Species-Specific Differences in the Operational RNA Code for Aminoacylation of $tRNA^{Pro}$ †

Catherine Stehlin,‡ Brian Burke,‡ Fan Yang,‡ Hongjian Liu,‡ Kiyotaka Shiba,§ and Karin Musier-Forsyth*,‡

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, and Department of Cell Biology, Cancer Institute, Japanese Foundation for Cancer Research, Kami-Ikebukuro, Toshima-ku, Tokyo 170-8455, Japan

Received February 16, 1998; Revised Manuscript Received April 6, 1998

ABSTRACT: An operational RNA code relates amino acids to specific structural features located in tRNA acceptor stems. In contrast to the universal nature of the genetic code, the operational RNA code can vary in evolution due to coadaptations of the contacts between aminoacyl-tRNA synthetases and the acceptor stems of their cognate tRNA substrates. Here we demonstrate that, for class II prolyl-tRNA synthetase (ProRS), functional coadaptations have occurred in going from the bacterial to the human enzyme. Analysis of 20 ProRS sequences that cover all three taxonomic domains (bacteria, eucarya, and archaea) revealed that the sequences are divided into two evolutionarily distant groups. Aminoacylation assays showed that, while anticodon recognition has been maintained through evolution, significant changes in acceptor stem recognition have occurred. Whereas all $tRNA^{Pro}$ sequences from bacteria strictly conserve A73 and C1·G72, all available cytoplasmic eukaryotic $tRNA^{Pro}$ sequences have a C73 and a G1·C72 base pair. In contrast to the *Escherichia coli* synthetase, the human enzyme does not use these elements as major recognition determinants, since mutations at these positions have only small effects on cognate synthetase charging. Additionally, *E. coli* $tRNA^{Pro}$ is a poor substrate for human ProRS, and the presence of the human anticodon−D stem biloop domain was necessary and sufficient to confer efficient aminoacylation by human ProRS on a chimeric $tRNA^{Pro}$ containing the *E. coli* acceptor−TψC stem−loop domain. Our data suggest that the two ProRS groups may reflect coadaptations needed to accommodate changes in the operational RNA code for proline.

The family of enzymes known as aminoacyl-tRNA synthetases catalyzes the specific attachment of amino acids to the 3′ end of cognate tRNA substrates. This reaction is essential for the accurate translation of the genetic code, which proceeds via codon−anticodon interactions at the opposite end of the L-shaped tRNA structure. The synthetase enzymes have been divided into two classes of 10 each on the basis of shared structural motifs in their catalytic domains (*1*, *2*). In recent years, the identification of nucleotide determinants, specific functional groups, and subtle tRNA structural features in each of the 20 aminoacylation systems has clarified how specific tRNA selection by the synthetases is achieved (*3*, *4*). Moreover, the ability to specifically aminoacylate acceptor stem-derived minihelices in at least 11 of the synthetase systems (*5*−*7*) provides strong evidence for an early "operational RNA code" based on acceptor stems that may have preceded the present day genetic code, which uses both domains of the L-shaped tRNA molecule (*8*). Experimental evidence also supports an early proposal that the nucleotide that follows the conserved CCA-3′ end (N73) serves as a "discriminator" base (*9*). In vitro and in vivo studies examining the effects of N73 mutations on aminoacylation provide support for the notion that N73 plays a key role in the operational RNA code for many amino acids (*10*).

Despite its important role in synthetase recognition, as elucidated primarily in *Escherichia coli* and *Saccharomyces cerevisiae* (*10*), the identity of N73 is not always conserved across taxa. For instance, U73 is conserved in all $tRNA^{Gly}$ sequences from bacteria, whereas the sequences of $tRNA^{Gly}$ from archaea and eucarya conserve an A at this position (*11*). In the glycyl system, it was shown that the *E. coli* and human enzymes do not cross-acylate their respective tRNAs (*12*). Moreover, it was determined that the species specificity can be switched by a single nucleotide exchange at position 73 (*13*). These facts imply that, in the glycine system, the operational RNA code in archaea and eucarya is distinct from that in bacteria, and this difference results in the barrier to cross-species tRNA recognition.

A similar conservation of discriminator base identities is observed in the sequences of lysine, proline, and histidine tRNAs (*11*). In the case of lysine, G73 is found in lysine tRNAs from archaea and eucarya and A73 is present in all known bacterial $tRNA^{Lys}$s. It had been shown previously that the A73 of $tRNA^{Lys}$ is an important recognition element for *E. coli* lysyl-tRNA synthetase (LysRS)[1] (*14*, *15*). We recently reported that an N-terminal truncated form of human LysRS can complement an *E. coli* null strain and that the

† This work was funded by Grant GM49928 from the National Institutes of Health (to K.M.-F.) and by a grant from the Ministry of Education, Science, and Culture, Japan (to K.S.). K.M.-F. also acknowledges the donors of the Petroleum Research Fund, administered by the American Chemical Society for partial support of this research.

* To whom correspondence should be addressed: Department of Chemistry, University of Minnesota, 207 Pleasant St. S.E., Minneapolis, MN 55455. Telephone: (612) 624-0286. Fax: (612) 626-7541. E-mail: musier@chemsun.chem.umn.edu.

‡ University of Minnesota.

§ Japanese Foundation for Cancer Research.

S0006-2960(98)00364-X CCC: $15.00
Published on Web 05/13/1998

purified human enzyme can efficiently cross-aminoacylate *E. coli* $tRNA^{Lys}$ and N73 variants of human $tRNA^{Lys}$ transcripts in vitro (*16*). Therefore, G73 is not a strong recognition element for human LysRS, and as in the glycine system, there are differences in the operational RNA code between *E. coli* and human systems.

