

Supporting Information

QUANTIFYING ELONGATION RHYTHM DURING FULL-LENGTH PROTEIN SYNTHESIS

Gabriel Rosenblum^{1,2}, Chunlai Chen², Jaskiran Kaur³, Xiaonan Cui², Haibo Zhang¹, Haruichi Asahara⁴, Shaorong Chong⁴, Zeev Smilansky³, Yale E. Goldman^{2*}, Barry S. Cooperman^{1*}.

¹Department of Chemistry, University of Pennsylvania, Philadelphia, PA, 19104-6323

²Pennsylvania Muscle Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6083

³Anima Cell Metrology, Inc., 75 Claremont Road, Suite 102, Bernardsville, NJ 07924-2270

⁴New England Biolabs Inc., 240 County Road, Ipswich, MA 01938

*Correspondence: goldmany@mail.med.upenn.edu, coopman@pobox.upenn.edu

SUPPORTING METHODS

Selection of FRET Events

Objective selection of FRET events. We found that the standard Hidden Markov Method software, HaMMMy,¹ although successful in finding relatively long FRET events (≥ 4 frames), often skipped shorter events (≤ 3 frames). This is most likely because the probability at any given time of a Phe-tRNA^{Phe}(Cy5.5) accommodation event is highly dependent on the time elapsed since the previous Phe-tRNA^{Phe}(Cy5.5) accommodation, invalidating the history-independence of Markov chains. Peptide segments containing 11 – 24 elongation cycles are thus not well described by the Markov model. Accordingly, we developed an ad-hoc algorithm which overcomes this difficulty by identifying FRET events within a trace when two criteria were satisfied simultaneously: 1) the intensity of the FRET signal was significantly higher than background; and 2) the donor and acceptor intensities were anti-correlated. In order to systematically capture FRET events (Fig. S2a), we developed an objective, MATLAB-based algorithm to automatically select for FRET events that satisfy the two criteria. Implementation of this algorithm required determination of background and threshold detection values, as described below.

Background level. For criterion #1, the background level was determined by first eliminating the major FRET intensity peaks (appearing in ≥ 3 consecutive frames) using Hidden Markov Method software, HaMMMy.¹ HaMMMy was set to search for two FRET states, designated zero (i.e., background) and non-zero FRET states. The MATLAB code used the HaMMMy-analyzed data to remove the non-zero FRET signals, in order to obtain

the background noise level, which is given by the HaMMY-filtered zero-FRET data. Visual inspection confirmed that the HaMMY-filtered zero-FRET data contained a negligible fraction (<5 events/trace in traces containing 300-1000 frames) of short-lived (<3 frames) FRET events. For criterion #2, the background level was determined by eliminating the major donor and acceptor peaks, which corresponded to major FRET peaks as identified by HaMMY.

Detection threshold value. Application of each criterion required evaluation of a threshold value to include most *bona fide* events and reject noise. The HaMMY-filtered data (zero FRET, criterion #1; zero donor and acceptor, criterion #2) were used to calculate a threshold for each trace, using an automated MATLAB routine. For both criteria, the number of data points (video frames) above background were determined as a function of threshold value. The threshold corresponded to that value giving an abrupt change in the number of data points as the threshold became more restrictive (Fig. S2d,e). This break point was found by searching for the largest change in slope, obtained by t-test (lowest P-value) when comparing two sets of data points along the selection curves (examples in Figs. S2d,e): set 1: $i-n$ to i and set 2: i to $i+n$ (where i is the central point and n is the window size). Varying the value of n in the range of 2-5 data points per window resulted in a stable region giving similar break point values that selected both short and long FRET events. Higher values of n produced more permissive thresholds, including many noise peaks.

For criterion #1 the value of the threshold was such that only 2-4% of the FRET data (the 98th to 96th percentiles, Fig. S2b) extended above the threshold in the great majority of the HaMMY-filtered traces. All FRET values exceeding the FRET intensity threshold were considered to have met criterion #1. For criterion #2, anti-correlated changes were identified by subjecting the donor and acceptor traces, separately, to a running t-test procedure, in which the average intensity in a 3-frame window is compared with the background. An anti-correlation event corresponded to the product of the donor and acceptor P-values falling below the threshold value (Fig. S2c). For the great majority of observed traces, anti-correlated FRET events could be distinguished from random anti-correlated fluctuations of the shot noise (Fig. S2c) when the threshold was set such that 7 – 9% of the product of multiplying the two P-values fell below the threshold value. A true FRET event satisfied both criterion #1 (FRET intensity > threshold intensity) and criterion #2 (P-value product < threshold value). The MATLAB algorithm is available upon request.

Visual selection of FRET events. Alternatively, events were selected visually using the same two criteria, but with background levels determined by eye. The mean values of segment translation times obtained with the automated algorithm are not statistically significant from mean values obtained by visual inspection (t-test, P-value > 0.4) (albeit with slight differences in s.e.m. values (Table 2), showing that the two approaches are in broad agreement.

Objective vs. visual selection of FRET events. The MATLAB algorithm and the visual inspection method both select traces that exhibit statistically significant differences between WT-s3 and M1-s3 and between M1-s3 and M2-s3. The differences between WT-s3 and M1+tRNA^{Arg}-s3 and between WT-s3 and M2-s3 are not statistically significant for either selection method.

A somewhat (~ 30%) larger number of events was selected visually, due to traces that are problematic for the present algorithm, either because the algorithm employs a threshold for one of the two criteria that is more restrictive than that used visually, or because of traces in which the donor signal exhibits baseline intensity changes that are not a result of energy transfer. On the other hand, the visual inspection method is subject to false negatives. Some FRET events that are shorter than the integration time (200 ms) produce weak signals that could be regarded as noise during visual inspection, but are picked up by the MATLAB code, as they satisfy both criteria

Distinguishing Single from Double Phe-tRNA^{Phe}(Cy5.5) Accommodations.

