Commentary

DID THE PRIMITIVE RIBOSOMAL RNA CODE PRIMITIVE RIBOSOMAL PROTEIN?

A. V. FINKELSTEIN

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Poustchino, Moscow Region, USSR

Received 20 July 1977

I have found that ribosomal 16 S RNA contains the nucleotide sequence which could code a fragment of the ribosomal protein S4, with relatively few errors (fig.1). The only insertion in the ribosomal 16 S RNA is UGA triplet which codes no amino acid residue at all.

The search for cognate protein and ribosomal RNA fragments was done as follows:

1. The adequate 'messenger' sequence was deduced from the known primary structure of proteins S4 [1] taking into consideration the degenerative nature of the code (fig.1).
2. A special computer program simulated the movement of this messenger along the known 16 S RNA sequence [2] in search of the most coinciding nucleotide sections.

The probability of obtaining 'by chance' of the 50 coincidences observed between sections of the 16 S RNA and the deduced messenger in which 55 positions are occupied by one base and 18 positions by two alternative bases (table 1) can be evaluated from the binomial distribution of random values [3].

$$\sum_{n \leqslant 55}^{n+m \geqslant 50} \sum_{m \leqslant 18} \left[\frac{55!}{n!\,(55-n)!} \left(\tfrac{1}{4}\right)^n \left(\tfrac{3}{4}\right)^{55-n} \frac{18!}{m!\,(18-m)!} \left(\tfrac{1}{2}\right)^{18} \right] \approx 2 \cdot 10^{-11} \qquad (1)$$

The coincidence between the primary structures of these sections is much better than the agreement, for example, between the structures of leucine and valine tRNA of *Escherichia coli* [4] (table 1). However, in evaluating the reliability of the results, the number of possible independent comparisons between the primary structures of sections must be taken into account. The relative shift of the sequences, the selection of starting points and lengths, the position of a deletion (or insertion) can be done in different ways. For a protein of 203 residues, an RNA of ≈ 1530 bases, a maximum fragment of 35 amino acid residues and an arbitrary position of deletion (or insertion) with a length up to three nucleotides, we could make no more than

$$[203 \cdot 1530 \cdot 35] \cdot [(34/2) \cdot (3 \cdot 2)] \approx 10^9 \text{ comparisons.}$$

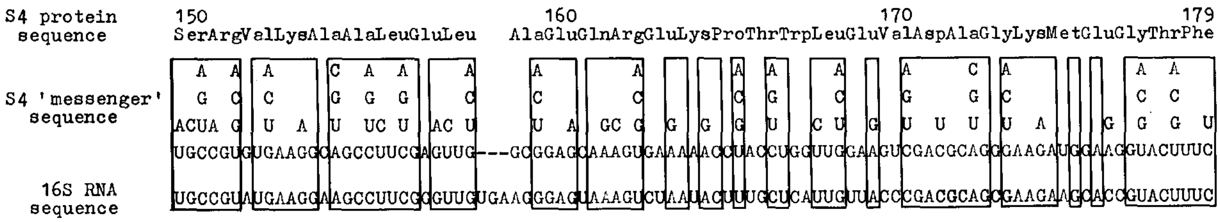

Fig.1. Cognate primary structures of protein S4 and ribosomal RNA fragments.

Table 1

Objects for comparison	Number of bases corresponding to one element of the sequence		Number of compared elements	Expected number of chance coinci-dences of elements	Number of displayed coinci-dences	Probability of chance coinci-dence	Deletions (insertions)		Number of sequences compared	Probability of chance coincidence
							Number	Bases		
	rRNA	S4 'mRNA'								
S4 'messenger' fragment and 16 S RNA fragment	1	1	55	14	33	$2 \cdot 10^{-11}$	(1?)	3	$< 10^9$ (deletion in S4 against the 'non-sense' codon UGA is considered an error)	< 0.02
	1	2	18	9	17					
	1	4	[not taken into account 17 $\equiv$	17 $\equiv$	17]				$< 10^7$ (deletion is not considered an error)	< 0.0002
	$tRNA^{Leu}$	$tRNA^{Val}$								
$tRNA^{Leu}$ and $tRNA^{Val}$	1	1	76	19	40	$6 \cdot 10^{-7}$	2	11	$2 \cdot 10^5$	0.1

The number of independent comparisons is much less than this upper limit. Nonetheless, the probability of chance coincidence of primary structures even for a complete run-through is very small and is not greater than the probability of chance coincidence between the primary structures of tRNAs, as evaluated in analogous manner.

Thus, the fragment of the ribosomal RNA gene and the fragment of the ribosomal protein S4 gene could have some common ancestor. Protein S4 is remarkable in that it seems to be one of the most important for assembly and structure formation of the ribosomal 30 S subparticle. It directly and independently of other proteins binds with ribosomal 16 S RNA [5,6], shielding great portion of the 5′-terminal part of the molecule [7] (including the 'cognate' 400–492 fragment!). The 16 S RNA·protein S4 complex can be supposed to be very 'ancient'.

Did the primitive ribosomal RNA itself serve as a template for the most primitive ribosomal proteins? Were the other primitive proteins [8,9] coded by the primitive ribosomal RNA (or, perhaps, by its complementary copy)? We hope that further 'molecular-paleontological excavations' in modern primary structures will give answers to these questions.

Acknowledgements

I would like to thank Dr T. A. Erokhina for help in the work and express my gratitude to Professor A. S. Spirin, Professor O. B. Ptitsyn, Dr L. P. Gavrilova and Dr V. E. Koteliansky for their interest and critical comments.

References

[1] Reinbolt, J. and Schiltz, E. (1973) FEBS Lett. 36, 250–252.
[2] Ehresman, C., Stiegler, P., Mackie, G. A., Zimmermann, R. A., Ebel, J. P. and Fellner, P. (1975) Nucleic Acids Res. 2, 265–277.
[3] Hudson, D. J. (1964) Statistics, CERN, Geneva.
[4] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Silver Spring.
[5] Schamp, H. W., Green, H. and Kurland, C. G. (1970) Molec. Gen. Genet. 109, 193–205.
[6] Mizushima, S. and Nomura, M. (1970) Nature 226, 1214–1218.
[7] Ungewickell, E., Garrett, R., Ehresmann, C., Stiegler, P. and Fellner, P. (1975) Eur. J. Biochem. 51, 165–180.
[8] Crick, F. H. C. (1968) J. Mol. Biol. 38, 367–379.
[9] Orgel, L. E. (1968) J. Mol. Biol. 38, 381–393.