In this report, we examine species-specific differences in $tRNA^{Pro}$ recognition. In the case of *E. coli* prolyl-tRNA synthetase (ProRS), it was previously established that the discriminator base, A73, and G72 are important acceptor stem recognition elements (*17*, *18*). These nucleotides constitute the operational RNA code for proline in the *E. coli* system. Analysis of available gene sequences for proline tRNAs shows the strict conservation of A73 and G72 in bacteria and of C73 and C72 in eucarya (*11*). Thus, the operational RNA code for proline appears to differ between bacteria and eucarya. In this work, we have chosen a representative synthetase from eucarya, human ProRS, and have begun to probe potential acceptor stem and anticodon domain recognition elements using site-directed mutagenesis. Aminoacylation assays using chimeric tRNA constructs were also carried out to identify the barrier to cross-species aminoacylation by human ProRS. Finally, to further our understanding of aminoacylation systems throughout evolution, we analyzed 20 sequences of ProRSs covering all three taxonomic domains. The results of this study contribute to our understanding of the relationship between the evolution of tRNA synthetases and the corresponding changes in tRNA acceptor stem recognition.

MATERIALS AND METHODS

Enzyme Purification and Activity Assays. Human ProRS purification was accomplished as described previously using plasmid pKS509 (*19*). This produces the C-terminal ProRS portion (codons 926−1440) of the human gluprolyl-tRNA synthetase fusion protein containing a histidine tag at the N terminus. Aminoacylation assays were carried out as described previously (*19*). The kinetic constants were derived from Lineweaver−Burk plots.

RNA Preparation. Preparation of all tRNAs was accomplished by in vitro transcription as described (*19*). Mutagenesis of tRNAs was accomplished either using the Kunkel method (*20*) or via overlap extension PCR (*21*). Following all mutagenesis procedures, the existence of only the desired mutations was verified by sequencing the entire $tRNA^{Pro}$ coding region. Bovine tRNA was purchased from Sigma.

Sequence Analysis. Multiple sequence alignments were performed using the PILEUP program provided by the Genetics Computer Group (Madison, WI). Sequences used for alignments were obtained from published ProRS sequences [*E. coli* (*1*), *Zymomonas mobilis* (*22*), *Drosophila melanogaster* (*23*), and *Homo sapiens* (*24*)], from genome databases [*Haemophilus influenzae* (*25*), *Mycoplasma genitalium* (*26*), *Mycoplasma pneumoniae* (*27*), *Synechocystis* sp. PCC6803 (*28*), *Methanococcus jannaschii* (*29*), *S. cerevisiae* (cytoplasmic and mitochondrial) (a web site of the *Saccharomyces* Genome Database), *Neisseria gonorrhoeae* and *Streptococcus pyogenes* (a web site of the University of Oklahoma's Advanced Center for Genome Technology, April 26, 1997, data release), *Methanobacterium thermoautotrophicum* (*30*), *Helicobacter pylori* (*31*), and *Archaeoglobus fulgidus* (*32*)], and from similarity searches in the Genbank database [*Chlamydia trachomatis* (L25105), *Mycobacterium tuberculosis* (Z95207), *Candida albican* (U86341), and *Caenorhabditis elegans* (U00037)].

To construct the evolutionary tree, two well-conserved regions (corresponding to the sequences of *E. coli* ProRS from G38 to D219 and from I404 to G535) in the alignment were used for the calculations. After all gaps and insertions were deleted manually from these regions, the 298 remaining residues were used for construction of a tree. The programs DISTANCE and KITSCH provided in the PHYLIP program package were used (*33*). Bootstrap analyses were performed with 100 replicates made by SEQBOOT (*33*).

RESULTS

As in the *E. coli* ProRS system (*18*), modified bases do not appear to contribute significantly to aminoacylation by human ProRS. Bovine and human tRNA sequences are generally very closely related or identical (*11*). An unmodified human $tRNA^{Pro}$ transcript was aminoacylated even more efficiently than bovine $tRNA^{Pro}$ present in a mixture of bovine tRNAs (data not shown). Therefore, using in vitro aminoacylation of human $tRNA^{Pro}$ mutant transcripts, we have elucidated a subset of the recognition elements used by the cognate human synthetase.

Acceptor Stem and Anticodon Mutations. Figure 1 shows the 10 single and 2 double mutations that were made in human $tRNA^{Pro}$ in this study. Aminoacylation results (decrease in k_{cat}/K_M relative to that of the wild-type transcript) are shown in parentheses and are also reported in Table 1. Single acceptor stem changes were made at positions 72 and 73, which were previously shown to be important for *E. coli* $tRNA^{Pro}$ aminoacylation by its cognate *E. coli* synthetase (*17*, *18*) (Figure 1, upper left). Interestingly, individual changes of C73 and C72 of human $tRNA^{Pro}$ to the corresponding nucleotides A73 and G72 present in the bacterial tRNA resulted in only minor changes in aminoacylation efficiency (−1.5- and +5.5-fold, respectively). As in the case of C72G, the double mutation C73A/C72G has a slightly positive effect on aminoacylation kinetics (1.5-fold). Substitutions C73 → G or C72 → A also had relatively small effects (−2.0- and +2.6-fold, respectively). In contrast, in the *E. coli* system, large decreases ranging from 29- to 185-fold were observed upon mutation of these two acceptor stem positions (*18*). The third base pair in the acceptor stem has been found to be critical for tRNA recognition by a number of synthetases. Although changes at this position in *E. coli* $tRNA^{Pro}$ had little effect on recognition by the bacterial synthetase (*18*), we wanted to determine whether changing the C3·G70 of human $tRNA^{Pro}$ to the G3·C70 found in *E. coli* $tRNA^{Pro}$ would affect human ProRS recognition. This single base pair transversion actually resulted in a slight (3.0-fold) increase in aminoacylation efficiency (Figure 1 and Table 1).

Previous work in the *E. coli* system showed that mutations at the anticodon positions G35 and G36 resulted in losses

[1] Abbreviations: AlaRS, alanyl-tRNA synthetase; AspRS, aspartyl-tRNA synthetase; GlyRS, glycyl-tRNA synthetase; IleRS, isoleucyl-tRNA synthetase; LysRS, lysyl-tRNA synthetase; ProRS, prolyl-tRNA synthetase; SerRS, seryl-tRNA synthetase.

