An RNA modification prevents extended codon-anticodon interactions from facilitating +1 frameshifting

Evelyn M. Kimbrough[1], Ha An Nguyen[1], Haixing Li[2,3], Jacob M. Mattingly[1,4], Nevette A. Bailey[2], Wei Ning[2], Howard Gamper[5], Ya-Ming Hou[5], Ruben L. Gonzalez, Jr.[2,*] and Christine M. Dunham[1,*]

[1]Department of Chemistry, Emory University, Atlanta, GA USA
[2]Department of Chemistry, Columbia University, New York, NY, USA
[3]Current address: Department of Physics, City University of Hong Kong, Kowloon, Hong Kong China
[4]Graduate Program in Biochemistry, Cell and Developmental Biology, Emory University, Atlanta, GA USA
[5]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA, USA

*Co-corresponding authors: Ruben L. Gonzalez, Jr., rlg2118@columbia.edu; Christine M. Dunham, christine.m.dunham@emory.edu.

Key words: ribosome, frameshift, near cognate, mRNAs, tRNAs, fidelity, translation

This file includes:
Supplementary Tables S1-S5
Supplementary Figures S1-S9

SUPPLEMENTAL TABLES

Supplemental Table 1: Fractional populations, rate constants, and transition rates characterizing the GS1⇄GS2 transitions measured by smFRET. Data from ribosome complexes containing non-slippery CCC-G proline codons and slippery CCC-C proline codons (that cause a +1 frameshift) with native $tRNA^{ProL}$, unmodified $tRNA^{ProL}$, unmodified $tRNA^{ProL}$ + m^1G37, or native $tRNA^{ProL}$ – m^1G37 in the P site.

Non-slippery CCC-G proline codon

P-site tRNA	GS1 (%)	GS2 (%)	K_{eq}	$k_{GS1\rightarrow GS2}$ (s^{-1})	$k_{GS2\rightarrow GS1}$ (s^{-1})
Native	51 ± 3	49 ± 3	0.97 ± 0.11	0.94 ± 0.17	0.96 ± 0.07
Unmodified	58 ± 2	42 ± 2	0.74 ± 0.05	0.67 ± 0.07	1.41 ± 0.23
Unmodified + m^1G37	56 ± 3	44 ± 3	0.78 ± 0.08	0.66 ± 0.06	1.28 ± 0.12
Native - m^1G37	56 ± 1	44 ± 1	0.78 ± 0.01	0.68 ± 0.14	1.08 ± 0.10

Slippery CCC-C proline codon

P-site tRNA	GS1 (%)	GS2 (%)	K_{eq}	$k_{GS1\rightarrow GS2}$ (s^{-1})	$k_{GS2\rightarrow GS1}$ (s^{-1})
Native	47 ± 5 [a]	53 ± 5	1.15 ± 0.23	1.19 ± 0.16	0.70 ± 0.18
Unmodified	29 ± 4	71 ± 4	2.52 ± 0.45	1.36 ± 0.24	0.32 ± 0.04
Unmodified + m^1G37	52 ± 2	48 ± 2	0.91 ± 0.09	0.88 ± 0.13	1.06 ± 0.03
Native - m^1G37	35 ± 2	65 ± 2	1.68 ± 0.16	1.21 ± 0.39	0.46 ± 0.16

[a] Mean ± s.d. of fractional population, equilibrium constants, and transition rates were determined from three independent data sets.

Supplemental Table 2: Cryo-EM data collection and model statistics for the 70S + unmodified P-tRNA^{ProL} + A-tRNA^{Val} dataset.

EMDB accession	**EMD-42495**		**EMD-42721**	
PDB ID	**8URM**		**8UXB**	
Name	**P-tRNA^{ProL} with A-tRNA^{Val}**		**P/E-tRNA^{ProL}**	
Data collection				
Microscope	Titan Krios			
Detector	Gatan K3			
Voltage (keV)	300			
Electron exposure ($e^-/Å^2$)	61.23			
Pixel size (Å)	1.069			
Defocus range (μm)	0.6-2.7			
frames per movie	50			
Micrographs (#)	10,949			
Initial particles (#)	2,314,417			
Final particles (#)	173,856		361,087	
Model refinement and validation statistics				
Composition (#)				
Atoms	148,566		145,221	
Residues	Protein: 5,715 Nucleotide: 4,727		Protein: 5,611 Nucleotide: 4,645	
Water	1,981		1,449	
Ligands	MG: 395		MG: 287	
Bonds (RMSD)				
Length (Å) (# > 4σ)	0.003 (5)		0.003 (9)	
Angles (°) (# > 4σ)	0.517 (121)		0.560 (58)	
MolProbity score	1.87		1.93	
Clash score	5.81		5.28	
Ramachandran plot (%)				
Outliers	0.00		0.33	
Allowed	3.79		7.22	
Favored	96.21		92.45	
Rotamer outliers (%)	2.44		1.73	
Cβ outliers (%)	0.00		0.02	
Peptide plane (%)				
Cis proline/general	1.1/0.0		0.0/0.0	
Twisted proline/general	0.0/0.0		0.0/0.0	
CaBLAM outliers (%)	2.08		4.89	
ADP (B-factors)	min/max/mean		min/max/mean	
Protein	9.53/124.59/43.32		100.02/457.94/166.84	
Nucleotide	6.74/217.41/53.01		99.88/673.91/162.94	
Ligand	6.42/64.39/27.62		88.16/251.69/121.83	
Water	5.29/123.08/29.11		98.43/657.06/153.54	
Resolution Estimates (Å)	Masked	Unmasked	Masked	Unmasked
d FSC (half maps; 0.143)	2.9	3.1	2.9	3.0
d 99 (full/half1/half2)	4.3/2.4/2.4	4.2/2.2/2.2	4.3/3.0/3.0	4.2/2.3/2.3
d model	2.3	2.4	2.6	2.4
d FSC model (0/0.143/0.5)	2.3/2.6/3.0	2.3/2.8/3.1	---/2.6/2.9	2.4/2.8/3.1
Map min/max/mean	-0.00/0.05/0.00		-0.00/0.05/0.00	
Model vs. Data				
CC (mask)	0.85		0.86	

Supplemental Table 3: Cryo-EM data collection and model statistics for the 70S-aa + unmodified tRNA^{ProL} dataset (normal and +1 frame classes).

