Nucleic Acids Research, 1999, Vol. 27, No. 24 **4743–4750**

Conservation of a tRNA core for aminoacylation

Ya-Ming Hou*, Hiromi Motegi[1], Richard S. A. Lipman, Christian S. Hamann and Kiyotaka Shiba[1]

Thomas Jefferson University, Department of Biochemistry and Molecular Pharmacology, 233 South 10th Street, Room 222, Philadelphia, PA 19107, USA and [1]Department of Cell Biology, Cancer Institute, Japanese Foundation for Cancer Research, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan

Received August 25, 1999; Revised and Accepted October 21, 1999

ABSTRACT

The core region of *Escherichia coli* tRNA^{Cys} is important for aminoacylation of the tRNA. This core contains an unusual G15:G48 base pair, and three adenosine nucleotides A13, A22 and A46 that are likely to form a 46:[13:22] adenosine base triple. We recently observed that the 15:48 base pair and the proposed 46:[13:22] triple are structurally and functionally coupled to contribute to aminoacylation. Inspection of a database of tRNA sequences shows that these elements are only found in one other tRNA, the *Haemophilus influenzae* tRNA^{Cys}. Because of the complexity of the core, conservation of sequence does not mean conservation of function. We here tested whether the conserved elements in *H.influenzae* tRNA^{Cys} were also important for aminoacylation of *H.influenzae* tRNA^{Cys}. We cloned and purified a recombinant *H.influenzae* cysteine-tRNA synthetase and showed that it depends on 15:48 and 13, 22 and 46 in a relationship analogous to that of *E.coli* cysteine-tRNA synthetase. The functional conservation of the tRNA core is correlated with sequence conservation between *E.coli* and *H.influenzae* cysteine-tRNA synthetases. As the genome of *H.influenzae* is one of the smallest and may approximate a small autonomous entity in the development of life, the dependence of this genome on G15:G48 and its coupling with the proposed A46:[A13:A22] triple for aminoacylation with cysteine suggests an early role of these motifs in the evolution of decoding genetic information.

INTRODUCTION

The core region of tRNA is largely known for its contribution to the formation of the tRNA L shape. This region connects the dihydrouridine (D) loop with the variable (V) loop and TΨC loop to stabilize the joining of two long helical stems of the tRNA. Because of the core, one long helical stem presents the CCA end while the other presents the anticodon end. The two ends of the L are well positioned for aminoacylation of tRNA and for participation of tRNA in the ribosomal machinery of protein synthesis. The importance of the core in tRNA structure is highlighted in the complexity of the core. Analysis of several tRNA crystal structures indicates that the top layer of the core is usually base 59, which is followed by the layer of the 15:48 base pair, the layer of the 21:[8:14] base triple, the layer of the 13:22 base pair, and then by the layers of base pairs in the D stem (1–8). While individual bases in each layer of the core are different from tRNA to tRNA, they co-vary in such a way as to preserve the general features of the core (9,10).

The significance of the core in tRNA aminoacylation has only recently been recognized. Studies of tRNA aminoacylation, whereby an amino acid is attached to the tRNA 3′-end by an aminoacyl-tRNA synthetase, have focused on nucleotides near the CCA end and the anticodon end (11). While these two ends provide specific bases for contact with many synthetases, their ability to present bases is controlled by the tRNA core. A few examples that shed light on the importance of the tRNA core in aminoacylation include tRNA^{Cys} (12,13), tRNA^{Ile} (14), tRNA^{Phe} (15,16), tRNA^{Ser} (17), and tRNA^{Arg} of *Escherichia coli* (17) and tRNA^{Phe} of *Saccharomyces cerevisiae* (18). In these examples, nucleotides in different layers of the core are implicated as determinants for the specificity of aminoacylation. The complexity of the tRNA core suggests that various layers of the core can contribute to tRNA aminoacylation in ways that are subtle but diverse.

The core of *E.coli* tRNA^{Cys} is of interest. We previously showed that two layers of the core are important for aminoacylation. One is the 15:48 layer and the other is the 13:22 layer (12,13,19,20). *Escherichia coli* tRNA^{Cys} has G15:G48 in the 15:48 layer. The structure of G15:G48 is distinct from the Pu15:Py48 that is commonly found in tRNA structures. Specifically, G15:G48 is accessible to the chemical probe dimethylsulfate (DMS) (12,13). This accessibility is a chemical signature of *E.coli* tRNA^{Cys} and it suggests that G15 of the 15:48 layer does not stack on A14 of the 8:14 layer directly below. Alteration of G15:G48 to G15:C48 eliminates the chemical signature and decreases the catalytic efficiency (k_{cat}/K_m) of aminoacylation by almost two orders of magnitude (12). The 13:22 layer of the tRNA is an unpaired A13:A22. Conversion of A13:A22 to a W·C pair also eliminates the chemical signature at G15 and decreases k_{cat}/K_m by almost two orders of magnitude (13). The significance of 15:48 and that of 13:22 in aminoacylation are coupled, both in structure and in function. Substitutions of 13:22 can alter the structure of 15:48 and inactivate aminoacylation (13).

*To whom correspondence should be addressed. Tel: +1 215 503 4480; Fax: +1 215 923 9162; Email: ya-ming.hou@mail.tju.edu

The structure in the core region of tRNA^{Cys}, free in solution, is not well understood. We have proposed that A13:A22 interacts with A46 in the V loop to form an A46:[A13:A22] base triple (19). The proposed 46:[13:22] triple can provide a basis for the structural and functional coupling between 13:22 and 15:48, as recent studies show that A46 is functionally associated with 13:22 and that it is also directly linked to G48 of 15:48 (19). However, in the crystal structure of the tRNA complexed with the elongation factor EF-Tu, A13:A22 forms a base triple with A9 (5). Although the significance of the 9:[13:22] base triple in aminoacylation is unclear at present, the discovery of this triple and its difference from the proposed 46:[13:22] triple emphasizes the difficulty in predicting the structure of the tRNA core.

