JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Quantifying Elongation Rhythm during Full-Length Protein Synthesis

Gabriel Rosenblum,[†,‡] Chunlai Chen,[‡] Jaskiran Kaur,[§] Xiaonan Cui,[‡] Haibo Zhang,[†] Haruichi Asahara,[⊥] Shaorong Chong,[⊥] Zeev Smilansky,[§] Yale E. Goldman,*[,‡] and Barry S. Cooperman*[,†]

[†]Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, United States

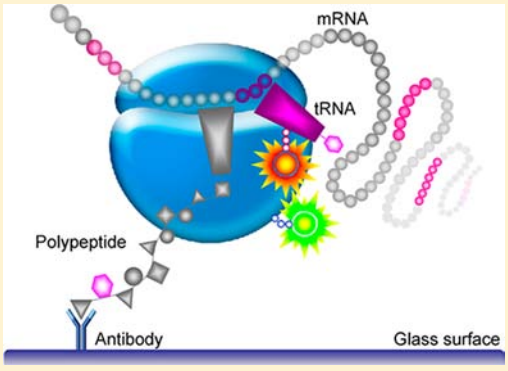
[‡]Pennsylvania Muscle Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6083, United States

[§]Anima Cell Metrology, Inc., 75 Claremont Road, Suite 102, Bernardsville, New Jersey 07924-2270, United States

[⊥]New England Biolabs Inc., 240 County Road, Ipswich, Massachusetts 01938, United States

Supporting Information

ABSTRACT: Pauses regulate the rhythm of ribosomal protein synthesis. Mutations disrupting even minor pauses can give rise to improperly formed proteins and human disease. Such minor pauses are difficult to characterize by ensemble methods, but can be readily examined by single-molecule (sm) approaches. Here we use smFRET to carry out real-time monitoring of the expression of a full-length protein, the green fluorescent protein variant Emerald GFP. We demonstrate significant correlations between measured elongation rates and codon and isoacceptor tRNA usage, and provide a quantitative estimate of the effect on elongation rate of replacing a codon recognizing an abundant tRNA with a synonymous codon cognate to a rarer tRNA. Our results suggest that tRNA selection plays an important general role in modulating the rates and rhythms of protein synthesis, potentially influencing simultaneous co-translational processes such as folding and chemical modification.



INTRODUCTION

Polypeptide elongation by ribosomes proceeds discontinuously,[1−3] with pauses regulating the rhythm of protein synthesis. Such pauses have been linked functionally to co-translational protein folding[4] and other co-translational processes such as chemical modification, ligand binding, oligomerization, interactions with chaperones and membranes (reviewed in refs 5 and 6), and programmed frame-shifts (reviewed in refs 7−9). Pausing elements include codon, adjacent codon-pair and tRNA isoacceptor usage, downstream and upstream mRNA secondary structure, nascent peptide and mRNA interactions with the ribosome, and unique tRNA−tRNA and tRNA−mRNA interactions.[10−15] Recent results suggest that internal Shine−Dalgarno[16] and sequences enriched in positively charged amino acids[17] are also capable of decreasing elongation rates. Pausing elements acting in unison sometimes result in very long pauses.[18,19]

During an elongation cycle, the ribosome oscillates between two different complexes (Figure 1). The pre-translocation (PRE) complex contains peptidyl-tRNA in the A-site (the tRNA entry site) and deacylated tRNA in the adjacent P-site. Transition to the post-translocation (POST) complex, in which the tRNAs are moved to the E(exit)- and P-sites, respectively, is catalyzed by the GTP form of elongation factor G (EF-G·GTP). The POST complex is converted to the PRE complex by the binding of aa-tRNA, as part of a ternary complex with the GTP bound form of elongation factor Tu (aa-tRNA·EF-Tu·GTP), to the cognate mRNA codon in the A-site, followed by transfer of the nascent peptidyl chain from the P-site tRNA to the aa-tRNA bound in the A-site.

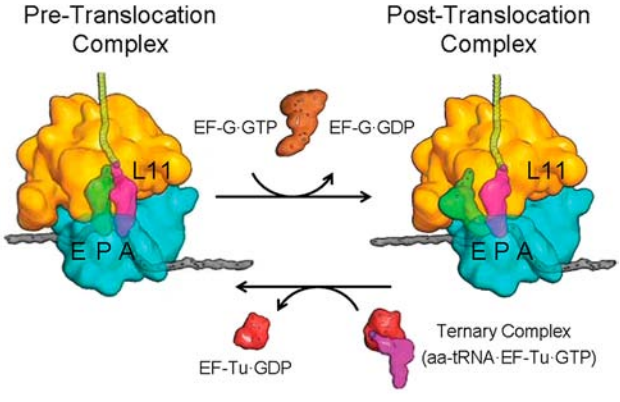


Figure 1. PRE and POST complexes. The ribosome shuttles back and forth between these two complexes, facilitated by EF-G·GTP and aa-tRNA·EF-Tu·GTP. The positions of tRNAs in the A, P, and E binding sites and of ribosomal protein L11 are indicated. L11 is close to tRNA in the A-site and much farther away from tRNA in the P-site.

Received: May 23, 2013
Published: July 3, 2013

   dx.doi.org/10.1021/ja405205c | J. Am. Chem. Soc. 2013, 135, 11322−11329

Characterizing the rhythm of protein synthesis requires determination of the rates of specific elongation cycles along the length of the protein chain. Because the dynamics of individual ribosomes are stochastic, it is not possible to synchronize their activities for more than a few turnovers, making it difficult to utilize ensemble methods to accurately identify and characterize other than quite long pauses. Single-molecule approaches overcome this limitation. Although a number of single-molecule multiple-turnover translation studies have recently been published, these studies employed non-physiological mRNAs that either incorporate an artificially long secondary structure, permitting use of an optical trap,[20] or monitor translation of mRNA sequences that encode a limited set of fluorescently labeled aminoacyl-tRNAs (aa-tRNAs).[21,22]

