MOLECULAR AND CELLULAR BIOLOGY, Aug. 1996, p. 4337–4348 Vol. 16, No. 8
0270-7306/96/$04.00+0

Isolation of a Novel Retinoic Acid-Responsive Gene by Selection of Genomic Fragments Derived from CpG-Island-Enriched DNA

MARY SHAGO[1,2] AND VINCENT GIGUÈRE[1,3]*

Molecular Oncology Group, Royal Victoria Hospital,[1] and Departments of Medicine, Oncology, and Biochemistry, McGill University,[3] Montréal, Québec, Canada H3A 1A1, and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A1[2]

Received 5 March 1996/Returned for modification 24 April 1996/Accepted 15 May 1996

One of the primary goals in transcription factor research is the elucidation of the genetic networks controlled by a factor or by members of a family of closely related factors. The pleiotropic effects of retinoic acid (RA) in the developing and adult animal are mediated by ligand-inducible transcription factors (RA receptors [RARs] and retinoid X receptors [RXRs]) that belong to the superfamily of nuclear receptors. Regulatory regions of RA effector genes contain RAR and RXR binding sites (RAR elements [RAREs] and RXR elements [RXREs]) that generally consist of direct or everted repeats of the core half-site motif, (A/G)G(G/T)TCA. In order to identify novel genes regulated by RA, we devised a selection strategy based on the premise that regulatory regions of a large number of housekeeping and tissue-specific genes are embodied within CpG island DNA. In this method, referred to as CpG-selected and amplified binding, fragments derived from the CpG island fraction of the murine genome are selected by a gel mobility shift assay using in vitro-transcribed and -translated RXR-RAR. Multiple rounds of selection coupled with amplification of the fragments by PCR enabled us to clone a population of CG-rich fragments of which approximately one-fifth contained consensus RAREs or RXREs. Twelve genomic fragments containing novel response elements are described, and the transcription unit associated with one of them, NN8-4AG, was characterized in detail. The mouse NN8-4AG transcript is upregulated by RA in F9 embryonal carcinoma cells and is homologous to an expressed sequence tag (EST41159) derived from a human infant brain cDNA library. Cloning of the murine NN8-4AG genomic sequence places the RXRE in the proximity of the transcription initiation sites of the gene. Although sequence analysis indicates that the EST41159 gene product is novel, a region of amino acid identity with sequences of a yeast polypeptide of, as yet, unknown function and the *Drosophila trithorax* protein suggests the presence of an evolutionarily and functionally conserved domain. Our study demonstrates that transcription factor binding sites and corresponding regulated genes can be identified by selecting fragments derived from the CpG island fraction of the genome.

Vitamin A (retinol) and its derivatives are essential for the embryonic development, growth, and reproduction of vertebrates and for the regulation of the proliferation and differentiation of many cell types in adult animals (46). The ability of the active retinol derivatives all-*trans* retinoic acid (atRA) and 9-*cis* RA (9cRA) to mediate these biological effects can be attributed to the regulation of specific target genes. Both atRA and 9cRA interact in the nucleus with ligand-inducible transcription factors that belong to the superfamily of steroid, thyroid, and retinoid receptors. Two classes of retinoid receptors have been defined: the RA receptors (RARα, -β, and -γ), which bind both atRA and 9cRA, and the retinoid X receptors (RXRα, -β, and -γ), which are capable of binding only 9cRA with high affinity (reviewed in references 16, 25, and 36). RARs and RXRs recognize short DNA elements that are present in the regulatory regions of responsive genes and that are termed RAR elements (RAREs) and RXR elements (RXREs). These elements generally consist of direct repeats of the core consensus hexamer, (A/G)G(T/G)TCA (for a review, see reference 25). Other sequences able to function as RAREs consist of inverted and everted repeats of the core hexamer motif as well as complex elements composed of multiple core motifs arranged in direct, inverted, or everted configurations with spacings of various lengths (43, 52, 53, 55).

Despite extensive knowledge of RA action at the molecular level, relatively few RA target genes have been identified (for reviews, see references 25 and 29). In addition to the characterization of known genes previously shown to be regulated by RA, differential and/or subtractive hybridization strategies have been used to isolate cDNAs derived from differentially expressed mRNAs of unknown genes from untreated versus RA-treated cells (14, 29). The major drawbacks of this method are that only a fraction of the cloned genes will be direct targets of the retinoid receptors and that a time- and effort-consuming step is required to identify the RAREs and RXREs associated with the transcriptional units encoding the regulated mRNAs. An alternative strategy would be to isolate RAREs and RXREs present in vertebrate genomes first and then to characterize the genes associated with these elements. However, the observation that retinoid receptor binding sites are relatively short and sometimes degenerate means that putative RAREs and RXREs are scattered throughout the genome, both in regions that play a role in transcription and in transcriptionally inert regions. In addition, to be biologically relevant, the search for retinoid receptor binding sites should concentrate not only on DNA fragments located in the regulatory region of genes but also on transcription units actively

* Corresponding author. Mailing address: Molecular Oncology Group, Royal Victoria Hospital, 687 Pine Ave. West, Montréal, Québec, Canada H3A 1A1. Phone: (514) 843-1479. Fax: (514) 843-1478. Electronic mail address: vgiguere@dir.molonc.mcgill.ca.

transcribed in the tissue or cells of interest. One approach directed towards the isolation of in vivo target genes is the immunoprecipitation of chromatin fragments with transcription factor-specific antibody (28). This approach has been used to isolate genomic targets of the thyroid hormone receptors (8, 9).

In an effort to isolate RA target genes, we have developed a novel selection strategy that takes advantage of a unique and striking feature of the vertebrate genome: the presence of CpG islands. CpG islands are G+C-rich regions that, unlike the majority of vertebrate genomic DNAs (bulk chromatin), do not show a reduced frequency of CpG dinucleotides and are not methylated at the C of these dinucleotides (3, 11). These regions are located in or near the promoter regions of genes and often include the first exon of the gene (3, 11). Moreover, Tazi and Bird (50) have shown that nonnucleosomal (NN) DNA, a fraction of the genome associated with transcribed genes, is predominantly derived from the CpG island fraction of the genome. These NN fragments can be isolated away from bulk chromatin with restriction enzymes containing CpG within their recognition sequences (50). Taken together, these observations suggest that CpG-island-derived NN fragments could provide an excellent source of genetic material to scan for biologically relevant binding sites for retinoid receptors or any other transcription factors.

Using CpG-island-enriched DNA derived from the nuclei of murine embryonal carcinoma (EC) cells as starting material, we utilized a combination of PCR amplification and gel mobility shift assays to select NN fragments that contain RAREs and RXREs. By analyzing the genomic DNA flanking each RARE, associated transcribed regions can be identified. As this isolation method is a modification of selected and amplified binding (SAAB) (12), we have termed this method CpG-SAAB. In this paper, we describe CpG-SAAB, the functional characterization of 12 genomic fragments containing RAREs and RXREs, and the identification of a novel RA-responsive gene associated with one of the RXRE-containing fragments. Our study demonstrates that transcription factor binding sites and corresponding regulated genes can be identified by selecting from fragments derived from the CpG island fraction of the genome.

MATERIALS AND METHODS

Isolation of nuclei and chromatin digestion. Nuclei were prepared as described by Dony et al. (20) and Tazi and Bird (50), with some modifications. Five 100-mm-diameter confluent plates of F9 EC cells were trypsinized, pelleted, and washed with cold phosphate-buffered saline. To gently lyse the cells, 7.5 ml of cold 0.025% Nonidet P-40 lysis buffer (10 mM triethanolamine [TEA], 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM dithiothreitol, and 0.025% Nonidet P-40) was added and pipetted up and down 15 times. After centrifugation at 200 × *g* for 4 min, the pellet was resuspended in 0.025% Nonidet P-40 lysis buffer, layered on a 0.8 M sucrose cushion dissolved in lysis buffer, and then pelleted again at 200 × *g* for 8 min. Nuclei were resuspended gently in 500 μl of storage buffer containing 75 mM KCl, 0.25 mM $MgCl_2$, 1 mM EDTA, 0.5 mM ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA), 0.15 mM spermine, 0.5 mM spermidine, 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) (pH 7.9), 10 mM β-mercaptoethanol, 0.23 M sucrose, and 20% glycerol. An average of 4×10^7 nuclei were obtained from five plates of F9 cells. Nuclei were stored at −70°C. As described by Tazi and Bird (50), 500-μl aliquots of nuclei were thawed, pelleted for 10 s, and gently resuspended in 1× digestion buffer (50 mM Tris-HCl [pH 8.0], 10 mM $MgCl_2$) and 400 U (each) of *Msp*I and *Hin*PI in a total volume of 1 ml. Digestion was followed by proteinase K treatment. Subsequent to phenol-chloroform extraction and ethanol precipitation, DNA was dissolved in 400 μl of Tris-EDTA.

Isolation and PCR amplification of NN fragments. Fragments for PCR amplification and gel mobility shift selection were prepared by running 10 to 30 μl of the chromatin digest on a 1.5% low-melt-agarose gel. Fragments migrating to between 75 and 140 bp were blunted with the Klenow fragment of DNA polymerase, phenol chloroform extracted, ethanol precipitated, and resuspended in 10 μl of Tris-EDTA. The fragments were then ligated overnight at 16°C to catch linkers (catch A [5′-GAGTAGAATTCTAATATCTC-3′] and its complement, catch B) by following a procedure outlined by Kinzler and Vogelstein (32). Overdigestion with *Xho*I in a total volume of 500 μl cuts between annealed linkers and results in NN fragments with a single catch linker on either end. The digest was passed through a Qiagen tip 5 column and ethanol precipitated, and fragments with the appropriate sizes were reisolated from a 1.5% low-melt-agarose gel (SeaKem). The resulting NN fragments, which were quantitated by ethidium bromide agarose plate quantitation, were resuspended in 10 μl of Tris-EDTA, typically resulting in a concentration of 2 to 5 ng/μl. To amplify and radiolabel the NN fragments for gel mobility shift selection, the catch A oligonucleotide was end labeled with [γ-^{32}P]ATP and polynucleotide kinase (New England Biolabs). PCR mixtures consisted of 15 ng of NN catch-linked fragment, 250 ng (each) of labeled catch A and unlabeled catch B oligonucleotides, PCR buffer, $MgCl_2$, and *Taq* polymerase (according to Perkin-Elmer Cetus instructions). Fifteen cycles of denaturation for 1 min at 94°C, annealing for 2 min at 50°C, and extension for 2 min at 72°C, with a 4-min denaturation at 94°C prior to the first cycle, were performed. Fragments with the appropriate sizes (115 to 190 bp) were then isolated on a 1.5% low-melt-agarose gel, resuspended in 10 μl of water, and stored at −20°C.