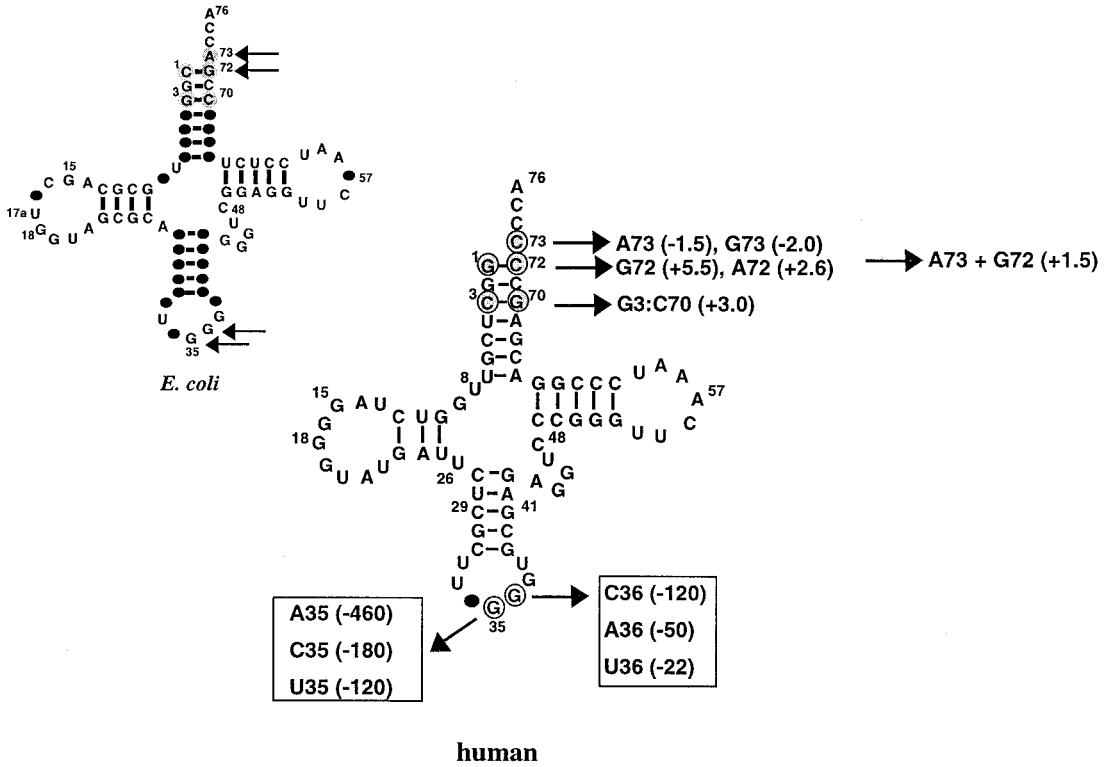

FIGURE 1: Sequence of human tRNA^{Pro} (unmodified) and variant transcripts prepared and tested in this study. Nucleotides that were tested are circled, and changes that were made are indicated by the arrows. The number in parentheses is the change (*x*-fold) in k_{cat}/K_M relative to that of the wild-type human transcript. The sequence and secondary structure of *E. coli* tRNA^{Pro} are shown for comparison (upper left). Nucleotides that are variable within each isoacceptor group are indicated by a solid dot. Conserved nucleotides in the acceptor stem domain that vary between the two species are shaded. In the case of the *E. coli* tRNA, arrows point to nucleotides that are important recognition elements for the cognate synthetase, as determined by in vitro aminoacylation experiments using mutant tRNA transcripts and purified *E. coli* ProRS (*18*).

Table 1: Effect of Single and Multiple Nucleotide Changes on Aminoacylation of Human tRNA^{Pro} Transcripts by Human ProRS[a]

mutant type	variant	k_{cat}/K_M (relative)	change in efficiency (*x*-fold)
human tRNA^{Pro}	wild-type	1	1
acceptor stem	C73A	0.67	−1.5
	C73G	0.50	−2.0
	C72A	2.6	2.6
	C72G	5.5	5.5
	C73A/C72G	1.5	1.5
	C3G/G70C	3.0	3.0
anticodon	G35A	0.0022	−460
	G35C	0.0056	−180
	G35U	0.0083	−120
	G36A	0.020	−50
	G36C	0.0083	−120
	G36U	0.045	−22

[a] Because of the relatively high K_M values (compared to tRNA concentrations) routinely used in the experiments, it was not practical to determine individual kinetic parameters for all of the mutants. Thus, only relative k_{cat}/K_M values, derived from Lineweaver−Burk plots, are reported. Concentrations of tRNA ranging from 0.5 to 40 μM were used, and the ProRS concentration was 25 or 50 nM. The results are averages of two to six determinations with average standard deviations of ±59%. The individual kinetic parameters measured for the wild-type tRNA were as follows: $k_{cat} = 0.032 \pm 0.007$ s^{-1} and $K_M = 5.0 \pm 3$ μM.

of aminoacylation efficiency ranging from 9- to 164-fold (*18*) (Figure 1, upper left). Therefore, all three mutations were tested at these two anticodon positions in human tRNA^{Pro}. At position G35, large decreases in aminoacylation efficiency ranging from 120- to 460-fold were measured (Figure 1 and Table 1). In fact, of all the positions we tested in this study, single mutations at this anticodon site had the largest impact on human ProRS recognition. Changes at G36 also resulted in significant losses in aminoacylation efficiency (−22- to −120-fold), although decreases were not quite as large as those observed at position G35 (Figure 1 and Table 1).

Cross-Species Aminoacylation by Human ProRS. Human ProRS weakly aminoacylates an *E. coli* tRNA^{Pro} transcript, with a 690-fold decrease in k_{cat}/K_M relative to that of the unmodified human transcript (Figure 2a). The relatively low efficiency of cross-acylation was surprising in light of the fact that the critical anticodon nucleotides G35 and G36 are present in *E. coli* tRNA^{Pro}, and all of the acceptor stem elements we tested were shown not to be important for the human synthetase. Although incorporation of individual and some multiple *E. coli* acceptor stem elements into human tRNA^{Pro} did not block aminoacylation by human ProRS (Table 1), we nevertheless wondered whether the relatively large number of species-specific differences in this domain contributed in a cooperative manner to prevent cross-acylation by the mammalian synthetase. To test this idea, we simultaneously incorporated five human acceptor stem nucleotides (C73, G1·C72, and C3·G70) into *E. coli* tRNA^{Pro} (Figure 2b). The presence of these human acceptor stem elements in the bacterial tRNA, however, did not result in improved cross-species aminoacylation by human ProRS (data not shown).

