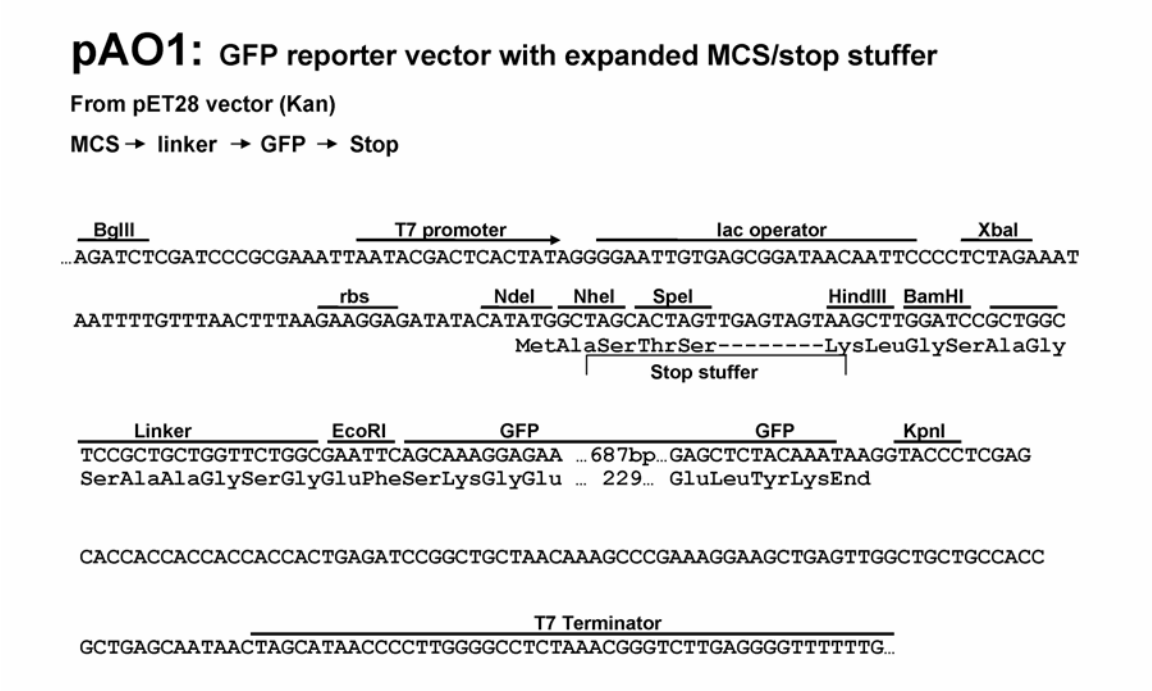
# Appendix A

## Expression vectors




Note: all vector maps list unique restriction sites only.

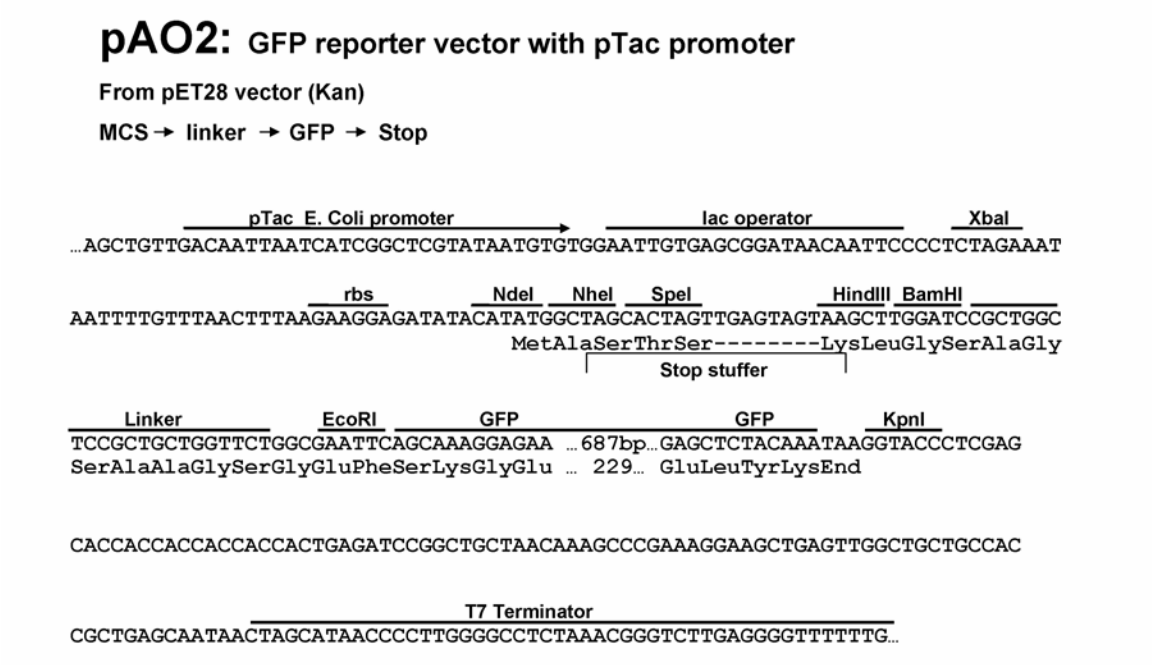
## Rationale

This vector was created to improve the utility of the GFP reporter vector described in Waldo et al. 1999 Nature Biotech. 17, 691. A new multiple cloning site (MCS) was cloned into the NdeI and BamHI restriction sites of the original vector. This MCS represents an improvement in two ways. First, the additional restriction sites add cloning flexibility and facilitate higher cloning efficiency, as cloning into NdeI at high efficiency is difficult (C.A.O, unpublished data). The second benefit of the new MCS is that the additional restriction sites represent “kill” sites to eliminate false negatives that result from self-ligation of single-cut plasmid contamination. For example, if the goal is to clone a library into the NheI and BamHI restriction sites, cutting the ligation product with SpeI or HindIII will eliminate self-ligants (DNA ligase must be heat inactivated first).

