

# Is the sequence-specific binding of aminoacyl-tRNAs by EF-Tu universal among bacteria?

Jared M. Schrader and Olke C. Uhlenbeck\*

Department of Molecular Biosciences, Northwestern University, Hogan 2-100, Evanston, IL 60208, USA

Received June 23, 2011; Revised July 19, 2011; Accepted July 21, 2011

## ABSTRACT

Three base pairs in the T-stem are primarily responsible for the sequence-specific interaction of tRNA with *Escherichia coli* and *Thermus thermophilus* EF-Tu. While the amino acids on the surface of EF-Tu that contact aminoacyl-tRNA (aa-tRNA) are highly conserved among bacteria, the T-stem sequences of individual tRNA are variable, making it unclear whether or not this protein–nucleic acid interaction is also sequence specific in other bacteria. We propose and validate a thermodynamic model that predicts the  $\Delta G^\circ$  of any tRNA to EF-Tu using the sequence of its three T-stem base pairs. Despite dramatic differences in T-stem sequences, the predicted  $\Delta G^\circ$  values for the majority of tRNA classes are similar in all bacteria and closely match the  $\Delta G^\circ$  values determined for *E. coli* tRNAs. Each individual tRNA class has evolved to have a characteristic  $\Delta G^\circ$  value to EF-Tu, but different T-stem sequences are used to achieve this  $\Delta G^\circ$  value in different bacteria. Thus, the compensatory relationship between the affinity of the tRNA body and the affinity of the esterified amino acid is universal among bacteria. Additionally, we predict and validate a small number of aa-tRNAs that bind more weakly to EF-Tu than expected and thus are candidates for acting as activated amino acid donors in processes outside of translation.

## INTRODUCTION

The complex of elongation factor Tu (EF-Tu) and GTP binds aminoacyl-tRNA (aa-tRNA), and the resulting ternary complex binds ribosomes and participates in a multistep decoding pathway (1–3). While all 43 bacterial elongator aa-tRNAs appear to bind EF-Tu•GTP with similar affinities (4–7), studies with misacylated tRNAs indicate that this uniformity is the result of offsetting variable contributions of the esterified amino acid and tRNA body to the total binding affinity (8–10). The side

chain of the amino acid fits into a large asymmetric pocket in EF-Tu, and the different amino acids contribute up to 2.8 kcal/mol to the total  $\Delta G^\circ$  of the  $\sim -10$  kcal/mol that is observed for a typical aa-tRNA (9,11). Protein and tRNA mutagenesis experiments established that while many of the contacts between EF-Tu and tRNA contribute to  $\Delta G^\circ$ , the interaction of three protein residues with three adjacent base pairs in the T-stem is primarily responsible for the 3.6 kcal/mol range of  $\Delta G^\circ$  contributed by the different tRNA bodies (12–16). Subsequent experiments analyzing the kinetics of decoding of tRNAs engineered to have different  $\Delta G^\circ$  values established that tight-binding aa-tRNAs release from EF-Tu•GDP more slowly and thereby lower the rate of peptide bond formation (17). This suggests that the observed uniform  $\Delta G^\circ$  of all aa-tRNAs binding to EF-Tu is the result of an evolutionary optimization by two opposing selective pressures. The T-stem sequence of each tRNA evolved to be tight enough to initially bind to EF-Tu, but not too tight to limit its rate of release from EF-Tu•GDP during ribosomal decoding.

The above principles guiding the sequence specificity of the EF-Tu•aa-tRNA interaction were deduced from experiments performed using either *Escherichia coli* or *Thermus thermophilus* EF-Tu and mutations made in four different tRNA bodies. In this article, we explore whether these principles can be extended to all of the tRNAs in all bacterial species. Thus, do EF-Tu proteins from other bacteria recognize their corresponding tRNAs using the same sequence-specific interactions with the three T-stem base pairs? If this is largely true, exceptions could lead to refinements of the recognition model or to the identification of individual tRNAs that do not bind EF-Tu. If this is not true for all bacteria, it may be possible to deduce when the sequence-specific recognition rules emerged in bacterial evolution.

## MATERIALS AND METHODS

### Sequence analysis and $\Delta G^\circ$ prediction

A total of 247 reviewed non-redundant (>90% identical) bacterial EF-Tu sequences were downloaded from Uniprot (<http://www.uniprot.org>) and aligned using

\*To whom correspondence should be addressed. Tel: +1 847 491 5139; Fax: +1 847 491 5444; Email: o-uhlenbeck@northwestern.edu

clustal X (18). A total of 37 706 bacterial tRNA sequences from 629 bacterial species were downloaded from a genomic tRNA database <http://gtrnadb.ucsc.edu/> (19). Separated tRNA<sub>CAU</sub> classes were downloaded from (20) while tRNA<sub>UCA</sub> classes were separated using Tfam 1.3 <http://tfam.lcb.uu.se/> (21). tRNAs in each anticodon class were subjected to a multiple sequence alignment using ClustalX. All sequences which were >94% identical were then removed using Jalview (22), yielding a total of 6113 non-redundant tRNAs. The major anticodon classes of tRNAs were then further analyzed to predict their  $\Delta G^\circ$  to EF-Tu.

$\Delta\Delta G^\circ$  values for single base pair mutations in the T-stems of three *E. coli* tRNAs were calculated relative to the sequence of yeast tRNA<sup>Phe</sup> (15). For each tRNA position (49–65, 50–64 and 51–63), the single base pair  $\Delta\Delta G^\circ$ s were then averaged among the three *E. coli* tRNAs yielding  $\Delta\Delta G^\circ_{\text{AVG}}$  (Table 2). To calculate the affinity of each bacterial tRNA, the  $\Delta G^\circ$  of yeast Phe-tRNA<sup>Phe</sup> was added with  $\Delta\Delta G^\circ_{\text{AVG}}$  (49–65),  $\Delta\Delta G^\circ_{\text{AVG}}$  (50–64) and  $\Delta\Delta G^\circ_{\text{AVG}}$  (51–63) based upon the sequence of each tRNA. The  $\Delta\Delta G^\circ_{\text{Predicted}}$  values were combined into 0.5 kcal/mol bins (23), and the number of tRNAs in each bin were fit to a Gaussian curve using Kaleidagraph (Synergy Software).

## Materials

Expression and purification of *T. thermophilus* EF-Tu, Yeast PheRS and *E. coli* PheRS were performed as described previously (9,12,15). tRNAs genes were chemically synthesized (IDT) in two fragments with at least 20 overlapping base pairs and extended using Taq DNA polymerase. tRNAs were transcribed by T7 RNA polymerase, gel purified and aminoacylated with [<sup>3</sup>H]-Phe as previously described (15).

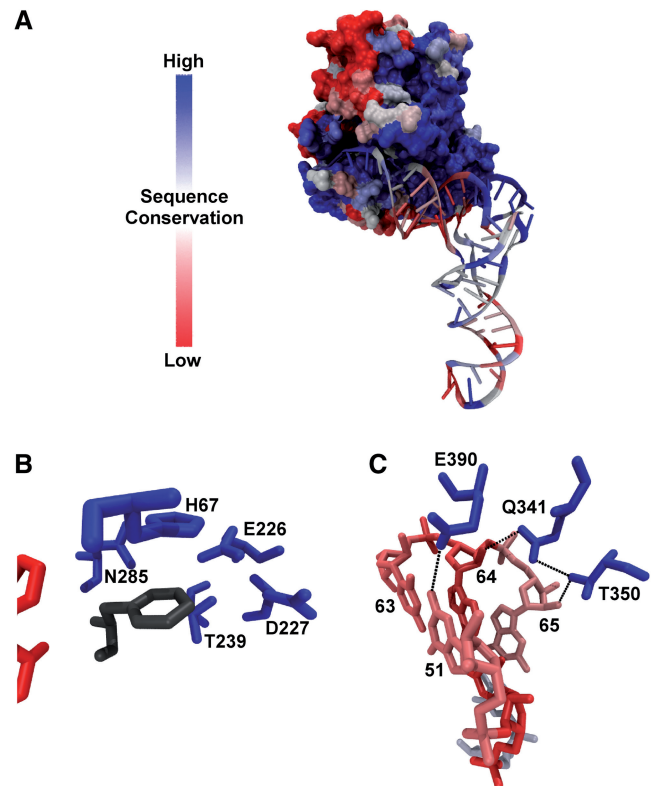
## EF-Tu-binding assay

The dissociation rate ( $k_{\text{off}}$ ) from EF-Tu•GTP was determined on ice in Buffer A (50 mM HEPES pH 7.0, 20 mM MgCl<sub>2</sub>, 0.5 M NH<sub>4</sub>Cl, 5 mM DTT, 20  $\mu$ M GTP, 3 mM phosphoenolpyruvate and 50  $\mu$ g/ml of pyruvate kinase) using a modified ribonuclease protection assay as described (15).  $K_D$  was calculated using a previously determined  $k_{\text{on}}$  of  $1.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (15).  $\Delta G^\circ$  was calculated using  $\Delta G^\circ = -RT \times \ln(K_D)$ . All measurements were performed in triplicate.

## RESULTS

### Sequence conservation in the interface between EF-Tu and aa-tRNA

Mutagenesis of *E. coli* and *T. thermophilus* EF-Tu have identified 15 amino acids in the interface with aa-tRNA that contribute to the  $\Delta G^\circ$  of binding, including 3 in the amino acid-binding pocket and 12 in the region that binds the tRNA (3,12–14,24) (S. J. Chapman, E. Y. Yikilmaz personal communication). In cases where the same mutation has been made in both proteins, the effects are very similar. Alignment of 247 bacterial EF-Tu sequences



**Figure 1.** Sequence conservation in the interface between EF-Tu and aa-tRNA. Percent amino acid identity among bacterial EF-Tu sequences mapped onto the structure of yeast Phe-tRNA<sup>Phe</sup> bound to *T. aquaticus* EF-Tu (11). Sequence conservation on tRNA using a bits scale (56) is similarly mapped. (A) Global structure of the ternary complex. (B) The esterified amino acid binding pocket with esterified phenylalanine is shown in grey. (C) The T-stem recognition sites.

indicate that 13 of these 15 thermodynamically important amino acids are universally (>99%) conserved and the remaining two are very (>80%) conserved but have substitutions of chemically similar amino acids in some species. When mapped on the structure of the ternary complex (Figure 1A), the extreme conservation of the part of the protein that contacts aa-tRNA is evident, including the binding pocket for the esterified amino acid (Figure 1B) and the sequence-specific recognition sites in the T-stem (Figure 1C). This high degree of conservation suggests that all bacterial EF-Tus may recognize aa-tRNAs in the same way.

