The EMBO Journal vol.12 no.7 pp.2959-2967, 1993

Selection for active *E.coli* tRNA^{Phe} variants from a randomized library using two proteins

Ellen Tinkle Peterson, Jutta Blank[1], Mathias Sprinzl[1] and Olke C.Uhlenbeck

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA and [1]Lehrstuhl für Biochemie, Universität Bayreuth, D-8580 Bayreuth, Germany

Communicated by J.A.Steitz

***In vitro* selection was used to isolate active *Escherichia coli* tRNA^{Phe} variants from randomized libraries. Functional tRNAs were first selected by multiple rounds of binding to *Escherichia coli* phenylalanyl-tRNA synthetase. These variants were then aminoacylated and selected for affinity to elongation factor-Tu. By randomizing potential recognition nucleotides, the importance of residues U20, G34, A35, A36 and U59, previously identified to be required for specific recognition by *E.coli* phenylalanyl-tRNA synthetase (FRS), was confirmed. However, the sequences of several active variants imply that the wild-type tertiary interactions G10-C25-U45 and A26-G44 are not required for recognition, as previously suggested. Selection of functional tRNAs from a second library randomized at positions normally involved in conserved tertiary interactions revealed new combinations of nucleotides at these positions, suggesting the presence of novel tertiary interactions. In both libraries, active sequences containing deletions were isolated. Taken together, it is clear that FRS is active with substrates having an unexpectedly broad sequence diversity. Finally, the potency of this method is illustrated by the identification of a second class of variants that was isolated by virtue of the presence of an impurity in the FRS preparation.**

Key words: aminoacyl-tRNA synthetase/elongation factor-Tu/recognition/selection/tRNA

Introduction

The three-dimensional structure of tRNA is partially maintained by nine different 'tertiary' interactions involving conserved and semi-conserved nucleotides (Quigley and Rich, 1976). In addition to this structural role, it is possible that these residues participate in sequence-specific contacts with aminoacyl-tRNA synthetases. For yeast phenylalanyl-tRNA synthetase (FRS), this is unlikely to be the case since replacement of tertiary interactions with combinations of residues present in other tRNAs did not alter aminoacylation rates (Sampson *et al.*, 1990). In contrast, FRS from *Escherichia coli* appears more sensitive than the yeast enzyme to substitutions at certain nucleotides involved in tertiary interactions. *In vitro* aminoacylation studies with point mutants of *E.coli* tRNA^{Phe} and purified *E.coli* FRS (Peterson and Uhlenbeck, 1992) revealed that in addition to substitutions in three single-stranded regions, replacement of two adjacent tertiary interactions in the central core (G10-C25-U45 and A26-G44) with the corresponding nucleotides present in other tRNAs did not lead to active tRNA variants. Since the mutant tRNAs seemed to fold normally by the criterion of cleavage by lead, it appeared that one or more functional groups in these two tertiary interactions were important for aminoacylation. While interaction of nucleotides in highly structured regions with synthetases is not unprecedented (Pütz *et al.*, 1991; Hayase *et al.*, 1992), it is possible that the local structure of the mutants is disrupted in a way that prevents correct orientation of the other nucleotides that contact the synthetase. Because it has proven difficult to mutate tertiary interactions in a rational fashion (Sampson *et al.*, 1990), further directed mutation in this region to find an active tRNA would be laborious and possibly futile. It is evident that a better method to identify active tRNA variants with substitutions in highly structured regions is needed.

In vitro selection has been used to successfully isolate RNA molecules from a pool of randomized sequences that bind a particular protein (Tuerk and Gold, 1990; Bartel *et al.*, 1991; Tsai *et al.*, 1991, 1992; Schneider *et al.*, 1992, 1993; Tuerk *et al.*, 1992). This technique permits the isolation of high-affinity sequences from a pool of heterologous RNAs by cycles of partitioning and amplification of the selected sequences. This method is ideal for identifying novel functional tRNA sequences since many variants can be tested in a single experiment, and mutants with multiple substitutions can be isolated. We will describe two types of *in vitro* selection to isolate active tRNA molecules from a randomized library based on their interaction with *E.coli* FRS. Initially, the libraries were enriched for those variants that bound FRS. Because elongation factor-Tu (EF-Tu)-GTP binds aminoacylated tRNAs, the variants that could be aminoacylated by FRS were then isolated by EF-Tu affinity chromatography. Since this method has not yet been used to examine tRNA-synthetase interactions, we wanted to establish first that the known recognition nucleotides could be identified by this technique. Thus, a library was designed to confirm the participation of the nucleotides determined previously in recognition. This library could potentially reveal mutations not tested by standard mutagenesis techniques that can substitute for the wild-type nucleotides. In contrast to yeast aspartyl-tRNA synthetase (Puglisi *et al.*, 1993), assessment of the activity of tRNA^{Phe} variants suggests that *E.coli* FRS is sensitive to large structural changes in tRNA. Thus, tRNA-like structures with alternative sequences can also be identified by this technique. Many of the conserved nucleotides were randomized in a second library designed to elucidate RNAs with novel sequences that maintain activity at a level similar to that of wild-type tRNA^{Phe}.

Results

Library design and selection

The two libraries derived from *E.coli* $tRNA^{Phe}$ are illustrated in Figure 1. Library A, which contains 14 randomized nucleotides or $\sim 2.7 \times 10^8$ sequences, was primarily designed to test the validity of this method. Nine of the randomized nucleotides were at positions previously identified as recognition nucleotides (Peterson and Uhlenbeck, 1992). Of the other five positions that were randomized, three (16, 17 and 60) were at positions where substitution does not significantly affect the rate of aminoacylation, and thus can be considered as controls. The role of the nucleotides at positions 32 and 38 had not previously been tested directly, and they were randomized to determine the extent of the contact made by FRS with the anticodon loop.

Library B contains 10 randomized positions (1.0×10^6 sequences) that are concentrated in one region of the folded structure where several tertiary interactions occur. Three completely conserved interactions, two between the D-loop and T-loop (18−55, 19−56), and one within the T-loop (54−58), were randomized to identify possible alternate combinations of residues that are allowed in active variants of $tRNA^{Phe}$. In addition, the neighboring single-stranded nucleotides at 16 and 17 in the D-loop, and 57 in the T-loop, were randomized to permit the creation of a new context that may be required for the formation of different tertiary interactions. Finally, the nucleotide at position 53 participating in the conserved base pair at the end of the T-stem was also randomized.

Both libraries were derived from a parent sequence that deviates slightly from wild-type *E.coli* $tRNA^{Phe}$. Although the C3−G70 base pair in the natural sequence has been maintained, in contrast to many of the constructs generated previously (Peterson and Uhlenbeck, 1992), the two base pairs at the top of the anticodon stem have been changed from G-C to A-U. This change was made in order to avoid the formation of an inactive conformation of $tRNA^{Phe}$ that appears to require the four contiguous G-C pairs (Dichtl *et al.*, 1993; E.T.Peterson and O.C.Uhlenbeck, unpublished data). This mutation results in only a modest decrease in k_{cat}/K_M for aminoacylation by FRS (Peterson and Uhlenbeck, 1992) and is therefore considered to be a near-wild-type derivative of the natural sequence. Library B also has an additional substitution, U60C, to introduce the lead binding site so that the structures of the variants can potentially be assessed by lead cleavage (Behlen *et al.*, 1990; Peterson and Uhlenbeck, 1992). This substitution has also been shown to have little effect on aminoacylation.

