The EMBO Journal vol.10 no.1 pp.71 – 81, 1991

Differentially expressed isoforms of the mouse retinoic acid receptor β are generated by usage of two promoters and alternative splicing

A.Zelent, C.Mendelsohn, P.Kastner, A.Krust, J.-M.Garnier, F.Ruffenach, P.Leroy and P.Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cédex, France

Communicated by P.Chambon

Using anchored PCR, three different cDNA isoforms of the mouse retinoic acid receptor β [mRAR-β1, mRAR-β2 (formerly mRAR-β0) and mRAR-β3], generated from the same gene by differential promoter usage and alternative splicing, were isolated. These three isoforms encode RAR proteins with different N-terminal A regions and identical B – F regions. The sequence encoding the first 59 amino acids of the mRAR-β3 A region is identical with the entire A region of mRAR-β1. However, the sequence of mRAR-β3 region A differs from that of mRAR-β1 by an additional 27 C-terminal amino acids encoded in an 81 nucleotide-long putative exon which is spliced in between the exons encoding the A and B regions of mRAR-β1. Both mRAR-β1 and β3 cDNAs differ entirely from mRAR-β2 in their 5′-untranslated (5′-UTR) and A region coding sequences. This N-terminal variability, in a region which was shown to be important for cell-type specific differential target gene *trans*-activation by other nuclear receptors, suggests that the three mRAR-β isoforms may be functionally distinct. The conservation of RAR-β isoform sequences from mouse to human, as seen by cross-hybridization on Southern blots or DNA sequence analysis, as well as their differential patterns of expression in various mouse tissues, corroborates this view. Additionally, the mRNA analysis data suggest that mRAR-β2, whose expression predominates in RA-treated embryonal carcinoma (EC) and embryonic stem (ES) cells, may be important during early stages of development. mRAR-β1 and β3, on the other hand, which are predominantly expressed in fetal and adult brain, may play some specific role in the development of the central nervous system.

Key words: EC cells/mouse embryo/mouse RAR-β cDNA isoforms/retinoic acid induction/5′-UTR

Introduction

The role of retinoic acid (RA) in cellular growth and differentiation, as well as in vertebrate development, has been extensively studied (for reviews see Roberts and Sporn, 1984; Maden, 1985; Brockes, 1989, 1990; Eichele, 1989; Summerbell and Maden, 1990). RA is now believed to play a fundamental role in the development of vertebrate nervous system (Durston *et al.*, 1989; Wagner *et al.*, 1990), limbs (Tickle *et al.*, 1982; Thaller and Eichele, 1987) and cranio-facial features (Langille *et al.*, 1989; Wedden *et al.*, 1987a,b, 1988), as well as in regeneration of amphibian limbs (Maden, 1982; Crawford and Stocum, 1988; Ragsdale *et al.*, 1989). In mammals, administration of pharmacological doses of RA and its analogues (retinoids) during pregnancy results in a wide range of cranio-facial and limb malformations (Morriss and Thorogood, 1978; Sulik and Dehart, 1988; Alles and Sulik, 1989; Satre and Kochar, 1989), as well as defects in brain development (Langman and Welch, 1967; Morriss, 1972; Lammer *et al.*, 1985), providing evidence, albeit indirect, on the role of RA in mammalian embryogenesis. In the adult organism, RA is also required for differentiation and maintenance of various epithelia (Lotan, 1980; Roberts and Sporn, 1984; Shapiro, 1985; Asselineau *et al.*, 1989; and references therein).

Although many ideas on the molecular mechanism of RA action have been elaborated in the past (Lotan, 1980; Chytil, 1984; Roberts and Sporn, 1984; Shapiro, 1985; Slack, 1987a,b; and references therein), isolation of cDNAs encoding a human RA receptor (hRAR) (Giguère *et al.*, 1987; Petkovich *et al.*, 1987), which belongs to the superfamily of nuclear hormone receptors (Evans, 1988; Green and Chambon, 1988; Beato, 1989), has greatly advanced our understanding of how a simple molecule such as RA can influence and affect a wide range of complex molecular and physiological processes. To date, three RARs (RAR-α, β and γ) have been shown to exist, both in mouse (mRARs) and human (hRARs) (Giguère *et al.*, 1987; Petkovich *et al.*, 1987; Benbrook *et al.*, 1988; Brand *et al.*, 1988; Krust *et al.*, 1989; Zelent *et al.*, 1989). They are proteins with modular structure, which on the basis of amino acid sequence homology with other nuclear receptors can be subdivided into six distinct regions, A – F (see Green and Chambon, 1988 and references therein). Mutagenesis studies, using steroid receptors as a model system, indicated that regions C and E are responsible for DNA and ligand binding (Evans, 1988; Green and Chambon, 1988; and references therein), respectively, and that the A/B region contains a *trans*-activation domain which may be both cell type- and promoter-specific (Tora *et al.*, 1988a,b, 1989; Tasset *et al.*, 1990); an activation function is also associated with region E, and at present the functions of the D and F regions are unclear. As described previously (Krust *et al.*, 1989; Zelent *et al.*, 1989), the amino acid sequences comprising regions B, C and E of RARs α, β and γ are highly conserved within a given species; in contrast, sequences of regions A, D and F are less or not conserved. However, the amino acid sequence of a given receptor subtype (α, β or γ) is nearly 100% conserved between mouse and human. This suggests that sequence differences in regions A, D and F among the various RAR subtypes are functionally important, and that each RAR may play a specific role in RA-dependent networks of transcriptional regulation of gene expression. These conclusions are corroborated by recent *in situ*

© Oxford University Press

hybridization studies which show distinct and spatio-temporally restricted patterns of expression for each of the RAR subtypes during mouse embryogenesis (Dollé *et al.*, 1989, 1990; Ruberte *et al.*, 1990, 1991).

The complexity of RARs is further illustrated by the recent finding of multiple cDNA isoforms of RAR-γ, both in mouse and human (Krust *et al.*, 1989; Giguère *et al.*, 1990; Kastner *et al.*, 1990). These isoforms diverge in DNA sequences located upstream of the A and B region boundary (A/B region junction), and contain the same B−F region coding sequences. Two major isoforms, in mouse and human, RAR-γ1 and γ2, have different 5′-untranslated region (5′-UTR) and A region sequences, and are differentially expressed in both mouse embryo and adult tissues. Five other mouse isoforms (mRAR-γ3 to γ7) lack A regions altogether and differ only in their 5′-UTRs.

This paper describes two novel isoforms of mRAR-β which, together with the initially described mouse and human RAR-β, contain different 5′-UTR and/or N-terminal A regions and are differentially expressed in the mouse embryo and in the adult animal. Interestingly, the two novel isoforms are preferentially expressed in both adult and fetal brain which suggests that they play a specific role in the central nervous system.

Results

Two novel mRAR-β isoforms, mRAR-β1 and β3

The RAR A regions and 5′-UTRs, which are encoded in exons different from the exon encoding the B region (Brand *et al.*, 1988), are highly divergent between the three RAR subtypes as well as between the isoforms of mouse and human RAR-γ (Krust *et al.*, 1989; Zelent *et al.*, 1989; Kastner *et al.*, 1990). Thus, our experimental strategy was aimed at isolating isoforms of mRAR-β which differ in the N-terminal A regions and 5′-UTRs. Using the anchored PCR technique (Loh *et al.*, 1989) and polyadenylated [poly(A)$^+$] RNAs obtained from various mouse tissues and cell lines, as well as from different stage embryos, two novel cDNA isoforms of mRAR-β were cloned (see Materials and methods for details). For reasons discussed below (see Discussion), they were designated as mRAR-β1 and mRAR-β3, and the RAR-β formerly described in human (Benbrook *et al.*, 1988; Brand *et al.*, 1988) and mouse (Zelent *et al.*, 1989) will hereafter be referred to as RAR-β2.

In total, ~2000 colonies were screened to give ~300 clones which hybridized with a ^{32}P-labelled B region oligonucleotide probe [nucleotides (nt) 317−340 in mRAR-β1 sequence, see Figure 1]. The most frequent clones contained either the mRAR-β2 region A sequences or intronic sequences lying immediately upstream of the B region-encoding exon (see below). Clones containing mRAR-β1 and β3 sequences were much less frequent. mRAR-β2 isoform cDNA clones were predominant because all PCR-amplified cDNAs originated mainly from sources which turned out to have high levels of mRAR-β2 mRNA (see Materials and methods). The high frequency of unspliced intronic sequence isolation could reflect the presence of relatively high levels of either unspliced or partially spliced primary transcripts. Note that analogous sequences, corresponding to the B region-encoding exon and the upstream intron, were also readily isolated for both mRAR-γ (Kastner *et al.*, 1990) and mRAR-α [Leroy *et al.*, 1991 (accompanying paper)].

Figure 1 shows the partial cDNA and deduced amino acid (up to the 15th amino acid of the B region) sequences of mRAR-β1. The ATG triplet encoded in nt 139−141 of mRAR-β1 (Figure 1) marks the N terminus of both mRAR-β1 and β3 (see below). It is surrounded by sequences which have been described by Kozak (1983, 1986) to be characteristic of eukaryotic initiation codons, whereas the in-frame ATG present downstream (nt 187−189 of mRAR-β1 in Figure 1) does not conform with these sequences, and hence is less likely to serve as an initiator of translation. Furthermore, multiple in-frame upstream termination codons (underlined in Figure 1) clearly define the 5′ border of the open reading frame (ORF). The mRAR-β1 A region (A1) sequence has no similarity with the A region of the previously described mRAR-β2 (A2 in Figure 2) and must, therefore, be encoded in exonic sequences (E1 in Figure 1) different from those encoding the A2 region (E3 in Figure 2). Note that the junction between the mRAR-β1 A1 and B regions corresponds exactly to the previously observed point of N-terminal divergence between various mouse or human RARs (Krust *et al.*, 1989; Zelent *et al.*, 1989), which has been referred to as the A/B region junction (indicated by a filled triangle in Figures 1−3).