There are significant differences in the D arm domain between tRNA^{Pro}s from the two species (Figure 2, compare panels a and h). While the *E. coli* tRNA has an "extra"

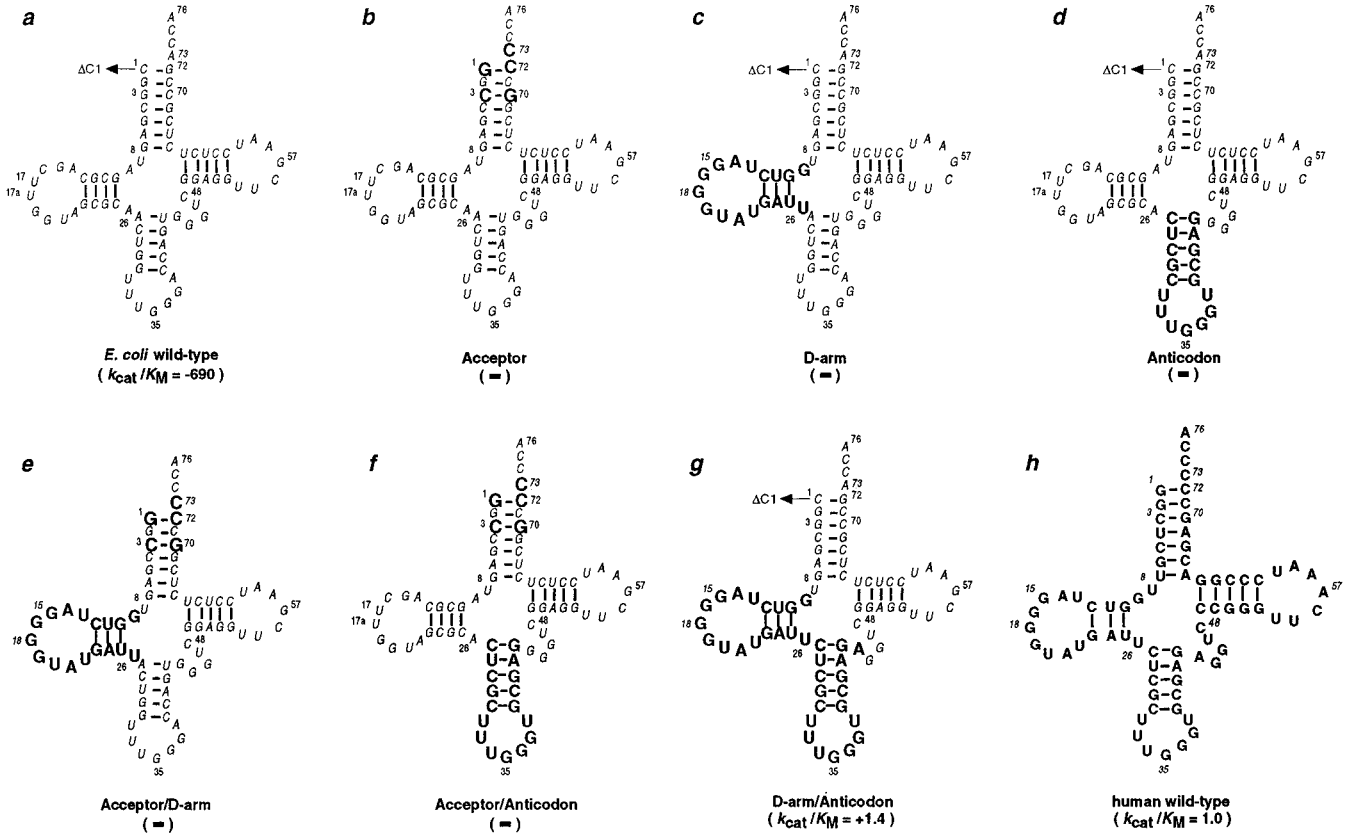

FIGURE 2: Wild-type *E. coli* (a) and human (h) tRNA^{Pro} and chimeric (b−g) *E. coli* and human tRNA^{Pro} variants tested as substrates for human ProRS. *E. coli* sequence elements are in a smaller font and italicized, while human sequence elements are in a larger and bold font. The arrow pointing to DC1 (*a*, *c*, *d*, and *g*) indicates that in these transcripts the first C was deleted to facilitate in vitro transcription, as described previously (*41*). k_{cat}/K_M is given relative to that of a wild-type human tRNA^{Pro} transcript, which was set at 1.0. A − indicates no improvement in k_{cat}/K_M relative to human ProRS aminoacylation of *E. coli* tRNA^{Pro}. In these assays, the tRNA concentration ranged from 0.5 to 36 μM, and the enzyme concentration was 50 or 100 nM.

nucleotide at position 17a and four G•C base pairs in the D stem, the human tRNA lacks nucleotide 17 and contains much weaker base pairing interactions in the D stem. As a result of these differences, the human tRNA is expected to have a shorter anticodon-D stem than the *E. coli* tRNA. To determine whether these structural differences resulted in the barrier to cross-species aminoacylation, we transplanted the human D domain from positions 9−26 into the framework of the *E. coli* tRNA. This domain swap, either alone (Figure 2c) or in combination with the five human acceptor stem elements (Figure 2e), did not result in improved cross-acylation. Thus, simply changing the sequence of the D arm or the length of the anticodon D stem does not confer charging by human ProRS. We next prepared *E. coli* tRNA^{Pro} variants containing the anticodon stem−loop domain (positions 27−43) of human tRNA^{Pro} alone (Figure 2d) or in the presence of the five human acceptor stem elements (Figure 2f). Once again, these chimeras were not amino-acylated by the human enzyme. Finally, we prepared a chimeric tRNA that contained both the anticodon and D domains of human tRNA^{Pro} (positions 9−44), while retaining the acceptor−TψC stem−loop of *E. coli* tRNA^{Pro} (Figure 2g). Despite the fact that it contained 13 out of 18 nucleotide changes in the acceptor stem relative to human tRNA^{Pro}, this variant was an excellent substrate for human ProRS (k_{cat}/K_M = 1.4 relative to that of wild-type human tRNA^{Pro}).

