Supporting Information

Woolstenhulme et al. 10.1073/pnas.1219536110

SI Experimental Procedures

Plasmid Construction. The two-hybrid selection uses three plasmids. The first, pSP100, has a pCDF origin and expresses tmRNA-DD and β-lactamase. The second, pTRG, contains the ColE1 origin and encodes the alpha subunit of RNA polymerase (RpoA) and Tet repressor (TetR) (Agilent Technologies). We amplified the first 117 codons of the *sspB* gene using the primers CCGCAAGAATTCAGATGGATTTGTCACAGCTAACACCACGTCG and GATCTCACTAGTTTACATGATGCTGGTATCTTCATCGTAGGCAGC. This PCR product subsequently was cloned at the 3′ end of the *rpoA* gene in pTRG with EcoRI and SpeI.

The cI control constructs used in Fig. 1*B* (cI-tag, cI-Glu-Pro-stop, and cI-nonstop) were expressed from the pBT vector [chloramphenicol resistance (CamR) and p15A origin; Agilent Technologies]. The cI-alone construct is simply the original pBT plasmid.

The primers TTGGCGCGGCCGCAGGGGAGCCAGCCGCAAACGACGAAAACTACGCTTTAGACGACTAAGATCTTAGGCG and CGCCTAAGATCTTAGTCGTCTAAAGCGTAGTTTTCGTCGTTTGCGGCTGGCTCCCCTGCGGCCGCGCCAA were used to amplify the cI gene and add the tmRNA tag (AANDENYALDD) to the C terminus of cI.

The primers TTGGCGCGGCCGCATCTGAACCGTGACTAAGATCTTAGGCG and CGCCTAAGATCTTAGTCACGGTTCAGATGCGGCCGCGCCAA were used to add the Glu-Pro-stop sequence.

The primers GATGATCGGCCGGCAGCCCGCCTAATGAGCGGGCTTTTTTTAGATCTGATGAT and ATCATCAGATCTAAAAAAAGCCCGCTCATTAGGCGGGCTGCCGGCCGATCATC were used to add the *trpA* terminator after the cI gene to create an mRNA transcript of defined length without a stop codon (cI-nonstop).

These three PCR products were ligated into the pBT plasmid following digestion with EagI and BglII.

cI clones were overexpressed for purification for MS analysis. The gene encoding the cI-stalling motif fusion was PCR amplified from the pBT plasmid using the following primers: GATATACCATGGGCAGCACAAAAAAGAAACCATTAACACAAG and GCAGCCGGATCCCCGGCGCGCCTAAGATCT. These PCR products were digested with NcoI and BamHI and cloned into the pET15b vector. The resulting plasmids and tmRNA-His_6 plasmid pCH201 (1) were used to transform BL21 (DE3) *Escherichia coli* cells. The cI protein was expressed, purified, and analyzed as described previously (2).

Library Construction. Twenty random codons were added to the 3′ end of the lambda cI protein by PCR with the primers GATAAAAATATTTCTAGATTTCAGTGCAATTTATCTCT and the reverse primer TTATGCAGATCTTTACTTACTTAN_{60}TGCGGCCGCGCCAAACGTCTC where "N" is an equal mixture of all four bases. The PCR product was cloned into pBT using XbaI and BglII, and the resulting plasmids were amplified in XL1-Blue. Thirty percent mutagenesis libraries of individual clones were created with the same scheme, except that the reverse primer contained the motif sequence, not N_{60}. The motif sequence was synthesized with phosphoramidite mixtures containing 70% of the original nucleotide and 10% of each of the other three nucleotides.

In Vitro Translation Constructs. All toeprinting DNA templates start with the following 5′ sequence, including a T7 promoter, ribosome binding site, and start codon (underlined): CTGTACATTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATG. The 3′ end of all templates includes the following primer binding site; the DNA primer in all of the experiments was NV1, GTTAATAAGCAAAATTCATTATAACC. To PCR amplify the cI clones for toeprinting, including 18 amino acids of cI and the 20-codon variable region, we used the 3′ primer GGTTATAATGAATTTTGCTTATTAACGGTAGCCAGCAGCATCCT, which places the end of the random sequence 40 nt from the NV1 primer binding site.

