Supporting Information

High-Efficiency "-1" and "-2" Ribosomal Frameshiftings Revealed by Force Spectroscopy

Te-Wei Tsai,[1,2] Haopeng Yang,[1] Heng Yin,[2] Shoujun Xu[2,*] & Yuhong Wang[1,*]

[1]Department of Biology and Biochemistry, [2]Department of Chemistry, University of Houston, Houston, TX 77204, USA

Table of Contents

1. Materials and experimental methods

Sequences of the mRNA and probing DNAs. All of the mRNAs and DNA probes are purchased from IDTDNA (Integrated DNA Technologies). The sequence of the mRNA containing the GA_7G motif was 5′-Bio-C AAC UGU UAA UUA AAU UAA AUU AAA AAG GAA AUA AAA AUG UUU GAA AAA AAG UAC GUA AAU CUA CUG CUG AAC UC-3′; the mRNAs containing the GGA6G and the GA3GUA2 (NS) motifs were the same as above except the replacements of GGA AAA AAG and GAA AGU AAG at the slippery sites, respectively; and the mRNA containing the dnaX stem loop was: 5′-Bio-GU UAA UUA AAU UAA AUU AAA AAG GAA AUA AAA AUG UUU GAA AAA AAG UAC GUA CCG GCA GCC GCU ACC CGC GGC GGC CGG UUG GUC UAC G-3′. The slippery motifs and the stem loop structure are underlined. Bio: biotin functionalized.

The sequences of the probing DNAs were as follows:

P15a: 3′-**G CAT TTA GAT GAC GA**G AAC TC/TEGBio/-5′;

P14: 3′-**CAT TTA GAT GAC GA**G AAC TC/TEGBio/-5′;

P13: 3′-**AT TTA GAT GAC GA**G AAC TC/TEGBio/-5′;

P12: 3′-**T TTA GAT GAC GA**G AAC TC/TEGBio/-5′;

P18: 3′-**C ATG CAT TTA GAT GAC GA**G AAC TC/TEGBio/-5′;

P15b: 3′-**T TTA GAT GAC GAC TT**C TCG AA/TEGBio/-5′;

P15c: 3′-**A GAT GAC GAC TTG AG**G AAC TC/TEGBio/-5′;

SLP15a: 3′-**GCA TGG CCG TCG GCG** TAC CCG/TEGBio/-5′;

SLP15b: 3′-**CCG TCG GCG ATG GGC**/TEGBio/-5′. The bases in bold are complementary bases with the mRNA. TEG: linker molecule (Integrated DNA Technologies). The probes were designed so that the ribosome complexes always formed 12-15 bp duplexes with the probes for optimal force resolution.

All other reagents were purchased from Sigma-Aldrich.

Formation of Pre_1 ribosome complex. All of the mixtures are in TAM_{10} buffer: 20 mM tris-HCl (pH 7.5), 30 mM NH4Cl, 70 mM KCl, 5 mM EDTA, and 7 mM BME (2-mercaptoethanol). Five mixtures were prepared: ribosome mix, TuWG mix, Tu0G mix, A mix, and A-Lys mix. The details are as follows. The ribosome mix contained 1 μM ribosome, 1.5 μM of IF1, 2, 3, 2 μM of mRNA, 4 μM of charged fMet-$tRNA^{fMet}$, and 4 mM of GTP. The TuWG mix contained 6 μM EF-Tu, 3 μM EF-G, 4 mM GTP, 4 mM PEP, and 0.02 mg/ml Pyruvate Kinase. The Tu0G mix contained no EF-G but all the rest components in TuWG. The A mix contained 100 mM Tris (pH 7.8), 20 mM $MgAc_2$, 1 mM EDTA, 4 mM ATP, 0.1 mg/ml total synthetase, 50 A_{260}/ml total tRNA, and 0.25 mM of phenylalanine, glutamic acid. The A-Lys mix contained 100 mM Tris (pH 7.8),

20 mM $MgAc_2$, 1 mM EDTA, 4 mM ATP, 0.1 mg/ml total synthetase, 2 A_{260}/ml $tRNA^{Lys}$, and 0.25 mM of lysine.

The five mixes were incubated at 37 ºC for 25 min. The ribosome, TuWG and A mixes were mixed with 1:2:2 ratio and then incubated at 37 ºC for 15 min. The resulting ribosome complex was purified via 1.1 M sucrose cushion. The ribosome complex was then incubated with the A-Lys and Tu0G mixes at 37 ºC for 2 min to form Pre_1. Pre_1 was then purified via 1.1 M sucrose cushion.

Formation of $Post_1$ ribosome complex. The Pre_1 (1 μM) was incubated with an EF-G solution that contained 2 μM EF-G, 4 mM GTP, 4 mM PEP, and 0.02 mg/ml Pyruvate Kinase at 37 ºC for 30 min. $Post_1$ was purified via 1.1M sucrose cushion.

Formation of $Post_2$ ribosome complex. The $Post_2$ was formed by incubating $Post_1$ (1 μM) with A-Lys and TuWG mixes in the ratio of 1:2:2 at 37 ºC for 30 min. $Post_2$ was purified via 1.1M sucrose cushion.

