Supplementary Information for:

Expanding Anfinsen's Principle: Contributions of Synonymous Codon Selection to Rational Protein Design

Ian M. Sander, Julie L. Chaney and Patricia L. Clark

Department of Chemistry & Biochemistry, University of Notre Dame
251 Nieuwland Science Hall, Notre Dame, IN 46556 USA

1) Materials and Methods

2) Figure S1

3) Figure S2

4) Table S1

Materials and Methods

Construction of YKB. An *E. coli* expression plasmid, pYKB, was created that encodes the N-terminal portion of Venus[15] (residues 1-155), the C-terminal portion of ECFP[15] (residues 155-236) and the N-terminal portion of Cerulean[15] (residues 1-173), each separated by a flexible 15 aa $(AGQ)_5$ linker[16], cloned between the *Sac*I and *Hind*III sites of plasmid pQE-2 (Qiagen). We also constructed the control plasmids pYK (encoding the N-terminus of Venus followed by the C-terminus of ECFP) and pKB (encoding the C-terminus of ECFP followed by the N-terminus of Cerulean). Each of these constructs includes an N-terminal His-tag.

Construction of YKB synonymous mutants. *Ykb* synonymous mutant sequences were selected to cover a wide range of codon rarity[21] without affecting mRNA secondary structure stability.[22] Specifically, amino acids residues 319-337 of YKB ("GDATYGKLTLKFICTTGK"), corresponding to residues 35-53 of the third, "B" half-domain, were used as input for synonymous mRNA sequence development using the codon usage frequencies of *E. coli* with a 17 codon sliding window. Codon rarity for each synonymous mutant was taken as the minimum %MinMax score[21] of all codon windows that included part of the synonymous mutation region. A total of 100 nucleotides centered on the synonymous mutation region were used as input for the mFold RNA folding algorithm[22] to identify those that did not significantly alter the predicted mRNA secondary structure of *ykb* ($\leq \pm 2$ kcal/mol).

The sequence changes used to construct each synonymous YKB mutant are provided in Table S1. *ykb* (+65) was used as template for all overlap extension PCR reactions. All synonymous mutants were cloned into pQE-2 between the *Sac*I and *Hind*III restriction sites.

YKB expression and conversion of fluorescence ratios to [YK]/[KB] molar ratios. *E. coli* was transformed with pYK or pKB. A single colony of each was grown overnight at 37°C in M9 4X complete media (160 µg/mL each of 20 common amino acids (Sigma, cell culture grade), 16 µg/mL p-aminobenzoic acid and p-hydroxybenzoic acid (MP Biomedical), 160 µg/mL uridine (Sigma), 160 µg/mL hypoxanthine (MP Biomedical), 1 mM MgSO4, 6% (w/v) Na2HPO4, 3% (w/v) KH2PO4, 0.5% (w/v) NaCl, 0.1% (w/v) NH4Cl, 2 µg/mL each D-biotin (MP Biomedical), nicotinamide (Sigma) and thiamine (MP Biomedical), and 200 ng/mL riboflavin (Sigma)) supplemented with 0.4% glucose and 100 µg/mL ampicillin, diluted 1:200 into 25 mL of fresh media and grown at 37°C. YKB expression was induced by addition of isopropyl ®-D-1-thiogalactopyranoside (IPTG) at a final concentration of 500 µM and grown at 37°C for 2 hours. Cells were harvested by centrifugation at 5000x*g* for 10 minutes and resuspended in 5 ml PBS (150 mM NaCl, 7.5 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.5 mM KCl, pH 7.5) per 1.0 OD_{600} of cells. Resuspended cells were stored at 4°C for at least 72 hours to allow chromophore formation to reach completion. After normalizing for soluble protein level (as determined by quantitative western blotting; see below), cells expressing each construct were mixed together in various ratios of YK:KB, and fluorescence emission spectra were measured from 450 to 550 nm using a PTI QM-6 fluorimeter (Birmingham, NJ). An excitation wavelength of 439 nm was used, and slit widths were 1 nm for excitation and 5 nm for emission. The emission spectrum of an equivalent number of cells expressing only the empty pQE-2 vector was subtracted from each sample spectrum. A standard curve was calculated from fluorescence emission intensities at 473 nm and 515 nm at each YK:KB ratio, and fluorescence emission ratio was plotted against the

[YK]/[KB] molar ratio and used to convert measured fluorescence ratios into molar ratios for formation of YK versus KB.