For the toeprints shown in Figs. 2 and 4, the 5′ primer was CTGTACATTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGTGTTCCGTTGTGGGGAAAGTTAT.

For toeprinting analysis of PPP stalling, three endogenous genes were PCR amplified from genomic DNA. The constructs were synthesized with the upstream sequence shown above, so that translation starts with the natural ATG codon. The NV1 primer binding site was added 51 nt downstream of the second Pro codon in the PPP motif in the gene. The primers used were

lepA: AGCTACCGGCCGTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGAAGAATATACGTAACTTTTCGATCATAGCTCAC and AGCTACCTCGAGGGTTATAATGAATTTTGCTTATTAACCCATGAGTCGATAATTAGTGCCTGCAACGG;
ligT: AGCTACCGGCCGTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGTCTGAACCGCAACGTCTGTTCTTTGCT and AGCTACCTCGAGGGTTATAATGAATTTTGCTTATTAACGGAGGCGTAAAGGGTGAACTCCGTCACCGC; and
amiB: AGCTACCGGCCGTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGATGTATCGCATCAGAAATTGGTTGGTAGC and AGCTACCTCGAGGGTTATAATGAATTTTGCTTATTAACAACCGCAGGCGTTTCAACGCGTTTCGCAAC.

Analysis of WPPP and FxxYxIWPPP was performed with the following dsDNA constructs: For WPPP, CTGTACATTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGACCATGATTACGAATTCGAGCTCATGGCCACCGCCATCGATTCGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGTTAATAAGCAAAATTCATTATAACC encodes MTMITNSSSWPPPSIRHASLALAVVLQRRVNKQNSL*; for the whole motif, the construct, CTGTACATTAATACGACTCACTATAGGGAGATTTTATAAGGAGGAAAAAATATGACCATGATTACGAATTCGAGCTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTTGGCCACCGCCATCGATTCGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGTTAATAAGCAAAATTCATTATAACC encodes MTMITNSSSLQKRLFQKYGIWPPPSIRHASLALAVVLQRRVNKQNSL*.

Kinetics. ***Materials.*** *E. coli* MRE600 tightly coupled 70S ribosomes were prepared as described previously (3). Overexpressed native initiation factor 1 (IF1) and IF3 and His-tagged IF2 were purified as described (4). Amino-terminally His-tagged release factor 1 (RF1) and aminoacyl-tRNA synthetases were expressed and purified as previously described (5). His-tagged elongation factor Tu (EF-Tu) and elongation factor G (EF-G) were purified over Ni-NTA resin, and the His-tag was removed by tobacco etch virus protease, followed by a second passage over a Ni-NTA column (6). $tRNA^{fMet}$, $tRNA^{Phe}$, $tRNA^{Arg}$, and $tRNA^{Glu}$ were purchased from Chemical Block. $tRNA^{Pro}$ and $tRNA^{Ala}$ were purified from bulk *E. coli* tRNA (Roche) with biotinylated

oligos Pro1 CCGAACGAAGTGCGCTACCAGGCTG3Bio-TEG, Pro2 CCCATGACGGTGCGCTACCAGGCTG3BioTEG, and Ala1 GCAAAGCAGGCGCTCTCCCAGCTGA3BioTEG as described (7). mRNA templates were synthesized by T7 RNA polymerase using DNA templates annealed to a short primer corresponding to the minimal promoter sequence. The mRNA transcripts have the following sequence: GGGUGUCUUGCG-AGGAUAAGUGCAUU**AUG**(X)UUUGCCCUUCUGUAGC-CA, in which the start codon is highlighted in bold and additional codons are inserted at the "X" site. UAA stop codons are used in termination motifs.

