

# ON CONCERTED ORIGIN OF TRANSFER RNAs WITH COMPLEMENTARY ANTICODONS

SERGEI RODIN,\* SUSUMU OHNO\*\* and ANDREI RODIN\*\*\*

\* *Institute of Cytology & Genetics, Siberian Branch of Russian Academy of Sciences, av. acad. Lavrentjeva 10, Novosibirsk 630090, Russia*

\*\* *Beckman Research Institute of the City of Hope, 1450 E. Duarte Road, Duarte, CA 91010-0269, U.S.A.*

\*\*\* *Center for Demographic and Population Genetics, University of Texas, P.O. Box 20334, Houston, TX 77225, U.S.A.*

(Received January 27, 1993)

**Abstract.** Pairs of antiparallely oriented consensus tRNAs with complementary anticodons show surprisingly small numbers of mispairings within the 17-bp- long anticodon stem and loop region. Even smaller such complementary distances are shown by illegitimately complementary anticodons, i.e. those with allowed pairing between G and U bases. Accordingly, we suppose that transfer RNAs have emerged concertedly as complementary strands of primordial double helix-like RNA molecules. Replication of such molecules with illegitimately complementary anticodons might generate new synonymous codons for the same pair of amino acids. Logically, the idea of tRNA concerted origin dictates very ancient establishment of direct links between anticodons and the type of amino acids with which pre-tRNAs were to be charged. More specifically, anticodons (first of all, the 2nd base) could selectively target 'their' amino acids, reaction of acylating itself being performed by another non-specific site of pre-tRNA or even by another ribozyme. In all, the above findings and speculations are consistent to the hypercyclic concept (Eigen and Schuster, 1979), and throw new light on the genetic code origin and associated problems. Also favoring this idea are data on complementary codon usage patterns in different genomes.

## 1. Introduction: Basic Premises

In the preceding paper (Rodin *et al.* 1993) we have introduced the hypothesis of paired emergence of primordial adapters with complementary anticodons supported by the results of consensus tRNA comparisons. Here, we describe this hypothesis in more detail and represent some new results and considerations that confirm and develop the main idea. The three basic premises promoted this research were as follows:

### 1.1. DINUCLEOTIDE RULES AND COMPLEMENTARY CODON - USAGE - PATTERN

As in all languages, any message in DNA (as well as RNA) language is written on the principle of nonrandomness (Air *et al.*, 1976; Grantham *et al.*, 1980, 1981; Nussinov, 1980, 1981, 1987; Blaisdell, 1985; Gouy, 1987; Ohno, 1987, 1988a, b; Trifonov, 1987; and many others). At the dinucleotide level of modern DNA construction, the universal grammatical rule has been formulated (Ohno, 1988b, 1991; Yomo and Ohno, 1989; Ohno and Yomo, 1991). This rule states that two dinucleotides, CG and TA, are consistently under represented and that these

\* Present address: Departamento de Sistemas Complejos, Instituto de Fisica, Universidad Nacional Autonoma de Mexico, Apdo Postal 20-364, Alvaro Obregon, 01000 Mexico D.F., Mexico.

## GENETIC CODE ( CODONS )

	5'	T	C	A	G	3'
		TTT PHE	<u>TCT</u> SER	<u>TAT</u> TYR	<u>TGT</u> CYS	T
		TTC PHE	TCC SER	<u>TAC</u> TYR	<u>TGC</u> CYS	C
T		<u>TTA</u> LEU	<u>TCA</u> SER	<u>TAA</u> TERM	<u>TGA</u> TERM	A
		<u>TTG</u> LEU	<u>TCG</u> SER	<u>TAG</u> TERM	<u>TGG</u> TRP	G
		<u>CTT</u> LEU	<u>CCT</u> PRO	<u>CAT</u> HIS	<u>CGT</u> ARG	T
		<u>CTC</u> LEU	CCC PRO	<u>CAC</u> HIS	<u>CGC</u> ARG	C
C		<u>CTA</u> LEU	<u>CCA</u> PRO	<u>CAA</u> GLN	<u>CGA</u> ARG	A
		<u>CTG</u> LEU	<u>CCG</u> PRO	<u>CAG</u> GLN	<u>CGG</u> ARG	G
		ATT ILE	<u>ACT</u> THR	AAT ASN	<u>AGT</u> SER	T
		ATC ILE	ACC THR	AAC ASN	<u>AGC</u> SER	C
A		<u>ATA</u> ILE	<u>ACA</u> THR	AAA LYS	<u>AGA</u> ARG	A
		<u>ATG</u> MET	<u>ACG</u> THR	<u>AAG</u> LYS	<u>AGG</u> ARG	G
		GTT VAL	<u>GCT</u> ALA	GAT ASP	GGT GLY	T
		GTC VAL	GCC ALA	GAC ASP	GGC GLY	C
G		<u>GTA</u> VAL	<u>GCA</u> ALA	GAA GLU	GGA GLY	A
		<u>GTG</u> VAL	<u>GCG</u> ALA	<u>GAG</u> GLU	GGG GLY	G
		Np, m	S	P, m	J	

Fig. 1. The genetic code with marked excessive (solid line) and deficient (wave line) dinucleotides. Punctuated is the AG dinucleotide, complementary to the excessive CT (see comment to Table I). Only one codon, CTA of leucine, is composed of excessive (CT) and deficient (TA) dinucleotides. Boxed are leucine codons and their complementary partners illustrating the fact that one amino acid is complementarily converted into different ones.

Np - nonpolar amino acids, P - polar amino acids, s - small size of amino acid residue, m - middle size, j - 'junk' column, including codons for amino acids of extreme features: smallest Gly, most reactive Cys, geometrically bulky Arg, as well as two additional codons for serine (unlinked to the main Ser tetrad by one substitution), and UGA used as the second Trp codon in mitochondria.

deficiencies are compensated by over representation of three dinucleotides, TG, CA and CT if we compare their frequencies with what is theoretically expected from given base composition. These phylogenetic preferences may have different bases. Of those, thermodynamic parameters favoring, for example, UG over GU independently of nearest neighbors in helices formed by RNAs (He *et al.*, 1991) are of particular interest with regard to origin of tRNAs.

Each trimer may be considered, in turn as a combination of two overlapping dimers. Consequently, CTG which is a combination of two excessive dimers, CT and TG, becomes the most over represented base trimer in coding as well as noncoding sequences (Ohno, 1988b; Yomo and Ohno, 1989; Ohno and Yomo, 1990, 1991). Not surprisingly, of 6 codons for Leu, CTG is the preponderant one. Accordingly, all 64 codons of the genetic code are marked in Figure 1 with regard to the frequency of dimers they contain. Remarkably, there is only one instance (codon CTA) combined by excessive (CT) and deficient (TA) dimers.

What is more intriguing is that both under represented dimers, CG and TA, are palindromes, and that two of the three excessive dimers, TG and CA, are complementary to each other. The group of CT and of moderately used dimers includes two self-complementary palindromes (GC and AT) and five pairs of dimers complementary to each other (Table I). It is especially worth noting that no complementary pair is formed by an excessive dimer and a deficient one. As a consequence of the above, DNA sequence of sufficient length is expected to contain nearly equal numbers of trimers complementary to each other (Ohno, 1991; Ohno and Yomo, 1991). Indeed, as it was first shown for the sample of *E. coli* 52 genes, glutamine CAG codon and its complementary partner, leucine CUG, appeared to be represented by nearly equal numbers (Alff-Steinberger, 1984). More recently, the observation has been made (Ohno, 1991; Ohno and Yomo, 1991) that, among all 64 base trimers ordered according their abundance, just complementary trimers occupy very often consecutive and closest ranks. This was attributed to the palindrome symmetry occurring within each strand as well as between two complementary strands of DNA providing a capability for both these strands to encode homologous oligopeptides (Yomo and Ohno, 1989; Ohno, 1991; Ohno and Yomo, 1991). Some considerations (*ibid*) furnish evidence that the abundance in palindromes is very ancient acquisition of genetic language so that even the universal assignment of 61 codons to 20 amino acids might be optimized to this pattern of primordial sequences.

If so, in the Genetic Code table itself, one can expect to find arguments in favor of the above supposition. Indeed, for every codon, its complementary partner does contain dinucleotides of the same degree of usage in modern DNA (and RNA) sequences, i.e. the complementary reflection of the genetic code table when codons of 1st (2nd) column transpose to those of 3rd (4th) (and *vice versa*) appears to be invariant to the universal grammatical rule and hence, to the corresponding codon usage (Figure 1).

TABLE I  
Palindromic relations intra- and inter-dinucleotides<sup>a</sup>

Deficient	$\begin{array}{c} \text{C G} \\         \\ \text{G C} \end{array}$		$\begin{array}{c} \text{T A} \\       \\ \text{A T} \end{array}$	self complementary
Excessive	$\begin{array}{c} \text{T G} \\        \\ \text{A C} \end{array}$	or	$\begin{array}{c} \text{C A} \\        \\ \text{G T} \end{array}$	
Excessive and moderate <sup>b</sup>	$\begin{array}{c} \text{C T} \\        \\ \text{G A} \end{array}$	or	$\begin{array}{c} \text{A G} \\        \\ \text{T C} \end{array}$	
Moderate	$\begin{array}{c} \text{G C} \\         \\ \text{C G} \end{array}$		$\begin{array}{c} \text{A T} \\       \\ \text{T A} \end{array}$	self complementary
	$\begin{array}{c} \text{T C} \\        \\ \text{A G} \end{array}$	or	$\begin{array}{c} \text{G A} \\        \\ \text{C T} \end{array}$	} homodimers
	$\begin{array}{c} \text{G T} \\        \\ \text{C A} \end{array}$	or	$\begin{array}{c} \text{A C} \\        \\ \text{T G} \end{array}$	
	$\begin{array}{c} \text{A A} \\       \\ \text{T T} \end{array}$	or	$\begin{array}{c} \text{T T} \\       \\ \text{A A} \end{array}$	
	$\begin{array}{c} \text{C C} \\         \\ \text{G G} \end{array}$	or	$\begin{array}{c} \text{G G} \\         \\ \text{C C} \end{array}$	

<sup>a</sup> There is a single case of palindromes formed by excessive (CT) and moderate (AG) dinucleotides and none formed by excessive and deficient dimers.

