Characterization of a functional promoter for the human retinoic acid receptor-alpha (hRAR-α)

Nigel J.Brand+, Martin Petkovich§ and Pierre Chambon*
Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11, rue Humann, 67085 Strasbourg Cédex, France

Received September 27, 1990; Accepted October 30, 1990 EMBL accession nos X56057, X56058

ABSTRACT

The three retinoic acid receptors RAR-α, β and γ are thought to mediate the effects of RA *in vivo*. We have determined here the exon organisation in the 5′ region of the human RAR-α (hRAR-α) gene, and have identified its promoter. This promoter drives the expression of promoterless β-globin or CAT reporter genes when transfected into HeLa, Cos-1 or mouse embryonal carcinoma (EC) P19.6 cells in culture. There are no TATA or CCAAT-box elements in this promoter, which appears to belong to the class of promoters made up of an initiator element preceded by several putative binding sites for the transcription factor Sp1. In addition, the hRAR-α promoter region contains a number of sequences that are similar to known enhancer elements. Notably, the hRAR-α promoter contains a sequence identical to a binding site for the Krox-20 transcription factors, a zinc finger-containing protein which is thought to play a role in the early development of the mouse central nervous system.

INTRODUCTION

Retinoic acid (RA) plays a major role in a variety of developmental processes (see 1–7 for references and reviews). In culture, certain murine and human teratocarcinoma cell lines respond to RA by undergoing differentiation forming, for example, cells which resemble primitive endoderm cells or neurons (1, 2, 8, 9). Often, these events are accompanied by specific changes in gene expression (for example, see 10–13 and refs therein). RA has been strongly implicated as an active morphogen in pattern formation, chiefly using the developing chick limb bud and the regenerating urodele amphibian limb blastema as models (4–7, 14, 15 and refs therein).

The role of RA in controlling such diverse processes has been strengthened by the identification by us and others of three related nuclear receptors for RA, termed retinoic acid receptors (RARs) α, β and γ, that are members of the steroid/thyroid hormone receptor superfamily of inducible transcriptional enhancer factors (16–18 and refs therein), and which bind RA selectively and with high affinity (19–24). The three RAR genes are expressed with varying degrees of tissue specificity during embryonic development and in adult tissues, and are found in a number of cultured cell lines (23–29). Alignment of RAR amino acid sequences and their comparison with other nuclear receptors has allowed the definition of six regions within the protein, termed A-F (23, 24, 30), including the two highly conserved regions C and E corresponding to the DNA binding and the ligand binding domains, respectively. In addition, recent evidence indicates that isoforms of mouse and human RAR-γ, which differ in the amino-terminal region A and 5′ untranslated region (5′-UTR), are generated through alternative splicing (23, 27, 31).

Previously, two partial hRAR-α complementary DNA (cDNA) sequences, which differed in the 5′ region, were published by ourselves (19) and Giguère et al. (20). Our hRAR-α cDNA clone was deduced to encode a protein of 432 amino acids, though an upstream in-frame termination codon was not seen (19). In contrast, the clone of Giguère et al. (20) encoded a 462 amino acid protein that began at a methionine codon beyond the 5′ border of our cDNA clone and was preceded by an in-frame TGA termination codon (20). Neither of the two cDNA clones was full-length in their 5′ region. We describe here the isolation and characterisation of cDNA and genomic DNA clones containing the sequences encoding the A region and the whole 5′-untranslated region (5′-UTR) of hRAR-α. This has led to the determination of the exon-intron organisation of the 5′ region of the hRAR-α gene and to the isolation of a functional promoter which resembles some RNA polymerase B (II) promoters that lack a TATA box (32, 33).

MATERIALS AND METHODS

Cloning and sequencing

Genomic clones corresponding to the hRAR-α gene were isolated from a library constructed from *Sau*3A fragments (partial digest) of human leucocyte DNA cloned into the *Bam*HI site of the vector EMBL3, supplied by Transgène S.A. (Strasbourg).

* To whom correspondence should be addressed

Present addresses: +Department of Cardiothoracic Surgery, National Heart and Lung Institute, London, SW3 6LY, UK and §Cancer Research Laboratory, Queens University, Kingston, Ontario K7L 3N6, Canada

Approximately 400,000 plaques were plated on a lawn of *Escherichia coli* Q_{358} on 150 mm diameter plates. Duplicate nylon filters (Biodyne) were laid on the plates and phage DNA was bonded to the filters according to the manufacturers instructions. Filters were hybridised according to Petkovich et al. (19), except that the hybridisation was carried out at 50°C for 36 hrs using a [^{32}P]-end labelled oligonucleotide MEO12 (5′-TCTGACCACTCTCCAGCACCAGCTTCCAGTTA-GTGGATATA-3′, see Fig. 2) from the 5′ end of the hRAR-α coding region, spanning codons 40 to 54. Filters were washed in 2×SSC/0.1% SDS at room temperature twice for 20 min. and exposed for 2 days to Kodak XAR-5 film with an intensifying screen at −80°C. Plaques corresponding to positive signals in duplicate were diluted in 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, replated on 90 mm plates, and re-screened as above. Phage DNA was prepared from positive plaques, digested with restriction enzymes and analysed by agarose gel electrophoresis followed by Southern blotting. The inserts of recombinant clones scored positive by Southern hybridisation using MEO12 as the probe were subsequently cloned into Genescribe-Z vectors pTZ18 or pTZ19 (USB) for further sequence analyses.

The screening of a randomly primed cDNA library constructed from polyA+mRNA obtained from the human breast cancer cell line T47-D was carried out as described previously (19), except that the end-labelled oligonucleotide MEO12 was used as a hybridisation probe.

RNA isolation and quantitative nuclease S1 mapping

A 614 bp *Bam*HI-*Sma*I fragment (see Fig. 1) from the subclone pRS7 containing part of hRAR-α exon 1 and sequences upstream was cloned in front of rabbit β-globin coding sequences in the polylinker of the expression vector pAL4 (a gift of J. White). 10 μg of either a construct containing the *Bam*HI-*Sma*I fragment inserted in the natural orientation (pAL4BS), the reverse orientation (pAL4SB) or the promoterless parent vector (pAL4) were transfected into HeLa cells in culture, together with 400 ng of a β-globin reference plasmid pG1B (34), containing the rabbit β-globin promoter from −109 and downstream coding sequences (see Fig. 1B), and carrier DNA (Bluescribe 13+) to a final concentration of 20 μg. Transfections were carried out as described previously (19). After 48 hrs, cytoplasmic RNA was prepared from transfected dishes using standard procedures (35).