In order to select RNA species that bind FRS, a technique to separate tRNA−synthetase complexes from free tRNA had to be chosen. Although several methods have been used to isolate tRNA−synthetase complexes (Lagerkvist *et al.*, 1966; Okamoto and Kawade, 1967; Seifert *et al.*, 1968; Yaniv and Gros, 1969; Goodman and Schwartz, 1988), nitrocellulose filter binding was chosen since it has been fairly well characterized for tRNA−synthetase interactions (Yarus and Berg, 1967, 1969, 1970; Bartmann *et al.*, 1975; Park *et al.*, 1989) and is amenable to manipulations required for *in vitro* selection (Tuerk and Gold, 1990). As has previously been reported by Bartmann *et al.* (1975), the $tRNA^{Phe}$−FRS complex is efficiently retained on nitrocellulose filters only below pH 6.0. This result is

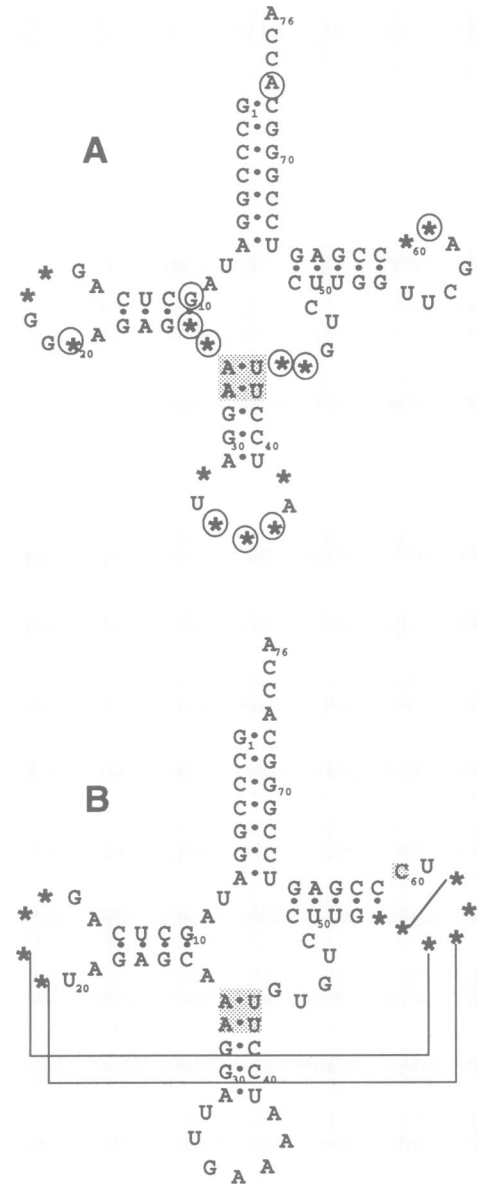

Fig. 1. The secondary structures of library A (**A**) and library B (**B**) derived from the sequence of *E.coli* $tRNA^{Phe}$. Randomized positions are indicated by an asterisk. The shaded nucleotides differ from the sequence of *E.coli* $tRNA^{Phe}$. The circled residues in (A) are required for recognition by *E.coli* FRS (Peterson and Uhlenbeck, 1992). The lines connect positions in (B) that form tertiary interactions in yeast $tRNA^{Phe}$.

partially explained by the observation that some synthetases, including *E.coli* FRS, bind their cognate tRNAs more tightly at low pH (Hélène *et al.*, 1969; Yaniv and Gros, 1969; Farrelly *et al.*, 1971; Kosakowski and Böck, 1971; Bartmann *et al.*, 1975). Indeed, the K_D for the interaction between *E.coli* $tRNA^{Phe}$ and FRS, determined in a neutral buffer by fluorescence titration (Bartmann *et al.*, 1974) or gel filtration (Bartmann *et al.*, 1975), is ~100 nM, while a much tighter interaction is suggested at pH 5.4 since the K_D for the *E.coli* $tRNA^{Phe}$ transcript is 1.1 nM, as determined by nitrocellulose filter binding (Figure 2).

The enzyme must exhibit specificity under the binding conditions to efficiently partition active tRNAs from inactive tRNAs. Several buffers were tested with the goal of finding conditions that maximize the difference in K_D between

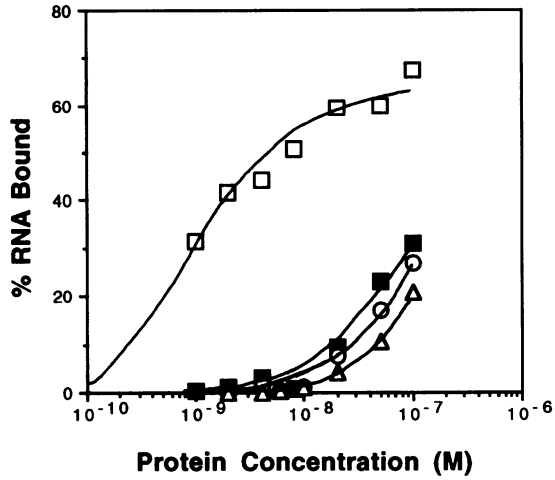

Fig. 2. The binding of tRNA transcripts to FRS in 2 mM DTT, 40 mM NaOAc (pH 5.4), 15 mM $MgCl_2$ and 25 mM KCl at 25°C using nitrocellulose filters. Wild type (open squares); A26G, G44A (closed squares); U59C (open circles); G34A (open triangles). All transcripts contain C3G, G70C. The binding curve for the wild-type transcript follows a simple binding equilibrium with K_D of 1.1 nM and retention efficiency of 62%.

wild-type tRNA^{Phe} and three mutants (U59C, A26G-G44A and G34A) known to have increased K_M values for aminoacylation. It was found that in sodium acetate buffer at pH 5.4, the addition of $MgCl_2$ and KCl to the binding buffer has only a modest effect on the K_D for the wild-type tRNA, but greatly improves discrimination between wild-type and mutant tRNAs. While little specificity is seen in acetate buffer and dithiothreitol (DTT) alone, the addition of 15 mM $MgCl_2$ and 25 mM KCl increases the K_D values of the mutants to ~100-fold the value of the wild type (Figure 2). Under these conditions, tRNA^{Phe} can also be aminoacylated with a k_{cat} ~2 orders of magnitude slower than observed under neutral conditions (data not shown), similar to what was seen by Kosakowski and Böck (1971). Thus, these conditions were considered satisfactory for selection experiments.

