J Mol Evol (1995) 40:519–530

JOURNAL OF MOLECULAR EVOLUTION

Structure, Function and Evolution of Seryl-tRNA Synthetases: Implications for the Evolution of Aminoacyl-tRNA Synthetases and the Genetic Code

Michael Härtlein, Stephen Cusack

European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL, 156X, 38042 Grenoble Cedex, France

Received: 20 October 1994

Abstract. Two aspects of the evolution of aminoacyl-tRNA synthetases are discussed. Firstly, using recent crystal structure information on seryl-tRNA synthetase and its substrate complexes, the coevolution of the mode of recognition between seryl-tRNA synthetase and $tRNA^{ser}$ in different organisms is reviewed. Secondly, using sequence alignments and phylogenetic trees, the early evolution of class 2 aminoacyl-tRNA synthetases is traced. Arguments are presented to suggest that synthetases are not the oldest of protein enzymes, but survived as RNA enzymes during the early period of the evolution of protein catalysts. In this view, the relatedness of the current synthetases, as evidenced by the division into two classes with their associated subclasses, reflects the replacement of RNA synthetases by protein synthetases. This process would have been triggered by the acquisition of tRNA 3′ end charging activity by early proteins capable of activating small molecules (e.g., amino acids) with ATP. If these arguments are correct, the genetic code was essentially frozen before the protein synthetases that we know today came into existence.

Key words: Aminoacyl-tRNA synthetases — tRNA — Genetic code — RNA world — Evolution

Introduction

The fidelity of modern protein biosynthesis rests on two key interactions, the specific attachment of amino acids to their cognate tRNA species and the accurate codon-anticodon interaction between mRNA and tRNA. The former process is catalyzed by the aminoacyl-tRNA synthetases which discriminate with remarkable selectivity amongst many structurally similar tRNAs and amino acids. The accuracy of the aminoacylation reaction is based on the existence of positive and negative elements in both tRNAs (''identity'' elements) and synthetases which permit both recognition and productive binding of cognate pairs and discrimination against or nonproductive binding of noncognate pairs. The latter is particularly important in the in vivo context where synthetases compete for uncharged tRNAs with sometimes overlapping identity sets. In spite of their common catalytic function, the synthetases have long been known to be diverse in subunit structure, polypeptide size, and amino acid sequence (Table 1). Quaternary structures include α, α_2, α_4, and $\alpha_2\beta_2$, and subunit polypeptide sizes range in *E. coli* from 334 residues (tryptophanyl-tRNA synthetase) to 951 (valyl-tRNA synthetase). Synthetases in higher eukaryotes are generally larger than their prokaryotic counterparts due to additional sequences dispensible for aminoacylation (Mirande et al. 1993). Amino acid sequence analysis (Eriani et al. 1990) and X-ray crystallography of seryl-tRNA synthetase (Cusack et al. 1990) have shown that the 20 aminoacyl-tRNA synthetases can be partitioned into two distinct structural classes of ten members each (Table 2). Each class is characterized by different short sequence motifs and a distinctive topology of the catalytic domain (Table 2). In addition there is

Based on a presentation made at a workshop—''Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code''—held at Berkeley, CA, July 17–20, 1994

Correspondence to: S. Cusack

Table 1. Size and subunit composition of aminoacyl-tRNA synthetases[a]

Enzyme	Prokaryotes		Lower eukaryotes		Higher eukaryotes		
Cys	α	461	—	—	—	—	
Glu	α	471	—	—	α	1,714	CX +PRO
Gln	α	550	—	809	—	—	CX
Arg	α	577	α	—	α	661	CX
Leu	α	860	α	1,090	α	—	CX
Ile	α	939	α	1,073	α	1,266	CX
Val	α	951	α	1,058	α	1,265	CX +Ef-1H
Trp	α_2	334	α_2	—	α_2	475	
His	α_2	424	α_2	526	α_2	509	
Tyr	α_2	424	α_2	—	α_2	—	
Ser	α_2	430	α_2	462	α_2	505	
Asn	α_2	465	α_2	492	α_2	552	
Lys	α_2	505	α_2	591	α_2	597	CX
Pro	α_2	572	α_2	—	α_2	1,714	CX +GLU
Asp	α_2	590	α_2	557	α_2	501	CX
Thr	α_2	642	α_2	734	α_2	724	
Met	α_2	677	α	751	α_2	—	CX
Ala	α_4	876	α	—	α	967	
Gly	$\alpha_2\beta_2$	689 + 303	α_2	687	α_2	687	
Phe	$\alpha_2\beta_2$	795 + 327	$\alpha_2\beta_2$	595 + 503	$\alpha_2\beta_2$	—	
			α (mito)	474			

[a] Subunit composition and size of aminoacyl-tRNA synthetases from prokaryotes (*E. coli*), lower eukaroyotes (*S. cerevisiae*), and higher eukaryotes (various). The table is an updated version of that found in Mirande et al. (1993). For higher eukaryotes, CX indicates that the synthetase is part of the multisynthetase complex, the prolyl- and glutamyl-tRNA synthetases being fused into a single polypeptide chain (Mirande et al. 1993). The valyl-tRNA synthetase of higher eukaryotes forms a separate complex with the heavy form of elongation factor 1 (eEf-1H).

now a clearer appreciation of the modular domain structure of the synthetases, which allows the identification of more closely related subclasses and explains the apparent wide diversity of synthetase subunit size (Cusack et al. 1991; Cusack 1993; Delarue and Moras 1993). These developments have revived speculation about the position of the synthetases in the evolution of the protein synthesis machinery and their potential role in the origin and evolution of the genetic code.

In this article we would like to review what is known about seryl-tRNA synthetase structure and function and the coevolution of seryl-tRNA synthetase and $tRNA^{ser}$ in different organisms. We will then go further back in time and comment on what sequence alignments tell us about the evolution of aminoacyl-tRNA synthetases. Finally, possible implications for the evolution of the genetic code will be discussed.

The Structure of Bacterial Seryl-tRNA Synthetases

The crystal structures of seryl-tRNA synthetase from *E. coli* (SerRSEC) and *Thermus thermophilus* (SerRSTT) have both been determined at 2.5-Å resolution (Cusack et al. 1990; Fujinaga et al. 1993). The *T. thermophilus* enzyme comprises 421 residues per subunit with a primary sequence identity of 39% to the *E. coli* enzyme (430 residues per subunit) and the three-dimensional structures are very similar (Fujinaga et al. 1993). The

Table 2. The two classes of aminoacyl-tRNA synthetases and their characteristics

Class 1		Class 2	
Glutamic acid[a]	(α)	Lysine[a]	(α_2)
Glutamine[a,b]	(α)	Asparagine	(α_2)
Arginine	(α)	Aspartic acid[a,b]	(α_2)
Leucine[c]	(α)	Proline	(α_2)
Isoleucine	(α)	Threonine[c]	(α_2)
Valine	(α)	Serine[a,b]	(α_2)
Cysteine	(α)	Histidine[c]	(α_2)
Methionine[a]	(α_2, α)	Glycine	($\alpha_2\beta_2$, α_2)
Tyrosine[a]	(α_2)	Phenylalanine[a]	($\alpha_2\beta_2$, α)
Tryptophan[a]	(α_2)	Alanine	(α_4, α)
Motifs (ϕ = hydrophobic)			
'ϕ-h-ϕ-G-h'		(1) 'gϕxxϕxxPϕϕ'	
'k-m-s-K-s'		(2) 'f-R-x-e'-loop-'h/rxxxFxxx(d/e)'	
		(3) 'gϕgϕgϕ(d/e)Rϕϕϕϕϕ'	
Fold of catalytic domain			
Rossmann fold (Tyr, Met, Gln, Trp)[a]		Antiparallel fold (Ser, Asp, Phe, Lys)[a]	
Activity			
2′ OH initial aa site		3′ OH initial aa site (exc Phe)[a]	

[a] Crystal structure of the synthetase has been determined
[b] Crystal structure of the cognate complex has been determined
[c] Crystal structure determination is in progress

