Supporting Information

for

A Convenient Method for Genetic Incorporation of Multiple Noncanonical Amino Acids into One Protein in *Escherichia Coli*

Ying Huang, William K. Russell, Wei Wan, Pei-Jing Pai, David H. Russell, and Wenshe Liu*

Contents

1. DNA sequences

tRNA (pylT)[1]:

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

The lpp promoter:

cccatcaaaaaaatattctcaacataaaaaactttgtgtaatacttgtaacgct

The rrnC terminator:

atccttagcgaaagctaaggatttttttta

AcKRS:

atgtcagataaaaaaccattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaagcaccat
gaggtctcaagaagtaaaatatacattgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaacagccagagc
attcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcgggcgaggatatcaataattttctcacaagatcaaccgaaa
gcaaaaacagtgtgaaagttagggtagtttctgctccaaaggtcaaaaaagctatgccgaaatcagtttcaagggctccgaagcctctg
gaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgcaaaatcaactccaaattcgtctgttcccgcatc
ggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaagtccagaggataaaatttctctgaatatggcaaag
cctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagcggctctataccaatgatagagaagactacctcggt
aaactcgaacgtgatattacgaaatttttcgtagaccggggttttctggagataaagtctcctatccttattccggcggaatacgtggaga
gaatgggtattaataatgatactgaactttcaaaacagatcttccgggtggataaaaatctctgcttgaggccaatggttgccccgactatt
ttcaactatgcgcgaaaactcgataggattttaccaggcccaataaaaattttcgaagtcggaccttgttaccggaaagagtctgacggc
aaagagcacctggaagaatttactatggtgaacttctttcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagt
ttctggactatctggaaatcgacttcgaaatcgtaggagattcctgtatggtctatggggatactcttgatataatgcacggggacctgga
gctttcttcggcagtcgtcgggccagtttctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgct
tgctcaaggttatgcacggctttaaaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaa

L11C:

atgaccaagaccccgccggcagcagttctgctgaaaaaagcggctggtatcaagtctggttccggtaagccgaacaaagacaaagt
gggtaaaatttcccgcgctcagctgcaggaaatcgcgcagaccaaagctgccgacatgactggtgccgacattgaagcgatgactcg
ctccatcgaaggtactgcacgttccatgggcctggtagtggaggactaa

GFP1Amber:

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcagtgga
gagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtca
ctactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgt
acaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtat
cgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacactaggtatacatcac
ggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatc
aacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacg
aaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgaactctacaaagagctccatcacc
atcaccatcactaa

GFP2Amber:

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcagtgga
gagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtca
ctactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgt
acaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtat
cgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacactaggtatacatcac
ggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatc
aacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacatagtctgccctttcgaaagatcccaacga
aaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgaactctacaaagagctccatcaccat
caccatcactaa

GFP3Amber:

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcagtgga
gagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtca
ctactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgt
acaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtat
cgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacactaggtatacatcac
ggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatc
aatagaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacatagtctgccctttcgaaagatcccaacga
aaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgaactctacaaagagctccatcaccat
caccatcactaa

GFP2Amber':

atggcatagagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtc
agtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaaca
cttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaag
gttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgtta
atcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacactaggtata
catcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagac
cattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacatagtctgccctttcgaaagatcc
caacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgaactctacaaagagctcca
tcaccatcaccatcactaa

2. Protein sequences

AcKRS:

msdkkpldvlisatglwmsrtgtlhkikhhevsrskiyiemacgdhlvvnnsrscrtarafrhhkyrktckrcrvsgedinnfltrst
esknsvkvrvvsapkvkkampksvsrapkplensvsakastntsrsvpspakstpnssvpasapapsltrsqldrveallspedki
slnmakpfrelepelvtrrkndfqrlytndredylgklerditkffvdrgfleikspilipaeyvermginndtelskqifrvdknlclr
pmvaptifnyarkldrilpgpikifevgpcyrkesdgkehleeftmvnffqmgsgctrenlealikefldyleidfeivgdscmvy
gdtldimhgdlelssavvgpvsldrewgidkpwigagfglerllkvmhgfknikrasrsesyyngistnl

