

Proofreading neutralizes potential error hotspots in genetic code translation by transfer RNAs

JINGJI ZHANG,¹ KA-WENG IEONG,¹ HARRIET MELLENIUS, and MÅNS EHREBERG

Department of Cell and Molecular Biology, Uppsala University, Uppsala 75124, Sweden

ABSTRACT

The ribosome uses initial and proofreading selection of aminoacyl-tRNAs for accurate protein synthesis. Anomalously high initial misreading in vitro of near-cognate codons by tRNA^{His} and tRNA^{Glu} suggested potential error hotspots in protein synthesis, but in vivo data suggested their partial neutralization. To clarify the role of proofreading in this error reduction, we varied the Mg²⁺ ion concentration to calibrate the total accuracy of our cell-free system to that in the living *Escherichia coli* cell. We found the total accuracy of tRNA selection in our system to vary by five orders of magnitude depending on tRNA identity, type of mismatch, and mismatched codon position. Proofreading and initial selection were positively correlated at high, but uncorrelated at low initial selection, suggesting hyperactivated proofreading as a means to neutralize potentially disastrous initial selection errors.

Keywords: ribosome; genetic code; accuracy; tRNA selection; proofreading; error hotspots

INTRODUCTION

Recent biochemical observations on the accuracy of initial codon selection by seven aminoacyl-tRNAs (Johansson et al. 2012; Zhang et al. 2015) have raised pertinent questions. The data set displays a remarkable variation in the ability of aminoacyl-tRNAs in a ternary complex with EF-Tu and GTP to discriminate between codon–anticodon interactions with fully matching base pairs and those with a single mismatch. As expected, discrimination of uracil:guanine (U:G) mismatches in the middle codon position was comparatively poor. It was, at the same time, not expected that discrimination against the very same type of U:G mispairing in middle codon position would vary by two orders of magnitude depending on the type of tRNA in the ternary complex. Indeed, our recent findings are in stark contrast to previous claims that the accuracy of codon reading by tRNAs is nearly uniform (Gromadski et al. 2006). We also identified potential error hotspots for codon misreading in the living cell, like Glu-tRNA^{Glu}_{UUC} misreading codons GGA and GAU with predicted initial selection errors in the 1% range (Zhang et al. 2015). It was gratifying that the very same codons, GGA and GAU, had anomalously high errors also in vivo, as shown in pioneering work by Farabaugh and collaborators (Manickam et al. 2014). They used β-galactosidase mutants in which an essential glutamic acid (Glu) codon GAA had

been replaced by near-cognate codons encoding other amino acids. Accordingly, the residual activity of the mutants depended on misreading of near-cognate codons by tRNA^{Glu} for errors above the background of the method, i.e., at error frequencies above the 10⁻⁶ to 10⁻⁵ range. It was at the same time clear that even in the low Mg²⁺ concentration limit our initial selection data (Zhang et al. 2015) fell short by factors of 10–100 in relation to the in vivo accuracy (Manickam et al. 2014), which includes the obligatory proofreading step (Thompson and Stone 1977; Ruusala et al. 1982; Gromadski and Rodnina 2004). A peculiar aspect that emerges through this comparison is that correspondence between in vivo and biochemical data requires proofreading to provide similar amplification factors as the total accuracy and initial selection change in the low accuracy range. Intuitively, however, one would expect initial selection and proofreading to co-vary, so high initial selection correlates with large factor of accuracy amplification by proofreading. The simplistic rationale is that the same type of codon–anticodon mismatch would be used for discrimination in both selection steps, which would lead to covariation of initial selection and proofreading.

In the present work, we explore the total accuracy of cognate in relation to all near-cognate codon reading by tRNA^{Glu}, tRNA^{Lys}, and tRNA^{Phe}, in each case along with the proofreading contribution to the total accuracy level. By varying the Mg²⁺ concentration we were able to calibrate the accuracy level from our biochemistry to the codon

¹Joint first authors

Corresponding author: ehreberg@xray.bmc.uu.se

Article published online ahead of print. Article and publication date are at <http://www.majournal.org/cgi/doi/10.1261/rna.055632.115>. Freely available online through the RNA Open Access option.

© 2016 Zhang et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

reading accuracy *in vivo* monitored by Farabaugh and collaborators. We discuss the present remarkable result that proofreading decreases sharply with decreasing initial selection in the high initial selection accuracy range, but remains constant as initial selection decreases further in its low accuracy range.

The present result in combination with previous work has made our cell-free system for protein synthesis *in vivo* compatible with respect to all major steps of protein synthesis, including the accuracy of codon selection, initiation (Pavlov et al. 2011), peptide bond formation (Johansson et al. 2008a), translocation (Borg and Ehrenberg 2015), termination (Indrisiunaite et al. 2015), and ribosome recycling (Borg et al. 2015). This opens for extensive integration between *in vitro* and *in vivo* experiments, including systems biology modeling of bacterial physiology based on the biochemistry of protein synthesis. The close correlation that now exists between *in vivo* and *in vitro* kinetics of the ribosome may also serve as a guideline for biochemical experiments of relevance for living cells.

RESULTS

Measuring total accuracy of tRNA selection

Here we have used a cell-free system for protein synthesis with *Escherichia coli* components of high purity, specific activity, and *in vivo*-like kinetics (Johansson et al. 2008a) to study the accuracy by which three aminoacyl-tRNAs (aa-tRNAs) select their cognate in relation to all near-cognate codons on the messenger RNA (mRNA) programmed ribosome for subsequent peptide elongation. Ribosomes discriminate between cognate and noncognate aa-tRNAs in two consecutive steps, initial selection and proofreading. In this process, an aa-tRNA in a ternary complex with elongation factor Tu (EF-Tu) and guanosine triphosphate (GTP) enters a ribosome with a tRNA free A-site programmed with an amino acid encoding base triplet (codon) (Fig. 1A). During initial selection, the aa-tRNA in the ternary complex is selected in a reaction that proceeds to GTP hydrolysis with high probability if the codon is cognate and with low probability if the codon is near-cognate to the aa-tRNA. Proofreading selection occurs after dissolution of the ternary complex by GTP hydrolysis on EF-Tu, and the aa-tRNA proceeds to peptide bond with high probability if the codon is cognate but is ejected with high probability in the proofreading stage if the codon is near-cognate (Fig. 1A). It is generally assumed that proofreading occurs before aa-tRNA accommodation in the A site, as depicted in Figure 1A (Gromadski and Rodnina 2004; Ogle and Ramakrishnan 2005). It can, however, not be excluded that in some cases noncognate aa-tRNAs are discarded from the A site after accommodation but before peptide bond formation. The total accuracy, A , by which an amino acid is incorporated is defined by the ratio between the k_{cat}/K_m values for peptide bond formation from a cognate (c) ternary complex and a noncognate (nc) ternary complex:

