Pergamon

0960-0760(94)E0074-U

J. Steroid Biochem. Molec. Biol. Vol. 50, No. 3/4, pp. 121–130, 1994
Copyright © 1994 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0960-0760/94 $7.00 + 0.00

The Genomic Organisation, Sequence and Functional Analysis of the 5′ Flanking Region of the Chicken Estrogen Receptor Gene

Pádraig V. Nestor,[1,2]* Rosemarie C. Forde,[1] Paul Webb[4] and Frank Gannon[1,2,3]

[1]Department of Microbiology and [2]National Diagnostics Centre, University College Galway, Ireland, [3]EMBO, Postfach 10.2209, Meyerhofstrasse 1, D 69012, Heidelberg, Germany and [4]Metabolic Research Unit, University of California, San Francisco, CA 94143, U.S.A.

The cDNA of many members of the nuclear receptor superfamily has been cloned. Recently more effort has been expended on the analysis of these genes at the genomic level and on the factors controlling their expression. The genomic organization of the chicken estrogen receptor gene is presented and compared to the other members of the superfamily of hormone receptor genes with emphasis on the relationship to the functional domains. The results show that the gene is divided into eight exons and that the position of the intron/exon boundaries are as in the human gene but different to the trout estrogen receptor gene. Primer extension and cDNA clone isolation was used to determine the transcription start site and 3.0 kb of 5′ flanking sequence was generated. There is striking sequence homology to the human estrogen receptor promoter and there is a well positioned "typical" TATA sequence, with potential candidate CAAT box sequences close to the start site of transcription. In transient transfection assays, subfragments of this region drove CAT expression in chicken embryo fibroblasts, and the level was increased further with the addition of forskolin, but not phorbal myristate acetate. Including sequences more distal to the cap site in promoter constructs, completely abolished the promoter activity and forskolin inducibility, indicating the presence of strong silencing activity.

J. Steroid Biochem. Molec. Biol., Vol. 50, No. 3/4, pp. 121–130, 1994

INTRODUCTION

Estrogens are key physiological mediators of female sexual diferentiation, development and maturation. They act by association with ligand specific intracellular receptors. This complex in turn binds to specific *cis* elements in inducible genes thereby eliciting an appropriate response at the transcriptional level ([1] and references therein). This role as transcription factor has been extended to a superfamily of nuclear receptor genes which now includes thyroid hormone, retinoic acid, progesterone, vitamin D, glucocorticoid, mineralocorticoid and androgen receptors, as well as some orphan receptors for which ligands have not yet been identified. The cloning of the cDNAs for these genes has highlighted the striking conservation and maintenance throughout evolution of areas of homology. Extensive "domain swapping", DNA and hormone binding assays have allowed these areas of conservation to be assigned particular functional roles and domains to be delineated [2].

The chicken estrogen receptor (cER) whose cDNA was cloned and partially sequenced [3, 4] was one of the first of this superfamily of genes to be analysed. It is expressed in a tissue specific manner, being found predominantly in the reproductive tissues and not in the brain, liver, kidney or lung [4]. The mechanisms for the control of the expression of the ER are not known. While the ER regulates its own level to some extent [5, 6] it is not clear whether this is an indirect mechanism using other hormones or peptides, or if there is a direct interaction between the receptors and the regulatory sequences of their own genes. Not only is the regulation of the gene important as a tissue specific transcription factor, but also from a practical aspect as a large percentage of breast tumours have elevated

*Correspondence to P. Nestor at the National Diagnostics Centre, University College Galway, Ireland.
Received 13 Sep. 1993; accepted 8 Mar. 1994

levels of ER, with objective responses expected from these on treatment with antiestrogens [7]. It appears therefore that not only is the ER involved in the normal growth and development of the female breast but it is intrinsically linked to the maintenance of the malignant state.

As part of an extended analysis of the control of the expression of the ER gene we have cloned the genomic gene for the ER and determined its exonic organization, transcription start site and the sequence of its 5′ flanking region. Furthermore, we successfully define the functional promoter by introducing subfragments of the 5′ flanking region into the CAT expression vector pBLCAT3 and transfecting these into chicken embryo fibroblasts (CEF) cells under a number of different conditions.

MATERIALS AND METHODS

A chicken λ gtWES library was screened initially with a cDNA clone (12.15) which corresponds to nearly all of the coding region of the ER gene (Fig. 1). These were kindly provided by Dr Andree Krust and Dr Pierre Chambon at Strasbourg.

Other subfragments of 12.15 also used in hybridizations are shown in Fig. 1. The Hind III fragment was cloned directly into digested pGEM4Z plasmids (Promega Corporation). An Msp 1 digest of 12.15 was repaired using the Klenow fragment and dNTPs and cloned into EcoR1 digested, dephosphorylated and repaired pGEM4Z. This treatment regenerated the EcoR1 site and allowed this fragment to be conveniently isolated for labelling. Appropriate DNA fragments were radioactively labelled in low melt agarose by the random priming method [8].

Preparation of chicken genomic DNA

A chicken liver, stored at −80°C was ground to a powder using a mortar and pestle and then homogenized further in a Potter homogenizer. Nuclei were lysed with the addition of SDS to 0.5% and proteinase K (Boehringer) to 10 mg ml^{-1}. After 2 h at 37°C the suspension was phenol/chloroform extracted twice. DNA was spooled from 70% ethanol and resuspended finally in H_2O to a concentration of 1 mg ml^{-1}.

Preparation of exon 2 specific library

200 μg of genomic DNA digested with EcoR1 was centrifuged at 26,000 rpm through a sucrose gradient (10–40%) prepared as described in Maniatis *et al.* [9] for 14 h at 15°C in an SW41 rotor (Beckman). 200 μl fractions were removed and ethanol precipitated. 30% over every 3rd fraction was analysed on a 0.6% agarose gel which was blotted and hybridized to 12.15. DNA from appropriate fractions (1–3 kb) was then ligated to EcoR1 digested, dephosphorylated λgt10 arms before being packaged *in vitro* and infected into *E. coli* C600 Hfl. This library (2×10^6 pfu) was screened with the Msp 1 fragment probe.

