days. The decay of fluorescence due to unfolding was fit to a single exponential, as described in the Results.

Results

Compatible Vectors for Comaintenance of GFP Fusions. To improve the utility of the GFP-based fragment complementation assay, we engineered two compatible vectors that can be comaintained in *E. coli* (Figure 2). The NGFP fusion vector, pET11a−link−NGFP, is based on the original GFP fragment expression plasmid used in this laboratory (pET11a−NZGFP), which has a ColE1 origin, confers resistance to ampicillin, and contains a T7 promoter and a terminator. The vector is different from the previous pET11a−NZGFP vector in two significant ways: (1) a hexahistidine tag was appended to the N-terminus of the NGFP fragment and (2) the DNA coding for the leucine zipper (at the C-terminus of the NGFP fragment), which was used to drive the interaction of the GFP fragments in our original implementation of the screen, was replaced with a linker region permitting in-frame subcloning of the gene coding for leucine zipper peptide or other proteins between *Xho*I and *Bam*HI sites. The linker contains a unique *Xma*I site to aid subcloning (i.e., ligation reactions can be digested to remove background after ligation). Introduction of unique restriction sites required extension of the linker region between the NGFP and the fused peptide (or protein) from GGSGSG to GGSGSGSS. The vector pET11a−Z−NGFP expresses NGFP with the leucine zipper peptide fused to its C-terminus.

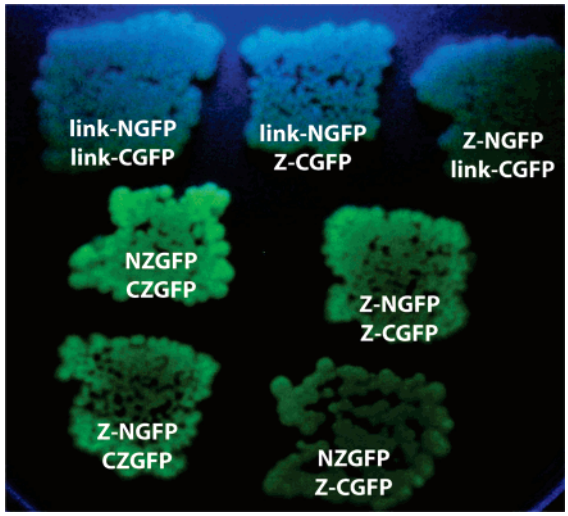


Figure 3. Cellular fluorescence upon GFP reassembly. (Top row) When either or both of the antiparallel leucine zipper peptides are replaced with short linker peptides, no GFP reassembly is observed. (Middle row) GFP reassembly occurs when the interacting peptides are fused to the GFP fragments, with the original fusion architecture (left) or the one reported here with longer linker sequences between the GFP fragments and zipper peptides. (Bottom row) Despite the differences in linker lengths, the fusions are interoperable.

The vector pMRBAD−link−CGFP is based on the p15A, kanamycin-resistance-conferring plasmid pMR101.[14] The T7lac promoter was replaced with the arabinose promoter, including the divergently transcribed *araC* regulator gene, from pBAD/HISa. Again, the DNA coding for the leucine zipper (at the N-terminus of CGFP) was replaced with a linker region for in-frame subcloning, this time between *Nco*I and *Aat*II sites, and a unique *Sph*I site in the linker is intended to aid subcloning. This again required extension of the linker between the fused peptide or protein and the CGFP fragment to TSGGSG from GGSG. The vector pMRBAD−Z−CGFP expresses CGFP with the leucine zipper peptide fused to its N-terminus after an initial Met-Ala-Ser.

Cotransformation or sequential transformation of BL21(DE3) *E. coli* with pET11a−Z−NGFP and pMRBAD−Z−CGFP with mild induction of the *lac*-controlled T7 polymerase (10 μM IPTG) and strong induction of the arabinose promoter (0.02−0.2% arabinose) on LB agar after 16 h at 37 °C and 24 h at room temperature results in cellular fluorescence in all colonies. The overnight cell-growth phase may be carried out at 30 °C, or the cells may be allowed to grow for 3 days at room temperature. If the Z−NGFP or Z−CGFP plasmids are replaced with link−NGFP or link−CGFP plasmids, respectively, then *no* cellular fluorescence is observed, proving that the antiparallel leucine zipper interaction is required for GFP reassembly (Figure 3).

We originally cloned the CZGFP fusion into pMR101,[14] which was comaintained with pET11a−NZGFP. Alone, pMR101−CZGFP expressed CZGFP fusion at a high level in BL21(DE3) *E. coli*, but pET11a−NZGFP and pMR101−CZGFP together did not result in fluorescent cells, probably due to the much larger amount of NZGFP fusion that was produced under these conditions. We believe that this overexpression of NZGFP with respect to CZGFP is due to the fact that both fusions were expressed from T7 promoters, but the

(14) Munson, M.; Predki, P. F.; Regan, L. *Gene* **1994**, *144*, 59.

copy number of the pMR101 plasmid is much lower than that of pET11a, resulting in much less template. Replacement of the T7lac promoter in pMR101−CZGFP with the arabinose promoter from pBAD/HISa created the plasmid pMRBAD−CZGFP, which results in cellular fluorescence in combination with pET11a−NZGFP. Since the CZGFP is expressed from a distinct promoter and induced separately, the template-level effect is ameliorated (although the *araBAD* promoter is weaker than the T7 promoter).

We also engineered vector pMRBAD−Z−CGFPtag, which expresses the Z−CGFP fusion with a biotinylation signal followed by a hexahistidine tag at its C-terminus. However, neither pET11a−NZGFP nor pET11a−Z−NGFP resulted in cellular fluorescence in combination with this plasmid (data not shown). Apparently, these fused tags do not permit GFP reassembly, perhaps due to the resulting proximity of the N-terminus of NGFP to the fusion in the folded complex (that is, the N- and C-termini of full-length GFP are close in space), or perhaps due to additional insolubility or misfolding caused by the large unstructured region. However, some variation in the linker lengths between the GFP fragments and the leucine zipper fusions is tolerated. Either pET11a−NZGFP (GGSGSG) or pET11a−Z−NGFP (GGSGSGSS) can be cotransformed with either pMRBAD−CZGFP (GGSG) or pMRBAD−Z−CGFP (TSGGSG), resulting in cellular fluorescence, where the sequences in parentheses represent the linkers between the zipper peptide and the GFP fragment (Figure 3).

Thus, the plasmids pET11a−link−NGFP and pMRBAD−link−CGFP can be used to rapidly subclone fusions to the NGFP and CGFP fragments, respectively, for facile interrogation of the interaction between the selected fusion proteins. Since the plasmids are comaintained and the fusions are independently expressed, the fluorescence phenotype of the cotransformed cells is homogeneous, permitting robust examination of libraries of fusions.

Antiparallel Leucine Zipper Libraries for Determining Interaction Requirements. We have established that the engineered antiparallel leucine zipper interaction is sufficient for GFP fragment reassembly, and that no fluorescence is observed in the absence of either or both of the zippers. We sought to determine how weak an interaction is required to permit GFP fragment reassembly. We opted for a combinatorial approach to this problem, wherein a library of leucine zippers fused to CGFP was screened against a constant Z−NGFP fusion. (Please note that throughout the text we refer to the fusions as X−YGFP, where X is the identity of the fused protein, and Y is the GFP fragment, N or C, to which the protein is fused.) The Z(EK)−CGFP library of fusions results in different Z−Z(EK) peptide−peptide interactions in combination with a constant Z−NGFP.

