Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome

Jingji Zhang, Ka-Weng Ieong, Magnus Johansson, and Måns Ehrenberg[1]

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

We used a cell-free system with pure *Escherichia coli* components to study initial codon selection of aminoacyl-tRNAs in ternary complex with elongation factor Tu and GTP on messenger RNA-programmed ribosomes. We took advantage of the universal rate-accuracy trade-off for all enzymatic selections to determine how the efficiency of initial codon readings decreased linearly toward zero as the accuracy of discrimination against near-cognate and wobble codon readings increased toward the maximal asymptote, the d value. We report data on the rate-accuracy variation for 7 cognate, 7 wobble, and 56 near-cognate codon readings comprising about 15% of the genetic code. Their d values varied about 400-fold in the 200–80,000 range depending on type of mismatch, mismatch position in the codon, and tRNA isoacceptor type. We identified error hot spots ($d = 200$) for U:G misreading in second and U:U or G:A misreading in third codon position by His-tRNA^{His} and, as also seen in vivo, Glu-tRNA^{Glu}. We suggest that the proofreading mechanism has evolved to attenuate error hot spots in initial selection such as those found here.

protein synthesis | genetic code | misreading | error hot spots | kinetics

The genetic code has 61 sense codons encoding the 20 canonical amino acids and three stop codons encoding termination of peptide elongation. The sense codons in the ORFs of mRNAs are translated on ribosomes by aminoacylated tRNAs (1). Rapid synthesis of the bacterial proteome requires that aminoacyl-tRNAs (aa-tRNAs) in ternary complex (T_3) with elongation factor Tu (EF-Tu) and GTP bind rapidly with large k_{cat}/K_m values to ribosomal aa-tRNA sites (A sites) programmed with cognate codons. High quality of the proteome requires that aa-tRNAs read noncognate codons with small k_{cat}/K_m values so the frequency of amino acid substitution (missense) errors is small (2). This means that high population genetic fitness requires sufficiently small missense error frequency for high proteome quality, yet not so small as to seriously reduce the speed of cognate codon reading by the universal rate-accuracy trade-off for all substrate-selective enzymatic reactions (2–4). The rate-accuracy trade-off prescribes the efficiency (k_{cat}/K_m) of cognate product formation to decrease with increasing accuracy of substrate selection, often in a linear fashion (5). The rate-accuracy trade-off depends on (*i*) the maximal possible discrimination between right and wrong substrate of an enzymatic selection step (the d value) and (*ii*) the fraction, α_d, of the d value that is implemented by the enzyme (2): As α_d increases toward 1 the efficiency of cognate product formation decreases toward zero.

The existence of maximal accuracy limits (d values) in amino acid discrimination by an amino acid-selecting protein was suggested years ago by Linus Pauling (6). He proposed that these d values would be very small for pairs of similar amino acids. For discrimination between valine and isoleucine he estimated a d value of 10, leading to the proposal of high intracellular amino acid substitution error frequency, which turned out not to be true (7, 8). We now know that Pauling greatly underestimated the d value by which the isoleucine-specific aa-tRNA synthetase (IleRS) discriminates against valine (9), but the notion of d-value limits for enzymatic selections has since continued to guide experimental research on replication, transcription, and genetic code translation.

In fact, the discovery of substrate proofreading was inspired by the challenge of how to transcend the accuracy limits postulated by Pauling (6). Here repeated use of the very same

d value can greatly enhance the accuracy of an enzyme (10–13), provided that discarding of noncognate substrate in the proofreading steps is driven by free energy dissipation (10, 14). Another case is Ninio's explanation (15) of error prone (*ram*) and hyperaccurate (*strA*) ribosomal mutants (16) as caused by varying utilization, α_d, of the d values of aa-tRNA selection on the mRNA programmed ribosome. By hypothesis, the *ram* mutations prolong the precatalysis residence times of ribosome-bound cognate and noncognate tRNAs by the same factor, thereby decreasing the fraction α_d of the d value used for codon selection. Furthermore, the *strA* mutations shorten the precatalysis tRNA residence times, thereby increasing the α_d values. The rate-accuracy trade-off would then imply that *strA* ribosomes have low and *ram* ribosomes have high efficiency of tRNA association to the ribosomal A site. It was demonstrated that among bacterial strains with greatly varying ribosomal accuracy phenotype, the wild-type strain grew most rapidly, whereas increasing accuracy above and decreasing accuracy below the wild-type level monotonously decreased the growth rate. It was proposed that the accuracy increase and decrease away from the wild-type level reduced the growth rate by reduced efficiency of protein elongation in the former and reduced quality of the proteome in the latter case (17).

The d value reflects the ability of an enzyme to sense the chemical difference between two substrates, whereas the expressed fraction, α_d, of the d value is tuned by their common shared structural elements. In the case of aa-tRNA selection in protein synthesis the difference between a cognate and a noncognate interaction comes from the codon–anticodon helix and may depend on a single, mismatched base pair. This minute difference between cognate and near-cognate substrates suggests that pushing the d value to ever higher values is evolutionarily compelling (2). It is, we suggest, the difficulty of evolving sufficiently high d values for genetic code translation that forced the evolution of proofreading of aa-tRNAs to

