

## SUPPLEMENTARY MATERIAL

### “Detecting protein-protein interactions with a GFP-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\*

\*email lynne.regan@yale.edu

## MATERIALS & METHODS

**Plasmid construction.** All enzymes for subcloning were purchased from New England Biolabs (Beverly, MA) unless noted. Digested plasmids and PCR products were purified by electrophoresis onto dialysis membrane before ligation. Ligated plasmids were transformed into DH10B *E. coli* by electroporation, and all DNA sequences were confirmed by the W.M. Keck Foundation Biotechnology Resource Laboratory (Yale University, New Haven, CT).

*Plasmids for fusion to CGFP.* The plasmid pMR101-CZGFP, which places CZGFP under the control of the T7lac promoter, was created by PCR amplification (Deep Vent polymerase) of the gene for CZGFP from pET11a-CZGFP using the oligonucleotides 5'-aataatccatggctagcgcacagctgg-3' and 5'-attattggatccttatcagttgtacagttcatccatgcc-3' (CZMRB.RV). This was followed by digestion of this gene and the plasmid pMR101 (p15A origin, *kan*<sup>R</sup>) with *Nco* I and *Bam*H I, and then ligation (T4 DNA ligase). The plasmid pMRBAD-CZGFP was created by PCR amplification of the *araBAD* promoter (including the *araC* gene) from pBAD/HISa (Invitrogen, Carlsbad, CA) using the oligonucleotides 5'-aataatccatgggtatatctccttctt-aaac-3' and 5'-aataatctgcagatcgatgcataatgtgcc-3', and PCR amplification of the CZGFP gene from pET11a-CZGFP using 5'-aataatccatggctagcgaacagctggagaagaaactg-3' and CZMRB.RV. (The former primer reverses a Glu4Ala mutation in pET11a-CZGFP that occurred spontaneously during the original cloning.) These fragments were digested with *Nco* I and *Pst* I, and *Nco* I and

*Bam*H I, respectively, and pMR101 was digested with *Bam*H I and *Pst* I. Ligation of these three fragments afforded the desired plasmid.

The plasmid pMRBAD-link-CGFP was created by PCR amplification of CGFP from pMRBAD-CZGFP using the oligonucleotides 5'-aataatccatggctaataatgcataatgggacgtcgggtggaag-cggtaagaatgg-3' and 5'-attattggatccttagttgtacagttcatcatgc-3' followed by digestion with *Nco* I and *Bam*H I. The *araBAD* promoter and pMR101 were prepared as above in a triple ligation with the CGFP fragment. The plasmid pMRBAD-Z-CGFP was engineered by annealing and Klenow extension of the oligonucleotides 5'- ttattaccatggcaagcgagcagctggaaaagaagtacaagccc-tggagaaaaaacttgctcagctg-3' and 5' - aataatgacgtctgcgcgagttttttccaatgcttggttttccattccagctgagcaagtttttc-3'. The resulting cassette and pMRBAD-link-CGFP were digested with *Nco* I and *Aat* II and ligated.

*Plasmids for fusion to NGFP.* The plasmid pET11a-link-NGFP was created by PCR amplification of NGFP from pET11a-NZGFP using the oligonucleotides 5' - aataatgctagccatcac-caccatcaccatggcgcgagcaaaaggagaagaactcttcaactgg-3' and 5' - caaggtgctctggctctggctcgagcaatccgggaattaaggatccattatt-3'. The former primer adds an N-terminal hexahistidine tag to the NGFP. This NGFP fragment and pET11a-NZGFP were digested with *Nhe* I and *Bam*H I and ligated. The plasmid pMRBAD-Z-NGFP was then created by Klenow extension of the oligonucleotides 5' - attattctcgagcgcctcaaaaaagaattgcaggcaaaaaagaacttgcgcagctgaagtgg-3' and 5' - aataatggatccttactgcgccagttccttttcagagcttgaactcccacttcagctgcgcaagttc-3'. This cassette and pMRBAD-link-NGFP were digested with *Xho* I and *Bam*H I and ligated.

