

Selection of CUG and AUG initiator codons for *Drosophila E74A* translation depends on downstream sequences

(translational control/protein isoforms/gene regulation)

LYNN BOYD* AND CARL S. THUMMEL

Howard Hughes Medical Institute, Department of Human Genetics, 5200 Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112

Communicated by Mario R. Capecchi, July 9, 1993

ABSTRACT Selection of a translation initiation site is thought to be determined by relative proximity to the 5' end and sequence context of a potential initiator codon. These guidelines seem insufficient to explain translation of the *Drosophila E74A* mRNA, whose 5' untranslated region is exceptionally long (1.8 kb) and contains many AUG triplets preceding the long open reading frame. In an effort to understand how the appropriate initiator codon is chosen, we have undertaken a study of *E74A* translation in transfected *Drosophila* cells. The results show that translation of the *E74A* protein utilizes at least three initiator codons: two minor forms of the protein are initiated at a CUG and an AUG, while the most abundant form initiates at a CUG. This main initiator CUG appears to be in a good context; however, it lies downstream of 17 AUG and 24 other CUG codons, several of which are also in good contexts. Unexpected results were obtained from sequence perturbations upstream and downstream of the main CUG initiator. Creating an AUG with a good context 72 bases 5' to the main CUG has only a modest inhibitory effect on initiation frequency at that CUG. Replacing sequences 44 bases 3' to the main CUG has an inhibitory effect on its use as an initiator as well as on the CUG 72 bases further upstream. These results indicate that factors other than context and relative proximity to the 5' end must be involved in initiator codon selection and may include elements such as secondary or tertiary structure of the RNA.

The scanning model for translation proposes that ribosomes initiate protein synthesis at the AUG codon that is in a good sequence context and closest to the 5' end of the mRNA (1). This model predicts that sequences upstream of a potential initiator codon can have inhibitory effects on its efficiency in initiating translation and that sequences downstream should have no effect. These predictions have been largely supported by analyses of naturally occurring RNAs and by manipulation of known sequences (1). However, there are mRNAs that do not conform to the established guidelines because they either ignore presumably favorable AUGs or they initiate translation at non-AUG codons (2–9).

Studies comparing the efficiency of initiation at non-AUG codons have found that they can be used at up to 50% the rate of an AUG *in vitro* and up to 15% in cultured cells (9–11). The use of non-AUG initiators seems to be dependent upon their context (9, 12). In addition, efficiency of initiation at a non-AUG start was enhanced by placing a sequence capable of forming a hairpin structure 14 nucleotides downstream (13). The secondary structure might inhibit ribosome transit and cause ribosomes to pause when positioned over the potential initiator, thus facilitating initiation. Although significant stimulation was observed in this artificial RNA, the use of this strategy in a naturally occurring mRNA has not been demonstrated.

This report describes the use of DNA transfections into *Drosophila* cultured cells to define regions of the *E74A* mRNA required for proper translation initiation. *E74A* is one of two nested transcription units encoded by the *Drosophila* ecdysone-inducible *E74* gene (14). *E74A* encodes a site-specific DNA-binding protein that is thought to function in a regulatory manner at the onset of metamorphosis (15). Extensive structural studies have identified only a single *E74A* mRNA, which contains an unusually long, 1.8-kb 5' leader (14, 16). Using a cDNA to express *E74A* protein in cultured cells, we found that the most abundant form of the *E74A* protein initiates at a CUG codon. The other two isoforms initiate at a CUG and an AUG codon. The relative frequency of initiation depends upon sequence elements >100 nucleotides downstream of the first CUG. These results suggest that features other than context, possibly RNA secondary or tertiary structures, may play a significant role in initiator codon selection. In addition, this study confirms previous observations that *Drosophila* genes can use non-AUG codons to initiate translation (17–19) and suggests that the use of these initiators may be more widespread than was previously thought.

MATERIALS AND METHODS

Expression Plasmids. *E74A* sequence from a full-length *E74A* cDNA (14) was inserted into the pPAC expression vector (20) at the *Bam*HI restriction site, 88 bp downstream from the actin 5C (*Act5C*) transcription start site. The *E74A* sequence included the entire 5' untranslated region (UTR) and the first 32 bases of the 1625-base 3' UTR. Point mutations were generated by PCR amplification using complementary oligonucleotides with the desired mutations and outlying primers as described (21). *Spe* I and *Xba* I restriction sites were used for subcloning the point mutations into the original pPAC-*E74A* construct. Point mutations were sequenced before use in transfection assays. The pAcLUC-*E74A* fusion construct was made by replacing the *Eco*RV-*Xba* I fragment from pPAC-*E74A* with an *Eco*RV-*Xba* I fragment from pAC-CAT/*E74A*/LUC (16), using an *Xba* I site in the luciferase coding region. The CUG point mutations were introduced into pAcLUC-*E74A* by using the *Hpa* I restriction sites at *E74A* nucleotides 1175 and 1856 (14).

