Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 13495–13500, November 1998
Biochemistry

Aminoacylation of tRNA in the evolution of an aminoacyl-tRNA synthetase

RICHARD S. A. LIPMAN AND YA-MING HOU*

Thomas Jefferson University, Department of Biochemistry and Molecular Pharmacology, 233 South 10th Street, BLSB 220, Philadelphia, PA 19107

Edited by Marianne Grunberg-Manago, Institute of Physico-Chemical Biology, Paris, France, and approved September 15, 1998 (received for review May 19, 1998)

ABSTRACT **Aminoacyl-tRNA synthetases catalyze aminoacylation of tRNAs by joining an amino acid to its cognate tRNA. The selection of the cognate tRNA is jointly determined by separate structural domains that examine different regions of the tRNA. The cysteine-tRNA synthetase of *Escherichia coli* has domains that select for tRNAs containing U73, the GCA anticodon, and a specific tertiary structure at the corner of the tRNA L shape. The *E. coli* enzyme does not efficiently recognize the yeast or human $tRNA^{Cys}$, indicating the evolution of determinants for tRNA aminoacylation from *E. coli* to yeast to human and the coevolution of synthetase domains that interact with these determinants. By successively modifying the yeast and human $tRNA^{Cys}$ to ones that are efficiently aminoacylated by the *E. coli* enzyme, we have identified determinants of the tRNA that are important for aminoacylation but that have diverged in the course of evolution. These determinants provide clues to the divergence of synthetase domains. We propose that the domain for selecting U73 is conserved in evolution. In contrast, we propose that the domain for selecting the corner of the tRNA L shape diverged early, after the separation between *E. coli* and yeast, while that for selecting the GCA-containing anticodon loop diverged late, after the separation between yeast and human.**

The genetic code is established by the 20 aminoacyl-tRNA synthetases, each of which activates an amino acid with ATP and transfers the activated amino acid to the 3′ end of the cognate tRNAs. The conservation of the genetic code suggests that aminoacyl-tRNA synthetases evolved early and were possibly among the first protein enzymes to emerge from the RNA world (1–3). The contemporary synthetases are modular enzymes that consist of separate structural domains (4–6). An investigation into the evolution of synthetase domains might provide an important insight into the conservation of the genetic code. We describe here the delineation of the evolution of synthetase domains for the cysteine-specific enzyme.

We considered the evolution of cysteine-tRNA synthetase from *Escherichia coli* to yeast to human, which are species that are widely separated from each other in evolution. We also considered that the evolution of a synthetase is codeveloped with the evolution of tRNA determinants that are necessary for the enzyme to aminoacylate its tRNA. The codevelopment of a synthetase with its cognate tRNAs is necessary for the preservation of the genetic code and is supported by studies of various tRNA-synthetase interactions (7–12). Thus, the *E. coli* enzyme has developed to recognize tRNA determinants of *E. coli* $tRNA^{Cys}$, while the yeast enzyme has developed to recognize determinants of yeast $tRNA^{Cys}$. If the *E. coli* enzyme does not recognize yeast $tRNA^{Cys}$, this would suggest that determinants for aminoacylation of $tRNA^{Cys}$ have diverged from *E. coli* to yeast and that the synthetase domains that interact with these determinants have in turn diverged as well. By investigating the ability of the *E. coli* enzyme to aminoacylate the yeast and human $tRNA^{Cys}$, and by identifying determinants of $tRNA^{Cys}$ that have changed, we would have the ability to establish the evolutionary relationships of different domains in cysteine-tRNA synthetase. Our previous studies show that the *E. coli* enzyme indeed fails to recognize efficiently the yeast tRNA (13). This finding provided evidence for the evolution of tRNA determinants. We therefore sought to elucidate tRNA determinants that have evolved by restoring determinants of the *E. coli* tRNA back to the yeast and human tRNAs and by determining whether they conferred recognition of the yeast and human tRNAs by the *E. coli* enzyme. Those that were successful in conferring recognition indicate determinants that have diverged, which in turn indicate that synthetase domains that interact with these determinants have concomitantly diverged from *E. coli* to yeast to human.

Cysteine-tRNA synthetase is a class I synthetase that consists of an N-terminal domain of a nucleotide-binding fold and a C-terminal domain rich in α-helices (14–15). The nucleotide-binding fold is characterized by the HIGH and KMSKS motifs and is responsible for activation of cysteine and transfer of the activated cysteine to the 3′ end of $tRNA^{Cys}$. The two-domain structure of cysteine-tRNA synthetase is proposed based on the known structures of related class I synthetases (16–21). This simplified two-domain model provides a framework for interaction with the two-domain structure of the tRNA L shape (22). In this model, the N-terminal nucleotide-binding fold is to recognize the acceptor stem of $tRNA^{Cys}$, while the C-terminal domain is to recognize tRNA determinants outside of the acceptor stem (22). Together, these two domains ensure the specificity of selecting the correct tRNA for aminoacylation.