Sequence Analyses. Previous genetic and biochemical studies have identified the genes for ProRSs from *E. coli* (*1*), *D. melanogaster* (*23*), and *H. sapiens* (*19*, *24*). Although the sequence identity between fly and human ProRSs is high (58%), the similarities between *E. coli* and fly or *E. coli* and human are very low (20 and 19%, respectively) compared to those of some other aminoacyl-tRNA synthetases [for instance, the sequence similarity between human and *E. coli* alanyl-tRNA synthetases (AlaRSs) is 41% (*34*)]. To gain further insight into the evolution of ProRSs, we compiled orthologous genes for *E. coli* and human ProRSs from other organisms by searching genome databases and the Genbank DNA database. Seventeen sequences showed high sequence similarities with either *E. coli* or human ProRS (data not shown). In total, 20 sequences were used for the analyses in this paper, including 11 from bacteria, 3 from archaea, and 6 from eucarya.

The ProRS sequences can be aligned so that 11 positions are completely conserved in their amino acid residue identities (Figure 3, shaded). Conserved residues are located not only in the class-defining motifs (residues R140, E142, R151, E154, G445, and R450 in the *E. coli* enzyme) but also in other regions (residues G38, E111, G464, G517, and G527 in the *E. coli* enzyme), indicating that ProRSs from *E. coli* and human, while displaying a low overall sequence similarity, have evolved from a common ancestral gene.

The existence of two distinct structural forms of ProRS was recently noted (*35*, *36*) and is supported by the sequence analysis carried out in this work. The alignment revealed that the 20 sequences are sorted into two groups on the basis

FIGURE 3: Alignment of 20 prolyl-tRNA synthetases. Amino acids that are identical for all of the sequences are shaded in gray. The locations of motifs 1−3 are indicated, along with predicted α-helices (cylinders) and β-sheets (arrows). Gaps in the sequences are indicated by dots, and numbers in black boxes represent residues that were omitted from the alignment. The numbers across the top and bottom refer to codon positions for the *E. coli* and human enzymes, respectively. Abbreviations are as follows: Ec, *E. coli*; Hi, *Ha. influenzae*; Ng, *N. gonorrhoeae*; Zm, *Z. mobilis*; Sy, *Synechocystis* sp. PCC6803; Sp, *St. pyogenes*; Mt, *My. tuberculosis*; Hp, *Hel. pylori*; Ct, *Ch. trachomatis*; Scm, *S. cerevisiae* (mitochondrial); Cam, *C. albican* (mitochondrial); Mg, *M. genitalium*; Mp, *M. pneumoniae*; Mj, *Me. jannaschii*; Mtt, *Met. thermoautotrophicum*; Af, *A. fulgidus*; Ce, *Ca. elegans*; Dm, *D. melanogaster*; Sc, *S. cerevisiae* (cytoplasmic); Hs, *H. sapiens.*

of insertion−extension patterns. One group has an insertion of 55−209 amino acids between motif 2 and motif 3, and the other group has an extension of 77−108 amino acids located at the C terminus that ends in a strictly conserved Y residue (Figure 3). We also noted that the identities of amino acid residues in some positions are conserved in a group-specific manner. For instance, in the predicted loop region of the β-strand−loop−β-strand of motif 2, R144 and G148 (in the *E. coli* enzyme) are completely conserved in the first group, whereas K and P are conserved at corresponding positions in the second group (Figure 3). Similarly, motif 3 has diverged between the two groups. While the first group has a GIGXXR consensus sequence, the second group has a GXXXR consensus sequence (Figure 3). The R residue in these motifs is known to be highly conserved among class 2 enzymes and interacts with the γ-phosphate of ATP in the yeast aspartyl-tRNA synthetase (AspRS) system (*37*).

Partitioning of the 20 ProRS sequences into two groups on the basis of an idiosyncratic insertion−extension pattern was further supported by phylogenetic analysis. Reasonably conserved regions in the alignment were used to compute a distance matrix of these sequences, and the phylogenetic tree was constructed from the distance relationships. As shown in Figure 4, two main branches are well-separated and this result was supported by high bootstrap confidence values.

The bipartite sorting of the sequences is consistent with taxonomic classification of the organisms from which the sequences were isolated. The "prokaryotic" group that includes *E. coli* ProRS mostly consists of sequences from bacteria. Two eukaryotic sequences in this prokaryotic group (*S. cerevisiae* and *C. albican*) are likely to represent mitochondrial ProRSs. This is supported by the fact that the genome of *S. cerevisiae* contains two ProRS sequences and one is in this first group while the other shows high sequence similarity to human cytoplasmic ProRS and is therefore in the second group. The second eukaryotic group contains sequences from primarily eucarya and archaea. Interestingly, this group also contains two sequences from the bacteria domain, those of *M. genitalium* and *M. pneumoniae* [and of the recently released *Borrelia burgdorferi* (*38*), which is not included in our analysis]. Because these organisms only have one gene each for ProRS in their genomes, these are ex-

