J Mol Evol (1987) 26:341–346

Journal of Molecular Evolution

Archetypical Features in tRNA Families

Krikor Nicoghosian,[1] Michel Bigras,[2] David Sankoff,[3] and Robert Cedergren[1]

[1] Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada
[2] Département d'Informatique et Recherche Opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7, Canada
[3] Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Summary. A compilation of known tRNA, and tRNA gene sequences from archaebacteria, eubacteria, and eukaryotes permits the construction of tRNA cloverleafs which show conserved structural elements for each tRNA family. Positions conserved across the three kingdoms are thought to represent archetypical features of tRNAs which preceded the divergence of these kingdoms.

Key words: Evolution — tRNA — Recognition sites — Statistics

Introduction

The comparison of homologous macromolecular sequences from different organisms leads to two types of inference. One, perhaps the better known, is phylogenetic inference; the other, which we consider here, is the reconstruction of primordial structures and functions (for example, Sankoff et al. 1973). It is of particular interest to investigate tRNA in this perspective, because of the important but still only partly understood sequence-dependent mechanism which accounts for the differential aminoacylation of the dozens of different tRNA molecules in any organism. Inference of the primordial states of these sequences could help clarify this mechanism and could even shed some light on the original establishment of the modern genetic code.

In this paper, we use the time-honored tactic of comparing sequences for the reconstruction of tRNA sequences as they existed in proto-organisms. The current availability of data on large numbers of tRNAs for different amino acid families in the three kingdoms (archaebacteria, eubacteria, and eukaryotes, Woese and Fox 1977) gives renewed impetus to the study of conserved features common to members of these families in the hope of deducing which present-day structural features are inherited from the organism ancestral to all three cell types. Since there is controversy over the definitive phylogeny of these kingdoms (i.e., over whether archaebacteria are more related to protoeukaryotes or protoeubacteria, Garrett 1985), we will simply search for features that are common to all three. Conserved structural features were obtained from a computer program that analyzes and records the distribution of nucleotides at each position of the tRNA molecule (Grosjean et al. 1982). These data were extracted from a tRNA sequence data bank compiled at the Université de Montréal which contains 478 sequences of which 128 are from mitochondrial genomes; it is roughly equivalent to the compilation of Sprinzl et al. (1985). The program is dependent on the use of aligned sequences. The first step was to construct a "composite" sequence distribution across all of the 350 sequences from the three kingdoms in the bank; this is a compilation of how many nucleotides of each kind occur at each sequence position. Though mitochondrial tRNA sequences are most likely related to the ancestral eubacterial tRNAs (Gray et al. 1982; Pace et al. 1986), they were all eliminated from the present analysis, since the presence of deletions and/or insertions make the alignment of much of this data with the other sequences difficult, and could thereby falsify the statistical analysis.

The next step was to find features in common within each tRNA family. A tRNA family is defined

Offprint requests to: K. Nicoghosian

Table 1. tRNA sequences used in this study

tRNA	Euk.	Eub.	Arch.	Chloro.	Phage	Total
All	132	87	63	52	16	350[a]
Alanine	4	3	5	2	—	14
Arginine	11	4	4	2	1	22
Asparagine	4	2	3	2	1	12
Aspartic acid	8	2	3	2	1	16
Cysteine	1	2	1	1	—	5
Glutamine	3	3	2	1	2	11
Glutamic acid	8	2	2	2	—	14
Glycine	8	6	8	2	1	25
Histidine	5	2	2	1	1	11
Isoleucine	4	5	2	3	1	15
Leucine	11	11	1	7	2	32
Lysine	7	2	3	—	—	12
Methionine (i)	12	8	6	4	—	30
Methionine (m)	2	2	2	4	—	10
Phenylalanine	9	7	1	4	—	21
Proline	6	4	4	2	2	18
Serine	9	7	3	2	3	24
Threonine	2	5	4	2	1	14
Tryptophan	4	2	1	2	—	9
Tyrosine	5	3	2	2	—	12
Valine	8	5	4	4	—	21

[a] The total number of tRNAs from the eukaryotes (euk.), eubacteria (eub.), archaebacteria (arch.), chloroplasts (chloro.), and T phages in our tRNA and tRNA gene sequence bank includes 348 sequences attributed to one of the tRNA families and two methionine tRNAs whose role (initiator or elongator) has not been elucidated

as all tRNAs which code for a given amino acid and thus includes data from tRNAs having different anticodon sequences. After compilation and analysis, cloverleaf diagrams (Lapalme et al. 1982) were prepared which highlight conserved characteristics of each family. Only common features within a family, which differ significantly from the composite of all tRNAs, are of interest as possible structural characteristics specific to the proto-tRNA for the corresponding amino acid.