For the *E. coli* cysteine enzyme, the tRNA determinants are U73 in the acceptor stem, the GCA anticodon, and a G15·G48 base pair at the corner of the tRNA L shape (23–25). The emphasis on U73 (the discriminator base of the tRNA) and the anticodon is a feature that is common to many other aminoacyl-tRNA synthetases. The emphasis on G15·G48, however, is unusual and has not been documented for many synthetases. In the genome database (26), most tRNAs have a purine·pyrimidine at 15·48 that forms a tertiary base pair known as the Levitt base pair (27), which has a structure distinct from that of a Watson–Crick base pair. The *E. coli* $tRNA^{Cys}$ and *Haemophilus influenzae* $tRNA^{Cys}$ are the only two that contain G15·G48. The mechanism of how G15·G48 contributes to *E. coli* aminoacylation with cysteine is unknown. Although a direct contact with a synthetase domain is possible, G15·G48 most likely acts by means of an indirect mechanism by modulating the presentation of U73 and the GCA anticodon to the synthetase. The indirect mechanism is supported by biochemical evidence (25) and by the crystal structure of the class I glutamine-tRNA synthetase complexed with its tRNA, which shows no direct contact between the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviation: DMS, dimethyl sulfate.
*To whom reprint requests should be addressed. e-mail: hou1@jeflin.tju.edu.

synthetase and the 15·48 base pair (17). Assuming that G15·G48 functions by an indirect mechanism, we proposed three domains in *E. coli* cysteine-tRNA synthetase, whereby an N-terminal domain interacts with U73, a C-terminal subdomain interacts with GCA, and another C-terminal subdomain senses the ability of G15·G48 to modulate the conformation of the tRNA structure.

The evolutionary relationship of the three domains is unknown. Although we predicted that the domain that senses G15·G48 might have diverged early because this determinant is not conserved among cysteine tRNAs, we could not have predicted the evolutionary relationship between the domain for interaction with U73 and the domain for interaction with the anticodon. Both U73 and the GCA anticodon are strictly conserved in evolution, which suggested that domains for these two determinants might have originated at the same time and developed independently under a strong selective pressure. By using the approach outlined above, we have distinguished the evolutionary development of the two domains and have established a model that relates these two domains with the domain that senses G15·G48. Our studies provide strong evidence for an ancient origin of the U73-interacting domain and suggest that it may have been the ancestor of the present-day cysteine-tRNA synthetases.

MATERIALS AND METHODS

Enzymes of Cysteine-tRNA Synthetase. The *E. coli* enzyme was purified from the overproducer strain pYM107/JM109, which harbors the gene in plasmid pYM107 (14). The yeast enzyme of *Saccharomyces cerevisiae* was partially purified by passing the cell extract from the protease-deficient strain BJ3501 (MATα, pep4::H1543, prb1-D1.6R, his3-D200, ura 3–52, can1) through a Sepharose (Pharmacia) CL-6B DEAE resin and by eluting the enzyme with a 0–0.5 M NaCl gradient (13). The human enzyme was also partially purified by passing the cell extract from cultured HeLa cell (5 g) through a Sepharose CL-6B resin in a procedure similar to that described for the yeast enzyme. In addition, we obtained the cDNA for a C-terminal truncated form of the human enzyme, which encodes residues M1 to R618 and deletes residues D619 through Q750. The deleted fragment (D619 to Q750) of the human enzyme, as shown in a multiple-sequence alignment, represents a C-terminal extension unique to the eukaryotic cysteine-tRNA synthetases. The cDNA for the truncated enzyme was cloned by Variagenics (Cambridge, MA) based on homology of a human-expressed sequence tag to the *E. coli* enzyme (28) and was fused behind the gene for bacterial thioredoxin in plasmid pTrx-Fus (Invitrogen). The fusion protein was partially purified by passing the cell extract through a MonoQ (fast protein liquid chromatography) (Pharmacia) resin and was eluted with a gradient of 0–0.5 N NaCl. Cleavage of the fusion with enterokinase released the truncated enzyme. Because the truncated enzyme with or without the fusion behaved similarly in the kinetics of aminoacylation (no more than a 3-fold difference in k_{cat}/K_m, unpublished observations), we used the fusion protein as the source of the truncated enzyme. Biochemical analysis indicated that the truncated enzyme had similar values of relative k_{cat}/K_m values as the full length enzyme from the HeLa cell extract.

Aminoacylation of tRNAs and of Microhelices with Cysteine. Conditions for aminoacylation were as described (23–25). The concentrations of tRNAs ranged from 0.5 to 100 μM, while those for microhelices ranged from 10 to 200 μM. All tRNAs were prepared as T7 transcripts and mutations were introduced by site-directed mutagenesis to the tRNA gene in plasmid pTFMa (a derivative of pUC18). Microhelices were made by T7 transcription from a primer based on a single-stranded DNA template (29). Mutations in microhelices were introduced by synthesizing the DNA templates with the desired substitutions.

Chemical Modifications of tRNAs. Conditions for modification of tRNAs with dimethyl sulfate (DMS) and with kethoxal were as described (23–25). For detection of the DMS modification, tRNAs were labeled at the 3′ end and sites of modification were identified by aniline scission. For detection of the kethoxal modification, primer extension (5′ labeled primer) with reverse transcriptase was used to identify sites of modification.

RESULTS AND DISCUSSION

Our general approach was to challenge the *E. coli* enzyme to aminoacylate the yeast and human cysteine tRNAs (Fig. 1), which share U73 and the GCA anticodon with the *E. coli* tRNA but differ in many other nucleotides (26). Because both the yeast and human tRNAs contain G15·*C*48 instead of G15·G48 (Fig. 1), we did not expect them to be efficient substrates for the *E. coli* enzyme (13). Once we confirmed that the *E. coli* enzyme did not

