Appendix C. Cloning Schemes and Plasmid Maps

TABLE OF CONTENTS

Bacterial expression strains.......... C-2

Abbreviated aECM protein names.......... C-2

aECM chronology.......... C-3

DNA and amino acid sequences.......... C-8

  Kt1, Kt3, and Kt5.......... C-8

  CS5, SC5, CS5-5TFI, and SC5-5TFI.......... C-16

  CS5-ELF.......... C-23

  RGD and RDG.......... C-31

Bacterial Expression Strains

AI	Isoleucine auxtrophic *E. coli* derived from BL21(DE3)
	Genotype: AI (B F^{-} *ompT hsdS*(r_B^- m_B^-)*gal dcm* □(DE3) *ilvD691*)
	Reference: Wang P, Tang Y, Tirrell DA. JACS 2003. 125:6900:6906.
AF	Phenylalanine auxotrophic *E. coli* derived from BL21(DE3)
	Genotype: *HsdS gal* (□c*I*ts857 *ind* 1 *S*am7 *nin*5 *lac*UV5-T7 gene *1*) *pheA*
	Reference: Yoshikawa E, Fournier MJ, Mason TL, Tirrell DA. Macromolecules 1994. 27:5471-5475.
BL21(DE3)	Standard protein expression strain purchased from Novagen
	Genotype: *HsdS gal* (□c*I*ts857 *ind* 1 *S*am7 *nin*5 *lac*UV5-T7 gene *1*)

Abbreviated aECM Protein Names

Kt1	aECM variant with terminal lysines and 1 cell-binding domain
Kt3	aECM variant with terminal lysines and 3 cell-binding domains
Kt5	aECM variant with terminal lysines and 5 cell-binding domains
CS5	aECM variant containing El-IK and CS5 cassettes
SC5	aECM variant containing El-IK and negative control SC5 cassettes
CS5-5TFI	CS5 variant with 5TFI replacement of I
SC5-5TFI	SC5 variant with 5TFI replacement of I
CS5-ELF	aECM variant containing ELF and CS5 cassettes
pN3-CS5	aECM variant containing ELF and CS5 cassettes with pN$_3$ replacement of F
RGD	aECM variant containing El-IK and RGD cassettes
RDG	aECM variant containing El-IK and negative control RDG cassettes

Note: The above lists as well as the DNA sequences and plasmid maps included in Appendix C are limited to those expression strains and proteins that have been investigated in this thesis; many more aECM proteins have been constructed. To aid future cloning, all of the generations of aECM proteins, including those that were not directly studied in this thesis, are discussed in the section "aECM Chronology."

aECM Chronology

Generation 1:

Designed by: Alyssa Panitch
Cloned by: Alyssa Panitch
First reference: Panitch A, Yamaoka T, Fournier MJ, Mason TL, Tirrell DA. Macromolecules 1999. 32:1701-1703.
Cloning information: A Panitch, Thesis, University of Massachusetts, Amherst, 1997.

Notes: These proteins contain $(VPGIG)_x$ multimers and 20 amino acids of the CS5 cell-binding domain without crosslinking sites or fusion tags. Dr. Panitch created several aECM variants with different ratios of elastin-like domains to cell-binding domains. She also created proteins that were entirely composed of elastin-like repeats. The sequence published in the Macromolecules article and the thesis is incorrect. The incorporated elastin-like multimer is repeated 5, not 4 times. Therefore, the proteins contain $(VPGIG)_{25}$ instead of $(VPGIG)_{20}$. Due to their hydrophobicity, these proteins did not stain well.

Generation 2:

Designed by: Eric Welsh
Cloned by: Eric Welsh
First reference: Welsh ER, Tirrell DA. Biomacrmolecules 2000. 1:23-30.
Cloning information: above reference and ER Welsh, Thesis, University of Massachusetts, Amherst, 1999.

Notes: Dr. Welsh added a T7 tag to the constructs cloned by Dr. Panitch as well as lysine residues at the C- and N-termini for site-specific crosslinking. Because his constructs were made directly from Generation 1, their sequences have also been published incorrectly, i.e., $(VPGIG)_{20}$ should read $(VPGIG)_{25}$. These proteins are currently referred to as Kt1, Kt3, and Kt5 by the aECM group.

Generation 3:

Designed by:	Ralf Weberskirch and Kathleen Di Zio
Cloned by:	Ralf Weberskirch
First reference:	
Cloning information:	R Weberskirch, Final Report, Caltech, 2002.

Notes: Dr. Weberskirch prepared two polylinker regions for the pET28a vector. By his nomenclature, they are PL1 and PL2. PL1 contains a T7 tag, a His tag, an enterokinase cleavage site, and an *XhoI* restriction site for gene insertion. PL2 is similar but does not contain the T7 tag. In future aECM generations, PL1 is referred to as RW. Dr. Weberskirch also prepared aECM proteins that contain the CS5 and CS1 cell-binding domains and the elastin-like sequence $(VPGIG)_2VPGMG(VPGIG)_2$. He designated this cassette EL in his stocks, although stocks prepared by all others in the aECM group use EL to designate $(VPGIG)_5$.

Generation 4:

Designed by:	Kathleen Di Zio and Sarah Heilshorn
Cloned by:	Kathleen Di Zio and Sarah Heilshorn
First reference:	Di Zio K, Tirrell DA. Macromolecules 2003. 36:1553-1558.
Cloning information:	K Di Zio, Thesis, University of Massachusetts, Amherst, 2002.

Notes: Dr. Di Zio used the pET28aRW vector to create aECM constructs with the CS1 and CS5 cell-binding domains and the elastin-like sequence $(VPGIG)_2VPGKG(VPGIG)_2$. Ms. Heilshorn assisted in cloning of a negative control protein containing a scrambled CS5 cell-binding domain. These proteins are currently referred to as CS1, CS5, and SC5, respectively. Dr. Di Zio refers to this elastin-like cassette as El-IK.

Generation 5:

Designed by: Julie Liu and Sarah Heilshorn
Cloned by: Julie Liu
First reference: Liu JC, Heilshorn SC, Tirrell DA. Biomacromolecules 2004. 5:4970504.
Cloning information: J Liu, Research Notebook, Caltech.

Notes: Ms. Liu used the pET28aRW vector to create aECM constructs with the El-IK elastin-like sequence interspersed with the RGD cell-binding domain, a negative control scrambled RDG cell-binding domain, the PHSRN cell-binding synergy site, and a negative control scrambled NHRPS cell-binding synergy site. These proteins are currently referred to as RGD, RDG, PHSRN, and NHRPS.

Generation 6:

Designed by: Nandita Sharma
Cloned by: Nandita Sharma
First reference:
Cloning information: N Sharma, Thesis, University of Massachusetts, Amherst, 2002.

Notes: Dr. Sharma used the pET28aRW vector to create aECM constructs with CS1 and CS5 cell-binding domains and the elastin-like sequence (VPGVG)$_2$VPGFG(VPGVG)$_2$. These vectors also contain the *pheS** gene which encodes the A294G mutant of *PheRS*. Dr. Sharma refers to this elastin-like casette as ELF and the vector that contains both the aECM gene and the *pheS** gene as pNS-CS5-ELF.

Generation 7:

Designed by:	Charles Liu, Julie Liu, and Sarah Heilshorn
Cloned by:	Charles Liu
First reference:	Liu CY, Apuzzo MLJ, Tirrell DA. Neurosurgery 2003. 52:1154-1165.
Cloning information:	above reference.

Notes: Dr. Liu prepared a new polylinker region for the pET28a vector designated cyl. This polylinker contains a T7 tag, a His tag, a cyanogen bromide cleavage site, and an *XhoI* restriction site for aECM gene insertion. He constructed several aECM variants with the elastin-like cassette El-IK, which he refers to as ELIK, and various DSL signaling domains. This new cloning scheme allows cell-binding domains to be exchanged in the pEC2 vector without disrupting the elastin-like multimers.

Generation 8:

Designed by:	Marissa Mock and Sarah Heilshorn
Cloned by:	Marissa Mock
First reference:	
Cloning information:	S Heilshorn, Thesis, Caltech, 2004.

Notes: Ms. Mock removed the vectors pET28aRW[CS5(El-IK)$_5$]$_3$ and pET28aRW[SC5(El-IK)$_5$]$_3$ from the BL21(DE3) expression system and transfected them into the AI expression system.

Generation 9:

Designed by: Marissa Mock, Isaac Carrico, and Sarah Heilshorn

Cloned by: Marissa Mock and Stacey Maskarinec

First reference:

Cloning information: M Mock, Research Notebook, Caltech.

Notes: Ms. Mock is using the pET28a-cyl vector to construct aECM variants with the ELF elastin-like sequence and the SC5, RGD, and RDG cell-binding domains. These vectors will also contain the *pheS** gene which encodes the A294G mutant of *PheRS*.

Strain/Plasmid Collection

Submitted by	Sarah Heilshorn	**Date**	May 10, 2004

Strain name	BL21(DE3) pLysS/pET28apKt1 BL21(DE3) pLysS/pET28apKt3 BL21(DE3) pLysS/pET28apKt5
Vector (kb)	pET28ap-Kt1 or pET28ap-T7K[CS5(VPGIG)$_{25}$]$_1$KAAK (5.764) pET28ap-Kt3 or pET28ap-T7K[CS5(VPGIG)$_{25}$]$_3$KAAK (6.664) pET28ap-Kt5 or pET28ap-T7K[CS5(VPGIG)$_{25}$]$_5$KAAK (7.564)
Cloning site	See plasmid map.

Construction

These constructs were initially designed and cloned by Alyssa Panitch. Eric Welsh later added the T7 tag to the constructs as well as lysine residues at the C- and N-termini for site-specific crosslinking. According to current DNA sequencing results, these sequences were initially published incorrectly: (VPGIG)$_{20}$ should read (VPGIG)$_{25}$ in all cases. Detailed cloning schemes can be found in the references below.

References

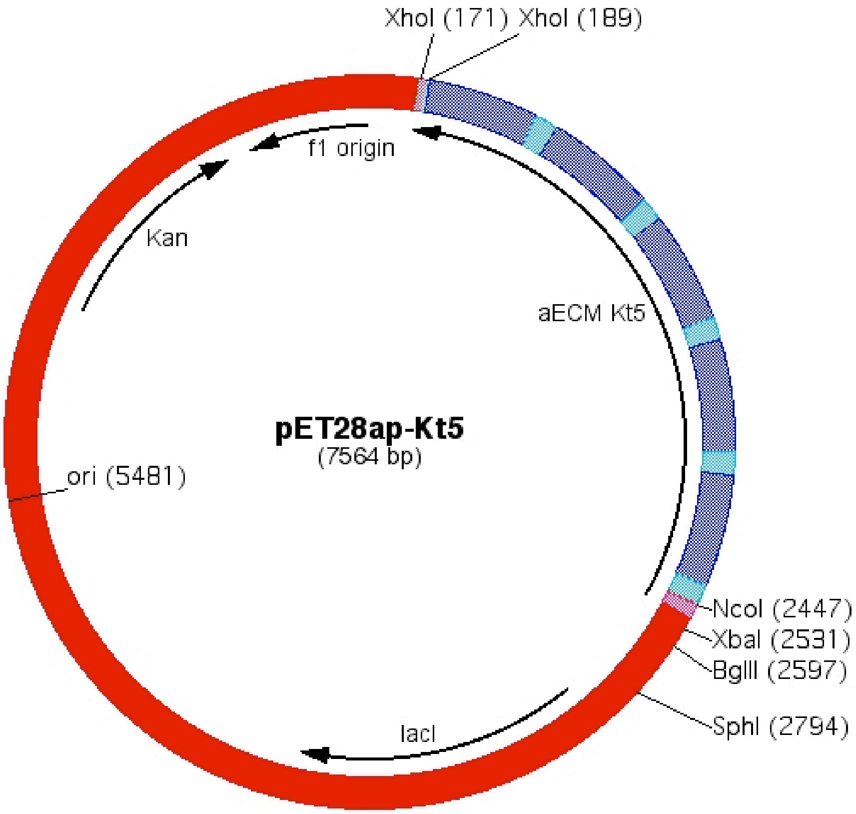
Panitch A, Yamaoka T, Fournier MJ, Mason TL, Tirrell DA. Macromolecules 1999. 32:1701-1703.
Panitch A, Thesis, University of Massachusetts, Amherst, 1997.
Welsh ER, Tirrell DA. Biomacrmolecules 2000. 1:23-30.
Welsh ER, Thesis, University of Massachusetts, Amherst, 1999.

Color codes for plasmid maps and DNA sequences

red	original pET28a vector
pink	T7 tag and C- and N-terminal lysines
green	ap polylinker designed by Alyssa Panitch
cyan	cell-binding domain
blue	elastin-like domain

Plasmid map

Only the plasmid map for pET28ap-Kt5 is shown. The other plasmid maps are similar.