The distribution of FRET events as a function of duration was clearly bimodal (Figs. S3a-b). The entire data set (Fig. S3a) was best fitted to the sum of two gamma distributions. A limited set of clear traces, each containing two adjacent longer FRET events (Phe-Phe doublets) and one or more short FRET events (single Phe-tRNA^{Phe} accommodations) were analyzed visually. The durations of the events of the shorter and longer components had average values of 0.6 ± 0.1 s and 3.6 ± 0.5 s, respectively. In order to determine a usable cutoff value to distinguish the long and short events and to determine how many

short events spilled over into the longer duration class and *vice versa*, the two gamma distributions obtained in Fig. S3b were each normalized to unity (Fig. S3c). The intersection between the two normalized gamma distributions at 1.2 s was used as the cutoff: events <1.2 s were assigned to single Phe-tRNA^{Phe}(Cy5.5)s on the 70S^{Cy3} and those ≥1.2 s were assigned to two consecutive accommodations. The longer events include the interval between the accommodations, corresponding to the sum of the times required for translocation to form a POST complex, binding Phe-tRNA^{Phe}(Cy5.5) to this POST complex, and the next translocation. These longer FRET events sometimes display brief excursions to low FRET efficiencies, which can be attributed to the waiting time between successful accommodations of the successive Phe-tRNA^{Phe}(Cy5.5)s. The average value of 3.6 s is longer than the total expected time, 2 s [corresponding to the sum of the average duration of a single Phe-tRNA^{Phe} accommodation (0.6 s) and the mean translation time per codon (1.4 s, Table 1)]. It is unclear at present to what extent the relatively long duration of the doublet events reflect: i) slowly translating UUC pairs² and/or ii) consecutive accommodations of two bulky Cy5.5-containing tRNAs.

The translation times presented in Fig. 5 and Table 1 depend on the validity of the intervals assigned to each trace, which in turn depend on correct assignments of shorter and longer events. For example, within events containing two long FRET events flanked by at least one shorter event (termed Pattern C, see below), if one of the longer FRET events ascribed to a Phe-Phe doublet were actually a relatively long single accommodation event, then a segment peptide assigned to s3 (14 aa) would actually

correspond to s2 (11 aa) or s4 (13 aa). As detailed below, such mis-assignments occur with low probability and have little effect on translation time values for segment s3.

From Fig. S3b, 4.1% of the single accommodation events have values above the 1.2 s cutoff mentioned above, so that ~4.1% of the events assigned as Phe-Phe doublets may actually correspond to single Phe accommodations. However, given the similarity of translation times for s2, s3, and s4 (19.8 ± 0.9 s, 23.7 ± 1.1 s, and 23.6 ± 1.7 s, respectively) the change in s3 translation time that would result from such mis-characterizations is only approximately 0.2 s. Alternatively, a relatively rapidly translated Phe-Phe doublet could be mistakenly assigned as a single Phe accommodation. Due to the broad distribution of the durations of doublet events, 11.2% of the normalized gamma distribution fitted to the long component falls below the 1.2 s cutoff value. Such mis-assignment would lead an s3 interval to be scored as s2 or s4, corresponding to an error of ~0.5 s. The sum of both potential errors is well within the uncertainty of s3 duration.

FRET Sequence Pattern Fitting.

The time intervals between two consecutive Phe-tRNA^{Phe} accommodations were assigned to particular sequence segments (Table 1) according to their temporal positions relative to the two characteristic Phe-Phe doublets near the center of the EmGFP sequence. Traces were selected that contained at least two FRET events, at least one of which was a Phe-Phe doublet, characterized by a long duration. Such traces displayed three characteristic patterns (Fig. S4). Pattern A contains the first doublet and the preceding single-Phe FRET events, and is terminated by Cy3 photobleaching just after the doublet. It provides

estimates of the times required for translation of segments s2 (11 aas) and, in many of the traces, of s1 (24 aas) as well. Pattern B contains the second doublet, appearing early in the trace, followed by one or more short FRET events, providing estimates of the times required for translation of segments s4 (13 aas) and, in many of the traces, of s5 (15 aas) as well. Pattern C contains the two consecutive doublets, with short events preceding or following the doublets, up to two on each side, providing estimates of the translation times of s3, with 14 aas, falling between the two doublets. Many of the traces also allowed estimates of the times required for translation of s1, s2, s4, and/or s5. The mean translation times of s1, s2, s4, and s5 derived from Patterns A, B and C are not significantly different (Fig. S4). Results from all three patterns were combined in constructing Figs. 4 and 5.

Calculating average codon usage and tRNA abundance for correlation with translation rates (Figure 4).

In Figure 4, the average translation rate in a segment is equal to the reciprocal of the average codon translation time in that segment, determined as described in Table 1. Such times are expected to be proportional to the inverse of either relative tRNA concentration³ or relative codon usage. Relative average tRNA concentrations and codon usages per segment were calculated from eq. 1A and eq. 1B, using isoacceptor $tRNA_i$ and $codon_i$ values (as percentages) obtained from⁴. In both cases, an average of the reciprocal values was first obtained, followed by taking the reciprocal of that value.; n is the number of codons in a segment.

$$\langle tRNA \rangle = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{[tRNA_i]}} \quad (eq.1A); \quad \langle codon \rangle = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{[codon_i]}} \quad (eq.1B)$$

SUPPORTING DATA

tRNA accommodation vs. tRNA sampling.