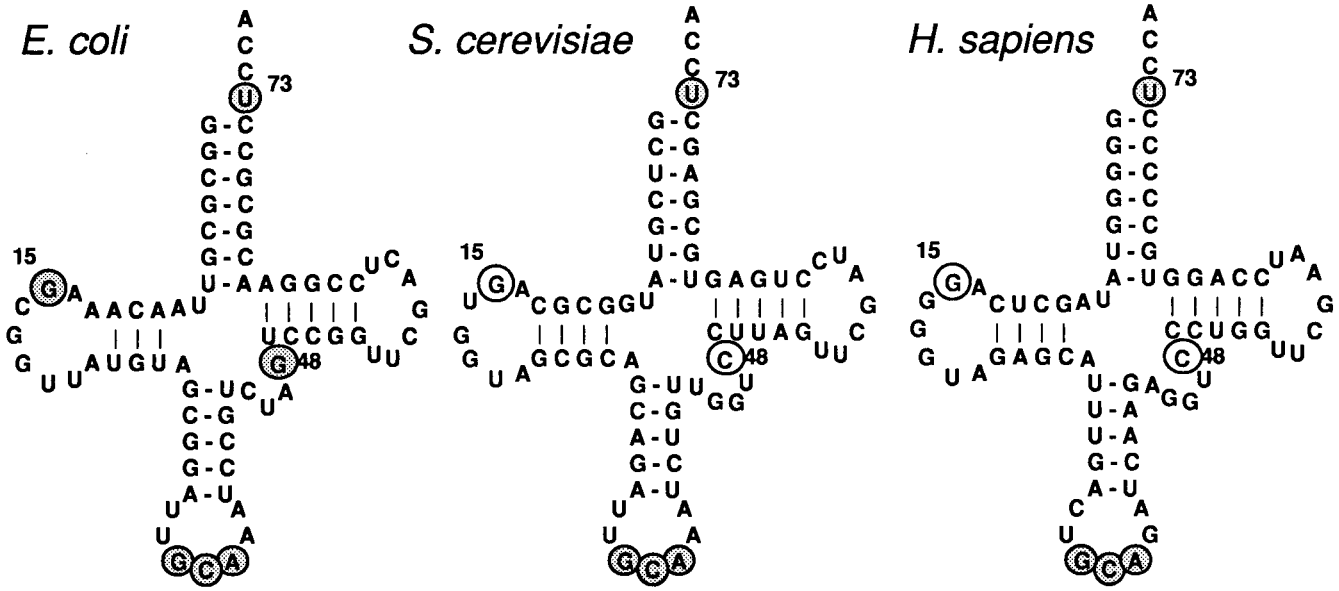


FIG. 1. Sequence and cloverleaf structure of $tRNA^{Cys}$ of *E. coli*, yeast (*S. cerevisiae*), and human. The U73 nucleotide, the GCA anticodon, and the G15·G48 tertiary base pair, which are important for aminoacylation of the *E. coli* tRNA by the *E. coli* enzyme, are highlighted by shaded circles. U73 and GCA are conserved in evolution and are important for aminoacylation of the yeast and human tRNAs by the respective homologous enzymes. In contrast, both the yeast and human tRNAs have G15·C48 (open circles) but do not specifically require this base pair for aminoacylation.

efficiently aminoacylate the heterologous tRNAs, we began to transfer elements of the *E. coli* tRNA that have changed back to the yeast and human tRNAs. Those that succeeded in converting the yeast and human tRNAs to be efficient substrates for the *E. coli* enzyme were interpreted as tRNA determinants that have diverged in the course of evolution.

Aminoacylation of Yeast and Human Cysteine tRNAs by the *E. coli* Enzyme. We compared the catalytic efficiencies of aminoacylation of the yeast and human tRNAs by the *E. coli* enzyme to that of the *E. coli* tRNA. The catalytic efficiency, expressed as k_{cat}/K_m, was the parameter that indicated the specificity and activity of aminoacylation. For example, we and others showed previously that the k_{cat}/K_m value of aminoacylation by the *E. coli* enzyme is reduced by 10^5-fold on substitution of U73 in the *E. coli* tRNA, that the value is reduced by 10^3-fold on substitution of the GCA anticodon, and that the value is reduced by 10^2-fold on substitution of G15·G48 with G15·C48 (13, 23, 24, 30). Thus, based on the criteria of k_{cat}/K_m, it is clear that U73 is the most important nucleotide for aminoacylation, followed by the GCA anticodon, followed by G15·G48. We showed in Table 1 that the k_{cat}/K_m values for the yeast and human tRNAs by the *E. coli* enzyme are two orders of magnitude below that of the *E. coli* tRNA. This finding confirmed the two heterologous tRNAs as poor substrates for the *E. coli* enzyme and raised the possibility that the determinants for tRNA aminoacylation have changed in the course of evolution. To test this possibility directly and to eliminate other possibilities such as the presence of negative determinants or alteration of tRNA structural framework for the yeast and human tRNAs, we next investigated whether introduction of structural features of G15·G48 to the yeast and human tRNAs would improve their catalytic efficiencies with the *E. coli* enzyme. We obtained mutants for the yeast and human +RNAs (yCys13 and hCys15, respectively) that show significant improvement in their k_{cat}/K_m with the *E. coli* enzyme (Table 1). The elucidation of these mutants is described below.

Introducing Structural Features of G15·G48 to the Yeast and Human tRNAs. We showed previously that recognition of G15·G48 in *E. coli* $tRNA^{Cys}$ by the *E. coli* enzyme depends on the structure at 15·48 but not on the sequence of 15·48 *per se* (24–25). The signature of this structure is that G15·G48 is accessible to two chemical probes that modify guanine. One is DMS, which modifies the N7 of G15, and the other is kethoxal, which modifies N1 and N2 of G15 and G48 (23–25). The accessibility to DMS and to kethoxal is unique to G15·G48 and has not been demonstrated in tRNAs containing G15·C48. These accessibilities therefore provide a convenient indicator for the structural features at 15·48 in *E. coli* $tRNA^{Cys}$ that are recognized by the *E. coli* enzyme.