Significance

The genetic code letters, codons, are in all life forms translated to protein building blocks, amino acids, on ribosomes with the help of tRNAs. For 15% of the total genetic code we have clarified how the rates of codon translation decrease as the accuracies of codon reading increase toward their maximal values. Our study is ground-breaking with respect to its large scope, its relevance for genetic code translation in living cells, and the precision of its error estimates. Our data suggest that the quality of the ensembles of proteins in cells is determined by "error hot spots" in code translation, where amino acid substitutions are frequent. Our study is essential for modeling of bacterial physiology and understanding genetic code evolution.
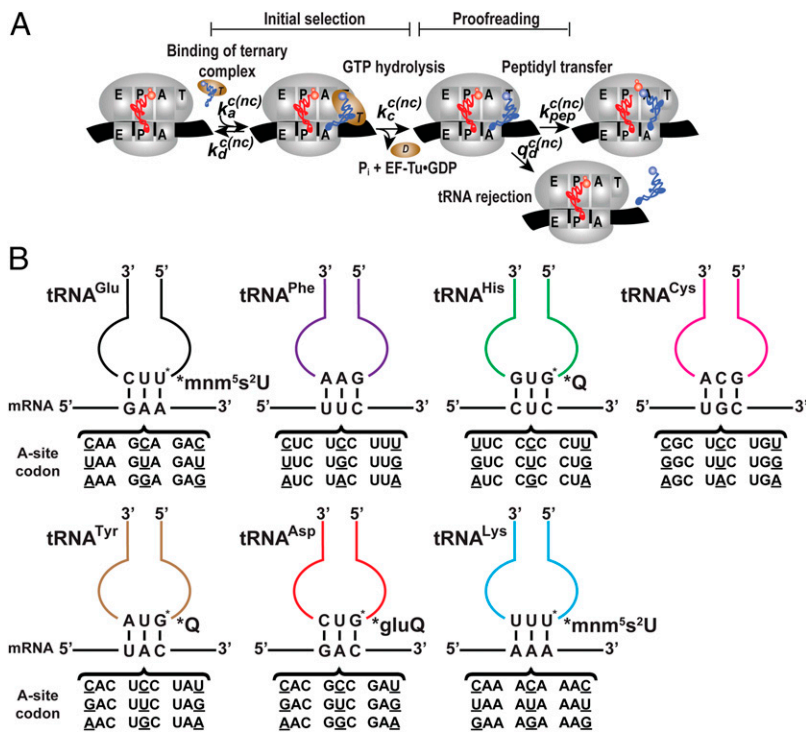
Fig. 1. tRNA selection on the ribosome. (*A*) Kinetic scheme of peptide bond formation on the mRNA programmed ribosome. aa-tRNAs can be rejected during initial selection or at the proofreading step. The two selection steps are separated by hydrolysis of EF-Tu–bound GTP. (*B*) The efficiency-accuracy trade-off in initial selection was evaluated for different aa-tRNAs reading all codons differing from their fully matched codons by one base change. Asterisk indicates modified base, its position in tRNA anticodon, and its chemical nature. tRNA^{Lys} data from ref. 5 are included in the article for comparison.

improve the accuracy of their codon selection. Ribosomal proofreading was first identified experimentally by Thompson and Stone (18), and later by Ruusala et al. (19). Here, substrate discarding in proofreading (Fig. 1*A*, rate constant q_d) is driven by hydrolysis of GTP on EF-Tu, with GTP shifted far above equilibrium with GDP. Accordingly, there is an initial selection of ternary complex, containing aa-tRNA, EF-Tu, and GTP. After GTP hydrolysis on EF-Tu, there is a proofreading step in which noncognate aa-tRNA dissociates with high and cognate aa-tRNA with low probability (Fig. 1*A*).

These brief historical remarks may serve to illustrate the high selection pressure for maximal d values and optimal tuning of the α_d fractions in genetic code translation. In fact, in-depth understanding of the coevolution of ribosomes, tRNAs, and tRNA modifications requires precise knowledge of the large set of d values and their expression levels that phenotypically define the accuracy of genetic code translation.

In this work, we have taken advantage of the linear trade-off between codon reading efficiency and the accuracy of initial codon selection (5) to estimate the d values of initial codon selection by ternary complexes containing six additional aa-tRNAs, in total now covering about 15% of all cognate and near-cognate codon readings during genetic code translation in bacteria. The d values specify the most important physicochemical aspect of the genetic code, namely the ability to rapidly translate the nucleic acid code to peptide sequence at a small error level. When d values are known, the cognate codon reading efficiencies and the near-cognate codon error frequencies can be predicted for any buffer condition in vivo or in vitro. Knowledge of the d values, finally, makes it possible to set the rate-accuracy trade-off in a population genetic context for predictions of, for instance, the optimal accuracy level of genetic code translation for maximal growth (2).

Results

Trade-Off Between Rate and Accuracy in Enzymatic Reactions.
Here we summarize aspects of enzymatic selection that are crucial for the present work (Fig. 1*A*). The efficiency of initial, cognate (c), or noncognate (nc) codon selection, $(k_{cat}/K_m)^{c,nc}$ is the association rate

constant, k_a, multiplied by the probability that ternary complex binding to the ribosome results in GTP hydrolysis:

$$(k_{cat}/K_m)^{c,nc} = k_a \frac{k_c^{c,nc}}{k_c^{c,nc} + k_d^{c,nc}} = \frac{k_a}{1 + k_d^{c,nc}/k_c^{c,nc}}. \quad [1]$$

The accuracy, A, of initial selection is the ratio between the cognate and noncognate k_{cat}/K_m values:

$$A = (k_{cat}/K_m)^c / (k_{cat}/K_m)^{nc} = \frac{1 + k_d^{nc}/k_c^{nc}}{1 + k_d^c/k_c^c}. \quad [2]$$