Results

Table 1 is a breakdown by kingdom of all tRNA families. All tRNA families have at least one representative member from the three kingdoms (phage and chloroplast tRNAs are considered as eubacterial) (LaRue et al. 1979, 1981).

Analysis of all 350 tRNAs gives the pattern in Fig. 1. Many of these constant positions have previously been noted. Their distribution coincides in large part (except for the CCA-terminus) with the positions that are constrained for reasons of tertiary structure. Thus $U^{8}R^{9}G^{10}$ in the tRNA composite may direct the folding back of the D loop in order that it be in an appropriate spatial relationship to permit the R^{15}–Y^{50} pairing with the T loop. $Y^{32}U^{33}$ and R^{37} would favor the 3′ stacked confirmation of the anticodon loop (see Rich and RajBhandary 1976). S^{30} to S^{40} delimits the extended anticodon (Yarus 1982).

The distribution of conserved nucleotides in the alanine family is surely one of the most distinctive. In this pattern 35 positions are highly constrained. Together with the 23 positions in the global tRNA composite this gives a very high total of 58 positions conserved out of 76. The conserved regions of G^{1} to S^{4} and S^{69} to A^{73} include the G^{3}–U^{70} base pair which is retained across all kingdoms.

The distributions for the arginine and serine tRNA families do not show large numbers of constraints. This may be due to the fact that arginine and serine are each coded for by six codons; it is possible that neither arginine nor serine tRNAs has a single ancestor or that there are subtle structural constraints based on the anticodon sequence. On the other hand, the leucine tRNA family, also using six codons, has considerably more conserved positions.

Even though a large number of positions seem constrained in the cysteine tRNA family, the low number of sequences available reduce the significance of these positions. The histidine tRNA family is characterized by the additional G at the 5′ terminus and several other scattered conserved positions. The lysine tRNA family has an unusually large number of conserved positions encompassing almost the entire D stem and aminoacyl stem, for a total of 60 positions out of 76. Analysis of the methionine tRNA family yields a pattern of many conserved positions. Among them, the three G–C pairs have been previously noticed and may be responsible for the altered anticodon loop conformation (Wrede et al. 1979) and initiation function (Seong and RajBhandary 1987).

The phenylalanine family shows the conserved D-loop region. This region has been considered to be the recognition site for the phenylalanine tRNA synthetase (McCutchan et al. 1978).

A comparison of tRNA family composites among themselves shows some resemblances. Thus, the conservation pattern of lysine tRNA is similar to Ala, Phe, and Thr; Thr to Ile, mMet, and Phe; Val to Ala, mMet, and Tyr. Many families share conserved nucleotides in the D-loop region (Cys, Ile, Leu, Lus, mMet, Phe, and Thr), while others have highly conserved aminoacyl stems (Ala, Cys, Glu, Gly, His, Ile, Lys, iMet, mMet, and Thr). T-loop and variable loop conservation can be seen in Ala, Asn, Asp, and Glu. Anticodon loop conservation is most prevalent in Lys, iMet, Phe, Tyr, and Val.