We did not expect that simply introducing G15·G48 to replace G15·C48 in the yeast and human tRNAs would confer the signature modification. Previous studies showed that the structural features of G15·G48 are contributed not only by G15·G48 but also by others in the D loop and variable region (24–25, 31). Features of other nucleotides that are necessary for the proper presentation of this base pair in the structural framework of the tRNA include: (*i*) maintenance of A13·A22 in the D stem and A46 in the variable loop, which have the potential to form an (A13·A22)·A46 base triple that may influence the base pairing of G15·G48; (*ii*) elimination of nucleotide 47 to allow interaction between (A13·A22)·A46 and G15·G48; and (*iii*) exploration of A21 or U21 to optimize the presentation of G15·G48. These features were derived from our modeling studies of *E. coli* $tRNA^{Cys}$ (refs. 24–25 and 31; C. S. Hamann & Y.-M.H., unpublished work). To transplant the structural features of G15·G48 in the *E. coli* tRNA to the yeast and human tRNAs, we created mutants that incorporated various aspects of features described in *i* to *iii*. We tested mutants for aminoacylation by the *E. coli* enzyme. For those that were efficient substrates for the *E. coli* enzyme, we examined whether they acquired the signature of chemical modifications that are characteristics of G15·G48 in *E. coli* $tRNA^{Cys}$.

The results are summarized in Table 2. In essence, the yeast mutant (yCys13) that harbored the complete set of necessary structural elements showed a strong aminoacylation activity by the *E. coli* enzyme and exhibited the signature of chemical modifications of G15·G48 as that of the *E. coli* tRNA. All other mutants that harbored only a partial set of the necessary elements were not efficient substrates for the *E. coli* enzyme. Specifically, the mutant yCys13, which contained G15·G48, A13·A22·A46, A21, and had a deletion of U47, showed a k_{cat}/K_m value only 3-fold lower than that of the *E. coli* tRNA. It also showed the accessibility to DMS and kethoxal in the same manner as that of the *E. coli* tRNA, indicating the acquisition of the correct structure of G15·G48. Thus, introduction of structural features of G15·G48 in the yeast tRNA conferred efficient aminoacylation by the *E. coli* enzyme. This result suggests that structural features of G15·G48 were the major elements that diverged from the *E. coli* to the yeast tRNAs.

However, simply recreating the same structural features of G15·G48 in the human tRNA was not sufficient to convert the tRNA to an efficient substrate for the *E. coli* enzyme. The human tRNA mutants that harbored the necessary structural elements for G15·G48 and that acquired the signature of chemical modifications of G15·G48 as that of the *E. coli* tRNA did not show a strong aminoacylation activity by the *E. coli* enzyme. For example, the mutant hCys11, which contained G15·G48, A13·A22·A46, A21 and had a deletion of U47 reacted with DMS and kethoxal as the *E. coli* tRNA. However, hCys11 was a poor substrate for the *E. coli* enzyme, with a k_{cat}/K_m of 0.03 relative to that of the *E. coli* tRNA. Similarly, the mutant hCys07, which was identical to hCys11 except for harboring U21, was also a poor substrate for the *E. coli* enzyme, with a k_{cat}/K_m of 0.08 relative to that of the *E. coli* tRNA. The failure of the human tRNA to gain efficient aminoacylation by the *E. coli* enzyme, despite having the structure of G15·G48, was in contrast to that of the yeast tRNA. This indicates that the human tRNA has diverged further from the yeast tRNA.

The Anticodon Loop of the Human tRNA. To identify the elements that have diverged from the yeast tRNA to the human tRNA, we focused on the GCA anticodon next. We showed that the anticodon is critical for aminoacylation in both the yeast and human tRNAs because substitution of the anticodon with CUA inactivates recognition by their respective enzyme (k_{cat}/K_m decreased by 250- to 1,000-fold). Thus, the anticodon is conserved in sequence and in function from

Table 1. Aminoacylation of tRNAs by the *E. coli*, yeast (*S. cerevisiae*), and human cysteine-tRNA synthetases (CysRS)

	tRNAs (k_{cat}/K_m)				
CysRS	eCys01	yCys01	hCys01	yCys13	hCys15
E. coli	1.00	0.01	0.01	0.28	1.71
Yeast	1.01	1.00	1.01	0.25	0.69
Human	0.61	0.24	1.00	0.19	0.31

Relative k_{cat}/K_m values compared to that of the homologous tRNA-synthetase interaction (underlined). The wild-type *E. coli* $tRNA^{Cys}$ (eCys01), yeast $tRNA^{Cys}$ (yCys01), and human $tRNA^{Cys}$ (hCys01) serve as the homologous substrates for the *E. coli*, yeast, and human enzymes, respectively. The yeast tRNA mutant yCys13 and the human tRNA mutant hCys15 are two variants that are efficient substrates for the *E. coli* enzyme. All relative k_{cat}/K_m values are the average of at least three to five independent measurements. Data obtained for the *E. coli* enzyme are derived from measurements of k_{cat} and K_m values and have an average standard deviation of less than 10%. Data obtained for the yeast enzyme are derived from ratios of initial rates under conditions where tRNA substrates are 7- to 10-fold below Km values. These data have an average standard deviation of 30%. Data for the human enzyme are obtained from the HeLa cell lysate and are obtained by ratios of initial rates. These data are comparable to those obtained from the thioredoxin fusion (not shown) and have an average standard deviation of less than 10%.