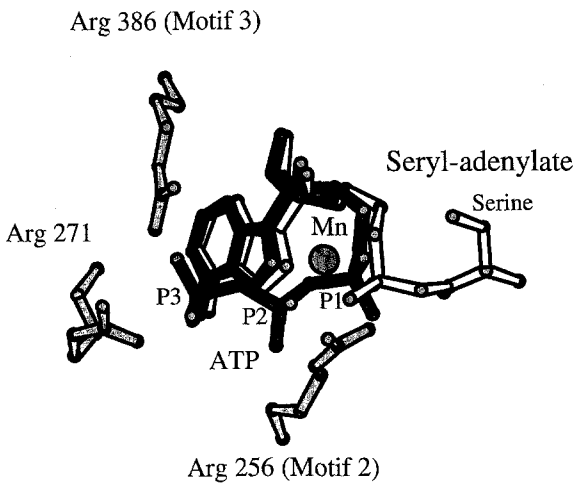

Fig. 1. Superposition of the conformations of ATP and Ser-AMP in the active site of SerRSTT. The α-, β-, and γ-phosphates of the ATP are labelled *P1, P2,* and *P3*. The position of the divalent cation Mn^{2+} is also shown together with the two class II conserved arginines, Arg-256 from motif 2 and Arg-386 from motif 3. Arg-271, another arginine which is, however, not strictly conserved, also interacts with the γ-phosphate of the ATP.

catalytic domain is based on a seven-stranded antiparallel β-sheet with two connecting helices. This antiparallel fold is characteristic of class II synthetases and has now been found in the aspartyl-tRNA synthetase (Ruff et al. 1991), phenylalanyl-tRNA synthetase (Mosyak et al. 1993), and lysyl-tRNA synthetase (Onesti et al. 1995). In addition seryl-tRNA synthetase has a remarkable additional N-terminal domain comprising a 60-Å-long solvent exposed, antiparallel coiled-coil (the "helical arm").

Structural Basis of the Activation of Serine by Seryl-tRNA Synthetase

The first step of the overall aminoacylation reaction catalyzed by seryl-tRNA synthetase is activation of serine by Mg^{2+}-ATP leading to the stable enzyme-bound intermediate, seryl-adenylate (Ser-AMP). A series of crystal structures at 2.3–3 Å resolution of binary and tertiary complexes involving seryl-tRNA synthetase, ATP, Mn^{2+} or Mg^{2+}, and seryl-adenylate (natural and analogues) have been determined, which provide the structural basis required to explain the specificity and mechanism of serine activation. The most extensive results are available for SerRSTT (Belrhali et al. 1994, 1995). Figure 1 shows a superposition of the conformations of ATP and Ser-AMP (enzymatically produced in the crystal) complexed with SerRSTT both in the presence of Mn^{2+}. These small substrates are bound by a network of hydrogen bond interactions in a deep hydrophilic cleft formed by the antiparallel β-sheet and surrounding loops of the synthetase catalytic domain. Four regions in the enzyme primary sequence are involved in the interactions including the motif-2 and -3 regions of class II synthetases. Apart from the specific recognition of the amino acid side chain, sequence alignments suggest that the interactions of the ATP and the adenylate are likely to be similar in all class II synthetases (Belrhali et al. 1994).

In the presence of a divalent cation (Mg^{2+} or Mn^{2+}) the ATP is found to be in an unusual 'bent' conformation (Fig. 1). The major divalent cation site bridges the α and β phosphates and also has two ligands in the protein, Glu-345 and Ser-348 (residue numbers are for SerRSTT). In addition two other Mn^{2+} sites are found on either side of the β-γ bridging phosphates, but without protein ligands. The position of the major divalent cation site differs from that attributed to it in the case of aspartyl-tRNA synthetase, where it is claimed to be between P_β and P_γ (Caverelli et al. 1993). The γ-phosphate interacts with Arg-386 in motif 3 and the α-phosphate with Arg-256 in motif 2, both absolutely conserved in all class II synthetases. Previous reports (Biou et al. 1994) that the ATP has an extended conformation can now be explained by the absence of Mg^{2+} binding in the crystallization medium which contains high concentrations of ammonium sulphate. If the ammonium sulphate is replaced by sodium citrate, the bent conformation of ATP is also observed in the presence of Mg^{2+}. ATP specificity is guaranteed by hydrogen bond interactions with three of the purine ring nitrogens. The Ser-AMP molecule is found in a very similar conformation to that already reported for two seryl-adenylate analogues (Belrhali et al. 1994). Serine specificity is ensured firstly by the interaction of the side-chain hydroxyl group with Thr-380 (in motif 3) of the enzyme and secondly by the small size of the side-chain pocket. The superposition of the enzyme-bound ATP and Ser-AMP structures provides strong support for an in-line displacement mechanism for serine activation (Fig. 1). The pentavalent transition state is presumably stabilized by the conserved motif-2 arginine (Arg-256 in SerRSTT) and possibly also by the divalent cation. These findings will be described in more detail elsewhere, together with the structural evidence for a similar mechanism for the synthesis of Ap_4A by the same enzyme (Belrhali et al. 1995).

The Structure of the SerRS-$tRNA^{ser}$ Complex

The crystal structure of the complex between seryl-tRNA synthetase and $tRNA^{ser}$(CGA) from *E. coli* has been determined at 4 Å resolution (Price et al. 1993) and between seryl-tRNA synthetase and $tRNA^{ser}$(GCA) from *T. thermophilus* at 2.9 Å resolution (Biou et al. 1994). Interestingly, the stoichiometry of the *T. thermophilus* complex is 1 tRNA:1 synthetase dimer (and the tRNA is partially disordered) although it is known from both solution studies and from the *E. coli* complex crystal structure that the synthetase can simultaneously bind two

tRNA molecules. The main conclusions from these crystal structures can be summarized as follows: (1) the tRNA binds across the two subunits of the dimer; (2) upon tRNA binding the helical arm of the synthetase is stabilized in a new orientation and curves between the TΨC loop and the long variable arm of the tRNA; (3) the synthetase makes several backbone contacts but few base-specific interactions; (4) contacts with the tRNA long variable-arm backbone extend until the sixth base-pair, explaining the need for a minimum length of the arm, but allowing longer arms (as in $tRNA^{selcys}$) to be accommodated; (5) bases 20A and 20B inserted into the D-loop in prokaryotic $tRNA^{ser}$ both play novel roles in tertiary interactions in the core of the tRNA. In particular the base of Gua-20B is stacked against the first base-pair of the long variable arm and thus defines the direction of the latter; and (6) disorder of the 3′ end of the tRNA in this crystal form prevents detailed description of the interactions of this part of the tRNA with the active site. (See Biou et al. 1994.)

Implication of the N-Terminal Domain of Seryl-tRNA Synthetase in Aminoacylation Activity and Specificity

To test the exact role of the synthetase helical arm in aminoacylation, two mutants of the *E. coli* synthetase have been constructed with a partial or full deletion of the helical arm (Borel et al. 1994). These mutants are not affected in the serine activation function, but have dramatically reduced aminoacylation activity. The kcat/Km value for the mutant enzyme with a full-arm deletion is reduced by more than 4 orders of magnitude, with a nearly 30-fold increased Km value for $tRNA^{ser}$. The only slightly truncated mutant form (16 amino acids of the tip of the arm replaced by a glycine) has an intermediate aminoacylation activity. The mutant synthetases have lost their specificity for $tRNA^{ser}$ and charge also noncognate type 1 tRNA(s). These results support the hypothesis that class II synthetases have evolved from an ancestral catalytic core enzyme by adding noncatalytic N-terminal or C-terminal tRNA binding (specificity) domains which act as determinants for cognate and antideterminants for noncognate tRNAs.

Cross-Dimer Binding During Aminoacylation

Deletion of the N-terminal part of *E. coli* SerRS does not affect the amino acid activation step of the reaction, but reduces aminoacylation activity by more than 3 orders of magnitude. (See above.) Heterodimers formed from two aminoacylation defective mutants, this N-terminal deletion and an active site mutant, restore charging activity: Aminoacylation activity in a mixture containing the heterodimers was compared to solutions containing the same concentrations of homodimers. It was eight times higher than the activities of the homodimer solutions and reached 50% of the theoretical value one would expect if 50% of the mixture would have been in the heterodimer form and supposing that a heterodimer contains one active site. Details of the complementation experiment will be published elsewhere. These results are in full agreement with the crystal structures of SerRS complexed with its cognate tRNA and provide functional evidence for tRNA cross-dimer binding originally suggested by Asahara et al. (1991).