<sup>b</sup> Furthermore, AG dinucleotide is also over represented, though not so considerably as its complementary partner, dinucleotide CT (Cocho, personal communication).

## 1.2. COMPLEMENTARY CODONS AND PROPERTIES OF THEIR AMINO ACIDS

Complementary reflection of the genetic code table columns links, in pairs, amino acids with contrasting properties such as size, shape, polarity, etc (Figure 1). This design looks reasonable at least because most modern proteins are nearly neutral in charge. Furthermore, from the very beginning of the genetic code evolution, it could have been selectively advantageous to fix alternating arrays of hydrophilic and hydrophobic amino acids which in turn define important secondary elements of proteins such as  $\beta$ -pleated sheets, slightly more complicated requirement having to be introduced in the case of  $\alpha$ -helix formation (Fitch and Upper, 1987; Ohno, 1987). The corresponding codons are of NRN and NYN configuration, respectively, where the 2nd bases are complementary to each other.

## 1.3. SYMMETRIC CLOVER-LEAF TOPOLOGY OF tRNAs

In a context of the above, what looks particularly impressive is perfect internal symmetry of tRNA when the anticodon is located almost precisely in the middle of the sequence, thus a uniform clover-leaf secondary structure being formed. Excluding its T $\psi$  – and D – arms and a variable segment between T $\psi$  – and anticodon stem, each sequence of tRNA is in fact a palindrome. Because of this symmetry, the complementary ('minus') image of any tRNA sequence appears to be very similar to 'plus' one, having, however, complementary anticodons (Eigen and Schuster, 1979).

All of the above led us to the main question: What if the first adapters had emerged by the shortest possible way: two complementary strands of a double helix becoming tRNA-like sequences with complementary anticodons?

Earlier Eigen and Schuster (1979) were tempted by the same possibility. They compared plus- and minus-chains of several pairs of modern tRNAs with complementary anticodons. Their results obtained on tRNAs of *E. coli* origin were negative in the sense that when two chains with complementary anticodons were both plus or minus type, less number of differences between them was registered than when one, plus, chain (written in 5'  $\rightarrow$  3' way) was compared with another, minus, chain (written in 3'  $\rightarrow$  5' way). Accordingly, Eigen and Schuster concluded that if even, initially, some of such tRNA pairs had emerged as independent replication units, subsequently they have lost their primordial similarity due to formation of first linkage groups, ancestral gene duplication and inevitable mutational divergence of daughter copies.

Yet, compared to 1979, vastly more sequenced tRNA sequences derived from a very wide range of species are available today. It was felt, therefore, that the above noted intriguing possibility should be reconsidered. We have compared only anticodon loops and stems of tRNAs rather than their complete length, since the latter condemns the results to be negative due to the smoothing influence of hypervariable positions, particularly of those in the aminoacyl stems (Rodin, 1991; Rodin and Kuznetsov, 1993). However, just for this reason, aminoacyl stem sequences

served as internal control. Two other (lateral) arms of the tRNA clover-leaf, D and T $\psi$ , include 5' and 3' control sites for eucaryotic RNA polymerase III, respectively, of which the site within T $\psi$  - loop and stem being particularly conservative. In turn, D-arms, especially of mitochondrial tRNAs, appeared to vary in their length often so considerably that can not be aligned unambiguously. As a first approximation, therefore, both D- and T $\psi$ -regions were disregarded in our analysis. Moreover, because of interference from mutational divergence of tRNAs, we have preferred to deal with consensus sequences representing main kingdoms.

## 2. Material and Methods

To process tRNA sequence data, the package VOSTORG designed for sequence analysis and phylogenetic trees building (Zharkikh *et al.*, 1991), and proprietary (A. Rodin) EMBL/GENBANK interface software have been used. In total, 896 tRNA gene sequences drawn from EMBL/GENBANK data bases and the compilation (Spizel *et al.*, 1989) have been collected.

The sequences of this set fell into 50 anticodon subsets. Of those, only 38 subsets consisting of a minimum 13 (Leu tRNA with anticodon CAG) to a maximum 37 (Met, CAT) tRNA genes, contained representatives of main kingdoms: archaebacteria, eubacteria, eukaryotes, mitochondria and chloroplasts. In turn, of these 38, only 24 subsets were sufficiently representative with regard to both complementary anticodons. Nevertheless, we also considered the remaining 14 anticodons, since there were 'quasi complementary' partners for them with one or more irregular bonds of G-U type within anticodon triplet. For example, one of six Leu anticodons, 5' UAG 3', was considered as capable of binding with the glutamine 5' CUG 3' and 5' UUG 3'.

In a few cases, 'synonymous' anticodons were taken together to give more reliable consensus reconstruction to the subsets. This provided additional six consensus tRNAs with anticodons having Y (pyrimidine: U or C) or R (purine: A or G) at 5' first position: Leu (YAG), Ser (RGA), etc. In all, 44 subsets were suitable to construct the corresponding consensus sequences (their full list is given in Rodin *et al.*, 1993).

Transfer RNAs, as evolutionary very old sequences, have accumulated so many mutations that their genealogical relationship between their families is not described by a tree-like graph. Quite the opposite, they represent an apparent 'bundle-like' phylogeny (Eigen *et al.*, 1988, 1989). Moreover, as we shall show further, it is the most early stage of tRNA evolution that unlikely can be approximated by an ordinary tree-like diagram in principle. Yet, computer simulations showed that the consensus for each tRNA family was close to the 'true' evolutionary precursor of that family (Eigen *et al.*, 1988, 1989). Therefore, instead of so doubtfully identified ancestral sequences, we have preferred to compare the consensus tRNAs.

The relatively short length, uniform clover-leaf secondary structure, and, hence, a possibility to take into account compensating base substitutions within stem regions,

occurrence of modified nucleotides in fixed position; all these characteristics facilitated an alignment of tRNAs. Nevertheless, for super control, we used Needleman-Wunch algorithm for aligning a pair of tRNAs (routine ALSEQ of VOSTORG package) in combination with the multiple alignment construction and manual editing (programs MULTAL and SEQED) of all the tRNAs aligned within each anticodon subset.

Consensus sequence reconstruction did not face any obstacles. In each position, the consensus records the most abundant base. When two or more bases occurred with equal (or nearly equal) frequencies, the multi potential symbols were assigned to such ambiguous positions (see, for example, Figures 2 and 3). However, such instances were surprisingly rare. Besides, in paired (stem) segments, we reduced the ambiguity more often than not following the requirement for bases to be complementary to each other, i.e. those symbols that maintained the internal chain complementarity have been chosen as more probable. As 'average', consensus tRNAs we built had assumed the stable structure except for Val (CAC and UAC), Clu (CUC), and Cys (GCA). The latter was particularly sensitive to sequence numbers and errors.

To count evolutionary distances between RNA sequences, base pairs within stem regions were considered as one position in (Eigen *et al.*, 1988, 1989). Such an approach seems motivated by analytical and computer simulations (Kimura, 1985). However, more well-grounded phylogenetic tree analyses showed that paired differences of this kind in stems appeared to have been fixated sequentially via rather long period of 'waiting for' the second substitution that compensated the former (Hancock *et al.*, 1988; Rodin, 1991; Rodin and Kuznetsov, 1993). Therefore, for both, direct and complementary distances (see below), we counted every such paired position as two independent ones.

### 3. Results and Discussion

#### 3.1. tRNAs WITH COMPLEMENTARY ANTICODONS: COMPLEMENTARY DISTANCE VERSUS DIRECT DISTANCE, AND ANTICODON LOOP AND STEM REGION VERSUS ACCEPTOR REGION

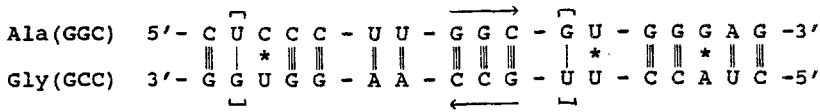
For every pair of anticodons, their consensus tRNAs were oriented in an antiparallel ('head-to-tail') manner. Then we calculated a number of base mispairings within: (1) anticodon arm (loop plus stem positions) and (2) acceptor stem, more variable and therefore serving as an internal control region. Very different results were obtained for these two domains of tRNA molecules.