Gel mobility shift assays. Gel mobility shift assays were performed with in vitro-transcribed and -translated RAR and RXR. The proteins were synthesized in rabbit reticulocyte lysate (RRL) with the TNT-T7 kit (Promega, Madison, Wis.). Typically, 2 μl of each RAR- and RXR-programmed RRL was combined with binding buffer (5 mM Tris-HCl [pH 8.0], 40 mM KCl, 6% glycerol, 1 mM dithiothreitol, and 0.05% Nonidet P-40), 2 μg of poly(dI-dC) (Pharmacia), 0.1 μg of denatured salmon sperm DNA, and 10 μg of bovine serum albumin in a total volume of 20 μl, and the mixture was put on ice for 10 min. One microliter of labeled NN fragments was added, and the reaction mixture was incubated for an additional 10 min at room temperature. The reaction mixture was loaded on a nondenaturing 3.5 to 4% polyacrylamide gel and run at 150 V for 2 h. The gel was dried and exposed with an intensifying screen for a few hours to overnight. Control binding reactions with a known RARE (βRARE, 5′-AGCTTGAAGG GTTCACCGAAAGTTCACTCGCA-3′ and 5′-AGCTTGCGAGTGAACTTT CGGTGAACCCTTCA-3′) were run alongside to provide a marker for protein-DNA complex migration. A section of the dried gel corresponding to the migration point of bound NN fragments was cut out, minced with a razor blade, and eluted with shaking overnight at 37°C in 600 μl of elution buffer (0.5 M ammonium acetate, 1 mM EDTA). After phenol-chloroform extraction, the selected NN fragments were precipitated with isopropanol and glycogen and resuspended in 10 μl of water. Half of the isolated NN fragments were used in a PCR amplification as described above. The selection for NN fragments that bound to RXR-RAR consisted of eight rounds of PCR amplification followed by gel mobility shift assay isolation.

Isolation of the RARβ2 genomic clone and preliminary testing of CpG-SAAB. As a control for the preliminary steps of CpG-SAAB, a genomic clone corresponding to the promoter region of the RARβ2 gene was isolated from a murine EMBL-3 SP6/T7 genomic library by using a probe derived from the 5′ end of the mouse RARβ2 cDNA. The RARβ2 gene is highly responsive to RA and contains a DR-5 RARE (βRARE) in its promoter (18, 49). Approximately 2,500 bp surrounding the βRARE were sequenced. The pattern of restriction enzyme recognition sites containing CpG dinucleotides within this region suggests that the βRARE is within a CpG island that spans approximately 1 kb. Flanking the βRARE are *Msp*I and *Hin*PI sites, which would enable the release of a 140-bp NN fragment containing the response element. Primers internal to the *Msp*I and *Hin*PI sites were designed and used as PCR primers to monitor the presence of the βRARE NN fragment through the preliminary steps of CpG-SAAB. With RARβ cDNA, F9 genomic DNA, NN fragments, or fragments eluted from the gel after the first round of selection being used as the template for PCR, a 140-bp fragment was visualized on an ethidium bromide-stained acrylamide gel (data not shown). No band was seen in the control lane without the template. The identity of this fragment was confirmed by Southern blotting with an internal probe (data not shown). These preliminary experiments established that the RARE of a known RA-responsive gene is contained within a CpG island, that the element can be released as a 140-bp NN fragment upon *Msp*I-*Hin*PI digestion of intact F9 nuclei, and that this NN fragment is bound and selected by in vitro-transcribed and -translated RAR-RXR in the first gel mobility shift selection of CpG-SAAB.

Cloning and sequencing of NN fragments. After the final PCR amplification, NN fragments were digested with *Eco*RI (which cuts within the catch linker) and ligated into pBluescriptKS (pKS; Stratagene) cut with *Eco*RI and treated with calf intestinal phosphatase (Boehringer Mannheim). Dideoxy sequencing reactions were performed with T7 DNA polymerase (Pharmacia) and the T3 and T7 primers. GC compressions were resolved with 7-deaza deoxynucleotides (Pharmacia) or by running sequencing reactions on a 40% formamide gel.

Recombinant plasmids. Plasmid TKLUC contains sequence −105 to +51 of the thymidine kinase (TK) promoter linked to the firefly luciferase gene. TKLUC plasmids containing NN fragments were constructed by releasing the NN fragments from pKS with *Sal*I and *Bam*HI and cloning them into the corresponding sites in TKLUC. To create NN-βRARE-TKLUC, a 140-bp *Msp*I-*Hin*PI fragment from the promoter region of the murine RARβ2 gene containing the βRARE was isolated, cloned directly into pKS (Stratagene) linearized with *Acc*I, and subsequently inserted into TKLUC as a *Sal*I-*Bam*HI fragment.

Transient transfection assays. P19 EC cells were cultured in α minimal essential medium containing 7.5% fetal calf serum. Cells were transfected with 1.5 μg of RSV-β-gal and 2 μg of reporter plasmid. β-Galactosidase and luciferase assays were performed as described previously (26).

Genomic library screening. Clone λ8-4-5 was isolated from a murine λ DASHII library of strain 129SV genomic DNA with a ^{32}P-labeled probe generated from NN fragment 8-4. Hybridization of duplicate nylon membranes was performed at 42°C overnight in a solution of 50% formamide, 1× Denhardt's solution, 5× SSPE (1× SSPE is 150 mM NaH_2PO_4 and 1 mM EDTA), 0.1% sodium dodecyl sulfate (SDS), 100 μg of denatured salmon sperm DNA per ml, and 10^6 cpm of probe per ml. Membranes were washed twice with 0.2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS for 20 min at 60°C. Genomic phage clones were analyzed in detail by restriction mapping and hybridization with NN8-4 and cDNA-derived fragments.

Sequence determination and analysis. Homology of NN8-4 and an expressed sequence tag (EST) from the EST database was discovered with the BLAST search program of the National Center for Biotechnology Information (2). Clone EST41159 (GenBank number R56376) was obtained through the IMAGE Consortium, Lawrence Livermore National Laboratory. EST41159 was subcloned into pKS, and both this cDNA and genomic subclones were sequenced with T7 DNA polymerase (Pharmacia). T3 and T7 primers, as well as specific primers derived from the cDNA or genomic sequences, were used to sequence both strands of the cDNA and genomic subclones. Similarity searches for putative amino acid sequences were performed with the BLASTX program (2).

Northern (RNA) analysis. F9 EC cells were grown on 0.1% gelatin in Dulbecco modified Eagle medium supplemented with 15% fetal bovine serum which had been stripped by using charcoal and anion exchange and treated with UV light for 1 h for retinoid depletion. The cells were treated with vehicle alone, retinoids at a concentration of 1 μM each, and/or cycloheximide (Cx) (final concentration, 10 μg/ml) for various time periods. Total RNA was extracted with TRIZOL (Gibco BRL). RNA samples were electrophoresed through a 1% formaldehyde–1× MOPS (morpholinepropanesulfonic acid)–1% agarose gel and transferred to a nylon membrane in 20% SSC. RNA was UV cross-linked to the membrane and hybridized to the appropriate random-primed probe in a solution containing 50% formamide, 5× Denhardt's solution, 5× SSPE, 0.1% SDS, and 100 μg of denatured salmon sperm DNA per ml. The membranes were washed once for 20 min at room temperature with 2× SSC–0.1% SDS and twice at 65°C with 0.2× SSC–0.1% SDS for 15 min each time. Autoradiography was carried out at −70°C with an intensifying screen. An oligonucleotide hybridizing to the 18S rRNA (5′-ACGGTATCTGATCGTCTTCGAACC-3′) was used as a control probe for loading.

RACE analysis. The rapid amplification of cDNA ends (RACE) system was used according to the instructions provided by the manufacturer (Gibco BRL). The primers GSP1, GSP2, and GSP3 were 5′-CAGCCTCTGGGCATTCTCCT CTGCCTTCTC-3′ in exon 3, 5′-CTTCCGAATCAGCTGTGTGGCAAACAG TCCC-3′ in exon 2, and 5′-CGCCACGCCCACGCAGAAGGAGAACAC-3′ in exon 1. First-strand synthesis was performed at 50°C.

Nucleotide sequence accession numbers. The GenBank accession numbers for the complete sequence of EST41159 and the genomic region flanking NN8-4 are U50383 and U50384, respectively.

RESULTS

Isolation of NN fragments containing RAREs by CpG-SAAB. Murine EC cell lines, such as F9 and P19, constitute excellent models to study RA action (31, 48). In addition, unlike many transformed cell lines, EC cells maintain their methylation status in culture (4). F9 EC cells were chosen as starting material for the development of the CpG-SAAB method. We first tested whether we could release CpG island DNA from F9 cell nuclei. Restriction enzymes which include a CpG in their recognition sites will preferentially digest within CpG island DNA. It has been demonstrated that chromatin digested by such CG-recognizing enzymes as *Msp*I and *Hin*PI will release oligonucleosomes predominantly derived from the CpG island fraction of the genome (5) as well as a population of fragments that are smaller than the DNA associated with a single nucleosome (50). *Msp*I and *Hin*PI (the recognition sites are CCGG and GCGC, respectively) have access to CpG island chromatin not only because of the open structure of the chromatin (relative to bulk chromatin) and the characteristic GC-rich composition of this chromatin but also because of the nonmethylated status of CpG islands. Since *Hin*PI is a methylation-sensitive enzyme, it will not digest within methylated bulk chromatin. While *Msp*I can cut methylated DNA, there appears to be a strong bias against such digestion in bulk chromatin, possibly due to the protection of methylated DNA by specifically bound factors (5, 15, 35). F9 cell nuclei subjected to digestion with a non-CpG-recognizing restriction enzyme, such as *Sau*3A (recognition site GATC), release a series of fragments which, after being radioactively end labeled and electrophoresed on an agarose gel, can be seen to correspond to the DNA fragments associated with a single nucleosome, a dinucleosome, and a trinucleosome, with this pattern continuing into high-molecular-weight DNA (Fig. 1A). In contrast, digestion with *Msp*I and/or *Hin*PI yields DNA fragments corresponding in size only to one to four nucleosomes and a

