An efficient cell-free protein synthesis platform for producing proteins with pyrrolysine-based non-canonical amino acids

Arnaz Ranji Charna[11,2] , Benjamin J Des Soye[1,2], Ioanni Ntai[2], Neil L. Kelleher[2,3,4], and Michael C Jewett[1,2,4,5*]

[1]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

[2]Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, USA

[3]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

[4]Center for Synthetic Biology, Northwestern University, Evanston, Illinois 60208, United States

[5]Member, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Northwestern University, 303 E. Superior, Chicago, IL 60611, USA

* Corresponding author: m-jewett@northwestern.edu

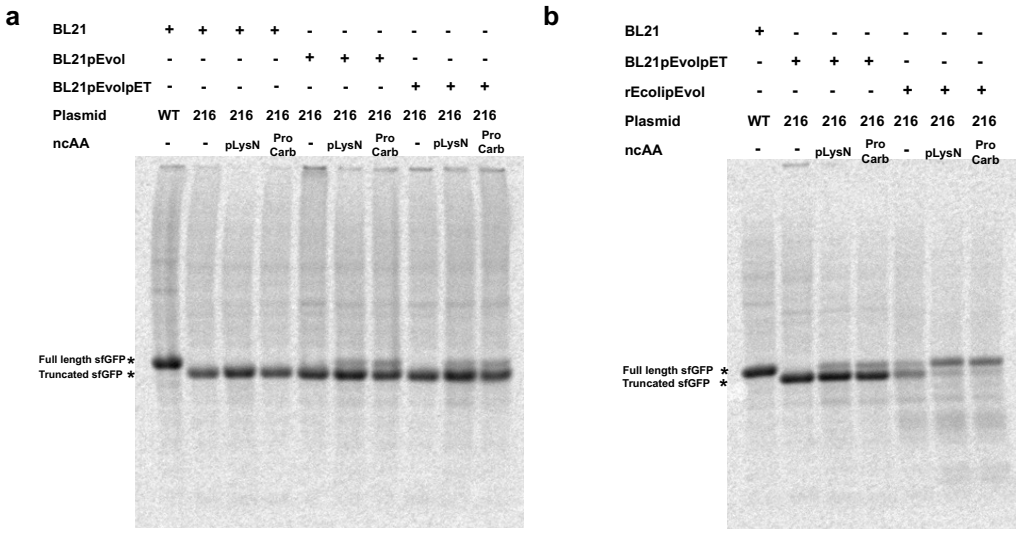
Tel: (+1) 847 467 5007; Fax (+1) 847 491 3728