##### 3.1.1. *Anticodon Loop and Stem*

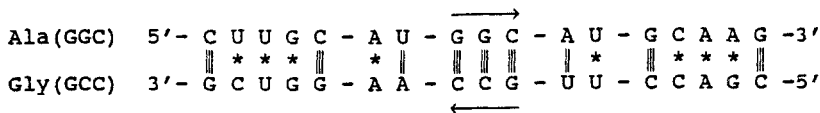
Figure 2a shows an example of surprisingly small number of mispairings within anticodon stem and loop region for two anti-parallelly oriented consensus tRNAs with anticodons GGC (Ala) and GCC (Gly). Similarly high degree of complemen-

## A) Anticodon Stem

## Consensus sequences

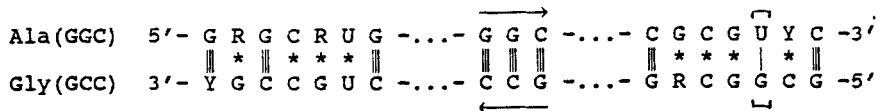


## B) Anticodon Stem

*E. coli*

## C) Acceptor Stem

## Consensus sequences



## D) Acceptor Stem

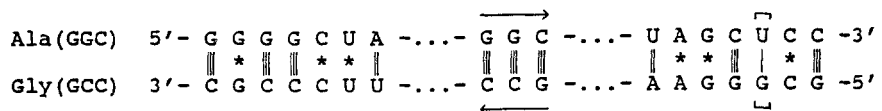
*E. coli*

Fig. 2. In all comparisons, regular base pairing are indicated by 3 lines for the G-C bond, 2 lines for the A-U bond, while only 1 line is used for illegitimate G-U pair. An asterisk is inserted between unmatched bases. Y denotes pyrimidine, U or C. R denotes purine, A or G. (A) Antiparallel comparison of consensus anticodon loops and stems between Ala tRNA (GGC) and Gly tRNA (GCC). (B) The same comparison between the same tRNAs derived from *E. coli*. (C) The antiparallel comparison of consensus acceptor stems between the same tRNAs as in A and B. (D) The same comparison as (C) but involving tRNAs derived from *E. coli*.

tarity was found in most of the other such pairs of consensus tRNAs with complementary anticodons. As far as the anticodon stem was concerned, an average number of mismatches for all pairs of consensus tRNAs (excluding a few unreliable ones) was 2.8 per 10 positions.

All these comparisons were integrated in the matrix of 'complementary distances'



TABLE II

Average distances (per the anticodon arm region) of consensus tRNAs representing 1st and 2nd columns of the table of anticodons<sup>a</sup>

Column	$\bar{D}_c$ complementary distance			$\bar{D}_d$ direct distance					
	III	IV	III+IV	I	II	I+II	III	IV	III+IV
Consensus tRNAs									
1-st column:									
1. Phe (GAA)	7.1	9.4	8.1	8.7	10.4	9.3	10.4	9.6	10.1
2. Leu (CAA)	7.2	8.8	7.9	8.1	10.2	8.9	9.2	9.0	9.1
3. Leu (UAA)	8.6	9.0	8.8	8.7	10.2	9.3	9.7	8.8	9.3
4. Leu (YAA)	7.2	8.6	7.9	7.8	9.2	8.3	9.1	8.8	9.0
5. Leu (CAG)	5.9	7.4	6.6	8.5	10.3	9.2	10.8	11.1	11.0
6. Leu (UAG)	5.3	7.3	6.2	8.1	9.7	8.7	8.4	9.2	8.8
7. Leu (YAG)	5.6	7.7	6.6	8.0	9.0	8.4	8.8	9.6	9.2
8. Leu (GAG)	5.3	7.3	6.2	9.2	9.1	9.2	9.5	10.5	10.0
9. Ile (GAU)	9.8	10.9	10.3	10.8	10.7	10.8	11.2	11.4	11.3
10. Met (CAU)	6.2	8.5	7.3	9.3	10.6	9.8	8.3	9.5	8.7
11. fMet (CAU)	6.6	8.5	7.4	8.5	10.6	9.3	9.0	9.3	9.2
12. Val (GAC)	7.5	9.0	8.2	11.1	10.9	11.0	10.2	11.1	10.6
13. Val (GAC)	7.3	7.7	7.5	10.1	11.3	10.6	10.5	10.9	10.7
14. Val (UAC)	6.7	7.4	7.0	10.0	8.8	9.6	8.7	9.1	8.9
2-nd column									
15. Ser (UGA)	6.6	7.1	6.8	9.6	9.7	9.7	8.4	8.9	8.7
16. Ser (GGA)	6.2	6.8	6.5	9.9	8.6	9.4	9.9	10.0	10.0
17. Ser (RGA)	7.6	7.2	7.4	9.9	9.5	9.8	9.5	9.5	9.5
18. Pro (UGG)	7.1	6.3	6.7	10.3	8.1	9.6	10.8	10.2	10.6
19. Pro (GGG)	6.9	6.9	6.9	10.5	8.4	9.8	11.0	11.1	11.0
20. Thr (UGU)	6.2	6.8	6.5	9.6	9.5	9.6	8.2	8.5	8.4
21. Ala (UGC)	6.7	6.3	6.5	10.0	8.4	9.1	8.8	8.8	8.8
22. Ala (GGC)	7.2	7.2	7.2	10.9	8.6	10.1	11.1	11.2	11.1

<sup>a</sup> Similar results are obtained for consensus tRNAs representing 3rd and 4th columns of the table of anticodons.

(see Rodin *et al.*, 1993: Table I), calculated on the basis of legitimate bonds A-U, G-C, plus the illegitimate 'wobbling' G-U bonds. In contrast, standard (direct) distances even between consensus tRNAs with 'uncomplementary' anticodons have surpassed on average the complementary ones (Table II). Yet, since the anticodons three bases were also included in distance evaluation, such comparisons were *a priori* favoring complementary distances. Therefore, clearly more impressive comparisons of complementary distances versus direct ones are shown in Table III, where the most proving is that mean number of mispairings between complementary anticodons appears to have been nearly twice less than mean number of mismatches between anticodons differing by only their 3rd base, that is the same as by 1st base of codons.

TABLE III  
Mean values ( $\bar{x}$ ) and standard deviation ( $6x$ ) of distances between consensus tRNAs

	Anticodon stem and loop		Acceptor stem <sup>a</sup>	
	$\bar{x}$	$6x$	$\bar{x}$	$6x$
I. Complementary distances				
All pairs of complementary anticodons <sup>b</sup>				
1st vs 3rd column: (1 × 3)				
max	4.37	2.02	7.44	1.71
min	3.9	1.56	7.68	1.78
(2 × 4)				
max	4.9	1.7	7.52	1.86
min	4.6	1.39	7.1	1.87
Only legitimate pairs:				
(1 × 3) + (2 × 4)				
max	5.9	1.97	7.67	2.21
min	5.1	1.3	8.07	2.06
Only illegitimate pairs:				
(1 × 3)	3.75	1.67	7.24	1.66
(2 × 4)	4.11	1.2	6.75	1.53
(2 × 3)	5.17	1.46	7.52	1.86
II. Direct distances (min)				
Pairs connected by one substitution in 3rd position of anticodons				
(1 × 1)	9.67	1.65	9.17	1.89
(2 × 2)	7.61	3.33	8.82	1.75
(3 × 3)	8.43	1.24	8.55	2.39
(4 × 4)	9.2	2.18	9.0	1.0
Pairs connected by one substitution in 2nd position of anticodons <sup>c</sup>				
(1 × 2)	7.5	1.73	8.2	1.47
(1 × 4)	7.7	2.17	9.5	2.87
(3 × 4)	8.0	2.07	7.7	1.49
(1 × 3) + (2 × 3) + (2 × 4) <sup>d</sup>	7.64	1.94	8.16	1.98

<sup>a</sup> For acceptor stem comparisons, the anticodon three bases have been included in order to make the results comparable with those for anticodon stem and loop; thus having 17-base-long region in both cases.

<sup>b</sup> For complementary distances, smaller  $\bar{x}$  values ('min') are obtained when unreliable consensus tRNAs were excluded; those for Val(CAC, UAC), Glu(CUC) and Cys(GCA).

<sup>c</sup> Relatively small mean values are probably due to secondary complementary relations connecting these anticodons (see text for detail).

<sup>d</sup> The pairs of complementary anticodons were excluded.

Likely because of mutations accumulated, the same tRNA pairs of *E. coli* do not display the above complementarity within the anticodon stem and loop region (Figure 2b). Anticodon stems (10 positions in sum) of their pairs had 7.2 mismatches on the average. However, in certain tRNA pairs (strikingly derived from primates) was also observed rather small number of mispairings comparable to that shown by consensus sequences. What is more, a few 'mixed' pairs, when consensus tRNA

were compared with modern ones, also displayed good complementarity in the anticodon stem (not shown).

Thus, with the anticodon triplets ideally complementary to each other, such pairs of antiparallely oriented tRNAs look like quite stable double helices capable of faithful replication; both strands serving as effective templates. Long ago we noticed conservation of two bases adjacent to the anticodon triplet (usually U from 5' and A from 3' end) can be attributed, among other reasons, to a very ancient constraint to facilitate better pairing between pre-tRNAs with complementary anticodons.

Therefore, at least with respect to anticodon domain of tRNAs, the verdict of Eigen and Schuster (1979) seems a bit premature. However, we do consider here extremely old sequences. It should be concluded, therefore, that the anticodon arm was under permanent strong pressure of negative selection inasmuch as the above complementary pattern was revealed not only for majority of consensus but even for some modern pairs of tRNAs with complementary anticodons. In accordance with the conclusion is that the all but one position of anticodon stem belong to the moderately variable class (Eigen *et al.*, 1989).

### 3.1.2. *Acceptor Stem*

As to more variable acceptor stem, none of the pairs of consensus tRNAs with complementary anticodons, both legitimate and illegitimate, showed the symmetrical pattern revealed for the anticodon stem (Figure 2c). The matrices of complementary and direct distances built for the acceptor stem also showed no essential difference. This is most convincingly reflected in Table III so that with regard to the acceptor stem, one may say, in the utmost, only about vague tendency for direct distances to exceed those of complementary ones.