Formation of $Post_3$ ribosome complex. Three different tRNA solutions were prepared: A-Tyr, A-Val, and A-Ser. All three solutions contained 100 mM Tris (pH 7.8), 20 mM $MgAc_2$, 1 mM EDTA, 4 mM ATP, 0.1 mg/ml total synthetase, 50 A_{260}/ml total tRNA, and 0.25 mM each of phenylalanine, glutamic acid, and lysine. The A-Tyr, A-Val, and A-ser also contained 0.25 mM tyrosine, valine and serine, respectively. Each one of these tRNA solutions was incubated with the ribosome mix and TuWG mix in the ratio of 2:1:2, and incubated at 37 ºC for 30 min. The resulting $Post_3$ complexes were purified by 1.1M sucrose cushion.

Formation of Pre_NS_1 and Post_NS_1 ribosome complexes. These complexes were prepared similarly as Pre_1 and $Post_1$, except the A-Lys mix is replaced by A-Ser mix to incorporate serine instead of lysine at the non-slippery codon.

Formation of Post_CGC_{1-2} ribosome complexes. These complexes were prepared in one step from the initiation complex in the presence of phenylalanine, glutamic acid, lysine, and without and with arginine for Post_CGC_1 and Post_CGC_2, respectively.

Formation of Post_GGA_1 and Post_SLP_1 ribosome complexes. These complexes were similarly as $Post_1$, which were generated by translocation of their corresponding Pre_1 complexes.

Formation of Post_GGA_{2-3} and Post_SLP_3 ribosome complexes. These complexes were prepared in one pot from the initiation complex, but were stopped at the corresponding codons. One tRNA solution was prepared containing: 100 mM Tris (pH 7.8), 20 mM $MgAc_2$, 1 mM EDTA, 4 mM ATP, 0.1 mg/ml total synthetase, 50 A_{260}/ml total tRNA, and 0.25 mM each of phenylalanine, glutamic acid (or glycine), lysine, and tyrosine or valine. The tRNA solution was

incubated with the ribosome mix and TuWG mix in the ratio of 2:1:2, and incubated at 37 ºC for 30 min. The resulting Post-complexes were purified by 1.1M sucrose cushion.

FIRMS measurements. The surface area of the sample well was 2×6 mm^2, coated with biotin then incubated with streptavidin. The ribosome complexes were immobilized on the surface via the 5′-end biotin on the mRNA. The probing DNAs were incubated with streptavidin-coated magnetic beads (M280, Invitrogen). For each FIRMS experiment, the probing DNA labelled with a magnetic bead is hybridized to the 3' side of the ribosome-mRNA complex. Magnetic signal of the sample was measured by an atomic magnetometer as a function of mechanical force. The force was provided by a centrifuge (5417R from Eppendorf), with the speed increasing by 100 rpm (revolution per minute) per step. The dissociation of the DNA-mRNA duplexes was indicated by a decrease in the magnetic signal, which occurred when the centrifugal force reached the dissociation force of the duplex. This is because the dissociated magnetic beads were removed from the sample. The typical force range in this work was 82 pN, after which the residual magnetic signal was taken as the background. FIRMS profiles were obtained by normalizing the overall magnetic signal decrease (B_0) to be 100% and then plotting the relative magnetic signal decrease (B/B_0) vs. the external force. The force values were calculated according to $m\omega^2 r$, in which m is the buoyant mass of M280 magnetic beads (4.6×10^{-15} kg). ω is the centrifugal speed, and r is the distance of the magnetic beads from the rotor axis (8 cm for 5417R). The typical force resolution was 3-4 pN in this work. The samples were at room temperature during approximately 2-3 hrs total measuring time. Each profile reported in this work was repeated at least three times to assure reproducibility.

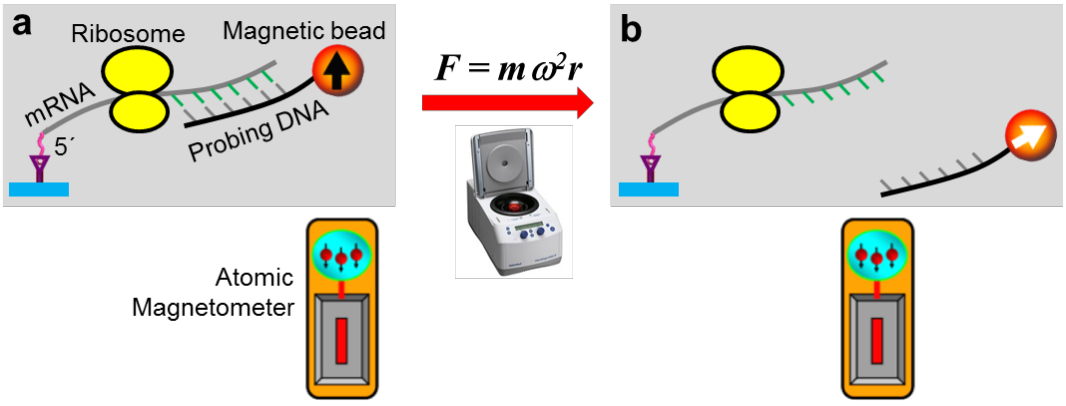

Figure S1. Scheme of the FIRMS method for probing frameshifting. **a**. The ribosome complexes are immobilized on the streptavidin-coated surface via the 5´-end biotin on the mRNA. The probing DNAs are labelled with magnetic beads. The uncovered mRNA on the ribosome

complexes forms duplexes of certain basepairs with the probing DNA. The number of bp of the duplex will thus reveal the position of the ribosome on the mRNA. The magnetic signal is measured by an atomic magnetometer. **b.** A mechanical force is applied to the sample by using a centrifuge. The sample magnetic signal is measured again after applying the force. If the force reaches the dissociation force of the DNA-mRNA duplex, a sharp decrease in the magnetic signal will be observed, because the dissociated DNA will be removed from the sample.

