Supporting Information for:

A versatile platform for single and multiple unnatural amino acid mutagenesis in *Escherichia coli*

Abhishek Chatterjee, Sophie B. Sun, Jennifer L. Furman, Han Xiao and Peter G. Schultz*

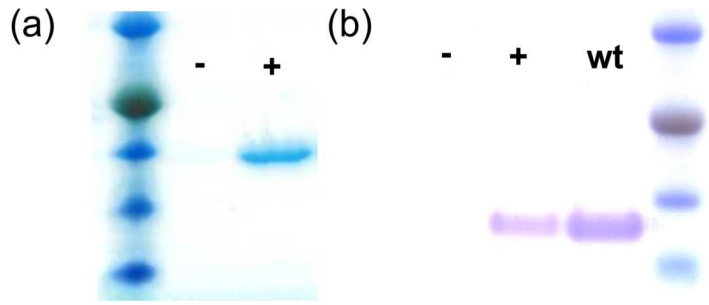

Figure S1: SDS-PAGE analysis of purified C-terminal $(His)_6$-tagged T4-lysozyme (T4L; a) and ketosteroid isomerase (KSI; b), expressed from pET101-T4L(Asn68TAG) and pET28-KSI(His78TAG), respectively, using pUltra-pAcF to suppress the amber nonsense codon, in the presence (+) or absence (-) of 1 mM pAcF in the growth medium. For KSI, the wild type (wt) protein expression level (without amber suppression) is also shown.

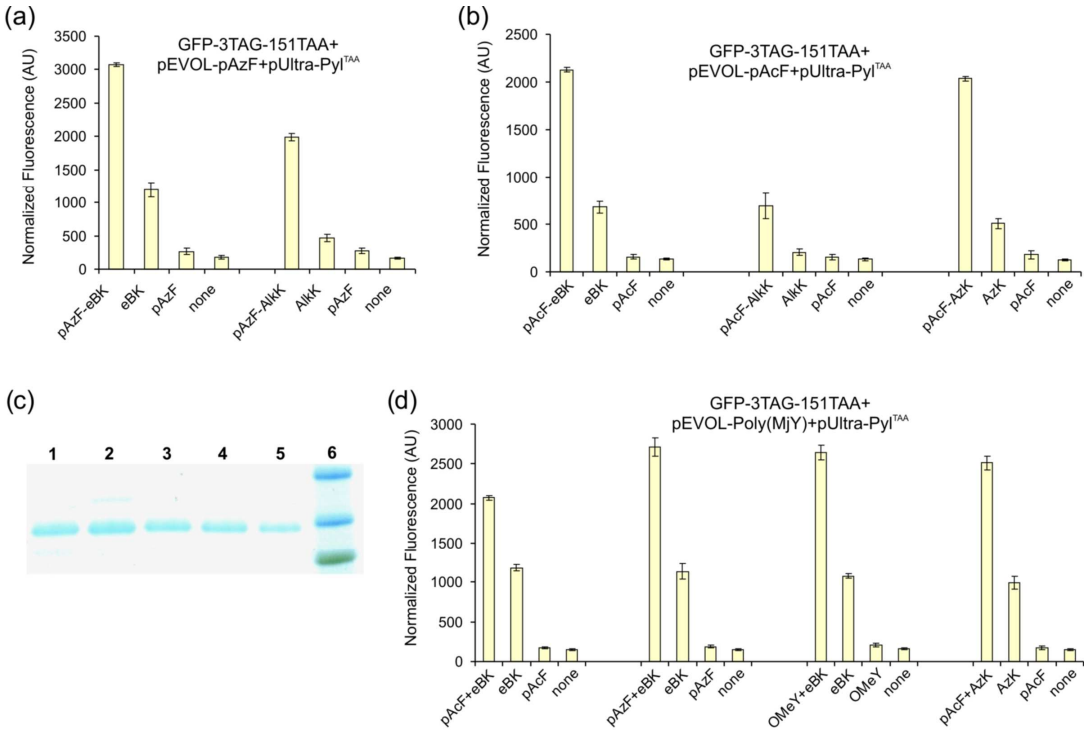

Figure S2: Expression of GFP from pET-GFP-3TAG-151TAA by dual-suppression using pEVOL-MjY (pAzF-specific, pAcF-specific or polyspecific) to suppress the TAG3 with pAzF, pAcF, or OMeY and pUltra-PylTAA to suppress TAA151 with eBK, AzK or AlkK. Normalized GFP fluorescence is reported in the presence of both UAAs, in the presence of either one of the two UAAs, and in the absence of both UAAs. (a) Incorporation pAzF and eBK, or pAzF and AlkK to TAG3 and TAA151, respectively, using pEVOL-pAzF and pUltra-Pyl. (b) Incorporation pAcF and eBK, pAcF and AlkK, or pAcF and AzK to TAG3 and TAA151, respectively, using pEVOL-pAcF and pUltra-Pyl. (c) SDS-PAGE analysis of purified GFP incorporating pAzF and eBK (1), pAcF and eBK (2), pAcF and AzK (3), pAzF and AlkK (4), or pAcF and AlkK (5). Lane 6 contains MW markers (d) Incorporation pAcF and eBK, pAzF and eBK, OMeY and eBK, or pAcF and AzK to TAG3 and TAA151, respectively, using pEVOL-polyspecific and pUltra-Pyl.