Table 2. Progressive modifications of yeast $tRNA^{Cys}$ (yCys01) and of human $tRNA^{Cys}$ (hCys01) to generate yCys13 and hCys15

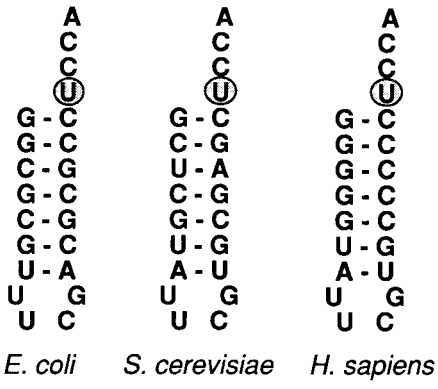
tRNA	15·48	13·22·46	47	21	32	37	Activity	DMS	KTX
eCys01	G·G	A·A·A	—	U	U	A	1.00	+	+
yCys01	G·C	C·G·G	U	A	U	A	0.01	−	−
yCys09	G·G	A·A·G	U	U	U	A	0.01	+	
yCys10	G·G	A·A·A	U	U	U	A	0.02	+	
yCys11	G·G	A·A·A	—	U	U	A	0.04	+	
yCys13	G·G	A·A·A	—	A	U	A	0.28	+	+
hCys01	G·C	C·G·G	U	A	C	G	0.02	−	−
hCys05	G·G	A·A·G	U	U	C	G	0.001	+	
hCys06	G·G	A·A·A	U	U	C	G	0.005	+	
hCys07	G·G	A·A·A	—	U	C	G	0.08	+	
hCys11	G·G	A·A·A	—	A	C	G	0.03	+	+
hCys12	G·G	A·A·A	—	U	U	A	0.97	+	+
hCys14	G·G	A·A·A	—	A	U	A	0.30		
hCys15	G·G	A·A·A	—	U	C	A	1.71	+	

Nucleotide sequences of *E. coli* $tRNA^{Cys}$ (eCys01) and those of the yeast and human tRNA variants. Activities are indicated as relative k_{cat}/K_m to that of the eCys01 tRNA by the *E. coli* enzyme and were determined by measurements of k_{cat} and K_m. The average of the standard deviations is 5%. Deletion of nucleotide 47 is indicated by "—", and chemical accessibility or inaccessibility to DMS and KTX (kethoxal) is indicated by "+" or "−".

E. coli to human. However, we noticed that the flanking sequences of the anticodon differ among the three tRNAs. In the anticodon loop, the yeast and *E. coli* tRNAs share the same sequence from U32 to A38, whereas the human tRNA differs by having C32 and G37 (Fig. 1). Of the two, G37 is of particular interest because it is immediately adjacent to the GCA anticodon. This comparison suggests that the anticodon loop of the human tRNA might provide the basis for the evolutionary difference between the yeast and the human tRNAs.

To test the significance of the anticodon loop, we introduced both U32 and A37 to hCys07 and hCys11 to create mutants hCys12 and hCys14, respectively. We tested the two mutants for aminoacylation by the *E. coli* enzyme. The results showed that both hCys12 and hCys14 were efficient substrates for the *E. coli* enzyme; their k_{cat}/K_m values were within 3-fold of that of the *E. coli* tRNA (Table 2). Thus, by introducing U32 and A37, which recreated the anticodon loop of the *E. coli* tRNA, we succeeded in converting the human tRNA to an efficient substrate for the *E. coli* enzyme. Of the hCys12 and hCys14 mutants, hCys12 had a better aminoacylation activity with the *E. coli* enzyme. We confirmed that this mutant maintained the signature of chemical modifications as that of the *E. coli* tRNA. This result verified that substitutions of nucleotides in the anticodon loop did not alter the structure of G15·G48. In the sequence framework of hCys12, we determined further that it was the substitution of G37 with A37 that was critical for aminoacylation by the *E. coli* enzyme. A mutant, hCys15, which maintained the G37A substitution but not the C32U substitution, retained the aminoacylation activity by the *E. coli* enzyme. Thus, a single nucleotide at position 37 represents a further diversification in the human tRNA from the *E. coli* and yeast tRNAs. Although this nucleotide is not part of the anticodon, its immediate proximity to the anticodon suggests a role in the development of the anticodon determinant for aminoacylation.

Aminoacylation of Microhelices Containing U73 by the *E. coli*, Yeast, and Human Enzymes. We demonstrated that the role of U73 as the major determinant for tRNA aminoacylation has been conserved from *E. coli* to yeast to human. The conservation from *E. coli* to yeast was shown previously (13). We now showed that the human enzyme is equally sensitive to substitution of U73. The k_{cat}/K_m value for the human enzyme for aminoacylation of the U73A variant of the human tRNA and of the *E. coli* tRNA is decreased 10^4- to 10^5-fold from that of the respective wild-type (not shown). These results provided preliminary evidence that U73 is critical for all three tRNAs. Here, we used the acceptor stem of each tRNA to demonstrate that U73 is the single determinant for aminoacylation for all three enzymes. For many synthetases, the acceptor stem of their cognate tRNAs contains sufficient information for specific aminoacylation (32). For example, we showed that the acceptor stem of the *E. coli* and of the yeast tRNAs, constructed in a microhelix motif, were substrates for aminoacylation by the *E. coli* and yeast enzymes (Fig. 2). Although these microhelix motifs have an overall 10^5-fold decrease in k_{cat}/K_m of aminoacylation relative to that of the full-length tRNAs, these motifs are specific substrates for the *E. coli* and yeast enzymes (13, 29). In both cases, aminoacylation depended on U73 such that the U73A substitution eliminated aminoacylation. Additionally, transfer of U73 to the microhelix of *E. coli* $tRNA^{Ala}$ conferred aminoacylation of the latter with cysteine by the *E. coli* enzyme (29). The use of the acceptor stem as substrate for aminoacylation allowed a closer examination of U73 and any possible effect of its flanking sequences. This experiment was relevant because while U73 is conserved in evolution, its flanking sequences differ among the three tRNAs,