We showed earlier that single-molecule fluorescence resonance energy transfer (smFRET) between fluorescently labeled tRNAs and ribosomes, labeled on protein L11 and programmed with short mRNAs, could be used to monitor aa-tRNA binding to the A-site as part of a PRE complex, which fluctuates between classical (high FRET) and hybrid (intermediate FRET) forms, and its subsequent translocation as peptidyl-tRNA to the P-site (low FRET).[23] We also showed that cell-free synthesis of full-length EmGFP could be achieved using reaction mixtures in which labeled ribosomes and/or labeled Phe-tRNA^{Phe} replaced the corresponding endogenous unlabeled components.[24] Building on these results, we describe here the development and application of an experimental platform using smFRET, coupled with an appropriately modified cell-free protein synthesis system,[24] that permits the real-time monitoring of the expression of a full-length protein, the green fluorescent protein variant Emerald GFP (EmGFP). This approach allows quantitative estimation of how different pausing elements, alone or in tandem, affect translation rates, knowledge which is critical for understanding how these elements are utilized in regulating the rhythm of protein synthesis. In the present work we demonstrate clear correlations between elongation rates and codon and isoacceptor tRNA usage, and provide a quantitative estimate of the effect on elongation rate of replacing a codon cognate to an abundant tRNA with a synonymous codon cognate to a rare tRNA. Our results also indicate that equilibration of PRE complexes between the classic and hybrid forms is incomplete in ribosomes actively engaged in multiple-turnover polypeptide synthesis.

EXPERIMENTAL SECTION

Reagents. 70S^{Cy3} (see ref 23) and Phe-tRNA^{Phe}(Cy5.5) (charging and labeling efficiencies 26% and 79%, respectively[25,26]) were prepared as described. 70S^{Cy3}, labeled at position 87 of ribosomal protein L11, is as active as unlabeled 70S ribosomes.[24] Phe-tRNA^{Phe}(Cy5.5) has higher activity in supporting cell-free protein synthesis (CFPS) than Phe-tRNA^{Phe}(Cy5).[24] Bulk *E. coli* tRNA (Roche) was deacylated by incubation in 50 mM Tris-HCl pH 9.0 at 37 °C for 3 h, followed by three ethanol precipitations.

EmGFP Sequence and Nomenclature. The conventional sequence numbering of GFP was used, with the chromophore residues located at positions 65−67. Wild-type EmGFP plasmid DNA was derived from the commercially available pREST-EmGFP plasmid (Life Technologies) as described.[24] This plasmid is optimized for use in *E. coli*. The HA-Tag sequence (ACC AGC TAC CCA TAC GAT GTT CCA GAT TAC GCT) was inserted at the N-terminus between the initiator ATG codon and the following codon. This extension and the point mutations of variants M1 and M2 were constructed using a QuikChange Lightning site-directed mutagenesis kit (Stratagene) with primers set according to the manufacturer's instructions. Overall, the construct contained the 11-residue HA-Tag sequence, a 39-residue insertion (His-Tag, Xpress epitope, and EK cleavage site), the 239-residue EmGFP sequence, and a 31-residue C-terminal extension to enable folding of the protein while it is associated with the ribosome.[27]

Buffers and Stock Solutions Used in Forming the CFPS$^{Cy3/Cy5.5}$ Kit. Anti-fade reagents were prepared in water except as otherwise indicated: glutathione (Fisher Scientific, 100 mM); ascorbate (Acros, 250 mM); protocatechuic acid (PCA, MP Biomedicals, 50 mM), recrystallized from warm water three times;[28] protocatechuate-3,4-dioxygenase (PCD, Sigma), prepared as previously described[29,30] and purged with argon; Trolox/*n*-propyl gallate/4-nitrobenzyl alcohol (TX/nPG/NBA) solution[31] dissolved in methanol and containing 100 mM of each. The glutathione/ascorbate, PCA, and TX/nPG/NBA solutions were each titrated with KOH to pH 7.0: 70S^{Cy3}, 0.5 μM in TAM$_{15}$ buffer; bulk deacylated tRNA, 218 μM; AA-F, 8.33 mM Tyr, 100 mM Met, and 16.66 mM of each of the other amino acids except for Phe.

Buffers. TAM$_{15}$ buffer: 50 mM Tris-HCl pH 7.5, 30 mM NH_4Cl, 70 mM KCl, 15 mM $Mg(OAc)_2$, 1 mM DTT. Buffer A: 25 mM Tris-OAc pH 7.5, 100 mM K-glutamate, 10 mM $Mg(OAc)_2$, 1 mM reduced glutathione.

Preparation of the CFPS$^{Cy3/Cy5.5}$ Kit. This kit contained Phe-tRNA^{Phe}(Cy5.5) as the sole source of Phe and 70S^{Cy3} in place of endogenous 70S ribosomes. The preparation described below resembled that described earlier,[24] with the major exception that the CFPS$^{Cy3/Cy5.5}$ kit was derived from the PURExpress kit (New England Biolabs), whereas the earlier kits were derived from the lysate-based RTS 100 *E. coli* HY kit (5prime, Inc.). The CFPS$^{Cy3/Cy5.5}$ kit was prepared by replacing the amino acids, bulk *E. coli* tRNAs, Phe tRNA synthetase (PheRS), and ribosomes in the standard kit with the following: (i) a mixture of all amino acids except Phe, denoted AA-F; (ii) deacylated bulk *E. coli* tRNAs; (iii) Phe-tRNA^{Phe}(Cy5.5); and (iv) 70S^{Cy3}. The release and recycling factors in the standard kit were omitted. The following components were pre-mixed at the indicated volumes and/or final concentrations (f.c.) in a total volume of 10 μL: modified PURExpress solution A lacking amino acids and tRNA (3.54 μL); modified PURExpress solution B lacking PheRS, ribosomes, and recycling and release factors (2.24 μL); AA-F (0.2 μL, f.c. 0.15 mM Tyr, 2 mM Met and 0.3 mM of each of the other amino acids except for Phe); Phe-RS inhibitor *N*-benzyl-2-phenylethylamine (0.2 μL, f.c. 3 mM); bulk deacylated tRNA (0.37 μL, f.c. 8 μM); 70S^{Cy3} (0.2 μL, f.c. 10 nM); buffer A (0.4 μL); water (0.35 μL); glutathione/ascorbate solution (0.2 μL, f.c. 2 and 5 μM, respectively); TX/nPG/NBA solution (0.2 μL, f.c. 2 mM each); and PCD (0.2 μL, f.c. 0.001 unit/μL). This pre-mixture was partially deoxygenated by being placed under vacuum for 3 min, followed by addition of PCA (1 μL, f.c. 5 mM), Cy5.5-Phe-tRNA^{Phe} (0.7 μL, f.c. 420 nM measured as ^{14}C-Phe), and the EmGFP plasmid (0.2 μL, f.c. 2 ng/μL). The final reaction mixture was incubated at room temperature for 1 min and injected into a pre-treated microscope glass chamber (see Surface Preparation) pre-heated to 30 °C and incubated for an additional 2 min prior to data acquisition.