Refolding of YKB from inclusion bodies. *E. coli* transformed with pYKB were grown, induced, harvested and resuspended in PBS as described above. Cells were lysed by three cycles of freezing at -80°C for 30 min followed by thawing in a room temperature water bath. After the first thaw, cells were treated with 1 mg/mL lysozyme. Lysates were treated with 25 µL 1 M $MgSO_4$ and 25 U RNase-free DNaseI (Ambion), centrifuged at 21000x*g* for 10 min and supernatant was discarded. Pellets were resuspended in PBS containing 6 M GdnHCl and shaken overnight at 4°C. Resuspended pellets were centrifuged at 21000x*g* and the supernatant was removed, incubated with freshly prepared dithiothreitol (DTT) to a final concentration of 1 mM, and heated to 95°C for 15 min. After heating, this chemically denatured YKB was rapidly diluted 1:100 into refolding buffer (35 mM KCl, 2 mM MgCl2, 50 mM Tris (pH 7.5), 5 mM DTT) at room temperature and incubated at 4°C overnight. Fluorescence emission spectra were measured as described above.

Expression level quantification. YKB expression levels were calculated by quantitative western blotting of whole cell lysates prepared for each synonymous mutant, using an anti-pentaHis mouse monoclonal antibody (Qiagen). The intensity of each YKB band was quantified using ImageJ. Results from at least three independent experiments on different blots were averaged and normalized to the intensity for YKB (+65).

Pulse-chase analysis of relative translation rate. *E. coli* bearing plasmids pYKB(+65) and pYKB(-54) were grown overnight at 37°C in M9 4X complete media supplemented with 0.4% glucose and 100 µg/mL ampicillin as described above. Overnight cultures were diluted 1:100 into fresh media and grown at 37°C to an OD600 of 0.6-0.8. An OD600-normalized 5 mL aliquot was centrifuged at 4000x*g* for 10 minutes at room temperature in a 50 mL conical tube. Cell pellets were washed by resuspension in 25 mL PBS followed by centrifugation. Washed cell pellets were resuspended in 5 mL of M9 4X complete media supplemented with 0.4% glucose lacking methionine, pre-warmed to 37°C. YKB expression was induced by addition of 1 mM final IPTG for 15 minutes. Induced cultures were pulsed with 200 µCi of ^{35}S-methionine (MP Biomedical) for 20 sec, followed by addition of 50 µg of unlabelled methionine. Aliquots (1 mL) were removed at indicated time points, and added to 10 µL 35 mg/mL ice-cold chloramphenicol. Each aliquot was centrifuged at 10,000xg for 10 min. The pellets were boiled in 6% SDS and 4.5 mM DTT for 20 minutes, and cellular proteins were separated by SDS-PAGE. A photographic image was collected with an IVIS Lumina imaging system (Perkin Elmer), and a luminescent image was collected using a phosphor screen. The luminescent image was quantified using ImageJ as described above for western blot quantification.

Calculation of average cognate [tRNA]. Relative charged tRNA concentrations for each codon were taken from Dong *et al*. (1996) *J. Mol. Biol.* **260**:649-663, and averaged for the percentage of total charged tRNA available for each codon in the synonymous mutant region.

Calculation of wobble velocity value. Numbers of tRNA genes in E. coli K12 were obtained from the Genomic tRNA Database (http://gtrnadb.ucsc.edu). NNNv, the wobble velocity value, for each codon in YKB was calculated using the formulas for bacterial codons provided.[20]

NNNv values were averaged over a sliding window of 17 codons. The lowest window value for each synonymous mutant is reported.

Table S1. YKB synonymous sequences.

