INDICATIONS FOR A COMMON EVOLUTIONARY ORIGIN SHOWN IN THE PRIMARY STRUCTURE OF THREE TRANSFER RNAs.

Thomas H. Jukes

Space Sciences Laboratory
University of California
Berkeley, California[1]

Received August 4, 1966

There is evidence that the evolutionary divergence of certain protein molecules from a common archetype has taken place by means of genetic duplication followed by functional differentiation. The extent of the differentiation is restricted by the constraints necessary to conserve the secondary and tertiary structure of the molecule in a manner that is compatible with its biological role. The process has included partial deletions and point mutations Up to the present, it has not been possible to examine these changes by direct comparison of genes, but indications of them are in the amino acid sequences of polypeptide chains. For example, the myoglobin sequence has been compared with the sequences of the α, β and γ polypeptide chains of hemoglobin, and as a result of this comparison their origin from a common archetype by divergent evolution is widely accepted (1). Chymotrypsinogen and trypsinogen are another example (2).

It is now possible for the first time to compare the primary structures of some functionally related nucleic acid molecules, all obtained from yeast, as a result of the discoveries of the base sequences of alanine transfer RNA (sRNA) (3), two serine sRNAs (4), and tyrosine sRNA (5). The comparison was made in a manner similar to that used for the protein sequences mentioned above. The essential procedure is the insertion of arbitrary gaps which infer

[1]Supported by grant NsG 479 by the National Aeronautics and Space Administration to the University of California, Berkeley.

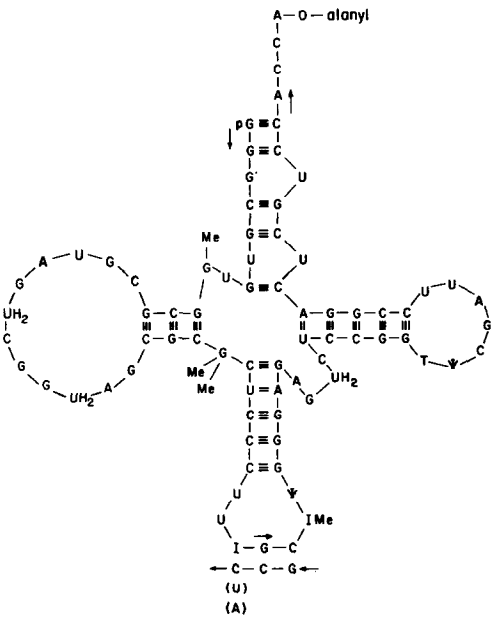
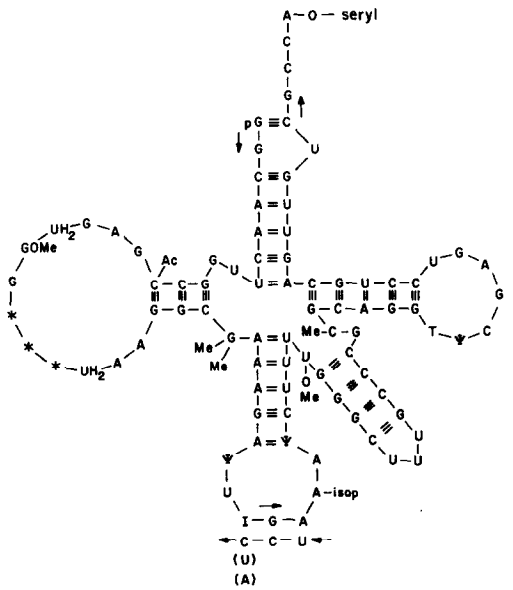
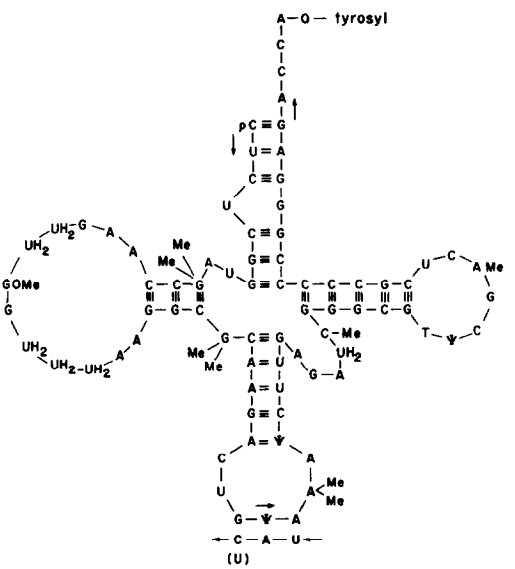