Surface Preparation. The sample flow chamber was formed on a PEG-biotin-coated glass coverslip as previously described.[23,32] The chamber (~3 μL) was filled with 0.5 mg/mL streptavidin (Sigma), incubated for 3 min, and washed with 20 μL of TAM$_{15}$ buffer. The chamber was then washed with 20 μL of biotinylated antibody against HA tag (Roche monoclonal, from rat IgG, 5 μg/mL) dissolved in TAM$_{15}$, incubated for 3 min, and finally washed with buffer A.

TIRF Microscopy Measurements. A custom-built objective-type total internal reflection fluorescence (TIRF) microscope was used that is based on a commercial inverted microscope (Eclipse Ti, Nikon) with a 1.49 NA 100× oil immersion objective (Apo TIRF, Nikon, Tokyo).[23] Single EmGFP spots appearing on the surface were detected using 488 nm laser excitation and 500−550 nm emission. A snapshot (0.3 s integration time) was taken every 10 s. Photobleaching was minimized by avoiding sample irradiation when data were not being recorded. Temporal accommodation of Phe-tRNA^{Phe}(Cy5.5) on 70S^{Cy3} was detected at 5 frames/s for 12 min, using constant 532 nm laser illumination. Emission signals were collected in the donor

channel (550−620 nm) and in the acceptor channel (685−735 nm) for sensitized emission. Two sets of anti-fading reagent mixtures were used: the PCA/PCD mixture was used in the cell-free translation of EmGFP experiments; the glucose oxidase/catalase mixture was used for experiments utilizing stalled ribosomes programmed with synthetic mRNA.

Data Analysis. Registration parameters for the donor and the acceptor channels were determined using 0.1 μm fluorescent microspheres (Tetraspeck, Life Technologies) and a custom-made program.[23] For each donor spot, a corresponding emission intensity trace was generated by Gaussian fitting to the acceptor channel, based on the registration parameters (rotation angle, ΔX and ΔY translation). Donor and acceptor traces that showed anti-correlation behavior were chosen (Supporting Information), and FRET efficiency was calculated as described in ref 23.

Statistical Analysis. The statistical significance of differences in the translational times of s3 for the various EmGFP variants was determined by performing an *F*-test of equality of variance, followed by a *t*-test, for the following comparisons: WT vs M1, M1 vs M1+tRNA^{Arg}, and WT vs M2 (see Figure 5); *t*-tests were also performed to determine the statistical significance of the correlation coefficients obtained by plotting the translational rates of the five wild-type EmGFP segments vs codon usage frequency or relative concentrations of tRNA isoacceptors (see Figure 4). In all cases the threshold *P*-value was set to 0.05. Gamma distribution functions were fitted to the duration histograms using MATLAB (Figures S5 and S6).

L11-tRNA FRET Efficiencies in the PRE Complexes. L11-tRNA FRET efficiencies in the PRE complexes were examined in the absence of EF-G (Figure 1) using initiation complexes (ICs) containing UUC-programed 70^{Cy3} premixed with the ternary complex (TC) EF-Tu·GTP·Phe-tRNA^{Phe}(Cy5.5). Initiation complexes were formed as previously described[23] using 5′ biotin-labeled mRNA (Dharmacon, Inc.) that had the following sequence: GGG AAU UCA AAA AUU UAA AAG UUA AUA AGG AUA CAU ACU *AUG UUC GUGUUC GUGUUC GUG UUC GUG*. The underlined sequence is a strong Shine−Dalgarno region. The italicized sequence codes for peptide MFVFVFV. Final concentrations of 1 nM IC and 10 nM TC were mixed in 20 μL of TAM$_{15}$ buffer and applied to a streptavidin-coated slide (see Surface Preparation). Unbound or unreacted reagents were washed away before video imaging was initiated. Single-molecule recording of intensity was carried out as indicated (see TIRF Microscopy Measurements), using an integration time of 100 ms. In order to diminish fluorophore photobleaching and blinking, measurements were carried out in the presence of an enzymatic deoxygenation system consisting of 3 mg/mL glucose, 100 μg/mL glucose oxidase (Sigma-Aldrich), 40 μg/mL catalase (Roche), and 1.5 mM TX (Trolox, Sigma-Aldrich, by dilution from a DMSO solution). FRET efficiencies were determined as previously described.[23]

RESULTS

smFRET Monitoring of EmGFP Synthesis. We first demonstrated that EmGFP synthesis was achieved on the slide surface, using TIRF microscopy to monitor the appearance of fluorescent spots corresponding to individual EmGFP molecules (Movie S1). Actively translating ribosomes were docked to the microscope slide by a surface-immobilized antibody against an N-terminal extension of EmGFP, a hemagglutinin (HA) affinity tag (Figure 2), a refinement over previous related work in which the entire population of ribosomes (including any non-translating ribosomes present) was attached to the surface.[27,33] We next used smFRET to monitor multiple events of binding and accommodation of individual fluorescent tRNAs [Phe-tRNA^{Phe}(Cy5.5)s] to ribosomes labeled with Cy3 at position 87 of protein L11 (denoted 70S^{Cy3}) during expression of EmGFP (Figure 3a,b, Figure S1 and Movie S2).

The translated EmGFP construct has a total of fourteen Phe residues, of which four occur in consecutive Phe-Phe doublets

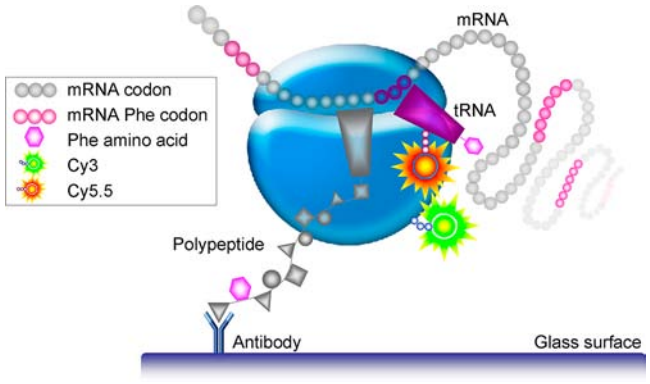


Figure 2. smFRET-TIRF monitoring of EmGFP expression. Antibodies against the HA-TAG at the N-terminus of the synthesized protein are immobilized to the PEG-coated glass surface. The reaction solution contains a reconstituted cell-free expression system that includes Phe-tRNA^{Phe}(Cy5.5) as the sole source of Phe for protein synthesis and 70S^{Cy3} replacing endogenous ribosomes. Two occurrences of a Phe-Phe codon doublet are located near the center of the EmGFP mRNA sequence. During Phe-tRNAPhe accommodation into the A-site, the tRNA-bound Cy5.5 is transiently held in close proximity to L11(Cy3) leading to FRET.[21,23] Donor and acceptor emission are recorded.