Note that efficient double digestions should result in none to very few self-ligants. An alternative method for elimination of self-ligants is to CIP (phosphatase) treat the digested vector; however, the digestion method may be simpler and/or more effective. For the intended purpose of the vector – to evolve protein solubility – false negatives should not affect an experiment. However, in some cases it may be beneficial to determine the faction of library that is non-functional. Both the original MCS and the new MCS include a stuffer with three stop codons, one in each frame, in order to help eliminate false positives. Also, in order to eliminate false positives, it is critical to gel purify the library PCR product since small fragments, such as primer-dimers that contain both restriction sites, will result in high GFP expression if in frame.

## Experimental Procedures

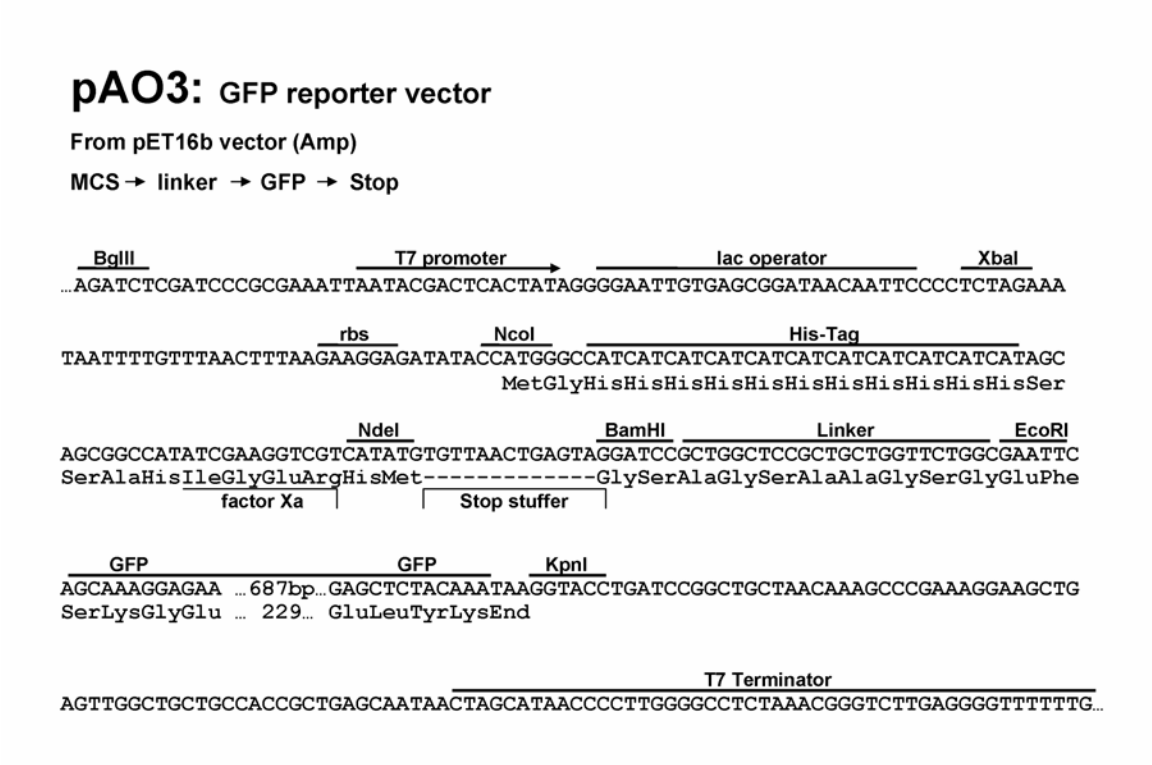
pAO1MCSS (5'-GCC GTA GAT CAT ATG GCT AGC ACT AGT TGA GTA AGC TTG GAT CCG ACT GG) and pAO1MCSA (reverse complement) were annealed and digested with NdeI and BamHI. The digestion product was purified using a desalting column (Prince-sep 20, Princeton Separations) and ligated into the NdeI and BamHI restriction sites of the GFP reporter vector described by Waldo et al (1999) Nature Biotech. 17, 691.




