THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 52, Issue of December 29, pp. 31178–31188, 1995
Printed in U.S.A.

A Novel Homeobox Protein Which Recognizes a TGT Core and Functionally Interferes with a Retinoid-responsive Motif*

(Received for publication, September 5, 1995, and in revised form, October 20, 1995)

Eric Bertolino, Bernard Reimund, Dunja Wildt-Perinic, and Roger G. Clerc‡

From Roche Ltd., Research Laboratories, CH-4070 Basel, Switzerland

We describe here a novel homeobox gene, denoted TGIF (5′TG3′ interacting factor), which belongs to an expanding TALE (three amino acid loop extension) superclass of atypical homeodomains. The TGIF homeodomain binds to a previously characterized retinoid X receptor (RXR) responsive element from the cellular retinol-binding protein II promoter (CRBPII-RXRE), which contains an unusual DNA target for a homeobox. The interactions of both the homeoprotein TGIF and receptor RXR*α* with the CRBPII-RXRE DNA motif occur on overlapping areas and generate a mutually exclusive binding *in vitro*. Transient cellular transfections demonstrate that TGIF inhibits the 9-*cis*-retinoic acid-dependent RXR*α* transcription activation of the retinoic acid responsive element. TGIF transcripts were detected in a restricted number of tissues. The canonical binding site of TGIF is conserved and is an integral part of several responsive elements which are organized like the CRBPII-RXRE. Hence, a novel auxiliary factor to the steroid receptor superfamily may participate in the transmission of nuclear signals during development and in the adult, as illustrated by the down-modulation of the RXR*α* activities.

Homeobox genes play a fundamental role in directing cellular differentiation processes and in determining cell fate. Over the past 10 years, the term homeodomain has evolved to define a superfamily of protein domains of ~60 amino acids with homology to the *Drosophila* homeotic proteins (15). Homeoproteins confer the specificity of action to a wide variety of transcription factors. They exert their action both by their DNA binding surfaces and by domains that are targets for protein: protein interactions with other transcription factors (16–18).

Regulated, tissue-specific, and developmental expression of eukaryotic genes results from the interplay of a variety of transcription factors, like the homeoproteins. They exert their effects on target genes by both activating and repressing transcriptional activities.

Vitamin A (retinol) and other retinoids, like the retinoic acid (RA),[1] were demonstrated to exert striking effects on cell proliferation and differentiation. Excessive intake as well as deficiency of vitamin A generate characteristic toxicity and malformation patterns in a number of organ systems. Retinoic acid as well as a number of small lipophilic hormones mediate their signals through ligand-activated transcription factors belonging to the large steroid/retinoid nuclear receptors superfamily (19). Two classes of retinoid receptors have been identified; the retinoic acid receptors (RARs) and the retinoid X receptors (RXRs) (20, 21). Homo- as well as hetero-, dimers of these receptors act in response to retinoids by binding to specific *cis*-acting retinoid-responsive promoter elements (22, 23), thereby generating a large diversity of transcriptional controls in the retinoid signaling pathways (24). The expression of several homeogenes was demonstrated to be differentially regulated by RA, and this suggests that homeogenes are likely to control the temporal and spatial modulation of the levels of endogenous retinoids (25).

Recently, the diversity of nuclear receptor-mediated control was found to be further extended by the synergy of other transcription factors. The interaction of retinoid receptors and transcription factors of the c-Jun and c-Fos family (AP-1), for example, can either repress or potentiate the retinoid-dependent transcription activation (26, 27). Therefore, there exist regulatory "cross-talk" pathways that allow modulation of the retinoid signal by the AP-1 signaling system (28).

There are two classes of cytoplasmic retinoid-binding proteins implicated in the transduction of the retinoid signal which also play an important role in retinoid homeostasis: the cellular retinoic acid-binding proteins, CRABPI and -II, and the cellular retinol-binding proteins CRBPI and -II (for review, see Ref. 29). CRBPII is expressed mostly in prenatal liver and in adult intestine (29) and is probably involved in the regulation of the vitamin A signaling pathway by controlling the intracellular transport and storage of retinol, a precursor of retinoic acid (30).

We show here the functional cloning of a new member of the homeobox gene superfamily, called TGIF, that belongs to a growing superclass of atypical homeodomains, whose hallmark is an extension of three amino acids between α helices 1 and 2. The TGIF homeoprotein recognizes a previously characterized retinoid response motif (CRBPII-RXRE) which consists of an unusual DNA target for a homeobox. TGIF can prevent the retinoid X (RXR) receptor from functioning as a transcriptional activator through interference with the previously characterized CRBPII-RXRE-responsive element (1, 2). The consensus binding site of TGIF was identified and is conserved in several CRBPII-RXRE-like responsive elements, suggesting a broader functional association of TGIF with this type of responsive element.

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

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X89750.

‡ To whom correspondence should be addressed: Roche Ltd., Rm. 69-209, Grenzacherstrasse 124, CH-4070 Basel, Switzerland. Tel.: 41-61-688-20-56; Fax: 41-61-688-45-75.

[1] The abbreviations used are: RA, retinoic acid; RAR, retinoic acid receptor; RARE, retinoic acid response element; RXR, retinoid X receptor; RXRE, retinoid X response element; PCR, polymerase chain reaction; COUP, chicken ovalbumin upstream promoter; CRBP, cellular retinol-binding protein; CRABP, cellular retinoic acid-binding protein; CAT, chloramphenicol acetyltransferase; GST, glutathione *S*-transferase; EMSA, electrophoretic mobility shift assay.

MATERIALS AND METHODS

Cloning and Sequencing of TGIF—A Southwestern protocol (31) was used to screen a λgt11 human embryonic liver cDNA library (Clontech, Catalog No. HL1005b) with a radiolabeled oligonucleotide containing three ligated copies of the CRBPII-RXRE (Fig. 1). The choice of the library was dictated by the high level of RXRα expression in liver. The screening of 10^6 plaques enabled the identification of a clone which was plaque-purified after two additional expression screening rounds. The positive clone was used as a template to obtain an end-labeled PCR probe with two primers (GATCCGAATTCGGAGATCCAGAATG and GGTCCAAGCTTACTCCACAGAGCTC). This probe was used to re-screen the human embryonic liver cDNA library and a 5′-methylmercuric hydroxide stretched λgt11 human placenta cDNA library (Clontech, Catalog No. HL1075b). Additional positive clones were obtained, and two overlapping λgt11 clones were analyzed further. Their cDNA inserts were isolated either by PCR or enzymatically and subcloned in the *Eco*RV site of the pSK$^-$ plasmid (Stratagene). These cDNA clones (pSK$^-$-TGIF), which once combined, contained the open reading frame (ORF) of TGIF, were enzymatically sequenced on both strands using the dideoxynucleotide termination procedure (U. S. Biochemical Corp.) and a Sequenase polymerase (U. S. Biochemical Corp.) using oligonucleotide primers. Both the PCR and the directly subcloned fragment showed an identical nucleotide sequence except for one error introduced by PCR on a wobble position which did not affect the corresponding amino acid sequence. This PCR fragment was used in the experiments described herein.

RNA Analysis—A 300base pair fragment representing the amino-terminal coding region of TGIF was ^{32}P-labeled by the random primers method and used to detect transcripts on a human tissues blot (Clontech) with 2 μg of oligonucleotide dT-selected RNA from different human tissues. Hybridizations were performed in 50% formamide, 5 × SSC, 5 × Denhart's solution, 50 mM sodium phosphate buffer (pH 6.8), 10 mM EDTA, 0.1% SDS, 10^6 cpm of DNA probe/ml, 200 μg of carrier DNA per ml for 12 h at 42 °C. Membranes were washed successively in 6 × SSC, 0.1% SDS and 1 × SSC, 0.1% SDS at room temperature. The amount of RNA loaded was scored by hybridizing the membranes to a control human β-actin cDNA probe.

Expression of TGIF for DNA Binding Studies or for Raising Polyclonal Rabbit Antibodies—A Y1089 lysogen strain expressing TGIF was isolated to express the β-galactosidase fusion protein. Extracts were prepared according to Ref. 31. The pSK$^-$-TGIF plasmid was digested with *Sau*3A or *Bst*UI-*Dra*I, and the DNA fragments encoding the TGIF homeodomain or the complete open reading frame were cloned into the pBSATG vector (32) in the *Eco*RV site (see Δ*TGIF* in Fig. 3*A*). A pBSATG-Oct plasmid (32) was linearized with *Pst*I for the expression of Oct-2 POU domain. *In vitro* transcription was performed using 1 μg of linearized pBSATG-derived templates in a 100-μl reaction containing 200 mM Tris-HCl (pH 7.5), 30 mM $MgCl_2$, 10 mM spermidine, 50 mM NaCl. The mixture was incubated for 1.5 h at 37 °C, phenol-chloroform-extracted, and ethanol-precipitated, and the pellets were resuspended in 10 μl of diethyl pyrocarbonate water. *In vitro* translation was carried out for 1 h at 30 °C in a 50-μl reaction containing 1 μl of the *in vitro* transcribed RNA, 10 μCi of L-[^{35}S]methionine (Amersham Catalog No. SJ1015), an amino acid mixture (minus methionine), and 35 μl of rabbit reticulocyte lysate both from Promega (Catalog No. L4960). The translation products were analyzed on SDS-polyacrylamide gel electrophoresis and showed the expected length.

The GST fusion proteins were constructed by ligating in-frame, respectively, in the *Bam*HI or in the *Eco*RI site from the pGEX-2T vector (Pharmacia): (i) a PCR DNA fragment encoding the TGIF homeodomain plus seven unrelated amino acids (LVPRGSI) at the amino terminus (see *GST-HD* in Fig. 3*B*), (ii) the complete open reading frame of TGIF contained in a *Bst*UI-*Dra*I-cut DNA fragment. The pGEX-TGIF full-length DNA fragment was deleted by *Sma*I and religated in-frame for the expression of the carboxyl-terminal part of TGIF lacking the homeodomain. 2 ml of a saturated overnight culture of *E. coli*, transformed with the pGEX-2T-TGIF vector, were grown in 50 ml of LB medium prior to isopropyl-1-thio-β-D-galactopyranoside (2 mM final) induction. After a 3-h induction, the bacteria were pelleted. Whole cell extracts were prepared by using a freeze-thaw method (0.4 M KCl buffer) (33). The GST-TGIF was purified by batch adsorption to glutathione-Sepharose 4B (Pharmacia Biotech Inc.). The amino-terminally truncated TGIF protein fused with the GST was isolated from a SDS-polyacrylamide gel electrophoresis, electroeluted in 50 mM (pH 8.9) $(NH_4)_2CO_2$ and 0.1% SDS solution, and dialyzed against phosphate-buffered saline to remove SDS before injecting the rabbits. Polyclonal antibodies were raised and tested for specificity according to standard techniques.