The structural and functional coupling of 15:48 and 13:22 in *E.coli* tRNA^{Cys} is unusual and this raises the question of whether further examples of such coupling might be found. The identification of additional examples will highlight the relationship between two layers of the core that jointly contribute to aminoacylation. It will also provide a basis to trace the history of this relationship in the development of the tRNA^{Cys} core. To test this possibility, we searched the database of tRNA genes (10) and found that only the tRNA^{Cys} of *Haemophilus influenzae* contains G15:G48 (Fig. 1; 21). This tRNA also contains A13:A22, and the A9 and A46 elements that are implicated as important for the structure of A13:A22 in *E.coli* tRNA^{Cys} (Fig. 1). However, the conservation of sequence elements between *H.influenzae* and *E.coli* tRNAs does not mean that they are functionally conserved, especially when these elements are in the tRNA core where base pairing interactions are complex and more difficult to predict than those of secondary structures. Here, we show that sequence conservation of the core between *H.influenzae* and *E.coli* tRNA^{Cys} has the significance of functional conservation. This significance is supported by the close sequence relationship between *H.influenzae* and *E.coli* cysteine-tRNA synthetases. Because the genome of *H.influenzae* is one of the smallest and probably was present during the early stages of the development of life, the dependence of this genome on the core of tRNA^{Cys} for aminoacylation suggests an ancient history of the core in aminoacylation with cysteine.

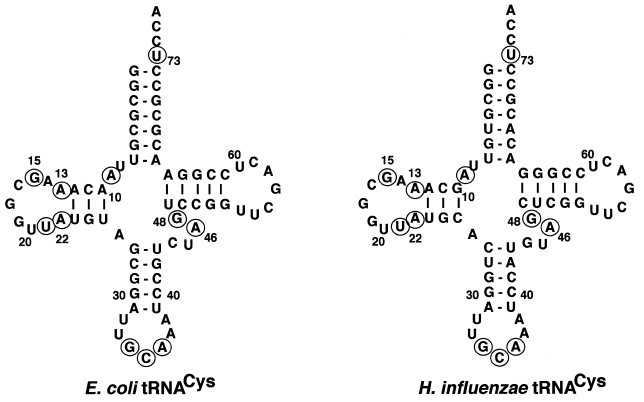

Figure 1. Sequence and cloverleaf secondary structure of (left) *E.coli* tRNA^{Cys} and (right) *H.influenzae* tRNA^{Cys}. Nucleotides that are important for aminoacylation of the *E.coli* tRNA are indicated by circles, including U73, the GCA anticodon, G15:G48, A13:A22, A9 and A46, and U21. These nucleotides are conserved in *H.influenzae* tRNA^{Cys} and are likely important for aminoacylation of the *H.influenzae* tRNA. Numbering of nucleotides is based on that of yeast tRNA^{Phe} (10).

MATERIALS AND METHODS

Construction of genes for tRNA^{Cys}

The gene for *H.influenzae* tRNA^{Cys} was identified from the genomic database (21). It was constructed by synthesizing 10 overlapping oligonucleotides, hybridizing these oligonucleotides in 66 mM Tris–HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM DTT and 1 mM ATP, and ligating them together with the pTFMa vector (a derivative of pUC18; 12) that was previously linearlized at the *Hin*dIII and *Bam*HI sites. The gene for *Bacillus subtilis* tRNA^{Cys} was constructed similarly. Each tRNA gene was built with a *Hin*dIII restriction site and the T7 promoter sequence at the 5′-end and *Bst*NI and *Bam*HI restriction sites at the 3′-end. Successful clones of a gene were identified by dideoxy sequencing analysis.

Preparation of tRNA transcripts and aminoacylation with cysteine

Both the *E.coli* and *H.influenzae* tRNA genes were constructed in plasmid pTFMa described above. Restriction of a pTFMa plasmid carrying a tRNA gene with *Bst*NI generated the template for T7 transcription. Mutants of tRNA genes were created by site-directed mutagenesis of the wild-type gene (22). Aminoacylation of T7 transcripts with cysteine was as described (12). In all routine procedures, we used purified T7 RNA polymerase for T7 transcription (23) and purified *E.coli* cysteine-tRNA synthetase for aminoacylation assays (24).

Chemical modification of tRNAs

Procedures for the DMS modification of N7 of G15 have been described (19,20). The site of modification was detected by aniline scission (25). An N7 of G15 accessible to DMS is the chemical signature of *E.coli* tRNA^{Cys}, and it indicates protrusion of the purine ring of G15 from that of A14 (12).

Cloning and expression of the gene for *H.influenzae* cysteine-tRNA synthetase

We designed two primers for amplification of the gene for *H.influenzae* cysteine-tRNA synthetase from *H.influenzae* genomic DNA (ATCC51907) (21) by PCR. The forward primer (5′-GCG GTC GAC ATG CTA AAA ATT TTC AAT ACC TTA A-3′) corresponds to the N-terminus of the gene, while the reverse primer (5′-CGC ACT AGT CTA TTG CTT ACG CCA AGT TGT CCC A-3′) corresponds to the C-terminus of the gene. The PCR product was cloned into the pTZ19R vector (Bio-Rad) and errors that arose during the PCR reaction were corrected by site-directed mutagenesis. The correct gene was excised by digestion with *Sal*I and *Spe*I, and subcloned into the corresponding sites of the pKS583 expression vector (26) to produce the construct pM392. In pM392, the gene for *H.influenzae* cysteine-tRNA synthetase is fused to a sequence at the 5′-end so that expression of the gene will result in an N-terminal $(His)_6$ tag (N-MRGSHHHHHHSS-GWVD). For protein expression, *E.coli* BL21(DE3) (27) was