This was expected. Indeed, following Eigen *et al.* (1989), each consensus tRNA was rewritten into its binary (R, Y) version, and then for each *i*-th site of the acceptor stem, we have computed the number of  $\delta_i$  bases differing from the consensus base. As a result, except for the 1st and its pairing partner, 72nd site, the other 12 positions in acceptor stem appeared to have  $\delta_i \cong 0.5n$ , (*n* is a number of sequences). This means in fact nearly complete randomization of divergence within this part of tRNA genes (Eigen *et al.*, 1989; Rodin, 1991).

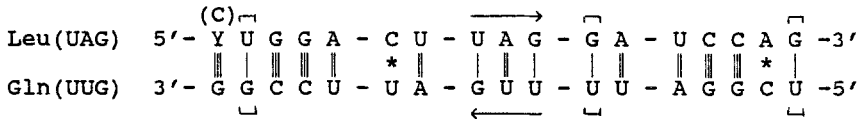
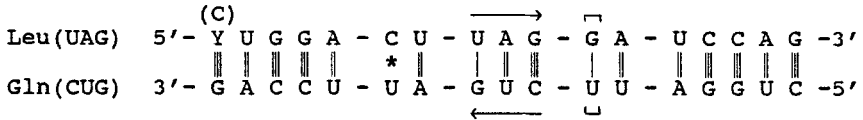
### 3.2. TRANSFER RNAs WITH QUASI-COMPLEMENTARY ANTICODONS

More often than not, for pairs of anticodons with an allowance of G-U binding, their consensus tRNAs showed even more perfect double helical-like structure compared to that for 'legitimate' pairs (Table III). For example (Figure 3a), as far as the 5-base-pair-long anticodon stem was concerned, the complementarity between Leu (UAG) and Gln (CUG) tRNAs was perfect. Only in the anticodon loop, was there one U\*C mispairing and one illegitimate G-U pairing. The complementarity between Leu (UAG) and Gln (UUG) was not bad either, showing one A\*C mispairing on the right arm of the stem and one C\*U mispairing and

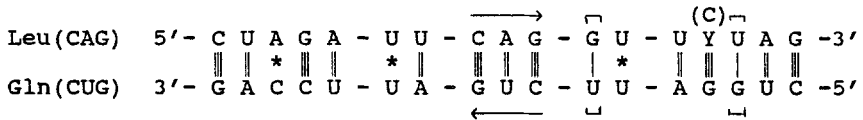
## A) Anticodon Stem

## Consensus sequences

## Illegitimate pairs:



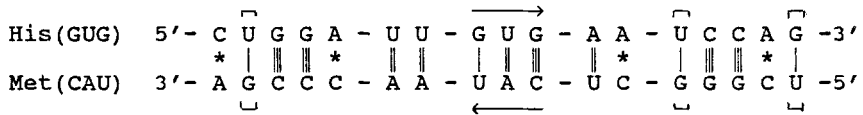
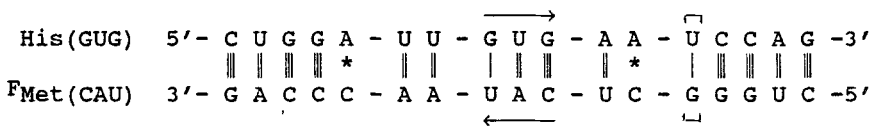
## Legitimate pair:



## B) Anticodon Stem

## Consensus sequences

## Illegitimate pairs:



## Legitimate pair:

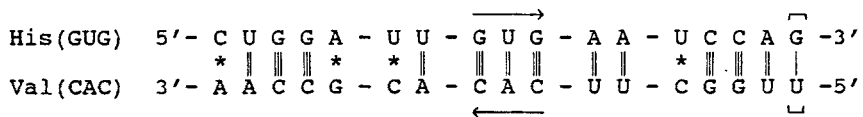


Fig. 3. The comparisons like that in Figure 2A, except that anticodons of each pair of tRNAs compared are not necessarily complementary unless one or two illegitimate G-U pairing is allowed (for comments see text).

one G-U pairing in the loop. So, complementarity between Leu (UAG) and Gln (CUG) tRNAs as an illegitimate pair surpassed that between Leu (CAG) and Gln (CUG) tRNAs that form the legitimate pair (Figure 3a).

Another instance of quasi-complementary anticodon pair involving G-U bond showing better complementarity than the legitimate pair is shown in Figure 3b. Between Val (CAC) and His (GUG) consensus tRNAs that form a legitimate pair, there were three mispairings and one G-U pair in the stem and one mispairing in the loop. Greater complementarity was seen for an illegitimate His (GUG) and Met (CAU) pair. Particularly striking was that between consensus His (GUG) and formyl-Met (CAU) tRNAs which is for the translation initiation in procaryotes, mitochondria and chloroplasts. Figure 3b shows that as far as the 5-bp-long stem was concerned, there was only one A\*C mispairing on the right arm and another A\*C was the only blemish on the loop. With regard to Val (CAC) tRNA, a legitimate partner of His (GUG) tRNA, it should be noted also that it is its GUG codon which is recognized sometimes by formyl-Met (CAU) tRNA as a starting codon of translation (Lewin, 1990). Perhaps, this coincidence is not random.

### 3.3. ON POSSIBLE CONCERTED EMERGENCE OF PRE-tRNAs FOR SYNONYMOUS CODONS

One of the interesting consequences of the above is that if we let the illegitimate pair to be replicated, it will soon produce a pair of 'daughter' sequences with new legitimately complementary anticodons. For example, the Leu (UAG) - Gln (CUG) pair shall yield a new pair with anticodons CUA and CAG. While the former is no longer of any use, UAG being a chain terminator, the latter shall recognize CUG which is the preponderant Leu codon. The latter, in turn, should form a complex with its legitimate complementary partner which is Gln tRNA with anticodon CUG. Similarly, the pair Leu (CAG) - Gln (UUG) generates again CUG (Gln) and CAA ('the novel' anticodon for Leu), while the Leu (UAG) - Gln (UUG) pair gives CUA and CAA, etc. It seems reasonable therefore to consider a capability of modern tRNA with UAG anticodon to recognize all six Leu codons (Lewin, 1990) as the unique trace marking back to its origin.

Thus, in the absence of any mutations, exclusively due to illegitimate 'wobbling' pairing of two quasi-complementary anticodons, Leu and Gln could gain novel synonymous triplets encoding them. In general, it is easy to conceive somewhat like a population of relative sequences capable of forming double helices with precisely or quasi-complementary anticodons as the most ancient prototypes of modern adapters, in accordance with the concept of molecular quasi-species (Eigen and Schuster, 1979).

But with what sequences has the process started? Let us assume that at the beginning the key role belonged to the sequences with anticodon having in its 5' and 3' sites complementary bases. Indeed, it is enough to assume that because of the internal complementarity of stem segments, even two identical sequences with such 'palindromic' triplet were probably capable of building the double helices, and then the subsequent acts of replication could produce the primary pool of complementary adapters.

In accordance with the above simple scheme of synonymous tRNA origin, a reduced complementarity should be expected for amino acids both encoded by small (of 1-2 members) series of synonymous triplets. The 10 mispairings detected for the pair Phe (GAA) – Glu (UUC) may serve an example (see Rodin *et al.*, 1993).

### 3.4. THE HYPOTHESIS OF DIRECT INTERACTION BETWEEN ANTICODON AND AMINO ACID IN FAVORABLY FOLDED PRIMORDIAL tRNAs

The concerted origin of tRNAs implies for their both, (+) and (-) strands to share very similar sequences except for anticodon three bases. Furthermore, for pairs of pre-tRNAs with palindromic anticodons like 5'CGG 3' – 5'CCG 3', 5'CAG 3' – 5'CUG 3' etc, only one difference of real principle, in the 2nd site of anticodon, is expected to have occurred originally. The starting repertoire of amino acids instructively involved into primitive translation was most probably smaller than today and mirrored their relative contents in the 'primordial soup'. In turn, yields of amino acids in the experiments imitating prebiotic synthesis correspond to the values expected from their chemical structure (Miller and Orgel, 1973; Eigen and Schuster, 1979). What seems rather eloquent is that of 24 palindromic anticodons (Figure 4), with the exception of two terminating UUA, UCA and for those captured by Arg and Pro, all the rest were assigned to amino acids produced as most frequent in the above experiments. First four places of this rank were occupied by Ala, Gly, Asp and Val, respectively (*ibid*). Remarkably, their tRNAs are subdivided on two pairs of legitimate complementary partners: Ala (5' GGC 3') – Gly (5' GCC 3') and Asp (5' GUC 3') – Val (5' GAC 3'), Ala (5' GGC 3') and Asp (5' GUC 3') being also in an illegitimate partnership. Besides, there are some other arguments in favor of just this group of four amino acids with their codons to be that with which evolution of translation started (for details, see: Eigen and Schuster, 1979). Logically, the above dictates very ancient establishment of direct links between anticodons and the type of amino acids with which pre-tRNAs were to be charged. If first adapters did emerge concertedly, they had to have identical anticodon and acceptor stems. Then, in an absence of specific synthetases, the primary recognition site on pre-tRNAs should have been their anticodons.