Library screening, subcloning and sequencing

Approximately 5×10^5 plaques were plated onto five, 21 × 21 cm Bioassay dishes and transferred in duplicate to Hybond N membrane. These filters were baked and prehybridized in 6× SSC, 5× Denhardts, 1% SDS and 100 μg ml^{-1} sheared salmon sperm DNA for 1 h at 65°C. This solution was replaced with 10 mls hybridization solution (prehybridization solution without the salmon sperm DNA) containing 2×10^6 cpm radioactive probe per filter and incubated at 65°C for 16 h. The probe was carefully removed and the filters were washed (50 mls per filter) in 2× SSC at room temperature, 2× SSC at 65°C and 0.2× SSC at 65°C (30 min each). Filters were then exposed to XAR5 (Kodak) film at −80°C for 48 h. After three successive rounds of screening positively hybridizing phages were isolated. Phages were propagated, DNA isolated and subcloned in EcoR1 fragments for fine mapping into pGEM4Z. Exonic sequences were identified by hybridization and restriction mapping and subcloned into M13 mp18 or 19 or sequenced directly from pGEM4Z using oligonucleotides within identified exons. Sequencing was performed by the dideoxy chain termination method using T7 DNA polymerase (Pharmacia). Sequence from the 5′ flanking region was derived from overlapping subclones in M13, and also directly on double stranded material using oligonucleotides based on the previously determined sequence. Intronic sequence was identified by comparison to the published cDNA sequences and identification of consensus splice motifs.

RNA isolation and primer extension

Total cellular RNA was isolated from the oviducts of laying chickens immediately after removal using the guanidine thiocyanate extraction method [10].

Primer extension was essentially as outlined by Ausubel *et al.* [11]. 10 ng of the oligonucleotide pExt (Fig. 3) was 5′ end labelled using T4 polynucleotide kinase and ^{32}P dATP. Labelled oligonucleotide and 50 μg of chicken total RNA (oviduct or liver) were coprecipitated in ethanol and then resuspended in 30 μl of hybridization buffer (50% formamide, 40 mM Pipes pH 6.5, 0.4 M NaCl, 1 mM EDTA) at 40°C for 12 h. After ethanol precipitation hybrids were extended using 40 units of AMV reverse transcriptase (Promega Corporation) at 40°C for 90 min. This was phenol extracted, ethanol precipitated and loaded on a 6% polyacrylamide sequencing gel and electrophoreses alongside a sequence derived from a HindII genomic subfragment (bp −1065 to +182) subcloned in M13 as template and pExt as primer.

Constructs, cell culture and transfection assays

PCR was used to generate two fragments of the promoter of the cER gene. These fragments were

subcloned into a repaired HindIII site immediately upstream of the CAT gene in the promoterless plasmid, pBLCAT3. pBL232 contained a 232 bp fragment, extending between oligonucleotides A ($_{-1}$AACTTTT-GAGAACTCCTC$_{-18}$) and B ($_{-232}$TTTCAGCGT-CCTTTCCC$_{-216}$), while pBL995 contained the 995 bp fragment between oligonucleotides A and C ($_{-995}$GAAGGTTTTCATTTAGC$_{-979}$).

Primary culture CEF cells were passaged in 1:1 ratio of Coons F-12/DMEM, without phenol red, with 10% foetal calf serum. For transfection, each 10 cm dish of cells was trypsinized and recovered in 0.5 ml electroporation buffer (phosphate buffered saline, 0.1% glucose and 10 μg/ml Biobrene (Applied Biosystems). This mix was added in an electroporation cuvette (Bio-Rad) to 10 μg of construct DNA and 2 μg of human β-actinhuman chorionic gonadotrophin (β-HCG, Hybritech Inc.). The β-HCG plasmid contains 4 kb of the human actin promoter spliced to the cDNA for HCG. The cells were exposed to 960 μF, 250 V current delivered by a Bio-Rad electroporation unit, recovered in 12 ml of growth medium and plated on 6 or 12 well microtitre plates, with a final density of approx. 10^6 cells per well. Hormone was added (10^{-8} M PMA and/or 10^{-5} M forskolin) 2 h after plating, and incubated for 3 days. Transiently transfected monolayers were washed twice with PBS, then covered with 200 μl of 0.2% Triton-X 100, 100 mM Tris, pH 7.8, and incubated at −70°C. After thawing, lysates were collected by gentle scraping and centrifuged at 15000 rpm for 20 min at 4°C, in eppendorf tubes. The cleared lysate was adjusted to 100 mM Tris, pH 7.8, 1 mM chloramphenicol, 8 μCi/ml tritiated acetyl Co-A (Du-Pont) in 250 μl final volume and incubated for 1 h at 37°C. Aliquots were assayed using the two-phase liquid scintillation system of Neuman *et al.* [12] and reaction products were monitored by scintillation counting in 3 ml scintillant (Beckman). β-HCG assays were carried out using a standard iodinated antibody kit (Tandem β-HCG, Hybritech inc.) using 100 μl of reserved growth medium. Transfection experiments were repeated at least four times, in duplicate.

RESULTS

Intron/exon boundaries

An initial screening of a total chicken genomic library in λ gt Wes with the probe for the cER (12.15), yielded two distinct λ clones (λA and λB). After subcloning the positively hybridizing fragments the exonic sequence within these were defined by mapping and hybridization. Appropriate fragments were sequenced in M13 and λA was shown to contain exon 1 and λB to contain exons 6, 7 and 8. An internal HindIII fragment of 12.15 (Fig. 1) which did not hybridize to these λ clones was then used to screen the same library and selected two other independent clones (C and D). Oligonucleotides positioned close to the already defined exons were used to define exonic sequences in these clones and were then used directly as primers to analyse the boundary sequences. Clone C and D were shown to contain exons 4 and 5 respectively. Finally an Msp 1 fragment of 12.15 (Fig. 1) which recognized none of these λ clones was used to screen the library and selected 1 other λ clone, E. This was shown to contain exon 3. Repeated rounds of screening with these probes and specific oligonucleotides failed to yield exon 2 from this library.

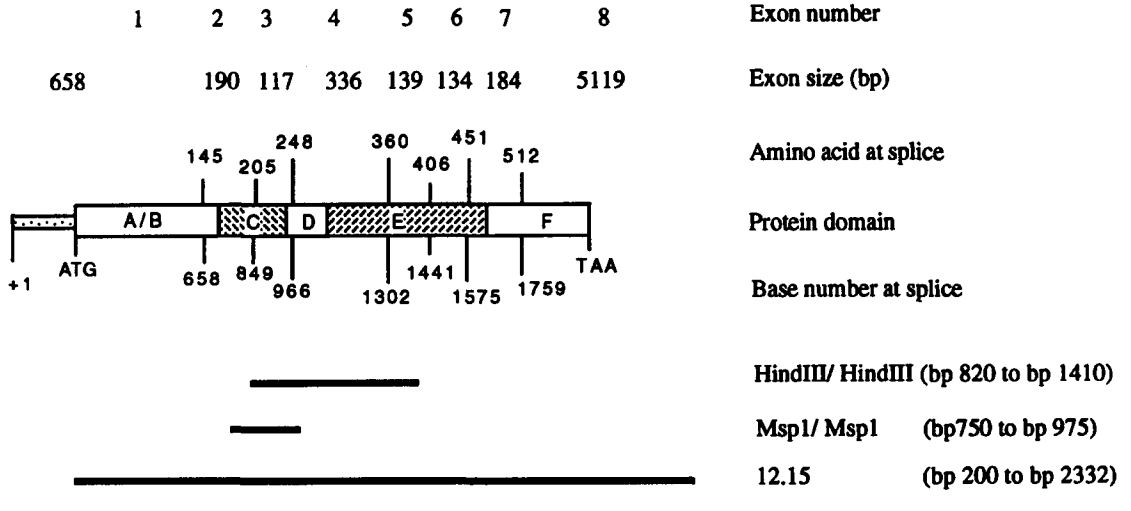

Fig. 1. Exon/intron organisation and domain structure of cER. This shows the cER mRNA from +1 to the ER open reading frame stop codon at +1992/4 and the 6 cER protein domains, A to F. The C domain is shown as a 66 amino acid core domain. The positions of the intron/exon boundaries in relation to these domains are shown. The exon size and the amino acid and base at the splice point are indicated. The heavy dark lines below gives the position of the cDNA subfragments used as probes (referred to in the text) to screen the genomic libraries.