## Rationale and Experimental Procedures

This vector was created in order to be able to express the GFP reporter vector in *E. coli* host strains. This was done in order to increase the throughput of directed evolution experiments using the GFP reporter screen. The original construct (Waldo et al. 1999 Nature Biotech. 17, 691) was cloned into pET28, which utilizes the T7 promoter and requires transformation of strains that have stable incorporation of the T7 RNA polymerase, such as BL21 (DE3) cells (DE3 represents a lysogen bearing the T7 RNA plolymerase gene). The benefit of an orthogonal promoter is the increased control of protein expression. The drawback is that BL21 (DE3) cells, the most common cells for T7-based expression, are engineered for protein expression (e.g., certain protease genes are negated) but are not suitable for cloning. Host strains such as DH5-α or XL10-gold are engineered for transformation of ligation products in part because they are negative

for the recombinase gene *recA*. Using the original GFP reporter vector requires transformation with a protein library ligation reaction of a host strain first, followed by collection and plasmid extraction (miniprep) of the transformation pool and, finally, transformation of expression strain with the supercoiled plasmid pool. Cloning into a vector that has protein expression driven by an *E. coli* promoter would allow expression screening after transformation of host strains. An example of a promoter recognized by *E. coli* is the hybrid pTac, which is a lac operator-driven promoter found in the pGEX series vectors from Amersham Biosciences. We replaced the T7 promoter sequence in pAO1 with pTac by cloning a PCR fragment containing pTac and the lac operator into the BglII and XbaI restriction sites of pAO1 to create pAO2. pAO2 was used to screen a random binary patterned library to determine the fraction of sequence space which is folded (unpublished data, C.A.O.)



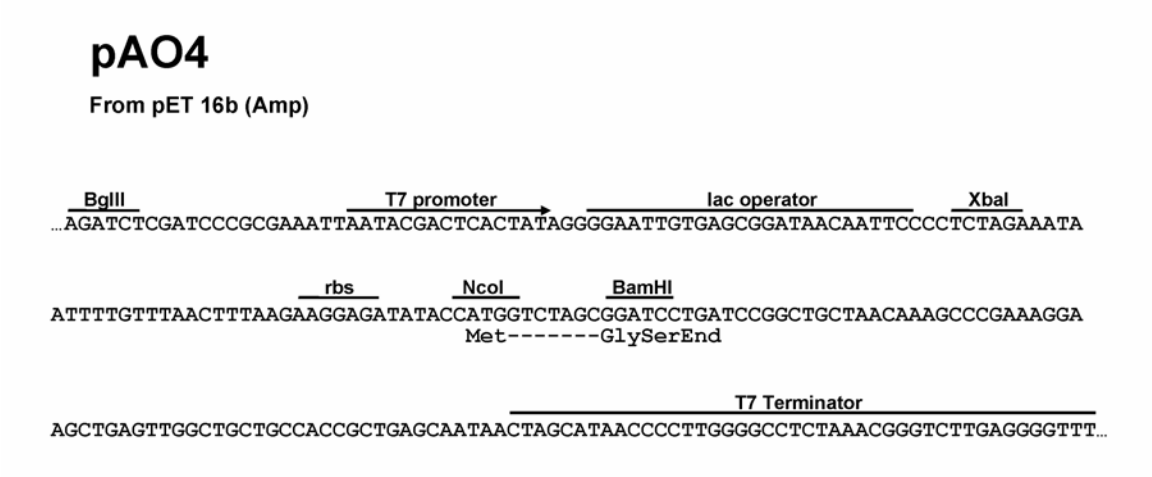

## Rationale

This vector contains the GFP reporter insert (from the vector described in Waldo et al. 1999 Nature Biotech. 17, 691) placed into pET16b. This was done to take advantage of the features of the pET16b vector, including the availability of NcoI for cloning or the N-terminal His-tag that is available when cloning into NdeI. An NcoI restriction site that was in the GFP gene sequence was eliminated. This vector was used for assessing the expression characteristics of the 10FnIII library (see Chapter 2).

## Experimental Procedures

In order to clone the GFP reporter fragment into the pET16 vector while at the same time eliminating the NcoI site found in GFP, we amplified GFP in two halves with four primers: GFPREP5NdeI (5’-GAA GGA GAT ATA CAT ATG TGT TAA CTG AGT

AGG) and GFPkoNcoIR (5’-GTG TTG GCC ACG GAA CAG GTA GTT TTC CAG TAG TGC); GFPkoNcoIF (5’-CTA CCT GTT CCG TGG CCA ACA CTT GTC ACT ACT CTG) and GFPREP3BclI (5’-GAC AGC TGA TCA GGT ACC TTA TTT GTA GAG CTC ATC CAT GC). The two fragments produced contained overlapping sequence, including the eliminated the NcoI site (eliminated by synonymous mutation). The two fragments were extended by 4 cycles of PCR without primers. Then primers GFPREP5NdeI and GFPREP3BclI were added to amplify the entire GFP fragment. This protocol introduces mutations faster than the standard site quickchange protocol. Quickchange requires amplification of the entire plasmid in both directions, followed by transformation of the amplified plasmids. After plasmid extraction and sequencing, it is important to cut the gene insert and place it into an un-amplified vector, as the amplified vector may bear deleterious mutations outside of the ORF, which is time consuming. Therefore, performing PCR in two steps may be a faster method for generating point mutations. The GFP PCR product was digested with NdeI and BclI and cloned into pET16b at the NdeI and BamHI restriction sites. BclI and BamHI have compatible overhangs, and cloning one into the other eliminates the restriction site. Therefore, the BamHI site contained in the GFP reporter was preserved and left available for cloning.