In contrast to EF-Tu, the sequences of the acceptor and T helices of tRNA which contact EF-Tu are highly variable among bacteria (Figure 1A). In particular, the three T-stem base pairs that are primarily responsible for sequence-specific binding are not very conserved (Figure 1C) (15). Using a genomic tRNA database (19) curated to remove duplicate sequences, 251 different combinations of the six residues were found in 6113 bacterial tRNAs. As summarized in Table 1, this variability in T-stem sequence is also present in each of the 45 individual tRNA classes defined by their anticodon sequence. At one extreme, the 167 bacterial tRNA<sub>UGU</sub><sup>Thr</sup> (tRNA<sub>anticodon</sub><sup>amino acid</sup>)

**Table 1.** Predicted  $\Delta G^\circ$  values of each tRNA anticodon class

tRNA anticodon class	Number of tRNAs	Number of 3-bp combinations	Measured $\Delta G^\circ$ Phe-tRNA <sup>X</sup> (kcal/mol) <sup>d</sup>	Mean Calculated $\Delta G^\circ$ Phe-tRNA <sup>X</sup> (kcal/mol)	$\sigma$	$R$
Ala CGC	58	20		-10.3	0.50	0.999
Ala GGC	92	23	-11.0	-10.0	0.53	0.996
Ala UGC	93	25		-10.0	0.52	0.998
Arg ACG <sup>a</sup>	149	31	-9.6	-9.4	0.32	0.990
Arg CCG <sup>a</sup>	161	45		-9.5	0.58	0.974
Arg CCU	207	51		-9.6	0.59	0.992
Arg UCU	226	41		-9.3	0.49	0.983
Arg UCG <sup>a</sup>	66	23		-9.4	0.32	0.987
Asn GUU	160	23	-9.4	-9.4	0.60	0.998
Asp GUC	75	16	-11.5	-11.0	0.33	0.988
Cys GCA	175	30	-10.1	-9.9	0.62	0.992
Gln CUG	83	32	-8.8	-8.9	0.50	0.984
Gln UUG	88	32		-9.1	0.56	0.997
Glu CUC	70	17		-10.3	0.50	0.998
Glu UUC	117	21	-12.2	-10.4	0.48	0.997
Gly CCC	78	10		-11.0	0.34	0.991
Gly GCC	76	13	-11.2	-11.0	0.39	0.980
Gly UCC	99	19		-10.8	0.68	0.981
His GUG	180	22		-10.7	0.56	0.997
Ile GAU	93	34	-9.2	-9.4	0.36	0.998
Leu CAA	237	26		-10.0	0.50	0.998
Leu CAG	146	22	-10.0	-9.9	0.51	0.981
Leu GAG	152	23		-10.4	0.60	0.989
Leu UAA	232	26		-10.3	0.43	1.000
Leu UAG	194	20		-10.2	0.49	0.941
Lys CUU	117	30		-9.4	0.86	0.989
Lys UUU	151	35	-10.0	-9.7	0.76	0.980
Met/fMet/Ile CAU	320	69		-9.5	0.94	0.970
Met CAU <sup>b</sup>	89	34	-9.9	-9.9	0.85	0.926
fMet CAU <sup>b</sup>	43	24	-8.9	-9.1	0.55	0.973
Ile CAU <sup>b</sup>	101	31		-9.3	0.48	0.984
Phe GAA <sup>a</sup>	118	34	-9.9	-9.4	0.82	0.943
Pro CGG	83	20		-9.8	0.36	0.993
Pro GGG	86	24		-9.8	0.56	0.995
Pro UGG <sup>a</sup>	141	25	-9.9	-9.6	0.54	0.994
Sec/Trp UCA	63	36		-8.2	1.2	0.706
Sec UCA <sup>c</sup>	56	15	-8.0	-8.2	1.2	0.706
Trp UCA <sup>c</sup>	7	6		-9.6	0.75	0.895
Ser CGA	162	20		-10.2	0.35	0.993
Ser GCU	176	20		-10.2	0.37	1.000
Ser GGA	170	17		-10.3	0.32	1.000
Ser UGA <sup>a</sup>	218	20	-9.6	-10.1	0.32	0.923
Thr CGU	148	48	-11.0	-10.0	0.67	0.995
Thr GGU	158	41		-10.1	0.52	0.999
Thr UGU	167	53		-9.8	0.57	0.982
Trp CCA <sup>a</sup>	154	37	-9.0	-9.4	0.40	0.995
Tyr GUA	139	31	-8.6	-9.2	0.38	0.997
Val CAC	64	25		-9.5	0.45	0.996
Val GAC	137	33		-9.7	0.41	0.998
Val UAC <sup>a</sup>	141	28	-9.3	-9.4	0.49	0.925

<sup>a</sup>Uses experimentally derived value for tRNAs containing G49C65 G50U64 G51C63.<sup>b</sup>tRNAs curated in (20).<sup>c</sup>tRNAs curated using tFAM 1.3 (21).<sup>d</sup>Data from (8).

contain 53 different six-base sequences, and at the other, the 78 tRNA<sup>Gly</sup><sub>CCC</sub> contain 10 different six-base sequences. This variability in T-stem sequences among bacterial tRNAs may mean that the rules for tRNA recognition are not universal among bacteria.

#### A thermodynamic model for predicting $\Delta G^\circ$ for any tRNA

To avoid making and assaying all of the many different T-stem sequences in the database, we took advantage of a

thermodynamic model to predict the affinity of each variant from a more limited set of data. Experiments measuring the binding affinities of single base pair mutations in the T-stems of three different tRNAs to either *E. coli* or *T. thermophilus* EF-Tu have been performed (15,17,25). Table 2 presents the  $\Delta\Delta G^\circ$  values for the different single base pair substitutions of the 49–65, 50–64 and 51–63 bp in the three tRNA backgrounds calculated using the sequence of yeast tRNA<sup>Phe</sup> as a reference.

**Table 2.**  $\Delta\Delta G^\circ$  values of single base pair mutations in three tRNAs

	tRNA <sup>Phe</sup> $\Delta\Delta G^\circ$ <sup>a</sup> (kcal/mol)	tRNA <sup>Leu</sup> $\Delta\Delta G^\circ$ <sup>ab</sup> (kcal/mol)	tRNA <sup>Thr</sup> $\Delta\Delta G^\circ$ <sup>c</sup> (kcal/mol)	$\Delta\Delta G^\circ_{AVG}$ (kcal/mol)
<b>49–65 CG</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>
49–65 AU	–0.5	–0.3	–0.8	–0.5
49–65 GC	–0.4	–0.2	–0.7	–0.4
49–65 UA	–0.1	N.D.	–0.2	–0.2
49–65 GU	–0.8	–0.9	–1.1	–0.9
<b>50–64 UA</b>	<b>0.0</b>	<b>0.0</b>	<b>N.D.</b>	<b>0.0</b>
50–64 AU	–0.1	0.1	N.D.	0.0
50–64 GC	–0.2	0.3	N.D.	0.0
50–64 CG	–0.5	0.1	N.D.	–0.2
50–64 GU	1.3	1.4	N.D.	1.4
50–64 UG	0.3	0.5	N.D.	0.4
<b>51–63 GC</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>
51–63 AU	0.8	1.4	1.0	1.1
51–63 CG	0.5	0.6	0.4	0.5
51–63 UA	0.9	1.0	1.2	1.0
51–63 GU	1.0	1.1	0.2	0.8
51–63 UG	1.0	N.D.	N.D.	1.0
51–63 AC	0.2	N.D.	N.D.	0.2

$\Delta\Delta G^\circ$  values are calculated relative to the sequence of tRNA<sup>Phe</sup> designated in bold.

<sup>a</sup>Data from (15).

<sup>b</sup>Data from (14,15).

<sup>c</sup>Data from (25).

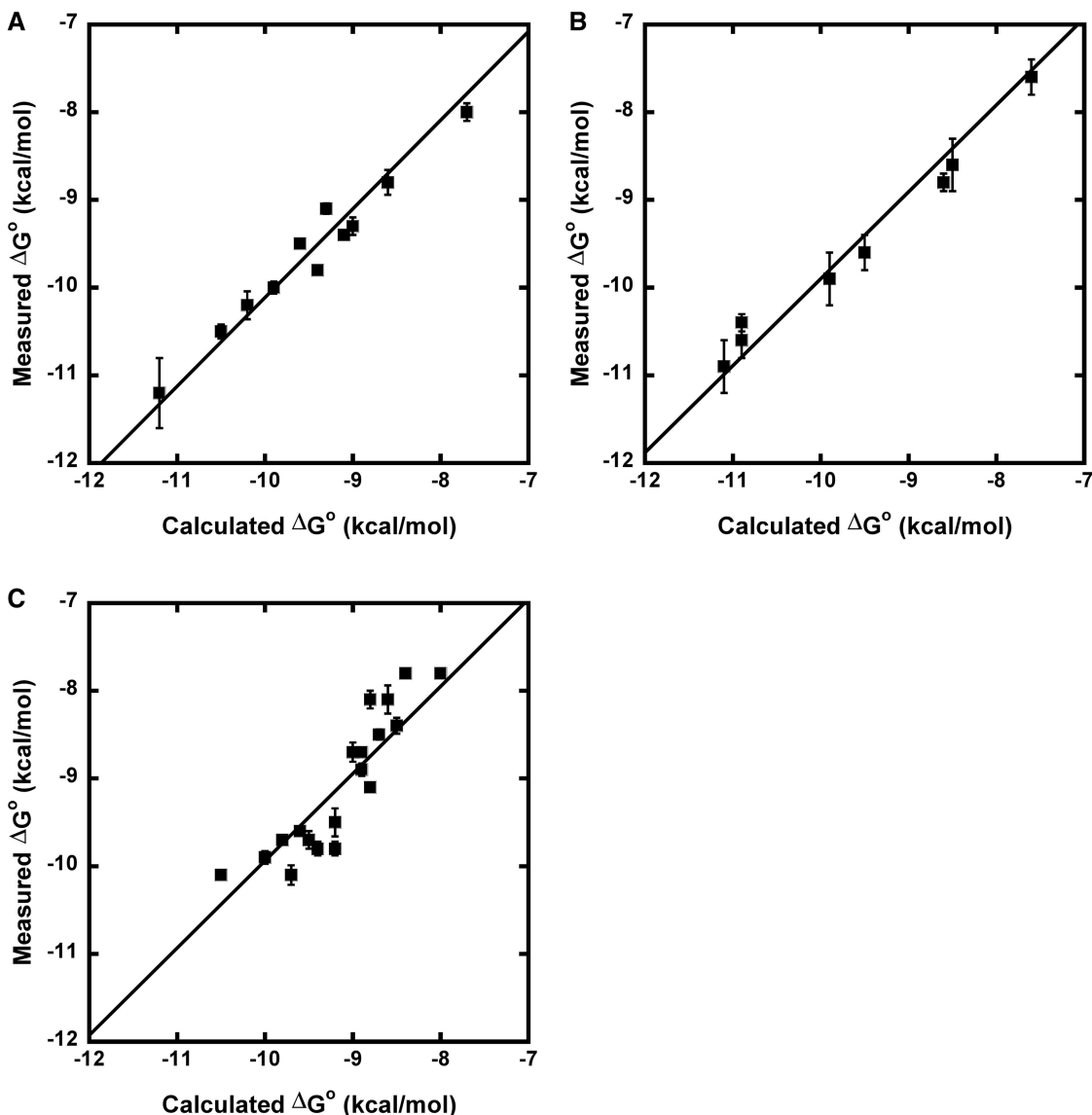
With the single exception of the G51–U63 pair in Thr-tRNA<sup>Thr</sup>, the  $\Delta\Delta G^\circ$  values agree closely for the three data sets, despite the fact that different EF-Tu proteins were used. This argues that the sequence dependence of tRNA binding is very similar in different tRNA bodies. These  $\Delta\Delta G^\circ$  values for each individual tRNA mutation were averaged to yield  $\Delta\Delta G^\circ_{AVG}$  values which describe the change in free energy for any given single base pair mutation. Depending on the sequence, the  $K_D$  of a tRNA can be affected by as much as 26-fold at each position.