Probing scheme in Figure 1c. The dissociation forces of DNA-mRNA duplexes will initially increase for longer DNAs; then it will remain constant when the probes reach the ribosome front because the mRNA has no more available nt to pair with the DNA. For example, P13 (13 complementary nt with the mRNA) will form 13-bp duplexes with both $Post_1$(-1) and $Post_1$(-2), while P14 (14 complementary nt with the mRNA) will form 13-bp duplex with $Post_1$(-1) but 14-bp duplex with $Post_1$(-2). Therefore, the appearance of a higher dissociation force from using P13 to P14 will unambiguously reveal the presence of $Post_1$(-2). Similarly, all ribosomal positions can be precisely determined.

Toe-printing assay. The mRNAs for toe-printings were transcribed and purified *in vitro* using the HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB). The beginning codons were: "ATG-TTT-GAA-AAA-AAG" for GA_7G and SLP mRNAs (Figure 3), but the SLP mRNA contained the dnaX stem loop. The "AAA" codon was replaced with "AGT" in AGU mRNA. The ribosome complexes were prepared similarly as the $POST_3$ complexes, in the presence of only amino acids F, E, and K. For AGU mRNA, amino acid S was included in addition to the other amino acids. The 5'-terminal of the 20-nt-long primer was labeled with Cy5 dye (described in the following) and was 62-nt downstream from the first nucleotide of the "AAG" codon in the frameshifting motif. The ribosome complexes (10 pmol) were annealed with the primer (20 pmol) in 1x reaction buffer of the AMV reverse transcriptase from NEB under the following condition: (rt, 5min)-(37 ºC, 5 min)-(ice, 5min). Then dNTP mixture (final concentration 0.5 mM), RNaseOUT (ThermoFisher, 60 units), DTT (final concentration 5 mM), and AMV reverse transcriptase (20 Units) were added. The mRNA extension reaction (total volume 20 μL) was continued under the following condition: (rt, 5 min)-(37 ºC, 30 min). The reverse transcriptase was deactivated by heating at 95 ºC for 10 min in the presence of 50 mM NaOH. Then the NaOH was neutralized with HCl, and the total volume was increased to 200 μL with 0.3M NaAc (pH 5.2). The solution was then extracted with equal volume phenol/chloroform, and precipitated with 100% ethanol. The pellets were collected by spin at 20K x g at 4 ºC for 30

min, washed with 70% ethanol 2 times, and resuspended with 10 μL loading solution. The cDNAs were separated on a 40 cm X 20 cm X 0.4 mm 15%-acrylamide gel contained 8 M of urea (40-45 W, 3-4 hours). The Gels were immediately scanned on the Strom 860 scanner (Molecular Dynamics) under high sensitivity setting.

5′-labeling of the primer. The unlabeled primer was ordered from IDTDNA. 0.6 nmol of the primer was mixed with T4 polynucleotide kinase and ATPγS according to the 5'-EndTag (Vector Laboratories) protocol. The reaction was continued at 37 ºC for 2 hrs. Then 50 μg of Cy5 Maleimide Mono-Reactive Dye (GE Healthcare) was dissolved in 5 μL of DMSO and was added in the primer reaction mix. The reaction was incubated at 30 ºC for 2-4 hours, followed by the precipitation procedure of the manufacture's protocol. The labeled primer was resuspended with small amount of nuclease-free water and purified from free dye via a size-exclusive column with Sephadex G-25 fine medium (GE Healthcare).

Recombinant protein expression with plasmid fused with GA_7G motif. The construct II containing the sequence "ATG TTT GAA AAA AAG" + "CG" + "CTC-N_{84}-ACA-N_{15}-TTA-AAA-N_{15}-CAC-N_{15}-ACT-AAA-N_{45}-$(CAT)_6$-TAA" was ordered from IDTDNA and sub-cloned into the pet 20b(+) vector between the NdeI and XhoI sites (Stratagene). The constructs I and III were mutated from construct II via single and double nucleotides deletion at the variation segment, respectively (QuickChange Lightning kit from Agilent Technology). These plasmids were transformed into the BL21-star cells (Life Technologies) and cultured. The cells were induced by 0.1 mM TPIG for 1 hour. Then 4x of the cell weight B-PER Complete protein extraction reagent (life technology) was added to lyse the cells. Then 10 ml of the clarified cell lysates were loaded on HisTrap HP 5 ml column (GE Healthcare Life Sciences). The His-tagged proteins were eluted via imidazole gradient and concentrated with Amicon Ultra filter (EMD Millipore). Typically, 500-1000 pmol of the 8.5 kDa protein was isolated (Figure S5). There was no significant yield difference comparing the different constructs, implying the frameshifting yield was high. However, the efficiencies could not be precisely determined because the proteins in the other reading frames were not detected on the gel.