The version of pET11a-NGFP used here is slightly different from the one in the previous report. QuikChange (Stratagene, La Jolla, CA) mutagenesis was used to reverse a Leu152Pro mutation that occurred spontaneously during cloning in the vector used in the previous report.

Note that the version of GFP used here is based of the sg100 variant (G. Palm, personal communication). Elsewhere, we have reviewed the use of other variants, topologies and fusion points in depth.<sup>1</sup>

**Construction of a library of antiparallel leucine zippers.** Each of the eight e and g “edge” positions in the leucine zipper peptide appended to the CGFP fragment was randomized to either Glu or Lys. A library cassette was created by Klenow extension of the oligonucleotides 5'- ttattaccatggcaagcgagcagctgraaaagraattacaagccctgraaaaaraacttgctcagctgraatgg-3' and 5'- taataagacgtctgcgcgagttt~~ttt~~tycaatgcttggtt~~ttt~~ccattycagctgagcaag-3'. (The randomized codons are underlined. R stands for an equimolar mixture of A and G; Y for an equimolar mixture of C and T.) To remove any cassettes with mismatches in the overlap region (italicized residues) caused by the randomized position, the cassette was subjected to PCR with the oligonucleotides 5'- ttattaccatggcaagcgagc-3' and 5'-aataatgacgtctgcgcgag-3'. The cassette and plasmid pMRBAD-link-CGFP were digested with *Nco* I and *Aat* II. After calf alkaline phosphatase treatment of the digested vector and agarose gel purification onto dialysis membrane, the fragments were ligated. The ligation was treated with *Sph* I to remove background, and the reaction was incubated at 80 °C for 20 min to denature the restriction enzyme. BL21(DE3) *E. coli* were transformed with pET11a-Z-NGFP and made competent for electroporation. These cells were then transformed with the treated pMRBAD-Z(EK)-CGFP library ligation.

**Analysis of the library.** After growth as described above, fluorescent colonies were picked with a sterile toothpick, swirled in 5 µL of sterile water and then restreaked onto screening media to confirm the phenotype. The cells in water were heated to 95 °C for 5 min and then subjected to PCR amplification of the Z(EK) region using *Taq* DNA polymerase (Promega, Madison, WI) and the oligonucleotides 5'-tagcggatcctacctgacgc-3' and 5'-ttcgggctt~~ttt~~gtagcagcc-

3'. The reactions were prepared for sequencing using the ExoSAP-IT kit (USB, Cleveland, OH) and sequenced in 96-well format by the Keck facility. Forty-eight positives and 48 negatives were sequenced.

**Surface plasmon resonance analysis of peptide-peptide interactions.** All peptides were chemically synthesized and HPLC purified by the Keck. SPR analysis was performed with a BIAcore 3000 instrument. The NZ peptide (NH<sub>2</sub>-CGGSGALKKELQANKKELAQLKWEL-QALKKELAQ-CO<sub>2</sub>H) was covalently coupled to a CM4 sensor chip through NHS-EDC/PDEA thiol chemistry via the N-terminal Cys, according to the manufacturer's instructions. NZ peptide injections were repeated until a signal of 500 RU above baseline was obtained, before capping of unused PDEA sites with free cysteine. A reference channel was also prepared, containing immobilized cysteine in place of peptide, and signal from the binding channel was subtracted from the reference channel.

Experimental CZ peptides were dissolved in HBS-EP-CM buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20, 0.5 mg mL<sup>-1</sup> carboxymethyl-dextran). (Addition of CM-dextran to HBS-EP was found to significantly improve response signal-to-noise.) Peptide concentrations were determined by measuring UV absorbance at 280 nm, assuming a calculated molar extinction coefficient of 5,690 M<sup>-1</sup> cm<sup>-1</sup>.<sup>2</sup> We verified that A<sub>280</sub> was the same in native buffer and 6 M guanidine for the NZ peptide. All experiments were performed at 25 °C. Experimental sample concentrations were between 1-500 μM. The injection sequence was: HBS-EP-CM buffer (1.5 min INJECT), CZ peptide (1 min KINJECT, followed by 10 min dissociation time), 1 M NaCl regeneration (1 min QUICKINJECT). Flow-rate was 40 μL min<sup>-1</sup>. All measurements were performed in duplicate.