Generation of Recombinant TGIF-Autographa californica Multiple Nuclear Polyhedrosis Virus for Baculovirus Expression of TGIF—A *Bam*HI-*Ssp*I DNA fragment, containing the complete coding region of TGIF, was digested from the pGEX-2T-TGIF plasmid and was cloned into Klenow-blunted *Bam*HI-*Nco*I sites in the pBlueBacHis vector, designed for generation of recombinant baculovirus (Invitrogen). *Spodoptera frugiperda* ovarian (Sf9) cells (2 × 10^6 Sf9 cells in TC-100 medium (Life Technologies, Inc.) without fetal calf serum) were co-transfected with 1 μg of wild type *A. californica* multiple nuclear polyhedrosis virus and 1 μg pBlueBacHis-TGIF. After 96 h, screening for recombinant virus in infected Sf9 cells was performed by serial dilution of the transfection medium supplemented with 5-bromo-4-chloro-3-indolyl β-D-galactoside for colorimetric test of positives. A culture of 10^8 Sf9 cells was infected with recombinant TGIF-*A. californica* multiple nuclear polyhedrosis virus (infection multiplicity = 5–10). After a 1-h infection in 25 ml of TC 100 medium with 5% fetal calf serum, the total volume was expanded up to 100 ml in a spinner flask. Cells were harvested 72 h after infection, washed once with phosphate-buffered saline buffer, and resuspended in 1 ml of high salt extraction buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM dithiothreitol, 600 mM KCl, and protease inhibitors, leupeptin, pepstatin, chymostatin, soybean trypsin inhibitor, aprotinin (all at 10 mg/ml), and phenylmethylsulfonyl fluoride (1 mM)). Whole cell protein extracts were obtained by lysing the cells with 30 strokes in a 2-ml glass Dounce homogenizer and by centrifuging twice at 10,000 × *g* for 30 min at 4 °C in order to remove cell debris.

Dimethyl Sulfate Methylation Interference and Electrophoresis Mobility Shift Assay (EMSA)—The EMSAs were performed with the different recombinant proteins as described in Ref. 34 and the probes shown in Fig. 1. For the G-specific dimethyl sulfate methylation interference experiment, the CRBPII-RXRE, cloned in the *Eco*RI site of pBluescript (Stratagene), was removed with *Hin*dIII-*Pst*I or *Bam*HI-*Hin*cII. This enabled the ^{32}P labeling of the coding and the noncoding strand with Klenow polymerase. DNA probe fragments were isolated by electroelution after polyacrylamide gel separation from the vector backbone and methylated by dimethyl sulfate (Fluka) for 2 min at room temperature (35). Methylated DNA probe (1.5–2 × 10^5 cpm/300 ng of DNA) was incubated with RXRα isolated from baculovirus or GST-TGIF isolated from bacterial cell extracts and analyzed in an EMSA. Complexed probe and free probe were separated in 6% nondenaturing polyacrylamide gel (acrylamide/bisacrylamide ratio 19:1), and both were isolated by electroelution and then ethanol-precipitated. Pellets were dissolved in 10% piperidine (DuPont NEN), and the DNA was cleaved at 90 °C for 30 min before lyophilization followed by two rounds of washing with 20 μl of water. The hydrolysis products were resolved on 8% denaturing acrylamide/urea gels. The dried gels were exposed to x-ray films (Fuji, Inc) with an intensifying screen at −70 °C for 12–36 h.

Cells and Extracts—COS-1 cells and U87 cells (glioblastoma, astrocytoma, grade III) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 200 mM glutamine, and penicillin and streptomycin. For the transfection experiments, 5 × 10^7 cells (5 ml of suspension) were plated in a 6-cm diameter Petri dish in order to reach 70% confluency at the transfection time. After transfection, cells were treated with 9-cis-RA (see *Transfection Experiments and CAT Assay*). Whole cell extracts were prepared by using a freeze-thaw protocol (0.4 M KCl buffer) (33).

Recombinant Plasmids for Transient Transfection Assays—The sequences of the DNA-responsive elements used in the reporter plasmids from this study are shown in Fig. 1. The CRBPII-RXRE promoter element (2, 12) (see also Fig. 1) or mutated/deleted responsive elements were inserted in the 5′ end of the TATA tk promoter sequence and the CAT (chloramphenicol acetyltransferase) gene reporter sequence from the pBLCAT5 vector (36) between the *Bam*HI and the *Hin*dIII sites. The pcDNA-TGIF recombinant plasmid consists of the TGIF cDNA sequence included between the *Bst*UI site (21 base pairs upstream of the initiation initiation codon) and the *Ssp*I site (position 1223 in Fig. 2*A*) which was cloned between the *Eco*RI and the *Eco*RV sites of the pcDNA vector (Invitrogen). The pSG5-RXRα recombinant contains the RXRα ORF inserted in the pSG5 vector (37); the TGIF and RXRα cDNAs were introduced in expression vectors having two different promoters (respectively from cytomegalovirus and SV40) in order to avoid competition between them and therefore to optimize the expression of both proteins. A pcDNA-βGal construct was included as an internal control to standardize the different transfection assays.

Transfection Experiments and CAT Assay—Cells were transfected by calcium phosphate co-precipitation (38) with 5 μg of one of the recombinant pBLCAT reporters in the presence or absence of 3 μg of pcDNA-TGIF and/or with various amounts of pSG5-RXRα depending

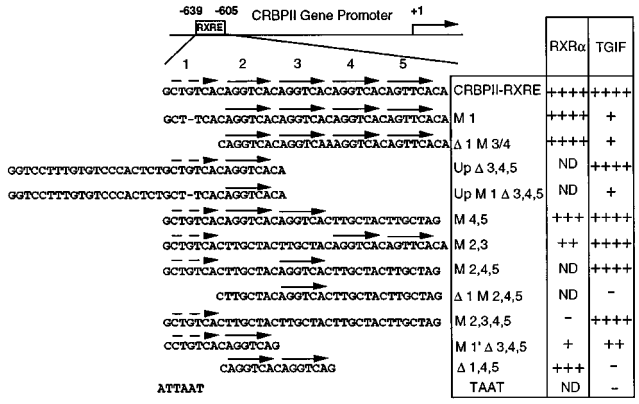

FIG. 1. **DNA probes used in this study and summary of the *in vitro* DNA binding of RXRα and TGIF.** The organization of the cellular retinol-binding protein promoter element (CRBPII-RXRE) located between positions −639 and −605 from the CRBPII gene is depicted. The *arrows* numbered from *2* to *5* indicate DR1 half-sites. The *dashed arrow numbered 1* indicates a divergent DR1 half-site. The nomenclature used for the identification of the DNA-responsive elements is Δ for deletions and *M* for nucleotide mutations or single nucleotide deletions. The probes, denoted *Up Δ3,4,5* and *Up M1 Δ3,4,5*, contain both upstream sequences flanking the CRBPII-RXRE (positions −670 to −639 in the CRBPII promoter (12)) and the wild type half-sites 1 and 2 for Up Δ3,4,5 and the half-site 1 mutated and the wild type half-site 2 for Up M1 Δ3,4,5. *In vitro* DNA affinities of RXRα and TGIF to the different probes are reported: ++++, very strong binding; +++, strong binding; ++, significant binding; +, weak binding; −, no binding; *ND*, not determined.

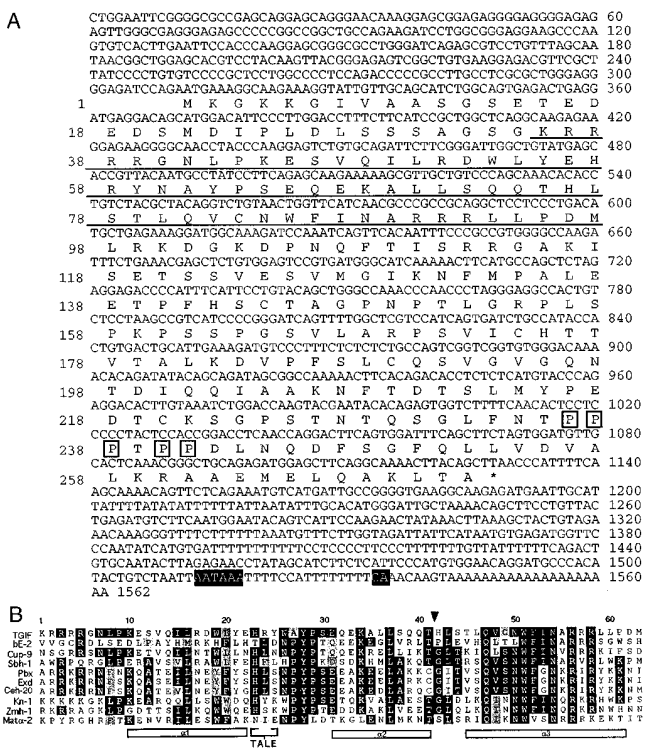

FIG. 2. **TGIF cDNA sequence and alignment of atypical homeodomains from the TALE (three amino acid loop extension) superclass.** *A*, the sequences of two independently isolated clones were combined to form a 1.562-kilobase-long cDNA fragment which contained the complete TGIF open reading frame. The deduced amino acid sequence is shown as *single-letter code*. The cDNA sequence (*numbering on the right*) encodes a 272-amino acid-long protein (*numbering on the left*). The atypical homeodomain sequence is *underlined*. The *boxed* prolines in the carboxyl-terminal region indicate a putative SH3 domain binding site (13). The polyadenylation consensus sequence is shown in *reverse lettering*. *B*, alignment of TGIF homeodomain with several TALE homeodomains. The alignment of the amino acid sequences, collected from the GenBank and EMBL data bases, was performed according to the algorithm Pile-up/Prettybox included in the GCG (University of Wisconsin) software package. Identical amino acids between the sequences are shown on *black background*, whereas the conserved amino acids are shown on *gray background*. For maximizing identities, 15 amino acids have been deleted in the bE-2 homeodomain at the position indicated by a *solid triangle*. The positions of the three α helices (α1, α2, and α3) are indicated by *open rectangles* as in Ref. 14. The position of the three amino acid loop extension (residues 23–25) is indicated by *TALE*.

on the cell type used (see Fig. 6). The endogenous TGIF levels were tested, and no significant amounts were detected either by Northern or by EMSA experiments.[2] In order to detect the TGIF-mediated transactivation regulation, we titrated out the TGIF and the RXRα activities to establish the optimal ratio of TGIF to RXRα. Optimal transactivation was obtained by using 1 μg of the RXRα effector plasmid for the U87 cells and 0.25 μg for the COS-1 cells. 0.5 μg of the pcDNA-βGal internal control was included in each experiment to standardize the transfection efficiency. Finally, calf thymus DNA was added as double-stranded carrier DNA to equalize the DNA concentrations in each precipitate. In several cases, we also used the pcDNA and/or the pSG5 vectors without insert in order to control for a plasmid-driven effect on the expression modulation of the reporter.

