Hypothesis

A possible novel interaction between the 3′-end of 18 S ribosomal RNA and the 5′-leader sequence of many eukaryotic messenger RNAs

David R. Sargan, Stephen P. Gregory and Peter H.W. Butterworth

Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, England

Received 11 August 1982

A novel interaction between the 5′-untranslated region of eukaryotic messenger RNAs and non-contiguous sequences in the 18 S ribosomal RNA is proposed. The small ribosomal RNA contains, at its 3′-terminus, a heavily conserved hairpin structure. It is suggested that mRNA 5′-leader sequence stabilises this structure by interacting with other conserved nucleotides which flank it. Sequences closely related to the required sequence (A–U–C–C–A–C–C) occur quite commonly in eukaryotic mRNAs and are often found immediately upstream from the AUG-codon. This interaction may have a role in the events which lead up to the initiation of protein synthesis.

Eukaryotic mRNA *Secondary structure, of 18 S rRNA* *Ribosome binding sites* *Protein synthesis, initiation*

1. INTRODUCTION

Eukaryotic messenger RNAs (mRNAs) do not contain the clearly defined 'Shine-Dalgarno' sequence [1] which is thought to act as a ribosome binding site in prokaryotic mRNAs. This conserved purine-rich sequence on the 5′-leader of the mRNA is thought to basepair with the sequence A–C–C–U–C–C–U at the 3′-end of 16 S ribosomal RNA (rRNA). A number of models have been proposed for recognition of mRNAs by eukaryotic ribosomes (review [2]). In one of these [3,4], an analogous basepairing is proposed between a pyrimidine-rich sequence of the mRNA leader and a conserved purine-rich sequence near to the 3′-terminus of 18 S rRNA (see fig.1a). An alternative, the 'scanning' model, suggests that the ribosome recognises the m^7-G cap structure present on most eukaryotic messengers to which it binds and subsequently moves along the mRNA sequence until it finds the first AUG-codon at which it initiates protein synthesis [5–7]. Unfortunately, neither of these mechanisms seems to be universally applicable (see [2] and section 4).

Here, we propose a novel type of interaction between some mRNA leaders and eukaryotic ribosomes in which a different, partially conserved sequence immediately adjacent or very close to the AUG-codon of the messenger stabilises a highly conserved structure in the 18 S rRNA. The type of interaction is somewhat similar to that proposed between the small nuclear RNA 'U1' and non-contiguous sequences in HnRNA, at splice junctions, which is thought to mediate HnRNA processing [8].

2. DETAILED DESCRIPTION OF THE MODEL

We have sequenced a gene for a known protamine of rainbow trout (in preparation) in which the mRNA leader is made up of only 14 nucleotides. This sequence, which must contain sufficient information for ribosome recognition, includes a sequence complementary to bases which flank the conserved hairpin close to the 3′-terminus of the small ribosomal RNA (fig.1a). This hairpin and its flanking nucleotides have been very highly conserved through evolution [4,9]. A survey of a number of other mRNAs has now shown that the required complementary sequence (A–U–C–C–A–C–C) is often present and, in

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$2.75

particular, that a related sequence is commonly found immediately 5′ to the AUG-codon initiating protein synthesis (see section 3). Supporting evidence for the involvement of the conserved hairpin in mRNA recognition comes from the finding that if exposed regions of mRNA in ribosomes are cross-linked, the entire structure is recovered with the 5′-mRNA leader [10]. If a basepairing mechanism is the sole method of interaction, then $\Delta G_{formation}$ for the structure in fig.1a, calculated according to [11], is -24.8 kcal . mol^{-1}, of which -10 kcal . mol^{-1} is contributed by the interaction between the 18 S rRNA and the mRNA. These figures are likely to be slightly reduced by steric constraints on the bonding molecules. It is also possible that a ribosomal protein mediates the interaction: hairpin structures in nucleic acids have often been implicated as protein recognition and binding sites.

3. OCCURRENCE OF THE SEQUENCE A–U–C–C–A–C–C IN EUKARYOTIC mRNAs

Kozak [2,12] has shown that the sequence X–X–X–C–A–X–X–A–U–G is quite highly conserved at the initiation codon of eukaryotic mRNAs. A closer look at the non-viral sequences reviewed by her [2] and 46 sequences or partial sequences published in [13–24] shows that other nucleotides near the AUG-codon are also under selective pressure (see table 1a). However, such a collection of mRNA sequences does not represent a random sample: it includes 19 globin mRNAs and several other sets of closely related genes. Such biased sampling could hinder the recognition of a consensus sequence using the data presented in table 1a and other compilations of mRNA sequences [2,12]. For instance, table 1b shows the strong bias in all positions in the globins, including bases at positions -5 and -6 which do not follow the proposed consensus. If selective pressure is low at these positions, this bias may result simply from a founder effect. Furthermore, some groups of mRNAs included in table 1a, such as those from *Dictyostelium discoideum*, do not fit the model at all, and consequently add 'noise' to this analysis. Likewise, many viral mRNAs do not show the consensus sequence (see table 1c). It is possible that viral systems have evolved special mechanisms to

