BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **245,** 64–69 (1998)
ARTICLE NO. RC988376

Striking Evolutionary Conservation of a cis-Element Related to Nuclear Receptor Target Sites and Present in TR2 Orphan Receptor Genes

Catherine Le Jossic and Denis Michel[1]
Laboratoire d'Endocrinologie Moléculaire de la Reproduction, UPRESA 6026 CNRS-Université de Rennes 1, Bat. 13, Campus de Beaulieu, 35042 Rennes France

Received February 24, 1998

A systematic scanning of nucleic acid databases for DNA elements made of combinations of RGGTCA nuclear receptor half sites, has revealed that identical 19 nucleotide-long motifs composed of two inverted RGGTCA sites with a spacing of 7 nucleotides (IR7), are present upstream of the regions coding for the human TR2 and of the sea urchin SpSHR2 orphan receptors. We have developed an experimental strategy based on PCR, to check if this IR7 could correspond to an unusually long cis-element, conserved along evolution and regulating the TR2 genes. We found that indeed IR7 is present in the 5′ untranslated region of TR2 genes from all species tested, including *Xenopus*, rainbow trout, zebrafish and mouse. The exact conservation throughout the animal kingdom of such a long, non repetitive and non coding genomic region, highly suggests that it should ensure important biological functions. In addition, this work has allowed the identification of a new, non coding, upstream exon in the mouse TR2 gene present in testicular TR2 mRNAs.

In striking contrast with the large size of the nuclear receptor superfamily, the DNA targets of these transcription factors appear poorly varied. For example, the major subfamily of nuclear receptors, whose most NH2-terminal zinc-finger contains a P box of the D/E category, recognises RGKTCA hexamers (where R= A or G; K= G or T), which are more frequently RGGTCA. This limited number of DNA targets for nuclear receptors raises the question of their specificity of action. In fact, most nuclear receptors bind to DNA as dimers, and their binding specificity is based on the relative orientation and spacing the two half sites (1). They can be arranged as direct (DR), inverted (IR) or everted (ER) repeats (2), spaced by nucleotide gaps of variable length. Some of these composite sites have already been unambiguously assigned to specific homo- or heterodimers of nuclear receptors (2), but many others remain orphan, whether the corresponding binding factors do not exist, or are not yet identified.

One possible approach to identify new targets for nuclear receptors is to screen gene regulatory regions for the presence of combinations of half sites. We have undertaken such a screening and found that inverted RGGTCA sites, separated by 7 nucleotides (IR7), are present in two related genes encoding nuclear receptors: the human TR2 and the sea urchin SpSHR2 orphan receptors. Surprisingly, in addition to the RGGTCA half sites, the 7 central bases are also identical between human and sea urchin. This unexpected observation prompted us to check for the possible biological relevance of the IR7 motifs.

MATERIALS AND METHODS

RT-PCR analyses. Testes from rainbow trout (*Oncorhynchus mykiss*), mouse (*Mus musculus*) and frog (*Xenopus laevis*), and total adult zebrafishes (*Danio rerio*) were roughly crushed under liquid nitrogen and immediately dissolved in the Trizol reagent (Gibco BRL) for RNA extraction. 2 μg of each total RNA were reverse-transcribed using random hexamers as primers and the MMLV reverse transcriptase (Gibco BRL). 1/20 of reverse-transcription reactions were used as templates for all PCR. The PCR primers used in this study were: acidic ribosomal phosphoprotein (PO): 5′-AAYGTGGGCTCCAAGCAGATG-3′ (sense) and CTGCTGAACATGCTGAACATCTC-3′ (antisense); TR2 primers: IR7, 5′-GGTCACGAACTCTGAC-3′ (sense, putative exon 1); mIR7, 5′-GGTCAatccggaTGAC-3′. The names of primers defined in coding regions correspond to the tripeptides encoded by their 3′ ends: NRC, 5′-GCAGWACTGGCAGCGGTT-3′ (antisense, exon 5); VTA: 5′-ATCCAGATCGTGACAGCAC-3′ (sense, exon 3); GDK: 5′-CCTGATGCTTTGTCTCCAC-3′ (antisense, exon 4). Hybridization temperatures were: 55°C for RT-PCR using the IR7 primer, and 59 °C for all others. Single amplifications of 30 cycles and RT-minus controls were done for all the PCRs presented in this study.