Exon	Intron Donor		Intron Acceptor	Exon
TAC AG Tyr Ar (144)	gtaagtagggaat	-1-	ctgctgctaacag	G CCA g Pro (146)
CAA G Gln G (208)	gtaataagctgct	-2-	tctggttttacag	GG CAC ly His (210)
GGT G Gly G (247)	gtatgtacactgc	-3-	tgtgtgtctctag	GA ATC ly Ile (249)
CCA G Pro G (359)	gtaaggaaaga	-4-	tttattctttcag	GA TTT ly Phe (361)
GAC AG Asp Ar (405)	gtcagtctgtgt	-5-	ggtattcacgcag	G AAT g Asn (407)
TCT G Ser G (450)	gtagtggtgtgc	-6-	tttcctactccag	GT GTG ly Val (452)
ATG AG Met Se (511)	gtaggagcaagca	-7-	taaaacccacag	C AAC r Asn (513)
$C/_{A}AG$	$gt^{a}/_{g}agt$		$t/_{cn11}{}^{c}/_{t}ag$	G
Donor splice consensus sequences.			Acceptor splice consensus sequences.	

Fig. 2. Intron exon junctions in the cER. This shows the nucleotide sequence of the intron/exon junctions in the cER. The relationship to the amino acid sequence at the splice site is also indicated. The consensus splice sequences [13, 14] for intron junctions are shown below for comparison.

By comparison of the restriction map of the cER cDNA and the signals obtained from Southern blotting of genomic DNA, a strategy to prepare a genomic library enriched for a 1.8 kb EcoR1 fragment which contained exon 2, was devised. Following the generation of this library and screening with the Msp1 fragment probe, a clone which contained exon 2 was isolated.

Complete sequencing of subclones which contained exons coupled with the use of oligonucleotides based on exon sequence close to introns allowed the sequence of each exon and that of the intron/exon boundaries to be ascertained. The intron position in relation to the ER protein domains and boundary sequences are presented in Figs 1 and 2. At all the boundaries good agreement was found with the consensus sequences for intron donor and acceptor sequences [13, 14] (Fig. 2). Alignment of the chicken intronic sequence with that available from the human ER (M. Ponglikitmongkol and P. Chambon, personal communication) shows that the degree of conservation is not influenced by the homology within the adjacent exons and that the sequence homology decreases rapidly in all introns (data not shown).

Three discrepancies between the cDNA sequences originally published were indicated by Maxwell *et al.*[4] and we confirm these. It was also found that the sequence at the 5′ UTR suggested by these authors, though incomplete, was in agreement with our results (Fig. 3) and shows two differences from the earlier report [3] (indicated with asterisks in Fig. 3).

Definition of the start point of transcription

A prerequisite to the identification of the promoter is the definition of the start point of transcription. Both Krust *et al.*[3] and Maxwell *et al.*[4] had indicated this site on the basis of the sequence of the longest cDNA they had isolated. To obtain a more exact start site of transcription we used an oligonucleotide (pExt) which is complementary to sequence towards the 5′ end of exon 1 (Fig. 3 and legend). When used as a primer with chicken oviduct RNA one strong band was found. No bands were visible using liver total RNA as a control, indicating that pExt is specific to cER messages. The use of pExt is a sequencing reaction, using a genomic fragment subcloned in M13 covering this area as a template, allowed the exact starting nucleotide to be determined. This shows that the transcription start site