The affinity of the library for FRS indicates that most of the molecules in the library bind quite poorly to FRS. As shown in Figure 3A for library A, the parent tRNA^{Phe} has a K_D of 4.4 nM, while the K_D values for both libraries are ~500 nM, which is comparable to the K_D of the weakest-binding mutant in Figure 2. In order to isolate members of the library which bind FRS tightly, 100 nM FRS and 1 μM RNA were used in the first two rounds of selection and amplification (Irvine *et al.*, 1991). At this stage, both libraries exhibit enrichment (Figure 3A). In order to continue the enrichment procedure, the protein concentration was reduced to 50 nM for the next two rounds, 5 nM for the fifth round, and 1 nM for the sixth round. In each round, the RNA concentration was in excess over the protein by at least 10-fold to ensure competition for binding sites. After six rounds of selection, the apparent K_D values for both libraries are reduced by ~2 orders of magnitude and are roughly comparable to wild type. However, since the binding curves for the libraries are broader than that for the wild type, the presence of species with differing affinities is suggested.

At different stages of enrichment, the libraries were also assayed for aminoacylation activity under conditions

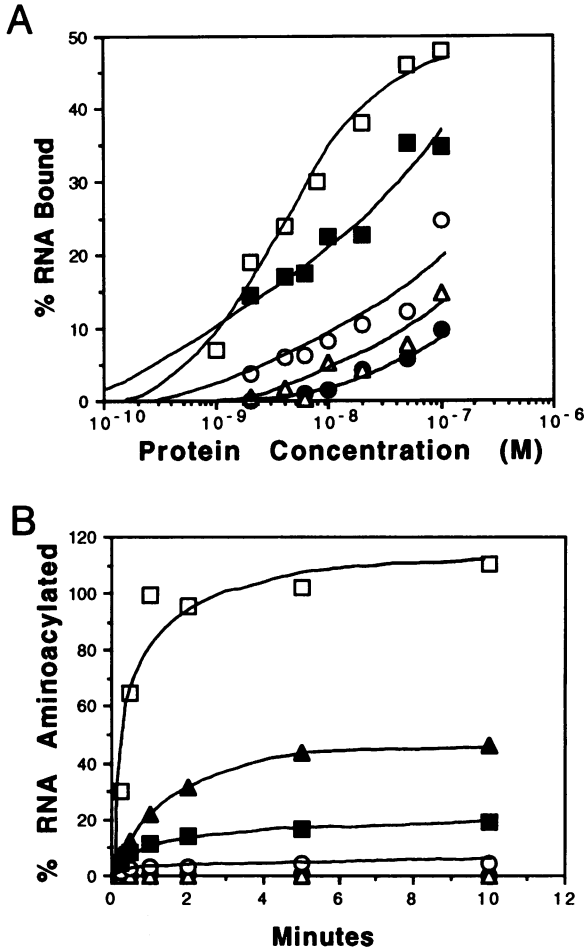

Fig. 3. Enrichment of library A by binding and aminoacylation compared with parent tRNA, C3G, G70C, G27A, G28A, C42U, C43U (open squares). (**A**) Binding of RNA libraries after 0 (closed circles), 2 (open triangles), 4 (open circles) and 6 (closed squares) rounds of selection by nitrocellulose filter binding to FRS. The binding curve for the parent tRNA follows a simple binding equilibrium with K_D of 4.4 nM and retention efficiency of 49%. (**B**) Aminoacylation of RNA libraries after 0, 2, 4, and 6 rounds of selection by FRS binding on nitrocellulose, and after selection by aminoacylation and EF-Tu binding (closed triangles). 100% aminoacylated is defined as 1120 pmol phe/A_{260}. The data for both binding and aminoacylation for library B are similar to those for library A.

previously used to evaluate tRNA substrates (Peterson and Uhlenbeck, 1992). As seen in Figure 3B for library A, little aminoacylation is observed until the fourth round of selection. After six rounds, the mixture is aminoacylated to ~20% the level of the parent tRNA. Thus, selection by binding at pH 5.4 also enriches the populations for aminoacylation at pH 7.5.

The affinity of ET-Tu–GTP for aminoacylated tRNA was exploited to further enrich the RNA populations for aminoacylation activity. A total of 30 pmol of RNA from each library was used in an aminoacylation reaction and applied to 15 μl of a Sepharose column containing 50 pmol of active immobilized *Thermus thermophilus* EF-Tu–GTP (Derwenskus *et al.*, 1984). While 70–100% of the wild-type phe-tRNA^{Phe} can be bound and eluted from the column, only 40–50% of the aminoacylated tRNAs in the library were retained on the column. This suggests that some tRNAs can be aminoacylated, but do not bind well to EF-Tu. Thus, the EF-Tu column appears to select on the basis

of variant sequence, as well as the presence of an amino acid. After the bound population was amplified with reverse transcriptase and polymerase chain reaction (PCR), the mixture aminoacylated to 50% the level of the parent $tRNA^{Phe}$ (Figure 3B), corresponding to an additional 2.5-fold enrichment for active variants in the seventh round of selection.

Analysis of individual variants

The amplified DNA from the seventh round of selection was cloned and RNAs from 15 library A clones and 23 library B clones were assayed for FRS binding. Twelve of the RNAs from library A and 17 from library B were found to have K_D values ≤ 15 nM, and the remainder bound considerably less well. This is consistent with the binding properties of the sixth-round libraries which suggested a substantial variation in K_D values, with some variants binding at least as well as the wild-type tRNA.

The tight-binding RNAs from both libraries were tested for aminoacylation activity under conditions which are subsaturating for the parent tRNA. Of the 12 RNAs from library A, six were found to have k_{cat}/K_M values within two-fold of the parental tRNA, three aminoacylated more slowly and aminoacylation was not detectable for the remaining three. Similarly, of the 17 RNAs from library B, 11 aminoacylated very well, four less well and two did not exhibit aminoacylation activity. The fraction of clones with aminoacylation rates within 2-fold of the parent (6 of 15 for library A, and 11 of 23 for library B) is consistent with the percentage of aminoacylatable substrates suggested by the activity of the libraries prior to cloning.

The sequences, K_D values and relative k_{cat}/K_M values for aminoacylation for all the tRNAs that showed tight binding activity are shown in Table I. Since many of the variants contained deletions and mutations outside the randomized regions, the sequences were aligned to maximize the number of residues identical to the parental tRNA. The fact that several of the independently selected clones have identical sequences (one sequence, A7, was found three times in