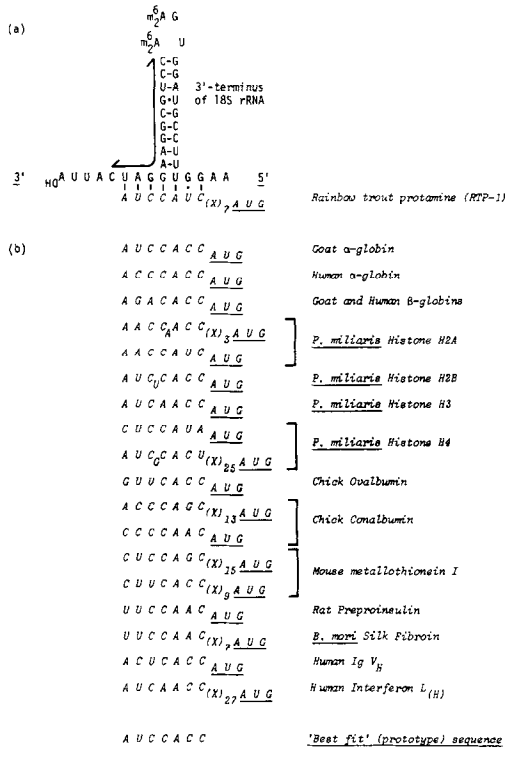

Fig.1. Possible secondary structure formed by bonding between the 3′-terminus of 18 S rRNA and the 5′-leader sequence of eukaryotic mRNAs. (a) The sequence of the highly conserved 3′-terminal loop and the flanking nucleotides of 18 S rRNA are from [9]. This has been aligned with a sequence within the 5′-leader of a rainbow trout protamine mRNA which has been derived from the DNA sequence of a cloned protamine gene and S_1-mapping studies (in preparation). The solid line beside the 18 S rRNA shows the sequence required for bonding with mRNA 5′-leaders in the model of [3,4]. (b) Some examples of sequences within the 5′-leaders of mRNAs are presented which show close homologies with the hypothetical consensus sequence. In some of these, the best fit is immediately adjacent to the functional AUG-codon; in others, it is found upstream from the functional AUG-codon or is derived by looping out of a single nucleotide. Examples of all types of fit are shown. 33 mRNAs were also drawn arbitrarily from those used to compile table 1a and were exhaustively searched for upstream consensus sequence homologies within 35 bases of the initiating AUG-codon. The criteria used to define upstream sites were that they should:

(i) Fit the consensus better than the -7 to -1 sequence in the same messenger;

(ii) Match perfectly at 5 or more of the 7 bases.

Twenty nine such sequences were found in 22 mRNAs.

take over the host ribosomes. Nonetheless, the overall impression of the data presented in table 1 is that there is significant sequence homology in this region.

In many cases where only a partial fit is found immediately adjacent to the AUG-codon, better fits can be found either a few nucleotides upstream from the site of the initiation of protein synthesis or by allowing one or more nucleotides to loop out during the interaction between mRNA and 18 S rRNA (see fig.1b). Such differences might influence translational efficiency of the mRNA. Analogous variations in the consensus TATA-box sequence of eukaryotic mRNA gene promoters are known to influence transcriptional efficiency; moreover, the presence of a TATA-box sequence alone is not sufficient to specify a promoter [25]. The consensus sequence we propose for the ribosome recognition of mRNAs may behave similarly.

4. IS THE HYPOTHESIS COMPATIBLE WITH OTHER MODELS FOR RIBOSOME BINDING TO mRNAs?

The interaction between mRNA and 18 S rRNA proposed here overlaps with the interaction proposed in [3]. Although it is possible that both types of interaction could occur one after another, the latter requires the melting of the conserved hairpin in the rRNA (see fig.1a) and would not be energetically favourable. De Wachter [26] has found that the required complementary sequence for the model in [3] is not conserved in mRNAs at a statistically significant level.

The 'scanning' model of ribosome binding pre-

Table 1

Sequence comparison of eukaryotic mRNAs immediately upstream from the initiating AUG-codon

			Position relative to AUG-codon									
		Base	−7	−6	−5	−4	−3	−2	−1	+1	+2	+3
(a)	All mRNAs	A	50	27	27	19	86	36	34	102	0	0
		C	20	15	32	65	5	42	52	0	0	0
		G	6	29	12	6	11	6	11	0	0	102
		U	19	24	31	12	0	18	5	0	102	0
		Consensus	*A*	*x*	$C/_U$	*C*	*A*	$C/_A$	*C*	*A*	*U*	*G*
(b)	Globin mRNAs	A	16	1	12	1	15	0	1	19	0	0
		C	1	1	6	18	4	15	18	0	0	0
		G	0	15	1	0	0	1	0	0	0	19
		U	2	2	0	0	0	3	0	0	19	0
		Consensus	*A*	*G*	*A*	*C*	*A*	*C*	*C*	*A*	*U*	*G*
(c)	Viral mRNAs	A	15	10	6	10	27	15	9	41	0	0
		C	7	3	13	17	7	9	14	0	0	0
		G	5	16	6	10	5	4	9	0	0	41
		U	14	12	16	4	2	13	9	0	41	0
		Consensus	$A/_U$	*x*	$C/_U$	*c*	*A*	*x*	*x*	*A*	*U*	*G*
Proposed consensus sequence:			*A*	*U*	*C*	*C*	*A*	*C*	*C*	*A*	*U*	*G*