Cell Transfections and Western Blots. The *E74A* expression plasmids were transfected into two *Drosophila* cell lines, Schneider line 2 and Kc167. Results from the two cell lines were indistinguishable; data shown are from Kc167 cells. Twelve micrograms of each DNA and 12 μ g of herring sperm DNA in 300 μ l of 250 mM CaCl₂ were added to 300 μ l of 2 \times Bes-buffered saline (BBS: 1 \times is 140 mM NaCl/0.75 mM Na₂HPO₄/25 mM Bes, pH 6.95). Five hundred microliters of this mix was added to $\approx 1 \times 10^7$ Schneider line 2 cells in 5 ml

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: UTR, untranslated region.

*Present address: Section of Genetics and Development, Cornell University, Ithaca, NY 14853.

of Schneider's medium (GIBCO) supplemented with 10% fetal bovine serum (GIBCO) or $\approx 1.5 \times 10^7$ Kc167 cells in M3 medium (GIBCO) supplemented with 5% fetal bovine serum. The cells were incubated at 25°C for 2 days and then harvested, washed once with phosphate-buffered saline, and suspended in 250 μ l of 10 mM Tris, pH 7.8/1 mM EDTA and 250 μ l of 2 \times SDS sample buffer. Twenty-five microliters of sample was loaded into each lane of an SDS/polyacrylamide gel. Western blot analysis was performed as previously described, using the same anti-E74A antibodies (22). For estimates of the relative amounts of the different isoforms, the films were scanned with a Molecular Dynamics laser scanning densitometer.

RESULTS AND DISCUSSION

Fig. 1 shows the sequence surrounding the proposed E74A translation initiation sites and indicates where the point mutations were generated. Expression plasmids containing the E74A coding region with various portions of the 5' UTR were transfected into cultured cells, and Western blots were used to assess the portion of the UTR required for proper translation initiation. Fig. 2a shows that the full-length, 1891-nucleotide leader yielded the same pattern of three E74A protein isoforms that had previously been observed *in vivo* (22). Removing 1727 nucleotides of the leader did not affect this pattern (*Spe* I), whereas removal of 1856 nucleotides of the leader yielded only the smallest isoform (*Hpa* I).

To determine whether the two larger E74A protein isoforms were due to initiation at CUG codons encoded between the *Spe* I and *Hpa* I restriction sites, we mutagenized the three CUG codons that lie in the appropriate reading frame (specific changes are shown in Fig. 1). CUG-1 and CUG-3 are good candidates for initiator codons since both of them conform to the consensus initiator context for *D. melanogaster* determined by Cavener and Ray (23) (CANNAUGN). In addition, both these codons have purines at positions -3 and +4, which have been shown to be important for efficient initiation in vertebrate systems (24).

Changing CUG-1 to a CUC codon eliminated the uppermost band on Western blots (Fig. 2b), suggesting that this isoform was due to translation initiation at CUG-1. Surprisingly, the same isoform was eliminated by changing CUG-2

to CUC. The context of this CUG does not match the consensus sequence as well as CUG-1 and, thus, would not be expected to function as efficiently as a start codon. As expected, a stop codon created by changing a UGG downstream of CUG-2 to UAG also eliminated the largest isoform. It is unclear why mutations at CUG-1 and CUG-2 have the same effect. Multiple exposures of Western blots show at least a 95% reduction in the amount of the largest isoform as a result of each mutation. Therefore, these two potential initiator codons cannot be functioning independently of one another. It is possible that one mutation affects a context or a secondary structure that favors initiation at the other codon. Alternatively, the presence of two potential initiator codons itself may favor initiation at one site. For example, the first codon may cause the ribosome to stall, which may enhance initiation at the second codon.