Table I. Variants with $K_D \leq 15$ nM

Variant no.	Sequence of randomized region						K_D (nM)	rel. k_{cat}/K_M
	D-loop	D-stem	Anticodon stem−loop	Variable loop	T-stem	T-loop		
Parent	16	22	26	44	49	54		
$tRNA^{Phe}$ [a]	UCGGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUUGG	UUCGAUU	4[b]	(1.0)[b]
Library A								
	** *	*	* * *** *	**		**		
A1	CGGGUA	GAGC	AAAGGAUUGAAACUCCUU	GCG C	CUUGG	UUCGAUU	2	1.5
A2	UUGGUA	GAGU	GAAGGGU GAA CUCCUU	AGGUC	CUUGG	UUCGAUG	4	1.3
A3	UUGGUA	GAGC	AAAGGGUUGAAA UCCUU	CCGUC	CUUGG	UUCGAUG	12	1.3
A4	ACGGUA	GAGU	GAAGGAGUGAAAAUCCUU	AGGUC	C GG	UUCGAAU	9	0.8
A5	CUGGUA	GAGC	AAAGAUUUGAA CUCCUU	GUGUC	CUUGG	UUCGAUC	2	0.8
A6	CAGGUA	GAGC	AAAGGAGUGAAAAUCCUU	AUGUC	C GG	UUCGAUC	7	0.5
A7 (3)	UAGGCA	GAGC	GAAGAUUCGAAA UCCUU	AGGUC	CUUGG	UUCGAUC	4	0.2
Library B								
	* * * *				*	***** C		
B1	GCAGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUCGG	UGCGAUC	1	1.4
B2 (2)	UUAGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUUGG	CGCGAUC	2	1.3
B3 (3)	UUAGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUUGG	CGCAAUC	1	1.3
B4	UAUGUA	GAGC	AAAGGUUUGAAAAUCCUU	GUGUC	CUC G	CCCG UC	1	1.2
B5	UUAGUA	GAGC	AAAGGAUUGAAAA CCUU	GUGUC	CUCGG	UGCGUUC	0.5	1.1
B6	UUAGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUUGG	GGCGGUC	7	0.7
B7	UUAGUA	GAGC	AAAGGAUUGAAAAUCCUU	GUGUC	CUUGG	CGCAGUC	8	0.6
B8	GUAGUA	GAGC	AGAGGAUUGAAAA CCUU	GUGUC	CUUGG	GGCGGUC	1	0.5
B9	UUGGUA	GAGC	AAAGGAUUGAAACUCCUU	GCGUC	CUCGG	CCCAGUC	3	0.3
B10	AA GUA	GAGC	AAAGGAUUGAAAA CCUU	GUGUC	CUUGG	GGCGGUC	12	0.2
B11	AAUGUA	GAGC	AAAGGAUUGAAAA CCUU	GUGUC	CUUGG	GCCGUUC	2	0.2
B12	UUGGUA	GAGC	AAAGGAUUGAAAAUCCCU	GUGUC	CUCGG	UCCGGUC	4	0.1
B13	UCAUAA	GAGC	AAGGAUUGAAAA CCUU	GUAUC	CUUGA	GCAGGUC	6	<0.01
Contaminant selected								
A8	A GG A	GAGA	GAAAGACUCUGAGUCCUU	AGGUC	CUUGG	UUCCGAC	15	<0.01
A9	UCGAUA	GAGU	GA G UU AUAUUCCUU	AGGUC	CUUGG	UUCCGACC	6	<0.01
A10	GAGGUA	GAGU	AAAGAAUUGGUAGUCCUU	ACGUC	CUUGG	UU AUC	12	<0.01
B14	GGCCUA	G	AAGGAUUGAAAAAAUCCUU	G GUC	CUUGU	UUUCCUC	0.5	<0.01

An asterisk represents the residues that were randomized in each library. The bold-faced nucleotides indicate sequence changes from the parent. The sequences that correspond to the primer binding sites at the 5′ and 3′ ends are not shown. The numbers in parentheses next to the variant number tell how many variants were found with the indicated sequence.

[a]Sequence of the parent $tRNA^{Phe}$ for library A. $tRNA^{Phe}$ nucleotides are substituted at the positions that were randomized for reference. The sequence of the parent for library B differs only by a U60C substitution.

[b]K_D and k_{cat}/K_M values for C3G, G70C, G27A, G28A, C42U, C43U variant.

library A and two sequences, B2 and B3, were found two and three times, respectively, in library B) suggests that the sequence complexity of the libraries after seven rounds of selection is significantly reduced.

The six different sequences from library A with the highest activity for both binding and aminoacylation are shown in cloverleaf form in Figure 4. It is striking that the wild-type nucleotides at four of the single-stranded residues previously identified as being important for aminoacylation by mutational studies (U20, G34, A35 and A36) are found in each of the six selected variants. There is also a strong (5 out of 6) bias for U59, a fifth single-stranded recognition nucleotide. As expected, little bias is seen in the three single-stranded 'control' positions 16, 17 and 60. Owing to the deletions often encountered in the anticodon loop, the importance of the two previously untested residues 32 and 38 is somewhat ambiguous. U32 is present in four of the six sequences, suggesting that its identity may be important. The deletions in the loop 3′ to the anticodon nucleotides make identification of the nucleotide that occupies position 38 difficult. In the three active variants without deletions in the anticodon loop in the region of nucleotide 38, two have wild-type A38.

We had previously shown that the substitution of residues involved in the two adjacent tertiary interactions G10–C25–U45 and A26–G44 with residues found in other tRNAs did not result in active substrates for FRS (Peterson and Uhlenbeck, 1992). However, several of the active variants in Figure 4 have nucleotides at these positions that had not previously been tested in combination. In addition to the wild-type sequence, the common phylogenetic variants G10–U25–G45, G26–A44 (variants A2, A4) and A26–A44 (A6), and the relatively rare combinations G10–C25–C45 (A1, A3) and A26–C44 (A3), were found to occupy the positions attributed to these tertiary interactions.

The eight sequences from library B with aminoacylation rates similar to the parent tRNA are shown in Figure 5. The variants exhibit several remarkable features with respect to the nucleotides found at the positions normally involved in conserved interactions. In the seven sequences where the T-loop residues can be unambiguously aligned, the nucleotides at the randomized positions 19 and 56 were always the wild-type G19 and C56. The neighbouring positions, normally forming an interaction between the conserved nucleotides G18 and Ψ55, are exclusively occupied by the nucleotides A18 and G55. Three of the four possible nucleotides are found at both positions 54 and 58 within the T-loop. Finally, all the active variants contained G53, thereby reforming the G53–C61 base pair.

The limited variety of nucleotides found in the single-stranded regions that were randomized in library B suggests that certain residues are preferred at these positions. Interestingly, there seems to be a preference for U at both positions 16 and 17 in this library. While the wild-type residue at position 16 is U (D in the modified tRNA), a C is the naturally occurring nucleotide at position 17. The nucleotide at position 57 is found to be either A or G in all the active isolates, in agreement with the requirement for a purine at this position in natural tRNA sequences.

Nucleotide deletions were often found in the variants from both libraries. This is presumed to result from the many rounds of amplification and is frequently seen in selection

