

DNA Sequences Homologous to the *Drosophila* opa Repeat Are Present in Murine mRNAs That Are Differentially Expressed in Fetuses and Adult Tissues

DENIS DUBOULE,* MARC HAENLIN, BRIGITTE GALLIOT, AND ELIANE MOHIER

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 67085 Strasbourg Cédex, France

Received 15 December 1986/Accepted 30 January 1987

A mouse embryonic cDNA containing two opa-like (CAX)_n repeats was isolated on the basis of its cross-hybridization with a *Drosophila* K10 cDNA. Such repeated sequences were present in different murine mRNAs, some of which were specifically expressed during fetal life or in different adult tissues. This suggests that, as already described for *Drosophila*, opa-like sequences are parts of proteins involved in ontogenic or cell-type-specific functions in vertebrates. However, unlike *Drosophila*, such repeated sequences were not found within the murine homeo-boxes containing genes of the Hox-1 complex.

Several *Drosophila* developmentally regulated genes contain one or more copies of the repeated motif, (CAX)_n, where X is either a G or an A and $n \leq 30$. These genes include homeotic genes such as *Antp* (17) and *Dfd* (16), genes involved in establishing dorsoventral polarity such as *K10* (F. Deryckère, personal communication) and *twist* (B. Thisse, personal communication), as well as other developmentally important genes such as *engrailed* (9), *ftz* (11), or *Notch* (20). Such repeats were also found within the *Ubx*, *Iab-2*, and *Iab-7* regions (16). In several cases, these sequences are a part of the coding region and encode stretches of glutamine residues. These repeated DNA sequences, called M repeats (17) or opa repeats (20), are present in rather high amounts within the *Drosophila* genome. Transcriptional studies reveal that the population of mRNAs containing opa-like sequences is different depending on the developmental stage, suggesting that genes containing such sequences are somehow involved in *Drosophila* development (20). For the homeotic genes, it has been hypothesized that such runs or polyamino acids could serve as linking or spacing units within a given homeotic gene product (11) as previously suggested concerning the hinge (poly-Gly) sequence found in the *Ubx* proteins (4). They may also function during the putative interactions between several such protein monomers in establishing spatial multimeric structure. We looked for the presence of homologous sequences in the mouse genome and their possible relation to recently described murine homeo-boxes. We report here that opa-like repeats also exist in the mouse genome and that they are expressed in mRNAs whose transcription is regulated in a tissue and ontogenetic manner, but, unlike the case of *Drosophila*, we show that these repeats are not commonly parts of murine proteins containing homeo-domains.

We used a *Drosophila* K10 cDNA clone (M. Haenlin and E. Mohier, unpublished data) containing opa-like sequences to screen a mouse embryonic cDNA library prepared in λ gt10 (12) from poly(A)⁺ RNA extracted (2) from 150 embryos at day 10. Three positive clones were isolated by hybridization under low-stringency conditions (37% formamide, 42°C, 0.1% sodium dodecyl sulfate, 5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]), and one of