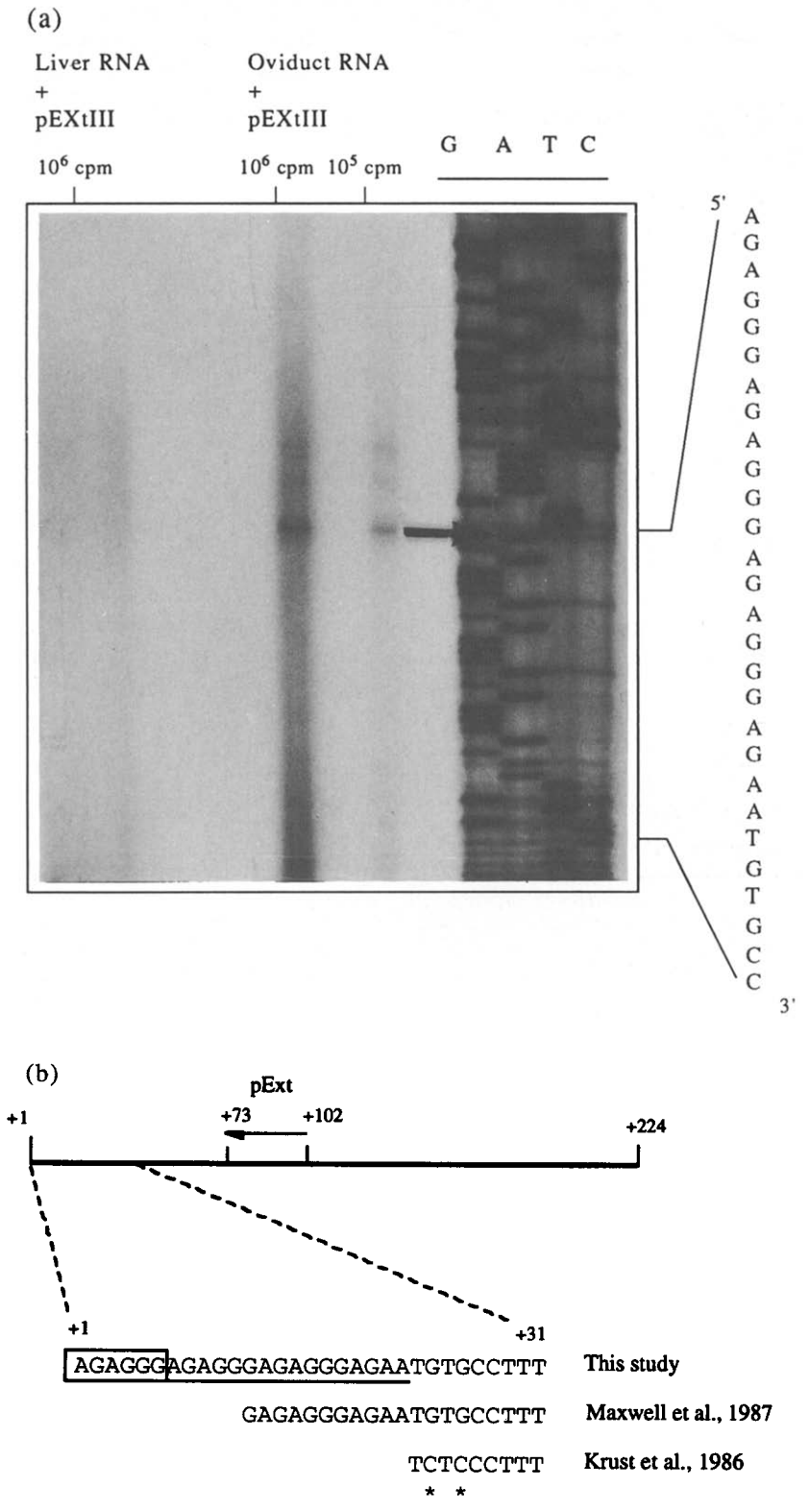

Fig. 3. Transcription start site of the cER and 5′ UTR sequence alignment. (a) shows the result of primer extension using pEXt oligonucleotide and oviduct or liver RNA. The A nucleotide arrowed is at +1 and the sequence from +1 to +28 is shown alongside. DNA sequence is derived from a genomic clone covering this area in M13 using pEXt as sequencing primer. (b) The position of the pEXt (5′GCAGCAATACATAGGTATTC-CTCTGATCTC3′) oligonucleotide used for primer extension is shown. The sequence up to the transcription start site is aligned to that previously published [3, 4] and the asterisks are used to indicate sequence differences. The polypurinic sequence (containing the AGAGGG repeat shown boxed) referred to in the text is underlined.

is 11 nucleotides further 5′ than originally thought with a consequent extension of the 5′ untranslated region sequence (Fig. 3).

Sequence of the 5′ flanking region

3000 bp of sequence was generated from the 5′ flanking region of the cER gene. Analysis of this sequence showed a well positioned, typical "TATA box" (−31 to −27 bp) and potential "CAAT box" sequences (−133 to −138 bp and −149 to −144 bp). Using the hexanucleotide half consensus sequence for an ERE or a G/PRE there were a number of perfect matches but searches for complete consensus hormone response elements failed to reveal any. The position of the best matched sequences with consensus cAMP (TGACGTCA) are at −238/35 (TGATTTCA) and −124/17 (TGACTCCA). The 5′ flanking sequence is not GC rich (~52% over 500 bp) and doesn't contain any long tracts of any nucleotide combination.

Functional assays

To analyse the functionality of the proposed cER promoter two subfragments 5′ to the start of transcription (−1 to 232 bp and −1 to 995 bp), were subcloned in the expression vector pBLCAT3 (called pBL232 and pBL995, respectively) and CAT activity was assayed in CEF cells. CAT levels were compared to that from pUC, and pBLCAT3 plasmids without an insert. pBL232 gave a 6-fold increase in activity (compare lanes 2 and 10 in Fig. 5) in the absence of any added inducers. This activity was potentiated to even higher levels (~50 fold pBLCAT3 alone, see lanes 2 and 11) on the addition of forskolin. Forskolin is used *in vitro* to stimulate cAMP production thereby activating the catalytic subunit of protein kinase A. CAT activity for pBL232 was not influenced however by the addition of the phorbol ester, PMA (which is used to activate protein kinase C) and it did not stimulate or inhibit the activation due to forskolin alone (lanes 12 and 13 in Fig. 5).

A parallel set of experiments with the pBL995 gave radically different results. CAT activity was close to background levels (compare lanes 6 to 9 with lanes 10 to 13 in Fig. 5) irrespective of the inducer added. This seems to imply a strong silencing activity in the region −232 to −995 bp.

DISCUSSION

The genomic organization for the hER [15], rainbow trout (rt) ER [16], hRARα [17] (incomplete), hRARγ [18], hAR [19–21], hGR [22, 23], mGR [24], cPR [25], hPR [26], c-erbA (cTRα) [27], c-erbA-1 (hTRα or THRA) [28], *Xenopus laevis* (Xl) TRβ [29] and hVitD (incomplete) [30] receptors, the hNGF1-B [31], COUP-TF [32], as well as the drosophila DHR3 [33] and EcR [34], genes have been previously determined. There is complete conservation of intron position and number between the cER, cPR, hPR, hER and hAR, i.e. 8 exons and 7 introns.

An extra intron has been described in one transcript of the hER [35] which divides the 5′ untranslated region into two exons. An intron in a similar position is also noted in the mER [36], the hGR [22, 23], the hVitD receptor [30], c-erbA-1 [28] and RAR [17, 18, 37, 38, 39, 40] genes. Analysis of sequence upstream of the hER has positioned this 5′ most untranslated exon at approx. −2 kb upstream from the start site of transcription of the first transcript. Having analysed 3.0 kb of sequence in the cER 5′ flanking region in this study a similar exon has not been identified in the cER.

Region A/B of the cER extending to the boundary of the DNA binding region at aa 173, is over 70% conserved with the hER and is encoded in 1 exon up to aa 145 (Fig. 1). This contrasts to the hRAR and TR genes where the intron divides the A/B region at exactly the N terminal end of the B domain, thereby delineating each domain into separate exons. A further variation amongst the vertebrates is the NGF1-B gene which has its A/B domain and CI finger encoded within one exon [31]. The A/B region varies in length amongst family members with 173 amino acids for cER, 602 for hMR, 416 for cPR, 88 for hRARα and a mere 16 for hVitD receptor. While this divergence could seem to imply a functionless first exon, this is not the case as in the hER, the A/B domain contributes as a cell and promoter specific transactivation domain with properties distinct from those described for the ER and GR hormone binding domains and elsewhere ([41] for discussion).

The DNA binding domain C, of the cER (66 amino acid core from aa 174–239) (Fig. 1) is 100% conserved between chicken and human and is encoded by two exons. It is an area of significant divergence of intron splice site positioning within this family of genes with at least five variations noted (see [42] for discussion) and in drosophila DHR3 and EcR genes there is no intron between the CI and CII fingers [33, 34]. This division of this domain is consistent with the model of the superfamily's evolutionary lineage recently proposed by Laudet *et al.*[42] and highlights further the structural and functional difference between the zinc fingers.