To express the mCherry protein, the sequence with His_6 tag was ordered from IDTDNA and inserted in the "-2" reading frame after the slippery site to replace the codons in construct II. All the other operations were the same except the induction time was 6 hours instead of 1 hour.

In vitro mRNA translation with PURExpress kit. The DNAs containing the sequence "T7 promoter" + "ATG TTT GAA AAA AAG" + "XX" + "mCherry " ("XX" are the variations in Table S1) were ordered from IDTDNA and transcribed into mRNA with the HiScribe T7 Quick High

Yield RNA Synthesis Kit (NEB). The mRNAs were purified with PureLink RNA Mini Kit (Ambion). The qualities of the mRNAs were confirmed by RNA Gels and 5 µg of the mRNA were used for one protein synthesis reaction with the PURExpress in vitro Protein Synthesis Kit (NEB). The mCherry incubation solutions were directly read with the Spectra-Max M5 plate-reader, with excitation/emission wavelengths at 584nm/612nm, respectively. Because the incubation conditions were exactly the same, the frameshifting efficiencies were compared directly from the fluorescence reading after subtracting with the background.

Table S1. Design of the recombinant protein constructs.

	DNA Sequence			Protein Sequence[c]	Frame-shifting	MW (kDa)
	5´-fragment	variation[a]	3´-fragment[b]			
I	ATG-TTT-GAA-AAA-AAG	_ _	CTC-N_{84}-ACA-N_{15}-TTA-AAA-N_{15}-CAC-N_{15}-ACT-AAA-N_{45}-$(CAT)_6$-TAA	MFEKK_LSDG	0	8.5
II		CG		MFEKKALSDG	-1	8.6
III		C_		MFEKKSLSDG	-2	8.6

a. "–" indicates null.

b. Red sections indicating the His_6 tags in "-2", "-1", and "0" frame, respectively.

c. The protein sequence refers to the peptide translated from the AUG codon at the beginning of the mRNA.

Table S2. The first seven amino acid sequences in the three ORFs of the three constructs. ORFs: open reading frames.

MW (kDa)	**Theoretical parameter**		**Construct I**		**Construct II**		**Construct III**	
	Solubility[a]	Instability[b]	Sequence[c]	Frame	Sequence	Frame	Sequence	Frame
8.5	Yes	30	MFEKK LS	0	MFEKK AL	-1	MFEKK SL	-2
6.5	Yes	15.2	MFEKK AQ	-1	MFEKK SA	-2	MFEKK PQ	0
4.5	Yes	14.6	MFEKK SS	-2	MFEKK RS	0	MFEKK AS	-1

a. Solubility is calculated by the PROSOII program at http://mips.helmholtz-muenchen.de/proso/ (Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D. 2007. Protein solubility: sequence based prediction and experimental verification. *Bioinformatics* **23:** 2536-2542).

b. Instability is calculated by the ProtParam tool at http://web.expasy.org/. Values less than 40 are considered stable (Guruprasad K, Reddy BV, Pandit MW. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering* **4:** 155-161).

c. Highlighted shows the different sequences in the different reading frames.

Table S3. The mRNA sequences and fluorescence readings of the mRNA-based translation with cell-free PURExpress kit.

Complex	**5´-Fragment[a]**	**Variation[b]**	**3´-Fragment**	**Fluorescence (a.u. x 10^4)**
I	ATG-TTT-GAA-AAA-AAG	– –	mCherry sequence	1.46 ± 0.012
II		CG		1.49 ± 0.014
III		C –		1.41 ± 0.008
IV	ATG-TTT-GAA-AGT-AAG	– –		1.51 ± 0.015
V	No mRNA			0.85 ± 0.009

a. Complexes I-III contain the same fragment with the GA_7G motif.

b. "–" indicates null.