Although transient FRET events are mainly obtained upon accommodation of Phe-tRNA^{Phe}(Cy5.5) when the 70S^{Cy3} ribosome reaches Phe codons, it is also possible for the ribosome to bind Phe-tRNA^{Phe}(Cy5.5) briefly at near-cognate or non-cognate codons, in so-called sampling events.⁵ In exploring this possibility we focused on s3. The mRNA encoding this segment of 14 amino acids contains 4 codons with single-base mismatches: 2 at the first base and 2 in the middle base. Less than 5% of traces display any short FRET events between the doublets, indicating that such events do not contribute significantly to the average translation times shown in Figs. 4 and 5. The paucity of detecting sampling events is likely due, in part, to their dwell times being much smaller than the time resolution, 200 ms/frame. In a previous study, conducted at much higher time resolution (10 ms/frame), the sampling dwell time was estimated to be ~ 70 ms⁵. Such brief encounters would generate FRET signals with lower apparent efficiency compared to longer lived binding events, merging them with background noise. In addition, in the present study the full complement of tRNAs was present, lowering the frequency of sampling, compared to the conditions of Geggier et al.⁵, which employed a single tRNA.

Classic and hybrid states in translating ribosomes.

Fitting the distribution presented in Fig. 6a to a single Gaussian centered at a FRET value of 0.65 (s.d. = 0.17) resulted in a reduced $\chi^2 = 2.6$, which is a measure of the goodness of fit. Attempts to fit this data to a mixture of two Gaussians resulted in FRET components with peaks < 0.1 apart, and without significant improvement to the fit. There was no marked improvement in the quality of the fit (reduced $\chi^2 = 2.8$) when it was constrained to two Gaussian FRET components centered at 0.44 and 0.66, based on the curves fitted to the data on stalled ribosomes in Fig. 6b. Relative areas under the Gaussians curves in this constrained fit were 0.05 ± 0.02 and 0.95 ± 0.02 , respectively, indicating that the 0.66 peak greatly predominates. Our results thus demonstrate that the PRE complexes in actively translating ribosomes spend much less time in the hybrid state ($\leq 5\%$), whereas stalled ribosomes (due to the absence of EF-G), populate the hybrid state for a greater proportion of time ($\sim 50\%$, Fig. 6b).

Correlating translation rates with codon usage and tRNA abundance.

The translation rates of WT EmGFP segments show positive correlations with average codon usage and average cognate tRNA abundance (Fig. 4), leading to the conclusion that the specific synonymous codons present in an mRNA exert significant control on translation rate. The introduction of the synonymous mutations in M1 leads to a significant decrease in the translational rate of s3, primarily due to the Arg CGG codon which is recognized by a rare tRNA^{Arg} isoacceptor. The synonymous mutation in M2 leads to a moderate decrease in translation rate (statistically insignificant from the WT), primarily due to the Arg CGA codon. Although this codon is rare, it is cognate to a

plentiful tRNA isoacceptor that contains Inosine (post-transcriptionally modified Adenosine) as the first anticodon base, leading to a distorted I-A interaction that may kinetically perturb the decoding process.^{2, 6} Comparison of the correlation plots of translation rate vs. either average codon usage (Fig. 4a) or tRNA abundance (Fig. 4b) for s1 – s5 for the WT and s2 – s4 for the M1 and M2 variants shows that the linear regression lines are in fair agreement with the correlations of most translated segments. M2-s3 shows a higher than expected translation rate in Fig. 4a. This may be explained by the lack of correlation between the rare CGA codon and the plentiful cognate tRNA isoacceptor concentration. The s3 translation rate in M1 vs. WT is approximately as expected for the decrease in both average codon usage and tRNA abundance. This result supports the notion that the decreased s3 rate in M1 largely or exclusively reflects a dependence on tRNA abundance.

Immobilization times of translating 70S^{Cy3} on the slide surface.

The CFPS reaction is initiated by addition of EmGFP plasmid. Due to Cy3 photobleaching time, the time elapsed from plasmid addition to data acquisition was optimized in order to maximize the number of traces that contained the characteristic two-doublet Phe signature. We found a 3 min incubation to be appropriate. It is likely that this time is required for: (i) transcribing the nascent mRNA, (ii) translation initiation and (iii) partial elongation of the nascent peptide. Upon data acquisition, 70S^{Cy3} harboring an HA-tagged nascent peptide outside of the ribosome tunnel diffuses until it encounters surface-bound HA-antibody. This overall process is composed of a sequence of stochastic steps, giving rise to a distribution of times for 70S^{Cy3} immobilization.

Although most 70S^{Cy3} immobilizes with a time constant of 15.4 ± 0.8 s, a smaller fraction immobilizes at longer times (time constant 170 ± 100 s), as in the case of the FRET trace presented in Figure 3. Such longer times may be due to ribosomes that initiate translation on an mRNA a) some time after it was transcribed, or b) transcribed later in the CFPS process.

mRNA sequence, internal Shine-Dalgarno sequences and secondary structure.

The expressed EmGFP mRNA sequence is shown below with the N-terminal extension marked in blue, the HA-Tag sequence bolded, the native EmGFP sequence in black, the C-terminal extension marked in red, segments s1 – s5 observed by the smTIRF bolded, and the Phe codons green. Putative internal Shine-Dalgarno-like (SD-like) sequences are underlined. Such sequences do not occur within the quantified sequence segments (s1-s5). Moreover, none of the SD-like sequences perfectly complement the anti-SD sequence in 16S rRNA: GGAGGU, which has been reported to induce the most significant pauses.⁷ Thus the variation in rates among segments s1 – s5 and mutants within s3 are not due to internal SD-induced pauses.

