Hypothesis

A POSSIBLE INVOLVEMENT OF CODING SEQUENCES IN mRNA–RIBOSOME INTERACTION IN EUKARYOTES

Gerald W. BOTH

Molecular and Cellular Biology Unit, CSIRO, PO Box 184, North Ryde, NSW 2113, Australia

Received 26 February 1979

1. Introduction

It is now evident that eukaryotic mRNAs differ from each other in terms of size, poly(A) content, m^7G-cap structure, coding sequence, length of 5′ and 3′ non-coding regions and whether or not they are derived from longer molecular weight precursors by splicing. Nevertheless, many viral and cellular mRNAs utilize the same components for protein synthesis and might therefore possess some common structural features recognized during the initiation process. In fact, it has been proposed that each mRNA may possess an intrinsic ability to initiate protein synthesis determined by its overall ability to interact with the protein synthesis machinery [1,2]. However, despite the accumulation of base sequence data for viral and cellular mRNAs, the initiator codon AUG remains the only feature common to all eukaryotic mRNAs [3]. It is clear that the initiation of protein synthesis is a complex, multi-step process and the importance of a particular interaction may vary for individual mRNAs. For example, in vitro, the requirement for the m^7G-cap during initiation depends to some extent on the mRNA and the cell-free system used to translate it [4]. Some mRNAs, e.g., polio and STNV, do not possess a cap yet can initiate protein synthesis. Other features of the RNA must be recognised in these cases. For prokaryotes, it is known that during initiation, a pyrimidine-rich region conserved in 16 S rRNA base-pairs with purine-rich sequences in the mRNAs 5′ – proximal to the AUG codon, thus contributing to the stability of the mRNA–ribosomes interaction [5,6]. Similarly, a purine-rich sequence near the 3′ terminus of eukaryote 18 S rRNA is highly conserved and by analogy, a similar interaction has been suggested for recognition between eukaryote mRNAs and ribosomes [7]. Although there is no evidence to support its existence to date, many molecular mechanisms in protein synthesis have been preserved between prokaryotes and eukaryotes [8]. However, the proposed interaction in eukaryotes occurs less consistently and at highly variable distances from the initiator codon. Some mRNAs lack a region of complementarity completely [7], another point which weakens the argument for the proposed interaction.

2. Base-pairing in the coding sequence

Recent studies on the binding of reovirus mRNA fragments to ribosomes [9] prompted a re-examination of the potential for 18 S rRNA–mRNA interaction during initiation. The results with reovirus mRNAs showed that while an interaction with the AUG codon was of primary importance for ribosome binding, the m^7G-cap and sequences in the coding sequence of the mRNA could also contribute significantly to the efficiency of the process. Since some of the sequences implicated were very pyrimidine-rich, the possibility that they could pair with the conserved purine-rich region of 18 S rRNA, i.e., . . . UAGGAAGGCGU . . . , was examined for these and other mRNAs. Regions where 4 or more base-pairs (including G/U pairs) could be formed were considered, and the most stable arrangement according to the rules in [10] are depicted in fig.1. On the 5′ side of the AUG are the regions of complementarity described [7] and in addition, those within mRNAs recently sequenced. As noted earlier and