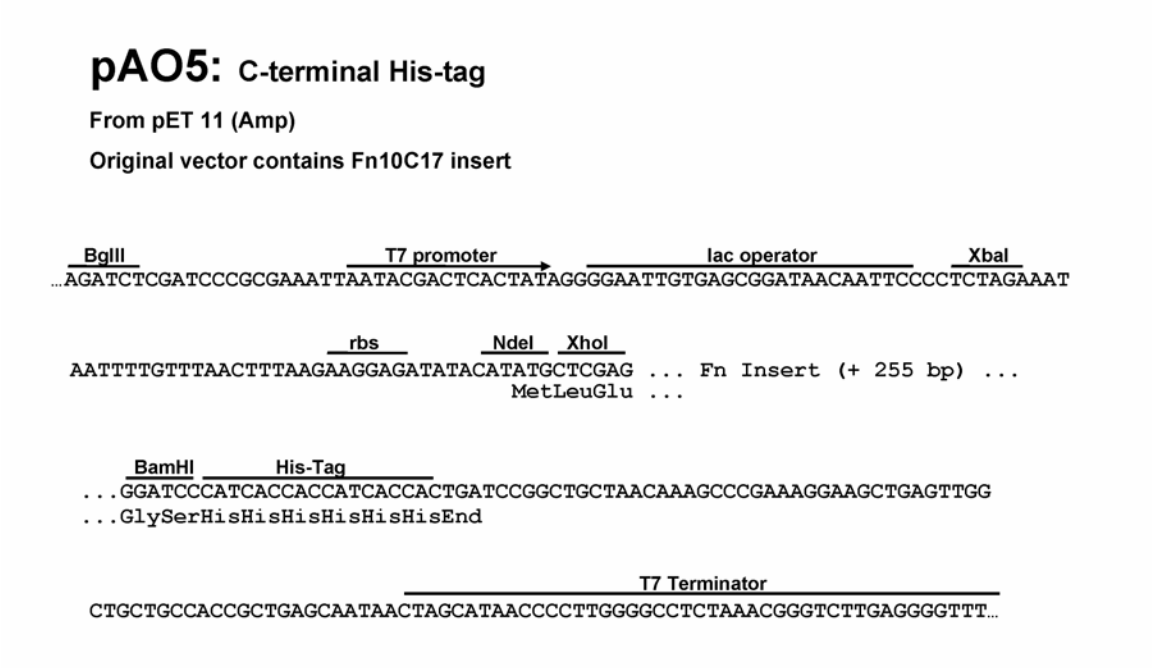



## Rationale

This vector was created for simple sub-cloning of genes from pAO3 into a non-GFP fusion context. This vector creates a modification of the pET16b ORF in that cloning into BamHI in-frame will result in a stop codon immediately following the Gly-Ser encoded by the BamHI restriction site. Cloning into the original BamHI site of pET16b results in the extension Gly-Ser-Gly-Cys-Stop at the C-terminus. We decided that the extraneous sequence containing the C-terminal cysteine was undesirable. An alternative method to eliminate extraneous sequence would be to include a stop codon in a 3’ primer preceding the BamHI restriction sequence. This is undesirable, as the purpose of this vector was to excise the inserts from pAO3 directly with NcoI and BamHI without a PCR step. Digesting a gene insert from a vector is desirable because one, a PCR step is eliminated, and two, the following ligations are much more efficient than ligations with digested PCR products. This is presumably due to higher digestion efficiency by restriction endonucleases with plasmid DNA rather than PCR products, since PCR products have short extensions adjacent to the cut site.

## Experimental Procedures

pAO4MCSS (5’-GGA GGA CCA TGG TCT AGC GGA TCC TGA TCA GGA AGG) and pAO4MCSA (reverse compliment of pAO4MCSS) were annealed, digested with NcoI and BclI, purified with QIaquick columns using buffer PN (Qiagen), and ligated into the NcoI and BamHI sites of pET16b. By cloning with BclI, we were able to retain a BamHI restriction site for 3’ cloning and achieve our goal of adding a stop codon immediately following a BamHI site, since BclI codes for a stop codon when in-frame.




## Rationale

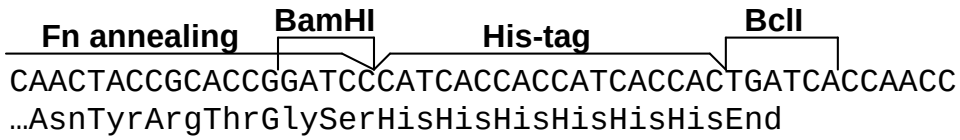
This plasmid was created to provide a simple expression vector that utilizes cloning with NdeI or XhoI and BamHI while creating a C-terminal His-tag with minimal extraneous sequence. BamHI is a common restriction site for 3’ cloning and was incorporated into the 10FnIII library (resulting in a Gly-Ser appendage as residues 89 and 90 of expression products, see Chapter 2). Also, the 10FnIII library contains a 5’ XhoI restriction site (which encodes residues 2 and 3; Leu-Glu). One commercial vector that places a His-tag at the C-terminus is pET21; however, this vector is not ideal for sub-cloning the 10FnIII library due to limitations with the available restriction sites. The C-terminal His-tag orientation is desirable, as the N-termini of fibronectin molecules are adjacent to the binding loops. An N-terminal His-tag may negatively affect binding and would require proteolytic cleavage, while a C-terminal His-tag should not affect protein function and could remain intact for biophysical characterization. This vector was designed for cloning fibronectin molecules without requiring PCR amplification with additional primers; i.e.,