AUG**ACCAGCUACCCAUACGAUGUCCAGAUUACGCU**CGGGGUUCUCAUCAUCAUCAUCA
 UCAUGGUAUGGCUAGCAUGACUGGUGGACAGCAAUGGGUCGGGAUCUGUACGACGAUG
 ACGAUAAGGAUCGAUGGGGAUCCGAAUUCGCCACCAUGGUGAGCAAGGGCGAGGAGCUG
 UUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUU
 CAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAG**UUCA**
UCUGCACCACCGGCAAGCUGCCCUGCCCUGGCCACCCUCGUGACCACCUUGACCUAC
GGCGUGCAGUGCUUC**GCCCGCUACCCGACCACAUGAAGCAGCACGAC**UUCUUC**AAGUC**
CGCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUC**AAGGACGACGGCAACU**
ACAAGACCCGCGCCGAGGUGAAGUUC**GAGGGCGACACCCUGGUGAACCGCAUCGAGCUG**
AAGGGCAUCGACUUC**AAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAACUA**
 CAACAGCCACAAGGUCUAUAUCACCGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACU
 UCAAGACCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAG
 AACACCCCAUCGGCGACGGCCCCGUGCUGCUGCCCAGACAACCACUACCUGAGCACCCA
 GUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCAUCACAUGGUCCUGCUGGAGUUCG

UGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGCUCGAGAAGGCGGCA
ACCGCAAGGCUGACCGUGUAGGCGUGCGUCGCGAUGUCGCACGCGUUAAGACUUUACU
GAACGAGAAGGCGAAGCUUGAUCCGGC

mRNA secondary structure within the EmGFP sequence was predicted using mfold,⁸ for a 41-nucleotide sliding window, as described by.⁹ Fig S8 shows the predicted stability of the mRNA secondary structure centered at each nucleotide of the EmGFP sequence. For clarity purposes, the X-axis is labeled with the amino acid sequence numbers. The sequence had a predicted global average of -8.94 ± 3.75 (SD) kcal/mole. The most stable secondary structures are predicted to reside close to both termini, far removed from the regions (s1 – s5) which are being measured by smFRET. Segment s5 has both the highest translation rate and most stable secondary structure of the five segments we have examined. Thus, a previously suggested negative correlation between translation rate and predicted mRNA structural stability⁹ is not observed for the intermediate mRNA stabilities present here. Although the synonymous mutations in s3 in variants M1 and M2 result in some small changes in the predicted mRNA secondary structure, the predicted stability of the mutant s3 segments is not significantly different from the global average. The s3 stabilities in both variants are below that predicted for s5. Thus, it is very unlikely that any mRNA structures formed within M1 and M2 are responsible for the decrease in the translation rates of s3 in these variants.