EMDB accession	**EMD-42541**		**EMD-42714**	
PDB ID	**8UTJ**		**8UX8**	
Name	**aa-tRNA^{ProL} (normal frame)**		**aa-tRNA^{ProL} (+1 frame)**	
Data collection				
Microscope	Titan Krios			
Detector	Gatan K3			
Voltage (keV)	300			
Electron exposure ($e^-/Å^2$)	56.07			
Pixel size (Å)	1.069			
Defocus range (μm)	-0.6-2.5			
frames per movie	50			
Micrographs (#)	4,467			
Initial particles (#)	1,081,450			
Final particles (#)	81,573		69,316	
Model refinement and validation statistics				
Composition (#)				
Atoms	146,138		146,547	
Residues	Protein: 5,777 Nucleotide: 4,642		Protein: 5,775 Nucleotide: 4,645	
Water	770		1,267	
Ligands	MG: 530		MG: 395	
Bonds (RMSD)				
Length (Å) (# > 4σ)	0.000 (0)		0.003 (10)	
Angles (°) (# > 4σ)	0.498 (46)		0.648 (11)	
MolProbity score	1.76		1.82	
Clash score	7.88		11.47	
Ramachandran plot (%)				
Outliers	0.00		0.02	
Allowed	4.72		3.65	
Favored	95.28		96.33	
Rotamer outliers (%)	7.56		0.15	
Cβ outliers (%)	0.00		0.00	
Peptide plane (%)				
Cis proline/general	1.1/0.0		1.1/0.0	
Twisted proline/general	0.0/0.0		0.0/0.0	
CaBLAM outliers (%)	2.10		1.90	
ADP (B-factors)	min/max/mean		min/max/mean	
Protein	88.89/470.97/165.43		20.94/188.80/60.02	
Nucleotide	89.73/744.38/176.47		28.29/324.14/69.12	
Ligand	80.65/355.47/134.43		3.45/96.09/42.25	
Water	83.22/659.80/178.17		14.28/227.08/62.08	
Resolution Estimates (Å)	Masked	Unmasked	Masked	Unmasked
d FSC (half maps; 0.143)	3.5	3.8	3.7	4.1
d 99 (full/half1/half2)	5.1/2.3/2.3	4.8/2.2/2.2	5.4/2.2/2.2	5.1/2.2/2.2
d model	2.5	2.7	2.6	2.9
d FSC model (0/0.143/0.5)	2.5/2.9/3.4	2.7/3.1/3.7	2.6/3.1/3.8	2.8/3.2/4.2
Map min/max/mean	-0.01/0.06/0.00		-0.01/0.06/0.00	
Model vs. Data				
CC (mask)	0.85		0.82	

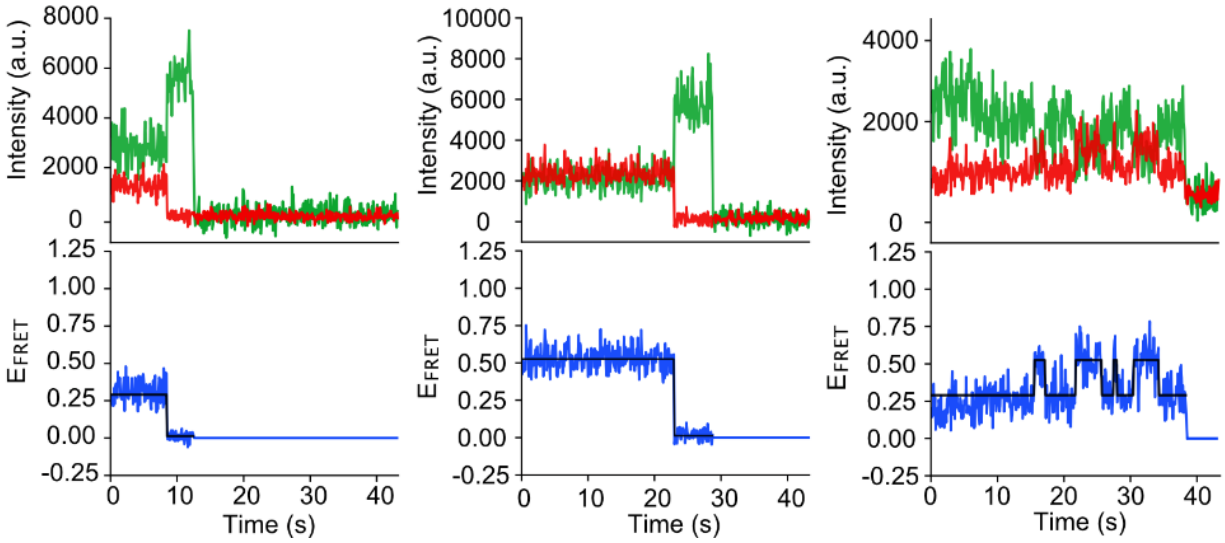
Supplemental Table 4: Cryo-EM data collection and model statistics for the 70S-aa + unmodified tRNA^{ProL} dataset (P/E and e*/E-tRNA^{ProL} classes).