The cDNA and deduced amino acid sequences of mRAR-β3 (Figure 1) diverge from those of mRAR-β1 and mRAR-β2 at exactly the same A/B region junction (Figures 1 and 2, see also the diagrams in Figure 3A). Interestingly, 81 nt upstream from this point of divergence, the cDNA sequence of mRAR-β3 (nt 130−445 of mRAR-β3 in Figure 1) becomes the same as the cDNA sequence of the 5′-UTR and A1 region of mRAR-β1 (nt 1−316 of mRAR-β1 in Figure 1). Therefore, the A region of mRAR-β3 (A3) has an additional amino acid sequence which distinguishes it from the A1 region of mRAR-β1. This sequence is probably encoded by an additional exon (E2 in Figures 1 and 3) which is spliced, 3′ and 5′, with the exon encoding the B region (E4 in Figures 1−3) and the exon encoding the A region amino acids common to both mRAR-β1 and β3 (E1 in Figures 1 and 3), respectively. The putative E1/E2 and E2/E4 (A/B) exon junctions in mRAR-β3 are indicated in Figure 1 by opened and filled triangles, respectively.

The 5′-UTR sequence of mRAR-β3 contains four short ORFs (sORFs) and can be folded, using Zuker's program for secondary structure prediction (Zuker and Stiegler, 1981; Jacobson *et al.*, 1984), into a stable secondary structure ($\Delta G \leq -116$ kcal/mol, data not shown). The 5′ border of the longest mRAR-β1 clone corresponds to nt 130 of mRAR-β3 (indicated by an open arrow in Figure 1). It is presently unknown whether both mRAR-β1 and β3 cDNAs are identical over their entire 5′-UTRs, since the isolated cDNA clones are most likely not full length (see below).

The sequences shown in Figure 1 were derived (Materials and methods) from independent mRAR-β1 and β3 PCR cDNA clones (three of each) which had their 3′ boundary in the B region and extended 5′ to various lengths as indicated by vertical bold arrows in Figure 1. All sequences shown in Figure 1 were found to be identical in at least two independent PCR amplified cDNA clones, except for the first 114 nt of mRAR-β3, which corresponded to a single clone. Subsequently, PCR was also used to amplify cDNAs containing the entire sequence encoding regions A−F of the two new mRAR-βs (see later sections on RNA analysis and

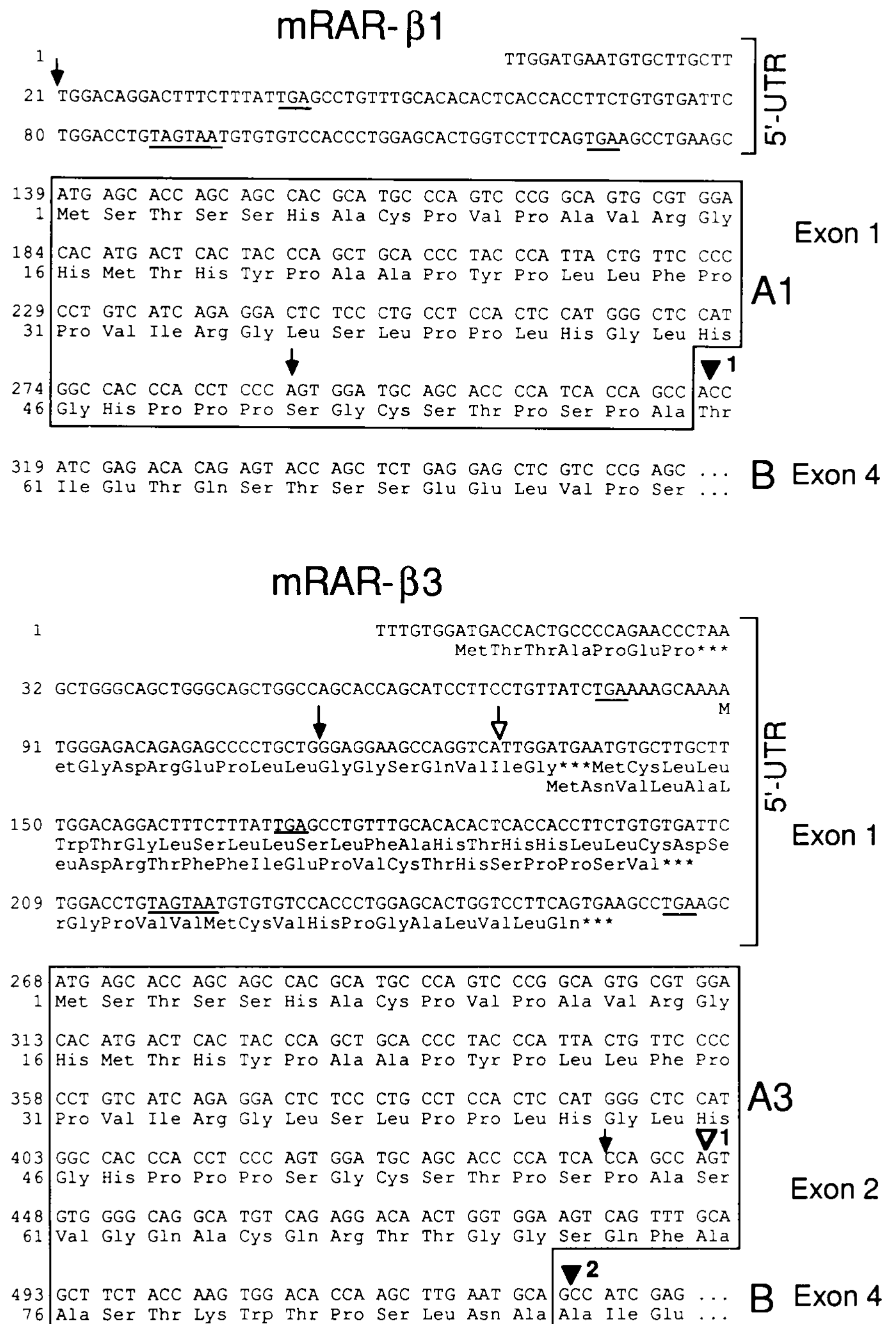

Fig. 1. Nucleotide and deduced amino acid sequences of mRAR-β1 and mRAR-β3 cDNA isoforms. Only the 5'-UTR (indicated with a bracket) and the A region sequences of the two receptors (A1 and A3, boxed), as well as the first 45 (mRAR-β1) and 9 (mRAR-β3) nt of their common B region sequence (indicated with a B) are shown. At the end of each sequence, dots are used to indicate that both cDNA clones can be extended further 3' to contain identical B–F region coding sequences (see Figure 3A and B). Nucleotides and amino acids are numbered on the left side of the figure. The junctions between exons 1 and 2 (identified in the right margin of mRAR-β3) is indicated with an open triangle. Filled triangle designates the A–B region junction which in the case of mRAR-β1 and mRAR-β3 is the boundary of exons 1 and 2, respectively, with exon 4. The numbers given to each exon–exon junction correspond to the number of the upstream exon (see also Figure 3A). Filled arrows indicate 5'-ends of additional independently isolated mRAR-β1 and mRAR-β3 cDNAs which were sequenced. The open arrow in mRAR-β3 indicates the 5'-end of the longest mRAR-β1 cDNA clone with respect to the colinear mRAR-β3 sequence. In both sequences all in-frame termination codons which lie upstream to the major ORF are underlined. Below the mRAR-β3 5'-UTR sequence, the deduced amino acid sequences of all possible upstream sORFs are indicated. Three successive asterisks indicate a stop codon at the end of the sORFs.

Materials and methods for details). Partial sequencing of these cDNAs showed that the nucleic acid sequences encoding regions B–F are identical for all the isoforms, and diverge upstream of the A/B region junction. Consequently, all the RAR-β cDNA isoforms must derive from the same mRAR-β gene.

