

## DNA Sequence-Enabled Reassembly of the Green Fluorescent Protein

Cliff I. Stains, Jason Porter, Aik T. Ooi, David J. Segal and Indraneel Ghosh\*

*Contribution from the Department of Chemistry University of Arizona, Tucson, Arizona 85721*

ghosh@email.arizona.edu

### Supplementary Materials

#### **Cloning of NGFP-ZnFingerA and CGFP-ZnFingerB Proteins**

**General Materials and Methods:** All restriction enzymes, Taq Polymerase, DNA ligase, dNTPS were obtained from New England Biolabs.

**Initial Cloning:** NGFP and CGFP coding DNA sequences were obtained by PCR amplification from plasmids which have been previously described<sup>1,2</sup> using the following primers (Figure S1).

NGFP-BamHI:  
GCTACGGGATCCATGGCTAGCAAAGGAGAA

NGFP-PstI:  
GCACGTCTGCAGACCTTGTTTGTCTGCCAT

CGFP-KpnI:  
CGTGCAGGTACCAAGAATGGAATCAAAGTG

CGFP-HindIII:  
CGACGTAAGCTTGGATCCTCAGTTGTACAG

**Figure S1.** Primers used for subcloning of GFP fragments into the pQE30 expression plasmid.

NGFP and CGFP fragments were digested with BamHI/PstI and KpnI/HindIII respectively and ligated into pQE30 (Qiagen) expression plasmids containing Zif268 (cloned with overlapping primers based upon published sequences)<sup>3</sup> and PBSII (gift from the Barbas Laboratory)<sup>4</sup> coding regions separated by a flexible 15 amino acid linker, sequences were confirmed by dideoxyoligonucleotide sequencing at the University of Arizona DNA Sequencing Facility.

**Protein Expression from pQE30 Plasmids:** Test protein expressions from these two plasmid constructs were unsuccessful in XL1-Blue (Stratagene), Top10 (Invitrogen), and BL21-Gold (DE3) (Novagen) cell lines; consequently a more robust system for protein expression was chosen which would also allow for the expression of both proteins within a single cell. The T7 promoter system (Novagen) contains plasmids specifically designed for this purpose and we chose the pETDuet-1 expression vector (Novagen) based on previous work which had shown that the expression of the same dissected GFP halves fused to leucine zippers produced adequate yields using a similar pET expression system.

NGFP-Zif268 BglII:

CCGCGGCGCGCGCGGAGATCTGATGGCTAGCAAAGGA

NGFP-Zif268 XhoI:

CGCGCGCGCGCCGGCTCGAGGTCCTTCTGCCGCAA

CGFP-PBSII EcoRI:

CCGCGCGGCCGGCGCGCAATTCGGAGAAGCCCTAT

CGFP-PBSII NotI:

GGCGGCGCGTGCGGCCGCTTATCAGTTGTACAGTTC

pETDuet-1 MCSI:

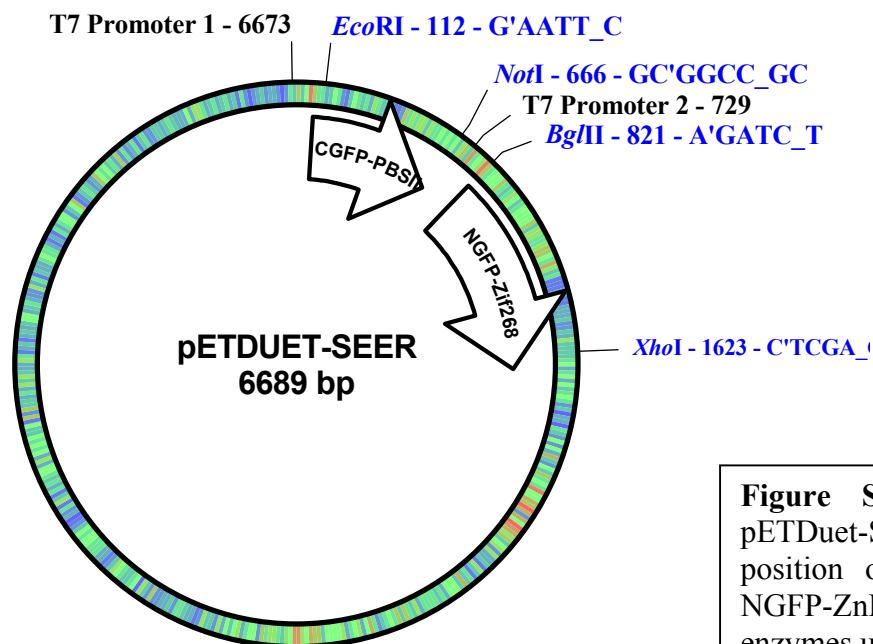
Fwd-ATGCGTCCGGCGTAGA Rev-GATTATGCGGCCGTGTACAA

pETDuet-1 MCSII:

Fwd-TTGTACACGGCCGCATAATC Rev-GCTAGTTATTGCTCAGCGG

**Figure S2.** Primers used for cloning of the SEER proteins into the pETDuet-1 expression plasmid as well as sequencing primers for both MCS of pETDuet-1.

**Cloning SEER Proteins into pETDuet-1:** Primers were used to amplify the NGFP-Zif268 (NGFP-ZnFingerA) and CGFP-PBSII (CGFP-ZnFingerB) genes from the above pQE30 plasmids by PCR amplification (Figure S2). These genes were then successively ligated into the pETDuet-1 expression plasmid using BglII/XhoI and EcoRI/NotI respectively yielding a plasmid containing CGFP-ZnFingerB in MCSI which contains an N-terminal His-tag and NGFP-ZnFingerA in MCSII which contains a C-terminal S-tag this plasmid was called pETDuet-SEER (see Figure S3), sequences were confirmed by dideoxyoligonucleotide sequencing at the University of Arizona DNA Sequencing Facility (see Figure S2) and are given below (see Figure S4).



**Figure S3.** A schematic of the pETDuet-SEER plasmid showing the position of the CGFP-ZnFingerB and NGFP-ZnFingerA genes, restriction enzymes used, and T7 promoter sites.

### Expression and Purification of NGFP-ZnFingerA (NGFP-Zif268) and CGFP-ZnFingerB (CGFP-PBSII)

**General Materials and Methods:** Buffer A is 10 mM Tris-HCl @ pH=7.5, 100 mM NaCl, 1 mM DTT, and 100  $\mu$ M ZnCl<sub>2</sub>. All reagents were obtained from Sigma unless otherwise noted. LB and 2xYT media were purchased from Becton Dickinson.

**Protein Expression:** BL21-Gold (DE3) cells (Novagen) were transformed with pETDuet-SEER using the standard heat shock protocol, plated on LB-Amp Agar plates, and grown overnight at 37 °C to obtain single colonies. Single colonies were picked and used to inoculate 2xYT media containing Amp and grown overnight with shaking at 37 °C. This overnight culture was used to inoculate a one liter 2xYT-Amp culture containing 100  $\mu$ M ZnCl<sub>2</sub> (EM Science) to a final O.D.<sub>600</sub> of 0.05. Cells were shaken at 37 °C until an O.D.<sub>600</sub> of 0.5-0.8 was reached at which time they were induced with 1 mM IPTG (Research Products International Corporation). Cells were induced for three hours after which they were pelleted at 3000 rcf and frozen overnight.

This yielded approximately 15 mg of CGFP-ZincFingerB of which 7.5 mg was purified by subsequent IMAC and 26 mg of NGFP-ZnFingerA of which 6 mg was purified by subsequent IMAC from a one liter culture.

**Figure S4.** A.) The DNA and amino acid sequence of CGFP-ZnFingerB. CGFP is shown in green, PBSII is shown in red, and the linker is shown in pink. B.) The DNA and amino acid sequence of NGFP-ZnFingerA. NGF is shown in green, Zif268 is shown in blue, and the linker shown in pink.