L11C:

mtktppaavllkkaagiksgsgkpnkdkvgkisraqlqeiaqtkaadmtgadieamtrsiegtarsmglvved

GFP1Amber:

mskgeelftgvvpilveldgdvnghkfsvsgegegdatygkltlkficttgklpvpwptlvttfsygvqcfsrypdhmkrhdffks ampegyvqertisfkddgnyktraevkfegdtlvnrielkgidfkedgnilghkleynynshk*vyitadkqkngikanfkirhni edgsvqladhyqqntpigdgpvllpdnhylstqsalskdpnekrdhmvllefvtaagithgmdelykelhhhhhh

GFP2Amber:

mskgeelftgvvpilveldgdvnghkfsvsgegegdatygkltlkficttgklpvpwptlvttfsygvqcfsrypdhmkrhdffks ampegyvqertisfkddgnyktraevkfegdtlvnrielkgidfkedgnilghkleynynshk*vyitadkqkngikanfkirhni edgsvqladhyqqntpigdgpvllpdnhylstk*salskdpnekrdhmvllefvtaagithgmdelykelhhhhhh

GFP3Amber:

mskgeelftgvvpilveldgdvnghkfsvsgegegdatygkltlkficttgklpvpwptlvttfsygvqcfsrypdhmkrhdffks ampegyvqertisfkddgnyktraevkfegdtlvnrielkgidfkedgnilghkleynynshk*vyitadkqkngikanfkirhni edgsvqladhyqk*ntpigdgpvllpdnhylstk*salskdpnekrdhmvllefvtaagithgmdelykelhhhhhh

GFP2Amber':

mak*skgeelftgvvpilveldgdvnghkfsvsgegegdatygkltlkficttgklpvpwptlvttfsygvqcfsrypdhmkrhdf fksampegyvqertisfkddgnyktraevkfegdtlvnrielkgidfkedgnilghkleynynshk*vyitadkqkngikanfkir hniedgsvqladhyqqntpigdgpvllpdnhylstqsalskdpnekrdhmvllefvtaagithgmdelykelhhhhhh

3. Construction of plasmids

The construction of pAcKRS-pylT-GFP1Amber, pET-L11C, pAcKRS-pylT-GFP2Amber, pAcKRS-pylT-GFP3Amber, and pAcKRS-pylT-GFP2Amber' all followed standard cloning and QuikChange site-directed mutagenesis procedures using Platinum Pfx (Invitrogen) and PfuTurbo (Stratagene) DNA polymerases. All the plasmid structures have been confirmed by DNA sequencing. All oligodeoxynucleotide primers were purchased from Integrated DNA Technologies, Inc.

Construction of pBK-AcKRS-pylT

M. barkeri pyrrolysyl-tRNA synthetase (MbPylRS) was amplified from *M. barkeri* genomic DNA that was purchased from ATCC by polymerase chain reaction (PCR) using two primers, GAGGAATCCCATATGGATAAAAAACCATTAG and CGTTTGAAACTGCAGTTACAGATTGGTTG. AcKRS (L266V, L270I, Y271F, L274A, C313F and D76G)[2] was subsequently generated by overlap extension PCR using MbPylRS as a template and eight oligodeoxynucleotide primers (GAGGAATCCCATATGGATAAAAAACCATTAG,
AAATTATTGATATCCTCGCCCGAAACCCTACATCGTTTGC,
GATGTAGGGTTTCGGGCGAGGATATCAATAATTTTC,
GTTGAAAATAGTCGGGGCAACCATTGGCCTCAAGCAG,
GCCCCGACTATTTTCAACTATGCGCGAAAACTCGATAGG,
CCGAACCCATCTGAAAGAAGTTCACCATAG,
CTATGGTGAACTTCTTTCAGATGGGTTCGG,
CGTTTGAAACTGCAGTTACAGATTGGTTG). Two restriction sites *Nde1* at 5' head and