Table 1. Kinetic parameters of aminoacylation of $tRNA^{Cys}$ of *H.influenzae* and *E.coli* by their respective cysteine-tRNA synthetases

	k_{cat}/K_m (M^{-1} s^{-1})	Relative
Haemophilus influenzae cysteine-tRNA synthetase		
H.influenzae $tRNA^{Cys}$	1.8×10^5	1.00
E.coli $tRNA^{Cys}$	2.6×10^5	1.4
The G15:C48 variant of *E.coli* $tRNA^{Cys}$	4.1×10^3	0.023
The A13:A22:G46 variant of *E.coli* $tRNA^{Cys}$	4.4×10^3	0.024
The A13U and U21A variant of *E.coli* $tRNA^{Cys}$	2.2×10^3	0.012
B.subtilis $tRNA^{Cys}$	2.5×10^5	1.4
Escherichia coli cysteine-tRNA synthetase		
H.influenzae $tRNA^{Cys}$	1.7×10^5	0.81
E.coli $tRNA^{Cys}$	2.1×10^5	1.00
The G15:C48 variant of *E.coli* $tRNA^{Cys}$	7.8×10^3	0.037
The A13:A22:G46 variant of *E.coli* $tRNA^{Cys}$	4.6×10^3	0.027
The A13U and U21A variant of *E.coli* $tRNA^{Cys}$	2.3×10^3	0.013
B.subtilis $tRNA^{Cys}$	3.4×10^5	1.6

The k_{cat}/K_m values were determined by individually measuring k_{cat} and K_m through the Michaelis–Menten equation and were an average of at least three determinations.

transformed by pM392, and transformed cells were grown in LB broth with 100 μg/ml ampicillin and induced at an OD_{600} of 0.5 with 1 mM IPTG for 4 h. The overexpressed *H.influenzae* cysteine-tRNA synthetase was purified using the B-PER 6×His Protein Purification kit as described by the manufacturer (Pierce Inc.).

RESULTS

Contribution of the core of *H.influenzae* $tRNA^{Cys}$ to aminoacylation

The core of *H.influenzae* $tRNA^{Cys}$ shares unusual features with that of *E.coli* $tRNA^{Cys}$. These features include G15:G48, A13:A22, A9 and A46 (Fig. 1). We wanted to determine if these features were important for aminoacylation so as to assess if the core of *H.influenzae* tRNA was functionally similar to that of *E.coli* tRNA. For this purpose, we cloned and expressed the genes for *H.influenzae* $tRNA^{Cys}$ and its cysteine-tRNA synthetase according to the *H.influenzae* genomic database (21). The tRNA gene was constructed by joining together several oligonucleotides that overlapped with each other to form the complete gene under the control of the T7 promoter. This allowed transcription of the gene by T7 RNA polymerase as the substrate for aminoacylation. Inspection of the tRNA gene shows that it has preserved U73 and the GCA anticodon that are important for aminoacylation with cysteine (28; Fig. 1). However, the tRNA gene encodes an unusual C27:U43 mismatch in the anticodon stem, the significance of which is not known.

The gene for *H.influenzae* cysteine-tRNA synthetase was constructed by PCR. It encodes 459 amino acids, which is shorter by two amino acids than the 461 amino acids encoded by the *E.coli* gene (24). The gene for the *H.influenzae* enzyme was fused with a sequence encoding a his tag at the 5′-end. We expressed the fusion gene in *E.coli* BL21(DE3) and purified the recombinant enzyme through a nickel-chelated resin for which the his tag has a high affinity. Analysis of the purified recombinant enzyme on an SDS gel indicated a major band (~80% purity) of molecular weight 53 kDa, which is consistent with the molecular weight expected from the deduced amino acid sequence of the gene.

Table 1 shows that the catalytic properties of aminoacylation of the *H.influenzae* enzyme are strikingly similar to those of the *E.coli* enzyme. First, the *H.influenzae* enzyme efficiently aminoacylated the T7 transcript of *H.influenzae* $tRNA^{Cys}$ with a k_{cat}/K_m (1.8×10^5 M^{-1} s^{-1}) that is comparable to that of the *E.coli* enzyme with the T7 transcript of *E.coli* $tRNA^{Cys}$ (2.1×10^5 M^{-1} s^{-1}). The similarity in the k_{cat}/K_m values is a result of similar k_{cat} and K_m for each tRNA by its homologous synthetases. The *E.coli* tRNA has a k_{cat} of 0.5 s^{-1} and a K_m of 2.38 μM. The *H.influenzae* tRNA has a k_{cat} of 0.52 s^{-1} and a K_m of 2.89 μM. Second, the *H.influenzae* enzyme and the *E.coli* enzyme were capable of cross-aminoacylation of each other's tRNA. The k_{cat}/K_m of aminoacylation of the *E.coli* tRNA by the *H.influenzae* enzyme is 1.4 relative to that of the homologous *H.influenzae* tRNA. Similarly, the k_{cat}/K_m of aminoacylation of the *H.influenzae* tRNA by the *E.coli* enzyme is 0.8 relative to that of the homologous *E.coli* tRNA. These similar k_{cat}/K_m values indicate that, despite differences in their primary structures, the two tRNAs share common structural motifs that are important for aminoacylation by the *H.influenzae* and *E.coli* enzymes.

Previous studies with the *E.coli* enzyme and its T7 tRNA transcript show that the enzyme does not depend on modified nucleotides that are absent from the T7 transcript (12). The k_{cat}/K_m of aminoacylation with the T7 transcript was only 3-fold below that of the tRNA containing modifications. This is insignificant compared to 2–5 orders of magnitude of major determinants that are important for aminoacylation. The k_{cat}/K_m value of the *H.influenzae* enzyme with its own T7 transcript and the k_{cat}/K_m of the enzyme with the T7 transcript of *E.coli* tRNA are similar to each other and to those of the *E.coli*

Table 2. Differences in the core among the cysteine-specific tRNAs

Type	D3V4			D4V5
	Eubacteria			Eukarya, some archaea
15:48	G:G	G:C	G:C	G:C
13:22	A:A	A:A	G:A	W-C
46	A	A	A	
9	A	A	G	
tRNA species	EC, HI	MP, MG, BB, BS, SA, SS, HP	MTb, SL	HS, SC, SP, MJ, MTh

EC, *Escherichia coli*; BB, *Borrelia burgdorferi*; BS, *Bacillus subtilis*; HI, *Haemophilus influenzae*; HP, *Helicobacter pylori*; HS, *Homo sapiens*; MG, *Mycoplasma genitalium*; MJ, *Methonococcus jannaschii*; MP, *Mycoplasma pneumoniae*; MTb, *Mycobacteria tuberculosis*; MTh, *Methanobacterium thermoautotrophicum*; SA, *Staphylococcus aureus*; SC, *Saccharomyces cerevisiae*; SL, *Streptomyces lividans*; SP, *Schizosaccharomyces pombe*; SS, *Synechocystis* sp.

enzyme. This suggests that the *H.influenzae* enzyme also does not emphasize modifications of its tRNA.