Running header: Cell-free protein synthesis of pyrrolysine based ncAAs

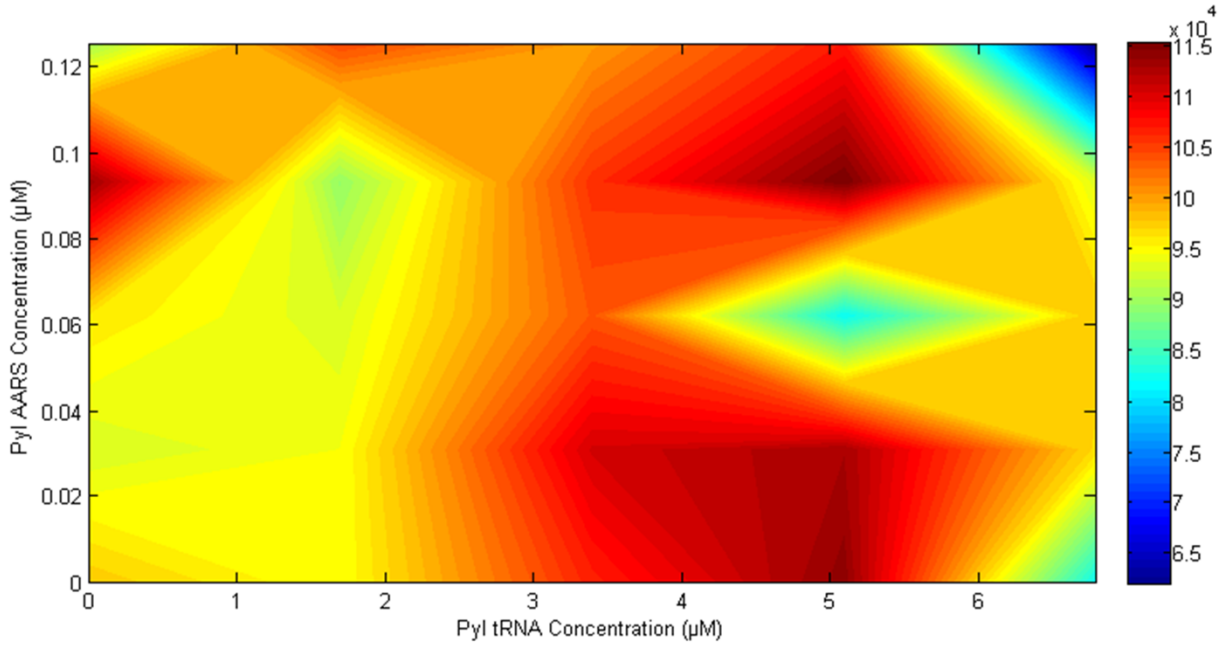
Supplementary Information

Supplementary Figures 1-2
Supplementary Tables 1-2
Supplementary Methods
Supplementary Sequences

Supplementary Figures



Supplementary Figure 1: Proteins produced in rEcoli extract have less truncated product than proteins produced in BL21 derived lysates. Autoradiogram of ^{14}C-leucine labeled proteins in (**a**) BL21 based extracts and (**b**) BL21 versus rEcoli extracts. Full length and truncated sfGFP proteins are indicated. WT indicates wt-sfGFP plasmid was used; 216 indicates sfGFP-T216 plasmid was used. Gels representative of three independent experiments.



Supplementary Figure 2: Addition of extra PylRS or Pyl $tRNA_{CUA}^{Pyl}$ does not significantly improve the incorporation of ncAAs in rEcoli based extracts. sfGFP yields (RFUs) of samples in rEcolipEvol extract are plotted using MatLab software with increasing PylRS concentration on the y-axis and increasing $tRNA_{CUA}$Pyl**.** Two independent reactions were performed for each data point.

Supplementary Tables

Supplementary Table 1. Monoisotopic masses of the different proteins in this work.

Figure	GFP species	mass (Da, Exper)	mass (Da, Theor)	error (ppm)	Shift from WT GFP (Da, Exper)	Shift from WT GFP (Da, Theor)
5a	WT GFP*	26847.52	26847.45	2.6	--	--
5a	Bl21pEvolpET extract: 1TAG + pLysN *	27010.63	27010.55	3.0	163.18	163.10
5a	Bl21pEvolpET extract: 1TAG + ProCarb*	26956.57	26956.51	2.2	109.12	109.05
5a	rEcolipEvol extract: 1TAG + pLysN	27141.7	27141.59	4.1	163.21	163.10
5a	rEcolipEvol extract: 1TAG + ProCarb *	26956.55	26956.51	1.5	109.10	109.05
5b	WT GFP*	26847.52	26847.45	2.6	--	--
5b	rEcolipEvol extract: 2TAG + pLysN	27318.78	27318.71	2.6	340.29	340.22
5b	rEcolipEvol extract: 3TAG + ProCarb	27306.71	27306.67	1.5	328.22	328.18