Although one might be tempted to equate con-

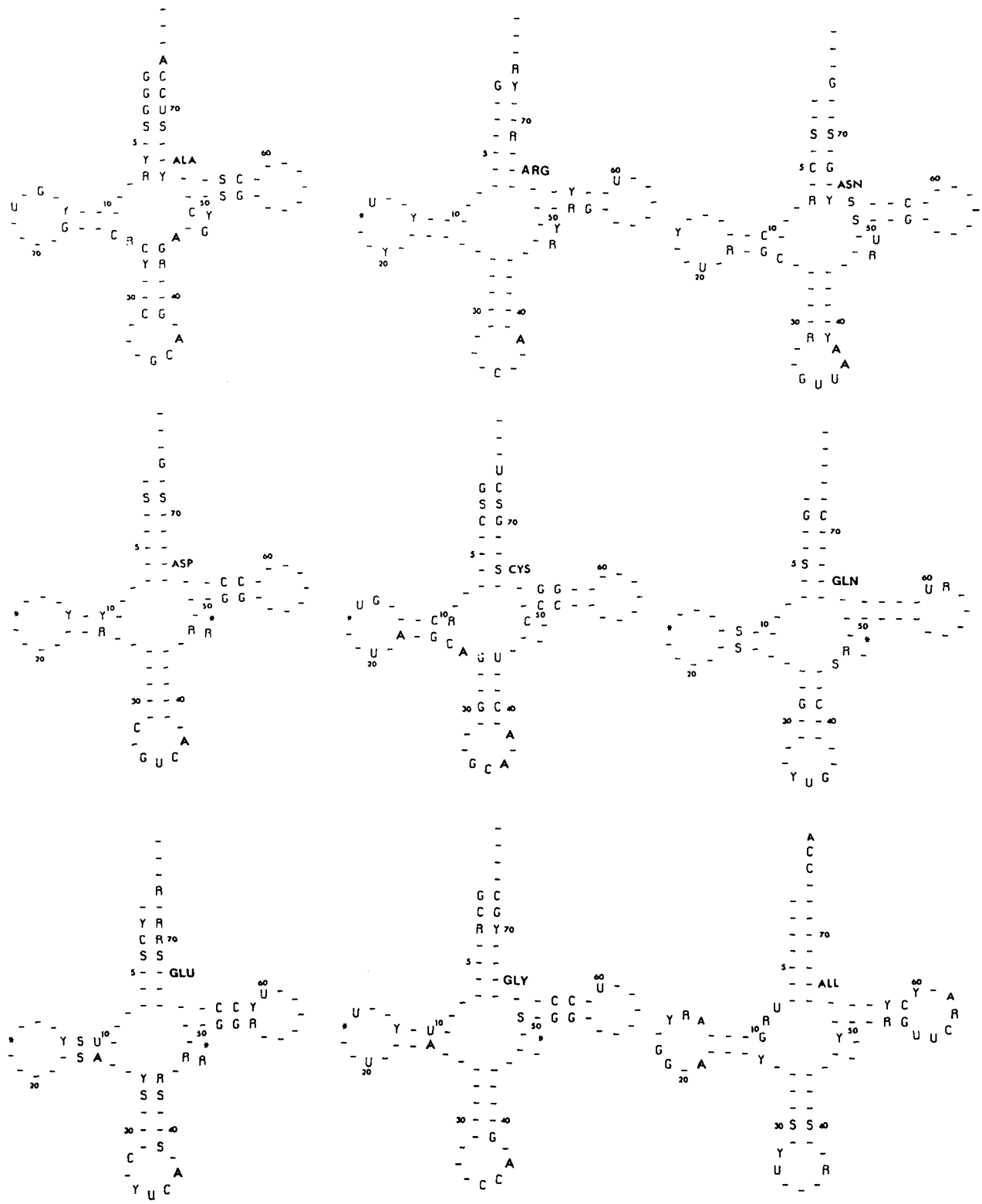

Fig. 1. Conserved features of tRNA families. The number of each of the four parent nucleotides was compiled for each position. The drawing marked A11 comprised the entire tRNA data set of 350 sequences. Letters A, C, G, and U represent the occurrence of these nucleotides in at least 90% of the sequences. Letters R, Y, and S represent at least 95% occurrence of A or G for R, C or U for Y, and G or C for S. In cases where two criteria were fulfilled, the letters A, C, G, and U took priority. Dashes were inserted in all other positions. The tRNA composites for the different families were done in the same manner except that in positions where both the A11 tRNA drawing and the tRNA family drawing would have the same symbol, a dash was used in the latter. In this way, only significant deviations from a more or less random distribution or from the global pattern from A11 tRNAs are emphasized. An asterisk denotes a position which is not always present in a tRNA sequence of a given family.