EMDB accession	**EMD-42840**		**EMD-42852**	
PDB ID	**8UZG**		**8V03**	
Name	**e*/E-tRNA^{ProL}**		**P/E-tRNA^{ProL}**	
Data collection				
Microscope	Titan Krios			
Detector	Gatan K3			
Voltage (keV)	300			
Electron exposure (e$^-$/Å^2)	56.07			
Pixel size (Å)	1.069			
Defocus range (μm)	-0.6-2.5			
frames per movie	50			
Micrographs (#)	4,467			
Initial particles (#)	1,081,450			
Final particles (#)	59,169		38,062	
Model refinement and validation statistics				
Composition (#)				
Atoms	144,201		144,507	
Residues	Protein: 5,526 Nucleotide: 4,642		Protein: 5,611 Nucleotide: 4,632	
Water	959		1,008	
Ligands	MG: 398		MG: 281	
Bonds (RMSD)				
Length (Å) (# > 4σ)	0.004 (0)		0.002 (0)	
Angles (°) (# > 4σ)	0.742 (67)		0.511 (85)	
MolProbity score	2.25		2.01	
Clash score	10.66		9.27	
Ramachandran plot (%)				
Outliers	0.48		0.51	
Allowed	10.13		8.91	
Favored	89.39		90.98	
Rotamer outliers (%)	1.53		0.00	
Cβ outliers (%)	0.18		0.00	
Peptide plane (%)				
Cis proline/general	0.0/0.0		0.0/0.0	
Twisted proline/general	0.0/0.0		0.0/0.0	
CaBLAM outliers (%)	5.59		4.93	
ADP (B-factors)	(min/max/mean)		(min/max/mean)	
Protein	112.32/427.55/193.37		34.31/323.50/91.27	
Nucleotide	108.97/648.37/193.92		33.41/438.34/90.57	
Ligand	77.02/329.00/145.11		26.45/134.35/60.08	
Water	98.52/642.00/220.98		10.22/452.34/122.88	
Resolution Estimates (Å)	Masked	Unmasked	Masked	Unmasked
d FSC (half maps; 0.143)	3.8	4.3	4.1	4.7
d 99 (full/half1/half2)	5.4/2.2/2.2	4.9/2.2/2.2	5.3/2.2/2.2	4.7/2.2/2.2
d model	2.8	2.8	2.7	3.0
d FSC model (0/0.143/0.5)	2.7/3.1/3.9	2.8/3.3/4.3	2.7/3.2/4.1	2.9/3.5/4.8
Map min/max/mean	-0.02/0.11/0.00		-0.04/0.16/-0.00	
Model vs. Data				
CC (mask)	0.85		0.70	

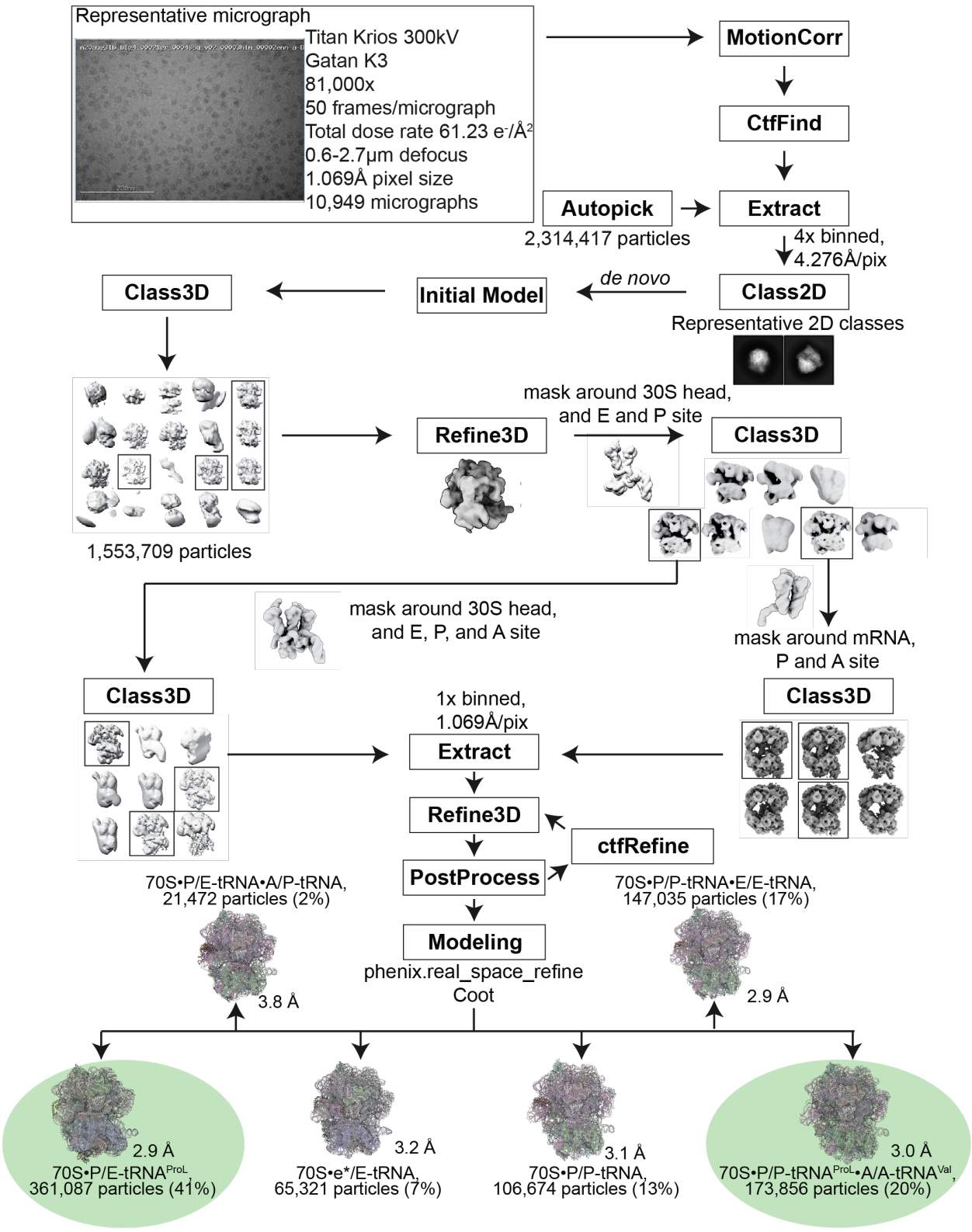
Supplemental Table 5: Oligonucleotides used in this study (5'-3'). The AUG start codon is shown in green, the proline and slippery codons are underlined, and the anticodon is bolded.