The  $\Delta\Delta G^\circ_{AVG}$  values derived from single base pair substitutions were usually able to accurately predict the  $\Delta G^\circ$  of multiple base pair substitutions. Figure 2 compares the measured  $\Delta G^\circ$  of multiple base pair substitutions in either Phe-tRNA<sup>Phe</sup>, Thr-tRNA<sup>Thr</sup>, or Val-tRNA<sup>Val</sup> with the values of  $\Delta G^\circ$  calculated by adding the  $\Delta G^\circ$  of the wild-type tRNA to the appropriate  $\Delta\Delta G^\circ_{AVG}$  values from Table 2. In each case, the majority of the experimental  $\Delta G^\circ$  values fit the predicted values within experimental error, yielding lines with unitary slope. While there are several discrepancies between the experimental and predicted values, the vast majority are associated with tRNA<sup>Thr</sup> where the very rare mismatched A52–C62 pair modifies the contribution of the adjacent 51–63 pair (25). In those cases where the A52–C62 pair in tRNA<sup>Thr</sup> is changed to the more common G52–C62 pair, the  $\Delta G^\circ$  values of multiple base pair mutants are more accurately predicted. Thus, it appears that the thermodynamic contributions of the 3 bp are independent of one another in most cases, permitting the  $\Delta G^\circ$  values of T-stems containing multiple base pair substitutions to be estimated using the relatively few  $\Delta\Delta G^\circ_{AVG}$  values.

### Calculating $\Delta G^\circ$ for all bacterial tRNAs

Of the 6113 non-redundant bacterial tRNA sequences, 5849 (96%) have T-stem sequences for which single base pair  $\Delta\Delta G^\circ_{AVG}$  values are available for all 3 bp. For each bacterial tRNA, we chose to calculate their phenylalanylated versions using the  $\Delta G^\circ = -10.1$  kcal/mol of Phe-tRNA<sup>Phe</sup> (15) and the appropriate  $\Delta\Delta G^\circ_{AVG}$  values associated with their T-stem sequence. By using a common esterified amino acid, facile comparison of the thermodynamic contributions of different tRNA bodies can be made both within a class and among classes. In addition, there is a considerable amount of experimental data for tRNAs misacylated with phenylalanine available for comparison (8,15). For each of the 45 different tRNA anticodon classes, calculated  $\Delta G^\circ$  values were grouped into bins of 0.5 kcal/mol, a bin size appropriate for the number of samples (23) and the error associated with the  $\Delta\Delta G^\circ$  values (8,15). Most of the resulting distributions of calculated  $\Delta G^\circ$  values fit well to a Gaussian distribution (Figure 3A), although several tRNA classes had  $\Delta G^\circ$  distributions with a significant second peak at a lower calculated  $\Delta G^\circ$  (Figure 3B). In many cases, the cause of this secondary peak can be attributed to the high preponderance of one six-base sequence. This particular sequence (G49C65, G50U64, G51C63) is one of the few that was previously found to be inaccurately calculated from the single base pair  $\Delta\Delta G^\circ$  values (Figure 2) (15). When the experimental  $\Delta G^\circ$  for this ‘non-additive’ T stem (15) was substituted for the calculated  $\Delta G^\circ$ , the fit of many tRNA classes was substantially improved (Figure 3B). The mean  $\Delta G^\circ$ ,  $R$  and  $\sigma$  values describing the distributions of all tRNA classes made after this correction are summarized in Table 1.

Nearly all bacterial tRNA anticodon classes have calculated  $\Delta G^\circ$  distributions that show an excellent fit to a Gaussian distribution ( $R > 0.9$ ). The breadths of the distributions ( $\sigma$ ) vary among the different tRNA classes but do not correlate with the number of T-stem sequences present in the class. The values of  $\sigma$  vary in a normal distribution between 0.3 and 0.8 kcal/mol, which is consistent with the error of the  $\Delta\Delta G^\circ_{AVG}$  values used for the calculation. This argues that most bacteria show a similar characteristic  $\Delta G^\circ$  for each individual tRNA class. A closer examination of the T-stems of tRNA<sup>Thr</sup><sub>UGU</sub> from different bacteria illustrates how the many different six residue sequences all calculate to a similar  $\Delta G^\circ$  value (Table 3). For example, while tRNA<sup>Thr</sup><sub>UGU</sub> from *Xanthomonas campestris* uses the weak-binding C49–G65, tight-binding C50–G64, and intermediate-binding C51–G63 base pairs to achieve a calculated  $\Delta G^\circ = -9.8$  kcal/mol, the tRNA<sup>Thr</sup><sub>UGU</sub> from *Leuconostoc citreum* uses a totally different combination of tight-binding G49–U65, intermediate-binding A50–U64 and weak-binding A51–U63 base pairs to give a nearly identical calculated  $\Delta G^\circ = -9.9$  kcal/mol. Thus, although these two tRNA<sup>Thr</sup><sub>UGU</sub> do not share any common nucleotides at the six positions, they have a similar calculated  $\Delta G^\circ$ . This strongly suggests that each T-stem evolved to reach its characteristic  $\Delta G^\circ$  and that all

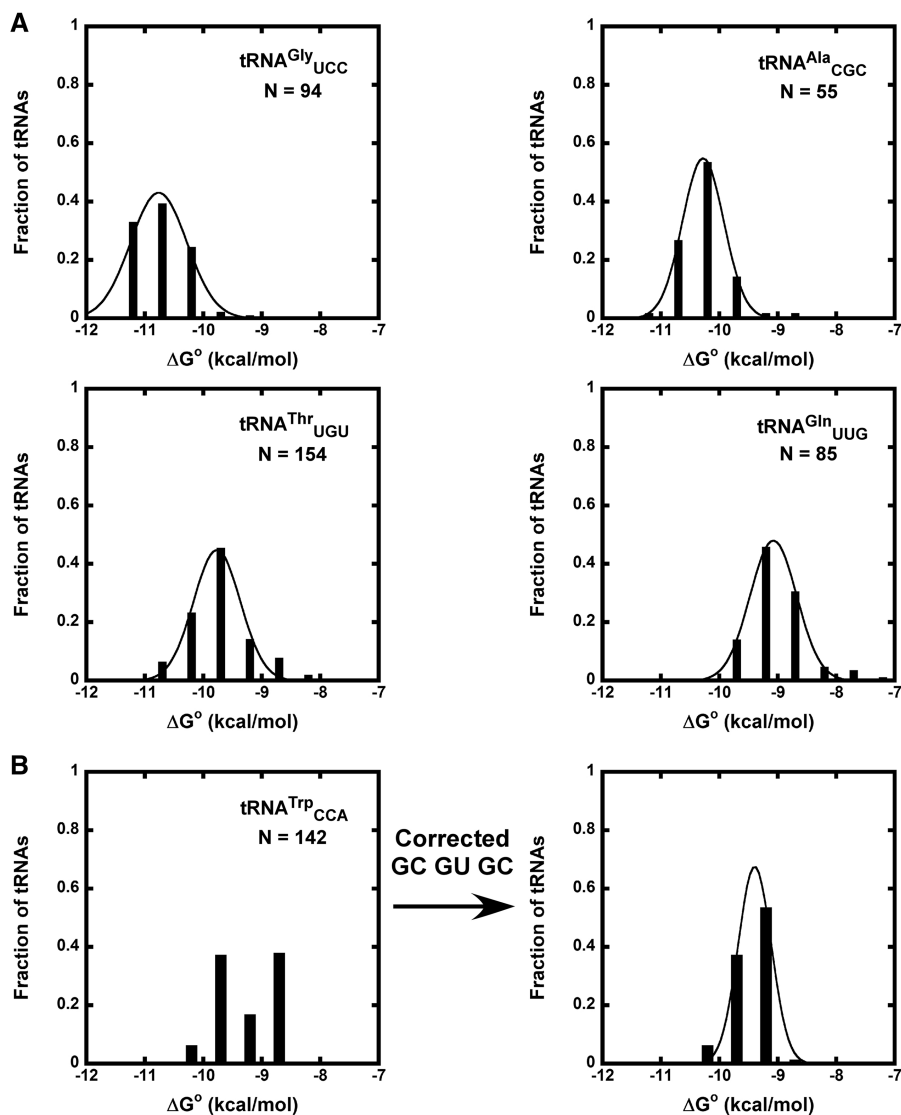


**Figure 2.** Calculation of  $\Delta G^\circ$  of multiple base pair mutants. Comparison of experimental  $\Delta G^\circ$  values of multiple base pair mutants to  $\Delta G^\circ$  values calculated from  $\Delta\Delta G^\circ_{\text{AVG}}$  values. Calculated  $\Delta G^\circ = \Delta G^\circ$  (wild-type aa-tRNA) +  $\Delta\Delta G^\circ_{\text{AVG } 49-65} + \Delta\Delta G^\circ_{\text{AVG } 50-64} + \Delta\Delta G^\circ_{\text{AVG } 51-63}$  (A) yeast Phe-tRNA<sup>Phe</sup> to *T. thermophilus* EF-Tu (15), (B) *E. coli* Val-tRNA<sup>Val</sup> to *E. coli* EF-Tu (17) and (C) Thr-tRNA<sup>Thr</sup> to *E. coli* EF-Tu (25).

bacteria use similar rules for the sequence-specific recognition of tRNA by EF-Tu.