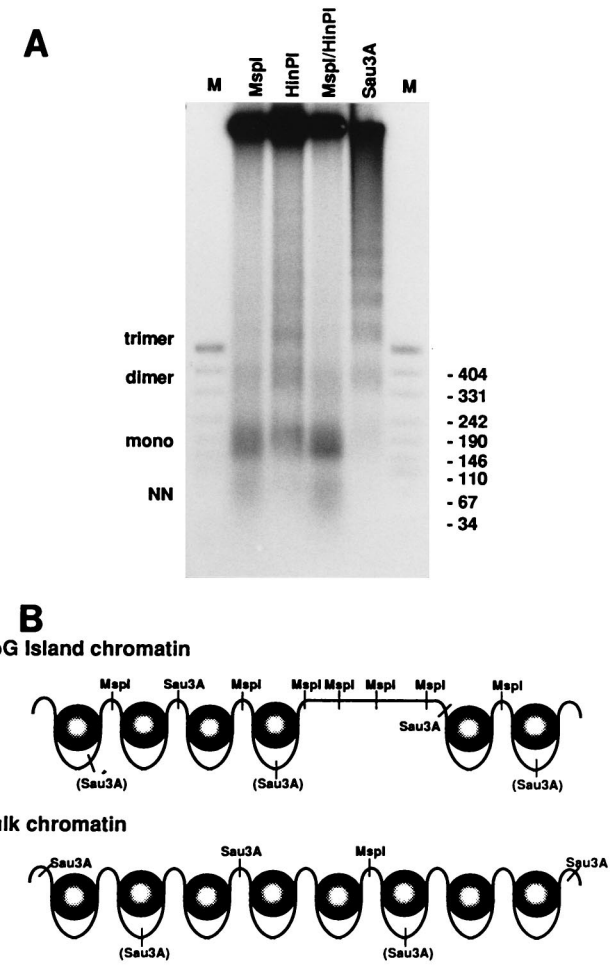


FIG. 1. NN fragments are released from CpG island chromatin upon digestion with CpG-recognizing enzymes. (A) Intact nuclei from F9 EC cells were subjected to digestion with restriction enzymes *Msp*I (CCGG), *Hin*PI (GCGC), or *Sau*3A (GATC) or a combination of both *Msp*I and *Hin*PI. Subsequent to isolation, DNA was end labeled and electrophoresed on a 1.5% agarose gel. An autoradiograph of the gel reveals differences in the digestion patterns caused by CpG-recognizing enzyme lanes. These fragments are (as illustrated in panel B) thought to be derived from the nucleosome-free regions within the CpG islands of activated or potentially activable genes. The outer lanes (M) consist of end-labeled fragments of pUC19 cut with *Hpa*II. The numbers on the right are marker band sizes (in base pairs). (B) CpG island chromatin is GC rich and contains an expected number of CpG dinucleotides, which are generally unmethylated. This type of chromatin contains an abundance of sites for restriction enzymes containing CpG in their recognition sites (such as *Msp*I). Bulk chromatin has a low GC content and is deficient in CpG dinucleotides, with those remaining being methylated. Very few CpG-containing recognition sites are present in bulk chromatin. Bracketed restriction enzyme sites in both types of chromatin indicate sites which are inaccessible because of blockage by nucleosomes. For details and data on the original experiments, see the report by Tazi and Bird (50).

correspondingly greater amount of high-molecular-weight DNA (Fig. 1A). As was predicted previously (50), a heterogeneous population of fragments that are smaller than a single nucleosome is released upon *Msp*I and/or *Hin*PI digestion. These fragments, termed NN fragments, are GC rich, are derived in large part from extended stretches of nucleosome-free DNA that are characteristic of transcribed (or potentially transcribable) genes, and are found predominantly within CpG islands (Fig. 1B) (50). Thus, the CpG-island-derived NN fragments constitute an ideal fraction of the genome in which to search for transcription factor binding sites associated with transcribed genes, in our case, RAREs and RXREs.

The strategy to isolate RARE-containing NN fragments is schematically represented in Fig. 2. Intact nuclei are first isolated, and the chromatin is subjected to digestion with *Msp*I and *Hin*PI. The DNA is then electrophoresed on a 1.5% agarose gel, and NN fragments migrating to between 75 and 140 bp are extracted from it. The upper size limit was chosen to allow separation of bound fragments from unbound fragments in the gel mobility shift assay selections. NN fragments are then end filled with the Klenow fragment of DNA polymerase I and ligated to blunt catch linkers (32). Overdigestion of the product of the ligation with *Xho*I, which cuts between ligated catch linkers, creates a population of fragments which have a single linker on either end, facilitating PCR amplification. As was described in Materials and Methods, PCR was performed with radiolabeled catch linkers. Binding reactions for gel mobility shift selections consisted of modified and radiolabeled NN fragments being incubated with in vitro-transcribed and -translated RXR-RAR. As is diagrammed in Fig. 2, parallel reactions with a known RARE are run alongside the NN fragment binding reaction as a marker for the bound fraction of fragments. A section of dried gel corresponding to the final position of the migration of bound NN fragments is cut out, and the DNA is eluted. Multiple rounds of amplification by PCR (with end-labeled catch oligonucleotide) and gel mobility shift selection are then performed to enrich the reaction mixture for NN fragments containing RAREs. By CpG-SAAB, a population of 120 RXR-RAR-selected NN fragments obtained from F9 EC cell nuclei was cloned into pKS for further analysis.

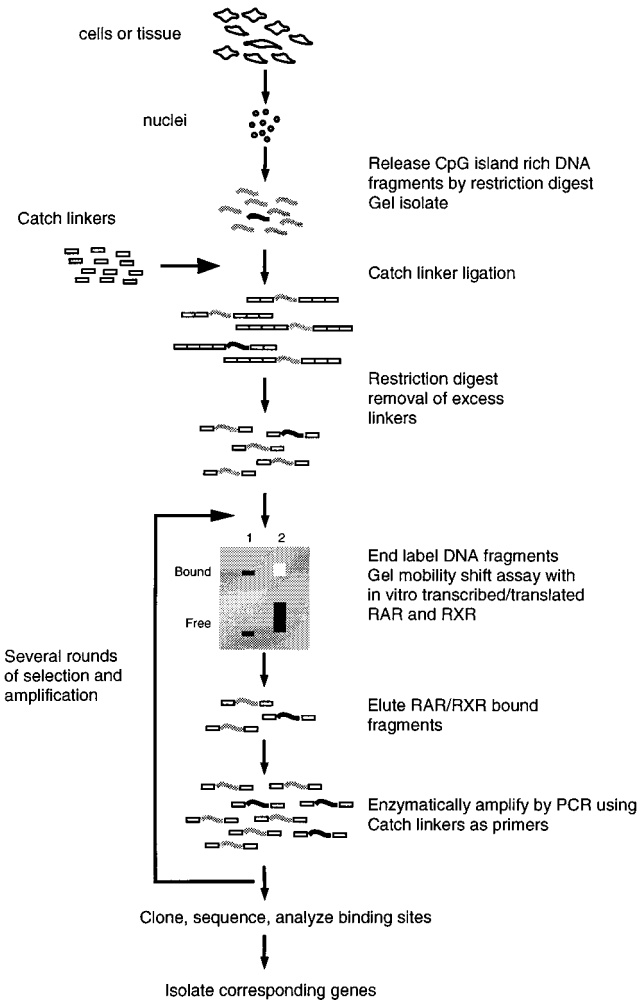


FIG. 2. CpG-SAAB for the selection of CpG-rich genomic fragments containing RAREs. Nuclei are purified from cells or tissue samples of interest. The CpG-island-rich NN fragments are released by restriction enzyme digests and gel isolated. Solid fragments represent those NN fragments containing an RARE or RXRE. Catch linkers are ligated to the collection of NN fragments, and excess linkers are removed by an *Xho*I digest. NN fragments are then selected in a gel mobility shift assay with RAR-RXR complexes produced by an in vitro transcription-translation system. The probes represented in the schematic of the gel mobility shift assay are βRARE as a control in lane 1 and NN fragments in lane 2. NN fragments bound to the heterodimer are eluted from the gel (the cutout gel slice is represented by the white rectangle) and enzymatically amplified with the catch linker oligonucleotides being used as primers. After several rounds of selection, DNA fragments are cloned into pKS and sequenced to search for the presence of a consensus RARE or RXRE. Putative RARE-RXRE-containing fragments are assessed for specific binding to RXR-RAR complexes and the ability to drive reporter gene expression in transient transfection assays. Fragments showing a positive response in these assays are then used as probes to isolate associated genes.

Physical analysis of RXR-RAR-selected NN fragments. The RXR-RAR-selected NN fragments obtained from F9 EC cell nuclei were sequenced and scanned for the presence of consensus RAREs and RXREs. As was discussed above, RARs and RXRs are members of a subfamily of the nuclear receptor superfamily that recognize direct, inverted, or everted repeats of the core consensus hexamer [(A/G)G(T/G)TCA] present in the regulatory regions of responsive genes. In general, elements containing direct repeats spaced by 1 nucleotide (DR1) function as RXREs, while DR2 and DR5 elements function as RAREs. We first performed a comparison of NN fragments that had not been subjected to gel mobility shift selection and NN fragments that had been subjected to selection. The results of this comparison are outlined in Table 1. No RARE and RXRE consensus motifs of any class were present in the 30 preselection NN fragments sequenced. Among the selected NN fragment RA response elements identified were five DR1s, one DR2, seven DR5s, and 13 complex elements containing combinations of direct and everted repeats. Of the 120 selected NN fragments sequenced, approximately one in five harbored a consensus RA response element. Since some of the NN-RA response element-containing fragments were isolated more than once, in total, 12 unique RA response elements were identified. The average sizes of the genomic fragments were 91 and 98 bp for pre- and postselection NN fragments, respectively. Non-CpG island bulk DNA has a G+C content of approximately 40% (3). As CpG islands typically have a base composition in excess of 50% G+C (17), averages of 61% for the preselection NN fragments and 59.5% for the selected NN fragments indicate that the majority of the fragments are probably derived from CpG island regions of the genome.

To ascertain whether, in general, the RA response element-containing NN fragments were derived from single-copy DNA, four of them were used to probe Southern blots of total mouse genomic DNA digested with various restriction enzymes. Three of these yielded the banding pattern expected of a sin-

TABLE 1. Characteristics of pre- and postselection NN fragments

Status of fragment	No. of fragments	Size[a]	GC content (%)	No. of fragments in the following class of RA response element[b]:			
				DR1	DR2	DR5	Complex
Preselection	30	91	61	0	0	0	0
Postselection	120	98	59.5	5	1	7	13

[a] Average sizes of the NN fragments in base pairs.
[b] DR1, DR2, and DR5 indicate NN fragments containing a single direct repeat of the core half-site motif spaced by 1, 2, or 5 nucleotides; complex refers to NN fragments containing at least one DR element plus one or more additional core half-site motifs arranged in various configurations.

gle-copy probe, while one resulted in multiple bands of equal intensities (three to four bands per digest) (data not shown). These results suggest that the NN fragments examined are derived from repeat-free genomic DNA, an observation that correlates well with previous studies indicating that the frequency of repeated sequences within CpG island regions is much lower than that expected for total genomic DNA (10, 17).