The precipitate was left on the cells for 16–20 h before the medium was changed. Cells were incubated for another 20–24 h in the presence of 10^{-7} M 9-*cis*-RA. Cells were harvested, and extracts were prepared. The normalized CAT activity assay was run as described previously (39). Each transfection experiment was repeated at least three times with different plasmid preparations. The percent of chloramphenicol acetylation was determined by thin layer chromatography followed by quantification in a PhosphorImager (Fuji). The values always agreed within 15% from one experiment to another.

PCR Binding Site Selection—Enrichment for binding sites from a random oligonucleotide pool (5′-GGCTGAGTCTGAACGGATCC(N_{15})-CCTCGAGACTGAGCGTCG-3′) was performed as described in Ref. 40. Binding reactions were carried out with purified GST-HD (see above). EMSAs were carried out as described in Ref. 34. After three rounds of enrichment, the selected oligonucleotides were cloned with the CloneAMP®pAMP System for Rapid Cloning of Amplification Products (Life Technologies, Inc., Catalog No. 18381-012).

The EMBL Data Library accession number for the human TGIF cDNA clone is X89750.

RESULTS

Isolation and Molecular Characterization of a New Homeobox Protein Binding to the Rat CRBPII-RXRE—The promoter region located between positions −639 and −605 of the rat cellular retinol-binding protein II (CRBPII) gene (12) was previously characterized as an optimal retinoid DNA-responsive element (CRBPII-RXRE) for RXRα, -β, and -γ (2, 41). This CRBPII-RXRE promoter element is composed of five almost perfectly conserved directly repeated half-sites with the consensus hexanucleotide sequence 5′AGGTCA3′. These hexamer half-sites are spaced by one nucleotide (Figs. 1 and 4) and generate a series of direct repeats (DR1). We have numbered these half-sites from 1 to 5 in Figs. 1 and 4. Half-sites 1 and 5 diverge from the canonical hexamer sequence.

To identify novel nuclear factors that specifically interact with the rat CRBPII-RXRE and which could interfere or synergize with RXRα or other retinoid receptor molecules on this promoter region, we have screened a human liver cDNA expression library cloned into the λgt11 phage vector with a radiolabeled double-stranded DNA probe consisting of 3 copies of the CRBPII-RXRE (see Fig. 1). Upon screening of a total of 10^6 plaques, two clones were isolated and their sequences were combined to form a 1562-nucleotide-long cDNA. This cDNA contains an open reading frame (ORF) encoding a protein of 272 amino acids, hereafter called TGIF for 5′TG3′ interacting factor (see Fig. 2*A*).

The initiation codon of TGIF occurred at the first in-frame ATG from the cDNA. The sequence context of this ATG con-

[2] E. Bertolino, B. Reimund, D. Wildt-Perinic, and R. G. Clerc, unpublished data.

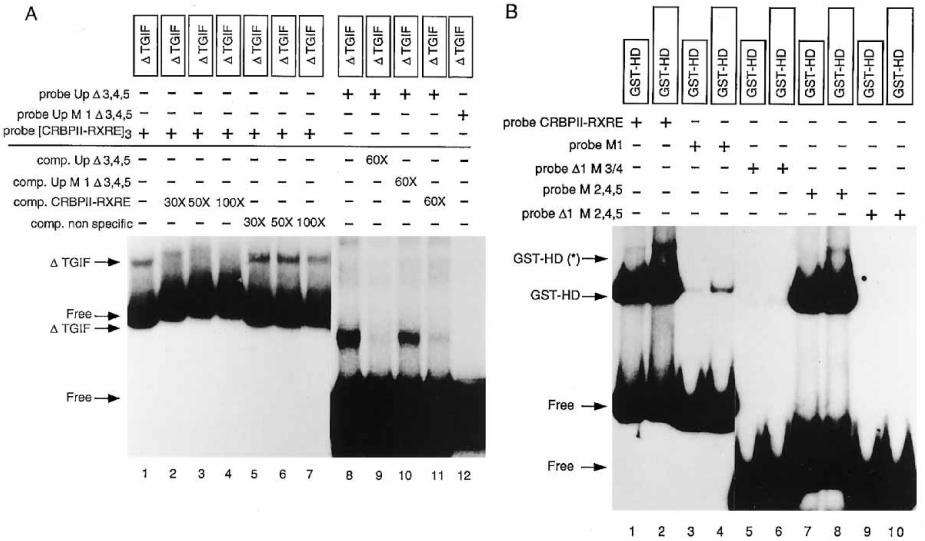

FIG. 3. **TGIF homeodomain interacts specifically with the retinoid-responsive element CRBPII-RXRE.** *A*, binding of a 106-amino acid-long TGIF polypeptide containing the TGIF homeodomain (Δ*TGIF*) to different probes (see Fig. 1). *[CRBPII-RXRE]*$_3$ stands for trimer of the CRBPII-RXRE. The protein·DNA complexes were specifically competed with an increasing molar excess (as indicated) of unlabeled competitor (*comp.*) oligonucleotides. *B*, binding of a bacterially expressed TGIF homeodomain-GST fusion protein (*GST-HD*) to different probes (see Fig. 1). * indicates a GST-HD oligomer. *Boxes* of different sizes represent two different concentrations of GST-HD. Total protein concentration in the binding reactions has been equilibrated with nonprogrammed cell extracts.

forms to that expected for an initiation codon (42). Furthermore, a nonsense codon TGA is found 100 base pairs upstream of this first ATG of the open reading frame. Another ATG codon containing an optimal context for initiation of translation occurs at position 372 of the cDNA in the reading frame (Fig. 2*A*). Nevertheless, the amino terminus of the protein has been assigned to the 5′-most ATG codon because *in vitro* translation of the full-length 272-amino acid-long cDNA has clearly demonstrated that the proximal ATG was exclusively utilized as an initiation codon.[2] Furthermore, a comparison of the mouse homologue to this human TGIF cDNA showed that the assigned initiation codon and its nucleotide context are fully conserved in both human and mouse cDNA sequences.[2] Analysis of the predicted reading frame encoding 272 amino acids revealed homology from amino acid positions 35–98 with homeodomains from different species. The carboxyl-terminal part of the new TGIF homeoprotein is rich in proline (Fig. 2*A*) and contains a putative SH3 binding domain (*X*PPP*X*PP *boxed* in Fig. 2*A*) (13). Proline-rich sequences were also implicated in transcriptional regulation suggesting a possible function for this homeoprotein (43, 44). A search of the sequences deposited in the latest release (88) of the GenBank and EMBL data bases revealed several entries with small DNA sequence stretches identical with the TGIF homeoprotein. These sequences, however, were generated by random DNA sequencing performed with human cDNA libraries. This implies that TGIF is a novel homeoprotein.

TGIF Belongs to a Large Superclass of TALE-atypical Homeodomains—The amino acid matches between the TGIF homeodomain and nine closely related homeodomains are shown in Fig. 2*B*. It shows a group of atypical homeodomains (14) whose hallmark is an extension or a deletion by several amino acids of the canonical Antennapedia consensus homeodomain. The TGIF homeodomain showed the highest degree of homology with the homeoboxes encoded by the bE2 alleles of *Ustilago maydis* (for maximizing identities, 15 amino acids have been deleted at the position indicated by the *solid triangle* in Fig. 2*B*), the *Saccharomyces cerevisiae* copper homeostasis CUP9 gene, and the yeast MAT α2 mating-type regulatory gene. Besides a conservation of the invariant amino acids in the DNA recognition α helix 3, the TGIF homeodomain sequence shares highly conserved residues in a 3-amino acid elongated loop between α helices 1 and 2 (consensus His^{23} and Lev^{24}). Another stretch of conserved residues among these homeodomains is located between positions 27 and 29 (Pro^{27}, Tyr^{28}, Pro^{29}). We suggest calling this growing group of atypical homeodomains the TALE homeodomain superclass for three amino acid loop extension (see Fig. 2*B* and "Discussion"). The highly conserved 3-amino acid-long loop structure is present in homeoproteins from different species ranging from yeast to human and must have an important role in the activities of these regulatory proteins.

TGIF Binds to a T/G-rich Region Located 5′ to Directly Repeated AGGTCA Half-sites—The DNA sequence-specific binding of TGIF was analyzed by testing a series of DNA probes by EMSA. To localize the domain of TGIF involved in DNA binding, a 106-amino acid-long amino-terminal fragment which contained the full-length homeodomain was synthesized *in vitro* in a rabbit reticulocyte lysate (Δ*TGIF* in Fig. 3*A*). As shown in Fig. 3*A*, *lane 1*, ΔTGIF generated a complex with the trimeric DNA probe used in the Southwestern gene screening. This complex was specifically competed by increasing concentrations of unlabeled specific CRBPII-RXRE DNA (Fig. 3*A*, *lanes 2–4*) but not by identical concentrations of a nonspecific oligonucleotide DNA (Fig. 3*A*, *lanes 5–7*).

To further analyze the DNA sequence requirements for TGIF, we tested the binding of the ΔTGIF polypeptide to a CRBPII promoter fragment overlapping only the upstream region of the CRBPII-RXRE probe (CRBPII promoter positions −659 to −624 and see also in Fig. 1, *Up Δ3,4,5*) (12). As shown in Fig. 3*A*, *lane 8*, the additional nucleotides upstream from the CRBPII-RXRE did not affect the DNA binding capability of ΔTGIF, indicating that half-sites 1 and 2 were sufficient. Specificity of the DNA binding was evaluated by competing this probe with three different unlabeled oligonucleotides. Strong competition was observed with the oligonucleotides Up Δ3,4,5 and CRBPII-RXRE (Fig. 3*A*, *lanes 9* and *11*). Identical amounts of the upstream CRBPII promoter fragment but mutated in half-site 1 (Fig. 1, *Up M1 Δ3,4,5*), could however compete only very weakly the ΔTGIF homeodomain·DNA complex (Fig. 3*A*, *lane 10*). These results suggest that the ΔTGIF domain binds specifically to the 5′ region of the CRBPII-RXRE which contains half-site 1. We concluded that ΔTGIF interacts strongly with a guanosine residue which is located at position −636 in the Up Δ3,4,5 probe and deleted in the Up M1 Δ3,4,5. The requirement for a guanosine residue at this position was confirmed by the lack of binding of TGIF to the labeled Up M1 Δ3,4,5 probe lacking guanosine −636 (Fig. 3*A*, *lane 12*). The full-length TGIF homeoprotein synthesized *in vitro* and the ΔTGIF (see above) had DNA binding properties identical with the CRBPII-RXRE probe (Fig. 5*A*, *lane 2*).