Oligonucleotides	
mRNA (fMet-Pro): GGGCCCUAAGGACAUAAAAAUGCCCCGUUAUCCUCCUGCUGC	IDT
$tRNA^{ProL}$: CGGCACGUAGUAGCGCAGCCUGGUAGCGCACCGUCAU**GGG**G UGUCGGGGGUCGGAGGUUCAAAUCCUCUCGUGCCGACCA	*in vitro* transcribed
mRNA (smFRET1_nonslippery): Biotin-GCAACCUAAAACUUACACAGGGGGGAAGGAGGUAAA AAUGCCCGUUCUAAGCACCACCACCACCACCACCAC	IDT
mRNA (smFRET1_slippery): Biotin-GCAACCUAAAACUUACAC AGGGGGGAAGGAGGUAAAAAUGCCCCGUUCUAAGCA CCA CCACCACCACCACCAC	IDT
mRNA (smFRET2): GCAACCUAAAACUCACACAGGGCCCUA AGGACAUAAAAAUGCCCCGUUAUCCUCCUGCUGCACUCG CUGCACAAAUCGCUCAACGGCAAUUAAGGA	*in vitro* transcribed
mRNA (smFRET2) probe (DNA): TGTGTAAGTTTTAGGTTG ATT TG-Biotin	IDT

SUPPLEMENTAL FIGURES

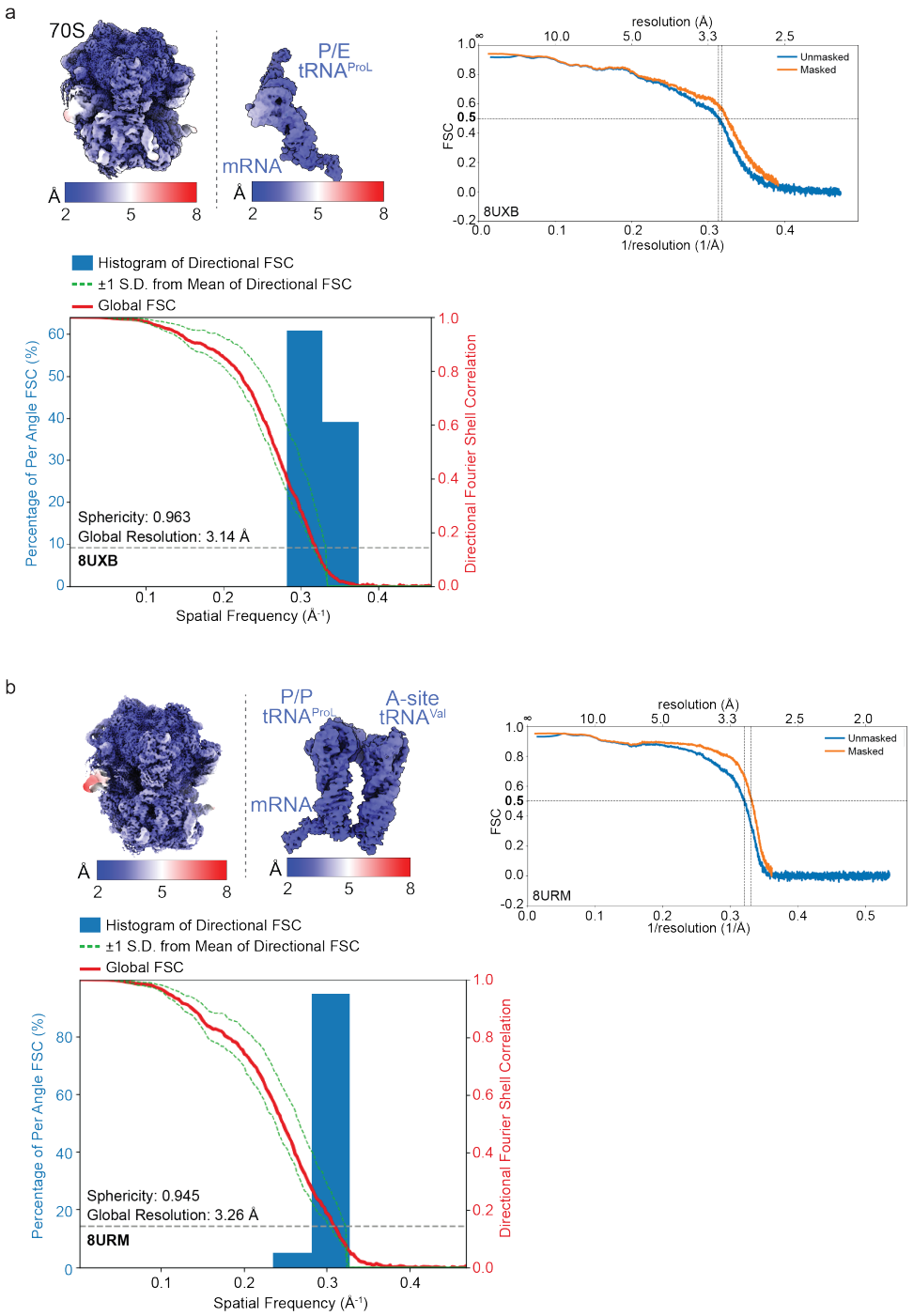


Supplementary Figure 1. Representative fluorescence intensity and E_{FRET} vs. time trajectories recorded for 70S ribosome complex containing a slippery CCC-C proline codon in the P site and an unmodified tRNA^{ProL} in the P site. Representative Cy3 (green) and Cy5 (red) fluorescence intensity vs. time trajectories and corresponding E_{FRET} (blue) vs. time trajectories for ribosomal complexes described in Figure 2B. The Viterbi paths (black) obtained from the hidden Markov modeling of the raw E_{FRET} *vs.* time trajectories are superimposed on the E_{FRET} trajectories. For the E_{FRET} *vs.* time trajectories, The FRET state at the higher E_{FRET} value of ~0.53 corresponds to GS1, and the FRET state at the lower E_{FRET} value of ~0.29 corresponds to GS2. Upon photobleaching of the Cy3 or Cy5 fluorophores, the E_{FRET} value drops to 0. Details of the smFRET data analysis, including how the fractional populations of GS1 and GS2 (% GS1 and % GS2, respectively), the rates of transitions between GS1 and GS2 ($k_{GS1\rightarrow GS2}$ and $k_{GS2\rightarrow GS1}$), and the corresponding equilibrium constant (K_{eq}) (Supplemental Figure S1) were calculated can be found in the Materials and Methods.