A 760 nt [^{32}P] 5′ end-labelled single-stranded probe was generated by extending an oligonucleotide primer M6 (5′-GC-ACCATTCTGTCTGTTTTGGG-3′, complementary to +39/+60 of the rabbit β-globin gene) annealed to a single-stranded template derived from pAL4BS (Fig. 1B). The probe was purified by gel electrophoresis and hybridised to RNA prepared form cells transfected as described above. S1 nuclease mapping was performed as described previously (36). The purified probe protects a fragment of 69 nt from the internal control plasmid pG1B. S1 nuclease digestion products were resolved on 6% acrylamide gels alongside dideoxy sequencing ladders primed with the oligonucleotide M6. Following autoradiography, all samples were re-loaded on a second gel so as to normalise loading with respect to the pG1B signal.

Expression in cultured cells and CAT assays

The vector pBLCAT3+ contains the bacterial chloramphenicol acetyl transferase (CAT) gene positioned upstream of SV40 polyadenylation signals (37). The 614 bp *Bam*HI-*Sma*I fragment of pRS7 was cloned into pBLCAT3+ that had been cut first with *Bgl*II, the ends made blunt by filling in with all four dNTPs and the Klenow fragment of DNA polI, and then cut with BamHI. The resulting recombinant pRARA12 (Fig. 1B), contains the RAR-α promoter fragment linked 5′ to the CAT gene. Typically, 2−4 μg of the recombinant or the parent vector were used to transfect cells along with 2 μg of an internal standard plasmid pCH110 (Pharmacia), in which the bacterial *lac*-Z gene encoding β-galactosidase is under the control of the SV40 promoter. Carrier DNA was added to a final total of 20 μg. Other plasmids transfected were the chimaeric expression plasmid RAR-β.ER.CAS, encoding a chimaeric receptor in which the DNA binding domain of hRAR-β has been replaced by the corresponding region of the human oestrogen receptor (21), the oestrogen-responsive reporter plasmid *vit-tk*-CAT and pSVCAT, in which the SV40 early promoter drives CAT expression. Transfections, culture and CAT assays were carried out as described previously (see 19). Where appropriate, hormones were added 20 hrs after transfection to 10^{-8} M final concentration (MCF-7, T47D) or 10^{-7} M (all other lines). Control plates were treated with an equal volume of vehicle (ethanol). All transfections were carried out in duplicate. HeLa cells were grown in Dulbeccos modified medium supplemented with 5% stripped foetal calf serum (FCS); the human breast cancer line MCF-7 was cultured with 10% FCS and 0.6 μg/ml insulin. The murine teratocarcinoma P19.6 was grown in Dulbeccos with 10% FCS. A second human breast cancer cell line used, T47D, was grown in RPMI containing 10% FCS and 0.6 μg/ml insulin but for transfection, the media was replaced with MCF-7 media for 4 hrs prior to and for 20 hrs following transfection.

RESULTS

Isolation of genomic and cDNA clones encoding the hRAR-α 5′ region

To isolate genomic sequences encoding the hRAR-α gene 5′ region, a human leucocyte genomic DNA library constructed in the lambda vector EMBL3 was screened using an oligonucleotide, MEO12, specific for the A region of hRAR-α (see Figs 1A and 2). This region exhibits no homology with either hRAR-β (21) or hRAR-γ (23) sequences. Four lambda clones, each containing inserts of at least 15 kb were isolated. Fig. 1A shows a restriction map for clone λ91.1, a second partially overlapping clone λ31.1, and plasmid sub- clones derived from them. A 1.2-kb BamHI-PstI region common to both genomic clones that hybridised to MEO12 (as determined by Southern blot analysis) was sub-cloned, giving the construct pBP31.1, and its DNA sequence was determined. A human cDNA library, constructed using polyA+ mRNA isolated from the human breast cancer cell line T47D (see 19) was also screened using the oligonucleotide MEO12 and a clone, T2/20, was isolated from it. This clone, which contains a ≃ 800 bp insert, (Fig. 2, in brackets) was sequenced. It contains part of region B of hRAR-α, all of region A and an extensive 5′- untranslated region (5′-UTR) that is 541 bp long.

Comparison of the DNA sequence of the genomic subclone pBP31.1 (Fig. 1A) with that of the cDNA clone T2/20 revealed that the two sequences were identical upstream of the codon for Thr_{60} over a distance of 540 bp (shown in upper case, Exon 2 in Fig. 2). At the 3′ side, the genomic clone diverges from the T2/20 sequence within the codon for Thr_{60}, where the DNA sequence reveals a potential splice donor site CCA/gtaagt (Fig. 2, intron sequence underlined and in lower case) resembling a