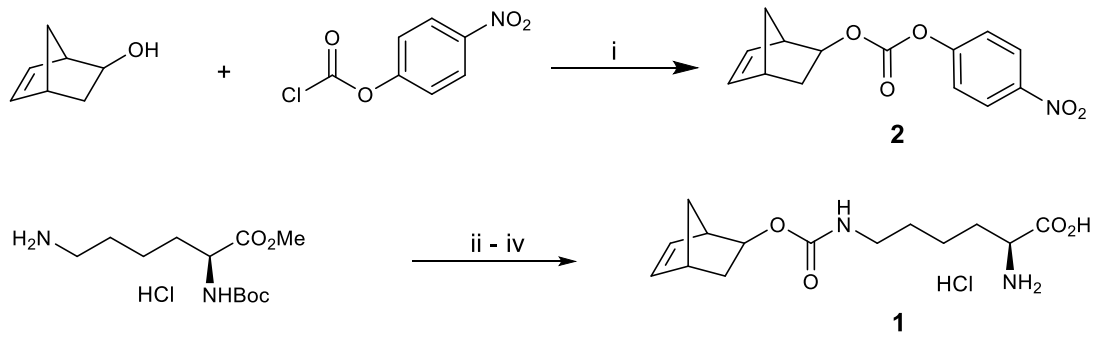
Supplementary Table 2. Oligonucleotide sequences used in this work.

Oligo name	Sequence
AR109	gatatccatatggataaaaaaccactaaacactctg
AR108	ggtagcgcggccgccaggttggtagaaatcccg
AR045	tggcggaaaccccgggaatc
GB1	gcttttagatcttaatacgactcactatagggagaccggctgatgagtccgtgaggacgaaacggtacccggtacc gtcggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgccagga agcttacatccgtcgaccaaaagc
T7500up	ccgaaggtaactggcttcagcagag
S2-f	tagaaaggtgaagaactgtttac
S2-r	catatgtatatctccttcttaaagttaaac

Supplementary methods

Synthesis of N^6-(5-Norbornen-2-yloxycarbonyl)-L-Lysine Hydrochloride

The synthesis of N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine (**1**) has previously been described[i], however the modified route shown in Scheme 1 was employed to avoid the use of phosgene. Condensation of p-nitrophenyl chloroformate with 5-norbornen-2-ol (mixture of isomers) in the presence of pyridine gave 5-norborn-5-en-2-yl (4-nitrophenyl) carbonate (**2**) in 28-35% isolated yield. Commercial N^2-Boc-L-lysine methyl ester hydrochloride was treated with **2** and triethylamine in DMF followed by sequential removal of the protecting groups to give compound **1** as the hydrochloride salt. Compound **1** is a ~ 3:1 mixture of the endo to exo isomers determined by ^{1}H NMR.



Scheme 1 - Reagents: i. pyridine/DCM, r.t.; ii. **2**, Et_3N, DMF; iii. LiOH, MeOH/H_2O; iv. 4M HCl/dioxane.

N^2-Boc-L-lysine methyl ester hydrochloride was purchased from ChemImpex Chemical Company and was used as received. All other reagents were purchased from Aldrich Chemical Company and were used as received. NMR spectra were obtained on Bruker Avance III Ag500 spectrometer at 500 MHz (proton) and 125 MHz (carbon). LCMS spectra were obtained using a Waters Acquity-H Single Quad UPLC-MS system, eluting on a Waters BEH C-18 column (50 x 2.1 mm, 1.7 micron) using 5 – 95% acetonitrile in water modified with 0.1% formic acid @ 0.8 mL/min over 2 minutes with a 1-minute hold at 95% acetonitrile in water.

5-norbornen-2-yl (4-nitrophenyl) carbonate (2):

