A single nucleotide is a sufficient 5′ untranslated region for translation in an eukaryotic in vitro system

Martin J.G. Hughes, David W. Andrews*

Dept. of Biochemistry, McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada

Received 26 June 1997

Abstract The 5′ untranslated region of an RNA molecule is thought to play an important role in the regulation of translation. Following a recent report that a single nucleotide is sufficient to act in this role in the unicellular organism *Giardia*, we show that this is also the case for a mammalian in vitro system. These results also demonstrate that an RNA can initiate translation from a start codon where an ideal translational consensus sequence is impossible.

Key words: Translation initiation; 5′ Untranslated region; Consensus sequence

1. Introduction

The leader of RNA molecules consists of three features important for efficient translational initiation. These include a cap structure, the start site sequence surrounding the initiating AUG (PuNNAUGG) [1–5] and an unstructured intervening 5′ untranslated region (5′-UTR) of at least 10 (and sometimes minimally 20) nucleotides [6,7]. Although there is good evidence for the importance of the purine in the −3 position (relative to the A of the initiation AUG), more recently the nucleotide at the +4 position has been shown to be of relatively less importance [5]. The leaders of most vertebrate cellular RNAs contain all three features [8]. However, if an AUG occurs closer to the 5′ cap than 10 nucleotides, it is not uniquely recognized as a start site and 'leaky scanning' occurs [6]. Leaky scanning leads to translational initiation from both this AUG and one or more AUG codons more distal to the cap structure. However, in this previous study, both AUG codons were positioned in an ideal 'consensus' with a purine in the −3 position as well as a G at the +4 position. It is not known if an AUG in a sequence that does not conform to the consensus sequence can cause leaky scanning if it is positioned closer to the cap structure than 10 nucleotides. In order for the first AUG to be positioned in a consensus start sequence there must be at least 3 nucleotides between the cap and the AUG. Because of the perceived importance of this start site sequence for translation initiation and because of the rarity of authentic leader sequences shorter than 10 nucleotides, translation initiation from AUG codons positioned less than 3 nucleotides from the cap structure has not been examined in detail.

A recent report suggesting that a single nucleotide 5′ UTR is sufficient to allow translation to initiate for the variant-specific surface protein of *Giardia* [9] was therefore surprising. *Giardia* are one of the simplest unicellular eukaryotes and very little is known about the regulation of translation in this organism. Therefore, it is possible that translation initiation close to the cap structure is peculiar to *Giardia*. To determine if translation initiation less than 3 nucleotides from the cap structure can occur in other eukaryotes we examined translation initiation close to the cap structure using a mammalian cell free system. Here we show that translation can initiate at an AUG codon one nucleotide 3′ of the cap structure, a position in the mRNA for which a consensus start site is impossible. This suggests that short 5′-UTRs can be used for the translation of mRNAs in multiple organisms. We also demonstrate that the AUG near the cap site need not be in a good consensus sequence for translation initiation in order for leaky scanning to occur.

2. Materials and methods

2.1. Plasmid constructs

Five clones were constructed which contained a combination of UTRs with one of two reporter genes, positioned behind the bacteriophage SP6 promoter (Fig. 1). The 5′-UTR sequences were based on the previously reported high expression UTK sequence [10]. The first (UTR-I) contained a single AUG start site in an ideal consensus context, 67 nucleotides downstream (+67) of the extreme 5′ end of the RNA. In the second (UTR-II), nucleotides were added at the 5′ end of the UTR, 3′ of the SP6 promoter, such that the transcript contained an AUG sequence at the +1 position of the RNA. The modifications mean that the AUG in the ideal consensus is now at +70. The frame of the polypeptide produced from the +1 position AUG start codon was set such that the translated product would read over the downstream start site and would translate a product of a size easily detectable by SDS-PAGE. The third (UTR-III) was identical to UTR-II except that a single nucleotide deletion shifts the frame of the encoded polypeptide initiating at position +1 such that the translated product consists of only 15 amino acids, with a stop codon occurring at nucleotide +47, within the 5′-UTR.

The two reporter genes encoded bovine pre-prolactin (PRL), or pre-prolactin fused to a double copy of an N-terminal epitope from the apoptosis related protein Bax (BPRL).

2.2. In vitro transcription and translation

RNA was produced from the plasmid templates in vitro using an SP6 polymerase based system (MBI Fermentas Inc., Hamilton, ON, Canada). The position of the 5′ end of the RNA molecules was confirmed by primer extension (performed using standard techniques, [11]).