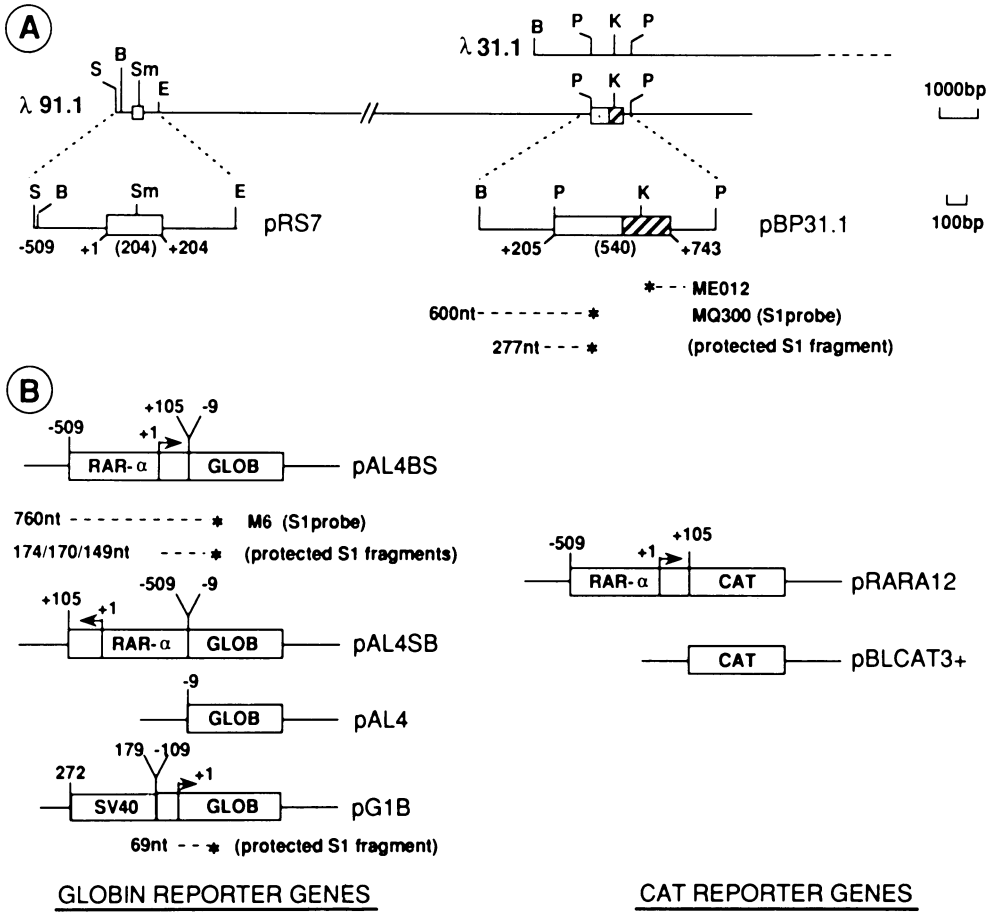

Figure 1. (A) Schematic representation of hRAR-α genomic clones. A lambda genomic clone λ91.1, isolated from a human leucocyte genomic DNA library constructed in EMBL 3 with an oligonucleotide ME012 specific for the A region of hRAR α (19), a partially overlapping clone λ31.1 and plasmid sub-clones derived from them are shown. Clone pBP31.1 contains exon 2 of hRAR-α, comprising 5'-UTR and region A (hatched box) of the receptor. Clone pRS7 contains exon 1 comprising the rest of the 5'-UTR and 5' flanking sequences. The exons are separated by an intron of at least 12-kb in length. Restriction enzyme sites are : B, BamHI; E, EcoRI; K, pnI; P, stI; S, SalI; Sm, SmaI. The 5' and 3' borders of the exons (determined by nuclease S1 mapping) are numbered relative to the beginning of exon 1 and their lengths are indicated in parentheses. The position of ME012 and a nuclease S1 probe extended from the oligonucleotide MQ300, together with the lengths of the probe and a fragment protected by it, are indicated. (B) Promoter-reporter constructs for testing the function of the hRAR-α promoter. The positions and lengths of end-labelled probes used in nuclease S1 mapping are indicated, as are the lengths of the S1 nuclease protected products. For descriptions of hRAR-α promoter constructs, which contain a 614 bp *Bam*HI-*Sma*I fragment comprising hRAR-α sequences from −509 to +105, see Materials and Methods. The globin reporter constructs are based on the vector pAL4 and contain rabbit β-globin sequences from −9. The reporter pG1B contains the SV40 72 bp enhancer linked 5' to the β-globin promoter and coding sequences from −109 (34). The CAT reporters are based on the promoterless vector pBLCAT3+ (37).

consensus donor sequence NAG/GTA$^{A}_{G}$GT (38). The isolation recently of cDNA clones for the RAR-α, β and γ from mouse (24, 27) and RAR-γ from human (23) indicates that the boundary between regions A and B is sharply defined and well conserved. In hRAR-α, for example, the A/B boundary is located between Ala_{59} and Thr_{60}, and correlates with previous observations by us and others (19, 21, 39) that this represents an exon boundary. Taken together, these data indicate that region A of hRAR-α is encoded on a separate exon from that encoding region B, and strongly suggests that this boundary occurs at an equivalent position in the β and γ receptors.

The genomic and cDNA sequences diverged completely at a point located 176 bp downstream from the 5' end of T2/20 and approximately 500bp from the 5' end of the pBP31.1 insert. In pBP31.1, this divergence from the cDNA sequence occurs at a sequence ttctttttctgcag/CAT (Fig. 2, underlined) that resembles a splice acceptor site (38). The identity of this splice site was confirmed by nuclease S1 mapping (36) using a 600 nt single-stranded end- labelled probe, extended from an antisense oligonucleotide MQ300 which hybridises within the 5'-UTR portion of the exon (see Fig. 1A and Fig. 2). A protected fragment of 277 nt in length was seen with MCF-7 or T47-D total RNA, but not with yeast tRNA (data not shown). This result confirmed the length of the exon 2 contained within pBP31.1 as being 540 bp.

These data indicated that the first amino acid coding exon (exon 2 encoding region A in Fig. 2) was preceded by at least one upstream exon. To isolate this exon, a 190 bp *Eco*RI-*Bst*II fragment from the extreme 5' end of T2/20 was end-labelled and used to isolate a 1.1-kb *Sal*I-*Eco*RI fragment of the genomic clone λ91.1. The fragment was cloned, giving the recombinant pRS7, and its DNA sequence was determined. The sequence revealed that this fragment indeed contained the 178 bp from the 5' end of T2/20 as a contiguous sequence, thus encoded in a single exon (see Fig. 2, exon 1). A putative splice donor site GAG/gtaatt (see above) (underlined in Fig. 2, intron sequence in lower case) was identified at the 3' side of this DNA sequence. In the genomic clone λ91.1 this 5'-UTR sequence encoded in exon 1 and present