near the center of the EmGFP sequence. Phe-Phe doublet accommodations are distinguishable from single Phe-tRNA^{Phe}(Cy5.5) accommodations by the longer duration of their FRET events (Figure 3a,b), with only minor overlap in durations (Supporting Information and Figure S3). Identification of the consecutive Phe-Phe doublets within an EmGFP expression trace by the long duration of their FRET events has two important consequences for this work. First, it permits assignment of time intervals between two consecutive Phe-tRNA^{Phe} accommodations to particular sequence segments (Table 1) according to their temporal positions relative to the two Phe-Phe doublets (Figures S4−S6). Second, it allows the effects of mutations within the 14 amino acids that separate the two Phe-Phe doublets, denoted segment 3 (s3), to be examined quantitatively.

Objective Selection of FRET Events. FRET events were distinguished from random fluctuations of the recordings either visually or by using custom software employing a MATLAB-based, ad-hoc algorithm. Although the standard Hidden Markov Method software, HaMMy,[34] was successful in finding relatively long FRET events (≥4 frames) it often skipped shorter events (≤3 frames). The ad-hoc algorithm, which is described in detail in Supporting Information, overcomes this difficulty, by identifying *bona fide* FRET events within a trace when two criteria are satisfied simultaneously: (1) the magnitude of the FRET signal is significantly higher than background and (2) the donor (Cy3) and acceptor (Cy5.5) intensities are anti-correlated (Figure 3b,c). For the most part, the visual inspection and algorithm approaches selected the same events and gave similar mean translation times and SEM values (Table 2). Except where noted, all data presented were obtained by the algorithm-selection method.

Segment Translation Times. The distributions of translation times in each segment could be well fit by a gamma distribution function (Figures S5 and S6), which models multi-step processes[35] and is skewed toward longer intervals relative to a Gaussian distribution. The gamma function fit estimates average translation times and SEM values that are very similar to the arithmetic mean and SEM values. For example, the two

dx.doi.org/10.1021/ja405205c | J. Am. Chem. Soc. 2013, 135, 11322−11329

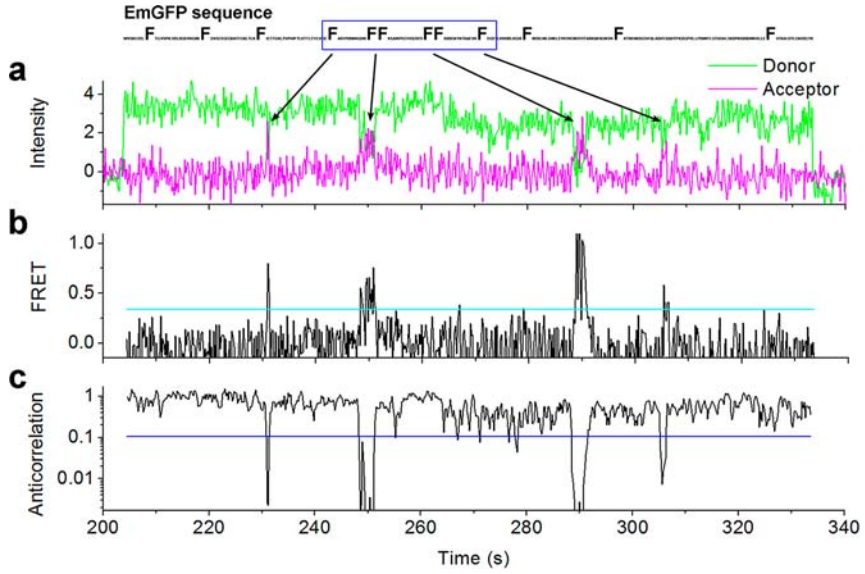


Figure 3. Objective selection of FRET events using an ad-hoc algorithm. (a) Donor (green) and acceptor (magenta) intensity channels. (b) Calculated FRET efficiency. The two longer FRET events were assigned to the Phe-Phe doublets (Supporting Information and Figure S3) in the native EmGFP sequence, and the relatively shorter events were assigned to Phe singlets. In this trace, the $70S^{Cy3}$ was immobilized on the surface 205 s after data acquisition commenced. Cy3-photobleaching was observed at 335 s. For the traces used in this work, Cy3 photobleaching times were exponentially distributed with a 63 ± 16 s time constant: i.e., the time interval from the appearance of Cy3 until it is photobleached. The bold cyan line represents a FRET threshold determined by the algorithm. (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. The blue bold line represents the threshold, as determined by the algorithm. For further details see Supporting Information and Figure S2.

Table 1. Translation Times and Rates of Wild-Type EmGFP Segments

segment	no. of amino acids in segment[1]	translation time (s)[2]	average codon translation rate constant (aa/s)[3]	sequence[4]
s1	24	46.6 ± 3.7(n = 31)	0.52 ± 0.04	F^{46}-ICTTGKLPVPWPTLVT TLTYGVQC-F^{71}
s2	11	19.8 ± 0.9(n = 75)	0.55 ± 0.03	F^{71}-ARYPDHMKQHD-$F^{83}F^{84}$
s3	14	23.7 ± 1.1(n = 90)	0.59 ± 0.03	$F^{83}F^{84}$- KSAMPEGYVQE**RTI**-$F^{99}F^{100}$
s4	13	23.6 ± 1.7(n = 63)	0.55 ± 0.04	$F^{99}F^{100}$-KDDGNYKTRAEVK-F^{114}
s5	15	20.3 ± 1.8(n = 32)	0.74 ± 0.07	F^{114}-EGDTLVNRIELKGID- F^{130}

[1]Excluding Phe. [2]Average translation times are calculated as mean ± SEM of the algorithm-selected events. Translation times between FRET events were measured, excluding the duration of the FRET events. [3]Calculated by dividing the number of amino acids in the segment by the average segment translation time. [4]EmGFP sequence with residues numbered such that the chromophore residues are $Thr^{65}Tyr^{66}Gly^{67}$. Each segment falls between Phe codons or Phe codon doublets. The $Arg^{96}Thr^{97}Ile^{98}$ sequence within s3, which is synonymously mutated in the experiments described in Figure 5, is bolded.