Furthermore, due to the same logic, it becomes clear why the central base of anticodons (and symmetrically, that of codons) is the most important in the genetic code. Indeed, first of all, it is the concerted origin of intra- and inter-complementary tRNAs-like hairpins which singles out the central base of palindromic trimers as 'future' anticodons located exactly in the center of each of these two sequences. In other words, the central base of codons and anticodons might become the most important 'simply' due to the concerted mechanism of tRNA origin. However, second, the specificity between the central base of anticodons and properties of the corresponding amino acids was most likely selected by their direct physico-chemical preferences. Third, one can see (Figures 1, 4) that complementary anticodons (and codons naturally) more often than not correspond to amino acids with contrasting features. Already emphasized in the Introduction, this design seems

very reasonable as providing the basic elements of secondarily ordered protein structure,  $\beta$ -pleated sheets and  $\alpha$ -helices. Again evidently, the concerted origin of tRNAs generates this pattern by the shortest way. Thus, the idea of concerted origin of tRNAs combines both specific physicochemical and evolutionary bases of the universal codon-to-amino acid assignment.

Nevertheless, the key point remains to clarify how exactly the anticodon region of pre-tRNA and amino acid might interact and select each other. The following findings may be relevant: (1) At least for arginine, an existence of rather specific amino acid binding site composed of ribonucleotides was shown and, by the way, this site preferred the L-amino acid (Yarus, 1988a); (2) The aminoacyl esterase activity was certainly demonstrated in ribozymes so that, by the principle of microscopic reversibility, primordial RNAs might also act like aminoacyl-tRNA-synthetases (Piccirilli *et al.*, 1992); (3) Ribozymes of unusually small size and trivial tRNA-like secondary structure have been reported recently (Pan and Uhlenbeck, 1992).

Accordingly, as in so many instances with RNAs, it seems rather tempting to speculate that first tRNAs might be able to combine their both informational (as adapters) and catalytic (as self-aminoacyl-synthetases) functions. To question more specifically, what if first tRNAs were able to fold on themselves to facilitate their own charging by amino acids? That is the anticodon (particularly its 2nd base) could provide a favorable targeting of amino acid as a substrate while the reaction of acylating itself could be performed either by another site of pre-tRNA or even by another non-specific ribozyme. At any way, emergence of first protein aminoacyl-tRNA-synthetases right as a product of the RNA world development, changed the situation radically and their coevolution with primordial transfer and ribosomal RNAs as well as with other protein components of the original translation machinery had to cause losing the properties speculated above for ancestral adapters.

Among indirect arguments in favor of self- or ribozyme-directed specific charging of ancestral tRNAs with amino acids, one of the most impressive is that the major site of many tRNAs for recognition by their synthetases is still its anticodon triplet accommodated by a specific binding pocket in these enzymes (Schulman and Pelka, 1988, 1989; Muramatsu *et al.*, 1988; Yarus, 1988b; RajBhandary, 1988; Rould *et al.*, 1991). When, for example, Met anticodon CAU was transplanted into *E. coli* tRNA<sup>Val</sup>, this chimeric tRNA became almost as good a substrate of MetRS (aminoacyl-tRNA synthetase) as the wild type Met tRNA, despite the fact that acceptor stems of tRNA<sup>Val</sup> and tRNA<sup>Met</sup> were quite different (Schulman and Pelka, 1988). This may be considered as one of peculiar relics indicating on starting period of the genetic code evolution.

### 3.5. SECONDARY COMPLEMENTARY PARTNERSHIP BETWEEN ANTICODONS

Among all complementary distances calculated there were included also a few cases of small number of mispairings shown by consensus tRNAs with anticodons that are not even quasi-complementary. The concept of complementary partners of partners should be introduced here (see also: Rodin *et al.*, 1993).

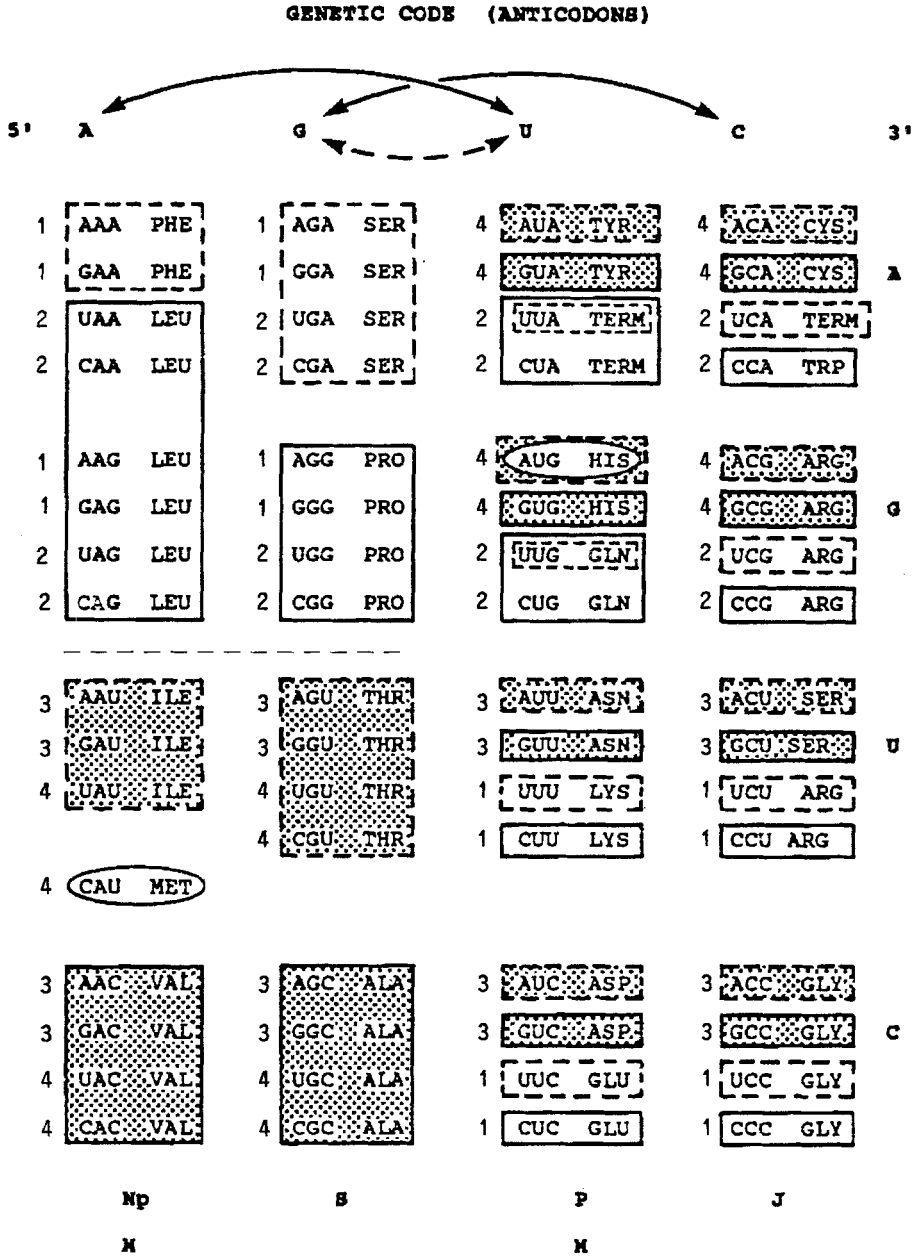


Fig. 4. Complementary partnership divides all 64 anticodons into two equal groups (shaded and unshaded, respectively), isolated from each other (shown by punctuated line for 1st and 2nd columns). Each group consists of two subgroups all 16 members of which may be obtained by the concerted origin of tRNAs for two replication cycles. Amino acid properties are abbreviated as in Figure 1.



All complementary relations between anticodons are schematically shown in Figure 4. It is easy to note that anticodons of 1st and 2nd columns differing from each other only by the 2nd base (A or G), as well as the corresponding amino acids, have the same list of complementary partners from the 3rd column. For short, such anticodons and amino acids will be called 'complementary twins', or simply 'twins'. One can see that Phe with Leu (1st column) and their twins Ser with Pro (2nd column) are complementarily linked with Gln, Lys, Glu and two 'stop' anticodons, UAA and CUA (Figure 4, unshaded group), while, similarly, Val, Ile, Met (and f-Met) from the 1st column appeared to be equivalent to Thr and Ala from the 2nd column in their complementary partnership that includes Tyr, His, Asn and Asp (Figure 4, shaded group).

The results of consensus tRNA comparisons are in a good agreement with complementary grouping of anticodons in Figure 4. Indeed, Tables II and III show that the 1st column is complementarily close to the 3rd column but not the 4th one, while the 2nd column is almost equally close to both its 'complementary' columns; the 4th (legitimately), and the 3rd (illegitimately).

Some concrete instances were especially significant (see also: Rodin *et al.*, 1993). For example, in 'upper' (unshaded in Figure 4) group Phe- and Leu- tRNAs from the 1st column as well as their 'twins', Ser- and Pro-tRNAs from the 2nd column, showed the lowest complementary distances with the same partners from the 3rd column. Relatively small direct distances were detected between these 'twins', too. However, these amino acids are considerably different in main properties. Although manifested not so distinctly, the similar tendency is observed for both complementary and direct distances between consensus tRNAs for 'lower' twin amino acids, Met, Val, and Thr, Ala, respectively. Quite the opposite, Phe- and Leu-tRNAs showed large direct distances with Ile-, Met- and Val-tRNAs in spite of the fact that all these amino acids are located in one and the same column (1st) and, therefore, share some common properties: nonpolarity, middle size, shape, etc.

Also as rather striking surprise was a very small complementary distance between Phe (GAA) from the 1st column and Arg (UCU) from the 4th column: only 3 mispairings. But, an expected small distance between Phe (GAA)'s twin Ser (GGA) of the 2nd column and Arg (UCU) also involved 3 mispairings. Exactly the same 'wrong' small complementary distance was seen between Val (GAC) from the 1st column and Gly (GCC) from the 4th column which was comparable to the expected small distance between Gly (GCC) and its legitimately complementary partner, Ala (GGC) from the 2nd column (Figure 2). Again, Ala (GGC) is twin of Val (GAC), and *vice versa* in relation to their complementary partners from 3rd column. In fact, the above are examples of secondary complementary partnership. Of all the exceptions when uncomplementary anticodons show still rather small complementary distance between their tRNAs, this was the most frequent type.