The proximal promoter, 5'-UTR and A region sequences of mRAR-β2 are highly conserved from mouse to human

The mouse cDNA and deduced amino acid sequences shown in Figure 2 (nt −120 to 639) have been derived from overlapping genomic and cDNA clones, isolated from

```
                           mRAR-β2

 -87             hRAR-β2 ──▶ A-------------------------          ┐
-120 TGTGAGAATCCTGGGAGTTGGTGATGTCAGACTGGTTGGGTCATTTGAAGGTTAGCAGC   │ RAR-β2
                         oooooooo       •••••••                   │ Promoter
 -61 -----T----------------------------------------------------T  │
 -61 CCGGGAAGG[GTTCAC]CGAAA[GTTCAC]TCGCATATATTAGGCAATTCAATCTTTCATTCC ┘
                                          =========
      ┌▶(+1)
  -2 -------------A---------------T--A--------------------------  ┐
  -2 GTGTGACAGAAGTGGTAGGAAGTGAGCTGCTCCGAGGCAGGAGGGTCTATTCTTTGCCA   │
                                                                  │
                                                       Pro        │
  58 -- ------------A--C---A---- A--T-TTTG-G-----------C---A----  │
  58 AAGGGGGGGACCAGAGTT CCCGTGCGCCGCGGCCACAAGACTGGGATGCAGAGGACGC   │
                                         sORF1 MetGlnArgThrA      │
         Ile          Phe              Ser                        │
 115 ----G-T---A------TTT------------T----------T--G-AA---A-T-G-  │
 116 GAGCCACCCGGGCAGGGAGCGTCTGGGCACCGGCGGGGTAGGA CCCGCGCGC TCCCG   │
     rgAlaThrArgAlaGlySerValTrpAlaProAlaGly***                    │ 5'-UTR
                                                                  │
 174 A----TT-TT--AA--A-TTA--T-G-A---------------TT-CT-----C-----  │
 173 GAGGC CTGCGCGGGCGTCGCCTGGAAGGGAGAACTTGGGATCGGTGCGGGAA CCCCC   │
                                                                  │
 233 ----C--------T--------A------------------------------------  │
 230 GCCCTGGCTGGATCGGCCGAGCGAGCCTGGAAAATGGTAAATGATCATTTGGATCAATT   │
                          sORF2 MetValAsnAspHisLeuAspGlnLe        │
 292 -----------------------------------------------------------  │ Exon 3
 289 ACAGGCTTTTAGCTGGCTTGTCTGTCATAATTCATGATTCGGGGCTGGGAAAAAGACCA   │
     uGlnAlaPheSerTrpLeuValCysHisAsnSer***                        │
                          sORF3 MetIleArgGlyTrpGluLysAspGl        │
 351 -----------------------------------------GG------------A---  │
 348 ACAGCCTACGTGCCAAAAAAGGGGCAGAGTTTGATGGAGTT CGTGGACTTTTCTGTGC   │
     nGlnProThrCysGlnLysArgGlyArgVal***                           │
                                                                  │
 410 CAT-T-------------------------------CATT-A----AA-G------     │
 406 GGCTCGCCTCCACACCTAGAGGATAAGCACTTTTGCAGA GCGCGGTGC GGAGAGATC   ┘

 469|--- --- --- --- --- --- --- --- --- --- --- --T --- --A ---|
 463|ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCC GGG CAG ATC|
   1|Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln Ile|
    |                                                           |
 514|--- --- --- --- --T --- --T --G --T --- --- --- --C --- --G|
 508|CTG GAT TTC TAC ACC GCG AGC CCT TCC TCC TGC ATG CTG CAG GAA| A2
  16|Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu Gln Glu|
    |                                                           |
    |                        Phe         Leu         Thr        |
 559|--A --- --- --- --A --- T-- --- --- --G --- --A A-- --- ---|
 553|AAG GCT CTC AAA GCC TGC CTC AGT GGA TTC ACC CAG GCC GAA TGG|
  31|Lys Ala Leu Lys Ala Cys Leu Ser Gly Phe Thr Gln Ala Glu Trp|
    |                           ▼3                              
 604|--- --T --- --C --- --- ---|--A --T --A --- --- --C --- ...
 598|CAG CAC CGG CAT ACT GCT CAA|TCC ATC GAG ACA CAG AGT ACC ...  B Exon 4
  46|Gln His Arg His Thr Ala Gln|Ser Ile Glu Thr Gln Ser Thr ...
```

Fig. 2. Alignment of the proximal promoter and 5′-UTR DNA sequences of mouse and human RAR-β2, as well as both nucleotide and deduced amino acid sequences of their A2 regions. The DNA sequences encoding the first seven amino acids of the B region are also shown. In order to obtain maximal alignment, gaps were introduced in either the human or mouse DNA sequence. The residues which differ in hRAR-β2 DNA or amino acid sequences from the corresponding mRAR-β2 sequences are indicated, and those which are identical with the mouse sequences are represented by dashes. The broken arrow indicates transcription initiation start site of hRAR-β2 mRNA (de Thé *et al.*, 1990) and marks the putative cap site (+1) of the mRAR-β2 mRNA. The mouse and human RAR-β2 promoter sequences (indicated by a bracket) contain conserved RAREs (boxed), as well as a putative TATA box, TRE and CRE sequence motifs (see text) designated with double line, filled and opened circles, respectively. The deduced amino acid sequences of three sORFs (sORF1–3), highly conserved in both mouse and human 5′-UTRs, are given. All other symbols and representations are as in Figure 1.

λEMBL3 mouse genomic DNA and λZapII RA-treated F9 cell cDNA libraries, respectively (see Materials and methods for details). The human DNA sequence was taken from the literature (de Thé *et al.*, 1987, 1990).

The ORF encoding mRAR-β2 begins with an ATG embedded in a relatively favourable Kozak sequence and is preceded by several in-frame termination codons (underlined in Figure 2). The A (A2 in Figure 2) as well as B–F region coding sequences of mRAR-β2 are highly conserved between mouse and human (Figure 2, and Zelent *et al.*, 1989). Interestingly, the sequence of the entire mRAR-β2 5′-UTR is almost as conserved, with respect to the corresponding human sequence, as that encoding the A2 region (86% and 90%, respectively). Both human and mouse mRAR-β2 5′-UTRs can be predicted to form (not shown) similar and stable secondary structures ($\Delta G \leq -223$ kcal/mol for mouse and ≤ -172 kcal/mol for human). In addition, the mRAR-β2 5′-UTR encodes three sORFs which are highly conserved at the nucleic and amino acid levels between the two species (sORFs 2 and 3 are 100% conserved at the DNA sequence level). Note that sORF2, sORF3 and the ORF encoding RAR-β2 are also in the same relative reading frames in both mouse and human DNAs.

Analysis of mouse genomic DNA clones revealed that, unlike mRAR-β3, the entire 5′-UTR and the A2 region of mRAR-β2 are most probably encoded in a single exon

(E3, Figures 2 and 3), which lies immediately downstream from the RA-inducible RAR-β gene promoter recently characterized in the human by de Thé *et al.* (1990) (Figures 2 and 3B). The 5′-flanking genomic DNA sequence of mRAR-β2 is indeed 98% identical to the published human sequence and contains an identical TATA box (underlined) as well as retinoic acid response (RARE, boxed) elements (Figure 2, +1 is indicated with respect to the human sequence as reported by de Thé *et al.*, 1990). Moreover, the 5′-flanking mRAR-β2 region exhibits RA-dependent promoter activity when transfected into cells in culture (C. Mendelsohn, unpublished results). At present, it is unknown what additional functional upstream regulatory elements may exist in this promoter, and to what degree they are conserved between human and mouse. Note, however, that the mRAR-β2 promoter sequence (nt −99 to −92, indicated by open circles in Figure 2) contains a 5′-TGATGTCA-3′ motif which closely resembles the consensus cAMP response element (CRE: 5′-TGACGTCA-3′, see Mitchell and Tjian, 1989). Additionally, there is a DNA sequence in both mouse and human promoters that is identical to the oestrogen response element (5′-TGGGTCA-3′) which is present (on the other strand) in the chicken ovalbumin gene promoter (Tora *et al.*, 1988b). This sequence (indicated by closed circles in Figure 2) closely resembles (on the other strand) the consensus palindromic AP1 binding site 5′-TGACTCA-3′ (see Mitchell and Tjian, 1989). Further experiments will show whether these sequence motifs constitute true *cis*-acting regulatory elements in the RAR-β2 promoter.

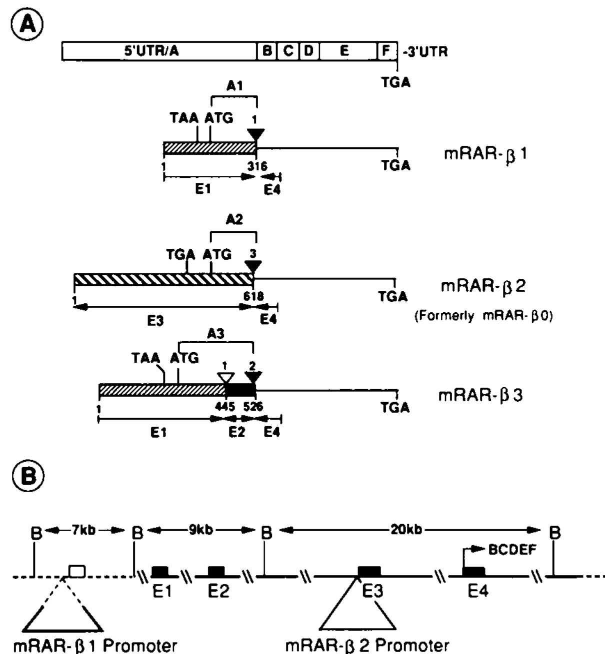

Fig. 3. (**A**) Schematic representation of the mRAR-β cDNA isoforms. The general organization of RAR sequence is shown above and mRAR-β1 to β3 are schematized below. The ORF containing B−F regions common to all RAR-β isoforms is indicated by a thin line. Regions where the sequences between the isoforms are identical are represented by an identical design. Both the 5′-UTR and A region sequences which differ between the three mRAR-β2 isoforms are indicated by boxes with different designs. Single- and double-headed arrows represent sequences which are contained (see text) within exons 1 to 4 (E1−E4) of the mRAR-β gene, as indicated. Absence of an arrowhead indicates that the exact 5′ (E1) or 3′ (E4) boundary of a given exon is not known. Numbers given below each scheme correspond to nucleotide numbers of mRAR-β1, β2 and β3 sequences as shown in Figures 1 and 2. Filled triangles indicate the A/B region junction and the open triangle indicates the E1/E2 boundary as in Figures 1 and 2. Initiation codons at the beginning of the A1, A2 and A3 regions (marked by brackets above the schemes representing mRAR-β1, β2 and β3 sequences, respectively), as well as the first downstream and upstream in frame termination codons are as indicated. (**B**) Preliminary schematic representation of the 5′ region of the mRAR-β gene. *Bam*HI (B) restriction enzyme fragments detected with various isoform and exon specific probes in Southern analysis (see Figure 4 and text), as well as their approximate sizes (in kb) are shown. The exons 1−4 (E1−E4) are indicated with black rectangles and their relative order in the mRAR-β gene is shown. However the exact distances between the exons are unknown (indicated by discontinuities in the solid line). The broken arrow points in the direction of exons which encode regions B−F. The sequences lying within the ~7 kb *Bam*HI fragment, believed to contain an additional 5′ exon(s) (opened rectangle, see text) which may encode mRAR-β3 and/or β1 upstream 5′-UTR sequences (see text), are indicated by a discontinuous line. Location of either the RA-inducible mRAR-β2 promoter, or the putative mRAR-β1 promoter, is as indicated.

