Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8380–8384, December 1987
Biochemistry

Structure of the chromosomal chicken progesterone receptor gene

CLARK S. HUCKABY, ORLA M. CONNEELY, WANDA G. BEATTIE, ALAN D. W. DOBSON, MING-JER TSAI, AND BERT W. O'MALLEY

Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030

Communicated by Klaus Hofmann, August 20, 1987 (received for review June 15, 1987)

ABSTRACT **We have isolated cosmid clones containing the chromosomal chicken progesterone receptor gene. The gene consists of eight exons and is ≈38 kilobases long. Individual exons correlate well with conserved functional domains of the receptor molecule. Alternative polyadenylylation in the second intron results in a putative non-hormone-binding protein. The cap site of the gene is heterogeneous over at least 14 base pairs and lies in a very G+C-rich region. The promoter lacks "TATA" and "CAAT" boxes, but CCGCCC motifs exist in the surrounding region.**

Many steroid hormone receptors have recently been cloned and analyzed at the cDNA sequence level (1–12). Specific regions of these molecules share extensive homologies (2–12). Functional analysis of some receptors has implicated the conserved regions in DNA binding and in steroid binding (13–18). A resulting concept is that steroid receptors are a family of evolutionarily related ligand-dependent factors that are important in the trans-activation of specific genes (19).

To better understand the evolution of steroid receptors in eukaryotes and to facilitate analysis of their transcriptional regulation, chromosomal genes must be characterized. In this paper we report the structure of a chromosomal gene for a steroid receptor. We have now isolated overlapping cosmid clones that contain the 38-kilobase (kb) transcription unit of the chicken progesterone receptor (PR) gene. Conserved functional domains are encoded by eight distinct exons. Transcription initiation sites spanning at least 14 base pairs (bp) are centered about 360 bp upstream from the first translation initiation codon of the first exon; the putative promoter region lacks a "TATA" box homology and is G+C rich.

MATERIALS AND METHODS

Materials. Enzymes were from Boehringer Mannheim or New England Biolabs, and radiochemicals were from ICN. Tissue and cloned DNAs were prepared according to standard procedures (20). Total and poly(A)$^+$ RNAs were prepared as described (21).

Filter Hybridization. Nitrocellulose filter-immobilized DNA was prehybridized at 68°C either 6 hr (Southern blots) or overnight (colony screening) in 6× SSC (1× SSC = 0.15 M sodium chloride/0.15 M sodium citrate, pH 7)/2 mM EDTA/0.25% (wt/vol) nonfat dry milk (22). Filters were hybridized with ^{32}P-labeled probes under the same conditions for 16 hr and then were washed at least 90 min in several changes of 1× SSC/0.1% (wt/vol) NaDodSO$_4$ at 68°C prior to autoradiography.

Cosmids. The chicken genomic cosmid library described (23) was screened according to standard procedures (24) using the nick-translated cDNA fragments given in Fig. 1. Cosmid DNA restriction endonuclease mapping by routine procedures (25) was verified by partial restriction digestion of linearized end-labeled cosmids. Selected restriction fragments were fine-mapped following subcloning into pGEM-2 or pUC-19 vectors.

Sequencing. Restriction fragments subcloned into M13mp18 or M13mp19 were sequenced by dideoxy chain termination (26). Some fragments were sequenced by partial chemical degradation (27).

S1 Mapping and Primer Extension. Single-stranded, uniformly-labeled probes for S1 mapping and reverse-extension experiments were prepared using M13 templates essentially according to Bergsma *et al.* (28) and Burke (29). Following hybridization with RNA (see Fig. 3), 300 μl of nuclease S1 buffer (50 mM sodium acetate, pH 4.5/0.3 M NaCl/2 mM zinc acetate) containing nuclease S1 (2 units/μl) was added, and the reaction was incubated at 37°C for 1 hr. The reaction was terminated by adding 15 μl of 0.5 M Tris, pH 8.0/0.25 M EDTA and extracting twice with phenol/chloroform/isoamyl alcohol, 25:24:1; nucleic acids were ethanol precipitated, resuspended in 10 mM Tris, pH 8.0/1.0 mM EDTA, adjusted to 2.0 M ammonium acetate, and reprecipitated.