*Corresponding author. Fax: (1) (905) 522-9033.
E-mail: andrewsd@fhs.mcmaster.ca

Abbreviations: Bax, Bcl-2 associated protein x; BPRL, 2 copies of the first 15 amino acids of Bax fused to the N-terminal of PRL; $BPRL_{0\ frame}$, product formed by translation of the BPRL gene from the authentic AUG codon; $BPRL_{+1\ frame}$, product formed by translation of the BPRL gene from an AUG situated 1 nucleotide from the 5′ cap of the RNA; PRL, bovine pre-prolactin; $PRL_{0\ frame}$, product formed by translation of the PRL gene from the authentic AUG codon; $PRL_{+1\ frame}$, product formed by translation of the PRL gene from an AUG situated 1 nucleotide from the 5′ cap of the RNA; UTK, high-expression 5′-UTR based on that of *Xenopus* β-globin

PII S0014-5793(97)00965-4

The different RNAs were added to a rabbit reticulocyte lysate (MBI Fermentas, Inc., Hamilton, ON, Canada) and allowed to translate for 1 hour, according to the manufacturers instructions. In this system, a 5′-cap structure is added to the mRNA by the lysate, enabling transcription to be performed without the addition of cap-analogue (data not shown). Following separation by SDS-PAGE, translation products were visualized and quantified using a phosphorimager (Molecular Dynamics, Sunnyvale, CA, USA).

2.3. Immunoprecipitation

The sequences encoding the amino terminus of Bax served to change the size of the proteins encoded as well as to permit immunoprecipitation of protein synthesized by translation of the mRNA in the 0-frame using Bax-specific antisera, using standard methods [12].

3. Results

RNA with UTR-I mediated efficient translation of a single protein product, corresponding to pre-prolactin ($PRL_{0\ frame}$), encoded by a predicted open reading frame of 229 amino acids. When RNA with UTR-II was translated, the major product was the same size as obtained from RNA with UTR-I. However, an additional band was also observed (compare Fig. 2, lanes 1 and 3). If the +1 site AUG was used as an initiation codon the predicted protein produced would contain 133 amino acids. There are no predicted open reading frames of this size in either the 0 or −1 frames. As can be seen, the novel polypeptide band in lane 3 (Fig. 2) migrates close to the expected size of a protein of 133 amino acids ($PRL_{+1\ frame}$). Relative translational efficiency suggests initiation at the +1 AUG is occuring between 30% and 50% as often as initiation at the favourable site at +70. If this new band indeed comes from a translation product due to initiation at position +1 of the RNA the band should disappear when RNA containing UTR-III was translated. In this RNA a single nucleotide deletion changes the translation frame for subsequent nucleotides. In the new frame translation terminates at nucleotide 47, after the synthesis of 15 amino acids (Fig. 1). A single protein product was obtained from this translation (Fig. 2, lane 2) which migrates at the same apparent molecular weight as the product translated from the RNA with UTR-I (Fig. 2, lane 1). Here, however, we observed that the product was translated approximately 20% less efficiently. We reasoned that this could be due to initiation at the +1 AUG site. Since the predicted product of translation initiating at this site would be a mere 1.7 kDa in size, we would not expect to visualize it using standard SDS-PAGE.

To confirm the identity of the $PRL_{0\ frame}$ and $PRL_{+1\ frame}$ bands, additional RNAs were translated. These RNAs encoded polypeptides with two copies of a Bax epitope at the amino terminus of the pre-prolactin molecule (BPRL). The sequence encoding the epitope was positioned such that the translation product would precipitate with antiserum to Bax only if translation was in the same frame as the 'authentic' pre-prolactin AUG start site. Since the +1 AUG is not in the same frame as PRL, translation products initiating at +1