A dry 100 mL flask was charged with 5-norbornylen-2-ol (0.441 g, 4 mmol), 4-nitrophenyl chloroformate (1.048 g, 5.20 mmol) and DCM (Volume: 24 ml). The flask was placed under nitrogen, cooled to 0 °C and pyridine (0.404 ml, 5.00 mmol) was added. The reaction was stirred at 0°C for 2 hours then allowed to warm to room temperature and stirred overnight. The reaction was diluted with DCM (30 mL) and washed with brine (3 x 50 mL). The organic phase was separated, dried over sodium sulfate, filtered and concentrated under vacuum. The residue was purified by flash column chromatorgraphy (Biotage Isolera, 50 g silica column, eluting with 0-100% DCM in hexanes over 10 CV @ 40 mL/min) to give a white solid determined by 1H NMR to be the product in a ~ 3:1 mixture of endo to exo isomers. ^{1}H NMR (500 MHz, CHLOROFORM-d) δ 8.25 – 8.32 (m, 2H), 7.35 – 7.43 (m, 2H), 6.44 (dd, *J* = 2.99, 5.52 Hz, 0.7H), 6.32* (dd, *J* = 2.84, 5.67 Hz, 0.2H), 6.07 (dd, *J* = 2.84, 5.67 Hz, 0.7H), 6.01* (dd, *J* = 3.31, 5.52 Hz, 0.2H), 5.38 (td, *J* = 3.27, 7.96 Hz, 0.7H), 4.75* (d, *J* = 6.94 Hz, 0.2H), 3.29 (br. S., 0.7H), 3.09* (br. S., 0.2H), 2.94 (br. S., 1H), 2.25 (ddd, *J* = 3.63, 8.43, 12.53 Hz, 0.7H), 1.83* (ddd, *J* = 2.52, 6.94, 12.93 Hz, 0.2H), 1.73 – 1.78* (m, 0.2H), 1.67 – 1.72* (m, 0.2H), 1.53 – 1.65 (m, 1.1H), 1.39 (d, *J* = 8.83 Hz, 0.7H), 1.14 (td, *J* = 3.15, 12.93 Hz, 0.7H) * indicates exo isomer.

N^2-t-butoxycarbonyl-N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine methyl ester:

A solution of N^2-Boc-L-Lysine methyl ester hydrochloride (0.757 g, 2.55 mmol) and 5-norbornen-2-yl (4-nitrophenyl) carbonate (0.772 g, 2.81 mmol) in dry DMF (Volume: 15 ml) was treated with triethylamine (1.066 ml, 7.65 mmol) and allowed to stir overnight at room temperature. The reaction was diluted with DCM (30 mL) and washed with water (5 x 25 mL). The organic phase was dried over sodium sulfate and filtered. The reaction was concentrated under vacuum and the residue taken up in DCM (3 mL). ~ 0.2 mL was loaded onto a 1g Biotage samplet and purified by

flash chromatographty (10 g KP-Sil column, 0-100 DCM in hexanes over 10 CV @ 12 mL/min followed by 0-10% methanol in DCM over 5 CV @ 12 mL/min) to give the product as a white waxy solid. With the column conditions determined, the remaining material was purified (Biotage Isolera, 25g KP-Sil column, 0-10% methanol in DCM over 10 CV @ 25 mL/min). The products from both columns were combined to give the product (0.6 g, 59%) as a waxy, white solid. ^{1}H NMR showed a ~ 3:1 mixture of isomers, endo to exo based on the starting carbonate. The product was slightly contaminated with DMF but was used as is in the next step. ^{1}H NMR (500 MHz, CHLOROFORM-d) δ 6.34 (dd, *J* = 2.84, 5.36 Hz, 0.7H), 6.24* (d, *J* = 2.52 Hz, 0.2H), 5.95 – 6.02 (m, 1H), 5.26 (d, *J* = 4.10 Hz, 0.7H), 5.11 (br. S., 0.7H), 4.64 (br. S., 1H), 4.30 (d, *J* = 4.73 Hz, 1H), 3.76 (s, 3H), 3.15 (br. S., 3H), 2.84 (br. S., 1H), 2.14 (ddd, *J* = 3.63, 8.35, 12.30 Hz, 0.7H), 1.80 (d, *J* = 5.36 Hz, 2H), 1.61 – 1.73 (m, 2H), 1.29 – 1.59 (m, 18H), 0.93 (d, *J* = 12.30 Hz, 0.7H) * indicates exo isomer. LCMS (*ESI+*) m/z calculated: 396; found: 397 (M+H), T_r = 2.16 min.