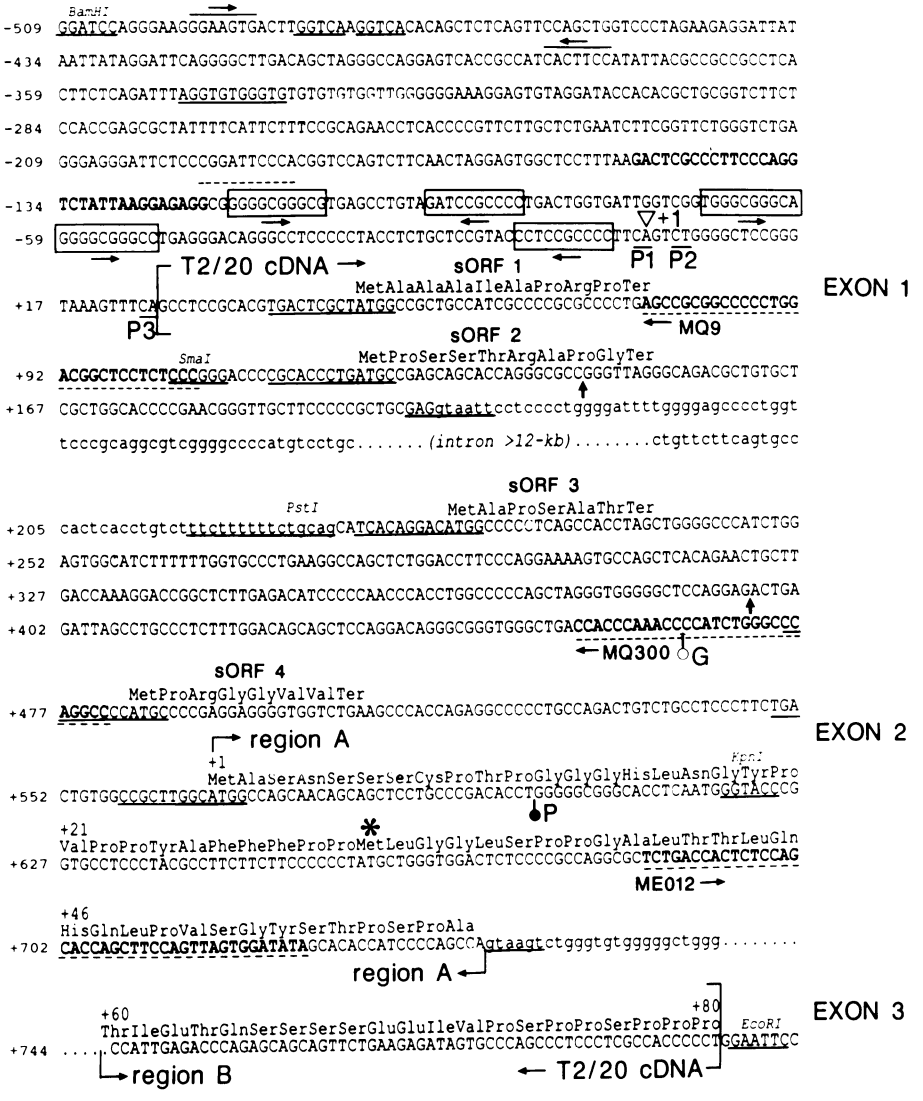

Figure 2. Nucleotide sequence of the human RAR-α promoter, exons 1 and 2 (upper case), and immediate flanking intronic regions (lower case), derived from genomic subclones pBP31.1, pRS7 and a cDNA clone T2/20. Sequence of part of exon 3, determined from T2/20, is also shown, starting at the polylinker *Eco*RI site for this clone (underlined) and the limits of the cDNA sequence are denoted by brackets. Other restriction sites, shown in Fig. 1A, are underlined. ↓P and ↓G represent the 5′ limits of hRAR-α cDNA sequences published previously by ourselves (19) and others (20). The nucleotide sequence is numbered relative to +1, the most 5′ of three transcription start-sites, P1-3, determined by nuclease S1 mapping (36). The sequences and orientation of oligonucleotides MQ9, MQ300 and MEO12, used as probes or for generating probes for nuclease S1 mapping analysis are denoted by a broken underline. The amino acid sequence of region A and part of region B of hRAR-α is shown and numbered relative to the first ATG. The 'Kozak' sequence (53) surrounding this ATG is underlined. A termination codon TGA upstream of and in frame with the ATG is underlined. Four short open reading frames (sORF), shown with their predicted amino acid sequences, are numbered according to proximity to the 5′ end, and their possible Kozak sequences are underlined. Possible binding sites for the transcription factor Sp1, based on a consensus 5′-$^{G}_{T}$GGGCGG$^{GGC}_{AAT}$-3′ (40), are boxed and their orientation indicated by arrows. A sequence 5′-GCGGGGGCGG-3′, identical to a binding site for *Krox*-20 (41) at −120/−111, is indicated by a broken overline. The sequence 5′-AGGTGTGGGTG-3′ at −347/−337, which resembles the SV40 enhancer GT-I motif (43), is denoted by a solid underline. Two copies of a sequence 5′-GGTCA-3′ beginning at −485 and −479 are also underlined. Finally, two sequences resembling elements which, in the Adenovirus E4 promoter, bind a factor E4TF1 (44) are shown by solid overlines, orientation denoted by arrows. Other symbols are described in the text.

in the cDNA clone T2/20, is separated from exon 2 by an intron which, by mapping and Southern blotting, is at least 12 kb in length (data not shown), and is preceded by a sequence of 535 bp (Fig. 2).

Identification of the hRAR-α promoter

The 5′ boundary of exon 1 contained within the genomic subclone pRS7 was determined by nuclease S1 mapping and primer extension analysis. A 614 nt-long [^{32}P]-end-labelled S1 nuclease probe, extended from the oligonucleotide MQ9 that is complementary to a sequence in exon 1 (Fig. 2), protected a major fragment of ≃ 80 nt length following hybridisation to MCF-7 or T47D polyA+ mRNA and nuclease S1 digestion. Primer extension from the oligonucleotide MQ9 hybridised to MCF-7 or T47-D polyA+ mRNA generated an extended fragment of approximately the same size (data not shown). These results suggested to us that the cDNA clone T2/20 contained the entire hRAR-α 5′-UTR.

To address the possibility that the genomic fragment contained in pRS7 might in fact contain the hRAR-α promoter, we transfected HeLa cells with constructs in which the 5′ *Bam*HI-*Sma*I fragment of the genomic subclone pRS7 (see Fig. 1a and

2), was inserted upstream of rabbit β-globin coding sequences in the vector pAL4 (see Fig. 1B). This fragment was cloned into pAL4 in either the natural or reverse orientations, giving the constructs pAL4BS and pAL4SB, respectively (Fig. 1B). Cytoplasmic RNA, prepared from HeLa cells transfected with either of these constructs or the parental vector, was hybridised to a labelled single-stranded DNA probe spanning the *Bam*HI-*Sma*I fragment and part of exon 1 of the β-globin gene (Fig. 1B; M6 probe), and subjected to nuclease S1 mapping. The results, normalised with respect to a co-transfected reference β-globin-containing plasmid, pG1B (34), are shown in Fig. 3. Three major start-sites of transcription, designated P1, P2 and P3, with P1 being the most 5′ of the three, were detected in pAL4BS-transfections, giving nuclease S1-protected fragments of 174, 170 and 149 nt in length, respectively (Fig. 3, lane 1). Conversely in pAL4SB, where the genomic *Bam*HI-*Sma*I fragment was inserted in the reverse orientation, there was no detectable transcription (same figure, lane 2), nor was transcription detected with the parental vector pAL4 (lane 3). The co-transfected control plasmid pG1B was expressed in all cases, as shown by the presence of the expected 69 nt protected fragment (see Fig. 3, lanes 1 to 3). pG1B contains one copy of the 72 bp repeat from the SV40 enhancer linked immediately upstream to the β-globin promoter, which accounts for the high level of expression seen with 400 ng of pG1B transfected when compared with 10 μg of the hRAR-α promoter construct co-transfected, and supports our previous observations that hRAR-α RNA is expressed at a low level (19). Identical results were obtained when Cos-1 cells were transfected with the same constructs (data not shown). Taken together, these data indicate that the *Bam*HI-*Sma*I fragment contains a promoter region that functions in HeLa and Cos-1 cells, when *cis*-linked to a β-globin reporter gene.

By comparison to a DNA sequence ladder, all three transcription start-sites P1, P2 and P3 (Fig. 2, bold underlines) map in the vicinity of the 5′ end of the cDNA clone T2/20 as indicated in Fig. 2. hRAR-α exon 1 is therefore ≃ 204 bp long (Fig. 1A and Fig. 2). Note that P3 is located immediately 5′ to the beginning of the T2/20 cDNA clone in agreement with the results of nuclease S1 mapping and primer extension analyses for the endogenous hRAR-α gene and its transcript (see above, data not shown), suggesting that this clone contains full 5′ cDNA sequences for a transcript that initiated at P3.