tRNA aminoacylation. Initiator $tRNA^{fMet}$ was aminoacylated with formylated radiolabeled [^{35}S]-methionine using MetRS and methionyl-tRNA formyltransferase as described (8). Pure tRNAs were charged by incubating the tRNA at 5 μM with the corresponding synthetase (~1 μM) in the presence of 2 mM ATP and 50 μM amino acid for 30 min at 37 °C in the following buffer: 100 mM Hepes·KOH (pH 7.6), 1 mM DTT, 10 mM KCl, 20 mM $MgCl_2$. The aminoacyl-tRNA was purified by extraction by phenol and $CHCl_3$, precipitated with ethanol, and resuspended in 2 mM NaOAc pH 5.0. The following tRNAs were aminoacylated as purified tRNAs: Pro-tRNA, Ala-tRNA, Arg-tRNA, Glu-tRNA, and Phe-tRNA. Bulk *E. coli* tRNA (Roche) was charged with a similar procedure in making the DTSA and DTSF aminoacyl-tRNA mixes for the DTS complex and related mutants, except the tRNA concentration was 100 μM, and the corresponding aminoacyl-tRNA synthetases and amino acids were added.

Ribosome complex formation. Initiation complexes were prepared by incubating 70S ribosomes (2 μM) with IF1, IF2, IF3, fMet-tRNA (3 μM each), and mRNA (6 μM) in polymix buffer with 2 mM GTP at 37 °C for 45 min. Then 100 μL of the resulting complex was reacted with 200 μL of preincubated mixture containing EF-Tu (15 μM), charged tRNA (2 μM), EF-G (2 μM), and GTP (2 mM) in polymix buffer for 37 °C for 5 min. The complexes then were purified over a 1-mL sucrose cushion [1.1 M sucrose, 20 mM Tris·HCl (pH 7.5), 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA] and spun at 260,000 × *g* in a TLA100.3 rotor for 2 h. The pellet was resuspended in polymix buffer [95 mM KCl, 5 mM NH_4Cl, 5 mM magnesium acetate, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 5 mM potassium phosphate (pH 7.5), 1 mM DTT], aliquoted, and stored at −80 °C.

Release assays. Approximately 25-nM ribosome nascent chain complexes (RNCs) were incubated with 5 μM RF1 in polymix buffer at 23 °C. The reaction was stopped at various time points by the addition of formic acid to a final concentration of 1%. Reactions with fast rate constants (>0.1 s^{-1}) were performed an RQF-3 quench-flow instrument (KinTek). Released peptides were separated from unreacted peptidyl-tRNA on cellulose TLC plates using electrophoretic TLC in pyridine-acetate buffer (pH 2.8) (9). The fraction of released peptide at each time point was quantified and plotted against time, and the data were fit with a single exponential equation. The reported rates are the averages of three separate experiments, with the SE given.

Peptidyl transfer assays. EF-Tu (40 μM) was incubated with 2 mM GTP in polymix buffer for 15 min at 37 °C to exchange GDP for GTP. This reaction then was combined with aminoacyl-tRNA to form a ternary complex, with final concentrations of 20 μM EF-Tu, 2 μM charged tRNA, and 2 mM GTP in polymix buffer. The complex was incubated for 10 min on ice and then reacted at 23 °C in polymix buffer with an equal volume of ~50 nM RNC. The reaction was stopped by the addition of KOH to a final concentration of 100 mM. Peptides were resolved on electrophoretic TLC and analyzed as above. The reported rates are the averages of three separate experiments, with the SE given.

Statistical Analysis of Peptide Libraries. We developed a likelihood ratio test to detect conserved motifs in sequences of surviving clones from the two-hybrid selection. Under the null hypothesis, we assume that the mutation rate is uniform across all bases in the sequence. Enriched motifs therefore are the locations that deviate from the overall frequencies across the entire 20-codon sequence. Specifically, we let $X_{i(k)}$ be the kth ordered base count at the ith position. We then define:

$$p_{(k)} = \frac{\sum_{i=1}^{N} X_{i(k)}}{\sum_{i=1}^{N}\sum_{j=1}^{4} X_{i(j)}} \quad \text{and} \quad r_{i(k)} = \frac{X_{i(k)}}{\sum_{j=1}^{4} X_{i(j)}},$$

where $p_{(k)}$ represents the estimate of the base frequency of the kth ordered base across all positions under the null hypothesis, and $r_{i(k)}$ estimates the kth ordered frequency at the ith at position. Assuming that the base positions are independent of each other and that the base occurrences across the sequences follow a multinomial distribution, we define our likelihood ratio test statistic for position i as

$$\Lambda_i = \frac{p_{(4)}^{X_{i(4)}} p_{(3)}^{X_{i(3)}} p_{(2)}^{X_{i(2)}} p_{(1)}^{X_{i(1)}}}{r_{i(4)}^{X_{i(4)}} r_{i(3)}^{X_{i(3)}} r_{i(2)}^{X_{i(2)}} r_{i(1)}^{X_{i(1)}}}.$$