N^2-t-butoxycarbonyl-N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine:

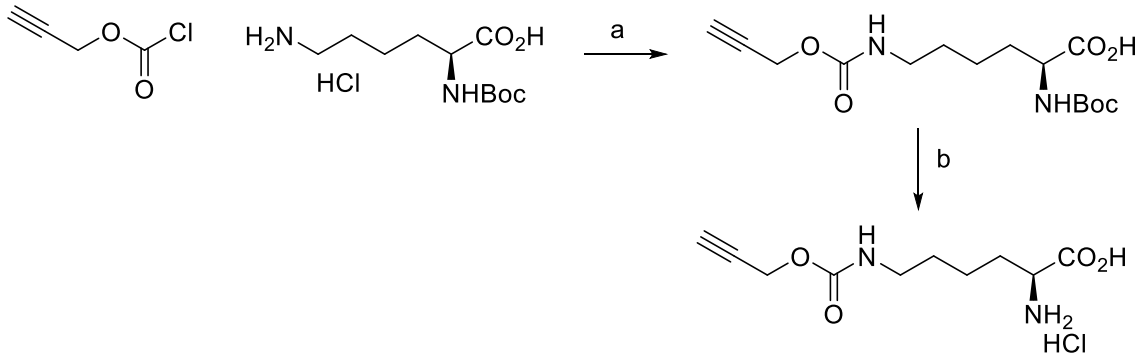
A solution of N^2-t-butoxycarbonyl-N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine methyl ester (0.60 g, 1.513 mmol) in methanol (8.07 ml) was treated with lithium hydroxide monohydrate (0.140 g, 3.33 mmol) in water (2.018 ml). The resulting mixture was stirred at room temperature for 2 hours then concentrated under vacuum to a volume of ~ 3 mL. Water (3 mL) was added and the resulting solution washed with methylene chloride (3 x 5 mL). The aqueous phase was acidified to ~ pH 4 with 1N HCl and extracted with ethyl ether (5 x 5 mL). The combined organic phases were dried over sodium sulfate, filtered and concentrated under vacuum to give the product (0.43 g, 74%) as a white solid. ^{1}H NMR (500 MHz, CHLOROFORM-d) δ 6.35 (d, *J* = 1.89 Hz, 0.7H), 6.25* (br. s., 0.2H), 5.99 (d, *J* = 2.52 Hz, 1H), 5.27 (d, *J* = 7.88 Hz, 1.3H), 4.57 - 4.83 (m, 1H), 4.13 - 4.41 (m, 1H), 3.16 (m, 3H), 2.85 (br. s., 1H), 2.15 (ddd, *J* = 3.63, 8.35, 12.30 Hz, 0.7H), 1.67 - 1.98 (m,

3H), 1.37 - 1.61 (m, 14H), 0.94 (d, *J* = 12.61 Hz, 0.7H) * exo isomer. LCMS (*ESI*+) m/z calculated: 382; found: 393 (M+H), T_r = 2.14 min.

N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine (1):

A solution of N^2-butoxycarbonyl-N^6-((5-norbornen-2-yl)oxy)carbonyl-L-lysine (0.425 g, 1.089 mmol) in DCM (4 mL) was treated with hydrogen chloride, 4.0M in dioxane (0.408 mL, 1.634 mmol) and stirred at room temperature for 1 hr. A white precipitate formed. LCMS showed ~ 50% conversion so additional hydrogen chloride, 4.0M in dioxane (0.408 ml, 1.634 mmol) was added. The reaction was stirred for 30 minutes at which point it was nearly complete by LCMS. The solvents were removed under vacuum and the white solid allowed to stand overnight to complete the reaction. Residual solvent and trace HCl were removed under high vacuum for 1 hour. Compound **1** was isolated as a white solid (0.35 g, 100%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 7.11* (t, *J* = 4.89 Hz, 0.2H), 6.92 - 7.02 (m, 0.7H), 6.29 - 6.37* (m, 0.2H), 6.27 (d, *J* = 2.84 Hz, 0.7H), 6.01* (dd, *J* = 3.15, 5.04 Hz, 0.2H), 5.94 (d, *J* = 2.21 Hz, 0.7H), 5.06 - 5.16 (m, 0.7H), 4.44 (d, *J* = 6.31 Hz, 0.2H), 3.82 (br. s., 1H), 3.05 (br. s., 0.7H), 2.92 (d, *J* = 5.36 Hz, 2H), 2.73 - 2.85 (m, 1.2H), 2.05 (ddd, *J* = 3.47, 8.28, 11.90 Hz, 0.7H), 1.76 (br. s., 2H), 1.19 - 1.64 (m, 6H), 0.80 (d, *J* = 12.30 Hz, 0.7H) ^{13}C NMR (125 MHz, DMSO-d_6) δ 171.1, 164.0*, 156.3, 156.2*, 140.9*, 138.2, 132.8*, 131.9, 74.1*, 74.0, 51.9, 47.2, 47.0*, 45.9*, 45.6, 41.7, 34.2*, 34.1, 29.7, 28.9, 21.6. * indicates exo isomer. LCMS (*ESI*+) m/z calculated: 282; found: 283 (M+H), T_r = 1.63 min.