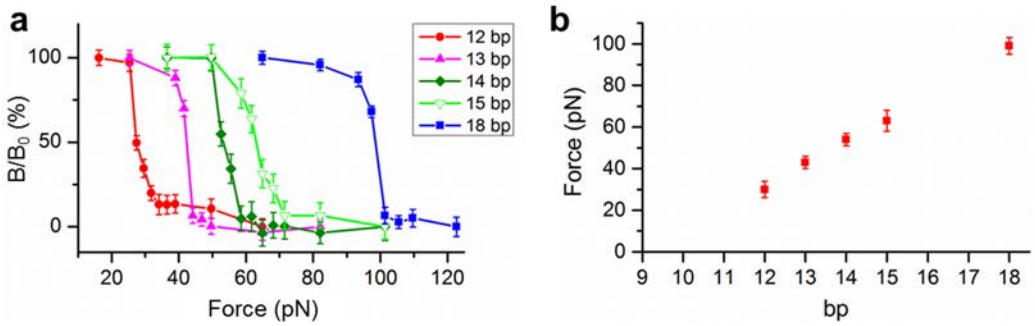

Figure S2. Force calibration for the DNA-mRNA duplexes. **a.** FIRMS profiles of the various duplexes formed between the mRNA containing GA_7G and the probing DNAs that have different numbers of complementary bases to the mRNA. The magnetic signal was normalized to the overall magnetic field decrease B_0, which was the magnetic signal difference between the initial magnetic signal and the final magnetic signal at the maximum force in each case. B/B_0: relative magnetic signal; bp: basepair. **b.** The correlation between the dissociation force and the number of bp. The plot shows that, on average, the dissociation force increases by approximately 11 pN per bp, which is consistent with the literature (Rief M, Clausen-Schaumann H, Gaub HE. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat Struct Biol* **6**: 346-349). The trend is also consistent with the literature (Sattin BD, Pelling AE, Goh MC. 2004. DNA base pair resolution by single molecule force spectroscopy. *Nucl Acids Res* **32**: 4876-4883), which showed near-linear trend to at least 80 pN when the number of bp increases for DNA duplexes. The dissociation force of 30 pN for the 12-bp duplex is consistent with multiple reports including our own previous work (Pope LH, Davies MC, Laughton CA, Roberts CJ, Tendler SJB, Williams PM. 2001. Force-induced melting of a short DNA double helix. *Eur Biophys J* **30**: 53-62. De Silva L, Yao L, Wang, Y, Xu S. 2013. Well-defined and sequence-specific noncovalent binding forces of DNA. *J Phys Chem* **117**: 7554-7558).