In a steady-state situation with equal concentrations of cognate and noncognate ternary complex the cognate GDP production rate would be A times that of the noncognate rate. Following Hopfield (12) we define the d values from the rate constants in Fig. 1*A* as

$$d = \frac{k_c^c k_d^{nc}}{k_c^{nc} k_d^c} = e^{\Delta\Delta G/RT}, \quad [3]$$

where $\Delta\Delta G$ is the difference in standard free energy between a noncognate and a cognate transition state for GTP hydrolysis, R is the gas constant, and T is the absolute temperature. With $a = k_d^c/k_c^c$ it follows that

$$A = \frac{1 + da}{1 + a} = \alpha_d d \sim \frac{a}{1 + a} d, \quad [4]$$

where α_d is the fraction of d that corresponds to A. According to Ninio (15), error-prone ribosome (*ram*) has smaller α_d and A whereas hyperaccurate (*strA*) ribosome has larger α_d and A than wild type (16). The universal trade-off between rate and accuracy of enzymatic reactions (2) follows from Eqs. **2** and **3** as (5)

$$(k_{cat}/K_m)^c = k_a \frac{d - A}{d - 1}. \quad [5]$$

When A is varied at constant k_a and d values, $(k_{cat}/K_m)^c$ decreases linearly from its largest value, k_a, toward zero as A increases from its
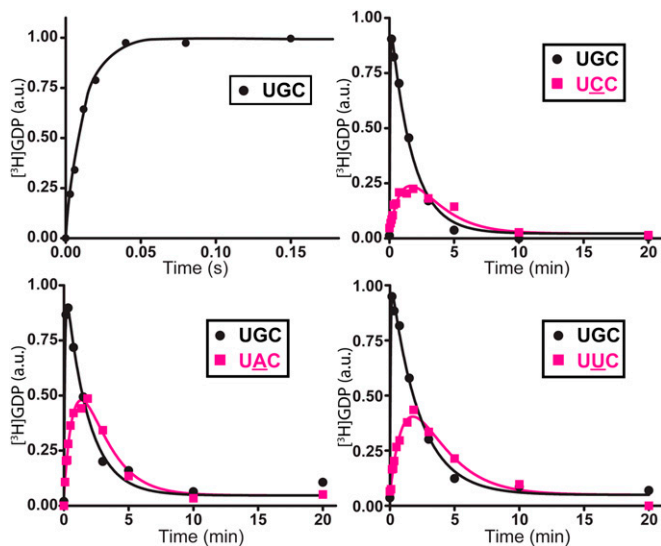
Fig. 2. Measurements of k_{cat}/K_m parameters for GTP hydrolysis during cognate or near-cognate codon reading. Time evolution of the level of [³H]GDP in response to [³H]GTP·EF-Tu·Cys-tRNA^{Cys} binding to 70S ribosomes programmed with cognate codon or near-cognate codons. For the cognate reaction in short time frame (*Top Left*), data were fitted to a single exponential function. Each near-cognate experiment was performed in parallel with a cognate experiment (in black). The very same ternary complex mixture was here used for both cognate and noncognate reactions, and both curves were jointly fitted with sharing of parameters to increase precision of the measurement (*SI Text*). In all experiments ribosomes were in excess over ternary complexes, and k_{cat}/K_m values were calculated from the apparent GTP-hydrolysis rate constant divided by the active ribosome concentration (here, 0.7 μM and 1.8 μM ribosomes were used for cognate and near-cognate reactions, respectively). The decrease in [³H]GDP level in the long time frame is due to spontaneous dissociation of [³H]GDP from EF-Tu followed by its rapid regeneration to [³H]GTP by pyruvate kinase. All experiments were performed in polymix buffer with the addition of 2 mM extra Mg(OAc)$_2$.

smallest value, 1, toward its largest value, *d*. Experimentally, selective variation of *A*, with other parameters unaltered, has been achieved by titrating the concentration of free Mg²⁺ from 1 to 8 mM (5). The *A* values of genetic code translation in vivo must be much smaller than the corresponding *d* values, because otherwise $(k_{cat}/K_m)^c$ would be prohibitively small.

Taking advantage of previously developed assays (5), we estimated the accuracy by which Cys-tRNA^{Cys}, Phe-tRNA^{Phe}, Glu-tRNA^{Glu}, His-tRNA^{His}, Tyr-tRNA^{Tyr}, or Asp-tRNA^{Asp} select their cognate codons compared with all near-cognate codons (Fig. 1*B*). The k_{cat}/K_m values for GTP hydrolysis were obtained from experiments in which preformed aa-tRNA·EF-Tu·[³H]GTP ternary complex was mixed with 70S initiation complex in excess, containing f[³H]Met-tRNA^{fMet} in the P site and an A-site codon cognate or near-cognate to the aa-tRNA in T₃. For cognate reactions quench-flow techniques were used to monitor the time dependence of the [³H]GDP concentration. A typical case is cognate UGC reading by Cys − tRNA$^{Cys}_{GCA}$ containing ternary complex, where the experimental data points in Fig. 2 (filled circles) are fitted to a single exponential function (black line). Most near-cognate reactions were monitored in experiments where reactions were manually started and quenched. Here [³H]GDP formation and [³H]GDP dissociation and rephosphorylation along with [³H]GTP exchange for unlabeled GTP in free T₃ were taken into account (Fig. 2 and *Experimental Procedures*). For high-precision estimates of $(k_{cat}/K_m)^{nc}$, parallel experiments were performed with the same ternary complex reading its cognate codon. The time dependence of [³H]GDP concentration in experiments where Cys − tRNA$^{Cys}_{GCA}$ reads near-cognate codons UCC, UAC or UUC is displayed in Fig. 2. Experimental data points associated with the cognate reading of UGC (filled circles) were fitted to single exponentials (black

lines) estimating the T₃ concentration and the rate of GDP exchange, k_{ex_GDP}, on EF-Tu off the ribosome. The data points associated with the near-cognate reactions (magenta filled squares) were fitted to exponential functions (magenta lines) which took into account the rate of GTP hydrolysis, as determined by $(k_{cat}/K_m)^{nc}$ multiplied by the concentration of active ribosomes, the T₃ concentration, k_{ex_GDP}, and the rate of GTP exchange on the ternary complex, k_{ex_GTP} (*Experimental Procedures*). Accordingly, the cognate and near-cognate [³H]GDP concentration curves had three parameters in common, which made their joint fitting advantageous and led to a relative error of the $(k_{cat}/K_m)^{nc}$ estimates of about 5%.