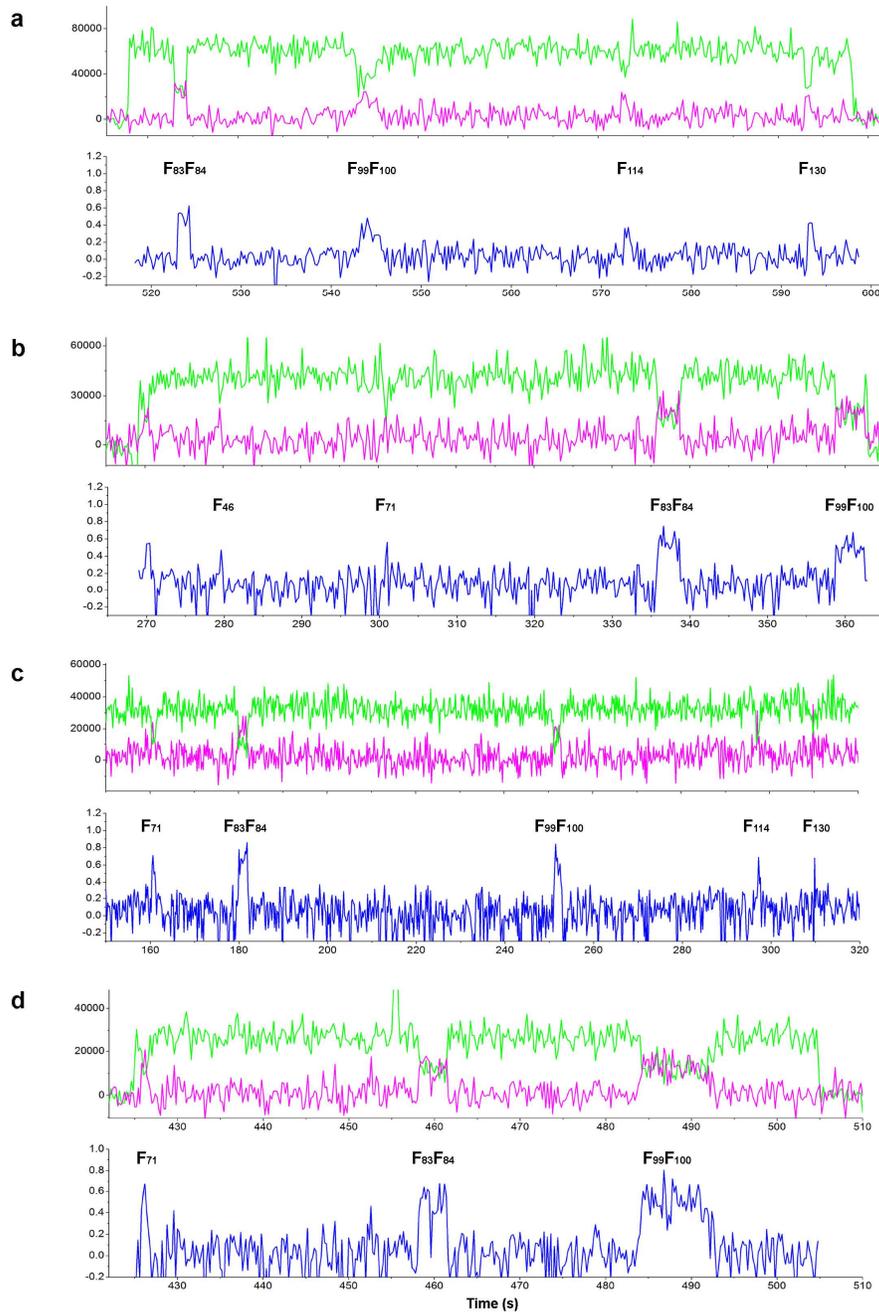


Figure S1. Single molecule trace gallery. The FRET traces in panels A-D show donor (green) and acceptor (magenta) emission and calculated FRET ratio (blue). The two longer FRET events in each trace were assigned to the Phe-Phe doublets ($F_{83}F_{84}$ and $F_{99}F_{100}$) in the EmGFP sequence, and the relatively shorter events were assigned to single Phe accommodations.

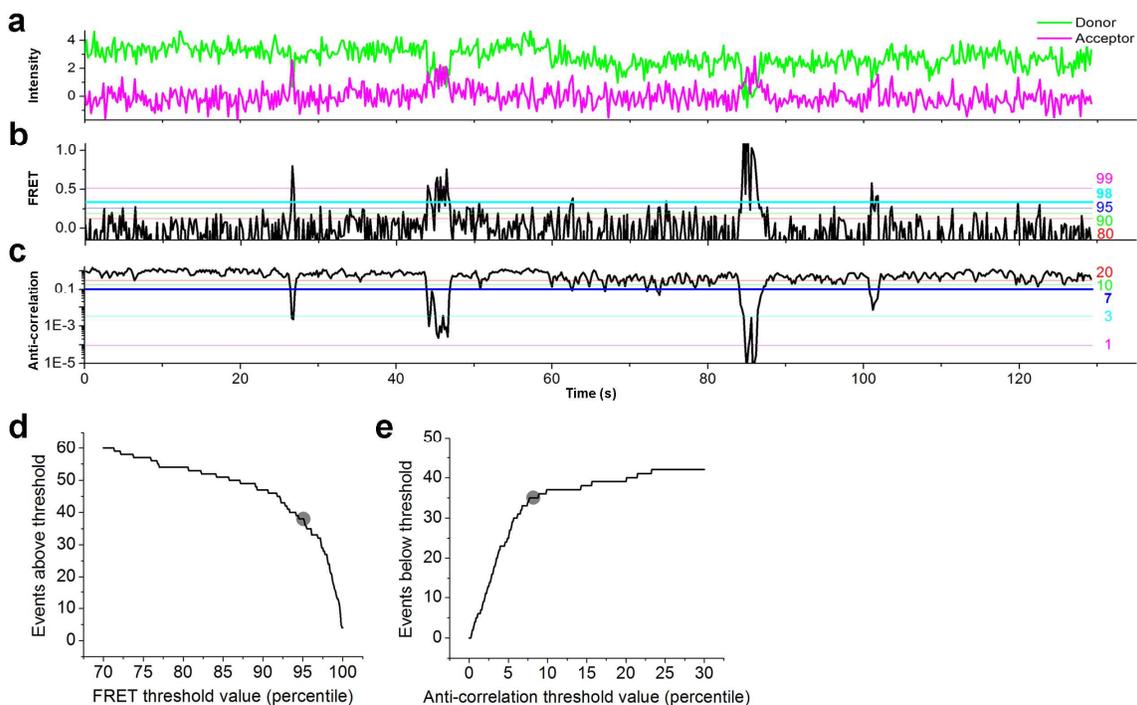


Figure S2. Objective selection of FRET events using a MATLAB algorithm. (a) Donor (green) and acceptor (magenta) intensity channels. (b) Calculated FRET efficiency and example thresholds set to reject various percentiles of the FRET values in the range 80% to 99%, as marked. The bold cyan line represents a FRET threshold selected by the algorithm. In this example, the threshold is set such that only 2% of the datapoints in the HaMMY-filtered FRET data are above the threshold. (c) The extent of anti-correlation was found by running t-tests on the intensity traces, and multiplying the P-values for the donor and acceptor channels at each time-point. Low values indicate statistically significant increases in acceptor signal occurring simultaneously with statistically significant decreases in donor signal. Thresholds set to accept various percentiles of the multiplied P-values time course in the range 20% to 1% are shown. The bold blue line represents the threshold, as determined by the algorithm, in this case accepting 7% of the data below this cutoff as anti-correlated events. (d) and (e) The number of data points

(video frames) above background were determined as a function of threshold value for both FRET (d) and anti-correlation (e). The break points in each case are indicated by a solid circle.