The mean  $\Delta G^\circ$  from the calculated distributions generally agree with the  $\Delta G^\circ$  values previously measured between *T. thermophilus* EF-Tu and *E. coli* tRNAs misacylated with phenylalanine (Table 1) (8). A clear exception is tRNA<sup>Glu</sup><sub>UUC</sub>, which has a predicted mean  $\Delta G^\circ$  that is 1.8 kcal/mol weaker than the experimental  $\Delta G^\circ$  of the *E. coli* Phe-tRNA<sup>Glu</sup><sub>UUC</sub>. While this may be a failure of our 3-bp model, it is also possible that the experimental value is incorrect since the unusually tight  $\Delta G^\circ$  of Phe-tRNA<sup>Glu</sup><sub>UUC</sub> could only be estimated by extrapolation of data obtained at higher ionic strengths or temperatures (8).

Different tRNA classes that are aminocylated by the same amino acid (termed isoacceptors) show very similar distributions centered about the same mean  $\Delta G^\circ$ . For example, similar mean  $\Delta G^\circ$  values were calculated for the three alanine isoacceptors: tRNA<sup>Ala</sup><sub>CGC</sub> (−10.3 kcal/mol), tRNA<sup>Ala</sup><sub>GGC</sub> (−10.0 kcal/mol) and tRNA<sup>Ala</sup><sub>UGC</sub> (−10.0 kcal/mol). Indeed, among all groups of isoacceptors, the mean calculated  $\Delta G^\circ$  values are generally within 0.3 kcal/mol, well within the error of the calculation. This high degree of similarity among the mean  $\Delta G^\circ$  values for isoacceptor tRNAs provides additional support to the model that the identity of the esterified amino acid largely drives the evolution of the T-stem sequence of the corresponding tRNA to have a common value of  $\Delta G^\circ$ .



**Figure 3.** Calculated  $\Delta G^\circ$  distributions for bacterial tRNAs. (A) Calculated  $\Delta G^\circ$  values for phenylalanylated tRNA<sup>Gly</sup><sub>UCC</sub>, tRNA<sup>Ala</sup><sub>CGC</sub>, tRNA<sup>Thr</sup><sub>UGU</sub>, and tRNA<sup>Gln</sup><sub>UUG</sub> fit to Gaussian curves. Precisions of fit ( $R$ ) and breadths ( $\sigma$ ) of distributions are in Table 1. (B) Calculated  $\Delta G^\circ$  for Phe-tRNA<sup>Trp</sup><sub>CCA</sub> before and after tRNAs containing the non-additive T-stem (G49C65 G50U64 G51C63) were substituted with the measured value.

The four tRNA classes with the lowest  $R$  values and/or highest  $\sigma$  values were tRNA<sub>CAU</sub>, tRNA<sub>UCA</sub>, tRNA<sup>Ser</sup><sub>UGC</sub> and tRNA<sup>Val</sup><sub>UAC</sub>. For two of these, the poor fit of the distribution simply reflects the fact that two or more types of tRNA share the same anticodon. Thus, tRNA<sub>CAU</sub> shows a very broad distribution of predicted  $\Delta G^\circ$  but contains both the initiator and elongator methionine tRNAs and tRNA<sup>Ile</sup><sub>CAU</sub> (20). However, if the tRNA<sub>CAU</sub> are separated into subclasses using other sequence elements that are known to define them (20), the resulting calculated  $\Delta G^\circ$  distributions for the tRNA<sup>fMet</sup><sub>CAU</sub>, tRNA<sup>Met</sup><sub>CAU</sub> and tRNA<sup>Ile</sup><sub>CAU</sub> subclasses are all much narrower (Figure 4A). As expected, those tRNA<sub>CAU</sub> identified as tRNA<sup>Ile</sup><sub>CAU</sub> are predicted to bind similarly to the other isoacceptor, tRNA<sup>Ile</sup><sub>GAU</sub>. Although the 41 tRNA<sub>CAU</sub> identified as initiator methionine tRNAs were found to have T-stem sequences predicted to bind EF-Tu similar to tRNA<sup>Ile</sup><sub>CAU</sub>

when both were phenylalanylated, the very weak binding of the esterified formyl methionine results in fMet-tRNA<sup>fMet</sup><sub>CAU</sub> binding quite weakly. Since tRNA<sup>fMet</sup><sub>CAU</sub> does not participate in elongation and is known to bind EF-Tu poorly (8,26,27), it is not surprising that it has evolved T-stem sequences that minimize association with EF-Tu (28). In addition to their weaker-binding T-stems, all but two of these tRNA<sup>fMet</sup><sub>CAU</sub> also contain a mismatched 1–72 bp which further destabilizes binding to EF-Tu (26). While the remaining tRNA<sub>CAU</sub> species include the elongator methionine tRNAs predicted to bind EF-Tu quite well, the distribution of calculated  $\Delta G^\circ$  values remains broad with a large shoulder of more weakly binding tRNAs (Figure 4A). Some of these weaker binding tRNA<sup>Met</sup><sub>CAU</sub> may actually be tRNA<sup>Ile</sup><sub>CAU</sub> since distinguishing the two relies on the poorly conserved identity elements for Tils, the enzyme which modifies C34 (20,29).

**Table 3.** Calculated  $\Delta G^\circ$  values of bacterial Phe – tRNA<sub>UGU</sub><sup>Thr</sup>

49–65	50–64	51–63	Calculated $\Delta G^\circ$ (kcal/mol)	Number of tRNAs
<b>GC</b>	<b>UA</b>	<b>CG</b>	-10.0	19
<b>GC</b>	<u>CG</u>	gu	-9.9	9
<b>GC</b>	gu	<u>GC</u>	-9.1	9
<b>AU</b>	<b>GC</b>	<u>GC</u>	-10.6	8
<u>GU</u>	<u>CG</u>	gu	-10.4	7
<u>AU</u>	<u>UA</u>	<b>CG</b>	-10.1	7
<u>GU</u>	<b>UA</b>	au	-9.9	7
<u>GU</u>	<u>CG</u>	au	-10.1	6
<u>GU</u>	gu	<u>GC</u>	-9.6	6
<b>GC</b>	<u>CG</u>	<u>GC</u>	-10.7	5
<u>GU</u>	<u>GC</u>	gu	-10.2	5
<b>AU</b>	<u>CG</u>	gu	-10.0	4
<b>GC</b>	<b>AU</b>	<u>GC</u>	-10.5	3
<b>GC</b>	<u>CG</u>	<b>CG</b>	-10.2	3
<b>GC</b>	<u>UA</u>	ug	-9.5	3
<b>AU</b>	<u>CG</u>	<b>CG</b>	-10.3	2
<u>GU</u>	<b>AU</b>	au	-9.9	2
<b>AU</b>	<b>UA</b>	gu	-9.8	2
<b>GC</b>	<b>UA</b>	gu	-9.7	2
cg	<b>AU</b>	<b>CG</b>	-9.6	2
<b>GC</b>	<b>UG</b>	<b>CG</b>	-9.6	2
<b>AU</b>	<b>AU</b>	au	-9.5	2
cg	<u>CG</u>	gu	-9.5	2
<b>AU</b>	gu	<b>CG</b>	-8.7	2
cg	gu	<b>CG</b>	-8.2	2
<b>GC</b>	<b>GC</b>	<b>CG</b>	-10.0	1
cg	<u>CG</u>	<b>CG</b>	-9.8	1
ua	<u>UA</u>	<b>CG</b>	-9.8	1
<b>GC</b>	<b>AU</b>	gu	-9.7	1
<b>AU</b>	<b>UA</b>	ug	-9.6	1

Calculated  $\Delta G^\circ$  values for 30 of the 53 most abundant T-stem sequences in bacterial tRNA<sub>UGU</sub><sup>Thr</sup> esterified with phenylalanine. Underlined tRNA base pairs bind tightly (**GC**), bold base pairs bind moderately (**GC**), and lowercase base pairs bind weakly (**gc**).

The broad calculated  $\Delta G^\circ$  distribution observed for the tRNA<sub>UCA</sub> sequence class (Figure 4B) is due to the presence of tRNA<sub>UCA</sub><sup>Trp</sup> from *Mycoplasmataceae* (30,31) and tRNA<sub>UCA</sub><sup>Sec</sup> which both decode the UGA codon. Using TFAM to sort these tRNAs (21), the tRNA<sub>UCA</sub><sup>Trp</sup> sequences give predicted  $\Delta G^\circ$  values similar to the more common tRNA<sub>CCA</sub><sup>Trp</sup> isoacceptor. tRNA<sub>UCA</sub><sup>Sec</sup> has an extended 8-bp acceptor stem that shifts its contact with EF-Tu by 1 bp such that the last base pair of the elongated acceptor-stem and the first two of the T-stem are recognized (32,33). By using these positions to calculate  $\Delta G^\circ$ , most tRNA<sub>UCA</sub><sup>Sec</sup> were predicted to bind weakly to EF-Tu, confirming previous experiments (32,33). Thus, similar to tRNA<sub>CAU</sub>, the non-Gaussian distribution of  $\Delta G^\circ$  values observed among tRNA<sub>UCA</sub> sequences can probably be explained by the presence of multiple tRNA classes and thus does not conflict with the six-base model but rather supports it.