Rate-Accuracy Trade-Off for 6 Cognate, 6 Wobble, and 48 Near-Cognate Codon Readings. To obtain rate-accuracy trade-off lines for the six ternary complexes of the present study, the original polymix buffer was supplemented with extra Mg²⁺ ions up to 10 mM. This increased the free Mg²⁺ concentration from 1.3 mM [polymix standard condition with 2 mM ATP/GTP and 10 mM phosphoenolpyruvate (PEP)] to about 8 mM (*Experimental Procedures*). Through this, the cognate k_{cat}/K_m values, measured as described in the previous section, increased monotonically toward asymptotes in the 75 and 200 μM⁻¹·s⁻¹ interval (Fig. 3*A*). The near-cognate k_{cat}/K_m values were more sensitive to the free Mg²⁺ concentration, increasing sharply with increasing [Mg²⁺] without reaching plateau values (Fig. 3*B*). The cognate and near-cognate k_{cat}/K_m values were used to construct trade-off lines by plotting
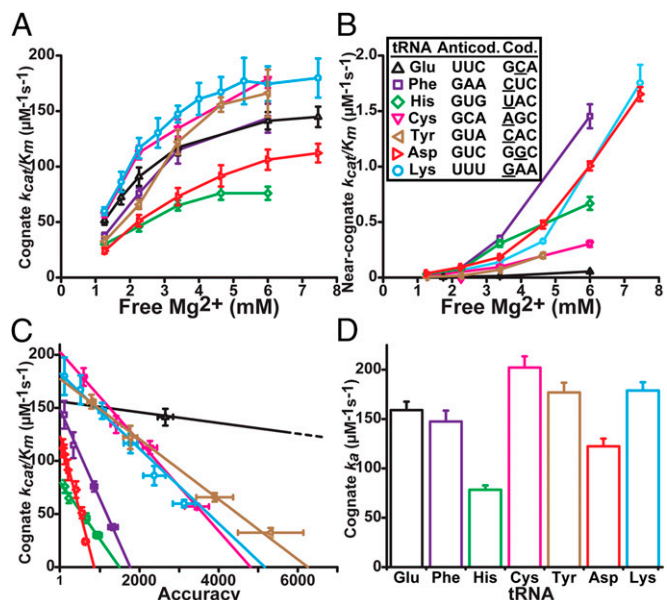

Fig. 3. The rate-accuracy trade-off. (*A*) Efficiency of cognate GTP hydrolysis, $(k_{cat}/K_m)^c$, for different tRNAs (see *B* for symbol legend) reading their fully matched codons at varying Mg²⁺ concentration. (*B*) Efficiency of noncognate GTP hydrolysis, $(k_{cat}/K_m)^{nc}$, for different tRNAs reading single-mismatch codons at varying Mg²⁺ concentration. (*C*) Efficiency of cognate GTP hydrolysis, $(k_{cat}/K_m)^c$, vs. the accuracy [calculated as the ratio $(k_{cat}/K_m)^c/(k_{cat}/K_m)^{nc}$] for different tRNAs reading single-mismatch codons as indicated in *B*. In each tRNA misreading case, cognate and noncognate (k_{cat}/K_m) values were measured at different Mg²⁺ concentrations as shown in *A* and *B*. The *x* intercept gives the maximal accuracy, *d*, for each misreading case, and the *y* = 1 line intercept gives the rate constant for association of each cognate tRNA to the ribosome. The complete rate-accuracy trade-off for Glu-tRNA^{Glu} misreading codon GCA (black triangles and black lines) is plotted in Fig. S1. (*D*) The rate constant, $\bar{k}_a$, for association of different aa-tRNAs in ternary complexes to ribosomes, estimated from the linear dependence between cognate GTP-hydrolysis efficiency and accuracy. Data in *A–C* represent weighted averages from at least two experiments ± propagated SD. Error bars in *D* represent the SD estimates from the parameter fitting procedure, where experimental errors, such as in *A* and *B*, are used as weights (*SI Text*). tRNA$^{Lys}_{UUU}$ data are from ref. 5.
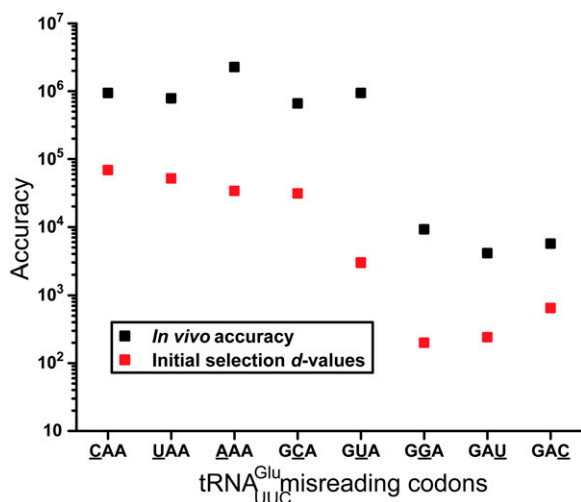
Fig. 4. Comparison between accuracy calculated from in vivo measurement and maximal accuracy of initial selection (d values) for tRNA^{Glu} misreading its near-cognate codons. Accuracy (black squares) is calculated as the inverse of in vivo error frequency measured by Manickam et al. (36). To compare with our in vitro measurements (d values), in vivo error frequency ($f_{in\ vivo}$) for each mistranslated codon was also normalized according to the abundance of tRNA^{Glu} and the competing cognate tRNA (1) or release factors (for U<u>AA</u>) in E. coli, assuming different tRNAs have similar efficiencies (k_{cat}/K_m values) for binding to the ribosome in vivo. $f_{nom} = 1/A = f_{in\ vivo}([tRNA^{Cognate}]/[tRNA^{Glu}])$. d values (red squares) are from Table S1.