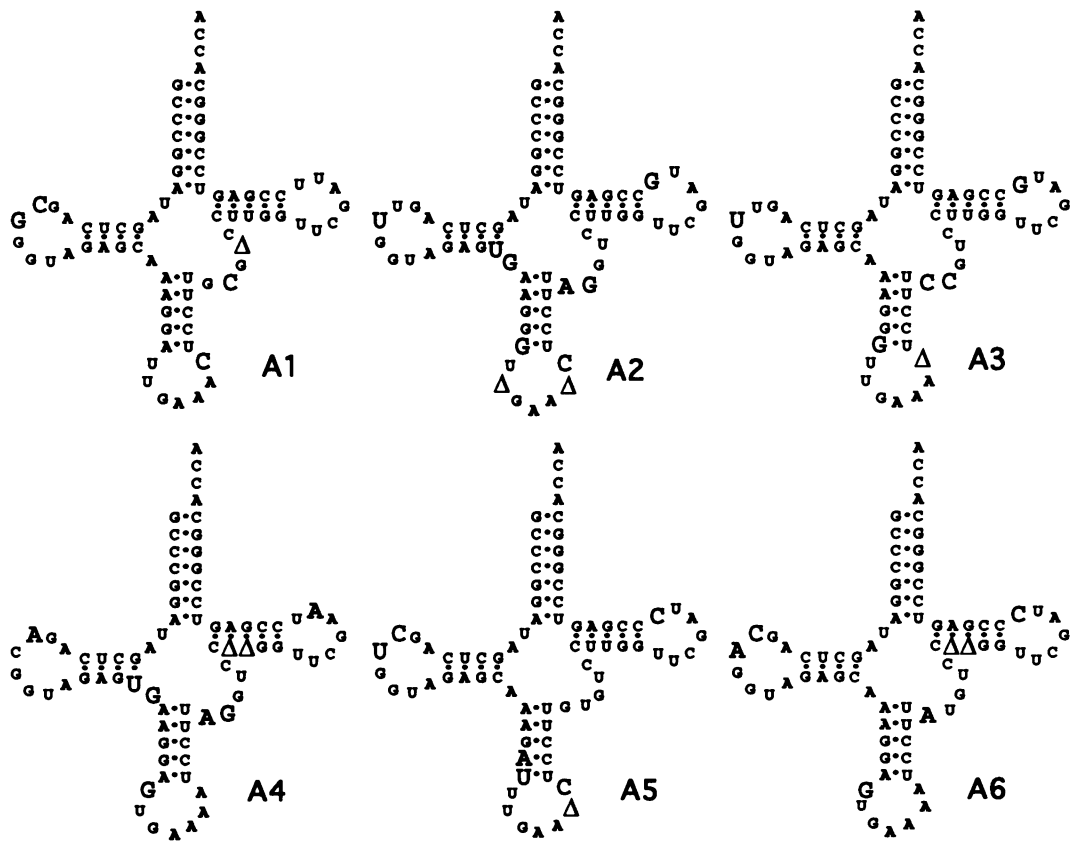

Fig. 4. The cloverleaf structures for the six active variants from library A. The nucleotides that differ from the parent are indicated by large capital letters. Deletions are represented by a Δ.

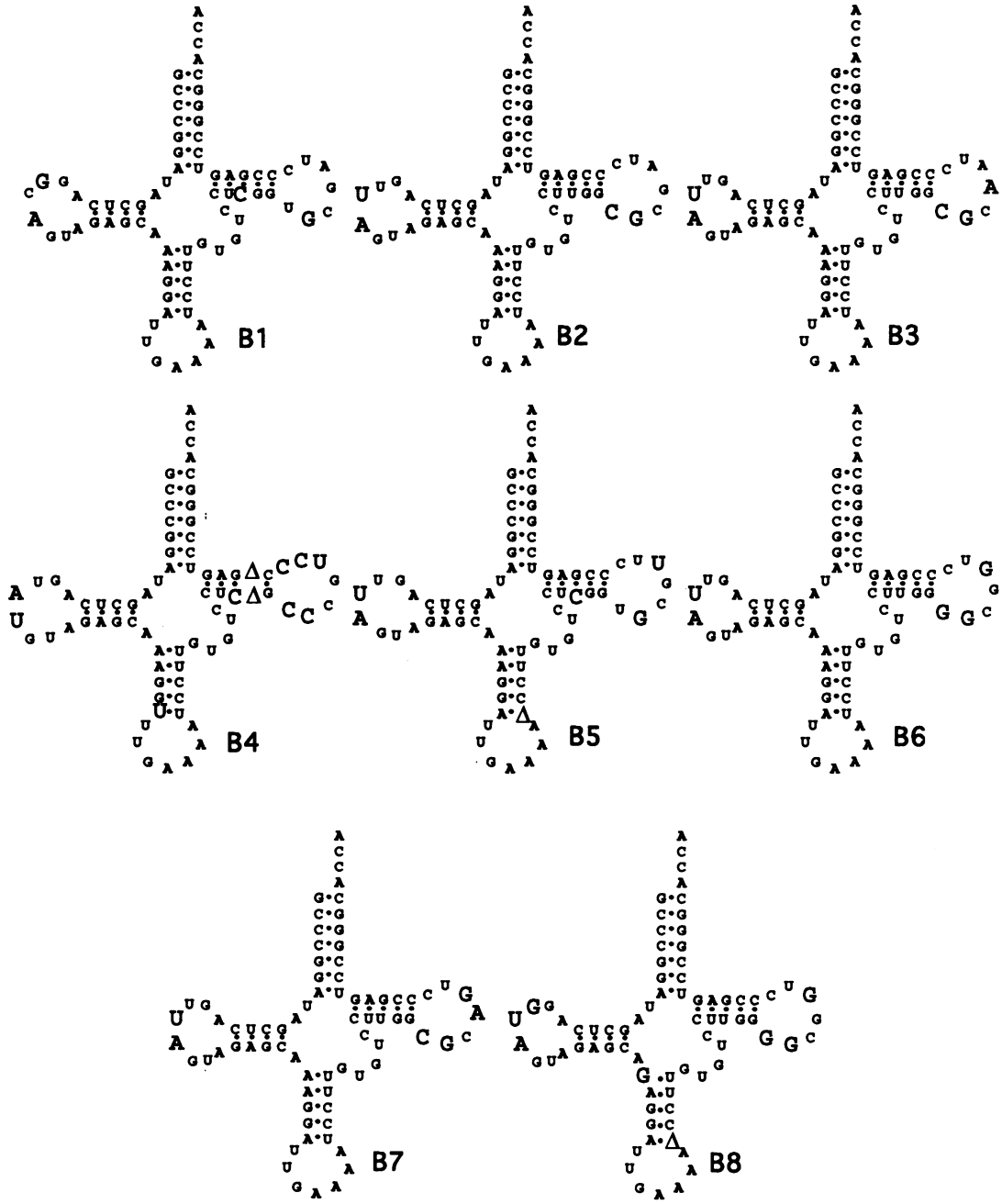

Fig. 5. The cloverleaf structures for the eight active variants from library B.

strategies of this type (Pan and Uhlenbeck, 1992). Since FRS contacts an extensive part of $tRNA^{Phe}$, including nucleotides in the anticodon loop, variable pocket and acceptor end, it is remarkable that deletions do not greatly interfere with activity. The most surprising examples are the variants with deletions in the anticodon stem–loop. Since the anticodon nucleotides themselves appear to be strictly required for activity, it is striking that the adjacent nucleotides are so frequently deleted. These deletions may be accommodated by nucleotide changes at adjacent regions of the molecule in some cases. Deletions were also found in the variable loop (variant A1) and in the T-stem and loop (A4, A6 and B4) of active isolates.

The five variants (A8, A9, A10, B13 and B14) that appeared to bind tightly to FRS, but could not be detectably aminoacylated, were further characterized. Four of the five sequences (A8, A9, A10 and B14) do not contain the phenylalanine (GAA) anticodon at the expected position thought to be essential for binding. Competition experiments indicated that these tRNAs do not compete with $tRNA^{Phe}$ for binding to FRS, but do compete with one another (data not shown). This suggests that they bind to an independent site. RNA excess binding experiments revealed that these four variants saturate at a 10-fold lower RNA to protein ratio than does $tRNA^{Phe}$ (data not shown). Thus, the independent site is at a lower concentration in the protein preparation. It is likely that these RNAs bind to an impurity in the FRS preparation used for selection. Another FRS preparation (kindly provided by Dr O.Lavrik, Institute of Bioorganic Chemistry, Novosibirsk, Russia) purified by a different

protocol, and unlikely to contain the same contaminants, did not bind variants of this class. While not yet identified, SDS gel analysis of the FRS sample used for selection reveals two polypeptides present at 5–10% of the total protein that are candidates for the selection of a distinct class of sequences.