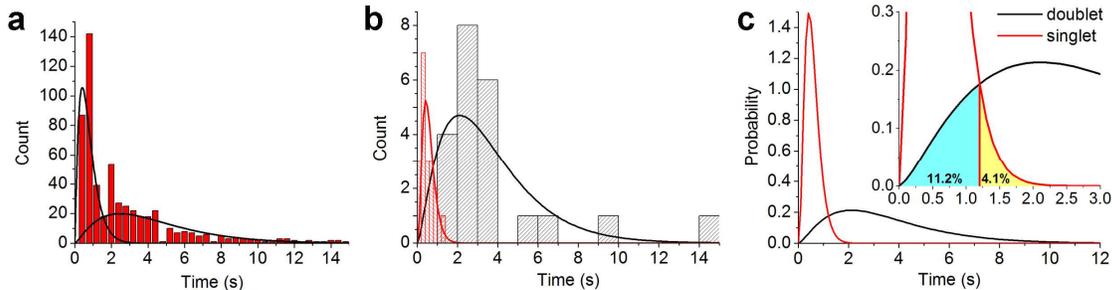


Figure S3. Duration of smFRET signals for Phe-Phe doublet and Phe singlet accommodations. (A) All traces were analyzed for the duration of FRET events, showing a best fit to two gamma distributions (reduced $\chi^2 = 3.7$ and 29.6 for fitting two and one gamma distribution, respectively. Reduced $\chi^2 = 7.6$ and 8.0 for fitting one or two exponentials, respectively). The data was selected automatically using the MATLAB algorithm. (B) A limited set of traces were chosen in which the two Phe-Phe doublets as well as single Phe accommodations were clearly identifiable in each trace. The durations of the FRET events were estimated and the distributions of FRET times for the doublets and the singlet accommodations were fitted separately to gamma distribution functions, showing average values of 3.6 ± 0.5 s and 0.6 ± 0.1 s, respectively (means \pm s.e.m.). (C) For assigning doublets and single Phe accommodation events, the area under each of the fitted curves from panel B was normalized to unity and the intersection of the two normalized curves defined a cutoff at 1.2 s. In the single accommodation curve, 4.1% of the area exceeds the intersection time, representing the expected proportion of relatively long single accommodation events. For the doublet curve, 11.2% of the events appear below the 1.2 s cutoff, indicative of short doublet events.

Segment	S1	S2	S3	S4	S5
	Pattern A			Pattern B	
Mean±CI95	46.0±4.6	20.4±1.4		24.3±2.6	20.2±2.2
% of traces	<i>n</i> = 28 (18.9%)			<i>n</i> = 30 (20.3%)	
	Pattern C				
Mean±CI95	47.2±5.9	19.5±1.2	23.7±1.1	22.9±2.1	20.5±3.1
% of traces	<i>n</i> = 90 (60.8%)				

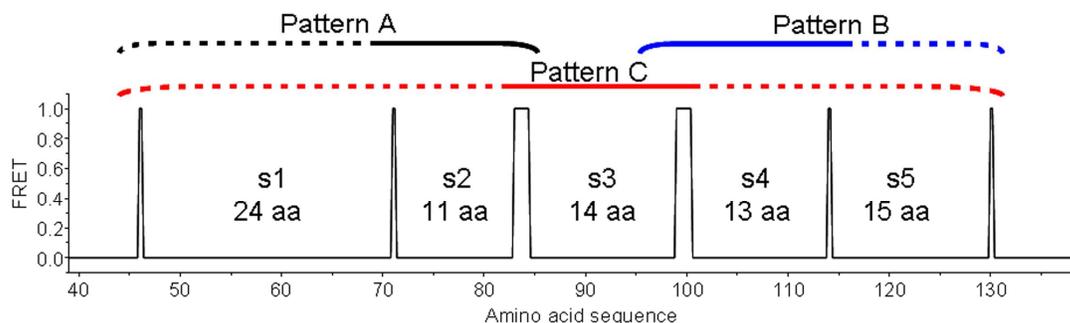


Figure S4. FRET pulse pattern fitting. An idealized FRET trace is shown, with FRET events located at F₄₆, F₇₁, F₈₃-F₈₄, F₉₉-F₁₀₀, F₁₁₄, and F₁₃₀. The two doublets are distinguished from the single Phe accommodation events by their relatively longer duration. Segment identities and numbers of amino acids within each segment are shown. The patterns (A, B and C), used to ascribe FRET events to their locations within the EmGFP sequence, are indicated. The Table, presenting the segment durations obtained from each pattern, shows that the durations of the intervals are consistent among the assigned patterns.

