Genomic organization of the human thyroid hormone receptor α (*c-erbA-1*) gene

Vincent Laudet*, Agnes Begue, Catherine Henry-Duthoit[1], Anita Joubel, Patrick Martin, Dominique Stehelin and Simon Saule
INSERM U 186/ CNRS UA 04 1160, Institut Pasteur de Lille, 59019 Lille Cedex and [1]INSERM U 124 IRCL, place de Verdun, 59045 Lille cedex, France

Received November 16, 1990; Revised and Accepted February 5, 1991 EMBL accession nos X55004, X55005,and X55066-X55074 (incl.)

ABSTRACT

The thyroid hormone receptor α (THRA or *c-erbA-1*) gene belongs to a family of genes which encode nuclear receptors for various hydrophobic ligands such as steroids, vitamin D, retinoic acid and thyroid hormones.These receptors are composed of several domains important for hormone-binding, DNA-binding, dimerization and activation of transcription. We show here that the human THRA gene is organized in 10 exons distributed along 27 kbp of genomic DNA on chromosome 17. The position of the introns in human THRA is highly conserved when compared to the chicken gene despite their differing lengths. The N-terminal A/B domain as well as the 5′ untranslated region is encoded by two exons. Interestingly, each of the putative zinc fingers of the receptor DNA-binding domain is encoded by one exon and the hormone-binding domain is assembled from three exons. The two last exons of the gene are alternatively spliced to generate two different messenger RNAs. In addition, we confirm that another gene, belonging to the nuclear receptor superfamily, *ear-1*, overlaps with the 3′ region of THRA in an opposite transcriptional orientation.

INTRODUCTION

The thyroid hormone receptor α (THRA) belongs to a family of ligand-activated transcription factors which regulate the expression of target genes by binding to specific cis-acting sequences (1-3). This family includes nuclear receptors for several hydrophobic ligands, such as glucocorticoids, estrogens, progesterone, vitamin D or retinoic acid and also for dioxin and ecdysone. In addition to these receptors for known ligands, several authors have described 'orphan receptors' which are putative receptors to ligands still to be identified (1). The total number of genes described in this growing family is near 30. The family of nuclear receptors has been classically divided in two subfamilies on the basis of structural homologies. One subfamily includes the genes encoding steroid hormone receptors and the other comprises the receptors for thyroid hormone and retinoic acid as well as several orphan receptors. To date, this subfamily contains three retinoic acid receptors, α, β and γ (4-7), two thyroid hormone receptors, α (THRA on chromosome 17q21) and β (THRB on chromosome 3p21) (8,9) and three *c-erbA* related genes named *ear-1*, *ear-2* and *ear-3* which are located on human chromosomes 17, 19 and 5 respectively (10,11). The THRA and THRB are similar in overall structure, being most related in the cysteine-rich DNA-binding and C-terminal hormone-binding regions and more divergent in the A/B domain which has been suggested to represent at least part of the transactivation domain. Nucleotide sequence comparison has revealed that the chicken THRA gene (12,13) is the cellular progenitor of the *v-erbA* oncogene found in the Avian Erythroblastosis Virus (AEV Refs : 12, 14-17). The *v-erbA* oncogene does not by itself cause tumors in animals, but enhances the transforming potential of oncogenes that induce sarcomas (18-19).

In addition to this diversity of thyroid hormone receptor (TR) genes, isolation of rat and human THR cDNA clones has revealed that a variety of TRs may be produced (8,12,20-23). Alternative splicing of the two THR genes leads to the generation of functionally different receptors (11, 24-28). In the case of the THRA gene, two receptor forms are generated which differ in their C-terminal region. One receptor (hTHRA1), is composed of 410 amino-acids, binds T3, and activates the transcription of target genes such as the growth hormone gene. The other receptor (hTHRA2), composed of 490 amino-acids, does not bind T3 and acts as a dominant negative regulator of TRs. Dominant negative regulation by hTHRA2 is exerted by suppression of the transactivation mediated by products of either the hTHRA1 or THRB genes (26,27). Recent studies have shown that the *v-erbA* product acts as a thyroid hormone independent repressor of target gene expression much like hTHRA2. The *v-erbA* product is able to block the thyroid hormone-mediated activation of a target gene when coexpressed with the normal receptor (18).

To understand the evolution of this superfamily of ligand-dependent transactivation factors and to facilitate the analysis of their transcriptional regulation, the chromosomal genes must be cloned. To date, knowledge of the complete genomic organization

* To whom correspondence should be addressed

of members of this family is restricted to members of the steroid hormone receptor subfamily (chicken progesterone [29], human estrogen [30] and human androgen receptor genes). For the thyroid hormone/retinoic acid receptor subfamily, only partial gene structures have been described for THRA and RAR β and γ genes (11, 33−36). In this paper, we show that the human THRA gene is composed of a transcriptional unit spanning 27 kbp split into 10 exons. In addition, we show that another gene, already described and termed *ear-1* or *REVerb* (11, 34) is detected in the same locus and overlaps in a tail-to-tail orientation with the 3′ region of THRA.

MATERIALS AND METHODS

Molecular Probes

Different fragments were derived from either *v-erbA* and THR cDNAs or genomic DNA and used as probes for cloning different parts of the locus:

The *v-erbA* probes used were *(i)* a 1.4 kbp HincII fragment representing the major part of the *v-erbA* gene and a part of the *v-erbB* gene and *(ii)* a 0.5 kbp SalI-SacI 3′ fragment of *v-erbA*. These two probes were isolated from the AEV provirus cloned in pKH47 (37).

The cDNA and genomic probes were *(i)* the entire 14C THRA cDNA *(ii)* an EcoRI-BstEII fragment of 14C containing the first 550 nucleotides of this cDNA, *(iii)* a 0.9 kbp EcoRI-KpnI fragment representing a large portion of exon 9 from λE2A2, *(iv)* a 1 kbp XhoI-KpnI fragment of the genomic clone λE2A3 containing parts of exons 8 and 9.

Library screening

The 14C and 14J cDNAs were isolated from a K562 cell cDNA library in λgt10 (provided by E.Canaani) using the 3′ *v-erbA* SalI-SacI probe and the 0.9 kbp EcoRI-KpnI genomic probe respectively. The 14E cDNA was isolated from a fetal liver cDNA library cloned in Charon 21, provided by J.J.Toole and screened with the 550 bp EcoRI-BstEII probe representing the 5′ part of 14C.