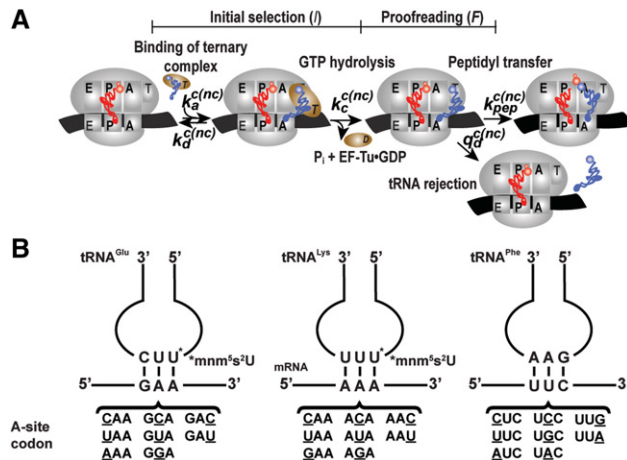


FIGURE 1. tRNA selection on the ribosome. (A) Kinetics scheme of tRNA selection on the mRNA-programmed ribosome. tRNAs are selected on the ribosome during initial selection (I) and proofreading (F) stages. These two stages were separated by GTP hydrolysis on EF-Tu. The total accuracy (A) of tRNA selection can be written as $A = I \cdot F$. (B) The total accuracy of tRNA selection was measured for Glu-tRNA^{Glu}, Lys-tRNA^{Lys}, and Phe-tRNA^{Phe} reading all possible single-mismatch codons, compared to their fully matched codons AAA (tRNA^{Lys}), GAA (tRNA^{Glu}), and UUC (tRNA^{Phe}). Mismatch codon positions are underlined.

$A = (k_{cat}/K_m)^c / (k_{cat}/K_m)^{nc}$. The universal definition of k_{cat}/K_m for enzymatic reactions is as the rate constant for substrate association multiplied by the probability that the first enzyme–substrate encounter leads to product formation rather than substrate dissociation. In a steady state situation with equal concentrations of competing cognate and noncognate substrates the ratio between cognate and noncognate product formation flows is equal to the accuracy, A . For the experiments we used Glu-tRNA^{Glu}_{UUC}, Lys-tRNA^{Lys}_{UUU}, and Phe-tRNA^{Phe}_{GAA} with cognate codons GAA, AAA, and UUC, respectively, and near-cognate codons as shown (Fig. 1B). We prepared ternary complex mixtures with either one of the three aa-tRNAs in the ternary complex (T_3) with EF-Tu and GTP. The mixtures in addition contained an energy regeneration system driven by phosphoenolpyruvate and other components, as described in Johansson et al. (2012). We prepared ribosome complex mixtures containing initiator tRNA charged with [³H]Met and formylated, f[³H] Met-tRNA^{Met}_{CAU}, in the initiation codon (AUG) programmed ribosomal P site, as described in Figure 1A (Johansson et al. 2012). The A site was programmed with any one of the cognate or near-cognate codons for the three aa-tRNAs (Fig. 1B). Incubation of the ribosome and ternary complex reactants for GTP hydrolysis in EF-Tu and subsequent peptide bond formation were started by mixing of the ribosome and ternary complex mixtures. The cognate reactions were in general fast and the reactions were carried out in a quench-flow instrument, as described in Johansson et al. (2012). The near-cognate reactions, however, were in general slow, so the experiments were carried out by hand.

In a typical set of experiments, a mixture with Glu-tRNA^{Glu}_{UUC} in a ternary complex with EF-Tu·GTP was rapidly mixed in the quench-flow instrument (for the cognate reactions) or by hand (for the near-cognate reactions) with a ribosome mixture containing ribosomes A-site programmed with the cognate Glu codon GAA or the near-cognate Gly codon GGA. The reaction was quenched after different incubation times and the extent of f[³H]Met-Glu formation was monitored by HPLC with on-line radiometry, as described in Johansson et al. (2012). The extent of dipeptide formation at different T₃ concentrations in excess is shown (Fig. 2A, black curves for the cognate and red curves for the near-cognate reactions). The cognate reactions took place in the 10-msec range and the near-cognate reactions in the 100-sec range. For both cognate and near-cognate reactions, the dipeptide formation rate increased with increasing ternary complex concentration (note the logarithmic time scale). The reaction rate, strictly defined as the inverse of the mean time for peptide bond formation, was plotted as a function of the ternary complex concentration (Fig. 2B, black curve in the cognate and red curve in the near-cognate case). The k_{cat}/K_m and K_m values for both cognate and near-cognate reactions were estimated by fitting the Michaelis–Menten expression (Fig. 2B, legend) to the experimental curve (Fig. 2B; Supplemental Table S1).

Calibration of cell-free codon selection accuracy to that in the living cell

Farabaugh and collaborators obtained estimates of the frequencies by which Glu-tRNA^{Glu} misreads codons CAA (Gln), UAA (stop), AAA (Lys), GCA (Ala), GGG (Gly), GGA (Gly), GAU (Asp), and GAC (Asp) in competition with the cognate tRNA readers of these codons in the living