mRAR-β1/β3 specific sequences are located upstream of the mRAR-β2 promoter in the mouse genome and cross-hybridize under stringent conditions with human DNA

In Southern blot analysis of *Eco*RI and *Bam*HI digested mouse genomic DNA, a probe corresponding to the whole 5′-UTR and A region of mRAR-β3 (see Materials and methods and figure legends for description of all probes) revealed three and two DNA fragments, respectively (Figure 4A, lanes E and B). Hybridization of the same blots with a ^{32}P-labelled mRAR-β1/β3 probe (containing their common exon 1 sequences, see Figure 1) and a mRAR-β3 specific probe (exon 2, see Figure 1) detected only the ~1.3 and ~17 kb *Eco*RI fragments, respectively (indicated by filled arrows in Figure 4A as E1 and E2, respectively; data not shown). In addition, the same two probes detected only the ~9 kb fragment in *Bam*HI digested mouse genomic DNA (not shown). This indicates that the ~2.6 kb *Eco*RI DNA fragment (indicated by an open arrow in Figure 4A), which is probably contained within the *Bam*HI fragment of ~7 kb, must be hybridizing with mRAR-β3 sequences located upstream to the 5′ border of the region over which the 5′-UTRs of mRAR-β1 and β3 are colinear (open arrow in mRAR-β3 sequence in Figure 1). Since there are neither *Eco*RI nor *Bam*HI restriction enzyme recognition sites in mRAR-β1/β3 5′-UTR and A region sequences, it is likely that the sequences comprising the 5′-UTR and A region of mRAR-β3 are encoded in more than the two exons (E1 and E2) discussed above. The putative additional exon(s) is represented with an open rectangle in Figure 3B. Further analysis is required to map, within the mRAR-β gene, the exact position of the all the mRAR-β1 and β3 5′-UTR and A region encoding exons.

Figure 4B shows the result of hybridization between a ^{32}P-labelled mRAR-β2 specific probe and the same Southern blot as above. Detection of unique bands in *Eco*RI and *Bam*HI digested mouse genomic DNA is consistent with the fact that mRAR-β2 A2 and 5′-UTR sequences are contained within a single exon (E3), whose sequence is devoid of recognition sites for these two restriction enzymes, and whose 5′-end corresponds to the cap site of the mRNA encoding the mRAR-β2 isoform (see previous sections). Interestingly, both a ^{32}P-labelled mRAR-β2 cDNA specific

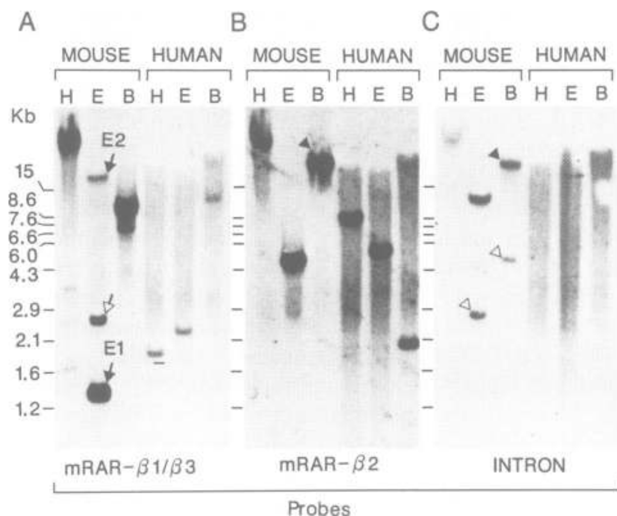

Fig. 4. Southern analysis of mouse and human RAR-β genomic DNA sequences. Both human and mouse genomic DNAs, digested with *Hin*dIII (H), *Eco*RI (E) and *Bam*HI (B) restriction enzymes, were hybridized with mRAR-β1/β3, mRAR-β2 and intron specific probes as shown in panels A, B and C, respectively. Positions of DNA size markers in a gel are indicated (in kb) on the left side of panel A and are the same for panels B and C. Filled triangle in B and C points to the ~20 kb *Bam*HI mouse genomic DNA fragment detected with the mRAR-β2 and intron specific probes, respectively (see Results). Open triangles in C designate bands corresponding to remnants of previous hybridization. Filled arrows in A point to *Eco*RI mouse genomic DNA fragments which were also detected with E1 and E2 exon (as indicated) specific probes (see text). The open arrow identifies the *Eco*RI mouse genomic DNA fragment which presumably hybridized with mRAR-β3 sequences lying upstream to the nucleotide marked with an open arrow in Figure 1. This fragment is thought to contain the additional mRAR-β exon(s) which would contain further upstream mRAR-β3 and/or possibly also mRAR-β1 5'-UTR sequences (see Figure 3B and text).

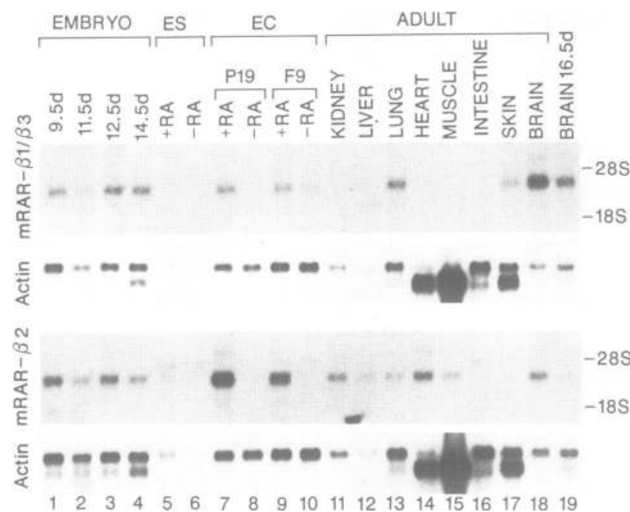

Fig. 5. Northern blot analysis of mRAR-β1/β3 and β2 mRNAs. Blots were hybridized with ^{32}P-labelled mRAR-β1/β3 and β2 specific probes as described in the text and Materials and methods, and as indicated on the left side of the figure. Except for lanes 5 and 6, where 15 μg of total ES cell (+ or -10^{-8} M RA for 48 h, as indicated) RNA was used, all other lanes represent 4 μg of poly(A)$^+$ RNA isolated from various sources as indicated in the text and in the figure at the top of each lane (1–19). The results of ^{32}P-labelled actin probe hybridization with mRAR-β1/β3 and β2 blots are shown below each blot in panels labelled actin. Positions of 28S (4712 bases, Hassouna *et al.*, 1984) and 18S (1869 bases, Raynal *et al.*, 1984) rRNAs in a given gel are indicated on the right side of the figure.

probe and an intron probe (412 bp of intronic sequences lying in the genomic DNA immediately upstream of the B region-encoding exon) hybridized to the same ~20 kb *Bam*HI fragment, which was not detected with the mRAR-β1/β3 probe (Figure 4A–C). The same DNA fragment also hybridized with an exon 4 specific probe (see Figure 3; data not shown). This indicates that both the mRAR-β2 promoter and the exon encoding the B region lie on the same *Bam*HI genomic DNA fragment (indicated with bold triangles in Figure 4B and C), which is different from the *Bam*HI fragments containing mRAR-β1/β3 specific sequences (Figure 4A). Therefore, both E1 and E2, as well as any other exons which may be located further upstream and encode additional mRAR-β1/β3 5'-UTR sequences, must be positioned in the mouse genomic DNA upstream from the mRAR-β2 promoter (see Figure 3B for a schematic representation). Consequently, there must be an additional promoter within the mRAR-β gene, which accounts for the expression of mRAR-β1/β3 isoforms, and which remains to be characterized.

The cDNAs for the possible human counterparts of mRAR-β1 and β3 (hRAR-β1 and β3), and the corresponding genomic DNA containing the 5' region of these putative human RAR-β isoforms, have not yet been characterized. However, it is likely that these counterparts exist, as is the case for mRAR-β2/hRAR-β2 (see above). Southern analysis showed specific cross-hybridization between restriction enzyme-digested human genomic DNA and a ^{32}P-labelled mRAR-β1/β3 specific probe, indicating that these sequences have indeed been conserved through evolution (Figure 4A). Interestingly, the ^{32}P-labelled intronic probe used above did not specifically cross-hybridize under the same stringency with human DNA (Figure 4C), even though the first 180 nt of human and mouse nucleic acid sequences immediately upstream of the acceptor splice site are 70% homologous (Dejean *et al.*, 1986; and data not shown). As expected, a ^{32}P-labelled mRAR-β2 specific probe also cross-hybridized with human genomic DNA (Figure 4B).