We have isolated four overlapping clones named λE2A3, λE2A2, λE2A1 and λE2A0 from three human genomic libraries. The λE2A3 clone was isolated from a human placental DNA library provided by Tom Maniatis (38) . This library was screened with the 1.4 kbp *v-erbA* probe. Because we were unable to isolate a phage containing the 5′ part of the THRA gene in this library, we decided to construct a new genomic DNA library using ML3 cells DNA partially digested with SauIIIA and inserted into the EMBL4 vector. From this library the λE2A2 and λE2A1 clones were isolated. The λE2A2 was obtained after screening with the entire 14C probe. The λE2A1 clone was isolated using the 5′ EcoRI-BstEII probe derived from the 14C cDNA. Finally, the λE2A0 clone was derived from DNA of ovarian carcinoma IGROV-2 cells, partially digested with Sau3A and inserted into the EMBL4 vector. The clone was isolated by screening with the 5′ EcoRI-BstEII probe. Cloning, phage and plasmid growth or isolation were performed using standard procedures (39).

Sequencing experiments

The different cDNA or genomic fragments were subcloned into the M13mp18 or M13mp19 sequencing vectors using standard techniques and an Applied Biosystem 370A automatic sequencer. Exon 8 and the 5′ half of 14C were sequenced using the base-specific chemical method described by Maxam and Gilbert (40).

PCR amplification on messenger RNA

Messenger RNA from MRC5 cells was extracted using guanidinium thiocyanate according to standard methods (39). RNA was retrotranscribed and amplified by PCR using the Perkin Elmer Cetus GeneAmp RNA PCR kit and the classical method of Wang et al.(41). Briefly, 500 ng of polyA+ RNAs were incubated for 15 min at 42°C in the buffer furnished by the supplier with 5mM $MgCl_2$, 1mM of each dNTP, 50 pmol of oligonucleotide A, 1 unit of RNAsin and 2.5 units of the reverse transcriptase from Perkin Elmer Cetus in a volume of 20 µl. The mixture was then denatured at 99°C for 5 min and ice-cooled for 5 min.The PCR amplification was then conducted in the same buffer at a final concentration of 2 mM $MgCl_2$ with 50pmol of primer, 2.5 units of TAQ polymerase. The PCR reaction was conducted in a Perkin Elmer Cetus DNA Thermal Cycler with the following 35 cycles : 1 min at 95°C, 90 sec at 60°C. During the last cycle, a synthesis step at 72°C was performed for 7 min. The samples were analyzed on agarose gels stained with ethidium bromide. Nested PCR was conducted using the same protocol and reagents on 2 µl of the primary amplification mixture as a target.

The oligonucleotides used were synthesised by Institut Pasteur de Lille Oligonucleotide Facility and were as follows:
(A) GTG GCC AGG ACC CTG TTG TG which corresponds to the beginning of 14J cDNA, (B) AAA AAC TGC CCA TGT TCT CCG which corresponds to the end of exon 7, (C) CCA GCT GAA TTC ACC CTC TGG CCG CCT GAG GCT which corresponds to the end of the coding part of 14C in exon 9 and (D) GCT CTG CTG CAG GCT GTG CTG CTA ATG T which was used for nested PCR and which corresponds to the end of exon 8. The sequences of the oligonucleotides are underlined in Fig.3

RESULTS

Cloning of THRA cDNAs

A K562 cell cDNA library was screened using a ^{32}P-labeled 0.5 kbp SalI-SacI probe which corresponds to the 3′ portion of *v-erbA*. Two positive clones were obtained from 4×10^6 plaques and designated 14C and 14E. The 14C clone was further characterized by restriction endonuclease mapping and sequence analysis. The restriction map is shown in Fig.1A. The 2288 bp

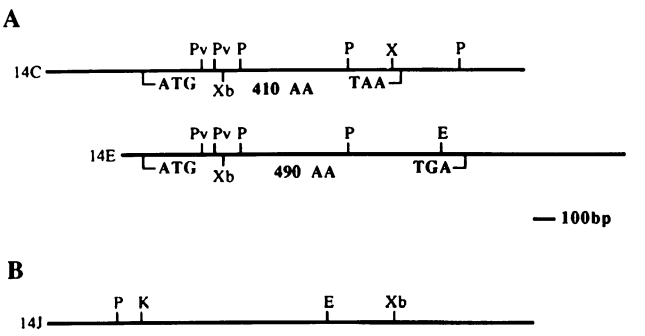


Fig. 1. Map of the three cDNA clones used in this study. (A) 14C and 14E clones have two thirds of their 5′ regions in common, as revealed by endonuclease mapping. (B) Map of the 14J cDNA. ATG : Initiation codon. TGA and TAA: Stop codon. Enzymes used : Pv, Pvu2; P, PstI ; X,XhoI ; Xb, XbaI ; E,EcoR1; K, KpnI.

sequence of 14C contains a long open reading frame encoding a protein of 410 amino acids with the ATG codon at position 467 and the TAG stop codon at position 1697. Comparison of the nucleotide sequence with various THR sequences shows that 14C is identical to the human thyroid hormone receptor hTHRA1 already described (25), although 14C is extended 394 nucleotides further in the 5′ direction.

Since the messenger RNA corresponding to hTHRA1 was described as a 5.0 kb species (11), we cloned the missing sequence using genomic probes from the THRA locus to screen the K562 cDNA library. Using a 0.9 kbp EcoRI-KpnI fragment of clone E2A2 (see materials and methods) a cDNA was obtained after screening 4×10^6 plaques. We isolated two identical clones which hybridized with the probe, representing a cDNA named 14J. This cDNA does not overlap with 14C and corresponds to the 3′ region of the 5.0 kb RNA species, as revealed by northern blot hybridization (data not shown) and PCR experiments (see later). The 14J clone was extensively characterized by restriction endonuclease mapping (Fig.1B) and DNA sequencing (see the sequence of exon 9 in Fig.3). Sequence analysis and comparison with the EMBL data bank of the 2381 bp cDNA did not reveal any homology with previously identified sequences.

By screening a human fetal liver cDNA library, we have characterized a third type of cDNA termed 14E, which was analyzed by extensive restriction enzyme digestions (Fig. 1A) and partial sequencing (not shown). This cDNA appears to be identical to the human testis cDNA (20) which encodes a product of 490 amino acids (hTHRA2). The 5′ two-thirds of 14E are identical to the 5′ region of 14C, but 14E diverges from 14C at position 1448 of 14C, where an alternative splice donor site was observed (11, 26 and Fig.3).