them, a 650-base-pair (bp) clone, was further analyzed (Fig. 1). The sequence of this clone revealed the presence of two repeated motifs homologous to the *Drosophila* Notch opa box (20). Both repeats were 163 bp long and were composed almost exclusively of CAG or CAA triplets. These two unusual sequences were therefore called opa-1 and opa-2, respectively, and the cDNA clone was called Mopa cDNA (Fig. 1). Because of the presence of these very large repeats, the three possible reading frames extended over more than 350 bp. Contrary to what has been found in some *Drosophila* cDNAs containing opa sequences (20), no termination codons were found either within these motifs or in the immediate proximity. In addition, Fickett's test (8) indicated that a portion of this Mopa cDNA clone had a high probability of encoding a protein product. After a stop codon (Fig. 2, position 420), coding sequence probability decreased significantly, suggesting that the Mopa cDNA represents the 3' end of an mRNA coding sequence. One of the possible translation products showed stretches of 34 (opa-1) and 32 (opa-2) glutamine residues (Fig. 2). The two boxes are separated by 13 amino acids, 7 of which were glutamine residues placed in a row (Fig. 2). When compared with the Notch opa box, each of these repeats showed a very high degree of homology as exemplified by opa-2 in which 84% of the nucleotides (Fig. 2, black dots) and 93% of the amino acids are identical to the Notch box, and in addition 95% of the nucleotide mismatches occur at the third position (Fig. 2, open dots). The lengths of opa-1 and opa-2 are similar to the poly-Gln sequence of Notch and other *Drosophila* cDNAs containing opa repeats (cDNA-S; cDNA-vs1) (20), as is the presence of a histidine (CAT) residue either within the box (opa-2) or at its border (opa-1) (Fig. 1, large stars). Outside the repeated structures, the putative Mopa protein is very rich in proline (13%) and glutamine (37%) residues that are often paired, as in the *Drosophila* *Antp* gene (17). A 3' poly(A) tail is also present starting 30 bp downstream from a potential polyadenylation signal (AATGAA; Fig. 2).

To see whether opa-like sequences are present in a high number of copies in the mouse genome and whether they are transcribed into mRNAs, we used a probe containing the two repeats plus about 150 bp of 3'-flanking region (Fig. 1, probe A). When this probe was hybridized under high-stringency conditions (50% formamide, 42°C) to a Southern

* Corresponding author.

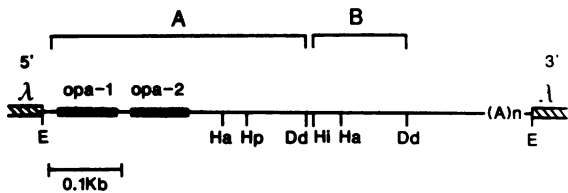


FIG. 1. Restriction map of the Mopa cDNA clone. This 600-bp cDNA clone represents the 3' part of a large transcript (see Fig. 4) as judged by the presence of a poly(A) tail. The following restriction sites are indicated: *EcoRI* (E); *DdeI* (Dd); *HaeIII* (Ha); *HpaII* (Hp); and *HinfII* (Hi). The two CAX repeats (opa-1 and opa-2) are located in the 5' part of this partial cDNA clone. The different probes used for Southern and Northern blot experiments are shown on the top (A and B).

blot of *EcoRI*-restricted mouse genomic DNA (20 μ g), numerous copies of opa-like sequences were found in the mouse genome (Fig. 3). However, only a small number of strongly hybridizing bands were observed (Fig. 3, arrowheads), suggesting that only a few repeats show high degrees of homology with the mouse opa sequences contained in the Mopa cDNA clone. The smear was presumably due to either shorter or less homologous sequences.

Poly(A)⁺ and poly(A)⁻ RNAs extracted from mouse embryos and fetuses from days 10 to 15 as well as total RNA from various organs were separated on methyl mercury gels (3) and transferred onto diazobenzylxymethyl (DBM) papers (1). Hybridization (19) at high stringency (50% formamide, 45°C) of probe A to fetal RNA (days 15 [15D], 13 [13D], and 11 [11D]) showed that several poly(A)⁺ RNAs contained sequences homologous to the opa box. Two of these transcripts were very abundant, while 10 others were distinguishable from the background smear (Fig. 4). No detectable

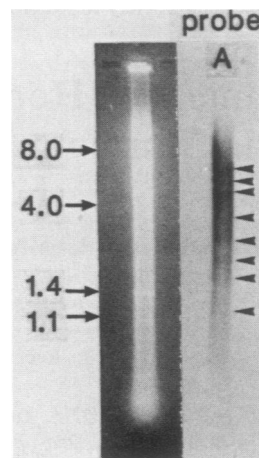


FIG. 3. Presence of opa-like repeats in the mouse genome. The hybridization of a mouse genomic blot with the opa-containing probe A shows that a large number of copies of opa-like repeats are present in the mouse genome as judged by the presence of a smear. However, only a restricted number of these sequences show a high degree of homology with the opa box and therefore appear as discrete bands (arrowheads). The filter was exposed for 3 days with Kodak X-Omat film and intensifying screens. Numbers on the left are kilobases.