The mRAR-β isoforms are differentially expressed

Northern blot hybridization, using ^{32}P-labelled probes specific for either mRAR-β1/β3 or mRAR-β2 sequences and various total and poly(A)$^+$ RNAs (Figure 5), was initially carried out to investigate if the mRAR-β isoforms are differentially expressed in the mouse. After appropriate exposure, blots were rehybridized with a ^{32}P-labelled actin probe (Zelent *et al.*, 1989) to check for the integrity of the RNAs loaded on the gel. The blot hybridized with the mRAR-β1/β3 specific probe, hybridized with the actin sequences to a lesser extent than the mRAR-β2-probed blot; this was most probably due to a less efficient transfer in the former case (compare upper and lower panels, lines labelled actin in Figure 5). The amount of actin mRNA in different tissues is variable (Erba *et al.*, 1988), and is usually not a good reference for equivalent loading of RNA samples on a given gel. However, since an equal amount of RNA (15 μg for total RNA, or 4 μg for poly(A)$^+$ RNA) was loaded in each lane, the differences in expression levels of a given isoform among various samples should roughly parallel the intensities of the signals detected on the Northern blot. Figure

5 shows the mRAR-β1/β3 (upper panel) and β2 mRNAs as single bands migrating with respect to 28S and 18S rRNAs at ~3.3–3.4 and 3.4–3.5 kb, respectively.

The resolution of the gel was not sufficient to differentiate clearly between mRNAs encoding mRAR-β1 and β3 isoforms with a probe spanning their common sequences. Therefore, we also carried out PCR analysis, using 5′ and 3′ oligonucleotide primers common to both isoforms (see Materials and methods and Figure 6B), on cDNAs synthesized from RNA samples containing mRAR-β1 and β3 transcripts as judged from Northern blot analysis. Southern blot hybridization of the amplified products with specific oligonucleotide probes (see Figure 6B for schematic representation) demonstrated that the oligonucleotides used as primers for amplification and probes for hybridization were highly specific, since neither cross-amplification nor cross-hybridization between various isoform cDNAs was observed (Figure 6A, lanes 1–16). Furthermore, Figure 6A clearly shows the expected size cDNA species corresponding to amplified mRAR-β1 (1477 bp) and mRAR-β3 (1558 bp) cDNAs. Since the same oligonucleotides were used in a given PCR reaction, and the difference in length between the two isoforms is small (81 bp), we believe that, for a given sample, the relative intensities of the two bands in a given lane are truly indicative of the relative abundance of the mRAR-β1 and β3 mRNA isoforms.

The results presented in Figures 5 and 6A, as well as in Table I, clearly show that the mRAR-β isoforms are differentially expressed in the mouse. Interestingly, mRAR-β1 and β3 displayed a more restricted pattern of expression than mRAR-β2 which was the only isoform which could be detected in RNAs from the adult kidney, liver, heart and skeletal muscles (Figure 5, lanes 11, 12, 14 and 15, respectively). Note the apparent complete absence of RAR-β transcripts in the adult intestine. The expression of mRAR-β1 and β3 in adult mouse tissues was confined primarily to the brain (Figure 5, lane 18), where mRAR-β1 mRNA was detected at a level 1.35 times higher than that of mRAR-β3 (Figure 6A and Table I). Both mRAR-β1 and β3 isoform mRNAs were present in the adult brain at levels apparently much higher than those of mRAR-β2 (see Table I). mRAR-β1 and β3 were also expressed at approximately equal levels (not shown) in the fetal brain, where only trace amounts of mRAR-β2 mRNA could be detected (Figure 5, lane 19 and Table I). Furthermore, both adult lung (Figures 5 and 6A, lanes 13 and 2, respectively) and skin (Figures 5 and 6A, lanes 17 and 3, respectively) predominantly contained mRAR-β3 isoform mRNA, although at a lower level than the brain. Northern blot analysis also showed that similarly to mRAR-β2, mRAR-β1 and/or β3 was induced by RA treatment (3.3×10^{-7} M for 24 h) of embryonal carcinoma (EC) cells (Figure 5, lanes 7 and 9), albeit to a lower level. PCR analysis indicated that mRAR-β1 mRNA was present at much higher levels than mRAR-β3 mRNA in RA-induced EC cells (Figure 6A, lanes 8–11 and Table I). Similarly, mRAR-β1 but not β3 transcripts were detected in RA-treated embryonic stem (ES) cells (Figure 6A, lane 7). Northern blot, and more clearly the PCR analysis, also showed that expression of mRAR-β2 was induced in RA-treated ES cells (Figures 5 and 6A, lanes 5 and 13, respectively). All three isoforms were expressed in the embryo at days 9.5, 11.5, 12.5, 14.5 post-coitum (p.c.) (Figures 5 and 6A, lanes 1–4 and 4–6, respectively);

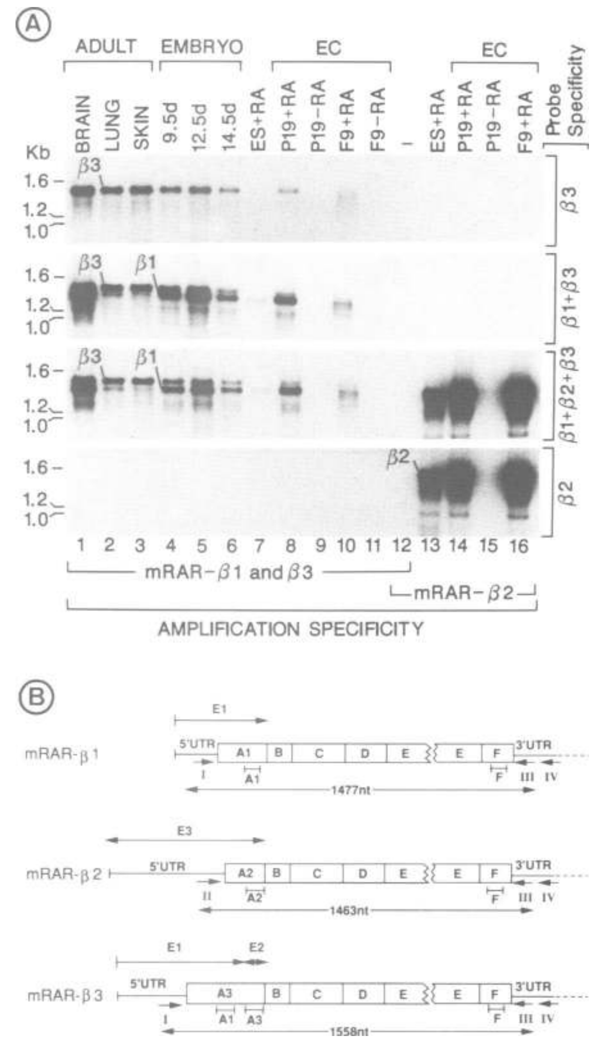

Fig. 6. PCR analysis of mRAR-β mRNA isoforms. (**A**) PCR was used to amplify mRAR-β1/β3 and mRAR-β2 cDNAs, lanes 1–11 and 13–16, respectively, synthesized (see Materials and methods) using total and poly(A)$^+$ RNAs from various sources as indicated at the top of each lane (1–16, see Materials and methods). Lane 12 corresponds to a negative control where cDNA was omitted from the PCR reaction. Specificities of oligonucleotides used as primers for amplification and as hybridization probes after electrophoresis and blotting of the amplified cDNAs (see also below) are indicated at the bottom and the right side of the figure, respectively. The bands corresponding to amplified mRAR-β1, β2 and β3 cDNA isoforms are identified in the figure as β1, β2 and β3, respectively. Positions of migration of molecular size markers (in kb) in each gel are shown on the left side of the figure. (**B**) Schematic representation of the experimental design for the above PCR analysis. Open boxes indicate the unique (A1–A3) and common (B–F) coding regions of each mRAR-β cDNA isoform and thin lines represent their 5′- and 3′-UTR sequences. Single- and double-headed arrows indicate E1 and E2–E3 exons, respectively, as in Figure 3A. Approximate locations of oligonucleotides used as primers for reverse transcription and PCR are indicated by short arrows numbered with Roman numerals IV and I–III, respectively. Note that oligonucleotides III and IV have the same sequence for all isoforms and oligonucleotides I and II are specific for mRAR-β1/β3 and mRAR-β2 isoform sequences, respectively. mRAR-β1, β2 and β3 specific ^{32}P-labelled oligonucleotide hybridization probes correspond to the regions indicated with small brackets as A1, A2 and A3, respectively. A ^{32}P-labelled oligonucleotide probe common to all the isoforms had a sequence from the F region of mRAR-β and is indicated in the diagram with a small bracket labelled F. Sequences of all the A1, A2, A3 and F oligonucleotides are available on request. Sizes (in nt) of the expected cDNAs are indicated within long arrows below the scheme representing each mRAR-β isoform.