Dissociation constants ( $K_D$ ) were obtained by fitting binding curves, assuming a simple

1:1 equilibrium binding model using the equation:

$$R_{eq} = \frac{R_{max}}{\left(\frac{K_D}{C_0} + 1\right)}$$

where  $C_0$  is the initial concentration of the CZ peptide in solution,  $R_{max}$  is the maximum (saturation) SPR signal, and  $R_{eq}$  is the observed equilibrium SPR signal. Here, we take the initial concentration of CZ peptide ( $C_0$ ) to be the same as the concentration at equilibrium due to vast excess.

**Circular dichroism of peptide-peptide interactions.** Peptides were dissolved in 50 mM sodium phosphate (pH 7.4) and their concentrations determined by UV absorbance. CD spectra were acquired for each individual peptide at 75  $\mu$ M, and for 1:1 mixtures of each of the CZ peptides with the NZ peptide (75  $\mu$ M each). The increase in helicity upon binding was determined from the difference between the CD signal at 222 nm for the arithmetic sum of the spectra from the individual peptides and the measured spectrum of the mixture of peptides ( $\Delta\Theta_{222}$ ). Since the peptides are initially present at the same concentration ( $C_0$ ), the apparent  $K_D$  can be estimated from:

$$K_D = C_0 \frac{(1 - f_{bound})^2}{f_{bound}}$$

where  $f_{bound}$  is the fraction bound at equilibrium. Here, we assume that only 1:1 binding occurs and that the ellipticity of each NZ/CZ dimer is the same upon full binding. The  $f_{bound}$  values were determined from  $\Delta\Theta_{222}/\Delta\Theta_{222}^{max}$ , where  $\Delta\Theta_{222}^{max}$  is a hypothetical maximum value selected such that the  $K_D$  for the wild-type CZ peptide would match the  $K_D$  determined from SPR ( $f_{bound} = 60\%$  in this case).

**Construction of TPR and ligand GFP fusions.** Individual TPR domains of Hsp Organizing Protein (HOP) were obtained by PCR amplification from a full-length HOP clone. TPR1 and TPR2A were amplified with primer sets 5'-aataatccatggctatggagcagggtcaatgagctgaagg-  
agaaaggc-3' and 5'-attattgacgtcccatattctgtaaaccctcttcagttgaggg-3', and 5'-aataatccatggctaagc-  
aggcactgaaagaaaaagagc-3' and 5'-attattgacgtcccttgctcctcaggattttctgctgctggtgc-3', respectively. These were digested with *Nco* I and *Aat* II before subcloning into pMRBAD-link-CGFP. A native *Nco* I site in the TPR2B gene sequence was removed through QuikChange mutagenesis before amplification with primers 5'-aataatccatggctgacctggctttggaggagaag-3' and 5'-attattgacg-  
tcccggtgtactgcgcatcaca-3'.

Sequences corresponding to the C-terminal 24 residues of human Hsc70 and Hsp90 were prepared by Klenow extension of primers 5'-aataatctcgagcgggggatttctggtggtgagctcctccctctgg-  
tggcttctcagggc-3' and 5'-attattgatcctcattaatcaacctcttcaatggtggccctgaggaagcaccaccagaggag-  
3', and 5'-aataatctcgagcagtgctgctgtaactgaagaaatgccacccttgaaggagatgacgacacat-3' and 5'-  
attattgatcctcattagtctacttctccatgcgtgatgtgctgcatctccttcaaggggtg-3', respectively. Reaction products were digested with *Bam*H I and *Xho* I, and ligated into pET11a-link-NGFP.

**SPR analysis of TPR-ligand interactions.** Individual TPR domains of HOP were cloned into plasmid pProEX-HTA (Invitrogen), and expressed in BL21(DE3) as N-terminal hexahistidine-tagged fusions. Expression products were purified on Ni-NTA agarose, and His<sub>6</sub>-tags removed by rTEV protease (Invitrogen) digestion followed by size exclusion chromatography over Superdex 75 (Amersham Biosciences). TPRs were dialyzed against water and stored as lyophilized powder at -80 °C. Peptides, biotinylated at the N-terminus and corresponding to the 24 C-terminal residues of human Hsc70 (bio-GGFPGGGAPPSGGASSGP-

TIEEVD-CO<sub>2</sub>H) and Hsp90 (bio-SAAVTEEMPPLGDDDDTSRMEEVD-CO<sub>2</sub>H), were synthesized and HPLC purified by the Keck facility.

