Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4537–4542, May 1996
Evolution

The presence of codon–anticodon pairs in the acceptor stem of tRNAs

(RNA world/double-stranded code/complementary tRNAs/aminoacyl-tRNA synthetases)

SERGEI RODIN*†‡, ANDREI RODIN§, AND SUSUMU OHNO*

*Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010; †Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, av. acad. Lavrentieva 10, Novosibirsk 630090, Russia; and §Human Genetics Center, University of Texas, SPH, P.O. Box 20334, Houston, TX 77225

Contributed by Susumu Ohno, December 28, 1995

ABSTRACT **A total of 1268 available (excluding mitochondrial) tRNA sequences was used to reconstruct the common consensus image of their acceptor domains. Its structure appeared as a 11-bp-long double-stranded palindrome with complementary triplets in the center, each flanked by the 3′-ACCD and NGGU-5′ motifs on each strand (D, base determinator). The palindrome readily extends up to the modern tRNA-like cloverleaf passing through an intermediate hairpin having in the center the single-stranded triplet, in supplement to its double-stranded precursor. The latter might represent an original anticodon–codon pair mapped at 1-2-3 positions of the present-day tRNA acceptors. This conclusion is supported by the striking correlation: in pairs of consensus tRNAs with complementary anticodons, their bases at the 2nd position of the acceptor stem were also complementary. Accordingly, inverse complementarity was also evident at the 71st position of the acceptor stem. With a single exception ($tRNA^{Phe}$–$tRNA^{Glu}$ pair), the parallelism is especially impressive for the pairs of tRNAs recognized by aminoacyl-tRNA synthetases (aaRS) from the opposite classes. The above complementarity still doubly presented at the key central position of real single-stranded anticodons and their hypothetical double-stranded precursors is consistent with our previous data pointing to the double-strand use of ancient RNAs in the origin of the main actors in translation—tRNAs with complementary anticodons and the two classes of aaRS.**

In the L-folded three dimensional structure of present-day tRNAs, their anticodon and 3′-terminal amino acid attachment site are separated by a distance of ≈70 Å so that tRNA is unable to be correctly aminoacylated by itself. The specific aminoacylation is provided enzymatically by 20 different species of aminoacyl-tRNA synthetases (aaRS), one for each amino acid and the cognate set of isoacceptor tRNAs.

On the one hand, to avoid the "chicken or egg" circular argument, direct interaction between anticodon and amino acid had to be hypothesized at 3′ half of proto-tRNAs (1–3). In addition, the genetic code could have already existed in the prebiotic RNA world, possibly by direct aminoacylation of anticodons helped by RNA synthetase-like activity inherent in group I introns of tRNA genes (3, 4). These anticodon–amino acid complexes supposedly operated as "coding" coenzymes of ribozymes (3). Whether or not they might serve primitive translation as adaptors remains unclear (3). Nevertheless, internal sequence periodicity indicates that tRNAs most likely evolved from very short hairpins (5), so that unavoidably, at some critical step of their elongation, the recognition of anticodon and that of the acceptor stem by aaRS came to be coupled.

Indeed, even modern tRNAs drastically truncated to the acceptor mini-helix proved to contain sufficient information to be properly charged by the correct amino acid (6). A key role in such recognition belongs to the 3′ unpaired 73rd base and first 3 bp of the acceptor stem (6, 7). The "reciprocal" truncations of aaRSs, such that in extreme cases the reduced enzyme cannot even physically extend to cover the anticodon, did not influence the rate and specificity of aminoacylation, as well (6). Accordingly, "the second code" placed into the acceptor stem of tRNAs was postulated. As more essential for aminoacylation (6, 8, 9), these observations led to the claim of the second code being older than the codon–anticodon code and, accordingly the "frozen accident" hypothesis (10), was revived for the particular algorithm of codon (anticodon)–amino acid assignment (6, 9).

Yet the codon assignment does not appear fortuitous. First, a general pattern of "similar codons for similar amino acids" does exist (11). Second, Corey–Pauling–Koltun models of stereochemical C4N complexes (2) showed the key-lock relations, though weak, between the anticodon (noncovalently complexed with the discriminator base) and cognate amino acid. Third, certain RNAs showed the capacity to preferentially bind with not only basic amino acids such as Arg (12, 13) but also Val and other aliphatic amino acids (14). At any rate, the idea of the acceptor code evolving independent of the codon–anticodon code encounters a formidable difficulty in explaining the subsequent linkage between the two codes.

Concordant evolution of the two codes becomes a possibility by an assumption that the proto-acceptor helix in proximity of the DCCA-3′ terminus originally contained the codon–anticodon pair. Statistical examination of modern tRNAs did reveal significant content of codons for Ala, Gly, Val, and Asp in the 3-4-5 positions of the acceptor stem (15, 16). A problem here was that most of the acceptor stem-specific identity elements mapped at 1-2-3 positions (6, 7). Therefore, we reexamined the acceptor in search of vestiges of the original codon-anticodon pair in 1-2-3 positions. For this purpose, we used constructed consensus tRNAs instead of present-day sequences, as discussed in (17, 18). The main results are as follows.