Functional analysis of RXR-RAR-selected NN fragments. Although the consensus sequences for RAREs and RXREs have been well established, it is important to ascertain whether the RA response elements associated with the NN fragments fulfill certain functional requirements, namely, the abilities to specifically bind RXR-RAR complexes and to confer RA responsiveness to a heterologous minimal promoter in a transient transfection system. The RARE-containing portions of the NN fragments are outlined in Table 2, along with the class of the binding site. A 140-bp *Msp*I-*Hin*PI genomic fragment located within a putative CpG island and containing the RARE from the promoter of the RARβ2 gene served in the functional assays as a control (see Materials and Methods). The relative binding affinities of the fragments are shown in Table 2. All of the NN fragments except for NN7-29, which contains a DR1 with one half-site of the suboptimal consensus sequence, bind RXR-RAR complexes at least as well as the genomic fragment containing the βRARE. NN fragments containing complex elements tend to bind with higher affinity to the retinoid receptors. Each of these fragments contains a perfect consensus direct repeat RARE or RXRE plus addi-

TABLE 2. Functional analysis of NN RA response elements

Name	Sequence[a]	Class[b]	Binding[c]	Fold[d] Induction	Total[e] Activity
NNβRARE	GGGAAGGGTTCACCGAAAGTTCACTCGCA	DR5	++	59	100
NN7-29	GGGTAGAGGTCACGTGTCAGTCCCG	DR1	+	4	25
NN7-51	TGACCTTGTGAATG(N15)TGAGCTAGTGGGGTCAACGAAAGGTCA	DR2, ER4, DR5	+++	55	111
NN7-71	AGGTCAAAAGTTCAAGCTCATCCAAGGATCAAAGTTA	DR2, DR0, DR5	+++	113	268
NN7-87	GGAGGTCAAACTCAGGTCA(N31)TGAGCC(N18)GGATCATT	DR5, 2HS	++	9	36
NN7-91	GGTTCGAGGTCAAGAAGGGGTCATG	DR5	+++	49	211
NN8-3	GGACAAGGGTCAGAGACAGGTCACTGTAG	DR5	+++	13	14
NN8-4	CGGACGGGGTCAAGGGTCAGAAGGC	DR1	++	5	75
NN8-29	GGGTCACACCTGAGGTCAGGATCACACGTG GGTTCACAGGGGTCACTGGGGTCAGCCTGGGGTCACAGTGACATCA	DR6, DR0, DR3, DR5	++	58	84
NN8-39	TGCCTTAGGTGAAAGGTCAAAGGTCATGCAGG	DR1	+++	5	11
NN8-44	AGTGACTCTGCAGCTGGGGTCATTCCAAGTTCACG	DR5, ER8	+++	32	93
NN8-49	GGTGAAAGGTCAAAGGTCATGCAGG	DR1	+++	4	9
NN8-66	GCCCAGAGGTCATAAGGTCACCAAGA	DR2	++	6	7

[a] Only the sequence of the RA response element(s) identified within each NN fragment is displayed.
[b] DR and ER denote pairs of core half-sites arranged as direct or everted repeats spaced by the indicated number of nucleotides; HS denotes individual core half-sites.
[c] Differential binding of RXR-RAR heterodimers to NN fragments in gel mobility shift assays is arbitrarily indicated as weaker than (+), the same as (++), or stronger than (+++) that to NN-βRARE.
[d] Fold induction denotes the RA inducibility of the RARE-containing NN fragments linked to the TK promoter.
[e] Total activity refers to the RA-induced luciferase activity generated by the NN fragments normalized against the basal activity of control TKLUC. The activity of the NN-βRARE-TKLUC construct was arbitrarily set at 100.

tional half-sites in various configurations which presumably cooperate to increase binding affinity.

A single copy of each NN fragment was cloned upstream of the TK promoter driving the luciferase reporter gene (TKLUC). Transient transfections were performed in P19 EC cells which were treated with 1 μM RA 16 h after transfection and harvested 24 h after treatment. Results seen with transient transfections of the constructs into F9 EC cells were qualitatively similar. Two values for each NN-RARE were calculated from this study (Table 2). Fold induction indicates the level of RA inducibility conferred to the TK promoter by each NN fragment. The total activity of the NN fragment is the luciferase activity detected upon RA treatment normalized against the basal activity of the control, TKLUC. The total activity of NN-βRARE-TKLUC was arbitrarily set at 100, and activities of the other reporters are relative to this control. Both values are calculated, because in some cases the basal activity of the fragment plus the TK promoter without RA treatment is much greater than that of TK alone. For example, although the fold induction (RA treatment versus no treatment) of NN8-4TKLUC is 5 while that of NN-βRARE-TKLUC is 59, the total RA-induced luciferase activities of the reporters are relatively similar (75 U for NN8-4TKLUC and 100 U for NN-βRARE-TKLUC). The higher basal activities of some of the NN fragments suggest the presence of additional binding sites for other transcription factors that could contribute to enhancer activity.

As an additional control, these functional assays were performed on a number of selected NN fragments with no recognizable consensus RA response elements (data not shown). In each case, these fragments did not bind specifically or at all to RRL programmed with RXR-RAR and conferred no RA-induced activation on TKLUC. Generally, the affinity of binding correlates well with the ability to confer RA induction upon a heterologous promoter, with a few notable exceptions. For example, fold induction values for DR1-containing fragments NN7-29, NN8-4, NN8-39, and NN8-49 are low. Such a result is not unexpected, considering that although heterodimer RXR-RAR binds with high affinity to a DR1 configuration, repression is thought to be the result (37). As was mentioned above, NN8-4 (and, to a lesser extent, NN7-29) is unusual in that the fragment confers a high level of total activity upon TKLUC. However, the fold induction value is low, because the basal level of transcription is greatly increased. Finally, a single copy of a DR2 element, such as that observed in NN8-66, generally confers activity more readily in transient transfections with the addition of exogenous receptor or if present in multiple copies (unpublished data).

Identification of a gene associated with the RARE of NN8-4. Following the successful isolation of novel RARE-containing genomic fragments by CpG-SAAB, we wished to identify transcription units associated with these genomic NN fragments. There exist a number of methods to identify transcribed sequences within a genomic region, including identification of conserved sequences by the screening of Northern blots and zoo blots, cDNA selection, exon trapping, and finally direct DNA sequencing and use of available software and DNA database searches, called software trapping (for a review, see reference 24). The last approach is becoming increasingly powerful as more efficient computer programs and search algorithms are developed and as databases, particularly those of ESTs, are expanded. The NN fragments were first analyzed with BLAST to search the data banks of the National Center for Biotechnology Information (2). NN fragment 8-4 was found to share a 42-bp-long stretch of homology with EST41159 (GenBank accession number R56376), an EST derived from a human infant brain library (7) (Fig. 3A). ESTs consist of partial sequences from randomly cloned novel cDNAs (13). The library was constructed so that cDNAs were cloned directionally into the vector, thus allowing identification of 5′ and 3′ ends. As shown in Fig. 3A, alignment of EST41159 with NN8-4 places the RXRE approximately 40 bp upstream of the 5′ end of cDNA EST41159. In addition, there are only four mismatches in the 42-bp overlap of the mouse genomic NN fragment and the human cDNA, suggesting a high level of conservation within this sequence. EST41159 cDNA was obtained from the IMAGE Consortium and was sequenced in its entirety (Fig. 3B). The cDNA is 2,520 nucleotides in length and includes a poly(A) tail and a consensus polyadenylation site. An open reading frame with a size of 412 amino acids commences at the extreme 5′ of the cDNA, suggesting that this cDNA is not full length.

Characterization of murine gene NN8-4AG encoding the homolog of human cDNA EST41159. To firmly establish the relationship between the murine NN8-4 genomic fragment and human EST41159 as well as to characterize the transcriptional unit associated with NN8-4 (referred to as NN8-4 associated gene, or NN8-4AG), a murine λ129 genomic library was screened with the NN8-4 fragment being used as a probe. Figure 4A shows a schematic of the NN8-4 fragment used as a probe to isolate λ genomic clone 8-4-5. The G+C content (67%) of NN8-4 suggested that it was derived from a CpG island. Sequence analysis of approximately 1 kb of DNA surrounding the RXRE of the NN8-4 fragment (Fig. 4B) revealed that the surrounding genomic sequence contains a putative CpG island that spans approximately 600 bp, as is demonstrated by the vertical bars representing CpG dinucleotides in Fig. 4A. The NN8-4AG CpG island sequence extends farther 5′ of the EST41159 sequence and contains two in-frame ATG codons (Fig. 4B). The homology of NN8-4AG and EST41159 continues downstream of the NN8-4 fragment border for an additional 12 amino acids before NN8-4AG diverges into an intronic sequence. A further nine exons of NN8-4AG were mapped onto λNN8-4-5 with the aid of the EST41159 probe (Fig. 4A) and were subsequently sequenced for comparison with human cDNA. Comparison of the amino acid sequence of the putative gene product coded for by the first ten exons of NN8-4AG with that of the EST41159 translation product reveals a level of amino acid identity of 86% (92% similarity) for the human and mouse clones (Fig. 4C). RACE was performed on mouse brain RNA with primers located in putative exons 1, 2, and 3 of the NN8-4AG transcript. A series of start sites was identified in the region of the RXRE, and the longest of these localized on the first G of the first half-site, 33 bp upstream of the 5′-most methionine start codon. No TATA box is evident upstream of these start sites, consistent with the presence of multiple transcriptional initiation sites (23). A perfect consensus CCAAT box is located 110 bp upstream of the cluster of start sites identified by RACE analysis. In addition, two GC boxes, or potential Sp1 binding sites, are located 12 and 56 bp upstream of the cluster of transcriptional start sites. These results place the NN8-4 RXRE within a CpG island in the vicinity of the multiple transcriptional start sites of NN8-4AG.

NN8-4 RXRE is bound by RAR-RXR heterodimers and RXR homodimers. As can be seen from Table 2, NN fragment 8-4 binds specifically to in vitro-transcribed and -translated RXR-RAR and confers RA-induced activation on TKLUC in transient transfection in P19 EC cells. It is highly likely that the consensus DR1 present within NN8-4 is responsible for this activity. To verify this notion, an oligonucleotide encoding this sequence was synthesized (NN8-4 RXRE), end labeled, and analyzed in gel mobility shift assays with in vitro-transcribed

A

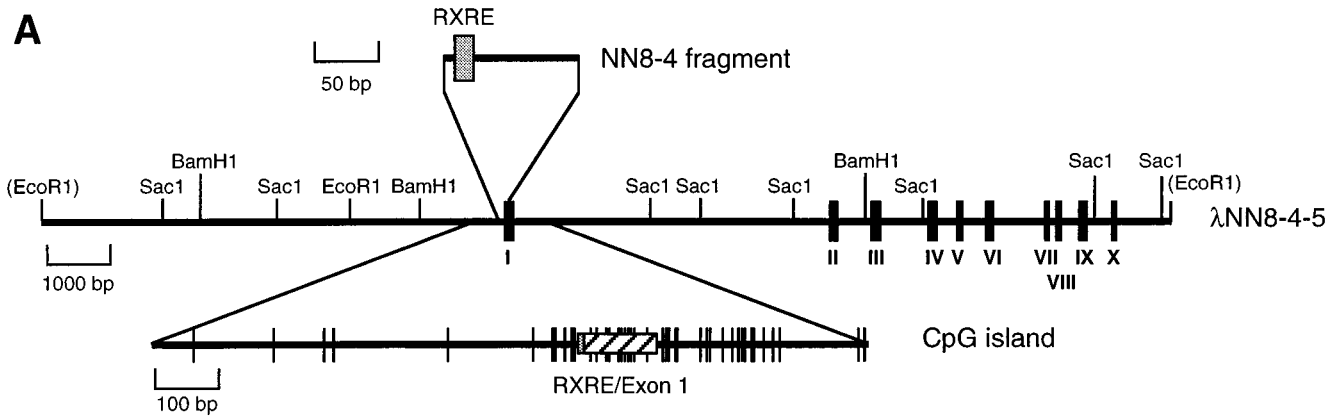
```
NN8-4      1 CCGGGACGGGGTCAAGGGTCAGAAGGCAGAGGCGTGCCCAAGATGGCGGCCCTCCATGTG

NN8-4     61 CGACGTGTTCTCCTTCTGCGTGGGCGTGGCGGACCGTGCCCGG
              | |||||||||||||||||||||||||||||| ||| || |||
EST41159   1  GCCGTGTTCTCCTTCTGCGTGGGCGTGGCGGGCCGCGCGCGG
```