The fifth sequence that is not efficiently aminoacylated, B13, contains the essential anticodon sequence. This variant will bind both independent preparations of FRS efficiently, suggesting that it binds to FRS and not an impurity. Furthermore, in contrast to the four variants described above, B13 competes with tRNA^{Phe} for binding (data not shown). Thus, this variant appears to bind FRS at low pH in a manner similar to the other variants that are fully active. Because this particular variant is not detectably aminoacylated and will not compete with tRNA^{Phe} under aminoacylation conditions (data not shown), it is possible that it was selected to bind FRS only at low pH and does not bind as well to FRS under physiological conditions.

Discussion

We have used an *in vitro* selection method to isolate variants of *E.coli* tRNA^{Phe} that are bound and aminoacylated by cognate phenylalanyl-tRNA synthetase. Initial mixtures of 10^6-10^8 sequences were first fractionated on the basis of their ability to bind *E.coli* FRS at low pH. After six rounds of enrichment and amplification, the variants that bound well were aminoacylated under neutral conditions and those that acquired an amino acid were selected on a *T.thermophilus* EF-Tu affinity column. The overall selection procedure was therefore based on the RNA interaction with FRS under two very different conditions, and the interaction with EF-Tu. The selection with EF-Tu was less stringent than for FRS since the protein is present on the column in a fairly high concentration. Finally, since transcription products after each round of amplification were selected by size, the active variants must also be approximately the same length.

The selection procedure resulted primarily in RNAs which bound FRS and aminoacylated with rates similar to the wild-type tRNA^{Phe} sequence. However, some of the individual RNAs isolated were more active, while a few had far less activity than the parent sequence both with respect to binding and aminoacylation. Thus, analysis of each individual variant is crucial for interpretation of the sequences isolated. The libraries were enriched >100-fold, as judged by a decrease in K_D. The identification of multiple identical clones in both libraries suggests that the sequence complexity of the selected mixture is significantly reduced as well. If we assume that there is no bias for the duplicate sequences during amplification and cloning, we estimate that the enriched libraries only contain a few hundred sequences. Taken together, the technique was clearly successful in obtaining active variants and thus could be used with other synthetases or with more complex libraries. Interestingly, some of the isolates were found to bind a protein other than FRS in the preparation, illustrating the potency of the method.

Our previous experiments have shown that *E.coli* FRS requires a folded tRNA structure to correctly position a number of recognition nucleotides that are far apart in the tertiary structure. It is therefore reasonable to assume that the active variants selected from libraries A and B must do the same. However, even with an understanding of the structural requirements, a straightforward interpretation of the variant structures from their sequences is complicated by the fact that most of the selected molecules have numerous substitutions and many have one or more deletions. Although we have previously argued that the tertiary structure of *E.coli* tRNA^{Phe} is similar to that of yeast tRNA^{Phe} (Peterson and Uhlenbeck, 1992), it is unclear whether the nucleotide interactions in the variants can be accurately deduced by simply aligning the sequences.

Because of the previous data available regarding substitutions at the recognition nucleotides, the sequences of the active tRNAs from library A are most easily interpretable. The selection of wild-type nucleotides at positions 20, 34, 35, 36 and 59, and the absence of selection for the wild-type nucleotides at positions 16, 17 and 60, is consistent with our previous mutagenesis experiments. Indeed, if the recognition nucleotides had not been known, the very limited number of selected active tRNAs would have led to the correct conclusion about the role of these positions in recognition. Interestingly, the requirement for the wild-type tertiary interactions G10–C25–U45 and A26–G44 suggested by our mutagenesis experiments seems to be refuted by the isolation of active tRNAs with changes in these regions. It appears that only specific combinations of these five nucleotides can lead to an active substrate for FRS. Other combinations previously tested, although quite common among other tRNAs, do not function in *E.coli* tRNA^{Phe}. In contrast to the designed mutants, some of the interactions in the selected variants are rarely observed in natural tRNAs. The most likely explanation for these results is that while the synthetase does not contact these nucleotides directly, the five wild-type nucleotides and the selected combinations form tertiary interactions that restrict the overall structure in a manner that is favorable for FRS. It is, however, important to remember that the tRNAs with novel nucleotide combinations contain many other changes that may have been co-selected to maintain the overall tRNA structure. Future experiments must therefore test the importance of the putative new tertiary interactions by mutation. It will also be interesting to see whether the novel combinations can be transplanted into the wild-type tRNA^{Phe}.

Since both yeast and *E.coli* FRS and *E.coli* EF-Tu are sensitive to certain structural changes in tRNA^{Phe} (Sampson *et al.*, 1990; Peterson and Uhlenbeck, 1992; I.Nazarenko, K.Harrington and O.Uhlenbeck, in preparation), it is not surprising that the active tRNAs from library B show selection for specific nucleotides at five of the 10 randomized positions. It is, however, surprising that the wild-type nucleotides are not always preferred. The selection of the wild-type G53 is easily understood by the fact that the highly conserved G53–C61 is essential for the folding of the T-loop (Quigley and Rich, 1976; Romby *et al.*, 1987; Behlen *et al.*, 1990) and while inversion at this base pair in yeast tRNA^{Phe} does not affect interaction with yeast FRS (Sampson *et al.*, 1992), the affinity for *E.coli* EF-Tu is reduced (I.Nazarenko, K.Harrington and O.Uhlenbeck, in preparation). The strong selection for the wild-type residues G19 and C56 makes it tempting to conclude that the tertiary Watson–Crick pair formed by these nucleotides is essential for synthetase and/or EF-Tu binding. Interestingly, mutational experiments in yeast tRNA^{Phe} indicate that this pair can be inverted to C19–G56 without significantly affecting folding, yeast FRS activity or *E.coli* EF-Tu binding (Behlen *et al.*, 1990;

Sampson *et al.*, 1990; I.Nazarenko, K.Harrington and O.Uhlenbeck, in preparation). Finally, the nucleotides A18 and G55 are found in all the active variants at the positions normally occupied by the highly conserved tertiary interaction between G18 and U55 in the wild-type tRNA. While this result suggests that a novel tertiary interaction forms between these two nucleotides, it is not immediately obvious why this combination is preferred over the wild-type pair. Because this tertiary interaction contains a modified nucleotide in the natural tRNA sequence, it is possible that A18–G55 compensates for the absence of this modification.