Table 1 shows evidence that G15:G48 and the proposed A46:[A13:A22] are important for aminoacylation by the *H.influenzae* enzyme. Specifically, the *H.influenzae* enzyme failed to efficiently aminoacylate two defective substrates for the *E.coli* enzyme, the G15:C48 and the G46:[A13:A22] variants of *E.coli* $tRNA^{Cys}$. The deficiency in k_{cat}/K_m of aminoacylation with the G15:C48 variant and that with the G46:[A13:A22] variant were quantitatively similar to those of the *E.coli* enzyme. Thus, the sequence conservation of G15:G48 and A13, A22 and A46 between *E.coli* and *H.influenzae* tRNAs has functional significance. Further, the structural coupling between G15:G48 and the proposed A46:[A13:A22] triple in *E.coli* $tRNA^{Cys}$ is preserved in the *H.influenzae* tRNA as well. Coupling in the *E.coli* tRNA has been demonstrated (13). Specifically, while substitution of A13 with U13 was functional, and while substitution of U21 with A21 was functional, the double mutant containing both the A13U and U21A substitutions was defective in *E.coli* $tRNA^{Cys}$. The defect of the double mutant was manifested in two features. One was the loss of the chemical signature of G15:G48 in the tRNA. The other was a decrease in the k_{cat}/K_m of aminoacylation by two orders of magnitude. The two features displayed by the double mutant are tightly coupled and this coupling is important for aminoacylation by the *E.coli* enzyme. Table 1 shows that the *H.influenzae* enzyme also emphasized the coupling as it failed to aminoacylate the double mutant containing the A13U and U21A mutations. The decrease in k_{cat}/K_m of aminoacylation with the double mutant is quantitatively similar to that for the *E.coli* enzyme.

The close parallel between the *E.coli* and *H.influenzae* enzymes is further demonstrated by their activity with the *B.subtilis* $tRNA^{Cys}$. Both reacted efficiently with this tRNA. Table 1 shows that the k_{cat}/K_m value of the *B.subtilis* tRNA for the *E.coli* enzyme relative to that of the *E.coli* tRNA is 1.6. Similarly, the k_{cat}/K_m value of the *B.subtilis* tRNA for the *H.influenzae* enzyme relative to that of the *H.influenzae* tRNA is 1.4. The *B.subtilis* tRNA shares the same D3V4 configuration (3 bp in the D stem and 4 nt in the V loop) as the *E.coli* and *H.influenzae* tRNAs (Table 2). However, the *B.subtilis* tRNA differs from the *E.coli* and *H.influenzae* tRNAs by having G15:C48 (instead of G15:G48), U46 (instead of A46) and A21 (instead of U21) in the tRNA core. The efficient activity of this tRNA with both the *E.coli* and *H.influenzae* enzymes provides an example of a tRNA core featuring G15:C48 that can be efficiently recognized by these two enzymes.

Close relationship between the *H.influenzae* and *E.coli* cysteine-tRNA synthetases

The parallel between the *H.influenzae* and *E.coli* cysteine-tRNA synthetases in aminoacylation suggests a close relationship of these two enzymes in evolution. This is supported by the observation that the two enzymes share 77% identity and 86% similarity throughout their sequences. To gain a better insight into the relatedness of these two enzymes in evolution, we performed multiple sequence alignments of these two enzymes with other members of the family of cysteine-tRNA synthetases.

Figure 2 shows a multiple sequence alignment of six cysteine-tRNA synthetase enzymes from *E.coli*, *H.influenzae*, *Mycoplasma genitalium*, *Helicobacter pylori*, *Mycobacteria tuberculosis* and *Homo sapiens*, respectively. This alignment shows that the N-terminal halves of these enzymes have preserved the HIGH and KMSKS motifs that are characteristics of the class I synthetases in crystal structures. In general, structural analysis has divided aminoacyl-tRNA synthetases into two classes (29,30). The class I enzymes share the HIGH and KMSKS motifs, which constitute the catalytic site and are responsible for activation of an amino acid with ATP to generate an aminoacyl-adenylate as the first step of aminoacylation. The catalytic site of the class I synthetases is embedded in a nucleotide-binding fold located in the N-terminal half of the enzyme (31). This nucleotide-binding fold is characterized by alternating β-strands and α-helices. The class II enzymes share three other motifs as the catalytic site, which is embedded in six anti-parallel strands (30). Although there is no crystal structure available for any cysteine-tRNA synthetase, the presence of HIGH and KMSKS designates the cysteine enzymes as class I synthetases (24).

