Supporting Information

Navon et al. 10.1073/pnas.1606518113

SI Methods

Bioinformatics. URSs were identified by counting the number of times each different sequence of a certain number (three, four, etc.) of amino acids (from here on called triplets or quadruplets, respectively) appears in a proteome or multiple proteome database (N_r). The expected number of each triplet (N_{ex}) was calculated by Eq. **S1**, which calculates the expected number of appearances of each triplet:

$$N_{ex} = F_{ex} \times N_{total}; F_{EX} = F_{Ai} \times F_{Aj} \times F_{Ak}, \quad \text{[S1]}$$

where F_{ex} is the expected frequency of each triplet composed of any three amino acids (*i*, *j*, and *k*, which can be different or identical), each with its own frequency in the proteome, and N_{total}, the total number of triplets in all of the proteomes used in the analysis. These equations can, of course, be modified for the analysis of larger sequences as well (quadruplets, quintuplets, etc.). A triplet was defined as a URS if the value of the ratio of the expected vs. real appearance number (N_{ex}/N_r) of a certain triplet is greater than 4 or at least 8 SD above the mean. The ratio of frequency values described above is close to unity when the expected number of times that a certain triplet appears is equal (or close to equal) to the number of times it actually appears in the database. If a sequence is completely absent ($N_r = 0$), its URS value becomes infinity, and thus is not included in graphical descriptions.

Embedding URSs into Target Proteins. Two protein targets were chosen for the study. The cyanobacterial *mntA* gene (NP_441238.1) was chosen because we had found that it was expressed at extremely high rates in a variety of vectors and strains of *E. coli*. GFP was chosen as the second target due to its intrinsic fluorescence, which enables efficient internal quantification at all levels of experimentation. Embedding of the URSs was performed at different sites of pET 20b$^+$ or pET 45b$^+$ expression vectors (Fig. S8*A*) or the pIVEX2.3-GFP plasmid (Fig. S8*B*) or pEGFP-N1 (Clontech; Fig. S8*C*) vector based on the minimal number of changes required (i.e., sites that already possessed at least one of the three residues in the URS). Multiple expression vectors were used in the study to ensure that changes in translation efficiency were not due to any single vector. Mutagenesis was executed by use of a QuikChange Site-Directed Mutagenesis Kit (Stratagene; Fig. S9), so that the total number of residues in the expressed protein was invariant. Mutagenesis was evaluated by sequencing (Macrogen or the Inter-Departmental Equipment Multi-Disciplinary Laboratories Unit, Technion). All of the constructs containing a mutated *mntA* gene were transformed to the BL21(DE3)pLysS *E. coli* expression strain. Experiments performed to assess in vivo translation efficiency were initiated with equal cells before addition of isopropyl β-D-1-thiogalactopyranoside. Analysis of expression levels was performed by SDS/PAGE, and immunoblotting with anti-His antisera rabbit anti-His probe (Santa Cruz Biotechnology) and alkaline phosphatase-affiniPure goat anti-rabbit IgG (Santa Cruz Biotechnology). Gels and blot band density were measured with GelAnalyzer software.

Quantification of the Effect of URSs on in Vitro Translation. In vitro cell-free transcription/translation was performed using the RTS 100 *E. coli* HY Kit (Roche/5 PRIME). The internal control of the commercial kit is a plasmid that expresses the fluorescent protein GFP. We have used this control to assess the effect of coexpression of the WT *mntA* or URS-embedded mutant genes along with GFP. The control GFP vector was added 30 min after the addition of the tested WT or mutant plasmid to the in vitro reaction. After 4.5 h of reaction, the reaction mixtures were kept at 4 °C for 24 h for GFP maturation. Fluorescence of the mature GFP was detected using an MD SpectraMax Plus 384 spectrophotometer (Hamilton Robotics) with excitation at 395 nm and emission at 508 nm.

Transfection to HeLa Cells. Transfection of HeLa cells was done on 80–100% confluent cultures. The cells were counted and seeded a day before transfection. The transfection was performed according to the original protocol from jetPRIME (Polyplus transfection) transfection reagent in a 24-well plate (250 ng of plasmid and 0.5 μL of transfection reagent). The transfection efficiency was estimated at 24 and 48 h after transfection by flow cytometry and fluorescence microscopy (Zeiss Axiovert 200 Inverted Cell Observer Microscope).

Flow Cytometry. Flow cytometry (fluorescence-activated cell sorting) was performed on HeLa cells that were transfected with pEGFP-N1 WT or mutated plasmid (Tables S2 and S3) that encoded to the EGFP. The experiments were done on a BD LSR II flow cytometer (BD Biosciences) at the Technion Lokey Life Science and Engineering (LS&E) Infrastructure Unit, and the results were analyzed using FCS Express 4 Image Cytometry software.