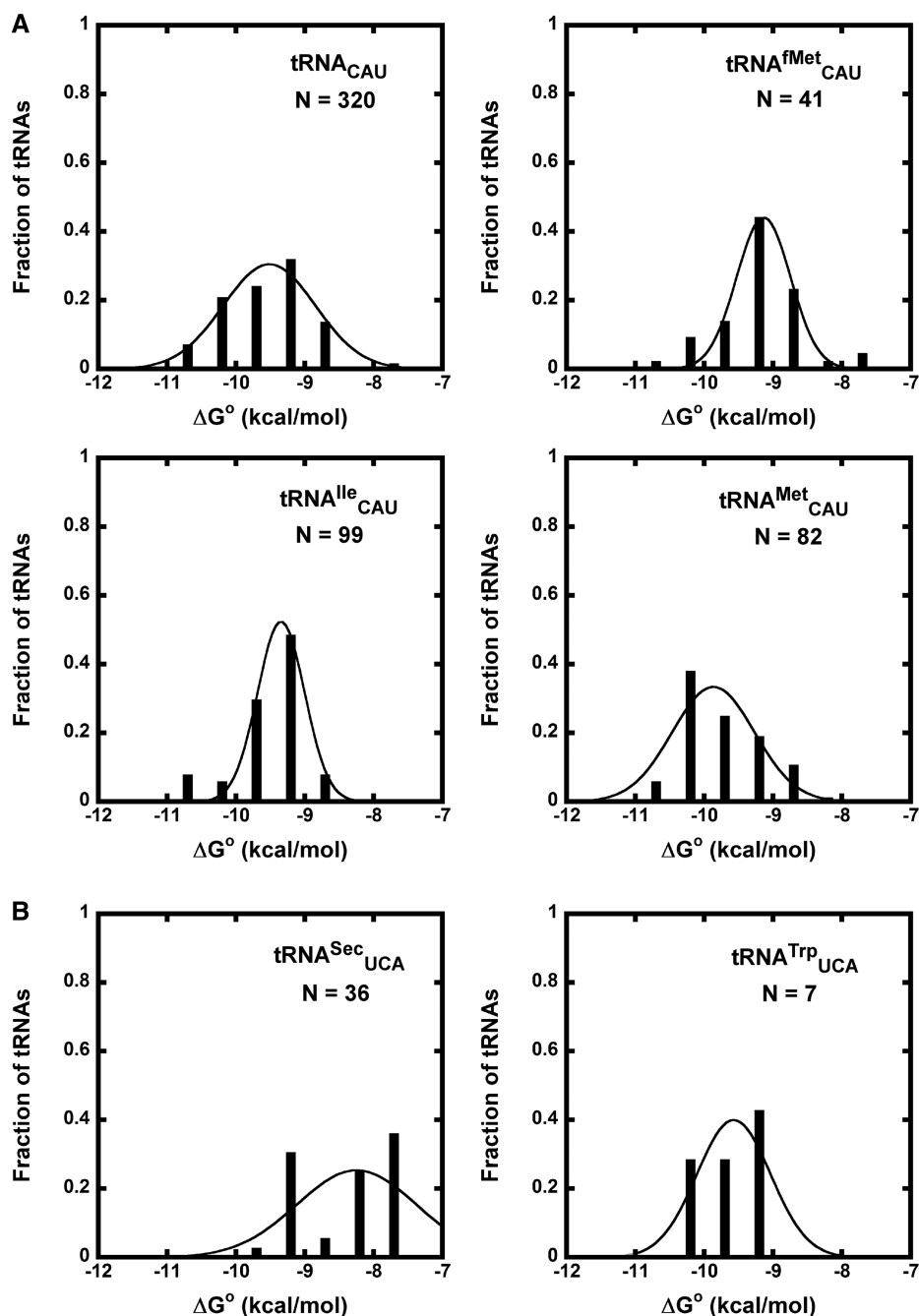
The two other tRNA classes, tRNA<sub>UGA</sub><sup>Ser</sup> and tRNA<sub>UAC</sub><sup>Val</sup>, each contain a small, distinct weaker affinity peak (Figure 5). This peak is unlikely to be explained by an unknown non-additive T-stem sequence since multiple T-stem sequences are present. It is possible that these two classes contain tRNAs with sequence elements outside of the three T-stem base pairs that strengthen

binding to EF-Tu, but no obvious elements were discerned. As will be discussed below, another possible explanation is that the calculated  $\Delta G^\circ$  values are correct, but these classes contain many tRNA species that do not solely participate in translation elongation. Additional experiments measuring the EF-Tu-binding properties of the weaker binding members of these two classes will be needed to understand their anomalously predicted  $\Delta G^\circ$  values.

#### Individual tRNAs with abnormal $\Delta G^\circ$ values

Of the 5849 calculated tRNA sequences, 25 have  $\Delta G^\circ$  values that are predicted to be tighter, and 19 have  $\Delta G^\circ$  values that are predicted to be weaker, than their corresponding mean  $\Delta G^\circ$  by >1.3 kcal/mol (Table 4). Given the large data set and the errors involved in the calculation, it is statistically unlikely that all of these tRNAs actually have anomalous  $\Delta G^\circ$  values. Nevertheless, this group of tRNAs, termed outliers, deserves further scrutiny. Indeed, one of the weaker outlier tRNAs is the tRNA<sub>UCC</sub><sup>Gly</sup> from *Staphylococcus* that functions as the glycine donor in the biosynthesis of pentapeptide crosslink in the cell wall (34–36). This tRNA was found to bind very poorly to *Staphylococcus* EF-Tu, suggesting it had evolved to avoid the translation machinery so that it could perform its specialized function (37). Those *Staphylococcus* species that contain this tRNA have a second copy of the tRNA<sub>UCC</sub><sup>Gly</sup> gene whose product binds EF-Tu more tightly and participates in translation. As expected, the predicted  $\Delta G^\circ$  value of this second tRNA is close to the mean of the tRNA<sub>UCC</sub><sup>Gly</sup> distribution. This successful identification of an aa-tRNA known to have a function outside of translation suggests that other outliers could have similar interesting functions. However, of the 44 outlier tRNAs, 41 possess T-stem sequences that have not been verified experimentally by inserting into tRNA<sup>Phe</sup> and therefore may have a  $\Delta G^\circ$  that was not predicted accurately by combining  $\Delta\Delta G^\circ_{AVG}$  values. Alternatively, it is possible that some of these tRNAs resemble *E. coli* tRNA<sup>Thr</sup> or tRNA<sup>Met</sup> and possess sequence elements outside of the six residues in the T-stem that contribute to  $\Delta G^\circ$ . Finally, the predicted  $\Delta G^\circ$  values of the outlier tRNA may be accurate, indicating that the tRNA actually possesses an unusual  $\Delta G^\circ$  for binding EF-Tu in the organism.

In order to test the above possibilities, 9 of the 44 outlier tRNAs were chosen and two chimeras were made for each (Figure 6). One chimera imported the six T-stem residues into tRNA<sup>Phe</sup>, thereby testing whether the  $\Delta\Delta G^\circ_{AVG}$  values were additive. The other chimera introduced the GAA anticodon and G3–U70 into the tRNA to permit aminoacylation with either Phe or Ala instead of the cognate amino acid. Since neither of these sequence changes affects EF-Tu affinity, this allows the outlier tRNA to be easily aminoacylated and its  $\Delta G^\circ$  to *T. thermophilus* EF-Tu to be measured (8,9). The weak-binding tRNA<sub>UCC</sub><sup>Gly</sup> from *Staphylococcus* was included as a positive control. A second positive control was tRNA<sub>UUC1</sub><sup>Glu</sup> from *Helicobacter pylori* that also binds weakly to EF-Tu (38). Although the function of this tRNA is unknown, there is another copy of the tRNA



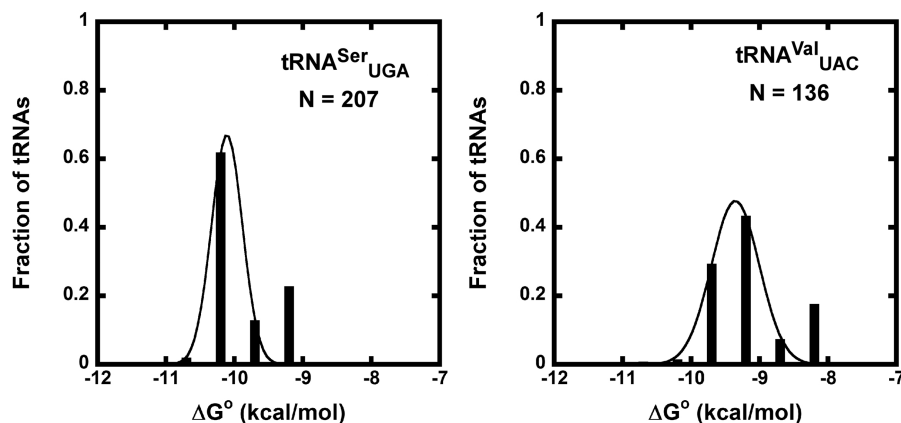
**Figure 4.** Anticodon classes containing multiple tRNA types. (A) Calculated  $\Delta G^\circ$  values of Phe-tRNA<sub>CAU</sub> sequences subdivided into tRNA<sub>CAU</sub><sup>fMet</sup>, tRNA<sub>CAU</sub><sup>Ile</sup> and tRNA<sub>CAU</sub><sup>Met</sup>. (B) Calculated  $\Delta G^\circ$  values of Phe-tRNA<sub>UCA</sub> subdivided into tRNA<sub>UCA</sub><sup>Sec</sup> and tRNA<sub>UCA</sub><sup>Trp</sup>.

gene (tRNA<sub>UUC2</sub><sup>Glu</sup>) in *Helicobacter* that binds EF-Tu quite well. The experimental  $\Delta G^\circ$  values for the two chimeras of the nine outlier tRNAs are compared with their predicted values in Table 4. Both the tRNA<sub>UCC</sub><sup>Gly</sup> from *Staphylococcus* and the tRNA<sub>UUC1</sub><sup>Glu</sup> from *Helicobacter pylori* showed weak  $\Delta G^\circ$  values similar to the predicted values and to the values that had been previously determined (38). In addition, we show that a tRNA<sub>CGC</sub><sup>Ala</sup> from *Sorangium cellulosum* and a tRNA<sub>CGU</sub><sup>Thr</sup> from *Bifidobacterium longum* that were predicted to bind weakly did indeed bind

similarly to their predicted values. Since both of these organisms contain an additional copy of the tRNA gene with a T-stem sequence that is predicted to bind EF-Tu normally, these two tRNAs are also candidates to have specialized functions that do not involve translation. Eight other tRNAs that may have specialized functions are listed in Table 4.