In general, both parallel and antiparallel coiled coils associate because of the interaction of hydrophobic residues that are buried at the peptide−peptide interface, as well as charge−charge interactions between "edge" positions (Figure 4).[15−17] It is not rigorously known what controls the orientation of the peptides with respect to each other, but antiparallelism can be favored by (1) judicious placement of charge−charge interactions that

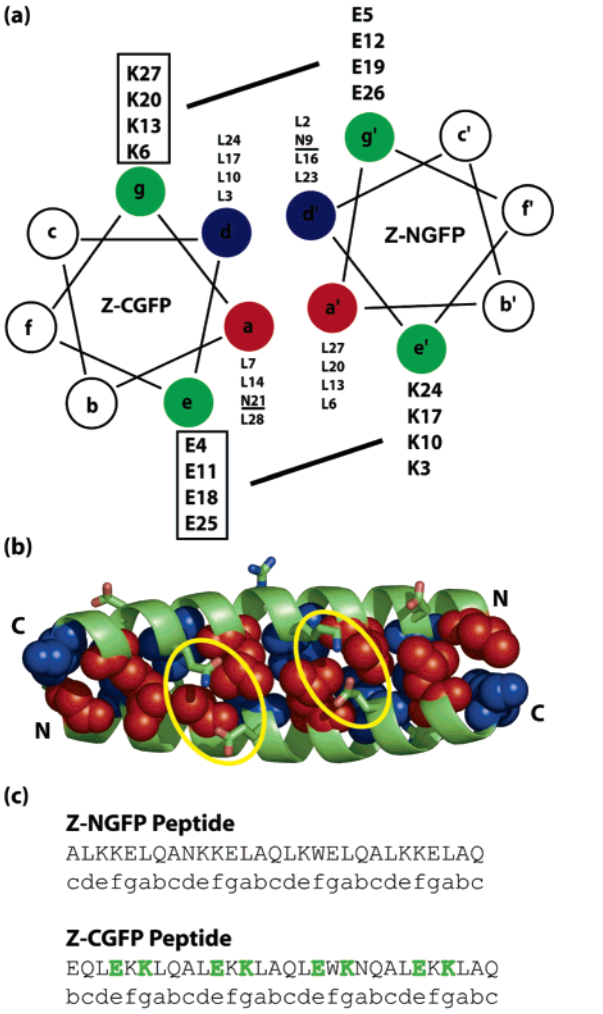


Figure 4. Antiparallel leucine zipper library. (a) A helical wheel diagram of the antiparallel leucine zipper interaction. Positions a and a′ (red) and d and d′ (blue) form the hydrophobic core, except for the N21−N9′ (underlined) interaction, which forms a hydrogen bond in the antiparallel orientation. The edge positions (e, e′, g, and g′, green) are charge-complementary in the control Z peptides. In the Z(EK) library, the e and g positions of the Z−CGFP peptide (boxed) were randomized between Glu and Lys. (b) A portion of the antiparallel leucine zipper from *T. thermophilus* SerRS (PDB file 1SER). Core positions are rendered as spheres, and edge position interactions are shown as sticks (circled in yellow). This panel was rendered with PyMOL. (c) Sequences of the zipper peptides with canonical helical positions noted below. The randomized positions are shown in green.

would result in charge−charge clashes in the parallel orientation and (2) the introduction of a polar residue into a canonically hydrophobic position, such that an interpeptide hydrogen bond is only possible in the antiparallel orientation.[18−20] Because the topology of our GFP fragment fusions likely requires antiparallelism for reassembly, we chose to hold the hydrophobic residues and buried polar residue constant, but to vary the charged Glu and Lys residues between Glu and Lys at random in the Z(EK)−CGFP peptide library. The codon RAA, where R represents an equimolar mix of both purine bases, codes for a 1:1 mix of Glu and Lys. This library was constructed using RAA codons for the eight charged edge positions in the Z peptide fused to CGFP, and there are therefore $2^8 = 256$ unique

(15) Alber, T. *Curr. Opin. Genet. Dev.* **1992**, *2*, 205.
(16) Adamson, J. G.; Zhou, N. E.; Hodges, R. S. *Curr. Opin. Biotechnol.* **1993**, *4*, 428.
(17) Lupas, A. *Trends Biochem. Sci.* **1996**, *21*, 375.
(18) Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603.
(19) Oakley, M. G.; Hollenbeck, J. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 450.
(20) McClain, D. L.; Binfet, J. P.; Oakley, M. G. *J. Mol. Biol.* **2001**, *313*, 371.

peptides in this library with between zero and eight charge−charge mismatches (mutations) from the Z peptide fused to NGFP.

The library was constructed by extension and amplification of synthetic oligonucleotides prepared with an equimolar mixture of purine base phosphoramidites at the requisite positions. Transformed cells were plated on LB agar supplemented with IPTG and arabinose and screened for fluorescence visually using a long-wave UV hand-held lamp. About 15% of the colonies were fluorescent. Fluorescent colonies (48) and dim colonies (48) were restreaked to confirm their phenotypes, and DNA was amplified from whole cells for sequencing.

Two of the sequences (one positive and one negative) appeared to result from mixed templates, and were discarded. Twenty of the negatives had an insertion, a deletion, or a large deleted region and were therefore also discarded. The remaining 47 positives and 27 negatives were analyzed more closely (see the Supporting Information for all the sequences). Figure 5 (top) shows the fraction of charge−charge mismatches seen in the positives (fluorescent colonies) and negatives (dim colonies), as compared to the distribution that would be expected if the colonies were selected at random and each randomized position had a 50% chance of being Glu (or Lys). At random, 4 ± 1 (standard deviation) charge mismatches would be expected. The negatives, which account for 85% of the total pool of clones and thus should be similar to the random distribution, had 4.0 ± 0.9 mutations; the positives had 1.8 ± 0.9 mutations. Specifically, all positives had three or fewer mutations, and all negatives had three or more mutations.

The parent CGFP zipper peptide with no edge position mismatches has a canonical net charge of zero; the NGFP zipper peptide to which the CGFP peptides in the library must bind has a canonical net charge of +2. The canonical net charges of the peptides that were positive by the screen were between −2 and +2 in 96% of the positive clones, whereas 44% of the negatives clones had a higher net charge load than this. It is worth noting that this statistic is affected by the fact that the possible net charges are quantized (0 mutations, net charge 0; 1 mutation, ±2; 2 mutations, 0 or ±4; etc.).

Mutations were not equally frequent at the eight positions (Figure 5, bottom). In fact, positions K6, E18, and K20 were essentially immutable (although a single positive had a K6E mutation, and it was the only mutation in that clone). These positions were mutated about half the time in the negatives, indicating a strong bias against mutations at these positions in positives. On the other hand, positions K13, E25, and K27 were mutated about half the time in the positives, and positions E4 and E11 were mutated about 15% of the time. However, among the negatives (and presumably the pool as a whole, therefore), position K13 is mutated almost 90% of the time, which makes it difficult to classify position K13 as being moderately frequently mutated (like E4 and E11) or frequently mutated (like E25 and K27) among positives. If the real frequency of mutation in all clones in the pool were 50% for a given position, then an observed frequency of 30−70% would be expected in 27 randomly selected clones 98+% of the time. This suggests that the naive pool is biased toward mutation at position 4 and against mutation at positions 3 and 8, presumably due to imperfect mixing of phosphoramidites during synthesis of the library template oligonucleotides.