Positions of elements in bases

	pET28ap-Kt1	pET28ap-Kt3	pET28ap-Kt5
Vector size	5764	6664	7564
Kan	4390-5202	5290-6102	6190-7002
f1 origin	5298-5753	6198-6653	7098-7553
lacI	1168-2247	2068-3147	2968-4047
ori	3681	4581	5481
aECM region	164-694	164-1594	164-2494

Culture conditions

Shake flask: 2xYT, induce with 2mM IPTG at OD 1
Fermentor: Terrific Broth, induce with 2.5 mM IPTG at OD 6-8

Full sequence   Only the sequence for pET28ap-Kt5 is shown. The others are similar.

```
  1  ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTA
     TAGGCCTATATCAAGGAGGAAAGTCGTTTTTTGGGGAGTTCTGGGCAAATCTCCGGGGTTCCCCAAT

 68  TGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGA
     ACGATCAATAACGAGTCGCCACCGTCGTCGGTTGAGTCGAAGGAAAGCCCGAAACAATCGTCGGCCT

                                           XhoI (171)              XhoI (189)
135  TCTCAGTGGTGGTGGTGGTGGTGCtcgac tta tta ctc gag ttt agc agc ttt ctc
     AGAGTCACCACCACCACCACCACGagctg aat aat gag ctc aaa tcg tcg aaa gag
                               777◀ ·   ·   E   L   K   A   A   K   E

191  gag cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg
     ctc gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc
768◀  L   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

242  tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc
     atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg
751◀  V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

293  gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc
     cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg
734◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

344  cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac
     gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg
717◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

395  acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat
     tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta
700◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

446  gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg
     cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc
683◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

497  cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc
     gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg
666◀  V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

548  gat gcc cgg aac gcc gat acc cgg cac ccc cgg gta cag gtg gta gtc aac
     cta cgg gcc ttg cgg cta tgg gcc gtg ggg gcc cat gtc cac cat cag ttg
649◀  I   G   P   V   G   I   G   P   V   G   P   Y   L   H   Y   D   V

599  gtc ttc acg cgg gat gtg gcc gat ctg gat ctc ttc acc gtc gag cgg cac
     cag aag tgc gcc cta cac cgg cta gac cta gag aag tgg cag ctc gcc gtg
632◀  D   E   R   P   I   H   G   I   Q   I   E   E   G   D   L   P   V

650  ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat
     ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta
615◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

701  gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg
     cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc
598◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

752  aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc
     ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg
581◀  V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

803  gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc
     cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg
564◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

854  cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac
     gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg
547◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V
```

```
  905 acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat
      tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta
 530◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

  956 acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg
      tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc
 513◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

 1007 aac gcc gat acc cgg cac ccc cgg gta cag gtg gta gtc aac gtc ttc acg
      ttg cgg cta tgg gcc gtg ggg gcc cat gtc cac cat cag ttg cag aag tgc
 496◀  V   G   I   G   P   V   G   P   Y   L   H   Y   D   V   D   E   R

 1058 cgg gat gtg gcc gat ctg gat ctc ttc acc gtc gag cgg cac ccc gat acc
      gcc cta cac cgg cta gac cta gag aag tgg cag ctc gcc gtg ggg cta tgg
 479◀  P   I   H   G   I   Q   I   E   E   G   D   L   P   V   G   I   G

 1109 cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac
      gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg
 462◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

 1160 gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat
      cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta
 445◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

 1211 gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg
      cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc
 428◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

 1262 aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc
      ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg
 411◀  V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

 1313 gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc
      cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg
 394◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

 1364 cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac
      gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg
 377◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

 1415 acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat
      tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta
 360◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

 1466 acc cgg cac ccc cgg gta cag gtg gta gtc aac gtc ttc acg cgg gat gtg
      tgg gcc gtg ggg gcc cat gtc cac cat cag ttg cag aag tgc gcc cta cac
 343◀  G   P   V   G   P   Y   L   H   Y   D   V   D   E   R   P   I   H

 1517 gcc gat ctg gat ctc ttc acc gtc gag cgg cac ccc gat acc cgg aac acc
      cgg cta gac cta gag aag tgg cag ctc gcc gtg ggg cta tgg gcc ttg tgg
 326◀  G   I   Q   I   E   E   G   D   L   P   V   G   I   G   P   V   G

 1568 gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc
      cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg
 309◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

 1619 cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac
      gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg
 292◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

 1670 acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat
      tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta
 275◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

 1721 gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg
      cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc
 258◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

 1772 cac ccc gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc
      gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg
 241◀  V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G
```

```
1823 gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc
     cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg
 224◄ I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

1874 cgg aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac
     gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg
 207◄ P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

1925 ccc cgg gta cag gtg gta gtc aac gtc ttc acg cgg gat gtg gcc gat ctg
     ggg gcc cat gtc cac cat cag ttg cag aag tgc gcc cta cac cgg cta gac
 190◄ G   P   Y   L   H   Y   D   V   D   E   R   P   I   H   G   I   Q

1976 gat ctc ttc acc gtc gag cgg cac ccc gat acc cgg aac acc gat gcc cgg
     cta gag aag tgg cag ctc gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc
 173◄ I   E   E   G   D   L   P   V   G   I   G   P   V   G   I   G   P

2027 aac acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc
     ttg tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg
 156◄ V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

2078 gat acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc
     cta tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg
 139◄ I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

2129 cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac
     gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg
 122◄ P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

2180 acc gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat
     tgg cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta
 105◄ G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

2231 acc cgg aac acc gat gcc cgg aac acc gat gcc cgg tac acc gat gcc cgg
     tgg gcc ttg tgg cta cgg gcc ttg tgg cta cgg gcc atg tgg cta cgg gcc
  88◄ G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

2282 aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc
     ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg
  71◄ V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

2333 gat gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc cgg gta
     cta cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg gcc cat
  54◄ I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   P   Y

2384 cag gtg gta gtc aac gtc ttc acg cgg gat gtg gcc gat ctg gat ctc ttc
     gtc cac cat cag ttg cag aag tgc gcc cta cac cgg cta gac cta gag aag
  37◄ L   H   Y   D   V   D   E   R   P   I   H   G   I   Q   I   E   E

                    NcoI  (2447)
2435 acc gtc gag gcc cat ggt ttt gcg acc cat ttg ctg tcc acc agt cat gct
     tgg cag ctc cgg gta cca aaa cgc tgg gta aac gac agg tgg tca gta cga
  20◄ G   D   L   G   M   T   K   R   G   M   Q   Q   G   G   T   M   S

                                                    XbaI  (2531)
2486 agc cat cat GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATC
     tcg gta gta CCATATAGAGGAAGAATTTCAATTTGTTTTAATAAAGATCTCCCCTTAACAATAG
   3◄ A   M   M

                                                   BglII  (2597)
2550 CGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCG
     GCGAGTGTTAAGGGGATATCACTCAGCATAATTAAAGCGCCCTAGCTCTAGAGCTAGGAGATGCGGC

2617 GACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCAC
     CTGCGTAGCACCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGCGGCTGTAGTG

2684 CGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCA
     GCTACCCCTTCTAGCCCGAGCGGTGAAGCCCGAGTACTCGCGAACAAAGCCGCACCCATACCACCGT

                                             SphI  (2794)
2751 GGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGC
     CCGGGGCACCGGCCCCCTGACAACCCGCGGTAGAGGAACGTACGTGGTAAGGAACGCCGCCGCCACG

2818 TCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGA
     AGTTGCCGGAGTTGGATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTCT
```

```
2885  TCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCA
      AGGGCCTGTGGTAGCTTACCGCGTTTTGGAAAGCGCCATACCGTACTATCGCGGGCCTTCTCTCAGT
2952  ATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTAT
      TAAGTCCCACCACTTACACTTTGGTCATTGCAATATGCTACAGCGTCTCATACGGCCACAGAGAATA
3019  CAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAG
      GTCTGGCAAAGGGCGCACCACTTGGTCCGGTCGGTGCAAAGACGCTTTTGCGCCCTTTTTCACCTTC
3086  CGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTT
      GCCGCTACCGCCTCGACTTAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAA
3153  GCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAA
      CGACTAACCGCAACGGTGGAGGTCAGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTT
3220  TCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCT
      AGAGCGCGGCTAGTTGACCCACGGTCGCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGA
3287  GTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGA
      CATTTCGCCGCCACGTGTTAGAAGAGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCT
3354  TGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCT
      ACTGGTCCTACGGTAACGACACCTTCGACGGACGTGATTACAAGGCCGCAATAAAGAACTACAGAGA
3421  GACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATC
      CTGGTCTGTGGGTAGTTGTCATAATAAAAGAGGGTACTTCTGCCATGCGCTGACCCGCACCTCGTAG
3488  TGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCT
      ACCAGCGTAACCCAGTGGTCGTTTAGCGCGACAATCGCCCGGGTAATTCAAGACAGAGCCGCGCAGA
3555  GCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGC
      CGCAGACCGACCGACCGTATTTATAGAGTGAGCGTTAGTTTAAGTCGGCTATCGCCTTGCCCTTCCG
3622  GACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTG
      CTGACCTCACGGTACAGGCCAAAAGTTGTTTGGTACGTTTACGACTTACTCCCGTAGCAAGGGTGAC
3689  CGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCG
      GCTACGACCAACGGTTGCTAGTCTACCGCGACCCGCGTTACGCGCGGTAATGGCTCAGGCCCGACGC
3756  CGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCG
      GCAACCACGCCTATAGAGCCATCACCCTATGCTGCTATGGCTTCTGTCGAGTACAATATAGGGCGGC
3823  TTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCT
      AATTGGTGGTAGTTTGTCCTAAAAGCGGACGACCCCGTTTGGTCGCACCTGGCGAACGACGTTGAGA
3890  CTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCT
      GAGTCCCGGTCCGCCACTTCCCGTTAGTCGACAACGGGCAGAGTGACCACTTTTCTTTTTGGTGGGA
3957  GGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAG
      CCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTC
4024  GTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCA
      CAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACATTCAATCGAGTGAGTAATCCGT
4091  CCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCA
      GGCCCTAGAGCTGGCTACGGGAACTCTCGGAAGTTGGGTCAGTCGAGGAAGGCCACCCGCGCCCCGT
4158  TGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGC
      ACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTGTCCACGGCCGTCG
4225  GCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCG
      CGAGACCCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCGCTGCTACTAGCCGGACAGCGAACGC
4292  GTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCG
      CATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGGTTTGCAAAGCCGC
4359  AGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAG
      TCTTCGTCCGGTAATAGCGGCCGTACCGCCGGGGTGCCCACGCGTACTAGCACGAGGACAGCAACTC
4426  GACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACG
      CTGGGCCGATCCGACCGCCCCAACGGAATGACCAATCGTCTTACTTAGTGGCTATGCGCTCGCTTGC
4493  TGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTG
      ACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTGTACTTACCAGAAGCCAAAGGCAC
4560  TTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGG
      AAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATACAAGGCCTAGACGTAGCGTCC
4627  ATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTG
      TACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCGCGACCGTAACTGGGACTCAC
```

```
4694  ATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGG
      TAAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGTTGCAAGGTCATTGGCCC

4761  CATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCAT
      GTACAAGTAGTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCCATAGTAATGGGGGTA

4828  GAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCC
      CTTGTCTTTAGGGGGAATGTGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCGGGAATTGTACCGG

4895  CGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGG
      GCGAAATAGTCTTCGGTCTGTAATTGCGAAGACCTCTTTGAGTTGCTCGACCTGCGCCTACTTGTCC

4962  CAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGG
      GTCTGTAGACACTTAGCGAAGTGCTGGTGCGACTACTCGAAATGGCGTCGACGGAGCGCGCAAAGCC

5029  TGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGAT
      ACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTA

5096  GCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGA
      CGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCGTCGGTACT

5163  CCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTG
      GGGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGAC

5230  AGAGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCG
      TCTCACGTGGTATATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGC

5297  CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCT
      GAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGA

5364  CACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
      GTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTT

5431  AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
      TTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGG

5498  CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGA
      GGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCT

5565  TACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
      ATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTA

5632  ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
      TGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTC

5699  TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
      AAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACG

5766  GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAG
      CGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTC

5833  CCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCC
      GGTGACCATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGG

5900  TAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
      ATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCT

5967  AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
      TTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGT

6034  AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
      TCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACT

6101  CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTACAT
      GCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTTGTTATTTTGACAGACGAATGTA

6168  AAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTA
      TTTGTCATTATGTTCCCCACAATACTCGGTATAAGTTGCCCTTTGCAGAACGAGATCCGGCGCTAAT

6235  AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTG
      TTAAGGTTGTACCTACGACTAAATATACCCATATTTACCCGAGCGCTATTACAGCCCGTTAGTCCAC

6302  CGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAG
      GCTGTTAGATAGCTAACATACCCTTCGGGCTACGCGGTCTCAACAAAGACTTTGTACCGTTTCCATC

6369  CGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCG
      GCAACGGTTACTACAATGTCTACTCTACCAGTCTGATTTGACCGACTGCCTTAAATACGGAGAAGGC

6436  ACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAA
      TGGTAGTTCGTAAAATAGGCATGAGGACTACTACGTACCAATGAGTGGTGACGCTAGGGGCCCTTTT
```

```
6503  CAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTT
      GTCGTAAGGTCCATAATCTTCTTATAGGACTAAGTCCACTTTTATAACAACTACGCGACCGTCACAA

6570  CCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTC
      GGACGCGGCCAACGTAAGCTAAGGACAAACATTAACAGGAAAATTGTCGCTAGCGCATAAAGCAGAG

6637  GCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATG
      CGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTACGCTCACTAAAACTACTGCTCGCATTAC

6704  GCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGT
      CGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTTGAAAACGGTAAGAGTGGCCTAAGTCAGCA

6771  CACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGAT
      GTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTCCCCTTTAATTATCCAACATAACTA

6838  GTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGT
      CAACCTGCTCAGCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACCTTGACGGAGCCACTCA

6905  TTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATT
      AAAGAGGAAGTAATGTCTTTGCCGAAAAAGTTTTTATACCATAACTATTAGGACTATACTTATTTAA

6972  GCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAATG
      CGTCAAAGTAAACTACGAGCTACTCAAAAAGATTCTTAATTAAGTACTCGCCTATGTATAAACTTAC

7039  TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATTGTA
      ATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTTTAACAT

7106  AACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGG
      TTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCC

7173  CCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGT
      GGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCA

7240  TTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAG
      AACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTC

7307  GGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAC
      CCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTG

7374  TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAG
      ATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTC

7441  AAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC
      TTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCG

7508  GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA
      CATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGGGTAAGCGGT
```