Seryl-tRNA Synthetases in Prokaryotes and Eukaryotes

Complete primary structures of several bacterial seryl-tRNA synthetases (*Escherichia coli, Coxiella burnetti, Bacillus subtilis,* and *Thermus thermophilus*) are known as well as for a lower eukaryote *S. cerevisiae.* For higher eukaryotes, partial but overlapping sequences from mouse, Chinese hamster, and human are available and can be assembled into a composite mammalian sequence (Table 3). Sequence identity scores were obtained after pairwise alignment of the sequences using the program BESTFIT (Smith and Waterman) with standard gap weight (3) and gap length weight (0.1). Compared to the *E. coli* enzyme, the highest degree of identity (61%) is with *C. burnetti* (a gram-variable obligatory intracellular bacterium), followed by the gram-positive *B. subtilis* (52%). The identity falls to 39% when compared with *T. thermophilus* possibly due to amino acid replacements which maintain enzyme activity and stability at high temperature. When the *E. coli* enzyme was compared with the lower (yeast) and higher (composite mammalian) eukaryote SerRS sequences identity scores of 31% and 37% respectively are obtained. The identity score for the comparison of the two eukaryotic enzymes is 50%. From immunochemical noncross reactivity between the eukaryotic and prokaryotic seryl-tRNA synthetases, Sidorik et al. (1991) proposed that prokaryotic and eukaryotic SerRS evolve from different ancestor genes. Due to the relatively high sequence identities (up to 37%), convergent evolution seems very unlikely. An alternative explanation for the absence of immunochemical noncross reactivity might be the fact that the less-conserved N-terminal (tRNA binding) domain is probably the most antigenic. Evidence for this comes from our observation that antiserum raised against the native *E. coli* enzyme only very inefficiently recognizes a N-terminal deletion mutant (catalytic domain mutant). A multiple sequence alignment (Table 3) and a similarity plot (Fig. 2) of aligned SerRS sequences also demonstrate that the most variability occurs in the N-terminal helical arm domain, which is involved in non-sequence-specific recognition of the long variable arm of $tRNA^{ser}$ (Biou et al. 1994). This can be interpreted as an example

Table 3. Multiple sequence alignment of seryl-tRNA synthetase (SRS) sequences from prokaryotes *Escherichia coli* (SRSEC), *Coxiella burnetii* (SRSCB), *Thermus thermophilus* (SRSTT), and *Bacillus subtilis* (SRSBS), a lower eukaryote *S. cerevisiae* (SRSSC), and a higher eukaryote (SRSHE) (mammalian assembled, see text) using the program PILEUP[a]

```
Srsec   1 .MLDPNLLR. ...NEPDAVA EKLARRGFKL .DVDKLGALE ERRKVLQVKT  ENLQAERNSR SKSIGQAKA RGEDIEPLRL EVNKL..... ..........
Srscb   1 .MLDPKILR. ...QNLEHVV EKLRRRGFEM .DSDTFLQLE NKRKEAQLAI  QSFQTKRNQL SKTIGMAKS KGENPELLMA EVSQL..... ..........
Srstt   1 .MVDLKRLR. ...QEPEVFH RAIREKGVAL .DLEALLALD REVQELKKRL  QEVQTERNQV AKRVP..KA PPEEKEALIA RGKAL..... ..........
Srsbs   1 .MLDTKMLR. ...ANFQEIK AKLVHKGEDL TDFDKFEALD DRRRELIGKV  EELKGKRNEV SQQVAVLKR EKKDADHIIK EMREV..... ..........
Srssc   1 .MLDINQFIE DKGGNPELIR QSQKARNASV EIVDEIISDY KDWVKTRFEL  DELNKKFNKL QKDIGLKFK NKEDASGLLA EKE..... .........K
Srshe   1 MVLDLDLFRV DKGGDPALIR ETQEKRFKDP GLVDQLVKAD SEWRRCRFRA  DNLNKLKNLC SKTIGEKMK KKEAVGDDES VPENVLNFDD LTADALRALK

Srsec  79 .......... .GEELDAAKA ELDALQAEIR DIALTIPNLP ADEVPVGKDE  NDNVEVSR.WG TPREFDFEVR .......... DHVTLGEMHS GLDFAAAVKL
Srscb  79  ......... .NDELKQEEA NFETIQKAFS DFQLAIPNLP HDSVPDGKSE  NDNREIRQ.WG APPGFDFTPK .......... DHTVLGERDN QLDFEAAAKL
Srstt  77  ......... .GEEAKRLEE ALREKEARLE ALLLQVPLPP WPGAPVG.GE  EANREIKR.VG GPPEFSFPPL .......... DHVALMEKNG WWE.PRISQV
Srsbs  80  ......... .GEEIKKLDE ELRTVEAELD TILLSIPNIP HESVPVGETE  DDNVEVRK.WG EKPSFAYEPK .......... PHWDIADELG ILDFERAAKV
Srssc  83 LTQQKK.... ..ELTEKEQQ EDKDLKKKVF ....QVGNIV HPSVVVSNDE  ENNELVRT.WK PEDLEAVGPI ASVTGKPASL SHHEILLRLD GYDPDRGVKI
Srshe 100 VSQIKKVRLL IDEAIQKCDG ERVKLEAWRF ENLREIGNLL HPSVPISNDE  DADNKVERIWG DCTVRKKY.. .......... SHVDLVVMVV GFEGEKGAVV

Srsec 158 TGSRFVVMKG QIARMHRALS QFMLDLHTEQ HGYSENYVPY LVNQDTLYGT  GQLPKFAGDL FHTRPLEEEA DT....SNYA LIPTAEVPLT NLVRGEIID.
Srscb 158 SGARFVVLRG SLARAHRALA QFMLDLHTDQ HGYEEVYVPY LVHEECLYGT  GQLPKFREEQ FQV......A GD....RNFF LVPTGEVPLV NLARDEIIE.
Srstt 154 SGSRSYALKG DLALYELALL RFAMDF.MAR RGFLPMTLPS YAREKAFLGT  GHFPAYRDQV WAI....... AE....TDLY LTGTAEVVLN ALHSGEILP.
Srsbs 159 TGSRFVFYKG LGARLERALY NFMLDLHVDE YNYTEVIPPY MVNRASMTGT  GQLPKFEEDA FKIR...... .E....EDYF LIPTAEVPIT NMHRDEILS.
Srssc 173 CGHRGYFFRN YGVFLNQALI NYGLQF.LAA KGYIPLQAPV MMNKELMSKT  AQPSEFDEEL YKVIDGE... ......DEKY LIATSEQPIS AYHSGEWFEK
Srshe 189 AGSRGYFLKG VLVFLEQALI QYALRT.LGS RGYIPIYTPF FMRKEVMQEV  AQLSQFNEEL YKVIGKGSEK SDDNSYDEKY LIATSEQPIA ALHRDEWL..

Srsec 253 .EDDLPIKMT AHTPCFRSEA GSYGRDTRGL IRMHQFDKVE MVQIVRP..E  DSMAALEEMT GHAEKVLQLL GLPYRKIILC TGDMGFGACK TYDLEVWIPA
Srscb 247 .APALPKKWV AQTPCFRSEA GSYGKDVRGM IRQHQFQKVE LVQLVQP..E  NSYQALEEIT RQAEKVLQLL ALPYRVVELC AGDLGFAAAK TYDLEVWLPS
Srstt 241 .YEALPLRYA GYAPAFRSEA GSFGKDVRGL MRVHQFHKVE QYVLTEASLE  ASDRAFQELL ENAEEILRLL ELPYRLVEVA TGDMGPGKWR QVDIEVYLPS
Srsbs 247 .GDSLPINYA AFSACFRSEA GSAGRDTRGL IRQHQFNKVE LVKFVKP..E  DSYEELEKLT NQAERVLQLL ELPYRVMSMC TGDLGFTAAK KYDIEVWIPS
Srssc 263 PQEQLPIHYV GYSSCFRREA GSHGKDAWGV FRVHAFEKIE QFVITEP..E  KSWEEFEKMI SYSEEFYKSL KLPYRIVGIV SGELNNAAAK KYDLEAWFPY
Srshe 286 RPKDLPIKYA GLSTCFRSEV GSHGRDTRGI FRVHQFEKIE QFVYSSPHDN  KSWEMFEEMI TTAEEFYQSL GIPYHIVNIV SGSLNHAASK KLDLEAWFPG

Srsec 350 QNTYREISSC SNVWDFQARR MQARC..RSK SDKKTRLVHT LNGSGLAVGR  TLVAVMENYQ QADGRIEVPE VLRPYM..NG LEYIG..... ..........
Srscb 344 QNKYREISSC SNCEDFQARR IQARW..RNP KTGKPELLHT LNGSGLAVGR  TLVAVMENYQ QADGHIRVPD ALKSYM..GG VD..... ..........
Srstt 340 EGRYRETHSC SALLDWQARR ANLRY..R.. .DPEGRVRYT LNNTALATPR  ILAMLLENHQ LQDGRVRVPQ ALIPYM...G KEVLEPCG.. ..........
Srsbs 344 QDTYREISSC SNFEAFQARR ANIRF..RRE AKGKPEHVHT LNGSGLAVGR  TVAAILENYQ QEDGSVVIPK VLRPYM..GN REVMKP.... ..........
Srssc 361 QKEYKELVSC SNCTDYQSRN LEIRCGIKKM GDREKKYVHC LNSTLAATQR  ALCCILENYQ TEDG.LVVPE VLRKYI.PGE PEFLPFVNEL PKNSTSSKDK
Srshe 386 SGAFRELVSC SNCTDYQARR LRIRYGQTKK MMDKVEFVHM LNATMCATTR  TICAILENYQ TEKG.IIVPE KLREFMPPGL QELIPFVKPA PIDQEPSKKQ

Srsec      .......... .......... .  430
Srscb      .......... .......... .  423
Srstt      .......... .......... .  423
Srsbs      .......... .......... .  425
Srssc      KKKN...... .......... .  462
Srshe      KKQHEGSKKE SKRGHPGEPA A  505
```