(a)102 eukaryotic, non-viral mRNA sequences and partial sequences drawn from [2]; 12–24 were aligned at their functional AUG-codons and base occurrence noted

(b)Alignment of 19 globin mRNA sequences used in the compilation in (a)

(c)Alignment of 41 viral mRNA sequences derived from [2]

sents fewer problems, but a number of mRNAs are known in which protein synthesis is initiated at a second or subsequent AUG-codon rather than the first the ribosome encounters downstream from the cap structure. The information as to which AUG-codon is to be used in these messengers is unlikely to reside solely in secondary structure as either exposed or basepaired AUG-codons may function as initiation sites [27]. In all of the 5 mRNA sequences of this type, for which unequivocal sequence information exists [2], stronger homologies (of 3–5 bases) to the proposed A–U–C–C–A–C–C consensus sequence are found immediately 5′ to the initiator AUG than to the unused AUG codons.

Clearly, the presence of the consensus sequence is not absolutely required for the translation of a mRNA. Rather, we suggest that the sequence and variations thereof may affect the rate and efficiency of translation from a particular AUG-codon. When Kozak showed that a purine in position –3 relative to AUG on an oligonucleotide is by itself sufficient to enhance binding to wheatgerm ribosomes, the rest of the oligonucleotide sequence was, fortuitously, in close agreement with our proposed consensus sequence [12]. More information may come from similar ribosome binding studies that use other synthetic oligonucleotides or messenger RNAs with mutations in the critical regions of their 5′-leader sequences.

ACKNOWLEDGEMENTS

We are grateful for the critical evaluation of these ideas by many of our colleagues, in particular Chris Konialis and Niall Dillon. The work of the authors' laboratory is supported by grants from the Medical Research Council, the Science and Engineering Research Council and the Wellcome Trust.

REFERENCES

[1] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[2] Kozak, M. (1981) Curr. Top. Microbiol. Immunol. 93, 81–123.
[3] Hagenbuchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell 21, 179–187.
[4] Azad, A.A. and Deacon, N.J. (1980) Nucleic Acids Res. 8, 4365–4376.
[5] Both, G.W., Furuichi, Y., Muthukrishnan, S. and Shatkin, A.J. (1976) J. Mol. Biol. 104, 637–658.
[6] Shatkin, A.J. (1976) Cell 9, 645–653.
[7] Kozak, M. (1980) Cell 22, 7–8.
[8] Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220–224.
[9] Van Charldorp, R. and Van Knippenberg, P.H. (1982) Nucleic Acids Res. 10, 1149–1157.
[10] Nakashima, K., Darzynkiewicz, E. and Shatkin, A.J. (1980) Nature 286, 226–230.
[11] Tinoco, I., Borer, P.N., Dengler, B., Levine, D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40–41.
[12] Kozak, M. (1981) Nucleic Acids Res. 9, 5233–5251.
[13] Roninson, I.B. and Ingram, V.M. (1982) Cell 28, 515–521.
[14] Martin, S.L., Zimmer, E.A., Davidson, W.S., Wilson, A.C. and Kan, Y.W. (1981) Cell 25, 737–741.
[15] Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Barrelle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653–668.
[16] Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) Nature 297, 335–337.
[17] Nakanishi, A., Teranishi, Y., Watanabe, Y., Natake, M., Noda, M., Kakidani, H., Jingiami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429–438.
[18] Torok, I. and Karch, F. (1980) Nucleic Acids Res. 81, 3105–3123.
[19] Ingolia, T.D. and Craig, E.A. (1981) Nucleic Acids Res. 9, 1627–1643.
[20] Nishioka, Y. and Leder, P. (1980) J. Biol. Chem 255, 3691–3694.
[21] Quinto, C., Quiroga, M., Swain, W.F., Nikovits, W.C., Standing, D.N., Pictet, R.L., Valenzuela, P. and Rutter, W.J. (1982) Proc. Natl. Acad. Sci. USA 79, 31–35.
[22] Kim, S., Davis, M., Sinn, E., Patten, P. and Hood, L. (1981) Cell 27, 573–581.
[23] Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Hirose, T., Inayama, S., Nakanishi, S. and Numa, S. (1982) Nature 295, 202–206.
[24] Schon, E.A., Wernke, S.M. and Lingrel, J.B. (1982) J. Biol. Chem. 257, 6825–6835.
[25] Tsai, M.J., Tsai, S.Y., Kops, L.E., Schulz, T.Z. and O'Malley, B.W. (1981) in: Developmental Biology Using Purified Genes, ICN–UCLA Symp. Mol. Cell Biol. (Brown, D.D. ed) vol. 23, pp. 313–329, Academic Press, New York.
[26] De Wachter, R. (1979) Nucleic Acids Res. 7, 2045–2054.
[27] Kozak, M. (1980) Cell 19, 79–90.