 Elsevier/North-Holland Biomedical Press

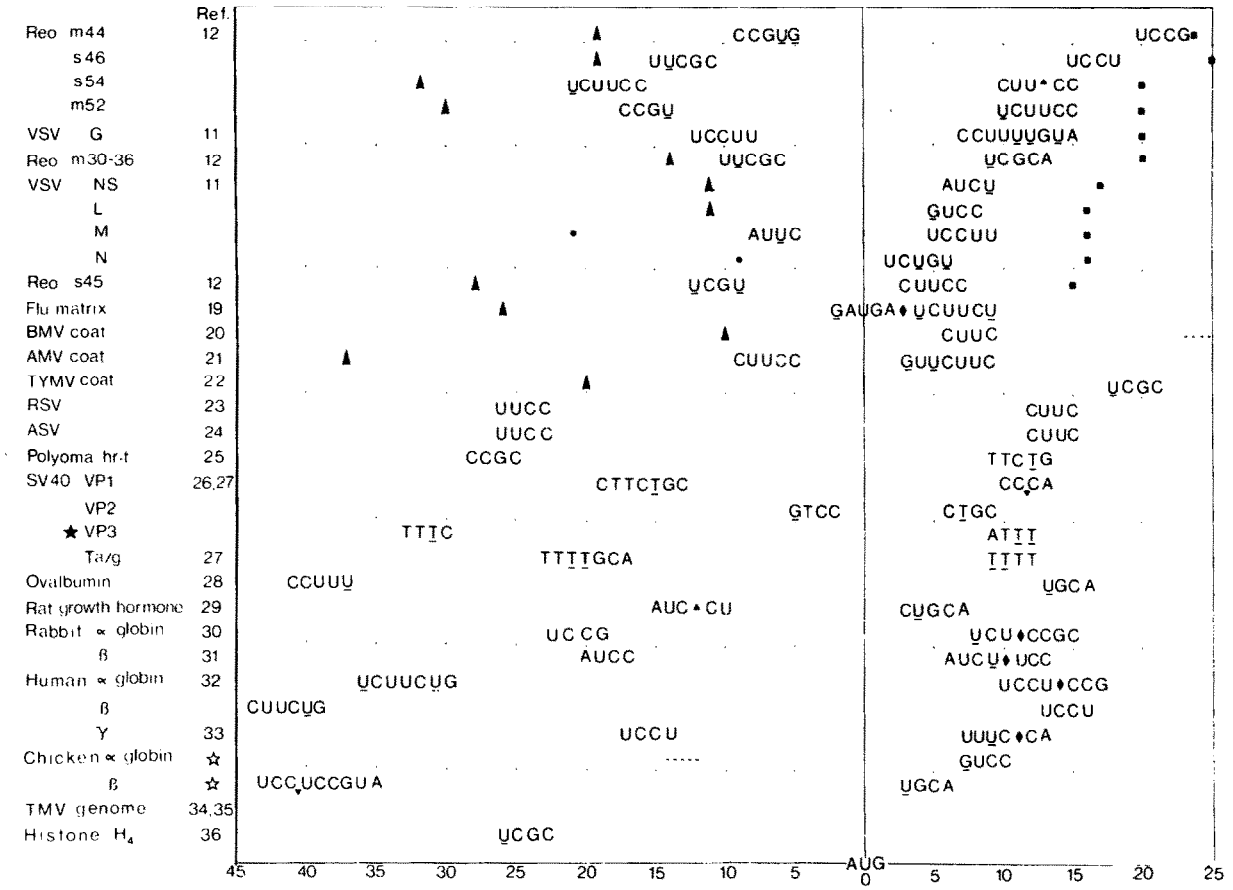

Fig.1. Possible sequences in mRNAs involved in base-pairing interactions with the 18 S rRNA sequence 3′ ... UAGGAAGGCGU ... 5′. mRNA sequences are aligned about the probable initiator codon. G/U base pairs are underlined; (▲) $m^7GpppN^{(m)}$-cap; (▲) 1 base bulge in the mRNA sequence; (▼) 1 base bulge in the rRNA sequence; (♦) 1 base loop out in the rRNA-mRNA double strand; (■) 3′ and (●) 5′ end of 40 S ribosome-protected mRNA sequence; (- - -) end of sequence determined; (☆) (R. I. Richards, J. Shine and J. R. E. Wells, personal communication); (★) The SV40 VP3 5′ complementary sequence may be altered when the precise location of splice points is known. A possible splice point occurs 3 bases from the AUG codon (G. Brownlee and M. Sleigh, personal communication).