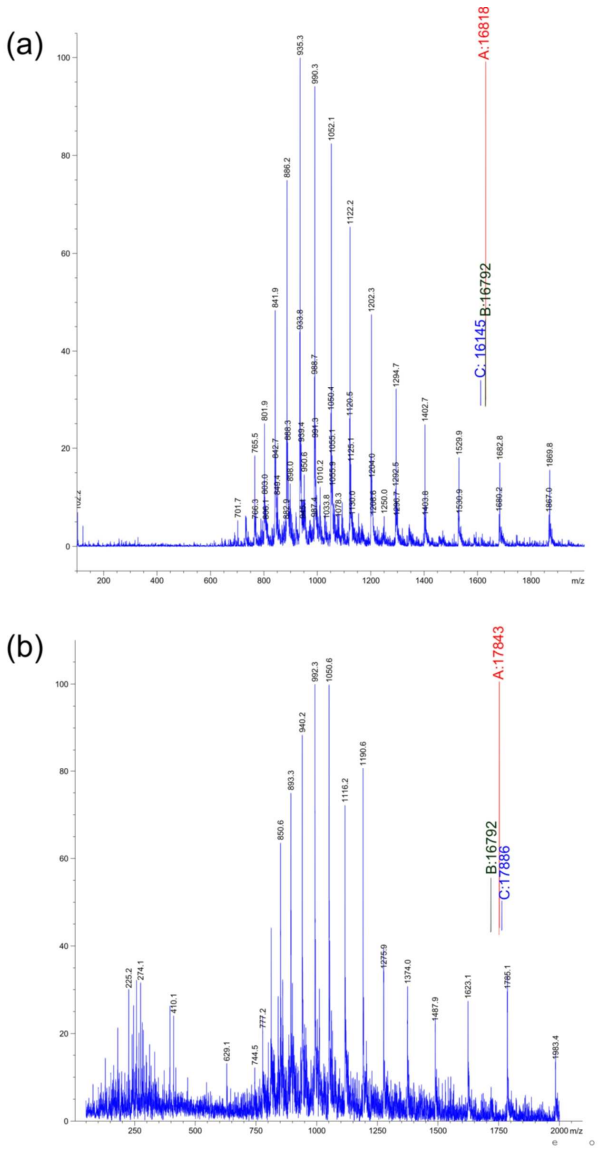

Figure S3: MS analysis of KSI-7-AzK-78-pAcF labeled with Alexa Fluor® 488 C5-Aminooxyacetamide alone (a) or together with Alexa Fluor® 594 DIBO Alkyne (b). Deconvolution spectrum is shown in the inset in both cases. In (a), the expected Alexa Fluor 488 labeled product (expected and observed MW 16818) is the major species with a minor (<5%) population of unlabeled protein (expected MW 16146, observed MW 16145) and labeled protein where the azido functionality of AzK is reduced to amine (<20%; expected and observed MW 16792). In (b), the expected Alexa Fluor 488 and 594 double labeled protein is the major species (expected and observed MW 17843), with a minor contamination (<20%) from Alexa Fluor 488-only labeled, azido-reduced protein (expected and observed MW 16792). A +43 Da adduct of the double labeled protein (MW 17886), presumably from mono-carbamylation by urea, is also observed.

(a) GFP-3*

ATGGCATAGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATG
TTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCA
ATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATG
TACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACA
CAAACTCGAATACAACTATAACTCACACTAGGTATAGATCACGGCAGACAAACAAAAGAATGGAATCAA
AGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAAT
ACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAA
AGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGC
ATGGATGAGCTCTACAAACTCGAGCACCACCACCACCACCACTGA

(b) GFP-3TAG-151TAA

ATGGCATAGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATG
TTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCA
ATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATG
TACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACA
CAAACTCGAATACAACTATAACTCACACTAAGTATACATCACGGCAGACAAACAAAAGAATGGAATCAA
AGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAAT
ACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAA
AGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGC
ATGGATGAGCTCTACAAACTCGAGCACCACCACCACCACCACTGA

(c) MbPylRS(opt)