E. coli cell (Manickam et al. 2014). For this, they used a β -galactosidase (β -gal) mutant in which the functional Glu codon (GAA) at position 537 had been reprogrammed to each one of the above codons, near-cognate to Glu-tRNA^{Glu}. Since Glu 537 is near-essential for β -gal function, β -gal activity could be used to estimate the tRNA^{Glu}-dependent missense error frequency of Glu insertion at position 537. The caveat of an apparent misreading, caused by a background of residual β -gal activity of the mutated enzyme variants themselves or of wild-type enzymes emerging from errors in transcription or aminoacylation, was checked in experiments with hyper-accurate and error-prone ribosome mutants (Manickam et al. 2014). To calibrate the error level in our biochemical system to that in the living cell, we measured k_{cat}/K_m values for peptide bond formation from Glu-tRNA^{Glu} containing a ternary complex on ribosomes programmed with all possible near-cognate codons. To derive the in vivo error frequencies, as measured by Farabaugh, from the present biochemical data, we used here the measured ratios of k_{cat}/K_m values for cognate and near-cognate peptide bond formation, along with previous estimates of tRNA iso-acceptor concentrations in the *E. coli* cell (Dong et al. 1996). This is motivated by previous findings that the variation of cognate k_{cat}/K_m values is similar in the cases we have studied (Johansson et al. 2011; Zhang et al. 2015). The accuracy of our system was tuned by different concentrations of free Mg²⁺ ions from 1.3 mM (standard polymix concentration) to 7.5 mM (addition of 10 mM extra Mg²⁺ in polymix) (Johansson et al. 2012). Since there is no proofreading for the cognate codon reading (Thompson and Stone 1977; Ruusala et al. 1982; Gromadski and Rodnina 2004), k_{cat}/K_m values are the same for GTP hydrolysis and peptide bond formation. At each Mg²⁺ concentration, we used the k_{cat}/K_m values for GTP hydrolysis from the ternary complex with Glu-tRNA^{Glu} on ribosomes programmed with Glu codon

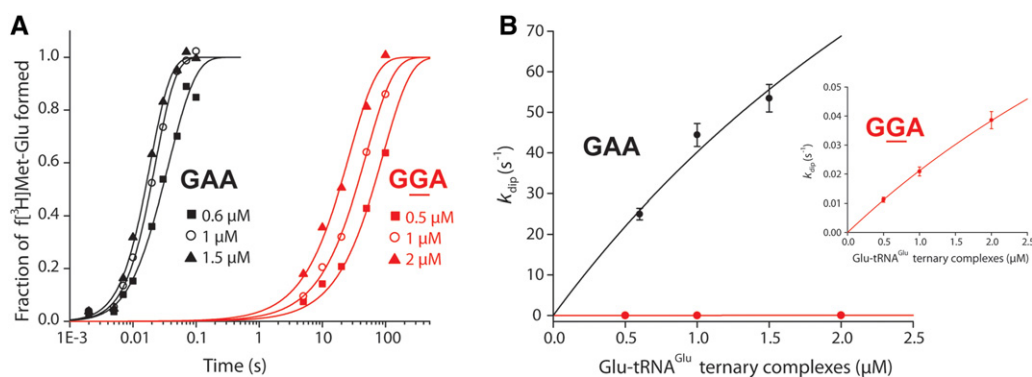


FIGURE 2. Measurements of cognate and near-cognate k_{cat}/K_m values for dipeptide formation. (A) Time evolution of dipeptide f[³H]Met-Glu formation. Ternary complexes EF-Tu·GTP·Glu-tRNA^{Glu} were reacted with 70S initial complexes programmed with f[³H]Met-tRNA^{fMet} in the P site and a cognate codon GAA (black) or near cognate codon GGA (red) in the A site. Reactions were performed at increasing complex concentration as indicated in the figure. Ternary complexes were in excess over ribosomes so that the rate of dipeptide formation k_{dip} was limited by ternary complex concentration. (B) Concentration dependence of the rate of dipeptide formation k_{dip} estimated from A. k_{cat}/K_m values were estimated by fitting the data into Michaelis–Menten model $k_{\text{dip}} = ((k_{\text{cat}}/K_m)[T_3]) / (1 + [T_3]/K_m)$. (Inset) Near-cognate reaction. Experiments were performed in polymix buffer with 2.3 mM free Mg²⁺.

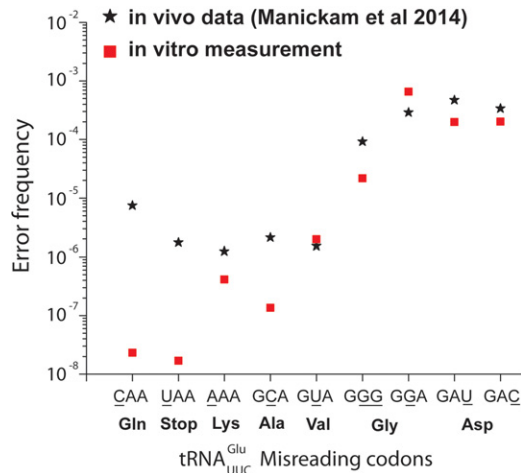


FIGURE 3. Comparison of in vivo and in vitro misreading error frequency. In vivo data (black stars) from the Farabaugh laboratory (Manickam et al. 2014) are based on induction of bioluminescence by mistranslation by tRNA^{Glu} ternary complex from *E. coli* strains with β -galactosidase mutants. In vitro measurements (red squares) were performed at 2.3 mM free Mg²⁺ and calibrated to in vivo condition according to the abundance of the two competing tRNA species in vivo (Dong et al. 1996) (see Materials and Methods) and assuming different ternary complexes as well as release factors have similar efficiencies for binding to ribosomes in the living cell. Mismatch codon positions are underlined.

GAA under the same condition from Zhang et al. (2015) as a proxy for in vivo reading of all codons near-cognate to the Glu codon by their respective cognate tRNAs and release factors (see Materials and Methods). The cognate and near-cognate k_{cat}/K_m values for fMet-Glu formation were measured at each Mg²⁺ concentration as described in the previous section (Supplemental Table S1). The best accuracy fit with in vivo accuracy was obtained at 2.3 mM free Mg²⁺ concentration, as shown in Figure 3. Comparison between our data from biochemical experiments (red squares) and in vivo data (black

stars) shows good correlation at error levels from 10⁻⁵ and higher. It is also clear that the error hotspots for Glu-tRNA^{Glu} reading of the near-cognate Gly (GGA) and Asp (GAU, GAC) have been independently identified in vivo (Manickam et al. 2014) and in the present work, highlighting the physiological relevance of our biochemistry. It is seen that when our error estimates go down below 10⁻⁶, the in vivo estimates remain in the 10⁻⁶ to 10⁻⁵ range, a result very likely due to an error background in this range (Manickam et al. 2014).