CysRS	Relative catalytic efficiencies (*k*cat/Km)		
E. coli	1.00	0.59	0.37
S. cerevisiae	3.7	1.00	1.10
H. sapiens	0.97	0.96	1.00

FIG. 2. Sequence and the secondary structure of microhelices that respectively contain the acceptor stem domain of *E. coli*, yeast, and human $tRNA^{Cys}$, followed by a UUCG tetraloop. The U73 nucleotide, which is circled and is conserved in evolution, is the single determinant for aminoacylation of microhelices. The activities of aminoacylation of microhelices are indicated by relative k_{cat}/K_m values. The value of the homologous synthetase–microhelix interaction is taken as 1.00, and all values for the heterologous interactions are relative to the homologous interaction. CysRS, cysteine-tRNA synthetase.

starting from the second and the third base pairs in the acceptor stem (Fig. 2).

We synthesized microhelices corresponding to the acceptor stems of the *E. coli*, yeast and human tRNAs, respectively (Fig. 2). We showed that these microhelices were functional substrates for their homologous enzymes and that the relative k_{cat}/K_m values of aminoacylation of these microhelices were comparable, within a range of 2- to 4-fold of each other. Most importantly, these microhelices were aminoacylated by the heterologous enzymes. For example, the *E. coli* enzyme aminoacylated the microhelices of the yeast and human tRNAs, and the yeast and human enzymes aminoacylated the microhelices of the *E. coli* tRNA and of each other's tRNAs (Fig. 2). These results demonstrate that the differences in the flanking sequences were not barriers for the dominance of U73 in the recognition by the three enzymes. Further, we showed that aminoacylation of microhelices in all cases depended on U73; substitution of U73 with A73 eliminated aminoacylation (k_{cat}/K_m decreased by 10^5-fold). This finding establishes that, despite variations in the flanking sequences, U73 maintains as the single determinant that is conserved in evolution.

The Evolutionary Relationships of tRNA Determinants. By using the *E. coli* enzyme as a reference, and by following its ability to aminoacylate its homologous and heterologous tRNAs, we obtained evidence for the evolution of tRNA determinants from *E. coli* to yeast to human. In the yeast tRNA, we show that the subset that determines the structural features of the 15·48 base pair has evolved. In the human tRNA, we show that two subsets have evolved: one that determines the structural features of 15·48 and the other that determines the 37 nucleotide adjacent to the anticodon sequence. However, in the course of evolution, the U73 determinant has been conserved both in sequence and in function.

We inferred that evolution of tRNA determinants is accompanied by evolution of synthetase domains that "read" these determinants. Based on the crystal structure of the class I glutamine enzyme complexed with its tRNA, we inferred that the "reading" of U73 and the GCA anticodon is largely through direct interaction while the "reading" of the 15·48 base pair is to sense the structural effect of the base pair on the overall tRNA structure. Thus, in Fig. 3, which is a schematic representation of the tRNA-synthetase complex for the cysteine enzyme, we suggest a time line for the development of synthetase domains. We suggest that a C-terminal subdomain that senses the structural effect of 15·48 has diverged early, between *E. coli* and yeast, but that another C-terminal subdomain that interacts with the GCA anticodon has diverged later, between yeast and human. The divergence of the subdomain for interaction with the anticodon loop is limited and is localized to residues that contact the 37 nucleotide. In contrast, the N-terminal domain that interacts with U73 has conserved in evolution. This N-terminal domain thus was more ancestral than the anticodon-binding domain. Although residues in the N-terminal domain that directly contact U73 may be just a few, we envision that these residues are preserved by a strong selective pressure. Multiple sequence alignment of the cysteine enzymes deduced from genomic database supports this hypothesis. Several residues immediately adjacent to the HIGH and KMSKS motifs are strictly conserved in evolution (Y.-M.H., Fiorentino, R., and R.S.A.L., unpublished work). These residues could constitute the primordial domain that interacts with the conserved U73 and that provided the early basis for the specificity of aminoacylation of the acceptor stem. The ability of the *E. coli*, yeast, and human enzymes to crossaminoacylate each other's acceptor stem, solely based on U73, provides additional support for this hypothesis. Thus, aminoacylation of the tRNA acceptor stem might have participated in early events of protein synthesis that predated the events on the ribosomes involving anticodon–codon interactions.

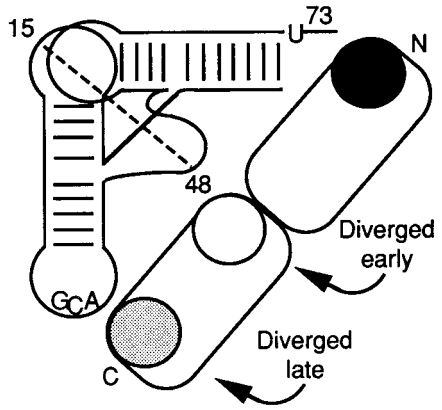


FIG. 3. Evolution of domains of cysteine-tRNA synthetase from *E. coli* to yeast to human. The domains are shown in the complex with the L-shaped structure of $tRNA^{Cys}$. The class I cysteine enzyme consists of an N-terminal domain and a C-terminal domain. The N-terminal domain provides the binding site for interaction with U73 in $tRNA^{Cys}$. The C-terminal domain contains two subdomains: one for sensing the structural effect of the 15·48 base pair on the tRNA structure and the other for interaction with the anticodon of the tRNA. The three sites are shown as circles in the *E. coli* enzyme. The binding site (closed circle) for interaction with U73 is conserved from *E. coli* to yeast to human. In contrast, the site (open circle) that senses the structural effect of 15·48 has diverged in the yeast enzyme, and the site (shaded circle) that interacts with the anticodon has partially diverged in the human enzyme. The evolution of the C-terminal domain of the yeast and human enzymes might include recruitment of determinants that reject noncognate tRNAs as a parallel mechanism to improve the specificity of aminoacylation.