Inspection of the hRAR-α promoter sequence reveals that it lacks obvious TATA or CCAAT box sequences upstream of the P1-P3 transcriptional start-sites (Fig. 2). A sequence TTC*A*GT-CT located around start-site P1 resembles a consensus sequence CTC*A*NTCT (where N equals any nucleotide, presumptive first nucleotide of the transcript underlined), recently proposed as a core initiator sequence for promoters transcribed by RNA polymerase B (II) which lack a TATA-box (32, 33). An equivalent sequence TTC*A*GCCT surrounding P3 also resembles the consensus, but P2 (GTCTGGGG) shows little similarity. It is possible that the 'P2' protected fragment is, in fact, a clipped form of the 'P1' fragment, resulting from over-digestion by S1 nuclease. The hRAR-α first exon and 5′ flanking sequences are extremely GC-rich and there are a number of sequences which resemble binding sites for the transcription factor Sp1, variants upon the consensus recognition sequence 5′-$^{G}_{T}$GGGCGG$^{GGC}_{AAT}$-3′ (40). Five such sites, three in one orientation, two in the other, are located from −117 to −4, and show 8 or 9 out of 10 matches with the consensus (see Fig. 2; boxed sequences, orientation denoted by arrows). The most distal of the Sp1-like sequence overlaps with a sequence 5′-GCGGGGGCGG-3′ at −120 to −111, denoted in Fig. 2 by a dotted overline. This sequence matches exactly a binding site reported recently for the transcription factors *Krox*-20 and *Krox*-24 (41, 42). *Krox*-20 protein was shown to bind *in vitro* to such a sequence located in the promoter of the mouse homeogene *Hox*-1.4. The site to which *Krox*-20 bound also overlapped with a sequence that was demonstrated to bind Sp1 (41, 42).

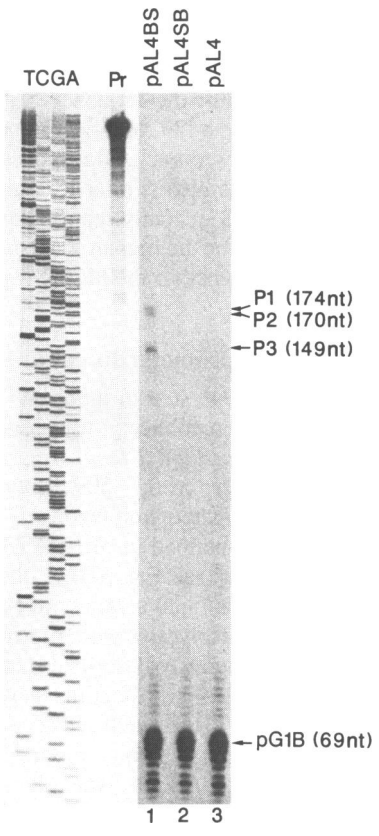

Figure 3. Determination of hRAR-α transcription start-sites by nuclease S1 mapping. A *Bam*HI-*Sma*I fragment from the genomic sub-clone pRS7 containing the promoter region for human RAR-α was cloned upstream of rabbit β-globin sequences in the expression vector pAL4. HeLa cells were transfected with recombinant plasmids (Fig. 1B) containing either the hRAR-α promoter fragment inserted in the natural (pAL4BS, lane 1) or reverse orientation (pAL4SB, lane 2), or the promoterless parent vector, pAL4 (lane 3). An internal reference β-globin plasmid, pG1B (34) was co-transfected in all cases to normalise for expression from the pAL constructs. Cytoplasmic RNA was prepared subsequently and analysed by nuclease S1 mapping (see Materials and Methods). Gels were exposed to Kodak-AR film at −80°C for up to 10 days to visualise the hRAR-α-specific protected fragments. Transcripts identified by nuclease S1 mapping that originate within the hRAR-α promoter are labelled by arrows (P1, P2 and P3, lane 1) and their sizes shown. An S1 nuclease protected fragment of 69 nt from pG1B seen in all samples is indicated. A dideoxy sequencing ladder, primed with the β-globin-specific oligonucleotide M6 used to generate the S1 probe, is shown alongside. Pr, undigested probe.

The 5′ flanking region of the hRAR-α gene contains a number of other short sequences bearing homologies to DNA elements thought or shown, in particular instances, to bind *trans*-acting protein factors. A sequence 5′-AGGTGTGGGTG-3′ beginning at −347 (underlined, Fig. 2), resembles the GT-I motif of the SV40 enhancer, known to bind the HeLa cell enhancer factor TEF-2 (43). In addition, there are two copies of a sequence 5′-GGAAGTG-3′, (overlined in Fig. 2, orientation denoted by arrows), beginning at positions −496 on the sense strand, and at position −378 on the anti-sense strand, very similar (6 matches out of 7) to the recognition sequence for the protein factor *E4TF1*

that binds to the Adenovirus E4 promoter (44). We note also that there are two copies of a sequence 5′-ATTATA-3′ arranged in tandem in a short AT-rich region from −439 to −423.

Interestingly, two copies of the sequence 5′-GGTCA-3′ are seen at positions −485 and −479, identical to half-palindromic oestrogen/thyroid hormone/retinoic acid response elements (45). This tandem arrangement is also similar to a tandem repeat of the sequence 5′-GTTCAC-3′ recently reported as a retinoic acid response element (RARE) for the human RAR-β promoter (46). We therefore investigated whether the hRAR-α promoter might be hormone-responsive.

Activity of the hRAR-α promoter in cultured cells

The hRAR-α promoter was tested in a variety of cell lines by linking it to the chloramphenicol acetyl transferase (CAT) gene from *E.coli*. The 614 bp *Bam*HI-*Sma*I fragment of pRS7, containing hRAR-α sequences from −509 through to +105 (see Fig. 1 and Fig. 2), was ligated immmediately upstream of a promoterless CAT gene contained in pBLCAT3+ (37), giving the recombinant pRARA12 (see Fig. 1B). Either pBLCAT3+ or pRARA12 were transfected into a variety of cultured cell lines and, after 48 hours, crude cell lysates were prepared and assayed for CAT activity using previously described methods (19). Fig. 4A and B shows the result of transfecting increasing amounts of either CAT construct into HeLa cells. In accordance with our results obtained using the β-globin reporter plasmid, the promoterless vector pBLCAT3+ was inactive, whereas pRARA12 CAT expression was clearly driven by the −509/+105 region of the hRAR-α promoter. When the amount of pRARA12 DNA transfected per dish was increased from 1 to 10 μg (Fig. 4A, lanes 1 through 4), there was a corresponding increase in the amount of CAT activity observed; parallel transfections with the parental vector yielded no visible CAT activity (lanes 5 to 8). CAT activity reached a plateau level at ≃ 5 μg of pRARA12 transfected (Fig. 4A and B). Accordingly, between 2−4 μg of either pRARA12 or pBL3CAT+ were used thereafter.