Western blot analysis of protein extracts prepared from both transfected tissue culture cells (Fig. 2a) and staged animals (22) revealed that the ratios of the E74A protein isoforms were the same in these two systems and that the middle protein isoform was the most abundant. We assume that the antibody used for these Western blots detected each isoform equally and that the steady-state level of the proteins was an accurate reflection of their relative initiation frequency (e.g., that the half-lives are the same for all isoforms). Changing CUG-3 to CUA eliminated the most abundant E74A isoform (Fig. 2b). This mutation also led to an increase in the smallest isoform. This suggests that in the wild-type E74A mRNA, CUG-3 engages most of the competent ribosomes and, in the absence of CUG-3, these ribosomes are free to continue scanning and initiate at the AUG. That the smallest isoform corresponds to a protein initiated at the AUG was demonstrated by changing that codon to AUC (Fig. 2b). The ACG codon immediately upstream of CUG-3 was also considered a candidate for a translation initiation site. Changing this codon to ACC resulted in only a slight decrease in the abundance of the middle isoform (data not shown).

The results from the CUG-3 mutation are in agreement with a process of ribosomal scanning. It is possible that ribosomes scan through this region of the mRNA and, due to leaky scanning, a small proportion initiate at CUG-1. Of the ribosomes that bypass CUG-1, a much larger proportion initiate at CUG-3 and a small pool is left to initiate at the

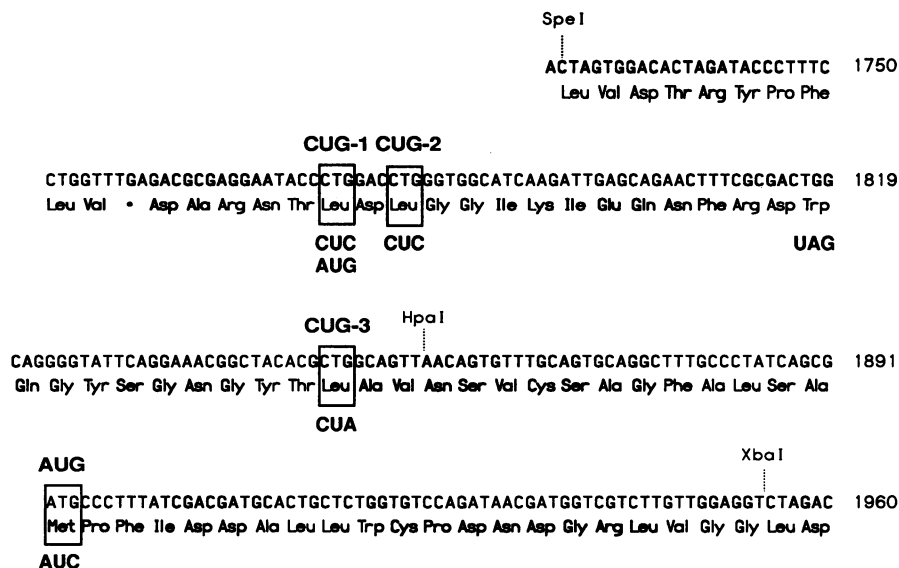


FIG. 1. Nucleotide sequence of a portion of the E74A mRNA (nucleotides 1726–1960) (14), showing the initiator codons and constructed point mutations. The CUG and AUG initiator codons are boxed. The sequences of mutated codons are listed below the DNA sequence. Although RNA sequences are shown for the mutations, manipulations and sequence analysis were performed at the DNA level. Three nucleotides in this sequence differ from the published version (14).