Synthesis of N^6-(propargyloxycarbonyl)-L-Lysine Hydrochloride



Conditions: a) THF, NaOH (aq), 0C, b) 4M HCl (dioxane), CHCl3, 0C

(S)-2-((tert-butoxycarbonyl)amino)-6-(((prop-2-yn-1-yloxy)carbonyl)amino)hexanoic acid. A solution of N2-Boc-L-lysine (1.232 g, 5 mmol) in THF (12.5 ml) and 1M sodium hydroxide (12.5 ml, 12.5 mmol) was cooled to 0 °C and treated with propargyl chloroformate (0.585 ml, 6.00 mmol) dropwise over 5 minutes. The resulting mixture was stirred at room temperature overnight. The reaction was cooled to 0°C and washed with ethyl ether (2 x 50 mL). The aqueous phase was acidified to ~ pH 4-5 with 1N HCl then extracted with ethyl acetate (3 x 50 mL). The combined extracts were dried over sodium sulfate, filtered and concentrated under vacuum to give the product as a colorless oil (1.42 g, 86%). 1H NMR (500 MHz, CHLOROFORM-d) δ 6.23 (br. s., 1H), 5.23 (d, J = 6.62 Hz, 1H), 4.97 (br. s., 1H), 4.71 (d, J = 1.89 Hz, 2H), 4.25 - 4.40 (m, 1H), 4.14 (bs, 1H), 3.24 (d, J = 5.67 Hz, 2H), 2.50 (br. s., 1H), 1.67 - 1.95 (m, 2H), 1.52 - 1.63 (m, 2H), 1.48 (s, 9H); 13C NMR (125 MHz, CDCl3) δ 155.99, 155.73, 80.46, 78.26, 74.68, 53.10, 52.53, 40.55, 31.64, 29.28, 28.34, 22.24, 14.23; LCMS (Method A) r.t. 1.21 min, *m/z* 329.5 (M+H).

(S)-2-amino-6-(((prop-2-yn-1-yloxy)carbonyl)amino)hexanoic acid HCl. A solution of (S)-2-((tert-butoxycarbonyl)amino)-6-(((prop-2-yn-1-yloxy)carbonyl)amino)hexanoic acid (1.420 g, 4.32 mmol) in chloroform (24 ml) was cooled to 0 °C. 4M hydrogen chloride (dioxane) (6.49 ml, 25.9

mmol) was added dropwise and the reaction allowed to warm to room temperature. After 1 hour LCMS showed no Boc cleavage. Additional HCl - dioxane (10 mL) was added and stirring continued for 3 days. The reaction was complete by LCMS. The volatiles were removed under vacuum and the resulting white powder dried for 2 hours @ 0.5 Torr. 1H NMR 500 MHz, DEUTERIUM OXIDE) δ 4.53 (s, 2H), 3.87 (t, *J* = 6.31 Hz, 1H), 3.02 (t, *J* = 6.62 Hz, 2H), 2.75 (br. s., 1H), 1.72 - 1.92 (m, 2H), 1.39 - 1.48 (m, 2H), 1.22 - 1.38 (m, 2H); 13C NMR (125 MHz, D2O) δ 172.67, 157.62, 100.00, 78.45, 75.44, 53.09, 52.29, 39.85, 29.39, 28.22, 21.32; LCMS (Method A) r.t. 0.84 min, *m/z* 229 (M+H).