Table 2. Translation Times Determined by Visual and Objective Selection of Events[a]

segment	selection method	WT	M1	M2	M1+R
s2	visual	19.8 ± 0.8 (n = 86)	20.1 ± 1.1 (n = 47)	19.4 ± 2.0 (n = 35)	21.2 ± 2.0 (n = 31)
	algorithm	19.8 ± 0.9 (n = 75, c = 66)	20.8 ± 1.3 (n = 39, c = 36)	19.1 ± 1.9 (n = 35, c = 28)	22.5 ± 2.2 (n = 29, c = 25)
s3	visual	23.7 ± 1.1 (n = 115)	30.3 ± 1.5 (n = 72)	26.5 ± 1.6 (n = 69)	24.0 ± 1.5 (n = 57)
	algorithm	23.7 ± 1.1 (n = 90, c = 84)	31.2 ± 1.6 (n = 54, c = 54)	26.1 ± 1.5 (n = 57, c = 54)	24.0 ± 1.8 (n = 48, c = 48)
s4	visual	23.6 ± 1.4 (n = 91)	22.6 ± 1.8 (n = 48)	20.0 ± 1.3 (n = 62)	20.4 ± 3.36 (n = 38)
	algorithm	23.6 ± 1.7 (n = 63, c = 57)	21.7 ± 2.6 (n = 32, c = 27)	22.1 ± 1.5 (n = 43, c = 41)	23.0 ± 1.4 (n = 28, c = 26)

[a]Mean and SEM values of translation times (s), for each of the segments were obtained by arithmetic calculation. The two sets of data were obtained either by visual selection of FRET events or objectively, using the MATLAB code as described in Supporting Information and Figure S2. The number of detected events is n. The number of events common to both the visually selected events and code-selected events is c. For WT-s1, the mean durations were 45.1 ± 3.3 and 46.5 ± 3.7 s for visual and algorithm selection, respectively. For WT-s5 the mean durations were 23.0 ± 1.8 and 20.3 ± 1.8 s for visual and algorithm selected events, respectively. For the other variants, the number of detected events in segments s1 and s5 were too low to yield reliable mean values. The data presented in the main text were obtained from code-selected FRET events, using arithmetic calculation of the mean.

estimates for the translation time of s3 within the native EmGFP sequence (WT-s3) are 22.6 ± 1.0 and 23.7 ± 1.1 s, respectively (Table 2). Such similarity is expected in analyzing processes containing multiple serial events, as in the present case for elongation cycles within a segment. The average translation rate per codon within a given segment, which is

dx.doi.org/10.1021/ja405205c | J. Am. Chem. Soc. 2013, 135, 11322−11329

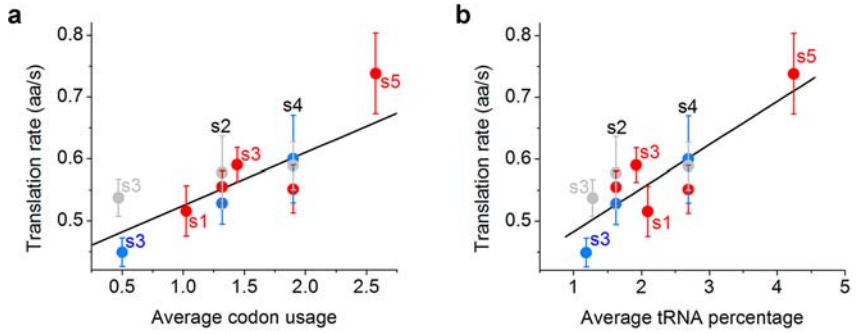


Figure 4. Correlating translation rates of EmGFP segments of WT and M1 and M2 variants (see Synonymous Mutations in s3) with codon usage (a) and tRNA abundance (b). Translation rates were calculated as described in Table 1 for WT (red) and M1 (blue) and M2 (gray) variants for the algorithm-selected FRET events. Frequency of codon usage (per 100 codons) and tRNA relative concentrations were reported by[10] and are given as percentages. Values of individual codons and tRNA relative concentrations were averaged over each segment using eqs 1A and 1B (Supporting Information). Predicted values of codon usage and tRNA percentage do not change between WT, M1, and M2, for segments s2 and s4, but are changed for s3. Reliable values for segments s1 and s5 were obtained only for WT. Error bars represent SEM.