Characterization of THRA locus

Four phages corresponding to the various cDNAs and named λE2A3, λE2A2, λE2A1 and λE2A0 were isolated from genomic libraries. The alignment of these four individual clones with THRA cDNAs is shown in Fig.2. We isolated subclones of the inserts corresponding to the different exons of THRA. Sequencing of these subclones demonstrated that the human THRA locus is split into 10 exons distributed over 27 kbp of genomic DNA (Fig.3). The exons differ greatly in size (169, 350, 68, 101, 148, 206, 147, 259, 3357, 841 bp) as do the introns (more than 10 kbp to 623 bp). Comparison of the intron/exon boundaries (Fig.3) shows that they all match a canonical splice consensus sequence (the alternative splice donor site of exon 9 included) (42).

Comparison of the 3′ region of the cDNA clones revealed several interesting features. First, the genomic sequence corresponding to the 3′ region of 14C includes an A-rich region (position 821−865 in exon 9, Fig.3) which probably served as an internal priming site during the generation of the poly-dT-primed cDNA library. In addition, there are no polyadenylation consensus signals evident in 14C suggesting that this cDNA represents a species incomplete in the 3′ region (see also ref. 25). Second, comparison of the 3′ region of 14E and 14J reveals two different types of putative polyadenylation sites (CATAAAA for 14E and ATTTTAAA or ATATTTTTA for 14J ; see arrowheads in Fig3).

Comparison of the genomic organization of THRA with the various cDNAs shows that although 14C and 14J cDNAs are not overlapping clones, they correspond to a messenger RNA of 5.0 kb yet to be described. Indeed, northern blot experiments show that both 14J and 14C specific probes hybridize with a 5.0

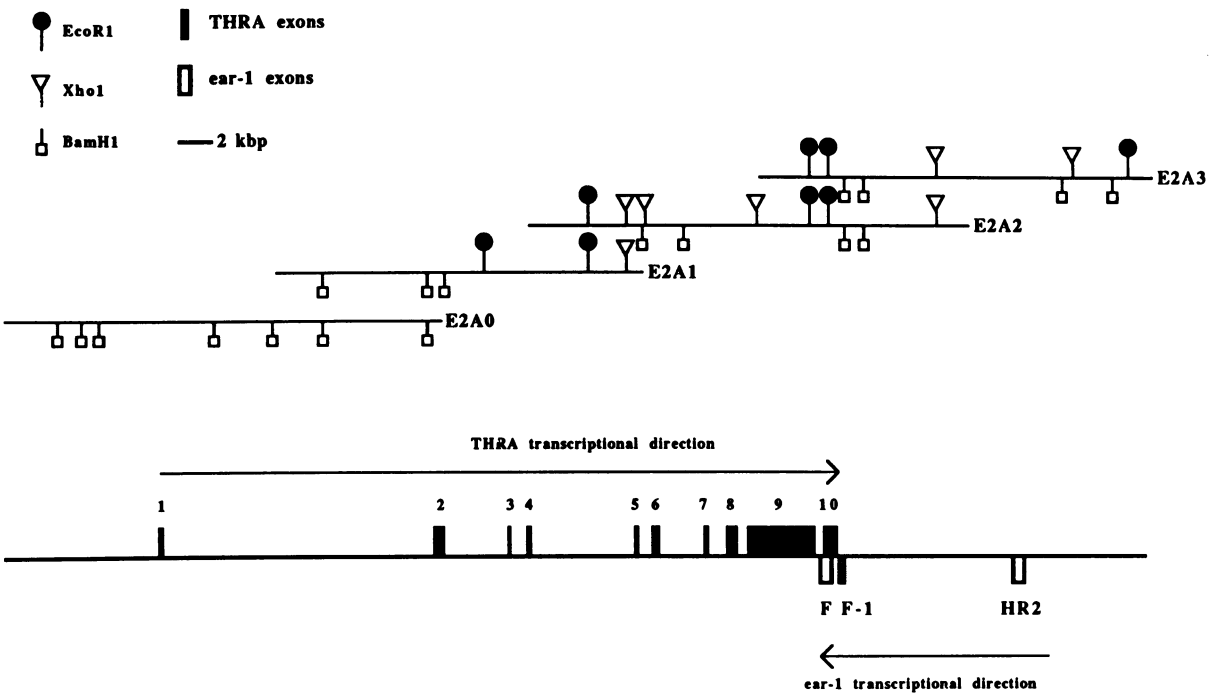
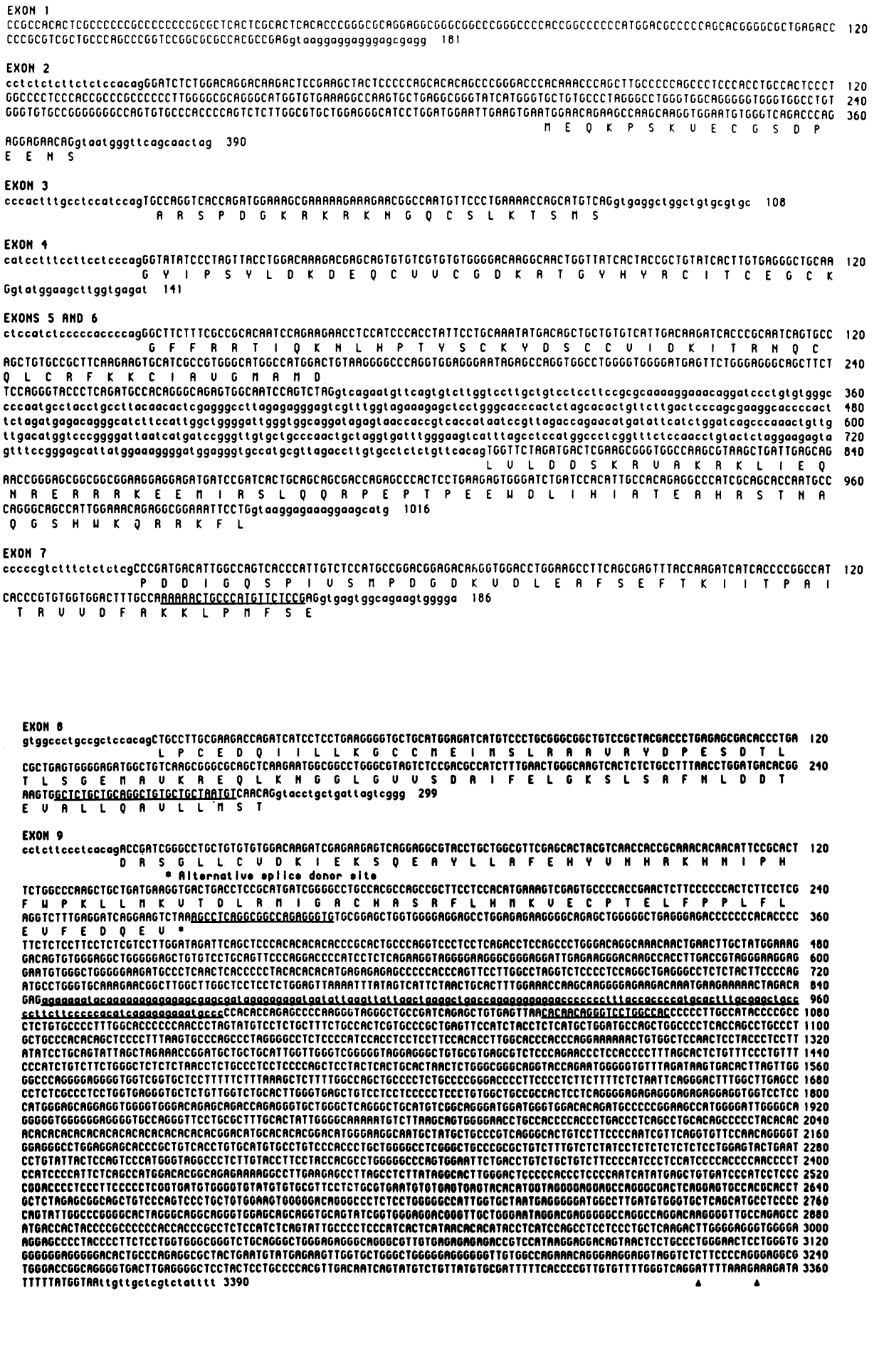