First Step Toward tRNA: The Double-Stranded Palindrome

Fig. 1*A* presents the common consensus of tRNA acceptor. The 3′ strand of the acceptor is in fact a 11-base-long palindrome with two flanking complementary repeats, the DCCA on the 3′ end and the NGGU on the 5′ end. The 5′ strand appears to be of the same palindromic structure if one assumes for the UGGd motif (Fig. 1*A*, in italics) to complement the universal unpaired 3′-DCCA tail. Thus, the central pair of triplets GSS (S is G or C) presents the only difference between the two strands. Fig. 1*B* shows how this structure can be produced by self-templating of the original 5′DCCA3′ tetranucleotide (or, with the same results, of its 5′-UGGd-3′

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.

Abbrevation: aaRS, aminoacyl-tRNA synthetase.

‡To whom reprint requests should be sent at the * address.

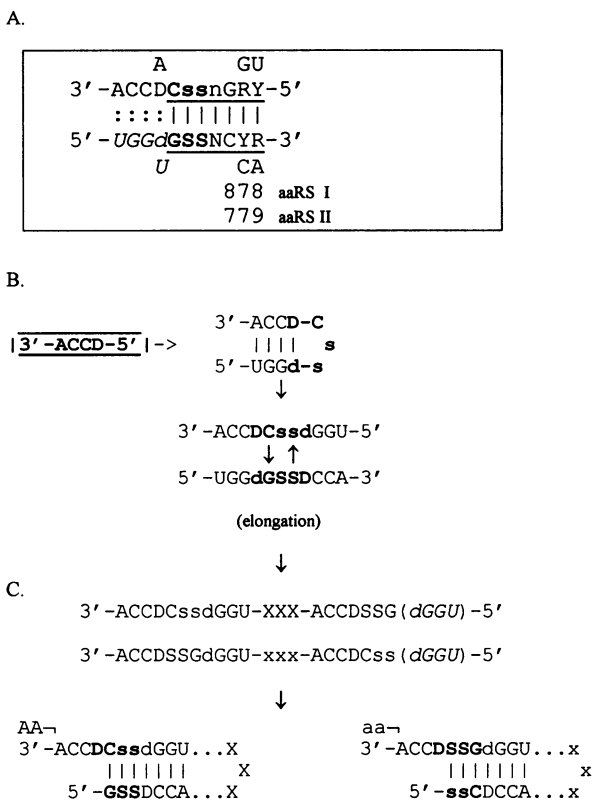


FIG. 1. (*A*) Common consensus acceptor stem reconstructed from 50 individual consensus sequences, each representing a particular group of tRNAs with the same anticodon as described (17, 18). A total of 1268 tRNA sequences were used (all retrieved from tRNA sequence database; ref. 19) covering viruses, archaebacteria, eubacteria, chloroplasts, and eukaryotes, but not mitochondria. The corresponding consensus sequences for 20 isoaccepting tRNA families (not shown) are very similar (with a few one-base exceptions) to the "ancestral" ones recently inferred by parsimony algorithms (20). Numbers under the structure indicate how often the consensus bases occur at 5–6-7 positions in the 10 isoacceptor tRNAs, recognized by class I and class II aaRSs, respectively. (*B*) Self-priming and self-templating of the original 3′-ACCD-5′ tetramer generates a palindrome with a central triplet coming as a fold-back hook. It replicates into the similar palindrome with the complementary triplet in the middle. (*C*) Self-templating elongation of the above both palindromes results in two helices with the strand-mirror position of internal complementary triplets and precisely the same flanking complementary repeats, ACCD and dGGU. Removal of the latter on the 5′ terminus gives the acceptor-like domain of the present-day tRNAs (see *A*) capable of aminoacylation.

complement), the internal loop coming as a fold-back hook. Due to the perfect symmetry of the palindromic templates, the process generates two equivalent hairpins with the mirror image location of the same central complementary triplets on the opposite strands (boldface type in Fig. 1*C*). Most of the present-day tRNAs begin with G at their 5′ end. Therefore we suppose that the two complementary triplets originally also were the palindromes GSC. Interestingly, the GGC and GCC compose the anticodon-codon pair for Ala and, symmetrically, the codon-anticodon pair for Gly. Their nearest 1-bp transition derivatives are 5′-GAC–GUC-3′ and 5′-GUC–GAC-3′ pairs encoding Val and Asp, respectively. These four, Ala, Gly, Val, and Asp, are the most preponderant amino acids among abiotically synthesized (21). Accordingly, they were thought to be the first amino acids recruited into translation. The next amino acids to gain prominence might have been Arg with anticodons CCG, GCG, and Pro with anticodons CGG and GGG.