Pst1 at 3' tail were inserted into the synthesized AcKRS gene. The gene was subsequently digested by *NdeI* and *PstI* restriction enzymes and cloned into the same two sites in a pBK plasmid[3] to afford pBK-AcKRS. In pBK-AcKRS, AcKRS is under the control of a constitutive *glnS* promoter. The pylT gene flanked by the *lpp* promoter at 5' end and the *rrnC* terminator at 3' end was constructed using overlap extension PCR of six oligodeoxynucleotides (CCCGGGATCCCCCATCAAAAAAATATTCTCAACAT, TTACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTG, ACTTTGTGTAATACTTGTAACGCTGAATCCGGAAACCTGATCATGTAGAT, CTAACCCGGCTGAACGGATTTAGAGTCCATTCGATCTACATGATCAGGTTT, TCAGCCGGGTTAGATTCCCGGGGTTTCCGCCACTGCCCATCCTTAGCGAA, and GAACCCAGATCTTAAAAAAAATCCTTAGCTTTCGCTAAGGATG). Two restriction sites, *BamH1* at 5' end and *Bgl11* at 3' end, were inserted in the synthesized DNA which was subsequently digested and cloned into the *BamHI* site in pBK-AcKRS to afford pBK-AcKRS-pylT. Using pBK-AcKRS-pylT together with 5 mM AcK to suppress an amber codon at position 149 of GFP_{UV} in BL21 cells was tested but did not give high amber suppression efficiency. We then decided to put AcKRS under control of a stronger promoter and constructed the plasmid pAcKRS-pylT-GFP1Amber.

Construction of pAcKRS-pylT-GFP1Amber

The pAcKRS-pylT-GFP1Amber plasmid contains genes encoding the AcKRS-pylT pair and GFP_{UV} that has one amber mutation at position 149 and a 6×His tag at the C-terminus (GFP1Amber). Both of AcKRS and GFP1Amber were under control of T7 promoters. AcKRS was PCR amplified from pBK-AcKRS-pylT by two oligodeoxynucleotides (GATATAACATGTCAGATAAAAAACCATTAGATG, and GTCGACCTGCAGTTACAGATTGGTTGAAATCCC). The amplified DNA was digested by *Pci1* end and *Pst1* restriction enzymes and cloned into *Nco1* and *Pst1* sites of the pETduet-1 vector which was purchased from Stratagene Inc. to afford pAcKRS. GFP1Amber was PCR amplified from pleiG-N149[4] (a gift from Dr. Peter G. Schultz) using two oligodeoxynucleotides (GAAGGAGATATACATATGAGTAAAGGAGAAG and GACTCGAGGGTACCTTAGTGATGGTGATGGTGATG), digested by *NdeI* and *KpnI* restriction enzymes, and then cloned into *NdeI* and *KpnI* restriction sites of pAcKRS to afford pAcKRS-GFP1Amber. pylT with the *lpp* promoter and the *rrnC* terminator was PCR amplified from pBK-AcKRS-pylT using two oligodeoxynucleotides (GCTAGATCTGGAAACCTGATGTAGATC and GATACTAGTTGGCGGAAACCCCGGG), digested by *SphI* restriction enzymes, and then cloned into the *SphI* site of pAcKRS-GFP1Amber to afford pAcKRS-pylT-GFP1Amber.

Construction of pAcKRS-pylT-GFP2Amber

Plasmid pAcKRS-pylT-GFP2Amber was derived from pAcKRS-pylT-GFP1Amber with an additional amber mutation at position 204 of the GFP_{UV} gene. The mutagenesis followed the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligodeoxynucleotides, CTTTCGAAAGGGCAGACTATGTCGACAGGTAATG and CATTACCTGTCGACATAGTCTGCCCTTTCGAAAG.

Construction of pAcKRS-pylT-GFP3Amber

Plasmid pAcKRS-pylT-GFP3Amber was derived from pAcKRS-pylT-GFP2Amber with an additional amber mutation at position 184. The mutagenesis followed the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligodeoxynucleotides, ATCGCCAATTGGAGTATTCTATTGATAATGGTCTGC and GCAGACCATTATCAATAGAATACTCCAATTGGCGAT.

Construction of pAcKRS-pylT-GFP2Amber'

Plasmid pAcKRS-pylT-GFP2Amber' was derived from pAcKRS-pylT-GFP1Amber. pAcKRS-pylT-GFP2Amber' contains the GFP2Amber' gene that has two amber mutations at positions 1 and 149 and a Met-Ala N-terminal leading dipeptide. The GFP2Amber' gene was PCR amplified from pAcKRS-pylT-GFP1Amber using Pfx DNA polymerase and two oligodeoxynucleotides, GATATACATATGGCATAGAGTAAAGGAGAAGAA and CATTACCTGTCGACATAGTCTGCCCTTTCGAAAG. The amplified DNA was digested by *NdeI* and *KpnI* restriction enzymes and subsequently cloned into a predigested pAcKRS-pylT-GFP1Amber at *NdeI* and *KpnI* sites to afford pAcKRS-pylT-GFP2Amber'.