We note that $-2\log(\Lambda_i)$ approximately follows a χ^2 distribution with three degrees of freedom. We evaluated the significance of each codon by summing the logged ratio statistics for all three positions in the codon and then appropriately comparing the codon-level statistic to a χ^2 distribution with nine degrees of freedom to obtain a P value.

1. Hayes CS, Bose B, Sauer RT (2002) Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in Escherichia coli. *Proc Natl Acad Sci USA* 99(6): 3440–3445.
2. Tanner DR, Cariello DA, Woolstenhulme CJ, Broadbent MA, Buskirk AR (2009) Genetic identification of nascent peptides that induce ribosome stalling. *J Biol Chem* 284(50): 34809–34818.
3. Shaw JJ, Green R (2007) Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol Cell* 28(3):458–467.
4. Brunelle JL, Youngman EM, Sharma D, Green R (2006) The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity. *RNA* 12(1):33–39.
5. Shimizu Y, et al. (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19(8):751–755.
6. Blanchard SC, Kim HD, Gonzalez RL, Jr., Puglisi JD, Chu S (2004) tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci USA* 101(35):12893–12898.
7. Yokogawa T, Kitamura Y, Nakamura D, Ohno S, Nishikawa K (2010) Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. *Nucleic Acids Res* 38(6):e89.
8. Walker SE, Fredrick K (2008) Preparation and evaluation of acylated tRNAs. *Methods* 44(2):81–86.
9. Youngman EM, Brunelle JL, Kochaniak AB, Green R (2004) The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* 117(5):589–599.

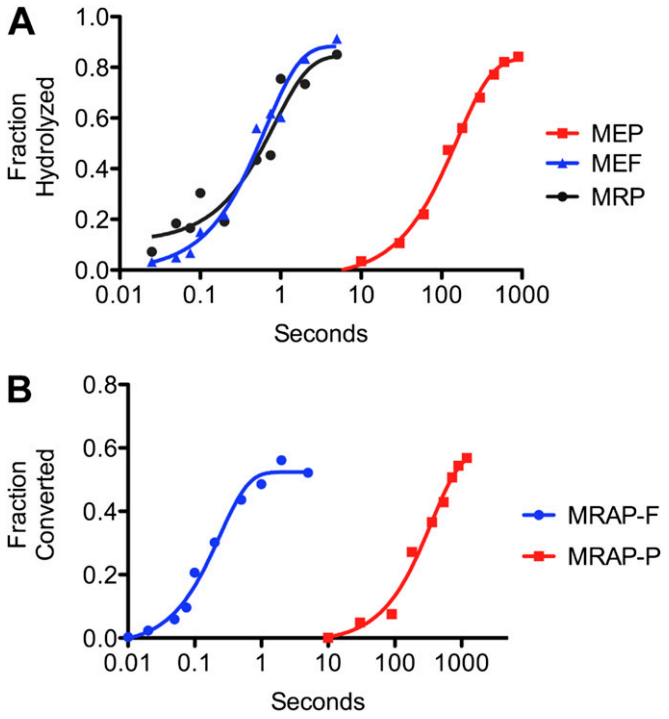

Fig. S1. Representative rate profiles from peptide release (*A*) and peptidyl transfer reactions (*B*). (*A*) Ribosome complexes containing a tripeptidyl-tRNA in the P site and a UAA stop codon in the A site were reacted with release factor 1. The fraction of the peptidyl-tRNA that is hydrolyzed is plotted against time. Curves for MEP (red), MEF (blue), and MRP (black) are shown. (*B*) Ribosome complexes containing a peptidyl-tRNA in the P site were reacted with a ternary complex of EF-Tu, GTP, and aminoacyl-tRNA. The fraction of tetrapeptide converted to pentapeptide is plotted against time. The reaction with Phe-tRNA is shown in blue, and Pro-tRNA is shown in red.