The earlier reported significant content of codons at 3-4-5 positions, just for the "primordial" tetrad of amino acids Ala, Gly, Val, and Asp (15, 16), does not seem convincing for the following reason. $tRNA^{Ala}$ and $tRNA^{Gly}$ sequences are so G:C rich that the Ala and Gly codons GCC and GGC (the most frequent at 3-4-5 positions) are *a priori* expected to occur by chance everywhere in the two tRNAs. Indeed, we did not find such preferential concentration of GCC and GGC at 3-4-5 site of the acceptor of $tRNA^{Ala}$ and $tRNA^{Gly}$ (though the significant difference did exist for 12 of 20 amino acids if the acceptor helix as a whole is compared with TΨC-, D-, and variable extra domains of modern tRNAs; unpublished data). Actually, according to the consensus palindrome structure (Fig. 1), the 5′-SDC-3′ might originally occupied the 3-4-5 positions, the DC coming from the uniform DCCA flanking motif. Note that all eight 5′-SDC3′ anticodons are assigned to Ala, Gly, Val, and Asp, prevailing among abiotically synthesized amino acids. Not surprisingly therefore, we meet these triplets at the 3-4-5 site, but in noncognate tRNAs as well.

Early Bifunctional Adaptors

Each of the palindromes in Fig. 1*C* may approximate the earliest prototype of tRNAs, the paired GNC at first three positions with the 3′ adjacent base determinator constituting a primitive code for a few primordial amino acids. To serve translation as bifunctional adaptors, these two short helices necessarily needed to be supplemented by a single-stranded proto-anticodon (22). The general outline of such duplication has been already suggested (ref. 5; see figure 5 in ref. 15). Alternatively, upon melting of the original double-stranded palindrome, one strand might form a separate proto-anticodon hairpin while the other simply directs synthesis of the opposite strand to rebuild the acceptor.

Either way, the two palindromes (Fig. 1*C*) were both capable of correct aminoacylation before appearance of their single-stranded anticodons. This conclusion comes from the antiparallel alignments of the tRNAs with complementary anticodons. Indeed, if such tRNAs did evolve from the two palindromes in Fig. 1*C*, then in "head-to-tail" alignment they should complement each other in the anticodon hairpin but not in the acceptor stem (Fig. 2*A*). This was exactly the difference we earlier found for consensus tRNAs (17, 18). At that time, the observed difference was interpreted to reflect a less constrained mutational erosion of the acceptor stems compared to the anticodon ones (17, 18). Actually, the difference might have been predetermined by the structure of the two original modules with mirror codon–anticodon-like pairs in the middle (Fig. 1*C*). The weak point of the above scenario is that despite independent elongation of the two palindromes, in the end they had to have resulted in the uniform cloverleaf structure.

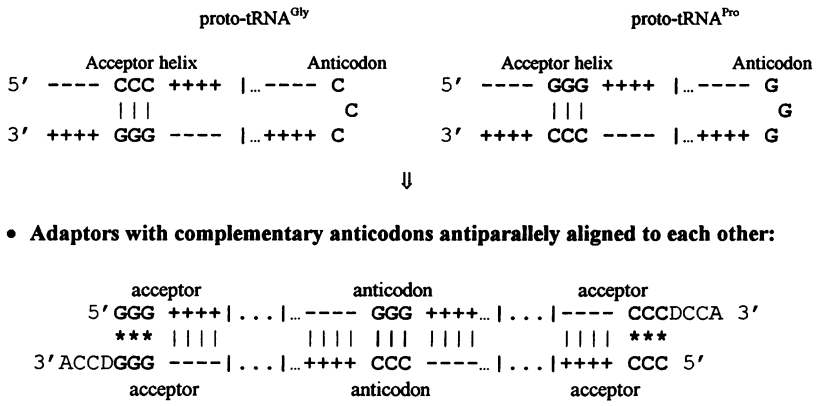
The alternative explanation might be that adaptors with complementary anticodons arose later as a simple complementary replica of already bifunctional molecule. This scenario predicts a perfect complementarity in both major stems, the acceptor and anticodon (Fig. 2*B*) that was not detected (17, 18). The most serious defect in this scenario is that the corresponding precursors of the complementary anticodons had to have occurred on the opposite strands of the two acceptors that made the two helices essentially indistinguishable for primordial tRNA synthetases, regardless of whether they were ribozymes or protein enzymes (Fig. 2*B*).

Correlated Second Base Complementarity of the Two Codes

For the most important central base of the genetic code, we unexpectedly found a surprising correlation between the real single-stranded anticodon and their hypothetical double-stranded codon–anticodon ancestor (Table 1). Literally, for consensus tRNAs with complementary anticodons, their bases in the 2nd position on the 5′ strand of the acceptor helix also turned out to be complementary with each other. Accordingly,

A:

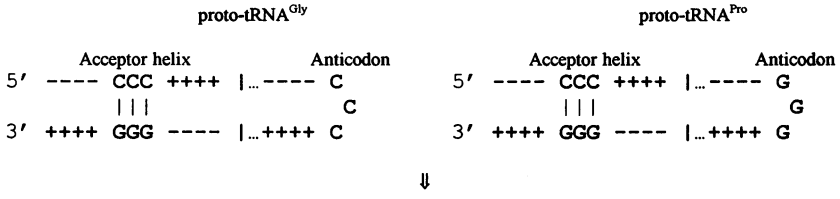
- **Complementary adaptors with duplicated anticodon/codon words:**



- **Adaptors with complementary anticodons antiparallely aligned to each other:**

B:

- **Complementary adaptors with duplicated anticodon/codon words:**



- **Adaptors with complementary anticodons antiparallely aligned to each other:**