Accuracy of genetic code translation varies over five orders of magnitude

Next, we expanded the total accuracy measurements to include also fMet-Lys and fMet-Phe formation from Lys-tRNA^{Lys} and Phe-tRNA^{Phe} containing a ternary complex respectively on ribosomes with cognate and all near-cognate codons for these two aa-tRNAs (Fig. 1B). All measurements were performed at 2.3 mM free Mg²⁺ concentration (addition of 2 mM extra Mg²⁺ in polymix) and, hence, under conditions relevant to the corresponding accuracy levels in the living cell (Fig. 3). As described above, the accuracy was in each case obtained as the ratio between the k_{cat}/K_m value for a particular tRNA isoacceptor reading its cognate codon and the k_{cat}/K_m value for the same tRNA reading a near-cognate codon. All k_{cat}/K_m values for dipeptide formation on Glu-tRNA^{Glu}, Lys-tRNA^{Lys}, and Phe-tRNA^{Phe} are summarized in Supplemental Table S1. The accuracy values are shown graphically in Figure 4 (white staples) and summarized in Table 1.

How proofreading correlates with initial selection

In bacterial translation, the total accuracy, A , is factorized into an initial selection part, I , and a proofreading part, F (Fig. 1A; Thompson and Stone 1977; Ruusala et al.

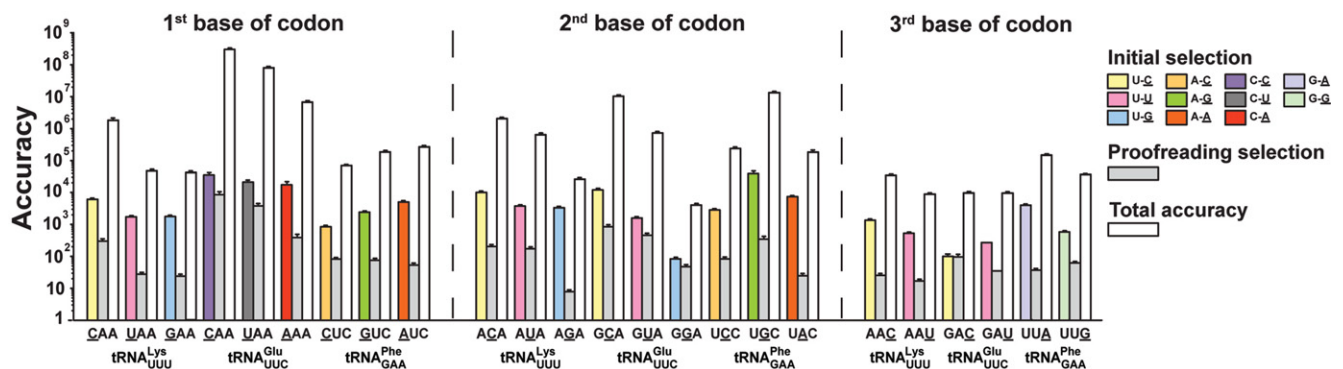


FIGURE 4. Accuracy of tRNA selection varies with mismatch codon positions, type of mismatch, and type of aa-tRNA. Total accuracy for single-base mismatch codons reading by tRNA^{Lys}_{UUU}, tRNA^{Glu}_{UUC}, and tRNA^{Phe}_{GAA} were summarized with respect to mismatch codon positions and mismatch identities. First, second, and third codon mismatch positions are shown in *left*, *middle*, and *right* panels, respectively. Mismatch identities are shown in different colors for initial selection. Gray area for each mismatch shows the contribution of proofreading (F) to total accuracy (with the white area), where proofreading was calculated from overall accuracy divided by initial selection (see Materials and Methods). Measurements were performed at 2.3 mM free Mg²⁺. Initial selection data were from Zhang et al. (2015).

TABLE 1. Initial selection (*I*), proofreading (*F*), and total accuracy (*A*) values for misreading of single-base mismatch codons by different tRNAs

tRNA	First base of codon			Second base of codon			Third base of codon				
	Mismatch	<i>I</i>	<i>F</i>	Mismatch	<i>I</i>	<i>F</i>	Mismatch	<i>I</i>	<i>F</i>	<i>A</i> × 10 ⁻³	
tRNA ^{Glu} _{UUC}	C-C	35,000 ± 7000	8600 ± 1900	300,000 ± 30,000	U-C	12,000 ± 1500	850 ± 130	U ^a -C	272	35	9.6 ± 1.0
	C-U	21,000 ± 3200	3800 ± 700	80,000 ± 8300	U-U	1600 ± 170	460 ± 70	U ^a -U	100 ± 17	97 ± 19	9.7 ± 1.0
	C-A	17,000 ± 4400	390 ± 110	6800 ± 700	U-G	84 ± 9	48 ± 7				
tRNA ^{Lys} _{UUU}	U-C	6000 ± 500	300 ± 60	1800 ± 300	U-C	10,000 ± 1000	200 ± 30	U ^a -C	1400 ± 130	26 ± 3	35 ± 3
	U-U	1700 ± 160	28 ± 4	48 ± 5	U-U	4000 ± 300	170 ± 25	U ^a -U	530 ± 50	17 ± 2	8.9 ± 0.8
	U-G	1800 ± 170	24 ± 4	43 ± 5	U-G	3300 ± 320	8 ± 1				
tRNA ^{Phe} _{GAA}	A-C	850 ± 80	82 ± 10	70 ± 6	A-C	2900 ± 250	83 ± 12	G-G	4000 ± 350	37 ± 5	150 ± 13
	A-G	2400 ± 220	76 ± 10	190 ± 20	A-G	39,000 ± 8400	340 ± 80	G-A	580 ± 50	63 ± 7	37 ± 2
	A-A	5000 ± 540	53 ± 8	270 ± 27	A-A	7400 ± 610	25 ± 4				

Total accuracy was measured at in vivo-like condition (2.3 mM free Mg²⁺ in polymix buffer). Accuracy values for initial selection (*I*) were taken from (Zhang et al. 2015). Proofreading factor (*F*) were calculated as $F = A/I$.

^aModifications on tRNA position 34 (third codon position) are indicated for tRNA^{Lys} (mm⁵-U) and tRNA^{Glu} (mm⁵-s²U).

1982): $A = I \times F$. Previously, initial selection values (I) were estimated from the ratio of k_{cat}/K_m values for GTP hydrolysis between cognate and near-cognate reactions (Johansson et al. 2012; Zhang et al. 2015). Using initial selection values obtained at a concentration of free Mg^{2+} of 2.3 mM (Zhang et al. 2015), we calculated the proofreading factor, F , from the initial selection, I , for every accuracy, A , through $F = A/I$ as shown graphically in Figure 4 and summarized in Table 1. There is a remarkably large variation in the accuracy of codon reading depending on tRNA identity, mismatch position in the codon, and type of mismatch. The accuracy varied by five orders of magnitude, from 4×10^3 to 3×10^8 , initial selection and proofreading by about three orders of magnitude from 100 to 40,000 and eight to 8000, respectively (Table 1; Fig. 4).