hybridization product was found in poly(A)⁻ RNA. In contrast to what is observed in *Drosophila* (20), the mouse opa RNAs did not show any developmentally specific expression, at least during the different fetal stages investigated (from days 11 to 15, Fig. 4). The presence of discrete bands in the poly(A)⁺ fractions indicated that only a limited popu-

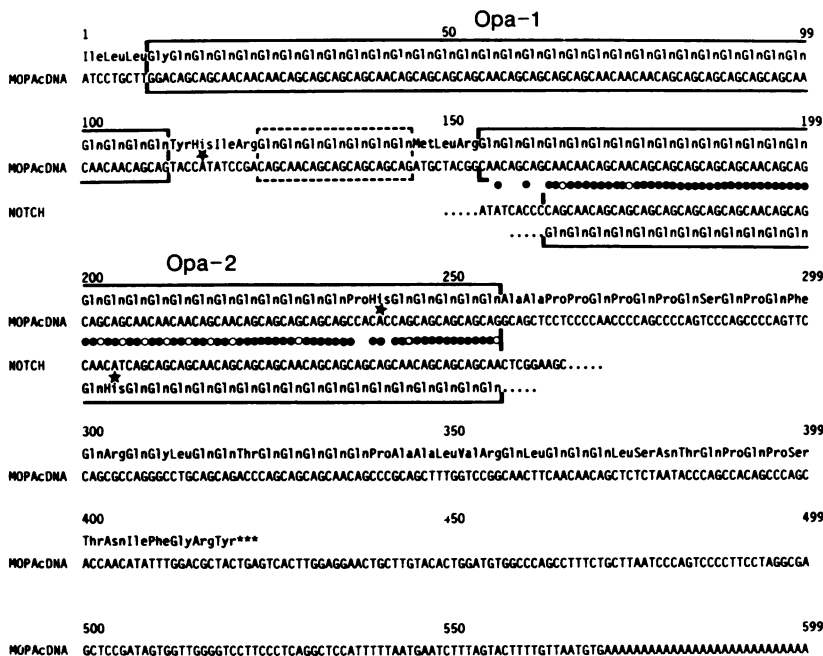


FIG. 2. Sequence of the Mopa cDNA clone and comparison with the *Drosophila* Notch opa box. The Mopa cDNA sequence is shown below one of the possible reading frames. The part of the sequence corresponding to opa-2 is aligned with the *Drosophila* Notch opa box sequence whose translation product is shown underneath. The homologies between these two boxes at the nucleotide level are indicated by closed circles, whereas nucleotides which differ without modifying the protein sequence are indicated by open circles. The stars show the positions of the histidine residues. Sequences were obtained by the Maxam and Gilbert technique (13) with double-stranded cDNA fragments end labeled at their *EcoRI*, *HpaII*, or *DdeI* sites.