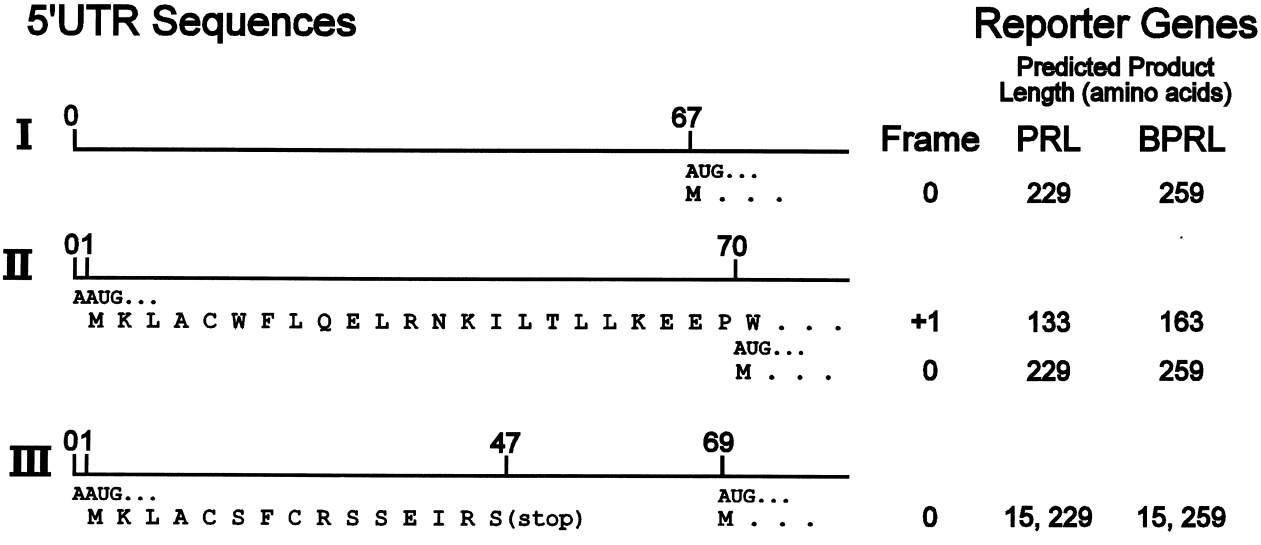

Fig. 1. Diagram of 5′ untranslated regions and reporter gene product lengths. DNA sequences encoding for three different 5′-UTR sequences (I, II and III) were inserted between an SP6 promoter and either of the two reporter genes, PRL or BPRL. The UTRs, based upon the high-expression UTK leader [10] were constructed by a combination of standard cloning techniques, oligonucleotide insertion and site-directed-mutagenesis [11]. In RNA with UTR-I the most 5′-AUG is located at nucleotide 67 (+67). This codon is located within an ideal consensus for translation initiation, which will result in a products of the indicated length. UTR-II differed from UTR-I in that additional nucleotides had been added to the 5′ end of the RNA. This introduced a second AUG, at position +1 in the UTR and shifted the second to +70. The N-terminal of the predicted amino acid sequence of the polypeptide translated from the +1 AUG is shown. The frame of the product translated from the upstream AUG is such that when fused with the PRL reporter it will read-through the reporter AUG start (+70) and continue translation to nucleotide 403 (in the +1 frame compared to the +70 AUG). UTR-III differs from II in that a single nucleotide has been removed at position +18, to change the frame of the product translated from the AUG in position +1. This creates a stop codon at position +47. The predicted amino acid sequence and position of the stop codon is shown. Each of the two reporter genes were fused to the 5′-UTRs. These were the genes coding for bovine pre-prolactin (PRL) and Bax-pre-prolactin (BPRL). BPRL consists of 2 copies of the N-terminal 15 amino acids of the apoptosis related protein Bax fused to the N-terminal of PRL. DNA encoding the Bax epitopes was inserted into the plasmids as oligonucleotides in such a way that a translated product (of the indicated number of amino acids) would be formed regardless of whether the coding sequence was in frame with the initiating AUG (0 frame), or in frame with that read from the AUG in position +1 (+1 frame). The number of amino acids of the predicted protein products from the various UTR-reporter gene fusions are shown. The identity of all clones was confirmed using Sequenase v2.0 (Amersham, Oakville, ON, Canada), and the 5′ end of the transcripts were mapped by primer extension, using a primer designed to anneal to the region corresponding to codons 4–11 of PRL.

