MOLECULAR AND CELLULAR BIOLOGY, Mar. 1997, p. 1281–1288 Vol. 17, No. 3
0270-7306/97/$04.00+0

Negative Protein 1, Which Is Required for Function of the Chicken Lysozyme Gene Silencer in Conjunction with Hormone Receptors, Is Identical to the Multivalent Zinc Finger Repressor CTCF

MARK BURCIN,[1]† RÜDIGER ARNOLD,[1] MARCUS LUTZ,[1] BURKHARD KAISER,[1]‡ DIETER RUNGE,[1]§ FRIEDRICH LOTTSPEICH,[2] GALINA N. FILIPPOVA,[3] VICTOR V. LOBANENKOV,[3] AND RAINER RENKAWITZ[1]*

Genetisches Institut der Justus-Liebig-Universität, D-35392 Giessen,[1] and Max-Planck-Institut für Biochemie, D-82152 Martinsried,[2] Germany, and Fred Hutchinson Cancer Research Center, Seattle, Washington 98104[3]

Received 18 October 1996/Returned for modification 2 December 1996/Accepted 23 December 1996

The transcriptional repressor negative protein 1 (NeP1) binds specifically to the F1 element of the chicken lysozyme gene silencer and mediates synergistic repression by v-ERBA, thyroid hormone receptor, or retinoic acid receptor. Another protein, CCCTC-binding factor (CTCF), specifically binds to 50-bp-long sequences that contain repetitive CCCTC elements in the vicinity of vertebrate c-*myc* genes. Previously cloned chicken, mouse, and human CTCF cDNAs encode a highly conserved 11-Zn-finger protein. Here, NeP1 was purified and DNA bases critical for NeP1-F1 interaction were determined. NeP1 is found to bind a 50-bp stretch of nucleotides without any obvious sequence similarity to known CTCF binding sequences. Despite this remarkable difference, these two proteins are identical. They have the same molecular weight, and NeP1 contains peptide sequences which are identical to sequences in CTCF. Moreover, NeP1 and CTCF specifically recognize each other's binding DNA sequence and induce identical conformational alterations in the F1 DNA. Therefore, we propose to replace the name NeP1 with CTCF. To analyze the puzzling sequence divergence in CTCF binding sites, we studied the DNA binding of 12 CTCF deletions with serially truncated Zn fingers. While fingers 4 to 11 are indispensable for CTCF binding to the human c-*myc* P2 promoter site A, a completely different combination of fingers, namely, 1 to 8 or 5 to 11, was sufficient to bind the lysozyme silencer site F1. Thus, CTCF is a true multivalent factor with multiple repressive functions and multiple sequence specificities.

Transcriptional repression is an important feature of gene regulation in prokaryotes as well as in eukaryotes. Despite the wealth of information on prokaryotic repression, eukaryotic repression has been analyzed only quite recently. In addition to the mechanism of repression conferred by competitive binding to a particular DNA target sequence, active repression is an important mechanism of transcriptional inhibition (for a review, see reference 14). Possible mechanisms for active repression have been summarized as involving a repression domain either (i) interfering directly with the preinitiation complex, (ii) binding to coactivators, or (iii) mediating repression by binding to corepressors, which in turn interfere with coactivators or the preinitiation complex. Such active repression has been found in several cases, and the corresponding regulatory sequences have been called silencers, analogous to enhancer elements, since they function independent of their position and orientation and since they are often organized in a modular structure (for a review, see reference 25).

Such a modular structure has been found for the chicken lysozyme silencer, located 2.4 kb upstream of the transcriptional start site. This silencer is comprised of two DNA response elements (F1 and F2) which synergistically repress gene activity. One silencer module (F2) is bound either by v-ERBA, the thyroid hormone receptor (T3R), or the retinoic acid receptor (RAR) (4). Functional tests revealed that ligand-free T3R, RAR, or v-ERBA acts as a silencer protein (3, 4). In the presence of T3R and the ligand (thyroid hormone [T3]), the two silencer modules can synergistically activate gene transcription (17). The second module (F1) of the chicken lysozyme silencer binds a ubiquitously expressed nuclear protein, termed negative protein 1 (NeP1). The biochemical characteristics of NeP1 have been determined previously (17), leading to identification of the most unusual property of binding quite a long stretch of DNA (about 50 bp) as a monomeric protein.