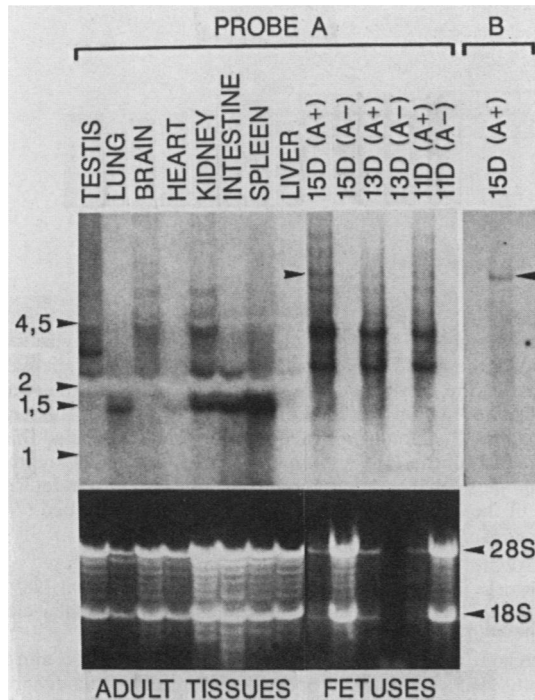


FIG. 4. Transcription of opa-like sequences in the mouse. Total RNA extracted from various adult organs as well as poly(A)⁻ and poly(A)⁺ RNA extracted from mouse fetuses at day 11 or 15 were separated on a methyl mercury gel, transferred onto DBM paper, and hybridized to the opa-containing probe A (Fig. 1). The bottom of the figure shows the UV pattern of the RNAs before the transfer. The top of the figure shows the hybridization pattern of RNAs extracted from the testes, lungs, brain, heart, kidneys, intestines, spleen, and liver and from the different fetuses from day 11 (D11) to day 15 (D15). The last lane on the right shows poly(A)⁺ RNA from day 15 fetuses hybridized to probe B which does not contain any opa-like repeat. In this last case, the same DBM paper was used after dehybridization of probe A. The pattern of opa-like repeats containing RNAs is very different from one adult tissue to another, but the differences seem to result from a combinatorial distribution within a set of transcripts (see the text). The pattern observed in the fetal mRNAs (15D [A+], 13D [A+], and 11D [A+]) is essentially identical for all the developmental stages studied but different from those observed for the adult tissues. No opa-like sequence-containing transcripts were found in the corresponding day 15 and 11 poly(A)⁻ RNAs, day 13 poly(A)⁻ RNA being degraded (see the UV pattern). The extreme right-hand lane shows the position of the fetal transcript corresponding to the Mopa cDNA. As expected, this mRNA was also detected when probed with the opa-containing probe (arrowhead). Numbers on the left are kilobases.

lation of embryonic mRNA contained opa-like sequences homologous to the *Drosophila* opa repeat. The hybridization of the same DBM paper with a part of the Mopa clone which did not contain either opa-1 or opa-2 (Fig. 1, probe B) showed that our cDNA clone was generated from a rather high-molecular-weight mRNA (Fig. 4, arrowhead). As expected, this transcript was present among the discrete bands observed when probe A was used (Fig. 4) and appeared in fetal mRNA preparations from all developmental stages.

Total RNA (30 μ g) extracted from different mouse adult tissues was treated as above and hybridized with the opa-containing probe A. The hybridization pattern proved to be specific for each of the tissues analyzed and seemed to result from a combinatorial arrangement of a limited number of transcripts (Fig. 4). For example, the lung showed a major

transcript of about 1.5 kilobases (kb) which was absent from the testes, brain, and liver but clearly present in the kidneys, intestines, and spleen. In addition to this RNA, the intestine contained a second slightly larger transcript (about 2 kb). This transcript was absent from the spleen but present in the testes and the kidneys. Further transcripts were present in kidneys, some of them being found in the brain and testes. One opa-like sequence containing RNA of about 4 kb appeared exclusively in the testes, whereas no strong transcripts were detected in the liver or in the heart (Fig. 4). The overall pattern obtained by mixing these different distributions does not fit exactly with the fetal pattern (see, for example, the 1.5-kb transcript, absent in embryonic RNAs, or the Mopa transcript [arrowhead], not found in RNA from adult organs). We therefore conclude that repeats homologous to the *Drosophila* Notch opa box are often found in murine transcripts which are subject to tissue-specific regulation and that some of them may be differentially expressed during fetal versus adult life. However, we cannot exclude the presence of such repeated sequences within RNAs ubiquitously transcribed, either in various tissues or during development.