B

```
                              30                                  60                                  90
GCCGTGTTCTCCTTCTGCGTGGGCGTGGCGGGCCGCGCGCGGGTCTCCGTGGAAGTCCGTTTCGTGAGCAGCGCCAAGGGAAAGGGGCTG
AlaValPheSerPheCysValGlyValAlaGlyArgAlaArgValSerValGluValArgPheValSerSerAlaLysGlyLysGlyLeu
                             120                                 150                                 180
TTTGCCACACAGCTCATCCGGAAGGGGGAGACCATCTTCGTAGAACGGCCCCTGGTGGCTGCACAGTTTCTCTGGAATGCACTTTATCGC
PheAlaThrGlnLeuIleArgLysGlyGluThrIlePheValGluArgProLeuValAlaAlaGlnPheLeuTrpAsnAlaLeuTyrArg
                             210                                 240                                 270
TACCGAGCCTGTGACCACTGCCTTAGGGCACTAGAGAAGGCAGAGGAGAATGCCCAGAGGCTGACCGGGAAACCAGGCCAGGTTCTGCCT
TyrArgAlaCysAspHisCysLeuArgAlaLeuGluLysAlaGluGluAsnAlaGlnArgLeuThrGlyLysProGlyGlnValLeuPro
                             300                                 330                                 360
CACCCAGAGCTGTGCACTGTGCGCAAAGACCTCCACCAGAACTGTCCCCATTGCCAAGTGATGTACTGCAGTGCAGAATGTCGGTTGGCA
HisProGluLeuCysThrValArgLysAspLeuHisGlnAsnCysProHisCysGlnValMetTyrCysSerAlaGluCysArgLeuAla
                             390                                 420                                 450
GCCACTGAGCAATACCACCAGGTCCTGTGCCCAGGCCCCTCCCAGGATGACCCCTTGCATCCTCTCAATAAGCTTCAGGAGGCATGGAGG
AlaThrGluGlnTyrHisGlnValLeuCysProGlyProSerGlnAspAspProLeuHisProLeuAsnLysLeuGlnGluAlaTrpArg
                             480                                 510                                 540
AGTATTCACTACCCACCTGAGACTGCAAGCATCATGTTGATGGCTAGGATGGTGGCCACAGTGAAGCAGGCGAAGGACAAGGACCGTTGG
SerIleHisTyrProProGluThrAlaSerIleMetLeuMetAlaArgMetValAlaThrValLysGlnAlaLysAspLysAspArgTrp
                             570                                 600                                 630
ATCAGACTCTTTTCCCAGTTTTGTAACAAAACAGCCAATGAAGAGGAGGAAATTGTCCATAAACTTCTGGGAGACAAATTCAAGGGCCAA
IleArgLeuPheSerGlnPheCysAsnLysThrAlaAsnGluGluGluGluIleValHisLysLeuGlyAspLysPheLysGlyGln
                             660                                 690                                 720
CTGGAACTTCTGCGGAGACTCTTCACAGAGGCCCTCTATGAGGAAGCAGTCAGCCAGTGGTTCACTCCAGATGGATTCCGGTCTCTCTTT
LeuGluLeuLeuArgArgLeuPheThrGluAlaLeuTyrGluGluAlaValSerGlnTrpPheThrProAspGlyPheArgSerLeuPhe
                             750                                 780                                 810
GCTCTTGTTGGGACCAATGGCCAAGGAATCGGGACCAGCTCCCTAAGCCAGTGGGTCCATGCCTGTGACACTCTGGAGTTGAAGCCTCAG
AlaLeuValGlyThrAsnGlyGlnGlyIleGlyThrSerSerLeuSerGlnTrpValHisAlaCysAspThrLeuGluLeuLysProGln
                             840                                 870                                 900
GACCGTGAGCACGTTGACGCCTTCATTGACCAGCTATACAAGGACATCGAGGCAGCAACTGGAGAGTTTCTTAACTGTGAAGGATCTGGC
AspArgGluHisValAspAlaPheIleAspGlnLeuTyrLysAspIleGluAlaAlaThrGlyGluPheLeuAsnCysGluGlySerGly
                             930                                 960                                 990
CTCTTTGTGCTTCAGAGCTGCTGCAACCACAGTTGTGTGCCCAATGCAGAGACCTCCTTTCCAGAAAACAACTTCCTTTTGCATGTCACT
LeuPheValLeuGlnSerCysCysAsnHisSerCysValProAsnAlaGluThrSerProPheProGluAsnAsnPheLeuLeuHisValThr
                            1020                                1050                                1080
GCTCTGGAGGATATTAAGCCAGGAGAGGAAATTTGTATCAGCTACTTGGACTGCTGTCAGCGGGAGCGCAGCCGCCACAGCCGCCACAAG
AlaLeuGluAspIleLysProGlyGluGluIleCysIleSerTyrLeuAspCysCysGlnArgGluArgSerArgHisSerArgHisLys
                            1110                                1140                                1170
ATCCTCAGGGAGAACTATCTATTTGTCTGTTCCTGTCCCAAATGCCTGGCAGAGGCTGATGAACCCAATGTGACCTCAGAAGAGGAAGAG
IleLeuArgGluAsnTyrLeuPheValCysSerCysProLysCysLeuAlaGluAlaAspGluProAsnValThrSerGluGluGluGlu
                            1200                                1230                                1260
GAAGAGGAGGAGGAGGAAGGAGAGCCAGAAGATGCAGAGCTGGGGGATGAGATGACTGATGTGTGATGTTGCCCTGCCCAGAAAGGG
GluGluGluGluGluGluGlyGluProGluAlaGluLeuGlyAspGluMetThrAspVal*
                            1290                                1320                                1350
CCCTGCCCTAGACCCTGCCAGAAAAGGGGGCTCTTCCCCCAGAGAAGTGGCTTGGAGGGAACTTCCCACTCCCATTGCCTGCTTTCCCCA
                            1380                                1410                                1440
TTCCAGCCCTCTCTGCTAGAGGGTAGGAGAGAGCCTGGATCTCTGGCCCCAACCCCCACCAGACCTCATGCCCTGGACACTGCTGCTGAG
                            1470                                1500                                1530
TTGGCTCAGACTCTGCACTGGCACTGAGCCTTTCACAACTGGCCTCCCCTTGAGGTTCCTCCACTCTAGGGTTTGAGGGGCTGGAATCAG
                            1560                                1590                                1620
GGCCAGGGCCTAACAGTGTTTCTTCCCCTCGGCCCCACGGCCCATACTCACCCCTTCACCTGAGGCTTCTCCCCTCTCAACTTGCTGTTG
                            1650                                1680                                1710
ATTTCTTTTGAGGGGTAAAAGAAAGACCGAGTTCATTGAAGCAGAGATGGGAGGTGGTGAGGCGTCCTCACCTCCCCCCACCACAGCTGT
                            1740                                1770                                1800
GGCTGGTAAGGCAACTGCTTGACCACTGGCCCAGCAAGCAGAGGCGGGCAGGAGGATCTGGGCATCCTGAGCCCCTTCCCTGAGGCTGTC
                            1830                                1860                                1890
TCCTGGGAATGCTGGACCAGGGGGATTACTTCCCCCACCCACACCCCCATTTGGACATCTGGGGAAGTGGAGATTAACCCTCCCCCAGAC
                            1920                                1950                                1980
TCGTGCCTTGCCTCTCAGGACCTGGCACAGTGGCTGGCCTTATGGAGTATGGAAAAGGCCTCTGCCTTCCTCAGGAGGGACACATAGTGG
                            2010                                2040                                2070
GTCATCACCCTTTCTCACACCTGCTTCCCAATGAGCACTCAGCAGCATGGCTGAGACCTGGGGCTTGACAGGCCTGCGCTCCTTAGATTA
                            2100                                2130                                2160
GGATGCTCTGGTATTCCTGGGGCCAGTTAAAATGGGTCAGTGAGGGGCTCTCCTGGCCTCAATCTCCATCTTTAGGGTCTCCACAGACTA
                            2190                                2220                                2250
ACATTGAAGTCTTCCAGAAACCTCGAGGAGCTGGGCTAGATACTTCGAGAGTAGCCCTGTCTTGTCCCCTCCTACCCCCAGATTAGCACC
                            2280                                2310                                2340
ACCCCTCTCACTCTTGGGTTGGATGCTATGGGAATTACAACCCATCACTCCCAATGCCTCCCTTTCCTGGAGGAGCCTGCTGCATTCTCC
                            2370                                2400                                2430
TGCTGAGGGATCCCAGAGAGGGCAGGTGCCCACCCCACAGGCCCTTTCCTGGACCATGGAGCCAGGTGAAACCTGTCTGGCCCCTTACCC
                            2460                                2490                                2520
ATTGGGTTCTTACTCCTTCACAACGGTCAGCACTCCTCTCCCAGCCCCCCTGCCCACTGTTCTCAATAAAATGTGAGTGGTGAAAAAAAA
```

FIG. 3. Homology of NN8-4 and human EST41159 as revealed by a BLAST search. (A) Alignment of the sequences of the NN8-4 fragment and EST41159. The RXRE in NN8-4 is underlined. (B) Nucleotide and predicted amino acid sequence of EST41159. The polyadenylation signal is underlined, and the termination codon is marked with an asterisk.

and -translated receptors. The NN8-4 RXRE–RXR-RAR complex is effectively inhibited by competition with excess unlabeled NN8-4 RXRE, βRARE (a DR5), and CRBPII RXRE (a DR1) (Fig. 5). A similar amount of unlabeled nonspecific competitor does not interfere with binding. When RXR alone is combined with labeled NN8-4 RXRE, a shifted band is visible, albeit with much less intensity than that of the band seen with both RXR and RAR. With the addition of ligand 9cRA, the amount of probe bound by RXR is increased to a level greater than that seen with the RXR-RAR heterodimer. This enhancement of RXR homodimer binding to a DR1 has been documented previously (38, 58). The RXRE coded for in NN8-4 interacts with receptors in a manner similar to that of previously characterized RXREs binding both the RXR-RAR heterodimer and RXR homodimer-9cRA with high affinity (37, 58).