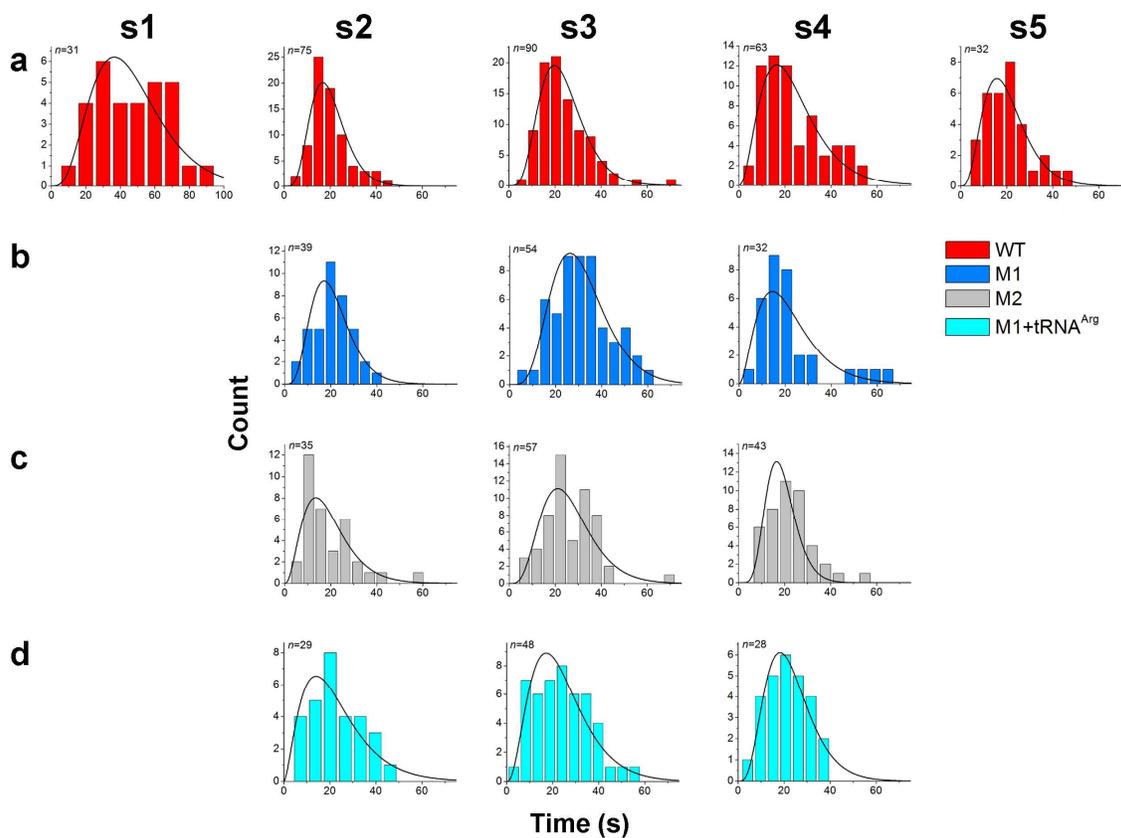


Figure S5. Distributions of segment translation times. (A) WT-plasmid; (B) M1-variant; (C) M2 variant; (D) M1 variant with 1 μM added tRNA^{Arg} . FRET events were selected by the MATLAB algorithm. The histograms were fit with gamma distribution functions.

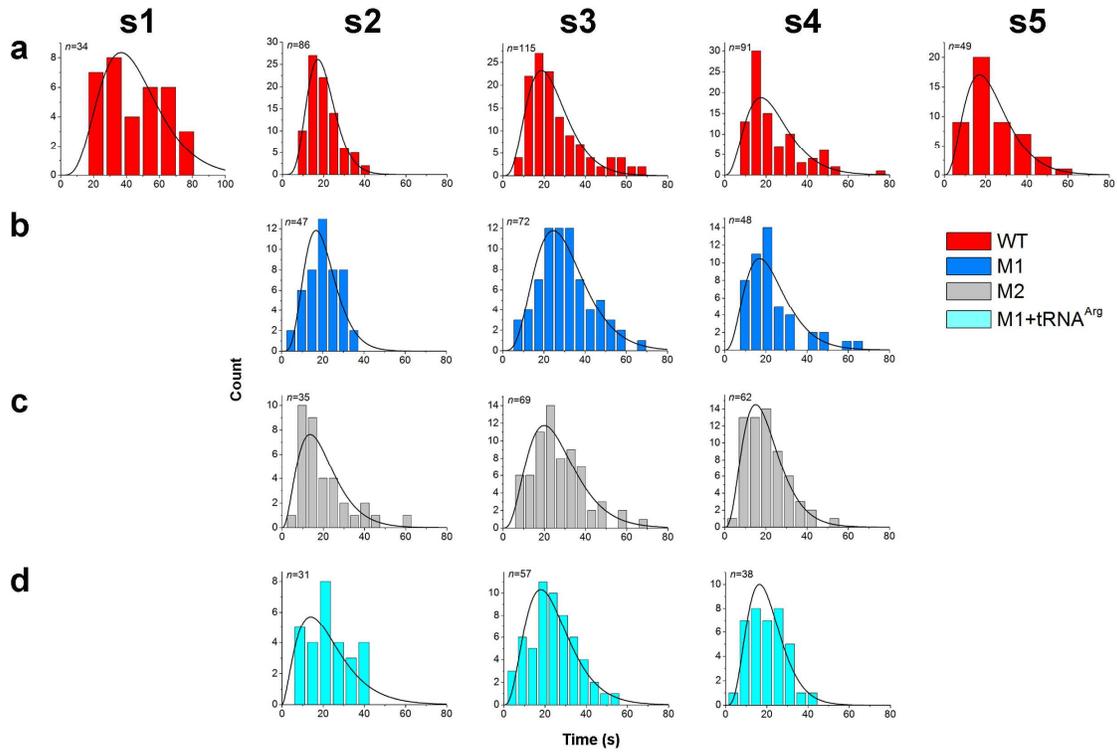


Figure S6. Distributions of segment translation times. Similar to Fig. S5 except the data presented were chosen by visual inspection of smFRET recordings. The means and s.e.m. values of the visually selected events are very similar to those selected objectively (Table 2).

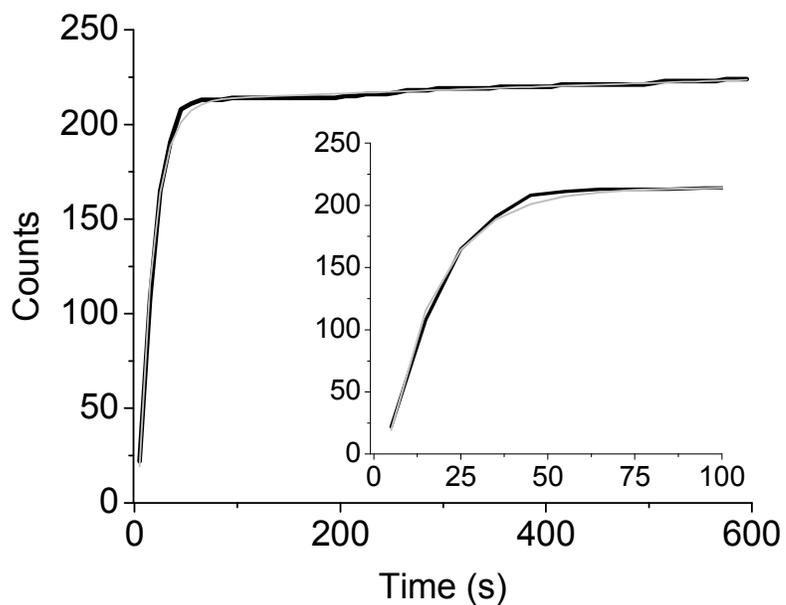


Figure S7. Accumulation of translating 70S^{Cy3} on the slide surface. Upon induction of the CFPS reaction with the addition of EmGFP plasmid, the solution is incubated for 3 minutes, followed by data acquisition. The accumulation of surface-bound 70S^{Cy3} from the start of data acquisition is biphasic, with the dominant (86%) and lesser (14%) phases proceeding with time constants of 15.4 ± 0.8 s, and 170 ± 100 s, respectively