For primer extension, hybridization reactions were ethanol precipitated, resuspended in H_2O, and then adjusted to 50 mM Tris, pH 8.2/50 mM KCl/5 mM $MgCl_2$/10 mM dithiothreitol/0.5 mM each dNTP in a final volume of 50 μl; 25 units of avian myeloblastosis virus reverse transcriptase was added. The sample was incubated at 42°C for 1 hr, adjusted to 0.3 M sodium acetate/5 mM EDTA, extracted with phenol/chloroform/isoamyl alcohol, and then ethanol precipitated twice as described for S1 samples. All samples were denatured in 80% (vol/vol) formamide at 95°C for 5 min prior to analysis on 8 M urea sequencing gels.

RESULTS

Isolation and Characterization of Cosmid DNA. Nick-translated cDNAs were used to screen a chicken genomic cosmid library, which identified four positive clones. Restriction mapping indicated that the overlapping cosmids represent 83 kb of contiguous genomic DNA (Fig. 1). Hybridization analysis with cDNA probes identified a discontinuous series of restriction fragments spanning 44 kb (Figs. 1 and 2*A*). There are about 26 kb and 13 kb of nonhybridizing 5′- and 3′-flanking sequences, respectively.

The Chicken PR Is Encoded by a Single-Copy Gene. High-stringency Southern blots (43) of chicken genomic DNA with cDNA probes identified bands corresponding to all of the cDNA-positive cosmid fragments and additional *Eco*RI fragments of 5.0 and 2.3 kb (Fig. 2*B*). The 5.0-kb band was not detected in all chicken genomes tested, whereas the 2.3-kb band was reproducible. The latter band could represent a related gene or allele; it is highly unlikely that it could represent a distinct second copy of a gene of this complexity. Selected genomic subclones hybridized only to the predicted

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PR, progesterone receptor.