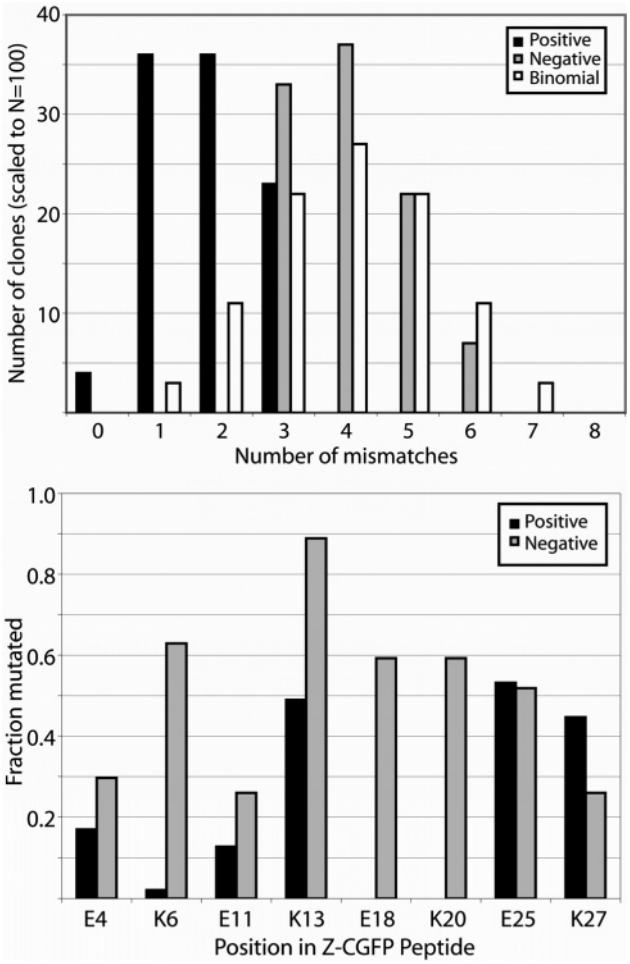


Figure 5. Distribution of mutations in the CZ(EK) library. (Top) Histogram of the number of mutations (charge mismatches) observed in positive and negative clones, as compared to the binomial distribution expected in the naive library. For clarity, the number of clones (47 positives, 27 negatives) was scaled to 100 for each group. (Bottom) Fraction of sites mutated by position among positive and negative clones. Since negative clones account for more than 80+% of the library, these should approximate the naive (unselected) library. A 50% mutation rate is expected at each position in the naive library. Positive clones have significantly fewer mutations at positions K6, E18, and K20, while positions E25 and K27 have nearly the same mutation rate among positives and negatives. One position, K13, has a significant number of mutations in the negatives, suggesting some library bias at this position, probably arising during oligonucleotide synthesis.

Any clone with fewer than three charge−charge violations binds sufficiently well to be scored positive in the screen, and any clone with more than three violations is scored negative. However, both positives and negatives had three mutations, but none of the positives with three mutations had mutations at any of the three "immutable" positions K6, E18, and K20. Taken together, the data suggest that mutations at positions K6, E18, and K20 are especially disruptive of the zipper peptide interaction.

Biophysical Analysis of Leucine Zipper Interactions. Ten clones from the CZ(EK) library were chosen for further analysis (Supporting Information). The corresponding 10 peptides (from 7 positives and 3 negatives), plus the "wild-type" peptide and a negative control with all 8 possible mutations, were synthesized, and their binding to the NZ peptide was assessed by SPR (see the Supporting Information for an example data set). We found that use of thiol-coupled NZ peptide and supplementation of the binding buffer with CM-dextran resulted in considerably

Table 1. Biophysical Analysis of Leucine Zipper Interactions

CZ(EK) Peptide[a]	Mutations[b]	Phenotype[c]	K_D (μM)[e]	CD K_D est.[h]
EQL**E**K**K**LQAL**E**K**K**LAQL**E**W**K**NQAL**E**K**K**LAQ	none	Pos	20	20
EQL**K**K**K**LQAL**E**K**K**LAQL**E**W**K**NQAL**E**K**K**LAQ	4	Pos	23	3
EQL**E**K**K**LQAL**E**K**K**LAQL**E**W**K**NQAL**K**K**K**LAQ	25	Pos	120	31
EQL**E**K**K**LQAL**K**K**K**LAQL**E**W**K**NQAL**E**K**K**LAQ	11	Pos	Tight[f]	50
EQL**K**K**K**LQAL**E**K**K**LAQL**E**W**K**NQAL**E**K**E**LAQ	4/27	Pos	Tight[f]	8
EQL**E**K**K**LQAL**E**K**E**LAQL**E**W**K**NQAL**K**K**K**LAQ	13/25	Pos	Weak[g]	600
EQL**K**K**K**LQAL**E**K**E**LAQL**E**W**K**NQAL**E**K**E**LAQ	4/13/27	Pos	Medium[g]	45
EQL**E**K**K**LQAL**E**K**E**LAQL**E**W**K**NQAL**K**K**E**LAQ	13/25/27	Pos	Very Weak[g]	1000
EQL**E**K**K**LQAL**E**K**E**LAQL**K**W**K**NQAL**K**K**K**LAQ	13/18/25	Neg	NB	2000
EQL**E**K**E**LQAL**E**K**E**LAQL**K**W**K**NQAL**E**K**K**LAQ	6/13/18	Neg	NB	3000
EQL**E**K**E**LQAL**E**K**E**LAQL**K**W**E**NQAL**E**K**K**LAQ	6/13/18/20	Neg	NB	2000
EQL**K**K**E**LQAL**K**K**E**LAQL**K**W**E**NQAL**K**K**E**LAQ	all eight	N/A[d]	NB	3000

[a] Sequence of CZ(EK) peptide tested for binding to Cys-NZ peptide (see Supporting Information). The E/K randomized positions are in bold, and underlined positions have mutations relative to the CZ peptide (first row). [b] Positions of E → K or K → E mutations in the given peptide relative to the CZ peptide, numbered from the N-termini. [c] Screen phenotype (Pos = fluorescent, Neg = nonfluorescent). [d] This peptide was synthesized as a hypothetical negative control, but it was not tested in the screen or isolated from the library. [e] Dissociation constants were determined from two independent trials of duplicate points. The error between trials was less than ±8% in each case. NB indicates that no binding was detected. [f] Although binding is observed, the data do not fit a monophasic binding model. [g] No saturating binding is detected within the limits of the experiment, but the binding curve can be resolved from the negatives. [h] Estimate of the apparent dissociation constant (μM) from the difference in ellipticity at 222 nm between the individual NZ and CZ peptides and a mixture of the peptides. The values are scaled using the SPR measurement of the wild-type CZ peptide.

less nonspecific binding to the chip surface than with neutravidin/biotin immobilization. However, due to limitations of peptide solubility and aggregation at higher concentrations, saturating binding could not be achieved for seven of the peptides (Table 1). Also, two of the peptides appeared to bind tightly, but their binding curves could not be fit to a monophasic binding model.