XhoI and BamHI can be used to digest fibronectin PCR product DNA from mRNA selection pools or any plasmids containing fibronectin clones. In addition to eliminating PCR steps, this strategy eliminates the need for cloning pool DNA into TOPO-TA plasmids for sequencing. For example, converged pool DNA at the end of a selection can be digested with XhoI and BamHI and ligated into the same restriction sites of pAO5. Pool members can be picked and plasmid prepped for sequencing, and the clones can be simultaneously assayed for the ability to be expressed in bacteria. If evolution is necessary with any of the clones, pAO5 plasmids bearing fibronectin DNA can be digested with NdeI and BamHI, and the purified insert can be placed into vectors pAO1, pAO2, or pAO3. Likewise, clones evolved using the GFP reporter screen can be cut with either NdeI or XhoI and BamHI and placed back into pAO5 again without requiring PCR amplification. Also, these NdeI-BamHI or XhoI-BamHI inserts would be compatible with sub-cloning into vectors pAO6-9 (see below).

## Experimental Procedures

This vector was created by amplification of fibronectins 10C17 and 10C19 from pCR4-TOPO using Fnlibrary5Nhe (5’-CTA TTT ACA ATT CAT ATG CTC GAG GTC GTC G) and Fnlibrary3His (GGT TGG TGA TCA GTG GTG ATG GTG GTG ATG GGA TCC GGT GCG GTA GTT G). The PCR products were digested with NdeI and BclI, purified (QIaquick, Qiagen), and ligated into the NdeI and BamHI restriction sites of pET11a, creating vectors pAO5-10C17 and pAO5-10C19.

Below is a diagram of primer Fnlibrary3His in the sense orientation:




BclI and BamHI have compatible overhangs. By digesting the PCR product with BclI and ligating it into the BamHI site of pAO5, the vector BamHI site is eliminated, while the new BamHI site encoded by the insert is preserved prior to the His-tag. Also, BclI encodes a stop, which is incorporated into the resulting vector.

## pAO6: N-terminal NusA fusion

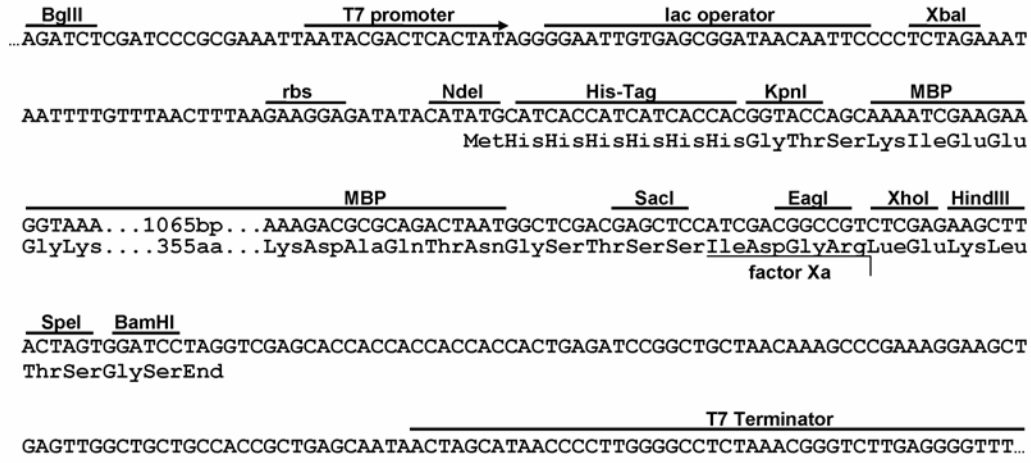
From pET 28 (Kan)

His-Tag → NusA solubility tag → fXa → MCS → STOP


## pAO7: N-terminal MBP fusion

From pET 28 (Kan)

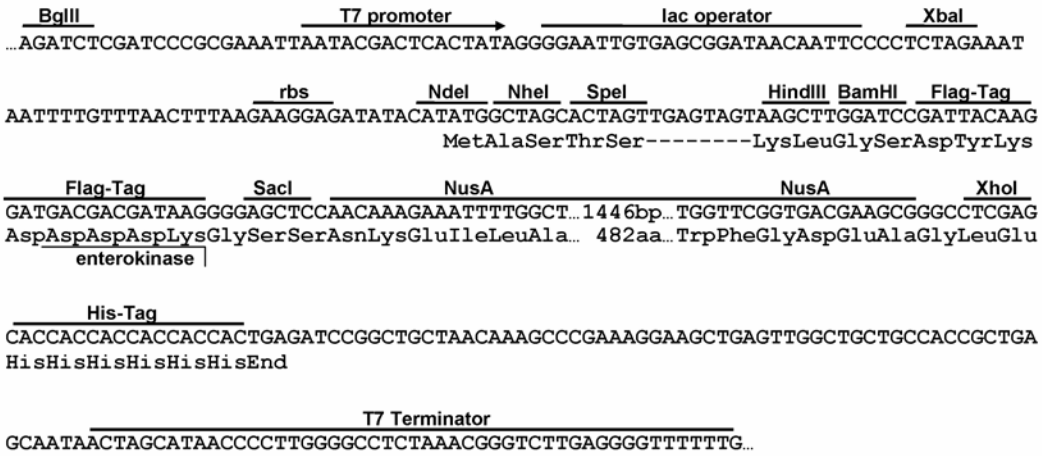
His-Tag → MBP solubility/affinity tag → fXa → MCS → STOP