Strain/Plasmid Collection

Submitted by	Sarah Heilshorn	**Date**	May 10, 2004

Strain name
BL21(DE3) pLysS/pET28aRW-CS5
BL21(DE3) pLysS/pET28aRW-SC5
AI pET28aRW-CS5
AI pET28aRW-SC5

Vector (kb)
pET28aRW-CS5 or pET28aRW-$[CS5(El\text{-}IK)_4]_3$ (6.443)
pET28aRW-SC5 or pET28aRW-$[CS5(El\text{-}IK)_4]_3$ (6.443)

Cloning site See plasmid map.

Construction

Ralf Weberskirch designed the pET28aRW vector which replaces the cloning region of the pET28a vector with a new polylinker region that includes a T7 tag, His tag, enterokinase cleavage site, *XhoI* restriction site, and stop codon. Kathleen Di Zio used the pET28aRW vector to create an aECM construct with the CS5 cell-binding domain and the elastin-like sequence $(VPGIG)_2VPGKG(VPGIG)_2$, which is referred to as El-IK. The SC5 domain is identical to the CS5 cell-binding domain, except the codons for REDV have been scrambled to encode RE<u>VD</u> as a negative control for cell culture experiments. Details on the cloning scheme can be found in the references below. Marissa Mock transfected these plasmids into the AI *E. coli* strain, which is isoleucine auxotrophic.

References

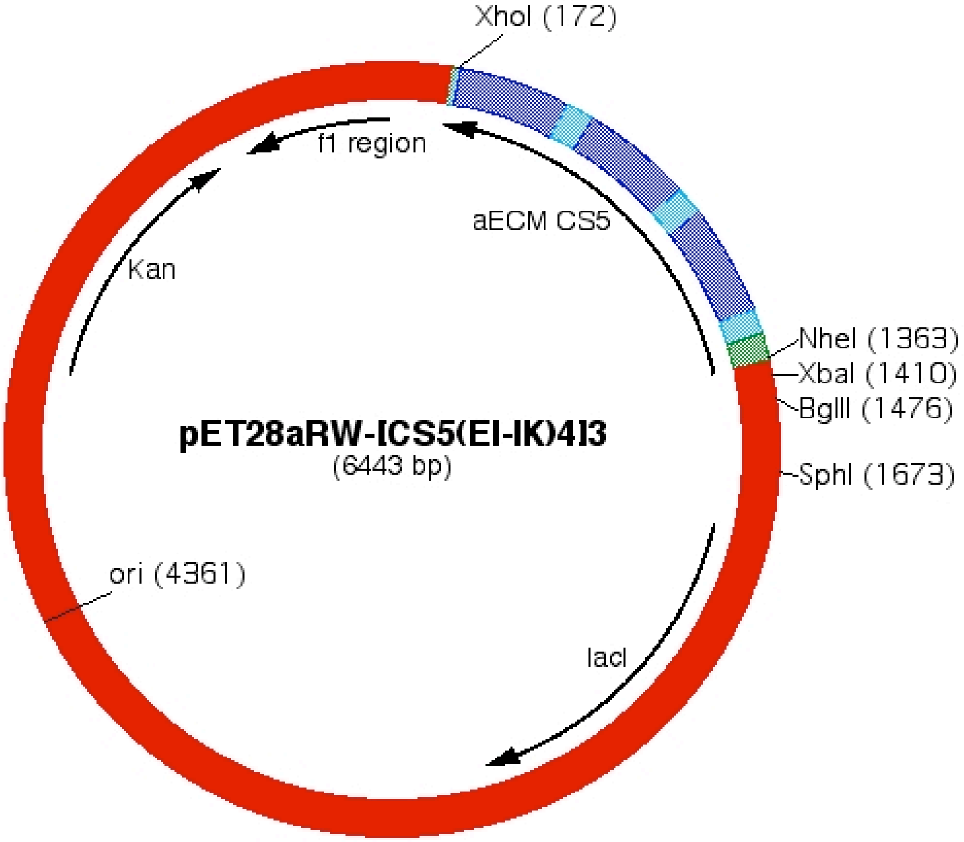
Di Zio K, Tirrell DA. Macromolecules 2003. 36:1553-1558.
Di Zio K, Thesis, University of Massachusetts, Amherst, 2002.

Color codes for plasmid maps and DNA sequences

red	original pET28a vector
green	RW polylinker designed by Ralf Weberskirch
cyan	cell-binding domain
blue	elastin-like domain

Plasmid map

Only the plasmid map for pET28aRW-CS5 is shown. The SC5 plasmid map is similar.



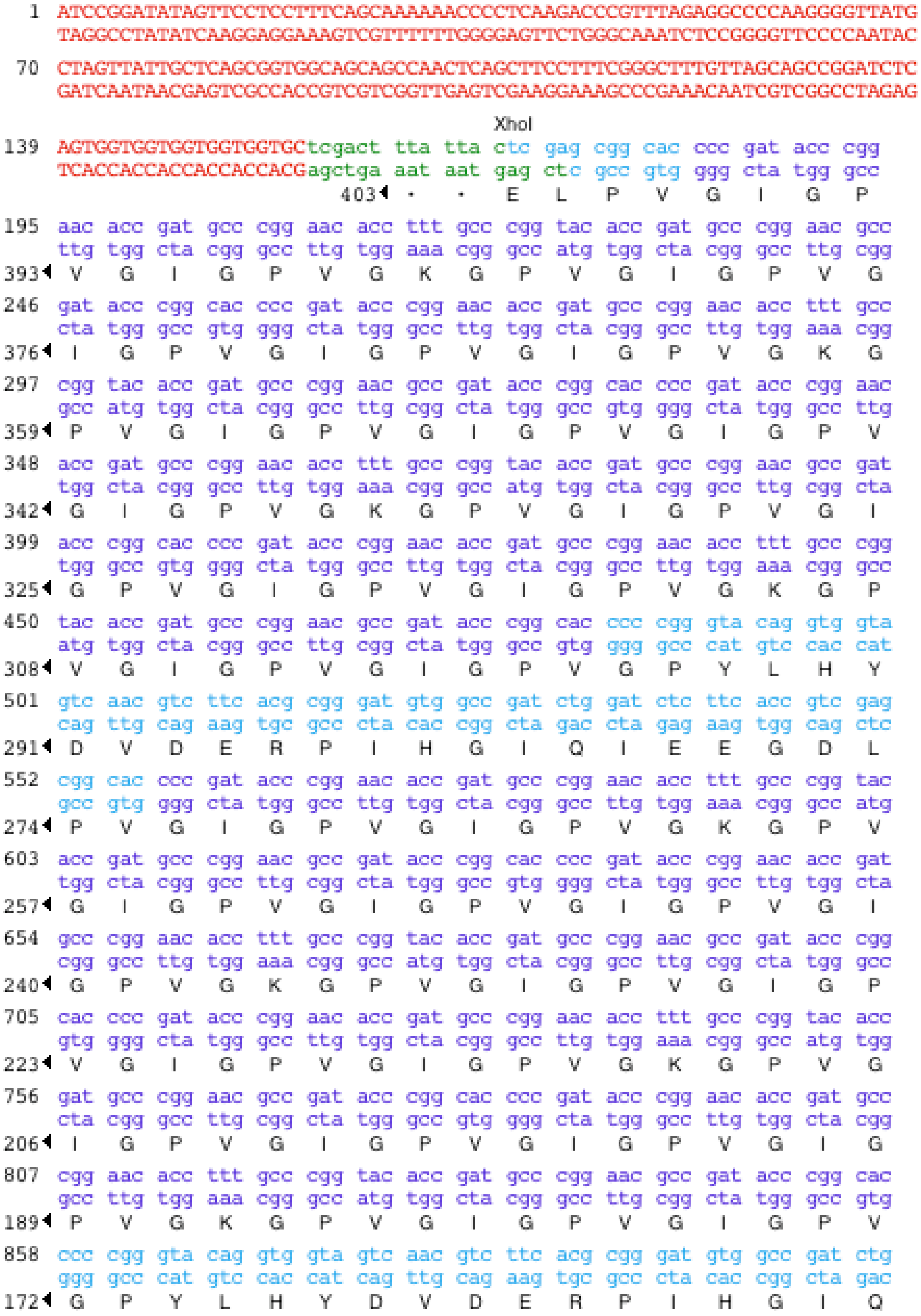
Positions of elements in bases

	pET28aRW-CS5	pET28aRW-SC5
Vector size	6443	6443
Kan	5070-5882	5070-5882
f1 origin	5978-6433	5978-6433
lacI	1848-2927	1848-2927
ori	4361	4361
aECM region	165-1373	165-1373

Culture conditions

Shake flask:	2xYT, induce with 2mM IPTG at OD 1
Fermentor:	Terrific Broth, induce with 2.5 mM IPTG at OD 6-8