Negative transcriptional regulation is also believed to play an important role in maintaining appropriate expression levels of the c-*myc* oncogene during normal cell growth stimulation and differentiation (24). A highly conserved and ubiquitously expressed transcriptional repressor, CTCF, has been shown to bind to a negative element upstream of the chicken *myc* promoter (21). This binding site (footprint V [FPV]) contains several CCCTC repeats (11, 16) and is bound by a single DNA binding domain of CTCF, comprising 11 Zn fingers (16). In contrast to the upstream position of the binding site in the case of the chicken c-*myc* gene, CTCF binds to sites within the coding regions of the mouse and human c-*myc* genes (11). These sites are known to function as RNA polymerase II pausing regions (9, 18, 26, 27). One common feature of each of the 50-bp-long CTCF binding sites in c-*myc* genes is a high GC content: 65, 85, and 87% in sites A and B of mammalian c-*myc* genes and site V of the chicken c-*myc* gene, respectively. Here we show, despite an extreme divergence between the GC-rich

* Corresponding author. Mailing address: Genetisches Institut der Justus-Liebig-Universität, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany. Phone: 49-641-99-35460. Fax: 49-641-99-35469.

† Present address: Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030.

‡ Present address: Abteilung Mykologie, Hans-Knöll-Institut für Naturstoff-Forschung e.V., D-07745 Jena, Germany.

§ Present address: Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15213.

CTCF binding sites in the c-*myc* genes and the AT-rich lysozyme silencer F1 sequence, which contains no CCCTC motif, that the F1 binding factor NeP1 is identical to CTCF. The high-affinity binding of CTCF to different sequences is apparently achieved by involving different sets of the 11 Zn fingers.

(This work contains parts of the Ph.D. thesis of M. Burcin.)

MATERIALS AND METHODS

Purification and sequencing of NeP1. NeP1 was purified from HeLa cell nuclear extract and prepared as described previously (1). Fractions eluting with 600 mM NaCl from an F1 DNA-affinity column were separated by sodium dodecyl sulfate (SDS)–7% polyacrylamide gel electrophoresis (PAGE) (20). The gel was stained with Coomassie brilliant blue, and the NeP1 band was excised and cleaved directly in the gel essentially as described elsewhere (8). After extensive washing with water, the gel pieces were minced, dried in a SpeedVac concentrator, resuspended with 200 μl of buffer (12 mM Tris, 0.5 mM EDTA, pH 8.5) containing endoprotease LysC (Boehringer Mannheim), and incubated at 37°C for 8 h at an enzyme-to-protein ratio of 1:10. The resulting peptide fragments were eluted twice with 60% acetonitrile–0.1% trifluoroacetic acid (each for 1 h), and the eluate was filtered through an Anotop filter (0.02-μm pore size; Merck) and concentrated by evaporation with the SpeedVac concentrator. The peptides were separated by reversed-phase high-pressure liquid chromatography (column, Superspher 60RP-select B; Merck). Eluents were 0.1% trifluoroacetic acid in water (solvent A) and 0.1% trifluoroacetic acid in acetonitrile (solvent B). The gradient used was 0 to 60% solvent B in 60 min; the flow rate was 300 μl/min. Eluted peptides were detected at 206 nm and sequenced in an ABI 492A pulsed liquid phase protein sequencer (Applied Biosystems, Weiterstadt, Germany) according to the manufacturer's instructions.

CTCF expression. The full-length cDNA of chicken CTCF was expressed in COS-1 cells with the plasmid pSG5-CTCF (16). A total of 2×10^6 to 3×10^6 COS 1-cells were transfected with 25 μg of DNA by the standard protocols. After cultivation for 48 h, the cells were collected, resuspended in 200 μl of lysis buffer (20 mM HEPES [pH 7.8], 400 mM KCl, 20% glycerol, 2 mM dithiothreitol), and frozen in cold methanol (−80°C). After thawing on ice and sedimentation of the cell debris (11,000 rpm, 4°C, 10 min), the supernatant was used for electrophoretic mobility shift assay (EMSA) (see below).

CTCF Zn finger deletions were generated (11) and expressed by in vitro transcription and translation using the TnT Kit (Promega). Quality of translation was monitored by SDS gel analysis of [^{35}-S]Met-labelled proteins, and similar quantities of the different deletions were judged after staining of the SDS-acrylamide gel. For the EMSA experiments, equal amounts of translation reaction products were used per lane.

Southwestern blot and Western blot analysis. After separation of crude nuclear proteins or NeP1-CTCF-enriched protein fractions by SDS-PAGE using a 7% polyacrylamide gel, the proteins were electroblotted onto a polyvinylidene difluoride membrane (Immobilon P; Millipore). For Southwestern blot analysis, the membrane, after de- and renaturation, was blocked overnight in SW buffer (20 mM HEPES [pH 7.6], 20 mM KCl, 5 mM $MgCl_2$, 10% [vol/vol] glycerol, 1 mM dithiothreitol) with 5% (wt/vol) milk powder (Carnation). The membrane was washed several times with SW buffer and preincubated with 4 ml of SW buffer containing 2 μg of poly(dI-dC) for 1 h. Incubation with the radioactive DNA (2.5×10^5 to 10×10^5 cpm/ml of buffer) was carried out at 16°C for 3 h in a rotating cylinder. After repeated washes in SW buffer, the membrane was air dried and autoradiographed. For Western blot analysis, the protein-loaded membrane was treated according to the manufacturer's protocol (Aurion).