A. CGFP-ZnFingerB:	1	ATGGGCAGCA	GCCATCACCA	TCATCACCAC
		M	G S S	H H H H H H
31	AGCCAGGATC	CGAATTCGGA	GAAGCCCTAT	GCTTGTCCGG AATGTGGTAA
	S Q D P	N S E	K P Y	A C P E C G K
81	GTCCTTCAGC	CAGCGCGCAA	ACCTGCGCGC	CCACCAGCGT ACCCACACGG
	S F S	Q R A N	L R A	H Q R T H T G
131	GTGAAAAACC	GTATAAATGC	CCAGAGTGCG	GCAAATCTTT TAGCCGCAGC
	E K P	Y K C	P E C G	K S F S R S
181	GATCACCTGA	CTACCCATCA	ACGCACTCAT	ACTGGCGAGA AGCCATACAA
	D H L T	T H Q	R T H	T G E K P Y K
231	ATGTCCAGAA	TGTGGCAAGT	CTTTCAGTCG	CAGCGATGTG CTGGTGCGCC
	C P E	C G K S	F S R	S D V L V R H
281	ACCAACGTAC	TCACACCGGT	GGGGGTGGCG	G TTCAGGCGG TGGGGGTTCT
	Q R T	H T G	G G G G	S G G G G S
331	GGTGGGGGTG	GTACCAAGAA	TGGAATCAAA	GTGAACTTCA AGACCCGCCA
	G G G G	T K N	G I K	V N F K T R H
381	CAACATTGAA	GATGGAAGCG	TTCAACTAGC	AGACCATTAT CAACAAAATA
	N I E	D G S V	Q L A	D H Y Q Q N T
431	CTCCAATTGG	CGATGGCCCT	GTCCTTTTAC	CAGACAACCA TTACCTGTCC
	P I G	D G P	V L L P	D N H Y L S
481	ACACAATCTG	CCCTTTCGAA	AGATCCCAAC	GAAAAGAGAG ACCACATGGT
	T Q S A	L S K	D P N	E K R D H M V
531	CCTTCTTGAG	TTTGTAACAG	CTGCTGGGAT	TACACATGGC ATGGATGAAC
	L L E	F V T A	A G I	T H G M D E L
581	TGTACAAC TG A			
	Y N *			