<u>Full sequence</u> Only the sequence for pET28aRW-CS5 is shown. The SC5 sequence is identical except for the shuffling of the REDV codons within the CS5 cell-binding domain to encode REVD in the SC5 construct.



```
  1  ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATG
     TAGGCCTATATCAAGGAGGAAAGTCGTTTTTTGGGGAGTTCTGGGCAAATCTCCGGGGTTCCCCAATAC

 70  CTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTC
     GATCAATAACGAGTCGCCACCGTCGTCGGTTGAGTCGAAGGAAAGCCCGAAACAATCGTCGGCCTAGAG

                                        XhoI
139  AGTGGTGGTGGTGGTGGTGCtcgact tta tta ctc gag cgg cac ccc gat acc cgg
     TCACCACCACCACCACCACGagctga aat aat gag ctc gcc gtg ggg cta tgg gcc
                            403◀ ·   ·   E   L   P   V   G   I   G   P

195  aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc
     ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg
393◀  V   G   I   G   P   V   G   K   G   P   V   G   I   G   P   V   G

246  gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc
     cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg
376◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   K   G

297  cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac
     gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg
359◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

348  acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat
     tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta
342◀  G   I   G   P   V   G   K   G   P   V   G   I   G   P   V   G   I

399  acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg
     tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc
325◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   K   G   P

450  tac acc gat gcc cgg aac gcc gat acc cgg cac ccc cgg gta cag gtg gta
     atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg gcc cat gtc cac cat
308◀  V   G   I   G   P   V   G   I   G   P   V   G   P   Y   L   H   Y

501  gtc aac gtc ttc acg cgg gat gtg gcc gat ctg gat ctc ttc acc gtc gag
     cag ttg cag aag tgc gcc cta cac cgg cta gac cta gag aag tgg cag ctc
291◀  D   V   D   E   R   P   I   H   G   I   Q   I   E   E   G   D   L

552  cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac
     gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg
274◀  P   V   G   I   G   P   V   G   I   G   P   V   G   K   G   P   V

603  acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat
     tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta
257◀  G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

654  gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg
     cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc
240◀  G   P   V   G   K   G   P   V   G   I   G   P   V   G   I   G   P

705  cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc
     gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg
223◀  V   G   I   G   P   V   G   I   G   P   V   G   K   G   P   V   G

756  gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc
     cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg
206◀  I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

807  cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac
     gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg
189◀  P   V   G   K   G   P   V   G   I   G   P   V   G   I   G   P   V

858  ccc cgg gta cag gtg gta gtc aac gtc ttc acg cgg gat gtg gcc gat ctg
     ggg gcc cat gtc cac cat cag ttg cag aag tgc gcc cta cac cgg cta gac
172◀  G   P   Y   L   H   Y   D   V   D   E   R   P   I   H   G   I   Q
```

```
 909 gat ctc ttc acc gtc gag cgg cac ccc gat acc cgg aac acc gat gcc cgg
     cta gag aag tgg cag ctc gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc
155◀  I   E   E   G   D   L   P   V   G   I   G   P   V   G   I   G   P

 960 aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc
     ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg
138◀  V   G   K   G   P   V   G   I   G   P   V   G   I   G   P   V   G

1011 gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc
     cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg
121◀  I   G   P   V   G   I   G   P   V   G   K   G   P   V   G   I   G

1062 cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac
     gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg
104◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

1113 acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat
     tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta
 87◀  G   K   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I

1164 acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg
     tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc
 70◀  G   P   V   G   I   G   P   V   G   K   G   P   V   G   I   G   P

1215 aac gcc gat acc cgg cac ccc cgg gta cag gtg gta gtc aac gtc ttc acg
     ttg cgg cta tgg gcc gtg ggg gcc cat gtc cac cat cag ttg cag aag tgc
 53◀  V   G   I   G   P   V   G   P   Y   L   H   Y   D   V   D   E   R

1266 cgg gat gtg gcc gat ctg gat ctc ttc acc gtc gag ttt atc atc atc atc
     gcc cta cac cgg cta gac cta gag aag tgg cag ctc aaa tag tag tag tag
 36◀  P   I   H   G   I   Q   I   E   E   G   D   L   K   D   D   D   D

                                                                 NheI
1317 gtg gtg gtg gtg gtg gtg gtg acc cat ttg ctg tcc acc agt cat gct agc
     cac cac cac cac cac cac cac tgg gta aac gac agg tgg tca gta cga tcg
 19◀  H   H   H   H   H   H   H   G   M   Q   Q   G   G   T   M   S   A

                                                 XbaI
1368 cat cat GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCA
     gta gta CCATATAGAGGAAGAATTTCAATTTGTTTTAATAAAGATCTCCCCTTAACAATAGGCGAGT
  2◀  M   M

                                             BglII
1435 CAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATC
     GTTAAGGGGATATCACTCAGCATAATTAAAGCGCCCTAGCTCTAGAGCTAGGAGATGCGGCCTGCGTAG

1504 GTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAA
     CACCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGCGGCTGTAGTGGCTACCCCTT

1573 GATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCC
     CTAGCCCGAGCGGTGAAGCCCGAGTACTCGCGAACAAAGCCGCACCCATACCACCGTCCGGGGCACCGG

                               SphI
1642 GGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAAC
     CCCCCTGACAACCCGCGGTAGAGGAACGTACGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGAGTTG

1711 CTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCATCGA
     GATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTCTAGGGCCTGTGGTAGCT

1780 ATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGT
     TACCGCGTTTTGGAAAGCGCCATACCGTACTATCGCGGGCCTTCTCTCAGTTAAGTCCCACCACTTACA

1849 GAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGT
     CTTTGGTCATTGCAATATGCTACAGCGTCTCATACGGCCACAGAGAATAGTCTGGCAAAGGGCGCACCA

1918 GAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTA
     CTTGGTCCGGTCGGTGCAAAGACGCTTTTGCGCCCTTTTTCACCTTCGCCGCTACCGCCTCGACTTAAT

1987 CATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAG
     GTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAACGACTAACCGCAACGGTGGAGGTC

2056 TCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAG
     AGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTTAGAGCGCGGCTAGTTGACCCACGGTC

2125 CGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGC
     GCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGACATTTCGCCGCCACGTGTTAGAAGAGCG
```

```
2194 GCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGC
     CGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCTACTGGTCCTACGGTAACGACACCTTCGACG
2263 CTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTC
     GACGTGATTACAAGGCCGCAATAAAGAACTACAGAGACTGGTCTGTGGGTAGTTGTCATAATAAAAGAG
2332 CCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTT
     GGTACTTCTGCCATGCGCTGACCCGCACCTCGTAGACCAGCGTAACCCAGTGGTCGTTTAGCGCGACAA
2401 AGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAA
     TCGCCCGGGTAATTCAAGACAGAGCCGCGCAGACGCAGACCGACCGACCGTATTTATAGAGTGAGCGTT
2470 TCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCA
     AGTTTAAGTCGGCTATCGCCTTGCCCTTCCGCTGACCTCACGGTACAGGCCAAAAGTTGTTTGGTACGT
2539 AATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAAT
     TTACGACTTACTCCCGTAGCAAGGGTGACGCTACGACCAACGGTTGCTAGTCTACCGCGACCCGCGTTA
2608 GCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGA
     CGCGCGGTAATGGCTCAGGCCCGACGCGCAACCACGCCTATAGAGCCATCACCCTATGCTGCTATGGCT
2677 AGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAG
     TCTGTCGAGTACAATATAGGGCGGCAATTGGTGGTAGTTTGTCCTAAAAGCGGACGACCCCGTTTGGTC
2746 CGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACT
     GCACCTGGCGAACGACGTTGAGAGAGTCCCGGTCCGCCACTTCCCGTTAGTCGACAACGGGCAGAGTGA
2815 GGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATT
     CCACTTTTCTTTTTGGTGGGACCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAA
2884 AATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTT
     TTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACATTCAA
2953 AGCTCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCC
     TCGAGTGAGTAATCCGTGGCCCTAGAGCTGGCTACGGGAACTCTCGGAAGTTGGGTCAGTCGAGGAAGG
3022 GGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGAC
     CCACCCGCGCCCCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTG
3091 AGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCC
     TCCACGGCCGTCGCGAGACCCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCGCTGCTACTAGCCGG
3160 TGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAAC
     ACAGCGAACGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGGTTTG
3229 GTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGT
     CAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGGGGTGCCCACGCGTACTAGCACGAGGACA
3298 CGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGC
     GCAACTCCTGGGCCGATCCGACCGCCCCAACGGAATGACCAATCGTCTTACTTAGTGGCTATGCGCTCG
3367 GAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCC
     CTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTGTACTTACCAGAAGCCAAAGG
3436 GTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAG
     CACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATACAAGGCCTAGACGTAGCGTC
3505 GATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGA
     CTACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCGCGACCGTAACTGGGACTCACT
3574 TTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCAT
     AAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGTTGCAAGGTCATTGGCCCGTA
3643 GTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACA
     CAAGTAGTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCCATAGTAATGGGGGTACTTGT
3712 GAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTA
     CTTTAGGGGGAATGTGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCGGGAATTGTACCGGGCGAAAT
3781 TCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCT
     AGTCTTCGGTCTGTAATTGCGAAGACCTCTTTGAGTTGCTCGACCTGCGCCTACTTGTCCGTCTGTAGA
3850 GTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTG
     CACTTAGCGAAGTGCTGGTGCGACTACTCGAAATGGCGTCGACGGAGCGCGCAAAGCCACTACTGCCAC
3919 AAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGAC
     TTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTG
3988 AAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCG
     TTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCGTCGGTACTGGGTCAGTGCATCGC
```

```
4057  ATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATATG
      TATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATATAC

4126  CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTC
      GCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCGAGAAGGCGAAGGAGCGAG

4195  ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
      TGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCC

4264  TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
      AATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTG

4333  CGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
      GCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCT

4402  CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC
      GCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGG

4471  CTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
      GAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCG

4540  GTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGC
      CACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCG

4609  TGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
      ACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTG

4678  CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
      GGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACAT

4747  GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATC
      CCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAG

4816  TGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
      ACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

4885  GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
      CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTA

4954  CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATG
      GGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTAC

5023  AACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAAC
      TTGTTATTTTGACAGACGAATGTATTTGTCATTATGTTCCCCACAATACTCGGTATAAGTTGCCCTTTG

5092  GTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGA
      CAGAACGAGATCCGGCGCTAATTTAAGGTTGTACCTACGACTAAATATACCCATATTTACCCGAGCGCT

5161  TAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCT
      ATTACAGCCCGTTAGTCCACGCTGTTAGATAGCTAACATACCCTTCGGGCTACGCGGTCTCAACAAAGA

5230  GAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGA
      CTTTGTACCGTTTCCATCGCAACGGTTACTACAATGTCTACTCTACCAGTCTGATTTGACCGACTGCCT

5299  ATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGC
      TAAATACGGAGAAGGCTGGTAGTTCGTAAAATAGGCATGAGGACTACTACGTACCAATGAGTGGTGACG

5368  GATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGC
      CTAGGGGCCCTTTTGTCGTAAGGTCCATAATCTTCTTATAGGACTAAGTCCACTTTTATAACAACTACG

5437  GCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGT
      CGACCGTCACAAGGACGCGGCCAACGTAAGCTAAGGACAAACATTAACAGGAAAATTGTCGCTAGCGCA

5506  ATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGA
      TAAAGCAGAGCGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTACGCTCACTAAAACTACTGCT

5575  GCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTC
      CGCATTACCGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTTGAAAACGGTAAGAGTGGCCTAAG

5644  AGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTAT
      TCAGCAGTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTCCCCTTTAATTATCCAACATA

5713  TGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGA
      ACTACAACCTGCTCAGCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACCTTGACGGAGCCACT

5782  GTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATT
      CAAAAGAGGAAGTAATGTCTTTGCCGAAAAAGTTTTTATACCATAACTATTAGGACTATACTTATTTAA

5851  GCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAATGTA
      CGTCAAAGTAAACTACGAGCTACTCAAAAAGATTCTTAATTAAGTACTCGCCTATGTATAAACTTACAT
```

```
5920  TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATTGTAAACG
      AAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTTTAACATTTGC

5989  TTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAA
      AATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTT

6058  TCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACA
      AGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGT

6127  AGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCC
      TCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGG

6196  CACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACC
      GTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGG

6265  CTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGA
      GATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCT

6334  AAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCG
      TTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGC

6403  CCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA
      GGCGCGAATTACGCGGCGATGTCCCGCGCAGGGTAAGCGGT
```

Strain/Plasmid Collection

Submitted by Sarah Heilshorn **Date** May 10, 2004

Strain name AF pNS-CS5-ELF

Vector (kb) pET28aRW-[CS5(ELF)$_5$]$_3$-*pheS** or pNS-CS5-ELF

Cloning site See plasmid map.

Construction

Ralf Weberskirch designed the pET28aRW vector which replaces the cloning region of the pET28a vector with a new polylinker region that includes a T7 tag, His tag, enterokinase cleavage site, *XhoI* restriction site, and stop codon. Nandita Sharma used the pET28aRW vector to create an aECM construct with the CS5 cell-binding domain and the elastin-like sequence (VPGIG)$_2$VPGFG(VPGIG)$_2$, which is referred to as ELF. PCR was used to amplify a region of the pKSS vector (Kast, 1994) that includes the gene encoding the A294G mutant for *E. coli PheRS* ▯-subunit. This PCR product was inserted into pUC19 at the *SmaI* site to create plasmid pUC-FS and subcloned into pET28aRW-[CS5(ELF)$_5$]$_3$ at the *SphI* site. This plasmid was transfected into the AF *E. coli* strain, which is phenylalanine auxotrophic. Details can be found in the references below.