The pattern of selected residues in library B might be best understood in terms of correlated changes. Since the randomized nucleotides are all located in one region of the tertiary structure of tRNA and few are needed for direct interaction with FRS or EF-Tu, perhaps the sequences were selected as a group in such a way that the general shape of the $tRNA^{Phe}$ was maintained. Thus, G19–C56 was selected with A18–G55 and other residues in this region to result in tRNA-like folding. Support for this view comes from comparing the results of randomizing positions 16 and 17 in the two libraries. Mutagenesis experiments indicate that changes at these positions in either yeast $tRNA^{Phe}$ or *E.coli* $tRNA^{Phe}$ do not affect folding (Behlen *et al.*, 1990) or FRS activity (Peterson and Uhlenbeck, 1992; Sampson *et al.*, 1992). Furthermore, specific nucleotides at these positions are unlikely to be required for EF-Tu binding since these positions vary among tRNA sequences. While no nucleotide bias was seen at positions 16 and 17 in library A, which contains wild-type interactions between the T-loop and D-loop, there is a strong preference for U residues at both positions in library B. This preference appears to be due to the identities of the proximal nucleotides since all the active variants from library B deviate from the wild type at nucleotides 18, 55 and 60. The function of the novel sequences at positions 54 and 58 may also depend on their context. It will be interesting to determine if the novel tertiary interactions suggested can indeed be superimposed on the wild-type structure.

In many respects, our results from *in vitro* selection experiments with RNA resemble those of others. Like with T4 DNA polymerase, U1-snRNP-A protein and R17 coat protein (Tuerk and Gold, 1990; Tsai *et al.*, 1991; Schneider *et al.*, 1992), the recognition nucleotides could be deduced from the sequences of the selected RNAs and rationalized with respect to the known binding site. In addition, proper RNA folding is important for the interaction and new folding principles are suggested by the sequences of our active variants, reminiscent of what was found for HIV RT and Rev protein (Bartel *et al.*, 1991; Tuerk *et al.*, 1992). Despite requiring interaction with two different proteins involved in translation, the sequence diversity obtained among the active tRNA sequences is much broader than found in biology. Many conserved nucleotides were substituted and deletions were found in spatially conserved regions without impairing the activity. It is not yet known if the variants are active in the other reactions required of tRNA prior to aminoacylation and following EF-Tu binding in the cell. The reactions at the levels of processing and translation may depend upon different characteristics of tRNA, and collectively define the features conserved in tRNA. An *in vivo* selection resulted in the identification of $tRNA^{Ala}$ variants that also have unexpected nucleotide changes (Hou and Schimmel, 1992). It will be interesting to determine to what extent tRNA can be altered and still maintain a level of activity similar to that of naturally occurring tRNA in light of the absence of significant phylogenetic variability among cytoplasmic tRNA species.

Materials and methods

Construction of libraries

The DNA oligonucleotides used as templates in the initial transcription were synthesized such that ~25% of each nucleotide was incorporated at the positions indicated by an asterisk in Figure 1. The template for library A consisted of the 18 nucleotide T7 promoter top strand (Milligan *et al.*, 1987) annealed to a 93-nucleotide oligo comprised of the T7 promoter and the complement of the tRNA sequence: 5′TGGTGCCCGGACTCGGNNTCGAACCAAGGACNNAAGGANTNNNANTCCTTNNCTCTNCCNNCTGAGNTATCCGGGCTATAGTGAGTCGTATTA3′, where the randomized nucleotides are indicated by N. The DNA template for the library B was constructed from a 57-mer containing the T7 promoter sequence and the tRNA sequence from nucleotides 1 to 40: 5′TAATACGACTCACTATAGCCCGGATAGCTCAGNNNNTAGAGCAAAGGATTGAAAATC3′ and a 48-mer containing the complement of the *E.coli* $tRNA^{Phe}$ (U60C) sequence from nucleotides 29 to 76: 5′TGGTGCCCGGACTCGNGANNNNNNCAAGGACACAAGGATTTTCAATCC3′. The two oligomers were annealed and extended to form double-stranded DNA with MMLV reverse transcriptase (United States Biochemical). Both templates were transcribed and tRNA purified as described by Sampson and Uhlenbeck (1988). The concentration of RNA was determined in water assuming 1400 pmol $tRNA/A_{260}$ unit.

Selection procedure

Each round of selection was initiated by heating the RNA in water for ~3 min at 90°C, then cooling the samples to room temperature. The RNA in the first six rounds of selection was then incubated for 10 min at room temperature under conditions of RNA excess in 15 mM $MgCl_2$, 25 mM KCl, 2 mM DTT and 40 mM sodium acetate (pH 5.4) with an appropriate concentration of *E.coli* FRS purified as described in Peterson and Uhlenbeck (1992). The binding reaction (100–500 μl) was then filtered through a nitrocellulose filter (Nitro ME, Micron Separations, Inc.) that had been soaked in the reaction buffer. The filter was then washed to remove unbound RNA with 5 ml of reaction buffer. The bound RNA was eluted by soaking the filter in 400 μl phenol and 200 μl 7 M urea for ~1 h at room temperature. The aqueous phase was chloroform extracted and the eluted RNA was ethanol precipitated in 0.3 M sodium acetate with 50 pmol reverse transcriptase (RT) primer of sequence 5′TGGTGCCCGGACTCGG3′. After precipitation, the RNA was dried, redissolved in 38.5 μl 0.25 mM EDTA and 2.5 mM Tris (pH 7.5), and annealed to 0.25 μM of additional primer by heating to 95°C for ~3 min and cooling on ice. The complementary DNA was extended from the primer with 0.2 mM dNTPs, 50 mM Tris (pH 8.3), 60 mM NaCl, 6 mM magnesium acetate, 10 mM DTT and 100 U MMLV reverse transcriptase in a final volume of 50 μl. Then 5–25 μl of the cDNA were amplified in a 40 μl PCR reaction as described by Pan and Uhlenbeck (1992) using the RT primer and a second primer of sequence 5′TAATACGACTCACTATAGCCCGGATAGCTCAG3′ containing the T7 polymerase promoter and the first 15 nucleotides of the tRNA sequence. The DNA from the PCR reaction was then either precipitated or used directly in the transcription reaction to generate RNA for the next round of selection. The optimal RNA and protein concentrations used in each round of FRS-binding selection were estimated according to Irvine *et al.* (1991) by measuring the approximate K_D of the libraries after 0, 2 and 4 rounds of selection (see Cloning and characterization of variants).

The seventh round of selection was initiated by aminoacylating 1 μM RNA after six rounds of selection in 15 mM $MgCl_2$, 25 mM KCl, 4 mM DTT, 2 mM ATP, 10 μM [^{3}H]phenylalanine, 30 mM Hepes (pH 7.5) and 1 U/ml FRS for 5 min at 37°C in a 25 μl reaction. The aminoacylated mixture was then applied to 15 μl of GTP-activated immobilized *T.thermophilus* EF-Tu (Derwenskus *et al.*, 1984) and incubated for 30 min at room temperature. The unbound tRNA was washed twice from the column material with 200 μl 50 mM Hepes (pH 7.1), 20 mM NH_4Cl, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and 95 μM GTP. The bound (aminoacylated) RNA was eluted in 200 μl 50 mM EDTA and 1 mM DTT, and precipitated with 0.3 M sodium acetate with 50 pmol of the RT primer.