The elucidation of the evolutionary relationship of three synthetase domains provides a new insight into their relative functions. It is reasonable to assume that the most conserved domain would play the most critical role in aminoacylation while the least conserved domain would play a lesser role. The time line in Fig. 3 is consistent with our biochemical studies that establish U73 as the most important nucleotide for aminoacylation of the *E. coli* tRNA by the *E. coli* enzyme. The significance of U73 is followed by the significance of the GCA anticodon, which is then followed by the significance of the 15·48 base pair. Thus, the functional significance of tRNA determinants supports the evolutionary relationship of synthetase domains that "read" them. These results reinforce the notion that the domain for interaction with U73 would have been the primordial synthetase, which was then combined with the domain for interaction with the anticodon. The joining of these two domains was then adjusted and determined by a third domain that senses the structural effect of the 15·48 base pair on the tRNA.

Redistribution of Synthetase Domains and Refinement of the Model. The model in Fig. 3 can be refined further to reflect the evolution of synthetase domains from *E. coli* to yeast to human. Table 1 shows that while the *E. coli* enzyme discriminates against the yeast and human tRNAs, the yeast and human enzymes can efficiently aminoacylate the *E. coli* tRNA, as well as the yCys13 and hCys15 mutants that have incorporated structural features of G15·G48 of the *E. coli* tRNA. The relative k_{cat}/K_m values for the yeast enzyme with the *E. coli* tRNA and with yCys13 and hCys15 are within 4-fold of that of the yeast tRNA. Similarly, the relative k_{cat}/K_m values for the human enzyme with the *E. coli* tRNA and with yCys13 and hCys15 are within 5-fold of that of the human tRNA. These values are small compared with the two orders of magnitude of decrease in k_{cat}/K_m for the *E. coli* enzyme with respect to the yeast and human tRNAs. Thus, to a first approximation, we suggest that while the subdomain that senses the structural effect of the 15·48 base pair in the *E. coli* enzyme is specific for G15·G48, those in the yeast and human enzymes appear to

have diminished specificity such that they can accommodate both G15·C48 of their tRNAs and G15·G48 of the *E. coli* tRNA. The diminished specificity of the domain for the 15·48 base pair in the eukaryotic enzymes suggests two aspects that can refine the model in Fig. 3. First, it suggests redistribution of functional significance from the three-domain structure of the *E. coli* enzyme to the two-domain structure of the eukaryotic enzymes. This redistribution highlights the acceptor stem and the anticodon as the two major points of synthetase-tRNA interactions. Second, the diminished specificity for the 15·48 base pair in the eukaryotic enzymes could arise from the presence of new determinants that are recruited to the eukaryotic tRNAs during the course of evolution. For example, some of these new determinants may function to prevent noncognate synthetase-tRNA interactions. Because the eukaryotic $tRNA^{Cys}$ molecules have G15·C48, which is common to many other tRNAs, the recruitment of negative determinants to reduce the emphasis on G15·C48 is plausible. Thus, the evolution of a given tRNA-synthetase interaction may be refined and the specificity may be improved in the context of coevolution of other pairs of tRNAs and synthetases.

The emphasis on the 15·48 base pair and on the structural features at the corner of the tRNA L for the *E. coli* enzyme may be a general feature among eubacterial and some archaeal cysteine enzymes. Although only *E. coli* and *H. influenzae* cysteine tRNAs contain G15·G48, and although all other eubacterial and archaeal cysteine tRNAs have G15·C48, further examination of details of tRNA sequences indicates that all of these tRNAs appear to have a common structure at the corner of the tRNA L, which is expected to be distinct from that of the eukaryotic cysteine tRNAs. Specifically, all eubacterial and archaeal cysteine tRNAs have a mismatch at 13·22 in the D stem, whereas all eukaryotic cysteine tRNAs have a Watson–Crick base pair at 13·22. As shown in present studies and in previous results, the mismatch of 13·22 is critical for the proper presentation of the 15·48 base pair to the *E. coli* enzyme, and conversion of the mismatch to a base pair disrupts the presentation (ref. 31; C. S. Hamann and Y.-M.H., unpublished work). This finding raises the possibility that the 13·22 mismatch common to all eubacterial and archaeal cysteine tRNAs is the underlying basis for the structural features of 15·48 and that the 13·22 mismatch may confer cross-recognition among eubacterial and archaeal cysteine enzymes. This possibility is supported by the ability of the *E. coli* enzyme to crosscharge *Mycobacterium tuberculosis* $tRNA^{Cys}$ and *Halobacterium volcanii* $tRNA^{Cys}$, both of which contain G15·C48 and an unpaired 13·22. The possibility is also supported by the ability of *H. volcanii* cysteine enzyme to crosscharge *E. coli* $tRNA^{Cys}$ (Y.-M.H. and K. Wan, unpublished work). However, the *H. volcanii* cysteine enzyme does not crosscharge human $tRNA^{Cys}$, indicating a sensitivity of the *H. volcanii* cysteine enzyme to the structure at the corner of the tRNA L reminiscent of the sensitivity of the *E. coli* enzyme. Although further experiments are necessary, these results rationalize the *E. coli* enzyme as an example of the eubacterial and some archaeal cysteine-tRNA synthetases.