Fig. 2. Genomic organization of THRA. In the upper part of the figure, the 4 genomic inserts are aligned with respect to THRA. The exon/intron distribution is shown below. Black boxes : THRA exons numbered from 1 to 10 ; open boxes : *ear-1* exons denoted F for final, F-1 for adjacent to final and HR2 for homologue of rat exon 2. The arrows give the transcriptional orientation of the two genes. The initiation codon of THRA is in exon 2, the two alternative stop codons in exon 9 and 10. The ATG of *ear-1* is in a upstream exon of RH2 and the stop codon is in exon F.


EXON 1
EXON 2
EXON 3
EXON 4
EXONS 5 AND 6
EXON 7
EXON 8
EXON 9
* Alternative splice donor site

```
EXON 10      * End of ear-1 exon F
acacaccacagaagCCAGCTCAGCTGTGAACTATTGGATTTGAGACAGGAACAGAACAAATCAGAGGGCCAGGGGAGGGTTGTGGGGGAGACAGAGTGGTTTAAATAGGGGAGGAGGGGA  120
AGTTCGGTGATGGGGGAGGGAGGCAGGTATTTACAAGAAGGCTCAGGGGGCCAGAGGCTCATCTTGGAATATTTTATAACAATATAAATAAGATTCTGGTTTGCTTTTCCTTTTCGTCTC  240
GTAAAGGAGAGAGAAGTGCAGAGTTCGATTCTGTACAAGGGGGCAGCGGCAGAAGGCCGGCCGGGCGGGTCACTGGGCGTCCACCCGGAAGGACAGCAGCTTCTCGGAATGCATGTTGTT  360
      E  R  E  V  Q  S  S  I  L  Y  K  G  A  A  A  E  G  R  P  G  G  S  L  G  V  H  P  E  G  Q  Q  L  L  G  M  H  V  V
CAGGGTCCGCAGGTCCGGCAGCTTGAGCAGCAGCTTGGTGAAGCGGGAAGTCTCCAAGGGCCGGTTCTTCAGCACCAGAGCCCGAAGAGCCCGCAGCAGCGTCTCCTGGAGCTGCTCCAC  480
 Q  G  P  Q  V  R  Q  L  E  Q  Q  L  G  E  A  G  S  L  Q  G  P  V  L  Q  H  Q  S  P  K  S  P  Q  Q  R  L  L  E  L  L  H
    Start of ear-1 exon F *
CGAAGCGGAATTCTCCATGCCCGAGCGGTCTGTGGGGAAGACGACAGCAGTGAGGCGGACTCCCCGAGCTCCTCTGAGGAGGAACCGGAGGTCTGCGAGGACCTGGCAGGCAATGCAGCC  600
 R  S  G  I  L  H  A  R  A  V  C  G  E  D  D  S  S  E  A  D  S  P  S  S  S  E  E  E  P  E  V  C  E  D  L  A  G  N  A  A
TCTCCCTGAAGCCCCCCAGAAGGCCGATGGGGAAGGAGAAGGAGTGCCATACCTTCTCCCAGGCCTCTGCCCCAAGAGCAGGAGGTGCCTGAAAGCTGGGAGCGTGGGCTCAGCAGGGCT  720
 S  P  *
GGTCACCTCCCATCCCGTAAGAGGACCTTCCCTTCCTCAGCAGGCCAAACATGGCCAGACTCCCTTGCTTTTTGCTGTGTAGTTGCCTCTGCCTGGGATGCCCTTCCCCCTTTCTCTGCC  840
TGGCAACATCTTAGTTGTCCTTTGAGGCCCCAACTCAAGTGTCACCTCCTTCCCCAGCTCCCCCAGGCAGAAATAGTTGTCTGTGCTTCCTTGGTTCATGCGTCTACTGTGACACTTATC  960
TGACTGTTTTATAATTAGTCGGGCATGAGTCTGTTTCCCAAGCTAGACTGTGTCTGAATCATGTCTGTATCCCCAGTGCCCGGTGCAGGGCCTGGCATAGAGTAGGTACTCCATAAAAGG 1080
                                                    * End of ear-1 exon F-1                                    ▲
TGTGTTGaattgaactgcgtctgcctcctcccccggtcaggcgagagcctgacctacCTGCAGAGACAAGCACCACCGCGGTGAAGAGGCCCAGCTCCTCCTCGGTAAGCGCCAGGGAGT 1100
TGAGCTTCTCGCTGAAGTCGAACATGGCACTGAGCAGGTCTCCCATGCCCATGGCACCAAGCTCCTGCAGGCTGTAGGTGGTGCGGCTTAGGAACATCACTGTCTGGTCCTTCACGTTGA 1320
                  * Start of ear-1 exon F-1
ACAACGAAGCAAAGCGCACcatcagcacctaggagggaaggggaacagtc 1370
```