[a] The Genbank/EMBL accession numbers for the used SRS sequences are: P09156 (SESEC), X75627 (SRSCB), P34945 (SRSTT), gi|467403 (SRSBS). The following overlapping partial sequences were used to construct the assembled mammalian sequence P26638 from mouse, TO3561 from human, and P26636 from Chinese hamster. PILEUP creates a multiple sequence alignment using a simplification of the progressive alignment method of Feng and Doolittle (1987)

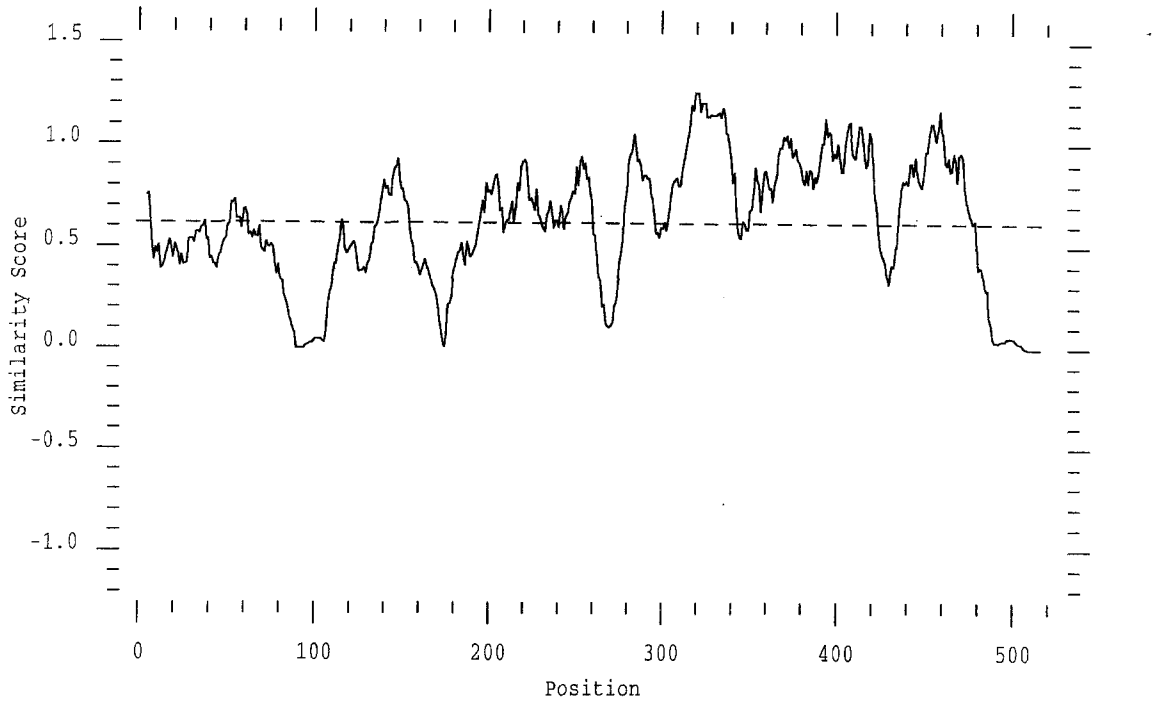

Fig. 2. Plot showing the similarity between the aligned SerRS sequences. The program PLOTSIMILARITY was used to calculate the average similarity among all aligned sequences at each position in the alignment, using a window of ten residues. The average similarity across the entire alignment is plotted as a *dotted line.*

of ongoing coevolution of isoacceptor tRNAs and tRNA binding domains in different organisms.

$tRNA^{ser}$ Structure and Identity in Evolution

$tRNA^{ser}$ from E. coli

Escherichia coli possesses five isoaccepting $tRNA^{ser}$s (with four distinct anticodons) in order to cope with the six codons for serine which are from two distinct codon groups. In addition the $tRNA^{sec}$ (an opal suppressor tRNA) necessary for the cotranslational incorporation of selenocysteine into proteins and an amber suppressor tRNA (*supD* gene product) are also specifically aminoacylated by SerRS. As a result there is no consistency in the anticodon bases. Not unsurprisingly therefore, it turns out that the anticodon is not recognized by seryl-tRNA synthetase, as was originally shown by Normanly et al. (1986, 1992). Amongst the 20 systems, this exceptional feature is shared only by leucyl- and alanyl-tRNA synthetase (Schulman 1991; Saks et al. 1994). The in vivo identity switch experiments of Normanly et al. (1992) have shown the importance for serine identity of the discriminator base and bases from the first three pairs of the acceptor stem as well as the D-stem base pair C11-G24. The latter, which is not directly in contact with the synthetase (Biou et al. 1994), seems to be a negative determinant to avoid mischarging with leucine and glutamate. Another special feature of serine iso-accepting tRNAs is that with the exception of those from animal mitochondria they possess a long variable arm of 14–20 nucleotides rather than the usual four or five. This feature is shared in prokaryotes by $tRNA^{tyr}$ and $tRNA^{leu}$ isoacceptors. Experiments by Himeno et al. (1990) and Normanly et al. (1992) have highlighted the importance of the long variable arm as an essential recognition element and this is also clearly apparent from the crystal structure of the complex (Biou et al. 1994). A conserved feature in prokaryotic $tRNA^{ser}$ isoacceptors is the absence of unpaired nucleotides at the base of the variable arm stem. The $tRNA^{leu}$ isoacceptors have one unpaired nucleotide at the 3′ end of the variable stem and $tRNA^{tyr}$ have two. These differences, which are correlated with the number of insertions in the D-loop (Table 4), may influence the orientation of the variable arm; the role of D-loop insertion G20B in orientating the long variable arm in $tRNA^{ser}$ has been clearly observed in the crystal structure of the SerRSTT complex (Biou et al. 1994). Himeno et al. (1990) have argued that these factors are crucial for the discrimination between long variable arm (type 2) tRNAs and have demonstrated this for $tRNA^{tyr}$ and $tRNA^{ser}$ (Himeno et al. 1990) and $tRNA^{leu}$ (Asahara et al. 1993). The contribution of discrete $tRNA^{ser}$ domains to aminoacylation by SerRS was studied by Sampson and Saks (1993). The long variable arm is the domain that makes the largest contribution to kcat/Km of aminoacylation. The acceptor stem is the second most important domain for recognition by SerRS, whereas the

Table 4. Characteristics of some type 2 tRNAs[a]

	E. coli				Yeast	Human
	$tRNA^{ser}$	$tRNA^{sec}$	$tRNA^{leu}$	$tRNA^{tyr}$	$tRNA^{ser}$	$tRNA^{ser}$
Unpaired bases at base of variable arm	0	0	1	2	1	1
D-loop	−17	−17	+17	−17	−17	−17
	+20A	+20A	+20A	+20A	+20A	+20A
	+20B			+20B		

[a] Characteristics of long variable arm (type 2) tRNAs which may influence the orientation of the long variable arm

anticodon stem/loop does not significantly contribute to it. Asahara et al. (1994) have studied intensively the role of the acceptor stem and variable arm in tRNA recognition by the synthetase in vitro and conclude that seryl-tRNA synthetase selectively recognizes $tRNA^{ser}$ on the basis of its characteristic tertiary structure rather than on the basis of sequence specificity. A similar conclusion was reached from the analysis of the crystal structure of the SerRSTT-$tRNA^{ser}$ complex (Biou et al. 1994).