To corroborate these SPR data, we also examined the difference in helicity between the NZ and each CZ peptide alone and mixed in solution (see the Supporting Information for an example data set). Assuming a 1:1 binding model and a constant maximum ellipticity for each fully associated NZ/CZ pair, it is possible to estimate the apparent dissociation constants on the basis of the fraction of peptide bound. This fraction was determined from the helicity difference scaled such that the wild-type K_D is the same as determined from SPR (see the Supporting Information, Materials & Methods). These values agree quite well with the SPR values (Table 1).

The positions of the charge−charge mismatches had a profound effect on the affinities of the peptides for the NZ peptide against which they were screened. The three peptides with a single mutation bound to NZ peptide with dissociation constants that varied over about an order of magnitude. One of the peptides with two mutations (at positions 4 and 27) bound about the same as the wild type, but the other (positions 13/25) bound much more weakly (30-fold). We estimate that the binding threshold for a positive phenotype in the screen is about 1 mM. The peptides from negative clones (with three or four mutations) and the negative control showed no detectable binding by SPR and almost no change in helicity upon mixing with the NZ peptide.

General Applicability of the Screen: Protein−Peptide Interactions. We also explored to what degree the screen would be useful for studying the interactions of larger proteins, in addition to the leucine zipper peptides with which we calibrated the system. The TPR is a 34 amino acid helix−turn−helix motif, most commonly found in tandem groups of three per domain, which is thought to mediate a variety of protein−protein interactions.[21] Hsp organizing protein (HOP) contains three independent 3-TPR domains: TPR1, TPR2A, and TPR2B. TPR1 binds to the C-terminus of human chaperone Hsc70 (Hsp70), and TPR2A binds to the C-terminus of Hsp90 (Figure 6).[22,23] There is no known ligand for TPR2B.

We fused TPR1, TPR2A, and TPR2B to CGFP (using pMRBAD−link−CGFP), and peptides from the C-termini of Hsc70 and Hsp90 to NGFP (using pET11a−link−NGFP). We then challenged each of the TPR domains with each of the C-terminal ligands, and also with a leucine zipper peptide as a negative control, using the fusion reassembly screen. To confirm the visual phenotypes, we also harvested cells from the screening agar plates and examined the fluorescence in cleared lysates for equal numbers of cells (Figure 6). Cells expressing the TPR1−CGFP fusion were 10-fold more fluorescent with coexpression of the cognate NGFP−Hsc70 peptide fusion than with the NGFP−Hsp90 peptide fusion, and much more fluorescent (35-fold) than with the leucine zipper peptide fusion. Likewise, coexpression of TPR2A−CGFP resulted in 3-fold more fluorescence with the cognate NGFP−Hsp90 fusion than with the NGFP−Hsc70 peptide, and background-level fluorescence was observed with the zipper peptide. Surprisingly, TPR2B−CGFP fusion resulted in cellular fluorescence with both NGFP−Hsc70 and NGFP−Hsp90 fusions, much greater than with the control zipper peptide.

To compare the results observed with the screen to the dissociation constants of the TPR−peptide interactions, we analyzed the TPR−peptide interactions in vitro using SPR (Table 2). The in vivo screen results correlated well with the interaction strengths for TPR1 and TPR2A with Hsc70 and Hsp90. The brightest colonies corresponded to the strongest interaction (TPR2A−Hsp90), medium brightness to interactions about 10-fold weaker than the TPR2A−Hsp90 interaction (TPR1−Hsc70 and TPR2A−Hsc70), and the least bright to the at least 300-fold weaker TPR1−Hsp90 interaction. However, the TPR2B interactions with both the Hsc70 and Hsp90 peptides were at least 10-fold weaker than the above interactions, leading to "medium" brightness, but both resulted in significant cellular

(21) D'Andrea, L.; Regan, L. *Trends Biochem. Sci.* **2003**, *28*, 655.

(22) Scheufler, C.; Brinker, A.; Bourenkov, G.; Pegoraro, S.; Moroder, L.; Bartunik, H.; Hartl, F. U.; Moarefi, I. *Cell* **2000**, *101*, 199.

(23) Brinker, A.; Scheufler, C.; Von Der Mulbe, F.; Fleckenstein, B.; Herrmann, C.; Jung, G.; Moarefi, I.; Hartl, F. U. *J. Biol. Chem.* **2002**, *277*, 19265.

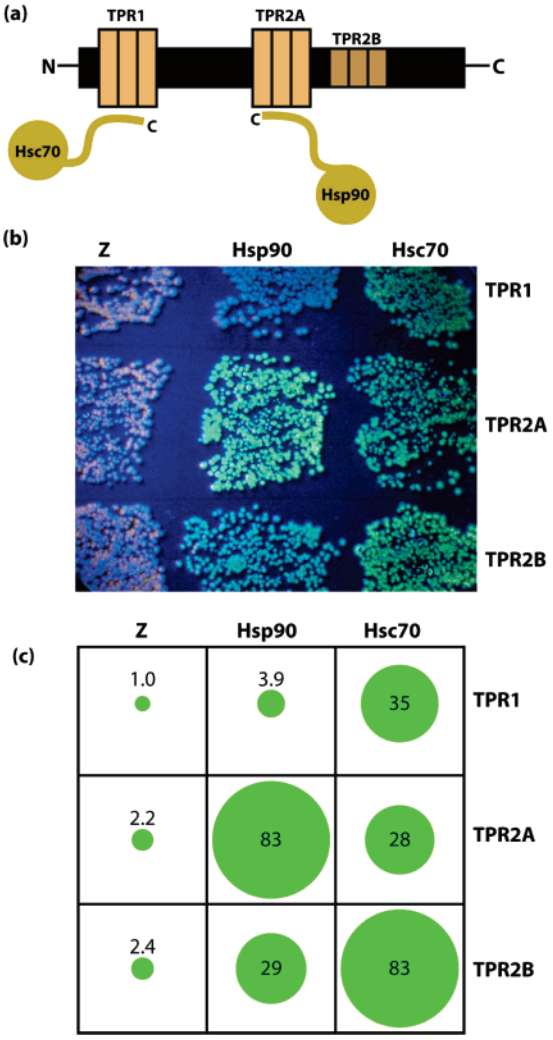


Figure 6. TPR−peptide interactions. (a) Schematic of known interactions between the TPR domains of HOP and the Hsc70 and Hsp90 ligands. TPR2B has no known ligand. (b) Interaction of TPR1, TPR2A, and TPR2B on NGFP with the Z peptide or C-terminal peptides from Hsc70 or Hsp90 in CGFP. (c) The amount of fluorescence was quantified from lysates of equal numbers of cells grown under the same conditions as in (b). The normalized relative fluorescence is shown, where the areas of the green circles are proportional to the fluorescence values. None of the cells with Z−CGFP are fluorescent. TPR1−NGFP/Hsc70−CGFP is 9-fold brighter than TPR1−NGFP/Hsp90−CGFP, and TPR2A−NGFP/Hsp90−CGFP is 3-fold brighter than TPR2A−CGFP/Hsc70−NGFP.