The results obtained with the help of point-mutated or par-

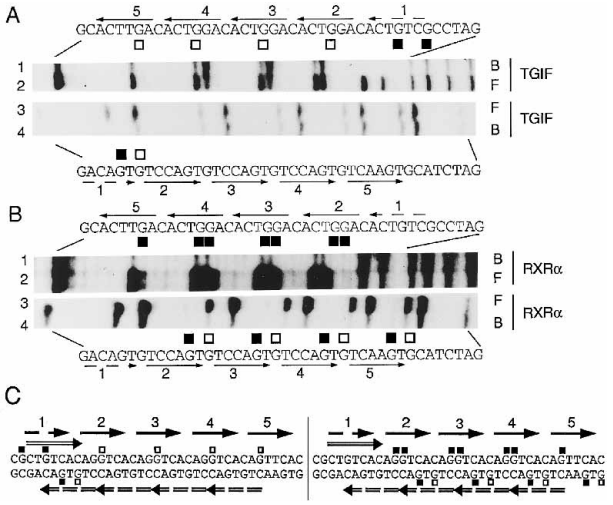

FIG. 4. **Overlapping binding of RXRα and TGIF to the CRBPII-RXRE.** G-specific methylation interference contacts of bacterially expressed GST-TGIF fusion protein (*A*) and of RXRα isolated from baculovirus-infected cell extracts with the partially methylated CRBPII-RXRE probe (*B*). In *A* and *B*, the transcribed strand pattern is shown on the *top* and the nontranscribed strand on the *bottom* (see Fig. 1 for nucleotide numbering), and the DR1 half-sites 1, 2, 3, 4, and 5 are indicated by *single arrows*. The *black squares* indicate strong methylation interferences, and the *open squares* mark weak interferences. *F* and *B* stand for free probe and bound probe, respectively. *C*, summary of the G-specific methylation interference contacts of TGIF (*left*) and RXRα (*right*). *Double arrows* indicate TGIF sites, *single arrows* indicate RXRα sites. *Double-dashed arrows* and *dashed arrows* depict, respectively, divergent TGIF sites and divergent RXRα half-sites. *Black* and *open squares* are as in *A* and *B*.

tially deleted responsive elements used as probes or competitors in EMSAs (Fig. 3, *A* and *B*) were confirmed by using a G-specific methylation interference protocol. To this end, we overexpressed in *Escherichia coli* a GST-TGIF fusion protein. Two complexes with different electrophoretic properties were visible in an EMSA after co-incubation of the GST-TGIF fusion protein and the CRBPII-RXRE probe. The GST-TGIF was probably processed by proteolytic cleavage into two DNA-binding polypeptides. Both polypeptides interfered, however, with identical residues in dimethyl sulfate mapping (data not shown). The CRBPII-RXRE·TGIF interactions indicated that the methylated G residues at position −639 and −636 from the transcribed strand (*solid squares* in Fig. 4*A*, *lanes 1* and *2*) and at position −634 from the nontranscribed strand (*solid square* in Fig. 4*A*, *lanes 3* and *4*) interfered strongly with the protein·DNA complex formation. This result is in agreement with that described in Fig. 3, *A* and *B*, in which we demonstrated that the deletion of residue Gly^{-636} in the Up M1 Δ3,4,5 DNA probe almost abolished TGIF binding. Interestingly, the residues located at position −629, −622, −615, and −609 on the coding strand, and −632 from the nontranscribed strand were slightly undermethylated, indicating a detectable but partial interference in the complex formation onto the half-sites 2, 3, 4, and 5. These results led us to conclude that the strongest interferences are restricted to four residues within an eight-nucleotide-long region which reads 5′GCTGTCAC3′ (*double arrow* in Fig. 4*C*). Inspection of the CRBPII-RXRE indicated that this sequence is partially conserved in four direct repeats overlapping half-sites 2, 3, 4, and 5 as shown in Fig. 4*C* (*double-dashed arrows*). These repeats which read 5′CTGTGAC3′ may well be weak binding sites for TGIF and could therefore explain the partial methylation interference in the TGIF·DNA complex detected on those repeats (Fig. 4*A*, *lanes 1*, *2*, *3*, and *4*).

To test the existence of these weak binding sites, *E. coli*-overexpressed glutathione *S*-transferase TGIF homeodomain (GST-HD) fusion protein was produced and used in an EMSA. In Fig. 3*B*, *lanes 1* and *2*, a second, slower migrating, complex was detected upon addition of larger amounts of GST-HD to the CRBPII-RXRE probe indicating that GST-HD bound to additional sites besides half-site 1. Probe M2,4,5 in which mutations were introduced in half-sites 2, 4, and 5 generated a weaker, retarded TGIF·DNA complex (Fig. 3*B*, *lanes 7* and *8*). Further, larger amounts of GST-HD allowed the detection of a weak binding to the probe M1 (Fig. 3*B*, *lanes 3* and *4*) and only residual DNA binding of GST-HD to the probes Δ1 M3/4 or Δ1 M2,4,5 was detectable (Fig. 3*B*, *lanes 5*, *6*, *9*, and *10*), suggesting the existence of weak TGIF binding sites with the DNA sequence 5′CTGTGAC3′ (*double-dashed arrows* in Fig. 4*C*). As summarized in Figs. 1 and 4*C*, the results obtained from both the G-specific methylation interference and a series of DNA binding assays, indicated that TGIF binds strongly to half-site 1 and weakly to sites located between the RXR half-sites 2, 3, 4, and 5 of the CRBPII-RXRE.

Since most homeodomains contact a 5′ATTAAT3′ recognition DNA sequence (15), we tested by EMSA whether or not TGIF had any affinity for a homeobox consensus DNA sequence. In fact, there is no detectable binding of the *in vitro*-expressed full-length TGIF protein to the 5′ATTAAT3′ consensus sequence motif (Fig. 5*A*, *lane 3*), in contrast to the Oct-2 POU domain (32) which recognized this probe specifically (Fig. 5*A*, *lanes 4–7*). In identical conditions, TGIF recognized the CRBPII-RXRE as shown in Fig. 5, *lane 2*.

Taken together, the results raise the question whether or not the non-TAAT recognition site of TGIF (5′GCTGTCAC3′) in the CRBPII-RXRE is the cognate TGIF binding site. To address this question, a random binding site selection was performed. The TGIF homeobox (GST-HD), expressed in *E. coli*, was challenged to determine the nucleotide preferences from a pool of oligonucleotides containing 15 randomized nucleotides. High affinity binding sites were selected in three successive rounds of EMSAs by testing three different GST-HD protein concentrations. A set of 42 sequences were analyzed and revealed a consensus core sequence which reads 5′TGTCA3′ as illustrated in Fig. 5*B*. The consensogram, derived from the 42 selected sequences, demonstrates that the 5′TGTCA3′ core sequence is extremely well conserved at each position (Fig. 5*C*). Furthermore, the two first nucleotides upstream of the core sequence are preferentially a G or a C residue, whereas the nucleotides A or T are over-represented downstream of the core sequence (Fig. 5*C*).

Interestingly, the TGIF binding site in the CRBPII-RXRE conforms well with the consensus binding site determined by random binding site selection experiment. As shown in Table I, the TGIF core binding site is also conserved in several retinoid/steroid receptor cognate binding sites from human, rat, mouse, and/or chicken gene promoters. The TGIF core binding site is flanked in all the described cases (Table I) with the consensus half-site motif (5′AGGTCA3′) recognized by the zinc finger-containing nuclear receptors. The TGIF core binding site and the 5′AGGTCA3′ motif form the two half-sites contained in imperfect direct or inverted repeats generally spaced by one nucleotide (Table I), as in the CRBPII-RXRE.

In Vitro Overlapping Binding of TGIF and RXRα on the CRBPII-RXRE—To evaluate the binding properties of RXRα to the CRBPII-RXRE, we have generated extracts from *S. frugiperda* Sf9 cells which overexpressed RXRα upon infection with a recombinant *A. californica* baculovirus. The dose-response curves corresponding to the quantification of the different probes complexed with increasing concentrations of RXRα in the EMSA indicated that the RXRα homo-cooperativity was dependent on the number of conserved hexanucleotide half-

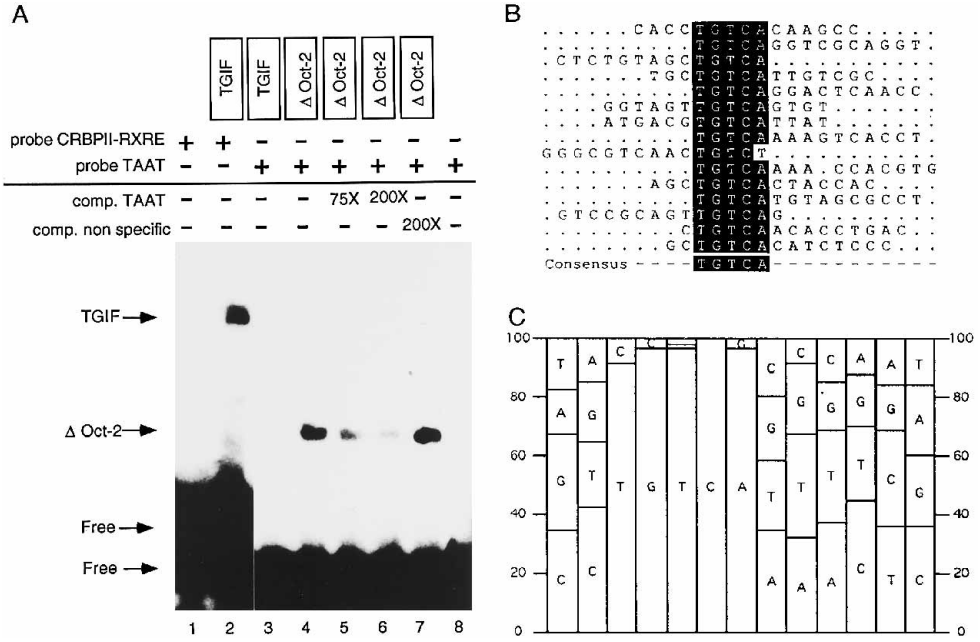

FIG. 5. **TGIF recognizes the TGT core binding site.** *A*, full-length TGIF and Oct-2 POU domain (Δ*Oct-2*) were expressed *in vitro* and were used in EMSAs. Specificity of binding was tested with specific and nonspecific unlabeled oligonucleotide competitors (*comp.*) which were added in excess over the probe as indicated. *B*, a representative set of sequences, isolated after three rounds of PCR binding site selection, is listed. The alignment of the selected nucleotide sequences was performed according to the algorithm Pile-up/Prettybox included in the GCG (University of Wisconsin) software package. The consensus sequence is shown in *reverse lettering*. *C*, consensogram derived from 42 different sequences which indicates the percent occurrence of each nucleotide at each position.