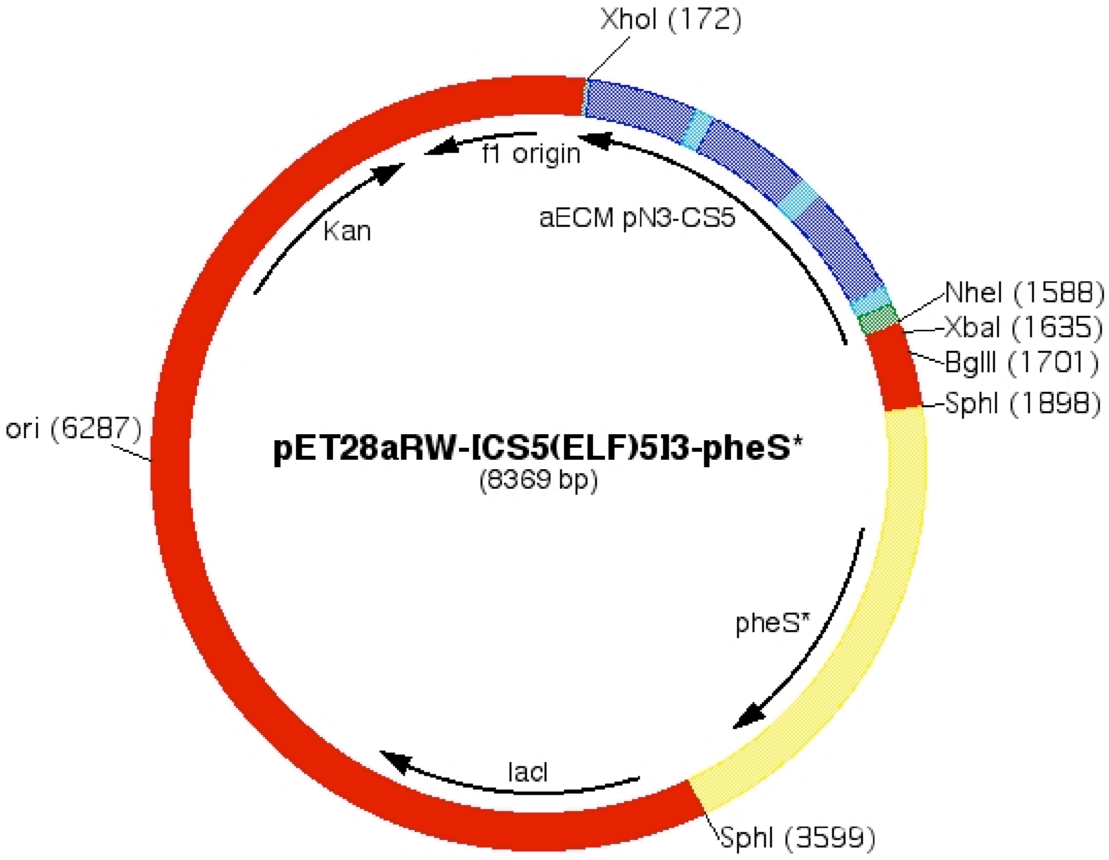
References

Kast P, Gene 1994. 138:109-114.
Sharma N, Thesis, University of Massachusetts, Amherst, 2002.

Color codes for plasmid maps and DNA sequences

red	original pET28a vector
green	RW polylinker designed by Ralf Weberskirch
cyan	cell-binding domain
blue	elastin-like domain
yellow	*pheS** gene subcloned from pUC-FS at *SphI* sites

Plasmid map



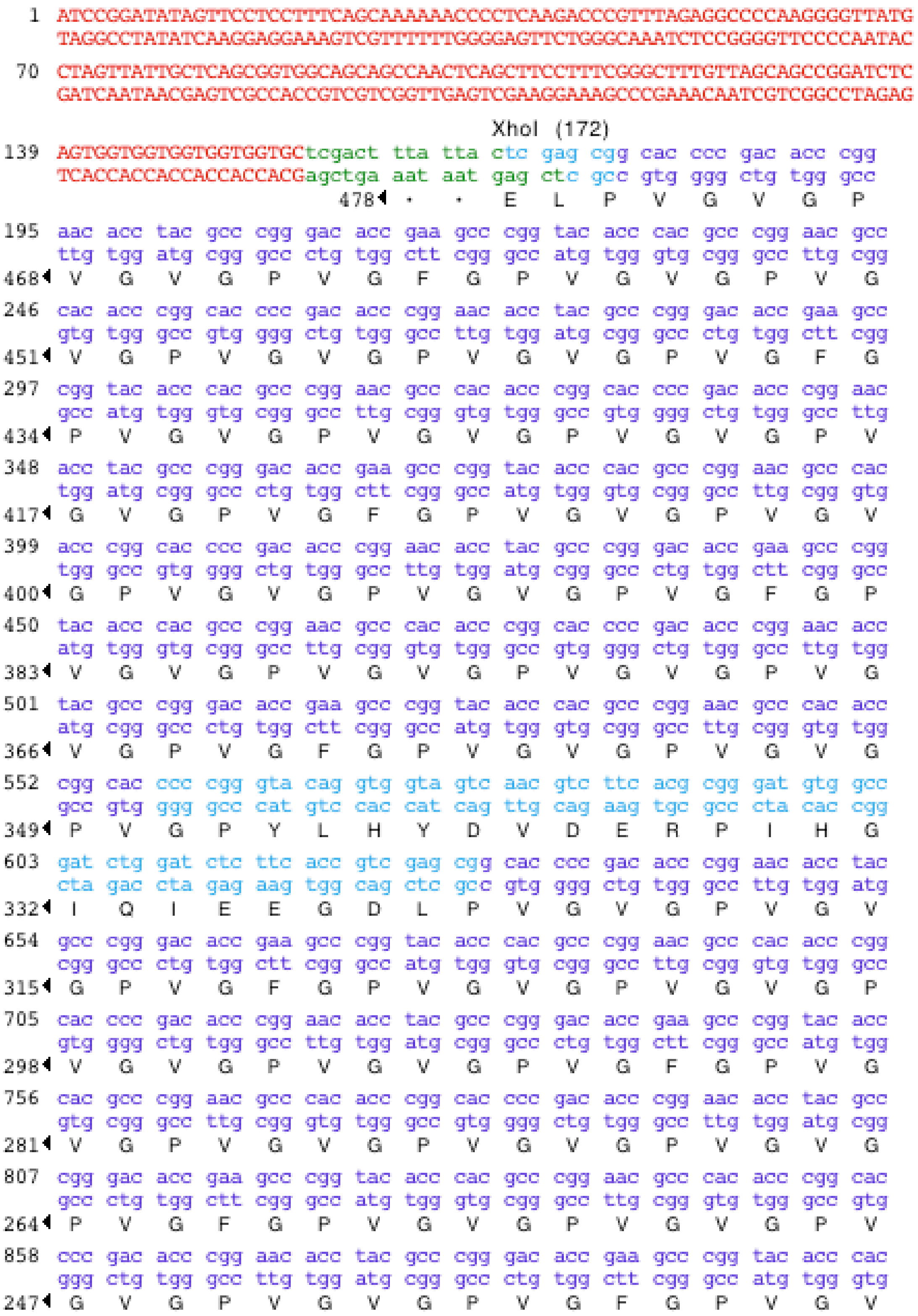
Positions of elements in bases

	pET28aRW-[CS5(ELF)$_5$]$_3$-pheS*
Vector size	8369
Kan	6996-7808
f1 origin	7904-8359
lacI	3774-4853
ori	6287
aECM region	165-1598
*pheS**	2358-3341

Culture conditions

For incorporation of noncanonical amino acids, grow in M9AA supplemented with all 20 amino acids. Cells can either be harvested and shifted into M9 with 19 natural amino acids + the noncanonical amino acid or grown until Phe is depleted from the original M9AA and supplementing with the noncanonical amino acid without a media shift.