[1] Corresponding author. Fax: 33 2 99 28 67 94. E-mail: Denis.Michel@univ-rennes1.fr.

0006-291X/98 $25.00

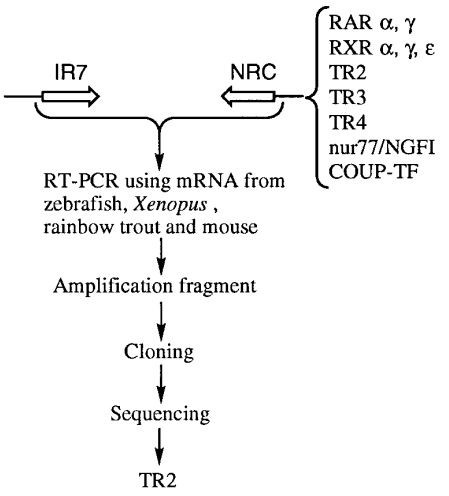
PCR fragment cloning. The zebrafish TR2 PCR fragment was blunted by the Klenow fragment and then inserted into the dephosphorylated EcoRV site of pBSK. For all other amplification products, 2 μl aliquots of PCR reactions were directly ligated to the pGEM-T vector (Promega Biotech.) according to the supplier protocol. DNA sequences were determined using the dideoxy method kit from Pharmacia.

RESULTS

Two orphan nuclear receptor genes are endowed with the same IR7 element. Search for nuclear receptor sites in the Genbank DNA database, by using the Lipman-Pearson alignment program (3), has revealed that 7 nucleotides-spaced inverted RGGTCA motifs are present in the proximal promoter region of the human TR2 gene (accession U19026) and in the 5′UTR of a splice variant of the sea urchin *Strongylocentrotus purpuratus* orphan steroid hormone orphan receptor SpSHR2 (accession U38281), which is precisely related to the mammalian TR2 (4). In addition, the 7 intervening nucleotides located between the two RGGTCA repeats are identical in the human and the sea urchin sequences, so that the region of complete nucleotide identity is 17 bp-long. Since, theoretically, there is no chance that a given DNA sequence longer than 14 nucleotides occurs in a vertebrate genome-sized random DNA, we postulated that the exact identity of the IR7 sequence, between very distantly related animal species, should not result from a mere coincidence, but from a selection mechanism. This hypothesis was all the more likely that the IR7 sequences are located in the vicinity of closely related genes, and in untranslated regions which are free from coding constraints.

Strategy used to check for the presence IR7 in TR2 genes from other species. To determine if the IR7 element does represent an important, evolutionary stable, cis-element of TR2 genes, we decided to look for its presence in other animal species. To this end, we have developed an experimental strategy based on the additional assumption that, as in sea urchin SpSHR2, IR7 may be located downstream of a transcription start site and then be present in the leader regions of the derived mRNAs. In such a case, it would be possible to detect TR2 messengers by RT-PCR, using as an upstream primer an oligonucleotide corresponding to the IR7 sequence and a downstream primer defined in the TR2 coding region (Figure 1A). This last primer has been deliberately chosen in a DNA-binding domain-coding region well conserved among many nuclear receptors, so that this primer can hybridize to a wide variety of nuclear receptor cDNAs (Figure 1B). It has been called NRC ("Nuclear Receptor Conserved region") since its sequence corresponds to a region encoding tripeptide NRC. RT-PCR with the IR7-NRC primer tandem have been performed using testis mRNA from adult *Xenopus,* rainbow trout and mouse, and mRNA from total

A



B

cDNA	Sequence	Access.
complementary to primer NCR	AACCGCTGCCAGTWCTGC	
Chicken RAR-gamma	----------	X73973
Chicken RAR-alpha	----------	X78335
Mouse RAR-alpha	----------	X57528
Zebrafish RAR-alpha	----------	L03399
Mouse Thyroid HR nur/77	----------	J04113
Rat NGFI-B	----------T	U17254
Drosophila DHR38	-----T----------	X89246
Drosophila Chorion F1	----------	X52591
Zebrafish RXR-epsilon	----------	U29942
Mouse RXR-gamma	----------T	M84819
Mouse TR2	----------	U30482
Human TR2	--------T--A------	U19026
Sea Urchin SpSHR2	----------	U38281
Mouse TR4	----------	U32939
Human TR4	--------T-----T---	L27586
Human COUP-TF	----AG-----A------	X16155
Mouse COUP-TF	----AG-----A------	X74134
Xenopus COUP-TF	--T-AG----------	X63092