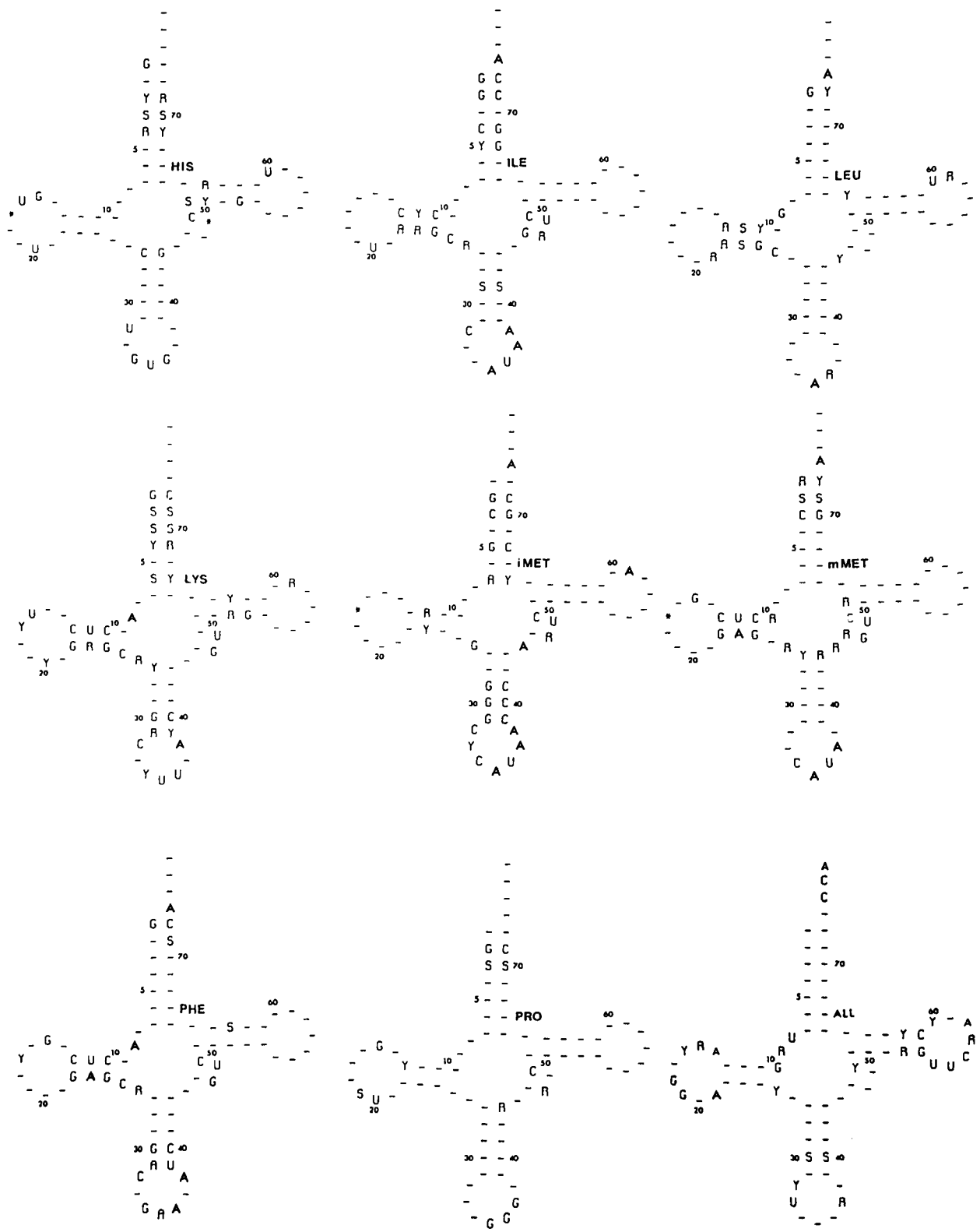

Fig. 1. *Continued*

served positions with recognition sites, the positions in Fig. 1 are not necessarily indicative of recognition sites for aminoacyl-tRNA synthetases, since the long evolutionary time scale implicit in using representatives from the three kingdoms allows extensive divergence in the coevolution of synthetase and tRNA in any given kingdom. The positions do, however, offer an indication of the primordial features