cognate k_{cat}/K_m values vs. normalized accuracy values: $A = (k_{cat}/K_m)^c/(k_{cat}/K_m)^{nc}$. This is illustrated for tRNA$^{Cys}_{GCA}$ misreading <u>A</u>GC, tRNA$^{Phe}_{GAA}$ misreading <u>C</u>UC, tRNA$^{Glu}_{UUC}$ misreading <u>G</u>CA, tRNA$^{His}_{GUG}$ misreading U<u>A</u>C, tRNA$^{Tyr}_{GUA}$ misreading <u>C</u>AC, and tRNA$^{Asp}_{GUC}$ misreading G<u>G</u>C (Fig. 3C). For each tRNA, different single-position mismatches resulted in different slopes of the straight lines, whereas the intercepts with the $A = 1$ vertical, k_a, representing the cognate rate constant for association of the specific aa-tRNA-containing ternary complex to the ribosome, was always the same (5). There was, however, a threefold variation in the cognate k_a values among the different tRNAs (Fig. 3D). With one trade-off line per aa-tRNA, we extrapolated lines for other single-position mismatch readings from one or a few additional data point(s) (Experimental Procedures). This allowed us to estimate d values for all possible single-position mismatches in all three positions for all six aa-tRNAs as summarized in Table S1 and shown graphically in Fig. S2, along with the previously obtained tRNA^{Lys} data (5).

Discussion

We previously developed an in vitro approach to study the rate-accuracy trade-off of initial codon selection by Lys-tRNA^{Lys} in complex with EF-Tu and GTP on the mRNA programmed ribosome (5). Here we used the same approach to study the rate-accuracy trade-off for all near-cognate readings in all three codon positions for six additional aa-tRNAs. The rate/accuracy data can in each case be summarized by the association rate constant, k_a, the same for all cognate and near-cognate codon readings by a specific aa-tRNA, and the d value. The k_a value is the maximal efficiency of codon reading (at the cost of no accuracy). The d value is the maximal accuracy (at the cost of zero rate of protein elongation). The d-value dataset in Table S1 and Fig. S2 encompasses about 15% of all cognate, wobble, and near-cognate codon readings in bacterial protein synthesis (1). It was obtained by varying the accuracy, A, by which each cognate codon is read in competition with each member of its set of near-cognate codons. When A was changed by increasing the free Mg^{2+} concentration from about 1 to 8 mM, the cognate codon reading efficiency (k_{cat}/K_m) increased linearly with decreasing A (Fig. 3C). Hence, the d value remained unaltered when the fraction, α_d, of the d value expressed as current

accuracy, A, changed. At low and high [Mg^{2+}] the ribosome is hyperaccurate and error-prone by a mechanism similar to that suggested by Ninio (15) for Gorini's ribosomal *straA* and *ram* mutants (16), respectively.

The dataset in Table S1 displays large variation in accuracy of initial codon selection with type of mismatch, identity of misreading tRNA, and mismatch position in the codon: the d values vary 400-fold in a 200–84,000 range, showing the accuracy of initial codon selection to be far from uniform (20). For any particular type of mismatch, the second codon position has in general the highest d value, whereas the third codon position has the lowest. In most cases, pyrimidine–purine misreadings have the lowest d values except those involving tRNA^{Asp} and tRNA^{Lys} (Table S1). In general, the same types of mismatches display similar d-value trends in different codon positions and for different tRNAs, for example, $d(U:C) > d(U:U) > d(U:G)$. The exception is first-position misreading by tRNA^{Lys}, where $d(U:U) = d(U:G)$ within experimental error.

Two of the lowest d values, 200 and 250, relate to second-position U:G mismatches: tRNA$^{Glu}_{UUC}$ reading G<u>G</u>A and tRNA$^{His}_{GUG}$ reading C<u>G</u>C. Both are associated with matched <u>G</u>:C pairs in the first codon position. When, in contrast, the same type of second-position misreadings are associated with U:A or A:U base pairs in first codon position (compare tRNA^{Glu} to tRNA^{Lys} and tRNA^{His} to tRNA^{Tyr} in Table S1) they have, respectively, 44- and 6-fold higher d values. Furthermore, the third-lowest d value among the second-position mismatches, 900, is associated with a U:G mismatch flanked by G:C pairs on both sides (tRNA^{Asp} reading G<u>G</u>C). Interestingly, a first-position G:C/C:G pair correlates with lower overall d values also in the third codon position (compare tRNA^{Glu} vs. tRNA^{Lys}, and tRNA^{His} and tRNA^{Asp} vs. tRNA^{Cys}, tRNA^{Tyr}, and tRNA^{Phe}), altogether suggesting a common pattern for how first-position G:C/C:G pairs affect the ribosome's ability to discriminate against some second- and third-position mismatches. It is conceivable that G:C/C:G base pairs in the first codon position confer stacking free energies that favor U:G mismatched rather than matched base pairs in the second codon position.

There is yet another twist to the story of U:G mismatches in second codon position. Recent crystal structures at high resolution (~3 Å) were obtained for ribosome complexes with P-site bound tRNA^{fMet}. The A site was filled with tRNA$^{Leu}_{GAG}$ in complex with its cognate CUC or near-cognate UUU codon with a first position G:U mismatch or, alternatively, with tRNA$^{Tyr}_{GUA}$ in complex with its cognate UAC or near-cognate UGC codon with a second-position U:G mismatch. Surprisingly, not only the cognate complexes but also first-position G:U and second-position U:G mismatched pairs were Watson–Crick-like (21, 22). Furthermore, the so-called monitoring 16S rRNA bases A1492, A1493, and G530 were all in their active form (23). There seemed, in other words, to be no structural basis of the type suggested by Ramakrishnan and coworkers (23) for discrimination against U:G mismatches in first or second codon position. We note that there is no direct relevance of these structures for tRNA selection on the ribosome, because accommodated A-site binding of tRNAs is likely not to be part of initial selection of ternary complex or proofreading selection of tRNAs. To explain their data the authors suggested that the Watson–Crick-like base pairing was due to formation of enolG:U interactions (22). As shown by Satpati and Åqvist (24), using molecular simulation tools it is not likely that enolG:U pairing plays a role in first-position misreading events but it may take part in second codon position misreading. They noted that the standard free energy cost of the keto to enol transition for G is high enough to keep even second codon position G:U misreading at a very low level. From these considerations it follows that the low d values associated with G:U mismatches in second codon position could to some extent be explained by enolG:U base pairing. The question of why middle-position U:G mismatches have much smaller d values when associated with tRNA$^{Glu}_{UUC}$ and tRNA$^{His}_{GUG}$ than with tRNA$^{Tyr}_{GUA}$, tRNA$^{Asp}_{GUC}$ and, in particular, tRNA$^{Lys}_{UUU}$ (Table S1), will, however, remain unanswered.