Name	Amino Acid Sequence[a]	Nucleotide Sequence
(+65)	GDATYGKLTLKFICTTG	GGCGATGCAACCTATGGTAAACTGACGCTGAAATTTATCTGCACCACCGGTAAA
(+37)	GDATYGKLTLKFICTTG	GGCGATGCAACCTATGGTAAACTAACACTGAAATTTATCTGCACCACCGGTAAA
(-22)	GDATYGKLTLKFICTTG	GGTGACGCAACATATGGAAAGCTCACTCTTAAGTTTATCTGCACGACCGGTAAA
(-42)	GDATYGKLTLKFICTTG	GGAGACGCAACGTACGGTAAGTTAACTCTTAAGTTCATATGTACCACTGGTAAA
(-54)	GDATYGKLTLKFICTTG	GGCGATGCTACATATGGGAAGCTAACACTAAAGTTTATATGTACAACTGGTAAA
(-54)	GDATYGKLTLKFICTTG	GGAGACGCAACGTACGGTAAGTTAACTCTTAAGTTCATATGTACCACTGGAAAG
(-64)	GDATYGKLTLKFICTTG	GGAGACGCTACATACGGAAAGTTAACACTTAAGTTCATATGTACTACCGGTAAA
(-76)	GDATYGKLTLKFICTTG	GGAGACGCTACATACGGAAAGTTAACACTTAAGTTCATATGTACTACCGGAAAG
(-100)	GDATYGKLTLKFICTTG	GGAGACGCTACATACGGAAAGCTAACACTAAAGTTCATATGTACAACAGGAAAG
(+65)	RYPDHMKQHDFFKSAMPE	CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
Y_cKB	RYPDHMKQHDFFKSAMPE	CGCTATCCCGATCATATGAAACAGCATGATTTCTTTAAGAGCGCGATGCCGGAA

[a] Sequences beginning "GDA..." corresponds to YKB residues 319-337; sequences beginning "RYP..." correspond to YKB residues 88-106.