It is often assumed that initial selection and proofreading have straightforward dependence on the very same type of codon–anticodon interaction to discriminate between cognate and noncognate codon reading. This would mean similar dependence of F and I on the same intrinsic discrimination parameter d and suggest that $\log_{10}(F)$ formally plotted as a function of $\log_{10}(I)$ would be a straight line with slope one. The parameter d is defined as the highest possible substrate selection accuracy of an enzyme state, as obtained when the rate constant of the forward reaction is much smaller than that of the discard reaction for both cognate and noncognate substrates (Ehrenberg and Blomberg 1980; Johansson et al. 2012; Zhang et al. 2015). What is observed, however, is a straight line with slope 1.6 at $\log_{10}(I)$ values larger than 3.5 and virtually constant $\log_{10}(F)$ at 1.7 in response to decreasing $\log_{10}(I)$ values smaller than 3.5 (Fig. 5A) (see Discussion and Supplemental Material).

The division of initial selection into a high and a low selection range was motivated by a comparison of the curve fitting with one line for the whole accuracy range (adjusted R^2 value

of 0.31; Supplemental Fig. S1; Supplemental Table S2), one constant and a line (adjusted R^2 value of 0.55; Fig. 5A; Supplemental Fig. S1; Supplemental Table S2), or two lines (adjusted R^2 value of 0.53; Supplemental Fig. S1; Supplemental Table S2). The hypotheses that (1) no linear correlation at all is better than the straight line correlation, (2) the straight line is a better model than the constant intersecting with a straight line, and (3) the constant and straight line model is better than a model with two intersecting straight lines, were all tested using the F distribution. The statistical analysis showed that hypotheses (1) and (2) could be rejected with P -values of 0.0108 and 0.00187, respectively, but that hypothesis (3) could not be rejected. These results and the adjusted R^2 -values (Supplemental Table S2) suggest that the best model is a constant and a straight line. For the full statistical analysis, see Supplemental Material.

DISCUSSION

In this work, we have collected a data set comprising the total accuracy and its partitioning in initial selection and proofreading by which each one of three of the 54 tRNA isoacceptors in *E. coli* discriminates between its cognate codon, containing three Watson–Crick base pairs, and all near-cognate codons containing one mismatched base pair (Table 1; Fig. 4). The accuracy of the data set was calibrated to the accuracy of codon reading in the living cell by varying the free Mg^{2+} concentration in the biochemical assays for best correspondence with the accuracy by which Glu-tRNA^{Glu} discriminates against codons near-cognate to the Glu codon (GAA). As mentioned above, the in vivo data set was obtained by careful measurement of the residual activity of β -galactosidase mutants expressed from open reading frames in which a Glu codon had been altered to codons near-cognate to Glu-tRNA^{Glu} (position 537) (Manickam et al. 2014). The

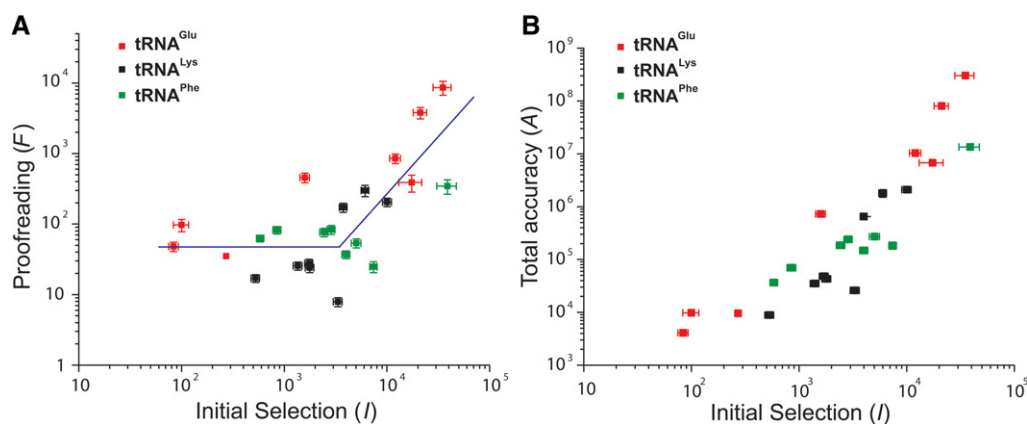


FIGURE 5. Enhanced proofreading on initial selection error hotspots. (A) Correlation between proofreading (F) and initial selection (I): a linear fit of $\log_{10}(F)$ versus $\log_{10}(I)$ for high I -values with a slope of 1.6 for I -values above $10^{3.5}$; at low I -values, in contrast, $\log_{10}(F)$ remains virtually constant at 1.7 as $\log_{10}(I)$ decreases further. A full description of the statistical analysis is found in the Supplemental Material. (B) Correlation between total accuracy (A) and initial selection (I). Data were obtained for all possible single-base mismatch codons reading by tRNA^{Glu}_{UUC}, tRNA^{Lys}_{UUU}, and tRNA^{Phe}_{GAA}. Measurements were performed at 2.3 mM free Mg^{2+} . Initial selection data were from Zhang et al. (2015).

best correspondence between *in vivo* and *in vitro* data was obtained for a free Mg^{2+} concentration of 2.3 mM in the polymix buffer previously developed for fast speed and high accuracy of mRNA translation (Johansson et al. 2012). At this Mg^{2+} concentration, the k_{cat}/K_m values for tRNA^{Glu}, tRNA^{Lys}, and tRNA^{Phe} were reduced about twofold in relation to their maximal values, as determined by their respective association rate constants (Fig. 6; Zhang et al. 2015). This may seem like a high kinetic cost (Gromadski and Rodnina 2004; Gromadski et al. 2006), but it must be borne in mind that increasing the Mg^{2+} concentration to speed up the cognate kinetics will activate near-cognate and noncognate ternary complexes as inhibitors of cognate protein synthesis (Johansson et al. 2008b). This means, in other words, that what appears as a gain in reality is a kinetic loss. We further note that the rate of translocation decreases sharply with increasing Mg^{2+} concentration (Borg and Ehrenberg 2015) and that, finally, 2.3 mM is close to estimates of the free, intracellular Mg^{2+} concentration (Alatossava et al. 1985).