Probe construction. The F1 sequence upstream of the chicken lysozyme gene was synthesized as complementary oligonucleotides of 66 bp in length, annealed, and end labelled with α-^{32}P-deoxynucleoside triphosphates with Klenow enzyme. The CTCF binding element FPV from the chicken c-*myc* promoter region was isolated from pFPV (16) by *Hin*dIII/*Eco*RI digestion. The resulting fragment of 88 bp was end labelled with α-^{32}P-deoxynucleoside triphosphates with Klenow enzyme, loaded onto a 5% polyacrylamide gel, cut out, and eluted in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.6) under gentle agitation for 15 h at room temperature.

EMSA. DNA-protein binding reactions for the EMSA were carried out in 40 μl of 1× SW buffer supplemented with 1 to 4 μg of salmon sperm DNA and 0.5 to 1.0 μg of poly(dI-dC) depending on the protein amounts. After preincubation for 15 min on ice, 15 to 40 fmol of the radiolabelled probe was added and incubated for 20 min at room temperature. DNA-protein complexes were analyzed on nondenaturing polyacrylamide gels (5% [wt/vol] acrylamide; 0.125% [wt/vol] bisacrylamide) in TBE buffer (90 mM Tris, 90 mM borate, 2 mM EDTA, pH 8.3). Electrophoresis was performed at 4°C with a field strength of 12 V/cm for 3 h.

DMS-DEPC interference assay. A single-stranded F1 oligonucleotide (sense or antisense) was end labelled with [γ-^{32}P]dATP with T4 polynucleotide kinase according to the supplier's instructions (Boehringer; New England Biolabs). The probes were heated for 2 min to denature the enzyme, and the DNA was annealed to the unlabelled complementary strand. The oligonucleotides were passed over a Sephadex G-50 column and ethanol precipitated. The probes were treated with dimethyl sulfoxide (DMS) or formic acid according to the work of Maniatis et al. (23). The diethylpyrocarbonate (DEPC) treatment was carried out as described by Sturm et al. (28). DMS- and DEPC-treated probes were used in a preparative EMSA reaction. After electrophoresis, the bands were excised and eluted in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.6) overnight at 4°C. After pelleting of the acrylamide by centrifugation, the DNA was precipitated and cleaved with 10% piperidine at 90°C for 30 min. A total of 10,000 cpm of each probe was applied to a 10% polyacrylamide sequencing gel.

Site-directed mutagenesis and cell culture. Mutated F1 sequences were generated with the Muta-Gene phagemid in vitro mutagenesis kit (Bio-Rad) as described by Kunkel (19). The chicken cell line HD3 was cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2% chicken serum, 100 μg of streptomycin per ml, and 100 U of penicillin per ml (Gibco). DNA transfections and chloramphenicol acetyltransferase (CAT) assays were carried out as described elsewhere (4, 17).

RESULTS

Purification and sequencing of NeP1 show identities with CTCF. Since only a few actively repressing transcription factors have been analyzed in detail (for a review, see reference 14) and since the modular structure of the lysozyme silencer shows interesting synergy in both repression and induction, we wanted to purify and characterize NeP1. HeLa cell nuclear extract was prepared and applied onto a Q-Sepharose column to enrich NeP1. The fractions eluting from 350 to 500 mM NaCl were further fractionated with a heparin-Sepharose column. NeP1 purification was controlled by EMSAs at each fractionation step with the F1-DNA sequence used as probe. Fractions with detectable NeP1 DNA binding activity were eluted with 900 mM NaCl with a step gradient. After dialysis, the protein fractions were applied to an F1-DNA-affinity column. Fractions eluting with 600 mM NaCl were further purified by SDS-PAGE. Analytical lanes were blotted, and NeP1 was identified by a Southwestern procedure (5). After separation on a preparative gel, the NeP1 band was cut out, blotted, and sequenced (8). The sequences obtained were LRY-TEEGK and (S)DLGVHLRK. Since the peptides are derived from a digestion with endoproteinase LysC, the peptide sequences given should be preceded by a lysine residue. A database search identified the protein as the transcription factor CTCF (11, 16). This finding was surprising and unexpected, since inspection of the long AT-rich NeP1 binding sequence F1 showed no apparent similarity to the previously characterized GC-rich CTCF binding sites in the c-*myc* gene proximal promoter regions. Therefore, we wanted to employ additional criteria to test whether indeed NeP1 is identical to CTCF.