confirmed with recent data, regions of potential pairing in the 5′ non-coding sequence occur at highly variable distances from the AUG codon and some, e.g., VSV N, NS and L, BMV coat and influenza matrix protein mRNAs lack complementary sequences completely. However, on the 3′ side of the AUG all mRNAs except TMV genome RNA (which lacks a complementary region on the 5′ side also) and sea urchin histone H4 are complementary to 4 or more bases in the 18 S rRNA sequence (fig.1). Although the distance of the complementary region from the AUG still varies somewhat, the variation is much smaller on this side. This is particularly evident for the globin mRNAs listed where complementary regions are spread between 14–45 and 3–17 bases on the 5′ and 3′ sides of the initiator triplet, respectively. For rabbit α-globin, the computer-generated secondary structure is drawn with the 3′ complementary sequence unpaired while the 5′ sequence is involved in intramolecular base pairings [30], consistent with the ideas proposed here.

In almost every case, the 3′ region of complemen-

tarity lies between 2–25 bases from the AUG codon, a region of mRNA which is part of the 40 S ribosome binding site for many messengers [9,11] (see fig.1). For those mRNAs where the extremity of the 40 S ribosomal subunit-protected region has been determined [11,12], the distance of the complementary region from the AUG codon correlates with the protected region of coding sequences, i.e., the shorter the protected region of coding sequences, the closer is the base-pairing site to the initiator codon (fig.1, top 11 lines). This suggests a functional interrelationship between the two and that the size of the ribosome binding site may vary depending on the structural features of the mRNA recognised by the attaching subunit. Thus, a base-pairing interaction between 18 S rRNA and mRNA is possible on either side of the AUG, but the required nucleotide sequence occurs more consistently and within a shorter distance on the 3′ side, compared to the 5′ side of the AUG codon. The proposed base-pairing, coupled with recognition of the m^7G-cap by a protein synthesis factor, may also explain, in part, the ability of synthetic ribopolymers m^7G cap–poly $(U)_n$, –poly $(U.C)_n$ and –poly $(A.C)_n$ to bind to wheat-germ 40 S ribosomes in vitro [2].

3. Interaction at the initiation codon

Of the 33 mRNAs considered in fig.1, there are only two exceptions, TMV genome RNA and histone H4 mRNA, which lack suitable complementary sequences. Hence this interaction cannot be obligatory for a stable mRNA–ribosome binding for all mRNAs. As noted [3], the AUG codon remains the only common feature of all mRNAs and its absence abolishes binding of mRNA fragments to ribosomes in vitro [9]. Hence, alignment of the initiator $Met.tRNA_f$ anti-codon sequence 3′ . . . UACC . . . 5′ [13,14] with the 5′ . . . AUG . . . 3′ codon may constitute the primary interaction between mRNA and ribosome. This interaction would presumably be stronger when the initiator sequence is . . . AUGG . . . and many mRNAs which show relatively weak, or no base-pairing at all, e.g., SV40 VP2, VP3 and T antigen mRNAs do contain the AUGG sequence. It has also been suggested [3] that the UAC codon in the 18 S rRNA sequence 3′ . . . AUUACUAG . . . 5′ may facilitate binding of the ribosome at the AUG site on the mRNA and this cannot be ruled out as an alternative to any role for the 18 S RNA in protein synthesis. However, of all the 16 possible sequences flanking the AUG codon only 10 are used so far and some of these, e.g., . . . CAUGG . . . are favoured (see table 1). In fact 26 of the 36 mRNAs examined here (including mouse α, β_A and β_B mRNAs [3]) contain the AUGG sequence. This bias argues for, if anything, interaction between the initiator tRNA anti-codon loop and AUG rather than one involving the UAC codon in 18 S rRNA.

Table 1
Sequences at the initiator AUG codon in eukaryotic mRNAs

Sequence	Occurrence
C AUG G	17
A AUG G	4
G AUG G	3
U AUG G	2
C AUG A	2
A AUG A	2
U AUG A	2
A AUG U	2
C AUG U	1
G AUG A	1

4. Discussion

Experiments with reovirus mRNAs have revealed that:

(1) The AUG codon is an important but not the sole recognition feature in an mRNA for ribosome binding;
(2) Additional features appear to be involved at the level of 40 S ribosome–mRNA interaction [9].