The calibration attempt endeavored here is a multiparameter problem. It is therefore unlikely that the present approach will provide perfect *in vivo* mimicking conditions. One reason is that there is a multitude of buffer components in the living cell that are absent in the polymix buffer, and for many of the common components the free concentrations have remained unknown. Another reason is that kinetics in the living cell proceeds under conditions of very high protein density (McGuffee and Elcock 2010), affecting the speed of all diffusion-controlled reactions, whereas kinetics in the test tube proceeds in dilute solution.

It should, at the same time, be stressed that calibration of speed and accuracy of cell-free and intracellular protein syn-

thesis and protein synthesis in the living cell have not been previously achieved. We deem such a calibration essential for two main reasons. The first is that when *in vitro* and *in vivo* kinetics match each other, then biochemical modeling can root bacterial physiology and population genetics in kinetically well-characterized steps of ribosomal protein synthesis and its auxiliary reactions. Such rooting will, not the least, facilitate deep understanding of resistance development against antibiotic drugs. The second reason is that when the biochemistry of bacterial protein synthesis is calibrated to its intracellular counterpart, the relevance of the functional observations in the test tube will be much easier to assess. When the relevance is ascertained it will become even more meaningful to discuss the structures of the ribosome and its auxiliary factors in terms of the evolutionary pressure that has shaped their kinetics within the constraints of physical law.

The accuracy of near-cognate codon reading displays an enormous variation (Table 1; Fig. 4), which we suggest to be at work also in the living cell, as seen here in a special case (Fig. 3). Indeed, from our data we propose that the average intracellular error frequency is dominated by a small number of error hotspots similar to those for tRNA^{Glu} misreading near-cognate codon U:G in second and U:U in third codon position. Another is likely to be tRNA^{His} misreading U:G in second or third codon position (Zhang et al. 2015). This unevenness also motivates qualification of the common wisdom that the error frequency of transcription is much lower than that of translation (Gout et al. 2013; Imashimizu et al. 2013). It now seems that there is strong, template context-dependent transcription error variation (Mellenius and Ehrenberg 2015), and this prediction in conjunction with the present data (Table 1) suggest that transcription errors dominate greatly in some and translation errors in other contexts.

A striking feature of our data set is that initial selection and proofreading are strongly correlated in the high but not in the low initial selection range (Fig. 5A). That is, a plot of the logarithm of the proofreading parameter, F , versus the logarithm of the initial selection, I , is a straight line with a steep slope for high I -values. At low I -values, in contrast, $\log_{10}(F)$ remains virtually constant as $\log_{10}(I)$ decreases further, so the total accuracy decreases much more slowly with I in this region (Fig. 5B). One consequence of this is that what initially appeared as a catastrophic decline in initial selection by tRNA^{Glu} reading G in second and U or C in third position (Zhang et al. 2015), can now be seen to have misreading frequencies below the “canonical” *in vivo* error frequency of 1/2000 (Parker 1989, 1992; Kurland 1992). To explain this we first note that the large local variation of $\log_{10}(F)$ as a function of $\log_{10}(I)$ is not caused by measurement errors, which are confined to a 15% range, but to a real variation as it exists in living cells. At high initial selection values, the total accuracy is so high that decreasing d -values in initial selection and, by hypothesis, proofreading (Ehrenberg and Blomberg 1980; Johansson et al. 2012; Zhang et al. 2015)

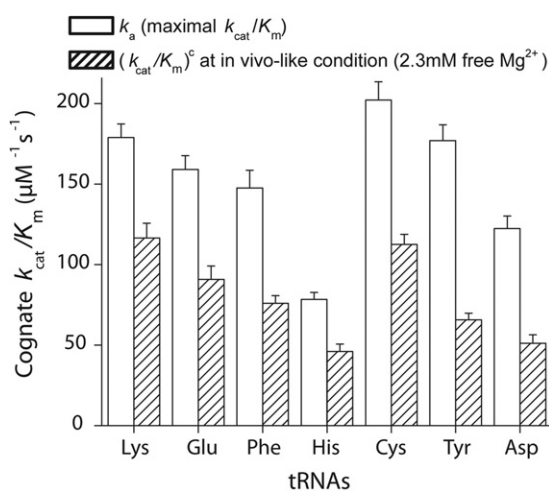


FIGURE 6. Kinetic loss for cognate reaction under *in vivo* condition. Cognate kinetic efficiency (k_{cat}/K_m)^c under *in vivo* condition (at 2.3 mM free Mg^{2+} , striped areas) for different tRNAs was compared to the association rate constants for binding of ternary complexes to the ribosome k_a , which is the maximal kinetic efficiency for dipeptide formation. All data shown here are from Zhang et al. (2015).

do not confer significant fitness reduction and are therefore tolerated. When, however, the d -values of initial selection and proofreading are reduced below an accuracy threshold, at which protein synthesis errors would greatly reduce the quality of the proteome and significantly decrease the fitness of the bacterial population, the evolutionary response has been to increase the expression of the d -values of proofreading. In a simple model this is done by reducing the ratio between the forward and discard rate constants in the proofreading branches (Supplemental Material). If taken at face value, the observation that the slope of the $\log_{10}(F)$ variation with $\log_{10}(I)$ is 1.6 (Fig. 5A) may suggest proofreading in two steps (Ehrenberg and Blomberg 1980) rather than one as previously suggested (Thompson and Stone 1977; Ruusala et al. 1982; Gromadski and Rodnina 2004; Gromadski et al. 2006). Indeed, two-step proofreading is a particularly attractive hypothesis since it would provide a natural explanation of how the expression of proofreading for noncognate substrates can be cranked up at low d -values with small concomitant losses of kinetic efficiency for the cognate reactions (Ehrenberg and Blomberg 1980). This being said, it is also clear that much more direct experimental evidence would be necessary to prove the existence of multistep proofreading of transfer RNAs in genetic code translation among bacteria.