## pAO8: C-terminal NusA fusion

**From pET28 vector (Kan)**

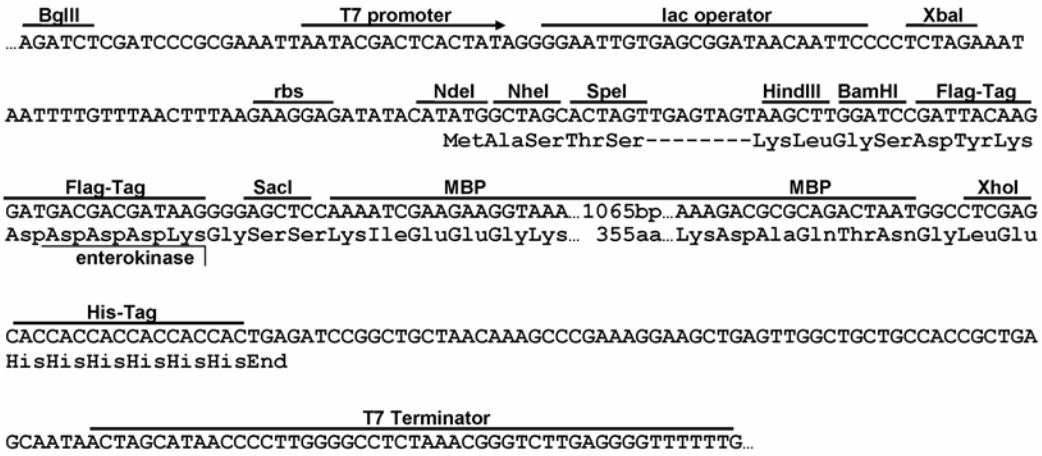
**MCS → FLAG-Tag/Enterokinase Cleavage site → NusA solubility Tag → His-Tag → STOP**




## pAO9: C-terminal MBP fusion

**From pET28 vector (Kan)**

**MCS → FLAG-Tag/Enterokinase Cleavage site → MBP solubility/affinity tag → His-Tag → STOP**



### Rationale (pAO6-9)

These four vectors were created to use the solubilizing properties of two proteins, NusA and MBP. Both proteins were incorporated into two vectors each to create both N-terminal and C-terminal fusions. Both proteins are used as solubilizing tags in commercial vectors (NusA: pET44, Novagen; MBP: pMal, NEB); however, we decided to create new vectors that reduced extraneous sequence, increased functionality, and improved the cloning strategy in relation to the fibronectin library and vectors that were constructed previously.

Cloning into the MCS of vectors pAO6 and pAO8 will generate NusA fusion proteins. Cloning into pAO7 and pAO9 will generate MBP fusions. pAO6 and pAO7 create N-terminal fusions and are identical other than the choice of the solubilizing protein tag. Cloning into pAO8 and pAO9 create C-terminal fusions and are also identical other than choice of solubilizing fusion. All four vectors contain a His-tag that is N-terminal in relation to the solubilizing protein tags in pAO6 and pAO7 and C-terminal to the solubilizing tags in pAO8 and pAO9.

All four vectors were tested with one protein, 10C17C25. MBP fusions produced slightly more soluble protein than NusA fusions (unpublished data, C.A.O.). However, the fraction of total protein expressed that was soluble was equal for both tags. MBP also has added value as an affinity tag. Vectors that generate fusions in both N- and C-terminal orientations were created to test whether or not orientation has an effect on expression. For 10C17C25 fusions, there was no difference in the amount of protein produced for either orientation.

The N-terminal fusions have the feature that cloning into the EagI restriction site while maintaining the factor Xa protease recognition sequence will allow for the enhanced expression of proteins that begin with any amino acid at the N-terminus after factor Xa processing. There is no extraneous sequence created by this cloning strategy, and an N-terminal Met is avoided (if desirable). This has applications for structural biology and native protein ligation. (Primers that introduce a stop codon preceding the 3' restriction sequence also eliminate extraneous C-terminal residues).

For some purposes it is not necessary to eliminate the solubilizing protein tag, such as the production of fibronectins for functional characterization (see Chapter 3). In this case, producing C-terminal fusions may be better than the N-terminal fusions due to the proximity of the fibronectin N-terminus and binding loops. The C-terminal fusion vectors contain an additional affinity/epitope tag, the Flag-tag, which acts as a linker between the MCS and the solubilizing tag. The C-terminal fusions produced by cloning into pAO8 or pAO9 are able to be cleaved with enterokinase, producing proteins that contain only a C-terminal Flag-tag.