Whatever the reason for these error-prone U:G mismatches in second codon position, their existence may have been the major physiological load that forced the evolution of proofreading in genetic code translation.

Another intriguing finding is that tRNA^{Phe}, tRNA^{Cys}, and tRNA^{Tyr} have very similar d values in their discrimination against third codon position mismatches (G:G and G:A; Table S1). Interestingly, these three tRNAs have the same base pairs (A:U) in the first but not in the second codon position. This may suggest a general decoding pattern such that the base pair in the first codon position greatly affects the level of discrimination against third codon position mismatches. This hypothesis, which may be a determinant of codon use patterns, will need further testing against larger datasets of translational d values.

There is a previous study of the accuracy of genetic code translation of a scope similar to that of the present one, albeit based on a model system of unclear relevance to authentic genetic code translation on the ribosome (25). The binding stabilities (association equilibrium constants) of matching and nonmatching anticodon–anticodon loops of native tRNAs were approximated by the lifetimes of these complexes as estimated off the ribosome with T-jump techniques. There are some obvious differences between their model system and initial codon selection on the ribosome. In the latter case, the tRNAs are in complex with EF-Tu as they enter the A site. Development of anticodon–codon contacts therefore requires that the tRNA bodies adopt bent, noncanonical conformations (26–28), expected to reduce the d values of codon selection (29). In the model system, in contrast, the interacting tRNAs are expected to have their canonical conformations. Furthermore, the ribosome-dependent d values are boosted by the participation of 16S rRNA in the recognition process (23), whereas the model system lacks such an accuracy enhancing feature. Activation of the monitoring bases A1492, A1493, and others increases the d values of ternary complex selection (29) by providing a geometrical component to codon–anticodon recognition (23). Their activation also creates a water-free environment around the codon–anticodon helix that amplifies the standard free energy difference between the presence and absence of a base-to-base H-bond (29). All this being said, there are also some remarkable similarities between our tabulated d values and the lifetime differences observed by Grosjean et al. (25). For instance, G:U or U:G mismatches in codon middle position resulted in lifetimes of intermediate length, from which it was suggested that such mismatches may lead to high translation errors. We arrived at the same conclusion from our d values, and it is possible that careful comparisons between the present dataset and that of Grosjean et al. (25) will facilitate understanding of the evolution of present-day code translation from more primitive systems, reminiscent of the model system studied by Grosjean et al. (25).

Another, yet related, aspect is how the difference between d values for different types of mismatches in a given codon position varies with the tRNA identity. For tRNA$^{Lys}_{UUU}$ misreading ACA, AUA, and AGA the d values are 25,000, 11,000, and 8,800, respectively, displaying a less than threefold variation with type of mismatch. For tRNA$^{Glu}_{UUC}$ misreading GCA, GUA, and GGA the d values are 31,000, 3,000, and 200, respectively, displaying a 150-fold variation with type of second-position mismatch, and there are more examples of such idiosyncrasies in the dataset of Table S1. An obvious candidate also for these large d-value variations are nearest-neighbor interactions of pairs of base pairs in the codon–anticodon helix. The understanding of stacking energies between pairs of base pairs in RNA:RNA and DNA:RNA double helices has increased greatly over recent years (30), but precious little is known about the stacking interactions in codon–anticodon helices on the ribosome.

We also note that third codon position wobble readings (i.e., U:G/G:U) are all less efficient than their completely matched codon counterparts, but the efficiency reduction is still within an order of magnitude ($1.3 < d < 9.0$, Fig. S2, Inset). Understanding the physical chemistry of the tRNA-dependent d-value variations will provide keys to the evolutionary constraints that have led to the present-day design of the tRNAs that translate the genetic code. This may

become possible by molecular computational techniques applied to high-resolution ribosomal structures as revealed by crystallography (31) and, more recently, cryo-electron microscopy (32).

The present finding that cognate association rate constants vary over a threefold range (Fig. 3D) is in line with previous observations of different cognate association rate constants for A-site binding of different aa-tRNAs in the absence of EF-Tu (33). It is likely that this variation pattern will be extended as our knowledge of cognate codon reading increases. Knowledge of the complete cognate codon reading pattern will be particularly important for in vivo modeling of tRNA concentration-dependent rates of cognate ternary complex binding to the ribosomal A site, error frequencies, and the origin of codon use patterns (34).

The accuracy of initial codon selection by a small subset of ternary complexes has previously been addressed with fast kinetics techniques (20, 35). Their experiments were performed at 20 °C, which allowed for estimation of individual rate constants on the pathway to GTP hydrolysis on EF-Tu in ternary complex with Phe-tRNA^{Phe} in response to cognate and a subset of near-cognate codons. One important finding in those studies is that the accuracy of initial selection of ternary complex is based not only on more rapid dissociation, but also on an apparently slower GTP hydrolysis of near-cognate than of cognate ternary complex. Those studies were based on combinations of experimental techniques, such as stopped flow and quench flow, and depended on fluorescence-labeled tRNAs, fluorescent mant-GTP, nonhydrolyzable analogs of GTP, and GTPase-deficient mutants of EF-Tu. The experiments were performed in buffers rendering both high fidelity (HiFi) and low fidelity (LoFi) ternary complex selection (20, 35). A direct comparison (Table S2) can be made between the d values for five near-cognate codon readings by tRNA^{Phe}-containing ternary complex (Table S1) and the corresponding d values calculated from tables 1 (HiFi) and 2 (LoFi) in ref. 20. A significant difference between the datasets is that their d values are about two orders of magnitude larger than ours. One reason for this discrepancy could be a strongly negative correlation between temperature and maximal accuracy of codon selection by ternary complex. In addition, their d-value datasets obtained under HiFi and LoFi conditions display large relative variations (Table S2), which suggests uncertain parameter estimates.