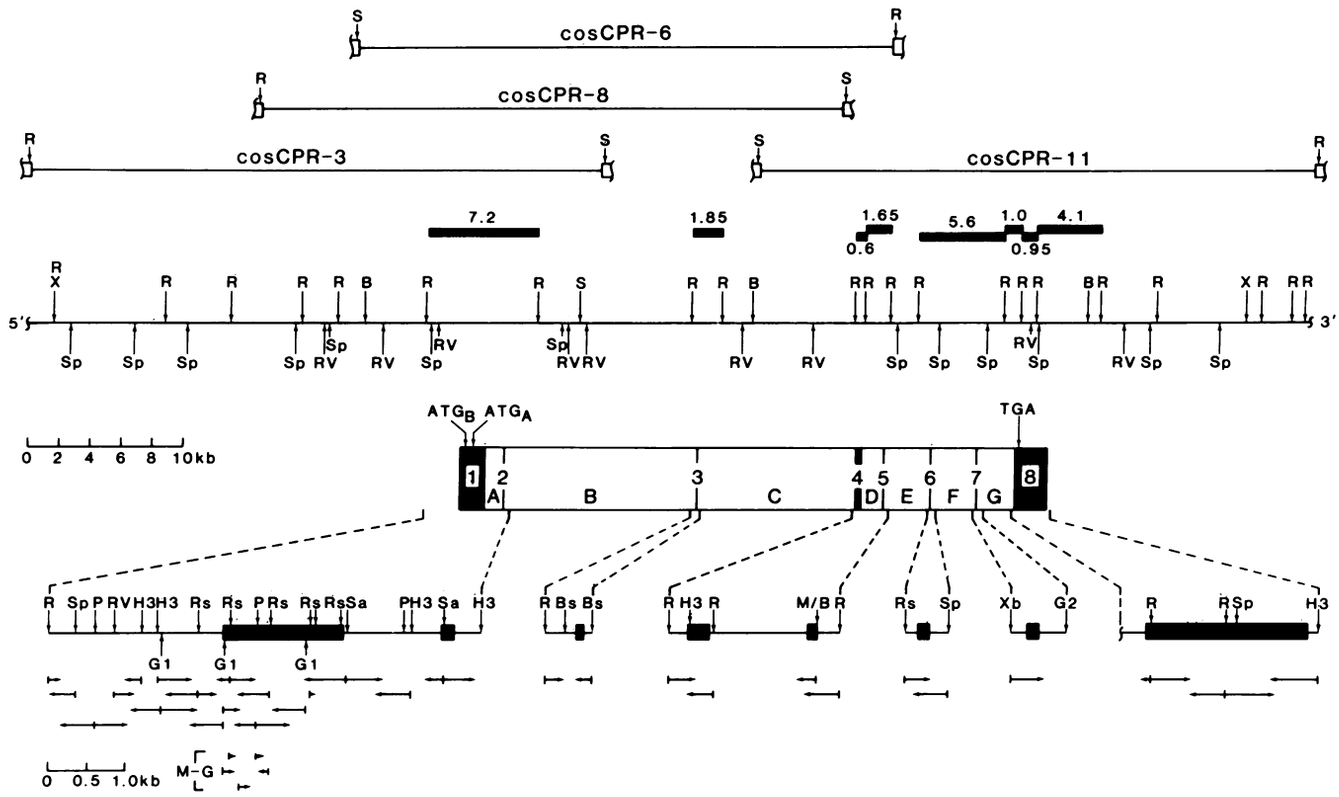

FIG. 1. Physical map and sequencing strategy of the chicken PR gene. At the top is shown the positions of cosmid clones obtained by screening with cDNA CPR-2 (ref. 7; identified clones cosCPR-3, -6, and -8) and the 3'-terminal 1.6-kb *Hin*dIII–*Bam*HI fragment of CPR-19 (ref. 12; identified clone cosCPR-11). Genomic inserts are shown as lines and vector sequences are shown as boxes. Below the cosmids is a restriction map of the region represented. The solid bars over the restriction map denote *Eco*RI fragments that hybridize to CPR-19 (see Fig. 2*A*). Below the map is shown the position of the transcription unit; exons (numbered 1–8) and introns (A–G) are indicated. Positions of the two initiation codons (ATG$_B$ and ATG$_A$) and the termination codon (TGA) are given. The sequencing strategy is presented below expanded maps of regions containing the exons; arrows depict the direction and extent of Maxam–Gilbert (M–G) or dideoxy (the remaining) sequencing experiments. Genomic restriction sites in exons that were not sequenced across here were crossed in the cDNA (12). B, *Bam*HI; Bs, *Bst*EII; G1, *Bgl* I; G2, *Bgl* II; H3, *Hin*dIII; P, *Pst* I; Pv, *Pvu* II; R, *Eco*RI; RV, *Eco*RV; Rs, *Rsa* I; S, *Sal* I; Sa, *Sac* I; Sp, *Sph* I; X, *Xho* I; Xb, *Xba* I. M/B denotes an *Mbo* I site at the 3' end of the cosCPR-6 insert that became a *Bam*HI site upon ligation into the cosmid vector.

fragments in genomic blots (Fig. 2 *C* and *D*), which provided no evidence for the presence of additional genes.

Fine Structure of the Gene. *Eco*RI or *Eco*RI–*Bam*HI cosmid fragments that hybridized to cDNA probes were subcloned into plasmid vectors for higher resolution restriction mapping and hybridization studies. Appropriate subfragments were inserted into bacteriophage M13 vectors and sequenced. The relevant portion of the sequencing strategy is given in Fig. 1. A comparison of the genomic sequence to that of the full-length cDNA (12) identified eight exons and seven introns (Fig. 1). The protein coding sequences of all exons agreed with the cDNA. The sequences of the splicing sites are given in Table 1.

Exon 1 encodes the two methionine residues that define the amino termini of the two forms of the receptor protein (ATG$_B$ and ATG$_A$ for the B and A forms, respectively; ref. 12), together with the entire nonconserved N-terminal domain of the molecule (see Fig. 1). The conserved DNA-binding domain is represented by exons 2 and 3; one putative "zinc finger" (31) is encoded in each of these exons. Exon 4 codes for the poorly conserved "hinge" region between the DNA- and hormone-binding domains, whereas exons 5, 6, and 7 and the first 154 bp of exon 8 represent the large hormone-binding domain. The termination codon and all of the long 3'-untranslated sequences are present in exon 8.