Cloning and characterization of variants

After seven rounds of selection, the cDNA from each library was amplified by PCR and purified on a native 15% 19:1 polyacrylamide gel. The double-

stranded DNA was then cloned into the *Sma*I site of pUC18. The colonies containing inserts were identified by colony hybridization with γ^{32}P-labeled T7 primer using a method adapted from Sambrook *et al.* (1989). The plasmids were sequenced with Sequenase 2.0 (United States Biochemical) using a protocol adapted from that provided by the manufacturer. In most cases, the sequence was determined by sequencing from primers directed towards both ends of the insert.

Binding activity of the individual variants was estimated using a protein excess binding assay. Template DNA for each variant was generated by PCR of the plasmid DNA with the same primers used during the selection. High specific activity (20 000 Ci/mol) RNA was made by transcription in the presence of [α^{32}P]CTP with 250 μM A, G and UTP, and 50 μM CTP in the standard transcription buffer. FRS at concentrations from 1 to 200 nM was incubated with RNA at a concentration at least 10-fold below the lowest protein concentration in 15 mM $MgCl_2$, 25 mM KCl, 2 mM DTT and 40 mM sodium acetate (pH 5.4) for 10 min at room temperature, and filtered through nitrocellulose presoaked in the reaction buffer. The unbound RNA was washed off the filter with 200 μl of the reaction buffer. The filters were dried and counted in toluene/PPO to determine the amount of RNA bound to the filter at the various protein concentrations. The data were fit to a retention efficiency and a K_D value assuming a simple binding equilibrium. Separate determinations gave K_D values within 2-fold.

k_{cat}/K_M for each variant was determined at 50 nM RNA, which is likely to be subsaturating, based on K_M = 330 nM for the parent tRNA (Peterson and Uhlenbeck, 1992). The template DNAs were prepared by PCR of miniprep DNA with the T7 primer and either 1201 or 1211 (New England Biolabs), depending on the orientation of the insert. The amplified DNA was precipitated and cut with *Bst*NI (New England Biolabs) to generate a discrete 3′ end. Unlabeled run-off transcripts were prepared and purified as described previously. The aminoacylation rate was determined in duplicate as described in Peterson and Uhlenbeck (1992). If no aminoacylation activity was observed with up to 3 U/ml FRS, the relative k_{cat}/K_M was estimated to be ≤ 0.01.

Acknowledgements

We thank the members of the Uhlenbeck laboratory, in particular Dave Long and Tao Pan, for helpful discussions and critical reading of the manuscript. This work was supported by NIH GM37552. We also thank the W.M.Keck Foundation for their generous support of RNA science on the Boulder campus.

References

Bartel,D.P., Zapp,M.L., Green,M.R. and Szostak,J.W. (1991) *Cell*, **67**, 529–536.

Bartmann,P., Hanke,T., Hammer-Raber,B. and Holler,E. (1974) *Biochemistry*, **13**, 4171–4175.

Bartmann,P., Hanke,T. and Holler,E. (1975) *J. Biol. Chem.*, **250**, 7668–7674.

Behlen,L.S., Sampson,J.R., DiRenzo,A.B. and Uhlenbeck,O.C. (1990) *Biochemistry*, **29**, 2515–2523.

Derwenskus,K.-H., Fischer,W. and Sprinzl,M. (1984) *Anal. Biochem.*, **136**, 161–167.

Dichtl,B., Pan,T., DiRenzo,A.B. and Uhlenbeck,O.C. (1993) *Nucleic Acids Res.*, **21**, 531–535.

Farrelly,J.G., Longworth,J.W. and Stulberg,M.P. (1971) *J. Biol. Chem.*, **246**, 1266–1270.

Goodman,R. and Schwartz,I. (1988) *Nucleic Acids Res.*, **16**, 7477–7486.

Hayase,Y., Jahn,M., Rogers,M.J., Sylvers,L.A., Koizumi,M., Inoue,H., Ohtsuka,E. and Söll,D. (1992) *EMBO J.*, **11**, 4159–4165.

Hélène,C., Brun,F. and Yaniv,M. (1969) *Biochem. Biophys. Res. Commun.*, **37**, 393–398.

Hou,Y.-M. and Schimmel,P. (1992) *Biochemistry*, **31**, 4157–4160.

Irvine,D., Tuerk,C. and Gold,L. (1991) *J. Mol. Biol.*, **222**, 739–761.

Kosakowski,H. and Böck,A. (1971) *Eur. J. Biochem.*, **24**, 190–200.

Lagerkvist,V., Rymo,L. and Waldenström,J. (1966) *J. Biol. Chem.*, **241**, 5391–5400.

Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8789.

Okamoto,T. and Kawade,Y. (1967) *Biochim. Biophys. Acta*, **145**, 613–620.

Pan,T. and Uhlenbeck,O.C. (1992) *Biochemistry*, **31**, 3887–3895.

Park,S.J., Hou,Y.-M. and Schimmel,P. (1989) *Biochemistry*, **28**, 2740–2746.

Peterson,E.T. and Uhlenbeck,O.C. (1992) *Biochemistry*, **31**, 10380–10389.

Puglisi,J.D., Pütz,J., Florentz,C. and Giegé,R. (1993) *Nucleic Acids Res.*, **21**, 41–49.

Pütz,J., Puglisi,J.D., Florentz,C. and Giegé,R. (1991) *Science*, **252**, 1696–1699.

Quigley,G.J. and Rich,A. (1976) *Science*, **194**, 796–806.

Romby,P., Carbon,P., Westhof,E., Ehresmann,C., Ebel,J.-P., Ehresmann,B. and Giegé,R. (1987) *J. Biomol. Struct. Dyn.*, **5**, 669–687.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) In Nolan,C. (ed), *Molecular Cloning—A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1.9–1.104.

Sampson,J.R. and Uhlenbeck,O.C. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1033–1037.

Sampson,J.R., DiRenzo,A.B., Behlen,L.S. and Uhlenbeck,O.C. (1990) *Biochemistry*, **29**, 2523–2532.

Sampson,J.R., Behlen,L.S., DiRenzo,A.B. and Uhlenbeck,O.C. (1992) *Biochemistry*, **31**, 4161–4167.

Schneider,D., Tuerk,C. and Gold,L. (1992) *J. Mol. Biol.*, **228**, 862–869.

Schneider,D., Gold,L. and Platt,T. (1993) *FASEB J.*, **7**, 201–207.

Seifert,W., Nass,G. and Zillig,W. (1968) *J. Mol. Biol.*, **33**, 507–511.

Tsai,D.E., Harper,D.S. and Keene,J.D. (1991) *Nucleic Acids Res.*, **19**, 4931–4936.

Tsai,D.E., Kenan,D.J. and Keene,J.D. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 8864–8868.

Tuerk,C. and Gold,L. (1990) *Science*, **249**, 505–510.

Tuerk,C., MacDougal,S. and Gold,L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6988–6992.

Yaniv,M. and Gros,F. (1969) *J. Mol. Biol.*, **44**, 17–30.

Yarus,M. and Berg,P. (1967) *J. Mol. Biol.*, **28**, 479–490.

Yarus,M. and Berg,P. (1969) *J. Mol. Biol.*, **42**, 171–189.

Yarus,M. and Berg,P. (1970) *Anal. Biochem.*, **35**, 450–465.

Received on March 18, 1993; accepted on April 7, 1993