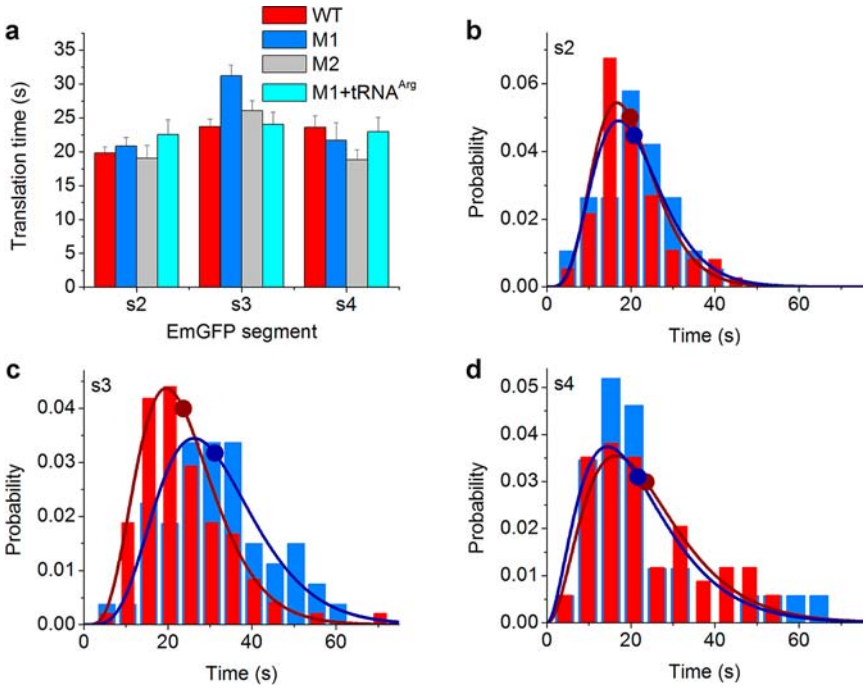


Figure 5. Translation times of segments of EmGFP variants. (a) Arithmetic mean translation times and SEM are presented. The segments are defined in Table 1. The variants contain synonymous mutations in which the native DNA sequence encoding $Arg^{96}Thr^{97}Ile^{98}$, CGC ACC ATC, was replaced by the synonomous sequences CGG ACA ATA (M1) or CGA ACA ATA (M2). M1+$tRNA^{Arg}$ refers to cell-free expression of the M1 variant in the presence of 1 μM added purified $tRNA^{Arg}$ (all isoacceptors). (b–d) Normalized distributions (bars), fitted gamma functions (lines), and average translation times (dots) of segments s2–s4 for the wild-type and M1 plasmids. Application of a statistical *t*-test yielded *P* values demonstrating that the observed differences in s3 between (i) WT and M1 ($P = 0.0001$); (ii) M1 and M1+$tRNA^{Arg}$ ($P = 0.003$); and (iii) M1 and M2 ($P = 0.02$) were significant.

equal to the number of translated amino acids/total translation time for each segment, varies from 0.52 ± 0.04 s^{-1} for s1 to 0.74 ± 0.07 s^{-1} for s5 (means ± SEM, Table 1). The overall average for s1 − s5, 0.57 ± 0.02 s^{-1}, measured at 30 °C, is in the range of values reported for (i) ensemble lysate-based cell-free protein synthesis (CFPS) (1.5 s^{-1}, 37 °C),[36] (ii) ensemble-reconstituted CFPS (0.3 s^{-1}, 25 °C),[37] (iii) single-molecule optical trap measurements under applied force, using a limited tRNA set (0.45 s^{-1}, room temperature),[20] and (iv) smTIRF measurements using a limited set of labeled tRNAs (0.1–1 s^{-1}, room temperature).[22] Importantly, the codon translational rates showed significant positive correlations with two variables that are themselves correlated:[10] average codon usage and average cognate tRNA abundance in *E. coli* (Figure 4).

Synonymous Mutations in s3. To test the effect on translation time of placing rare codons within the s3 sequence, three consecutive wild-type codons (CGC ACC ATC) encoding $Arg^{96}Thr^{97}Ile^{98}$ were replaced by three rare synonymous codons that also code for ArgThrIle (codons CGG ACA ATA,[10,38] termed variant M1). The average translation time of M1-s3 was 31.2 ± 1.6 s, a significant increase (*t*-test, $P = 0.0001$) of 7.5 ± 1.4 s relative to the value of 23.7 ± 1.1 s for WT-s3 (Figure 5a,c). Replacement of the CGG codon at residue 96 in M1 by CGA (mutant M2), also a rare codon but cognate to a more plentiful isoacceptor $tRNA^{Arg}$

dx.doi.org/10.1021/ja405205c | J. Am. Chem. Soc. 2013, 135, 11322–11329

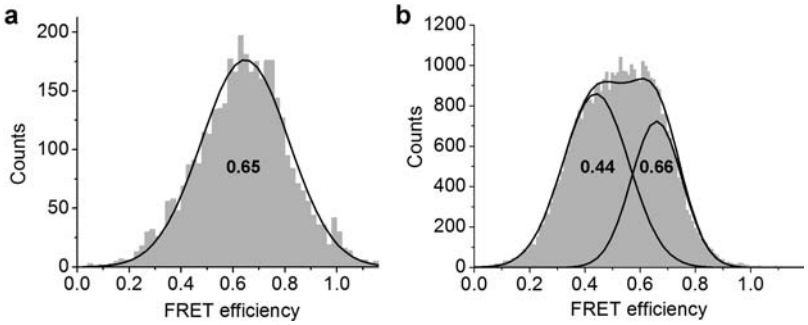


Figure 6. Distribution of FRET efficiencies of $70S^{Cy3}$ and Phe-$tRNA^{Phe}$(Cy5.5). (a) FRET efficiencies observed during EmGFP translation were fitted to a single Gaussian centered at 0.65 ± 0.17 (mean ± SD, reduced χ^2 = 2.6). (b) FRET efficiencies of PRE complexes were determined for a UUC-programmed $70S^{Cy3}$ in the absence of EF-G (Supporting Information). Two Gaussians resulted in a statistically significant better fit than a single Gaussian (reduced χ^2 = 3.9 and 9.1 for a double and a single Gaussian fit, respectively), as determined by an F-test ($P < 10^{-6}$). The two Gaussians are centered at 0.44 ± 0.12 and 0.66 ± 0.09, with corresponding proportions of 61% and 39%, respectively. The sum of the two Gaussian components is also shown. The component with high FRET efficiency (0.66) was assigned to the classical state, and the component with lower FRET (0.44) was assigned to the hybrid state.[23].

(Table S1), led to a statistically insignificant change (*t*-test, P = 0.2) of s3 translation time (M2-s3 = 26.1 ± 1.5 s) relative to WT, but to a significantly more rapid s3 translation time (*t*-test, P = 0.02) relative to M1. As expected, neither substitution had a significant effect on the translation times of the neighboring segments (Figure 5b,d). These results confirm that the Phe-Phe doublets are correctly assigned, providing support for the reliability of our procedure that identifies sequence segments based on their temporal positions relative to the long FRET pulses. The introduction of synonymous mutations changed the translation rates, although the strong correlations between such rates and both codon usage and tRNA percentage were retained. (Figure 4). For WT-s3, the average translation time per codon is 1.4 s. Assuming that the wild-type CGC codon has a similar translation time leads to the conclusion that the 7.5 ± 1.4 s increase in s3 seen on CGG substitution is due to an approximately 6-fold increase [(1.4 + 7.5)/1.4] in the translation time, or 6-fold decrease of translation rate, for Arg incorporation into the nascent peptide. As CGC replacement by CGG is not predicted to significantly alter the energetics of mRNA secondary structure (Supporting Information and Figure S8), we interpret such slowing as being due to the increased time spent by the ribosome waiting for a low abundance cognate tRNA to bind.[10,39] Supporting this interpretation, addition of extra $tRNA^{Arg}$ (1 μM, all Arg isoacceptors) during translation of the M1 mutant restored the translation time to a value (24.0 ± 1.8 s, Figure 5a) indistinguishable from that seen with wild-type (*t*-test, P = 0.88).