Exon 4 encodes one of the less well conserved areas, the D domain (Fig. 1). This is very hydrophillic being composed largely of arginine and lysine residues and may, as a result, be exposed at the surface of the protein. Interestingly in the cER this exon contains 9 amino acids of the DNA binding domain (aa 248 to 256) and this intron position is the most highly conserved amongst all the members of the superfamily. In the rabbit progesterone [43] and rat glucocorticoid [44] receptor this region encodes a constitutive nuclear localization signal (NLS) and also a domain associated with interactions with Hsp90 [45]. This strict intron

position conservation reinforces the intolerance of mutation in this area and highlights the importance of, and the conservation of, the functional domains to either side.

Exon 4 also contains part of the hormone binding domain (aa 296–359) which encodes the $\tau 2$ transactivation domain in the GR. An extra intron, positioned at the beginning of this region in the c-erbA [27], c-erbA-1 (hTRα or THRA) [28] and XlTRβ [29] genes delineates the E domain. This intron is also present in the trout ER [16], but is not found in the cER. This would indicate that this feature may have been lost from the cER and hER genes early in evolution and reaffirms the evolutionary relatedness of this family of genes.

The highly conserved E domain (aa 296–546) (Fig. 1) is encoded by 5 exons (4–8). The ligand binding activity of all receptors, the major ligand inducible transactivation function [41] in the hER (and presumably the cER), a second nuclear localization signal (NLS) of the GR [44] as well as the area of interaction with HSp 90 [45] encoded with this domain are almost certainly not confined to one particular exon but composed of amino acids from different exons appropriately juxtaposed after hormone binding, dimerisation or other intracellular events [46].

In strict contrast to these, the dimerization domain has been well defined as an area of heptad repeated hydrophobic amino acids in the mER [47, 48] and a similar candidate repeated structure is present in the cER and hER. The N terminal portion of this domain is thought to form part of the dimer interface while the C terminal part may be responsible for giving the complex stability. It is interesting to note that the intron position in the ER genes separates these functional subdomains (exons 7 and 8 in the cER). This region has also been shown to be intimately involved in binding ligand in the mER [48]. Furthermore, it has been implicated in heterodimerization between the RAR and TR [49]. The very high conservation of intron position in this region, amongst this superfamily reflects further its multifunctional importance.

Finally, exon 8 extends from amino acid 541 through the translation stop codon to the end of the 3′ UTR (our results, manuscript in preparation). The 3′ UTR encoded as one large exon is a recurring theme in the steroid hormone receptor family.

Analysis of 5′ flanking sequences

To explore the potential regulatory signals which control the transcription of the cER, 3.0 kb of 5′ flanking sequence was generated. The sequence of the 5′ end of the mRNA transcribed from this gene had previously been published [3, 4], but as can be seen from Fig. 3 there are differences in sequence and length between the two original publications. Primer extension indicated that the transcription start site was another 11 bp more 5′ to the most 5′ nucleotide published by Maxwell *et al.* [4] which extends the 5′ UTR to 224 bp. Chicken liver RNA was used in parallel reactions as a negative control and gave no visible bands even after prolonged exposure. To confirm that a mRNA existed with this start point, a chicken oviduct cDNA library in λ Zap (Stratagene) was screened with a 250 bp fragment at the 5′ end of 12.15. This yielded a clone which extended to the new predicted start point (data not shown).

Interestingly the extended sequence at the start of the 5′ UTR makes up a stretch of 22 purines with an AGAGGG repeat (underlined in Fig. 3). Similar (albeit longer) sequences have been implicated in unusual DNA structures referred to as the "H form" of DNA sometimes associated with the control regions of genes. A similar sequence is not present at the cap site of other receptor genes and its significance and function in the cER, if any is unknown [50].

Analysis of the 5′ flanking sequence shows a well positioned potential TATA box and CAAT box sequences. Alignment of human and chicken ER 5′ flanking sequences shows extensive homology over two stretches in particular (-405 to -290 and -157 to -1, both shown boxed in Fig. 4) separated by areas of intermediate homology. When the mER is included in these promoter alignments the overall homology is reduced as the mER promoter sequence is derived from the 5′ flanking region of a transcript which is homologous to transcript 2 of the hER [35, 36]. A more valid alignment would involve the region at the 3′ end of intron 1 in the mouse, but this is not available.

When the cER and hER 5′ flanking regions are compared, their transcription start sites do not align and consequently candidate TATA box sequences [51] are not at the same position (Fig. 4). These genes may have evolved different transcription start sites and different sequences and positions for "TATA factor" binding or alternatively the start site currently ascribed for the hER [51] is incorrect. In the latter case the TATGA sequence aligning to the chicken TATTA sequence is in the appropriate position. It is unlikely that this TATGA sequence is active with the current hER + 1 as it is positioned at $-46/42$ and this would be an unusual distance to the Cap site.

Although there is noticeable conservation of modular structure in the receptor proteins, a striking feature of receptor promoters characterized to date is the distinct lack of conserved features. Presumably this reflects the different parameters influencing each receptor in the many environments in which they are active and the means of achieving a differential response. The human and mouse RAR$\beta 2$ genes and the cER are distinctive in that they show good candidate TATA motif sequences. A survey of the promoters for other members of this family of genes shows that candidate TATA and CAAT sequences have not been identified for many others including the cPR [25], rat AR [52], hAR [53], hRAR$\alpha 1$ [17], mRAR$\alpha 1$ [38], mRAR$\alpha 2$ [54] or the