The 5' End of the Gene. S1 mapping with probes upstream from ATG$_B$ suggested that genomic sequences were represented in oviduct mRNA up to the vicinity of the *Bgl* I site located about 350 bp 5' to ATG$_B$ (data not shown). A 24-mer oligonucleotide matching the sequence of the bottom strand (178 to 201 bp downstream from the *Bgl* I cleavage site) was used to prime M13 clones and generate single-stranded, uniformly labeled probes terminating at upstream restriction sites (Fig. 3*A*). The probes were hybridized to RNA preparations and then either cleaved with nuclease S1 or extended with reverse transcriptase (Fig. 3*B*). Oviduct-specific S1-protected fragments of about 210–223 nucleotides corresponded to extension products from two separate primers and indicate that the mRNA cap sites are concentrated about 9–22 bp 5' to the *Bgl* I cleavage site. The primer extension products terminate mainly at the purine residues of the noncoding (top) strand. The principal cap site (+1) is the guanosine located 22 bp 5' of the *Bgl* I cleavage site and 365 bp 5' of ATG$_B$ (Fig. 3).

Genomic sequences flanking the cap sites are presented in Fig. 4. The region that includes all of exon 1 is very G+C rich, especially just downstream from the cap sites, where some sequences are over 80% G+C. Upstream the G+C content decreases progressively through 60% at about −200 and averages about 40% from −400 to −1000. No obvious TATA or "CAAT" boxes were found in the putative promotor region. However, five simian virus 40-like CCGCCC motifs are in the region. Only one such element is located upstream (on the bottom strand) at position −109; the remaining four are in the 5'-untranslated mRNA leader at +97 and +144 (bottom strand) and also at +174 and +178 (top strand). An unusual sequence element not shown in Fig. 4 is a polypurine

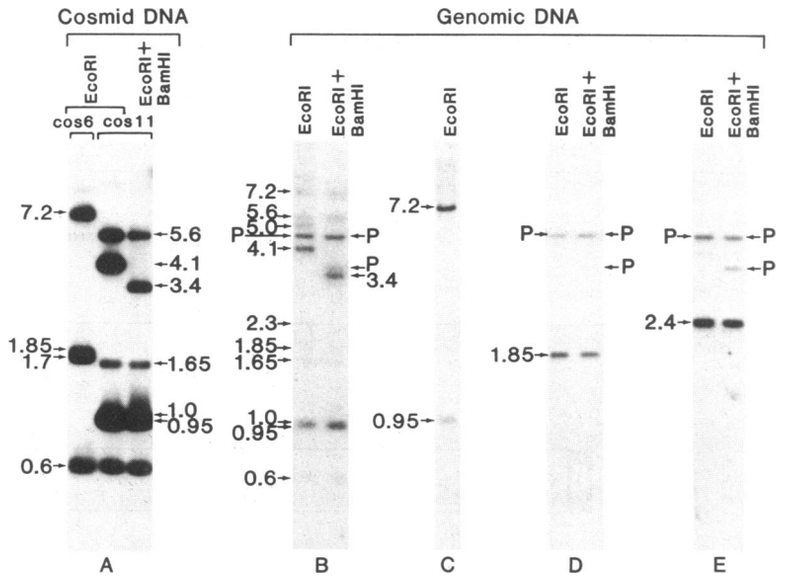

FIG. 2. Southern blot analysis of cosmid (*A*) and genomic (*B–E*) DNA. The DNA was digested with *Eco*RI or *Eco*RI and *Bam*HI (as indicated at the top), resolved on agarose gels, blotted to nitrocellulose, and then incubated with the following probes: (*A*) Plasmid-free cDNA CPR-19 (12). (*B*) CPR-19 in the plasmid. (*C*) Plasmids containing the 0.95-kb *Eco*RI fragment of cosCPR-11 (part of exon 8) and the 2.5-kb *Hin*dIII–*Sac* I subfragment of the 7.2-kb *Eco*RI cosmid fragment (exon 1 and 5′ flank). (*D*) Plasmid containing the 1.85-kb *Eco*RI fragment of cosCPR-6 (exon 3). (*E*) pOV-2.4 (30), a chicken ovalbumin subclone used as a control to identify plasmid contamination bands (P) seen in *B*, *D*, and *E*.

tract consisting of 34 consecutive direct repeats of GAGAG from −1397 to −1228.

A cDNA Variant Contains Only Exons 1 and 2. Several cDNA clones from oligo(dT)-primed libraries (12) appear to derive from mRNAs that terminated in the 5′ end of the long intron separating exons 2 and 3. They contain sequences that match those of exons 1 and 2, followed by 23 bp of intron sequence, an in-frame termination codon, and then the sequence ACTAAA 124 bp further downstream (Fig. 5). The latter sequence probably acts as an alternative polyadenylylation signal; the cDNAs terminate in long poly(A) tracts.