Structural Dynamics of PRE Complexes during Elongation. Productive binding events, involving cognate codon−anticodon interactions and leading to incorporation of Phe into the nascent chain, denoted "accommodations", predominate over codon-independent interactions[40] (see Supporting Information). Accommodation events during EmGFP synthesis exhibit a distribution of FRET efficiencies with a single peak centered at 0.65 (Figure 6a). By contrast, in stalled $70S^{Cy3}$ PRE complexes (i.e., in the absence of EF-G), peptidyl-$tRNA^{Phe}$(Cy5.5) bound in the A-site fluctuates between classic and hybrid states, characterized by FRET efficiencies of 0.66 and 0.44, respectively, which are present in roughly equal amounts (Figure 6b, Supporting Information, and references[23,41,42]). These results indicate that, during elongation of EmGFP under the present conditions, the classic state predominates, possibly due to rapid translocation before the classic/hybrid distribution can equilibrate. A short-lived hybrid state is, however, a likely intermediate during conversion of the classic state PRE complex to POST complex[23] (Supporting Information). It remains to be determined to what extent the low occupancy of the hybrid state obtained with Phe-$tRNA^{Phe}$ is representative of other charged tRNAs.

■ DISCUSSION

Synonymous mutations that are presumed to induce local changes in translation rates have been shown to result in protein misfolding and human disease,[43] emphasizing the regulatory role that can be played by tRNA selection at specific locations along the mRNA sequence. In this paper we provide quantitative evidence that translation proceeds along the sequence of full-length EmGFP with a variable rhythm, strongly influenced by cognate isoacceptor tRNA abundance (Figures 4 and 5). Indeed, since EmGFP mRNA has relatively few internal Shine−Dalgarno sequences[16] and lacks strong predicted mRNA secondary structures (Supporting Information and Figure S8), it is likely that cognate isoacceptor tRNA abundance is the predominant modulator of translation rates for specific EmGFP mRNA segments. Such modulation might be rather general for *E. coli* mRNAs, given the over-representation of clusters of rare codons found in the *E. coli* genome.[44]

A direct demonstration of the influence of tRNA availability on local translation rate is provided by the approximately 6-fold slower translation of codon CGG within s3 of M1 mRNA, coding for a rare $tRNA^{Arg}$, as compared with CGC within s3 of WT mRNA, coding for a more plentiful $tRNA^{Arg}$, and the elimination of this difference when the total concentration of $tRNA^{Arg}$ isoacceptors is increased. Although the rates and yields of EmGFP expression for the WT, M1, M2, and other variants with synonymous codons in positions 96−98 (Figure S9) are not significantly affected by the locally reduced rate, reducing or accelerating the translation rate of short segments in other regions of the coding sequence, more critical for protein folding, might lead to more marked effects on overall EmGFP expression. An interesting candidate for further examination is the s1-s2 region. Translation of this region, comprising residues 47−82, occurs while EmGFP β-strands 1 and 2 (residues 12−21 and 25−36) and a portion of strand 3 (residues 41−42), which form part of the folding nucleus of GFP proteins,[45] are

dx.doi.org/10.1021/ja405205c | J. Am. Chem. Soc. 2013, 135, 11322−11329

expected to emerge from the ribosomal exit tunnel.[46] We speculate that the relatively slow translation rates of s1 and s2 (Figure 4) increase the probability for correct co-translational folding of the emerging folding nucleus, thereby increasing overall folding efficiency.

While ensemble methods are capable of detecting rather long pauses that are found in specific proteins,[2,3,13,16,18] the single-molecule experimental platform presented here can reveal relatively moderate attenuations, resulting from sequences enriched in rare codons or other pausing elements. Such elements, which affect the rate of polypeptide chain elongation and the concomitant vectorial emergence of the nascent polypeptide from the exit tunnel, are widely distributed in mRNAs and peptide sequences, suggesting that proteins are quite generally translated with variable rates and rhythms. Future work will quantify translation rate effects of additional pausing elements, singly or in tandem, providing further mechanistic details regarding coupling between translational rhythm and protein functionality, and allow comparisons of the relative magnitudes of modulation by different pausing elements. We are mindful that great caution will be needed in using quantitative results found in such *in vitro* studies to rationalize *in vivo* results and trends, given the large disparity between *in vivo* and optimized *in vitro* polypeptide elongation rates (8−15 s^{-1} [47] vs 1.5 s^{-1},[23] respectively, at 37 °C).

Although the present approach for identifying specific segments within the EmGFP sequence relies on the fortunate placement of two consecutive Phe doublets, it can be ported straightforwardly to other protein sequences through the use of pairs of different labeled isoacceptor tRNAs that are encoded by consecutive codons. In addition, it will generally be possible to reduce segment size by introducing additional codons cognate to labeled isoacceptor tRNAs via site-specific mutations that do not affect protein function. Such mutations would increase the relative changes in segment translation times due to pausing elements, thus improving the precision with which such changes could be measured. Efforts in these directions are underway.

■ ASSOCIATED CONTENT

Supporting Information

Supporting methods, results, figures, and two movies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

goldmany@mail.med.upenn.edu; cooprman@pobox.upenn.edu

Notes

The authors declare the following competing financial interest(s): Drs. Goldman and Cooperman are consultants for Anima Cell Metrology.

■ ACKNOWLEDGMENTS

This work was supported by NIH grant GM080376 to B.S.C. and Y.E.G., a Human Frontier Science Program postdoctoral fellowship to G.R., and an American Heart Association Postdoctoral Fellowship to C.C. Drs. Goldman and Cooperman are consultants for Anima Cell Metrology. We are thankful for helpful discussions on FRET event selection with Professors Philip Nelson (University of Pennsylvania) and David Talaga (Montclair State University).