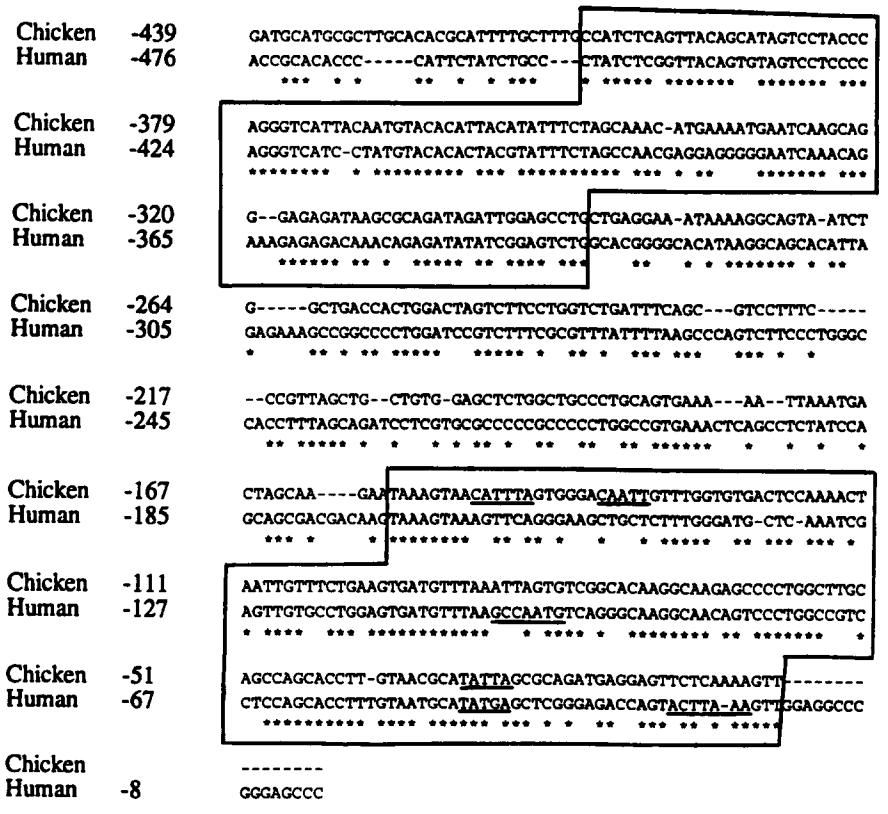

Fig. 4. Alignment of the cER and hER 5′ flanking regions. This shown an alignment of the cER and hER flanking regions. Dashes were introduced to improve the alignment and the asterisks represent conserved sequence. The areas showing highest homology and referred to in the text are boxed. The candidate TATA and CAAT sequences are underlined. TATA and CAAT sequences for the hER are taken from Green *et al.* [51]. The human sequence (TATGA) aligning to the chicken TATA box is also shown underlined.

hGR [22, 23], although InR elements have in some cases (hRARα1 and mRARα1). Recent reports using the hPR to transactivate genes showed that, in the absence of a TATA sequence, transcription was not inhibited but the correct transcription start site was not always used [55]. This may explain the heterogeneity of the transcription start sites described in many cases. The functional implications of this are not known.

The complex regulatory interactions between different hormones might suggest a direct interaction between the regulatory region of the gene and the receptor encoded by that gene. Analysis of the sequence of the cER 5′ flanking region failed to reveal any perfect consensus sequence for hormone response elements. Several half EREs were found but examination of the sequence on either side revealed no resemblance to the expected dyad symmetrical TGACCT spaced 3 nucleotides away. This was equally true with G/PREs. This lack of complete hormone response elements is consistent with most other members of this family of genes. Notable exceptions are the hRARβ [56], mRARα2 [54] and -β2 [57] genes. These have a complete functional retinoic acid response element in their promoters and have been shown to be responsive to retinoic acid (RA). The failure to identify response elements may have indicated simply that not enough sequence data was available to identify hormone response elements (HREs) active over longer stretches of DNA. Alternatively these incomplete HREs could be active individually or cooperatively with reiterated copies as has been demonstrated in the chicken ovalbumin gene [58, 59].

Functional analysis

This is the first report of a functional ER promoter and as CEF cells are thought not to express ER, it would support the idea that ER is not directly involved in its own regulation, in this species at least. Instead these studies have shown an indirect mechanism for the regulation of ER levels mediated in part at least by a phosphorylated dependent protein. This clearly categorizes the cER, amongst the family of cAMP inducible genes. This is supported by the recent elucidation of a functional CREB element in the promoter for the RARβ2 gene which enhances RA dependent activation [60]. Although there are two imperfect candidate cAMP response elements close to the start of transcription, their involvement has not yet been shown. While it is noted that the pBLCAT3 plasmid has shown some response to forskolin previously, there

is very little induction in CEF cells (compare lanes 2 and 3 in Fig. 5), which would seem to exclude this as a source of the induction seen here.

The lack of sensitivity to added phorbal esters implies a lack of involvement of a protein kinase C dependent transcription factor, and as there was no net increase in activity on the addition of PMA and forskolin in combination it would seem there is no direct cooperativity between these kinases or interaction between transcription factors stimulated by them.

The inclusion of 5′ flanking sequence more distal to the cap site in promoter constructs, completely abolished the promoter activity and forskolin inducibility. This indicated a strong silencing activity between bp −232 and −995. Having examined a panel of known silencing sequences one candidate matching sequence was identified which showed similarity to a silencer characterized in the human IL-3 gene [61]. However this sequence is at bp −194 to −186 in the cER and is not contained within the larger construct displaying the silencing activity. This may suggest a novel sequence mediating this activity. It is reasonable to speculate that the region between −277 and −407, which shows even higher sequence homology (>85% over the 150 bp shown boxed in Fig. 4) between the cER and hER than the area proximal to the cap site, harbours important functional elements involved in the regulation of this gene. Extensive studies are now underway on this promoter to define fully the elements mediating the response to cAMP and to characterize the silencing elements. Parallel studies with the hER promoter will demonstrate the degree of conservation of the functional elements in these genes.

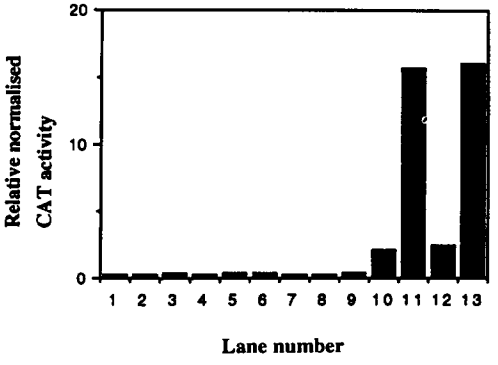

Fig. 5. Transfection in CEF cells. This shows the relative normalized CAT activity of pBL332 and pBL995 after the addition of forskolin and/or PMA compared to the plasmid pUC or the CAT vector pBLCAT3 without an insert. Lane 1, pUC; Lane 2, pBLCAT3, no added inducer; Lane 3, pBLCAT3 + forskolin; Lane 4, pBLCAT3 + PMA; Lane 5, pBLCAT3 + PMA + forskolin; Lane 6, pBL995, no added inducer; Lane 7, pBL995 + forskolin; Lane 8, pBL995 + PMA; Lane 9, pBL995 + PMA + forskolin; Lane 10, pBL232, no added inducer; Lane 11, pBL232 + forskolin; Lane 12, pBL232 + PMA; and Lane 13, pBL232 + PMA + forskolin.

Acknowledgements—The authors would like to acknowledge the financial assistance of the Cancer Research Advancement Board, Eolas and the Health Research Board of Ireland.

