Primers used for cloning

Primer	Oligonucleotide Sequence (5'→3')
*Ma*PylRS NdeI	GTTGTTCATATGACCGTGAAGTACACCGACGCCCAG
126/129-Rev	CAGGTTTGGCGCCAGCATCGGGCG
126/129-For	CGCCCGATGCTGGCGCCAAACCTGNNKAGCGTGNNKCGTGACCTGCGCGATCACACCGAC
166-Rev	CAGCATGGTGAACTCCTCCAG
166/168-For	CTGGAGGAGTTCACCATGCTGNNKCTGNNKGATATGGGTCCGCGCGGCGAC
227/228(IP)-Rev	CTCATGCACGTCATGGGCCGCATCCAGCGGAATTGGACCCACAGCCGCGCTGCAC
227/228(IP)-For	GTGCAGCGCGGCTGTGGGTCCAATTCCGCTGGATGCGGCCCATGACGTGCATGAG
*Ma*PylRS PstI	GGTGGTCTGCAGTTAATTGATTTTGGCACCATTCAGGTAG
HRpNEUmaRS-For	GAGACCCAAGCTGGCTAGCGCCACCATGACCGTGAAGTACACCGACGC
HRpNEUmaRS_Rev	GGCTGATCAGCGGGTTTAAAGCGGCCGCTTAATTGATTTTGGCACCATTC
206W-For	CAGGAAGAGAGCGACGTGtggAAGGAAACCATCGACGTG
206W-Rev	CACGTCGATGGTTTCCTTccaCACGTCGCTCTCTTCCTG
MbRS-349W-For	GAGATTCCTGTATGGTCTGGGGGGATACTCTTGATATAATG
MbRS-349W-Rev	CATTATATCAAGAGTATCCCCCCAGACCATACAGGAATCTC
HRpEVOL-MbRS-For	CTAACAGGAGGAATTACTAGTATGGATAAAAAACCATTAGAT
HRpEVOL-MbRS-Rev	CAATGATGATGATGATGATGGTCGACTTATAGATTGGTTGAAATC

Reagents and molecular biology

Primers were synthesized and purified by Integrated DNA Technologies (IDT), and plasmids were sequenced by GENEWIZ. All molecular biology reagents were either obtained from New England Biolabs or Vazyme. His-HRP antibody was obtained from ProteinTech Group. pBAD-ubiquitin (6TAG) was used as previously described[1]. Primers HRpNEUmaRS-For and HRpNEUmaRS-Rev were used to clone the MaBzK-RS to pNEU-maPylT for mammalian expression system. Primers 206W-For, 206W-Rev, HRpNEUmaRS-For, and HRpNEUmaRS-Rev were used to introduce mutation 206W in maPylRS WT and mutants. Primers MbRS-349W-For, MbRS-349W-Rev, HRpEVOL-MbRS-For, and HRpEVOL-MbRS-Rev were used to construct pEVOL-MbRS (349W).

sfGFP (151TAG)

MGSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPT
LVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDT
LVNRIELKGIDFKEDGNILGHKLEYNFNSHNVXITADKQKNGIKANFKIRHNVEDGSVQL
ADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELY
KGSHHHHHH

Red: amber codon TAG at 151th position.

Ubiquitin (6TAG)

MQIFVXTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQ
KESTLHLVLRLRGGHHHHHH

Red: amber codon TAG at 6th position

EGFP (182TAG)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPT
LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDT
LVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQL
ADHXQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELY
KHHHHHH

Red: amber codon TAG at 182th position

MaPylRS-WT

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLYSVMRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLVDMGPRGDATEV
LKNYISVVMKAAGLPDYDLVQEESDVYKETIDVEINGQEVCSAAVGPHYLDAAHDVHEP
WSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

MaBzKRS2 (aka MaPylRS-206W)

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLYSVMRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLVDMGPRGDATEV
LKNYISVVMKAAGLPDYDLVQEESDVWKETIDVEINGQEVCSAAVGPHYLDAAHDVHE
PWSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

Red: mutated residues.

MaBzKRS2-227I/228P

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLYSVMRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLVDMGPRGDATEV
LKNYISVVMKAAGLPDYDLVQEESDVWKETIDVEINGQEVCSAAVGPIPLDAAHDVHEP
WSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

Red: mutated residues.

MaBzKRS (aka MaBzKRS-Hit 2)

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLTSVRRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLHDMGPRGDATEV
LKNYISVVMKAAGLPDYDLVQEESDVYKETIDVEINGQEVCSAAVGPIPLDAAHDVHEP
WSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

Red: mutated residues.