As well known, one of the three polypeptide chain terminators, UAG still remains a Trp codon in the mitochondrial coding system (Jukes, 1966; Bibb *et al.*, 1981); except for that of plants (Wissinger, *et al.*, 1992). This afforded us with an opportunity

to compare the consensus anticodon loop and stem of Ser (UGA) tRNA with those of mitochondrial Trp (UCA) tRNA. There were only five mispairings in 17. Remarkably, also observed was the 'wrong' low complementary distance between Ser (UGA) and Trp (CCA) (only 3 mispairings). This eloquent coincidence appears to seem not only as a new (among many others) evidence of 'senseful past' of the UCA triplet (Jukes, 1966), but also indicates once again on the paired origin of tRNAs. The Trp tRNA with anticodon CCA and its complementary partner, Pro (UGG) tRNA showed 5 mispairings. Because of its immense size and complexity, however, tryptophane is not likely to have been present at the beginning. More likely, all 4 codons beginning with UG encoded cysteine. Indeed, in *E. coli* as well as in mammals, UGA may be recognized by a unique species of tRNA known as seleno-cysteyl-tRNA<sup>(Ser)</sup><sub>Sec</sub>, thereby encoding selenocysteine (Leinfelder *et al.*, 1989).

All the above is incompatible to classic theory of molecular evolution by gene duplication followed by mutational divergence of 'daughter' copies and, conversely, is just what is predicted by the alternative model of concerted origin of tRNAs with complementary anticodons. Consistent to this are data showing that minimization of chemical distances between pairs of neighboring amino acids in the code table could play only a subsidiary role in shaping the evolution of codon-to-amino acid assignment (Wong, 1975, 1980, 1981).

### 3.6. ON THE PHYLOGENETIC DIAGRAM FOR CONSENSUS tRNAs

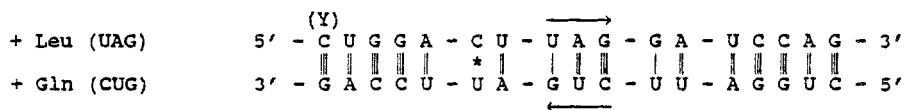
In nonenzymatic replication of RNAs, very often G bases are incorporated incorrectly opposite template T and presumably U bases (Orgel, 1992). In correct replication (in terms of Watson-Crick rules), G-U pairs should produce A\*C mispairing which then, in the next cycle, lead back to G-U pairing.

Besides, although in codon-anticodon interaction, the 'wobble' pairing is generally allowed only in the 3rd site of codons (Crick *et al.*, 1976), for example, Leu tRNA with UAG anticodon is able to recognize not only CUA but also the remaining five Leu codons, including UUG and UUA (Lewin, 1990).

Accordingly, allowing for G-U pairing, one can be convinced that any pair of pre-tRNAs with complementary or quasi-complementary anticodons may generate 16 anticodons (Rodin *et al.*, 1993). So, all 64 anticodons can be divided on 4 subgroups (enumerated in Figure 4) each consisting of 16 anticodons linked inside by the concerted mechanism of their tRNA origin. In turn, with regard to the most important 2nd base, these subgroups form the already discussed two complementarily isolated groups (shaded and unshaded in Figure 4). If we assume also that errors like transversions might take place during replication in 3rd site of anticodons (and, symmetrically, in 1st site of codons), then genetic code could really start with a single pair of complementary pre-tRNAs. Furthermore, this assumption is necessary since every such subgroup of 16 triplets does not cover completely the series of synonymous codons.

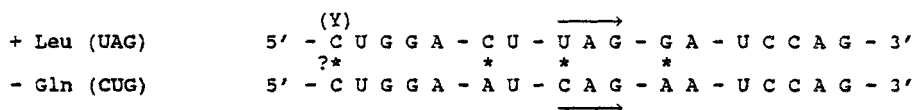
Alas, it is difficult to test the above by tRNA phylogeny reconstruction, since

A. Initial consensus tRNAs with complementary anticodons anti-parallelly oriented to each other:



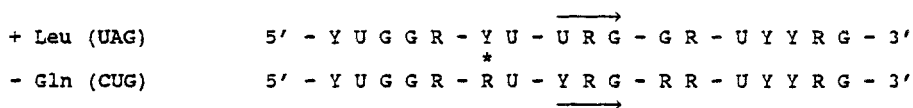
Complementary distance  $D_c = 1$   
 Direct distance  $D_d = 6$

B. The same pair of tRNAs, but one in plus- and another in inverted minus-configuration, parallelly oriented to each other:



Direct distance (max)  $D_d = 4$

C. The same pair of tRNAs but rewritten in the degenerated alphabet (Y U R G) because of bipotential pairing of G with C and U:



Direct distance (min)  $D_d = 1$

Fig. 5. An example of rewriting consensus tRNAs with complementary anti-codons in the degenerated alphabet. Due to such a transformation of consensus tRNA sequences, the corresponding phylogenetic trees may be optimized by standard methods in order to approach their most parsimonious topology.

existing theory of gene evolution implies tacitly that only one of two strands of any replicating unit is valuable to function and gradually evolve. In contrast, concerted tRNA origin supposes that both the strands should be considered as functionally and evolutionarily active. Consequently, the concerted origin should be far more rapid process producing at once numerous substitutions in G-U paired positions. We suppose it was one of the main reasons why all attempts failed to resolve a phylogeny between different tRNA families by a tree-like graph. In other words, the alternative bundle-like topology of tRNA trees (Eigen *et al.*, 1988, 1989) may reflect the rapid bundle-like concerted emergence of first complementary adapters.

On the first view, it seems sufficient to follow Eigen and Schuster (1979), i.e. to write for any 'plus' sequence its complementary ('minus') image and then to

invert the latter (see Figure 5a). Figure 5b shows, however, this approach is inadequate since direct distances between plus- and inverted minus-chains may be considerably overestimated in comparison with 'true' complementary distances. Such an overestimation is rooted in ambiguous replication of irregular G-U base pairs.

To lower, at least in part, these discrepancies, the consensus tRNAs were rewritten in the reduced alphabet as illustrated in Figure 5c. In result, one deals now with minimal and maximal direct instead of complementary distances. Though the reduced alphabet certainly smooths over differences between pairs of tRNAs with complementary and uncomplementary anticodons, we have chosen, as a compromise, the following heuristic two-step approach. On first step, the integral 'chimeric' matrix of distances was composed. In the upper left and low right quarters of the matrix are direct distances calculated for pairs of consensus tRNAs representing 1st + 2nd and 3rd + 4th columns of the table of anticodons, respectively, while in the upper right quarter and, symmetrically, in the low left quarter are the corresponding complementary distances. On second step, we used plus- and inverted minus-tRNAs to optimize trees by means of the VOSTORG package options (Zharkikh *et al.*, 1991). The topology of divergence was rather insensitive to the methods of tree building and optimization. One of the most 'parsimonious' trees (31 mutations in sum) is shown in Figure 6. In contrast, the matrix of maximal direct distances yielded, as expected, notably different trees with larger sum of mutations; from starting 161 to optimal 146 (not shown). Nevertheless, the tendency for complementary anticodons to fall into the same clusters again was observed, though being considerably diluted by tRNAs for synonymous codons.

Generally, Figure 6 shows almost ideal accordance with complementary partnership subdividing anticodons on four clusters. Only six positions are 'incorrect' in the tree (Figure 6, underlined): those of tRNAs for Cys (GCA), Tyr (GUA), Glu (UUC, YUC), Lys (UUU), and Arg (UCG). Among other possible reasons, this may be attributed:

(1) for Cys (GCA) and Tyr (GUA) – to doubtful consensus sequence for the former;

(2) For Glu (UUC, YUC) and Lys (UUU) – to their complementary partnership with Phe (GAA) because, as we mentioned before, a secondary and therefore reduced complementary distance is expected for small series of synonymous codons;

(3) For Arg (UCG) – to possible origin of its primordial adapter from that for other arginine codons and (or) lack of its legitimate complementary partner, Ser (CGA) in our set of consensus tRNAs.