Also shown in Figure 2 are secondary structural elements of the six enzymes that are modeled according to known structures of class I synthetases. This modeling focuses on the N-terminal half and shows the characteristic alternation of β-strands and α-helices. The modeling was based on the alignment of the *E.coli* enzyme in a previously published structure-based multiple sequence alignment of all class I synthetases (32). In the previous alignment, because secondary structure

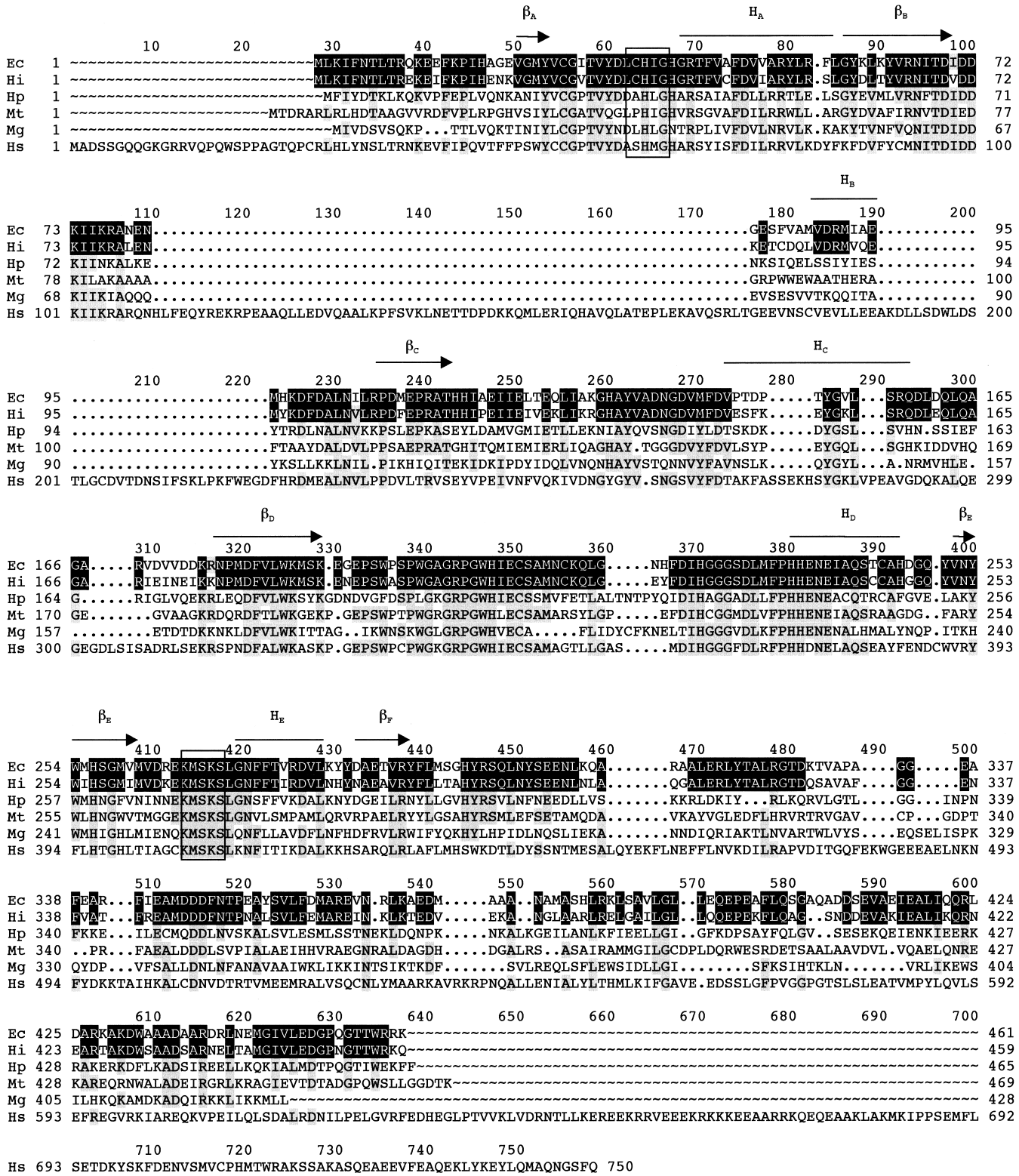

Figure 2. Alignment of six sequences of cysteine-tRNA synthetase. The alignment was achieved through the program CLUSTAL W. The conserved HIGH and KMSKS motifs that are characteristics of the class I synthetases are boxed. Secondary structural elements of β-strands (indicated as β) and α-helices (indicated as α) that are modeled from a structurally based multiple sequence alignment of all class I synthetases are indicated at the top of the sequence alignment. Residues that are conserved among these six sequences are shaded, whereas those that are common between the *E.coli* and *H.influenzae* enzymes are shown in black. Dots indicate gaps in the alignment, while numbers at the ends of each column of the alignment indicate the amino acid positions in each sequence. Abbreviations and EMBL accession numbers: Ec, *E.coli*, 1786737; Hi, *H.influenzae*, 1174501; Hp, *Helicobacter pylori*, 2507427; Mt, *Mycobacteria tuberculosis*, 3122884; Mg, *Mycoplasma genitalium*, 1351147; Hs, *Homo sapiens*, 1711625.

elements in known crystal structures of several class I synthetases were superimposable, they were used as guidelines to model sequences in other class I enzymes where structural information is not yet available. This established predictions for secondary structure elements in the *E.coli* enzyme. In the present alignment, the sequence of the *E.coli* enzyme and its modeled secondary structure elements were used to align and model all other sequences.

The alignment in Figure 2 shows that the *E.coli* enzyme is most closely related to the *H.influenzae* enzyme. In the N-terminal domain of the *E.coli* enzyme, which ends after the secondary structure element β_F, 244 of the 291 amino acids are identical to those of the *H.influenzae* enzyme. This accounts for an 84% identity. By comparison, a similar calculation with the *M.tuberculosis* enzyme and with the human enzyme shows 128 and 130 residues of identity, respectively. These account for ~44% identity for each. The strong similarity between the *E.coli* and *H.influenzae* enzymes is also evident in the phylogenetic analysis in Figure 3. This analysis used the neighbor joining distance method (33) to calculate the relatedness of the N-terminal half of 15 cysteine-tRNA synthetases. These include the six enzymes in Figure 2, five additional enzymes of eubacteria (*B.subtilis*, *Aquifex aeolicus*, *Treponema pallidum*, *Borrelia burgdorferi* and *Mycoplasma pneumoniae*), two additional enzymes of eukarya (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) and two enzymes of archaea (*Archaeoglobus fulgidus* and *Pyroccocus horikoshii*). Clearly, even among these 15 enzymes that are widely separated in evolution, the *E.coli* and *H.influenzae* enzymes are most similar in their N-terminal halves.