Single-Molecule Translation Experiments. Mutant 30S subunits labeled with Cy3B and 50S subunits labeled with Black Hole Quencher 2 (BHQ2; a FRET quencher), initiation factor 2 (IF2), elongation factor Tu (EF-Tu), elongation factor G (EF-G), elongation factor Ts (ET-Ts), and S1 were prepared as described previously for our single-molecule experiments (1–3). All experiments were conducted in a Tris-based polymix buffer consisting of 50 mM Tris·acetate (pH 7.5), 100 mM potassium chloride, 5 mM ammonium acetate, 0.5 mM calcium acetate, 5 mM magnesium acetate, 0.5 mM EDTA, 5 mM putrescine-HCl, and 1 mM spermidine. All single-molecule experiments had 4 mM GTP and were performed at 22 °C.

Total tRNAs from *E. coli* were purchased from Sigma–Aldrich, charged with tRNA synthetase mixture purified from *E. coli* S30 extract, and purified for single-molecule experiments (2, 4–7).

The mRNA constructs used in this study contain, from the 5′ end, the following elements: (*i*) 21 bases that hybridize to a DNA oligo with a 3′ biotin to immobilize the mRNA via a biotin–Neutravidin interaction, (*ii*) a 5′ UTR and Shine–Dalgarno sequence based on the bacteriophage *T4 gene 32*, and (*iii*) a sequence of the *T4 gene 32* with the particular URS inserted at the location specified in the main text. The mRNAs were transcribed in vitro from plasmids purchased from GenScript according to previously published protocols (7, 8).

The 30S preinitiation complexes (PICs) were formed as described (3) by incubating the following at 37 °C for 5 min: 0.25 μM Cy3B-30S preincubated with stoichiometric S1, 1 μM IF2, 1 μM fMet-tRNA^{fMet}, 1 μM mRNA hybridized with stoichiometric 3′ biotin DNA oligos, and 4 mM GTP to form 30S PICs in the Tris-based polymix buffer.

Before use, we preincubate an SMRT Cell V3 from Pacific Biosciences, a ZMW chip, with 1 mg/mL Neutravidin solution in 50 mM Tris·acetate (pH 7.5) and 50 mM KCl at room temperature for 5 min. The cell is then washed with buffer 6 [50 mM Tris·acetate (pH 7.5), 100 mM potassium chloride, 5 mM ammonium acetate, 0.5 mM calcium acetate, 5 mM magnesium acetate, and 0.5 mM EDTA]. After washing, 40 μL of buffer 6 is left in the cell to keep the cell surface wet. We then dilute the

30S PICs with our Tris-based polymix buffer containing 1 μM IF2 and 4 mM GTP down to 25–75 nM PIC concentration, depending on the specific mRNA used. The diluted PICs are then loaded into the SMRT cell at room temperature for 3 min to immobilize the 30S PICs into the ZMW wells. We wash away excessive unbound material with our Tris-based polymix buffer containing 1 μM IF2, 4 mM GTP, 1 mM Trolox, and a PCA/PCD oxygen scavenging system (2.5 mM 3,4-dihydroxybenzoic acid and 250 nM protocatechuate deoxygenase) (9). After washing, 20 μL of the washing mix is left in the cell to keep the surface wet and to remove oxygen in the liquid. We then form ternary complexes between total charged *E. coli* tRNAs and EF-Tu(GTP) as described (3).

Before starting an experiment, the cell is loaded into a PacBio RS sequencer modified by Pacific Biosciences for observing single-molecule fluorescence from the ZMW wells (10). To begin the elongation experiment, the instrument illuminates the SMRT cell with a green laser and then immediately delivers 20 μL of a delivery mixture containing 200 nM BHQ2-50S, 1 μM IF2, 160 nM EF-G, 2.5 μM total tRNA ternary complexes, 4 mM GTP, 1 mM Trolox, and the PCA/PCD oxygen scavenging system in our Tris-based polymix buffer onto the cell surface. In all experiments, data were collected at 10 frames per second (100-ms exposure time) for 5 min.

Data analysis for all experiments is conducted with MATLAB (MathWorks) scripts written in-house as used in previous experiments (11) The FRET states are assigned as previously described based on a hidden Markov model-based approach and visually corrected (12). All lifetimes were fitted to a single-exponential distribution using maximum-likelihood parameter estimation in MATLAB.

1. Dorywalska M, et al. (2005) Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic Acids Res* 33(1):182–189.
2. 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.
3. Marshall RA, Dorywalska M, Puglisi JD (2008) Irreversible chemical steps control intersubunit dynamics during translation. *Proc Natl Acad Sci USA* 105(40):15364–15369.
4. Pavlov MY, Ehrenberg M (1996) Rate of translation of natural mRNAs in an optimized in vitro system. *Arch Biochem Biophys* 328(1):9–16.
5. Lesley SA (1995) Preparation and use of E. coli S-30 extracts. *Methods Mol Biol* 37:265–278.
6. Johansson M, Chen J, Tsai A, Kornberg G, Puglisi JD (2014) Sequence-dependent elongation dynamics on macrolide-bound ribosomes. *Cell Reports* 7(5):1534–1546.
7. Tsai A, Kornberg G, Johansson M, Chen J, Puglisi JD (2014) The dynamics of SecM-induced translational stalling. *Cell Reports* 7(5):1521–1533.
8. McKenna SA, et al. (2007) Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* 2(12):3270–3277.
9. Aitken CE, Marshall RA, Puglisi JD (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J* 94(5):1826–1835.
10. Chen J, et al. (2014) High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proc Natl Acad Sci USA* 111(2):664–669.
11. Chen J, Petrov A, Tsai A, O'Leary SE, Puglisi JD (2013) Coordinated conformational and compositional dynamics drive ribosome translocation. *Nat Struct Mol Biol* 20(6):718–727.
12. Aitken CE, Puglisi JD (2010) Following the intersubunit conformation of the ribosome during translation in real time. *Nat Struct Mol Biol* 17(7):793–800.