NeP1 and CTCF have identical properties. Initially we asked whether the F1 binding NeP1 protein copurifies with CTCF. The NeP1 purification was controlled at each fractionation step by binding to the F1 DNA sequence. Binding was tested by EMSA (not shown) as well as by Southwestern analysis (Fig. 1A). Typical for Southwestern experiments is the detection of several nonspecific bands in addition to the specific band. Binding specificity was determined by competition experiments with the unlabelled F1 DNA and as a control with a nonspecific unlabelled DNA. These competitions revealed that proteins migrating at an apparent molecular mass of 130 kDa bind specifically to the F1 element (5). The same apparent molecular weight is found for CTCF (21) (Fig. 1B). In order to assess whether the F1 binding NeP1 protein copurifies with CTCF, we performed a Western blot analysis. A band recognized by immunoblotting with an anti-CTCF antibody (Fig. 1B) comigrates with a band identified by the F1 fragment in the Southwestern experiment. Moreover, the NeP1 Southwestern band and the CTCF Western band are present in identical chromatography fractions throughout the entire purification procedure (Fig. 1). Therefore, NeP1 has the properties of CTCF in affinity chromatography on the F1-containing column.

EMSA experiments with the F1 DNA and nuclear extracts

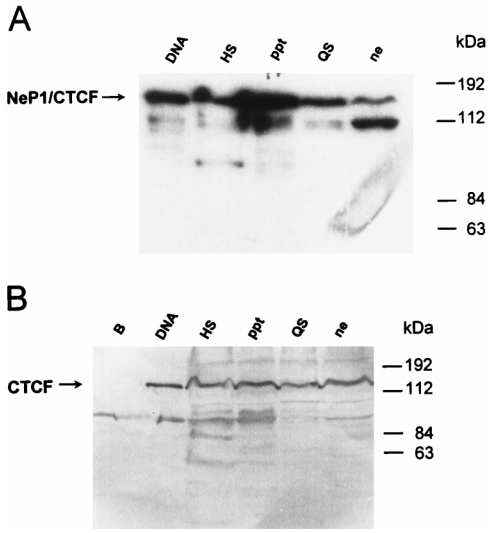

FIG. 1. Purification of a protein detected by NeP1-specific DNA and by a CTCF-specific antibody. (A) Autoradiogram of a Southwestern experiment with SDS-PAGE-separated and blotted proteins visualized with the ^{32}P-labelled F1 DNA. The band marked "NeP1/CTCF" is the specific band (5) seen in unfractionated nuclear extract (ne), after Q-Sepharose fractionation (QS), subsequent ammonium sulfate precipitation (ppt), and further fractionation by heparin Sepharose (HS) and DNA-affinity column (DNA). (B) Western blot analysis of fractionated nuclear extract (see panel A) after SDS-PAGE and blotting with a polyclonal antibody against chicken CTCF (16). As a negative control, bacterial protein extract was used (lane B).

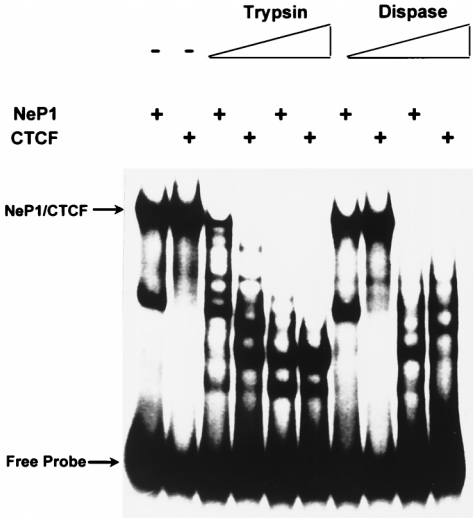

FIG. 2. Proteolytic band shift assay. Affinity-purified NeP1 or nuclear extract from COS cells expressing chicken CTCF was used for band shift reactions with the F1 probe. Lanes marked with trypsin or dispase indicate the protein treatment with increasing amounts of the respective protease. Affinity-purified NeP1 shows an additional band of higher mobility which is frequently seen and which very likely is caused by a degradation during the purification procedure.

from several species revealed the same mobility of the specifically retarded complex (17). The detailed comparison between chicken and human nuclear extracts in proteolytic clipping assays demonstrated a high conservation of this protein in the two species (17), a finding recently demonstrated for CTCF on the amino acid level as well (11). Therefore, we asked whether the cloned and expressed chicken CTCF would generate a band shift with the F1 sequence comparable to that of HeLa nuclear extract. This is clearly the case, since an identical position is found for both retarded complexes (Fig. 2, lanes 1 and 2). Affinity-purified NeP1 shows an additional band of higher mobility which is frequently seen and which very likely is caused by a degradation during the purification procedure.

We next used protease digestion to determine the identity of purified NeP1 with cloned CTCF. Treatment of both affinity-purified NeP1 and cloned chicken CTCF expressed in COS cells with two different proteases (trypsin and dispase) revealed a set of partial proteolytic degradation products binding to the F1 sequence (Fig. 2). All of these products were found for both NeP1 and CTCF, providing further evidence for the identity of NeP1 and CTCF.