Table I. Relative abundance of mRAR-β isoform mRNAs in mouse embryo and adult tissues

	Embryo			EC		Adult			Embryo
	9.5 d	12.5 d	14.5 d	P19+RA	F9+RA	Lung	Skin	Brain	Brain 16.5 d
mRAR-β1 / mRAR-β3	1.35	1.90	2.03	5.70	5.80	0.35	0.20	1.35	–
mRAR-β1+β3 / mRAR-β2	1.06	1.82	3.32	0.23	0.30	6.26	>15	8.59	14.08
mRAR-β1 / mRAR-β2	0.61	1.19	2.22	0.20	0.26	1.62	>2.5	4.93	–
mRAR-β3 / mRAR-β2	0.45	0.63	1.10	0.03	0.04	4.64	>12.5	3.66	–

For a given sample (as indicated at the top of each column in the table), the apparent mRAR-β1/mRAR-β3 ratio was calculated using values obtained from densitometric scanning of mRAR-β1 and β3 specific bands in the autoradiograms shown in the two middle horizontal panels in Figure 6A (β1+β3, and β1+β2+β3 probes, lanes 1–11). In both cases, the calculated ratios were nearly the same; the average of the two ratios is given. The apparent mRAR-β1+β3/mRAR-β2 ratio was computed using values obtained from densitometric scanning of the Northern blot autoradiograms (Figure 5). Note that before calculating the ratio indicated in each column, the values obtained from scanning corresponding lanes in the autoradiogram of the blot which was hybridized with the mRAR-β1/β3 specific probe had to be corrected by a factor reflecting the difference between the two blots shown in Figure 5, with respect to both blot intensities of actin hybridization (blotting efficiency), and differences in specific activity (assuming that the RAR-β1/β3 and RAR-β2 probes hybridized with the same efficiency), length and G–C content of the two ^{32}P-labelled probes used (see text and Materials and methods for details). On average, the correction factor, by which the values corresponding to the integrated scanning intensities of the mRAR-β1/β3 specific bands were multiplied, was ~1.7. The last two ratios (mRAR-β1/mRAR-β2 and mRAR-β3/mRAR-β2) were then derived using the value from the above ratios. Dashes in the last column of the table indicate that the corresponding ratios were not calculated. Error is estimated to be within 10% of the given values.

however, the amounts of mRAR-β2 and mRAR-β1/β3 transcripts decreased and increased respectively during the course of embryogenesis (Table I).

Discussion

All mRAR-β isoforms differ only in their 5′-UTR and/or A region sequences and derive from the same gene by multiple promoter usage and alternative splicing

At least three 5′-divergent cDNA (mRAR-β1, β2 and β3) isoforms of mRAR-β exist in the mouse. Their corresponding mRNAs are generated, from a primary transcript(s) of a single gene, by differential splicing of at least three exons (E1, E2, E3) encoding the 5′-UTR and A regions with the exons encoding the common B–F regions (see Figure 3B). All three mRAR-β cDNAs described here were shown in transient co-transfection experiments to encode RA-inducible *trans*-acting enhancer factors (not shown). Isoforms of mRAR-γ (Giguère *et al.*, 1990; Kastner *et al.*, 1990) and α [Leroy *et al.*, 1991 (accompanying paper)], which vary in their A region and/or 5′-UTR sequences, have also been isolated. The observation that the exons encoding the A regions of mRAR-β1 and β2 are in the same 5′ to 3′ order (Figure 3B) as exons encoding mRAR-γ1 and γ2 (Kastner *et al.*, 1990) was the initial rationale for the nomenclature used here to designate mRAR-β isoforms. The observation that the amino acid sequences comprising the A regions of mRAR-β1 and β2 show a higher degree of homology with mRAR-γ1 and γ2 A regions (31% and 27%, respectively), than vice versa (23% and 16%), further supports this nomenclature. Moreover, there is a high degree of sequence similarity (51%) between the mRAR-β1 and mRAR-α1 (previously called mRAR-α0, Zelent *et al.*, 1989) A region sequences and, interestingly, the exon encoding the A region of mRAR-α1 is also the furthest 5′-located coding exon within the mRAR-α gene [Leroy *et al.*, 1991 (accompanying paper)]. We note also that all three A regions of mRAR-α1, β1 and γ1 have a high content of proline residues (20%, 24% and 16%, respectively), which suggests some functional importance for prolines in these domains.

Our preliminary analysis of mRAR-β genomic DNA clones as well as Southern blot hybridization data indicate that mRAR-β1/β3 and mRAR-β2 isoforms must be expressed from two different promoters in the mRAR-β gene. The high degree of RAR-β2 promoter sequence conservation from mouse to human (at least in the region from −87 to −1; see de Thé *et al.*, 1990; Sucov *et al.*, 1990), suggests that the RA-dependent expression pattern of RAR-β2 mRNA is the same during both human and mouse development and in adult tissues. The genomic sequences containing either the mouse or human RAR-β1/β3 promoter(s) have not yet been isolated, and it is unknown whether the RA induction of mRAR-β1 and mRAR-β3 expression is occurring at the level of transcription and/or mRNA maturation/stabilization. In this respect, it will be interesting to see whether the mRAR-β1/β3 promoter(s) also contains a RARE and how its sequence compares with that of the RAR-β2 promoter RARE.

Our present findings bring out some interesting suggestions. First, it appears that all of the mRAR genes (α, β and γ) may have the same or very similar genomic organization. Therefore, one can expect to find more than one promoter in each RAR gene. Likewise, since seven different isoforms of mRAR-γ (Kastner *et al.*, 1990) and of mRAR-α [Leroy *et al.*, 1991 (accompanying paper)] were identified, it is likely that additional mRAR-β isoforms remain to be discovered. Second, the isoforms of the three RAR subtypes which show amino acid sequence similarity in their A region (such as mRAR-α1, mRAR-β1 and mRAR-γ1 as well as mRAR-β2 and mRAR-γ2) may be functionally related.

Are the three mRAR-β isoforms exerting specific functions?

The conservation of RAR-β isoform DNA sequences from mouse to human, as well as their differential spatio-temporal patterns of expression in mouse embryo and adult tissues, argues strongly in favour of each isoform having an important and specific function in development and/or adult animal. In this respect it is important to recall that the A/B region of the progesterone (Tora *et al.*, 1988a) and oestrogen (Tora *et al.*, 1988b, 1989; Tasset *et al.*, 1990) receptors have been shown to contain transcriptional activation functions which exhibit cell and promoter specificities. Our present data on the distribution of mRAR-β isoform RNAs (as judged from blotting analysis) are corroborated by recent *in situ* hybridization studies performed using a riboprobe containing B–F region sequences common to all three isoforms, which showed a restricted spatio-temporal pattern of mRAR-β gene expression during mouse development (days 7.5–14.5 p.c.; Dollé *et al.*, 1990; Ruberte *et al.*, 1991). With the exception of fetal skin, liver and heart, mRAR-β gene expression was detected by *in situ* hybridization in specific regions of all fetal organs positive for either mRAR-β1/β3 or β2 isoforms in the adult animal (see Figure 5). The lack of detection of mRAR-β gene transcripts in the above tissues through day 14.5 p.c. could indicate either that the mRAR-β gene is expressed in these organs at later stages, or that its level of expression is too low to be detected by *in situ* hybridization. Interestingly, high levels of mRAR-β1/β3 isoform mRNAs are restricted primarily to the adult and 16.5 day p.c. fetal brains, which suggests that these isoforms are important in some specific aspects of mammalian neurodevelopment. Likewise, mRAR-β3 expression is apparently high in comparison with that of other mRAR-β isoforms in the adult mouse lung and skin. In this respect, we note that mRAR-β transcripts were specifically detected by *in situ* hybridization in bronchial epithelia during development (Dollé *et al.*, 1990).

mRAR-β2 is clearly the predominant isoform in RA-treated EC and ES cells, and its level of expression decreases in comparison with that of either mRAR-β1 or β3 as embryogenesis progresses (see Table I). This temporal pattern of expression, which is reminiscent of the one observed for mRAR-γ1 and γ2 (Kastner *et al.*, 1990), suggests that the mRAR-β2 isoform may play an important role early in development. In this respect, we note that expression of the mRAR-β gene is seen in discrete locations early in development (7.5 days p.c.) at a time when RAR-α and γ gene expression is not clearly apparent (Ruberte *et al.*, 1991).

Further *in situ* hybridization studies, using isoform specific probes, are necessary to investigate the differential expression patterns of the three mRAR-βs in the mouse embryo as well as in adult tissues. Studies employing homologous recombination should also be most helpful in revealing the possible specific function of each mRAR-β isoform.

Possible involvement of the specific 5′-UTRs in differential expression of the mRAR-β isoforms

It is remarkable that the DNA sequence of the entire RAR-β2 5′-UTR is highly conserved between mouse and human. This suggests that the RAR-β2 5′-UTR has some important and specific function(s) which has been evolutionarily preserved. Although its nature remains to be determined, it should be noted that both mouse and human RAR-β2 5′-UTRs encode three conserved upstream sORFs and can form stable secondary structures (see Results). The presence of upstream sORFs in the 5′-UTR of the yeast *GCN4* gene mRNA has been shown to be important for tight translational control of its expression (see Hinnebusch, 1990 for review). Although analogous regulatory mechanisms have not yet been found in higher eukaryotes, the presence of extensive secondary structures and upstream sORFs in long 5′-UTRs of eukaryotic mRNAs generally tends to impede translation (Rao *et al.*, 1988; Clemens, 1989; Kozak, 1989; Horiuchi *et al.*, 1990; Waterhouse *et al.*, 1990; and references therein).