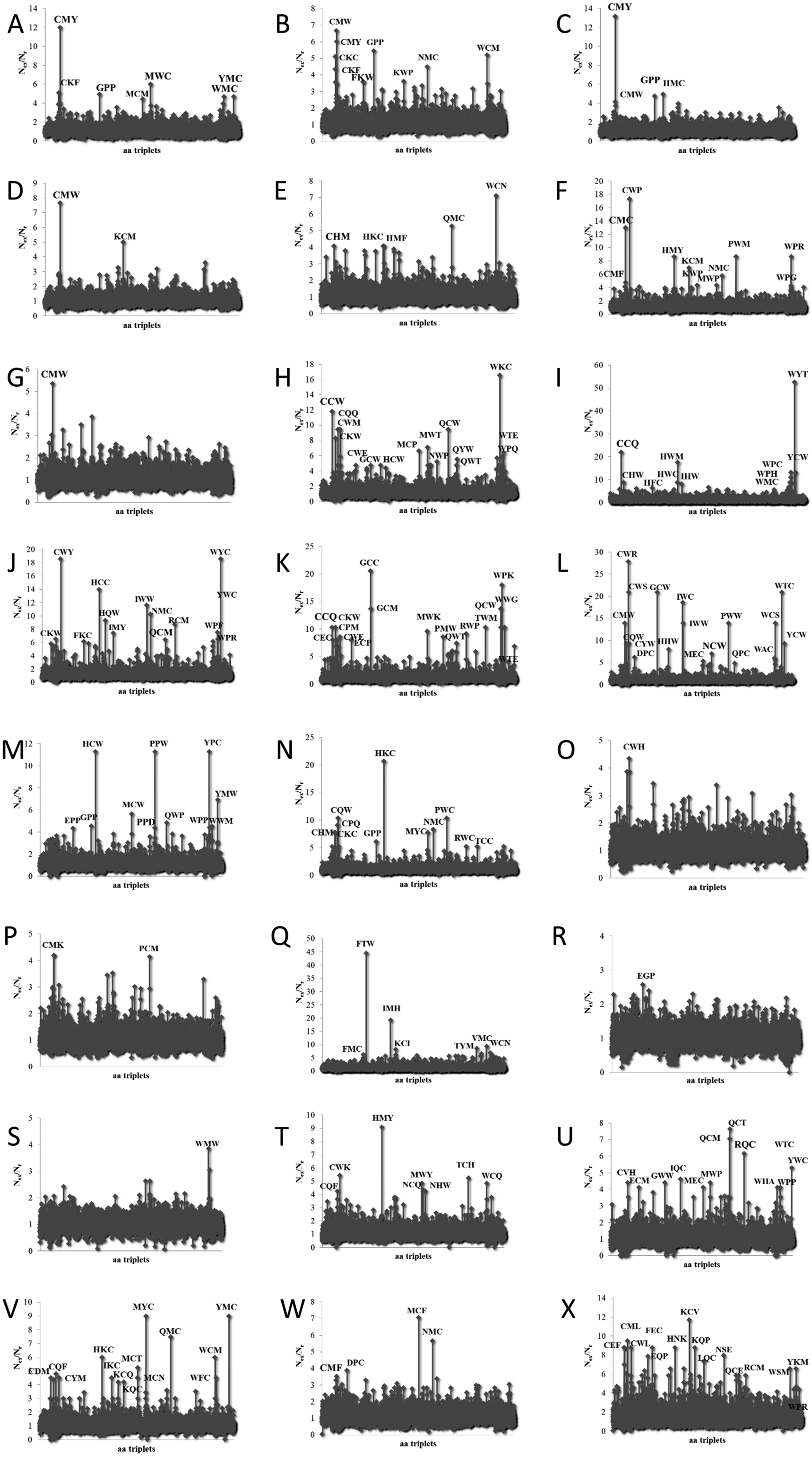

Fig. S1. Identification of URSs in various microorganisms. The ratio of expected appearance to actual appearance (N_{ex}/N_r) of all triplets in the proteomes (number of proteomes in database in parentheses) of 25 microorganism genera identifies their URSs: *Shigella* (seven proteomes) (*A*); *Salmonella* (17 proteomes) (*B*); *Klebsiella* (three proteomes) (*C*); *Pseudomonas* (17 proteomes) (*D*); *Legionella* (four proteomes) (*E*); *Heamophilus* (eight proteomes) (*F*); *Vibrio* (12 proteomes) (*G*); *Campylobacter* (10 proteomes) (*H*); *Helicobacter* (nine proteomes) (*I*); *Neisseria* (eight proteomes) (*J*); *Rickettsia* (13 proteomes) (*K*); *Staphylococcus* (19 proteomes) (*L*); *Streptococcus* (44 proteomes) (*M*); *Listeria* (four proteomes) (*N*); *Bacillus* (24 proteomes) (*O*); *Mycobacterium* (22 proteomes) (*P*); *Chlamydia* (six proteomes) (*Q*); *Clostridium* (25 proteomes) (*R*); *Candida* (three proteomes) (*S*); *Treponema* (two proteomes) (*T*); *Mycoplasma* (five proteomes) (*U*); *Deinococcus radiodurans* (one proteome) (*V*); *Haloarcula marismortui* (one proteome) (*W*); and *Thermus thermophilus* (two proteomes) (*X*). In the following genera, there are also triplets that do not appear at all in the proteomes (and thus do not appear in the above graphs): *Campylobacter*: CHW, CWC, HWW, and WCH; *Helicobacter*: CCW, CWW, HMW, MWM, and WPR; *Rickettsia*: CWC, CWT, HMC, HMW, HWW, MCW, MWM, PHW, QWP, WEW, WHW, WWC, WWP, and WWW; *Neisseria*: CHM, CMC, and CMH; *Listeria*: CCW, CHW, CQC, CWC, CWP, CWW, PCC, QPC, WCH, WCW, WCY, WPC, and WRC; *Staphylococcus*: CWH, CWW, WCC, and WCW; *Chlamydia*: CKC, CKW, CMC, CMW, CMY, CNW, CQW, CWH, CWP, CYM, DCM, HCM, HKW, HMW, HMY, HNW, MCT, MCW, MMW, MWA, MWF, MWQ, NCM, NMW, NWC, NWW, PWW, QCM, QHW, QWC, WFM, WHC, WHM, WHW, WHY, WKW, WMC, WMW, WNW, WPM, WPW, WPY, WVM, WWH, WWP, WYW, YMC, YMW, and YWC; *Mycoplasma*: CAW, CMR, CMW, CWC, HCW, HWC, WCC, WCH, and WHC; and *Treponema*: CMW, QWW, and WWM.

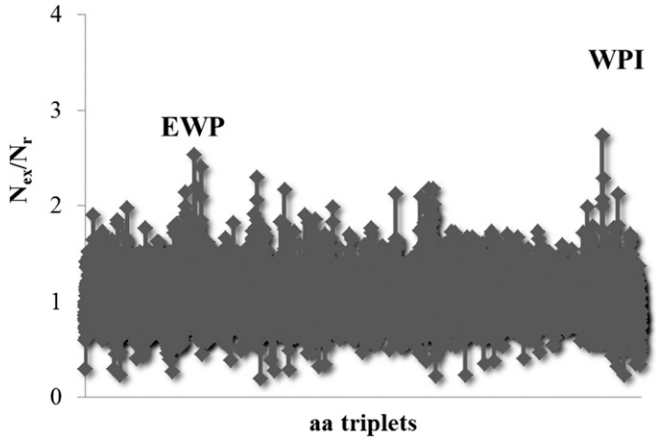

Fig. S2. Identification of URSs in the human proteome. The ratio of expected appearance to actual appearance (N_{ex}/N_r) of all 2.5×10^7 triplets in the human proteome is shown. EWP and WPI are the only two triplet amino acid sequences that are significantly above the typical triplet usage.