Since previous publications have identified F1 and site V as strong and specific binding sites for NeP1 and CTCF, respectively, we wondered whether F1 and site V show similar affinities to the purified NeP1 protein. Therefore, we carried out gel retardation assays using both response elements and the affinity-purified NeP1 protein (Fig. 3). Both DNA elements generate similarly retarded bands, showing only a slight difference in migration between the DNA probes caused by the different probe lengths. Specific competition with the F1 sequence is identical for both retarded complexes, and nonspecific competition does not affect either complex (Fig. 3), indicating that there is no gross difference in binding affinity to these two binding sites. Therefore, CTCF and NeP1 bind equally well each other's cognate binding sites.

Taken together with our data on NeP1 peptide sequencing, all of the above tested parameters argue strongly for the identity of NeP1 and CTCF.

Protein contact sites within the F1 sequence. We wondered whether the apparent divergence between F1 and other CTCF binding sites would be reflected in a different contact pattern visualized by a methylation and carbethoxylation interference assay. In this assay, the F1 sequence was partially methylated at the N-7 positions of the guanines or partially carbethoxylated at the N-7 positions of the adenines. Subsequent EMSA with the purified NeP1 fraction separated the protein-bound DNA from the unbound DNA. The methylation or carbethoxylation of each of the two DNA strands was visualized after piperidine reaction and separation of the reaction products on a sequencing gel (Fig. 4). Several positions throughout the binding site, when modified, were found to interfere with protein binding

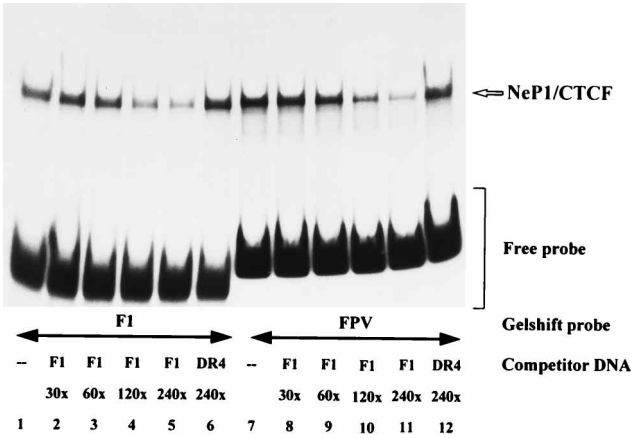

FIG. 3. Affinity-purified NeP1 binds with similar affinities to the F1 and the FPV probes. The band shift reaction was analyzed in the presence or absence of specific (F1) or nonspecific competitors with the indicated amounts. Nonspecific competition was tested with a binding site for the thyroid hormone receptor (DR4).

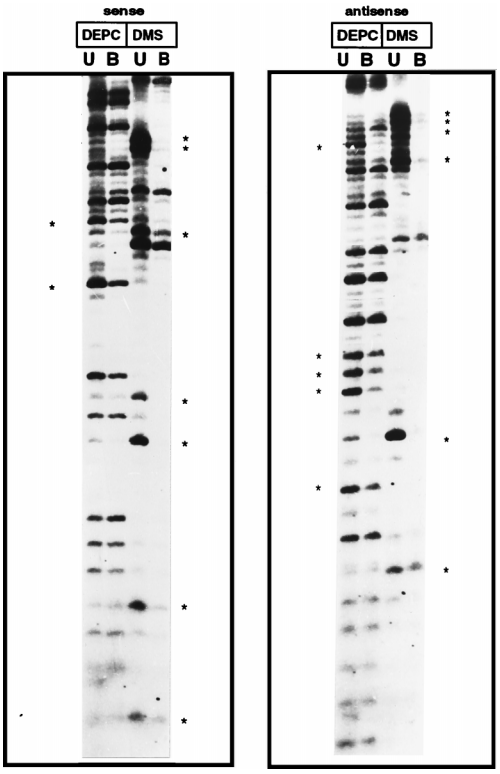

FIG. 4. Specific methylated or carbethoxylated nucleotides in the F1 sequence interfere with protein binding. Either of the ^{32}P-end-labelled strands (sense or antisense) was annealed with the complementary strand and partially methylated (DMS) or carbethoxylated (DEPC). The unbound (U) and the protein-bound (B) DNA were analyzed after a gel shift reaction with purified NeP1. The modified nucleotides interfering with binding are indicated by asterisks.

(labelled with an asterisk in Fig. 4). The distribution of these interfering sites is indicated on the DNA double helix (Fig. 5A). A large region is covered by these sites, which are found on the back as well as on the front of the DNA. A gap in contact sites is seen in the vicinity of the site preferentially cut by DNase I after NeP1 binding in vitro (4). Similarities and differences in these interference patterns among the three sequences (F1 and c-*myc* sites A and V) are summarized and discussed below (see Fig. 7).