SPR was performed as described for the leucine zipper interactions except as noted. Biotinylated peptides were dissolved in buffer HBS-EP (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20) at a concentration of 5 mg mL<sup>-1</sup> and bound to a CM4 sensor chip surface via NHS-coupled neutravidin. Lyophilized TPR samples were dissolved in buffer HBS-EP, and stock protein concentrations determined by measurement of UV absorbance at 280 nm assuming calculated molar extinction coefficients ( $\epsilon_{\text{TPR1}} = 14,770 \text{ M}^{-1} \text{ cm}^{-1}$ ;  $\epsilon_{\text{TPR2A}} = 12,920 \text{ M}^{-1} \text{ cm}^{-1}$ ;  $\epsilon_{\text{TPR2B}} = 11,880 \text{ M}^{-1} \text{ cm}^{-1}$ ).<sup>2</sup> TPR samples were prepared by dilution into HBS-EP at concentrations between 1-500  $\mu\text{M}$ . The injection sequence was: HBS-EP buffer (1 min INJECT), TPR sample (1 min KINJECT followed by 10 min dissociation), 1 M NaCl regeneration (1 min QUICKINJECT). A reference channel, containing NHS-coupled neutravidin capped with biotin, was subtracted from the experimental channel.

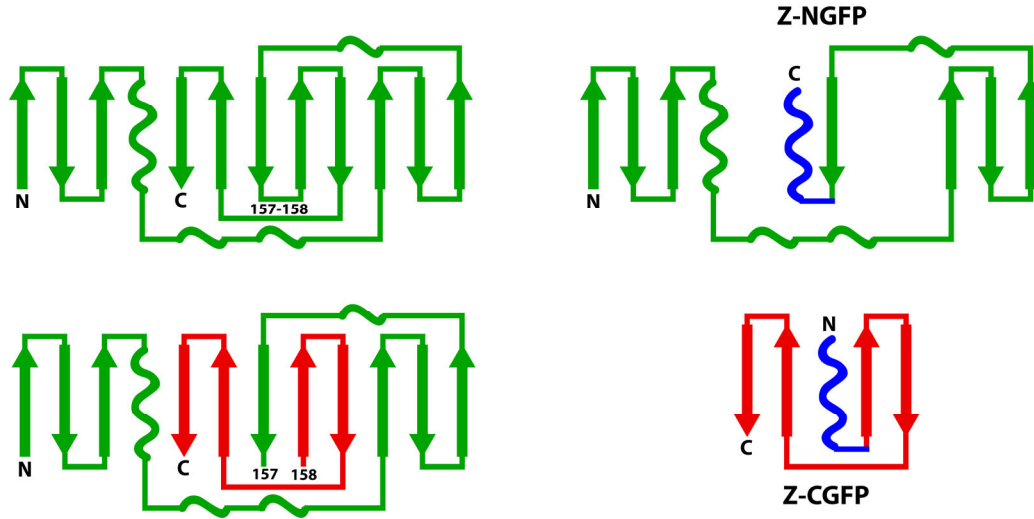
**Persistence of refolded GFP.** The Ztev cassette was created by Klenow extension of the oligonucleotides: 5'-ttattaccatggcaagcgagcagctggaaaagaagtacaagccctggagaaaaaacttgctcagctg-gaatgaaaaaccaag-3' and 5'- taataagacgtcccctgaaaatacaggttttcgccgcctgcgcgagttttttccaatgcttgg-ttttccattccag-3' with subcloning into pMRBAD-link-CGFP as described above to yield pMRBAD-Ztev-CGFP.