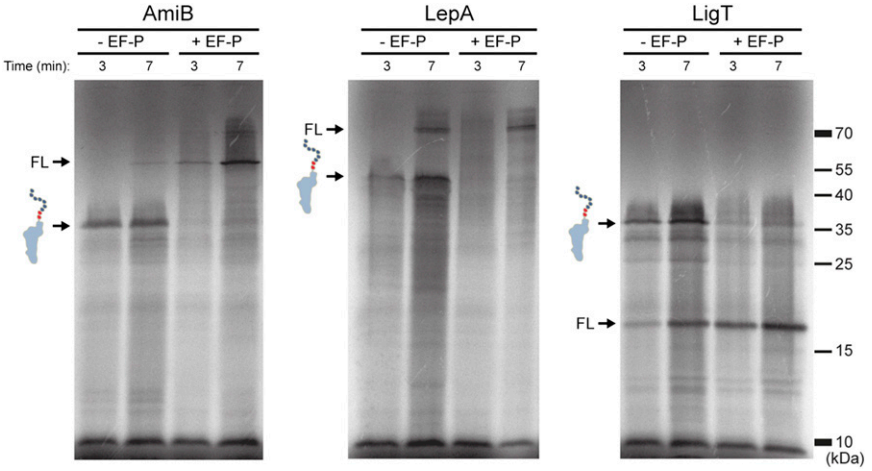

Fig. S2. Three endogenous *E. coli* genes, *amiB*, *lepA*, and *ligT,* were translated in vitro with [^{35}S]-methionine in the absence or presence of EF-P. The reactions were stopped after 3 or 7 min. Peptidyl-tRNAs are labeled with a cartoon, and the full-length, released protein is labeled "FL."

Table S1. Sequence of 10 stalling motifs

	Full 20mer peptide sequences
Elongation	
E1	SLKVVRQTYYPPRLSRSPPM*
E2	SVETGRVRFLLEHGPPIACI*
E3	EQVINLGPDEEWGATRKCVH*
E4	CEIKGYLLPLKIAPYSSLAK*
E5	KFGGTISCMQSLRDILELAA*
E6	FYGLLSDGGGKKRVNIPWSL*
Termination	
T1	LTKKGWEKREELLWILFHGT*
T2	GGIRGSYVLRTPNGGFWNSG*
T3	RVIIQTEEVWIKKQAKHDTS*
T4	RPHQRFVIPHVGFD*

Peptide sequences from the two-hybrid selection that show clear stalling sites with toeprinting assays. The underlined codon is found in the P site in toeprinting analyses. The full sequence is shown, but it is likely that not all these residues are essential for stalling. * corresponds to a stop codon.