ATGGATAAAAAACCGCTGGACGTTCTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCTGC
ATAAAATTAAACACCACGAAGTGTCACGTTCGAAAATCTATATCGAAATGGCGTGCGGTGATCATCTGGT
GGTTAACAATAGCCGTTCTTGTCGCACCGCGCGTGCCTTTCGCCATCACAAATACCGCAAAACGTGCAAA
CGTTGTCGCGTGTCAGATGAAGACATTAACAATTTCCTGACCCGTAGTACGGAATCCAAAAACTCAGTGA
AAGTTCGCGTCGTGAGTGCTCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGTCTCCCGTGCCCCGAAAC
CGCTGGAAAACTCAGTGTCGGCAAAAGCTTCCACCAATACGAGCCGCTCTGTTCCGTCGCCGGCAAAAA
GCACCCCGAACAGCTCTGTCCCGGCAAGCGCACCGGCACCGTCTCTGACGCGTAGTCAGCTGGATCGCGT
GGAAGCCCTGCTGTCCCCGGAAGACAAAATCTCACTGAATATGGCAAAACCGTTTCGTGAACTGGAACC
GGAACTGGTTACCCGTCGCAAAAACGATTTCCAACGTCTGTATACGAATGATCGCGAAGACTACCTGGGT
AAACTGGAACGTGATATCACCAAATTTTTCGTGGACCGCGGCTTTCTGGAAATCAAATCTCCGATTCTGA
TCCCGGCTGAATATGTTGAACGCATGGGTATTAACAATGATACCGAACTGAGTAAACAGATTTTTCGTGT
GGATAAAAACCTGTGCCTGCGGCCGATGCTGGCACCGACGCTGTATAATTACCTGCGTAAACTGGATCGC
ATTCTGCCGGGTCCGATTAAAATCTTTGAAGTGGGCCCGTGTTATCGTAAAGAATCGGATGGCAAAGAAC
ACCTGGAAGAATTTACCATGGTTAACTTCTGCCAAATGGGCAGCGGTTGTACGCGCGAAAATCTGGAAGC
GCTGATCAAAGAATTCCTGGATTACCTGGAAATCGACTTCGAAATCGTCGGTGATTCTTGCATGGTGTTTG
GCGATACCCTGGACATCATGCATGGTGACCTGGAACTGAGTTCCGCTGTTGTCGGTCCGGTCAGCCTGGA
TCGTGAATGGGGCATTGACAAACCGTGGATCGGCGCGGGTTTTGGCCTGGAACGCCTGCTGAAAGTTATG
CACGGCTTCAAAAACATCAAACGTGCGTCTCGCTCGGAATCGTATTACAACGGCATCTCAACCAATCTGT
AATAA

(d) KSI-7TAA-78TAG

ATGAACCTGCCGACCGCCTAAGAAGTCCAGGGTCTGATGGCGCGTTTTATTGAACTGGTCGATGTGGGCG
ACATTGAAGCAATTGTGCAGATGTACGCGGACGATGCGACCGTTGAAAGCCCGTTTGGTCAACCACCGAT
TCATGGCCGTGAACAGATTGCGGCGCACTACCGTCAGTGGCTGGGCGGGGGTAAACTGCGTGTTTGCCTG
ACCGGTCCAGTGCGTGCTAGTTAGAACGGCTGCGGCGCGATGCCGTTGCGTAAAGAGTGGGTTTGGAATG
GTCAGCCTTGCGCAACGGATGTTATTCTGGTTATGCGCTTCGATGAACACGGTCGCATCCAGACCGAACA
GCGCTATTGGAGCGAAGTGAATCTGAGTGTACGCGAACCGCAGGGCAGTCTCGAGCACCACCACCACCA
CCACTGA

Figure S4: Nucleotide sequences of GFP-3*, GFP-3TAG-151TAA, MbPylRS(opt) and KSI-7TAA-78TAG

Table S1: ESI-MS analyses of purified proteins incorporating single/double UAAs

Protein	UAAs	Expected MW (Da)	Observed MW (Da)
wtGFP	none	27710	27711
GFP(Y151TAG)	pAcF	27736	27737
GFP(Y151TAG)	OMeY	27724	27726
GFP(Y151TAG)	pAzF	27735	27737
GFP(Y151TAG)	pNO_2F	27739	27741
GFP(Y151TAG)	Tryptophan	27733	27733
GFP(Y151TAG)	Proline	27644	27644
GFP-3*	pAcF	28071	28073
GFP-3*	OMeY	28035	28036
T4L (N68TAG)	pAcF	19742	19743
KSI(H78TAG)	pAcF	16007	16006
GFP(3TAG151TAA)	pAcF+eBK	28084	28083
GFP(3TAG151TAA)	pAzF+eBK	28083	28082
GFP(3TAG151TAA)	OMeY+eBK	28072	28071
GFP(3TAG151TAA)	pAcF+AzK	28097	28097
GFP(3TAG151TAA)	pAcF+AlkK	28066	28064
GFP(3TAG151TAA)	pAzF+AlkK	28065	28064
KSI(7TAA78TAG)	pAcF+eBK	16133	16134
KSI(7TAA78TAG)	pAcF+AzK	16146	16146