Construction of pET-L11C

Plasmid pET-L11C contains the L11C gene whose transcription is under control of T7 promoter. The gene was PCR amplified from the genomic DNA of *E. coli* using two primers, GGAGATATACATATGACCAAGACCCCGCCGGCA and GTCGTCGGTACCTTAGTCCTCCACTACCAG. The amplified DNA was digested by *NdeI* and *KpnI* restriction enzymes and cloned into *NdeI* and *KpnI* sites in pET30a (Stratagene Inc.) to afford pET-L11C.

4. Protein expression and purification

Expression and purification of GFP_{UV}

To express different GFP_{UV} variants, *E. coli* BL21 cells was transformed with pAcKRS-pylT-GFP1Amber, pAcKRS-pylT-GFP2Amber, pAcKRS-pylT-GFP3Amber or pAcKRS-pylT-GFP2Amber' together with or without pET-L11C. The cells transformed with one plasmid were grown in LB media containing 100 μg/mL ampicillin and induced with the addition of 500 μg/mL IPTG when OD_{600} reached 0.6. 5 mM AcK and 5 mM nicotinamide were subsequently added into the media in 30 min after induction. The cells were then let grow overnight or 10 h at 37 degree. The protein expression in cells transformed with two plasmids followed exactly same procedures except the addition of 25 μg/mL kanamycin into media to force cells to maintain pET30-L11C. The GFP_{UV} expression in cells transformed with either one or two plasmids in media with no addition of AcK also followed the same procedures. Cells were harvested by centrifugation (4500 r.p.m., 20 min, 4 degree) and resuspended in 20 mL of lysis buffer (50 mM HEPES, pH 7.4, 500 mM NaCl, 10 mM DTT, 10% glycerol, 0.1% Triton X-100, 5 mM imidazole, and 1μg/mL lysozyme). The resuspended cells were sonicated and the lysate was clarified by centrifugation (10200 r.p.m., 60 min, 4 degree). The supernatant was decanted and loaded to Ni-NTA superspeed agarose

(Qiagen Inc.) column on FPLC. The column was washed by 5× bed volume of buffer A that contained 50 mM HEPES, pH 7.5, 300 mM NaCl, 5 mM imidazole and then eluted by running a gradient that changed from buffer A to buffer B in 10× bed volume. Buffer B contained 50 mM HEPES, pH 7.5, 300 mM NaCl, 250 mM imidazole. Proteins were concentrated by Amicon (Millipore, NMWL 10 KDa) and analyzed by 12% SDS-PAGE.

Wild type GFP_{UV} was purified from BL21 cells transformed with pREP.[5] It does not have a 6×His tag and it does not require induction. The expression of wild type GFP_{UV} was simply done by growing transformed cells in LB medium overnight. The cells were collected and lysed following same procedures discussed above. The supernatant of the cell lysate was fractionated by the addition of 70% ammonium sulfate. The precipitate was then resolubilized. The protein concentration was determined using the extinction coefficient at 397 nm.

Protein concentration determination

The concentration of GFP-1AcK was determined by Pierce BCA protein assay kits. Based on this concentration, we calculated the extinction coefficient of GFP-1AcK at 397 nm as 27000 $cm^{-1}M^{-1}$. The concentration of other proteins was determined by BCA protein assays and confirmed by simple calculation based on their absorbance at 397 nm.

5. Mass spectrometry analysis

LC-ESI-MS analysis of intact protein

An Agilent (Santa Clara, CA) 1200 capillary HPLC system was interfaced to an API QSTAR Pulsar Hybrid QTOF mass spectrometer (Applied Biosystems/MDS Sciex, Framingham, MA) equipped with an electrospray ionization (ESI) source. Liquid chromatography (LC) separation was achieved using a Phenomenex Jupiter C4 microbore column (150 × 0.50 mm, 300 Å) (Torrance, CA) at a flow rate of 10 μL min^{-1}. The proteins were eluted using a gradient of (A) 0.1% formic acid versus and (B) 0.1% formic acid in acetonitrile. The gradient timetable was as follows: 2% B for 5 min, 2-30% in 3 min, 30-60% in 44 min, 60-95% in 8 min, followed by holding the gradient at 95% for 5 min, for a total run time of 65 min. The MS data were acquired in positive ion mode (500-2000 Da) using spray voltage of +4900 V. BioAnalyst software (Applied Biosystems) was used for spectral deconvolution. A mass range of m/z 500-2000 was used for deconvolution and the output range was 10000-50000 Da using a step mass of 0.1 Da and a S/N threshold of 20.