To verify the residues critical for NeP1 binding, we generated clustered point mutations within the F1 sequence. Three different mutants (mut1 to mut3) containing mutations in two adjacent nucleotides (Fig. 5A and 7) were designed. Gel retardation analysis with the purified NeP1 fraction revealed that all three mutant F1 fragments show a severely reduced binding affinity: mut2 retains a residual binding activity, whereas the specific binding of mut1 and mut3 is not detectable (Fig. 5B).

Mutations which specifically eliminate protein binding to the F1 sequence reduce synergistic repression of the silencer mediated by v-*erbA*. The natural sequence arrangement of F1 within the chicken lysozyme silencer is such that it is adjacent to the T3R binding site F2 (4). In order to test whether the contribution of the F1 sequence to repression depends specifically on the ability of NeP1-CTCF to bind to this sequence, we cloned the mutated F1 sites (F1mut1 through F1mut3) and a double mutation (F1mut1+mut3) next to the F2 site upstream of the tkCAT reporter gene. Transfection of these F1-F2 constructs into chicken HD3 cells showed, as predicted by our previous data (4), a strong repression mediated by the presence of both F1 and F2 elements in response to v-*erbA* binding to the F2 element sequence (Fig. 5C). While the wild-type F1-F2 combination leads to a synergistic increase in repression, the F1 element alone does not noticeably affect reporter gene activity (Fig. 5C) except in the case of a multimerized F1 element (4). Remarkably, all three mutations in the F1 sequence which eliminated or decreased NeP1-CTCF binding result in practically complete loss of synergistic repression (Fig. 5C). Therefore, CTCF binding to the element F1 is required for efficient transcriptional repression in response to v-*erbA* binding in the vicinity of the F2 element.

Different sets of CTCF Zn fingers bind the divergent CTCF binding sequences. Having shown that NeP1 factor binding to the F1 lysozyme silencer element is CTCF, we asked how this protein could recognize this DNA sequence, which has so little similarity to the sequence of CTCF binding sites determined previously in c-*myc* genes (see Fig. 7) (11, 16). Since a single Zn finger can recognize and make contacts with three to four consecutive base pairs (7, 10), the arrangement of CTCF-contacting bases in the F1 DNA segment compared to that in two other CTCF binding sites (summarized in Fig. 7) suggests that within the 11-Zn-finger domain of CTCF different groups of individual fingers may be used to create different DNA interaction subdomains capable of specific binding to different nucleotide sequences. If so, then different fingers deleted sequentially from either end of the 11-Zn-finger region should contribute specifically to the binding to divergent sequences. This hypothesis was tested by using the full-length DNA binding domain of CTCF and its 11 serially truncated forms (11) for EMSA with the F1 sequence as a DNA probe. The amino-terminally deleted CTCF forms containing Zn fingers 2 to 11, 3 to 11, and 4 to 11 generated an amount of complexed DNA similar to that of the full-length DNA binding domain containing Zn fingers 1 to 11 (Fig. 6A). Even the deletion with the remaining fingers 5 to 11 showed only a slight reduction of binding activity in contrast to that exhibited by fingers 6 to 11, which have lost almost all of the F1 binding activity. Similarly, the carboxy-terminal deletions show that some particular fingers are dispensable: fingers 1 to 10 show wild-type binding, whereas fingers 1 to 9 and 1 to 8 display a gradual loss in generating a DNA complex. Only fingers 1 to 7 and the more severe deletions 1 to 6 and 1 to 5 have lost detectable DNA binding activity. Therefore, fingers 1 to 8 or fingers 5 to 11 are sufficient to bind to the F1 DNA fragment, indicating that fingers 5 to 8 are necessary. This particular type of finger contribution to DNA binding by CTCF is remarkably different from that described previously with the site V sequence and with the c-*myc* site A sequence (11). For example, and also for control and comparison, c-*myc* site A was tested here as well, showing that fingers 4 to 11 have lost almost all binding activity and that the C-terminal deletion of the single finger number 11 (construct 1 to 10) also abrogates DNA binding (Fig. 6B).

Thus, the tested sequences require different sets of the CTCF Zn fingers (Fig. 7). This finding explains how a particular DNA binding protein efficiently and specifically binds highly divergent DNA sequences.