In *Drosophila*, stretches of monotonic amino acids, especially poly-Gln, are frequently associated with homeo-box sequences (see, for example, references 10 and 16), suggesting that a possible coordinate involvement of these two repeated sequences is required to achieve the corresponding protein functions. To determine whether such opa-like sequences are also present within murine genes containing homeo-boxes, we used probe A to screen a panel of overlapping cosmid clones containing five homeo-genes (*Hox1-2* to *Hox1-6*), members of the Hox-1 complex (5, 7, 14) (Fig. 5A). In this way we isolated a 3-kb *EcoRI-BamHI* restriction fragment containing such an opa-like sequence. This repeat was located about 2.5 kb downstream from *Hox1-4* and about 10 kb upstream from *Hox1-5* (previously called *Hox1-3* and *Hox1-4*, respectively, in reference 7). This fragment was subcloned into pEMBL8, and the sequence of the corresponding subfragment revealed the presence of a small opa-like element (Fig. 5B). However, preliminary evidence indicated that this small repeat of 10 CAX triplets in a row (Fig. 5) is not part of either the *Hox1-4* or *Hox1-5* coding sequence. We are currently looking for the presence of a possible opa repeat-containing gene between these two transcription units. No other such (CAX)_n sequences were detected within the Hox-1 complex. Thus, unlike *Drosophila*, murine genes containing homeo-boxes (at least those from the Hox-1 cluster) do not contain opa-like (M repeat) sequences. Whether this reflects differences in the mode(s) of action of the respective proteins or in their overall function remains to be determined.

Although the presence of repetitive sequences within eucaryotic mRNAs has been well documented (8, 15), their functions are still unclear. For the opa repeat, these repeated sequences may be used as linking or spacing units between different protein domains or be required for the correct establishment of a multimeric protein structure (10). It was recently shown that the RNA sequence encoding the human glucocorticoid receptor contains an opa-like sequence which is not present in its mouse counterpart, strongly suggesting that, in this case, no specific function can be attributed to that sequence (15). However, the results presented here clearly indicate that these sequences are not randomly distributed among the RNA population, but instead are preferentially detected in mRNAs whose transcription seems to be regulated in an adult tissue-specific manner. In

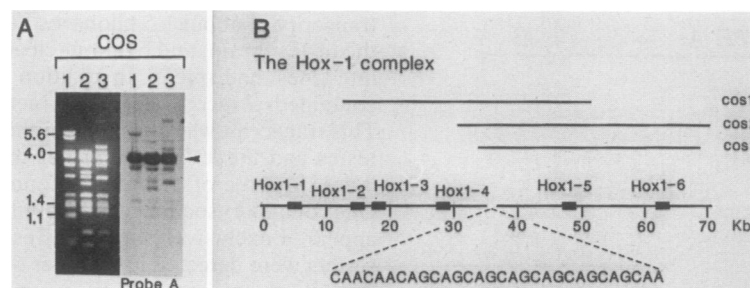


FIG. 5. Presence of opa-like sequences in the mouse homeo-box Hox-1 complex. (A) A 2- μ g sample of DNA from cosmids 1, 2, and 3 (7) was double digested with *Eco*RI and *Bam*HI, separated on an agarose gel, transferred onto a nitrocellulose membrane, and hybridized to probe A (Fig. 1). The left panel shows the UV pattern, and the right panel shows the hybridization pattern. A strongly hybridizing band (arrowhead) was observed at the same position in all three cosmid DNAs, suggesting the presence of an opa-like repeat within the region of overlap of these three cosmid clones. No other significant hybridization was observed within the homeo-box-containing gene cluster, thus excluding the presence of such (CAX)_n repeats in the Hox-1 murine homeo genes. (B) Subcloning and sequencing studies revealed that this repeat is composed of 10 CAX triplets in a row. The exact location of this sequence in the Hox-1 complex was determined by restriction mapping and comparison with previously reported mapping data on the Hox-1 complex (7). The orientation of the repeat was not determined and is therefore given arbitrarily. Digestion, electrophoresis, and hybridization of the cosmid clones were as previously described (7).

addition, these sequences are also present among mRNAs transcribed during fetal development. It is not yet known whether such RNAs are involved in important developmental processes as reported for *Drosophila*.