In support, the following arguments may be added: (1) Removing Cys (GCA) from the set replaces its 'twin' Tyr (GUA) into another subtree, 'correct' for it; (2) Contrary to Lys (UUU), its nearest neighbor in the table of anticodons, Lys (CUU), appears to be localized in the tree properly, perhaps just due to its complementary partnership with Leu (GAG) from the large series; (3) The pathways leading from the tree root to Arg (UCG), Tyr (GUA), and Cys (GCA) are of

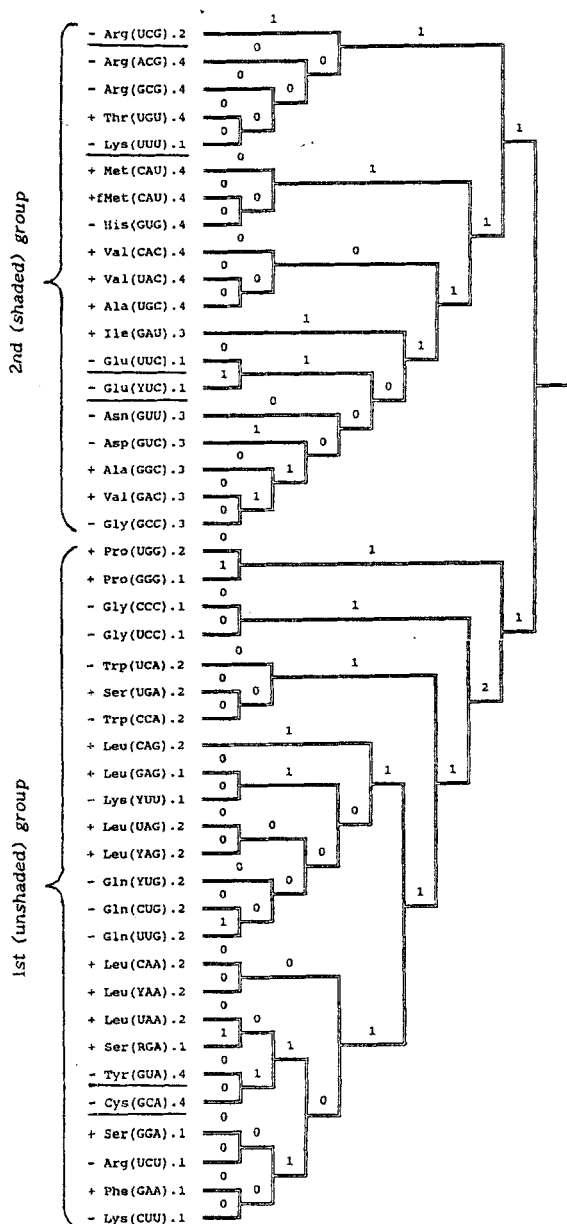


Fig. 6. One of the most 'parsimonious' trees of consensus tRNAs constructed by VOSTORG (TREPRT routine, UPGMA method) package (Zharkikh *et al.*, 1991). tRNAs with anticodons from 1st and 2nd columns in Figure 4 were of plus type, while those from 3rd and 4th columns, of minus type. All 44 tRNAs were compared in 5' → 3' orientation; the length of sequences compared was 17 bases (those of anticodon stem and loop); the total number of substitutions is 31. All other the most 'parsimonious' tree variants of the same total length (31 substitutions) including those built by different ways (Zharkikh's method of unique substitutions (Zharkikh 1977, 1985; see also: Mirkin and Rodin, 1984), neighbor joining, *etc.*) had in fact the same structure with an accuracy to rotation of zero-length branches.

relatively large length so that their transposition in another subtree influences very slightly on the total number of mutations.

### 3.7. CONCERTED tRNA ORIGIN AND COMPLEMENTARY CODON-USAGE-PATTERN

Among numerous aspects of codon usage problem, one just related to complementary codons was passed over in silence except for a very few studies (Alff-Steinberger, 1984; Ohno, 1991; Ohno and Yomo, 1991). Regarding the palindromes within mRNA sequences one more premise can be verified. All other things being equal, the more bases of RNA sequence are paired in stems, the more stable will be its secondary folding. But since two complementary mRNA stem regions strands do not necessarily share the same translation frame, only for a statistically rich ensemble of various mRNAs can one hope to detect nearly equal quantities of codons complementary to each other.

To test this possibility, we have processed published data on codon usage (Wada *et al.*, 1990; Ikemura, 1992). The results, brought together in Table IV, show as expected comparable summed contents of complementary codon groups, the most contrasting exclusion being shown by mitochondrial genes. However, literally for each species analyzed, the following 'nuances' have to be noted: (1) a total codon number of the unshaded group is obviously less than that of the shaded group, and (2) codons of 2nd and 4th columns compared to those of 1st and 3rd ones, are considerably under represented, the most challenging deficiency is shown by 4th column codons with central G base.

The first peculiarity can be attributed to three stop-codons because of: (i) all these three are localized in the unshaded group; (ii) curiously, their complementary partners, leucine codons UUA and CUA, as well as serine UCA occupy usually last ranks in usage; (iii) of deviations from universal genetic code in mitochondria, the most well-known is usage of the UGA by tryptophan.

Not surprisingly, mitochondrial UGA has left the region of minor concentrations. Besides, what is really remarkable is that its complementary partners, legitimate UCA (for serine) and especially illegitimate UUA (for leucine) also have become extremely frequent. Furthermore, leucine UUA occupied the absolute rank no. 1 among all codons (both in nucleus and mitochondria) on average twice surpassing CUG which is the most preponderant codon in basic genome. In turn, this dramatic decrease of CUG (Leu) content in mitochondrial genes is accompanied by falling of CAG (Gln) usage. Generally, one can observe qualitatively concerted changes in complementary codon usages occurred in (T, A) - rich mitochondrial genes in comparison with coson-usage preferences in basic genomes (see: Wada *et al.*, 1990).

The second peculiarity is in a sharp conflict with what is expected, were 'palindromic' secondary structure of mRNAs the leading reason of this symmetry in complementary codon group usages. Indeed, the reverse excess of codons with the central G base is expected since due to their ability to recognize both C and U

TABLE IV

Summed usage of codon (anticodon) groups, complementary and quasi-complementary to each other calculated for data from (Wada *et al.*, 1990; Ikemura, 1992)<sup>a</sup>

Columns	1st	2nd	3rd	4th	1 + 2	3 + 4	2 - 4	3 - 1
2nd site of: codons	U	C	A	G	U + C	A + G	C - G	A - U
(anticodons)	(A)	(G)	(U)	(C)	(A + G)	(U + C)	(G - C)	(U - A)
1st group <sup>b</sup>								
ECO (937)	135.9	76.4	156.9	37.4	212.3	194.3	39.0	21.0
YSC (581)	134.0	105.7	181.2	59.3	239.7	240.5	46.4	47.2
MZE (62)	141.7	108.4	158.8	62.8	250.1	221.6	45.6	17.1
DRO (266)	118.2	101.9	171.3	58.4	220.1	229.7	43.5	53.1
FSB (76)	137.6	104.7	152.6	72.7	242.3	225.3	32.0	15.0
XEL (111)	120.1	96.5	187.9	84.0	216.6	271.9	12.5	67.8
CHK (210)	122.7	90.1	190.3	75.8	212.8	266.1	14.3	67.6
HUM (1490)	134.0	101.1	171.2	86.2	235.1	257.4	14.9	37.2
YSC, Mt (20)	188.9	97.0	120.4	57.9	285.9	178.3	39.1	-68.5
2nd group <sup>c</sup>								
ECO (937)	157.2	150.2	143.7	142.5	307.4	286.2	7.7	-13.5
YSC (581)	144.5	122.6	170.6	82.5	267.1	253.1	40.1	26.1
MZE (62)	140.9	155.0	126.4	106.5	295.9	232.9	48.5	-14.5
DRO (266)	132.8	134.0	159.8	123.9	266.8	283.7	10.1	27.0
FSB (76)	139.6	136.4	139.2	117.1	276.0	256.3	19.3	- 0.4
XEL (111)	133.7	129.8	145.9	101.9	263.5	247.8	27.9	12.2
CHK (210)	140.5	130.2	146.9	103.2	270.7	250.1	27.0	6.4
HUM (1490)	130.8	126.7	144.7	105.1	257.5	249.8	21.6	13.9
YSC, Mt (20)	185.7	106.1	175.3	69.0	291.8	244.3	37.1	-10.4

<sup>a</sup> As in (Wada *et al.*, 1990; Ikemura, 1992), codon usage summed is listed as frequency/1000.

<sup>b</sup> This group includes complementary codons (anticodons) unshaded in Fig. 4.

<sup>c</sup> This group includes complementary codons (anticodons) shaded in Fig. 4.

Species are abbreviated as in (Wada *et al.*, 1990):

ECO - *Escherichia coli*, YSC - *Saccharomyces cerevisiae*, MZE - Maize, DRO - *Drosophila melanogaster*, FSB - Fish (Group B), XEL - *Xenopus laevis*, CHK - chicken, HUM - Human, YSC, Mt - *Saccharomyces cerevisiae* mitochondria.

In brackets are shown numbers of genes summed for each of these species.

bases, they could in principle compensate excessive codons with central U base when the latter are not balanced by legitimate A bases.

The contradiction is safely removed due to the simple conversion of this C excess - G deficiency in codon usage to the mirror G excess - C deficiency in anticodon usage. Accordingly, quantity of the central G base containing antocodons becomes more than sufficient to compensate completely those with the central legitimate C from the 4th column. Moreover, as two last columns of Table IV show, in the majority of cases excessive anticodons with the central G base could balance those with the second U base when the latter surpasses a number of legitimately complementary anticodons with the second A base.

These results can be evaluated as a new argument for the view that in the course of coevolution between codon-usage patterns and tRNA content (Bulmer, 1987) the former had to be optimized to the second rather than *vice versa*! Thus, it seems possible to speculate that from the very beginning of genetic code evolution complementary codon usage pattern has been determined, at least partly, by the concerted origin of pre-adapters with complementary anticodons.

### 3.7.1. Prospects

Taken together, the findings described throw new light on many controversies of genetic code origin and associated topics: the hypothesis of direct recognition (Woese *et al.*, 1966, Woese, 1967) versus that of frozen accident (Crick, 1968); bundle-like versus tree-like phylogeny of tRNAs (Eigen *et al.*, 1989); the ambiguity reduction scenario (Fitch, 1966; Fitch and Upper, 1987); the codon reassignment processes (Osawa and Jukes, 1988, 1989; Lehman and Jukes, 1988; Szathmary, 1991; Szathmary and Zintzaraz, 1992), and generally, the theory of RNA quasi-species and hypercycles as most probable precursors of the universal genetic coding system (Eigen and Schuster, 1979).