DISCUSSION

Transcriptional repression in eukaryotes is achieved by a variety of mechanisms. One class of mechanisms involves specific domains of DNA binding proteins actively repressing gene transcription (for a recent review, see reference 14). These repression domains may either act directly on the preinitiation complex, or on activation domains of transcriptional activators including their coactivators, or bind to corepressors. Combi-

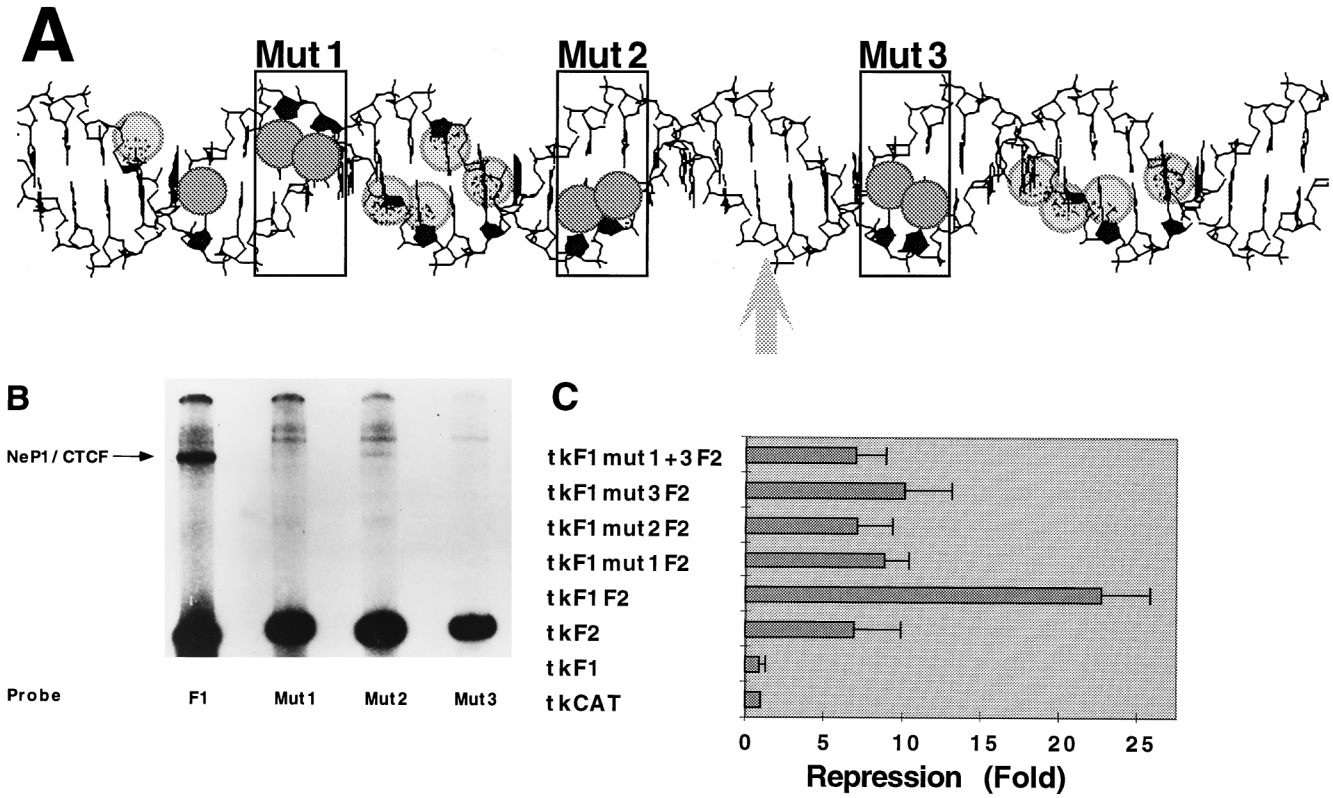

FIG. 5. Mutations within the F1 sequence interfere with binding and function. (A) The F1 sequence DNA helix model is shown with the interfering nucleotides (from Fig. 4) indicated by shaded circles. Dark shading indicates the interfering base modifications facing the front of the helix, whereas the light shading indicates a location behind the helix. The sugar residues of interfering nucleotides are filled. The three different mutations with two nucleotides exchanged each are indicated (mut1, mut2, and mut3), and the DNase I-hypersensitive site seen in footprint reactions (4) (arrow) is shown. (B) Band shift reactions with purified NeP1 and the wild-type F1 probe or the three different F1 mutations. (C) CAT activity of the indicated reporter constructs after transfection into HD3 cells is expressed relative to that of the reporter plasmid tkCAT. Reporter gene activity was set to 1 and showed a standard deviation of ±0.1.

nations of these may be used as well. Such a combination very likely mediates the repression exerted by the unliganded thyroid hormone receptor (TR). For this receptor, it has been shown that in the absence of ligand the carboxy-terminal domain is able to interact with TFIIB, a component of the preinitiation complex. The presence of ligand abrogates repression and binding (2). In addition, in analyzing transcriptional repression in vitro, binding of TR to TATA-binding protein (TBP) was shown to inhibit the formation of a functional preinitiation complex (13). Again, this interaction was alleviated by the presence of ligand. In contrast to these direct effects of TR on the preinitiation complex, corepressors have been identified as interacting with TR (6, 15).