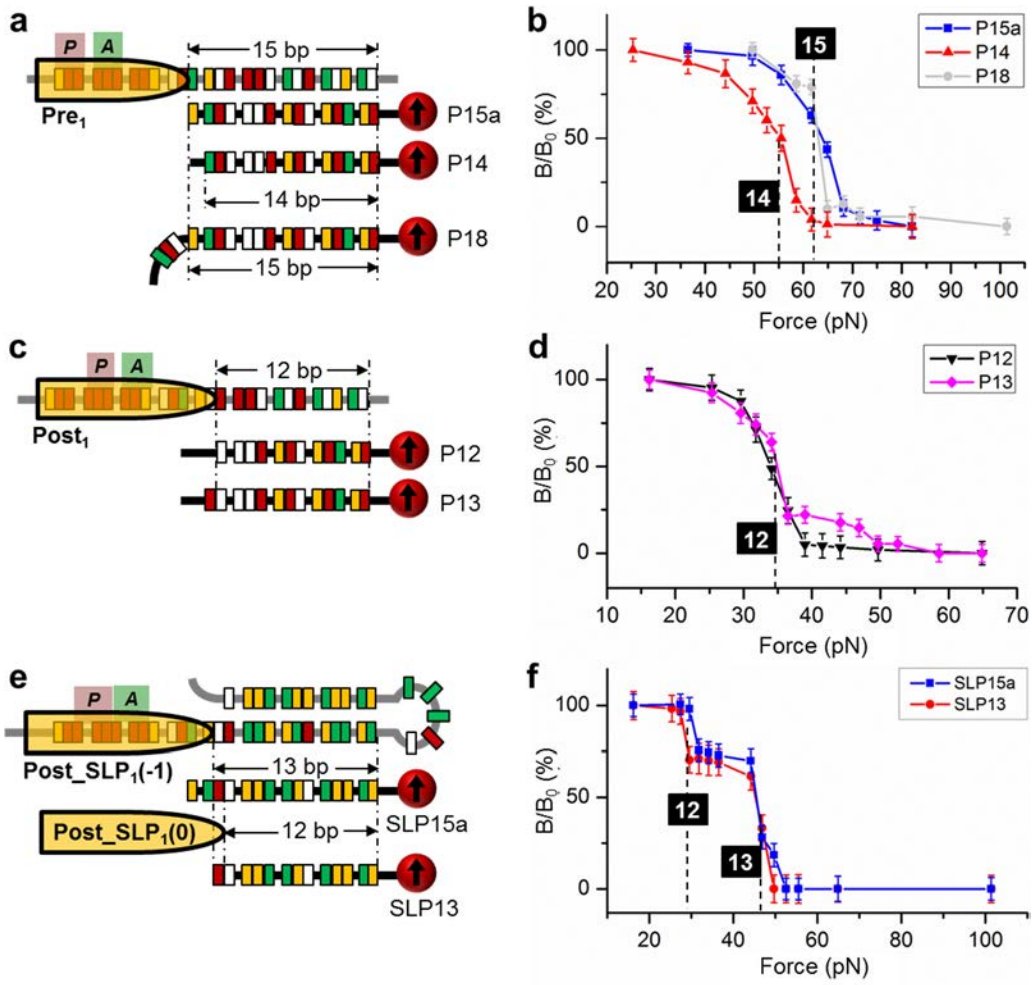

Figure S3. Use of multiple DNA probes to precisely determine the ribosome reading frame. **a, b.** Scheme and results of using DNAs P14, P15a, and P18 to verify the position of ribosome in the Pre_1 complex. The dissociation force increase from P14 to P15a indicates the minimum of 15-bp duplexes formed between P15a and Pre_1. The constant dissociation force from P15a to P18 indicates the maximum of 15-bp duplexes. Therefore the position of ribosome in Pre_1 is precisely determined to have 15 complementary bases with P15a. **c, d.** Scheme and results of using P12 and P13 to verify the position of ribosome in the $Post_1$ complex promoted by EF-G•GDPCP. The overlap of the main feature for P13 and the feature for P12 indicates this dissociation force corresponds to 12-bp DNA-mRNA duplexes, confirming the main product to be the "0" reading frame. **e, f.** Scheme and results of using SLP15a and SLP13 probes for the $Post\_SLP_1$ complex. The overlapping profiles indicate the higher binding force belonged to the 13-bp duplexes. The lower binding force feature thus corresponds to the 12-bp duplex, i.e., the "0" reading frame in $Post\_SLP_1$. Therefore, both $Post\_SLP_1$(-1) and $Post\_SLP_1$(0) were present, with the former being the main product at approximately 63%.

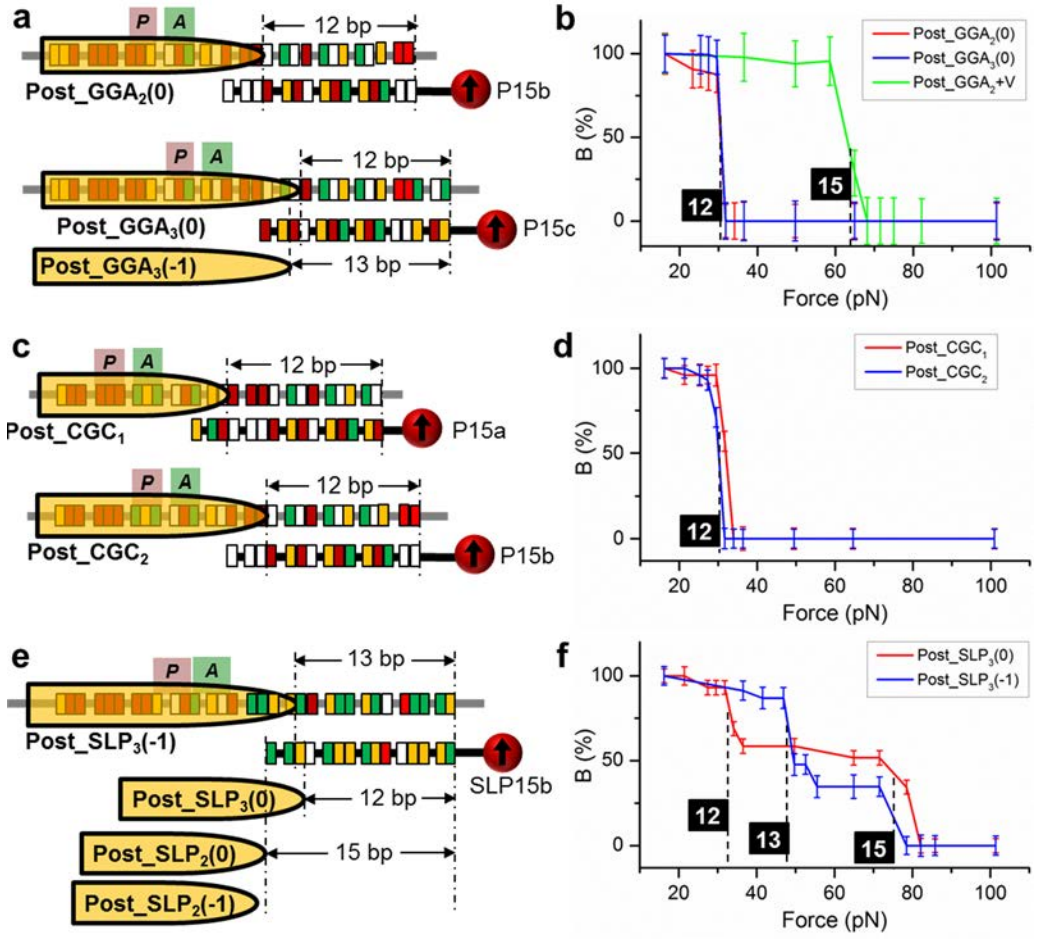

Figure S4. Control experiments to confirm the roles of the slippery sequence and the stem loop. **a, b.** Scheme and results for probing the second and third translocation products when GGA replaced GAA in the GA_7G motif. The appearance of only 12-bp dissociation force in the red and blue traces indicated only the "0" frame product in Post_GGA_2 and Post_GGA_3, respectively. When the downstream aminoacyl tRNA for the "-1" frame translation, no Post_GGA_3(-1) was formed, because no 13-bp duplex was detected. The 15-bp corresponds to the unreacted Post_GGA_2(0). **c, d.** Scheme and results for probing the translocation products when CGC replaced the AAG in the GA_7G motif. The appearance of only 12-bp dissociation force in both Post_CGC_1 and Post_CGC_2 showed that only the "0" frame products existed in the first two translocation steps. **e, f.** Scheme and results for the GA_7G motif coupled with the stem loop. The Post_SLP_3(0) trace showed ~40% 12-bp duplex and 60% 15-bp duplex. The former was the desired "0" frame product. The latter was the unreacted Post_SLP_2(-1). This is because only the amino acyl tRNA corresponding to the "0" frame was present. Similarly, the Post_SLP_3(-1) contained ~60% the desired 12-bp duplex, and the remaining 40% unreacted Post_SLP_2(0). The "-1" frameshifting percentage was similar to that in Post-SLP_1 (Figure 2f).