B

```
   1 TTGACCTTGAACTTCTGGTCCTCCCTGCCCTCACCTCCTGAATGCAGAGGTTATAGGCAT
  61 AGTCGGAATGCTGGTTTTAAGCAGTGTTGTGGCTCAAAGTTCTGCACTCATGTGTTCAAA
 121 GCAAATCTCTCCCAAACATAGAATAAAAATGTTGCCACCAGAGGGAGGGAAAAATTTTAA
 181 ACCGAGTGAGCCCAGCATCCAACCAGGATGGAGGGTCAAAGCAAAAATAAGACAGTCTTC
 241 TCTGGGTCTAAAAACTTTGTTCGGTGTTGGAGGTAATGACGAGGGGCTGCTTAAGCAGGA
 301 AGAGGGCTGCAAACTTTTGGCCAGCCTGGGATACATAGTGAGACCCTGTCTCAAACACCA
 361 GACAGTGGGAAGAGCAACTATGGACACTTAAGAACACAGCTAGGATAGGCTGGAGGGTGG
 421 GGGGTGGGGAGTAGGGGGTCTTCCCCTGAGCTACTCCATCAGCCGTCATTCATGCTGTCC
 481 ATCCGACATAAGCTACCGAGTCAGACACAGTCTACCAGAAGTCCAAGCACAAAGTCAGGC
 541 CAGGTTCCACCCAATCAAGCAACCAAGACACCAAAAGCCTTGAATTGCTAGCCACGTCCC
 601 AGGGCGGGGTAAGGTGGGGAAGGGAGCGCTCGCTGGCTCTGTGGGCGGGGCCTCCCGGGA
 661 CGGGGTCAAGGGTCAGAAGGCAGAGGCGTGCCCAAGATGGCGGCCTCCATGTGCGACGTG
                                   MetAlaAlaSerMetCysAspVal
 721 TTCTCCTTCTGCGTGGGCGTGGCGGACCGTGCCCGGGGTTCCGTGGAAGTCCGTTATGTG
     PheSerPheCysValGlyValAlaAspArgAlaArgGlySerValGluValArgTyrVal
 781 GACAGCATCAAGGTGAGGCCAGGTCGGACCGGGCCCGGAGAGCCACGCGGGGGAACCCAC
     AspSerIleLys
 841 GCTCACATTCTGGAGTCCGGGCCCCTCGGCGCCAGTCTGGTTGCTGCTTGGGCCCGGGGT
 901 GCAGAGTCGGTCCTCGTGCCCAGGCTCGTGGCGATATCGCGCGGGGGGAGGGAGGAAAGC
 961 CCGGTCCACTGGCGGGTGGCCGTGCTGCTGCTGTCCTGGGTGGAGTTGAGGCTCTGGGTA
1021 GACCTGGGGTTGAGGGATATCCTGGGCCAGATTAAAGCCAAGGCTATCTTGGTTAGGCTG
1081 GGGCTTAGTTAGCTACTATTCTGGCGGTAGTGAGGGAGTACTTAGCCGAGG
```

C

```
mNN-8AG     1 MAASMCDVFSFCVGVADRARGSVEVRYVDSIKGKGLFATQLIRKGETIFIERPLVAAQPL
                ||||||||| ||| |||||:| | ||||||||||||||||||||:||||||||||
hEST41159       .....AVFSFCVGVAGRARVSVEVRFVSSAKGKGLFATQLIRKGETIFVERPLVAAQPL

mNN-8AG    61 WNALYQYRACDHCLRALEKAEENAQRLTGKPSQILPHPELCSVRKDLHQNCPHCQVMYCS
              |||||  |||||||||||||||||||||||||||| |:|||||||:|||||||||||||||
hEST41159       WNALYRYRACDHCLRALEKAEENAQRLTGKPGQVLPHPELCTVRKDLHQNCPHCQVMYCS

mNN-8AG   121 AECRLAAAEQYHQILCPGPSHD-PRHPLNKLQEAWRGVHYPPETASIMLMARMVATVKQA
              |||||| |||||:|||||| | | ||||||||||| :|||||||||||||||||||||||
hEST41159       AECRLAATEQYHQVLCPGPSQDDPLHPLNKLQEAWRSIHYPPETASIMLMARMVATVKQA

mNN-8AG   181 ~DKDHWVRLFNHFCSRTANQEQAIVHKLLKGKFKDQLELLLGLFKEALYEEALSLWFTPE
               ||| |:|||: ||::|||:|: ||||||  ||| |||||  || |||||||:| ||||:
hEST41159       KDKDRWIRLFSQFCNKTANEEEEIVHKLLGDKFKGQLELLRRLFTEALYEEAVSQWFTPD

mNN-8AG   240 GFRSLFALVGTNGQGIATSSLSQWVHACDALDWTPQDREQLDTFIDQLYKDIEAATGEFL
              |||||||||||||||| |||||||||||| |:  |||| :| ||||||||||||||||||
hEST41159       GFRSLFALVGTNGQGIGTSSLSQWVHACDTLELKPQDREHVDAFIDQLYKDIEAATGEFL

mNN-8AG   300 NCEGSGLFVLQSC..
              |||||||||||||
hEST41159       NCEGSGLFVLQSCCN ...
```

FIG. 4. Characterization of the murine NN8-4-associated gene (NN8-4AG) and its putative CpG island. (A) Schematic representation of the genomic region flanking NN8-4. NN fragment 8-4, used as a probe to isolate genomic clone λNN8-4-5, is shown at top. The shaded box indicates the position of the RXRE. On the λNN8-4-5 schematic, exons are indicated with solid boxes. Below the λNN8-4-5 schematic is a representation of the 1 kb surrounding NN8-4, with the RXRE and exon 1 being indicated as shaded and striped boxes, respectively, and CpG dinucleotides being represented by vertical bars. (B) Nucleotide sequence of the CpG island associated with NN8-4AG. The amino acid sequence of the first exon is indicated below the DNA sequence. The nucleotide corresponding to the farthest 5′ transcriptional start site as determined by RACE analysis is indicated by a star above the sequence. The RXRE and the CCAAT and GC boxes are underlined. (C) Comparison of the putative amino acid sequence of the first 10 exons of the murine NN8-4AG gene product and the EST41159 protein sequence. Identical amino acids are indicated by a vertical line, and similar amino acids are shown with two dots.

The NN8-4AG transcript is directly upregulated by 9cRA in F9 cells. Our functional analyses suggest that the RXRE associated with NN8-4AG may be biologically relevant. As the human cDNA and murine NN8-4AG sequences obtained from genomic clones were highly conserved, we used EST41159 to probe RNA derived from F9 EC cells at various stages of RA treatment. A transcript with a size of approximately 2.75 kb was present at low basal levels and was upregulated in the presence of RA (Fig. 6). Addition of either atRA or 9cRA results in a four- to fivefold increase in the expression of NN8-4AG after 24 h of treatment (Fig. 6, lanes 4 and 8). Although the basal level of NN8-4AG mRNA varied by about twofold in F9 EC cells under these culture conditions, it is evident that upregulation occurred more quickly in the presence of atRA (Fig. 6, lane 2). However, it is likely that this atRA-induced gene expression is not a primary effect, as pretreatment with protein synthesis inhibitor Cx negates the increase in expression seen at 12 h of treatment with the retinoid (Fig. 6, lane 11). No additional protein synthesis is necessary for the upregulation upon 9cRA treatment, as 12 h levels with Cx and 9cRA are similar to those with 9cRA alone (Fig. 6, lane 10). Although NN8-4AG is responsive to both atRA and 9cRA, it appears that activation with these ligands occurs via two different pathways. The Northern blot was stripped and probed with genes known to be directly regulated by RA, RARβ, and CRABPII (Fig. 6). As has been demonstrated

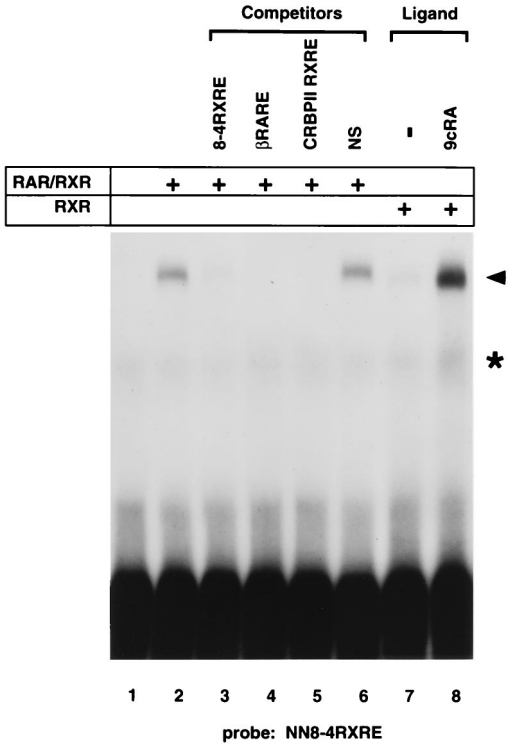


FIG. 5. Binding of RXR-RAR heterodimers and RXR homodimers to the NN8-4 RXRE. End-labeled NN8-4RXRE was incubated with RRL (total, 4 μl) programmed with human RARα1 and/or murine RXRβ mRNA. The probe was also incubated with the same amount of unprogrammed lysate as the control (lane 1). Cold competitors were used at a molar excess of 150-fold. 9cRA was added to the binding reaction mixture to a final concentration of 1 μM. The arrowhead indicates the specific complexes, while the asterisk indicates a non-specific (NS) complex.

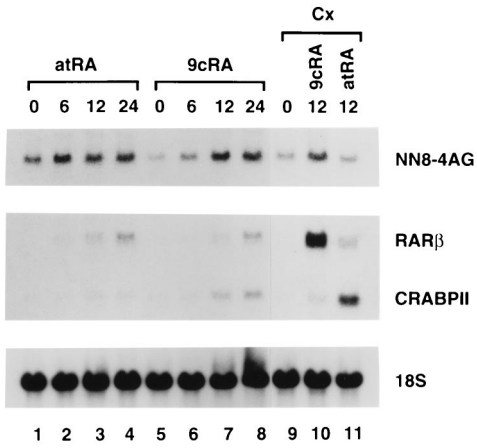


FIG. 6. Northern analysis of NN8-4AG mRNA. EST41159 was used to probe total RNA derived from F9 EC cells which had been grown in Dulbecco modified Eagle medium supplemented with charcoal-stripped and UV-treated serum. RNA was isolated at the indicated times (in hours, indicated above the gels) after RA treatment. Retinoids were added to a concentration of 1 μM. Cx was added to a final concentration of 10 μg/ml 1 h prior to RA treatment. For time 0 samples, Cx was added 1 h before the cells were harvested. RARβ and CRABPII, genes known to be induced by RA, were used as controls. An oligonucleotide hybridizing to the 18S rRNA was used to normalize for RNA loading.

previously, RARβ is upregulated in response to both atRA and 9cRA (19, 21). 9cRA activates expression of CRABPII more effectively than atRA (21, 44). Although expression levels of both mRNAs at 12 h with retinoid and Cx treatment indicate that upregulation occurs with no additional protein synthesis, RARβ is superinduced by a combination of 9cRA plus Cx while CRABPII is superinduced only in the presence of atRA and Cx. This observation provides further evidence that the expression of these two genes is controlled by distinct retinoid signaling pathways.