Finally, of pressing items in the agenda prompted by the idea of tRNA concerted origin, the following three may be noted:

(1) Complementary partnership of tRNAs, both legitimate and illegitimate, implies that composition and replication connections in primordial molecular quasi-species might have been more complicated than the fittest master RNA sequence with a long 'tail' of its nearest mutational derivatives. Accordingly, the question how the quasi-species and hypercycles could have played a role of earliest information integrator needs to be reexamined.

(2) The question how compatible is the idea of tRNA concerted origin to capture or even swapping of codons, probably directed by mutation pressure  $CG \rightleftharpoons TA$ , waits for a detailed multifaceted analysis. Here we should note only that, for example, captures of terminating UGA, UAA and UAG by amino acids Trp, and Gln, respectively, might have been realized merely due to a single base substitution in the 1st or 3rd site of anticodons of the corresponding ancestral tRNAs. The same could be resulted immediately, just by the concerted mechanism (Rodin *et al.*, 1993: Figure 3). Anyhow, the above reassignments do not change the complementary partnership between anticodons. Certainly more interesting instances of swapping are those involving codons different in their central base like, for example, AAG, AAA for Lys and AGG, AGA for Arg (Szathmary, 1991). Indeed, both these amino acids as well as their anticodons (CUU, UUU and CCU, UCU respectively) are in fact 'twin' ones connected through one and the same complementary partners Ser (AGA, GGA). Thus, again the reassignment of anticodons (and, naturally, codons) appears to be invariant in relation to their complementary partnership. What is more, because many synonymous codons belong to different sets complementarily linked inside (Figure 4), logically the concerted origin of tRNAs implies subsequent, though local, codon-to-amino acid reassignments.



(3) The concerted origin of tRNAs as well as many other arguments suggest that initial variety of amino acids involved into translation was considerably smaller than today. However, the resolving capacity of phylogenetic sequence analysis is too limited to decipher the original codon-to-amino acid assignment. So, further analysis of the genetic code hidden complementary symmetry in combination with more detailed investigation of hypercyclic models and phylogenetic comparisons on extended sets of relevant sequences will help at least to exclude the most incredible versions.

### Acknowledgements

We thank Drs. G. Cocho, J. Holmquist, W.-H. Li, W. Fitch, M. Waterman, M. Eigen and E. Szathmari for critical remarks and useful comments. Two of us (S. and A. Rodin) thank the Consejo Nacional de Ciencia y Tecnologia, Instituto de Fisica of Mexico, and the programs of the Russian Academy of Sciences 'Human Genome' and 'Priority Research in Genetics', for financial support of this research.

### References

- Air, G. M., Blackburn, E. H., Coulson, A., Galibert, F., Sanger, F., Sedat, J., and Ziff, E. B.: 1976, *J. Mol. Biol.* **107**, 445.
- Alf-Steinberger, C.: 1984, *Nucl. Acids Res.* **12**, 2235.
- Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W., and Clayton, D. A.: 1981, *Cell* **26**, 167.
- Blaisdell, B. E.: 1985, *J. Mol. Evol.* **21**, 278.
- Bulmer, M.: 1987, *Nature* **325**, 728.
- Crick, F. H. C.: 1968, *J. Mol. Biol.* **38**, 367.
- Crick, F. C. R., Brenner, S., Klug, A., and Pieczenic, G.: 1976, *Origins Life* **7**, 389.
- Eigen, M. and Schuster, P.: 1979, *Hypercyclic: A Principle of Natural Self-Organization*, Springer-Verlag, Heidelberg.
- Eigen, M., Winkler-Oswatitsch, R., and Dress, A.: 1988, *Proc. Natl. Acad. Sci. USA* **85**, 5913.
- Eigen, M., Lindemann, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress, A., and Von Haeseler, A.: 1989, *Science* **244**, 673.
- Fitch, W. M.: 1966, *J. Mol. Biol.* **16**, 1.
- Fitch, W. M. and Upper, K.: 1987, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 759.
- Grantham, R., Gautier, C., and Gouy, M.: 1980, *Nucl. Acid Res.* **8**, 1893.
- Gouy, M., Jacob, M., and Mercier, R.: 1981, *Nucl. Acid Res.* **9**, 54.
- Gouy, M.: 1987, *Mol. Biol. Evol.* **4**(4), 426.
- Hancock, J. M., Tautz, D., and Dover, G. A.: 1988, *Mol. Biol. Evol.* **5**, 393.
- He, L., Kierzek, R., SantaLucia, J., Water, A. E., and Turner, D. H.: 1991, *Biochemistry* **30**, 11124.
- Ikemura, T.: 1992, In *Transfer RNA in Protein Synthesis*, 87.
- Jukes, T. H.: 1966, *Molecules and Evolution*, Columbia Univ. Press. New York.
- Kimura, M.: 1985, *J. Genet.* **64**, 7.
- Lehman, N. and Jukes, T. H.: 1988, *J. Theor. Biol.* **135**, 203.
- Leinfelder, W., Statman, T. C., and Bock, A.: 1989, *J. Biol. Chem.* **264**, 9720.
- Lewin, B.: 1990, *Genes IV*, Oxford University Press.
- Miller, S. L. and Orgel, L. E.: 1973, *The Origins of Life on Earth*, Englewood Cliffs, N. J., Prentice Hall.

- Mirkin, B. G. and Rodin, S. N.: 1984, *Graphs and Genes*, Springer-Verlag, Berlin Heidelberg New York Tokyo.
- Muramatsu, T.: 1988, *Nature* **336**, 179.
- Nussinov, R.: 1980, *Nucl. Acid Res.* **8**, 4545.
- Nussinov, R.: 1981, *J. Mol. Evol.* **17**, 237.
- Nussinov, R.: 1987, *J. theor. Biol.* **125**, 219.
- Ohno, S.: 1987, *J. Mol. Evol.* **25**, 325.
- Ohno, S.: 1988a, In *The Semiotics of Cellular Communication in the Immune System*, NATO ASI Series, **H23**, Springer-Verlag, Berlin Heidelberg, 131.
- Ohno, S.: 1988b, *Proc. Natl. Acad. Sci. USA* **85**, 9630.
- Ohno, S.: 1991, In *Evolution of Life: Fossils, Molecules, and Culture*, Eds.: Osawa, S. and Honjo, T., Springer, Tokyo, 97.
- Ohno, S. and Yomo, T.: 1990, *Proc. Natl. Acad. Sci. USA* **87**, 1218.
- Ohno, S. and Yomo, T.: 1991, *Electrophoresis* **12**, 103.
- Orgel, L. E.: 1992, *Nature* **358**: 203-209.
- Osawa, S. and Jukes, T. H.: 1988, *Trends Genet.* **4**, 191.
- Osawa, S. and Jukes, T. H.: 1989, *J. Mol. Evol.* **28**, 271.
- Pan, T. and Uhlenbeck, O. C.: 1992, *Nature* **358**: 560.
- Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F., and Cech, T. R.: 1992, *Science* **256**: 1420.
- RajBhandary, U. L.: 1988, *Nature* **336**, 112.
- Rodin, S.: 1991, *Idea of Coevolution*, Nauka, Novosibirsk (in Russian)
- Rodin, S., Ohno, S., and Rodin, A.: 1993, *Proc. Natl. Acad. Sci. USA* **90**, 4723.
- Rodin, S. and Kuznetsov, I. B.: 1993, *Genetica*, **8**, 1256 (in Russian).
- Rould, M. A., Perona, J. J., and Steitz, T. A.: 1991, *Nature* **352**, 213.
- Schulman, L. H. and Pelka, H.: 1988, *Science* **242**, 765.
- Schulman, L. H. and Pelka, H.: 1989, *Science* **246**, 1595.
- Sprinzel, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R.: 1989, *Nucl. Acid Res. (Suppl)* **17**, r1.
- Szathmary, E.: 1991, *J. Mol. Evol.* **32**, 178.
- Szathmary, E. and Zintzaraz, E.: 1992, *J. Mol. Evol.* **35**, 185.
- Trifonov, E. N.: 1987, *J. Mol. Biol.* **194**, 643.
- Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T., and Ikemura, T.: 1990, *Nucl. Acid Res. (Suppl)* **18**, 2367.
- Wissinger, B., Brennicke, A. and Schuster, W.: 1992, *Trends Genet.* **8**, 322.
- Woese, C. R.: 1967, *The Genetic Code*, Harper and Row, New York.
- Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M., and Saxinger, W. C.: 1966, *Cold Spring Harbor Symp. Quant. Biol* **31**, 723.
- Wong, J. T-F.: 1975, *Proc. Natl. Acad. Sci. USA* **72**, 1909.
- Wong, J. T-F.: 1980, *Proc. Natl. Acad. Sci. USA* **77**, 1083.
- Wong, J. T-F.: 1981, *Trends Biochem. Sci.* **6**, 33.
- Yarus, M.: 1988a, *Science* **240**: 1751.
- Yarus, M.: 1988b, *Cell* **55**, 739.
- Yomo, T. and Ohno, S.: 1989, *Proc. Natl. Acad. Sci. USA* **86**, 8452.
- Zharkikh, A. A.: 1977, In *Mathematical Models of Evolution and Selection*, Ed. V. A. Ratner, ICIIG-Press, Novosibirsk, 5.
- Zharkikh, A. A.: 1985, in *Structure and Evolution of Genomes*, Itogi nauki i tehniki, Molekuljarnaja biologija, VINITI-Press, Moscow, **21**, 123.
- Zharkikh A. A., Rzhetsky A. Y., Morosov P. S., Sitnikova, T.L., and Krushkal J. S.: 1991, *Gene* **101**, 251.