The *E.coli* and *H.influenzae* enzymes also show the greatest similarity in their C-terminal halves. The two enzymes share 111 amino acids in common, which is 65% identity. The lower identity in the C-terminal half (65%) than that of the N-terminal half (84%) is expected, as the C-terminal halves of class I synthetases are generally more diverse. Because of the diversity of the C-terminal half among members of the class I synthetases, and because of the lack of a crystal structure for cysteine-tRNA synthetase, we did not perform structural modeling for the C-terminal half. For comparison, the identity in the C-terminal half between the *E.coli* and *M.tuberculosis* enzymes or between the *E.coli* and human enzymes is even lower, at 10 and 16%, respectively. Thus, while the *H.influenzae* enzyme maintains a relatively strong similarity in the C-terminal half to that of the *E.coli* enzyme (65%), the *M.tuberculosis* and human enzymes do not (10 and 16% identity). The similarity in the C-terminal halves of the *E.coli* and *H.influenzae* enzymes is significant. While the N-terminal half of class I synthetases provides contact with the acceptor stem of their tRNAs (7), the C-terminal half provides contact with other regions, such as the tRNA core. The close relationship between the *E.coli* and *H.influenzae* enzymes in the C-terminal halves suggests a rationale for their ability to recognize conserved features of the tRNA core.

DISCUSSION

The sequence of the tRNA core in *E.coli* $tRNA^{Cys}$ is strikingly similar to that of the *H.influenzae* $tRNA^{Cys}$. The two sequences share the D and TΨC loops in common and they differ only in the 10:25 bp in the D stem and nucleotide 44 in the V loop

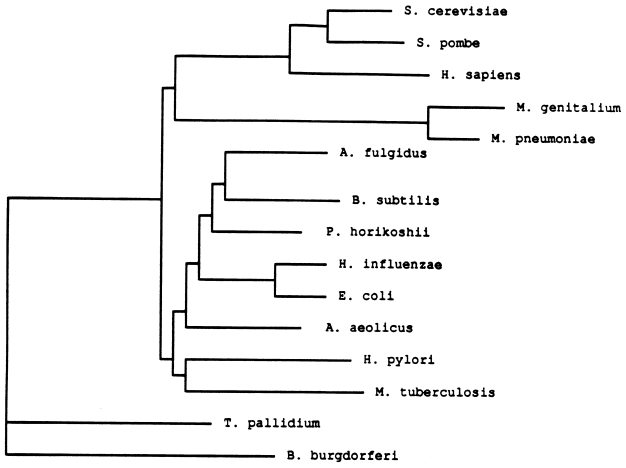

Figure 3. Unrooted phylogeny of the N-terminal half of cysteine-tRNA synthetases constructed by using the neighbor joining method (33) according to the sequence alignment of Figure 2. Abbreviations and EMBL accession numbers: A. aeolicus, *Aquifex aeolicus*, 2983551; A. fulgidus, *Archaeoglobus fulgidus*, 3122884; B. subtilus, *Bacilus subtilis*, 549024; B. burgdorferi, *Borrelia burgdorferi*, 2688519; E. coli, *E.coli*, 1786737; H. influenzae, *Haemophilus influenzae*, 1174501; H. pylori, *Helicobacter pylori*, 2507427; H. sapiens *Homo sapiens*, 1711625; M. genitalium, *Mycoplasma genitalium*, 1351147; M. pneumoniae, *Mycoplasma pneumoniae*, 2500964; M. tuberculosis; *Mycobacteria tuberculosis*, 3122884; P. horikoshii, *Pyrococcus horikoshii*, 3257043; S. cerevisiae; *Saccharomyces cerevisiae*, 1730840; S. pombe, *Schizosaccharomyces pombe*, 13516321; T. pallidum, *Treponema pallidum*, 3322350.

(Fig. 1). We present evidence that the sequence conservation of the core has significance for functional conservation, as manifested by the coupling relationship between G15:G48 and A13:A22 and the contribution of the coupling to aminoacylation with cysteine. These studies provide an example of conservation of the tRNA core in aminoacylation. Because the structure of the tRNA core is complicated and contains many details, conservation of the core in aminoacylation was not necessarily anticipated. For example, while the core of yeast $tRNA^{Phe}$ and that of *E.coli* $tRNA^{Phe}$ are both important for aminoacylation (15,18), the details of the core are different. Yeast $tRNA^{Phe}$ emphasizes G20 in the D loop, whereas *E.coli* $tRNA^{Phe}$ emphasizes U20. Thus, the core does not need to conserve sequence in order to conserve function. In principle, there can be many ways to formulate a core. In this respect, conservation of the core sequence and function between the *E.coli* and *H.influenzae* $tRNA^{Cys}$ is significant. Previous studies have indicated that tRNA elements that are important for aminoacylation can be conserved in evolution. However, these elements are located in the anticodon and the acceptor stem, where the tRNA structure is simpler and more straightforward. The role of the anticodon in the conservation of aminoacylation is not surprising, because it is conserved in evolution and is usually involved in recognition by synthetases. Conservation of nucleotides in the acceptor stem includes examples such as the G3:U70 base pair of $tRNA^{Ala}$ (34), A73 of $tRNA^{Phe}$ (15,18), U73 of $tRNA^{Cys}$ (28) and U73 or A73 of $tRNA^{Gly}$ (35).