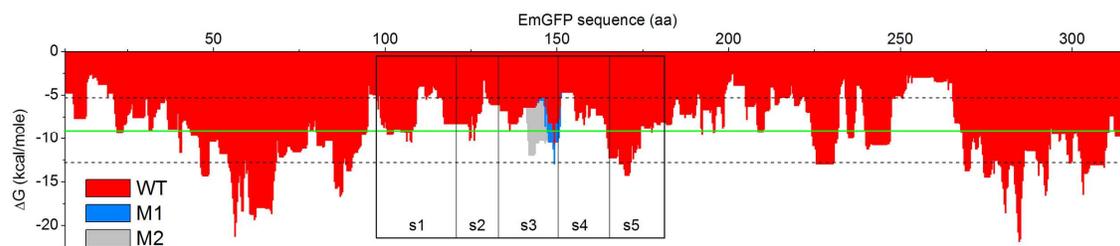


Fig. S8. Predicted stability of mRNA secondary structures. Free energies of secondary structure in the mRNAs were predicted using mfold⁸ using a 41-nucleotide sliding window as described in Kudla et al.⁹. The global average (green line) and standard deviation ranges (dashed lines) were calculated to be -8.94 ± 3.75 kcal/mole. The nucleotide sequence on the X-axis is given as amino acid sequence numbers for clarity. The EmGFP segments observed by smFRET are in boxes.

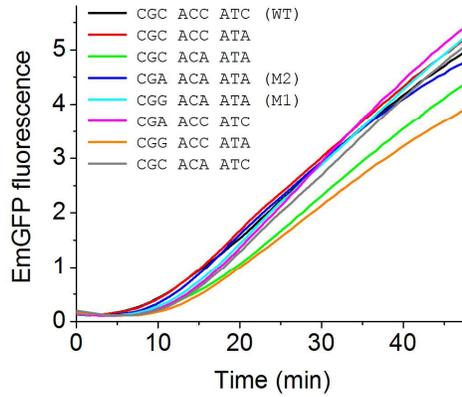


Fig. S9. Ensemble cell-free expression of EmGFP with synonymously modified coding sequences. The sequence of Arg⁹⁶Thr⁹⁷Ile⁹⁸, located in s3, was subjected to the indicated synonymous mutations. Lysate-based cell-free protein synthesis was initiated by the addition of 20 ng plasmid DNA.¹⁰ EmGFP fluorescence was continuously monitored using a plate reader with $\lambda_{\text{ex}} = 486 \text{ nm}$ and $\lambda_{\text{em}} = 535 \text{ nm}$. Little variation was observed in the time courses of the various samples. These results imply that slowing in segment s3 due to the synonymous mutations located at positions 96-98 do not markedly affect the overall rate of accumulation of active EmGFP.

Table S1. Codon usage and relative tRNA concentrations

tRNA	Codon recognition (5'-3')	Anticodon (5'-3')	Codon usage	Relative tRNA concentration ^a
Arg2	CGU, CGC, CGA	ICG ^b	3.11, 2.22, 0.13	7.37
Arg3	CGG	CCG	0.17	0.99
Thr1	ACC, ACU	GGU	2.65, 1.39	0.16
Thr3	ACC, ACU	GGU	2.65, 1.39	1.70
Thr4	ACA, ACU, ACG	UGU	0.35, 1.39, 0.75	1.42
Ile1	AUC, AUU	GAU	3.67, 2.14	5.39
Ile2	AUA	CAU	0.09	N/A

Data was obtained from ⁴. The frequency of codon usage and the relative tRNA concentrations are given in percentage of the total tRNA.

^a the concentrations of all tRNA isoacceptors for each aa are treated collectively.

^bThe tRNA isoacceptor Arg2 is the only bacterial tRNA isoacceptor that undergoes post-transcriptional modification of adenosine to inosine in the 5'-anticodon base.

SUPPLEMENTARY REFERENCES

- (1) McKinney, S. A.; Joo, C.; Ha, T. *Biophys J* 2006, 91, (5), 1941-51.
- (2) Letzring, D. P.; Dean, K. M.; Grayhack, E. J. *RNA* 2011, 16, (12), 2516-28.
- (3) Fluitt, A.; Pienaar, E.; Viljoen, H. *Comput Biol Chem* 2007, 31, (5-6), 335-46.
- (4) Dong, H.; Nilsson, L.; Kurland, C. G. *J Mol Biol* 1996, 260, (5), 649-63.
- (5) Geggier, P.; Dave, R.; Feldman, M. B.; Terry, D. S.; Altman, R. B.; Munro, J. B.; Blanchard, S. C. *J Mol Biol* 2010, 399, (4), 576-95.
- (6) Murphy, F. V. t.; Ramakrishnan, V. *Nat Struct Mol Biol* 2004, 11, (12), 1251-2.
- (7) Li, G. W.; Oh, E.; Weissman, J. S. *Nature* 2012, 484, (7395), 538-41.
- (8) Zuker, M. *Nucleic Acids Res* 2003, 31, (13), 3406-15.
- (9) Kudla, G.; Murray, A. W.; Tollervey, D.; Plotkin, J. B. *Science* 2009, 324, (5924), 255-8.
- (10) Rosenblum, G.; Chen, C.; Kaur, J.; Cui, X.; Goldman, Y. E.; Cooperman, B. S. *Nucleic Acids Res* 2012, 40, (12), e88.