Table S2: DNA oligomers used in this study

CDF-Mlu-F	GGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCG
CDF-Bsa-R	CGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTG
CDFi-F	GTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCTTTCACCTGCA GGATCCGGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTG AGAAGCACACGGTCACACTGC
CDFi-R	GAAAGCGGCCGCACCTCCTTTGTGAAATTGTTATCCGCTCACAATTCCAC ACATTATACGAGCCGATGATTAATTGTCAACAGCTCCCTAATGCAGGAGT CGCATAAGGGAGAGCGTCG
rrnB-Not-F	AATAATTGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGG
rrnB-Sbf-R	AATAATTCCTGCAGGCTAACAAGGATTATGGGGATGATTGATGACCGGG AGC
proK-F	aaataatCCTGCAGGtaattccgcttcgcaacatgtgagcaccgg
proK-R	AATAAATccatggCAAATTCGACCCTGAGCTGCTCGAGCATGC
MjYRS-F	aacaactGCGGCCGCatggacgaatttgaaatgataaagagaaacacatctgaaattatc
MjYRS-R	aacaactGCGGCCGCttataatctctttctaattggctctaaaatctttataagttcttc
ScWRS-F	aacaactGCGGCCGCATGAGCAACGACGAAACTGTAGAGAAAGTC
ScWRS-R	aacaactGCGGCCGCTTACTTCTTTTCTTGCTTAGTTTTTGGCTTAGGTGCG
MbPylRS- F	aataattGCGGCCGCatggataaaaaaccattagatgttttaatatctgcg
MbPylRS-R	aataattGCGGCCGCttatagattggttgaaatcccattatagtaagattcgg
MbPylOptRS-R	aataattgcggccgcttaTTACAGATTGGTTGAGATGCCGTTGTAATACGATTCCGA GCG
MbPylOptRS-F	aataattgcggccgcATGGATAAAAAACCGCTGGACGTTCTGATCTCCGC
PhPRS-F	aacaactGCGGCCGCATGGTGGAGAGGAAGAGGTGGAGTGAGG
PhPRS-R	aacaactGCGGCCGCttaATAGGTTCTAGCGAATCTAGCTATAAACTTCGC
Trp-tRNA-R	CCCcccgcTTTAGAGgcgggAGCTCTACCATTGAGCCACCGCTTCaatgcggggcgc atcttactgcgcagatac
Trp-tRNA-F	GAGCTcccgcCTCTAAAgcgggGGGtTGCAGGTTCAATTCCTGcCCGTTTCACC Aaattcgaaaagcctgctcaacg
Pro-tRNA-F	CTGCGGCGcTctaGACGCCGTGACCCGAGTTCAAATCTCGGCGGCCCCACC Aaattcgaaaagcctgctcaacg
Pro-tRNA-R	CGGGTCACGGCGTCtagAgCGCCGCAGGATCACCAGGCTACCCCACGGCCC Caatgcggggcgcatcttactgcgcagatac

Pyl-tRNA-F	AtGGACTcTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCAaattcgaaaagcctgctcaacg
Pyl-tRNA-R	CCCGGCTGAACGGATTTAGAGTCCATTCGATCTACATGATCAGGTTTCCaatgcggggcgcatcttactgcgcagatac
MmPylT-TAA-F	gaatggactttaaatccgttcagccgggttag
MmPylT-TAA-R	cggctgaacggatttaaagtccgttcgatctacatg
MmPylT-U25C-F	GTAGATCGAACGGACTTTAAATCCG
MmPylT-U25C-R	TTTAAAGTCCGTTCGATCTACATGATC
U2Ev-F	aataattAAGCTTgcaatttatctcttcaaatgtagcacctgaagtcagccc
U2Ev-R	aataattCTCGAGgcagtgtgaccgtgtgcttctcaaatgcctgagg
gfpTAGm-151.153	CTCGAATACAACTATAACTCACACtagGTAtagATCACGGCAGACAAACAAAAGAATGG
gfpTAGm-3	GGAATTCAGGAGCCCTTCACCATGgcatagAGTAAAGGAGAAGAACTTTTCACTGGAG
gfpTAAm-151	CTCGAATACAACTATAACTCACACtaaGTATACATCACGGCAGACAAACAAAAGAATGG
tacI-SalI-F	aataattGTCGACggGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG
tRNA-XhoI-R	CAAATTCGACCCTGAGCTGCTCGAGCATGC