We tested whether RA, oestrogen or factors present in whole serum might regulate the activity of the hRAR-α promoter in HeLa cells. Treatment of HeLa cells in which the human oestrogen receptor (hER) expression vector HEO (47) was co-transfected with either pRARA12 or pBLCAT3+ and then cultured with hormone-stripped medium supplemented with 20 mM oestradiol (E2) did not alter CAT expression from pRARA12 (data not shown). HeLa cells were also co-transfected with 2 μg of pRARA12 and 2 μg of either the wild-type hRAR-α expression vector RAR-αθ or the wild-type hRAR-β expression vector RAR-βθ (21). The transfected cells were subsequently treated with 10^{-7} M RA or vehicle (ethanol) (see Materials and Methods). Fig. 4C represents the results of a typical experiment. The slight variability between the signals seen for RA versus vehicle-treated cells (for example, compare lanes 9 and 10) was not significant, as judged from several independent experiments. As a RA positive control, the plasmid RAR-β.ER.CAS (21), expressing a chimaeric receptor in which the putative DNA binding domain of RAR-β (domain C) has been replaced by the equivalent region from the hER, was used to *trans*-activate the ER-responsive plasmid vit-tk-CAT in the presence or absence of RA (Fig. 4C, compare lanes 13 and 14). HeLa cells were also treated with 15% whole (i.e. not stripped) serum for 24 hours following transfection. This also had no effect upon the level of CAT activity in pRARA12 transfections (data not shown).

We transfected a variety of other cell types with pRARA12

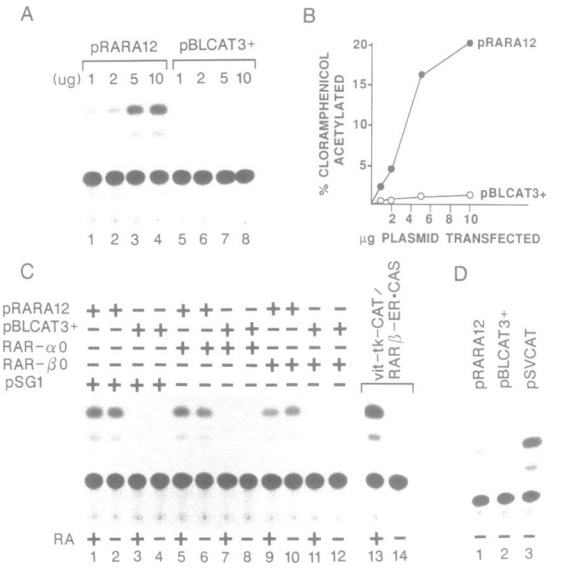

Figure 4. Activity of hRAR-α promoter-driven CAT reporter genes in cultured cells. (A) HeLa cells were transfected with the indicated amount of either the recombinant pRARA12 (Fig. 1B) (lanes 1−4) containing the hRAR-α promoter linked to the *E.coli* CAT gene or the promoterless parent vector pBLCAT3+ (37) (lanes 5−8), 2 μg of the β-galactosidase indicator plasmid pCH110, and carrier DNA (BSM+) to a final level of 20 μg. Transfection, cell culture and assays for CAT activity were performed as detailed in Materials and Methods. Transfection efficiency was normalized by measuring the amount of β-gal activity in 1/5 of the extract prepared from one 90 mm dish of transfected cells (see Materials and Methods) ; the equivalent of 30 units of β-gal activity for each extract were assayed for CAT activity. (B) Graphic representation of the data from (A). The percentage of chloramphenicol acetylated, calculated by cutting out the regions corresponding to acetylated and non-acetylated forms of ^{14}C-chloramphenicol and scintillation counting, is plotted against the amount (in μg) of pRARA12 (solid circles) or pBLCAT3+ (open circles) transfected. (C) HeLa cells were transfected with either 2 μg of pRARA12 (lanes 1 and 2, 5 and 6, 9 and 10) or 2 μg pBLCAT3+ (lanes 3 and 4, 7 and 8, 11 and 12), together with either the wild-type RAR-α expression vector RAR-αθ (21) et al., 1988) (lanes 5−8), the wild-type RAR-β expression vector RAR-βθ (21) (lanes 9−12) or the parental expression vector pSG1 (a gift of Stephen Green) (lanes 1−4). The β-gal expression vector pCH110 and BSM was co-transfected as described in Brand et al (21), and the equivalent of 50 β-gal units were assayed in each case. As a RA *trans*-activation control, HeLa cells were co-transfected with 100 ng of RAR-β-ER.CAS (21) and 1 μg of the reporter *vit-tk*-CAT (lanes 13 and 14); 10 β-gal units from each transfection were assayed. Transfections were done in duplicate and one of each pair was treated with 10^{-7} M RA (odd numbers) or vehicle alone (even numbers), as described in Materials and Methods. The results shown represent a typical experiment. (D) Expression from the hRAR-α promoter in the mouse EC cell line P19.6. 4 μg of either pRARA12 (lane 1) or pBLCAT3+ (lane 2) were transfected with pCH110 and BSM to 20 μg. 1 μg of the SV40-driven CAT reporter pSVCAT was also transfected (lane 3). 50 β-gal units were assayed in each case.

or pBLCAT3+. Some murine embryonal carcinoma (EC) cells such as the F9 and P19 lines are responsive to RA, undergoing terminal differentiation when treated with nano- or micromolar concentrations of RA (for example, see 1, 9, 48). Weak expression of pRARA12 was seen in transfected P19.6 cells, with no visible activity from the promoterless pBLCAT3+ (Fig. 4D, compare lanes 1 and 2). The SV40-driven reporter pSV-CAT was also transfected and was strongly expressed in P19.6 cells (Fig. 4D, lane 3). pSV-CAT was also expressed strongly in the human breast cancer cell lines MCF-7 and T47D, but no CAT activity was observed when these lines were transfected with pRARA12 (data not shown). This was a surprising result given that hRAR-α is expressed in MCF-7 and T47D cell lines (19). The hRAR-α promoter may be positively regulated by *trans*-

acting factors present in T47D and MCF-7 cells that bind to some far upstream element not represented in the pRARA12 clone.

DISCUSSION

We report here the isolation of a cDNA clone which extends much further in the 5′ direction the sequence of the hRAR-α mRNA which has been previously reported (19, 20) and we demonstrate that region A and the 5′-UTR are encoded in two exons. This cDNA clone, T2/20, is co-linear with the hRAR-α cDNA sequences reported by us (19) and by Giguère et al. (20), but extends it by 573 bp and 438 bp, respectively. Previously, we have isolated a cDNA clone which was predicted to code for a 432 amino acid protein, starting from the methionine codon marked by an asterisk in Fig. 2, and which diverged from T2/20 (our unpublished data) at the point denoted as ↓P in Fig. 2, which does not correspond to an exon-intron boundary. It is likely that this clone corresponded to an artifact generated during the construction of the cDNA library, since the 252 bp sequence present upstream of ↓P was found to be the sequence located between the two vertical arrows indicated in Fig. 2 (positions 145−396), but in the reverse orientation. We have no explanation for this 'specific' artefact. We note, however, that analysis of hRAR-α 5′-UTR, using the RNA secondary structure prediction program of Zuker (49, 50) indicates that this region in the mRNA may fold into at least two complex hairpin structures with $\Delta G \leq -100$ Kcal (data not shown).