MaBzKRS-Hit 9

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLYSVGRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLWDMGPRGDATE
VLKNYISVVMKAAGLPDYDLVQEESDVYKETIDVEINGQEVCSAAVGPIPLDAAHDVHE
PWSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

Red: mutated residues.

MaBzKRS-206W

MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPS
RHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPM
LAPNLTSVRRDLRDHTDGPVKIFEMGSCFRKESHSGMHLEEFTMLNLHDMGPRGDATEV
LKNYISVVMKAAGLPDYDLVQEESDVWKETIDVEINGQEVCSAAVGPIPLDAAHDVHEP
WSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN

Ma-tRNA^{Pyl}

GGGGGACGGTCCGGCGACCAGCGGGTCTCTAAAACCTAGCCAGCGGGGTTCGACGCC
CCGGTCTCTCGCCA

Synthetase library construction and BzKRS mutant selection

To identify an efficient synthetase for the incorporation of BzK, the primers *Ma*PylRS NdeI to *Ma*PylRS PstI were used to randomize the active site residues Y126, M129, and V168 of MaPylRS. Residue H227 and Y228 were mutated to Ile and Pro, respectively. The selection of an orthogonal synthetase for BzK incorporation was carried out by following the procedures described previously.[1] To identify the most efficient mutants, we applied 100 μg/mL of chloramphenicol in positive selections. A total of 200 colonies were subsequently screened in the presence and absence of 2 mM BzK, yielding 9 hits that exhibited BzK-dependent phenotype. Sequencing of these hits converged on two mutants (hit-2 and hit-9). They were recloned to pEVOL plasmid with primers HRpEVOL-For and HRpEVOL-Rev followed by measuring the incorporation efficiency using plasmid pBAD-sfGFP (151TAG). The incorporation efficiencies of the hits were compared by reading the green fluorescence intensity (excitation at 485 nm, emission at 528 nm) normalized to OD at 600 nm. As shown in Figure S1, although hit-9 had higher fluorescence intensity in the prescence of BzK, it also had higher background fluorescence in the absence of BzK. Hit-2 (named as MaBzKRS) showed high efficiency and low background, and was thus used in subsequent experiments.

Protein expression and purification

pBAD-ubiquitin (6TAG) or pBAD-sfGFP (151TAG) was co-transformed with pEvol-BzKRS into DH10b *E. coli* competent cells. The transformants were plated on LB agar plate supplemented with 35 μg/mL chloramphenicol and 50 μg/mL ampicillin. A single colony was picked and incubated into 2 mL 2xYT. After incubation at 37°C for 6 h, the cell culture was diluted 50 times and cultured at 37°C. When OD_{600} reached 0.6-0.8, the cell culture was added with 0.2% arabinose with or without 2 mM BzK, and then incubated at 37°C for 8 h. The cell pellets were harvested and resuspended in 15 mL lysis buffer containing 50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8.0 and then lysed by ultrasonication. The soluble fractions were collected and incubated with Ni^{2+}-NTA resin at 4°C for 2 h with constant mechanical rotation. The mixture was loaded into column and eluted with 300 μL elution buffer (20 mM Tris, 250 mM NaCl, 500 mM imidazole, pH 8.0). The eluates were buffer exchanged into 100 μL of PBS buffer. Proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie Brilliant Blue and Western blot with 1:10000 anti-his antibody (cat# HRP-66005) or 1:2000 anti-benzoylation antibody (PTM biolabs, Anti-benzoyllysine mouse mAb).

FACS analysis of BzK incorporation into HEK-293T cells

HEK-293T cells were plated with 1×10^5 cells per well in a 24-well plate and grown at 37°C in a CO_2 incubator overnight. The plamid pNEU-BzKRS (0.5 μg) and pcDNA 3.1EGFP (182TAG) (0.5 μg) with 3 μL polyethylenimine (PEI) in 0.5 mL DMEM media. Six hours post transfection, 2 mM BzK were added and cultured for 24-48 h. The cells were washed with PBS and the live cells were analyzed by flow cytometry using a FACS instrument (BD Biosciences, San Jose, CA, USA). The FlowJo (Tree Star Inc., Ashland, OR, USA) was used for the data analysis.