Figure 1. Comparison of possible secondary structures for yeast alanine, serine, and tyrosine sRNAs (see text). *indicates suggested deletion of bases in serine sRNA. The proposed anticodons are shown in each case in complementary juxtaposition to messenger coding triplets, in conformation with the "wobble" hypothesis (12).





that deletions of genetic material have taken place during evolution as shown by alignment of homologous portions of the sequences. The identity of these portions is presumed to have been preserved either due to the essential nature of their function, or because the time lapse following duplication is not yet sufficient for the accumulation of enough randomly-occurring point

mutations to produce differences in base composition. Both of these phenomena are readily perceptible when the globin chains (6) or the chains of cytochrome c (7) are compared.

A secondary structure proposed for yeast alanyl transfer RNA (sRNA) by Holley, et al. (3), subsequently modified by Cantor et al. (8), was used to provide an evolutionary model for comparing the yeast alanyl, seryl and tyrosyl sRNAs. The comparison is shown in Figure 1. The model includes the same four helical regions that are shown in the "cloverleaf" conformation by Holley et al (3), except that one of the regions has been shortened by one base pair, as explained below. It is suggested that one evolutionary deletion is present in each of the three sRNAs, extending from residues 47 to 58 in alanine sRNA, 18 to 22 in serine sRNA, and 48 to 58 in tyrosine sRNA. The numbering system includes all the gaps, to enable comparisons to be made at homologous sites. The three sRNAs possess the following features in common in the proposed model:

Helical Regions	Connecting Regions	Loops
a:1-7, 77-83	8-9, 28, 46	c:13-24
b:10-12, 25-27	51-53, 58-59	e:34-40
d:29-33, 41-45	84	h:65-71
f:47-50, 54-57		
g:60-64, 72-76		

Helical region f and the UUU sequence upon which it loops back have evidently been deleted from alanine and tyrosine sRNAs during evolution and a short connecting region exists in place of the deleted portion. The postulated deletion of a trinucleotide from the "c" loop in serine sRNA implies that this loop originally contained 12 residues, just as in the case of the other two sRNAs, and draws attention to homology between the c loops in all three sRNA molecules. The c loop in alanine sRNA is shown in expanded form, omitting the complementary pairing between residues 13 and 24 (3), to compare it more closely with the corresponding loops in the other two sRNAs.

The three sRNAs are compared for evolutionary homology in their base

sequences in Table 1, in which the proposed helical regions, connecting regions, loops, and deletions are all aligned. The most obvious region of correspondence in all three sRNAs is in residues 64 to 72 which includes the sequence that Zamir et al, (9) found to be common to all yeast sRNAs. Elsewhere the homology for all three appears only in the secondary structure (Figure 1). This is not surprising in view of the extensive incidence of point mutations that can evidently occur during evolution (6,7). There is, however, a region of homology in the primary sequences of serine and tyrosine sRNAs in the region extending from residues 22 to 33 and reoccurring in the complementary sequences 10-12 and 41-44. This indicates that the common archetype of the cistrons for these two sRNAs probably underwent gene duplication at a time subsequent to the gene duplication which separated the archetype of alanine sRNA from the common archetype of serine and tyrosine sRNAs.

A problem exists, of course, in replacing a complementary base pair in a helical region by another complementary base pair. This involves changing one member of the pair by a point mutation and then subsequently inserting the complement of the replaced base by another point mutation at some distance from the first replacement in the cistron. There are indications in the "a" helical region of how this could proceed. Each a helical region (Figure 1) contains one or two Us opposite to a G or another U. Replacement of these by the appropriate complementary base (C or A) in point mutations, or replacement of the base opposite the U by an A would bring about complementarity. Such a procedure would enable the helical regions in two sRNAs to evolve into different base compositions while preserving their helical character. The comparison is in agreement with the concept of divergent evolution from an archetype. A much more recent gene duplication, followed by only 3 single-base changes, has apparently led to the separation of the two yeast serine sRNAs. The archetype of the three sRNAs in table 1 appears to have been transcribed from a piece of DNA that contained 84 base pairs, excluding from the numbering system the terminal -CCA-OH that is added following transcription. Analyses

Table I. Homologous alignment of base sequences in yeast alanine, serine and tyrosine transfer RNAs (3, 4, 5).