The conservation of the core in aminoacylation between *E.coli* $tRNA^{Cys}$ and *H.influenzae* $tRNA^{Cys}$ separates these two tRNAs from all other cysteine-specific tRNAs. Although the

two tRNAs also share the important U73 and the GCA anticodon for aminoacylation (28), these elements are conserved in all cysteine-specific tRNAs (10). Thus, the core is the only region that provides determinants to be used to discern the relatedness of different species in the family of cysteine-specific tRNAs. Table 2 summarizes the analysis of the tRNA core specific for cysteine tRNAs. This analysis shows that the *E.coli* and *H.influenzae* tRNAs belong to a unique subclass of eubacterial tRNAs. In this subclass, they share G15:G48 and they contain A13:A22, A9 and A46. All other eubacterial tRNAs have G15:C48. While some have A13:A22, others have G13:A22. However, the *E.coli* and *H.influenzae* tRNAs share the common D3V4 feature with all eubacterial tRNAs. In contrast, eukaryotic tRNAs have G15:C48 and display the D4V5 feature.

We show here that the *B.subtilis* $tRNA^{Cys}$ is an efficient substrate for both the *E.coli* and *H.influenzae* enzymes. The *B.subtilis* tRNA is a representative of the major subclass of eubacterial tRNAs. It has G15:C48 and A13:A22 but lacks A46 so that it is unlikely to form the proposed A46:[A13:A22] base triple. The efficient aminoacylation of this tRNA by the *E.coli* and *H.influenzae* enzymes is unexpected. Data presented in Table 1 and data from previous studies (12,13) indicate that the *E.coli* enzyme is sensitive to G15:C48. Additionally, two representative members of the eukaryotic tRNAs, the *S.cerevisiae* and human tRNAs, are poor substrates for the *E.coli* enzyme (28). Both of these tRNAs have G15:C48. Efficient aminoacylation of the *B.subtilis* tRNA thus suggests that the effect of G15:C48 in this tRNA is compensated for by other features in the core, such as lack of the 46:[13:22] base triple. Recent studies indicate that members of the other subclass of eubacterial tRNAs featuring G15:C48 and G13:A22 in the D3V4 configuration are also efficiently aminoacylated by the *E.coli* enzyme (T.W.Christian and Y.M.Hou, in preparation). Thus, while G15:G48 in the tRNA core is important for aminoacylation for the *E.coli* and *H.influenzae* enzymes, there is a certain sequence context in which G15:C48 can function. The details of the sequence context remain to be elucidated. However, we have not found examples of eukaryotic tRNAs in which G15:C48 is recognized by the *E.coli* enzyme. It appears that the functional context for G15:C48 with respect to the *E.coli* enzyme is achievable only in the eubacterial D3V4 configuration and not in the eukaryotic D4V5 configuration.

The separation of eubacterial tRNAs from eukaryotic tRNAs based on the core is analogous to the separation of eubacterial cysteine-tRNA synthetase enzymes from eukaryotic enzymes shown in Figures 2 and 3. Specifically, Figure 3 shows that separation of the eubacterial enzymes from the eukaryotic enzymes can be easily visualized in the phylogenetic tree. Inspection of the human enzyme in Figure 2 (as a representative of the eukaryotic enzymes) shows that it is distinguished from eubacterial enzymes by three features, a long N-terminus, a long C-terminus, and a long insertion domain after the secondary structure element β_B in the catalytic site. Thus, there is a correlation between the structure of a tRNA core and the structure of a cysteine-tRNA synthetase.

The evolutionary relationships of cysteine-specific archaeal tRNAs and synthetases to those of eubacteria and eukarya are not clear at present. Some archaeal tRNAs have the D4V5 configuration (Table 2), whereas others have the D3V5 configuration. Although the known archaeal synthetases are more related to the eubacterial group (Fig. 3), other archaea lack even an identifiable cysteine-tRNA synthetase (36,37). Additional sequences of archaeal tRNAs and synthetases are necessary to gain insights into their aminoacylation with cysteine.

The close parallel between *E.coli* and *H.influenzae* enzymes for aminoacylation of their tRNAs has evolutionary implications. The genome of *H.influenzae* is one of the smallest of autonomous organisms (38). It is certainly the smallest of the genomes of the γ-proteobacteria (*H.influenzae*, *E.coli* and *Synechocystis* sp.) that have been completely sequenced (39–41). Although this genome consists of only 1.83×10^6 base pairs, which is about one-third the size of that of *E.coli* (4.67×10^6 base pairs), it nonetheless must encode all the components necessary to sustain a cellular life. As with the genome of *E.coli*, the genome of *H.influenzae* encodes only one $tRNA^{Cys}$ and aminoacylation of this tRNA is dependent on coupling between A13:A22 and G15:G48. The significance of this coupling thus may be traced back to the common ancestor of *H.influenzae* and *E.coli* in their evolutionary lineages. This suggests an ancient history of the coupling and the 15:48 and 13:22 motifs that comprise the coupling in the development of aminoacylation with cysteine.

We have recently proposed that the coupling is mediated by an A46:[A13:A22] adenosine base triple (20) in the ligand-free *E.coli* $tRNA^{Cys}$. The crystal structure of the tRNA bound to EF-Tu shows an A9:[A13:A22] base triple. It seems certain that A13:A22 is in an adenosine base triple, either with A46 or with A9. This provides an important new insight into the role of adenosine base triples in tRNA aminoacylation. Previously, the significance of adenosine base triples or adenosine-rich motifs in the general principles of RNA folding and RNA structure was well recognized. For example, early phylogenetic analysis of the 16S rRNA secondary structure indicates that adenosine-rich motifs appear at a higher frequency than any other nucleotides at the universally conserved positions (42,43). Additionally, adenosine-rich motifs account for >60% of the unpaired regions of the RNA that presumably provide the binding sites for ribosomal proteins (42). The recent crystal structure of the P4–P6 domain of the *Tetrahymena* group I intron shows that adenosine-rich motifs are central to the helical packing of the RNA (44,45). These motifs also provide the binding sites for metal ions and for ATP in RNA structures (46,47). The functional importance and diversity of adenosine-rich motifs thus suggests an ancient history of these motifs that may be the basis for the primordial 'RNA world'. Our discovery that an adenosine base triple has an ancient role in tRNA aminoacylation further contributes to this idea.