In addition to potential stable secondary structures, we note that the hRAR-α 5′-UTR contains also several small open reading frames (sORF1 to 4, Fig. 2). It has been proposed that the formation of secondary structures and the presence of small open reading frames in the 5′-UTR of mRNA is involved in the control of efficiency of translation. For example, the GC-rich 5′-UTR of the transcript for human platelet-derived growth factor 2, related to the viral *v*-sis oncogene (c-*sis*/PDGF-2), forms secondary structures and impedes translation *in vivo* (51). Similarly, the 5′-UTR of poliovirus RNA contains stable secondary structures (52). In yeast genes, such as GCN4, sORFs in the 5′-UTR actually regulate translation, but this has not yet been demonstrated in vertebrates (for review and references, see 53). Interestingly, the 5′-UTR of RAR-α from mouse (mRAR-α) is highly similar to that of hRAR-α, exhibiting ≃ 80% homology at the DNA level, and extensive conservation has also been found in the 5′-UTRs of human and mouse RAR-β as well as RAR-γ (unpublished results from our laboratory). This suggests that the RAR 5′-UTR plays some role in regulating RAR expression at the translational level that has been conserved in evolution.

Comparison of the sequences of hRAR-α genomic DNA clones with that of the T2/20 cDNA clone shows that region A of the receptor and 363 bp of the adjacent 5′-UTR are encoded in one exon (Fig. 2, exon 2). The 3′ splice donor site for this exon, predicted by comparison of genomic and cDNA clones, splits codon Thr_{60} of hRAR-α, at a position noted previously by us and others to constitute a possible exon boundary (19, 21, 39), and which corresponds also to the boundary between regions A and B observed in murine and human RARs (23, 24). In both the mouse and human, alternative splicing at this junction in the RAR-γ transcript generates a number of mRNAs which may encode distinct receptor isoforms, containing common B to F region sequences, but different A regions and/or 5′-UTRs (23, 27, 31).

The remainder of the 5′-UTR of the T2/20 cDNA clone is encoded in a single exon (exon 1, Fig. 2). A 614-bp *Bam*HI-*Sma*I genomic fragment encompassing the 5′ portion of this cDNA, and about 500 bp of 5′ flanking sequences, functions as a promoter when linked to promoterless β-globin or CAT reporter plasmids (Fig. 1B), and expressed transiently in cultured cells (Figs 3 and 4). Using the β-globin reporter construct and nuclease S1 mapping, we have shown that RNA is initiated at three main sites (P1, P2 and P3); P1 being the most 5′ of the three (see Figs 2 and 3) corresponds therefore to the start of exon 1 for the hRAR-α gene (Fig. 2, +1). The hRAR-α 5′ flanking promoter region is GC-rich, and contains several potential binding sites for the transcription factor Sp1 (40). Sequences corresponding to TATA or CCAAT boxes upstream of the three transcription start-sites were not observed. However, 8-bp sequences surrounding the start-sites P1 and P3 resemble the recently described consensus initiator sequence present in some promoters that lack a TATA-box (32, 33). This contrasts with the promoter for the human RAR-β receptor that has recently been characterised (46). The hRAR-β promoter contains one major transcription start-site, preceded by a TATA sequence at −30 and a further upstream CCAAT box sequence. The hRAR-β sequences from −87 to +155 show no nucleotide sequence homology with the −509/+105 region of the hRAR-α gene.

The hRAR-α upstream promoter region contains several short sequences that are similar to known *cis*-acting enhancers or to sites shown to bind transcription factors. A sequence, 5′-GCGGGGGCGG-3′, which in the mouse homeobox gene *Hox*-1.4 binds the zinc-finger factor *Krox*-20 (41), overlaps with the most distal of the possible Sp1-binding sites (Fig. 2, dotted overline). *Krox*-20 has been implicated in early central nervous system development and is expressed in rhombomeres of the developing hindbrain (41, 54). It will be of interest to determine whether *Krox*-20 also binds to the above sequence sequence, in the context of the hRAR-α promoter, since RAR-α and *Krox*-20 have been recently shown to be expressed in the same rhombomeres of the mouse embryo (29). A direct repeat of a sequence 5′-GGTCA-3′ at −488 to −474 in the hRAR-α promoter is identical to half of a palindromic sequence that is present in responsive elements of oestrogen receptor (ER) and some responsive elements of the thyroid hormone and retinoic acid recep tors (45). However, the CAT construct pRARA12, containing the hRAR-α promoter fragment, was not responsive to oestradiol and hER, nor to RA and hRAR-α or hRAR-β, as determined by transient expression assays in HeLa cells. The sequence around the proximal repeat in the hRAR-α promoter, 5′-CAAGGTCAC-3′, exhibits a 7 out of 9 match to the sequence 5′-APuPuGTTCAC-3′ whose repetition constitutes a retinoic acid response element (RARE) in the hRAR-β promoter (46). This sequence does not function as a RARE within the hRAR-α promoter as determined by co-transfecting HeLa cells with the −509/+105 promoter-driven CAT construct pRARA12 and expression vectors for either cloned human RAR-α or β in the presence of 10^{-7} M RA (Fig. 4C). In this respect, we note that it has been reported that RA has no effect upon the expression of hRAR-α (55), nor mouse RAR-α (24, 56).

Given the pleiotropic effects of RA during development, it seems likely that the expression of receptors for RA must be finely regulated. The selective distribution of RAR-α transcripts during mouse embryogenesis (25, 26, 29), the finding that 5′-UTR sequences in RAR-α are highly conserved in evolution, together with complex alternative splicing patterns in the same region (our unpublished observations; see also 27) suggest that expression

of the RAR-α gene may be regulated both transcriptionally and post-transcriptionally.

ACKNOWLEDGEMENTS

We are indebted to Transgène SA for the gift of the human genomic library, S. Green for providing the T47D cDNA library and pSG1, J. White for pAL4, and G. Schutz and B. Luckow for pBLCAT3$^+$. We thank our colleagues of the retinoic acid receptor group for helpful discussions, particularly Ph. Kastner, A. Krust, P. Leroy and A. Zelent. We also thank A. Staub and F. Ruffenach for oligonucleotides, M. Acker and the culture staff for cells, C. Werlé and B. Boulay for preparing the figures and members of the secretariat for typing this manuscript. This work was supported by INSERM (grants CNAMTS), CNRS (AIV), the Ministère de la Recherche et de l'Enseignement Supérieur and the Fondation pour la Recherche sur le Cancer. N.J.B. was the recipient of a long-term EMBO fellowship; M.P. was supported by the Medical Research Council of Canada.