Z-NGFP/Ztev-CGFP and Z-NGFP/Z-CGFP complexes (which have a His<sub>6</sub>-tag on the N-terminus of the NGFP fragment) were expressed from BL21(DE3) *E. coli* with growth in LB medium. After overnight growth of 1 L of culture at 37 °C, cells were grown at 25 °C for 4 more days with addition of 0.2 % arabinose and 15  $\mu\text{M}$  IPTG. Protein was then purified according to the manufacturer's instructions for Qiagen Ni-NTA agarose with some modifications. Briefly,

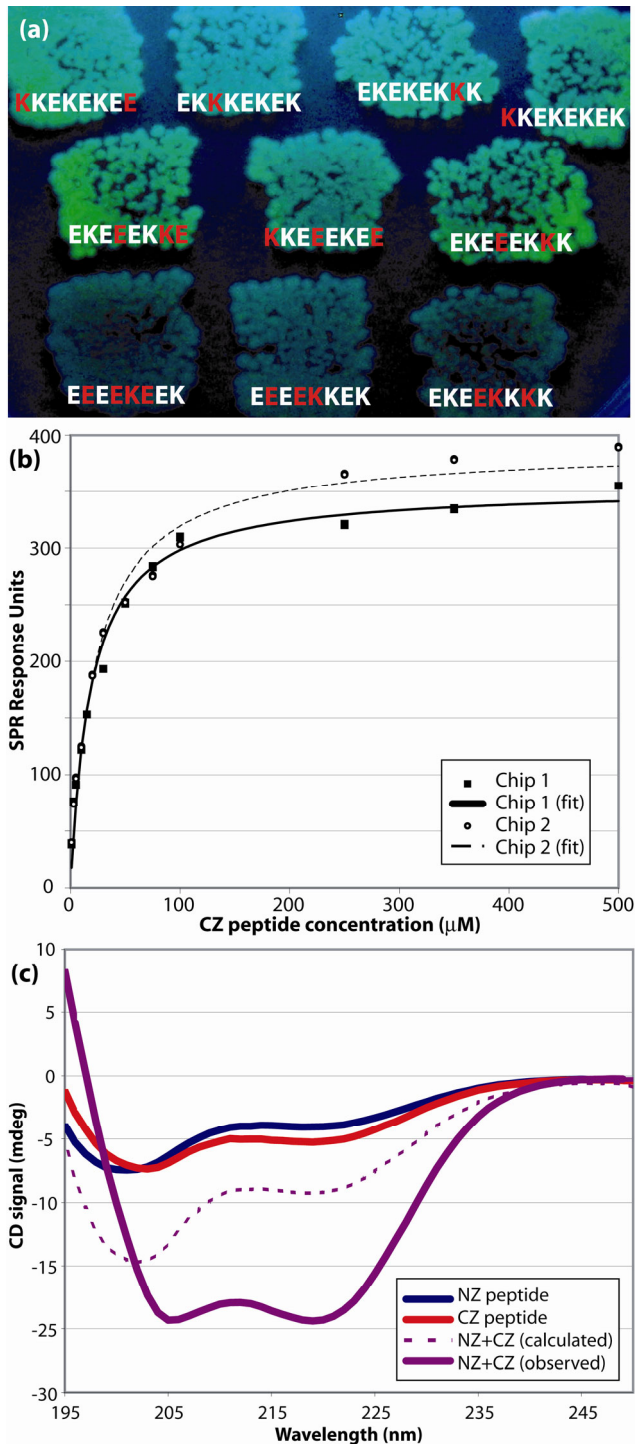
lysis was accomplished in 25 mL of lysis buffer (50 mM Tris-HCl pH 8, 200 mM NaCl, 10 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) by 2 h treatment on ice with DNase I, RNase I and 1 mg mL<sup>-1</sup> hen egg white lysozyme, and addition of 5 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub> and 0.1% Triton X-100. After brief sonication, the lysate was cleared by centrifugation at 30,000 g for 1 h, and the soluble fraction was mixed with 3 mL of Ni-NTA agarose (Qiagen) slurry and incubated at 4 °C for 1 h. The bound matrix was washed twice with 15 mL of wash buffer (same as lysis buffer, but 20 mM imidazole) and eluted in six 1 mL fractions with elution buffer (same as lysis buffer, but 250 mM imidazole). Purified complexes were then exchanged into GFP buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM dithiothreitol) using a PD10 column (Amersham Biosciences) for storage in the dark at 4 °C.