In the present dataset, tRNA^{Glu} and tRNA^{His} stand out as severe misreaders of mismatches in second or third positions of their near-cognate codons. That is, tRNA$^{Glu}_{UUC}$ misreads GGA (Gly) with d value 200, GAU (Asp) with d value 240, and GAC (Asp) with d value 650, whereas tRNA$^{His}_{GUG}$ misreads CGC (Arg) with d value 250 and CAA (Gln) with d value 200 (Table S1). Concerning the misreading profile of tRNA^{Glu}, in vivo accuracy data, based on the residual activity of a beta-galactosidase mutant with its Glu codon at position 537 altered to neighboring codons have been obtained (Fig. 4) (36). The enzymatic activity for these mutants was reduced by many orders of magnitude by the insertion of amino acids other than Glu at this position, but residual activity was restored by mistranslation of these codons by tRNA^{Glu}. This assay did not measure the accuracy of initial codon selection of tRNA^{Glu}, but the total accuracy by which in each case the tRNA cognate to the mutated Glu codon competed with tRNA^{Glu}. It is, however, reassuring for the relevance of our biochemistry that this authentic in vivo assay system identified the very same error hot spots, GGA, GAU, and GAC, as found here for the misreading of near-cognate codons by tRNA^{Glu} (Fig. 4). It follows from the figure that the d values for initial selection of Glu instead of Gly and Asp are about two orders of magnitude too small to be compatible with the in vivo data from Manickam et al. (36). The difference, we suggest, is eliminated by the accuracy amplification provided by proofreading of tRNA after hydrolysis of GTP on EF-Tu. Indeed, we suggest that the existence of pronounced hot spots for initial selection errors made by tRNA^{Glu}, tRNA^{His}, and perhaps other tRNAs yet to be found rationalizes the evolution of proofreading selection of aa-tRNAs. The argument here is that translation errors in the percentage range will reduce the functional quality of the proteome to the extent that it will pay off to invest in a

proofreading mechanism that slows down peptide bond formation and costs energy (2).

In summary, we have found that the maximal accuracy (d value) of near-cognate codon selection for a significant subset of ternary complexes displays a 400-fold variation in the 200–80,000 range. Owing to the linear dependence of efficiency vs. accuracy this finding also suggests a 400-fold current accuracy variation irrespective of where on the trade-off line the translation machinery is actually working in vivo (Fig. 3C). We have identified initial selection error hot spots for tRNA^{Glu} and tRNA^{His} by their misreading of U:G in second and U:U or G:A in third codon position with d values equal to 200. We have found perfect correspondence between the tRNA^{Glu}-caused error hot spots found here and those identified in living *Escherichia coli* bacteria, bearing witness to the physiological relevance of the present dataset. From our results we suggest that proofreading has evolved in bacterial protein synthesis to minimize damage on the bacterial proteome by amino acid substitution errors due to initial tRNA selection error hot spots as identified here and maybe others that remain to be found from larger datasets. We are optimistic that the present dataset and its future extensions will serve as an inspiration and testing ground for theoretical approaches to explain the accuracy of codon reading and its idiosyncratic variation with tRNA type and codon context. We do hope that the linear trade-off lines for efficiency and accuracy of genetic code translation will further our understanding of its determinants in the living cell.

Experimental Procedures

Reagents and Buffer Conditions. Purified translation components were prepared as described previously (ref. 37 and references therein). Native tRNA^{Glu} and tRNA^{Phe} were from Chemical Block; tRNA^{Tyr} was from Sigma-Aldrich; and total tRNA, used for measurement of tRNA^{His}, tRNA^{Cys}, and tRNA^{Asp}, was from Roche.

[^{3}H]Met and [^{3}H]GTP were from Perkin-Elmer or Biotrend. Other chemicals were either from Sigma-Aldrich or Merck. All experiments were performed at 37 °C in polymix buffer containing 95 mM KCl, 5 mM NH$_4$Cl, 0.5 mM CaCl$_2$, 8 mM putrescine, 1 mM spermidine, 5 mM potassium phosphate, 1 mM DTE, and 5 mM Mg(OAc)$_2$ and supplemented with energy regeneration components containing 1 mM ATP + 1 mM GTP for the ribosome mixture or 2 mM ATP for the ternary complex mixture, 10 mM PEP, 50 μg/mL pyruvate kinase, and 2 μg/mL myokinase. Extra Mg(OAc)$_2$ was added to adjust the free Mg^{2+} concentration in the reaction. Free Mg^{2+} varies from 1.3 mM to 7.4 mM with addition of 0–10 mM extra Mg(OAc)$_2$ assuming that one ATP or GTP molecule chelates one Mg^{2+} and PEP chelates Mg^{2+} with a K_d value of 6 mM (38).

Measurement of k_{cat}/K_m for Cognate and Near-Cognate GTP Hydrolysis. Ribosome and ternary complex mixtures were prepared as described in ref. 5 except here [^{3}H]GTP instead of [^{3}H]GDP was used for preparing ternary complexes. Ribosome mixture contained 1 μM 70S ribosomes, 1.5 μM f[^{3}H]Met-tRNA^{fMet}, 1.5 μM mRNA, 1.5 μM IF1, 0.5 μM IF2, and 1.5 μM IF3. Ternary complex mixture contained 5 μM tRNAs, 0.5 μM EF-Tu, 0.5μM [^{3}H]GTP, 0.2 mM amino acid, and 1.5 units/μL aa-tRNA synthetase.