REFERENCES

1. Gronemeyer H.: Transcription activation by estrogen and progesterone receptors. *Ann. Rev. Genet.* 25 (1991) 89–123.
2. Miesfeld R. L.: The structure and function of steroid receptor proteins. *Crit. Rev. Biochem. Molec. Biol.* 24 (1989) 101–117.
3. Krust A., Green S., Argos P., Kumar V., Walter P., Bornert J. M. and Chambon P.: The chicken oestrogen receptor sequence: homology with verbA and the human oestrogen and glucocorticoid receptors. *EMBO J* 5 (1986) 891–897.
4. Maxwell B. L., McDonnell D. P., Conneely O. M., Schulz T. Z., Greene G. L. and O'Malley B. W.: Structural organization and regulation of the chicken estrogen receptor. *Molec. Endocr.* 1 (1987) 25–35.
5. Read L. D., Greene G. L. and Katzenellenbogen B. S.: Regulation of estrogen receptor messenger ribonucleic acid and protein levels in human breast cancer cell lines by sex steroid hormones, their antagonists and growth factors. *Molec. Endocr.* 3 (1989) 295–304.
6. Shapiro D. J., Barton M. C., McKearin D. M., Chang T-C., Lew D., Blume J., Nielsen D. A. and Gould L.: Estrogen regulation of gene transcription and mRNA stability. *Rec. Prog. Hormone Res.* 45 (1989) 29–64.
7. Miller W. R.: Endocrine treatment for breast cancers: biological rationale and current progress. *J. Steroid Biochem. Molec. Biol.* 37 (1990) 467–480.
8. Feinberg A. P. and Vogelstein B.: A technique for radiolabelling DNA fragments to high specific activity. *Analyt. Biochem.* 132 (1983) 6–13.
9. Maniatis T., Fritsch E. F. and Sambrook J.: *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.
10. Chirgwin J. J., Przybyla A. E., MacDonald R. J. and Rutter W. J.: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18 (1979) 5294–5299.
11. Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Smith J. A., Seidman J. G. and Struhl K. (Eds): *Current Protocols in Molecular Biology.* Greene Publishing Associates and Wiley—Interscience. Massachusetts General Hospital and Harvard Medical School.
12. Neumann J. R., Morency C. A. and Russian K. O.: A novel rapid assay for chloramphenicol acetyltransferase gene expression. *Biotechniques* 5 (1987) 444–447.
13. Mount S. M.: A catalogue of splice junction sequences. *Nucl. Acid Res.* 10 (1982) 459–472.
14. Shapiro M. B. and Senepathy P.: RNA splice junctions of different classes of eucaryotes: statistics and functional implications in gene expression. *Nucl. Acid Res.* 15 (1987) 7155–7174.
15. Ponglikitmongkol M., Green S. and Chambon P.: Genomic organisation of the human oestrogen receptor gene. *EMBO J.* 7 (1988) 3385–3388.
16. Le Roux M-G., Thézé N., Wolff J. and Le Pennec J-P.: Organisation of a rainbow trout estrogen receptor gene. *Biochem. Biophys. Acta* 1172 (1993) 226–230.
17. Brand N. J., Petkovich M. and Chambon P.: Characterization of a functional promoter for the human retinoic acid receptor—alpha (hRARα). *Nucl. Acid Res.* 18 (1990) 6799–6806.
18. Lehmann J. M., Hoffman B. and Pfahl M.: Genomic organization of the retinoic acid receptor gamma gene. *Nucl. Acid Res.* 19 (1991) 573–578.
19. Lubahn D. B., Brown T. R., Simental J. A., Higgs H. N., Migeon C. J., Wilson E. M. and French F. S.: Sequence of the intron/exon junctions of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity. *Proc. Natn. Acad. Sci. U.S.A.* 86 (1989) 9534–9538.
20. Brinkmann A. O., Faber P. W., van Rooij H. C. J., Kuiper G. G. J. M., Ris C., Klaasen P., van der Korput J. A. G. M., Voorhorst M. M., van Laar J. H., Mulder E. and Trapman J.: The human androgen receptor: domain structure, genomic organisation and regulation of expression. *J. Steroid Biochem.* 34 (1989) 307–310.