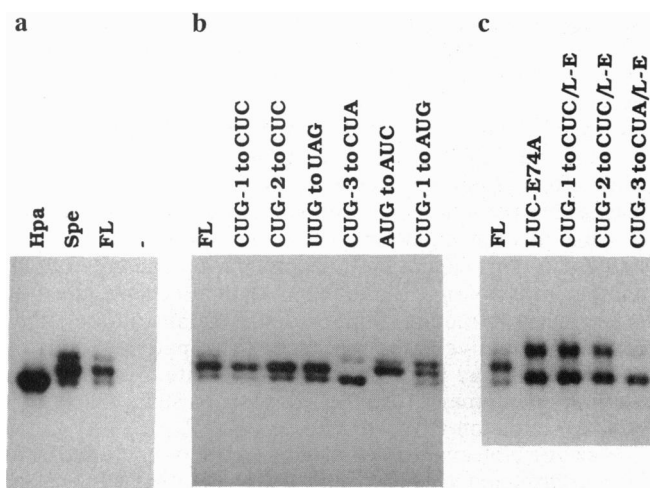


FIG. 2. Western blots of extracts from cultured cells transfected with wild-type and mutagenized cDNA expression constructs. Full-length, deleted, or mutated *E74A* cDNAs were inserted downstream from the *Act5C* promoter and transfected into *Drosophila* cultured cells. All constructs included only the first 32 bases of the 1625-base *E74A* 3' UTR. (a) The full-length *E74A* leader (FL), a deletion of nucleotides 1–1727 (*Spe*), a deletion of nucleotides 1–1856 (*Hpa*), and a control transfection with no *E74A*-expressing DNA (–). (b) Point mutations introduced into the *E74A* cDNA with the full-length leader were used to identify initiator codons. Mutations are depicted in Fig. 1. (c) Replacement of the *E74A* AUG and 20 codons with the AUG and 17 codons from firefly luciferase changes the pattern of initiator codon selection. The fusion construct with the wild-type full-length *E74A* leader (LUC-*E74A*) and with mutations at CUG-1, CUG-2, and CUG-3 were used. /L-E indicates that these mutations were inserted into the luciferase-*E74A* fusion construct. The mobility of the two bands is less than expected, most likely due to the basic amino acids encoded in the luciferase sequence.

AUG. To further test whether initiation of these protein isoforms was in agreement with the scanning hypothesis, we replaced CUG-1 with an AUG codon (Fig. 2*b*). Because of the favorable context of CUG-1 and what has been observed with similar mutations (3, 4, 12, 24), it was expected that the level of the largest isoform would increase and the amounts of the smaller forms would decrease severalfold. When adjusted for the amount of total protein, there was a 2-fold increase in the amount of the largest isoform. However, only a 20–30% decrease in the smaller forms was detected. Thus, an AUG at this position does not appear to efficiently recruit scanning ribosomes.

There are at least four possible explanations for this scenario. The simplest explanation is that the C at position –4 is critical for providing a good context for translation initiation. We have tested this possibility by changing the four nucleotides upstream from CUG-1 to be identical to those upstream from CUG-3 and saw no increase in the level of the largest isoform (data not shown). Second, there could be some previously undefined aspect of the sequence context surrounding CUG-3 but not CUG-1. Third, it is possible that some secondary or tertiary structure of the mRNA promotes initiation at CUG-3 during scanning. Fourth, initiation at CUG-3 could be independent of scanning. For instance, an internal ribosome entry site could direct ribosomes to CUG-3. Consistent with this possibility, we have identified sequences that promote internal ribosome entry in the *E74A* 5' leader (L.B., C.S.T., and P. Sarnow, unpublished results). However, this activity requires sequences upstream of nucleotide 1628, sequences that are not required for efficient initiation at CUG-3.

If the favoritism for CUG-3 is operating via some structure in the mRNA, it should be possible to dissect where those sequences reside. It is clear that nucleotides 1–1727 are not

required (Fig. 2*a*). As an initial step in identifying sequences important for preferential initiation at CUG-3, nucleotides 1890–1955 of *E74A* were replaced with 51 nucleotides from firefly luciferase. This resulted in the replacement of the *E74A* AUG and 20 downstream codons with the AUG and 17 downstream codons from luciferase. This construct yielded two proteins detectable with anti-*E74A* antibodies on Western blots (Fig. 2*c*), indicating that only two start sites were utilized. To determine which initiator was bypassed in the fusion mRNA, the CUG-1, CUG-2, and CUG-3 mutations were each inserted into the fusion construct. The CUG-1 and CUG-2 mutations had no effect on the detected protein products. However, the CUG-3 mutation eliminated most of the larger isoform (Fig. 2*c*). Thus, replacing nucleotides 1890–1955 with foreign sequences negatively affected initiation at CUG-1 or CUG-2. This was not expected, since the sequence replaced was >100 nucleotides downstream from the affected initiator codons. Also, there was more of the smaller isoform than the larger one, indicating that the fusion construct resulted in a decrease in the frequency of initiation at CUG-3 relative to initiation at the AUG.

The changes in initiation at the CUG codons in the luciferase-*E74A* fusion could result from changes in a previously favorable secondary or tertiary structure(s). This is consistent with the demonstration that downstream secondary structure can exert a positive effect on initiation at non-AUG codons (13). Indeed, a potential helix exists 16 nucleotides downstream of CUG-3. Determination of whether higher-order RNA structures are formed and whether they play a role in *E74A* initiator codon selection will require structural studies of the *E74A* mRNA.