TABLE I
Alignment of promoters containing the TGIF core binding sequence

Binding site[a]	Gene of origin[b]	Position	Sequence[c]	References
RXR/RAR/HNF4/ARP-1	mCRBPII	−658/−636	GC**TGTCA**CAGGTCACAGTTCACA	(1)
RXR/RAR	rCRBPII	−639/−605	GC**TGTCA**CAGGTCACAGGTCACAGGTCACAGTTCACA	(1, 2)
COUP/ER	mLactoferrin	−350/−330	AG**TGTCA**CAGGTCAAGGTAAC	(3, 4)
COUP/ARP-1	cOvalbumin	−91/−60	TCTATGG**TGTCA**AAGGTCAAACTTCTGAAGGG	(3, 5–7)
RARβ	hCompl. factor H	−149/−127	TGCCC**TGTCA**GTGACCT	(8)
TR/RXR	hMyosin heavy chain	−158/−136	GCTGTCCTCC**TGTCAC**CTCCAGA	(9–11)
TR	rMyosin heavy chain	−149/−134	TGTCCTCC**TGTCA**CCT	(10)

[a] Nuclear factors binding to the corresponding sequences.
[b] r, rat; m, mouse; c, chicken; h, human, x, *Xenopus*.
[c] Half-sites which are recognized by the nuclear factors are underlined. The TGIF core binding sequence is shown in bold.

sites. The divergent half-site 1, although poorly bound by RXRα, did also contribute to the stabilization of the RXRα binding to the CRBPII-RXRE DNA (see Fig. 1 for a summary of the DNA affinities of RXRα).[2] To further investigate how many half-sites were occupied on the CRBPII-RXRE, we mapped the binding sites of RXRα by deoxyguanosine G-specific dimethyl sulfate interference. Fig. 4*B*, *lanes 1–4*, shows a typical experiment, where a complex generated by protein extracts obtained from Sf9 cells infected with a recombinant baculovirus expressing RXRα has been mapped onto the CRBPII-RXRE. On both strands, all G residues contained within the half-sites 2, 3, 4, and 5 were undermethylated. Residues from half-sites 2, 3, and 4 were slightly more undermethylated than residues from half-site 5, indicating a slightly predominant occupancy of the proximal (relative to half-site 1) repeat (Fig. 4*B*, *lanes 1* and *2*). This was consistent with the weaker M2,3·RXRα complex (Fig. 1) and also correlates well with the recent description of the mouse CRBPII-RXRE in which only the half-sites 1, 4, and 5 from the rat CRBPII-RXRE were conserved, allowing weaker RXR binding (1). In the region corresponding to half-site 1, none of the strands were undermethylated. This, however, contrasted with the EMSA experiments reported in Fig. 1, in which specific binding to this site was detected upon addition of a 10-fold larger amount of RXRα to the binding reactions.

Repression of RXRα-mediated Transcription Activation—To investigate the function of TGIF in the control of the cellular retinol-binding protein promoter, a series of retinoid-responsive reporter plasmids (pBLCAT5-CRBPII-RXREs) were transfected into either human glioma U87 or COS-1 cell lines together with vectors expressing RXRα (pSG5-RXRα) and/or TGIF (pcDNA-TGIF). The choice of the cell lines was dictated by the low levels of endogenous RXR's activities. Cells were treated with 10^{-7} M 9-*cis*-RA to selectively induce RXRα-dependent transcription activation (45, 46). Fig. 6 provides evidence that TGIF acts as a repressor of the RXRα-dependent transcriptional activation. In these experiments, a weak constitutive activity was observed by endogenous RXRα upon transfection of various reporter plasmids (Fig. 6, *A* and *B*, *lanes 1*, *5*, *9*, and *13*). As described earlier (2), transfection of RXRα generated a 9-*cis*-RA-dependent activation (4–5-fold in U87 cells and 10-fold in COS-1 cells) of the CRBPII-RXRE reporter plasmid (Fig. 6, *A* and *B*, *lanes 2*).

Co-transfection of TGIF and RXRα expression vectors resulted in a repression (3-fold in U87 cells and 2-fold in COS-1 cells) of the 9-*cis*-RA-inducible expression of the CRBPII-RXRE reporter plasmid (Fig. 6, *A* and *B*, compare *lane 2* with *lane 4*). The difference of repression activities detected in those cell lines suggest that the TGIF and RXRα ratio is very important

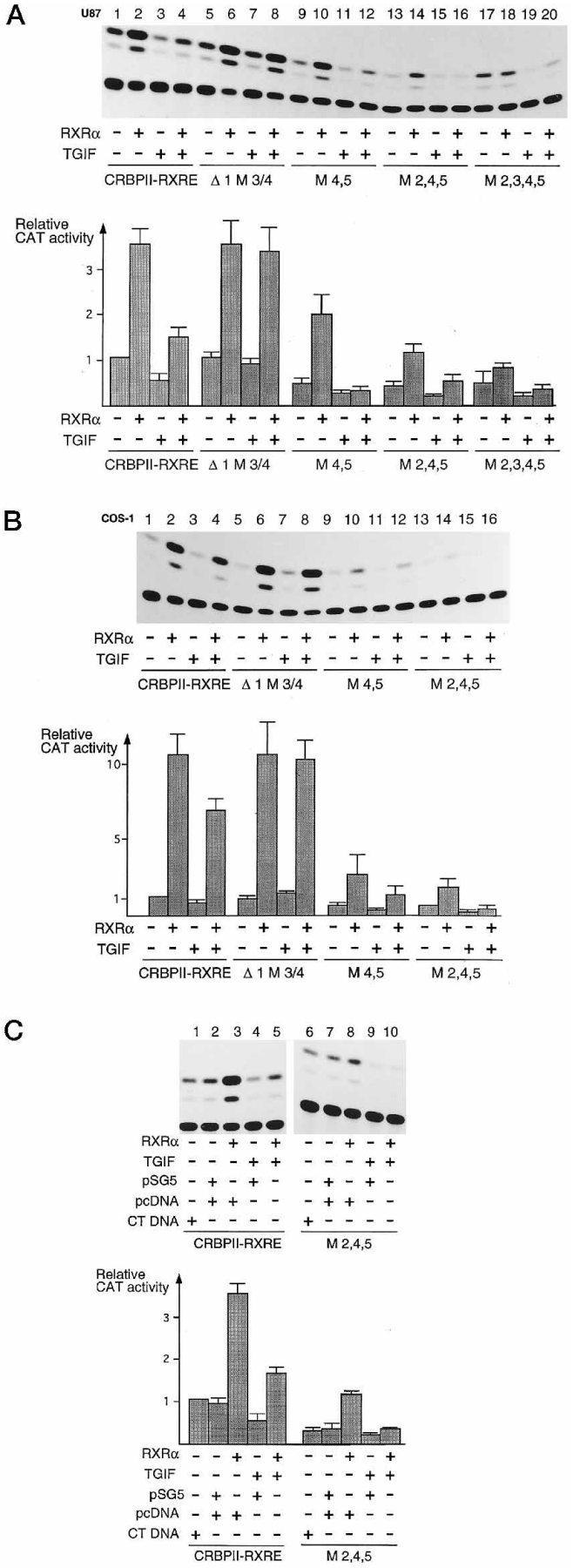

FIG. 6. **Repression of both basal and RXRα-dependent transactivation by TGIF in U87 cells and COS-1 cells.** *A*, 5 μg of the pBLCAT5 reporter gene constructs containing several CRBPII-RXREs promoter sequences (see Fig. 1) were co-transfected in U87 cells with either 1 μg of pSG5-RXRα effector plasmid and/or 3 μg of pcDNA TGIF effector plasmid, as indicated (+). *B*, COS-1 cells were co-transfected with the different reporter gene constructs and effector plasmids as in *A* except that 0.25 μg of RXRα effector plasmid were transfected. *C*, control experiments were performed in U87 cells by co-transfecting the reporter plasmids together with the pSG5 and pcDNA plasmids with or without insert in the same conditions as described in *A*. All experiments were performed in presence of 10^{-6} M 9-*cis*-retinoic acid for induction of RXRα-specific transcriptional activation. The *top part* of *A*, *B*, and *C* depict the characteristic thin layer chromatography pattern of nonacetylated and acetylated ^{14}C-labeled chloramphenicol (CAT) obtained in the different CAT assays. Relative expression of the CAT reporter gene has been quantified, and the values with standard deviations represent the mean of three independent experiments.

(see "Materials and Methods"). Further, transfection of the TGIF expression vector alone resulted in a decrease of the constitutive activity (3-fold in U87 and 2-fold in COS-1 cells), indicating that TGIF also repressed the endogenous RXRα-dependent activation (Fig. 6, *A* and *B*, compare *lanes 3* with *lanes 1*) which may be linked to the cooperative binding of RXRα on the CRBPII-RXRE and the saturation of the RXR binding sites.

Deletion of the TGIF binding site (half-site 1) in pBLCAT-Δ1 M3/4 did not affect the RXRα transcriptional activation in the presence of TGIF suggesting that the repression (Fig. 6, *A* and *B*, *lanes 5–8*) was mediated by the first DNA repeat of the CRBPII-RXRE element. To rule out a possible competition between the promoters contained within the expression vectors pSG5-RXRα and pcDNA-TGIF, we co-transfected pSG5-RXRα and pcDNA devoid of TGIF sequences (Fig. 6*C*, *lane 3*). No repression was observed under these conditions, nor did the co-transfection of both expression vectors devoid of cDNA sequences result in a reduction of the constitutive activity (Fig. 6*C*, *lane 2*).