Protein digestion

GFP_{UV} variants were dissolved in 25 mM Ammonia bicarbonate, and denatured at 90 degree for 15 min. Trypsin (Sigma) or proteinase Asp-N (Roche) was dissolved in 0.01% TFA (pH 3). This solution was added to the denatured GFP_{UV} solution (w/w=1:50) and incubated at 37 degree overnight.

Tandem-MS of proteolytic peptides

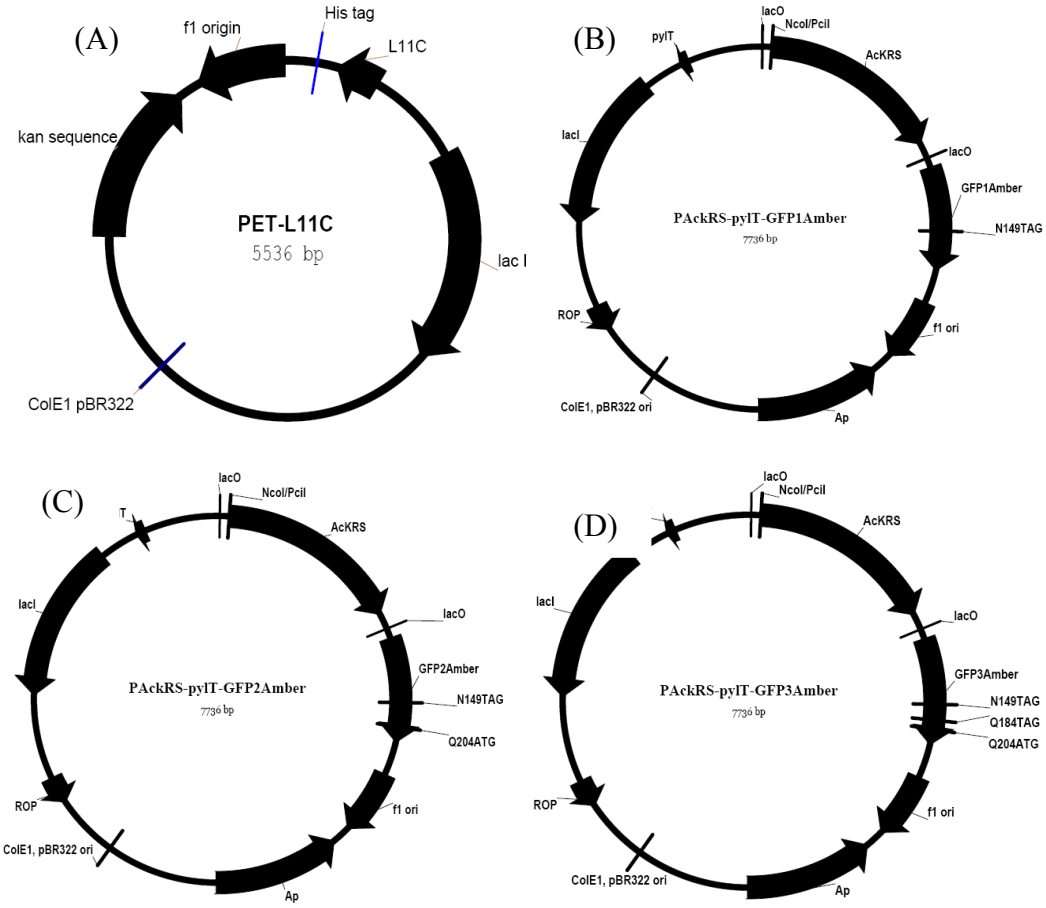
Peptides from tryptic and proteinase Asp-N digests were mixed 1:1 (v/v) with matrix (5 mg mL^{-1} α-cyano-4-hydroxycinnamic acid, 50% (v/v) acetonitrile, 10 mM ammonium

dihydrogen phosphate, 1% TFA) and 1 μL of the mixture was spotted onto a stainless steel target plate. Mass spectra and tandem MS spectra were collected using an Applied Biosystems 4800 Tof/Tof (Framingham, MA). Collision induced dissociation tandem MS spectra were acquired using air at the medium pressure setting and at 2 kV of collision energy. Tandem MS data was manually interpreted using the Data Explorer™ software package (Applied Biosystems, Framingham, MA).