This report adds to a rapidly growing list of mRNAs that yield more than one protein product via initiation at non-AUG codons (3–8). Although non-AUG codons are generally weak initiators (9, 11, 12), it appears that the CUG in *E74A* may take advantage of cis-acting factors that enhance initiation. Apparently, these cis factors can be downstream of the initiator, reminiscent of the complex features associated with the control of transcription. By modulating the activity of these cis elements (e.g., with RNA-binding proteins), the cell could regulate the level of initiation at various codons. The furnishings for this type of initiation are not specific to *Drosophila*, since the same three protein isoforms were synthesized in a rabbit reticulocyte *in vitro* translation system (data not shown).

The possibility that the multiple *E74A* protein isoforms generated by the use of different initiator codons may be functionally significant is supported by an analysis of the *E74* gene from other *Drosophila* species. A comparison of sequences from *D. melanogaster*, *D. pseudoobscura*, and *D. virilis* shows that conservation of sequences downstream and including CUG-3 occurs in all three species (C. W. Jones, personal communication). Positions –3, –2, and –1 are also conserved. In *D. virilis*, the sequence conservation decreases dramatically upstream of CUG-3. In *D. pseudoobscura*, CUG-1, CUG-2, and the open reading frame are highly conserved, consistent with this region representing a translated, protein-coding region. These observations seem particularly informative given that the sequence of most of the 5' UTR is highly diverged amongst the three species. The conservation of these sequences supports the assertion that these CUG codons are used for translation initiation *in vivo*. A specific balance of the three isoforms may be important for the proposed regulatory function of the *E74A* proteins. Our results indicate that this balance may be achieved via differential translation initiation.

We thank Lisa Urness for providing the *Spe* I 5'-leader-deletion construct, C. Weldon Jones for generously sharing his unpublished results, and Lucy Cherbas for advice on tissue culture transfection.

This research was supported by the Howard Hughes Medical Institute.

1. Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241.
2. Oh, S.-K., Scott, M. P. & Sarnow, P. (1992) *Genes Dev.* **6**, 1643–1653.
3. Lock, P., Ralph, S., Stanley, E., Boulet, I., Ramsay, R. & Dunn, A. R. (1991) *Mol. Cell. Biol.* **11**, 4363–4370.
4. Saris, C. J. M., Domen, J. & Berns, A. (1991) *EMBO J.* **10**, 655–664.
5. Acland, P., Dixon, M., Peters, G. & Dickson, C. (1990) *Nature (London)* **343**, 662–665.
6. Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W. & Eisenmann, R. N. (1988) *Cell* **52**, 185–195.
7. Florkiewicz, R. Z. & Sommer, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3978–3981.
8. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R. & Charnay, P. (1990) *Mol. Cell. Biol.* **10**, 3456–3467.
9. Mehdi, H., Ono, E. & Gupta, K. C. (1990) *Gene* **91**, 173–178.
10. Dasso, M. C. & Jackson, R. J. (1989) *Nucleic Acids Res.* **17**, 6485–6497.
11. Peabody, D. S. (1989) *J. Biol. Chem.* **264**, 5031–5035.
12. Kozak, M. (1989) *Mol. Cell. Biol.* **9**, 5073–5080.
13. Kozak, M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8301–8305.
14. Burtis, K. C., Thummel, C. S., Jones, C. W., Karim, F. D. & Hogness, D. S. (1990) *Cell* **61**, 85–99.
15. Urness, L. D. & Thummel, C. S. (1990) *Cell* **63**, 47–61.
16. Boyd, L. (1993) Ph.D. thesis (Univ. of Utah, Salt Lake City).
17. Sugihara, H., Andrisani, V. & Salvaterra, P. M. (1990) *J. Biol. Chem.* **265**, 21714–21719.
18. Bellen, H. J., Kooyer, S., D'Evelyn, D. & Pearlman, J. (1992) *Genes Dev.* **6**, 2125–2136.
19. Desimone, S. M. & White, K. (1993) *Mol. Cell. Biol.* **13**, 3641–3649.
20. Krasnow, M. A., Saffman, E. E., Kornfeld, K. & Hogness, D. S. (1989) *Cell* **57**, 1031–1043.
21. Higuchi, R. (1990) in *Recombinant PCR*, eds. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, San Diego), pp. 177–183.
22. Boyd, L., O'Toole, E. & Thummel, C. S. (1991) *Development* **112**, 981–995.
23. Cavener, D. R. & Ray, S. C. (1991) *Nucleic Acids Res.* **19**, 3185–3192.
24. Kozak, M. (1986) *Cell* **44**, 283–292.