ACKNOWLEDGEMENTS

We thank Dr Caryn Evilia for sequence alignment, Remy Fiorentino for analysis of databases, and Lijuan Mei for assistance with construction of some of the mutants presented in this work. This work was supported in part by grant number GM 47935 from the NIH (to Y.-M.H.), a grant from the Jefferson Medical College (to Y.-M.H.), and a grant from the Ministry of Education, Science, and Culture of Japan (to K.S.).

REFERENCES

1. Biou,V., Yaremchuk,A., Tukalo,M. and Cusack,S. (1994) *Science*, **263**, 1404–1410.
2. Cusack,S., Yaremchuk,A. and Tukalo,M. (1996) *EMBO J.*, **15**, 6321–6334.
3. Kim,S.H., Suddath,F.L., Quigley,G.J., McPherson,A., Sussman,J.L., Wang,A.H., Seeman,N.C. and Rich,A. (1974) *Science*, **185**, 435–440.
4. Nissen,P., Kjeldgaard,M., Thirup,S., Polekhina,G., Reshetnikova,L., Clark,B.F. and Nyborg,J. (1995) *Science*, **270**, 1464–1472.
5. Nissen,P., Thirup,S., Kjeldgaard,M. and Nyborg,J. (1999) *Structure*, **7**, 143–156.
6. Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F. and Klug,A. (1974) *Nature*, **250**, 546–551.
7. Rould,M.A., Perona,J.J., Soll,D. and Steitz,T.A. (1989) *Science*, **246**, 1135–1142.
8. 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.
9. Klug,A., Robertus,J.D., Ladner,J.E., Brown,R.S. and Finch,J.T. (1974) *Proc. Natl Acad. Sci. USA*, **71**, 3711–3715.
10. Sprinzl,M., Horn,C., Brown,M., Ioudovitch,A. and Steinberg,S. (1998) *Nucleic Acids Res.*, **26**, 148–153.
11. Giege,R., Sissler,M. and Florentz,C. (1998) *Nucleic Acids Res.*, **26**, 5017–5035.
12. Hou,Y.M., Westhof,E. and Giege,R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6776–6780.
13. Hou,Y.M. (1994) *Biochemistry*, **33**, 4677–4681.
14. Nureki,O., Niimi,T., Muramatsu,T., Kanno,H., Kohno,T., Florentz,C., Giege,R. and Yokoyama,S. (1994) *J. Mol. Biol.*, **236**, 710–724.
15. Peterson,E.T. and Uhlenbeck,O.C. (1992) *Biochemistry*, **31**, 10380–10389.
16. McClain,W.H. and Foss,K. (1988) *J. Mol. Biol.*, **202**, 697–709.
17. Asahara,H., Himeno,H., Tamura,K., Nameki,N., Hasegawa,T. and Shimizu,M. (1994) *J. Mol. Biol.*, **236**, 738–748.
18. Sampson,J.R., DiRenzo,A.B., Behlen,L.S. and Uhlenbeck,O.C. (1989) *Science*, **243**, 1363–1366.
19. Hamann,C.S. and Hou,Y.M. (1997) *Biochemistry*, **36**, 7967–7972.
20. Hamann,C.S. and Hou,Y.M. (1997) *Bioorg. Med. Chem.*, **5**, 1011–1019.
21. 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., Merrick,J.M. *et al*. (1995) *Science*, **269**, 496–512.
22. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
23. Grodberg,J. and Dunn,J.J. (1988) *J. Bacteriol.*, **170**, 1245–1253.
24. Hou,Y.M., Shiba,K., Mottes,C. and Schimmel,P. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 976–980.
25. Ehresmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.P. and Ehresmann,B. (1987) *Nucleic Acids Res.*, **15**, 9109–9128.
26. Shiba,K., Stello,T., Motegi,H., Noda,T., Musier-Forsyth,K. and Schimmel,P. (1997) *J. Biol. Chem.*, **272**, 22809–22816.
27. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
28. Lipman,R.S. and Hou,Y.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 13495–13500.
29. Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) *Nature*, **347**, 203–206.
30. Cusack,S., Berthet-Colominas,C., Hartlein,M., Nassar,N. and Leberman,R. (1990) *Nature*, **347**, 249–255.
31. Brunie,S., Zelwer,C. and Risler,J.L. (1990) *J. Mol. Biol.*, **216**, 411–424.
32. Landes,C., Perona,J.J., Brunie,S., Rould,M.A., Zelwer,C., Steitz,T.A. and Risler,J.L. (1995) *Biochimie*, **77**, 194–203.
33. Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
34. Hou,Y.M. and Schimmel,P. (1989) *Biochemistry*, **28**, 6800–6804.
35. Hipps,D., Shiba,K., Henderson,B. and Schimmel,P. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 5550–5552.
36. Bult,C.H., White,O., Losen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G. *et al*. (1996) *Science*, **273**, 1058–1073.
37. Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H., Dubois,J., Aldredge,T. *et al*. (1997) *J. Bacteriol*., **179**, 7135–7155.
38. Koonin,E.V. and Mushegian,A.R. (1996) *Curr. Opin. Genet. Dev.*, **6**, 757–762.
39. Blattner,F.R., Plunkett,G., Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. (1997) *Science*, **277**, 1453–1474.
40. 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**, 185–209.
41. 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.
42. Gutell,R.R., Weiser,B., Woese,C.R. and Noller,H.F. (1985) *Prog. Nucleic Acid Res. Mol. Biol.*, **32**, 155–216.
43. Woese,C.R., Winker,S. and Gutell,R.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
44. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) *Science*, **273**, 1696–1699.
45. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) *Science*, **273**, 1678–1685.
46. Jiang,F., Kumar,R.A., Jones,R.A. and Patel,D.J. (1996) *Nature*, **382**, 183–186.
47. Cate,J.H., Hanna,R.L. and Doudna,J.A. (1997) *Nature Struct. Biol.*, **4**, 553–558.