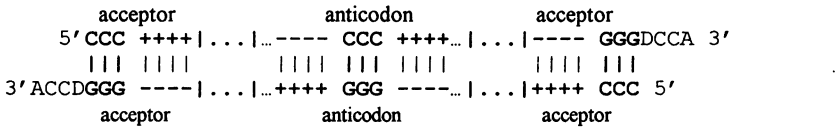


FIG. 2. (*A*) Protoadapters originated from the two palindromes shown in Fig. 1*C*. They resemble mitochondrial tRNAs with lost TΨC- or (and) D-arm earlier also considered as intermediate adaptor (23). Subsequent elongation reproduces the tRNA-like cloverleaf molecule with periodically repeated UGGU and ACCA motifs along the sequence (see also ref. 24), for example the invariant 5′-URG-3′ at 8–10 positions, complementing the preceding CCA of the acceptor palindrome, the same 5′-UGG-3′ triplet at 17–19 positions of the D loop, the frequent 5′-UGGNC-3′ motif of the extra loop between the anticodon and TΨC stem. The scheme predicts for antiparallelly aligned modern tRNAs with complementary anticodons to display a significant base pairing in the anticodon stem and loop, in contrast to the acceptor stems, where impairing must occur invariably at the 2nd position and frequently at the 3rd one. It is supported by the frequency distribution of complementary base pairs along the acceptor helix. (*B*) Hypothetical pairs of complementary protoadaptors with the opposite strand location of their anticodon precursors in the acceptor. Note that the two acceptors are in fact indistinguishable for synthetases.

inverse complementarity was seen between bases of the 71st position on its 3′ strand. Of all 32 such pairs, there were only 3 exceptions (underlined in Table 1). These three, however, were examples of the opposite strand location of anticodons unrecognizable by synthetases (Fig. 2*B*). Were the two codes completely independent of each other, then the expected (due to chance) value of base complementarity at the 2nd position of acceptor (predominantly G/C, see Table 1) would be 0.5, while the authentic value was $29/32 = 0.91$, the probability of this or higher value being about $1/10^6$! If one takes into account a few cases of T base occurrence at the 2nd position, the random chance to have 29/32 complementary bases here, in parallel with complementary single-stranded anticodons, will be even smaller. The 3rd position showed the similar tendency, albeit not as clear. This is expected if the anticodon-like triplet originally occupied the 1-2-3 positions, since the 1st (5′) and 3rd (3′) bases of anticodon correspond to the more degenerated 3rd (3′) and 1st (5′) bases of codon, respectively. All other bases in the acceptor (including base determinator at the 73rd impaired position) appeared to be not complementarily correlated with the anticodon at all.

The Duplicated Code and the Double-Strand Origins of tRNAs and aaRSs

With a single exception ($tRNA^{Phe}$ - $tRNA^{Glu}$ pair), the above 2nd base-associated correlation was especially impressive for the pairs of tRNAs with complementary anticodons recognized by aaRS from the opposite classes (Table 1). Furthermore, for 15 such pairs, out of the 18, the central purine base (G or A) in the anticodon was accompanied by the presence of purine (invariably G) at the 2nd position of the acceptor 5′ strand, and symmetrically the same parallelism occurred for the central pyrimidine bases (mostly C in the acceptor). It means that anticodon-like triplets were mostly concentrated on the 5′ acceptor strand of individual tRNAs. Exceptions were pairs of tRNAs with complementary anticodons recognized by aaRSs from the same class (see Table 1).

The above finding is of particular interest as completing our hypothesis that genes of the two adaptor molecules, tRNAs with complementary anticodons and the two types of aaRSs, could have originated in pairs, each from the complementary strands of primordial RNAs (26). Vestiges of the double-strand

Table 1. Correlated complementarity of the second bases in anticodon and acceptor

Class I × Class I pairs	Class I × Class II pairs	Class II × Class II pairs
Leu (C, TAA) ⇔ Gln (G, TTA*)	Phe (C, GAA) ⇔ Glu (C, TTC)	Phe (C, AAA) ⇔ Lys (R, TTT)
Leu (C, CAA) ⇔ Gln (G, TTG)	Leu (G, AAG) ⇔ Lys (C, CTT)	Ser1 (G, GGA) ⇔ Gly (C, TCC)
Leu (G, GAG) ⇔ Glu (C, CTC)	Ile (G, AAT) ⇔ Asn (C, ATT)	Pro (G, GGG) ⇔ Gly (C, CCC)
Leu (C, TAG) ⇔ Gln (G, CTA*)	Ile (G, GAT) ⇔ Asp (C, ATC)	Thr (C, GGT) ⇔ Ser2 (G, GCT)
Leu (Y, CAG) ⇔ Gln (G, CTG)	Met (G, CAT) ⇔ His (Y, GTG)	Thr (C, GGT) ⇔ Gly (C, GCC)
Ile (G, TAT) ⇔ Tyr (S, GTA)	Val (G, AAC) ⇔ Asn (C, GTT)	Ala (G, AGC) ⇔ Ser2 (G, GCT)
Val (T, TAC) ⇔ Tyr (S, GTA)	Val (G, GAC) ⇔ Asp (C, GTC)	Ala (G, GGC) ⇔ Gly (C, GCC)
	Val (G, CAC) ⇔ His (Y, GTG)	
	Ser1 (G, AGA) ⇔ Arg (C, TCT)	
	Ser1 (G, TGA) ⇔ Trp (S, TCA*)	
	Ser1 (G, CGA) ⇔ Arg (C, TCG)	
	Pro (G, RGG) ⇔ Arg (C, CCT)	
	Pro (G, TGG) ⇔ Trp (S, CCA)	
	Pro (G, CGG) ⇔ Arg (C, CCG)	
	Thr (C, TGT) ⇔ Cys (S, GCA)	
	Thr (C, CGT) ⇔ Arg (G, ACG)	
	Ala (G, TGC) ⇔ Cys (S, GCA)	
	Ala (G, CGC) ⇔ Arg (T, GCG)	