## Experimental Procedures

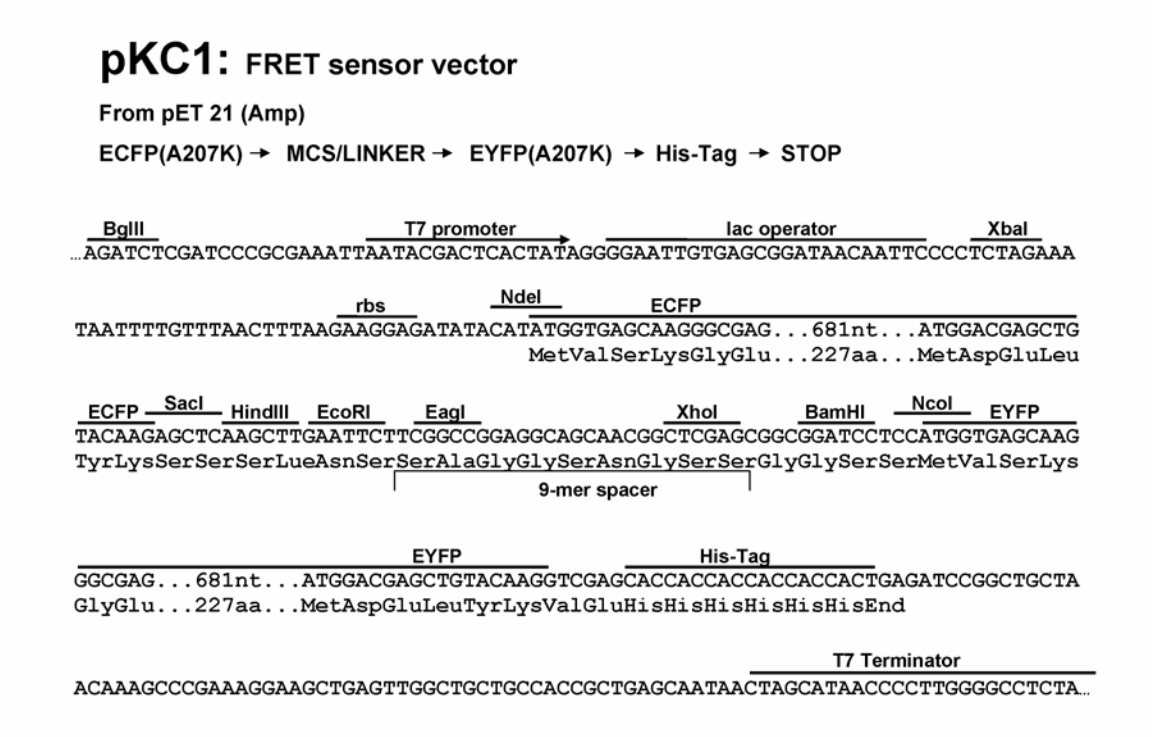
NusA was amplified directly from E. coli (XL10-gold), and MBP was amplified from pDW363C (Ja et al. 2005 JBC 280, 32057). For pAO6, NusA was amplified first with primers NusA5N1 (5'-CAT CAT CAC CAC GGT ACC AGC AAC AAA GAA ATT TTG GCT GTA GTT GAA GC) and NusA3N1 (5'-CTC GAG ACG GCC GTC GAT GGA GCT CGT CGA GCC CGC TTC GTC ACC GAA CCA GC). To create pAO7, MBP was first amplified with primers MBP5N1 (5'-CAT CAT CAC CAC GGT

ACC AGC AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG) and MBP3N1 (5’-CTC GAG ACG GCC GTC GAT GGA GCT CGT CGA GCC ATT AGT CTG CGC GTC TTT CAG G). The NusA and MBP fragments were both extended by amplification with primers SOLTAG5N (5’- GAG TAG GCT TCA CAT ATG CAT CAC CAT CAT CAC CAC GGT ACC AGC) and SOLTAG3N (5’-CAT TAC GTC GAC CTA GGA TCC ACT AGT AAG CTT CTC GAG ACG GCC GTC GAT GG). The final PCR products were digested with NdeI and SalI. We cloned both fragments into pAO1, which was digested with NdeI and XhoI to remove all components of the GFP reporter ORF. SalI and XhoI have compatible overhangs, and the result of both ligations is that the original vector XhoI site is eliminated while preserving the internal XhoI site encoded within primer SOLTAG3N.

To create pAO8, NusA was amplified first with NusA5C1 (5’-AC GAC GAT AAG GGG AGC TCC AAC AAA GAA ATT TTG GCT GTA GTT GAA GC and NusA3C (5’-GAT TAG CTC GAG GCC CGC TTC GTC ACC GAA CCA GC). In order to create pAO9, MBP was first amplified with primers MBP5C1 (5’-AC GAC GAT AAG GGG AGC TCC AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG) and MBP3C (5’-GAT TAG CTC GAG GCC ATT AGT CTG CGC GTC TTT CAG G). Both PCR products were extended further with SOLTAG5C (5’-CTA AGC GGA TCC GAT TAC AAG GAT GAC GAC GAT AAG GGG AGC TCC) and the same 3’ primers, NusA3C and MBP3C.

The final NusA and MBP PCR products were digested with BamHI and XhoI and were ligated into the same restriction sites in pAO1. Therefore, pAO8 and pAO9 utilize the same MCS that was used for pAO1 and replaces the liker-GFP sequence with FLAG-

NusA or FLAG-MBP. By cloning into XhoI as the 3’ restriction site, both vectors utilize a C-terminal His-tag present in the original pET vector.