DISCUSSION

We have cloned and characterized the chromosomal chicken PR gene, demonstrating that (*i*) the single-copy 38-kb transcription unit consists of eight exons, (*ii*) conserved functional domains of the receptor protein are encoded by distinct exons, (*iii*) the promoter resides in a very G+C-rich region and lacks TATA and CAAT boxes, and (*iv*) transcription initiation is heterogeneous.

The overall structure of the gene is unusual. The long terminal exons contain a large amount of noncoding sequence and together comprise 77.5% of the sequence of the longest major mRNA (there are three major mRNAs: 1.8, 3.6, and 4.5 kb; see ref. 12). The shortest exon (exon 3, 117 bp) is flanked by the two largest introns (introns B and C of 12.4 and 10.2 kb, respectively).

In support of the hypothesis of Gilbert (33), division of the gene into exons is correlated with the proposed functional domains of the receptor molecule. Most strikingly, each of the two zinc fingers that have been postulated for the chicken PR DNA-binding domain (referred to as C1 in ref. 12) is encoded by a distinct small exon. Exons for the first six "fingers" of *Xenopus* transcription factor IIIA are similarly distinct (34) and suggest that such domains may duplicate and evolve as separate units. Poorly conserved polypeptide regions flanking C1 are encoded by distinct exons (exons 1

Table 1. Sequences of intron boundaries

Intron	Approximate size, kb	Splice position	Sequence		
			Exon	Intron	Exon
A	1.24	1568/1569	TAC GTG CG Tyr Val Ar	GTAAGGGCGC...GCTTTTCAAG	G CCA GAC g Pro Asp
B	12.4	1720/1721	ATG GAA G Met Glu G	GTACTATATC.....TGTAG	GG CAG CAC ly Gln His
C	10.2	1837/1838	CTG GGA G Leu Gly G	GTAGCTGCTT...TTTTTTCCAG	GT CGA AAA ly Arg Lys
D	1.26	2140/2141	CTA CCA G Leu Pro G	GTAATAGCGG...TCATTTCCAG	GA TTT CGG ly Phe Arg
E	2.9	2285/2286	CTG AAT GA Leu Asn Gl	GTAAGTAGAC...TGGGTGTAAG	A CAG AGG u Gln Arg
F	2.8	2416/2417	AAT ACA A Asn Thr I	GTAAGCTCTT...TTCTTTTCAG	TT CCT TTG le Pro Leu
G	2.5	2574/2575	CAT GAT His Asp	GTAGTTTCAG...TTAAATCCAG	CTA GTG Leu Val

The splice position given is relative to +1 as determined in Fig. 3. The position in the cDNA sequence in Conneely *et al.* (12) can be obtained by subtracting 52 from each number.