The *E. coli*, yeast, and human cysteine enzymes are representatives of two of the three kingdoms of modern living organisms, the eubacteria and eukarya. In the third kingdom, the archaea, the cysteine enzyme has been identified in the genomic database of some (such as *Archaeoglobus fulgidus*) but not in others (such as *Methanococcus jannaschii* or *Methanobacterium thermoautotrophicum*) (33–35). The evolutionary relationship of the domain structure of the *A. fulgidus* cysteine enzyme can be established by studying its ability to crosscharge *E. coli*, yeast, and human tRNAs, as we describe here for the *E. coli* enzyme. However, the absence of an identifiable cysteine enzyme from *M. jannaschii* or *M. thermoautotrophicum* raises the possibility of an entirely different lineage for aminoacylation of $tRNA^{Cys}$ in these organisms. How this lineage was developed and how it is related to that developed from *E. coli* to yeast to human will remain important questions in the pursuit of understanding the origin and evolution of life.

We thank Vince Stanton and Martin Zillmann of Variagenics for providing the clone of the human tRNA gene and the cDNA clone of the C-terminal truncated human cysteine-tRNA synthetase. We thank Howard Gamper for discussion, Charlie Williams for confirming the sequence of human $tRNA^{Cys}$, and Christian Hamann for critically reading the manuscript. This work was supported in part by National Institutes of Health Grant GM47935 (to Y.-M.H.), a grant from Thomas Jefferson University, and a grant from the Lucille P. Markey Charitable Trust.

1. Moras, D. (1992) *Trends Biochem. Sci.* **17,** 159–164.
2. Carter, C. W., Jr. (1993) *Annu. Rev. Biochem.* **62,** 715–748.
3. Schimmel, P. & Ribas de Pouplana, L. (1995) *Cell* **81,** 983–986.
4. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203- 206.
5. Cusack, S., Hartlein, M. & Leberman, R. (1991) *Nucleic Acids Res.* **19,** 3489–3498.
6. Nagel, G. M. & Doolittle, R. F. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8121–8125.
7. Hipps, D., Shiba, K., Henderson, B. & Schimmel, P. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 5550–5552.
8. Nair, S., Ribas de Pouplana, L., Houman, F., Avruch, A., Shen, X. & Schimmel, P. (1997) *J. Mol. Biol.* **269,** 1–9.
9. Quinn, C. L., Tao, N. & Schimmel, P. (1995) *Biochemistry* **34,** 12489–12495.
10. Shiba, K., Schimmel, P., Motegi, H. & Noda, T. (1994) *J. Biol. Chem.* **269,** 30049–30055.
11. Shiba, K., Suzuki, N., Shigesada, K., Namba, Y., Schimmel, P. & Noda, T. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 7435–7439.
12. Wakasugi, K., Quinn, C. L., Tao, N. & Schimmel, P. (1998) *EMBO J.* **17,** 297–305.
13. Hou, Y.-M., Sterner, T. & Bhalla, R. (1995) *RNA* **1,** 707–713.
14. Hou, Y.-M., Shiba, K., Mottes, C. & Schimmel, P. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 976–980.
15. Eriani, G., Dirheimer, G. & Gangloff, J. (1991) *Nucleic Acids Res.* **19,** 265–269.
16. Risler, J. L., Zelwer, C. & Brunnie, S. (1981) *Nature (London)* **292,** 384–386.
17. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246,** 1135–1142.
18. Brick, P., Bhat, T. N. & Blow, D. M. (1989) *J. Mol. Biol.* **208,** 83–98.
19. Doublie, S., Bricogne, G., Gilmore, C. & Carter, C. W. Jr. (1995) *Structure* **3,** 17–31.
20. Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S.-i., Kigawa, T., Miyazawa, T., Yokoyama, S. & Morikawa, K. (1995) *Science* **267,** 1958–1965.
21. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P. & Yokoyama, S. (1998) *Science* **280,** 578–582.
22. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8763–8768.
23. Hou, Y.-M., Weshof, E. & Giegé, R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6776–6780.
24. Hamann, C. S. & Hou, Y.-M. (1997) *Biochemistry* **36,** 7967–7972.
25. Hamann, C. S. & Hou, Y.-M. (1997) *Bioorg. Med. Chem.* **5,** 1011–1019.
26. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998) *Nucleic Acids Res.* **26,** 148–153.
27. Levitt, M. (1969) *Nature (London)* **224,** 759–763.
28. Cruzen, M. E. & Arfin, S. M. (1994) *DNA Sequence* **4,** 243–248.
29. Hamann, C. S. & Hou, Y.-M. (1995) *Biochemistry* **34,** 6527–6532.
30. Komatsoulis, G. A. & Abelson, J. (1993) *Biochemistry* **32,** 7435–7444.
31. Hou, Y.-M. (1994) *Biochemistry* **33,** 4677–4681.
32. Martinis, S. A. & Schimmel, P. (1995) in *tRNA: Structure, Biosynthesis and Function*, eds. Söll, D. and RajBhandary, U. (Am. Soc. Microbiol., Washington, DC), pp. 349–370.
33. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelsen, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al.* (1997) *Nature (London)* **390,** 364–379.
34. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) *Science* **273,** 1058–1073.
35. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) *J. Bacteriol.* **179,** 7135–7155.