Fig. 3. Sequence of the 10 exons of THRA gene.Exons are typed in upper case letters except for the 147 bp junction between 14C and 14J cDNAs in exon 9 (844–961) which is in underlined, lower case letters. Each exon is numbered individually. Intronic sequences are in lower case letters. Oligonucleotides A (exon 9), B (exon 7), C (exon 9) and D (exon 8) are underlined. Putative polyadenylation signals are shown by arrowheads located under the first base of the site. For the genomic portion containing exon 10, the *ear-1* exons 'F' and 'F-1' are shown in italics. The translation product of these two exons is not shown since it is identical to that in reference 11. The beginning of the THRA exon 10 is underlined. The alternative splice donor site located 128 bp downstream from the beginning of exon 10 is indicated.

kb mRNA species in MRC5 cells (data not shown). To exclude the possibility that the two probes recognize two different RNAs of the same size, we performed PCR experiments which showed that the 14C and 14J sequences both contributed to the same 5.0 kb transcript. It was difficult to directly amplify the missing 147 nucleotide junction between 14C and 14J because this fragment was entirely encoded by exon 9 and we could not exclude the possibility that genomic DNA contaminating the RNA preparation would give the expected signal. We therefore chose an indirect strategy depicted in Fig.4A wherein RNAs isolated from MRC5 cells expressing the 5.0 kb mRNA were retrotranscribed using oligonucleotide primer A. This oligonucleotide corresponds to the very 5′ part of the 14J cDNA and was chosen to be specific for these sequences (an extensive computer search has revealed no significant homology with 14C). The single strand DNA molecules obtained were used as templates for PCR experiments using oligonucleotides B and C corresponding to sequences of exons 7 and 9. The expected 556 bp band was obtained (data not shown) and was reamplified by nested PCR using primer C already described and primer D corresponding to the end of exon 8 and yielded the expected 307 bp product. This band co-migrates with the product of the same PCR reaction conducted with MRC5 RNAs retrotranscribed with oligonucleotide C (Fig. 4, panel B). To exclude the possibility that the observed bands were due to carryover contamination (49), we performed control amplifications where no target DNAs were added. As expected, this control did not reveal any band after amplification with either oligonucleotides B and C (not shown) or C and D (Fig.4, panel B). This experiment shows that after priming in the 14J sequences we have been able to amplify parts of 14C. The use of oligonucleotides which correspond to different exons (namely 7, 8 and 9) for primary as well as nested PCR prevents the amplification of genomic DNA contaminating the RNA preparation. We thus conclude that 14C and 14J sequences both contribute to the 5.0 kb mRNA species.

As these experiments do not exclude the possibility that a small intron could lie in the 147 bp missing fragment between 14C and 14J, we performed amplification experiments of that genomic DNA fragment and of MRC5 mRNA retrotranscribed with

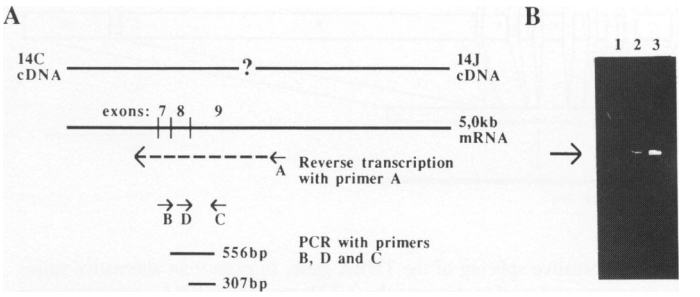


Fig. 4. (A) Strategy for the PCR experiments. Arrows show the oligonucleotides used : A and C correspond to sequences in exon 9, and B and D are from exons 7 and 8, respectively. This strategy prevents the amplification of genomic DNA. (B) visualization on 2% agarose gel stained with ethidium bromide of the nested PCR using primers D and C. The arrow shows the 307 bp band. Lane 1 : negative control without target ; lane 2 : amplification of the 307 bp band using as target the MRC5 RNAs retrotranscribed with primer A ; lane 3 : amplification conducted with an EcoRI-digested 14C cDNA as target.

oligonucleotide A. In the two samples we show that the bands obtained after amplification using primers flanking the 147 bp fragment displayed the same size, suggesting that there is no intron lying within this fragment (data not shown).

Alternative splicing in the THRA locus

The two products that can be generated from the THRA locus from the two mRNAs species differ in their C-terminal regions. The two products may be generated through an alternative splicing event in exon 9 (Fig.5). The 410 amino acid product may be generated by a messenger RNA which begins at exon 1 and terminates at the 3′ end of exon 9. The stop codon used for this product lies in the 3′ region of exon 9 (251 bp) (see Fig.3). The 490 amino acid product may be generated from the RNA produced by an alternative splicing event that uses a splice donor site present at 128 bp after the beginning of exon 9 and the splice acceptor site of exon 10. The stop codon used to

terminate this product is located 364 bp downstream from the beginning of exon 10. This event would replace the 40 final amino acids of the hTHRA1 with 120 amino acids specific for hTHRA2.

A nuclear receptor gene in an inverted orientation overlaps with THRA at its 3' end

A region of homology with the human *ear-1* cDNA is positioned in an inverted orientation relative to THRA (11, 34). This region corresponds to the most 3' exon of the human *ear-1* gene (named F here) which overlaps with exon 10 of THRA over 1028 bp (263 nucleotides of coding region for both genes ; see fig.2 and 3). The overlapping region does not extend in the 3' untranslated part of exon 9 (data not shown).

We have also located two others *ear-1* exons. The first (denoted 'F-1' in Fig.2) is immediately adjacent to the human exon F and corresponds to exon 7 of *REVerbA*, the rat homologue of *ear-1* (34). This exon is separated by 50 bp from the end of the THRA exon 10. The second (denoted HR2 for 'Homologous to Rat exon 2' in Fig.2) corresponds to the final exon of the A/B domain of *ear-1*. This region is homologous to exon 2 of the *REVerbA* described in rats (34) (see Fig.6). The precise location of the splice acceptor site of this exon has not been determined since there is no sequence reminiscent of a canonical splice acceptor consensus in this region (Fig.6)

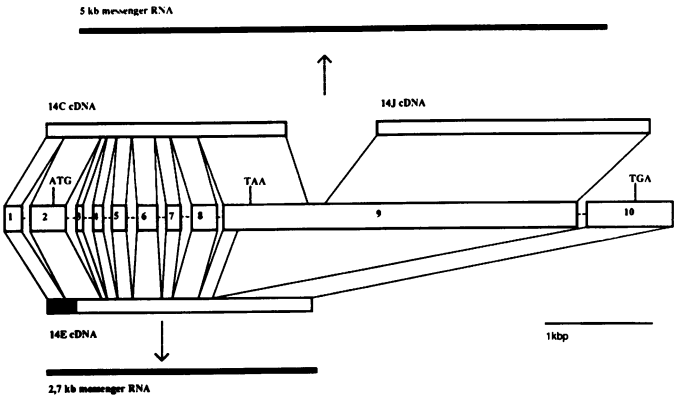


Fig. 5. Alternative splicing of the THRA gene. In exon 9 an alternative splice site is present and used to generate the 2.7 kb messenger RNA containing exon 10. The black boxes at the beginning of 14E show sequences absent from our clone but identified in the literature (20). The 5.0 kb messenger RNA contains sequences from exon 1 to 9. The 2.7 kb mRNA contains sequences from exon 1 to 8, the 128 first nucleotides of exon 9 and exon 10. The 5.0 kb RNA encodes the 410 amino acid product and the 2.7 kb RNA species encodes the 490 amino acid product.The black box at the beginning of the 14E cDNA represents missing sequences in our clone.