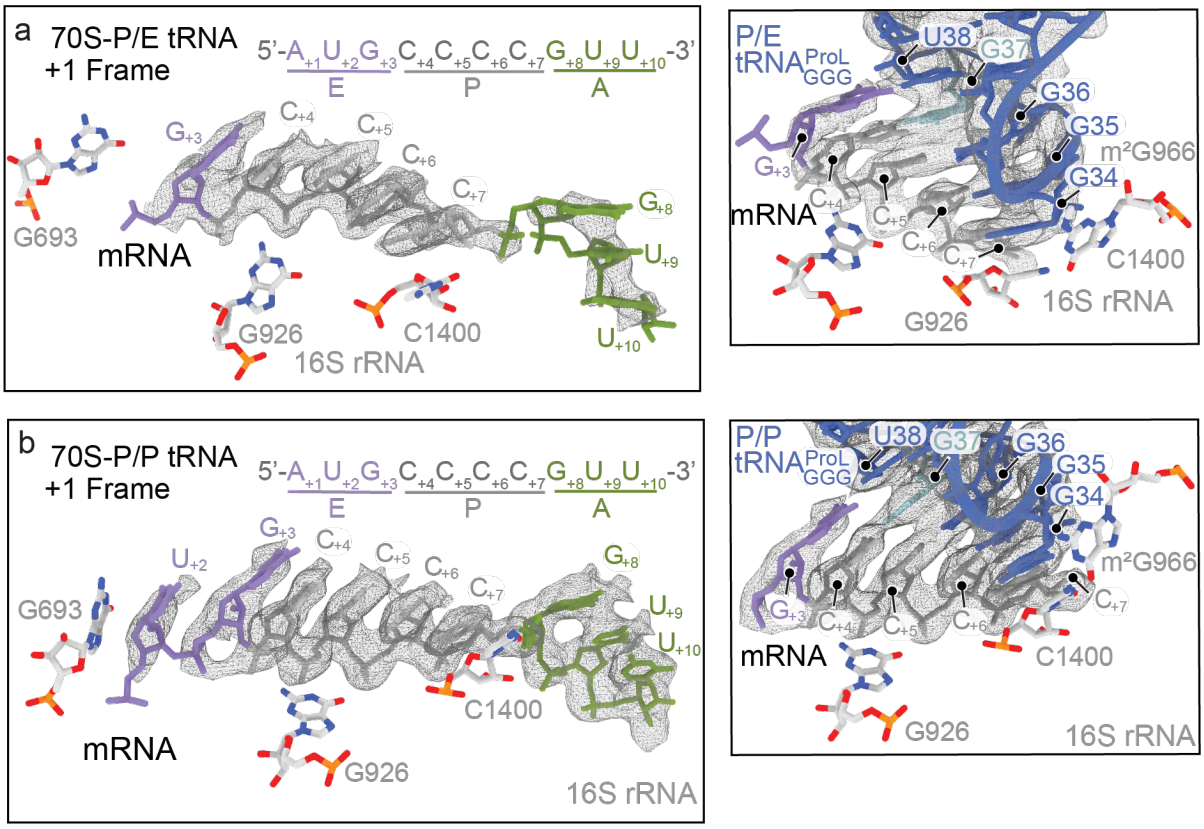


Supplementary Figure 2. Cryo-EM processing workflow for the 70S + unmodified tRNA^{ProL} + tRNA^{Val} complex (POST–). Representative cryo-EM micrograph and flow chart of data processing and refinement with representative 2D class averages and 3D classifications shown.