Shown are pairs of complementary anticodons and the respective 2nd bases in the 5′ acceptor strand of the corresponding consensus tRNA genes. To use all 32 pairs, the tRNA genes, not tRNA themselves, are presented in the table, and three "nonsense" anticodons (marked by asterisks) are assigned to Gln (TTA, CTA) and Trp (TCA) according to ref. 25. There are two unlinked groups of such pairs for serine designated Ser1 and Ser2, respectively. The 2nd bases in the acceptor and anticodon are in boldface type. Underlined are the three exceptions when the complementarity of the anticodons is not accompanied by the complementarity of the 2nd bases in the 5′ strand of the corresponding acceptor helices.

origin were still imprinted in their sequences (Fig. 3). Yet, the complementarity between consensus tRNAs with complementary anticodons was detected in their antiparallelly aligned anticodon loop and stem regions, but not in acceptors (17, 18). The reason for the above then became clear. In contrast, the complementarity between class I and class II aaRSs was shown

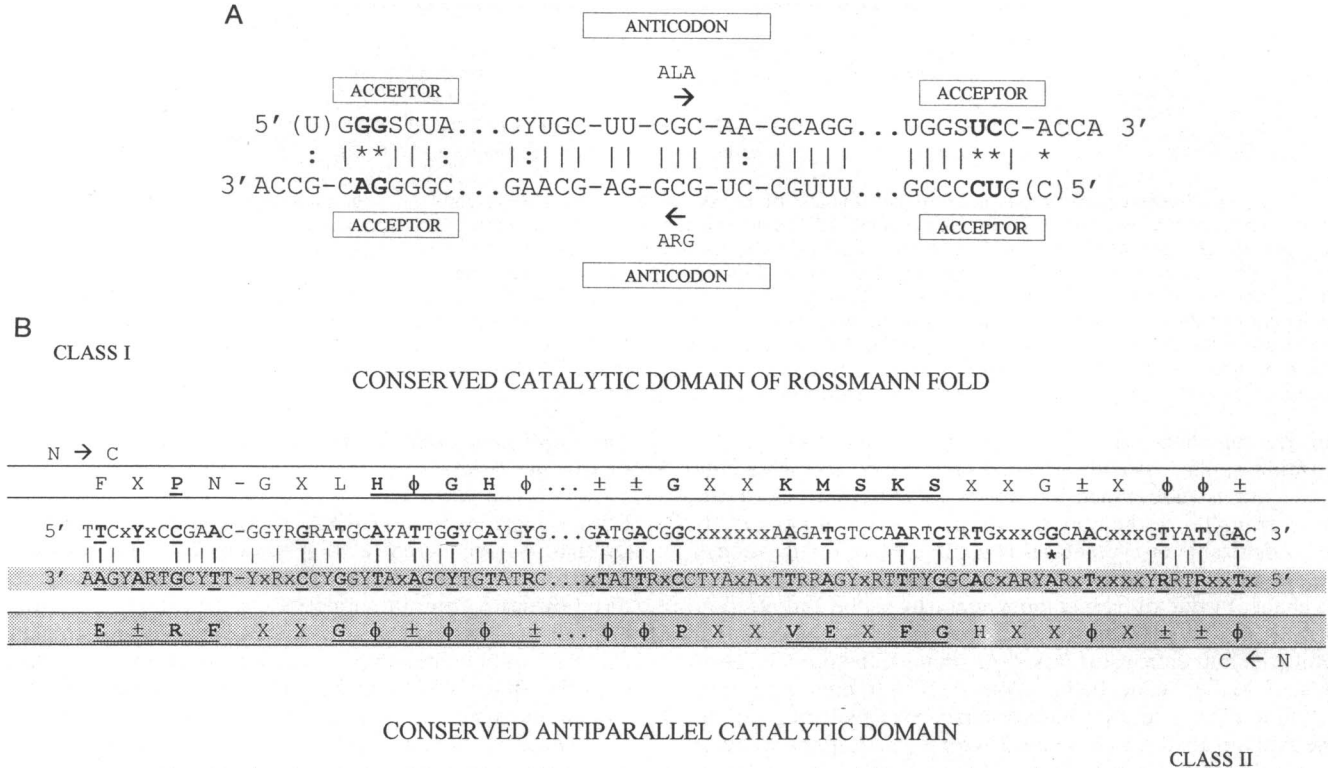


FIG. 3. The two antiparallel alignments, one (*A*) of the consensus tRNAs with complementary anticodons (the pair of $tRNA^{Ala}$ and $tRNA^{Arg}$ is shown) and the other (*B*) of the consensus signature motifs of the two aaRSs from the opposite classes. (*A*) Shown in boldface type are base mismatches at the 2nd and 3rd paired positions of the corresponding acceptors. Other examples of base complementarity between such tRNAs within their head-to-tail aligned anticodon stem and loop regions (in contrast to their acceptor stems) are presented in refs. 17 and 18. (*B*) Sequence complementarity revealed for class-defining signature motifs of the two aaRSs within their catalytic domains. Amino acids are given in the conventional single-letter symbols, boldface underlined letters showing the most conserved amino acid residues. The two HIGH- and KMSKS-containing signature motifs of the class I complement the motifs 2 and 1 of the class II, respectively (for more detail see ref. 26). In the center is shown a "head-to-tail" alignment of the corresponding gene fragments with underlined central bases of codons. Amino acid groups are designated as follows: ±, polar residue; ϕ, hydrophobic residue; X, any residue. The two complete alignments of individual aaRSs (for each amino acid) from the two classes and statistical evidence of their nonrandom sense–antisense similarity are presented in ref. 26.

by their signature motifs within the catalytic domains that recognize ATP, amino acid and the tRNA 3′-terminal half of the acceptor, but not the anticodon (26). The correlated 2nd base complementarity (Table 1) filled the gap admirably.