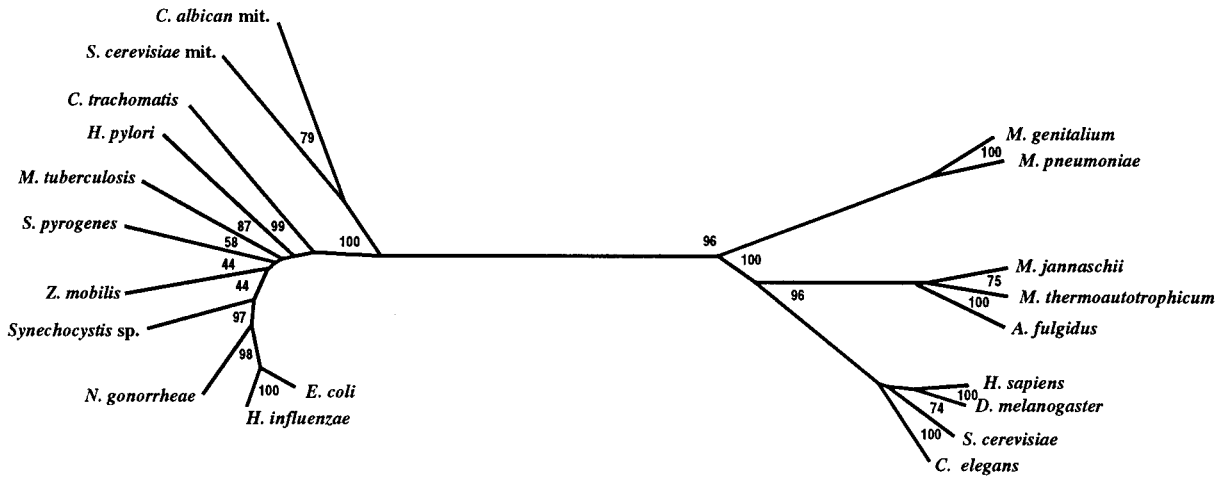

FIGURE 4: Phylogenetic tree of 20 prolyl-tRNA synthetase sequences. The distance matrix used to construct the tree was computed using the Dayoff PAM matrix (*33*). Branch lengths are proportional to relative sequence divergence. Numbers show bootstrap values (in percentage) with 100 replicates.

amples of "crossovers" of synthetase genes from one taxonomic domain to another (*35*). The ProRS sequence from *Thermus thermophilus* is also sorted into the eukaryotic group (*36*, *39*). Similar crossovers are found in glycyl-tRNA synthetases (GlyRSs) from *M. genitalium*, *M. pneumonia*, *My. tuberculosis*, *Mycobacterium leprae*, *T. thermophilus*, *B. burgdorferi*, and *Treponema palliduma* (*35*, *40*; K. Shiba, unpublished) and isoleucyl-tRNA synthetases (IleRSs) from *My. tuberculosis*, *B. burgdorferi*, *Tr. palliduma*, *T. thermophilus*, and *Thermotoga maritim* (*35*; K. Shiba, unpublished).

DISCUSSION

Changes in tRNA^{Pro} Recognition through Evolution. In previous in vitro work, we characterized *E. coli* tRNA^{Pro} recognition by *E. coli* ProRS and found that critical recognition elements included acceptor stem positions A73 and G72 and anticodon nucleotides G35 and G36, as well as certain core elements that determine the tRNA tertiary structure (*18*, *41*−*43*). We now find that changes in the anticodon nucleotides at positions G35 and G36 have even greater effects on aminoacylation by the cognate human synthetase than they do in the *E. coli* system (*18*) (Figure 1). A 3.5 Å resolution crystal structure of *T. thermophilus* ProRS complexed with *T. thermophilus* tRNA^{Pro} shows that there are numerous base-specific interactions between the second and third anticodon bases and the synthetase's anticodon binding domain (*39*). As mentioned earlier, the *T. thermophilus* enzyme belongs to the eukaryotic group and is very similar to human ProRS. Despite the fact that the bacterial tRNA has these critical anticodon elements, it is a poor substrate for the mammalian enzyme. This was surprising given that the acceptor stem does not appear to contribute significantly to human ProRS recognition (Figure 1 and Table 1). In particular, individual or multiple mutations at five human tRNA^{Pro} acceptor stem positions that differ between the two species had only minor effects on aminoacylation. In fact, most acceptor stem changes tested resulted in small but positive changes in the overall catalytic efficiency. Notably, when the single mutation C72G is made in human tRNA^{Pro}, aminoacylation by the human synthetase is increased 5.5-fold (Table 1). Taken together, the lack of efficient *E. coli* tRNA^{Pro} aminoacylation by the human enzyme and the lack of critical acceptor stem nucleotides in the human tRNA suggest that tRNA structural elements (specific nucleotides and/or conformational features) outside the anticodon domain are likely to be important for human ProRS recognition; however, the location of these secondary recognition elements has diverged through evolution. Alternatively, *E. coli* tRNA^{Pro} may contain sequence elements outside the acceptor stem that serve to block human ProRS recognition.

A series of chimeric tRNA^{Pro} constructs was prepared to identify sequence elements or structural features outside the anticodon trinucleotide that may be required by the human enzyme. Transplantation of 5 nucleotides near the top of the acceptor stem that differ between the human and *E. coli* tRNA did not affect human ProRS recognition (Figure 2b). This was not too surprising, since our mutational analysis failed to identify important recognition elements in this domain (Table 1 and Figure 1). As a result of differences in the D domain, the anticodon-D stem is 1 base pair shorter in the human tRNA than in the *E. coli* tRNA^{Pro}, and 3 of the 9 key tertiary interactions found in most nonmitochondrial tRNAs (*44*) differ between the two species. Transplanting the individual D and anticodon domains of human tRNA^{Pro} into the *E. coli* tRNA^{Pro} framework failed to improve cross-acylation (despite the fact that the former construct now contains the same number of base pairs in the D stem as wild-type human tRNA^{Pro}) (Figure 2c−f). Transplanting both domains simultaneously, however, resulted in a chimeric tRNA substrate that was aminoacylated as efficiently as human tRNA^{Pro} (Figure 2g). When the CCA-3′ end that is conserved in all tRNAs is excluded, 13 out of 15 acceptor stem nucleotides differ between this active chimeric construct and human tRNA^{Pro} (Figure 2h). Correct positioning of the acceptor stem into the active site of human ProRS appears to depend primarily on specific nucleotides or conformational features within the human D and anticodon domains. In the cocrystal structure of *T. thermophilus* ProRS complexed to tRNA^{Pro}, the acceptor stem does not enter the active site and is disordered (*39*). As a result, only synthetase interactions with the well-ordered anticodon stem−loop are clearly observed (*39*). As in the case of *T. thermophilus* ProRS, which is 37% identical to human ProRS, the human enzyme is likely to interact extensively with the anticodon domain.