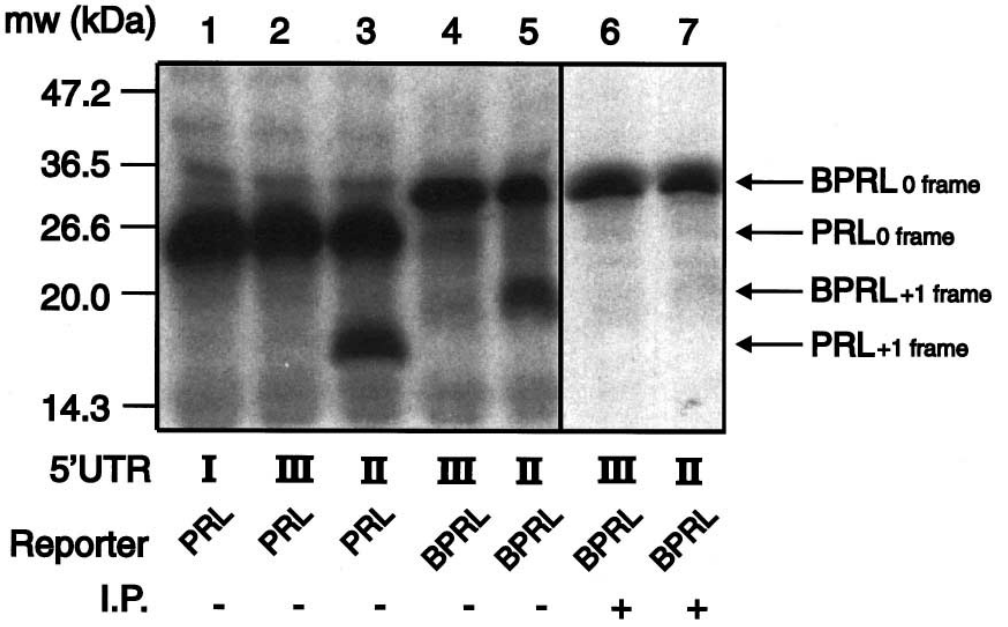

Fig. 2. Translation can initiate at an AUG codon with only 1 nucleotide of 5′-UTR. RNAs, transcribed in vitro (MBI Fermentas Inc., Hamilton, ON, Canada) and consisting of the leader/reporter fusions (see Fig. 1) as indicated, were translated in a rabbit reticulocyte lysate system (MBI Fermentas Inc., Hamilton, ON, Canada) in the presence of [^{35}S]methionine. Products of the translation were separated by SDS-PAGE on a 10% gel, and visualized using a phosphorimager (Molecular Dynamics, Sunnyvale, CA, USA). Pre-prolactin ($PRL_{0\ frame}$), translated from ribosomes initiating at the authentic start site is the major product in lanes 1–3. An additional band can be seen in lane 3, corresponding to the predicted translation product from the +1 AUG in the UTR-II/PRL RNA ($PRL_{+1\ frame}$). In lanes 4 and 5 the addition of 30 amino acids of Bax sequence to the N-terminal of PRL causes the $PRL_{0\ frame}$ product to decrease in mobility ($BPRL_{0\ frame}$). Similarly, the $PRL_{+1\ frame}$ product seen with UTR-II/PRL can be seen to decrease in mobility with the alternate reporter, BPRL ($BPRL_{+1\ frame}$). When the translation products shown in lanes 4 and 5 were immunoprecipitated (I.P.) with anti-Bax antisera [12], and the precipitated products analysed, only the $BPRL_{0\ frame}$ products could be seen (lanes 6 and 7), demonstrating that the $BPRL_{+1\ frame}$ protein was not produced in frame with the AUG situated at +70 of leader III, but was, instead, produced from translation initiating at +1.

would not be precipitable with antiserum to Bax. As can be seen in lane 4 (Fig. 2), when RNA with UTR-III and the BPRL coding region was translated the only observed product is of higher apparent molecular weight than the PRL product, as expected ($BPRL_{0\ frame}$). Furthermore, when RNA with the same coding region but containing UTR-II was translated two bands resulted both of which were of slightly slower mobility than the corresponding bands from RNA with UTR-II, encoding only PRL (compare Fig. 2, lane 3 with lane 5). As expected only the more slowly migrating species were immunoprecipitable with the Bax antiserum (compare Fig. 2, lanes 4 and 5 with 6 and 7) confirming that the additional translation product obtained from RNA with UTR-II ($BPRL_{+1\ frame}$) is not in the same frame as, and is thus unrelated to, translation of $BPRL_{0\ frame}$. This result is consistent with the band labelled $BPRL_{+1\ frame}$ resulting from translation initiation at nucleotide 2 of the RNA.