The *E. coli* SelC gene product ($tRNA^{sec}$) is a key molecule in the cotranslational incorporation of selenocysteine in certain enzymes (Böck et al. 1991), but is initially charged with serine by seryl-tRNA synthetase. It is a structurally untypical tRNA and with 95 nucleotides is the longest tRNA known to date (Baron et al. 1993). However its overall L-shape resembles that of canonical *E. coli* $tRNA^{ser}$ with comparable distances between the CCA end and the anticodon of the molecules, indicating that the unusual extra base-pair in the acceptor stem does not provoke a severe structural distorsion. The extra arm is composed of eight base-pairs (exclusively G · C and G · U), with no unpaired nucleotides at its base (Table 4). On the other hand, compared to canonical $tRNA^{ser}$, there is no D-loop insertion at position 20b, which plays a crucial role in determining the orientation of the variable arm in $tRNA^{ser}$. (See above.) This difference, combined with the eight-base-pair acceptor stem, may explain why $tRNA^{sec}$ has a 100-fold decrease in charging activity as compared to $tRNA^{ser}$ (Baron and Böck 1991), since the other $tRNA^{ser}$ identity elements, apart from C11:G24, are present.

$tRNA^{ser}$ from S. cerevisiae

There are significant differences between $tRNA^{ser}$ from yeast and from *E. coli,* notably in the acceptor stem sequences and the determinants of the orientation of the long variable arm (Table 4). These differences could explain the low cross-aminoacylation of yeast $tRNA^{ser}$ with *E. coli* SerRS. On the other hand, it has been demonstrated that yeast seryl-tRNA synthetase expressed in *E. coli* recognizes bacterial serine-specific tRNA in vivo (Weygand-Durasevic et al. 1993). A model for yeast $tRNA^{ser}$ based on chemical probing experiments has been proposed (Dock-Bregeon et al. 1989) and also the complex with yeast SerRS has been studied in solution (Dock-Bregeon et al. 1990); as in *E. coli* the long variable arm of the tRNA makes contact with the synthetase.

Higher Eukaryotic Cytoplasmic $tRNA^{ser}$

Animal, plant, and yeast serine isoaccepting tRNAs have the same features regarding the orientation of the long variable arm (Achsel and Gross 1993). The system most extensively studied is that of human. Achsel and Gross (1993) identified major and minor identity elements by an identity switch from human $tRNA^{val}$ to $tRNA^{ser}$ in vitro. The discriminator base (G) and the long extra arm are absolutely required for aminoacylation by human seryl-tRNA synthetase. This is in contrast to *E. coli* $tRNA^{ser}$ where the discriminator base is unimportant (at least in vitro), whereas identity determinants in the acceptor stem are required. The long extra arms of human $tRNA^{ser}$ and $tRNA^{sec}$ function as major identity elements in an orientation-dependent but not sequence-specific manner (Wu and Gross 1993). The exchange of the discriminator base A73 for G is alone sufficient to convert human $tRNA^{leu}$ into a serine acceptor in vitro (Breitschopf and Gross 1994). Human amber suppressor serine tRNA functions in vivo with the cognate enzyme from yeast, but not with that from *E. coli,* indicating that not all determinants specifying the $tRNA^{ser}$ identity are conserved in evolution (Weygand-Durasevic et al. 1994).

Animal Mitochondrial $tRNA^{ser}$

In general mitochondrial synthetases appear to be capable of charging eubacterial as well as mitochondrial tRNA, whereas their eubacterial counterparts do not charge efficiently cognate mitochondrial tRNAs (Kumazawa et al. 1991). This unilaterality in aminoacylation may imply that, in response to simplifications in the species number and structural elements of animal mitochondrial tRNAs, the mitochondrial synthetases have evolved simpler tRNA recognition mechanisms possibly reminis-

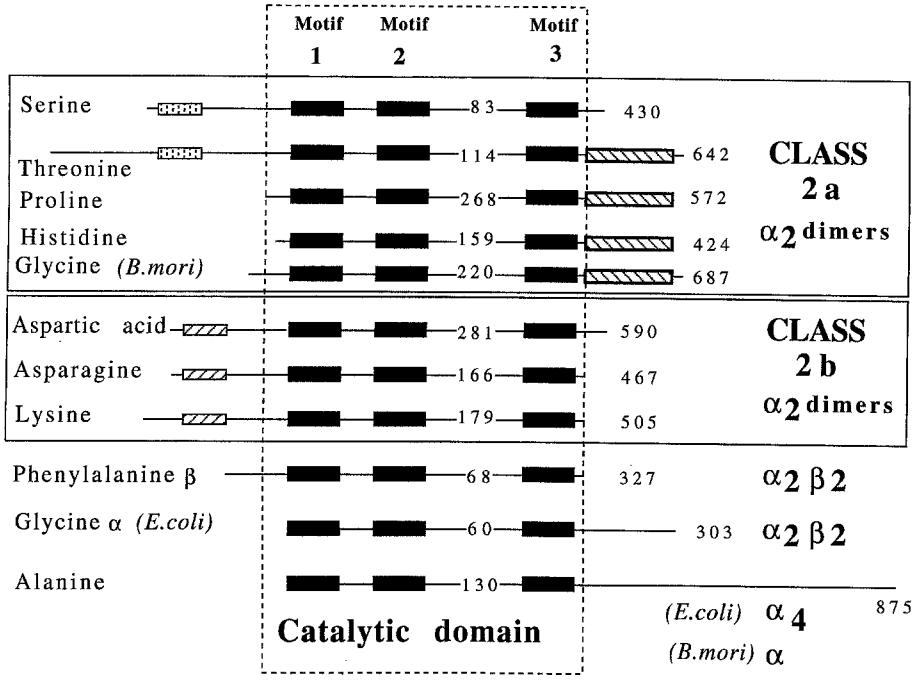

Fig. 3. Domain structure of class II aminoacyl-tRNA synthetases. Homologous domains attached to the N- or C-terminal of the catalytic domain have similar *shading*. The sequence *numbering* refers to the *E. coli* enzyme except where otherwise stated.

cent of ancestral forms. Interestingly, bovine mitochondrial seryl-tRNA synthetase not only charges cognate *E. coli* tRNA species, but also extensively misacylates noncognate *E. coli* tRNA species (Kumazawa et al. 1991). Unlike all other serine-specific tRNAs, only those from animal mitochondria do not possess a long variable arm, and in some cases (those for the AGY codons) the D-stem/loop is even missing (Steinberg et al. 1994). We have shown that in bacteria the long variable arm of the tRNA interacts with the N-terminal helical arm domain of the synthetase. If the mitochondrial seryl-tRNA synthetase has to recognize tRNAs without a long variable arm it is possible that it does not have a N-terminal helical arm and has developed other methods of specific recognition (e.g., for the mitochondrial $tRNA^{ser}$[AGY] it has been shown that the T-loop is the main recognition site for the mitochondrial seryl-tRNA synthetase; Ueda et al. 1992). This might explain the apparent promiscuity of this enzyme with *E. coli* tRNAs, since it is reminiscent of the situation with the N-terminal deletion mutants of *E. coli* SerRS. (See above and Borel et al. 1994.) The animal mitochondrial seryl-tRNA synthetases are clearly enzymes of great interest; the first indications of whether or not they have a helical arm might come when the sequence of such an enzyme is determined.

The Evolution of Aminoacyl-tRNA Synthetases

The accumulation of primary sequence and structural information on aminoacyl-tRNA synthetases over recent years puts us now in a position to try to draw up evolutionary trees of these enzymes for each of the two classes. This should clearly be done using structurally based sequence alignments as far as possible rather than relying on automatic "blind" methods of multiple sequence alignment which can lead to avoidable errors. Structurally based sequence alignments for the ten class II synthetases have been presented previously (Cusack et al. 1991; Cusack 1993; Belrhali et al. 1994) and will soon be published for class I synthetases (Risler et al. private communication). Here we would like to summarize where we are with regard to class II synthetases.