	Ala	Ser	Tyr		Ala	Ser	Tyr		Ala	Ser	Tyr
1	pG	pG	pC		C	A	A		C	C	G
	G	G	U		C	G	G		C	A	G
	G	C	C		C	A	A		G	G	C
	C	A	U		U	Ψ	C		*G	G	G
	G	A	C		*U	U	U		*T	T	T
	U	C	G		I	I	G		*Ψ	Ψ	Ψ
	G	U	G		G	G	Ψ		*C	C	C
	*U	U	U		C	A	A		*G	G(A)	G
	GMe	G	A		*IMe	Aisop	AMe_2		*A	A	AMe
10	*G	G	GMe_2	40	Ψ	A	A	70	U	G(A)	C
	*C	C	C		G	Ψ	Ψ		*U	U	U
	G	CAc	C		G	C	C		*C	C	C
	C	G	A		G	U	U		C	C	G
	G	A	A		A	U	U		G	U	C
	U	G	G		G	U	G		G	G	C
	A	UH_2	UH_2		A	UOMe	A		A	C	C
	G	GOMe	UH_2		*G	G	G		C	A	C
	UH_2	G	GOMe		-	G	A		U	G	C
	C	-	G		-	G	-		C	U	G
20	G	-	UH_2	50	-	C	-	80	G	U	G
	G	-	UH_2		-	U	-		U	G	G
	*UH_2	UH_2	UH_2		-	U(C)	-		C	U	A
	*A	A	A		-	U	-		C	C	G
	G	A	A		-	G	-	84	A	G	A
	C	G	G		-	C	-				
	*G	G	G		-	C	-			C	
	*C	C	C		-	C	-			C	
	*GMe_2	GMe_2	GMe_2		UH_2	G	UH_2			A	
	C	A	C		*C	CMe	CMe			OH	
30	U	A	A	60	U	G	G				

Aisop = 6-aminoisopentenyl adenosine (10,11);other abbreviations are conventional ones. The number system includes the gaps. Residues 36-38 are the proposed anticodons (3, 4, 5) and 64-68 are the "invariant" sequence (9). The identity of bases in all 3 molecules is emphasized by *. It is assumed that the modified bases are formed from A, G, U, and C, e.g. GMe_2 from G, and I from A. Base changes that differentiate the second from the first serine sRNA (4) are in parentheses; this differentiation may have arisen following gene duplication.

---

of other sRNAs, however, may lead to comparisons which will reveal the presence of other deletions, and if so, an original length greater than this could be indicated. If the sRNAs indeed evolved from a common origin by gene

duplications, this might suggest that the genetic code evolved from a small number of anticodons that underwent differentiation as the number of sRNAs became increased by gene duplications. Each duplication could give rise to a new anticodon.

REFERENCES

1. Ingram, V.M., 1963. The Hemoglobins in Genetics and Evolution (Columbia Univ. Press, N.Y.) Chapter 6, p. 132.

2. Hartley, B.S., J.R. Brown, D.L. Kauffmann and L.B. Smillie. 1965. Nature 207:1157.

3. Holley, R.W., J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrell, J.R. Penswick and A. Zamir. 1965. Science 147:1462.

4. Zachau, H.G., D. Dütting and H. Feldman. 1966. Angew. Chemie, International Edn. 5:422.

5. Madison, J.T., G.E. Everett and H. Kung. 1966 Abstracts, Cold Spring Harbor Symp. Quant. Biol. 31:43.

6. Zuckerkandl, E. and L. Pauling. 1965, in Evolving Genes and Proteins (Academic Press, N.Y.), p. 97.

7. Margoliash, E. and E.L. Smith, 1965, in Evolving Genes and Proteins (Academic Press, N.Y.), p. 221.

8. Cantor, C., S. Jaskunas and I. Tinoco, Jr.. 1966. J. Mol. Biol. (in the press).

9. Zamir, A., R.W. Holley and M. Marquisee. 1965. J. Biol. Chem. 240:1267

10. Hall, R.H., M.J. Robins, L. Stasiuk and R. Thedford. 1966. J. Am. Chem. Soc. 88:2614.

11. Biemann, K. and S. Tsunakawa. 1966. (cited by Zachau, et al. 1966).

12. Crick, F.H.C. 1965. Codon-anticodon pairing - the wobble hypothesis; manuscript.