B. NGFP-ZnFingerA: 1 ATGGCAGATC TGATGGCTAG CAAAGGAGAA GAACTCTTCA  
M A D L M A S K G E E L F T

41 CTGGAGTTGT CCAATTCTT GTTGAATTAG ATGGTGATGT TAACGGCCAC  
G V V P I L V E L D G D V N G H

91 AAGTTCTCTG TCAGTGGAGA GGGTGAAGGT GATGCAACAT ACGGAAAAC  
K F S V S G E G E G D A T Y G K L

141 TACCCTGAAG TTCATCTGCA CTA CTGGCAA ACTGCCTGTT CCATGGCCAA  
T L K F I C T T G K L P V P W P T

191 CACTAGTCAC TACTCTGTGC TATGGTGTTT AATGCTTTTC AAGATACCCG  
L V T T L C Y G V Q C F S R Y P

241 GATCATATGA AACGGCATGA CTTTTTCAAG AGTGCTATGC CCGAAGGTTA  
D H M K R H D F F K S A M P E G Y

291 TGTACAGGAA AGGACCATCT TCTTCAAAGA TGACGGCAAC TACAAGACAC  
V Q E R T I F F K D D G N Y K T R

341 GTGCTGAAGT CAAGTTTGAA GGTGATACCC TTGTTAATAG AATCGAGTTA  
A E V K F E G D T L V N R I E L

391 AAAGGTATTG ACTTCAAGGA AGATGGCAAC ATTCTGGGAC ACAAATTGGA  
K G I D F K E D G N I L G H K L E

441 ATACA ACTAT AACTCACACA ACGTTCCCAT CATGGCAGAC AAACAAGGTC  
Y N Y N S H N V P I M A D K Q G L

491 TGCAGGGCGG TTCAGGCGGT GGGGGTTCTG GCGGGGGTGG GTACCCCGGG  
Q G G S G G G G S G G G G Y P G

541 GAACGCCCTT ACGCTTGCCC AGTGGAGTCC TGTGATCGCC GCTTCTCCCG  
E R P Y A C P V E S C D R R F S R

591 CTCCGACGAG CTCACCCGCC ACATCCGCAT CCACACAGGC CAGAAGCCCT  
S D E L T R H I R I H T G Q K P F

641 TCCAGTGCCG CATCTGCATG CGCAACTTCA GCCGCAGCGA CCACCTCACC  
Q C R I C M R N F S R S D H L T

691 ACCCACATCC GCACCCACAC AGGCGAAAAG CCCTTTGCCT GCGACATCTG  
T H I R T H T G E K P F A C D I C

741 TGGAAGAAAG TTTGCCAGGA GCGATGAACG CAAGAGGCAT ACCAAGATCC  
G R K F A R S D E R K R H T K I H

791 ACTTGCGGCA GAAGGACCTC GAGTCTGGTA AAGAAACCGC TGCTGCGAAA  
L R Q K D L E S G K E T A A A K

841 TTTGAACGCC AGCACATGGA CTCGTCTACT AGCGCAGCTT AA  
F E R Q H M D S S T S A A \*

**Purification by metal affinity chromatography:** Cells were re-suspended in Buffer A and lysed using standard sonication protocols. NGFP-ZnFingerA was found entirely in inclusion bodies whereas a relatively small amount of CGFP-ZnFingerB was found in the soluble fraction (with the remainder residing in inclusion bodies). Consequently both proteins were purified under denaturing conditions as follows. Inclusion bodies obtained from above were solubilized in Buffer A containing 6 M Urea and incubated on ice for one hour. The resulting solution was diluted to 4 M Urea with Buffer A and clarified by centrifugation at 18,000 rcf for 20 minutes. This lysate was passed over Ni-NTA agarose beads (Qiagen) and eluted with Buffer A containing 4 M Urea and increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM sequentially). NGFP-ZnFingerA eluted in the 2 mM imidazole fractions whereas CGFP-ZnFingerB eluted in the 50-500 mM fractions (a mixture of both proteins was found in the 10-20 mM fractions). Fractions of CGFP-ZnFingerB which still contained small amounts of NGFP-ZnFingerA were further purified by dialyzing into Buffer A with 4 M Urea and 2 mM imidazole and re-exposure to the same IMAC column. NGFP-ZnFingerA eluted in the flow-through, 2 mM imidazole, whereas CGFP-ZnFingerB eluted in the previously observed fractions.

### Mass Spectroscopy (MALDI) of the SEER Proteins

The refolded SEER proteins were analyzed by MALDI-MS analysis. MALDI mass spectra were acquired on a Bruker Reflex-III MALDI/TOF the masses obtained were within 0.1% of the calculated masses.

NGFP-ZnFingerA MH<sup>+</sup> calculated: 32686; found: 32648

CGFP-ZnFingerB MH<sup>+</sup> calculated: 21294; found: 21273

### Refolding Experiments

**General Materials and Methods.** All spectra were taken on a Photon Technology International (PTI) spectrofluorometer with excitation and emission wavelengths of 468 nm and 505 nm respectively. Slit widths were set to 5 nm for excitation and 10 nm for emission. All refolding experiments were conducted using 3.5K MWCO Slide-A-Lyzer Dialysis Cassettes (Pierce). All DNA constructs used in refolding are shown in Figure S5 and were obtained HPLC purified

*Zif268-3-PBSII:*

GCGTAGCGTGGGCGTAA**GTGTGGAAACACCG**

*Zif268-10-PBSII:*

GCGTAGCGTGGGCGTAGGACGATA**GTGTGGAAACACCG**

*Zif268:*

GCGTAGCGTGGGCGTAGGACGATAC**CCTATGTGCCACCG**

*PBSII:*

GCGTAC**CCTATGTGC**TAGGACGATAG**GTGTGGAAACACCG**

**Figure S5.** DNA targets used in the refolding experiments. **Blue**, **Red**, and **Pink** indicate the respective Zif268, PBSII, and decoy DNA binding sites. Numbers between the zinc finger names indicate the distance between binding sites in base pairs.