As described in Cusack et al. (1993), the three class II motifs can now be identified in all class II synthetases following recent results on the glycyl, alanyl, and phenylalanyl systems.[1] Figure 3 shows schematically the domain structure of class II aminoacyl-tRNA synthetases based on the multiple alignments. Identification of homologous domains N-terminal or C-terminal to the catalytic domain permit separation of the homodimeric class II synthetases into class IIa (the enzymes for serine,

[1] Examination of the sequence of glycyl-tRNA synthetase from human and a putative GlyRS sequence from yeast suggests a different position of motif 1 than that proposed on the basis of the *Bombyx mori* enzyme alone (Cusack 1993). In this new alignment, the normally conserved motif 1 proline is replaced by a serine or a threonine. Furthermore, the position of motif 1 in the large β-subunit of phenylalanyl-tRNA synthetase has also been modified from that shown in Cusack (1993) and Mosyak et al. (1993) in the light of further refinement of the crystal structure of PheRS from *T. thermophilus* (Safro and Mosyak private communication). Neither of these corrections changes the basic conclusions presented in Cusack (1993).

threonine, glycine [except *E. coli*], and histidine) and class IIb (the enzymes for aspartic acid, asparagine, and lysine) (Cusack et al. 1991; Delarue and Moras 1993; Cusack 1993). These putative tRNA binding domains are apparently conserved throughout evolution for a given system. In the case of aspartyl-tRNA synthetase, the N-terminal domain is involved in specific anticodon recognition (Cavarelli et al. 1993) and a similar domain topology is found in other oligonucleotide binding proteins (Murzin 1993). As expected a very similar fold has been found for the N-terminal domain in lysyl-tRNA synthetase (Onesti et al. personal communication). The C-terminal domain characteristic of class IIa has not yet been proved to be an anticodon binding domain but is a very strong candidate; its absence from seryl-tRNA synthetase is consistent with the fact that only this class IIa synthetase does *not* recognize the anticodon. The region between motifs 2 and 3 is also a site for the insertion of extra domains that are even more idiosyncratic and thus probably more recent acquisitions; the example of aspartyl-tRNA synthetase shows that this insertion domain is not conserved even between the yeast and bacterial enzymes (Delarue et al. 1994). Glycyl-tRNA synthetase from *E. coli,* and the phenylalanyl-tRNA synthetases (except that from yeast mitochondria, which is monomeric) have an $\alpha_2\beta_2$ composition, the α-subunits having the characteristic class II motifs (Table 1 and Fig. 3). The remarkable structure of the phenylalanyl-tRNA synthetase from *T. thermophilus* shows that both the α- and β-subunits contain class II antiparallel fold, although only that of the α-subunit is apparently catalytically active (Mosyak et al. 1993).

Clearly it is of great interest to use this growing wealth of structural and sequence information on aminoacyl-tRNA synthetases to try to reconstruct the evolution of these enzymes, despite the difficulty of defining a time axis in such problems (Benner et al. 1993). In the above paragraph the modular nature of the class II synthetases has been emphasized. Consideration of this leads to the almost inevitable conclusion that the common ancestor to these enzymes consisted of a class II catalytic domain on to which were subsequently added N- or C-terminal anticodon binding domains and more recently (since the evolution of different cell types), insertions between motif 2 and motif 3 (and elsewhere). If we consider only the extended sequences around the three motifs, that is, the only regions that can be aligned amongst *all* class II synthetases, we can calculate, using, for example, the program CLUSTALV (Higgins and Sharp 1988), an unrooted "evolutionary" tree (Fig. 4). This shows that even without putting any information in about the extra domains, the subclass divisions into class IIa, class IIb, and others (GlyRSEC, PheRS, AlaRS) is clearly reproduced.

This tree poses many questions. Amongst these are the different origin of the $\alpha_2\beta_2$ GlyRS from *E. coli* and the α_2 GlyRS from other organisms (Cusack 1993). Similarly, why did yeast mitochondrial PheRS simplify to a monomeric enzyme, or alternatively, why did the $\alpha_2\beta_2$ PheRS become more complex with the addition of a second catalytic-like domain (Mosyak et al. 1993)? Another question concerns the significance of the subclasses. The tree suggests that at some earlier time, there were significantly fewer synthetases. For instance, class IIb synthetases AspRS and LysRS have apparently diverged from a common ancestor by gene duplication and adaptation of the amino acid site and the tRNA specificity. (Note: Without loss of generality we do not consider AsnRS in the following argument.) What was the function of this common ancestor? Was it such an inaccurate synthetase that it could not tell the difference between aspartic acid and lysine and charged a tRNA with either of them? Or was it, on the contrary, such a good enzyme that it could act as a double synthetase with specific LysRS and AspRS activity? Both possibilities seem unlikely. Did the common ancestor have just one of the two activities then? This possibility begs the question of what performed the synthetase function for the other amino acid. Obviously the problem is compounded if one considers the common ancestor of all the ten class II synthetases. Very recently an interesting piece of information has emerged which may shed light on the nature of the common ancestor of class II synthetases. Whereas many enzymes with the same catalytic fold as class I aminoacyl-tRNA synthetases are known (kinases, dehydrogenases, etc.), until now, no other enzyme has been

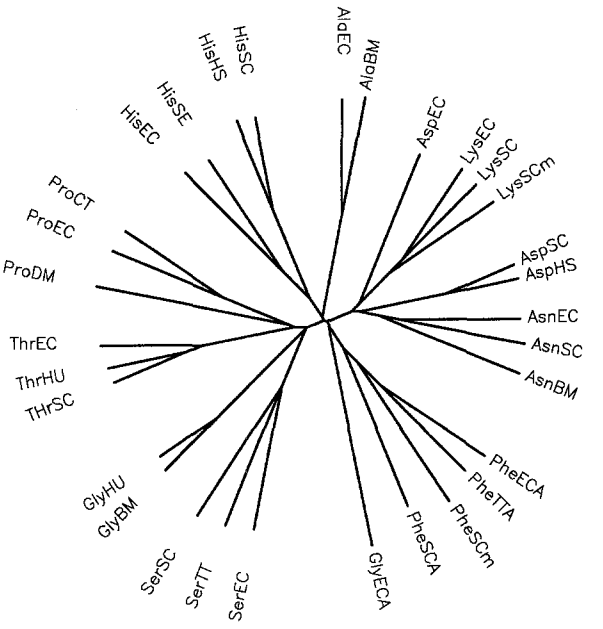

Fig. 4. Unrooted tree of class II aminoacyl-tRNA synthetases based on alignment of 140 residues from extended regions around the three class II motifs. Abbreviations used: *GlyECA:* α-subunit of glycyl-tRNA synthetase from *E. coli, TT: T. thermophilus SC(m): S. cerevisiae* (mitochondrial), *HU:* human, *DM: Drosophila melanogaster, BM: Bombyx mori, SE: Streptococcus equisimilis, CT: Chlamydia trachomatis.* Figure prepared using the program CLUSTALV (Higgins and Sharp 1988).

found with the antiparallel ATP binding fold characteristic of class II synthetases. However it has recently been shown (Artymiuk et al. 1994) that biotin synthetase (BirA) from *E. coli,* whose crystal structure is known (Wilson et al. 1992), has such a fold, although there is no obvious sequence homology. Biotin synthetase activates biotin with ATP to form an intermediate biotin-AMP. The same enzyme then transfers the biotin to lysines of certain target proteins (acetyl-CoA carboxylase); BirA thus has a close functional similarity to an aminoacyl-tRNA synthetase. Although convergent evolution cannot be ruled out, this finding permits one to hypothesize that BirA and class II synthetases had a common ancestor which was capable of activating small molecules with ATP. Class II synthetases would then have evolved when an activation domain, specialized to an amino acid, also became able to transfer the amino acid to the 3′ end of a tRNA molecule. This is clearly consistent with the view, based on sequence alignments (e.g., Cusack et al. 1993) and on experiments with mini-RNA helices (Schimmel et al. 1993), that early synthetases acquired tRNA binding specificity firstly through acceptor stem interactions with the activation domain (via the motif 2 loop in the case of class II synthetases) and later by addition of extra tRNA binding domains (notably anticodon recognition domains).