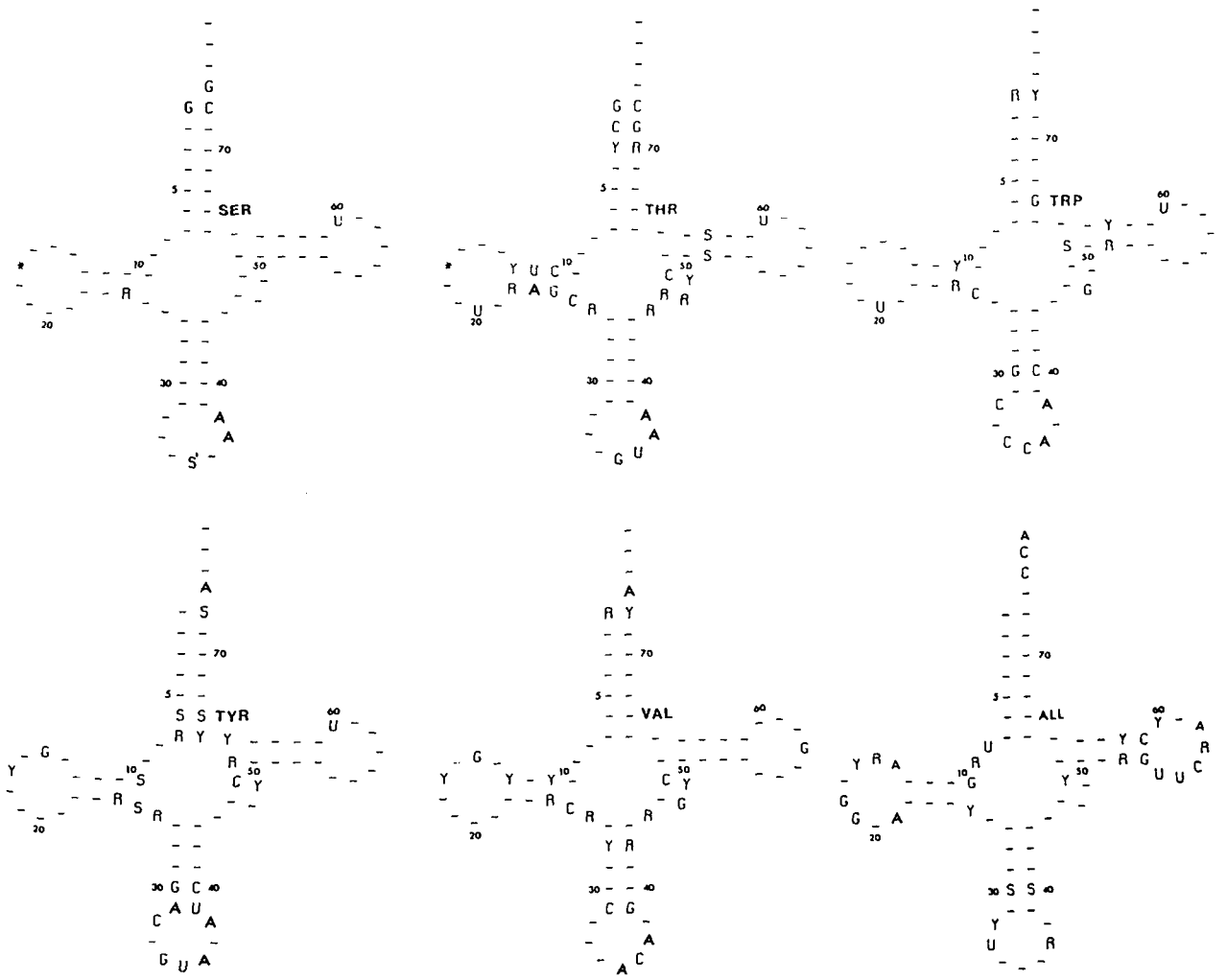

Fig. 1. *Continued*

in each tRNA family to the extent that each family had a unique ancestor.