We thank P. Chambon for continuing interest, M. S. Featherstone, G. Albrecht, and G. Richards for helpful discussions, and P. Sondermeyer (Transgene S.A.) and M. LeMeur for advice in preparing cDNA libraries and DBM papers, respectively. We thank C. Werlé and B. Boulay for the illustrations and the secretarial staff for typing the manuscript.

This work was supported by grants from the INSERM and CNRS. D.D. was the recipient of an EMBO long-term fellowship.

LITERATURE CITED

- Alwine, J. C., D. J. Kemp, and G. R. Stack. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* **74**:5350-5354.
- Auffray, C., and F. Rougeon. 1980. Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107**:303-314.
- Bailey, J. M., and N. Davidson. 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. *Anal. Biochem.* **70**:75-85.
- Beachy, P. A., S. L. Helfand, and D. S. Hogness. 1985. Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature (London)* **313**:545-551.
- Colberg-Poley, A. M., S. D. Voss, K. Chowdhury, C. L. Stewart, E. F. Wagner, and P. Gruss. 1985. Clustered homeo boxes are differentially expressed during murine development. *Cell* **43**:39-45.
- Davidson, E. H., and J. W. Posakony. 1982. Repetitive sequence transcripts in development. *Nature (London)* **297**:633-635.
- Duboule, D., A. Baron, P. Mähl, and B. Galliot. 1986. A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J.* **5**:1973-1980.
- Fickett, J. W. 1982. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res.* **10**:5303-5318.
- Kuner, J., M. Nakanishi, Z. Ali, B. Drees, E. Gustavson, J. Thies, L. Kauvar, T. Kornberg, and P. O'Farrell. 1985. The engrailed locus of *Drosophila melanogaster*: molecular cloning. *Cell* **42**:309-316.
- Laughon, A., S. B. Carrol, F. A. Storfer, P. D. Riley, and M. P. Scott. 1985. Common properties of proteins encoded by the Antennapedia complex genes of *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* **50**:253-262.
- Laughon, A., and M. P. Scott. 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature (London)* **310**:25-31.
- Le Bouc, Y., D. Dreyer, F. Jaeger, M. Binoux, and P. Sondermeyer. 1986. Complete characterization of the human IGF-I nucleotide sequence isolated from a newly constructed adult liver cDNA library. *FEBS Lett.* **186**:108-112.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**:560-564.
- McGinnis, W., C. P. Hart, W. J. Gehring, and F. H. Ruddle. 1984. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell* **38**:675-680.
- Miesfeld, R., S. Rusconi, P. J. Godowski, B. A. Maler, S. Okret, A. C. Wikström, J. A. Gustafsson, and K. R. Yamamoto. 1986. Genetic complementation of the glucocorticoid receptor deficiency by expression of cloned receptor cDNA. *Cell* **46**:389-399.
- Regulski, M., K. Harding, R. Kostriken, F. Karch, M. Levine, and W. McGinnis. 1985. Homeo-box genes of the Antennapedia and Bithorax complexes of *Drosophila*. *Cell* **43**:71-80.
- Schneuwly, S., A. Kuroiwa, P. Baumgartner, and W. J. Gehring. 1986. Structural organisation and sequence of the homeotic gene Antennapedia of *Drosophila melanogaster*. *EMBO J.* **5**:733-739.
- Sutcliffe, J. G., R. J. Molner, F. H. Bloom, and R. A. Lerner. 1982. Common 82-nucleotide sequence unique to brain RNA. *Proc. Natl. Acad. Sci. USA* **79**:4942-4946.
- Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA* **76**:3683-3687.
- Wharton, K. A., B. Yedvobnick, V. G. Finnerty, and S. Artavanis-Tsakonas. 1985. OPA: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in *D. melanogaster*. *Cell* **40**:55-62.