In the case of the human enzyme, our experiments suggest that synthetase interactions with the D arm are also likely. The fact that transplantation of the individual D and anticodon domains failed to improve aminoacylation by the human enzyme suggests that these interactions occur in a cooperative fashion. Thus, our analysis using the chimeric tRNA constructs shows that sequence and/or conformational differences between human and *E. coli* $tRNA^{Pro}$ in the D-anticodon domain result in the barrier to cross-species aminoacylation.

On the basis of the high-resolution crystal structures of two class II synthetases complexed with their tRNA substrates, the variable loop between the two β-strands of motif 2 makes important major groove acceptor stem contacts. For example, in the case of yeast AspRS, the motif 2 loop contacts G73 and the first base pair (U1·A72) (*45*). The much longer motif 2 loop of *T. thermophilus* seryl-tRNA synthetase (SerRS) interacts with the acceptor stem major groove down to the fourth base pair (*46*). In the case of *E. coli* ProRS, the motif 2 loop sequence is VRPRF, and this sequence is highly conserved within the prokaryotic group of enzymes (Figure 3). The corresponding loop sequence in human ProRS is different at all five positions (FKHPQ) but is also highly conserved among the eukaryotic group (Figure 3). By analogy to other class II synthetases, the motif 2 loop of *E. coli* ProRS may make base specific contacts to A73 and G72. Although our experiments provide no evidence of base-specific interactions with acceptor stem elements in the case of human ProRS, our results do not rule out the existence of important backbone contacts. The eukaryotic group may have evolved a different motif 2 loop sequence to adapt to the changes in the eukaryotic tRNA acceptor stem, where backbone interactions may be more important than base-specific recognition. Experiments are underway to test this hypothesis.

Class II Aminoacyl-tRNA Synthetase Evolution and the Relationship to Acceptor Stem Changes. The class II aminoacyl-tRNA synthetases show a wide range of sequence divergence across species. For instance, the primary sequence of AlaRS has been shown to be highly conserved among humans, insects, plants, and *E. coli* (*34*). This sequence similarity is extended to AlaRSs from the archaea domain (K. Shiba, unpublished observation). There is also a significant degree of homology among the primary structures of all known SerRSs (*47*). In contrast to the relatively conserved nature of AlaRS and SerRS proteins, no significant sequence similarities have been observed between *E. coli* and human GlyRSs, and on the basis of primary structure, it is not even clear that they share a common ancestor (*12*). Comparison of 21 GlyRS sequences showed that they could be sorted into two groups, one containing synthetases primarily from archaea and eucarya and the other containing sequences from bacteria (*35*). We show here that ProRS sequences can also be sorted into two groups. In the case of ProRSs, however, 11 residues are strictly conserved in all 20 sequences examined, indicating that in this case the two groups (bacteria and archaea−eucarya) diverged from a common ancestor.

Another synthetase system that has been studied in some detail is that of LysRS (*16*). While a high degree of sequence similarity is observed among LysRSs from eucarya, bacteria and the crenarchaeota subdomain of archaea (or eocytes) (*16*), an open reading frame that is similar in sequence to the apparently predominant group of class II LysRSs could not be identified from the genomes (*Me. jannaschii*, *Met. thermoautotrophicum*, and *A. fulgidus*) in the euryarchaeota subdomain of archaea (*29*). Recently, the existence of a completely different "archaeal-type" class I LysRS has been confirmed in these archaea genomes and in spirochaetes (*48*, *49*), and this discovery is the first example of a class switch among tRNA synthetases.

Recent studies of class II GlyRS (*13*, *50*) and LysRS (*16*) revealed a species-specific operational RNA code in these systems. For example, all $tRNA^{Gly}$s from bacteria have U73, while those from archaea and eucarya have A73. Therefore, the partitioning of enzymes is consistent with the conservation of N73, which is an important recognition element for both mammalian and bacterial GlyRSs. In the lysyl system, the A73 that functions as an important element for *E. coli* LysRS is also strictly conserved in bacteria. On the other hand, G73 which is conserved in the archaea and eucarya lysine tRNAs is not a critical element for the human enzyme. Results from these class II systems are in contrast to those obtained in the alanyl system in which a common operational RNA code, namely, recognition of a G3·U70 wobble base pair, is used among *E. coli*, human, and yeast AlaRSs (*34*, *51*, *52*). Previous studies have also shown that the mechanisms by which class II AspRSs (*53*, *54*) and phenylalanyl-tRNA synthetases (*55*, *56*) recognize their cognate tRNA substrates have diverged somewhat among different species.

The relationship between synthetase evolution and tRNA recognition is addressed in this work for the prolyl system. In accordance with the important role of N73 and N72 in *E. coli* ProRS recognition, all $tRNA^{Pro}$ sequences from bacteria strictly conserve an A and a G, respectively, at these acceptor stem positions (*11*). On the other hand, in eucarya, all available cytoplasmic $tRNA^{Pro}$ sequences have C73 and C72. However, in the human system, base-specific recognition of acceptor stem elements does not appear to occur. This is in contrast to the glycyl system in which the positional location of the operational RNA code appears to be conserved between *E. coli* and humans, although the identity of the important nucleotides is different (*13*). In the prolyl system, anticodon recognition has been maintained, whereas the operational RNA code has diverged through evolution. Whether the two groupings of ProRS are strictly correlated with two types of acceptor stem recognition remains to be determined. Proline tRNAs from archaea strictly conserve A73 and C72 and may represent an intermediate in evolution. In conclusion, previous studies together with the new observations reported here support the notion that there is a relationship between the development of the operational RNA code based on tRNA acceptor stems and the evolution of aminoacyl-tRNA synthetases, which appear to have co-adapted to accommodate changes in acceptor stem recognition.

ACKNOWLEDGMENT

We thank Dr. Stephen Cusack for providing data for *T. thermophilus* prolyl-tRNA synthetase prior to publication. We also thank Drs. Paul Schimmel and Sunghoon Kim for helpful comments on the manuscript.