Table S2. Pausing at the top 100 tripeptide motifs in the *E. coli* proteome

Rank	Motif	Pause score	Error	% error	No. of hits	Rank	Motif	Pause score	Error	% error	No. of hits
1	GGT	5.88	0.33	5.7	302	51	GTD	4.30	0.30	6.9	166
2	GGS	5.68	0.24	4.2	311	52	GGQ	4.27	0.31	7.4	194
3	GAG	5.44	0.26	4.7	521	53	GGH	4.25	0.38	9.0	103
4	WGP	5.38	1.09	20.2	25	54	CGY	4.24	0.91	21.5	34
5	GTC	5.31	0.81	15.2	25	55	GTN	4.23	0.56	13.2	94
6	MGT	5.30	0.78	14.7	85	56	MGI	4.23	0.40	9.5	126
7	GGA	5.28	0.30	5.6	436	57	GPT	4.21	0.41	9.7	118
8	GGG	5.28	0.24	4.6	444	58	GTY	4.21	0.36	8.6	79
9	GTI	5.21	0.47	8.9	221	59	GAI	4.20	0.21	5.1	402
10	GSD	5.13	0.57	11.1	186	60	TGT	4.19	0.33	7.8	235
11	GTT	5.09	0.66	13.0	225	61	GAS	4.19	0.17	4.1	375
12	GCG	5.05	0.42	8.3	128	62	DGP	4.19	0.38	9.1	111
13	GDG	5.05	0.46	9.1	233	63	GYG	4.16	0.31	7.4	193
14	GGV	4.96	0.24	4.8	429	64	EGY	4.16	0.59	14.3	151
15	GPN	4.94	0.59	12.1	79	65	GTV	4.16	0.25	6.1	299
16	GGF	4.93	0.35	7.1	221	66	GSS	4.16	0.29	7.0	230
17	GDC	4.91	0.65	13.2	33	67	GVG	4.16	0.19	4.6	418
18	GAT	4.87	0.27	5.6	370	68	GTA	4.15	0.42	10.1	328
19	GGD	4.85	0.31	6.3	244	69	WGA	4.15	0.49	11.9	78
20	GGI	4.78	0.20	4.3	357	70	VYG	4.13	0.86	20.8	158
21	CTF	4.73	1.12	23.6	22	71	GKS	4.13	0.21	5.2	299
22	GTG	4.71	0.24	5.1	362	72	APP	4.12	0.41	10.0	107
23	GVT	4.70	0.24	5.2	338	73	GTW	4.11	0.44	10.7	46
24	GWT	4.70	0.69	14.7	57	74	GSV	4.10	0.35	8.5	325
25	GST	4.68	0.35	7.5	222	75	GAW	4.09	0.67	16.4	95
26	GPG	4.68	0.30	6.3	167	76	GPI	4.08	0.34	8.4	101
27	GGM	4.64	0.41	8.9	182	77	GAD	4.08	0.23	5.7	356
28	GCI	4.64	0.86	18.5	71	78	RR*	4.08	0.83	20.3	33
29	VA*	4.55	0.93	20.5	26	79	GTR	4.07	0.33	8.2	174
30	GSG	4.54	0.20	4.3	352	80	GAA	4.07	0.19	4.6	570
31	GGY	4.53	0.34	7.6	172	81	GCT	4.07	0.45	11.1	56
32	AGP	4.53	0.94	20.8	158	82	GWG	4.07	0.52	12.8	64
33	GAK	4.52	0.53	11.8	270	83	GFG	4.06	0.32	7.9	289
34	GTS	4.52	0.29	6.4	200	84	AGT	4.06	0.47	11.5	321
35	GSY	4.51	0.41	9.0	112	85	GIC	4.05	0.41	10.1	53
36	LS*	4.50	0.95	21.2	21	86	GSI	4.04	0.26	6.4	224
37	GTH	4.49	0.65	14.5	76	87	GVM	4.03	0.70	17.4	167
38	GDT	4.46	0.48	10.8	234	88	GWA	4.02	0.68	17.0	79
39	GGK	4.45	0.25	5.7	266	89	GSP	4.02	0.41	10.3	138
40	WGI	4.44	0.72	16.1	46	90	GIV	4.02	0.21	5.3	397
41	GIT	4.44	0.26	5.8	314	91	GSA	4.02	0.28	6.9	303
42	GIF	4.43	0.73	16.5	151	92	GTL	4.01	0.16	4.0	477
43	PPP	4.39	0.42	9.6	81	93	GIQ	4.01	0.30	7.5	138
44	GVY	4.38	0.57	13.0	154	94	DGS	4.00	0.29	7.4	257
45	GGC	4.37	0.40	9.2	74	95	GGN	4.00	0.44	11.0	207
46	GTF	4.36	0.40	9.1	138	96	KGT	3.99	0.36	9.1	132
47	GPQ	4.35	1.05	24.1	90	97	DGA	3.99	0.62	15.5	250
48	PPD	4.33	0.49	11.3	67	98	GEG	3.99	0.25	6.3	270
49	WGT	4.32	0.67	15.5	39	99	GCE	3.99	0.43	10.8	69
50	EPP	4.32	0.77	17.8	56	100	GKT	3.97	0.18	4.4	386

Pause scores were calculated for all 8,000 tripeptides using published ribosome-profiling data from *E. coli* MG1655 (1). The 100 tripeptides with the highest pause scores are listed. The pause score is the ribosome density at the three codons of the tripeptide motif divided by the median density for the opening reading frame. Tripeptides were excluded if they had fewer than 20 occurrences (hits) or had pause scores with higher than 25% error. * corresponds to a stop codon.

1. Li GW, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484(7395):538–541.