FIG. 1. Scheme of the strategy used to determine the status of the IR7 motif in nuclear receptor genes from evolutionary distant vertebrate classes. (**A**) RT-PCR amplifications were done using as a downstream primer an oligonucleotide to a sequence widely conserved among receptors and species (see panel 1B), and as an upstream primer, an oligonucleotide of IR7 sequence. (**B**) Sequence conservation, among various nuclear receptors, of the cDNA region used to define the NRC primer used in the present study.

adult zebrafishes. As shown in figure 2, lanes IR7/NRC, they yielded clearly visible amplification products, about 600 nucleotide-long. Two bands were obtained from fishes, the larger one more intense for the zebrafish. When the IR7 oligonucleotide is replaced by another one, containing the two half sites with identical spacing and orientation but with different central bases, these PCR products are no longer obtained (Figure 2, lanes mIR7/NRC), indicating that the nature of the central bases is also conserved between all the tested species.

To confirm their precise relationships with TR2

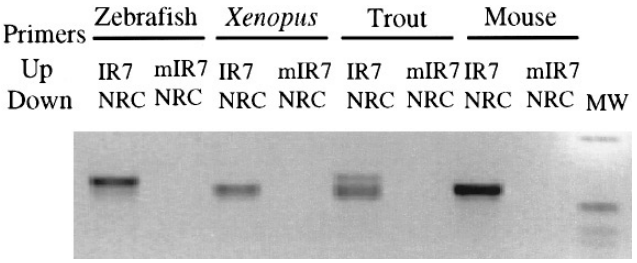


FIG. 2. The DNA fragments amplified by the IR7-NRC PCR appeared of comparable size in ethidium bromide-stained agarose gel, with the exception of fishes, where bands of higher molecular weight were obtained, of particularly strong intensity in zebrafish. After insertion into plasmid, all cloned fragments turned out to correspond to TR2 cDNAs. When changing the IR7 primer to an oligonucleotide still containing the inverted half-sites, but separated by 7 different nucleotides, PCRs performed under the same conditions failed to amplify any fragment.

genes, the exact content of the PCR products was then determined. After cloning and sequencing, all these amplified DNA fragments turned out to correspond to TR2 cDNAs. Although no fish TR2 gene sequences were so far available in genetic databases, the PCR fragments obtained from zebrafish and rainbow trout are obviously analogous to TR2 cDNAs, as unambiguously demonstrated by sequence comparison of the coding region (data not shown). Besides, the coding region of the mouse PCR fragment is 100% identical to the mouse TR2 cDNA sequences previously reported (5). The 5′ sequences are less well conserved (Figure 3), probably because they are non-coding, or coding for the NH2 termini of nuclear receptor that are generally not well conserved. However, the fact that these cDNA fragments were primed at their 5′ end by the IR7 primer, indicates that an IR7 motif is present in all of them and in the same orientation. Although we cannot exclude that the actual cDNA sequence may slightly differ from that of the IR7 primer, the fact that PCR hybridization steps were done at temperatures higher than the melting temperature of the IR7 oligonucleotide, indicates that this possibility of differences should be very limited. In the case of the mouse TR2 cDNA,

FIG. 3. Sequence alignment of the IR7-NRC fragments with the 5′ upstream region of the human TR2 gene (accession U19026). Sequences are well conserved in the coding region but not in the 5′UTR, except at the level of the IR7 element, as demonstrated by the common feature of all cDNAs to have been primed by an oligonucleotide defined in the IR7 sequence. Although a few base substitutions cannot be excluded at the level of the IR7 primer, this possibility should be very limited considering the low melting temperature of the 16-mer IR7 oligonucleotide used. In addition, one must notice that the most 3′ two bases of the IR7 domains of cloned cDNAs are C and Y (C/T), in accordance with the expected structure of nuclear receptor half-sites, though they were not included in the primer sequence (see Materials and Methods). In the case of the mouse, the IR7 sequence was confirmed by direct genomic sequencing. The initiation codons are underlined.