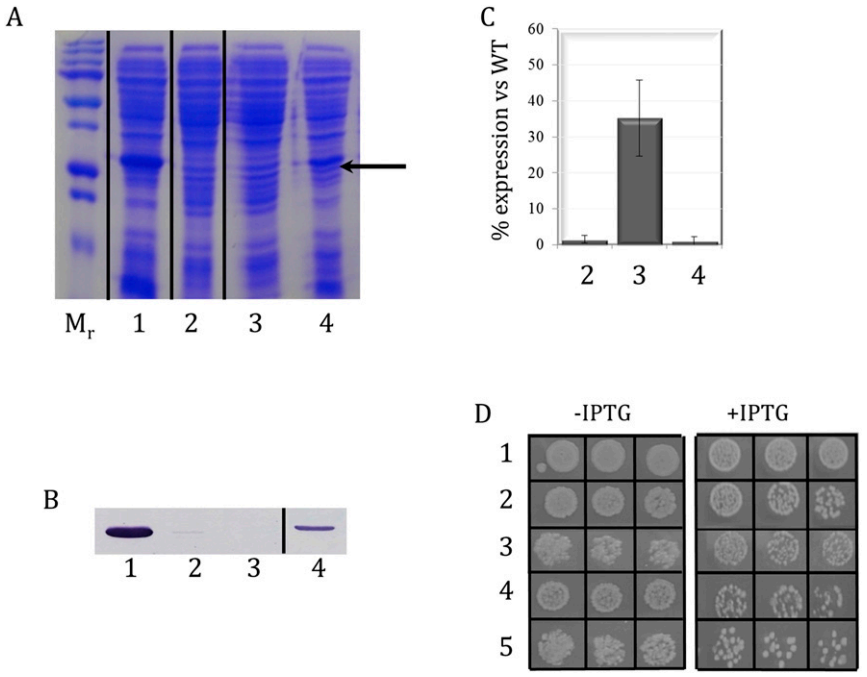

Fig. S3. Embedded URSs inhibit translation of GFP in vivo. (*A*) SDS/PAGE analysis of whole-cell [BL-21(DE3)pLysS *E. coli*] content following expression of WT-GFP (lane A) or mutant GFP (samples 2–4) for 4 h. Embedded URSs were CMY, CMYW, or the non-URS MYC. Data are representative of two to four independent experiments. The arrow indicates the position of the mature GFP protein. (*B*) Immunoblot of SDS/PAGE in *A*, using anti-His_6 antibodies. (*C*) Fluorescent quantification of GFP mutant protein expression analyzed in *A* and *B* compared with the fluorescence from WT-GFP. Results shown in *C* are the mean ± SD from two to four independent experiments. (*D*) Expression of proteins with embedded URSs compromises cell growth. Equal amounts of BL21(DE3)pLysS *E. coli* cells containing WT-MntA or MntA with embedded URSs (Table S1) were diluted with fresh LB for a series of 10-fold dilutions. Four microliters of each dilution was plated on fresh LBA (*Left*) or LBA with 2 mM IPTG (*Right*). Far fewer colonies are seen in the presence of IPTG when URS mutations are expressed. Vertical black lines indicate noncontiguous samples from a larger gel (*A*), immune blot (*B*) or agarose plates (*D*). IPTG, isopropyl β-D-1-thiogalactopyranoside; LBA, Luria–Bertani medium with ampicillin.