21. Kuiper G. G. J. M., Faber P. W., van Rooij H. C. J., van der Korput J. A. G. M., Ris-Stalpers C., Klaasen P., Trapman J. and Brinkmann A. O.: Structural organisation of the human androgen receptor gene. *J. Molec. Endocr.* **2** (1989) R1–R4.
22. Encio I. J. and Detera—Wadleigh S. D.: The genomic structure of the human glucocorticoid receptor. *J. Biol. Chem.* **266** (1991) 7182–7188.
23. Zong J., Ashraf J. and Thompson E. B.: The promoter and first, untranslated exon of the human glucocorticoid receptor gene are GC rich but lack consensus glucocorticoid receptor element sites. *Molec. Cell Biol.* **10** (1990) 5580–5585.
24. Strähle U., Schmidt A., Kelsey G., Stewart A. F., Cole T. J., Schmid W. and Schütz G.: At least three promoters direct expression of the mouse glucocorticoid receptor gene. *Proc. Natn. Acad. Sci. U.S.A.* **89** (1992) 6731–6735.
25. Huckaby C. S., Conneely O. M., Beattie W. G., Dobson A. D. W., Tsai M. J. and O'Malley B. W.: Structure of the chromosomal chicken progesterone receptor gene. *Proc. Natn. Acad. Sci. U.S.A.* **84** (1987) 8380–8384.
26. Misrahi M., Venencie P. Y., Saugierveber P., Sar S., Dessen P. and Milgrom E.: Structure of the human progesterone receptor gene. *Biochem. Biophys. Acta* **1216** (1993) 289–292.
27. Zahraoui A. and Cuny G.: Nucleotide sequence of the chicken proto-oncogene c-erb A corresponding to domain 1 or v-erbA. *J. Biochem.* **166** (1987) 63–69.
28. Laudet V., Begue A., Henry-Duthoit C., Joubel A., Martin P., Stehelin D. and Saule S.: Genomic organization of the human thyroid hormone receptor α (cerbA-1) gene. *Nucl. Acids Res.* **19** (1991) 1105–1112.
29. Shi Y-B., Yaoita Y. and Brown D. D.: Genomic organization and alternative promoter usage of the two thyroid hormone receptor β genes in *Xenopus laevis*. *J. Biol. Chem.* **267** (1992) 733–738.
30. Hughes M. R., Malloy P. J., Kieback D. G., Kesterson R. A., Pike J. W., Feldman D. and O'Malley B. W.: Point mutations in the human vitamin D receptor gene associated with hypocalcemic rickets. *Science* **242** (1988) 1702–1705.
31. Watson M. A. and Milbrandt J.: The NGFI-B gene, a transcriptionally inducible member of the steroid receptor gene superfamily: genomic structure and expression in rat brain after seizure induction. *Molec. Cell Biol.* **9** (1989) 4213–4219.
32. Ritchie H. H., Wang L-H., Tsai S., O'Malley B. W. and Tsai M-J.: COUP-TF gene: a structure unique for steroid/thyroid receptor superfamily. *Nucl. Acids Res.* **18** (1990) 6857–6862.
33. Koelle M. R., Seagraves W. A. and Hogness D. S.: DHR3: A *Drosophila* steroid receptor homolog. *Proc. Natn. Acad. Sci. U.S.A.* **89** (1992) 6167–71.
34. Koelle M. R., Talbot W. S., Seagraves W. A., Bender M. T., Cherbas P. and Hogness D. S.: The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67** (1991) 59–77.
35. Keaveney M., Klug J., Dawson M. T., Nestor P. V., Neilan J. G., Forde R. C. and Gannon F.: Evidence for a previously unidentified upstream exon in the human oestrogen receptor gene. *J. Molec. Endocr.* **6** (1991) 111–115.
36. White R., Lees J. A., Needham M., Ham J. and Parker M.: Structural organization and expression of the mouse estrogen receptor. *Molec. Endocr.* **1** (1987) 735–744.
37. Krust A., Kastner Ph., Petkovich M., Zelent A. and Chambon P.: A third human retinoic acid receptor, hRAR-γ. *Proc. Natn. Acad. Sci. U.S.A.* **86** (1989) 5310–5314.
38. Leroy P., Krust A., Zelent A., Mendelsohn C., Garnier J-M., Kastner Ph., Dierich A. and Chambon P.: Multiple isoforms of the mouse retinoic acid receptor α are generated by alternative splicing and differential induction by retinoic acid. *EMBO J.* **10** (1991) 59–69.
39. Zelent A., Mendelsohn C., Kastner P., Krust A., Garnier J-M., Ruffenach F., Leroy P. and Chambon P.: Differentially expressed isoforms of the mouse retinoic acid receptor β are generated by usage of two promoters and alternative splicing. *EMBO J.* **10** (1991) 71–81.
40. Kastner Ph., Krust A., Mendelsohn C., Garnier J-M., Zelent A., Leroy P., Staub A. and Chambon P.: Murine isoforms of retinoic acid receptor γ with specific patterns of expression. *Proc. Natn. Acad. Sci. U.S.A.* **87** (1990) 2700–2704.
41. Bocquel M. T., Kumar V., Stricker C., Chambon P. and Gronemeyer H.: The contribution of the N- and C-terminal regions of steroid receptors to activation of transcription is both receptor and cell specific. *Nucl. Acids Res.* **17** (1989) 2581–2595.
42. Laudet V., Hänni C, Coll J., Catzeflis F. and Stéhelin C.: Evolution of the nuclear receptor gene family. *EMBO J.* **11** (1992) 1003–1013.
43. Guiochon-Mantel A., Loosfelt H., Lescop P., Sar S., Atger M., Perrot-Applanat M. and Milgrom E.: Mechanisms of nuclear localization of the progesterone receptor: evidence for interaction between monomers. *Cell* **57** (1989) 1147–1154.
44. Picard D. and Yamamoto K. R.: Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *EMBO J.* **6** (1987) 3333–3340.
45. Chambraud B., Berry M., Redeuilh G., Chambon P. and Baulieu E-E.: Several regions of the human estrogen receptor are involved in the formation of receptor—heat shock protein 90 complexes. *J. Biol. Chem.* **265** (1990) 20,686–20,691.
46. Webster N. J. G., Green S., Tasset D., Ponglikitmongkol M. and Chambon P.: The transcriptional activation function located in the hormone—binding domain of the human oestrogen receptor is not encoded in a single exon. *EMBO J.* **8** (1989) 1441–1446.
47. Fawell S. E., Lees J. A., White R. and Parker M. G.: Characterization and co-localization of steroid binding and dimerization activities in the mouse estrogen receptor. *Cell* **60** (1990) 953–962.
48. Lees J. A., Fawell S. E., White R. and Parker M. G.: A 22 amino-acid peptide restores DNA—binding activity to dimerization—defective mutants of the estrogen receptor. *Molec. Cell Biol.* **10** (1990) 5529–5531.
49. Spanjaard R. A., Darling D. S. and Chin W. W.: Ligand-binding and heterodimerization activities of a conserved region in the ligand-binding domain of the thyroid hormone receptor. *Proc. Natn. Acad. Sci. U.S.A.* **88** (1991) 8587–8591.
50. Michel D., Chatelain G., Herault Y. and Brun G.: The long repetitive polypurine/polypyrimidine sequence $(TTCCC)_{48}$ forms DNA triplex with PU-PU-Py base triplets *in vivo*.*Nucl. Acid Res.* **20** (1992) 439–443.
51. Green S., Walter P., Kumar V., Krust A., Bornert J-M., Argos P. and Chambon P.: Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* **320** (1986) 134–139.
52. Baarends W. M., Themmen A. P. N., Blok L. J., Mackenbach P., Brinkmann A. O., Meijer D., Faber P. W., Trapman J. and Grootegoed J. A.: The rat androgen receptor gene promoter. *Molec. Cell Endocr.* **74** (1990) 75–84.
53. Tilley W. D., Marcelli M. and McPhaul M. J.: Expression of the human androgen receptor gene utilizes a common promoter in diverse human tissues and cell lines. *J. Biol. Chem.* **265** (1990) 13,776–13,781.
54. Leroy P., Nakshatri H. and Chambon P.: Mouse retinoic acid receptor α 2 isoform is transcribed from a promoter that contains a retinoic acid response element. *Proc. Natn. Acad. Sci. U.S.A.* **88** (1991) 10,138–10,142.
55. Thomson A. A., Ham J., Bakker O., Parker M. G.: The progesterone receptor can regulate transcription in the absence of a functional TATA box element. *J. Biol. Chem.* **265** (1990) 16,709–16,712.
56. de Thé H., del Mar Vivanco-Ruiz M., Tiollais P., Stunnenberg H. and Dejean A.: Identification of a retinoic acid responsive element in the retinoic acid receptor β gene. *Nature* **343** (1990) 177–180.
57. Sucov H. M., Murakami K. K. and Evans R. M.: Characterization of an autoregulated response element in the mouse retinoic acid receptor type β gene. *Proc. Natn. Acad. Sci. U.S.A.* **87** (1990) 5392–5396.
58. Tora L., Gaub M.-P., Mader S., Dierich A., Bellard M. and Chambon P.: Cell specific activity of a GGTCA half-palindromic oestrogen-responsive element in the chicken ovalbumin gene promoter. *EMBO J.* **7** (1988) 3771–3778.
59. Kato S., Tora L., Yamauchi J., Masushige S., Bellard M. and Chambon P.: A far upstream estrogen response element of the ovalbumin gene contains several half-palindromic 5′-TGACC-3′ motifs acting synergistically. *Cell* **68** (1992) 731–742.
60. Kruyt F. A. E., Folkers G., van den Brink C. E. and van der Saag P. T.: A cyclic AMP response element is involved in retinoic acid—dependent RARβ2 promoter activation. *Nucl. Acid Res.* **20** (1992) 6393–6399.
61. Mathey-Prevot B., Andrews N. C., Murphy H. S., Kreissman S. G. and Nathan D. G.: Positive and negative elements regulate human interleukin-3 expression. *Proc. Natn. Acad. Sci. U.S.A.* **87** (1990) 5046–5050.