Division of labor between the two classes of aaRS is evident in that (*i*) the hydrophobic amino acids are activated mostly by the class I enzymes, while the small polar ones are aciviated by the class II enzymes, and (*ii*) the charged amino acids are equally distributed in the two classes, though the largest of them are preferred by class I synthetases (27). Fig. 3 and Tables 1 and 2 show how congruent the two complementarities are. Those aaRSs attaching amino acids onto the 2′-OH group of the tRNA terminal ribose are all (except for PheRS) from class I, whereas those specific for 3′-OH are all from class II (28–30). Accordingly, the two enzymes approach the cognate tRNAs from the opposite sides and, in general, the corresponding three-dimensional complexes "tRNA-aaRS" look like mirror images of one another (29, 30). These two modes of class-specific OH-group recognition could be an inborn rationality which we formulated as the mirror enzymes "have to" recognize the opposite sides of mirror substrates (26). The rationality correlates with the genetic code structure: in a binary R–Y approximation, anticodons with complementary second bases and cognate amino acids are almost ideally balanced with regard to the aaRS class membership (Table 2), *viz.* 18 xYx anticodons for 6 amino acids recognized by class I aaRSs correspond to 18 xRx anticodons for 5 amino acids recognized by class II aaRSs, and reciprocally, 14 xRx anticodons (4 amino acids) recognized by class I aaRSs are complemented by 14 xYx anticodons (5 amino acids) recognized by class II aaRSs, in total compiling 32 anticodons (10 amino acids) of class I per 32 anticodons (10 amino acids) of class II. What is more, this balance is now extended to the correlated base complementarity at the 2nd position of acceptor (shaded in Table 2).

Metabolically, the RNA world was sufficiently rich (31) in order to synthesize proteinous amino acids with ribozyme help. Since the first synthetases were likely ribozymes as well, it seems reasonable to assume for them a greater binding affinity to basic amino acids (Arg, Lys, His) (4). Not surprisingly, it is this group of tRNA–aaRS complexes which fits the "rule" better than any other complex: tRNAs with complementary anticodons are recognized by the complementary synthetases. This rule did not apply well at all when complexes were grouped on the basis of the precursor–product relationship of their amino acids in biosynthesis accordingly the coevolution theory of the genetic code (32). Compared to randomized codes, the first group of "prebiotically" synthesized amino acids, and the second group of their biosynthetic products (33), were not significantly associated either with a complementarity of their anticodons or with the complementarity of their acceptor 2nd bases, or with the two complementing classes of cognate aaRSs (Fig. 4). However, it should be noted here that the precursor–product component of the genetic code might be specifically optimized in relation to the first (5′) base of codons (3, 34), while the complementary component under discussion is with regard to their second base.

Concluding Remarks

All attempts have thus far failed to decipher the acceptor code in terms of simple words resembling the anticodons. This failure is expected, because long coevolution of tRNAs and protein aaRSs did not conserve the acceptor's prototypic code

Table 2. Distribution of the 20 aaRSs into two classes correlated with the 2nd base in anticodon and in the acceptor stem 5′ strand of cognate tRNAs (adopted from ref. 27)