Supplementary Sequences

pY71-GB1f (Transzyme Plasmid). In green is the transzyme sequence. In blue is the Kan resistance sequence.

GGATCCTGCAGTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC
AGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG
TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGG
GGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG
GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG
AAACGAATTCAGATCTTAATACGACTCACTATAGGGAGACCGGCTGATGAGTCCGTGAGGACGAAAC
GGTACCCGGTACCGTCGGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTT
AGATTCCCGGGGTTTCCGCCAGGAAGCTTACATCCGTCGACCGGCTGCTAACAAAGCCCGAAAGGA
AGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGT
CTTGAGGGGTTTTTTGCTGAAAGCCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCA
ATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT
CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACAT
CAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACG
ACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATT
ACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGA
CGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACA
CTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTC
CCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGA
AGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC
TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCT
GATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCG
CGGCTTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAG

CAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACA
CAACGT

pEVOL-Pyl. In green is Pyl aaRS.

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCG
GAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCAAAATCTATATTGAAATGGCA
TGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGCAGGACTGCAAGAGCGCTCAGG
CACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGT
TCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCCTAC
CAGAACGAAAAAGGCAATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGAATAC
AGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTTTCACCTGCGATACCGGTTTCCACC
CAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAGCATTTCTACAGGAGC
AACTGCATCCGCACTGGTAAAAGGGAATACGAACCCCATTACATCCATGTCTGCCCCTGTT
CAGGCAAGTGCCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAAC
CCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGC
TCTCTCGCAGAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTATCTGG
GGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATC
CCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
TCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACC
TTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAATTTTTGAAATA
GGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTG
AACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACT
TCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGATTCCTGCATGGTCTATGGGGA
TACCCTTGATGTAATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCG
CTTGACCGGGAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGC
CTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTA
TAACGGGATTTCTACCAACCTGTAAGTCGACCATCATCATCATCATTGAGTTTAAACGG
TCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCA
GAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCA
CCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCT
CCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGA
CTGGGCCTTGTTTGTGAGCTCCCGGTCATCAATCATCCCCATAATCCTTGTTAGATTATCAA
TTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTTAATATCATACGCCGTTATACGTTGTT
TACGCTTTGAGGAATCCCATATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGG
GCTCTGGATGTCCAGGACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGC
AAAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGCA
GGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTT
CGGATGAGGATCTCAATAAGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAAAAGT
CAAGGTCGTTTCTGCCCCTACCAGAACGAAAAAGGCAATGCCAAAATCCGTTGCGAGAGC
CCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTTTCA
CCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAA
TATCAAGCATTTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAACCCCAT
TACATCCATGTCTGCCCCTGTTCAGGCAAGTGCCCCCGCACTTACGAAGAGCCAGACTGA
CAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTC
AGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAAGACCTGCAGCAGATCTACGCG
GAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACA
GGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGG
CATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGA

GACCCATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGA
TCCAATAAAAATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACAC
CTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAAT
CTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCG
ATTCCTGCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAACTTTCCTC
TGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCCTGGATAGG
GGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGA
GCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAACTGCAGTTTCAAA
CGCTAAATTGCCTGATGCGCTACGCTTATCAGGCCTACATGATCTCTGCAATATATTGAGTT
TGCGTGCTTTTGTAGGCCGGATAAGGCGTTCACGCCGCATCCGGCAAGAAACAGCAAACA
ATCCAAAACGCCGCGTTCAGCGGCGTTTTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCG
CAACATGTGAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATG
CCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGC
ACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGG
ATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGAAACCTGATCATGTAG
ATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCAAATTCGA
AAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTT
TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAA
GGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTT
GTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAAT
CGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGG
GGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGA
TTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACC
GTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCA
CTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACAC
TATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATC
AGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTT
TAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGA
AATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGAT
TTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGG
TAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATT
TTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTC
TGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGC
CAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATC
AGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGA
CATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTG
AAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATA
CAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCG
AGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAA
CAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAA
GCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACC
GGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGG
TAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGG
CAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCT
AAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGA
GTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAG
CAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATA
TTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGAT
ATAAGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGACAA

CTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCC
CCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGAC
CGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTT
GGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGAC
GCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCA
GGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCG
ACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAG
CTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAA
TGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAG
TTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCG
CGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCA
CTCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGA
GATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTT
GGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAG
GGGATCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATT
GTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAA
CCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAA
CGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGT
CACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTT
TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTAGATC
T

pY71-sfGFP-1xTAG. In green is the amber codon.

TCTTTTTCGGCGATTTATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACATATGagcAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTC
TGGTGGAACTGGATGGCGATGTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGATG
CCACGATTGGCAAACTGACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCGA
CGCTGGTGACCACCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCA
CGATTTCTTTAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATG
GCAAATATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAA
AGGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCCATA
ATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCATAACGTT
GAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTG
CTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCGaacGAAAAAGGCTA
GCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTACGTGGAGCCATCCGCAGTT
CGAAAAATAAGTCGACCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCT
GAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTTGCTGAAAGCC
AATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATA
CCATATTTTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAAACTCACCGAGGCAGTTTCCATAGAT
GGCAAGATCCTTGATATCGTCTGGCGATTCGGACTCGTTCCAAACATCATACAACCTATTAGTTTCCC
CTTCGTCAAAATAACGTTTATCAAGTTGGATATCACATTGATGACGACTGGAATCGGGTGAGGTTGCC
AAAGGCCTTTAAGGCT

pY71-sfGFP-2xTAG. In green are the amber codons.

TCTTTTTCGGCGATTTATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACATATGTAGAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATT

CTGGTGGAACTGGATGGCGATGTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGAT
GCCACGATTGGCAAACTGACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCG
ACGCTGGTGACCACCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTC
ACGATTTCTTTAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGAT
GGCAAATATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGA
AAGGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCCAT
AATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCATAACGT
TGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGT
GCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCGaacGAAAAAGGCT
AGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTACGTGGAGCCATCCGCAGT
TCGAAAAATAAGTCGACCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTTGCTGAAAGC
CAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAAT
ACCATATTTTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAAACTCACCGAGGCAGTTTCCATAGA
TGGCAAGATCCTTGATATCGTCTGGCGATTCGGACTCGTTCCAAACATCATACAACCTATTAGTTTCC
CCTTCGTCAAAATAACGTTTATCAAGTTGGATATCACATTGATGACGACTGGAATCGGGTGAGGTTGC
CAAAGGCCTTTAAGGCT

pY71-sfGFP-3xTAG. In green are the amber codons.

TCTTTTTCGGCGATTTATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACATATGTAGAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATT
CTGGTGGAACTGGATGGCGATGTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGAT
GCCACGATTGGCAAACTGACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCG
ACGCTGGTGACCACCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTC
ACGATTTCTTTAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGAT
GGCAAATATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGA
AAGGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCCAT
AATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCATAACGT
TGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGT
GCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCGTAGGAAAAAGGC
TAGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTACGTGGAGCCATCCGCAG
TTCGAAAAATAAGTCGACCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG
CTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTTGCTGAAAG
CCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAA
TACCATATTTTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAAACTCACCGAGGCAGTTTCCATAG
ATGGCAAGATCCTTGATATCGTCTGGCGATTCGGACTCGTTCCAAACATCATACAACCTATTAGTTTC
CCCTTCGTCAAAATAACGTTTATCAAGTTGGATATCACATTGATGACGACTGGAATCGGGTGAGGTTG
CCAAAGGCCTTTAAGGCT