MATERIALS AND METHODS

Reagents and buffer conditions

The purified *E. coli* translation system, including 70S ribosomes purified from MRE600 strain, synthetic mRNAs, initiation factors, elongation factors, and $f[{}^3\text{H}]\text{Met-tRNA}^{\text{Met}}$, were prepared according to Johansson et al. (2008a) and references therein. *E. coli* tRNA^{Glu}, tRNA^{Lys}, and tRNA^{Phe} were from Chemical Block. $[{}^3\text{H}]\text{Met}$ were from PerkinElmer, and all other chemicals were from Merck or Sigma-Aldrich. All experiments were performed in polymix buffer (Jelenc and Kurland 1979) containing 95 mM KCl, 5 mM NH₄Cl, 0.5 mM CaCl₂, 8 mM putrescine, 1 mM spermidine, 5 mM potassium phosphate, 1 mM dithioerythritol, 10 mM phosphoenolpyruvate (PEP), 5 mM Mg(OAc)₂, 1 mM ATP, and 1 mM GTP, with Mg(OAc)₂ additions between 0 and 10 mM, and supplemented with 1 μg/mL pyruvate kinase and 0.1 μg/mL myokinase for energy regeneration. As PEP chelates Mg²⁺ with a K_d -value of 6 mM (Wold and Ballou 1957) and assuming that one ATP or GTP molecule chelates one Mg²⁺, the free Mg²⁺ concentration in the reaction was estimated to be between 1.3 mM [without addition of Mg(OAc)₂] to 7.5 mM [with 10 mM addition of Mg(OAc)₂]. Free Mg²⁺ concentration was estimated to be 2.3 mM when 2 mM of extra Mg(OAc)₂ was added to polymix buffer.

Measurement of cognate and near-cognate dipeptide formation

Final concentrations after mixing equal volume of ribosome and ternary complex mixtures are given below. All kinetics measurements

were performed at 37°C. Ribosome mixture, containing ribosomes programmed with mRNA with mismatch A-site codon, were prepared by incubating 70S ribosomes (at concentrations at least 2× lower than ternary complex concentrations and between 0.2 and 0.5 μM), $f[{}^3\text{H}]\text{Met-tRNA}^{\text{Met}}$ (1.5× ribosome concentration), mRNA (2× ribosome concentration), IF1 (1.5× ribosome concentration), IF2 (0.5× ribosome concentration), and IF3 (1.5× ribosome concentration) in polymix buffer with varying addition of Mg(OAc)₂ at 37°C for 15 min. Ternary complex mixture, containing ternary complexes EF-Tu·GTP·Glu-tRNA^{Glu}, EF-Tu·GTP·Lys-tRNA^{Lys}, or EF-Tu·GTP·Phe-tRNA^{Phe}, was prepared by incubating the corresponding tRNA (tRNA^{Glu}, tRNA^{Lys}, or tRNA^{Phe}; varying concentrations between 0.5 and 15 μM), EF-Tu (4 μM in excess over the highest tRNA concentration in a ternary complex titration experiment), amino acid (Glu, Lys, or Phe; 0.2 mM), aa-tRNA synthetase (GluRS, LysRS, or PheRS; 1.5 units/μL), and EF-Ts (1.5 μM) in polymix buffer with addition of varying concentrations of Mg(OAc)₂ at 37°C for 15 min. Equal volumes of the ternary complex and ribosome mixtures were rapidly mixed in a temperature controlled quench-flow apparatus (RQF-3; KinTeck Corp.) (for cognate reactions) or manually (for near-cognate reactions). The reaction was stopped at different incubation times by rapidly quenching with formic acid (17% final concentration).

The extent of $[{}^3\text{H}]\text{dipeptide}$ formed at different time points was quantified by C18 reversed-phase HPLC equipped with a β-RAM model 3 radioactivity detector (IN/US Systems) (Johansson et al. 2012). The rate of cognate and near-cognate peptide bond formation ($k_{\text{dip}}^{\text{c/nc}}$) was estimated at different Mg²⁺ concentrations by fitting the data into a single exponential model:

$$\text{dip}(t) = [\text{Rib}] \cdot (1 - e^{-k_{\text{dip}}^{\text{c/nc}} t}) + \text{bg},$$

where $\text{dip}(t)$ is the time evolution of dipeptide formed, and the plateau $[\text{Rib}]$ is the active ribosome concentration in the reaction.

Estimate of k_{cat}/K_m for cognate and near-cognate dipeptide formation

For cognate reactions, we used k_{cat}/K_m -values for GTP hydrolysis from Zhang et al. (2015) as the k_{cat}/K_m -values for dipeptide formation, since there was no proofreading during tRNA selection (Thompson and Stone 1977; Ruusala et al. 1982; Gromadski and Rodnina 2004). As a control experiment, we only measured the k_{cat}/K_m -value for dipeptide formation for tRNA^{Glu} reading its cognate codon GAA at 2.3 mM free Mg²⁺ condition.

Ternary complexes, varied from 0.5 to 1.5 μM for cognate reaction and 0.5 to 15 μM for near-cognate reaction, are always in excess over ribosome complexes (0.2–0.5 μM); the rate of dipeptide formation was then limited by the concentration of ternary complexes $[T_3]$, and it follows Michaelis–Menten kinetics. At each Mg²⁺ concentration, the rate of cognate or near-cognate dipeptide formation $k_{\text{dip}}^{\text{c/nc}}$ was given by

$$k_{\text{dip}}^{\text{c/nc}} = \frac{(k_{\text{cat}}/K_m)^{\text{c/nc}}_{\text{dip}} \cdot [T_3]}{1 + ([T_3]/K_m)}.$$

The efficiency for cognate or near-cognate dipeptide formation $(k_{\text{cat}}/K_m)^{\text{c/nc}}_{\text{dip}}$ and the K_m -value were estimated by fitting $k_{\text{dip}}^{\text{c/nc}}$ to the experimental data.

Estimate of total accuracy (A) and proofreading (F) factor

Total accuracy (A) is defined as the ratio between the kinetic efficiency for cognate and noncognate reaction for dipeptide formation.

$$A = \frac{(k_{\text{cat}}/K_m)_{\text{dip}}^{\text{c}}}{(k_{\text{cat}}/K_m)_{\text{dip}}^{\text{nc}}} = I \cdot F.$$

The data of initial selection I were taken from Zhang et al. (2015). The proofreading parameter F was then calculated as $F = A/I$.

Calibration of in vitro accuracy to the accuracy in vivo

Since different ternary complexes as well as release factors had very similar binding efficiency to the ribosome (Freistroffer et al. 2000; Johansson et al. 2011; Zhang et al. 2015), we calibrated our in vitro measurement according to the ratio of the abundance of tRNA^{Glu} and the tRNA cognate to the mismatch codon (Dong et al. 1996) or abundance of the release factors (Bremer and Dennis 2008) for measurement on stop codon UAA.