**Circular dichroism of refolded GFP under denaturing conditions.** Far-UV CD (195-260 nm) spectra recorded in the same buffer used for kinetics with 0, 1.5, 4.5 and 6 M urea. CD spectra were acquired immediately, with completion within 20 minutes of mixing with urea.

## FIGURES AND TABLES

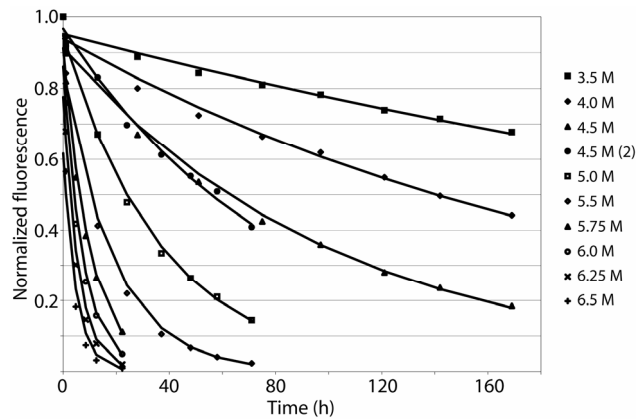


**Supple. Fig. 1. Schematic of GFP dissection and fusion.** A schematic of the secondary structure topology of GFP is shown on the top left. In this work, GFP is dissected in a loop between residues 157 and 158 (bottom left). Fusion (here, of the leucine zipper peptides) is made to the C-terminus of NGFP [i.e., GFP(1-157)] and the N-terminus of CGFP [i.e., GFP(158-238)]. Arrows represent  $\beta$ -strands, squiggles represent  $\alpha$ -helices and lines represent loops.

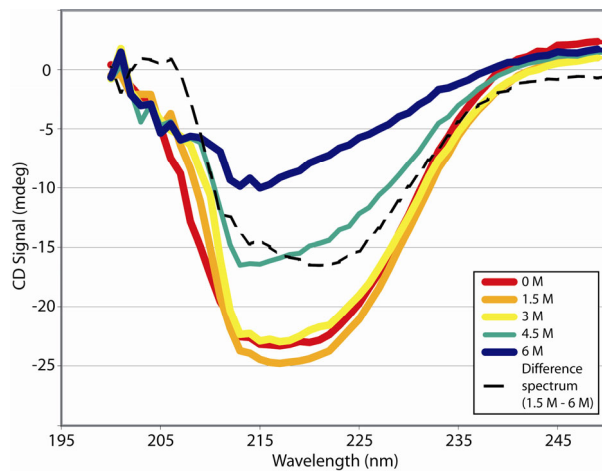


**Supple. Fig. 2. Analysis of positive and negative clones.** (a) The ten clones chosen for SPR analysis are shown. The sequences of the e and g positions are noted, with mutations from the original peptide highlighted in red. The top two rows were scored as positives, and the bottom row was scored as negatives. (b) Example SPR data sets for the “wild-type” CZ peptide (with no charge-charge mismatches). Independent data sets from two different chips are shown. Each data set consists of two duplicate points per concentration. Because the loading density of Cys-NZ peptide on the chips is not exactly the same, the  $R_{\text{max}}$  (saturation response) values differ slightly. The fit  $K_D$  values differ by less than  $\pm 8\%$ . (c) Example CD data set to estimate peptide binding. The CD signal for 75  $\mu\text{M}$  NZ peptide, 75  $\mu\text{M}$  CZ peptide and a mixture of the two peptides at

75  $\mu\text{M}$  each is shown. The broken line is the arithmetic sum of the individual CZ and NZ spectra.



**Supple. Fig. 3. 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.



**Supple. Fig. 4. Circular dichroism spectrum of Z-NGFP/Z-CGFP complex.** At low concentrations of urea, there is substantial CD signal at 222 nm, indicative of high  $\alpha$ -helical content. That signal begins to diminish at 4.5 M urea, and the remaining CD signal at 6 M urea has a weak minimum at 215 nm, indicative

of  $\beta$ -sheet. The dashed line is a difference spectrum between 1.5 M and 6 M urea, emphasizing the loss of signal at 222 nm. Spectra were acquired within 15 min of addition of urea, at which time more than 80% of fluorescence remains even at 6 M urea. Therefore, the loss of CD signal here is almost entirely due to denaturation of the antiparallel leucine zippers.