Full sequence



```
  1 ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATG
    TAGGCCTATATCAAGGAGGAAAGTCGTTTTTTGGGGAGTTCTGGGCAAATCTCCGGGGTTCCCCAATAC

 70 CTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTC
    GATCAATAACGAGTCGCCACCGTCGTCGGTTGAGTCGAAGGAAAGCCCGAAACAATCGTCGGCCTAGAG

                                            XhoI  (172)
139 AGTGGTGGTGGTGGTGGTGCtcgact tta tta ctc gag cgg cac ccc gac acc cgg
    TCACCACCACCACCACCACGagctga aat aat gag ctc gcc gtg ggg ctg tgg gcc
                          478◀ ·   ·   E   L   P   V   G   V   G   P

195 aac acc tac gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc
    ttg tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg
468◀ V   G   V   G   P   V   G   F   G   P   V   G   V   G   P   V   G

246 cac acc cgg cac ccc gac acc cgg aac acc tac gcc cgg gac acc gaa gcc
    gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg ctt cgg
451◀ V   G   P   V   G   V   G   P   V   G   V   G   P   V   G   F   G

297 cgg tac acc cac gcc cgg aac gcc cac acc cgg cac ccc gac acc cgg aac
    gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg
434◀ P   V   G   V   G   P   V   G   V   G   P   V   G   V   G   P   V

348 acc tac gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac
    tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg
417◀ G   V   G   P   V   G   F   G   P   V   G   V   G   P   V   G   V

399 acc cgg cac ccc gac acc cgg aac acc tac gcc cgg gac acc gaa gcc cgg
    tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg ctt cgg gcc
400◀ G   P   V   G   V   G   P   V   G   V   G   P   V   G   F   G   P

450 tac acc cac gcc cgg aac gcc cac acc cgg cac ccc gac acc cgg aac acc
    atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg
383◀ V   G   V   G   P   V   G   V   G   P   V   G   V   G   P   V   G

501 tac gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc
    atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg
366◀ V   G   P   V   G   F   G   P   V   G   V   G   P   V   G   V   G

552 cgg cac ccc cgg gta cag gtg gta gtc aac gtc ttc acg cgg gat gtg gcc
    gcc gtg ggg gcc cat gtc cac cat cag ttg cag aag tgc gcc cta cac cgg
349◀ P   V   G   P   Y   L   H   Y   D   V   D   E   R   P   I   H   G

603 gat ctg gat ctc ttc acc gtc gag cgg cac ccc gac acc cgg aac acc tac
    cta gac cta gag aag tgg cag ctc gcc gtg ggg ctg tgg gcc ttg tgg atg
332◀ I   Q   I   E   E   G   D   L   P   V   G   V   G   P   V   G   V

654 gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg
    cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc
315◀ G   P   V   G   F   G   P   V   G   V   G   P   V   G   V   G   P

705 cac ccc gac acc cgg aac acc tac gcc cgg gac acc gaa gcc cgg tac acc
    gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg
298◀ V   G   V   G   P   V   G   V   G   P   V   G   F   G   P   V   G

756 cac gcc cgg aac gcc cac acc cgg cac ccc gac acc cgg aac acc tac gcc
    gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg
281◀ V   G   P   V   G   V   G   P   V   G   V   G   P   V   G   V   G

807 cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg cac
    gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg
264◀ P   V   G   F   G   P   V   G   V   G   P   V   G   V   G   P   V

858 ccc gac acc cgg aac acc tac gcc cgg gac acc gaa gcc cgg tac acc cac
    ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg
247◀ G   V   G   P   V   G   V   G   P   V   G   F   G   P   V   G   V
```

```
  909 gcc cgg aac gcc cac acc cgg cac ccc gac acc cgg aac acc tac gcc cgg
      cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc
  230◀ G   P   V   G   V   G   P   V   G   V   G   P   V   G   V   G   P

  960 gac acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg cac ccc
      ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg
  213◀ V   G   F   G   P   V   G   V   G   P   V   G   V   G   P   V   G

 1011 cgg gta cag gtg gta gtc aac gtc ttc acg cgg gat gtg gcc gat ctg gat
      gcc cat gtc cac cat cag ttg cag aag tgc gcc cta cac cgg cta gac cta
  196◀ P   Y   L   H   Y   D   V   D   E   R   P   I   H   G   I   Q   I

 1062 ctc ttc acc gtc gag cgg cac ccc gac acc cgg aac acc tac gcc cgg gac
      gag aag tgg cag ctc gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg
  179◀ E   E   G   D   L   P   V   G   V   G   P   V   G   V   G   P   V

 1113 acc gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg cac ccc gac
      tgg ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg
  162◀ G   F   G   P   V   G   V   G   P   V   G   V   G   P   V   G   V

 1164 acc cgg aac acc tac gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg
      tgg gcc ttg tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc
  145◀ G   P   V   G   V   G   P   V   G   F   G   P   V   G   V   G   P

 1215 aac gcc cac acc cgg cac ccc gac acc cgg aac acc tac gcc cgg gac acc
      ttg cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg
  128◀ V   G   V   G   P   V   G   V   G   P   V   G   V   G   P   V   G

 1266 gaa gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg cac ccc gac acc
      ctt cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg ctg tgg
  111◀ F   G   P   V   G   V   G   P   V   G   V   G   P   V   G   V   G

 1317 cgg aac acc tac gcc cgg gac acc gaa gcc cgg tac acc cac gcc cgg aac
      gcc ttg tgg atg cgg gcc ctg tgg ctt cgg gcc atg tgg gtg cgg gcc ttg
   94◀ P   V   G   V   G   P   V   G   F   G   P   V   G   V   G   P   V

 1368 gcc cac acc cgg cac ccc gac acc cgg aac acc tac gcc cgg gac acc gaa
      cgg gtg tgg gcc gtg ggg ctg tgg gcc ttg tgg atg cgg gcc ctg tgg ctt
   77◀ G   V   G   P   V   G   V   G   P   V   G   V   G   P   V   G   F

 1419 gcc cgg tac acc cac gcc cgg aac gcc cac acc cgg cac ccc cgg gta cag
      cgg gcc atg tgg gtg cgg gcc ttg cgg gtg tgg gcc gtg ggg gcc cat gtc
   60◀ G   P   V   G   V   G   P   V   G   V   G   P   V   G   P   Y   L

 1470 gtg gta gtc aac gtc ttc acg cgg gat gtg gcc gat ctg gat ctc ttc acc
      cac cat cag ttg cag aag tgc gcc cta cac cgg cta gac cta gag aag tgg
   43◀ H   Y   D   V   D   E   R   P   I   H   G   I   Q   I   E   E   G

 1521 gtc gag ttt atc atc atc atc gtg gtg gtg gtg gtg gtg gtg acc cat ttg
      cag ctc aaa tag tag tag tag cac cac cac cac cac cac cac tgg gta aac
   26◀ D   L   K   D   D   D   D   H   H   H   H   H   H   H   G   M   Q

                          NheI  (1588)
 1572 ctg tcc acc agt cat gct agc cat cat GGTATATCTCCTTCTTAAAGTTAAACAAAATTA
      gac agg tgg tca gta cga tcg gta gta CCATATAGAGGAAGAATTTCAATTTGTTTTAAT
    9◀ Q   G   G   T   M   S   A   M   M

        XbaI  (1635)                                                        BglII
 1632 TTTCTAGAGGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGATCGA
      AAAGATCTCCCCTTAACAATAGGCGAGTGTTAAGGGGATATCACTCAGCATAATTAAAGCGCCCTAGCT

 1701 GATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC
      CTAGAGCTAGGAGATGCGGCCTGCGTAGCACCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCG

 1770 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGG
      GATATAGCGGCTGTAGTGGCTACCCCTTCTAGCCCGAGCGGTGAAGCCCGAGTACTCGCGAACAAAGCC

                                                             SphI  (1898)
 1839 CGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGccgaacgacc
      GCACCCATACCACCGTCCGGGGCACCGGCCCCCTGACAACCCGCGGTAGAGGAACgtacggcttgctgg

 1908 gagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgt
      ctcgcgtcgctcagtcactcgctccttcgccttctcgcgggttatgcgtttggcggagaggggcgcgca
```

```
1977 tggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgca
     accggctaagtaattacgtcgaccgtgctgtccaaagggctgacctttcgcccgtcactcgcgttgcgt
2046 attaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttg
     taattacactcaatcgagtgagtaatccgtggggtccgaaatgtgaaatacgaaggccgagcatacaac
2115 tgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctcgaa
     acaccttaacactcgcctattgttaaagtgtgtcctttgtcgatactggtactaatgcggttcgagctt
2184 attaaccctcactaaagggaacaaaagctggtaccgggccccccctcgaggtcgacggtatcgataagc
     taattgggagtgatttcccttgttttcgaccatggcccgggggggagctccagctgccatagctattcg
2253 ttgatatcgaattcccccgggaccaaaatggcaagtaaaatagcctgatgggataggctctaagtccaa
     aactatagcttaagggggccctggttttaccgttcattttatcggactaccctatccgagattcaggtt
2322 cgaaccagtgtcaccactgacacaatgaggaaaaccatgtcacatctcgcagaactggttgccagtgcg
     gcttggtcacagtggtgactgtgttactccttttggtacagtgtagagcgtcttgaccaacggtcacgc
                                       1▶M  S  H  L  A  E  L  V  A  S  A
2391 aaggcggccattagccaggcgtcagatgttgccgcgttagataatgtgcgcgtcgaatatttgggtaaa
     ttccgccggtaatcggtccgcagtctacaacggcgcaatctattacacgcgcagcttataaacccattt
  12▶K  A  A  I  S  Q  A  S  D  V  A  A  L  D  N  V  R  V  E  Y  L  G  K
2460 aaagggcacttaacccttcagatgacgaccctgcgtgagctgccgccagaagagcgtccggcagctggt
     tttcccgtgaattgggaagtctactgctgggacgcactcgacggcggtcttctcgcaggccgtcgacca
  35▶K  G  H  L  T  L  Q  M  T  T  L  R  E  L  P  P  E  E  R  P  A  A  G
2529 gcggttatcaacgaagcgaaagagcaggttcagcaggcgctgaatgcgcgtaaagcggaactggaaagc
     cgccaatagttgcttcgctttctcgtccaagtcgtccgcgacttacgcgcatttcgccttgacctttcg
  58▶A  V  I  N  E  A  K  E  Q  V  Q  Q  A  L  N  A  R  K  A  E  L  E  S
2598 gctgcactgaatgcgcgtctggcggcggaaacgattgatgtctctctgccaggtcgtcgcattgaaaac
     cgacgtgacttacgcgcagaccgccgcctttgctaactacagagagacggtccagcagcgtaacttttg
  81▶A  A  L  N  A  R  L  A  A  E  T  I  D  V  S  L  P  G  R  R  I  E  N
2667 ggcggtctgcatccggttacccgtaccatcgaccgtatcgaaagtttcttcggtgagcttggctttacc
     ccgccagacgtaggccaatgggcatggtagctggcatagctttcaaagaagccactcgaaccgaaatgg
 104▶G  G  L  H  P  V  T  R  T  I  D  R  I  E  S  F  F  G  E  L  G  F  T
2736 gtggcaaccgggccggaaatcgaagacgattatcataacttcgatgctctgaacattcctggtcaccac
     caccgttggcccggcctttagcttctgctaatagtattgaagctacgagacttgtaaggaccagtggtg
 127▶V  A  T  G  P  E  I  E  D  D  Y  H  N  F  D  A  L  N  I  P  G  H  H
2805 ccggcgcgcgctgaccacgacactttctggtttgacactacccgcctgctgcgtacccagacctctggc
     ggccgcgcgcgactggtgctgtgaaagaccaaactgtgatgggcggacgacgcatgggtctggagaccg
 150▶P  A  R  A  D  H  D  T  F  W  F  D  T  T  R  L  L  R  T  Q  T  S  G
2874 gtacagatccgcaccatgaaagcccagcagccaccgattcgtatcatcgcgcctggccgtgtttatcgt
     catgtctaggcgtggtactttcgggtcgtcggtggctaagcatagtagcgcggaccggcacaaatagca
 173▶V  Q  I  R  T  M  K  A  Q  Q  P  P  I  R  I  I  A  P  G  R  V  Y  R
2943 aacgactacgaccagactcacacgccgatgttccatcagatggaaggtctgattgttgataccaacatc
     ttgctgatgctggtctgagtgtgcggctacaaggtagtctaccttccagactaacaactatggttgtag
 196▶N  D  Y  D  Q  T  H  T  P  M  F  H  Q  M  E  G  L  I  V  D  T  N  I
3012 agctttaccaacctgaaaggcacgctgcacgacttcctgcgtaacttctttgaggaagatttgcagatt
     tcgaaatggttggactttccgtgcgacgtgctgaaggacgcattgaagaaactccttctaaacgtctaa
 219▶S  F  T  N  L  K  G  T  L  H  D  F  L  R  N  F  F  E  E  D  L  Q  I
3081 cgcttccgtccttcctacttcccgtttaccgaaccttctgcagaagtggacgtcatgggtaaaaacggt
     gcgaaggcaggaaggatgaagggcaaatggcttggaagacgtcttcacctgcagtacccatttttgcca
 242▶R  F  R  P  S  Y  F  P  F  T  E  P  S  A  E  V  D  V  M  G  K  N  G
3150 aaatggctggaagtgctgggctgcgggatggtgcatccgaacgtgttgcgtaacgttggcatcgacccg
     tttaccgaccttcacgacccgacgccctaccacgtaggcttgcacaacgcattgcaaccgtagctgggc
 265▶K  W  L  E  V  L  G  C  G  M  V  H  P  N  V  L  R  N  V  G  I  D  P
3219 gaagtttactctggtttcggcttcgggatggggatggagcgtctgactatgttgcgttacggcgtcacc
     cttcaaatgagaccaaagccgaagccctacccctacctcgcagactgatacaacgcaatgccgcagtgg
 288▶E  V  Y  S  G  F  G  F  G  M  G  M  E  R  L  T  M  L  R  Y  G  V  T
3288 gacctgcgttcattcttcgaaaacgatctgcgtttcctcaaacagtttaaataaggcaggaatagatta
     ctggacgcaagtaagaagcttttgctagacgcaaaggagtttgtcaaatttattccgtccttatctaat
 311▶D  L  R  S  F  F  E  N  D  L  R  F  L  K  Q  F  K  ·
```

```
3357 tgaaattcagtgaactgtggttacgcgaatgggtgaacccggcgattgatagcgatgcgctggcaaatc
     actttaagtcacttgacaccaatgcgcttacccacttgggccgctaactatcgctacgcgaccgtttag

3426 aaatcactatggcgggcctggaagttggggggatccactagttctagagcggccgccaccgcggtggagc
     tttagtgataccgcccggaccttcaacccccctaggtgatcaagatctcgccggcggtggcgccacctcg

3495 tccaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaa
     aggttaagcgggatatcactcagcataatgttaagtgaccggcagcaaaatgttgcagcactgaccctt

                                     SphI  (3599)
3564 aaccctggcgttacccaacttaatcgccttgcatgCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCT
     ttgggaccgcaatgggttgaattagcggaacgtacGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGA

3633 CAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCA
     GTTGGATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTCTAGGGCCTGTGGT

3702 TCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGA
     AGCTTACCGCGTTTTGGAAAGCGCCATACCGTACTATCGCGGGCCTTCTCTCAGTTAAGTCCCACCACT

3771 ATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCG
     TACACTTTGGTCATTGCAATATGCTACAGCGTCTCATACGGCCACAGAGAATAGTCTGGCAAAGGGCGC

3840 TGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGA
     ACCACTTGGTCCGGTCGGTGCAAAGACGCTTTTGCGCCCTTTTTCACCTTCGCCGCTACCGCCTCGACT

3909 ATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCT
     TAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAACGACTAACCGCAACGGTGGA

3978 CCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG
     GGTCAGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTTAGAGCGCGGCTAGTTGACCCAC

4047 CCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTC
     GGTCGCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGACATTTCGCCGCCACGTGTTAGAAG

4116 TCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAG
     AGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCTACTGGTCCTACGGTAACGACACCTTC

4185 CTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTT
     GACGGACGTGATTACAAGGCCGCAATAAAGAACTACAGAGACTGGTCTGTGGGTAGTTGTCATAATAAA

4254 TCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGC
     AGAGGGTACTTCTGCCATGCGCTGACCCGCACCTCGTAGACCAGCGTAACCCAGTGGTCGTTTAGCGCG

4323 TGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTC
     ACAATCGCCCGGGTAATTCAAGACAGAGCCGCGCAGACGCAGACCGACCGACCGTATTTATAGAGTGAG

4392 GCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCA
     CGTTAGTTTAAGTCGGCTATCGCCTTGCCCTTCCGCTGACCTCACGGTACAGGCCAAAAGTTGTTTGGT

4461 TGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCG
     ACGTTTACGACTTACTCCCGTAGCAAGGGTGACGCTACGACCAACGGTTGCTAGTCTACCGCGACCCGC

4530 CAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATA
     GTTACGCGCGGTAATGGCTCAGGCCCGACGCGCAACCACGCCTATAGAGCCATCACCCTATGCTGCTAT

4599 CCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAA
     GGCTTCTGTCGAGTACAATATAGGGCGGCAATTGGTGGTAGTTTGTCCTAAAAGCGGACGACCCCGTTT

4668 CCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCT
     GGTCGCACCTGGCGAACGACGTTGAGAGAGTCCCGGTCCGCCACTTCCCGTTAGTCGACAACGGGCAGA

4737 CACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATT
     GTGACCACTTTTCTTTTTGGTGGGACCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAA

4806 CATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTA
     GTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACAT

4875 AGTTAGCTCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCC
     TCAATCGAGTGAGTAATCCGTGGCCCTAGAGCTGGCTACGGGAACTCTCGGAAGTTGGGTCAGTCGAGG

4944 TTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTA
     AAGGCCACCCGCGCCCCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCAT

5013 GGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATC
     CCTGTCCACGGCCGTCGCGAGACCCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCGCTGCTACTAG

5082 GGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC
     CCGGACAGCGAACGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGG

5151 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTC
     TTTGCAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGGGGTGCCCACGCGTACTAGCACGAG
```

```
5220 CTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGC
     GACAGCAACTCCTGGGCCGATCCGACCGCCCCAACGGAATGACCAATCGTCTTACTTAGTGGCTATGCG

5289 GAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGT
     CTCGCTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTGTACTTACCAGAAGCCA

5358 TTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATC
     AAGGCACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATACAAGGCCTAGACGTAG

5427 GCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGA
     CGTCCTACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCGCGACCGTAACTGGGACT

5496 GTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGG
     CACTAAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGTTGCAAGGTCATTGGCC

5565 GCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATG
     CGTACAAGTAGTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCCATAGTAATGGGGGTAC

5634 AACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGC
     TTGTCTTTAGGGGGAATGTGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCGGGAATTGTACCGGGCG

5703 TTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGAC
     AAATAGTCTTCGGTCTGTAATTGCGAAGACCTCTTTGAGTTGCTCGACCTGCGCCTACTTGTCCGTCTG

5772 ATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGAC
     TAGACACTTAGCGAAGTGCTGGTGCGACTACTCGAAATGGCGTCGACGGAGCGCGCAAAGCCACTACTG

5841 GGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGC
     CCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCG

5910 AGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGT
     TCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCGTCGGTACTGGGTCAGTGCA

5979 AGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATA
     TCGCTATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTAT

6048 TATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTC
     ATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCGAGAAGGCGAAGGAG

6117 GCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
     CGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTA

6186 ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG
     TGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTC

6255 GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
     CTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTT

6324 TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAG
     AGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTC

6393 CTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
     GAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCC

6462 AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCT
     TTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGA

6531 GGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTC
     CCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAG

6600 CAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
     GTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCAT

6669 TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
     ACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACC

6738 TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAAC
     ATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTG

6807 CACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
     GTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCT

6876 AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
     TCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCA

6945 CATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGG
     GTACTTGTTATTTTGACAGACGAATGTATTTGTCATTATGTTCCCCACAATACTCGGTATAAGTTGCCC

7014 AAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTC
     TTTGCAGAACGAGATCCGGCGCTAATTTAAGGTTGTACCTACGACTAAATATACCCATATTTACCCGAG
```

```
7083  GCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGT
      CGCTATTACAGCCCGTTAGTCCACGCTGTTAGATAGCTAACATACCCTTCGGGCTACGCGGTCTCAACA
7152  TTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGA
      AAGACTTTGTACCGTTTCCATCGCAACGGTTACTACAATGTCTACTCTACCAGTCTGATTTGACCGACT
7221  CGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCA
      GCCTTAAATACGGAGAAGGCTGGTAGTTCGTAAAATAGGCATGAGGACTACTACGTACCAATGAGTGGT
7290  CTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTG
      GACGCTAGGGGCCCTTTTGTCGTAAGGTCCATAATCTTCTTATAGGACTAAGTCCACTTTTATAACAAC
7359  ATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATC
      TACGCGACCGTCACAAGGACGCGGCCAACGTAAGCTAAGGACAAACATTAACAGGAAAATTGTCGCTAG
7428  GCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATG
      CGCATAAAGCAGAGCGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTACGCTCACTAAAACTAC
7497  ACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGG
      TGCTCGCATTACCGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTTGAAAACGGTAAGAGTGGCC
7566  ATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTT
      TAAGTCAGCAGTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTCCCCTTTAATTATCCAA
7635  GTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCG
      CATAACTACAACCTGCTCAGCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACCTTGACGGAGC
7704  GTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATA
      CACTCAAAAGAGGAAGTAATGTCTTTGCCGAAAAAGTTTTTATACCATAACTATTAGGACTATACTTAT
7773  AATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAA
      TTAACGTCAAAGTAAACTACGAGCTACTCAAAAAGATTCTTAATTAAGTACTCGCCTATGTATAAACTT
7842  TGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATTGTA
      ACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTTTAACAT
7911  AACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCC
      TTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGG
7980  GAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGG
      CTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACC
8049  AACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
      TTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTA
8118  GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGG
      CCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCC
8187  AACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG
      TTGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCC
8256  AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACA
      TTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGT
8325  CCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA
      GGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGGGTAAGCGGT
```