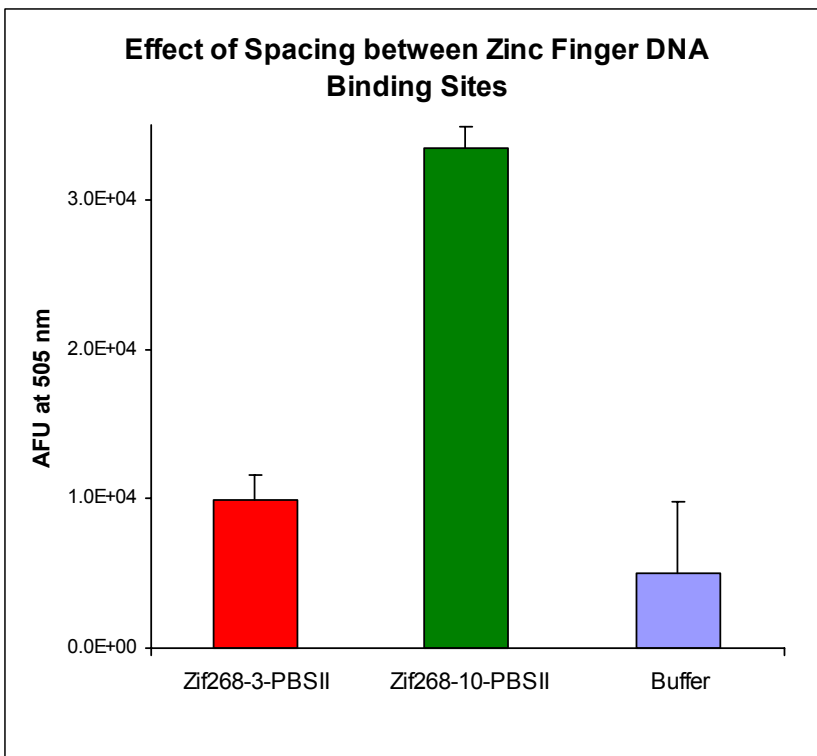
from IDT. Oligos were annealed in 1x BamHI Buffer (NEB) using the following procedure: heating to 95 °C for 7 min, cooling to 56 °C at a rate of 1 °C/min, equilibrating at 56 °C for 5 min, and finally cooling to 25 °C at a rate of 1 °C/min using a Techne Genius thermocycler. All refolding experiments were conducted at 4 °C. The theoretical extinction coefficients for NGFP-ZnFingerA and CGFP-ZnFingerB at 280 nm are 17210 and 7680 M<sup>-1</sup> cm<sup>-1</sup> respectively.

**Initial Refolding Experiments:** Initial refolding experiments were conducted using SEER proteins in Buffer A containing 4 M Urea obtained by IMAC from the 10 mM imidazole wash. The concentration of each half in this was determined to be 15 μM by UV absorbance at 280 nm. Samples were refolded as follows, 4 μM of *Zif268-10-PBSII* double stranded DNA was added to 1 mL of SEER proteins (15 μM each) and dialyzed into Buffer A in a stepwise manner (Buffer A containing 2 M Urea, 1 M Urea, 0.5 M Urea, and twice into Buffer A with no Urea) over a period of two days. A negative control was also performed in which no DNA was added to the SEER proteins. Precipitate was observed in the negative control sample but not in the sample containing *Zif268-10-PBSII* DNA, indicating that the proteins were significantly more soluble in the presence of DNA (this was confirmed by later observations with every DNA target tested including Herring Sperm DNA). Fluorescence excitation and emission spectra of these samples were always taken two days (48 hours) after exchange of Urea.

**GFP Reassembly as a Function of DNA Concentration:** The concentration of each SEER protein was kept constant at 5 μM while the concentration of *Zif268-10-PBSII* DNA was varied between 5, 10, and 20 μM in 250 μL total of Buffer A containing 4 M Urea. Samples were refolded as before into Buffer A over a period of two days and emission spectra of each sample were taken two days after refolding.

**Specificity of the GFP Reassembly Process:** Samples containing 5 μM of each SEER protein and 2.5 μM of each DNA target *Zif268-10-PBSII*, 2.5 μM of *Zif268*, 2.5 μM of *PBSII*, and 15.4 μg of Herring Sperm (Invitrogen) DNA were prepared in 250 μL total Buffer A containing 4 M Urea. These samples along with a negative control (no DNA) were refolded as described above over a period of two days. Fluorescence emission spectra of each sample were taken two days after refolding.

**Effect of Spacing between Zinc Finger Binding Sites:** Separate samples containing 5 μM of each SEER protein and 2.5 μM or *Zif268-3-PBSII* (three base pair separation between binding sites) and *Zif268-10-PBSII* in 250 μL of Buffer A containing 4 M Urea were refolded as stated above over a period of two days. The fluorescence emission spectra of each sample were taken and are shown in Figure S6.



**Figure S6.** Fluorescence of equimolar SEER proteins (5  $\mu$ M) in the presence or absence of (2.5  $\mu$ M) DNA with spacing (3 and 10) between binding sites.

## References

1. Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2000**, *122*, 5658-5659
2. Magliery, T. J.; Wilson, C. G. M.; Pan, W. L.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2005**, *127*, 146-157
3. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809-817
4. a) Blancafort, P.; Segal, D. J.; Barbas, C. F., 3rd *Mol. Pharmacol.* **2004**, *66*, 1361-1371, b) Segal, D. J. *Methods* **2002**, *26*, 76-83, c) Pabo, C.O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, *70*, 313-340