The overall similarity between individual members of most tRNA families is strongly suggestive that their origin is by divergence from a unique tRNA ancestor (or a closely related population of tRNA genes), rather than by origination from ancestors of other tRNA families. Amino acid specificity changes, although common enough in the laboratory (Schulman and Pelka 1985), seem not to have been common in the evolution of many tRNA families (however see Cedergren et al. 1980). Also, these data imply that, if indeed a single cell "progenote" was the common ancestor of the three major kingdoms (Fox et al. 1980), structure of the tRNA was largely determined in this ancestor. Thus the cloverleaf, the length, and the aminoacyl-tRNA recognition pattern could be almost a primordial property of tRNA (Kingo et al. 1986). This may be quite surprising in view of the fact that mitochondrial tRNAs lacking T and D loops have been found (Wolstenholme et al. 1987).

Acknowledgments. We thank numerous colleagues who have provided us with tRNA sequences over the years. This work was supported by NSERC of Canada.

References

Cedergren RJ, LaRue B, Sankoff D, Lapalme G, Grosjean H (1980) Convergence and minimal mutation criteria for evaluating early events in tRNA evolution. Proc Natl Acad Sci USA 77:2791–2795

Fox GE, Stackenbrandt E, Hespell RB, Gibson J, Maniloff S, Dyer TA, Wolfe RS, Bolch WE, Tanner RS, Magnan LJ, Zablen LB, Blahemore R, Gupta R, Bonen L, Lewis BJ, Stahl A, Livehrsen KR, Chen KN, Woese CR (1980) The phylogeny of prokaryotes. Science 209:457–463

Garrett RA (1985) The uniqueness of archaebacteria. Nature 318:233–235

Gray MW, Sankoff D, Cedergren RJ (1984) On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res 12:5837–5852

Grosjean H, Cedergren RJ, McKay W (1982). Structure in tRNA data. Biochimie 64:387–397

Kinjo M, Hasegawa T, Nagano K, Ishikara H, Ishigami M (1986)

Enzymatic synthesis and some properties of a model primitive tRNA. J Mol Evol 23:320–327

Lapalme G, Cedergren RJ, Sankoff D (1982). An algorithm for the display of nucleic acid secondary structure. Nucleic Acids Res 10:8351–8356

LaRue B, Cedergren RJ, Sankoff D, Grosjean H (1979). Evolution of methionine initiator and phenylalanine transfer RNAs. J Mol Evol 14:287–300

LaRue B, Newhouse N, Nicoghosian D, Cedergren RJ (1981) The evolution of multi-isoacceptor tRNA families. J Biol Chem 256:1539–1543

McCutchan T, Silverman S, Kohli J, Söll D (1978) Nucleotide sequence of phenylalanine transfer RNA from *S. pombe*: implications for transfer RNA recognition by yeast phenylalanyl-tRNA synthetase. Biochemistry 17:1622–1631

Pace NR, Olsen GJ, Woese CR (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45: 325–326

Rich A, RajBhandary UL (1976) Transfer RNA: molecular structure, sequence and properties. Annu Rev Biochem 45: 805–860

Sankoff D, Morel C, Cedergren RJ (1973). The evolution of 5S RNA and the non-randomness of base replacement. Nature New Biol 245:232–234

Schulman LA, Pelka H (1985) *In vitro* conversion of a methionine to a glutamine-acceptor tRNA. Biochemistry 24:7309–7314

Seong BL, RajBhandary UL (1987) *Escherichia coli* formylmethionine tRNA: mutation in GGG-CCC sequence conserved in anticodon stem of initiater tRNAs affect initiation of protein synthesis and conformation of anticodon loop. Proc Natl Acad Sci USA 84:334–338

Sprinzl M, Moll J, Meissner, Hartman T (1985) Compilation of tRNA sequences. Nucleic Acids Res 13:r1–r104

Woese CR, Fox G (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090

Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc Natl Acad Sci USA 84:1324–1328

Wrede P, Woo N, Rich A (1979) Initiator tRNAs have a unique anticodon loop conformation. Proc Natl Acad Sci USA 76: 3289–3293

Yarus M (1982) Translational efficiency of tRNA: uses of an extended anticodon. Science 218:646–652

Received May 6, 1987/Accepted May 30, 1987