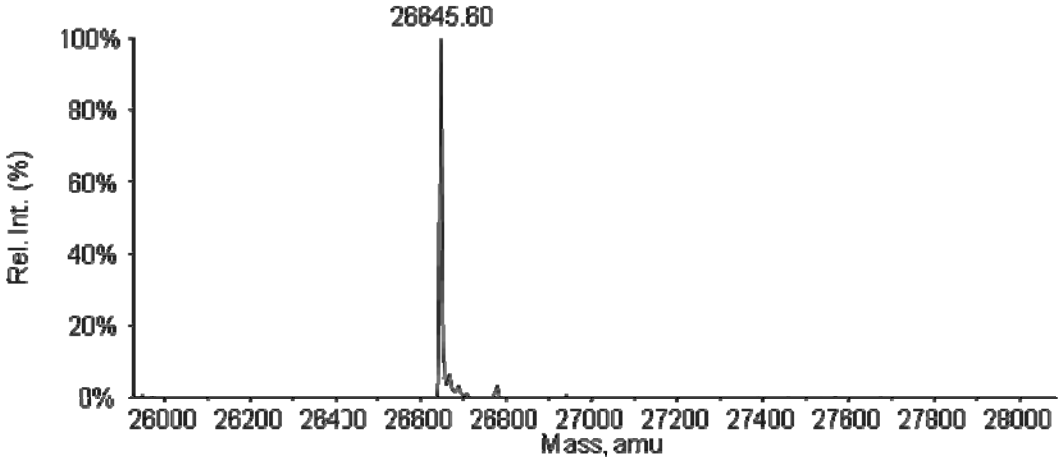
References:

[1] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459.
[2] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat Chem Biol* **2008**, *4*, 232.
[3] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[4] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew Chem Int Ed Engl* **2009**, *48*, 4052.
[5] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotech.* **2002**, 20, 1044