Implications for the Evolution of the Genetic Code

It has been long recognized that synthetases represent an evolutionary chicken-and-egg paradox (Hopfield 1978), since proteins cannot be made accurately without highly specific synthetases, but synthetases are themselves accurately made proteins. The arguments presented in the previous paragraphs suggest that the hypothesis that the first aminoacyl-tRNA synthetases were protein enzymes runs into great difficulties. The fact that class II synthetases may have evolved from an *existing* protein with ATP-mediated activation activity also suggests that synthetases were not necessarily amongst the first protein enzymes; the same argument can be made for class I synthetases.

Fortunately, the RNA world hypothesis (Gilbert 1986), provides a way out of this impasse, despite its difficulties (Joyce and Orgel 1993). A series of discoveries in the natural world as well as the "artificial" RNA world (i.e., that accessible by the in vitro selection of RNA molecules) go some way toward demonstrating the chemical plausibility of template directed polypeptide synthesis by RNA catalysts exclusively. These discoveries include (1) the demonstration that extensively deproteinized large ribosomal subunits retain peptidyl transferase activity, suggesting that the catalytic activity resides in the RNA component (Noller et al. 1992); (2) the demonstration of the aminoacyl-esterase activity of the *Tetrahymena* ribozyme (Piccirilli et al. 1992); (3) the in vitro selection of RNAs with specific ATP (Sassanfar and Szostak 1993) or polar (Yarus 1989) and nonpolar (Mayfield and Yarus 1994) amino acid binding sites; (4) the in vitro selection of an RNA capable of aminoacylating itself (Illangasekare et al. 1995). It is beyond the scope of this article to discuss how, why, and to what sophistication polypeptide synthesis may have evolved in the RNA world. See, for instance, Orgel 1989; Gibson and Lamond 1992.) However, in relation to the question as to whether the present-day protein synthetases have anything to tell us about the origin of the genetic code, it is important to ask what the genetic code looked like when this replacement process occurred. Had the 20-amino-acid genetic code already evolved at that time and functioned in a more-or-less rigorous way or did a simpler code with fewer amino acids exist? We are not yet in a position to answer these questions, although we will speculate briefly about how RNA synthetases might have been replaced by protein synthetases. The partition of modern synthetases into two mutually exclusive and structurally different classes might, if the modern genetic code had already evolved at the time, only reflect two different pathways in the replacement process. Using the arguments developed above, the key to the replacement process would have been the acquisition by existing amino acid activation domains (which must have been useful in their own right) of tRNA 3′ end binding. It is then relatively easy to imagine that (1) the development of an efficient aminoacylation enzyme would soon make the corresponding RNA synthetase redundant, and (2) depending on the amino acid specificity (e.g., polar or nonpolar), different activation domains might have existed, thus explaining the occurrence of the two classes of synthetase catalytic domain, and (3) gene duplication and evolution of the amino acid binding site and tRNA specificity would give rise to the observed evolutionary tree of protein synthetases. It cannot be excluded that a glutaminyl-tRNA synthetase did not exist in an RNA world since such an enzyme is not found in several organisms existing today.

If the evolutionary relationship between protein synthetases only reflects the process of replacing existing RNA synthetases and not the coevolution of the synthetases with the genetic code, then we have to look elsewhere for the origins of the genetic code. However, if we now return to the special case of serine, we can argue that *ongoing* evolution of the genetic code can depend on the nature of the synthetase. The fact that the anticodon is *not* recognized by seryl-tRNA synthetase gives some flexibility to the genetic code and allows limited codon rearrangements. An example is the cytoplasmic genetic code in *Candida cylindrea,* where a leucine codon (according to the universal genetic code) becomes a serine codon (Kawaguchi et al. 1989). Another example is the use of a very special opal suppressor serine-accepting tRNA ($tRNA^{sec}$) in the process of cotranslational incorporation of selenocysteine, which

has been called the 21st amino acid (Böck et al. 1991). Anticodon recognition, as occurs in most synthetase systems, on the other hand, leads to tRNA evolution by the mixing of tRNA isoacceptor pools without affecting the genetic code, as pointed out by Saks and Sampson (submitted). This type of tRNA recognition, in which the synthetase recognizes directly the amino acid and the anticodon (and thereby indirectly also the corresponding codons), might have been used in evolution to contribute to a freezing of the genetic code.

Acknowledgments. We would like to thank all our colleagues at the Grenoble Outstation of EMBL who participate in the work on aminoacyl-tRNA synthetases for contributions to the ideas developed in this paper. We would also like to thank Peter Artymiuk, Silvia Onesti, Mike Yarus, Kiyotaka Shiba, Paul Schimmel, Mark Safro, and Jean-Loup Risler for communication of results prior to publication.

References

Achsel T, Gross HJ (1993) Identity determinants of human $tRNA^{ser}$: sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. EMBO J 12:3333–3338

Artymiuk PJ, Rice DW, Poirrette AR, Willett P (1994) A tale of two synthetases. Nat Struc Biol 1:758–760

Asahara H, Himeno M, Shimizu M (1991) A recognition model of $tRNA^{ser}$ by seryl-tRNA synthetase in *E. coli.* Chem Lett 3:363–366

Asahara H, Himeno H, Tamura K, Hasegawa T, Watenabe K, Shimizu M (1993) Recognition nucleotides of *E. coli* $tRNA^{leu}$ and its elements facilitating discrimination from $tRNA^{ser}$ and $tRNA^{tyr}$. J Mol Biol 231:219–229

Asahara H, Himeno H, Tamura K, Nameki N, Hasegawa T, Shimizu M (1994) *Escherichia coli* seryl-tRNA synthetase recognizes $tRNA^{ser}$ by its characteristic tertiary structure. J Mol Biol 236:738–748

Baron C, Böck A (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific $tRNA^{sec}$ of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. J Biol Chem 266:20375–20379

Baron C, Westhof E, Böck A, Giege R (1993) Solution structure of selenocysteine-inserting $tRNA^{sec}$ from *Escherichia coli.* Comparison with canonical $tRNA^{ser}$. J Mol Biol 231:274–292

Belrhali H, Yaremchuk A, Tukalo M, Larsen K, Berthet-Colominas C, Leberman R, Beijer B, Sproat B, Als-Nielsen J, Grübel, Cusack S (1994) Crystal structures at 2.5 Å resolution of seryl-tRNA synthetase complexed with two analogs of seryl adenylate. Science 263:1432–1436

Belrhali H, Yaremchuk AD, Tukalo M, Berthet-Colominas C, Rasmussen B, Bösecke P, Diat O, Cusack S (1995) The structural basis for seryl-adenylate and Ap_4A synthesis by seryl-tRNA synthetase. Structure (in press)

Benner SA, Ellington AD, Tauer A (1989) Modern metabolism as a palimpsest of the RNA world. Proc Natl Acad Sci USA 86:7054–7058

Benner SA, Cohen MA, Gonnet GH, Berkowitz DB, Johnsson KP (1993) Reading the palimpsest: contemporary biochemical data and the RNA world. In: Gesteland RF, Atkins JF (eds) The RNA World. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, pp 27–70

Biou V, Yaremchuk A, Tukalo M, Cusack S (1994) The 2.9 A crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{ser}$. Science 263:1404–1410

Böck A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F (1991) Selenocysteine: the 21st amino acid. Mol Microbiol 5:515–520

Borel F, Vincent C, Leberman R, Härtlein M (1994) Seryl-tRNA synthetase from *Escherichia coli:* implication of its N-terminal domain in aminoacylation activity and specificity. Nucleic Acids Res 22: 2963–2969

Breitschopf K, Gross HJ (1994) The exchange of the discriminator base A73 for G is alone sufficient to convert human $tRNA^{leu}$ into a serine acceptor in vitro. EMBO J 13:3166–3169

Cavarelli J, Rees B, Ruff M, Thierry JC, Moras D (1993) Yeast $tRNA^{asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. Nature 362:181–184

Caverelli J, Eriani G, Rees B, Ruff M, Boeglin M, Mitschler A, Martin F, Gangloff J, Thierry JC, Moras D (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. EMBO J 13:327–337

Cusack S, Berthet-Colominas C, Härtlein M, Nassar N, Leberman R (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5Å. Nature 347:249–255

Cusack S, Härtlein M, Leberman R (1991) Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. Nucleic Acids Res 19:3489–3498

Cusack S (1993) Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases: an update. Biochimie 75:1077–1081

Delarue M, Poterszman A, Nikonov S, Garber M, Moras D, Thierry JC (1994) Crystal structure of a prokaryotic aspartyl-tRNA synthetase. EMBO J 13:3219–3229