Table 2. SPR Analysis of TPR−Peptide Interactions

domain[a]	Hsp90[b]	Hsc70[b]
TPR1	NSB[c]	12
TPR2A	1.3	18
TPR2B	300	500

[a] Indicates the TPR domain derived from human HOP. [b] Dissociation constants (K_D) (μM) for TPR domains from 24 amino acid peptides derived from the C-termini of Hsc70 and Hsp90. [c] No saturating binding within the limits of the experiment.

fluorescence. This suggests that the screen is useful for discriminating strongly bound ligands from weakly bound ones for a given protein, but that the absolute amount of fluorescence does not directly relate to the protein−ligand dissociation constant for an arbitrary protein−ligand pair.

Is the Peptide−Peptide Interaction Required To Maintain the Association of the Refolded GFP? Although a peptide−peptide or protein−peptide interaction is clearly required for the reassembly of the GFP, we wondered whether the peptide−peptide interaction plays a role in maintaining the reassembled complex. At one extreme, one could imagine that the peptide−peptide complex nucleates the refolding process but is not required to hold the complex together; at the other extreme, maintenance of a stable complex might require the peptides to remain constantly bound. The mechanism has important implications for the generality of the method, because, if the latter is true, the screen is likely to be much more sensitive to the nature of the fusion (linker length, position of fusion, etc.).

To investigate this issue, we constructed a Z−CGFP fusion with a recognition site for TEV protease inserted between the peptide and CGFP fragment. This Ztev−CGFP fusion also tests the dependence of reassembly on linker length (since the linker is effectively lengthened by nine amino acids). Cells expressing both Z−NGFP/Z−CGFP and Z−NGFP/Ztev−CGFP fusions fluoresce under similar conditions of induction and growth on LB agar. Since the Z−NGFP fusion has a hexahistidine tag at its N-terminus, we reasoned that, if these reassembled complexes are stable, we would be able to purify them using Ni−NTA agarose, and indeed this was possible. In mild buffer, there is no significant loss of fluorescence of the purified complexes over the course of months.

It should be noted, in contrast, that overexpression of either the Z−NGFP or Z−CGFP fusion alone results in totally insoluble protein that is found in the pellet after centrifugation of the lysate (not shown). Even upon coexpression, at short times after induction, both Z−NGFP and Z−CGFP are found exclusively in the insoluble fraction. Only after growth at room temperature for several days is the complex found in the soluble fraction, which is roughly the time scale over which cellular fluorescence evolves. Indeed, purification of the individual fusions from cells reported by this laboratory and the Kerppola laboratory required the extraction of protein from the insoluble fraction using strong denaturant (urea or guanidium−HCl).[10,12]

Each of the Z−NGFP/Z−CGFP and Z−NGFP/Ztev−CGFP complexes was subjected to overnight cleavage at room temperature with TEV protease, and the fluorescences before and after the cleavage were compared (Figure 7). SDS−PAGE analysis indicates no changes in the apparent masses of the Z−NGFP or Z−CGFP fusion overnight. However, the Ztev−CGFP fusion is essentially quantitatively cleaved to the Ztev peptide and CGFP fragment. Strikingly, there is *no change* in the fluorescence of either the Z−NGFP/Z−CGFP or the Z−NGFP/(Ztev−)CGFP complex over the course of the overnight cleavage reaction. This indicates that the leucine zipper peptide (and therefore the peptide−peptide interaction) is not necessary to maintain the reassembled GFP.

Kinetics of Unfolding of the Refolded GFP. We sought to estimate the stabilities of the Z−NGFP/Z−CGFP and cleaved Z−NGFP/CGFP complexes by monitoring the loss of fluorescence upon urea denaturation. However, we found that the urea-induced unfolding of these complexes was irreversible with or without DTT reductant. That is, room-temperature incubation of the complexes in 8 M urea resulted in complete loss of fluorescence over a few hours. Dilution to 6 or 3 M urea resulted in less than 5% recovery of fluorescence over as long as 10 days. Since these are not equilibrium conditions, it is not possible to measure the free energy of unfolding. We also observed that loss of fluorescence below 4 M urea was extremely slow, requiring days or longer to see a significant change. This com-

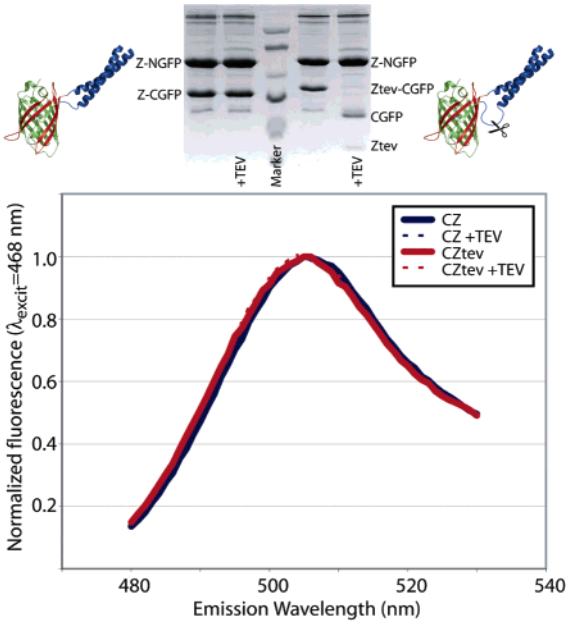


Figure 7. Persistence of GFP fluorescence in the absence of peptide−peptide interaction. (Top) Purified, fluorescent Z−NGFP/Z−CGFP and Z−NGFP/Ztev−CGFP complexes were subjected to overnight incubation with and without TEV protease. Z−NGFP and Z−CGFP are unaffected, but Ztev−CGFP is nearly quantitatively cleaved, freeing the Ztev peptide from the complex. (Bottom) The fluorescence spectrum of both complexes is the same whether the Ztev peptide has been cleaved off, proving that the peptide−peptide interaction is not necessary to maintain the reassembled GFP complex. For clarity, the fluorescence emission values were scaled to the maximum values for the uncut samples, since the concentrations of the two complexes were slightly different.

bination of slow and irreversible unfolding over a broad range of urea concentrations makes it possible to study the kinetics of urea-induced unfolding in real time. Typically, such experiments are performed using stopped-flow techniques, which allow access to fast unfolding events and permit analysis of reaction times before back-reaction (i.e., folding) becomes significant for proteins that unfold reversibly. Here this is not necessary.

Kinetic traces for urea denaturation, observed by fluorescence ($\lambda_{excitation}$ = 468 nm, $\lambda_{emission}$ = 505 nm), were followed over three time regimes: every 12 h for 72 h (0−7.2 M urea), every 24 h for 168 h (0−4.5 M urea), and every 4 h for 48 h (5.75−6.5 M urea). The fluorescence decays were fit to a simple exponential decay equation:

$$F = F_0 e^{-kt}$$

where F is the observed fluorescence, F_0 is the initial fluorescence, k is the decay constant, and t is time. The data and fits for the kinetic traces of Z−NGFP/CGFP are shown (Supporting Information). For each time regime, only those traces that could be fit resulting in an error in k of less than about 10% were included.

These k values represent the rate constants for the dissociation of the complex at given concentrations of urea denaturant. To estimate the rate constant for dissociation in buffer (i.e., zero denaturant), the k values were fit to the empirical equation[24]

$$\ln k = \ln k_0 + m[\text{urea}]$$

(24) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman & Co.: New York, 1999.