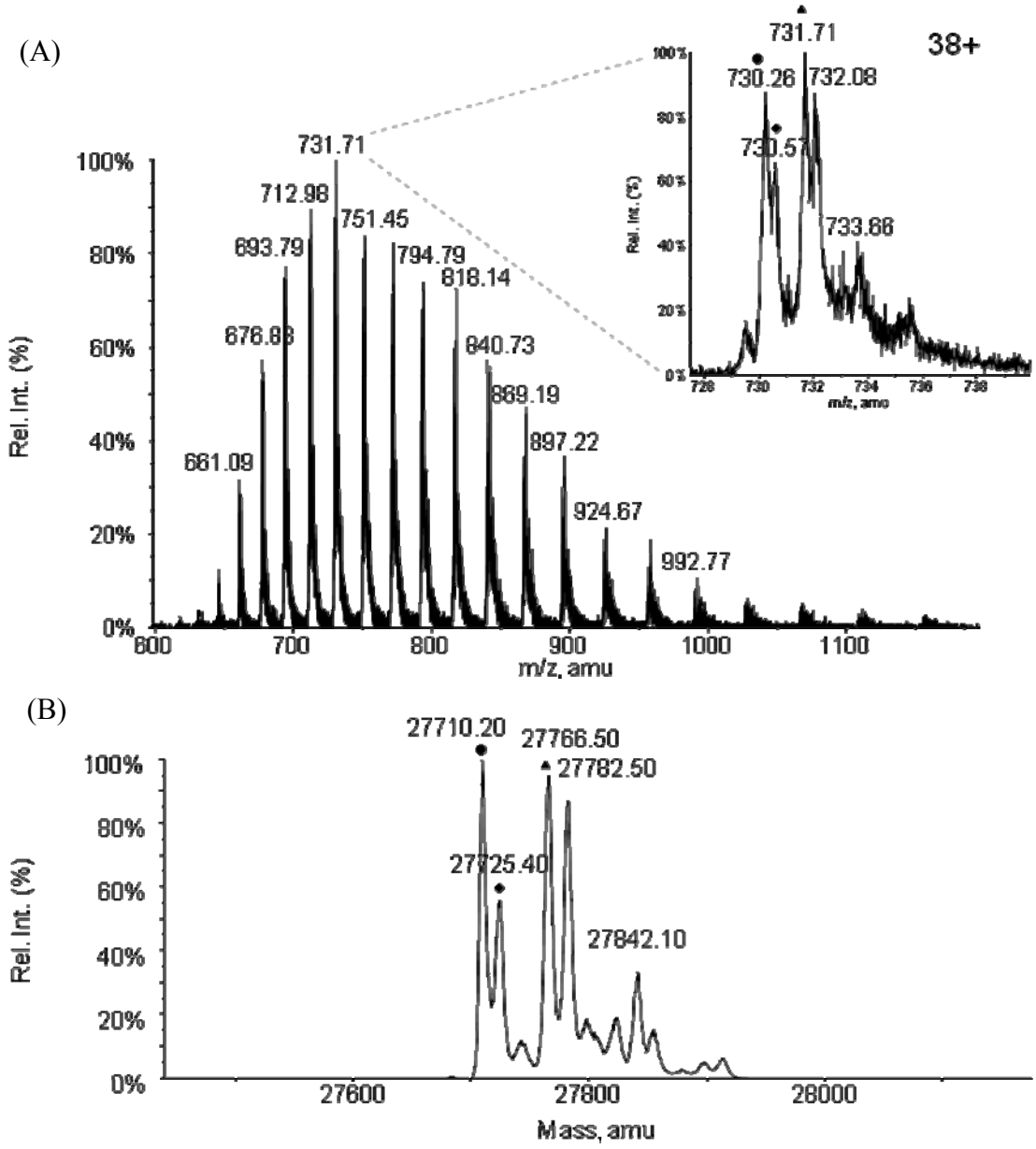
6. Supplementary Figures



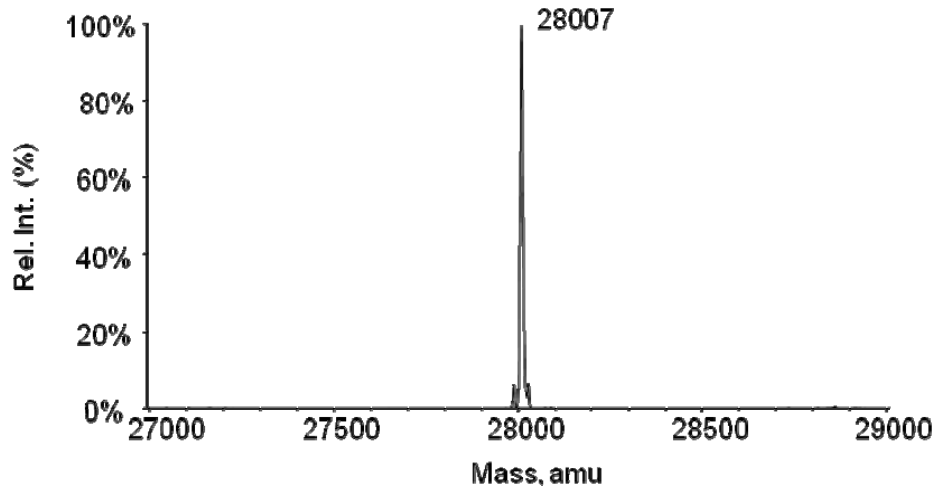
Supplementary Figure 1. Plasmid maps for (**A**) pET-L11C, (**B**) pAcKRS-pylT-GFP1Amber, (**C**) pAcKRS-pylT-GFP2Amber, and (**D**) pAcKRS-pylT-GFP3Amber.



Supplementary Figure 2. Deconvoluted ESI-MS spectrum of wild type GFP_{UV}.



Supplementary Figure 3. (**A**) ESI-TOF MS of GFP_{UV} expressed in cells transformed with pAcKRS-pylT-GFP1Amber and pET-L11C and grown in LB medium with non addition of AcK. (**B**) Deconvoluted MS.



Supplementary Figure 4. Deconvoluted ESI-MS spectrum of GFP-2AcK'.