Similarly to mRAR-β2, the 5′-UTR(s) of mRAR-β1/β3 (presently it is not known if the mRAR-β1 and β3 5′-UTRs remain colinear up to the cap site) also encode(s) upstream sORFs and can form stable secondary structure(s). Thus, the different mRAR-β isoform 5′-UTRs may contribute to differential stabilities and/or translation efficiencies of their mRNAs. This would increase the combinatorial possibilities for controlling the spatio-temporal expression of the RAR-β isoforms.

Conclusion

It is clear from this and other studies [Giguère *et al.*, 1990; Kastner *et al.*, 1990; Leroy *et al.*, 1991 (accompanying paper)] that all three RAR genes exhibit a complex pattern of organization and expression. These properties are shared with other developmentally important loci that encode known transcription factors (see for example Cho *et al.*, 1988; Robertson, 1988; Zavortink and Sakonju, 1989; Bermingham *et al.*, 1990), most notably thyroid hormone receptors (TR) which belong to the same superfamily of nuclear receptors as RARs (see for reviews Evans, 1988; Green and Chambon, 1988). The TR-β gene [which in the human mapped to the same chromosome as hRAR-β (Brand *et al.*, 1988)] may encode several N-terminally different isoforms of which two were shown to have differential patterns of expression (Hodin *et al.*, 1989; Forrest *et al.*, 1990). Further work is obviously necessary to characterize the promoters and exon/intron organization of the mRAR-β gene and to establish the possible specific functions of the various isoforms. Nevertheless, it is clear that the increasing complexity of the RAR gene family structure and expression could provide the diversity of RA-inducible enhancer factors which may be required to account for the highly pleiotropic effects of RA during vertebrate development and in the adult animal.

Materials and methods

Cloning of mRAR-β1 and β3 isoforms

Both mRAR-β1 and β3 cDNA isoforms were isolated using the anchored PCR technique (Loh *et al.*, 1989), following previously described methods (Kastner *et al.*, 1990). G-tailed cDNAs, derived from poly(A)$^+$ RNAs isolated from various stage mouse embryos (8.5, 9.5, 11.5 and 14.5 days p.c.), adult tissues (muscle and skin) and RA-responsive EC cells (P19 and F9, untreated and treated for 24 h with 3.3×10^{-7} M RA), were used in two rounds of 30 cycle amplification. Two nested mRAR-β B region oligonucleotide primers (first primer 5′-TAAAGCTTGGATCCCCGAGG-AGGAGGAAGTGGAGATGGT; second primer 5′-TAGGATCCT-GGGCTCGGGACGAGCTCCT, the non-specific linker sequences are underlined) and 5′ primers as described by Loh *et al.* (1989) were used.

Both rounds of amplification were carried out using annealing steps at 45°C for the first five cycles, and then at 55°C for the next 25 cycles. In both cases, duration time was 1 min and 30 s. Denaturation and extension were in each case for 20 s at 94°C and for 3 min at 72°C, respectively, except in the first cycle where denaturation was at 95°C for 1 min. At the end of each round of 30 cycles, further extension at 72°C for 10 min was included. Before subcloning, the cDNAs from individual PCR reactions which utilized the above indicated G-tailed cDNAs were pooled. All buffer conditions, purification (both after the first and second rounds of PCR), sequencing, subcloning and analysis of PCR amplified products were as previously described (Kastner *et al.*, 1990). Six independent clones were obtained for both mRAR-β1 and β3, and three of each were sequenced on both strands using Sequenase (Stratagene) and standard dideoxy sequencing methodology (Ausubel *et al.*, 1987).

In order to obtain mRAR-β2 cDNA clone, whose 5'-UTR and A region sequence appear in Figure 2, a λZapII cDNA library constructed from RA-treated (3.3×10^{-7} M for 48 h) F9 cell poly(A)$^+$ RNA was screened under stringent conditions with ^{32}P-labelled mRAR-β2 cDNA (Zelent *et al.*, 1989). Several clones were obtained representing two independent isolates. One clone was used to obtain the sequence shown in Figure 2. A mouse genomic DNA clone which contained the mRAR-β2 promoter as well as exon 3 (E3) sequence, was obtained using a ^{32}P-labelled mRAR-β2 specific probe (nt 125–498 in Figure 2) from an L cell genomic DNA library constructed in λEMBL3 vector. Screening conditions were as described for isolation of cDNA clones (see above and Zelent *et al.*, 1989). The promoter and E3 sequence shown in Figure 2 have also been derived from sequencing of the corresponding regions in this clone.

Southern blotting

Human and mouse genomic DNAs were isolated from lymphocytes and C57BL/c mouse liver, respectively, using standard methodology (Ausubel *et al.*, 1987). DNAs (15 μg) were digested thoroughly with *Hin*dIII, *Bam*HI and *Eco*RI (New England Biolabs) according to the supplier's suggestions. After electrophoresis in 0.8% agarose and 1 × TAE (Ausubel *et al.*, 1987) at 3 V/cm for ~6 h, DNAs were transferred to Schleicher and Schuell BAS 85 nitrocellulose membranes as suggested by the manufacturer. mRAR-β1/β3 and β2 ^{32}P-labelled probes were as described for Northern analysis. The specific activity of the three probes was ~10^9 c.p.m./μg of DNA. Hybridization was as for Northern blots (see below) except that 40% formamide was used. The most stringent wash was carried out at 65°C in 0.5 × SSPE (Ausubel *et al.*, 1987)/0.1% SDS/0.03% NaPP$_i$ for 20 min. Blots were then exposed for 24 h to Kodak XAR-5 film at −80°C using two intensifying screens.

Northern blot analysis

Isolation of poly(A)$^+$ RNAs, Northern blotting and hybridization conditions were as described (Zelent *et al.*, 1989). Probes, consisting of an mRAR-β3 fragment (nt 1–500, Figure 1) and mRAR-β2 (nt 125–498, Figure 2) sequences, were labelled by random priming with [^{32}P]dCTP (3000 Ci/mmol, Amersham) to ~5×10^8 and 10^9 c.p.m./μg of DNA, respectively. All probes were used in hybridization at 2.5×10^6 c.p.m./ml. After the most stringent wash (Zelent *et al.*, 1989), blots were exposed for 8 days at −80°C using two intensifying screens and Kodak XAR-5 film. Actin probe was as previously described (Zelent *et al.*, 1989).

PCR analysis of RNA (see Figure 6B)

All cDNAs were synthesized using 1.3 μg of poly(A)$^+$ (see Figure 5) or 5 μg of total (ES cell) RNAs, an oligonucleotide primer (IV) from the 3'-UTR of mRAR-β (5'-CTTGCATTTTCAATCTGGAACTGAAGGTCCTG, see Zelent *et al.*, 1989) and Mo-MLV reverse transcriptase (BRL) under conditions recommended by the supplier. For PCR amplification, 1.5 μl out of 25 μl of reverse transcriptase reaction was used directly. The 5' primers for amplification were 5'-ATCGGCCGTGATTCTGGACCTGTAGTAA (I) for mRAR-β1 and β3 (nt 73–93 and 202–222, respectively, Figure 1) and 5'-ATCGGCCGTGGACTTTTCTGTGCGGCTCG (II) for mRAR-β2 (nt 389–411, Figure 2). The 3' oligonucleotide primer (III) was from their common 3'-UTR and had a sequence (5'-ATGGATCCGGGAATGTCTGC-AACAGCTGGA) located immediately upstream of the primer used for reverse transcription. Amplification conditions were as described in the previous section. Aliquots of 15 μl from each reaction were then loaded onto 1.5% agarose gels and amplified cDNAs were electrophoresed in 1 × TAE at ~3 V/cm until the bands corresponding to amplified mRAR-β1 and β3 cDNAs, when visualized under short-wave UV light after ethidium bromide staining, were clearly resolved. Nucleic acids were then transferred to BAS 85 nitrocellulose as recommended by the supplier. Hybridization oligonucleotides (A1, A2, A3 and F), which were as indicated in Figure 6B, were labelled to ~10^8 c.p.m./μg with ^{32}P using T4 polynucleotide kinase (New England Biolabs) and [γ^{32}P]ATP (>5000 Ci/mmol, Amersham) and standard reaction conditions (Ausubel *et al.*, 1987). All probes were used in hybridization solution at 10^6 c.p.m./ml. Hybridization was as described for Northern and Southern blotting (see above), except that no formamide was used and the incubation temperature was 50°C. The most stringent wash was carried out for 20 min at 60°C in 2 × SSPE/0.1% SDS/0.03% NaPP$_i$. All blots were exposed for 20 h at room temperature and without screens to Kodak XAR-5 film. Both Northern and PCR data were quantified using a GS300 scanning densitometer from Hoefer Scientific Instruments (San Francisco, CA).

Acknowledgements

We thank A.Dierich for gifts of embryonic stem cell RNA, A.Staub for oligonucleotide synthesis, the cell group for cells, C.Reibel for technical assistance, C.Werlé and B.Boulay for illustrations and the secretarial staff for typing the manuscript. This work was supported by the Institut National de la Santé et de la Recherche Médicale (Grant CNAMTS), the Centre National de la Recherche Scientifique, the Fondation pour la Recherche Médicale Française and the Association pour la Recherche contre le Cancer. A.Z. was the recipient of a fellowship from the American Cancer Society. C.M. was the recipient of a fellowship from NIH (GM 13597).

References

Alles,A.J. and Sulik,K.K. (1989) *Teratology*, **40**, 163–171.