DISCUSSION

Genomic organization of THRA

In these studies, we determined the exon/intron organization of the human THRA gene. The THRA gene is split into 10 exons distributed along 27 kbp of chromosome 17 and overlaps in its 3' region with the structurally related *ear-1* gene. In addition, we provide evidence that the 410 amino acid product of THRA is encoded by a large messenger RNA which contains a lengthy 3' untranslated region.

The four phage inserts aligned to determine this structure were isolated from three different genomic libraries (see Material and Methods). In these libraries (derived from human placental, ML3 and IGROV-2 cells DNA) no rearrangement of the THRA gene has been described. By extensive endonuclease mapping and sequencing, we determined that the phage sequences overlap and have no broad differences in their common regions. Moreover, the structure we determined in this paper is confirmed by previous results obtained in the study of an EcoRI RFLP site present in the human THRA gene (44).

We have cloned two cDNAs, 14C and 14J, both of which hybridize to a 5.0 kb mRNA species. We demonstrate by PCR experiments that the two cDNAs may be attributed to the same messenger RNA. Indeed, after specific reverse transcription of 14J 5' related sequences we were able to amplify 14C 3'-related sequences. Thus, we conclude that the 410 amino acid product of THRA is translated from an RNA of 5.0 kb which contains a long 3' untranslated region of at least 3106 nucleotides. This feature is common among nuclear receptor genes. For example, the human glucocorticoid receptor contains a 3' untranslated

```
                                    * REV-ERB SPLICE ACCEPTOR
REV-erb 200                 TTTGAAAGA AAGGTGGTGT TATCACCTAC ATTGGCTCCA GCGGATCCTC CCCCAGCCGG ACCAGCCCAG AGTCCCTCTA TAGTGACAGC TCCAATGGCA GCTTCCAGTC
HR2       1  ccccctttcc ttttgttctc tctccaatgg CATCACCTGC ATTGGCTCCA GTGGCTCCTC CCCAAGCCGC ACCAGCCCTG AATCCCTCTA TAGTGACAAC TCCAATGGCA GCTTCCAGTC
EAR-1   647                 ACAACAACA CAGGTGGCGT CATCACCTAC ATTGGCTCCA GTGGCTCCTC CCCAAGCCGC ACCAGCCCTG AATCCCTCTA TAGTGACAAC TCCAATGGCA GCTTCCAGTC
                                                 I  T  Y   I  G  S   S  G  S  S   P  S  R   T  S  P   E  S  L  Y   S  D  N   S  N  G   S  F  Q  S
                                    --------- EAR-1 SPLICE ACCEPTOR

REV-erb 309  CCTGACTCAA GGTTGTCCCA CATACTTCCC ACCATCACCT ACTGGCTCCC TCACCCAGGA CCCTGCCCGC TCATTTGGCA CTGTGCCACC CAGCCTCAGT GATGATAGCT CCCCTTCTTC
HR2       121  CCTGACCCAA GGCTGTCCCA CCTACTTCCC ACCATCCCCC ACTGGCTCCC TCACCCAAGA CCCGGTTGCC TCCTTTGGGA GCATTCCACC CAGCCTGAGT GATGACGGCT CCCCTTCTTC
EAR-1   756  CCTGACCCAA GGCTGTCCCA CCTACTTCCC ACCATCCCCC ACTGGCTCCC TCACCCAAGA CCCGGCTCGC TCCTTTGGGA GCATTCCACC CAGCCTGAGT GATGACGGCT CCCCTTCTTC
               L  T  Q   G  C  P   T  Y  F  P   P  S  P   T  G  S   L  T  Q  D   P  A  R   S  F  G   S  I  P  P   S  L  S   D  D  G   S  P  S  S

                                                                                                                    M  E  D   S  S  R   V  S  P   S  K  G  T   S  N  I
REV-erb 429  TGCTTCATCA TCATCATCATC TTCCTCCTCC TCCTTCTATA ATGGGAGCCC CCCCGGAAGT CTACAAGTGG CCATGGAAGA CAGCAGCCGA GTGTCCCCCA GCAAGGGCAC CAGCAACAT
HR2       241  --CT-CGTCT TCCTCGTCGTC ATCCTCCTCC TCCTTCTGTG ATGGGAGCCC CCCTGGGAGT CTACAAGCGG CCATGGAGGA CAGCAGCCGA GTGTCCCCCA GCAAGAGCAC CAGCAACAT
EAR-1   876  --CT-CATCT TCCTCGTCGTC ATCCTCCTCC TCCTTCTATA ATGGGAGCCC CCCTGGGAGT CTACAAGTGG CCATGGAGGA CAGCAGCCGA GTGTCCCCCA GCAAGAGCAC CAGCAACAT
                  S  S   S  S  S  S   S  S  S   S  F  Y   N  G  S  P   P  G  S   L  Q  V   A  M  E  D   S  S  R   V  S  P   S  K  S  T   S  N  I

               T  K  L   N  G  M   V  L  L
REV-erb 549  CACTACCAAG CTTAACGGCA TGGTACT  575
HR2       358  CACCAgtgag tacttccagc tgatttt  384
EAR-1   996  CACCAAGCTG AATGGCATGG TGTTACT  1022
               T  K  L   N  G  M   V  L  L
                  * SPLICE DONOR
```

Fig. 6. Sequence of the HR2 exon of *ear-1* compared with the human *ear-1* cDNA (11) and the rat *REV-erb* cDNA. The initiation codon used to generate the *REV-erb* gene product is underlined as is the splice donor site. Intronic sequences are in lower case letters (see also Ref 47).

region of 2.3 kb (45) and the rat THRB gene possesses a similar 2.7 kb region. As numerous reports have suggested that sequences present in the 3′ non coding parts of various mRNAs may play a regulatory role in the turnover of these molecules (see 44 and references therein), it is tempting to speculate that secondary structures in the long, 3′-untranslated region of the THRA 5.0 kb mRNA may positively or negatively modulate its degradation.