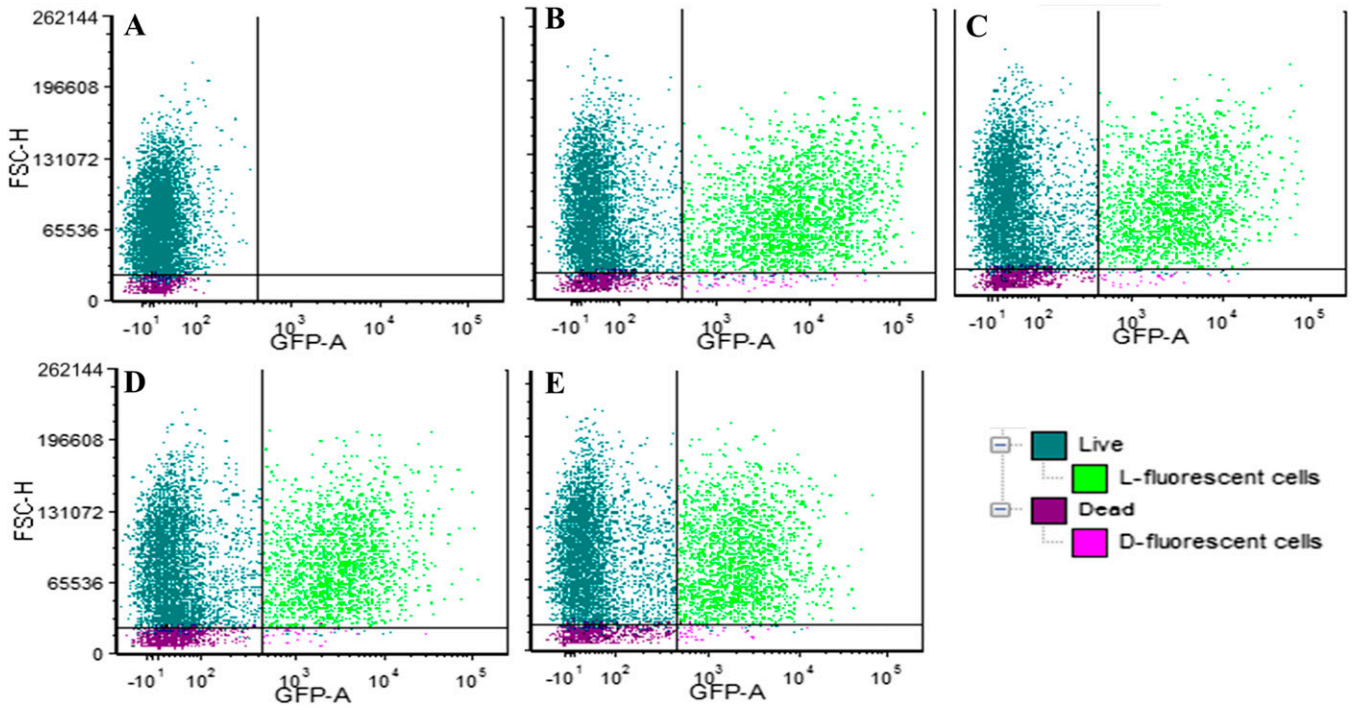

Fig. S4. Proteins with embedded *E. coli* URSs are expressed at normal levels in HeLa cells. FACS analysis 48 h after transfection of HeLa cells with plasmids expressing WT-eGFP or URS-embedded eGFP. Control, no plasmid (*A*); WT (*B*); and MYC (non-URS mutation), CMY, and CMYW embedded at position 1 (*C*–*E*), respectively. In each panel, the results are divided into four quadrants showing the number of cells: nonfluorescent live cells (*Upper Left*), fluorescent live cells (*Upper Right*), nonfluorescent dead cells (*Lower Left*), and fluorescent dead cells (*Lower Right*). FSC-H, forward scatterer-height.

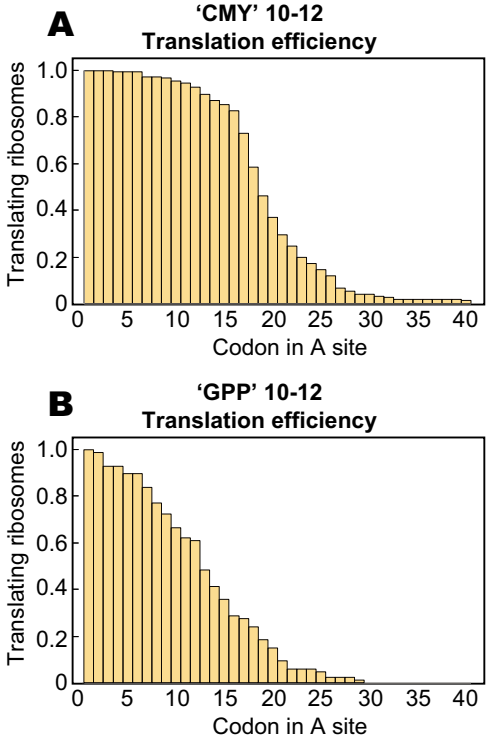

Fig. S5. (*A*) Ribosomes translate the $C_{10}M_{11}Y_{12}$ peptide sequence at a normal rate and efficiency until codon 15. Translation slows starting at codon 15, leading to significant rates of translational arrest beginning at codon 16. By codon 30, only 4% of ribosomes are still translating. (*B*) Ribosomes translate a similar sequence with $G_{10}P_{11}P_{12}$ at a normal rate until immediately after the incorporation of P_{12}. Translation rates decrease drastically and immediately, and by codon 30, 100% of ribosomes have ceased translation.