Evolutionary conservation of EST41159 protein. The putative gene product of the human homolog of NN8-4AG, EST41159, was analyzed with the BLASTX program and the databases of the National Center for Biotechnology Information (2). The predicted protein sequence of EST41159 showed significant similarity with that of a yeast (*Saccharomyces cerevisiae*) putative polypeptide of, as yet, unknown function (accession number P38890). The similarity corresponds to amino acids 308 to 381 in the human gene product and 333 to 403 in the yeast protein, with a P value of 8.95×10^{-7} (Fig. 7). This region also has a high degree of identity with the C-terminal region of the *Drosophila* trithorax protein (39), a member of a group of proteins which are thought to participate in chromatin remodeling (41).

DISCUSSION

To be physiologically relevant, the search for transcription factor binding sites should concentrate not only on DNA fragments located in the regulatory region of genes but also on transcription units actively transcribed in the tissue or cells of interest. The previous demonstration that active chromatin located within CpG islands can be separated from bulk inactive chromatin and that a specific fraction of nucleosome-free DNA associated with these islands contains the promoter region of transcribed genes (50) provides an elegant way to identify pertinent genomic fragments. These small DNA fragments (NN fragments) represent a highly purified fraction of the genome and can, in theory, be selected further for their ability to bind to transcription factor complexes. On the basis of these premises, we have developed a method for the isolation of RXR-RAR binding sites from CpG island DNA, isolated 12 RA response elements, and characterized in detail a novel RA-responsive gene associated with one of them.

The advantages of the CpG-SAAB method are numerous for exploring the molecular mechanisms of retinoid receptors and other transcription factors. First, in contrast to such subtraction screening strategies as differential display PCR, only genes directly regulated by the receptors are cloned. Second, because NN fragments are derived from active chromatin, genes primed to respond to retinoids in the appropriate tissue or cell line will be preferentially identified. Third, the promoter and/or the regulatory region of these genes is likely to be identified first, thus providing molecular tools in the form of natural promoters and enhancers to study the action of retinoid receptors. Fourth, since this strategy targets the promoters of genes, the 5′ ends of the associated transcripts, which are regions of cDNA that are often difficult to isolate, are in many cases likely to be cloned first. Finally, CpG-SAAB can be applied to any combination of transcription factor and cell line or tissue of interest. A specific set of responsive genes is expected to be identified for each combination of transcription factor and starting chromatin.

Although not all genes are associated with CpG islands, all housekeeping and a significant fraction (40%) of tissue-specific genes are (3). Since CpG island DNA composes approximately 2% of the genome, the bulk of genomic DNA is excluded from the selection. In the case of the retinoid receptors, at least, it

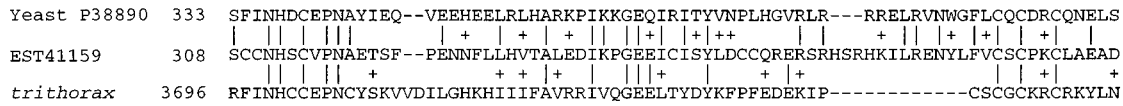


FIG. 7. Conservation of the predicted amino acid sequence of human EST41159 relative to the sequences of a yeast putative protein and the *Drosophila trithorax* gene product. Amino acids that are identical are indicated by a vertical lines, while conservative changes are designated by plus signs. The GenBank accession numbers for the putative yeast and trithorax proteins are P38890 and Z31725, respectively.

may be prudent to avoid the use of naked DNA as starting material for binding site selection, as members of a major *Alu* repeat subfamily and major (gamma) satellite DNA repeats contain RAREs (45, 54). Although some of these repeats may play a role in regulating the expression of nearby genes, these elements may also interfere with the isolation of other relevant binding sites because of their number and affinity for the receptors. According to preliminary analysis of a number of the RARE-containing NN fragments by genomic Southern blotting, these fragments are derived from single-copy genomic DNA, a situation previously documented for the CpG island fraction of the genome (17). In addition, molecular and genetic analyses of retinoid action have shown that RXR-RAR complexes appear to play a pleiotropic role in the control of embryonic development and in homeostasis in the adult animal (46). Retinoid receptors are therefore likely to be involved in the general maintenance of the cellular machinery and, by inference, could regulate a large number of housekeeping genes.

We chose to use murine EC cell lines as starting material for CpG-SAAB. F9 EC cells differentiate into blastocyst-like endoderm cells upon RA treatment and provide a model system in which the genetic cascade following retinoid treatment mimics early development. CpG-SAAB enabled us to isolate a population of DNA fragments containing a variety of RAREs and RXREs. The sequencing of 120 clones resulted in the discovery of 26 that contained functional RAREs and RXREs. Fragments that did not contain any obvious consensus RA response elements were likely selected by proteins present in the RRL, and in practice, their number could be reduced by introducing an additional purification step, such as immunoprecipitation with specific receptor antibodies.

The majority of the selected NN fragments containing the response elements bind the RXR-RAR heterodimer with an affinity comparable to that of the fragment containing the potent βRARE. Analysis of these fragments shows that complex elements are relatively common but also provides further evidence that the most potent RAREs are of the DR5 type (Table 2). Surprisingly, low total transcriptional activity was generated when NN8-3, a fragment containing a consensus DR5 with half-sites of sequence (GGGTCA and AGGTCA), was tested in the presence of RA. On the other hand, NN7-91, which contains a DR5 with the same half-site sequences but in the opposite order, confers a high level of RA-induced transcriptional activity on TKLUC. A number of factors could be responsible for this difference in activities. The composition of the half-sites, spacer, and nucleotides 5′ of the upstream half-site may be critical in the determination of transcriptional activity imparted by the receptors. For example, several of the nuclear receptors which bind as monomers have been shown to require an A-T-rich region extending 5′ from the core half-site (27, 30, 57). If a monomeric receptor binding site is embodied within a RARE, monomeric nuclear receptors may interfere with the binding of RXR-RAR heterodimer complexes (51). Finally, the context of the RARE within the genomic fragment itself may be critical, as the sequences flanking the RAREs may contain binding sites for additional transcription factors which may modulate receptor activity (40). The cell-type-specific activity of the NN-RAREs may be imparted by the presence or absence of these factors within a particular cell type. The identification of a large collection of natural RAREs and RXREs by CpG-SAAB and their in-depth characterization will allow us to test these parameters in a systematic manner.

NN fragments containing functionally competent RAREs identified by CpG-SAAB can subsequently be used to isolate large flanking genomic DNA fragments containing the associated CpG island. Cross et al. (17) have demonstrated that cloned CpG islands can be used to detect transcripts of expressed genes at a relatively high frequency. We indeed observed that a genomic fragment containing NN7-91 detects both a conserved locus in several species by zoo blot analysis and a transcript in F9 EC cell RNA (data not shown). In addition to these traditional methods of searching for transcripts, we are also focusing on sequence and database analysis. Particularly useful to this type of project are the EST databases. By software trapping, we have been able to quickly identify a transcript associated with one of the retinoid receptor-selected NN fragments (NN8-4). While the proximity of the RXRE contained in NN8-4 to the transcriptional start site of the associated gene (NN8-4AG) allowed for the identification of this transcribed sequence, sequencing of genomic DNA flanking the remaining selected NN fragments (1 to 2 kb in each direction) would greatly increase the probability of finding a match.

We have shown that the NN8-4AG transcript is differentially upregulated by atRA and 9cRA (Fig. 6). Treatment of F9 EC cells with either atRA or 9cRA leads to an induction of the expression of the NN8-4AG gene, but activation by atRA seems to require the synthesis of an additional factor(s). We have demonstrated that the NN8-4 fragment contains a DR1 RXRE that is recognized with high affinity by RXR-RAR heterodimers and by RXR homodimers in the presence of 9cRA and is able to confer RA responsiveness to the TK promoter (Fig. 5 and Table 2). It can therefore be assumed that the 9cRA response is mediated via the DR1 RXRE, either through RXR homodimers (58), RXR-RAR heterodimers, or a heterodimer complex in which RXR acts as a 9cRA-responsive partner to another member of the nuclear receptor family (for examples, see references 33, 42, and 56). Since 9cRA can also bind RAR, it is possible that ligand occupation by both of the receptors is required for the 9cRA response. Conflicting results concerning the ability of RXR to bind ligands in the context of an RXR-RAR heterodimer have been obtained in studies with receptor-specific ligands (6, 22, 34, 44). Further study will be required to address the nature of the dimer mediating the 9cRA response of NN8-4AG. It is not yet known whether the protein synthesis-dependent response to atRA also involves the DR1 element.

In spite of sequence analysis and database searches, the function of the novel gene product coded for by EST41159 and NN8-4AG is unknown. This result is not unexpected, as up to 50% of ESTs from the human brain are unique and do not

show significant similarity to proteins with known biological functions (1). However, the observed similarity for the C-terminal domain of EST41159, a putative yeast protein, and the *Drosophila trithorax* gene product does delineate an evolutionarily conserved domain. This region of the *trithorax* gene product has been previously shown to share similarity with a number of proteins, some that are known to be associated with chromatin and others with unknown functions (47). The region of similarity we have delineated is not as extensive as that previously defined, however; like those of the other homologous proteins, the motif lies at or near the C terminus of otherwise nonhomologous proteins. The yeast putative protein (P38890) mentioned in this study and EST41159 share regions of homology throughout their entire sequences, suggesting they are more closely related to one another.

The CpG-SAAB method described here allowed us to isolate a subset of retinoid receptor binding sites and identify a novel RA-responsive gene with an as-yet-unknown function. It is hoped that the characterization of a large number of genes associated with these elements will help chart the genetic pathways implicated in the early response of EC cells to RA and improve our understanding of the mechanisms regulating cellular differentiation induced by hormonal stimuli.

ACKNOWLEDGMENTS

We thank A. Matthyssen, G. Flock, and A. Moraitis for technical assistance, J. Trasler for the 18S oligonucleotide sequence and hybridization protocol, and R. Heyman for providing 9cRA. We thank R. Sladek and L. D. B. McBroom and other members of the Giguère laboratory as well as A. Nepveu for comments on the manuscript.

This research was supported by grants from the National Cancer Institute of Canada and the Medical Research Council of Canada (MRC). V.G. is an MRC Scientist, and M.S. is a recipient of an MRC studentship.