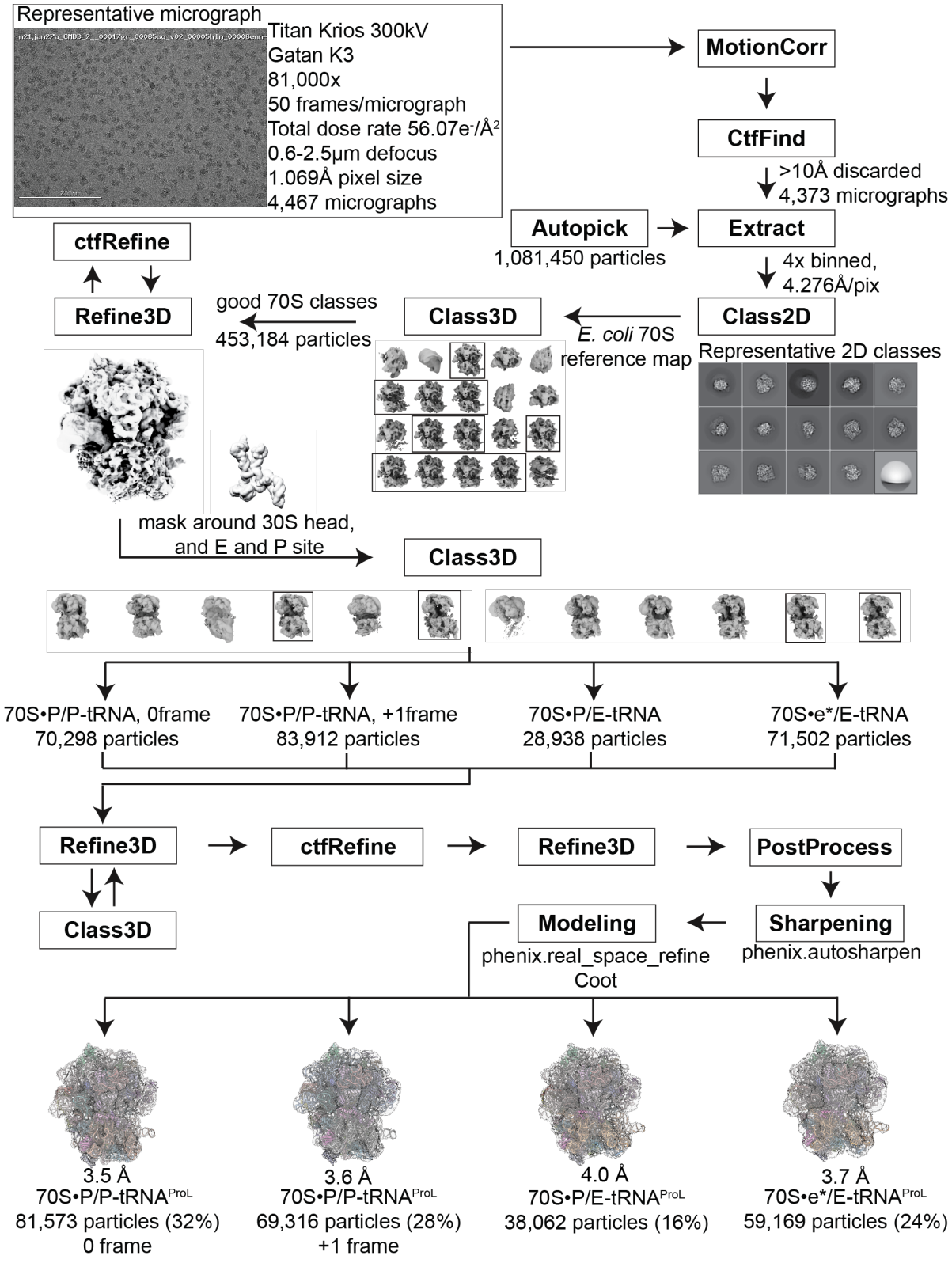
All initial 3D classes containing tRNA were selected and combined for 3D refinement, a round of ctf refinement, and a second 3D refinement. The output from the second 3D refinement underwent an unaligned, masked 3D classification (the mask was designed around the 30S head (16S rRNA nucleotides 900-1200, E/E tRNA and P/P tRNA). Three high-resolution classes were isolated, and each class underwent 3D refinement and a subsequent round of CTF refinement and 3D classification to remove any low-resolution particles. The 3D classification yielded two classes, one with E, P, and A-site tRNA and one with only E-site tRNA. The class with E-, P-, and A-site tRNA underwent another focus mask classification (mask of the 30S head (16S rRNA nucleotides 900-1200, E-tRNA, P-tRNA, and A/-tRNA), this yielded 3 high resolution classes. The class with only E-site tRNA underwent a different focus mask classification (mask of the mRNA, P-, and A-site tRNA) and this yielded 3 high resolution classes (excluding classes with empty ribosomes or ribosomes bound to Hibernation Promotion Factor). After a final 3D refinement, the resulting maps were post processed and autosharpened. The two populations of particles highlighted with a green circle are the particles that had high quality maps that enabled us to build molecular models into. The other four populations had poor map quality around the tRNAs that prevented us from building models. Iterative rounds of modeling in Coot and real-space refinements in Phenix were used to build final models into the maps. PDB codes 8UXB and 8URM.



Supplementary Figure 3. Local resolution maps for the classes isolated from the 70S + unmodified tRNA^{ProL} + tRNA^{Val} complex dataset (POST–). Filtered maps colored by the estimated resolution for the 70S, tRNA, and mRNA (left panel) and half map Fourier shell correlation (FSC) curves (right panel) for the 3D reconstructions containing: a, the P/E-tRNA in the +1 frame (PDB code 8UXB) and b, P/P-tRNA^{ProL} and tRNA^{Val} with the mRNA in the +1 frame (PDB code 8URM).