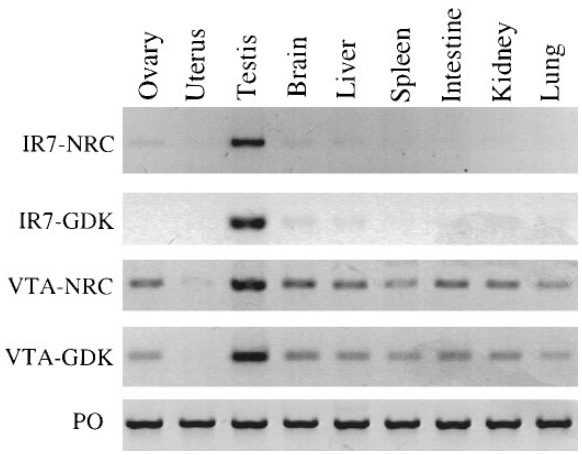


FIG. 4. Tissue distribution of IR7-containing TR2 mRNAs in the adult mouse. The same reverse-transcriptions mixtures were used as templates for PCR amplification of TR2-containing TR2 mRNAs or of total TR2 mRNAs, by using an upstream primer corresponding to the IR7 sequence or defined in the constitutive exon 3 (VTA) respectively. These two primers were coupled to the same downstream primers, chosen in exon 4 (GDK) or 5 (NRC). This comparative RT-PCR analysis shows that IR7-containing TR2 mRNAs are more restricted to the testis than total TR2 mRNAs, that are well detectable in other tissues. The ubiquitous acidic ribosomal phosphoprotein (PO) mRNA was used as an internal control.

the exact conservation of the IR7 motif has been confirmed by genomic sequencing (data not shown).

Although the NRC primer can theoretically prime the synthesis of various nuclear receptor cDNAs (Figure 1B), all the amplified fragments that we obtained correspond exclusively to TR2. This result, reproduced for all the tested species, strongly suggests that IR7 is specific of the TR2 mRNA.

Presence of IR7-NRC RT-PCR amplifiable mRNA in the adult mouse testis. The great power of amplification of RT-PCR approaches often leads to the detection of very low abundance mRNA species without biological relevance. To exclude such a possibility in this study, we wanted to estimate the relative abundance of IR7-containing TR2 mRNAs. For this purpose, we have tested mRNAs extracted from various mouse tissues previously analysed for their content in TR2 by either northern blot or in situ hybridization, by RT-PCR using as an upstream primer an oligonucleotide corresponding to the IR7 sequence. We have also measured total TR2 mRNAs from the same amounts of input cDNA templates, by using couples of primers defined in coding exons. Figure 4 shows that the TR2 mRNA species amplifiable by IR7-NRC (IR7-exon 5) or IR7-GDK (IR7-exon 4) RT-PCR, is abundant in the testis. After the same number of cycles, starting from the same templates and using the same downstream primers, amplifications with an upstream primer defined in exon 3 (VTA) yielded detectable amplification products in all tissues with the exception of the uterus, and with a higher level in the testis, in agreement with the TR2 mRNA tissue distribution previously reported (6). This result does not prove definitely that all testicular TR2 mRNA molecules contain the IR7 domain, but indicates that the IR7-containing TR2 mRNAs represent an important part of total TR2 mRNAs in the testis.

The IR7-NRC strategy has allowed to identify an additional exon in the mouse TR2 gene. Sequence comparison of the mouse IR7-NRC fragment with the TR2 gene sequence previously reported (5) (accession = U28269) pointed to a striking discrepancy at the level of the 5′UTR region. As shown in figure 5A, while the coding regions are completely identical, a sharp boundary is observed 20 nucleotides upstream of the translation initiation codon. Our interpretation is that the DNA region upstream of this point, and carrying the IR7 element, may correspond to an additional non-coding exon, perhaps optional. Alternatively, the most 5′ cDNA sequence previously proposed may correspond to intronic sequences. The latter possibility is supported by the fact that the DNA sequence precisely located at the limit of homology closely relates to the 3′ cis-elements of major introns, with a terminal YAG trinucleotide, preceded by a polypyrimidine stretch (7). Our attempts at amplifying RT-PCR products by using an upstream primer defined in the 5′ region specific of the TR2 cDNA previously reported (5) have so far failed, starting from all the mouse tissues tested in Figure 4 (data not shown).