Thirty-two outlier tRNAs are the products of the only gene copy for the anticodon class present in the genome of the organism. This was unexpected because these tRNAs





**Figure 5.** Anticodon classes with poor fit to a Gaussian distribution. Calculated  $\Delta G^\circ$  values of Phe – tRNA<sup>Ser</sup><sub>UGA</sub> ( $R = 0.923$ ) and Phe – tRNA<sup>Val</sup><sub>UAC</sub> ( $R = 0.925$ ).

presumably are involved in decoding and thus should be subject to the same selective pressures on EF-Tu binding as all other tRNAs. For five of these tRNAs, the two chimeras were made and their  $\Delta G^\circ$  values determined (Table 4). Insertion of the T-stems of four of these variants into yeast tRNA<sup>Phe</sup> led to  $\Delta G^\circ$  close to the predicted values. However, the T-stem chimera from tRNA<sup>Arg</sup><sub>CCU</sub> from *Rubrobacter xylophilus* bound 0.4 kcal/mol more tightly than predicted, indicating that it is not an outlier but has a non-additive T stem. Consistent with this interpretation, the T-stem sequence from this tRNA (G49C65, G50U64, A51U63) is similar to the very abundant non-additive T stem discussed above. While the measured  $\Delta G^\circ$  values of the other four yeast tRNA<sup>Phe</sup> chimeras were close to the predicted values, the measured  $\Delta G^\circ$  values of the anticodon chimeras were quite different in three cases. Thus, *Bradyrhizobium japonicum* tRNA<sup>Leu</sup><sub>GAG</sub> binds 0.7 kcal/mol tighter, *Corynebacterium urealyticum* tRNA<sup>Thr</sup><sub>UGU</sub> binds 0.7 kcal/mol weaker, and *Mesorhizobium loti* tRNA<sup>Thr</sup><sub>UGU</sub> binds 1.4 kcal/mol tighter than the same six base sequences in tRNA<sup>Phe</sup>. In all three cases, these affinities are much closer to the mean  $\Delta G^\circ$  of their corresponding anticodon class, indicating that these tRNAs are also not outliers. Presumably, these three tRNAs have additional elements outside their T stem that affect the affinity to EF-Tu. Identification of these elements will require additional experiments, although candidate elements can be identified in the sequence. For example, the unexpectedly tight-binding tRNA<sup>Thr</sup><sub>UGU</sub> from *Mesorhizobium loti* contains a rare U5G68 U6G67 motif, which selectively binds a divalent metal ion in the major groove (39,40) and thereby could stabilize its interaction to EF-Tu. Finally, only one of the five selected outlier tRNAs from single copy genes actually has an anomalous  $\Delta G^\circ$  value. Both chimeras of the tRNA<sup>Arg</sup><sub>CCU</sub> from *Fusobacterium nucleatum* bind significantly tighter to EF-Tu than tRNA<sup>Arg</sup><sub>CCU</sub> from other bacteria. While this tRNA is likely used in translation, it is unclear why the affinity has been selected to be so tight.

## DISCUSSION

We developed a simple thermodynamic model to predict  $\Delta G^\circ$  of EF-Tu to different tRNA sequences. Our model assumes that the contribution of each of the three T-stem base pairs contributes independently to binding. This permits calculation of  $\Delta G^\circ$  for different T-stem sequences by summing the experimental  $\Delta\Delta G^\circ_{AVG}$  for the single base pair substitutions at the three sites. The model accurately predicts experimental  $\Delta G^\circ$  values for tRNAs containing multiple T-stem base pair changes irrespective of the tRNA body or EF-Tu used. When the model is used to predict  $\Delta G^\circ$  values for a large set of tRNAs from different bacteria, the tRNA anticodon classes have predicted  $\Delta G^\circ$  distributions which fit well to a Gaussian curve and have a width consistent with the experimental errors of the  $\Delta\Delta G^\circ_{AVG}$  values. Their predicted mean  $\Delta G^\circ$  values are characteristic for each anticodon class despite the fact that their sequences vary considerably among bacteria. Since the variable affinity of tRNAs for EF-Tu is to compensate for the variable affinity contributed by the esterified amino acid, it appears that this evolutionary adaptation is maintained in all bacteria and uniform tRNA recognition rules are used.

However, many tRNA classes showed a strong secondary peak in their distribution. The most common reason for this is that for a certain sequence (G49C65, G50U64, G51C63), the  $\Delta\Delta G^\circ_{AVG}$  values at the three sites are not additive, so their  $\Delta G^\circ$  values were not accurately predicted by our simple model. Indeed, when experimental data was substituted for this common combination of six residues, all of the secondary peaks in the  $\Delta G^\circ$  distributions were substantially shifted into the main distribution. The presence of certain non-additive sequences is not surprising considering that the three ‘specificity’ base pairs are adjacent and that the protein contacts the 49–65 and 50–64 bp indirectly via the 2' OH groups. Stacking between certain combinations of base pairs may form unusual helix geometries that alter binding. It is also possible that certain helical sequences selectively bind divalent ions that could influence  $\Delta G^\circ$  (39–41).

**Table 4.** tRNAs with anomalous calculated  $\Delta G^\circ$  values

tRNA	Species	Mean Isoacceptor $\Delta G^\circ$ (kcal/mol)	Calculated $\Delta G^\circ$ (kcal/mol)	tRNA Chimera $\Delta G^\circ$ (kcal/mol)	tRNA <sup>Phe</sup> Chimera $\Delta G^\circ$ (kcal/mol)
Another gene copy present					
Ala CGC	<i>Sorangium cellulosum</i>	-10.3	-9.0	-9.3 ± 0.6	-8.9 ± 0.6
Arg ACG	<i>Sorangium cellulosum</i>	-9.2	-10.7		
Arg ACG	<i>Pseudomonas aeruginosa</i>	-9.2	-10.7		
Asn GUU	<i>Bradyrhizobium japonicum</i>	-9.4	-10.7		
Gln UUG	<i>Mesorhizobium loti</i>	-9.1	-7.7		
Glu UUC	<i>Thiomicrospira denitrificans ATCC 33890</i>	-10.4	-8.1		
Glu UUC	<i>Helicobacter pylori</i>	-10.4	-8.0	-8.6 ± 0.04	-8.4 ± 0.1
Gly GCC	<i>Acaryochloris marina MBIC11018</i>	-11.0	-9.4		
Gly UCC	<i>Staphylococcus species</i>	-10.8	-9.5	-9.6 ± 0.3	-9.7 ± 0.2
Ile GAU	<i>Mycobacterium vanbaalenii PYR-2</i>	-9.4	-10.7		
Ile GAU	<i>Leuconostoc mesenteroides ATCC 8294</i>	-9.4	-8.1		
Thr CGU	<i>Bifidobacterium longum</i>	-10.0	-8.1	-8.3 ± 0.5	-8.3 ± 0.3
Only gene copy present					
Arg ACG	<i>Aquifex aeolicus</i>	-9.2	-10.7		
Arg CCG	<i>Geobacter sp FRC32</i>	-9.5	-11.0		
Arg CCG	<i>Candidatus Ruthia magnifica Cm Calyptogenia magnifica</i>	-9.5	-11.0		
Arg CCG	<i>Candidatus Vesicomysocius okutanii HA</i>	-9.5	-11.0		
Arg CCG	<i>Bifidobacterium longum</i>	-9.5	-11.0		
Arg CCU	<i>Rubrobacter xylanophilus</i>	-9.6	-8.0	-9.1 ± 0.4	-8.4 ± 0.1
Arg CCU	<i>Fusobacterium nucleatum</i>	-9.6	-11.0	-10.8 ± 0.6	-10.9 <sup>a</sup>
Arg UCU	<i>Clavibacter michiganensis</i>	-9.3	-10.7		
Arg UCU	<i>Lactobacillus sakei</i>	-9.3	-10.7		
Arg UCU	<i>Zymomonas mobilis</i>	-9.3	-10.7		
Arg UCU	<i>Acidiphilium cryptum JF-5</i>	-9.3	-10.7		
Arg UCU	<i>Chloroflexus aurantiacus J 10 fl</i>	-9.3	-10.7		
Arg UCU	<i>Opitutus terrae PB90</i>	-9.3	-10.7		
Arg UCU	<i>Renibacterium salmoninarum ATCC 33209</i>	-9.3	-10.7		
Arg UCU	<i>Lactobacillus reuteri F275</i>	-9.3	-10.7		
Arg UCU	<i>Bacteroides fragilis NCTC 9434</i>	-9.3	-10.6		
Arg UCU	<i>Cytophaga hutchinsonii ATCC 33406</i>	-9.3	-10.6		
Asn GUU	<i>Borrelia afzelii Pko</i>	-9.4	-8.0		
Gln CUG	<i>Lactobacillus acidophilus NCFM</i>	-8.9	-10.3		
Gln UUG	<i>Solibacter usitatus Ellin6077</i>	-9.1	-7.7		
Gln UUG	<i>Candidatus Pelagibacter ubique HTCC1063</i>	-9.1	-7.6		
Leu GAG	<i>Bradyrhizobium japonicum</i>	-10.4	-9.1	-9.8 ± 0.6	-9.4 <sup>a</sup>
Lys CUU	<i>Ehrlichia ruminantium Gardel</i>	-9.5	-11.0		
Pro CGG	<i>Lactobacillus acidophilus NCFM</i>	-9.9	-11.2		
Thr GGU	<i>Ehrlichia canis Jake</i>	-10.1	-8.6		
Thr GGU	<i>Mycoplasma genitalium</i>	-10.1	-8.2		
Thr GGU	<i>Mycoplasma gallisepticum</i>	-10.1	-8.2		
Thr UGU	<i>Corynebacterium urealyticum DSM 7110</i>	-9.8	-11.2	-10.5 ± 0.3	-11.2 <sup>a</sup>
Thr UGU	<i>Mesorhizobium loti</i>	-9.8	-8.6	-10.0 ± 0.7	-8.7 ± 0.4
Thr UGU	<i>Candidatus Pelagibacter ubique HTCC1063</i>	-9.8	-8.2		
Thr UGU	<i>Dinoroseobacter shibae</i>	-9.8	-8.2		
Tyr GUA	<i>Borrelia afzelii Pko</i>	-9.2	-10.5		

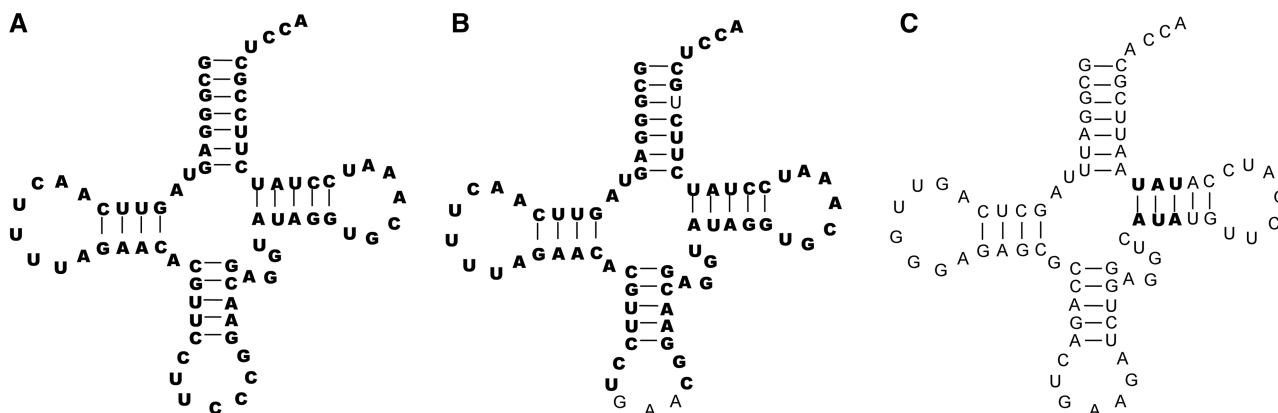
Both calculated and measured  $\Delta G^\circ$  are for phenylalanylated tRNAs.