Class I synthetases (HIGH + KMSKS motifs, charge of 2′ OH)						Class II synthetases (3 signature motifs, charge of 3′ OH)					
		The 2nd base						The 2nd base			
		5′ Acceptor		Anticodon				5′ Acceptor		Anticodon	
LeuRS	α	**C**	3	**A**	6	LysRS	α2	**C**	1	**T**	2
		T	1					**R**	1		
		G	2								
IleRS	α	**G**	3	**A**	3	AspRS	α2	**C**	2	**T**	2
MetRS	α (α)	**G**	1	**A**	1	AsnRS	α2	**C**	2	**T**	2
ValRS	α	**G**	3	**A**	4	SerRS	α2	**G**	4	**G**	4
		T	1			SerRS	α2	**G**	2	**C**	2
CysRS	α	**S**	2	**C**	2	ThrRS	α2	**C**	4	**G**	4
ArgRS	α	**C**	4	**C**	6	ProRS	α2	**G**	4	**G**	4
		T	1								
		G	1								
TrpRS	α2	**S**	2	**C**	2	HisRS	α2	**Y**	2	**T**	2
TyrRS	α2	**S**	2	**T**	2	AlaRS	α4 (α)	**G**	4	**G**	4
GluRS	α	**C**	2	**T**	2	GlyRS	α2β2 (α2)	**C**	4	**C**	4
GlnRS	α	**G**	4	**T**	4	**PheRS**	**α2β2 (α, α2)**	**C**	2	**A**	2
		G	15	C	10			C	15	G	10
		G	2	A	8			Y	2	T	8
		C	12	T	2			G	12	A	2
		T	3	T	6			G	3	G	6
				A	*6*					*C*	*6*
		Y ≈	15	Y	18			R	15	R	18
		R ≈	17	R	14			Y	17	Y	14

Shown are oligomeric structures of the enzyme, the consensus 2nd bases of the 5′ acceptor strand and the anticodon, and their numbers, respectively. All 32 pairs of complementary anticodons are used, as in Table 1. PheRS is marked in boldface type as a member of the class II aminoacylating on the 2′ OH group. (S = G or C.) Underlined are the cases of "wobble" complementarity between the 2nd bases G and U. In italics are shown A*C pairs.

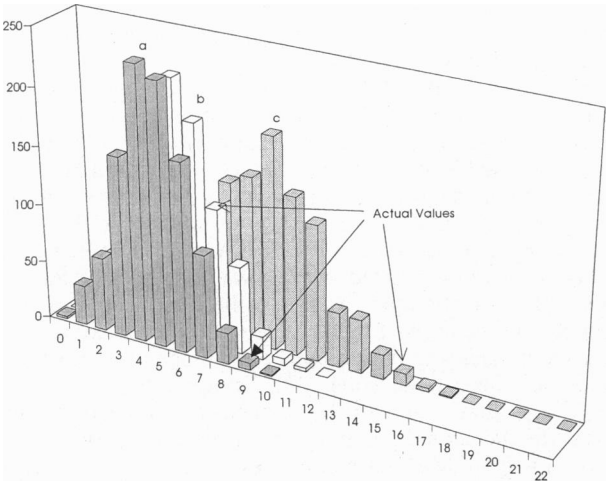


FIG. 4. Simulational testing of the hypothesis that tRNAs with complementary anticodons are recognized by the complementary aaRSs, assuming that (plot a) according to ref. 4, the first amino acids recruited into translation were the basic ones (Arg, His, Lys; 10 anticodons in total); (plot c) according to the coevolutionary hypotheses of precursor–product relations between amino acids (32), Arg, Lys, His, Asn, Trp, Cys, Gln, Tyr, Phe, and Met (22 anticodons in total) could have entered translation secondarily, as products of the "inventive" biosynthesis, from the early group of prebiotically synthesized amino acids (33). Plot b represents the difference between the plots c and a. Each plot depicts the distribution of the following random variable: the number of anticodon complementary pairs (where the first anticodon belongs either to set a or b or c, as described above), such that the first and second anticodons of the pair are recognized by the aaRSs of the different classes (i.e., class I and class II or *vice versa*, respectively). For each set we have opted for 1000 randomly generated "codes" with the basic internal code structure (number of codons for each amino acids, degeneracy patterns, etc.) remaining intact. It appeared to be virtually impossible to treat the aforementioned distribution(s) analytically due to the complexity of the underlying statistical model (the explicit distribution lies somewhere in between the hyperheometrical and binomial ones). For the simulated distribution (plot a), the actual value, being 9 out of 10, is highly nonrandom (assuming quasinormality, it fits into the right 2.5% distribution tail). The actual value for plot c also appears to be highly nonrandom, but only due to the basic amino acid component (i.e., the same plot a subset). Indeed, when factored out, the remaining amino acids show the distribution with a perfectly random actual value (plot b). Similarly, within each of these two groups [i.e., separately for "precursor" and "product" amino acids (33)], the actual numbers of the corresponding anticodon pairs recognized by the complementing synthetases also appeared to be statistically indistinguishable from the values expected due to chance.

in its original form. More distinct traces of compliance to the primordial code could be imprinted in the ribozyme synthetases. Indeed, the concerted origin of tRNAs from a common pair of short hairpins with complementary anticodons logically implies the similar cohesive origin for their RNA synthetases. If so, the ribozymes had to be equally distributed on the two complementary classes. We think that the two classes of protein aaRSs simply inherited the two classes of equal distributions (32:32 for anticodons and 10:10 for amino acids) from their RNA forerunners. Among numerous data in support of this viewpoint, the most notable is a parallel observed for ribozymes and aaRSs. First, the *Tetrahymena* group I self-splicing intron uses molecular handles on the RNA minor groove similar to those used by AlaRS from class II when the latter recognizes the $tRNA^{Ala}$ acceptor stem (35). Second, the member of class I TyrRS and the P5abc ribozyme subdomain both recognize the group I intron's catalytic core (36). Moreover, the above convergence suggests that originally the group I intron's catalytic core had tRNA-like structure (37). Third, the group I introns are able to directly and specifically bind Arg right to its double-stranded codon–anticodon pair within the conserved P7 helix (12).