Strain/Plasmid Collection

Submitted by	Sarah Heilshorn	**Date**	May 10, 2004

Strain name	BL21(DE3) pLysS/pET28aRW-RGD BL21(DE3) pLysS/pET28aRW-RDG
Vector (kb)	pET28aRW-RGD or pET28aRW-[RGD(El-IK)$_4$]$_3$ (6.407) pET28aRW-RDG or pET28aRW-[RDG(El-IK)$_4$]$_3$ (6.407)
Cloning site	See plasmid map.

Construction

Ralf Weberskirch designed the pET28aRW vector which replaces the cloning region of the pET28a vector with a new polylinker region that includes a T7 tag, His tag, enterokinase cleavage site, *XhoI* restriction site, and stop codon. Julie Liu used the pET28aRW vector to create an aECM construct with the RGD cell-binding domain and the elastin-like sequence (VPGIG)$_2$VPGKG(VPGIG)$_2$, which is referred to as El-IK. The RDG domain is identical to the RGD cell-binding domain, except the codons for RGD have been scrambled to encode R<u>DG</u> as a negative control for cell culture experiments. Details on the cloning scheme can be found in the reference below.

References

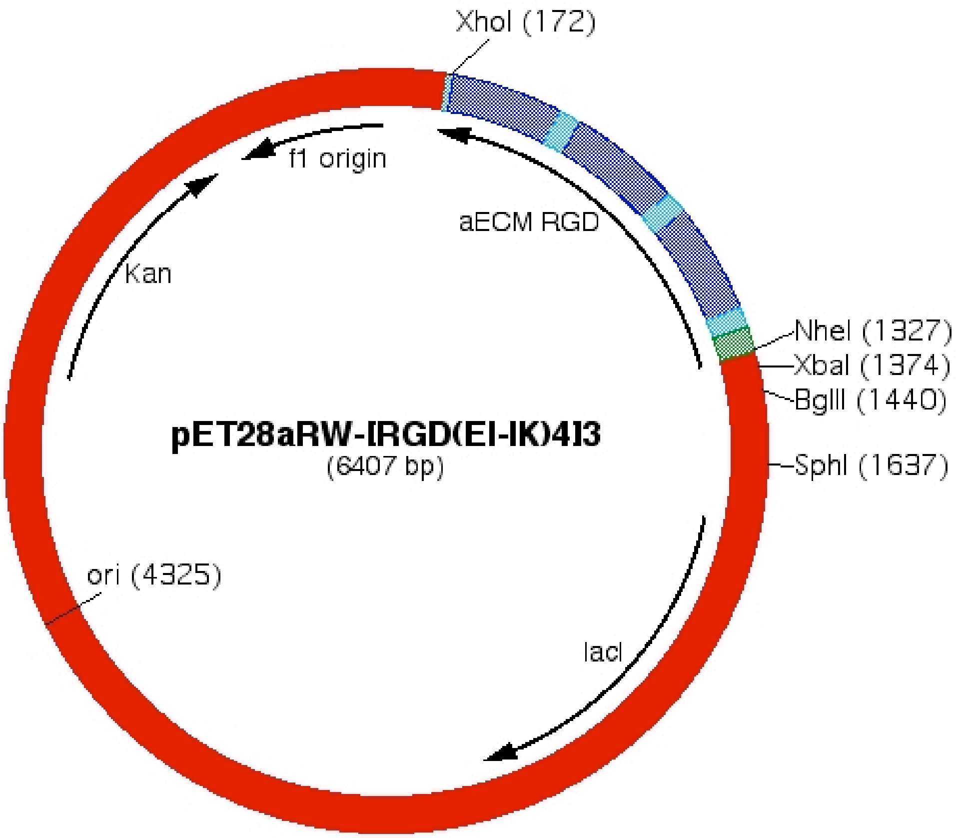
Liu JC, Heilshorn SC, Tirrell DA. Biomacromolecules 2004. 5:4970504.

Color codes for plasmid maps and DNA sequences

red	original pET28a vector
green	RW polylinker designed by Ralf Weberskirch
cyan	cell-binding domain
blue	elastin-like domain

Plasmid map

Only the plasmid map for pET28aRW-RGD is shown. The RDG plasmid map is similar.



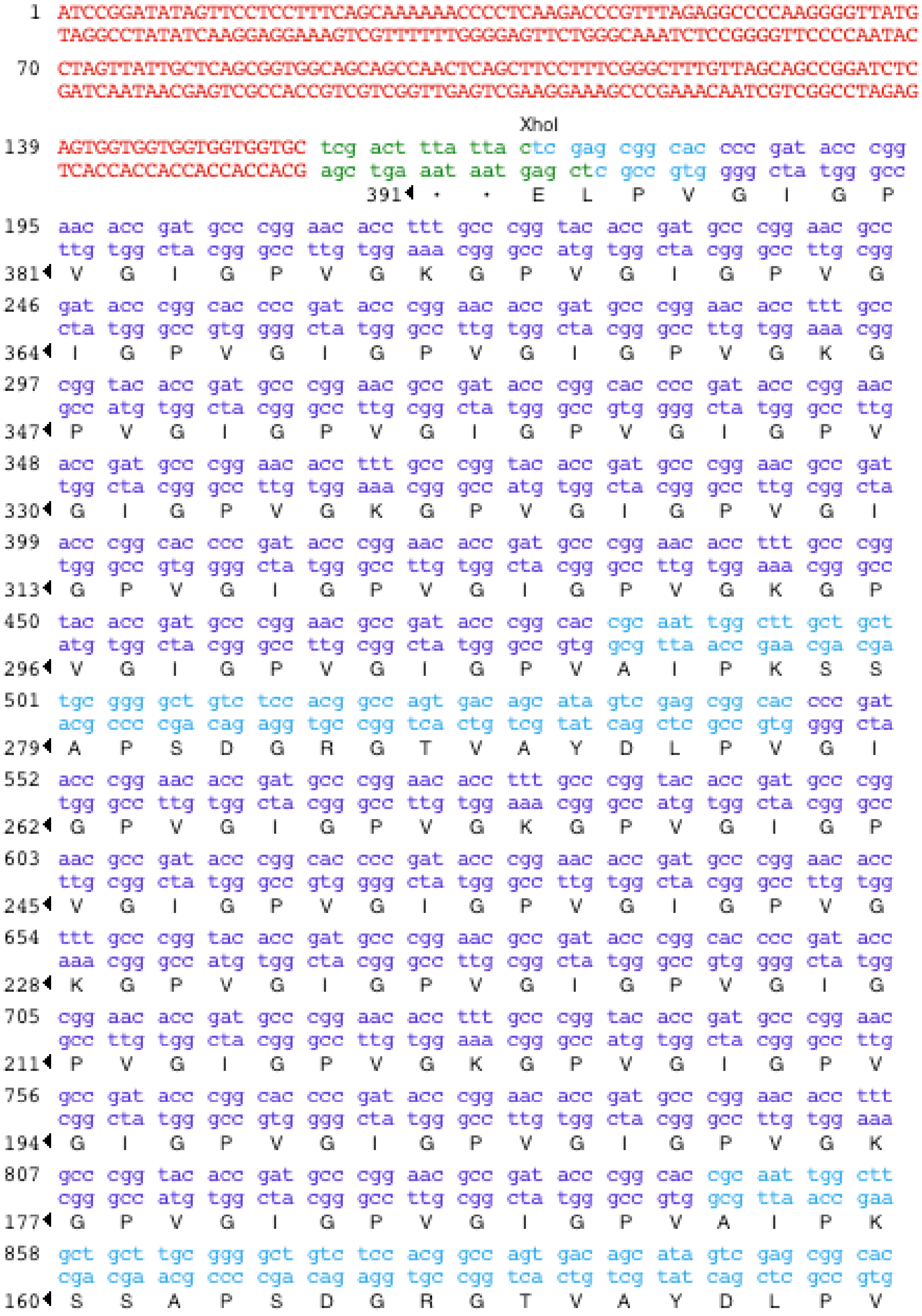
Positions of elements in bases

	pET28aRW-RGD	pET28aRW-RDG
Vector size	6407	6407
Kan	5034-5846	5034-5846
f1 origin	5942-6397	5942-6397
lacI	1812-2891	1812-2891
ori	4325	4325
aECM region	165-1337	165-1337

Culture conditions

Shake flask: 2xYT, induce with 2mM IPTG at OD 1
Fermentor: Terrific Broth, induce with 2.5 mM IPTG at OD 6-8