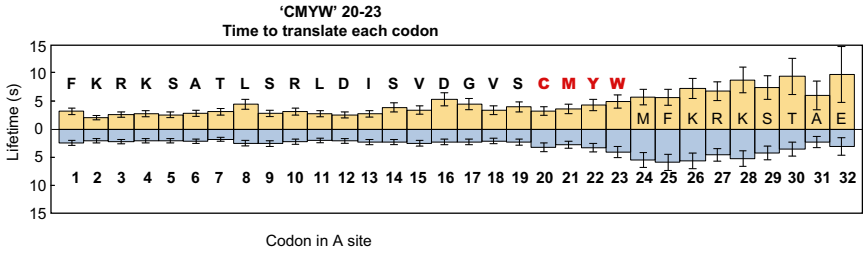

Fig. S6. Decrease in translation rates for CMYW (20–23) is shifted by 10 codons relative to the CMYW (10–13) sequence. Conformational state lifetimes during translation of the $C_{20}M_{21}Y_{22}W_{23}$ mRNA sequence are shown. The increases in both rotated and nonrotated state lifetimes are shifted by 10 codons in the 3′ direction, relative to the $C_{10}M_{11}Y_{12}W_{13}$ sequence.

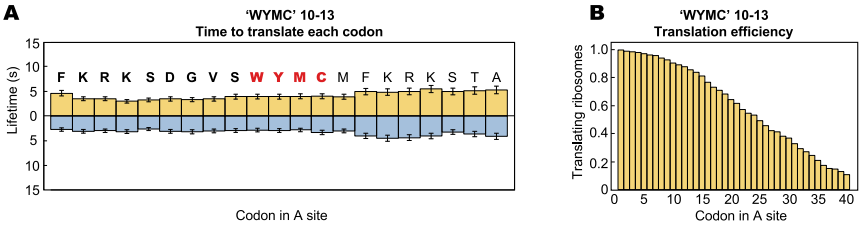

Fig. S7. Translation of the WYMC sequence proceeds normally. (*A*) Translating ribosomes are unperturbed by the presence of the scrambled WYMC sequence between positions 10 and 13. Translation rates are constant over the course of the entire mRNA sequence, similar to translation of canonical mRNAs, such as gp32, and absent the large increase in conformational state lifetimes that results from the translation of the CMYW sequence. (*B*) Efficiency of translation of the WYMC mRNA is similar to the efficiency of translation of gp32; 36% of ribosomes translating the WYMC mRNA are still translating at codon 30, compared with 22% of ribosomes that were still translating at codon 30 of the gp32 sequence.