<sup>a</sup> $\Delta G^\circ$  measured in (15).

Since only two such non-additive T-stem sequences were found out of 47 experimentally tested, it is unlikely that more than four additional ones will be found among the remaining 100 T stems that have not been measured.

The poor fit of two distributions, tRNA<sub>CAU</sub> and tRNA<sub>UCA</sub>, actually supports the six base model because the tRNA classes contain a mixture of two or three tRNA types with different EF-Tu-binding properties. When these two tRNA classes are sorted into their individual subclasses, the model is fairly effective at predicting characteristic  $\Delta G^\circ$  values. However, two anticodon classes, tRNA<sub>UAC</sub><sup>Val</sup> and tRNA<sub>UGA</sub><sup>Ser</sup>, have significantly poorer

Gaussian fits due to the presence of a minor peak with weaker affinity. These minor peak tRNAs have several different six base sequences and do not appear in any particular set of organisms. It is possible that these tRNAs contain sequence elements outside the T stem that contribute to  $\Delta G^\circ$ , although no pattern of unusual sequences was discerned. A final possible explanation is that members of these particular anticodon classes also participate in non-translational functions that could place additional selective pressure on their sequence. For example, in addition to its function in translation, Ser-tRNA<sub>Ser</sub> often participates in the pentapeptide crosslink synthesis for



**Figure 6.** Chimeras to experimentally test outlier tRNAs. (A) Sequence of *Staphylococcus aureus* tRNA<sup>Gly</sup><sub>UCC</sub>. (B) Chimera containing GAA anticodon and G3 U70 mutation to enable aminoacylation with either Phe or Ala. (C) Chimera with the *S. aureus* T-stem bases inserted into yeast tRNA<sup>Phe</sup>.

peptidoglycan production (35,42,43) and has also been shown to participate in the biosynthesis of valanimycin (44). Perhaps for those tRNAs that have such dual roles, their T stems have evolved to be weaker for EF-Tu binding to allow a greater fraction of the molecules to be available for the non-translational biosynthetic enzymes (16,45,46). More experiments will be needed to determine whether the additional biosynthetic functions of aa-tRNAs negatively affect EF-Tu affinity.

Our thermodynamic model also identified 44 rare 'outlier' tRNAs with predicted EF-Tu affinities that are either much tighter or weaker than expected and thus are candidates for unusual function. However, since many tRNAs contain T stems that had not been experimentally verified, it was important to prepare and measure the affinities of chimeras to confirm each putative outlier. Two examples of tRNAs that were known to bind EF-Tu poorly were successfully predicted: a tRNA<sup>Gly</sup><sub>UCC</sub> present in all *Staphylococcus species* and tRNA<sup>Glu</sup><sub>UUC1</sub> from *Helicobacter pylori*. We additionally verified that tRNA<sup>Ala</sup><sub>CGC</sub> from *Sorangium cellulosum* and tRNA<sup>Thr</sup><sub>CGU</sub> from *Bifidobacterium longum* bind EF-Tu poorly and, because these organisms have a second copy of the tRNA gene which is predicted to bind normally, these tRNAs are candidates to participate in a function that does not involve translation. At least four other such candidates of unusual function were found (Table 4).

We also predicted many outlier tRNAs which were the only genomic copy of their respective anticodon class. When five chimeras from this group of tRNAs were tested, one contained a non-additive T-stem sequence and three contained elements outside the T stem which resulted in an affinity that was closer to the mean  $\Delta G^\circ$  for their respective anticodon classes. Therefore, four of the five predicted outlier tRNAs that were tested did not actually have anomalous  $\Delta G^\circ$  values and thus were not outliers. This suggests that many of the other predicted outliers from single copy genes in Table 4 do not actually have anomalous  $\Delta G^\circ$  values, supporting the view that the selective pressure to maintain the

appropriate  $\Delta G^\circ$  value must be very strong. However, we found one predicted outlier from a single copy gene, tRNA<sup>Arg</sup><sub>CCU</sub> from *Fusobacterium nucleatum*, which binds to EF-Tu 1.2 kcal/mol more tightly than the mean  $\Delta G^\circ$  value for tRNA<sup>Arg</sup><sub>CCU</sub>. It has been shown that aa-tRNAs with tighter than cognate affinities to EF-Tu release slowly into the ribosome after GTP hydrolysis (17), suggesting that this tRNA<sup>Arg</sup><sub>CCU</sub> will cause slow translation of AGG codons and therefore may perform a unique regulatory function in *Fusobacterium*. It is possible that by slowing the translation rate of AGG codons, which compose only 0.23% of codons in *Fusobacterium* ORFs, tRNA<sup>Arg</sup><sub>CCU</sub> could be used to modulate translational efficiency of mRNAs enriched in AGG codons or it could be used to induce specific pauses to facilitate cotranslational protein folding (47,48). Another possible explanation for the tight affinity of this tRNA is that under conditions where the EF-Tu•GTP concentration is low, such as encountered during starvation, this tRNA<sup>Arg</sup> would still function efficiently, possibly to translate AGG codons in starvation-induced genes. Indeed, it is known that this arginine isoacceptor is selectively charged upon starvation in *E. coli* (49). Future experiments will be needed to uncover the role of this tight-binding tRNA.

We have provided evidence that the mechanism for specific recognition of aa-tRNAs by EF-Tu is conserved in all bacteria. Indeed, the  $\Delta G^\circ$  values of individual tRNA classes are remarkably uniform among bacteria, especially considering the broad range of growth conditions encountered between organisms. Because the entire aa-tRNA interface of EF-Tu is almost universally conserved, bacterial tRNAs use the same three T-stem base pairs employed by *E. coli* to subtly modulate the affinities to EF-Tu in a way that compensates for the variable affinities of the esterified amino acid. This universality of the recognition rules was difficult to anticipate because of the extreme variability of bacterial T-stem sequences. Because two of the three specificity contacts involved interactions with the backbone of the RNA with unpredictable sequence specificity, the rules had to be determined empirically. As a result of quite variable

contributions to  $\Delta G^\circ$  found at the three sites, multiple different T-stem sequences can give similar  $\Delta G^\circ$  values. The sequence-specific recognition of helices may be used by other RNA-binding proteins. For example, the many contacts between the helical regions of tRNA and aa-tRNA synthetases could contribute to sequence-specific binding as some experiments have suggested (50).

While all bacteria use the same mechanism of recognition of aa-tRNAs by EF-Tu, an analysis of protein and tRNA sequences indicates that eukaryotic and archaeal EF-1 $\alpha$  recognize aa-tRNAs in a different manner. While no ternary complex structures with EF-1 $\alpha$  are available, a GTP-bound form of an archaeal protein (51) permits an accurate structure-based sequence alignment with bacterial EF-Tu sequences. Although some of the residues that make up the amino acid-binding pocket in bacterial EF-Tu (H67, E226, D227, T239 and N285) are present in archaea (N285, D227), others are quite different from bacteria (Q226, V239 and F67). These changes are likely to alter the specificity of the pocket for different esterified amino acids (24,52). Although two of the three archaeal EF-1 $\alpha$  residues (S350, D390 and V341) expected to interact with the T stem (13) are fairly similar to bacteria (T350 and E390), the third (Q341) is not. It is therefore likely that the combination of the three changes would be sufficient to alter the subtle hydrogen-bonding pattern used to achieve sequence-specific tRNA binding. In addition, if the bacterial  $\Delta\Delta G_{AVG}^\circ$  values are used to predict the  $\Delta G^\circ$  of archaeal tRNAs, it is clear that archaeal tRNAs often fit poorly to the corresponding bacterial distributions. For example, while bacterial tRNA<sup>Trp</sup> binds rather weakly with a mean  $\Delta G^\circ = -9.4$  kcal/mol (Table 1), applying our model yields a distribution of archaeal tRNA<sup>Trp</sup> with two completely distinct peaks: one at  $-10.2$  kcal/mol which is significantly tighter and the other at  $-8.7$  kcal/mol which is significantly weaker. This suggests the way that archaeal EF-1 $\alpha$  interacts with aa-tRNAs differs substantially from bacteria. Although no appropriate structure is available to permit alignment, it is likely that eukaryotic EF-1 $\alpha$  also recognizes aa-tRNAs differently from bacteria. The differences in archaea and eukaryotic EF-1 $\alpha$  are not entirely surprising as their translation systems contain many distinct features from bacteria, especially in initiation (53). While relatively little biochemical data is available, eukaryotic EF-1 $\alpha$  also appears to require the presence of the esterified amino acid and GTP to bind tRNA (54,55). However, it is clear that additional quantitative data will be needed to understand how different archaeal or eukaryotic tRNAs interact with EF-1 $\alpha$ .

## ACKNOWLEDGEMENTS

The authors would like to thank Margaret Saks and Jonathan Widom for insightful discussions.

## FUNDING

National Institutes of Health (Grant number GM037552 to O.C.U.). Funding for open access charge: GM037552.

*Conflict of interest statement.* None declared.