Delarue M, Moras D (1993) The aminoacyl-tRNA synthetase family: modules at work. Bioessays 15:675–687

Dock-Bregeon AC, Garcia A, Giegé R, Moras D (1990) The contacts of yeast $tRNA^{ser}$ with seryl-tRNA synthetase studied by footprinting experiments. Eur J Biochem 188:283–290

Dock-Bregeon AC, Westhof E, Giegé R, Moras D (1989) Solution structure of a tRNA with a large variable arm: yeast $tRNA^{ser}$. J Mol Biol 206:707–722

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motif. Nature 347:203–206

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Fujinaga M, Berthet-Colominas C, Yaremchuk AD, Tukalo MA, Cusack S (1993) Refined crystal structure of the seryl-tRNA synthetase from *Thermus thermophilus* at 2.5 A resolution. J Mol Biol 234:222–323

Gibson TJ, Lamond AI (1990) Metabolic complexity in the RNA world and implications for the origin of protein biosynthesis. J Mol Evol 30:7–15

Gilbert W (1986) The RNA world. Nature 319:618

Härtlein M, Madern D, Leberman R (1987) Cloning and caracterization of the gene for *Escherichia coli* seryl-tRNA synthetase. Nucleic Acids Res 15:1005–1017

Hartman H (1978) Speculations on the evolution of the genetic code. II. Orig Life 9:133–136

Higgins DG, Sharp PM (1988) CLASTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Himeno H, Hasegawa T, Ueda T, Watanabe K, Shimuzu M (1990) Conversion of aminoacylation specificity from $tRNA^{Tyr}$ to $tRNA^{Ser}$ in vitro. Nucleic Acids Res 18:6815–6819

Hopfield JJ (1978) Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. Proc Natl Acad Sci USA 75:4334–4338

Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. Science 267:643–647

Joyce GF, Orgel LE (1993) Prospects for understanding the origin of the RNA world. In: Gesteland RF, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, NY, pp 1–25

Kawaguchi Y, Honda H, Taniguchi-Morimura J, Iwasaki S (1989) The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea.* Nature 341:164–169

Khan AS, Wilcox AS, Polymeropoulos, Hopfins JA, Stevens TJ, Robinson M, Orpana AK, Sikela JM (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAc. Nature Genet 2:180–185

Kumazawa Y, Himeno H, Miura K, Watanabe K (1991) Unilateral aminoacylation specificity between bovine mitochondria and eubacteria. J Biochem 109:421–427

Lehman H, Jukes TH (1988) Genetic code development by stop codon takeover. J Theor Biol 135:203–214

Lunel C, Buttin G, de Saint Vincent BR (1992) A seryl-tRNA synthetase gene is coamplified with the adenylate deaminase 2 gene in coformycin resistant Chinese hamster fibroblasts. Nucleic Acids Res 20:2597

Mayerfeld I, Yarus M (1994) An RNA pocket for an aliphatic hydrophobe. Nature Struct Biol 1:287–292

Mirande M, Lazard M, Kerjan P, Bec G, Agou F, Quevilo S, Waller JP (1993) The aminoacyl-tRNA synthetase family: an evolutionary view of their structural organisation. In: Nierhaus KH, Franceschi F, Subramanian AR, Erdmann VA, Wittmann-Liebold B (eds) The translation apparatus: structure, function, regulation, evolution. pp 657–668

Miseta A, Woodley CL, Greenberg JR, Slobin LI (1991) Mammalian seryl-tRNA synthetase associates with mRNA in vivo and has homology to elongation factor 1 alpha. J Biol Chem 266:19158–19161

Mosyak L, Reshetnikova L, Safro M (1993) Phenylalanyl-tRNA synthetase from *Thermus thermophilus* has four antiparallel folds of which only two are catalytically functional. Biochimie 75:1091–1098

Murzin AG (1993) Oligonucleotide/oligosaccharide binding-fold: common structural and functional solution for non-homologous sequences. EMBO J 12:861–867

Noller HF, Hoffarth V, Zimniak L (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. Science 256: 1416–1419

Normanly J, Ogden RC, Horvath SJ, Abelson J (1986) Changing the identity of a transfer RNA. Nature 321:213–219

Normanly J, Ollick T, Abelson J (1992) Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. Proc Natl Acad Sci USA 89:5680–5684

Onesti S, Miller AD, Brick P (1995) The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli.* Structure 3:163–176

Orgel LE (1989)The origin of polynucleotide-directed protein synthesis. J Mol Evol 29:465–474

Piccirilli JA, McConnell TS, Zaug AJ, Noller HF, Cech TR (1992) Aminoacyl esterase activity of the Tetrahymena ribozyme. Science 256:1420–1424

Price S, Cusack S, Borel F, Berthet-Colominas C, Leberman R (1993) Crystallization of the seryl-tRNA synthetase: $tRNA^{ser}$ complex of *Escherichia coli.* FEBS Lett 324:167–170

Rich A (1965) Evolving genes and proteins. ed. Bryson V, Vogel HG Academic, New York, pp 453–468

Rogg H, Staehelin M (1971) Nucleotide sequences of rat liver serine-tRNA1. Products of digestion with pancreatic ribonuclease. Eur J Biochem 21:235–242

Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D (1991) Class II aminoacyl-tRNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{asp}$. Science 252:1682–1689

Sassanfar M, Szostak JW (1993) An RNA motif that binds ATP. Nature 364:550–552

Saks ME, Sampson JR, Abelson JN (1994) The transfer RNA identity problem: a search for rules. Science 263:191–197

Saks ME, Sampson JR, (submitted) Patterns and modes of tRNA sequence evolution. J Mol Evol

Sampson JR, Saks ME (1993) Contributions of discrete $tRNA^{ser}$ domains to amino-acylation by *E. coli* seryl-tRNA synthetase: a kinetic analysis using model RNA substrates. Nucleic Acids Res 21:4467–4475

Schimmel P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci USA 90:8763–8768

Schulman LH (1991) recognition of tRNAs by aminoacyl-tRNA synthetases. Prog Nucleic Acid Res Mol Biol 41:23–47

Sidorik LL, Gudzera OI, Dragovoz VA, Tukalo MA, Beresten SF (1991) Immuno-chemical non-cross-reactivity between eukaryotic and prokaryotic seryl-tRNA synthetases. FEBS Lett 292:76–78

Steinberg S, Gautheret G, Cedergren R (1994) Fitting the structurally diverse animal mitochondrial tRNAs to common three-dimensional constraints. J Mol Biol 236:982–989

Ueda T, Yotsumoto Y, Ikeda K, Watanabe K (1992) The T-loop region of animal mitochondrial tRNAser(AGY) is a main recognition site for homologous seryl-tRNA synthetase. Nucleic Acids Res 20: 2217–2222

Wetzel R (1978) Aminoacyl-tRNA synthetase families and their significance to the origin of the genetic code. Orig Life 9:39–50

Weygand-Durasevic I, Johnson-Burke D, Söll D (1987) Cloning and characterization of the gene coding for cytoplasmic seryl-tRNA synthetase from *Saccharomyces cerevisiae.* Nucleic Acids Res 15: 1887–1904

Weygand-Durasevic I, Ban N, Jahn D, Söll D (1993) Yeast seryl-tRNA synthetase expressed in *Escherichia coli* recognizes bacterial serine-specific tRNA *in vivo.* Eur J Biochem 214:869–877

Weygand-Durasevic I, Nalaskowska M, Söll D (1994) Coexpression of eukaryotic $tRNA^{Ser}$ and yeast seryl-tRNA synthetase leads to functional amber suppression in *Escherichia coli.* J Bacteriol 176:232–239

Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW (1992) *E. coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. Proc Natl Acad Sci USA 89:9257–9261

Wu XQ, Gross HJ (1993) The long extra arms of human tRNA ((Ser)Sec) and tRNA(Ser) function as major identity elements for serylation in an orientation-dependent, but not sequence specific manner. Nucleic Acids Res 21:5589–5594

Yarus M (1989) Specificity of arginine binding by the Tetrahymena intron. Biochemistry 28:980–988

Yokogawa T, Suzuki T, Ueda T, Mori M, Ohama T, Kuchino Y, Yoshinari S, Motoki I, Nishikawa K, Osawa S (1992) Serine tRNA complementary to the nonuniversal serine codon CUG in Candida cylindracea: evolutionary implications. Proc Natl Acad Sci USA 89:7408–7411