A. MntA:

1 2

MAATLSR**LDISV**DGV**SVTYN**NARLALYNATCTVEPGTITALVGPNGSGKSTLFKSIM
GFLQPSQGRVRIGGFSVQKAQKQQLMAYV

3

PQADEVDWNFPVSVFDVVMMG**RYGYM**NVLRIPSAKDRRLVMESLERVGMVKYRDRQI
GELSGGQKKRAFLARALAQEGKVILLDEPFTGVDVKTEKGMIDLLMELRDEGHTILI
STHDLASISTFCDHTILLNRTILAQGKTEETFTKENLELTFGGLPMLSLNQMFESTE
VDAHHHHHH

B. GFP

1

MTKGEELFTGVV**PILV**ELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPW
PTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF
EGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNI
EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAG
ITHGVDELYQPGGGSHHHHHH

C. eGFP

1

MVSKGEELFTGV**VPILV**ELDGDVKDPPVATMVSKGEELFTGVVPILVELDGDVK
DPPVATMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYK
TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVN
FKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLL
EFVTAAGITLGMDELYKGFP

Fig. S8. MntA, GFP, and eGFP sequences with positions of URS or non-URS mutations according to Table S1. Positions in the original sequence that were mutagenized are underlined and in bold.

Oligos used for *mntA* in pET 20b^{+}.

PCR oligos, Endonuclease site is underlined		
Name	Sequence:	Endonuclease:
mntA-F	5'-CATATGGCAGCCACCCTATCTCGTCTTGATAT-3'	NdeI
mntA- R	5'-CTCGAGCGCATCTACCTCTGTCGATTCAAACA-3'	XhoI
Mutagenesis primers: the mutation site is underlined		
Name	Sequence:	
CMY-1F	5'-GCAGCCACCCTATCTCGTCTTGATATA**TGCATGTAT**GGTGTCAGCGTTA-3'	
CMY-1R	5'-TAACGCTGACACCATACATGCATATATCAAGACGAGATAGGGTGGCTGC-3'	
CMYW-1F	5'-AGGAGATATACATATGGCAGCCACCCTATCTCGTCTTGAT**TGGTGCATGTAT**GGTGTAGCGTTACCTATA-3'	
CMYW-1R	5'-TATAGGTAACGCTGACACCATACATGCACCAATCAAGACGAGATAGGGTGGCTGCCATATGTATATCTCCT-3'	
WCMY-1F	5'-TATACATATGGCAGCCACCCTATCTCGTCTTGATATA**TGCATGTATTG**GGTCAGCGTTACCTATAACAA-3'	
WCMY-1R	5'-TTGTTATAGGTAACGCTGACCCAATACATGCATATATCAAGACGAGATAGGGGGCTGCCATATGTATA-3'	
CMY-2F	5'- GATATATCCGTCGATGGTGTCAGC**TGCATG**TATAACAACGCCCGTTTGGCCCTT-3'	
CMY-2R	5'-AAGGGCCAAACGGGCGTTGTTATACATGCAGCTGACACCATCGACGGATATATC-3'	

Fig. S9. (Continued)

CMYW-2F	5'-ATATCCGTCGATGGTGTCTGGTGCATGTATAACAACGCC-3'
CMYW-2R	5'-GGCGTTGTTATACATGCACCAGACACCATCGACGGATAT-3'
WCMY-2F	5'-GTCGATGGTGTCAGCTGCATGTATTGGAACGCCCGTTTGG-3'
WCMY-2R	5'-CCAAACGGGCGTTCCAATACATGCAGCTGACACCATCGAC-3'
CMY-3F	5'-GTTTGATGTGGTGATGATGGGGCGCTGCATGTATATGAATGTGTTGCGTATTCCTAG-3'
CMY-3R	5'-CTAGGAATACGCAACACATTCATATACATGCAGCGCCCCATCATCACCACATCAAAC-3'
CMYW-3F	5'-TTTGATGTGGTGATGATGGGGTGGTGTATGTATATGAATGTGTTG-3'
CMYW-3R	5'-CAACACATTCATATACATACACCACCCCATCATCACCACATCAAA-3'
WCMY-3F	5'-GATGGGGCGCTGTATGTATTGGAATGTGTTGCGTATT-3'
WCMY-3R	5'-AATACGCAACACATTCCAATACATACAGCGCCCCATC-3'

Fig. S9. (Continued)

Oligos used for GFP in pIVEX2.3-GFP.