The involvement of a short sequence in the coding region of mRNAs capable of base-pairing with the conserved 18 S rRNA sequence is consistent with these observations. It is suggested that along with the m^7G-cap and protein synthesis factors, this interaction may increase the rate and stability of binding and serve to align the initiator $Met.tRNA_f$ carried on the 40 S subunit with the AUG codon. This idea could be tested by comparing mRNA binding to 40 S and 80 S ribosomes in the presence and

absence of an oligonucleotide complementary to 3′ . . . UAGGAAGGCGU . . . 5′, or part of it. A similar experiment has been done for the prokaryotic situation [15]. By choosing appropriate mRNAs, e.g., VSV M and NS mRNAs (see fig.1), it should be possible to inhibit binding of one or both mRNAs depending whether the complementary region on the 3′ or 5′ side of the AUG is important. Similarly an oligonucleotide complementary to the rRNA sequence 3′ $_{HO}$AUUACUAG . . . 5′ could be used to determine whether the UAC codon has a role to play in mRNA–ribosome recognition [3].

It must also be considered that this purine-rich sequence has no role in initiation but is conserved as a recognition signal for processing of 18 S rRNA from a larger molecular weight precursor [17]. Since the 28 S rRNA forms part of the same precursor [17], it might be expected to have a similar sequence for recognition by the processing enzyme. Determination of the sequence of ~100 bases at the 3′ ends of Vero cell and chicken 28 S rRNAs shows extensive sequence conservation between the two, but no sequence similar to that conserved in 18 S rRNA (unpublished results). This suggests that the purine-rich sequence may have some other function.

5. Concluding remarks

Comparison of available eukaryotic mRNA sequences has revealed that all mRNAs differ from each other except for the AUG codon. Despite this, ribosomes are able to translate a variety of cellular and viral mRNAs. Even heterologous mRNAs are translated with fidelity in vitro, e.g., the three reovirus *m* mRNAs can be translated in cell-free extracts of wheatgerm with relative efficiencies similar to that achieved during viral infection of mouse L cells [16]. Since the sequences of the ribosome binding sites of the *m* mRNAs are clearly different [12], the ribosomes and protein synthesis factors of wheat and mouse L cells must be able to cope with this variation in recognition sequences. This could be achieved if a protein factor or rRNA had the ability to recognise a variety of sequences in mRNA, yet still perform the same functions. Thus, while a sequence like 3′ . . . UAGGAAGGCGU . . . 5′ may be conserved in 18 S rRNA, only part of it might be used in the recognition of a particular mRNA binding site, the portion used being dependent on the mRNA sequence. Where the possibility for multiple interactions existed, the permissible one would be that which assured that the AUG and the Met.tRNA$_f$ were aligned for initiation. This proposal is analogous to, but more flexible than, the situation described for prokaryotes [5,6] where at least some part of the sequence . . . CCUCC . . . is required for mRNA–ribosome interaction [7]. The ability to recognise multiple structures in mRNAs need not be confined to this example, but could be an intrinsic part of other components of protein synthesis, e.g., initiation factor eIF-2 which appears to recognise internal mRNA sequences [18]. Thus the efficiency and rate of initiation for a particular mRNA would be an inherent property determined by the overall strength of the multiple interactions occurring during the initiation process.

This paper does not consider contributions, if any, to mRNA–ribosome interaction made by the respective secondary structures of the mRNA and the ribosomes. These may permit efficient interaction of protein synthesis components in ways not yet understood.

Acknowledgement

The author wishes to thank Dr G. G. Brownlee for critically reading the manuscript.