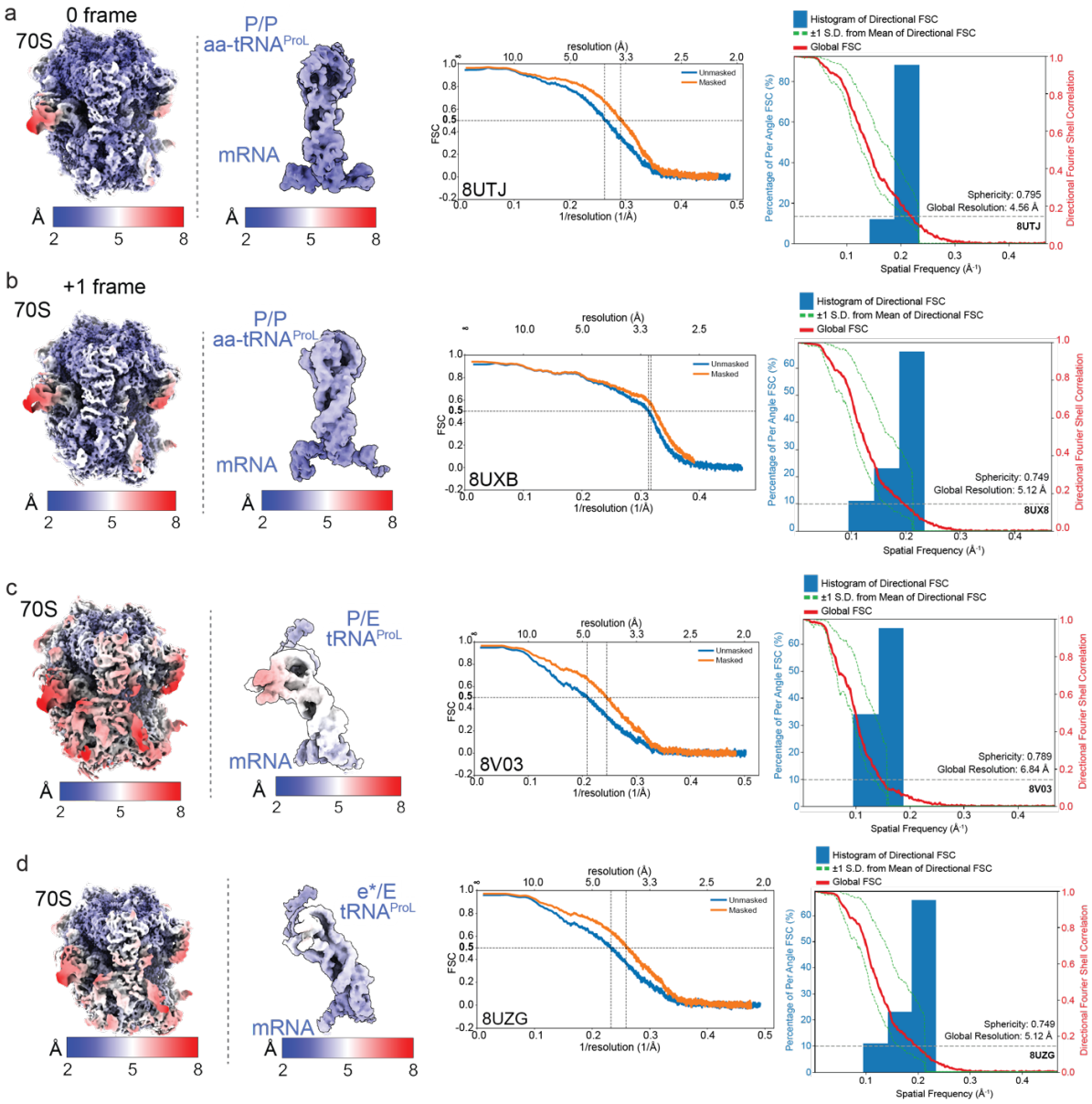


Supplementary Figure 4. Representative map quality of mRNA and tRNA for the 70S + unmodified tRNA^{ProL} + tRNA^{Val} complex dataset (POST–). a, (left) Map quality of the mRNA in the 70S + unmodified P/E-tRNA^{ProL} structure in the +1 frame. 16S rRNA nucleotides that surround the E, P and A sites are shown for context. (right) Map quality of both the mRNA and the tRNA in the same complex in the P/E site. Shown is the Phenix autosharpened cryo-EM map with a threshold of 5.36 (both panels), range -1.36-11. (PDB code 8UXB) b, (left) Map quality of the mRNA in the in the 70S + unmodified P/P-tRNA^{ProL} + A-site tRNA^{Val} structure in the +1 frame. 16S rRNA nucleotides that surround the E, P and A sites are shown for context. (right) Map quality of both the mRNA and the tRNA in the same complex in the P site. Shown is the Phenix autosharpened cryo-EM map with a threshold of 4.51 (both panels), range -1.24-15.1 (PDB code 8URM).

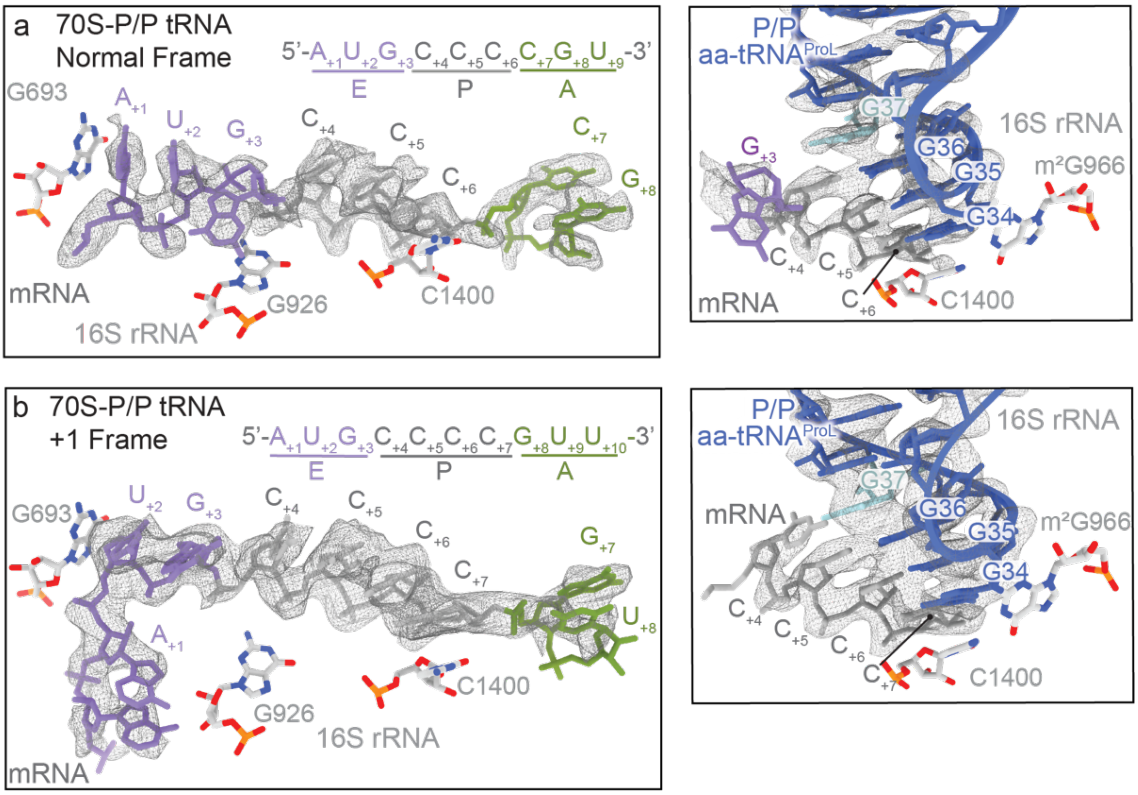


Supplementary Figure 5. Cryo-EM processing workflow for the 70S + unmodified aa-tRNA^{ProL} complex (POST). Representative cryo-EM micrograph and flow chart of data processing and refinement with representative 2D class averages and 3D classifications shown.

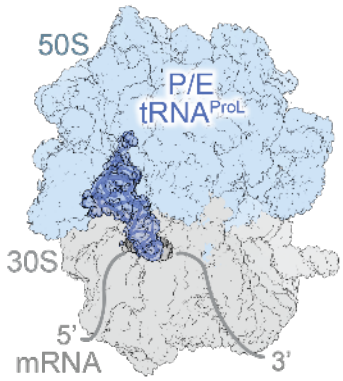
All initial 3D classes containing tRNA were selected and combined for 3D refinement, a round of CTF refinement, and a second 3D refinement. The output from the second 3D refinement underwent an unaligned, masked 3D classification (the mask was designed around the 30S head (16S rRNA nucleotides 900-1200, E- and P-site tRNAs). Four high resolution classes were isolated, and each class underwent 3D refinement and a subsequent round of CTF refinement and 3D classification to remove any low-resolution particles. The 3D classification for the e*/E-tRNA containing class yielded a small subset of P/E-tRNA containing particles that were incorporated into the P/E-tRNA class. After a final 3D refinement, the resulting maps were post processed and autosharpened. Iterative rounds of modeling in Coot and real-space refinements in Phenix were used to build final models into the maps. PDB codes 8UX8, 8UTJ, 8V03 and 8UZG.