Calibrated in vitro accuracy

$$= \frac{(k_{\text{cat}}/K_m)_{\text{dip}}^{\text{c}} \cdot [\text{tRNA}^{\text{c}} \text{ or release factors 1 and 2}]^{\text{in vivo}}}{(k_{\text{cat}}/K_m)_{\text{dip}}^{\text{nc}} \cdot [\text{tRNA}^{\text{Glu}}]^{\text{in vivo}}}.$$

Error frequency is calculated as the inverse of the calibrated in vitro accuracy, which is shown as (■) in Figure 3.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Dr. Philip Farabaugh for valuable discussions. This work was supported by grants from the Knut and Alice Wallenberg Foundation (RiboCORE), the Swedish Research Council, and the Human Frontier Science Program.

Received December 14, 2015; accepted March 7, 2016.

REFERENCES

- Alatossava T, Jutte H, Kuhn A, Kellenberger E. 1985. Manipulation of intracellular magnesium content in polymyxin B nonapeptide-sensitized *Escherichia coli* by ionophore A23187. *J Bacteriol* **162**: 413–419.
- Borg A, Ehrenberg M. 2015. Determinants of the rate of mRNA translocation in bacterial protein synthesis. *J Mol Biol* **427**: 1835–1847.
- Borg A, Pavlov M, Ehrenberg M. 2015. Complete kinetic mechanism for recycling of the bacterial ribosome. *RNA* **22**: 10–21.
- Bremer H, Dennis PP. 2008. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *Ecosal Plus* **3**. doi: 10.1128/ecosal.5.2.3.
- Dong HJ, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* **260**: 649–663.
- Ehrenberg M, Blomberg C. 1980. Thermodynamic constraints on kinetic proofreading in biosynthetic pathways. *Biophys J* **31**: 333–358.
- Freistroffer DV, Kwiatkowski M, Buckingham RH, Ehrenberg M. 2000. The accuracy of codon recognition by polypeptide release factors. *Proc Natl Acad Sci* **97**: 2046–2051.
- Gout JF, Thomas WK, Smith Z, Okamoto K, Lynch M. 2013. Large-scale detection of in vivo transcription errors. *Proc Natl Acad Sci* **110**: 18584–18589.
- Gromadski KB, Rodnina MV. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* **13**: 191–200.
- Gromadski KB, Daviter T, Rodnina MV. 2006. A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol Cell* **21**: 369–377.
- Imashimizu M, Oshima T, Lubkowska L, Kashlev M. 2013. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res* **41**: 9090–9104.
- Indrisiunaite G, Pavlov MY, Heurgue-Hamard V, Ehrenberg M. 2015. On the pH dependence of class-1 RF-dependent termination of mRNA translation. *J Mol Biol* **427**: 1848–1860.
- Jelenc PC, Kurland CG. 1979. Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proc Natl Acad Sci* **76**: 3174–3178.
- Johansson M, Bouakaz E, Lovmar M, Ehrenberg M. 2008a. The kinetics of ribosomal peptidyl transfer revisited. *Mol Cell* **30**: 589–598.
- Johansson M, Lovmar M, Ehrenberg M. 2008b. Rate and accuracy of bacterial protein synthesis revisited. *Curr Opin Microbiol* **11**: 141–147.
- Johansson M, Jeong KW, Trobro S, Strazewski P, Åqvist J, Pavlov MY, Ehrenberg M. 2011. pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. *Proc Natl Acad Sci* **108**: 79–84.
- Johansson M, Zhang JJ, Ehrenberg M. 2012. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc Natl Acad Sci* **109**: 131–136.
- Kurland CG. 1992. Translational accuracy and the fitness of bacteria. *Annu Rev Genet* **26**: 29–50.
- Manickam N, Nag N, Abbasi A, Patel K, Farabaugh PJ. 2014. Studies of translational misreading in vivo show that the ribosome very efficiently discriminates against most potential errors. *RNA* **20**: 9–15.
- McGuffee SR, Elcock AH. 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol* **6**: e1000694.
- Mellenius H, Ehrenberg M. 2015. DNA template dependent accuracy variation of nucleotide selection in transcription. *PLoS One* **10**: e0119588.
- Ogle JM, Ramakrishnan V. 2005. Structural insights into translational fidelity. *Annu Rev Biochem* **74**: 129–177.
- Parker J. 1989. Errors and alternatives in reading the universal genetic code. *Microbiol Rev* **53**: 273–298.
- Parker J. 1992. Variations in reading the genetic code. In *Transfer RNA in protein synthesis* (ed. Hatfield DK, Lee BJ, Pirtle RM), pp. 191–267. CRC Press.
- Pavlov MY, Zorzet A, Andersson DI, Ehrenberg M. 2011. Activation of initiation factor 2 by ligands and mutations for rapid docking of ribosomal subunits. *EMBO J* **30**: 289–301.
- Ruusala T, Ehrenberg M, Kurland CG. 1982. Is there proofreading during polypeptide synthesis? *EMBO J* **1**: 741–745.
- Thompson RC, Stone PJ. 1977. Proofreading of the codon-anticodon interaction on ribosomes. *Proc Natl Acad Sci* **74**: 198–202.
- Wold F, Ballou CE. 1957. Studies on the enzyme enolase. I. Equilibrium studies. *J Biol Chem* **227**: 301–312.
- Zhang J, Jeong KW, Johansson M, Ehrenberg M. 2015. Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome. *Proc Natl Acad Sci* **112**: 9602–9607.



RNA

A PUBLICATION OF THE RNA SOCIETY

Proofreading neutralizes potential error hotspots in genetic code translation by transfer RNAs

Jingji Zhang, Ka-Weng leong, Harriet Mellenius, et al.

RNA 2016 22: 896-904 originally published online April 18, 2016
Access the most recent version at doi:[10.1261/ma.055632.115](https://doi.org/10.1261/ma.055632.115)

Supplemental Material <http://rnajournal.cshlp.org/content/suppl/2016/04/18/rna.055632.115.DC1.html>

References This article cites 28 articles, 13 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/22/6/896.full.html#ref-list-1>

Open Access Freely available online through the *RNA* Open Access option.

Creative Commons License This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A promotional banner for a webinar. The background is green with a molecular structure. The text reads "Webinar on lncRNAs presented by lead authors" in white. To the right is a dark grey button with "View on demand" in white. Further right is the "EXIQON" logo in white on a green background.

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