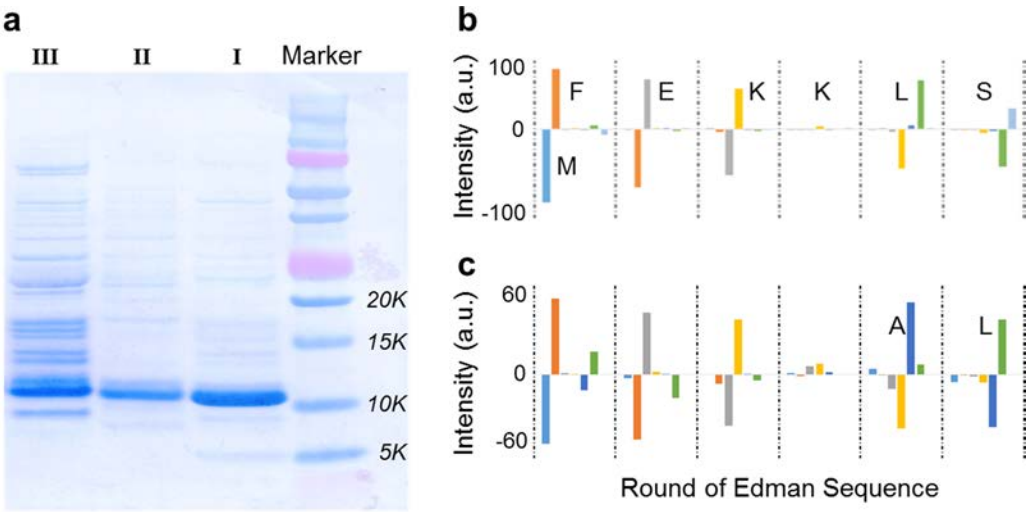

Figure S5. In cell translational frameshifting verified by sequencing. **a**. The SDS-PAGE image of the recombinant proteins (all the expected sequences shown in Table S1). The 8.5 kDa proteins migrated near the 10K marker. **b.** Edman sequencing of the 8.5 kDa protein in construct I. **c.** Edman sequencing of the 8.5 kDa protein in construct II. The signal is the differential value of the amino acid intensities of the current cycle minus the previous cycle; therefore for the two consecutive lysine residues, the second signal was near zero.

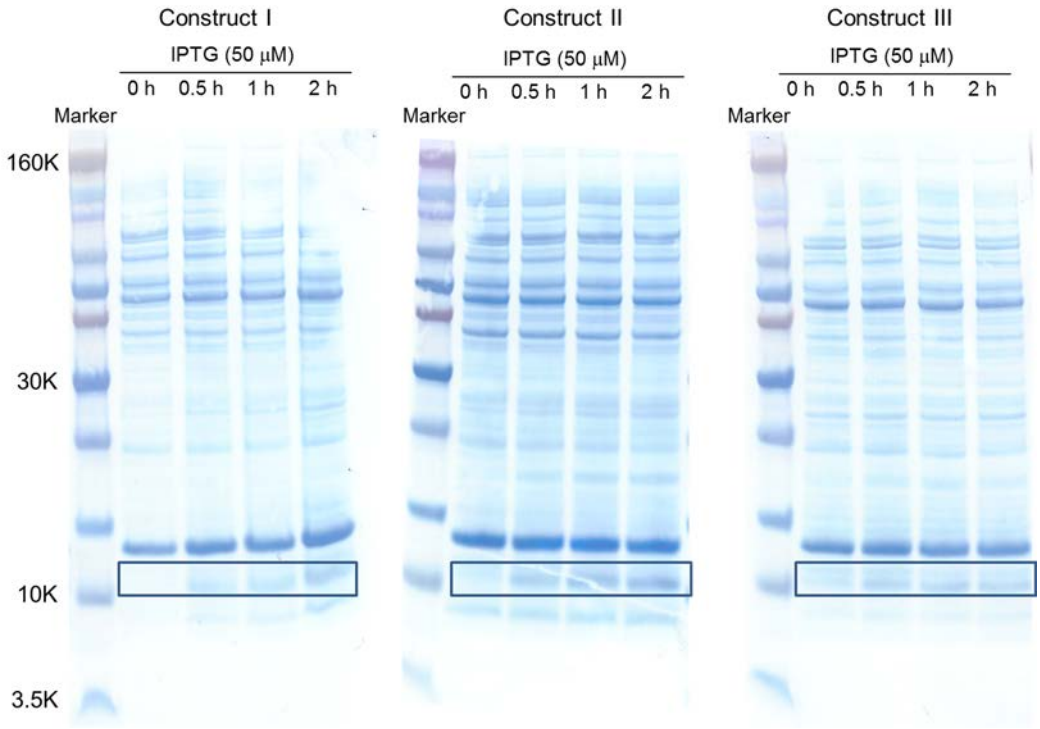

Figure S6. The SDS-PAGE images of IPTG induced protein expression in constructs I-III. The boxes framed the protein of interested in crude cell lysates. The purified proteins were displayed in Figure 5 of the text. For constructs I and II, the proteins were expressed more with longer time. For construct III, the similar protein band was not responding to the IPTG induction, which agreed with the N-terminal sequencing result that the targeted protein was not expressed successfully.