To further document this repressor activity, we transfected reporter plasmids which would reduce or no longer support the cooperative binding of RXRα onto the CRBPII-RXRE (see Fig. 1). Mutations of two or more RXRα binding sites (*M4,5*, *M2,4,5*, and *M2,3,4,5* in Fig. 1) reduced or abolished the RXRα-mediated activation as well as the constitutive activity. With the pBLCAT-M4,5 reporter (Fig. 6, *A* and *B*, *lanes 9–12*), the constitutive activity in both U87 and COS-1 cells was reduced about 2-fold below the activity measured with the wild type CRBPII-RXRE. The RXRα-mediated transactivation was about 3-fold in Fig. 6, *A* and *B*, *lanes 9* and *10*. With a 3-fold reduced RXRα-dependent transactivation using the pBLCAT-M2,4,5 reporter, we could observe an even higher TGIF-dependent inhibitory effect (Fig. 6, *A* and *B*, *lanes 14* and *16*).

Transfection of the reporter plasmids pBLCAT-M4,5 and pBLCAT-M2,4,5 resulted in a proportionally stronger repression (5-fold in U87 cells and 3-fold in COS-1 cells) of the RXRα activity (Fig. 6, *A* and *B*, *lanes 10* and *12*). We also tested a reporter plasmid in which all RXRα binding sites were mutated (M2,3,4,5). As expected, no RXRα-mediated transactivation was detectable in U87 cells (Fig. 6*A*, *lanes 17* and *18*). Using the same reporter plasmid, TGIF did induce a slight decrease of the signal (Fig. 6*A*, *lanes 19* and *20*). This transcription repression could be directed toward the basal transcription activity or toward endogenous transcription activating proteins that can interact with the TGIF response element (see Table I), suggesting that TGIF is a general transcription repressor.

Mutually Exclusive Binding of TGIF and RXRα on the CRBPII-RXRE—The inspection of the G-specific methylation interference patterns generated by both TGIF and RXRα (Fig. 4, *A* and *B*) indicated that their interactions occurred at contiguous-adjacent areas on the CRBPII-RXRE retinoid-responsive ele-

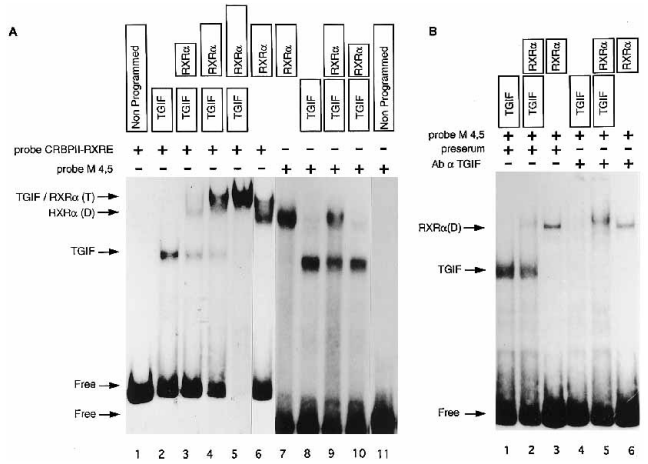

FIG. 7. **Mutually exclusive binding between TGIF and RXRα.** *A*, TGIF and RXRα from baculovirus-infected cells were used in EMSAs with the CRBPII-RXRE and the M4,5 probes (see Fig. 1). The sizes of the boxes are a graphical representation of the protein concentrations. *D* and *T* stand for dimer and trimer/tetramer, respectively. *B*, TGIF was incorporated in excess over RXRα in binding reactions with the M4,5 probe. Antibodies (serum dilution 1/20) directed against GST-TGIF (see "Materials and Methods") were co-incubated as indicated. Total protein concentration of the binding reactions has been equilibrated with non-programmed cell extracts and/or preserum.

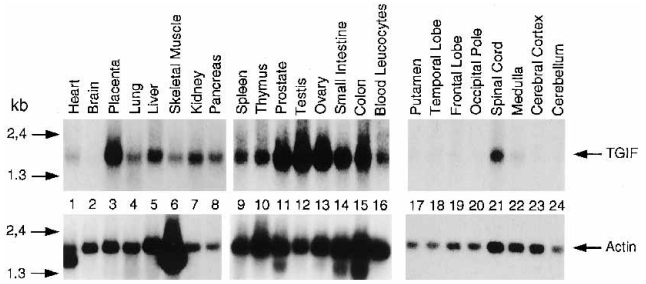

FIG. 8. **TGIF mRNA is expressed in a restricted number of human adult tissues.** A 300-bp-long probe corresponding to the amino-terminal coding region of TGIF was used to detect a single TGIF transcript on Northern blots containing 2 μg of oligo(dT)-selected RNAs from different human tissues. TGIF mRNA is highly expressed in placenta, liver, kidney, testis, and ovary tissues. It is less expressed in lung, pancreas, thymus, prostate, small intestine, colon, blood leukocyte, and spinal cord tissues. It is almost not detectable by Northern blots in brain and muscle tissues. Migration of RNA size standards is indicated (*kb*). Hybridization by a human β-actin probe was used to score the amount of RNA loaded.

ment. This raised the possibility that these two proteins recognize overlapping DNA binding sites, *i.e.* half-site 1 for TGIF and half-sites 2 and 3 for RXRα. To test whether or not TGIF and RXRα generated mutually exclusive DNA binding *in vitro*, co-incubation of these factors with the CRBPII-RXRE was performed in EMSA. A recombinant baculovirus was generated to overexpress full-length TGIF in Sf9 cells. TGIF complexed the probe specifically (Fig. 7*A*, *lane 2*) as already shown in Fig. 3. Nonprogrammed Sf9 cell extract (Fig. 7*A*, *lane 1*) did not. Two complexes with different electrophoretic migration were visible upon incubation of RXRα from baculovirus-infected Sf9 cells, and the rat CRBPII-RXRE probe (Fig. 7*A*, *lane 6*). The fastest migrating complex (*RXRα(D)* in Fig. 7) represents an RXRα homodimer because it co-migrated with an RXRα·M4,5 DNA complex (Fig. 7*A*, *lane 7*) in which 2 out of 4 RXRα sites are mutated, allowing only the formation of an RXRα dimer on half-sites 2 and 3. The slower migrating complex in the same lane (Fig. 7*A*, *lane 4*) most likely represents the binding of three or four RXRα molecules. This is in agreement with what has been suggested previously in Ref. 2.

As shown in Fig. 7*A*, co-incubation of a constant amount of TGIF and increasing amounts of RXRα with the CRBPII-RXRE probe, led to the disruption of the TGIF·CRBPII-RXRE complex, suggesting an incorporation of TGIF in the larger complex called "TGIF/RXRα(T)" or an exclusive binding between TGIF and RXRα. In conditions where equal amounts of TGIF and RXRα were co-incubated with the CRBPII-RXRE probe, the TGIF and the dimeric RXRα(D) complexes were supershifted into the larger complex TGIF/RXRα(T) (Fig. 7*A*, compare *lane 4* with *lanes 2* and *6*). This result (*lane 4*) was obtained in conditions where the probe concentration was limiting and the TGIF/RXRα protein concentrations identical, in order to allow TGIF to bind on half-site 1 and to force the RXRα molecules onto the free distal half-sites (3, 4, and 5) of the CRBPII-RXRE probe.

The M4,5 probe was used in EMSA to evaluate possible steric hindrance between RXRα and TGIF on their high affinity half-sites 1, 2, and 3 (Fig. 7*A*, *lanes 7–11*). Using this probe, RXRα generated a homodimer complex to half-sites 2 and 3, which was equally intense as the complex obtained with TGIF (Fig. 7*A*, compare *lanes 7* and *8*). Co-incubation of TGIF and RXRα with the M4,5 probe gave rise to two complexes (Fig. 7*A*, *lane 9*) migrating at the level of the respective protein·DNA complexes but with reduced intensity (Fig. 7*A*, *lanes 7* and *8*). This result suggests that RXRα and TGIF shared the M4,5 probe without binding simultaneously on the same probe molecules. There were neither intermediary complexes nor a supershift of the complexes. A stronger TGIF signal was recovered by diluting out RXRα in a co-incubation with the M4,5 probe (Fig. 7*A*, *lane 10*).

A similar experiment was performed in Fig. 7*B* with an excess of TGIF over RXRα. Both proteins could complex the M4,5 probe, respectively, in Fig. 7*B*, *lanes 1* and *3*. Co-incubation of TGIF and RXRα with the M4,5 probe led to a disruption of the RXRα dimeric complex (Fig. 7*B*, *lane 2*). Specific antibodies raised against TGIF (Fig. 7*B*, *lanes 4–6*) neutralized the binding of TGIF to its recognition site (Fig. 7*B*, compare *lanes 1* and *4*) but did not affect RXRα binding (Fig. 7*B*, compare *lanes 3* and *6*). Complete RXRα binding could however be restored upon neutralization of the TGIF DNA binding (Fig. 7*B*, compare *lanes 5* and *2*). The slightly supershifted complex RXRα(D) observed in *lanes 2* and *5*, where TGIF and RXRα were co-incubated, was not due to the presence of TGIF in the large complex because the band retardation is identical with or without neutralization of DNA binding of TGIF (compare both lanes). Partially degraded TGIF protein could slow down complex migration. In this case, there would be neither a sterical hindrance by RXRα nor a DNA binding neutralization by the antibodies which were not directly raised against the DNA binding domain.

Attempts to co-immunoprecipitate RXRα with specific antibodies to TGIF failed, suggesting that no direct protein-protein interaction between TGIF and RXRα occurred.[2] These results demonstrated that the presence of TGIF prevents RXRα from binding to the DNA recognition half-sites 2 and 3 on the CRBPII-RXRE, leading to a disruption of the RXR cooperative binding. Further, they support the notion that TGIF prevents RXRα from functioning as a transcriptional activator by interacting with its cognate responsive element.

Tissue-specific Expression—As shown in Fig. 8, poly(A)$^+$ RNA from different adult human tissues were probed in Northern blots and revealed a single TGIF transcript of 2 kilobases. TGIF mRNA is highly expressed in the placenta, liver, kidney, testis, and ovary (Fig. 8, respectively, *lanes 3*, *5*, *7*, *12*, and *13*). It is weakly expressed in the small intestine and is almost not detectable on Northern blots in heart, brain, skeletal muscle, and peripheral blood leukocytes (Fig. 8, *lanes 14*, *1*, *2*, *6*, and

16). However, inspection of different subregions of the human brain revealed subtle signal variations. The mRNA corresponding to TGIF is fairly well expressed in the spinal cord (Fig. 8, *lane 21*), but it is almost not detectable in the cerebral cortex and in the cerebellum (Fig. 8, *lanes 23* and *24*). Interestingly, TGIF mRNA co-localizes with RXRα mRNA in adult liver, placenta, and kidney (21).