REFERENCES

1. **Adams, M., A. R. Kerlavage, C. Fields, and J. C. Venter.** 1993. 3400 new expressed sequence tags identify diversity of transcripts in human brain. Nat. Genet. **4:**256–267.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Antequera, F., and A. Bird.** 1993. Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. USA **90:**11995–11999.
4. **Antequera, F., J. Boyes, and A. P. Bird.** 1990. High levels of de novo methylation and altered chromatin structure at CpG islands. Cell **62:**503–514.
5. **Antequera, F., D. Macleod, and A. P. Bird.** 1989. Specific protection of methylated CpGs in mammalian nuclei. Cell **58:**509–517.
6. **Apfel, C. M., M. Kamber, M. Klaus, P. Mohr, S. Keidel, and P. K. LeMotte.** 1995. Enhancement of HL-60 differentiation by a new class of retinoids with selective activity on retinoid X receptor. J. Biol. Chem. **270:**30765–30772.
7. **Bento Soares, M., M. de Fatima Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis.** 1994. Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. USA **91:**9228–9232.
8. **Bigler, J., and R. N. Eisenman.** 1994. Isolation of a thyroid hormone-responsive gene by immunoprecipitation of thyroid hormone receptor-DNA complexes. Mol. Cell. Biol. **14:**7621–7632.
9. **Bigler, J., and R. N. Eisenman.** 1995. Novel location and function of a thyroid hormone response element. EMBO J. **14:**5710–5723.
10. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature (London) **321:**209–213.
11. **Bird, A. P.** 1987. CpG islands as gene markers in the vertebrate nucleus. Trends Genet. **3:**342–347.
12. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
13. **Boguski, M. S., T. M. J. Lowe, and C. M. Tolstoshev.** 1993. dbEST—database for "expressed sequence tags." Nat. Genet. **4:**332–333.
14. **Bouillet, P., M. Oulad-Abdelghani, S. Vicaire, J.-M. Garnier, B. Schuhbaur, P. Dollé, and P. Chambon.** 1995. Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal carcinoma cells and characterization of a novel mouse gene, Stra1 (mouse LERK-2/Eplg2). Dev. Biol. **170:**420–433.
15. **Boyes, J., and A. Bird.** 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell **64:**1123–1134.
16. **Chambon, P.** 1994. The retinoid signaling pathway: molecular and genetics analyses. Semin. Cell Biol. **5:**115–125.
17. **Cross, S. H., J. A. Charlton, X. Nan, and A. P. Bird.** 1994. Purification of CpG islands using a methylated DNA binding column. Nat. Genet. **6:**236–244.
18. **de Thé, H., M. del Mar Vivanco-Ruiz, P. Tiollais, H. Stunneberg, and A. Dejean.** 1990. Identification of a retinoic acid responsive element in the retinoic acid receptor β gene. Nature (London) **343:**177–180.
19. **de Thé, H., A. Marchio, P. Tiollais, and A. Dejean.** 1989. Differential expression and ligand regulation of retinoic acid receptor α and β genes. EMBO J. **8:**429–433.
20. **Dony, C., M. Kessel, and P. Gruss.** 1985. Post-transcriptional control of *myc* and p53 expression during differentiation of the embryonal carcinoma cell line F9. Nature (London) **317:**636–639.
21. **Durand, B., M. Saunders, P. Leroy, M. Leid, and P. Chambon.** 1992. All-trans and 9-cis retinoic acid induction of CRABPII transcription is mediated by RAR-RXR heterodimers bound to DR1 and DR2 repeated motifs. Cell **71:**73–85.
22. **Forman, B. M., K. Umesono, J. Chen, and R. M. Evans.** 1995. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. Cell **81:**541–550.
23. **Fromm, M., and P. Berg.** 1983. Transcription in vivo from SV40 early promoter deletion mutants without repression by large T antigen. J. Mol. Appl. Genet. **2:**127–135.
24. **Gardiner, K., and R. J. Mural.** 1995. Getting the message: identifying transcribed sequences. Trends Genet. **11:**77–79.
25. **Giguère, V.** 1994. Retinoic acid receptors and cellular retinoid binding proteins: complex interplay in retinoid signaling. Endocr. Rev. **15:**61–79.
26. **Giguère, V., M. Shago, R. Zirngibl, P. Tate, J. Rossant, and S. Varmuza.** 1990. Identification of a novel isoform of the retinoic acid receptor γ expressed in the mouse embryo. Mol. Cell. Biol. **10:**2335–2340.
27. **Giguère, V., M. Tini, G. Flock, E. S. Ong, R. M. Evans, and G. Otulakowski.** 1994. Isoform-specific amino-terminal domains dictate DNA-binding properties of RORα, a novel family of orphan nuclear receptors. Genes Dev. **8:**538–553.
28. **Gould, A. P., J. L. Brookman, D. I. Strutt, and R. A. H. White.** 1990. Targets of homeotic gene control in *Drosophila*. Nature (London) **348:**308–312.
29. **Gudas, L. J., M. B. Sporn, and A. B. Roberts.** 1994. Cellular biology and biochemistry of the retinoids, p. 443–520. *In* M. B. Sporn, A. B. Roberts, and D. S. Goodman (ed.), The retinoids, 2nd ed. Raven Press, New York.
30. **Harding, H. P., and M. A. Lazar.** 1993. The orphan receptor Rev-ErbAα activates transcription via a novel response element. Mol. Cell. Biol. **13:**3113–3121.
31. **Jones-Villeneuve, E. M. V., M. McBurney, K. A. Rogers, and V. I. J. Kalnins.** 1982. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glia cells. J. Cell Biol. **94:**253–262.
32. **Kinzler, K. W., and B. Vogelstein.** 1989. Whole genome PCR: application to the identification of sequences bound by regulatory proteins. Nucleic Acids Res. **17:**3645–3653.
33. **Kliewer, S. A., K. Umesono, D. J. Noonan, R. A. Heyman, and R. M. Evans.** 1992. Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. Nature (London) **358:**771–774.
34. **Kurokawa, R., J. DiRenzo, M. Boehm, J. Sugarman, B. Gloss, G. M. Rosenfeld, R. A. Heyman, and C. K. Glass.** 1994. Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. Nature (London) **371:**528–531.
35. **Lewis, J. D., R. R. Meehan, W. J. Henzel, I. Maurer-Fogy, P. Jeppesen, F. Klein, and A. Bird.** 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell **69:**905–914.
36. **Mangelsdorf, D. J., K. Umesono, and R. M. Evans.** 1994. The retinoid receptors, p. 319–349. *In* M. B. Sporn, A. B. Roberts, and D. S. Goodman (ed.), The retinoids, 2nd ed. Raven Press, New York.
37. **Mangelsdorf, D. J., K. Umesono, S. A. Kliewer, U. Borgmeyer, E. S. Ong, and R. M. Evans.** 1991. A direct repeat in the cellular retinol binding protein type II gene confers differential regulation by RXR and RAR. Cell **66:**555–561.
38. **Medin, J. A., S. Minucci, P. H. Driggers, I. J. Lee, and K. Ozato.** 1994. Quantitative increases in DNA binding affinity and positional effects determine 9-cis retinoic acid induced activation of the retinoid X receptor β homodimer. Mol. Cell. Endocrinol. **105:**27–35.
39. **Mozer, B. A., and I. B. Dawid.** 1989. Cloning and molecular characterization of the *trithorax* locus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:**3738–3742.
40. **Nakshatri, H., and P. Chambon.** 1994. The directly repeated RG(G/T)TCA motifs of the rat and mouse cellular retinol-binding protein II genes are promiscuous binding sites for RAR, RXR, HNF-4, and ARP-1 homo- and heterodimers. J. Biol. Chem. **269:**890–902.
41. **Orlando, V., and R. Paro.** 1995. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. Curr. Opin. Genet. Dev. **5:**174–179.

42. **Perlmann, T., and L. Jansson.** 1995. A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. Genes Dev. **9:**769–782.
43. **Richard, S., and H. H. Zingg.** 1991. Identification of a retinoic acid response element in the human oxytocin promoter. J. Biol. Chem. **266:**21428–21433.
44. **Roy, B., R. Taneja, and P. Chambon.** 1995. Synergistic activation of retinoic acid (RA)-responsive genes and induction of embryonal carcinoma cell differentiation by an RA receptor α (RARα)-, RARβ-, or RARγ-selective ligand in combination with a retinoid X receptor-specific ligand. Mol. Cell. Biol. **15:**6481–6487.
45. **Rudert, F., S. Bronner, J. M. Garnier, and P. Dollé.** 1995. Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm. Genome **6:**76–83.
46. **Sporn, M. B., A. B. Roberts, and D. S. Goodman (ed.).** 1994. The retinoids, 2nd ed. Raven Press, New York.
47. **Stassen, M. J., D. Bailey, S. Nelson, V. Chinwalla, and P. J. Harte.** 1995. The *Drosophila trithorax* proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech. Dev. **52:**209–223.
48. **Strickland, S., and V. Mahdavi.** 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell **15:**393–403.
49. **Sucov, H. M., K. K. Murakami, and R. M. Evans.** 1990. Characterization of an autoregulated response element in the mouse retinoic acid receptor type β gene. Proc. Natl. Acad. Sci. USA **87:**5392–5396.
50. **Tazi, J., and A. Bird.** 1990. Alternative chromatin structure at CpG islands. Cell **60:**909–920.
51. **Tini, M., R. A. Fraser, and V. Giguère.** 1995. Functional interactions between retinoic acid-related orphan nuclear receptor (RORα) and the retinoic acid receptors in the regulation of the γF-crystallin promoter. J. Biol. Chem. **270:**20156–20161.
52. **Tini, M., G. Otulakowski, M. L. Breitman, L.-T. Tsui, and V. Giguère.** 1993. An everted repeat mediates retinoic acid induction of the γF-crystallin gene: evidence of a direct role for retinoids in lens development. Genes Dev. **7:**295–307.
53. **Umesono, K., V. Giguère, C. K. Glass, M. G. Rosenfeld, and R. M. Evans.** 1988. Retinoic acid and thyroid hormone induce gene expression through a common responsive element. Nature (London) **336:**262–265.
54. **Vansant, F., and W. F. Reynolds.** 1995. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. Proc. Natl. Acad. Sci. USA **92:**8229–8233.
55. **Vasios, G., S. Mader, J. D. Gold, M. Leid, Y. Lutz, M. P. Gaub, P. Chambon, and L. Gudas.** 1991. The late retinoic acid induction of laminin B1 gene transcription involves RAR binding to the responsive element. EMBO J. **10:**1149–1158.
56. **Willy, P. J., K. Umesono, E. S. Ong, R. M. Evans, R. A. Heyman, and D. J. Mangelsdorf.** 1995. LXR, a nuclear receptor that defines a distinct retinoid response pathway. Genes Dev. **9:**1033–1045.
57. **Wilson, T. E., T. J. Fahrner, M. Johnson, and J. Milbrandt.** 1991. Identification of the DNA binding site for NGFI-B by genetic selection in yeast. Science **252:**1296–1300.
58. **Zhang, X.-K., J. Lehmann, B. Hoffmann, M. I. Dawson, J. Cameron, G. Graupner, T. Hermann, P. Tran, and M. Pfahl.** 1992. Homodimer formation of retinoid X receptor induced by 9-cis retinoic acid. Nature (London) **358:**587–591.