**Supple. Table 1.**

Sequences of peptides scored positive and negative in the screen. Positions of mutations are underlined. Only the eight randomized positions, **bold** in the full sequence that follows, are listed in the table:

EQLE**EK**KLQAL**EKK**LAQL**E**WKNQAL**EKK**LAQ

Mutations	Positives	Negatives
0	EKEKEKEK EKEKEKEK	
1	EKE <u>EE</u> EKEK <b>K</b> KKEKEKEK EKEKEK <b>KK</b> EKE <u>EE</u> EKEK EKE <u>EE</u> EKEK <b>E</b> EEKEKEK EKEKE <b>KK</b> EKEKE <b>KK</b> E <b>KK</b> KEKEK EKEKE <b>KK</b> EKE <u>EE</u> EKEK EKE <u>EE</u> EKEK EKEKE <u>EE</u> EKE <u>EE</u> EKEK EKEKE <b>KK</b> EKEKE <u>EE</u> EKEKE <b>KK</b>	
2	EKE <u>EE</u> E <b>KK</b> <b>K</b> K <u>EE</u> EKEK <b>K</b> KKEKE <b>KK</b> E <b>KK</b> KEKE <u>EE</u> E <b>KK</b> KEKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> <b>K</b> KKEKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> EKE <u>EE</u> EKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> E <b>KK</b> KEKE <u>EE</u> E <b>KK</b> KEKE <u>EE</u> EKEKEK <u>EE</u> EKE <u>EE</u> E <b>KK</b> EKE <u>EE</u> E <b>KK</b> EKEKEK <u>EE</u> EKE <u>EE</u> E <b>KK</b>	
3	EKE <u>EE</u> E <b>KK</b> E E <b>KK</b> <u>EE</u> EKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> E <b>K</b> KKEKEK <u>EE</u> <b>K</b> KKEKEK <u>EE</u> EKE <u>EE</u> E <b>KK</b> E EKE <u>EE</u> E <b>KK</b> E EKE <u>EE</u> E <b>KK</b> E <b>K</b> K <u>EE</u> EKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> E <b>K</b> KKEKEK <u>EE</u>	<u>EE</u> E <b>KK</b> EK EKE <u>EE</u> E <b>KK</b> EKKEKKEK EKE <u>EE</u> E <b>KK</b> <u>EE</u> E <b>KK</b> E <b>KK</b> <u>EE</u> EKE <u>EE</u> EKE <u>EE</u> E <b>KK</b> EKE <u>EE</u> E <b>KK</b> EKE <u>EE</u> E <b>KK</b>
4		<u>EE</u> EKEEK <u>EE</u> E <b>KK</b> KK <u>EE</u> EKEEK <u>EE</u> E <b>KK</b> KEK EKKEK <b>KK</b> <u>KE</u> EEEEEK <u>EE</u> E <b>KK</b> E <u>EKE</u> KE <u>EE</u> <u>EE</u> K <b>KK</b> KK <u>EKE</u> KK <u>EE</u>
5		<u>KE</u> EEFEK <u>KE</u> KEFEK <u>KE</u> EEKEK <u>KK</u> KEK <b>KK</b> <u>EE</u> KEK <b>KK</b> <u>KE</u> EEEEE
6		<u>KE</u> EEFEK <u>KE</u> EEKE <u>EE</u>

## REFERENCES

1. Magliery, T. J.; Regan, L., Reassembled GFP: Detecting protein-protein interactions and protein expression patterns. In *Green Fluorescent Protein: Properties, Applications and Protocols*, 2nd ed.; Chalfie, M.; Kain, S. R., Eds. John Wiley & Sons: Hoboken, NJ, 2004; in press.
2. Gill, S. C.; Hippel, P. H., *Anal. Biochem.* **1990**, *182*, 319.