DISCUSSION

Our studies have revealed that a novel homeoprotein, TGIF, recognizes an unusual DNA sequence for homeoproteins only reported in a restricted number of examples. As described in the case of RXRα, the TGIF binding on a CRBPII gene promoter element had as a consequence a functional interference.

TGIF is a member of a growing family of homeoproteins which is characterized by the requirement of insertions and/or deletions in their sequences for maximizing identities in amino acid alignments (14). A substantial number can be classified in a novel group of atypical homeodomains characterized by the presence of three additional amino acids between helix 1 and 2. This three amino acid loop extension could be determined on the basis of structural comparisons between the atypical α2 and the typical Engrailed homeodomains (47). We suggest calling this group of atypical homeodomains the TALE (three amino acid loop extension) homeodomain superclass. Four classes, Kn, PBC, HAC-ATYP, and M-ATYP (according to the general nomenclature described in Ref. 48) can be grouped in the TALE superclass. These four classes share a three amino acid extended loop, and this structural conservation suggests that their members have common biological features. The high divergence of the TGIF homeodomain with the homeodomain of members of these four classes, the closest being the HAC-ATYP class, suggests that TGIF may be integrated in a new class.

The TGIF homeodomain shares with the other atypical TALE homeodomains highly conserved residues in the extended loop between helices 1 and 2 (amino acids 23 to 31, in Fig. 2*B*). The large number of these TALE homeodomains allows us to predict at which position the three amino acids were inserted during evolution on the basis of amino acid conservation between the typical and atypical homeodomains. We propose that the insertion of these three amino acids occurred carboxyl-terminally to amino acid 22, thereby not affecting the spacing between the highly conserved residues Asn^{23} and Tyr^{25} in the classical as well as that between residues Asn^{26} and Tyr^{28} (in Fig. 2*B*) in the atypical homeodomains. Furthermore, as indicated in Fig. 2*B*, the two first amino acids from the three amino acid insertion (His^{23} and Leu^{24}) are well conserved in this TALE superclass of homeodomains, suggesting thereby an important role of these residues in the function of this superclass of homeoproteins.

In contrast to most homeoproteins which specifically interact with the target consensus sequence 5′AATTA3′, TGIF together with a1/α2, caudal (Cad) and the thyroid nuclear factor 1 (TTF1) display high affinity for non-ATTA consensus sequence elements (15, 49–51), as demonstrated for TGIF by the binding site selection. TGIF homeodomain contains 15 rare amino acids located mainly in helix 1 and helix 3 which occur less than 5 times (1.5%) at each position among 346 homeodomains (14), *e.g.* Cys^{49}. However, 4 out of these rare 15 amino acids (residues Pro^{9}, Trp^{19}, Asn^{50}, and Ile^{53}) were strongly represented in the group of the atypical homeodomains. Amino acids Asn^{50} and Ile^{53}, located in the TGIF recognition helix 3 (see Fig. 2), have been described as critical residues for DNA sequence-specific binding (40, 52, 53). Interestingly, Asn^{50} was conserved at the same position in the α2 homeodomain and Ile^{53} in the typical a1 homeodomain. Furthermore, TGIF shares with a1/α2 not only these residues involved in specific DNA binding, but also the affinity for the TGT core DNA binding site (47, 49). The "Asn^{51} alignment rule" was defined on the basis of the alignment of the A residue ($\binom{TGT}{ACA}$ and $\binom{AAT}{TTA}$) in the DNA recognition sites contacted by Asn^{51}, a highly conserved amino acid in all homeodomains (Asn^{54} for TALE homeodomains). The alignment of the DNA recognition sequences from TGIF, α2, and Engrailed according to the above-mentioned rule showed that residue Arg^{57} (in Fig. 2*B*) from the α2 homeodomain (47) which contacts the central G nucleotide in the TGT core ($\binom{TGT}{ACA}$) is also conserved in TGIF. In contrast to the Engrailed homeodomain which recognizes an AAT core, this amino acid sequence conservation in TGIF and α2 might well reflect their unusual DNA binding behavior.

Comparison of the rat and mouse CRBPII-RXREs indicates that the TGIF DNA recognition site is present in both species and that in the mouse it is flanked in the 3′ direction by a direct repeat spaced by one nucleotide (DR1) composed of half-sites 4 and 5 (Table I). The DR1 hexamer half-sites 4 and 5 consists of a weaker binding site for RXRα, for RXR:RAR heterodimer, and for the hepatocyte nuclear factor HNF-4, and a stronger binding site for the apolipoprotein A1 regulatory protein 1 (ARP-1) (1). The TGIF consensus binding site is moreover conserved in the promoters from the mouse lactoferrin gene, the chicken ovalbumin gene, the human complement factor H gene and human/rat myosin heavy chain genes (Table I). These TGIF target promoter sequences were shown to be adjacent or overlapping to steroid/retinoid receptors recognition sites in these promoters. The sites are bound by the COUP-transcription factor (TF), estrogen receptor (ER), RARβ, ARP-1, and/or thyroid receptor (TR) and are composed of half-sites which together with the consensus binding site of TGIF site are arranged as imperfect direct or inverted repeats spaced by one nucleotide as in the CRBPII-RXRE. The spacing is of two nucleotides in the human/rat myosin heavy chain TRE. Furthermore, the TGIF consensus recognition sequence is a COUP-TF natural half-site (3). The presence of a TGIF binding site in the CRBPII-RXRE could be seen as fortuitous. However, the conservation of the TGIF binding site contiguous or overlapping to several steroid/retinoid receptor binding sites argues strongly in favor of a functional relevance for this TGIF site. This observation prompted us to further study the functional interaction of TGIF with the CRBPII-RXRE in the context of RXR dependent transcription activation. We describe in this paper that two regulatory factors, belonging to two different families of transcription factors, are interfering functionally upon binding on their respective DNA targets. The mutually exclusive binding between TGIF and RXRα on the rat CRBPII-RXRE leads to the repression of the RXRα-dependent transcription activation. A weak inhibitory effect is also directed toward the endogenous RXR transcription activity. As demonstrated herein on the rat CRBPII-RXRE, the exclusive binding of TGIF with the retinoid receptor could also influence the steroid/retinoid receptor homodimer and heterodimer-mediated transactivation on the homologous mouse promoter region of the CRBPII gene. Possibly, TGIF could also modulate the activity of the mouse lactoferrin, chicken ovalbumin, the human complement factor H, and human/rat myosin heavy chain gene promoters.

Several homeoproteins (a1/α2, Extradenticle, Engrailed, PBX, HOX) have been shown to interact cooperatively, thereby changing the DNA target site specificity by conferring strong binding on sites for which the single proteins show only weak affinity (16, 54–58). This could also hold true for TGIF when several molecules are interacting with the CRBPII-RXRE. There are in half-site 1 two superposed and inverted TGIF binding sites, a high and a low TGIF affinity site, followed in

the 3′ direction by directly repeated low affinity TGIF sites (see Figs. 3*B* and 4). Furthermore, in the mouse CRABPII promoter (at position −658 to −650), a putative TGIF low affinity binding site (5′GCTGTGAC3′) is overlapping with a DR1 binding site for RXR (CRABPII/RARE2) (59). Sequences reading TGTGA were also found in additional RAREs and EREs, but the functional importance of these possible weak TGIF binding site has still to be proven.

Although the mechanisms involved in transcriptional activation have been extensively investigated over the past years, much less is known, however, of the processes governing transcriptional repression. Transcription repression can be achieved by different mechanisms. It can be brought about by the action of a repressor directly blocking a DNA-responsive site for a transcriptional activator. The repressor also can directly inhibit transcription by neutralizing the activation domain of a transcriptional activator or by titrating out activating factors (60). Recent reports focused on the repression of eukaryotic transcription and in particular on the exclusive binding of transcription factors on contiguous or overlapping DNA sites (61–63 and references cited therein). For example, DAX 1, a novel orphan member of the nuclear hormone receptor superfamily, acts as a dominant negative regulator of RAR-mediated transcription by competing for the RAR DNA sites (64). Another experiment has been carried out with transgenic mice overexpressing an isoform of RARβ4 lacking the A domain, which is important for activation of the CRBPII promoter (65). These animals were clearly predisposed to hyperplasia and neoplasia (66). These *in vivo* disorders have been proposed to result from the competition of RARβ4 with other RARs for retinoic acid (RA) response elements contained within the CRBPII promoter and thereby affecting indirectly intracellular RA signaling.

Some inhibitory factors need hormone induction to actively repress transcription and thereby interfere, in the presence of a ligand, with transcription activators by occupying adjacent or overlapping sites (28 and references cited therein). For example, the ligand-dependent effect of RAR-driven AP-1 (c-Fos-/c-Jun heterodimer) transrepression can be dissociated from the ligand-mediated RAR transcription activation (27). Similar to the situation observed with TGIF/RXRα, the AP-1/RAR mutual repression occurs by exclusive binding on an identical site within the osteocalcin promoter (67). It would be of interest to test whether TGIF's inhibitory activity could be influenced by a ligand. Nevertheless, this hypothesis is quite unlikely since no homeoprotein was reported so far to act in a ligand-dependent fashion. It is interesting to visualize the convergence of two different regulatory pathways, a ligand-dependent and a ligand independent one, to control the regulation of the CRBPII gene through its RXRE sites. The consequence of such an interference of two different classes of factors on an overlapping responsive site can either be the enhanced repression or the transactivation of the target promoters (68, 69).

A single factor can either function as a repressor or as an activator of transcription, as described in the AP1/RAR example (see above). The switching of RAR from an activator to a repressor of retinoid-dependent transcription can be obtained merely by changing its relative positioning in the heterodimeric complex with RXR, depending on the spacing between the half sites (23). RAR:RXR heterodimers activate transcription in a ligand-dependent manner by binding on directly repeated half-sites spaced by 5 nucleotides (DR5). RAR occupies the downstream half-site. In contrast, RAR:RXR heterodimers do not activate transcription when bound to a DR1. RAR is inhibiting RXR transactivation by binding on the upstream half-site and thereby blocks the binding of the ligand to the RXR. Although the blocking of 9-*cis*-RA binding to RXR by TGIF was not studied, mutually exclusive binding seems to be the major mechanism leading to transcription inhibition by TGIF. However, the possibility that TGIF is an activator itself should not be excluded. The switching of TGIF to an activator would not depend on alternate positioning on the CRBPII-RXRE binding site but on other regulatory proteins present in the cell (for examples, see Ref. 60 and references cited therein).