Asselineau,D., Bernard,B.A., Bailly,C. and Darmon,M. (1989) *Dev. Biol.*, **133**, 322–335.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology.* Wiley, New York.

Beato,M. (1989) *Cell*, **56**, 335–344.

Benbrook,D., Lernhardt,E. and Pfahl,M. (1988) *Nature*, **333**, 669–672.

Bermingham,J.R., Jr, Martinez-Arias,A., Petit,M.G. and Scott,M.P. (1990) *Development*, **109**, 553–566.

Brand,N., Petkovich,M., Krust,A., Chambon,P., de Thé,H., Marchio,A., Tiollais,P. and Dejean,A. (1988) *Nature*, **332**, 850–853.

Brockes,J.P. (1989) *Neuron*, **2**, 1285–1294.

Brockes,J.P. (1990) *Nature*, **345**, 766–768.

Cho,K.W.Y., Goetz,J., Wright,C.V.E., Fritz,A., Hardwicke,J. and DeRobertis,E.M. (1988) *EMBO J.*, **7**, 2139–2149.

Chytil,F. (1984) *Pharmacol. Rev.*, **36**, 93S–100S.

Clemens,M.J. (1989) *Curr. Opinion Cell Biol.*, **1**, 1160–1167.

Crawford,K. and Stocum,D.L. (1988) *Development*, **104**, 703–712.

DeJean,A., Bougueleret,L., Grzeschik,K.H. and Tiollais,P. (1986) *Nature*, **322**, 70–72.

de Thé,H., Marchio,A.,Tiollais,P. and Dejean,A. (1987) *Nature*, **330**, 667–670.

de Thé,H., del Mar Vivanco-Ruiz,M., Tiollais,P., Stunnenberg,H. and Dejean,A. (1990) *Nature*, **343**, 177–180.

Dollé,P., Ruberte,E., Kastner,P., Petkovich,M., Stoner,C.M., Gudas,L. and Chambon,P. (1989) *Nature*, **342**, 702–705.

Dollé,P., Ruberte,E., Leroy,P., Morriss-Kay,G. and Chambon,P. (1990) *Development*, **110**, in press.

Durston,A.J., Timmermans,J.P.M., Hage,W.H., Hendriks,H.F.H., de Vries,N.H., Heideveld,M. and Nieuwkoop,P.D. (1989) *Nature*, **340**, 140–144.

Eichele,G. (1989) *Trends Genet.*, **5**, 246–251.

Erba,H.P., Eddy,R., Shows,T., Kedes,L. and Gunning,P. (1988) *Mol. Cell. Biol.*, **8**, 1775–1789.

Evans,R.M. (1988) *Science*, **240**, 889–895.

Forrest,D., Sjöberg,M. and Vennström,B. (1990) *EMBO J.*, **9**, 1519–1528.

Giguère,V., Ong,E.S., Segui,P. and Evans,R.M. (1987) *Nature*, **330**, 624–629.

Giguère,V., Shago,M., Zirngibl,R., Tate,R., Rossant,J. and Varmuza,S. (1990) *Mol. Cell. Biol.*, **10**, 2335–2340.

Green,S. and Chambon,P. (1988) *Trends Genet.*, **4**, 309–314.

Hassouna,N., Michot,B. and Bachellerie,J.P. (1984) *Nucleic Acids Res.*, **12**, 3563–3583.

Hinnebusch,A.G. (1990) *Trends Biochem. Sci.*, **15**, 148–152.

Hodin,R.A., Lazar,M.A., Wintman,B.I., Darling,D.S., Koenig,R.J., Larsen,P.R., Moore,D.D. and Chin,W.W. (1989) *Science*, **244**, 76–79.

Horiuchi,T., Macon,K.J., Kidd,V.J. and Volanakis,J.E. (1990) *J. Biol. Chem.*, **265**, 6521–6524.

Jacobson,A.B., Good,L., Simonetti,J. and Zuker,M. (1984) *Nucleic Acids Res.*, **12**, 45–52.

Kastner,P., Krust,A., Mendelsohn,C., Garnier,J.M., Zelent,A., Leroy,P.,

Staub,A. and Chambon,P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2700–2704.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.
Kozak,M. (1986) *Cell*, **47**, 481–483.
Kozak,M. (1989) *Mol. Cell. Biol.*, **9**, 5134–5142.
Krust,A., Kastner,P., Petkovich,M., Zelent,A. Chambon,P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5310–5314.
Lammer,G.J., Chen,D.T., Hoar,R.M., Agnish,N.D., Benke,P.J., Braun,J.T., Curry,C.J., Fernhoff,P.M., Grix,A.W., Lott,I.T., Richard,J.M. and Sun,S.C. (1985) *New Engl. J. Med.*, **313**, 837–841.
Langille,R.M., Paulsen,D.F. and Solursh,M. (1989) *Differentiation*, **40**, 84–92.
Langman,J. and Welch,G.W. (1967) *J. Comp. Neurol.*, **131**, 15–26.
Leroy,P., Krust,A., Zelent,A., Mendelsohn,C., Garnier,J.-M., Kastner,P., Dierich,A. and Chambon,P. (1991) *EMBO J.*, **10**, 59–69.
Loh,E.Y., Elliott,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) *Science*, **243**, 217–220.
Lotan,R. (1980) *Biochim. Biophys. Acta*, **605**, 33–91.
Maden,M. (1982) *Nature*, **295**, 672–675.
Maden,M. (1985) *Trends Genet.*, **1**, 103–104.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Morris,G.M. (1972) *Am. J. Anat.*, **113**, 241–250.
Morris,G.M. and Thorogood,P.V. (1978) In Johnson,M.H. (ed.), *Development in Animals*. Elsevier North-Holland, Amsterdam, Vol. 3, pp. 363–411.
Petkovich,M., Brand,N.J., Krust,A. and Chambon,P. (1987) *Nature*, **330**, 444–450.
Ragsdale,C.W., Petkovich,M., Gates,P.B., Chambon,P. and Brockes,J.P. (1989) *Nature*, **341**, 654–657.
Rao,C.D., Pech,M., Robbins,K.C. and Aaronson,S.A. (1988) *Mol. Cell. Biol.*, **8**, 284–292.
Raynal,F., Michot,B. and Bachellerie,J.P. (1984) *FEBS Lett.*, **167**, 263–268.
Roberts,A.B. and Sporn,M.B. (1984) In Sporn,M.B., Roberts,A.B. and Goodman,D.S. (eds), *The Retinoids*. Academic Press, New York, Vol. 2, pp. 209–286.
Ruberte,E., Dollé,P., Krust,A., Zelent,A., Morriss-Kay,G. and Chambon,P. (1990) *Development*, **108**, 213–222.
Ruberte,E., Dollé,P., Chambon,P. and Morriss-Kay,G. (1991) *Development*, **111**, in press.
Satre,M.A. and Kochhar,D.M. (1989) *Dev. Biol.*, **133**, 529–536.
Shapiro,S.S. (1985) In Sherman,H.I. (ed.), *Retinoids and Cell Differentiation*. CRC Press, Boca Raton, FL, pp. 30–55.
Slack,J.M.W. (1987a) *Nature*, **327**, 553–554.
Slack,J.M.W. (1987b) *Trends Biochem. Sci.*, **12**, 200–204.
Sucov,H.M., Murakami,K.K. and Evans,R.M. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5392–5396.
Sulik,K.K. and Dehart,D.B. (1988) *Teratology*, **37**, 527–537.
Summerbell,D. and Maden,M. (1990) *Trends Neurosci.*, **13**, 142–147.
Tasset,D., Tora,L., Fromental,C., Scheer,E. and Chambon,P. (1990) *Cell*, **67**, 1177–1187.
Thaller,C. and Eichele,G. (1987) *Nature*, **327**, 625–628.
Tickle,C., Alberts,B., Wolpert,L. and Lee,J. (1982) *Nature*, **296**, 564–566.
Tora,L., Gronemeyer,H., Turcotte,B., Gaub,M.-P. and Chambon,P. (1988a) *Nature*, **333**, 185–188.
Tora,L., Gaub,M.-P., Mader,S., Dierich,A., Bellard,M. and Chambon,P. (1988b) *EMBO J.*, **7**, 3771–3778.
Tora,L., White,J., Brou,C., Tasset,D., Webster,N., Scheer,E. and Chambon,P. (1989) *Cell*, **59**, 477–487.
Wagner,M., Thaller,C., Jessell,T. and Eichele,G. (1990) *Nature*, **345**, 819–822.
Waterhouse,P., Khokha,R. and Denhardt,D.T. (1990) *J. Biol. Chem.*, **265**, 5585–5589.
Wedden,S.E. (1987a) *Development*, **99**, 341–351.
Wedden,S.E., Lewin-Smith,M.R. and Tickle,C. (1987b) *Dev. Biol.*, **122**, 78–89.
Wedden,S.E., Ralphs,J.R. and Tickle,C. (1988) *Development*, **103**, 31–40.
Zavortink,M. and Sakonju,S. (1989) *Genes Dev.*, **3**, 1969–1981.
Zelent,A., Krust,A., Petkovich,M., Kastner,P. and Chambon,P. (1989) *Nature*, **339**, 714–717.
Zuker,M. and Stiegler,P. (1981) *Nucleic Acids Res.*, **9**, 133–148.

Received on September 17, 1990

Note added in proof

The sequence data in this paper are available from EMBL under the following accession numbers: X56569(RAR-β1), X56573(RAR-β2) and X56574 (RAR-β3).