## REFERENCES

- Krab, I.M. and Parmeggiani, A. (1998) EF-Tu, a GTPase odyssey. *Biochim. Biophys. Acta*, **1443**, 1–22.
- Pape, T., Wintermeyer, W. and Rodnina, M.V. (1998) Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. *EMBO J.*, **17**, 7490–7497.
- Schrader, J.M., Saks, M.E. and Uhlenbeck, O.C. (2010) *Ribosomes: Structure, Function, and Dynamics*. Springer Wien, New York.
- Louie, A. and Jurnak, F. (1985) Kinetic studies of Escherichia coli elongation factor Tu-guanosine 5'-triphosphate-aminoacyl-tRNA complexes. *Biochemistry*, **24**, 6433–6439.
- Louie, A., Ribeiro, N.S., Reid, B.R. and Jurnak, F. (1984) Relative affinities of all Escherichia coli aminoacyl-tRNAs for elongation factor Tu-GTP. *J. Biol. Chem.*, **259**, 5010–5016.
- Ott, G., Schiesswohl, M., Kiesewetter, S., Forster, C., Arnold, L., Erdmann, V.A. and Sprinzl, M. (1990) Ternary complexes of Escherichia coli aminoacyl-tRNAs with the elongation factor Tu and GTP: thermodynamic and structural studies. *Biochim. Biophys. Acta*, **1050**, 222–225.
- Asahara, H. and Uhlenbeck, O.C. (2005) Predicting the binding affinities of misacylated tRNAs for Thermus thermophilus EF-Tu.GTP. *Biochemistry*, **44**, 11254–11261.
- Asahara, H. and Uhlenbeck, O.C. (2002) The tRNA specificity of Thermus thermophilus EF-Tu. *Proc. Natl Acad. Sci. USA*, **99**, 3499–3504.
- Dale, T., Sanderson, L.E. and Uhlenbeck, O.C. (2004) The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry*, **43**, 6159–6166.
- LaRiviere, F.J., Wolfson, A.D. and Uhlenbeck, O.C. (2001) Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science*, **294**, 165–168.
- Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B.F. and Nyborg, J. (1995) Crystal structure of the ternary complex of Phe-tRNA<sup>Phe</sup>, EF-Tu, and a GTP analog. *Science*, **270**, 1464–1472.
- Sanderson, L.E. and Uhlenbeck, O.C. (2007) Directed mutagenesis identifies amino acid residues involved in elongation factor Tu binding to yeast Phe-tRNA(Phe). *J. Mol. Biol.*, **368**, 119–130.
- Sanderson, L.E. and Uhlenbeck, O.C. (2007) Exploring the specificity of bacterial elongation factor Tu for different tRNAs. *Biochemistry*, **46**, 6194–6200.
- Sanderson, L.E. and Uhlenbeck, O.C. (2007) The 51-63 base pair of tRNA confers specificity for binding by EF-Tu. *RNA*, **13**, 835–840.
- Schrader, J.M., Chapman, S.J. and Uhlenbeck, O.C. (2009) Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis. *J. Mol. Biol.*, **386**, 1255–1264.
- Roy, H. and Ibba, M. (2008) RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. *Proc. Natl Acad. Sci. USA*, **105**, 4667–4672.
- Schrader, J.M., Chapman, S.J. and Uhlenbeck, O.C. (2011) Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proc. Natl Acad. Sci. USA*, **108**, 5215–5220.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.*, **37**, D93–D97.
- Silva, F.J., Belda, E. and Talens, S.E. (2006) Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Res.*, **34**, 6015–6022.
- Taqi, H., Cui, Y. and Ardell, D.H. (2007) TFAM 1.0: an online tRNA function classifier. *Nucleic Acids Res.*, **35**, W350–W353.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

23. Sturges, H.A. (1926) The choice of a class interval. *J. Am. Stat. Assoc.*, **21**, 65–66.
24. Roy, H., Becker, H.D., Mazauric, M.H. and Kern, D. (2007) Structural elements defining elongation factor Tu mediated suppression of codon ambiguity. *Nucleic Acids Res.*, **35**, 3420–3430.
25. Saks, M.E., Sanderson, L.E., Choi, D.S., Crosby, C.M. and Uhlenbeck, O.C. (2011) Functional consequences of T-stem mutations in *E. coli* tRNAThrUGU in vitro and in vivo. *RNA*, **17**, 1038–1047.
26. Seong, B.L. and RajBhandary, U.L. (1987) Mutants of *Escherichia coli* formylmethionine tRNA: a single base change enables initiator tRNA to act as an elongator in vitro. *Proc. Natl Acad. Sci. USA*, **84**, 8859–8863.
27. Fischer, W., Doi, T., Ikehara, M., Ohtsuka, E. and Sprinzl, M. (1985) Interaction of methionine-specific tRNAs from *Escherichia coli* with immobilized elongation factor Tu. *FEBS Lett.*, **192**, 151–154.
28. Stortchevoi, A., Varshney, U. and RajBhandary, U.L. (2003) Common location of determinants in initiator transfer RNAs for initiator-elongator discrimination in bacteria and in eukaryotes. *J. Biol. Chem.*, **278**, 17672–17679.
29. Ikeuchi, Y., Soma, A., Ote, T., Kato, J., Sekine, Y. and Suzuki, T. (2005) molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Mol. Cell.*, **19**, 235–246.
30. Simoneau, P., Li, C.M., Loechel, S., Wenzel, R., Herrmann, R. and Hu, P.C. (1993) Codon reading scheme in *Mycoplasma pneumoniae* revealed by the analysis of the complete set of tRNA genes. *Nucleic Acids Res.*, **21**, 4967–4974.
31. Inamine, J.M., Ho, K.C., Loechel, S. and Hu, P.C. (1990) Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. *J. Bacteriol.*, **172**, 504–506.
32. Rudinger, J., Hillenbrandt, R., Sprinzl, M. and Giege, R. (1996) Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *EMBO J.*, **15**, 650–657.
33. Baron, C. and Bock, A. (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *J. Biol. Chem.*, **266**, 20375–20379.
34. Bumsted, R.M., Dahl, J.L., Soll, D. and Strominger, J.L. (1968) Biosynthesis of the peptidoglycan of bacterial cell walls. X. Further study of the glycyl transfer ribonucleic acids active in peptidoglycan synthesis in *Staphylococcus aureus*. *J. Biol. Chem.*, **243**, 779–782.
35. Petit, J.F., Strominger, J.L. and Soll, D. (1968) Biosynthesis of the peptidoglycan of bacterial cell walls. VII. Incorporation of serine and glycine into interpeptide bridges in *Staphylococcus epidermidis*. *J. Biol. Chem.*, **243**, 757–767.
36. Roberts, R.J. (1972) Structures of two glycyl-tRNAs from *Staphylococcus epidermidis*. *Nat. New Biol.*, **237**, 44–45.
37. Giannouli, S., Kyritsis, A., Malissov, N., Becker, H.D. and Stathopoulos, C. (2009) On the role of an unusual tRNA<sup>Gly</sup> isoacceptor in *Staphylococcus aureus*. *Biochimie*, **91**, 344–351.
38. Cathopoulos, T.J., Chuawong, P. and Hendrickson, T.L. (2008) Conserved discrimination against misacylated tRNAs by two mesophilic elongation factor Tu orthologs. *Biochemistry*, **47**, 7610–7616.
39. Varani, G. and McClain, W.H. (2000) The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.*, **1**, 18–23.
40. Ott, G., Arnold, L. and Limmer, S. (1993) Proton NMR studies of manganese ion binding to tRNA-derived acceptor arm duplexes. *Nucleic Acids Res.*, **21**, 5859–5864.
41. Keel, A.Y., Rambo, R.P., Batey, R.T. and Kieft, J.S. (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure*, **15**, 761–772.
42. Villet, R., Fonvielle, M., Busca, P., Chemama, M., Maillard, A.P., Hugonnet, J.E., Dubost, L., Marie, A., Josseume, N., Mesnage, S. *et al.* (2007) Idiosyncratic features in tRNAs participating in bacterial cell wall synthesis. *Nucleic Acids Res.*, **35**, 6870–6883.
43. Lloyd, A.J., Gilbey, A.M., Blewett, A.M., De Pascale, G., El Zoeiby, A., Levesque, R.C., Catherwood, A.C., Tomasz, A., Bugg, T.D., Roper, D.I. *et al.* (2008) Characterization of tRNA-dependent peptide bond formation by MurM in the synthesis of *Streptococcus pneumoniae* peptidoglycan. *J. Biol. Chem.*, **283**, 6402–6417.
44. Garg, R.P., Gonzalez, J.M. and Parry, R.J. (2006) Biochemical characterization of VlmL, a Seryl-tRNA synthetase encoded by the valanimycin biosynthetic gene cluster. *J. Biol. Chem.*, **281**, 26785–26791.
45. Francklyn, C.S. and Minajigi, A. (2010) tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett.*, **584**, 366–375.
46. RajBhandary, U.L. and Soll, D. (2008) Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics. *Proc. Natl Acad. Sci. USA*, **105**, 5285–5286.
47. Zhang, G., Hubalewska, M. and Ignatova, Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.*, **16**, 274–280.
48. Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
49. Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M. and Pan, T. (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.*, **6**, 151–157.
50. Perona, J.J. and Hou, Y.M. (2007) Indirect readout of tRNA for aminoacylation. *Biochemistry*, **46**, 10419–10432.
51. Kobayashi, K., Kikuno, I., Kuroha, K., Saito, K., Ito, K., Ishitani, R., Inada, T. and Nureki, O. (2010) Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1 $\alpha$  complex. *Proc. Natl Acad. Sci. USA*, **107**, 17575–17579.
52. Sato, A., Watanabe, Y., Suzuki, T., Komiyama, M., Watanabe, K. and Ohtsuki, T. (2006) Identification of the residues involved in the unique serine specificity of *Caenorhabditis elegans* mitochondrial EF-Tu2. *Biochemistry*, **45**, 10920–10927.
53. Benelli, D. and Londei, P. (2011) Translation initiation in Archaea: conserved and domain-specific features. *Biochem. Soc. Trans.*, **39**, 89–93.
54. Dreher, T.W., Uhlenbeck, O.C. and Browning, K.S. (1999) Quantitative assessment of EF-1 $\alpha$ .GTP binding to aminoacyl-tRNAs, aminoacyl-viral RNA, and tRNA shows close correspondence to the RNA binding properties of EF-Tu. *J. Biol. Chem.*, **274**, 666–672.
55. Gromadski, K.B., Wieden, H.J. and Rodnina, M.V. (2002) Kinetic mechanism of elongation factor Ts-catalyzed nucleotide exchange in elongation factor Tu. *Biochemistry*, **41**, 162–169.
56. Juhling, F., Morl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. and Putz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.*, **37**, D159–D162.