<u>Full sequence</u> Only the sequence for pET28aRW-RGD is shown. The RDG sequence is identical except for the shuffling of the RGD codons within the RGD cell-binding domain to encode RDG in the RDG construct.



```
  1 ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATG
    TAGGCCTATATCAAGGAGGAAAGTCGTTTTTTGGGGAGTTCTGGGCAAATCTCCGGGGTTCCCCAATAC

 70 CTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTC
    GATCAATAACGAGTCGCCACCGTCGTCGGTTGAGTCGAAGGAAAGCCCGAAACAATCGTCGGCCTAGAG

                                         XhoI
139 AGTGGTGGTGGTGGTGGTGC tcg act tta tta ctc gag cgg cac ccc gat acc cgg
    TCACCACCACCACCACCACG agc tga aat aat gag ctc gcc gtg ggg cta tgg gcc
                             391◀ ·   ·   E   L   P   V   G   I   G   P

195 aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc
    ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg
381◀ V   G   I   G   P   V   G   K   G   P   V   G   I   G   P   V   G

246 gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc
    cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg
364◀ I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   K   G

297 cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac
    gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg
347◀ P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

348 acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat
    tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta
330◀ G   I   G   P   V   G   K   G   P   V   G   I   G   P   V   G   I

399 acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg
    tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc
313◀ G   P   V   G   I   G   P   V   G   I   G   P   V   G   K   G   P

450 tac acc gat gcc cgg aac gcc gat acc cgg cac cgc aat tgg ctt gct gct
    atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg gcg tta acc gaa cga cga
296◀ V   G   I   G   P   V   G   I   G   P   V   A   I   P   K   S   S

501 tgc ggg gct gtc tcc acg gcc agt gac agc ata gtc gag cgg cac ccc gat
    acg ccc cga cag agg tgc cgg tca ctg tcg tat cag ctc gcc gtg ggg cta
279◀ A   P   S   D   G   R   G   T   V   A   Y   D   L   P   V   G   I

552 acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg
    tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc
262◀ G   P   V   G   I   G   P   V   G   K   G   P   V   G   I   G   P

603 aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc
    ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg
245◀ V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G

654 ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc gat acc
    aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg
228◀ K   G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G

705 cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc cgg aac
    gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg
211◀ P   V   G   I   G   P   V   G   K   G   P   V   G   I   G   P   V

756 gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac acc ttt
    cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa
194◀ G   I   G   P   V   G   I   G   P   V   G   I   G   P   V   G   K

807 gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac cgc aat tgg ctt
    cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg gcg tta acc gaa
177◀ G   P   V   G   I   G   P   V   G   I   G   P   V   A   I   P   K

858 gct gct tgc ggg gct gtc tcc acg gcc agt gac agc ata gtc gag cgg cac
    cga cga acg ccc cga cag agg tgc cgg tca ctg tcg tat cag ctc gcc gtg
160◀ S   S   A   P   S   D   G   R   G   T   V   A   Y   D   L   P   V
```

```
 909 ccc gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat
     ggg cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta
143◀  G   I   G   P   V   G   I   G   P   V   G   K   G   P   V   G   I

 960 gcc cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg
     cgg gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc
126◀  G   P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P

1011 aac acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac ccc
     ttg tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg ggg
109◀  V   G   K   G   P   V   G   I   G   P   V   G   I   G   P   V   G

1062 gat acc cgg aac acc gat gcc cgg aac acc ttt gcc cgg tac acc gat gcc
     cta tgg gcc ttg tgg cta cgg gcc ttg tgg aaa cgg gcc atg tgg cta cgg
 92◀  I   G   P   V   G   I   G   P   V   G   K   G   P   V   G   I   G

1113 cgg aac gcc gat acc cgg cac ccc gat acc cgg aac acc gat gcc cgg aac
     gcc ttg cgg cta tgg gcc gtg ggg cta tgg gcc ttg tgg cta cgg gcc ttg
 75◀  P   V   G   I   G   P   V   G   I   G   P   V   G   I   G   P   V

1164 acc ttt gcc cgg tac acc gat gcc cgg aac gcc gat acc cgg cac cgc aat
     tgg aaa cgg gcc atg tgg cta cgg gcc ttg cgg cta tgg gcc gtg gcg tta
 58◀  G   K   G   P   V   G   I   G   P   V   G   I   G   P   V   A   I

1215 tgg ctt gct gct tgc ggg gct gtc tcc acg gcc agt gac agc ata gtc gag
     acc gaa cga cga acg ccc cga cag agg tgc cgg tca ctg tcg tat cag ctc
 41◀  P   K   S   S   A   P   S   D   G   R   G   T   V   A   Y   D   L

1266 ttt atc atc atc atc gtg gtg gtg gtg gtg gtg gtg acc cat ttg ctg tcc
     aaa tag tag tag tag cac cac cac cac cac cac cac tgg gta aac gac agg
 24◀  K   D   D   D   D   H   H   H   H   H   H   H   G   M   Q   Q   G

                 NheI                                              XbaI
1317 acc agt cat gct agc cat cat GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGA
     tgg tca gta cga tcg gta gta CCATATAGAGGAAGAATTTCAATTTGTTTTAATAAAGATCT
  7◀  G   T   M   S   A   M   M

                                                               BglII
1379 GGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGA
     CCCCTTAACAATAGGCGAGTGTTAAGGGGATATCACTCAGCATAATTAAAGCGCCCTAGCTCTAGAGCT

1448 TCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCG
     AGGAGATGCGGCCTGCGTAGCACCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGC

1517 CCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTA
     GGCTGTAGTGGCTACCCCTTCTAGCCCGAGCGGTGAAGCCCGAGTACTCGCGAACAAAGCCGCACCCAT

                                                  SphI
1586 TGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGG
     ACCACCGTCCGGGGCACCGGCCCCCTGACAACCCGCGGTAGAGGAACGTACGTGGTAAGGAACGCCGCC

1655 CGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTC
     GCCACGAGTTGCCGGAGTTGGATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAG

1724 GAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGT
     CTCTAGGGCCTGTGGTAGCTTACCGCGTTTTGGAAAGCGCCATACCGTACTATCGCGGGCCTTCTCTCA

1793 CAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTAT
     GTTAAGTCCCACCACTTACACTTTGGTCATTGCAATATGCTACAGCGTCTCATACGGCCACAGAGAATA

1862 CAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCG
     GTCTGGCAAAGGGCGCACCACTTGGTCCGGTCGGTGCAAAGACGCTTTTGCGCCCTTTTTCACCTTCGC

1931 GCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTG
     CGCTACCGCCTCGACTTAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAACGAC

2000 ATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGC
     TAACCGCAACGGTGGAGGTCAGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTTAGAGCG

2069 GCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCG
     CGGCTAGTTGACCCACGGTCGCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGACATTTCGC

2138 GCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGAT
     CGCCACGTGTTAGAAGAGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCTACTGGTCCTA

2207 GCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCC
     CGGTAACGACACCTTCGACGGACGTGATTACAAGGCCGCAATAAAGAACTACAGAGACTGGTCTGTGGG
```

```
2276  ATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGT
      TAGTTGTCATAATAAAAGAGGGTACTTCTGCCATGCGCTGACCCGCACCTCGTAGACCAGCGTAACCCA

2345  CACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGG
      GTGGTCGTTTAGCGCGACAATCGCCCGGGTAATTCAAGACAGAGCCGCGCAGACGCAGACCGACCGACC

2414  CATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCC
      GTATTTATAGAGTGAGCGTTAGTTTAAGTCGGCTATCGCCTTGCCCTTCCGCTGACCTCACGGTACAGG

2483  GGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGAT
      CCAAAAGTTGTTTGGTACGTTTACGACTTACTCCCGTAGCAAGGGTGACGCTACGACCAACGGTTGCTA

2552  CAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTA
      GTCTACCGCGACCCGCGTTACGCGCGGTAATGGCTCAGGCCCGACGCGCAACCACGCCTATAGAGCCAT

2621  GTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTT
      CACCCTATGCTGCTATGGCTTCTGTCGAGTACAATATAGGGCGGCAATTGGTGGTAGTTTGTCCTAAAA

2690  CGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAAT
      GCGGACGACCCCGTTTGGTCGCACCTGGCGAACGACGTTGAGAGAGTCCCGGTCCGCCACTTCCCGTTA

2759  CAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCC
      GTCGACAACGGGCAGAGTGACCACTTTTCTTTTTGGTGGGACCGCGGGTTATGCGTTTGGCGGAGAGGG

2828  CGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCG
      GCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGC

2897  CAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTT
      GTTGCGTTAATTACATTCAATCGAGTGAGTAATCCGTGGCCCTAGAGCTGGCTACGGGAACTCTCGGAA

2966  CAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTT
      GTTGGGTCAGTCGAGGAAGGCCACCCGCGCCCCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAA

3035  TATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTG
      ATAGTACGTTGAGCATCCTGTCCACGGCCGTCGCGAGACCCAGTAAAAGCCGCTCCTGGCGAAAGCGAC

3104  GAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGT
      CTCGCGCTGCTACTAGCCGGACAGCGAACGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCA

3173  CACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGT
      GTGACCAGGGCGGTGGTTTGCAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGGGGTGCCCA

3242  GCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAA
      CGCGTACTAGCACGAGGACAGCAACTCCTGGGCCGATCCGACCGCCCCAACGGAATGACCAATCGTCTT

3311  TGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAAC
      ACTTAGTGGCTATGCGCTCGCTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTG

3380  ATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTAT
      TACTTACCAGAAGCCAAAGGCACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATA

3449  GTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCG
      CAAGGCCTAGACGTAGCGTCCTACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCGC

3518  CTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACA
      GACCGTAACTGGGACTCACTAAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGT

3587  ACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGG
      TGCAAGGTCATTGGCCCGTACAAGTAGTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCC

3656  TATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGC
      ATAGTAATGGGGGTACTTGTCTTTAGGGGGAATGTGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCG

3725  CCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGC
      GGAATTGTACCGGGCGAAATAGTCTTCGGTCTGTAATTGCGAAGACCTCTTTGAGTTGCTCGACCTGCG

3794  GGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGC
      CCTACTTGTCCGTCTGTAGACACTTAGCGAAGTGCTGGTGCGACTACTCGAAATGGCGTCGACGGAGCG

3863  GCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTA
      CGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACAT

3932  AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGC
      TCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCGTCG

4001  CATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTA
      GTACTGGGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACAT

4070  CTGAGAGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGC
      GACTCTCACGTGGTATATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCG
```

```
4139 GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
     CGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAG
4208 ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
     TGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTC
4277 GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTG
     CGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGAC
4346 ACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
     TGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCC
4415 CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG
     GCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGC
4484 CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGG
     GGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAAGCCACATCC
4553 TCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTA
     AGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCAT
4622 ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
     TGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCT
4691 TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
     AATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGAT
4760 GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
     CTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAA
4829 GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAA
     CTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTT
4898 AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
     TTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTG
4967 GTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATG
     CAATTCCCTAAAACCAGTACTTGTTATTTTGACAGACGAATGTATTTGTCATTATGTTCCCCACAATAC
5036 AGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATAT
     TCGGTATAAGTTGCCCTTTGCAGAACGAGATCCGGCGCTAATTTAAGGTTGTACCTACGACTAAATATA
5105 GGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCC
     CCCATATTTACCCGAGCGCTATTACAGCCCGTTAGTCCACGCTGTTAGATAGCTAACATACCCTTCGGG
5174 GATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTC
     CTACGCGGTCTCAACAAAGACTTTGTACCGTTTCCATCGCAACGGTTACTACAATGTCTACTCTACCAG
5243 AGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGAT
     TCTGATTTGACCGACTGCCTTAAATACGGAGAAGGCTGGTAGTTCGTAAAATAGGCATGAGGACTACTA
5312 GCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCA
     CGTACCAATGAGTGGTGACGCTAGGGGCCCTTTTGTCGTAAGGTCCATAATCTTCTTATAGGACTAAGT
5381 GGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGT
     CCACTTTTATAACAACTACGCGACCGTCACAAGGACGCGGCCAACGTAAGCTAAGGACAAACATTAACA
5450 CCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGAT
     GGAAAATTGTCGCTAGCGCATAAAGCAGAGCGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTA
5519 GCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTT
     CGCTCACTAAAACTACTGCTCGCATTACCGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTTGAA
5588 TTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAG
     AACGGTAAGAGTGGCCTAAGTCAGCAGTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTC
5657 GGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATC
     CCCTTTAATTATCCAACATAACTACAACCTGCTCAGCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAG
5726 CTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGAT
     GATACCTTGACGGAGCCACTCAAAAGAGGAAGTAATGTCTTTGCCGAAAAAGTTTTTATACCATAACTA
5795 AATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAG
     TTAGGACTATACTTATTTAACGTCAAAGTAAACTACGAGCTACTCAAAAAGATTCTTAATTAAGTACTC
5864 CGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGT
     GCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCA
5933 GCCACCTGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATT
     CGGTGGACTTTAACATTTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAA
```

```
6002 TTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAG
     AAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTC

6071 TGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAAC
     ACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTG

6140 CGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCG
     GCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGC

6209 TAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGT
     ATTTCGTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCA

6278 GGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCT
     CCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGA

6347 GCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA
     CGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGGGTAAGCGGT
```