DISCUSSION

The present work started from the initial hypothesis that the presence of identical IR7 elements in the human TR2 and sea urchin SpSHR2 genes is not fortuitous. Experiences show that this assumption was valid and more useful than expected, since in addition to reveal that the presence of IR7 elements in TR2 genes is generalized among metazoan species, this work has allowed the identification of a new exon in the mouse TR2 gene. Since a TR2 mouse mRNA lacking this IR7-containing exon has been described, the exact proportion of the IR7-containing mRNA species remains to be determined. However, our comparative RT-PCR experiments already allow to conclude that TR2 mRNAs containing IR7 are relatively abundant in the testis.

Contrary to most cis-elements, whose length rarely exceeds 8-9 nucleotides and can tolerate certain base changes, IR7 is very long and with apparently stringent sequence requirements. To our knowledge, the conservation degree of the IR7 motif is higher than those of all cis-elements so far known. When we began this work, the IR7 element was absent from nucleotide sequence databases, with the exception of the human TR2 and sea urchin SpSHR2 genes. Very recently, a

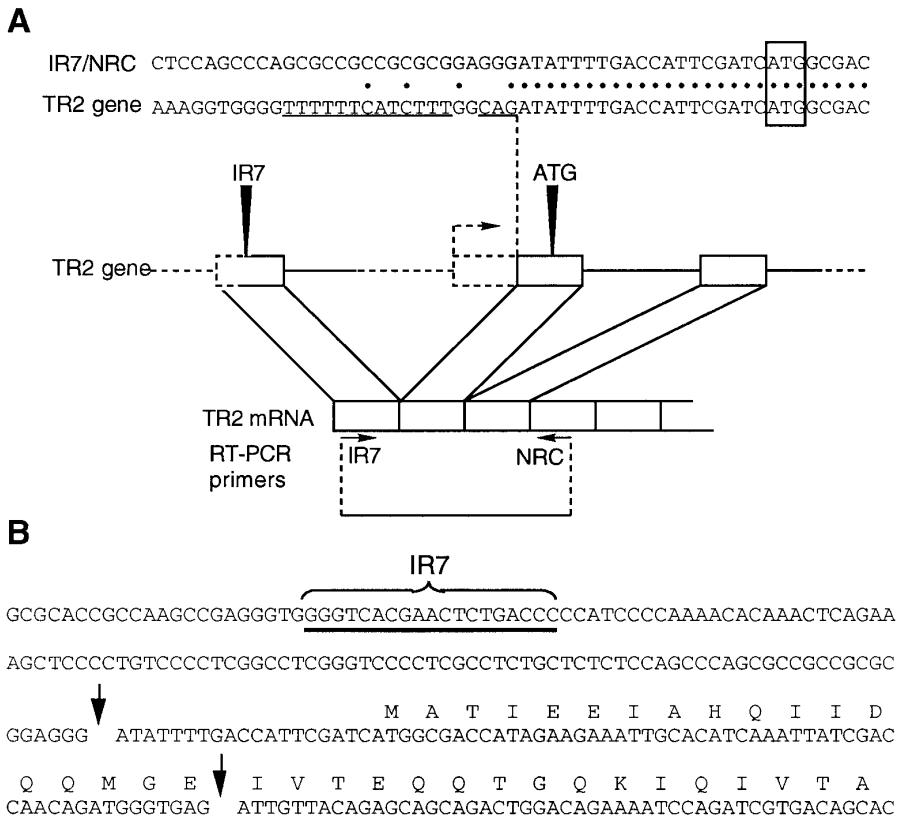


FIG. 5. Organisation of the upstream region of the mouse TR2 gene. (**A**) Sequence comparison between the mouse IR7-NRC fragment (upper line) with the mouse TR2 cDNA previously reported (accession U28269)(bottom line). The PCR fragment sequence is in perfect agreement with that of Lee et al. (6) all along the coding region, but not in the 5′UTR. The TR2 gene may be regulated by alternative promoters, one upstream of the IR7-containing exon, and another upstream of the first coding exon (arrow). An alternative explanation would be that the bottom sequence corresponds to an intron, since typical intronic 3′ cis-element (underlined) are present in 3′ of the non homologous region. Our RT-PCR analyses suggest that this intron separates a non coding exon carrying the IR7 element and a second exon carrying the translation initiation codon (boxed). Besides, the mouse cDNA sequence obtained by IR7-NRC RT-PCR is more closely related to that of the 5′ end of the human TR2 gene (accession U19026). (**B**) 5′ region of the mouse TR2 cDNA obtained by IR7-NRC PCR and genomic sequencing. The exon boundaries are deduced from sequence comparison with the TR2 genomic sequences and indicated with arrows.

novel cDNA sequence carrying the same IR7 sequence has been introduced in databases. This DNA sequence (accession AF008302) corresponds to a mRNA from the salamander axolotl, named developmental orphan receptor-1, and obviously related to the mammalian TR2 genes. Once again, in this case, the IR7 element is present in the 5′UTR. This additional report confirms that IR7 is astonishingly conserved along animal evolution and further emphasises the close association of this motif with TR2 genes. Generally, cis-elements are not restricted to particular genes, but widely distributed, and the pattern of expression of each gene is specified by a combination of cis-elements rather than by a single one. By contrast, the association of IR7 with a single gene, namely TR2, and its unusual high sequence stringency, are very atypical features.