Cognate reactions and fast near-cognate reactions (e.g., tRNA^{Glu} misreading GA<u>U</u>) were measured in a quench-flow instrument (RQF-3; KinTek Corp.). Most near-cognate GTP hydrolysis reactions were very slow and experiments were performed manually; also, a cognate reaction was performed in parallel to the near-cognate reaction to increase precision of the measurement as described in ref. 5.

To estimate active ribosome concentration, a ternary complex mixture was prepared with the corresponding cognate tRNA with the addition of EF-Ts [or with corresponding release factor (3 μM) when stop codons were used] and reacted with the ribosome complex in equal volumes as described in ref. 5.

1. Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol* 260(5):649–663.
2. Ehrenberg M, Kurland CG (1984) Costs of accuracy determined by a maximal growth rate constraint. *Q Rev Biophys* 17(1):45–82.
3. Johansson M, Lovmar M, Ehrenberg M (2008) Rate and accuracy of bacterial protein synthesis revisited. *Curr Opin Microbiol* 11(2):141–147.
4. Lovmar M, Ehrenberg M (2006) Rate, accuracy and cost of ribosomes in bacterial cells. *Biochimie* 88(8):951–961.
5. Johansson M, Zhang J, Ehrenberg M (2012) Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc Natl Acad Sci USA* 109(1):131–136.
6. Pauling L (1957) *The Probability of Errors in the Process of Synthesis of Protein Molecules* (Birkhäuse, Basel), pp 597–602.
7. Loftfield RB (1963) The frequency of errors in protein biosynthesis. *Biochem J* 89:82–92.
8. Parker J (1989) Errors and alternatives in reading the universal genetic code. *Microbiol Rev* 53(3):273–298.
9. Fersht AR (1977) Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry* 16(5):1025–1030.
10. Ehrenberg M, Blomberg C (1980) Thermodynamic constraints on kinetic proofreading in biosynthetic pathways. *Biophys J* 31(3):333–358.
11. Freter RR, Savageau MA (1980) Proofreading systems of multiple stages for improved accuracy of biological discrimination. *J Theor Biol* 85(1):99–123.
12. Hopfield JJ (1974) Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci USA* 71(10):4135–4139.
13. Ninio J (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57(5):587–595.
14. Kurland CG (1978) The role of guanine nucleotides in protein biosynthesis. *Biophys J* 22(3):373–392.
15. Ninio J (1974) A semi-quantitative treatment of missense and nonsense suppression in the strA and ram ribosomal mutants of Escherichia coli. Evaluation of some molecular parameters of translation in vivo. *J Mol Biol* 84(2):297–313.
16. Gorini L (1971) Ribosomal discrimination of tRNAs. *Nat New Biol* 234(52):261–264.
17. Tubulekas I, Hughes D (1993) Suppression of rpsL phenotypes by tuf mutations reveals a unique relationship between translation elongation and growth rate. *Mol Microbiol* 7(2):275–284.
18. Thompson RC, Stone PJ (1977) Proofreading of the codon-anticodon interaction on ribosomes. *Proc Natl Acad Sci USA* 74(1):198–202.
19. Ruusala T, Ehrenberg M, Kurland CG (1982) Is there proofreading during polypeptide synthesis? *EMBO J* 1(6):741–745.
20. Gromadski KB, Daviter T, Rodnina MV (2006) A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol Cell* 21(3):369–377.
21. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G (2012) A new understanding of the decoding principle on the ribosome. *Nature* 484(7393):256–259.
22. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G (2013) New structural insights into the decoding mechanism: Translation infidelity via a G·U pair with Watson-Crick geometry. *FEBS Lett* 587(13):1848–1857.
23. Ogle JM, et al. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292(5518):897–902.
24. Satpati P, Åqvist J (2014) Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res* 42(20):12876–12884.
25. Grosjean HJ, de Henau S, Crothers DM (1978) On the physical basis for ambiguity in genetic coding interactions. *Proc Natl Acad Sci USA* 75(2):610–614.
26. Schmeing TM, Voorhees RM, Kelley AC, Ramakrishnan V (2011) How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* 18(4):432–436.
27. Smith D, Yarus M (1989) tRNA-tRNA interactions within cellular ribosomes. *Proc Natl Acad Sci USA* 86(12):4397–4401.
28. Smith D, Yarus M (1989) Transfer RNA structure and coding specificity. II. A D-arm tertiary interaction that restricts coding range. *J Mol Biol* 206(3):503–511.
29. Satpati P, Bauer P, Aqvist J (2014) Energetic tuning by tRNA modifications ensures correct decoding of isoleucine and methionine on the ribosome. *Chemistry* 20(33):10271–10275.
30. Mellenius H, Ehrenberg M (2015) DNA template dependent accuracy variation of nucleotide selection in transcription. *PLoS One* 10(3):e0119588.
31. Selmer M, et al. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313(5795):1935–1942.
32. Hashem Y, et al. (2013) High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. *Nature* 494(7437):385–389.
33. Fahlman RP, Dale T, Uhlenbeck OC (2004) Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol Cell* 16(5):799–805.
34. Quax TE, et al. (2013) Differential translation tunes uneven production of operon-encoded proteins. *Cell Reports* 4(5):938–944.
35. Gromadski KB, Rodnina MV (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* 13(2):191–200.
36. 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(1):9–15.
37. Johansson M, Bouakaz E, Lovmar M, Ehrenberg M (2008) The kinetics of ribosomal peptidyl transfer revisited. *Mol Cell* 30(5):589–598.
38. Wold F, Ballou CE (1957) Studies on the enzyme enolase. I. Equilibrium studies. *J Biol Chem* 227(1):301–312.

BIOCHEMISTRY