Thus, protein translation from RNAs with UTR-II or UTR-III can start following a 5′-UTR of only 1 nucleotide, demonstrating that in a mammalian in vitro translation system, translational initiation requires less than 2 nucleotides of UTR, and does not require a 'consensus' sequence surrounding the start codon. Furthermore, since two products are obtained (one from the +1 frame and the other from the 0 frame) we conclude that initiation is also occuring from the AUG at position +70, confirming 'leaky scanning' in UTR-II, and presumably, UTR-III.

Many of the features of in vitro translation systems have proven to accurately reflect translation in vivo. Translational efficiency, either as determined by the sequence surrounding an AUG start site, or at a non-AUG site has been shown to have a similar hierarchy both in vivo and in vitro [2,13,14]. The involvement of secondary structure in 5′-UTRs has been shown to be of equal importance in both cases [15–17], and in different eukaryotic systems [18]. The effect of short leader sequences has been examined both in vitro [6] and in vivo [7], though not directly compared. We predict, therefore, that the in vitro results obtained here have relevance to the in vivo situation.

4. Concluding remarks

Our results demonstrate that translation can be initiated in a position which would, according to accepted beliefs, be very unfavourable – there is no consensus sequence, and the 5′-UTR is exceptionally short. Since the variant-specific surface protein VSP1267 has also been shown to initiate translation in vivo with such a short 5′-UTR in *Giardia*, we suggest that this is a general phenomenon and that the criteria for translational initiation may not be as rigid as regularly reported. In addition, we note that a prokaryotic mRNA (that encoding for *Streptomyces fradiae* aminoglycoside phosphotransferase) has also been reported to have a 1–2 nucleotide 5′-UTR [19]. Furthermore, translation of mitochondrial mRNAs is well known to begin at the 5′ end [20]. Thus initiation very close to the 5′ end of the message may be a feature of translation in all organisms. However, since eukaryotic, prokaryotic and mitochondrial translation initiation operate by different mechanisms, it is also possible that the mechanisms responsible for initiation at the 5′ end of the RNA are also different. The

results described here using an ultra-short leader were obtained from in vitro experiments. Whilst this approach offers the advantage of a defined system (including an easily definable transcription start site), it remains to be determined how efficient initiation close to the 5′ end of an mRNA is in vivo.

Acknowledgements: This project was funded by a grant from Connaught Laboratories, Toronto, ON, Canada. D.W.A. is in receipt of an MRC Scientist award from the MRC of Canada.

References

[1] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[2] Kozak, M. (1986) Cell 44, 283–292.
[3] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148.
[4] Kozak, M. (1987) J. Mol. Biol. 196, 947–950.
[5] Grünert, S. and Jackson, R.J. (1994) EMBO J. 13, 3618–3630.
[6] Kozak, M. (1991) Gene Express. 1, 111–115.
[7] Slusher, L.B., Gillman, E.C., Martin, N.C. and Hopper, A.K. (1991) Proc. Natl. Acad. Sci. USA 88, 9789–9793.
[8] Kozak, M. (1994) J. Mol. Biol. 235, 95–110.
[9] Mowatt, M.R., Aggarwal, A. and Nash, T.E. (1991) Mol. Biochem. Parasitol. 49, 215–227.
[10] Falcone, D. and Andrews, D.W. (1991) Mol. Cell. Biol. 11, 2656–2664.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Press, Cold Spring Harbour, NY.
[12] Zhu, W., Cowie, A., Wasfy, G.W., Penn, L.Z., Leber, B. and Andrews, D. (1996) EMBO J. 15, 4130–4141.
[13] Kozak, M. (1989) Mol. Cell. Biol. 9, 5073–5080.
[14] Peabody, D.S. (1989) J. Biol. Chem. 264, 5031–5035.
[15] Kozak, M. (1986) Proc. Natl. Acad. Sci. USA 83, 2850–2854.
[16] Kozak, M. (1989) Mol. Cell. Biol. 9, 5134–5142.
[17] Pelletier, J. and Sonenberg, N. (1985) Cell 40, 515–526.
[18] Vega Laso, M.R., Zhu, D., Sagliocco, F., Brown, A.J.P., Tuite, M.F. and McCarthy, J.E.G. (1993) J. Biol. Chem. 268, 6453–6462.
[19] Janssen, G.R., Ward, J.M. and Bibb, M.J. (1989) Genes Dev. 3, 415–429.
[20] Montoya, J., Ojala, D. and Attardi, G. (1981) Nature 290, 465–470.