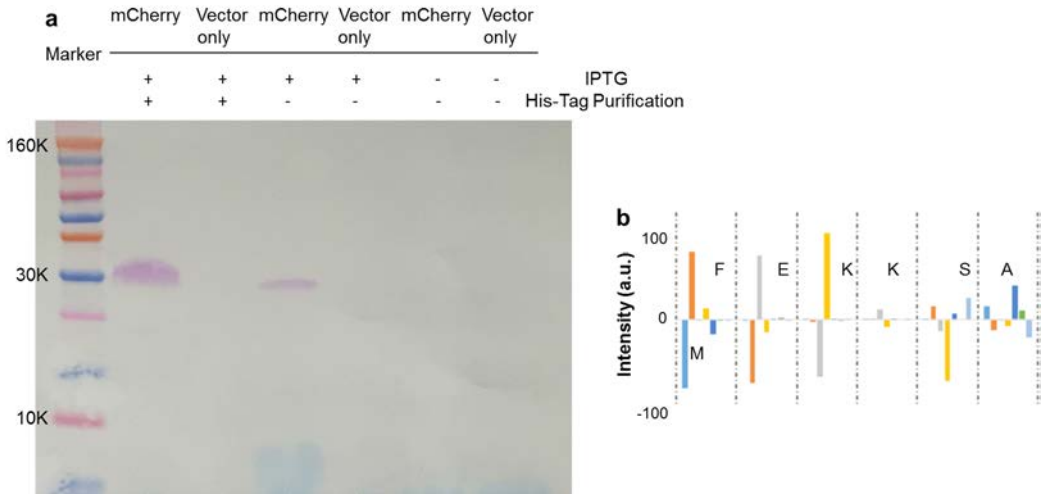

Figure S7. Observation of the mCherry fluorescent protein expression in the "-2" reading frame. **a.** The SDS-PAGE image of the recombinant expression of mCherry protein after the FPLC purification. The ~30 kDa protein was visible without staining. Lanes 2, 4 and 6 showed the mCherry expression was dependent on IPTG induction; and lanes 3, 5 and 7 showed no protein expression with the empty vector. **b.** Edman sequencing of the mCherry protein, which shows that the peptide in the "-2" frame was translated.

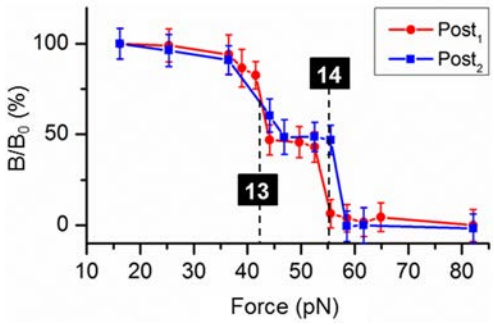

Figure S8. Overlaying FIRMS profiles of $Post_1$ probed by P15a and $Post_2$ probed by P15b. The overlapping profiles showed that the reading frame distribution was probably preserved in the second translocation step. In other words, normal translocation proceeded in the second translocation step, in contrast to the "-1" and "-2" frameshiftings occurred in the first translocation step.

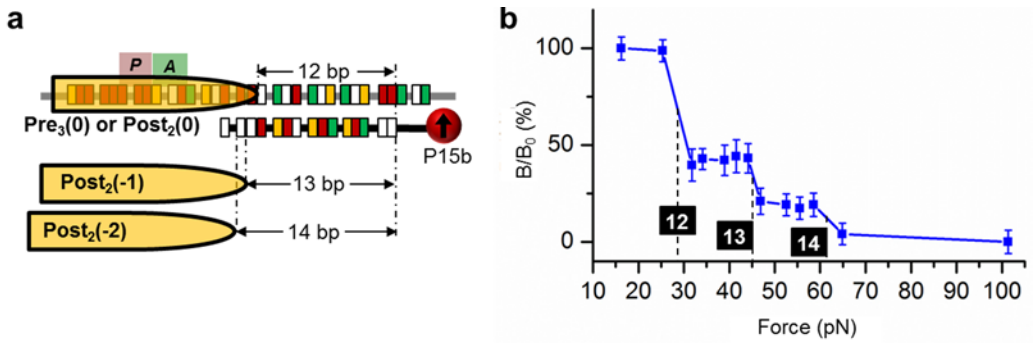

Figure S9. Formation of the "0" frame ribosome complex prior to $Post_3(0)$. **a.** Detection scheme. Probing DNA P15b was used to distinguish the different reading frames following the second translocation step. The "0" reading frame, in the form of either $Post_2(0)$ or $Pre_3(0)$, will be shown as 12-bp duplex. **b.** FIRMS profile. The main product is 12-bp duplexes, indicating the presence of the "0" reading frame.