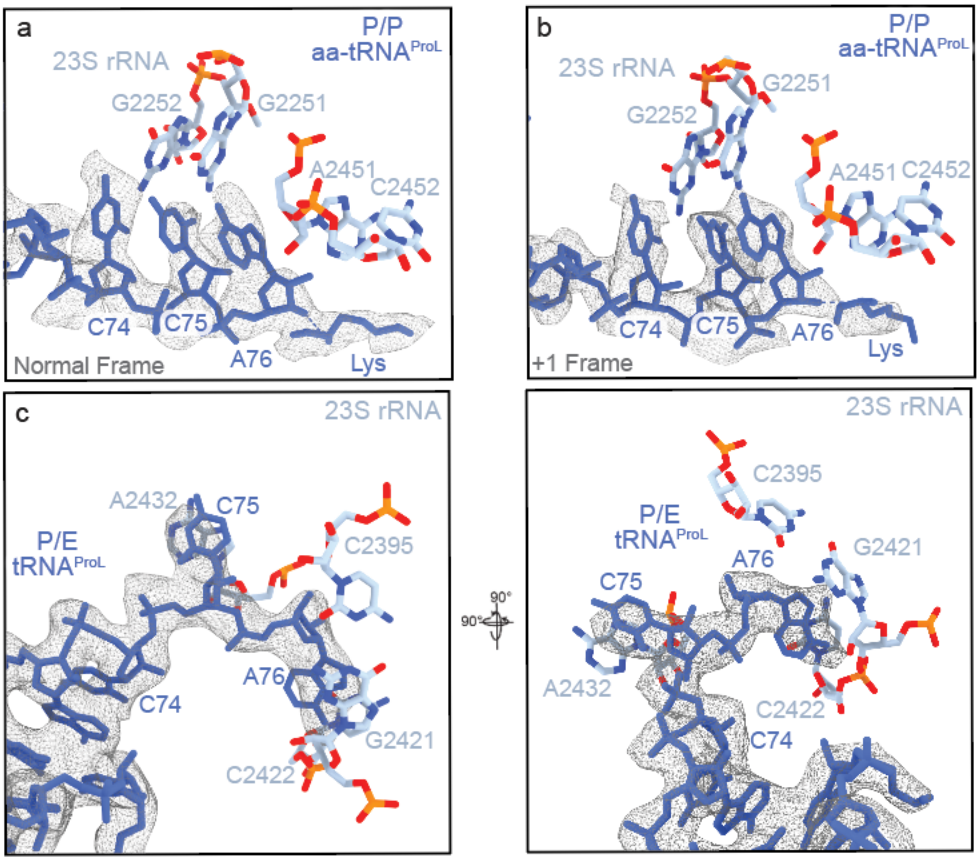
Supplementary Figure 6. Local resolution maps for the classes isolated from the 70S + unmodified aa-tRNA^{ProL} complex dataset. Filtered maps colored by the estimated resolution for the 70S, tRNA, and mRNA (left panel) and half map Fourier shell correlation (FSC) curves (right panel) for the final 3D reconstructions containing: a, P-tRNA in the normal frame; b, P-tRNA in the +1 frame; and c, P/E tRNA. PDB codes 8UX8, 8UTJ, 8V03 and 8UZG.



Supplementary Figure 7. Representative map quality of mRNA and tRNA from the 70S + unmodified aa-tRNA^{ProL} complex dataset. a, (left) Map quality of the mRNA in the 70S-P-site aa-tRNA^{ProL} structure in the normal frame. 16S rRNA nucleotides that surround the E, P and A sites are shown for context. (right) Map quality of both the mRNA and the tRNA in the same structure in the P site. Shown is the Phenix autosharpened cryo-EM map with a threshold of 4.58 (left) or 5.96 (right), range -4.69-14.9. b, (left) Map quality of the mRNA in the 70S-P-site aa-tRNA^{ProL} complex in the +1 frame. 16S rRNA nucleotides that surround the E, P and A sites are shown for context. (right) Map quality of both the mRNA and the tRNA in the same structure in the P site. Shown is the Phenix autosharpened cryo-EM map with a threshold of 4.58 (left) or 5.96 (right), range -4.69-14.9. Shown is the Phenix autosharpened cryo-EM map with a threshold of 5.28 (left) or 7.21 (right), range -2.11-16.3. PDB codes 8UX8 and 8UTJ.



Supplementary Figure 8. 4.0 Å cryo-EM ribosome structure of unmodified tRNA^{ProL} on a slippery CCC-C proline codon in a P/E state (16% particles).



Supplementary Figure 9. Aminoacylation status for unmodified P/P-, P/E-, and e*/E-tRNA^{ProL} and interactions with 23S rRNA nucleotides. a,b, 3'-CCA-Lys and (c) 3'-CCA end from the 70S-tRNA^{ProL} structures and corresponding interactions with 23S rRNA. The Lys-CCA tail of P-site, unmodified aa-tRNA^{ProL} in the normal frame (a) or +1 frame (b) forms Watson-Crick interactions between C74 and C75 and 23S rRNA nucleotides G2252 and G2251 of the peptidyl transferase center, respectively. b, The 3'-CCA end of the unmodified P/E-tRNA^{ProL} and (c) unmodified e*/E-tRNA^{ProL} no longer interact with the peptidyl transferase center and instead interact with 23S rRNA nucleotides in the E site. Shown are Phenix autosharpened map with a threshold of a, 5.45 (range -4.69-14.9), b, Phenix autosharpened map with a threshold of 6.34 (range -2.11-16.3), c, Phenix autosharpened map with a threshold of 4.91 (range -1.47-14).