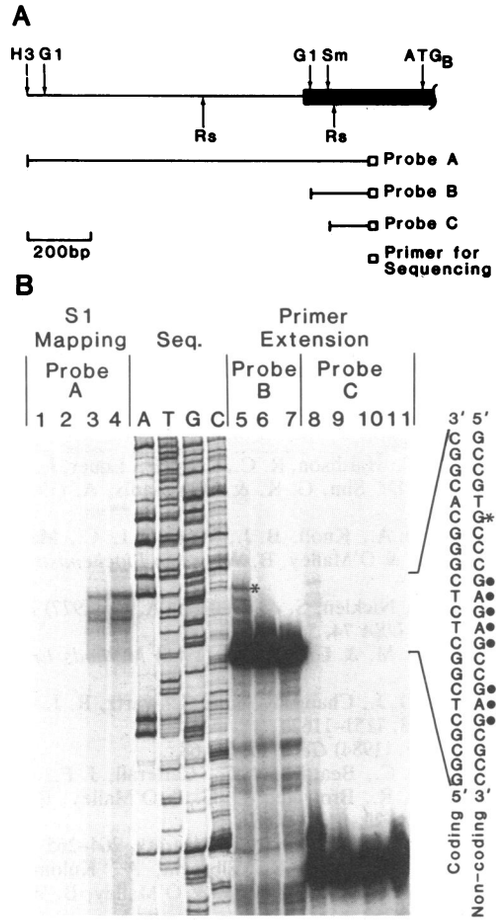

FIG. 3. Mapping the cap sites of the gene. (*A*) Restriction map of 5'-end region to show probe positions. The thick bar is exon 1. Restriction sites are as in Fig. 1 with the addition of *Sma* I (Sm). The primer for sequencing and template-directed probe synthesis was the oligonucleotide 5'-AGCGATCTGAGCGTCAGCCGTCTT-3'. Sequencing was by dideoxy chain termination with deoxyadenosine 5'-[α-^{35}S]thiotriphosphate, using an M13mp19 clone of the 2.5-kb *Hin*dIII–*Sac* I subfragment of the 7.2-kb *Eco*RI cosmid fragment. Probe A was synthesized by priming the above template, filling-in with Klenow enzyme in the presence of [^{32}P]dCTP, cleaving at the *Hin*dIII site, and isolating the probe in a strand-separating acrylamide gel. Probes B and C were prepared similarly, except the template was an M13mp18 subclone of the 410-bp *Bgl* I–*Pst* I fragment near the 5' end of exon 1; our subcloning methodology changed the *Bgl* I site to a *Sma* I site at the vector–insert junction. Probes B and C were prepared simultaneously by partial *Sma* I digestion of the filled-in template, and 2×10^5 (probe A) or 5×10^4 (probes B and C) cpm were hybridized with 20 μg of RNA (see below) for 16 hr in 20 μl of 10 mM Hepes, pH 7.0/0.6 M NaCl/2 mM EDTA/60% (vol/vol) formamide at 54°C (probe A) or 65°C (probes B and C). (*B*) Autoradiograph of sequencing gel showing products of S1 mapping, sequencing, and primer extension, as indicated. Lanes: 1 and 10, yeast tRNA; 2, 6, and 9, total hen liver RNA; 3, 5, and 8, total oviduct RNA from chickens given a 24-hr secondary estrogen treatment (32); 4, poly-(A)$^+$ RNA from hen oviduct; 7 and 11, yeast tRNA without reverse transcriptase. The sequence of both strands in the cap site region is shown at the right; an asterisk indicates the guanosine designated nucleotide +1, which corresponds to the most prominent extension product (also indicated on autoradiograph). Dots indicate nucleotides corresponding to less prominent cap sites.

and 4). The most highly conserved portion of the putative hormone-binding domain (C2) is encoded mostly by exon 5 plus about 11 codons of exon 6. The protein-coding regions of exons 5 through 8 (which represent the putative hormone-binding domain), together with their flanking introns, are relatively uniform and moderate in size as compared to the

```
-300:  TGAAAAACAC AGCAAACGTC TGAAAGAGAG GGTGAAGAAC CAGGAAGGAG
-250:  AGGACATATC TTCCTCCCCG CGCTCCGTAA CAGCGATTTG CCCAGCCTGA
-200:  GTGCAGCCAG TATTGCACTC CTCGACCAGG ACGGACACCT GCAAGAGCCG
-150:  GGAGACGGGG TGAGATGAGG CTGCGAAGGC AGGAAGCGGC CGGGCGGCTA
-100:  AGAGCGGAGC CGAGAGGGAA CGGGGAGCGG TGGGAGGGTC GGAGCAGGCG
 - 50:  GAGCGAAGGA GGGGCCAGGC GTGCCGGGAG CTGCCGGAGG GAGTCGCCGT
          *  .....  ...
   1:  GCCCGAGAGC CGAGCGCCGA AGGGGCGAGA AGGCGGCTGC CTGCCGGGGG
  51:  CAGCGGGGAA GCGGGAGGGG CGAACCCGGG CGCCTCTCAG CAGTACGGGC
 101:  GGCGGCGGGA CGGGGAGCGT CCAGGGGAGC GCGGCGAGGA GCAGGGCGGG
 151:  CAGCGGGGGA CGAGGCGCCG TCCCCGCCCG CCCTCTCGCT GCCCCTCGGA
```

FIG. 4. Sequence of the cap site region from positions −300 to +200. An asterisk indicates nucleotide +1; dots indicate less prominent cap sites downstream (see Fig. 3). The boxed guanosine at position +53 corresponds to the first nucleotide of the cDNA sequence in ref. 12. The CCGCCC motif at position −109 is underlined.

rest of the gene. Evolutionarily, it will be very informative to learn whether other genes in this family have a similar structure.