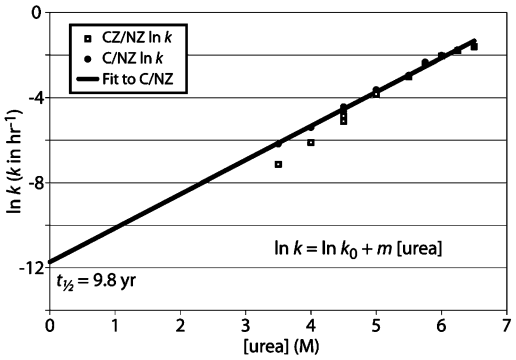


Figure 8. Rate of dissociation of reassembled GFP. Fluorescence decays of CGFP/Z−NGFP (from TEV scission of Ztev−CGFP/Z−NGFP) in various concentrations of urea were fit to single-exponential functions (Supporting Information). The natural logarithms of the decay constants (solid circles) are plotted versus urea concentration, and extrapolated back to zero urea (aqueous solution). The $t_{1/2}$ of the dissociation of the reassembled GFP is predicted to be about a decade in the absence of denaturant. At less than approximately 5 M urea, the apparent rates of dissociation of Z−CGFP/Z−NGFP (open squares) are reduced, presumably due to involvement of the peptide−peptide interactions (see the text).

where k_0 is the rate constant in buffer and m is the slope of the fit (Figure 8, filled circles). This extrapolation affords a value for $k_0 = (8.0 \pm 2.0) \times 10^{-6}$ h^{-1}, which corresponds to $t_{1/2}$ = 9.8 years. Therefore, the complex is extremely kinetically stable (i.e., inert), and the reassembly of the GFP is an essentially irreversible process under these conditions. This k_0 value also confirms that the bait−prey interaction is not necessary to maintain the complex under the screening conditions.

We also carried out the same urea denaturation experiment for the Z−NGFP/Z−CGFP complex (Figure 8, open squares). Above 5 M urea, the k values for the fluorescence decay were essentially identical to the values obtained for the Z−NGFP/CGFP complex. However, below 5 M urea, the apparent k values were markedly lower for the Z−NGFP/Z−CGFP complex, resulting in a nonlinear relationship between ln k and urea concentration. We speculated that this nonlinearity might be the result of the peptide−peptide interaction, which would only be expected to be significant at lower urea concentrations.

Because GFP is almost entirely a β-sheet structure, and the peptides form an α-helical coiled coil upon association, we examined the complex at different concentrations of urea (0−6 M) by circular dichroism spectroscopy (Supporting Information). Generally, α-helices give rise to strong CD minima at 222 and 208 nm, and the β-sheet gives rise to a weaker minimum around 215 nm. Below 4.5 M urea, significant negative signals were observed at 222 and 208 nm; above 4.5 M urea, there was a weak negative peak around 215 nm. These CD spectra were acquired in triplicate over the course of about 20 min, over which time more than 80% of fluorescence is still observed, even at 6 M urea. Thus, the observed loss of CD signal corresponds to the loss of α-helical content in a fully assembled complex. This supports the view that the peptide−peptide interaction is responsible for the slower dissociation of the Z−NGFP/Z−CGFP complex. It implies both that (1) while the peptide−peptide interaction is not necessary to kinetically stabilize the complex under screening conditions, it nonetheless does stabilize it, and (2) the peptides continue to associate after the complex has formed.

Discussion

A Useful Set of Vectors for Examining Protein−Protein Interactions in Bacteria. Here we present two compatible vectors that can be comaintained in *E. coli* for expression of the GFP fragment fusions. The vectors pET11a−link−NGFP and pMRBAD−link−CGFP are independently maintained in *E. coli* and possess promoters that can be independently induced, and they result in transformants that faithfully report the interaction of the fused proteins or peptides. These vectors have convenient sites for subcloning bait and prey proteins fused to the GFP fragments, as well as sites that can be cleaved by restriction enzymes in the linker regions to aid in subcloning and identification of legitimate clones. They are useful in any *E. coli* strain that expresses T7 polymerase (i.e., lysogenized with DE3 lamboid phage), such as BL21(DE3). Furthermore, a hexahistidine tag appended to the N-terminus of the NGFP fusion permits direct purification of interacting proteins from positive clones.

The plasmids created intermediately between the final "link" vectors and the original pET11a vectors differ slightly in the lengths of the peptide linkers between the GFP fragments and the interacting peptides, but interestingly, they are interoperable. This is important because it suggests that the exact fusion topology is not critical, which significantly expands the usefulness of the screen. If only a single, exact fusion were competent to allow GFP reassembly, a large number of false negatives would result because of these steric restrictions. Further evidence for this notion comes from the insertion of a TEV protease cleavable linker between the CGFP fragment and the zipper peptide, extending the linker by nine amino acids. This complex also results in GFP fragment reassembly. Furthermore, the Kerppola group has fused parallel leucine zippers to the N-termini of nearly the same fragments of the highly related YFP and CFP proteins, which leads to reassembly.[12] Since the topology of the interacting proteins is entirely different in this case (we have fused antiparallel leucine zippers to the C-terminus of NGFP and the N-terminus of CGFP), the data taken together support our hypothesis that the interacting proteins are required to bring the GFP fragments into close proximity, but they are not required to precisely align the fragments.

It is worth noting that addition of purification tags to the C-terminus of the CGFP fragment prevented reassembly. This suggests that it might be possible to improve the screen by optimizing the sequences at the unfused termini, or perhaps even through random mutagenesis in the way that Stemmer and colleagues improved the properties of GFP using DNA shuffling.[25]

Interaction of Antiparallel Leucine Zippers. Parallel leucine zippers are ubiquitous dimerization domains, found in transcription factors such as Fos and Jun. The basis for the interaction of this type of coiled coil has been studied extensively, including by library methods employing protein fragment reassembly of DHFR.[6,9,15−17,26] In general, the association of leucine zippers is controlled by the burial of hydrophobic residues and the interaction of so-called edge positions, whose opposite charges on opposite peptides provide a Coulombic "peptide Velcro".[27,28] Recent work indicates that the positions of the charge−charge interactions, and not just the number of favorable interactions, contribute to affinity.[6,26]

Leucine zippers have also been found to associate in an antiparallel orientation, and the later acknowledgment of this fact suggests that some presumed parallel interactions may in fact be antiparallel.[19] It is considerably less well understood what controls the orientation of leucine zippers than what controls affinity. Three clear factors that favor antiparallelism are topological constraints, edge−edge interactions that are only favorable in the antiparallel orientation, and polar residues in the hydrophobic core that can only hydrogen bond to each other in the antiparallel orientation.[18,20] Antiparallel leucine zipper-like interactions are found in the four-helix bundle Rop and the yeast seryl-tRNA synthetase, and antiparallel leucine zippers have also been designed.

The topology of the fusions to the GFP fragments (i.e., C-terminal to NGFP and N-terminal to CGFP) suggested that an antiparallel leucine zipper might be effective in driving the fragment reassembly, and this turned out to be correct. The engineered zipper peptides exploited both the burial of a polar residue (an Asn) and charge−charge interactions that would only be favorable in the antiparallel orientation. Therefore, when we sought to create a library of zipper interactions of differing affinities to measure the limits of the screen, we decided to base the library on the peptide Velcro concept, randomizing the edge positions on one peptide between Glu and Lys. The 256 different Z(EK) peptides affixed to the CGFP fragment varied from eight charge−charge matches to Z−NGFP (i.e., the wild type), to eight mismatches, and all positional possibilities in between. To a first approximation, we expected that fewer mismatches would lead to higher affinity, although we also anticipated that there might be some positional preferences for the matches and mismatches, on the basis of work in parallel zippers.