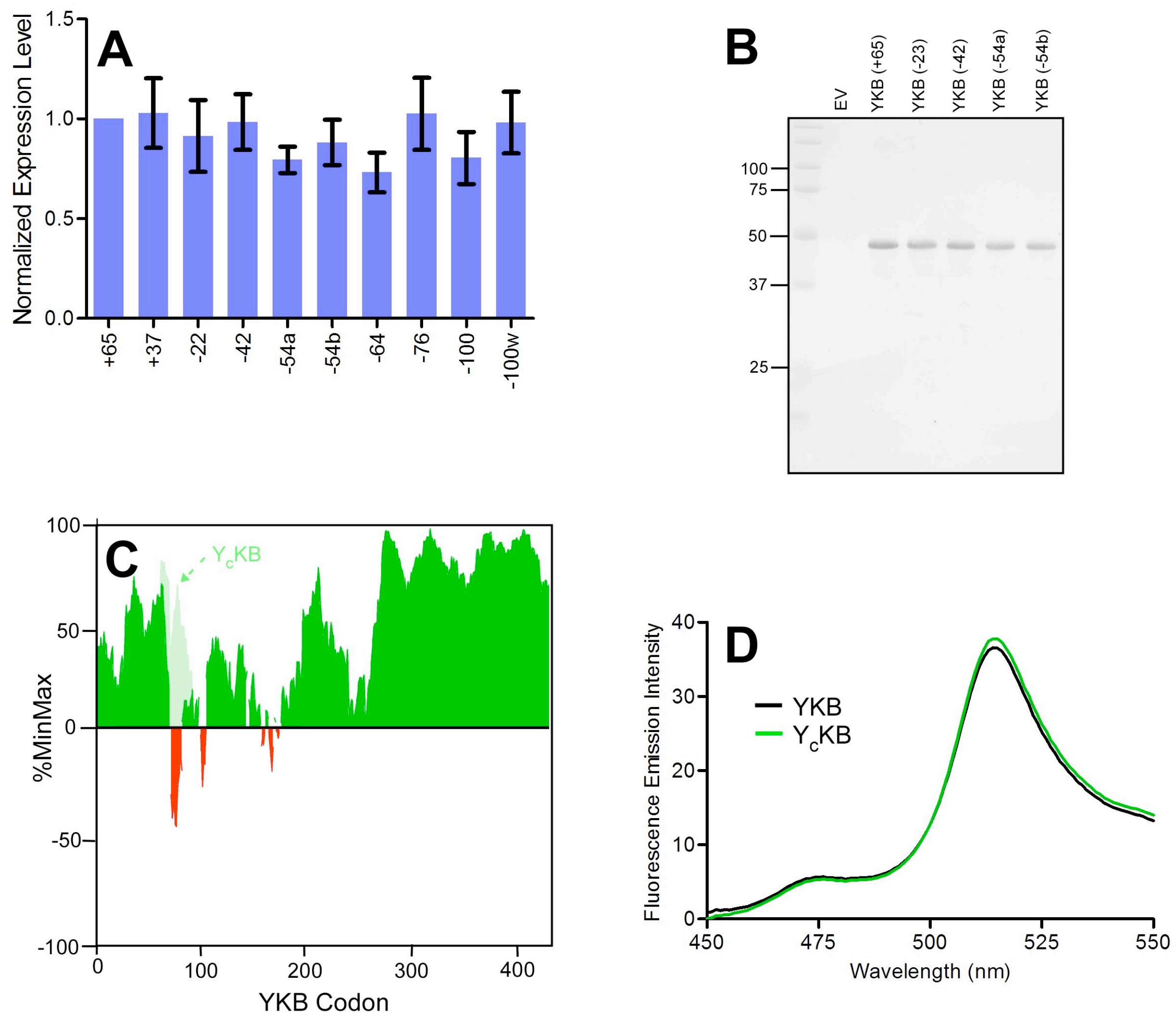

Figure S1. Control experiments for YKB expression in *E. coli*. (A) Total cellular protein levels for YKB synonymous mutants are similar. Error bars represent standard error of at least three replicates. (B) Representative anti-His5 western blot of whole cell lysates from *E. coli* expressing the indicated YKB synonymous mutants. (C) Relative codon usage (22) for YKB (+65) versus YcKB (+65) (light green). (D) Fluorescence emission spectra of *E. coli* expressing YKB (+65) (black) or YcKB (+65) (green).

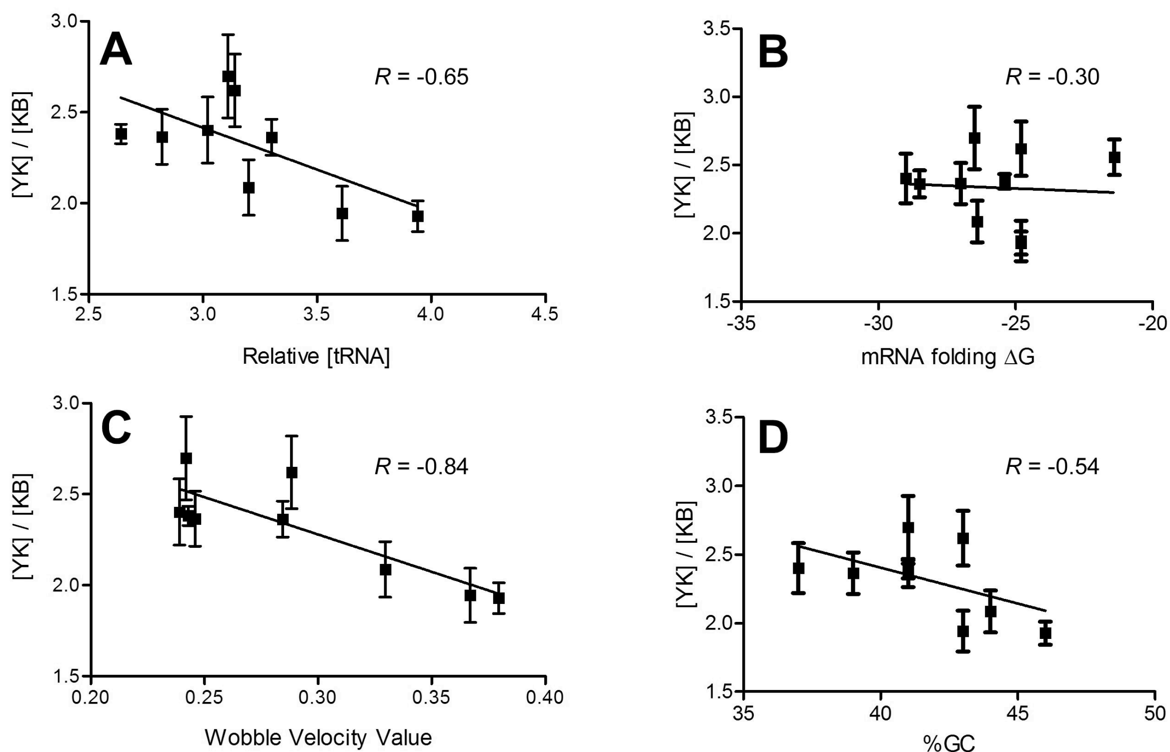

Figure S2. Correlations between [YK]/KB] molar ratios and other mRNA sequence and structural features. (A) Correlation with average cognate tRNA concentration[23] in the synonymous codon mutation region (Pearson correlation coefficient R = -0.65, p-value = 0.05). (B) Correlation with predicted mRNA stability[22] of 100 nt centered on the synonymous mutation region (R = -0.30, p-value = 0.4). (C) Correlation with wobble base velocity value[21] in the synonymous mutation region (R = -0.84, p-value = 0.007). (D) Correlation with %GC content of the synonymous switch region. (R = -0.54, p-value = 0.1).