■ REFERENCES

(1) Guisez, Y.; Robbens, J.; Remaut, E.; Fiers, W. *J. Theor. Biol.* **1993**, *162*, 243−252.
(2) Varenne, S.; Buc, J.; Lloubes, R.; Lazdunski, C. *J. Mol. Biol.* **1984**, *180*, 549−576.
(3) Zhang, G.; Hubalewska, M.; Ignatova, Z. *Nat. Struct. Mol. Biol.* **2009**, *16*, 274−280.
(4) Kimchi-Sarfaty, C.; Oh, J. M.; Kim, I. W.; Sauna, Z. E.; Calcagno, A. M.; Ambudkar, S. V.; Gottesman, M. M. *Science* **2007**, *315*, 525−528.
(5) Fedorov, A. N.; Baldwin, T. O. *J. Biol. Chem.* **1997**, *272*, 32715−32718.
(6) Kramer, G.; Boehringer, D.; Ban, N.; Bukau, B. *Nat. Struct. Mol. Biol.* **2009**, *16*, 589−597.
(7) Baranov, P. V.; Gesteland, R. F.; Atkins, J. F. *RNA* **2004**, *10*, 221−230.
(8) Dinman, J. D. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 661−673.
(9) Farabaugh, P. J. *Annu. Rev. Genet.* **1996**, *30*, 507−528.
(10) Dong, H.; Nilsson, L.; Kurland, C. G. *J. Mol. Biol.* **1996**, *260*, 649−663.
(11) Irwin, B.; Heck, J. D.; Hatfield, G. W. *J. Biol. Chem.* **1995**, *270*, 22801−22806.
(12) Letzring, D. P.; Dean, K. M.; Grayhack, E. J. *RNA* **2011**, *16*, 2516−2528.
(13) Sorensen, M. A.; Pedersen, S. *J. Mol. Biol.* **1991**, *222*, 265−280.
(14) Chen, C.; Zhang, H.; Broitman, S. L.; Reiche, M.; Farrell, I.; Cooperman, B. S.; Goldman, Y. E. *Nat. Struct. Mol. Biol.* **2013**, *20*, 582−588.
(15) Kudla, G.; Murray, A. W.; Tollervey, D.; Plotkin, J. B. *Science* **2009**, *324*, 255−258.
(16) Li, G. W.; Oh, E.; Weissman, J. S. *Nature* **2012**, *484*, 538−541.
(17) Charneski, C. A.; Hurst, L. D. *PLoS Biol.* **2013**, *11*, e1001508.
(18) Giedroc, D. P.; Theimer, C. A.; Nixon, P. L. *J. Mol. Biol.* **2000**, *298*, 167−185.
(19) Ramachandiran, V.; Kramer, G.; Horowitz, P. M.; Hardesty, B. *FEBS Lett.* **2002**, *512*, 209−212.
(20) Wen, J. D.; Lancaster, L.; Hodges, C.; Zeri, A. C.; Yoshimura, S. H.; Noller, H. F.; Bustamante, C.; Tinoco, I. *Nature* **2008**, *452*, 598−603.
(21) Chen, C.; Stevens, B.; Kaur, J.; Smilansky, Z.; Cooperman, B. S.; Goldman, Y. E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16980−16985.
(22) Uemura, S.; Aitken, C. E.; Korlach, J.; Flusberg, B. A.; Turner, S. W.; Puglisi, J. D. *Nature* **2010**, *464*, 1012−1017.
(23) Chen, C.; Stevens, B.; Kaur, J.; Cabral, D.; Liu, H.; Wang, Y.; Zhang, H.; Rosenblum, G.; Smilansky, Z.; Goldman, Y. E.; Cooperman, B. S. *Mol. Cell* **2011**, *42*, 367−377.
(24) Rosenblum, G.; Chen, C.; Kaur, J.; Cui, X.; Goldman, Y. E.; Cooperman, B. S. *Nucleic Acids Res.* **2012**, *40*, e88.
(25) Kaur, J.; Raj, M.; Cooperman, B. S. *RNA* **2011**, *17*, 1393−1400.
(26) Pan, D.; Qin, H.; Cooperman, B. S. *RNA* **2009**, *15*, 346−354.
(27) Katranidis, A.; Atta, D.; Schlesinger, R.; Nierhaus, K. H.; Choli-Papadopoulou, T.; Gregor, I.; Gerrits, M.; Buldt, G.; Fitter, J. *Angew. Chem., Int. Ed.* **2009**, *48*, 1758−1761.
(28) Crawford, D. J.; Hoskins, A. A.; Friedman, L. J.; Gelles, J.; Moore, M. J. *RNA* **2008**, *14*, 170−179.
(29) Aitken, C. E.; Marshall, R. A.; Puglisi, J. D. *Biophys. J.* **2008**, *94*, 1826−1835.
(30) Patil, P. V.; Ballou, D. P. *Anal. Biochem.* **2000**, *286*, 187−192.
(31) Dave, R.; Terry, D. S.; Munro, J. B.; Blanchard, S. C. *Biophys. J.* **2009**, *96*, 2371−2381.
(32) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507−516.
(33) Uemura, S.; Iizuka, R.; Ueno, T.; Shimizu, Y.; Taguchi, H.; Ueda, T.; Puglisi, J. D.; Funatsu, T. *Nucleic Acids Res.* **2008**, *36*, e70.
(34) McKinney, S. A.; Joo, C.; Ha, T. *Biophys. J.* **2006**, *91*, 1941−1951.
(35) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: Cambridge, England, 1995.

(36) Underwood, K. A.; Swartz, J. R.; Puglisi, J. D. *Biotechnol. Bioeng.* **2005**, *91*, 425−435.
(37) Takahashi, S.; Tsuji, K.; Ueda, T.; Okahata, Y. *J. Am. Chem. Soc.* **2012**, *134*, 6793−6800.
(38) Sharp, P. M.; Li, W. H. *Nucleic Acids Res.* **1986**, *14*, 7737−7749.
(39) Ikemura, T. *Mol. Biol. Evol.* **1985**, *2*, 13−34.
(40) Geggier, P.; Dave, R.; Feldman, M. B.; Terry, D. S.; Altman, R. B.; Munro, J. B.; Blanchard, S. C. *J. Mol. Biol.* **2010**, *399*, 576−595.
(41) Blanchard, S. C.; Kim, H. D.; Gonzalez, R. L., Jr.; Puglisi, J. D.; Chu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12893−12898.
(42) Munro, J. B.; Altman, R. B.; O'Connor, N.; Blanchard, S. C. *Mol. Cell* **2007**, *25*, 505−517.
(43) Sauna, Z. E.; Kimchi-Sarfaty, C. *Nat. Rev. Genet.* **2011**, *12*, 683−691.
(44) Clarke, T. F. t.; Clark, P. L. *PLoS One* **2008**, *3*, e3412.
(45) Huang, J. R.; Hsu, S. T.; Christodoulou, J.; Jackson, S. E. *HFSP J.* **2008**, *2*, 378−387.
(46) Kramer, G.; Ramachandiran, V.; Hardesty, B. *Int. J. Biochem. Cell Biol.* **2001**, *33*, 541−553.
(47) Talkad, V.; Schneider, E.; Kennell, D. *J. Mol. Biol.* **1976**, *104*, 299−303.