Both of these hypotheses were in fact borne out. Z(EK) peptides with sufficient affinity to drive GFP assembly had zero to three mismatches, and those that were insufficient to drive reassembly had three or more mismatches. Mutations in positions K6, E18, and K20 were exceptionally rare, and mutations in any of these positions were what discriminated the positives from the negative among those peptides with three total mutations. Therefore, library statistics suggested that mutations at these three positions are especially disruptive to the engineered antiparallel leucine zipper employed here.

Biophysical analysis of the corresponding peptides demonstrated that the position of the mutations profoundly affected the affinity of the library peptides for the NZ peptide. Indeed, the affinities for the peptides with one mutation varied over an order of magnitude. One of these with a mutation at position 4, and one of the peptides with two mutations (4/27), bound at least as tightly as the wild-type peptide (with no charge−charge mismatches). Thus, the edge position charge−charge interactions are not all equivalent. The statistically determined mutation rates among the positive clones and the low impact of mutations in positions 4 and 27 suggest that the charge−charge interactions at the ends of the peptides contribute less to affinity than more internal interactions. Further exploration of this issue is merited.

Since each mutation has up to about a 10-fold effect on the

(25) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. *Nat. Biotechnol.* **1996**, *14*, 315.
(26) Arndt, K. M.; Pelletier, J. N.; Muller, K. M.; Alber, T.; Michnick, S. W.; Pluckthun, A. *J. Mol. Biol.* **2000**, *295*, 627.
(27) Oshea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.* **1993**, *3*, 658.
(28) Kohn, W. D.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1998**, *283*, 993.

dissociation constants, this suggests that each charge−charge interaction is an average of 0−1.35 kcal mol^{-1} more stable than the associated charge−charge mismatch (i.e., K−E or E−K versus E−E or K−K). Systematic mutations of parallel coiled coils show that E−K interactions are about 0.4 kcal mol^{-1} more stable than E−Q or Q−K pairs, and that an E−K interaction is about 1.0 kcal mol^{-1} more stable than a K−K or E−E repulsion.[29,30] Also, it has been shown that an antiparallel coiled coil with no charge−charge mismatches is 2.1 kcal mol^{-1} more stable than a parallel coiled coil with two charged mismatches, again suggesting approximately 1 kcal mol^{-1} of stability for each attractive interaction relative to a repulsive one.[20]

Two factors complicate the interpretation of these data. First, both the NZ and CZ peptides apparently homodimerize at moderately high concentrations (as determined from α-helical character in CD spectra above about 200 μM, not shown). We did not attempt to measure the affinities of these homotypic interactions, but if a significant fraction of any CZ peptide were homodimerized in solution, then the effective concentration of monomer would be reduced and the apparent K_D would be overestimated by SPR. (This assumes that only monomer is competent to bind, and that the concentration of CZ peptide is much greater than that of the NZ peptide on the chip.) High-affinity NZ−CZ interactions can be measured with confidence as long as the analyte concentration is far below the K_D for homodimerization of CZ. However, higher order oligomerization may partially account for the irregular binding curves of some of the peptides as determined by SPR.

Second, it is conceivable that some of the NZ−CZ pairs could associate in a parallel fashion exclusively or in addition to the antiparallel orientation. It is not clear if the GFP fragments could reassemble if the peptides bound in parallel, but the fact that no peptides with 6−8 mutations were positive in the screen suggests that either this is topologically forbidden for reassembly or the peptides cannot interact this way. It is certain that the orientation could not be distinguished by SPR or CD. (However, if the affinities for parallel and antiparallel association were significantly different, a biphasic binding curve might be observed, and this may be part of what is occurring with the peptides with mutations in position 11 or positions 4 and 27.) The complete lack of binding of the negative control peptide suggests that favorable charge−charge interactions alone do not favor parallel association in this system, since the negative control peptide would be fully charge-complementary in the parallel orientation. This is surprising, considering that buried polar interactions and edge Coulombic interactions were found to contribute about equally to coiled coil orientation in a similar peptide system.[20]

Beyond Peptide−Peptide Interactions. Previously, we had only demonstrated the ability of small, strongly interacting peptides to drive the reassembly of the GFP fragments. To demonstrate that our dissection topology was compatible with larger proteins as well, we fused the three TPR domains of HOP to the CGFP fragment, and known peptide ligands for two of those domains to NGFP. For each TPR domain, cellular fluorescence intensity correctly reported the relative strengths of the TPR−ligand interactions. The Hsc70 and Hsp90 ligand peptides both end in the sequence EEVD-CO_2^-, which is known to be a strong determinant of binding for both TPR1 and TPR2A, although specificity determinants between the domains lie N-terminal to this in the peptide.[23] The zipper peptide ends with the sequence ELAQ-CO_2^-.

However, the absolute affinities of the TPR−ligand interactions did not quantitatively correspond to cellular fluorescence. TPR2B has no known peptide ligand, and the interactions of the Hsc70 and Hsp90 peptides with TPR2B have dissociation constants at least 25-fold higher than those of the cognate TPR1−Hsc70 and TPR2a−Hsp90 interactions. Both TPR2B−Hsc70 and TPR2B−Hsp90 interactions led to cellular fluorescence significantly above background (the zipper peptide control). Since the K_D values of the TPR2B−peptide interactions are near or below the threshold for reassembly established above, the screen is correctly reporting, qualitatively, a weak interaction with the EEVD-CO_2^- peptides. We believe that the expression levels, solubilities, and folding reversibilities of the fusions contribute significantly to the cellular fluorescence (see the next section), which complicates the relationship between cellular fluorescence and K_D. Indeed, TPR2B expresses at a much higher level than TPR1 or TPR2A under similar conditions of overexpression (data not shown).

After our original report, the Kerppola group and then the Michnick group demonstrated the reassembly of virtually the same YFP, CFP, and GFP fragments in mammalian cells with larger proteins as bait and prey. Hu et al. showed that the interaction of full-length Fos and Jun could be detected in COS-1 cells, regardless of which protein was fused to the N-termini of the GFP fragments. Interactions of Rel family proteins and domains (IκBα and NF-κB), which are unrelated to leucine zipper proteins, were also detected this way.[12] Remy and Michnick showed that protein kinase B/Akt fused C-terminal to NGFP could be used to isolate ligands fused to the N-terminus of CGFP from a cDNA library in COS-1 cells.[31] This further demonstrates the ability of the screen to qualitatively identify protein−protein interactions in vivo, but we would caution that false negatives and false positives must be controlled carefully in any library approach, due to complications from solubility and expression levels.

Mechanism of the Screen. The dissection of GFP leads to two nearly insoluble proteins, even when fused to highly soluble leucine zipper peptides. Strikingly, however, cells containing both GFP fragments fused to leucine zipper peptides that interact even weakly ($K_D \approx 1$ mM) become fluorescent by accumulating soluble, reassembled GFP complex, which is competent to autocatalyze chromophore formation. This suggests a number of questions about the mechanism of the GFP reassembly (Figure 9). How do the interacting proteins participate in complex formation and reassembly of GFP? How do the properties of the fused proteins affect the reassembly? What accounts for the detection of such weak interactions?