An alternative polyadenylylation signal resides near the 5' end of intron B, and this results in a truncated mRNA. cDNA clones representing this mRNA species have been found and are postulated to correspond to a 1.8-kb band in RNA blots (12). The putative protein product of this variant mRNA would contain only the N-terminal domain and half of the DNA-binding domain of the receptor; it would lack hormone-binding activity. If actually present in cells, such truncated receptors might compete with normal receptor forms for available steroid-regulatory elements on target genes. Alternatively, such molecules may exist as dangerous cellular variants if they retain any biologic activity, since the repressive hormone-binding regulatory domain is absent.

The promoter of the chicken PR gene appears to belong to a class of eukaryotic promoters that lack TATA boxes and are exceptionally G+C rich (35–41). An important regulatory element in such genes may be the "GC" box whose core sequence is CCGCCC and is functional on either strand (42). In contrast to genes with TATA boxes, initiation of transcription from G+C-rich promoters is heterogeneous (35–41). These promoters are frequently associated with genes that are expressed at low levels in a non-tissue-specific manner (39–41). Although the chicken PR gene is expressed

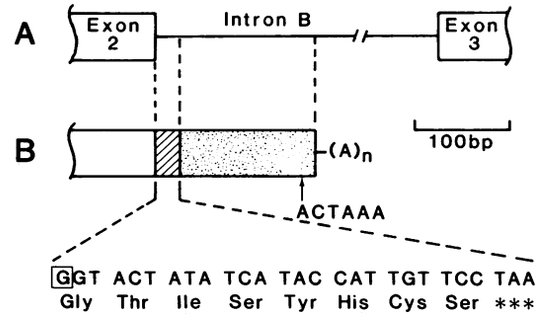

FIG. 5. One class of variant cDNAs results from alternative polyadenylylation in intron B. (*A*) Diagram of intron B and flanking exons. (*B*) Diagram of variant cDNA aligned with sequences in exon 2 and intron B. Open box, sequences normally present in exon 2; hatched box, additional coding sequence in the variant; stippled box, the 3'-noncoding sequence of the variant. The alternative polyadenylylation signal (ACTAAA) is indicated, and the novel coding sequence of the variant is shown. The boxed guanosine is the last nucleotide in the normal second exon.

at relatively low levels, its expression appears to be highly specific to the oviduct, where its mRNA levels are potentiated by estrogen and attenuated by progesterone (7).

Interestingly, the putative promoter of the human estrogen receptor contains both TATA and CAAT-like elements and is not particularly G+C rich (4). The apparently different promoter types of that gene and the chicken PR suggest that the regulation of different steroid receptor genes could occur by fundamentally different mechanisms. This question, however, will require direct functional analysis of a variety of such promoters.

We thank Drs. B. L. Maxwell and D. R. Sargan for tissue RNAs, C. Wilkinson for plasmid DNA preparation, and C. McCarthy for manuscript preparation. This work was supported by National Institutes of Health Grant HD08188. C.S.H. was supported by National Institutes of Health Postdoctoral Fellowship GM10332.

1. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Ors, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Nature (London)* **318,** 635–641.
2. Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wikström, A.-C., Gustaffson, J.-A. & Yamamoto, K. R. (1986) *Cell* **46,** 389–399.
3. Walter, P., Green, S., Greene, G., Krust, A., Bornert, J. M., Jeltsch, J. M., Staub, A., Jensen, E., Scrace, G., Waterfield, M. & Chambon, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7889–7893.
4. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320,** 134–139.
5. Green, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. & Shine, J. (1986) *Science* **231,** 1150–1154.
6. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M. & Chambon, P. (1986) *EMBO J.* **5,** 891–897.
7. Conneely, O. M., Sullivan, W. P., Toft, D. O., Birnbaumer, M., Cook, R. G., Maxwell, B. L., Zarucki-Schulz, T., Greene, G. L., Schrader, W. T. & O'Malley, B. W. (1986) *Science* **233,** 767–770.
8. Jeltsch, J. M., Krozowski, Z., Quirin-Stricker, C., Gronemeyer, H., Simpson, R. J., Garnier, J. M., Krust, A., Jacob, F. & Chambon, P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5424–5428.
9. Loosfelt, H., Atger, M., Misrahi, M., Guiochon-Mantel, A., Meriel, C., Logeat, F., Benarous, R. & Milgrom, E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9045–9049.
10. Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlansky, F., Guiochon-Mantel, A., Galibert, F. & Milgrom, E. (1987) *Biochem. Biophys. Res. Commun.* **143,** 740–748.
11. McDonnell, D. P., Mangelsdorf, D. J., Pike, J. W., Haussler, M. R. & O'Malley, B. W. (1987) *Science* **235,** 1214–1217.
12. Conneely, O. M., Dobson, A. D. W., Tsai, M.-J., Beattie, W. G., Toft, D. O., Huckaby, C. S., Zarucki, T., Schrader, W. T. & O'Malley, B. W. (1987) *Mol. Endocrinol.* **1,** 517–525.
13. Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. (1986) *Cell* **46,** 645–652.
14. Danielsen, M., Northrop, J. P. & Ringold, G. M. (1986) *EMBO J.* **5,** 2513–2522.
15. Kumar, V., Green, S., Staub, A. & Chambon, P. (1986) *EMBO J.* **5,** 2231–2236.
16. Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. (1987) *Cell* **49,** 39–46.
17. Godowski, P. J., Rusconi, S., Miesfeld, R. & Yamamoto, K. R. (1987) *Nature (London)* **325,** 365–368.
18. Miesfeld, R., Godowski, P. J., Maler, B. A. & Yamamoto, K. R. (1987) *Science* **236,** 423–427.
19. Green, S. & Chambon, P. (1986) *Nature (London)* **324,** 615–617.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1–545.
21. Kulomaa, M. S., Weigel, N. L., Kleinsek, D. A., Beattie, W., Conneely, O. M., March, C., Zarucki-Schulz, T., Schrader, W. T. & O'Malley, B. W. (1986) *Biochemistry* **25,** 6244–6251.
22. Johnson, D. A., Gautsch, J. W., Sprotsman, J. R. & Elder, J. H. (1984) *Gene Anal. Tech.* **1,** 3–8.
23. Simmen, R. C. M., Tanaka, T., Ts'ui, K. F., Putkey, J. A., Scott, M. J., Lai, E. C. & Means, A. R. (1985) *J. Biol. Chem.* **260,** 907–912.
24. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
25. Colbert, D. A., Knoll, B. J., Woo, S. L. C., Mace, M. L., Tsai, M.-J. & O'Malley, B. W. (1980) *Biochemistry* **19,** 5586–5592.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
27. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–550.
28. Bergsma, D. J., Chang, K. S. & Schwartz, R. J. (1985) *Mol. Cell. Biol.* **5,** 1151–1162.
29. Burke, J. F. (1984) *Gene* **30,** 63–68.
30. Woo, S. L. C., Beattie, W. G., Catterall, J. F., Dugaiczyk, A., Staden, R., Brownlee, G. G. & O'Malley, B. W. (1981) *Biochemistry* **20,** 6437–6446.
31. Berg, J. M. (1986) *Nature (London)* **319,** 264–265.
32. Baez, M., Sargan, D. R., Elbrecht, A., Kulomaa, M. S., Zarucki-Schulz, T., Tsai, M.-J. & O'Malley, B. W. (1987) *J. Biol. Chem.* **262,** 6582–6588.
33. Gilbert, W. (1978) *Nature (London)* **271,** 501.
34. Tso, J. Y., Van Den Berg, D. J. & Korn, L. J. (1986) *Nucleic Acids Res.* **14,** 2187–2200.
35. Singer-Sam, J., Keith, D. H., Tani, K., Simmer, R. L., Shively, L., Lindsay, S., Yoshida, A. & Riggs, A. D. (1984) *Gene* **32,** 409–417.
36. McGrogan, M., Simonsen, C. C., Smouse, D. T., Farnham, P. J. & Schimke, R. T. (1985) *J. Biol. Chem.* **260,** 2307–2314.
37. Velerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, Th. M., van der Voorn, L., van Ormondt, H. & van der Eb, A. J. (1985) *EMBO J.* **4,** 437–443.
38. Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2147–2151.
39. Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L. & Luskey, K. L. (1984) *Cell* **38,** 275–285.
40. Luskey, K. L. (1987) *Mol. Cell. Biol.* **7,** 1881–1893.
41. Yamaguchi, M., Hirosae, F., Hayashi, Y., Nishimoto, Y. & Matsukage, A. (1987) *Mol. Cell. Biol.* **7,** 2012–2018.
42. Kadonaga, J. T., Jones, K. A. & Tijan, R. (1986) *Trends Biochem. Sci.* **11,** 20–23.
43. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.