References

[1] Lodish, H. F. (1974) Nature 251, 385–388.
[2] Both, G. W., Furuichi, Y., Muthukrishnan, S. and Shatkin, A. J. (1976) J. Mol. Biol. 104, 637–658.
[3] Baralle, F. E. and Brownlee, G. G. (1978) Nature 274, 84–87.
[4] Shatkin, A. J. (1976) Cell 9, 645–653.
[5] Shine, J. and Dalgarno, L. (1975) Nature 254, 34–38.
[6] Steitz, J. A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734–4738.
[7] Hagenbuchle, O., Santer, M. and Steitz, J. A. (1978) Cell 13, 551–563.
[8] Weissbach, H. and Ochoa, S. (1976) Ann. Rev. Biochem. 45, 191–216.
[9] Kozak, M. and Shatkin, A. J. (1978) Cell 13, 201–212.
[10] Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973) Nature New Biol. 246, 40–41.
[11] Rose, J. K. (1978) Cell 14, 345–353.

[12] Kozak, M. (1977) Nature 269, 390–394.
[13] Piper, P. W. and Clark, B. H. C. (1974) Nature 247, 516–518.
[14] Simsek, M., Rajbhandary, U. L., Boisnard, M. and Petrissant, G. (1974) Nature 247, 518–520.
[15] Taniguchi, T. and Weissmann, C. (1978) Nature 275, 770–772.
[16] Both, G. W., Lavi, S. and Shatkin, A. J. (1975) Cell 4, 173–180.
[17] Choi, Y. C., Nazar, R. N. and Busch, H. (1975) in: The Cell Nucleus (Busch, H. ed) vol. 3, p. 109, Academic Press, New York.
[18] Kaempfer, R., Rosen, H. and Israeli, R. (1978) Proc. Natl. Acad. Sci. USA 75, 650–654.
[19] Both, G. W. and Air, G. M. (1979) Eur. J. Biochem. in press.
[20] Dasgupta, R., Shih, D. S., Saris, C. and Kaesberg, P. (1975) Nature 256, 624–628.
[21] Koper-Zwarthoff, E. C., Lockard, R., Alzner-DeWeerd, B., Rajbhandary, U. L. and Bol. J. F. (1977) Proc. Natl. Acad. Sci. USA 74, 5504–5508.
[22] Guilley, H. and Briand, J. P. (1978) Cell 15, 113–122.
[23] Haseltine, W. A., Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 989–993.
[24] Shine, J., Czernilofsky, A. P., Friedrich, R., Bishop, J. M. and Goodman, H. M. (1977) Proc. Natl. Acad. Sci. USA 74, 1473–1477.
[25] Friedmann, T., LaPorte, P. and Esty, A. (1978) J. Biol. Chem. 253, 6561–6567.
[26] Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weissman, S. M. (1978) Science 200, 494–502.
[27] Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voore, A., Van Heuverswyn, H., Van Herreweghe, J., Volchaert, G. and Ysebaert, M. (1978) Nature 273, 113–120.
[28] McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. and Brownlee, G. G. (1978) Nature 273, 723–728.
[29] Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. and Goodman, H. M. (1977) Nature 270, 486–494.
[30] Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M. and Salser, W. A. (1978) Cell 15, 43–54.
[31] Efstratiadis, A., Kafatos, F. C. and Maniatis, T. (1977) Cell 10, 571–585.
[32] Chang, J. C., Temple, G. F., Poon, R., Neumann, K. H. and Kan, Y. W. (1977) Proc. Natl. Acad. Sci. USA 74, 5145–5149.
[33] Chang, J. C., Poon, R., Neumann, K. H. and Kan, Y. W. (1978) Nucl. Acids Res. 5, 3515–3522.
[34] Richards, K., Guilley, H., Jonard, G. and Hirth, L. (1978) Eur. J. Biochem. 84, 513–519.
[35] Jonard, G., Richards, K., Mohier, E. and Gerlinger, P. (1978) Eur. J. Biochem. 84, 521–531.
[36] Grunstein, M. and Grunstein, J. E. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 1083–1092.