One thing that is certain is that fusion to interacting proteins is required to initiate the reassembly reaction. If the interacting leucine zipper peptides are replaced by short linker peptides, or zipper peptides with incompatible charged residues, or if one of the zippers is replaced by a TPR domain, no GFP reassembly occurs. However, the interaction between the bait and prey need not be particularly strong, as in the case of weakly interacting antiparallel leucine zipper peptides (up to about 1 mM) or

(29) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Eng.* **1994**, *7*, 1365.
(30) Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849.
(31) Remy, I.; Michnick, S. W. *Methods* **2004**, *32*, 381.

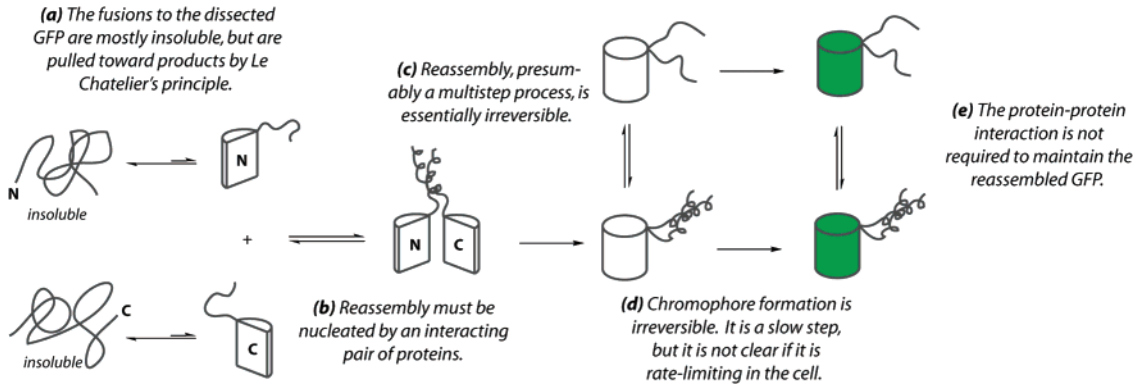


Figure 9. Mechanism of the GFP reassembly screen. Most of the fusions to the dissected GFP are insoluble. Interaction between the dissected GFP portions only occurs if they are fused to interacting proteins, which must nucleate the reassembly reaction, presumably from the small fraction of soluble fusion. Reassembly is essentially irreversible, which effectively pulls more of the fusions into solution, by Le Chatelier's principle. The protein−protein interaction is not necessary to maintain the reassembled complex. Therefore, the screen is dependent upon the expression level and solubility of the fusions, as well as the strength of the protein−protein interaction, but weak and transient interactions can be detected due to the irreversibility of the reassembly reaction.

TPR2B−EEVD peptide interactions (about 0.5 mM). The accumulation of soluble complex and the detection of such weak interactions are probably due to the fact that GFP reassembly is essentially an irreversible process. Extrapolation back to aqueous solution suggests that, even in the absence of interacting proteins, the half-life of the reassembled complex is about a decade, and it is likely longer if the protein−protein interaction is possible after reassembly. This irreversible step effectively pulls the solubilization equilibrium toward soluble products, and traps weak interactions. Presumably, the screen can also be used to trap transient interactions, since the bait−prey interaction is not necessary to maintain the complex, at least in vitro.

It is worth noting that that our independent biophysical measurements of the affinities of the peptides and TPRs do not directly address the affinities of those species when fused to the GFP fragments. Specifically, fusion to the GFP fragments may well perturb the affinities of the fused proteins for each other. Here, we measured the parameters that we thought would be relevant to the potential user of our system, the affinity of the free species and the correlation of those affinities to screen phenotype.

One clear role of the interacting proteins is to nucleate the formation of the complex. However, since most of the material is insoluble upon translation, the physical properties of the fused proteins have an effect on the screen, as well. Greater solubility and expression of the GFP fragment fusions will probably lead to greater cellular fluorescence, within limits. Likewise, extremely low expression of one of the fusions will lead to less cellular fluorescence. Moreover, at least in the case of the protein fused C-terminal to NGFP, it is likely that the fused protein must be able to refold, since fusions of proteins C-terminal to unfolded proteins lead to aggregation. (This is the principle behind a screen for folded, soluble protein.[32]) It is possible that both fused proteins might be improperly folded upon translation, and that the screen is effectively trapping the small amount of soluble fusion in which the interacting proteins properly refold. Therefore, the amount of cellular fluorescence may be related to both the physical properties of the fused proteins and the affinities of those proteins for each other. This implies that the screen is most useful for qualitative identification of interacting proteins, although a semiquantitative relationship between fluorescence and affinity may exist for a series of highly related proteins (e.g., peptide ligands of the same protein).

One important feature of the screen is that the reassembly reaction is not especially dependent upon the geometry of the interacting proteins. Fusions with different linker lengths between GFP fragments and interacting peptides, as well as fusions of a variety of proteins in different primary orientations (i.e., N-terminal versus C-terminal), are competent for reassembly. There are likely to be limits to this tolerance beyond the small number of cases that have been explored, and indeed, we think that this screen will be more demanding of direct interactions (as opposed to interactions through a complex) than yeast two-hybrid analysis, as a result. On the other hand, it is perhaps equally surprising that CD spectroscopy clearly demonstrates that the protein−protein interaction can be maintained in some cases after GFP reassembly, as with the antiparallel leucine zippers examined here.

Conclusion

We have engineered a pair of compatible plasmids that greatly facilitate the use of GFP fragment reassembly as a screen for protein−protein interactions in bacteria. The vectors allow facile subcloning of the genes of interest as fusions to the GFP fragments, and their compatibility and independent transcriptional control afford faithful reporting of interactions. Cellular fluorescence is affected by the physical properties of the fused proteins, so care must be taken in correlating phenotype with affinity, although this is possible with a related series of proteins. The screen can detect weak ($K_D \approx 1$ mM) and probably transient interactions due to irreversibility of the reassembly reaction, and it is remarkably tolerant of the nature of the fragment fusions.

Acknowledgment. We gratefully thank Chung Wang, Institute of Molecular Biology, Academia Sinica, Taiwan, for his gift of the full-length clone of HOP. Velcro is a registered trademark of Velcro Industries B.V. T.J.M. is an NIH Postdoctoral Fellow (Grant GM065750). C.G.M.W. is a James Hudson Brown−Alexander Brown Coxe Postdoctoral Fellow (Yale University School of Medicine). This work was supported in part by NIH Grants GM62413 and GM57265 (L.R.) and GM35208 (A.D.H.) and a fellowship from the Leukemia and Lymphoma Society of America (I.G.).

(32) Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. *Nat. Biotechnol.* **1999**, *17*, 691.

Supporting Information Available: Description of the materials and methods used in this study; figures showing the schematic of GFP dissection and fusion, analysis of positive and negative clones, rate of dissociation of reassembled GFP, and circular dichroism spectrum of the Z−NGFP/Z−CGFP complex; and a table showing the sequences of peptides scored positive and negative in the screen (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA046699G