REFERENCES

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) *Nature 347*, 203−206.
2. Cusack, S., Härtlein, M., and Leberman, R. (1991) *Nucleic Acids Res. 19*, 3489−3498.
3. McClain, W. H. (1993) *J. Mol. Biol. 234*, 257−280.
4. Saks, M. E., Sampson, J. R., and Abelson, J. N. (1994) *Science 263*, 191−197.
5. Hamann, C. S., and Hou, Y.-M. (1995) *Biochemistry 34*, 6527−6532.
6. Quinn, C. L., Tao, N., and Schimmel, P. (1995) *Biochemistry 34*, 12489−12495.
7. Martinis, S. A., and Schimmel, P. (1996) in *Escherichia coli and Salmonella* (Neidhardt, F., Ed.) 2nd ed., pp 887−901, ASM Press, Washington, DC.
8. Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8763−8768.
9. Crothers, D. M., Seno, T., and Söll, D. G. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3063−3067.
10. Hou, Y.-M. (1997) *Chem. Biol. 4*, 93−96.
11. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
12. Shiba, K., Schimmel, P., Motegi, H., and Noda, T. (1994) *J. Biol. Chem. 269*, 30049−30055.
13. Hipps, D., Shiba, K., Henderson, B., and Schimmel, P. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5550−5552.
14. McClain, W. H., and Foss, K. (1988) *J. Mol. Biol. 202*, 697−709.
15. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992) *Nucleic Acids Res. 20*, 2335−2339.
16. Shiba, K., Stello, T., Motegi, H., Noda, T., Musier-Forsyth, K., and Schimmel, P. (1997) *J. Biol. Chem. 36*, 22809−22816.
17. McClain, W. H., Schneider, J., and Gabriel, K. (1994) *Nucleic Acids Res. 22*, 522−529.
18. Liu, H., Peterson, R., Kessler, J., and Musier-Forsyth, K. (1995) *Nucleic Acids Res. 23*, 165−169.
19. Heacock, D., Forsyth, C., Shiba, K., and Musier-Forsyth, K. (1996) *Bioorg. Chem. 24*, 273−289.
20. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488.
21. Horton, R. M., and Pease, L. R. (1991) in *Directed Mutagenesis* (McPherson, M. J., Ed.) pp 217−247, IRL Press, New York.
22. Peekhaus, N., Tolner, B., Poolman, B., and Kramer, R. (1995) *J. Bacteriol. 177*, 5140−5147.
23. Cerini, C., Kerjan, P., Astier, M., Gratecos, D., Mirande, M., and Semeriva, M. (1991) *EMBO J. 10*, 4267−4277.
24. Fett, R., and Knippers, R. (1991) *J. Biol. Chem. 266*, 1448−1455.
25. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995) *Science 269*, 496−512.
26. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., and Kelley, J. M. (1995) *Science 270*, 397−403.
27. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C., and Herrmann, R. (1996) *Nucleic Acids Res. 24*, 4420−4449.
28. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996) *DNA Res. 3*, 109−136.
29. Bult, C. J., White, O., Olsen, G. J., Zhou, L. X., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1996) *Science 273*, 1058−1073.
30. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., and Reeve, J. N. (1997) *J. Bacteriol. 179*, 7135−7155.
31. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., and Venter, J. C. (1997) *Nature 388*, 539−547.
32. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., and Venter, J. C. (1997) *Nature 390*, 364−370.
33. Felsenstein, J. (1993) *PHYLIP, Phylogeny Inference Package*, version 3.5c, University of Washington, Seattle, WA.
34. Shiba, K., Ripmaster, T., Suzuki, N., Nichols, R., Plotz, P., Noda, T., and Schimmel, P. (1995) *Biochemistry 34*, 10340−10349.
35. Shiba, K., Motegi, H., and Schimmel, P. (1997) *Trends Biochem. Sci. 22*, 453−457.
36. Cusack, S., Yaremchuk, A., and Tukalo, M. (1998) in *The Many Faces of RNA* (Eggleston, D. S., Prescott, C. D., and Pearson, N. D., Eds.) pp 55−65, Academic Press, London.
37. Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J.-C., and Moras, D. (1994) *EMBO J. 13*, 327−337.
38. Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., and Venter, J. C. (1997) *Nature 390*, 580−586.
39. Cusack, S., Yaremchuk, A., Krikliviy, I., and Tukalo, M. (1998) *Structure 6*, 101−108.
40. Mazauric, M.-H., Reinbolt, J., Lorber, B., Ebel, C., Keith, G., Giegé, R., and Kern, D. (1996) *Eur. J. Biochem. 241*, 814−826.
41. Liu, H., and Musier-Forsyth, K. (1994) *Biochemistry 33*, 12708−12714.
42. Yap, L.-P., and Musier-Forsyth, K. (1995) *RNA 1*, 418−424.
43. Yap, L.-P., Stehlin, C., and Musier-Forsyth, K. (1995) *Chem. Biol. 2*, 661−666.
44. Giegé, R., Puglisi, J. D., and Florentz, C. (1993) *Prog. Nucleic Acid Res. Mol. Biol. 45*, 129−205.
45. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) *Science 252*, 1682−1689.
46. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *EMBO J. 15*, 2834−2842.
47. Lenhard, B., Filipic, S., Landeka, I., Skrtic, I., Söll, D., and Weygand-Durasevic, I. (1997) *J. Biol. Chem. 272*, 1136−1141.
48. Ibba, M., Bono, J. L., Rosa, P. A., and Söll, D. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14383−14388.
49. Ibba, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R., and Söll, D. (1997) *Science 278*, 1119−1122.
50. Nameki, N., Tamura, K., Asahara, H., and Hasegawa, T. (1997) *J. Mol. Biol. 268*, 640−647.
51. Hou, Y.-M., and Schimmel, P. (1989) *Biochemistry 28*, 6800−6804.

52. Ripmaster, T. L., Shiba, K., and Schimmel, P. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 4932−4936.
53. Nameki, N., Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992) *Biochem. Biophys. Res. Commun. 189*, 856−862.
54. Becker, H. D., Giegé, R., and Kern, D. (1996) *Biochemistry 35*, 7447−7458.
55. Nazarenko, I. A., Tinkle-Peterson, E., Zakharova, O. D., Lavrik, O. I., and Uhlenbeck, O. C. (1992) *Nucleic Acids Res. 20*, 475.
56. Peterson, E. T., and Uhlenbeck, O. C. (1992) *Biochemistry 31*, 10380−10389.

BI980364S