While the RXRα, RARs, HNF-1, and ARP-1 seem to be the major players in the mouse and rat CRBPII gene regulation (1, 2), it is tempting to speculate that this regulation could be modulated also by TGIF. The possibility that TGIF can regulate the transcription of the CRBPII gene suggests that it may synergize with these factors playing an important role in retinoid homeostasis. TGIF may interfere functionally in a similar manner with the members of the retinoid/steroid receptors superfamily which regulate transcription on responsive elements from gene promoters shown in Table I containing the canonical binding site of TGIF.

Acknowledgments—We thank I. Hampele for the excellent experimental contribution. We thank W. Hunziker for providing the recombinant RXRα baculovirus, A. Hayes, M. Brockhaus, and H. Jacot for their help with the tissue culture and the antibody production and J. Pomerantz for his random binding site selection protocols. The expert assistance of T. Rohrer with the preparation of the figures is greatly acknowledged. We thank P. Matthias, F. Monsma, and J. R. Pink for their critical comments on the manuscript. Finally, we are grateful to O. Civelli, M. Steinmetz, and to C. Köhler for encouragement and support.

REFERENCES

1. Nakshatri, H., and Chambon, P. (1994) *J. Biol. Chem.* **269,** 890–902
2. Mangelsdorf, D. J., Umesono, K., Kliewer, S. A., Borgmeyer, U., Ong, E. S., and Evans, R. M. (1991) *Cell* **66,** 555–561
3. Cooney, A. J., Tsai, S. Y., O'Malley, B. W., and Tsai, M. J. (1992) *Mol. Cell. Biol.* **12,** 4153–4163
4. Liu, Y., and Teng, C. T. (1991) *J. Biol. Chem.* **266,** 21880–21885
5. Ladias, J. A., and Karathanasis, S. K. (1991) *Science* **251,** 561–565
6. Wang, L. H., Tsai, S. Y., Cook, R. G., Beattie, W. G., Tsai, M. J., and O'Malley, B. W. (1989) *Nature* **340,** 163–166
7. Sagami, I., Tsai, S. Y., Wang, H., Tsai, M. J., and O'Malley, B. W. (1986) *Mol. Cell. Biol.* **6,** 4259–4267
8. Muñoz-Cánoves, P., Vik, D. P., and Tack, B. F. (1990) *J. Biol. Chem.* **265,** 20065–20068
9. Flink, I. L., and Morkin, E. (1990) *J. Biol. Chem.* **265,** 11233–11237
10. Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M. (1991) *Cell* **65,** 1255–1266
11. Zhang, X. K., Lehmann, J., Hoffmann, B., Dawson, M. I., Cameron, J., Graupner, G., Hermann, T., Tran, P., and Pfahl, M. (1992) *Nature* **358,** 587–591
12. Demmer, L. A., Birkenmeier, E. H., Sweetser, D. A., Levin, M. S., Zollman, S., Sparkes, R. S., Mohandas, T., Lusis, A. J., and Gordon, J. I. (1987) *J. Biol. Chem.* **262,** 2458–2467
13. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994) *Cell* **76,** 933–945
14. Bürglin, T. R. (1994) in *Guidebook to Homeobox Genes* (Duboule, D., ed) pp. 27–71, Oxford University Press, Oxford, UK
15. Gehring, W. J., Affolter, M., and Bürglin, T. (1994) *Annu. Rev. Biochem.* **63,** 487–526
16. Stark, M. R., and Johnson, A. D. (1994) *Nature* **371,** 429–432
17. Lai, J. S., Cleary, M. A., and Herr, W. (1992) *Genes & Dev.* **6,** 2058–2065
18. Kutoh, E., Stromstedt, P. E., and Poellinger, L. (1992) *Mol. Cell. Biol.* **12,** 4960–4969
19. Stunnenberg, H. G. (1993) *Bioessays* **15,** 309–315
20. Chambon, P., Zelent, A., Petkovitch, M., Mendelsohn, C., Leroy, P., Krust, A., Kastner, P., and Brand, N. (1991) in *Retinoids: 10 Years On* (Saurat, J. H., ed) pp. 10–27, Karger, Basel
21. Mangelsdorf, D. J., Borgmeyer, U., Heyman, R. A., Zhou, J. Y., Ong, E. S., Oro, A. E., Kakizuka, A., and Evans, R. M. (1992) *Genes & Dev.* **6,** 329–344
22. Heery, D. M., Pierrat, B., Gronemeyer, H., Chambon, P., and Losson, R. (1994) *Nucleic Acids Res.* **22,** 726–731
23. Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M. G., Heyman, R. A., and Glass, C. K. (1994) *Nature* **371,** 528–531
24. Leid, M., Kastner, P., and Chambon, P. (1992) *Trends Biochem. Sci.* **17,** 427–433
25. Conlon, R. A. (1995) *Trends Genet.* **11,** 314–319
26. Nicholson, R. C., Mader, S., Nagpal, S., Leid, M., Rochette-Egly, C., and Chambon, P. (1990) *EMBO J.* **9,** 4443–4454
27. Chen, J. Y., Penco, S., Ostrowski, J., Balaguer, P., Pons, M., Starrett, J. E., Reczek, P., Chambon, P., and Gronemeyer, H. (1995) *EMBO J.* **14,** 1187–1197
28. Pfahl, M. (1993) *Endocrin. Rev.* **14,** 651–658
29. Chytil, F., and Stump, D. G. (1991) in *Retinoids: 10 Years On* (Saurat, J. H., ed) pp. 38–45, Karger, Basel

30. Blomhoff, R., Green, M. H., Berg, T., and Norum, K. R. (1990) *Science* **250,** 399–404
31. Singh, H., Clerc, R. G., and LeBowitz, J. H. (1989) *BioTechniques* **7,** 252–261
32. Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D., and Sharp, P. A. (1988) *Genes & Dev.* **2,** 1570–1581
33. Kumar, V., and Chambon, P. (1988) *Cell* **55,** 145–156
34. Yang, N., Schüle, R., Mangelsdorf, D. J., and Evans, R. M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3559–3563
35. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 449–560
36. Boshart, M., Kluppel, M., Schmidt, A., Schutz, G., and Luckow, B. (1992) *Gene (Amst.)* **110,** 129–130
37. Green, S., Issemann, I., and Sheer, E. (1988) *Nucleic Acids Res.* **16,** 369
38. Chen, C. A., and Okayama, H. (1988) *BioTechniques* **6,** 632–638
39. Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044–1051
40. Pomerantz, J. L., and Sharp, P. A. (1994) *Biochemistry* **33,** 10851–10858
41. Yu, V. C., Delsert, C., Andersen, B., Holloway, J. M., Devary, O. V., Naar, A. M., Kim, S. Y., Boutin, J. M., Glass, C. K., and Rosenfeld, M. G. (1991) *Cell* **67,** 1251–1266
42. Kozak, M. (1986) *Cell* **44,** 283–292
43. Catron, K. M., Zhang, H., Marshall, S. C., Inostroza, J. A., Wilson, J. M., and Abate, C. (1995) *Mol. Cell. Biol.* **15,** 861–871
44. Tjian, R., and Maniatis, T. (1994) *Cell* **77,** 5–8
45. Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., Stein, R. B., Eichele, G., Evans, R. M., and Thaller, C. (1992) *Cell* **68,** 397–406
46. Levin, A. A., Sturzenbecker, L. J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Kratzeisen, C., Rosenberger, M., Lovey, A., and Grippo, J. F. (1992) *Nature* **355,** 359–361
47. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell* **67,** 517–528
48. Bürglin, T. R. (1995) in *Biodiversity and Evolution* (Arai, R., Kato, M., and Doi, Y., ed) The National Science Foundation, Tokyo
49. Goutte, C., and Johnson, A. D. (1988) *Cell* **52,** 875–882
50. Guazzi, S., Price, M., De-Felice, M., Damante, G., Mattei, M. G., and Di-Lauro, R. (1990) *EMBO J.* **9,** 3631–3639
51. Dearolf, C. R., Topol, J., and Parker, C. S. (1989) *Nature* **341,** 340–343
52. Ekker, S. C., von-Kessler, D. P., and Beachy, P. A. (1992) *EMBO J.* **11,** 4059–4072
53. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989) *Cell* **59,** 553–562
54. Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F., and Dahlquist, F. W. (1994) *Biochemistry* **33,** 15309–15320
55. Krumlauf, R., and Gould, A. (1992) *Trends Genet.* **8,** 297–299
56. Mak, A., and Johnson, A. D. (1993) *Genes & Dev.* **7,** 1862–1870
57. van-Dijk, M. A., and Murre, C. (1994) *Cell* **78,** 617–624
58. Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995) *Mol. Cell. Biol.* **15,** 3989–3997
59. Durand, B., Saunders, M., Leroy, P., Leid, M., and Chambon, P. (1992) *Cell* **71,** 73–85
60. Johnson, A. D. (1995) *Cell* **81,** 655–658
61. Biggin, M. D., and Tjian, R. (1989) *Cell* **58,** 433–440
62. Barberis, A., Superti-Furga, G., and Busslinger, M. (1987) *Cell* **50,** 347–359
63. Desplan, C., Theis, J., and O'Farrell, P. H. (1988) *Cell* **54,** 1081–1090
64. Zanaria, E., Muscatelli, F., Bardoni, B., Strom, T. M., Guioli, S., Guo, W., Lalli, E., Moser, C., Walker, A. P., McCabe, E. R., Meitinger, T., Monaco, A. P., Sassone-Corsi, P., and Camerino, G. (1994) *Nature* **372,** 635–641
65. Nagpal, S., Saunders, M., Kastner, P., Durand, B., Nakshatri, H., and Chambon, P. (1992) *Cell* **70,** 1007–1019
66. Berard, J., Gaboury, L., Landers, M., De-Repentigny, Y., Houle, B., Kothary, R., and Bradley, W. E. (1994) *EMBO J.* **13,** 5570–5580
67. Schüle, R., Umesono, K., Mangelsdorf, D. J., Bolado, J., Pike, J. W., and Evans, R. M. (1990) *Cell* **61,** 497–504
68. Komuro, I., Schalling, M., Jahn, L., Bodmer, R., Jenkins, N. A., Copeland, N. G., and Izumo, S. (1993) *EMBO J.* **12,** 1387–1401
69. Grueneberg, D. A., Natesan, S., Alexandre, C., and Gilman, M. Z. (1992) *Science* **257,** 1089–1095