PCR oligos, Endonuclease site is underlined	
Name	Sequence:
GFP-F	5'-GATATACCATGACTAAAGGTGAAGAACTT-3'
GFP-R	5'-CCCCGGGTTGGTACAGTTCATCCACGCCATGTGT-3'
Mutagenesis primers: the mutation site is underlined	
Name	Sequence:
CMY-1F	5'GAAGAACTTTTCACTGGAGTTGTCTGCATGTATGTTGAATTAGATGGTGATGTTAATGGGCACAAA-3'
CMY-1R	5'- TTTGTGCCCATTAACATCACCATCTAATTCAACACAATACATGACAACTCCAGTGAAAAGTTCTTC-3'
CMYW-1F	5'-TGAAGAACTTTTCACTGGAGTTGTCATGTATTGTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTC-3'
CMYW-1R	5'- GAAAATTTGTGCCCATTAACATCACCATCTAATTCAACACAATACATGACAACTCCAGTGAAAAGTTCTTCA-3'
MYC-1F	5'-GGAGTTGTCTGCATGTATTGGGAATTAGATGGTGATG-3'
MYC-1R	5'-CATCACCATCTAATTCCCAATACATGCAGACAACTCC-3'

Fig. S9. (Continued)

Oligos used for eGFP in pEEGFP-N1.

PCR oligos, Endonuclease site is underlined		
Name	Sequence:	Endonuclease:
eGFP -F	5'-GCAGTCGACGGTACAGCTAGCAACCATGGTGAGCAA GG-3'	NheI
eGFP-R	5'-GCCGCTTTATGGATCCAGCTCGTCCATG-3'	BamHI
eGFP-F	5'-GCAGTCGACGGTACAGCTAGCAACCATGGTGAGCAA GG-3'	NheI
eGFP-R	5'-CACGCTGAACTTGGGATCCTTTACGTCGCCGTCCAG-3'	BamHI
Mutagenesis primers: the mutation site is underlined		
Name	Sequence:	
CMY -1F	5'- CGGGGTGGTGTGCATGTATGTCGAGCTGGACGGCG -3'	
CMY -1R	5'- CGCCGTCCAGCTCGACATACATGCACACCACCCCG -3'	
CMYW-2F	5'- CGGGGTGGTGATGTATTGTGTCGAGCTGGACGGC -3'	
CMYW-2R	5'- GCCGTCCAGCTCGACACAATACATCACCACCCCG -3'	
T3-3F	5'- GGTGTGCATGTATTGGGAGCTCGACGGCGACGTAAACG -3'	
T3-3R	5'- CGTTTACGTCGCCGTCGAGCTCCCAATACATGCACACC -3'	

Fig. S9. Primers used in this study for URS and non-URS mutation embedding.

Table S1. *E. coli* strains with vectors expressing proteins with embedded URSs used in this study

Protein	Site*	URS	Expression vectors
MntA	—	—†	pET 20b+, pET 45b+
MntA	1	CMY	pET 20b+, pET 45b+
MntA	2	CMY	pET 20b+, pET 45b+
MntA	3	CMY	pET 20b+, pET 45b+
MntA	1	WCMY	pET 20b+, pET 45b+
MntA	1	CMYW	pET 20b+, pET 45b+
MntA	2	WCMY	pET 20b+, pET 45b+
MntA	2	CMYW	pET 20b+, pET 45b+
MntA	3	WCMY	pET 20b+, pET 45b+
MntA	3	CMYW	pET 20b+, pET 45b+
GFP	—	—†	pIVEX2.3-GFP, pET 45b+
GFP	1	CMY	pIVEX2.3-GFP, pET 45b+
GFP	1	MYC‡	pIVEX2.3-GFP, pET 45b+
GFP	1	CMYW	pIVEX2.3-GFP, pET 45b+
EGFP	—	—†	EGFP-N1
EGFP	1	CMY	EGFP-N1
EGFP	1	MYC‡	EGFP-N1
EGFP	1	CMYW	EGFP-N1

*Site refers to different positions within the entire encoding region of the gene.
†WT.
‡Non-URS mutation.

Table S2. FACS analysis of expression of EGFP or EGFP-URS mutants in HeLa cells

Cell type	No plasmid†	WT	MYC	CMY	CMYW
% Live cells*	92.8 ± 0.2	88.7 ± 1.3	88.5 ± 0.7	85.6 ± 1.8	88.8 ± 1.0
% Fluorescent live cells	0.02 ± 0.01	29.1 ± 0.9	23.5 ± 0.7	24.4 ± 1.1	23.8 ± 0.7
% Dead cells	7.4 ± 0.2	11.3 ± 1.3	11.5 ± 0.7	14.4 ± 1.8	11.2 ± 1.0
% Fluorescent dead cells	0.0 ± 0.0	6.2 ± 0.9	3.3 ± 0.8	3.0 ± 0.8	5.0 ± 0.7

*Classification performed as described in *Methods*.
†Quantification based on at least three independent experiments.

Table S3. FACS analysis of all fluorescent HeLa cells expressing EGFP or EGFP-mutants

Cell type	WT†	MYC	CMY	CMYW
% Live cells*	97.4 ± 0.3	98.0 ± 0.7	98.2 ± 0.3	97.4 ± 0.3
% Dead cells	2.6 ± 0.3	2.0 ± 0.7	1.8 ± 0.3	3.0 ± 0.3

*Classification performed as described in *Methods*.
†Quantification based on at least three independent experiments.