REFERENCES

1. Strickland, S. and Mahdavi, V. (1978) Cell *15*, 393–403.
2. Hogan, B.L.M., Taylor, A. and Adamson, E. (1981) Nature *291*, 235–237.
3. Roberts, A.B. and Sporn, M.B. (1984) in Sporn, M.B., Roberts, A.B. and Goodman, D.S. (eds), 'The Retinoids', Academic Press, Florida, vol. 2, pp. 209–286.
4. Maden, M. (1985) Trends in Genetics *1*, 103–107.
5. Brockes, J.P. (1989) Neuron *2*, 1285–1294.
6. Eichele, G. (1989) Trends in Genetics *5*, 246–251.
7. Brockes, J. (1990) Nature *345*, 766–768.
8. Andrews, P.W. (1984) Developmental Biology *103*, 285–293.
9. Sharma, S. and Notter, M.F.D. (1988) Developmental Biology *125*, 246–254.
10. Wang, S.-Y. and Gudas, L.J. (1983) Proc. Natl. Acad. Sci. USA *80*, 5880–5884.
11. Marotti, K.R., Brown, G.D. and Strickland, S. (1985) *108*, 26–31.
12. LaRosa, G.J. and Gudas, L.J. (1988) Mol. Cell Biol. *8*, 3906–3917.
13. Vasios, G.W., Gold, J.D., Petkovich, M., Chambon, P. and Gudas, L.J. (1989) Proc. Natl. Acad. Sci. USA 86, 9099–9103.
14. Thaller, C. and Eichele, G. (1987) Nature *327*, 625–628.
15. Slack, J.M.W. (1987) Nature *327*, 553–554.
16. Evans, R.M. (1988) Science *240*, 889–895.
17. Green, S. and Chambon, P. (1988) Trends in Genetics *4*, 309–314.
18. Beato, M. (1989) Cell *56*, 335–344.
19. Petkovich, M., Brand, N.J., Krust, A. and Chambon, P. (1987) Nature *330*, 444–450.
20. Giguère, V., Ong, E.S., Segui, P. and Evans, R.M. (1987) Nature *330*, 624–629.
21. Brand, N., Petkovich, M., Krust, A., Chambon, P., de Thé, H., Marchio, A., Tiollais, P. and Dejean, A. (1988) Nature *332*, 850–854.
22. Benbrook, D., Lernhardt, E. and Pfahl, M. (1988) Nature *333*, 669–672.
23. Krust, A., Kastner, P., Petkovich, M., Zelent, A. and Chambon, P. (1989) Proc. Natl. Acad. Sci. USA *86*, 5310–5314.
24. Zelent, A., Krust, A., Petkovich, M., Kastner, P. and Chambon, P. (1989) Nature *339*, 714–717.
25. Dollé, P., Ruberte, E., Kastner, P., Petkovich, M., Stoner, C.M., Gudas, L.J. and Chambon, P. (1989) Nature *342*, 702–705.
26. Dollé, P., Ruberte, E., Leroy, P., Morriss-Kay, G. and Chambon, P. (1990) Development, in press.
27. 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.
28. Ruberte, E., Dollé, P., Krust, A., Zelent, A., Morriss-Kay, G. and Chambon, P. (1990) Development *108*, 213–222.
29. Ruberte, E., Dollé, P., Chambon, P. and Morriss-Kay, G. (1990) Development, in press.
30. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M. and Chambon, P. (1986) EMBO J. *5*, 891–897.
31. Giguère V., Shago, M., Zirngibl, R., Tate, P., Rossant, J. and Varmuza, S. (1990) Mol. Cell. Biol. *10*, 2335–2340.
32. Smale, S.T. and Baltimore, D. (1989) Cell *57*, 103–113.
33. Smale, S.T., Schmidt, M.C., Berk, A.J. and Baltimore, D. (1990) Proc. Natl. Acad. Sci. USA *87*, 4509–4513.
34. Wasylyk, C. and Wasylyk, B. (1986) EMBO J. *5*, 553–560.
35. Gough, N.M. (1988) Anal. Biochem. *173*, 93–95.
36. Zenke, M. Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) EMBO J. *5*, 387–397.
37. Luckow, B. and Schütz, G. (1987) Nucleic Acids Res. *15*, 5490.
38. Shapiro, M.B. and Senapathy, P. (1987) Nucleic Acids Res. *15*, 7155–7174.
39. Dejean, A., Bougueleret, L., Grzeschik, K-H. and Tiollais, P. (1986) Nature *322*, 70–72.
40. Dynan, W.S., Sazer, S., Tjian, R. and Schimke, R.T. (1986) Nature *319*, 246–248.
41. Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Duboule, D. and Charnay, P. (1990) EMBO J. *9*, 1209–1218.
42. Lemaire, P., Vesque, C., Schmitt, J;, Stunnenberg, H., Frank, R. and Charnay, P. (1990) Mol. Cell. Biol. *10*, 3456–3467.
43. Davidson, I., Xiao, J.H., Rosales, R., Staub, A. and Chambon, P. (1988) Cell *54*, 931–942.
44. Jones, N.C., Rigby, P.W.J. and Ziff, E.B. (1988) Genes & Development *2*, 267–281.
45. Umesono, K., Giguere, V., Glass, C.K., Rosenfeld, M.G. and Evans, R.M. (1988) Nature *336*, 262–265.
46. de Thé, H., del Mar Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H. and Dejean, A. (1990) Nature *343*, 177–180.
47. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J.-R. and Chambon, P. (1987) Cell 51, 941–951.
48. Jones-Villeneuve, E.M.V., McBurney, M.W., Rogers, K.A. and Kalnins, V.I. (1982) J. Cell Biol. *94*, 253–262.
49. Zuker, M. and Stiegler, P. (1981) Nucleic Acids Res. *9*, 133–148.
50. Jacobson, A.B., Good, L., Simonetti, J. and Zuker, M. (1984) Nucleic Acids Res. *12*, 45–52.
51. Rao, C.D., Pech, M., Robbins, K.C. and Aaronson, S.A. (1988) Mol. Cell. Biol. 8, 284–292.
52. del Angel, R.M., Papavassiliou, A.G., Fernandez-Tomas, C.,Silverstein, S.J. and Racaniello, V.R. (1989) Proc. Natl. Acad. Sci. USA *86*, 8299–8303.
53. Kozak, M. (1989) J. Cell Biol. *108*, 229–241.
54. Wilkinson, D.G., Bhatt, S., Chavrier, P., Bravo, R. and Charnay, P. (1989) Nature *337*, 461–464.
55. de Thé, H., Marchio, A., Tiollais, P. and Dejean, A. (1989) EMBO J. *8*, 429–433.
56. Hu, L. and Gudas, L. (1990) Mol. Cell. Biol. *10*, 391–396.