Sequence diversity at 1–3 positions of the modern acceptors is certainly insufficient to reliably discriminate all 20 amino acids (22). A statistical survey of primary tRNA sequences in the hope of uncovering at least weak correlation with aaRS class membership also failed (38). The two assignments of amino acids—to anticodons and to the aaRS classes—are not ideally connected as well (27). The list of idiosyncratic relations between the genetic code and the two adaptor molecules can be continued. Nevertheless, the correlation becomes evident only when we focus upon the complementary relations between the habitually single-stranded anticodons, their aminoacyl-carrying double-stranded predecessors and the two classes of aaRSs. Therefore we conclude that the two codes for amino-acylation of tRNSs and codon–anticodon interaction did not arise independently, but rather the two were originally one and the same.

We thank Profs. P. Schimmel and E. Szathmary for interest and useful comments.

1. Hopfield, J. J. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4334–4338.
2. Shimizu, M. (1982) *J. Mol. Evol.* **18,** 297–303.
3. Szathmary, E. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9916–9920.
4. Weiner, A. M. & Maizels, N. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7383–7387.
5. Bloch, D., McArthur, B. & Mirrop, S. (1985) *BioSystems* **17,** 209–225.
6. Schimmel, P., Giege, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8763–8768.
7. Normanly, J. & Abelson, J. (1989) *Annu. Rec. Biochem.* **58,** 1029–1049.
8. Hou, Y.-M. & Schimmel, P. (1988) *Nature (London)* **333,** 140–145.
9. De Duve, C. (1988) *Nature (London)* **333,** 117–118.
10. Crick, F. H. C. (1968) *J. Mol. Biol.* **38,** 367–380.
11. Woese, C. R. (1967) *The Origins of the Genetic Code* (Harper & Row, New York).
12. Yarus, M. (1988) *Science* **240,** 1751–1758.
13. Yarus, M. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 205–217.
14. Majerfeld, I. & Yarus, M. (1994) *Nat. Struct. Biol.* **1,** 287–292.
15. Moller, W. & Janssen, G. M. C. (1990) *Biochimie* **72,** 361–368.
16. Moller, W. & Janssen, G. M. C. (1992) *J. Mol. Evol.* **34,** 471–477.
17. Rodin, S., Ohno, S. & Rodin, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4723–4727.
18. Rodin, S., Ohno, S. & Rodin, A. (1993) *Origins Life Evol. Biosphere* **23,** 393–418.
19. Steinberg, S., Misch, A. & Sprinzl, M. (1993) *Nucleic Acids Res.* **21,** 3011–3015.
20. Di Giulio, M. (1995) *Origins Life Evol. Biosphere* **25,** 549–564.
21. Miller, S. L. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 17–27.
22. Saks, M. E. & Sampson, J. R. (1995) *J. Mol. Evol.* **40,** 509–518.
23. Okimoto, R. & Wolstenholme, D. R. (1990) *EMBO J.* **9,** 3405–3411.
24. Hartman, H. (1995) *J. Mol. Evol.* **40,** 541–544.
25. Osawa, S. & Jukes, T. H. (1989) *J. Mol. Evol.* **28,** 271–278.
26. Rodin, S. & Ohno, S. (1995) *Origins Life Evol. Biosphere* **25,** 565–589.
27. Eriani, G., Cavarelli, J., Martin, F., Ador, L., Rees, B., Thierry, J.-C., Gangloff, J. & Moras, D. (1995) *J. Mol. Evol.* **40,** 499–508.
28. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203–206.
29. Ruff, M., Kishnaswamy, S., Boeglin, M., Poterzman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991) *Science* **252,** 1682–1689.
30. Moras, D. (1992) *Trends Biochem. Sci.* **17,** 159–164.
31. Benner, S. A., Cohen, M. A., Gonnet, G. H., Berkowitz, D. B. & Johnsson, K. P. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 27–70.
32. Wong, J. T.-F. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1909–1912.
33. Wong, J. T.-F. (1981) *Trends Biochem. Sci.* **6,** 33–36.
34. Szathmary, E. & Zintzaras, E. (1992) *J. Mol. Evol.* **35,** 185–189.
35. Strobel, S. A. & Cech, T. (1995) *Science* **267,** 675–679.
36. Mohr, G., Caprara, M. G., Guo, Q. & Lambowitz, A. (1994) *Nature (London)* **370,** 147–150.
37. Guo, G. & Lambowitz, A. (1992) *Genes Dev.* **6,** 1357–1372.
38. Nicholas, H. B. & McClain W. H. (1995) *J. Mol. Evol.* **40,** 482–486.