Fluorescence confocal microscopy of HEK-293T cells

HEK-293T cells were plated in a 6-well plate at densities of 3×10^5 cells per well and grown at 37°C in a CO2 incubator. After 24 h of incubation, the plamid pNEU-BzKRS (1.5 μg) and pcDNA 3.1 EGFP (182TAG) (1.5 μg) with 9 μL polyethylenimine (PEI) in 2 mL DMEM media and incubated for 6 h. Then the cells were treated with or without 2 mM BzK. After incubation for 48 h, the cells were visualized on a Nikon EclipseTi confocal microscope. Each stain was performed on three individual groups, and the photographs in the figures are representative of each group. The cells

were then rinsed thrice with PBS and lysed with RIPA buffer. The samples were separated on Western blot with 1:2000 anti-GFP monoclonal antibody (Proteintech #HRP-66005) or 1:2000 anti-benzoylation antibody (PTM biolabs, Anti-benzoyllysine mouse mAb). Anti-GAPDH antibody (Proteintech #HRP-60004) was used as a reference protein.

Mass spectrometric analysis

Mass spectrometric measurements were performed as previously described.[2] Briefly for electrospray ionization mass spectrometry, mass spectra of intact proteins were obtained using a QTOF Ultima (Waters) mass spectrometer, operating under positive electrospray ionization (+ESI) mode, connected to an LC-20AD (Shimadzu) liquid chromatography unit. Protein samples were separated from small molecules by reverse phase chromatography on a Waters Xbridge BEH C4 column (300 Å, 3.5 µm, 2.1 mm x 50 mm), using an acetonitrile gradient from 30-71.4%, with 0.1% formic acid. Each analysis was 25 min under constant flow rate of 0.2 mL/min at RT. Data were acquired from m/z 350 to 2500, at a rate of 1 sec/scan. Alternatively, spectra were acquired by Xevo G2-S QTOF on a Waters ACQUITY UPLC Protein BEH C4 reverse-phase column (300 Å, 1.7 µm, 2.1 mm x 150 mm). An acetonitrile gradient from 5%-95% was used with 0.1% formic acid, over a run time of 5 min and constant flow rate of 0.5 mL/min at RT. Spectrum were acquired from m/z 350 to 2000, at a rate of 1 sec/scan. The spectra were deconvoluted using maximum entropy in MassLynx.

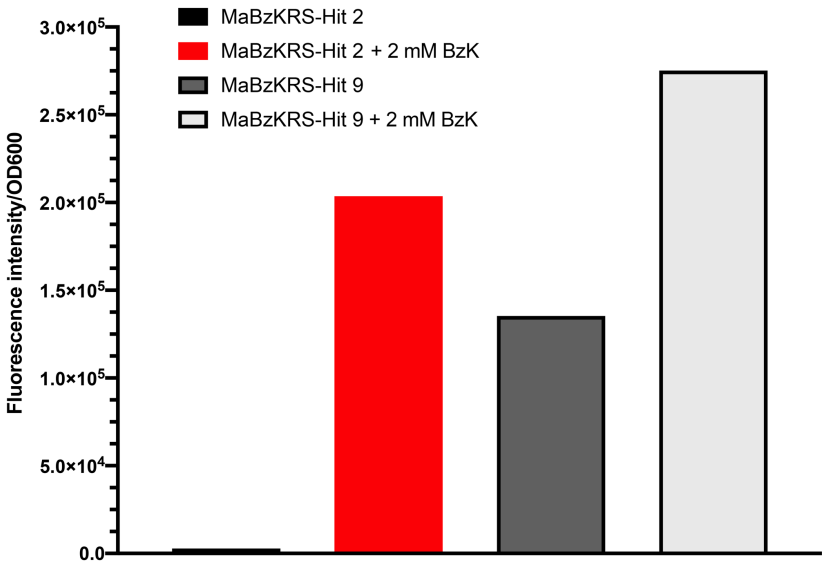

Figure S1. Comparison of the BzK incorporation efficiency of identified hits. Cells expressing pEvol-hitRS and pBAD-sfGFP (151TAG) were induced at 30 °C for 6 h.

Supplementary References

1. (a) Liu, J.; Zheng, F.; Cheng, R.; Li, S.; Rozovsky, S.; Wang, Q.; Wang, L., Site-Specific Incorporation of Selenocysteine Using an Expanded Genetic Code and Palladium-Mediated Chemical Deprotection. *J. Amer. Chem. Soc.* **2018,** *140* (28), 8807-8816; (b) Liu, J.; Cheng, R.; Wu, H.; Li, S.; Wang, P. G.; DeGrado, W. F.; Rozovsky, S.; Wang, L., Building and Breaking Bonds via a Compact S-Propargyl-Cysteine to Chemically Control Enzymes and Modify Proteins. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (39), 12702-12706.
2. Liu, J.; Chen, Q.; Rozovsky, S., Utilizing Selenocysteine for Expressed Protein Ligation and Bioconjugations. *J. Amer. Chem. Soc.* **2017,** *139* (9), 3430-3437.