The putative polyadenylation sites of 14J do not correspond to any known consensus sequence for polyadenylation (47). We nonetheless feel we have identified the authentic terminus of the RNA because the corresponding site in the genomic DNA, contains no A-rich sequence that may be used for internal priming during library construction. The role of these alternative polyadenylation sites on the stability or on the translation rates of the various THRA mRNAs remains to be addressed (46, 47).

The rat and human *ear-1* genes

Nuclear receptor genes are generally highly conserved during evolution especially among mammals. However, whereas the initiation codon of the rat *REV-erbA* gene is located within exon 2 which corresponds to our human HR2 exon (34), the ATG initiation codon proposed for the human sequence lies in the preceeding exon (11, 49). Indeed, the human product described (11) is 106 amino acid longer at its 5′ end than the rat product (34) and the 5′ portions of the two published cDNAs have no homology upstream of the HR2 exon. This observation is interesting since, to our knowledge, this represents the major variation between rat and human nuclear receptor genes in a domain (A/B) which is generally implicated in trans-activation.

Comparison with other nuclear receptor genes

The genomic organization of the human THRA gene shares similar structural features with its avian homologue (33). The structure of the two genes is similar within their coding region, with the initiation codon at the end of an exon encoding 17 amino acids in both cases (the entire length of this initiation exon is 202 nucleotides for chicken exon 3, and 350 nucleotides for human exon 2). The following 5 exons are of similar size for the two species and their boundaries are located at similar positions. No sequencing data of the 3′ region of the chicken gene are available in the literature, and RNase mapping experiments suggest that the equivalent of human exon 10 does not exist or is not transcribed in the chicken (13). The major differences between the two genes are as follows : *(i)* The 5′ untranslated regions share no homology and are split into three exons in the chicken and two in the human genes. Nevertheless, in the chicken as in humans, the most 5′ exon is far from the next one (4.5 kbp in chickens, 11 kbp in humans) *(ii)* The size of the 5′ two-thirds of the gene is 21 kbp for humans and 12 kbp for chickens, the difference being a consequence of the relative sizes of introns. For example, the intron lying between the two zinc finger-coding exons is 80 bp in the chicken and 5 kbp in the human gene. Since the complete chicken gene has yet to be cloned, we cannot compare the two entire genes. A comparison is also possible with the partially determined organization of the 3′ region of the rat THRA gene. In this region, human and rat genes possess a nearly identical organization. In particular, the 263 nucleotide overlap found between the coding regions of *ear-1* and THRA genes is perfectly conserved between rats and humans (11, 34). This argues in favor of an important role for the unusual overlapping structure.

When THRA is compared with other nuclear receptor genes of the steroid hormone receptor sub-familly, several observations can be made. First, the steroid hormone receptor genes, such as the chicken progesterone receptor (29), the human estrogen receptor genes (30) or the human androgen receptor gene (31, 33) appear longer than human THRA. For instance, the human estrogen receptor gene is more than 140 kbp in length (30). Second, like all known nuclear receptor genes except *nur77* (a growth factor inducible 'orphan receptor' gene ; see 43), the two zinc finger domains of THRA are encoded by separate exons. In both THRA and retinoic acid receptors genes an intron is located one amino acid after the last cysteine residue of the first zinc finger C1 (2, 5, 35). In contrast, the intron is located in a different position in the middle of the region lying between the two zinc fingers in the steroid hormones receptor genes and *nur77* (2, 29, 30, 43). Third, the A/B domain of THRA is split in two exons (exons 2 and 3), whereas in steroid hormone receptors gene (human estrogen receptor and chicken progesterone receptor genes) and *nur77*, this domain is encoded by a single exon (29, 30, 43). Fourth, the organization of the 5′ region of the thyroid hormone/retinoic acid receptor genes seems variable : the RARα and γ 5′ region are split into several exons (36, 50) like THRA, whereas the same portions (*i.e.* 5′ untranslated region and A/B domain) of RARβ are composed of only one exon (V. Giguere, personal communication). Finally, as pointed out by Ponglikitmongkol et al. (30), there is an intron between domain D and E in THRA which is not present in estrogen or progesterone receptor genes. In these latter genes, the exon encoding domain D exhibits a mosaic structure encoding the end of C2, D and the first 63 amino acids of E. This difference may be explained by the probable loss in the steroid hormone receptor sub-family, of an intron which is still present in THRA genes. A comparison of the genomic structures of various nuclear receptor genes reveals a common mode of organization which appears to have evolved through gene duplication from a common ancestral progenitor to give rise to the three sub-families presently known : steroid hormone receptor genes, thyroid hormone/retinoic acid receptor genes and the *nur77* family. Additional work on other nuclear receptor genes, including determination of their genomic organization, is needed to better understand the evolutional history of this superfamily of genes.

ACKNOWLEDGEMENTS

We thank Eli Canaani, John Toole and Tom Maniatis for providing human cDNA or genomic libraries, Jean Coll and Jean Marc Vanacker for helpful discussions, Brigitte Quatannens for the excellent technical assistance, Marie-Christine Bouchez and Nicole Devassine for piloting the word processor apparatus and Thierry Dupressoir for help in figure preparation. We are also grateful to Ruth Globus and Philip Boucher for critical reading of the manuscript.

V.L. holds a fellowship from the French Ministère de la Recherche et de la Technologie. We thank the Association de la Recherche contre le Cancer and Ligue Nationale contre le Cancer for financial support.

REFERENCES

1. Evans,R.M.(1988) *Science*, **240**, 889–895.
2. Green,S. and Chambon,P. (1988) *TIG*, **4**, 309–314.
3 .Beato,M. (1989) *Cell*, **56**, 335–344.
4 .Petkovich,M., Brand,N.J., Krust,A. and Chambon,P. (1987) *Nature*, **330**, 444–450.
5. Brand,N., Petkovich,M., Krust,A., Chambon,P., de Thé,H., Marchio,A., Tiollais,P. and Dejean,A. (1988) *Nature*, **332**, 850–853.