An important point would be to identify IR7-specific factors. In fact, widely spaced inverted half-sites have already been identified as alternative targets for retinoic acid and estrogen receptors. This is the case for palindromic half-sites separated by 9 nucleotides (IR9) (8) or longer spacers (9), or, for everted half-sites, 6 nucleotides apart (ER6) (10). Besides, widely spaced half sites in direct (9) or inverted (11) orientations, have also been shown to act as a complex element for estrogen receptors (ERE). But in all these cases, the only function attributed to the central nucleotides of dimer nuclear receptor sites is the regulation of half-site spacing. The striking conservation of the 7 central nucleotides in the TR2 IR7 element suggests that they may play additional, sequence-dependent functions, as already shown for the RXR-RLD1 heterodimer (12). Whether or not IR7 belongs to a novel class of cis element, related or not to classical nuclear receptor sites, remains to be established, as is its possible involvement in the TR2 gene regulation. Although the precise function of TR2 orphan receptors remains to be determined, one may anticipate that this very ancient nuclear re-

ceptor (13) should be of major biological importance, as strongly suggested by the presence of a very specific cis element in its gene, unchanged during evolution from sea urchin to human.

REFERENCES

1. Glass, C. K. (1994) *Endocrine Rev.* **15,** 391–407.
2. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M. (1995) *Cell* **83,** 835–839.
3. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U S A* **85,** 2444–2448.
4. Kontrogianni-Konstantopoulos, A., Vlahou, A., Vu, D., and Flytzanis, C. N. (1996) *Dev. Biol.* **177,** 371–382.
5. Lee, C.-H., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Wei, L.-N. (1995) *Genomics* **30,** 46–52.
6. Lee, C.-H., Chang, L., and Wei, L.-N. (1996) *Mol. Reprod. Dev.* **44,** 305–314.
7. Tarn, W.-Y., and Steitz, J. A. (1997) *Trends Biochem.* **22,** 132–137.
8. Lee, M.-O., Hobbs, P. D., Zhang, X.-K., Dawson, M. I., and Pfahl, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5632–5636.
9. Kato, S., Sasaki, H., Suzawa, M., Masushige, S., Tora, L., Chambon, P., and Gronemeyer, H. (1995) *Mol. Cell. Biol.* **15,** 5858–5867.
10. Forman, B. M., Casanova, J., Raaka, B. M., Ghysdael, J., and Samuels, H. H. (1992) *Mol. Endo.* **6,** 429–442.
11. Kato, S., Tora, L., Yamauchi, J., Masushige, S., Bellard, M., and Chambon, P. (1992) *Cell* **68,** 731–742.
12. Apfel, R., Benbrook, D., Lernhardt, E., Ortiz, M. A., Salbert, G., and Pfahl, M. (1994) *Mol. Cell. Biol.* **14,** 7025–7035.
13. Laudet, V. (1997) *J. Mol. Endocrinol.* **19,** 207–226.