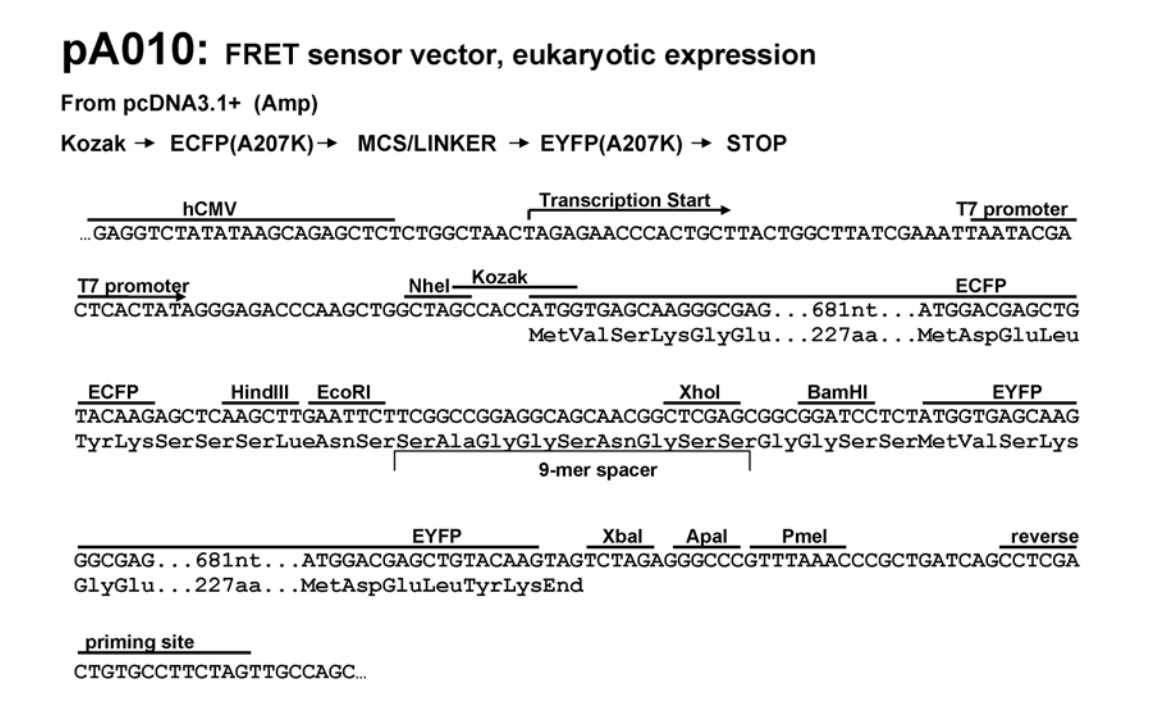

## Rationale

pKC1 was created to provide a versatile FRET sensor vector for cloning selection products (see Chapter 3).

## Experimental Procedures

We obtained ECFP and EYFP from Dr. Jose Aberola-Ila. The ECFP and EYFP clones contain the A207K mutation, which eliminates dimerization. ECFP was amplified in two steps with primers ECFP5NdeI (5’-GGA GAT ATA CAT ATG GTG AGC AAG GGC GAG) and ECFP3-1 (5’-CCT CCG GCC GAA GAA TTC AAG CTT GAG CTC TTG TAC AGC TCG TCC ATG CCG AG), followed by ECFP3-2 (5’-CAT GGA GGA TCC GCC GCT CGA GCC GTT GCT GCC TCC GGC CGA AGA ATT CAA G). EYFP was amplified with primers EYFP5BamHI (5’-AGC GGC GGA TCC TCC ATG GTG AGC

AAG GGC GAG G) and EYFP3SalI (5’-TCG TCG GTC GAC CTT GTA CAG CTC GTC CAT GCC CAG AG). ECFP was digested with NdeI and BamHI. EYFP was digested with BamHI and SalI. pET21a (Novagen) was digested with NdeI and XhoI. The three fragments were ligated together to produce pKC1. Cloning into the XhoI site of pET21 with SalI eliminates that restriction site while preserving an internal restriction site encoded by ECFP3-2. The ligation results in the utilization of an in-frame His-tag at the C-terminus that is encoded by the pET21 vector.




## Rationale

pAO10 was created to provide a convenient FRET sensor vector for expression in mammalian cell culture. It was designed to allow for simple shuttling of DNA from pKC1 by cutting and pasting from pKC1 into pAO10. Note that not all restriction sites in the pKC1 linker are available in pAO10 due to limitations of the host vector (pcDNA3.1+, Invitrogen). The reason why this is beneficial is that simply amplifying the entire ORF of pKC1 plasmids bearing functional inserts is that amplification may prove difficult as CFP and YFP have identical primer annealing sequences at the 5’ and 3’ ends. Also, digesting and ligating DNA from plasmid is more efficient than digesting and ligating DNA amplified by PCR.

## Experimental Procedures

ECFP was amplified in two steps with primers ECFP5NHE (5’-GGTG GTG CTA GCC ACC ATG GTG AG) and ECFP3-1, followed by ECFP3-2 (see pKC1, above). The ECFP PCR product was digested with NheI and BamHI and ligated into the same restriction sites in pcDNA3.1+. ECFP5NHE encodes a Kozak consensus sequence. EYFP was amplified with primers EYFP5BamHIm (5’-GGT GGT GGA TCC TCT ATG GTG AGC AAG GGC GAG G) and EYFP3XbaI (5’-GGT TGG TCT AGA CTA CTT GTA CAG CTC GTC CAT GCC CAG). EYFP5BamHIm is different from the primer used to create pKC1 in that a Kozak sequence was eliminated by synonymous mutation. The EYFP PCR product was digested with BamHI and XbaI and ligated into the pcDNA3.1+ vector bearing ECFP to create pAO10.