6. Zelent,A., Krust,A., Petkovich,M., Kastner,P. and Chambon,P. (1989) *Nature*, **339**, 714–717.
7. Krust,A., Kastner,Ph., Petkovich,M., Zelent,A. and Chambon,P. (1989) *Proc.Natl.Acad.Sci.USA*, **86**, 5310–5314.
8. Weinberger,C., Thompson,C.C., Ong,E.S., Lebo,R., Gruol,D.J. and Evans, R.M. (1986) *Nature*, **324**, 641–646.
9. Spurr,N.K., Solomon,E., Jansson,M., Sheer,D., Goodfellow,P.N., Bodmer,W.F. and Vennström,B. (1984), *EMBO J.*, **3**, 159–163.
10. Miyajima,N., Kadowaki,Y., Fukushige,S., Shimizu,S., Semba,K., Yamanashi,Y., Matsubara,K., Toyoshima,K. and Yamamoto,T. (1988) *Nucl.Acids Res.*, **16**, 11057–11074.
11. Miyajima,N., Horiuchi,R., Shibuya,Y., Fukushige,S., Matsubara,K, Toyoshima,K. and Yamamoto,T. (1989) *Cell*, **57**, 31–39.
12. Sap,J, Munoz,A., Damm,K, Goldberg,Y., Ghysdael,J., Leutz,A., Beug,H. and Vennström, B. (1986) *Nature*, **324**, 635–640.
13. Forrest,D., Sjöberg,M. and Vennström,B. (1990) *EMBO J.*, **9**, 1519–1528.
14. Saule,S., Roussel,M., Lagrou,C. and Stéhelin,D. (1981) *J.Virol.*, **38**, 409–419.
15. Debuire,B., Henry,C., Benaissa,M., Biserte,G., Claverie,J.M., Saule,S., Martin,P. and Stéhelin,D. (1984) *Science*, **224**, 1456–1459.
16. Kahn,P., Frykberg,L., Brady,C., Stanley,I., Beug,H., Vennström,B. and Graf,T. (1986) *Cell*, **45**, 349–356.
17. Gandrillon,O., Jurdic,P., Benchaibi,M., Xiao,J.H., Ghysdael,J. and Samarut,J. (1987) *Cell*, **49**, 687–697.
18. Gandrillon,O., Jurdic,P., Pain,B., Desbois,C., Madjar,J.J., Moscovici,M.G., Moscovici, C. and Samarut,J. (1989) *Cell*, **58**, 115–121.
19. Damm,K., Thompson,C.C. and Evans,R.M. (1989) *Nature*, **339**, 593–597.
20. Benbrook,D. and Pfahl,M. (1987) *Science*, **238**, 788–791.
21. Thompson,C.C., Weinberger,C., Lebo,R. and Evans,R.M. (1987) *Science*, **237**, 1610–1614.
22. Koenig,R.J., Warne,R.L., Brent,G.A., Harney,J.W., Reed Larsen,P. and Moore,D.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5031–5035.
23. Murray,M.B., Zilz,N.D., McCreary,N.L, MacDonald,M.J. and Towle,H.C. (1988) *J.Biol.Chem.* **263**, 12770–12777.
24. Mitsuhashi,T., Tennyson,G.E. and Nikodem,V.M. (1988) *Proc.Natl.Acad.Sci USA*, **85**, 5804–5808.
25. Nakai,A., Sakurai,A., Bell,G.I. and DeGroot,L.J. (1988) *Molec.Endocrinol.*, **2**, 1087–1092.
26. Izumo,S. and Madhavi,V. (1988) *Nature*, **334**, 539–543.
27. Koenig,R.J., Lazar,M.A., Hodin,R.A., Brent,G.A., Reed Larsen,P. , Chin,W.W. and Moore,D.D.(1989) *Nature*, **337**, 659–661.
28. Hodin,R.A., Lazar,M.A., Wintman,B.I., Darling,D.S., Koenig,R.J., Reed Larsen,P., Moore,D.D. and Chin,W.W. (1989) *Science*, **244**, 76–79.
29. Huckaby,C.S., Conneely,O.M., Beattie,W.G., Dobson,A.D.W., Tsai,M.J. and O'Malley, B.W. (1987) *Proc.Natl.Acad.Sci.USA*, **84**, 8380–8384.
30. Ponglikitmongkol,M., Green,S. and Chambon,P. (1988) *The EMBO Journal*, **7**, 3385–3388.
31. Kuiper,G.G.J.M., Faber,P.W., van Rooij,H.C.J., van der Korput,J.A.G.M., Ris-Stalpers,C., Klaassen,P., Trapman,J. and Brinkmann,A.O. (1989). *J. Mol. Endocr.*, **2**, R1-R4.
32. Tilley,W.D., Marcelli,M. and McPhaul,M.J. (1990). *J. Biol. Chem.* **265**, 13776–13781.
33. Zahraoui,A. and Cuny,G. (1987) *Eur.J.Biochem.*, **166**, 63–69.
34. Lazar,M.A., Hodin,R.A., Darling,D.S. and Chin,W.W. (1989) *Mol.Cell.Biol.* **9**, 1128–1136.
35. Dejean,A., Bougueleret,L., Grzeschik,K.H. and Tiollais,P. (1986), *Nature*, **322**, 70–72.
36. 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.
37. Vennstom,B., Fanshier,L., Moscovici,C. and Bishop,J.M. (1980) *J.Virol.*, **336**, 575–585.
38. Lawn,R.M., Fritsch,E.F., Parker,R.C., Blake,G. and Maniatis,T. (1978) *Cell*, **15**, 1157–1174.
39. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) Molecular Cloning : A laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, New York
40. Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.* **65**, 499–560.
41. Wang,A.M., Doyle,M.V.and Mark,D.F. (1989) *Proc.Natl.Acad.Sci. USA*, **86**, 9717–9721.
42. Breathnach,R. and Chambon,P. (1981) *Annu.Rev.Biochem.*, **50**, 349–383.
43. Ryseck,R.P., Macdonald-Bravo,H., Mattei, M.G., Ruppert,S. and Bravo, R. (1989) *EMBO J.*,**8**, 3327–3335.
44. Mathieu-Mahul,D., Xu,D.Q., Saule,S., Lidereau,R., Galibert,F., Berger,R., Mauchauffé,M. and Larsen,C.J. (1985) *Hum. Genet.*, **71**, 41–44.
45. Hollenberg,S.M., Weinberger,C., Ong,E.S., Cerelli,G., Oro,A., Lebo,R., Thompson,B., Rosenfeld,M.G. and Evans,R.M. (1985) *Nature*, **318**, 635–641.
46. Jackson,R.J. and Standart,N. (1990) *Cell*, **62**, 15–24.
47. Wickens,M. (1990)*Trends Biol. Sci.*,**15**,277–281.
48. Higuchi,R and Kwok, S. (1989) *Nature*, **339**, 237–238.
49. Lazar,M.A., Jones,K.E. and Chin,N. (1990) *DNA*, **9**, 77–83.
50. Brand,N.J., Petkovich,M. and Chambon,P. (1990) *Nucl. Acids Res.*, **18**, 6799–6806