We have previously identified the functional TR binding site F2 within the lysozyme gene silencer at −2.4 kb upstream of the transcription start site (4) and have found that binding of another protein, called NeP1, to 50 bp of DNA next to the F2 site is required to mediate efficient transcriptional repression by unliganded TR (4, 17).

The mechanism for synergistic repression by TR and NeP1 is not known. Synergy mediated by cooperative binding to naked DNA might be excluded, since, at least in vitro, the DNA binding affinity is not changed for TR or NeP1 by the presence or absence of the respective neighboring factor (1). Another possible mechanism involves chromatin remodelling by induction of a particular DNA conformation such as bending accompanied by nucleosomal repositioning or disruption. Indeed, a strong DNA bend induced by NeP1 binding to the F1 element has been shown previously (1). Moreover, binding of different TR complexes together with NeP1 did not change the magnitude of the DNA flexure angle; rather, the position of the bend center shifted within a range of about 20 bp. Simultaneously, the bending orientation was moved in the case of the NeP1–TR-RXR complex (1). We think that NeP1-induced DNA bending may occur in conjunction with repositioning of nucleosomes resulting in altered chromatin structure. A role for nucleosome assembly in TR silencing has been found in *Xenopus* oocytes (30). Nucleosome redistribution caused by NeP1 may facilitate interaction of TR with both DNA and its protein partners connecting TR to the transcription preinitiation complex.

Independent of the lysozyme silencer analysis, a nuclear factor mediating active repression on the c-*myc* gene has been identified and analyzed (11, 16, 21). This factor, CTCF, is characterized by a DNA binding domain consisting of 11 zinc fingers and by two repression domains. These repression domains can be transferred to the Gal4-DNA binding domain and reduce transcription (11).

Here we show that NeP1 is identical to CTCF, despite extreme divergent binding sites. This identity is suggested by many lines of evidence. (A) Both proteins have an identical apparent molecular mass of 130 kDa (Fig. 1). (B) Identical peptide sequences after independent, binding-site-specific purification were generated. (C) Both proteins bind with similar affinity to the respective binding site of the other protein (Fig. 2 and 3). In addition, a large DNase I footprint area of about

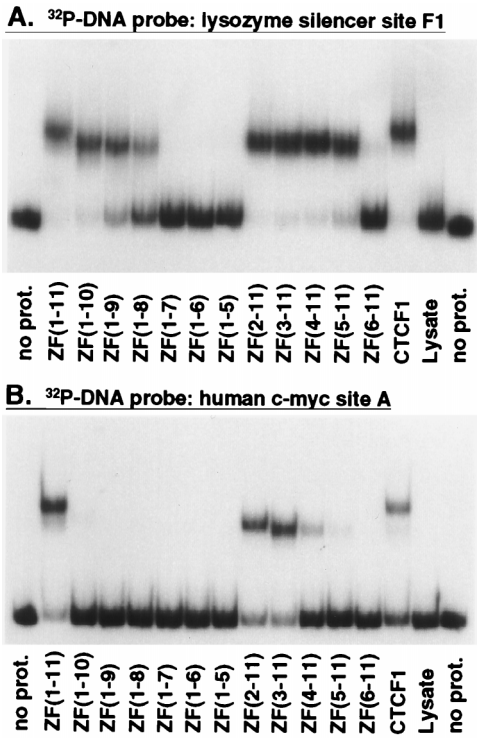

FIG. 6. F1 and c-*myc* site A are bound by different sets of Zn finger deletions of CTCF. The DNA binding domain of CTCF containing all 11 Zn fingers, ZF (1-11), or different N-terminal and C-terminal deletions leaving the indicated sets of Zn fingers intact were in vitro transcribed and translated, and equal amounts of each protein were incubated with the F1 probe (A) or with the c-*myc* site A probe (B). Retarded complexes were resolved after gel electrophoresis. Control lanes included the CTCF DNA binding domain synthesized from the pCITE/CTCF1 plasmid (16), TnT lysate not primed with a template (Lysate), or no protein.

50 bp (4, 21) is generated, identical DNA bending was found for the affinity-purified NeP1 and for the cloned CTCF (1), and DNA binding was shown to be dependent on Zn ions or on the Zn finger domain (16, 17). Furthermore, a ubiquitous distribution has been observed for both, and the proteins were found to be very conserved between different species (16, 17). (D) There are apparently no other CTCF gene family members because, on a Southern blot containing human, mouse, chicken, and frog DNA probed with human CTCF cDNA at low stringency, genomic CTCF appears as a single copy locus (11b). In addition, in total RNA samples from different human tissues the CTCF mRNA is represented by a single Northern blot band. Moreover, only one single chromosome locus was found to contain human genomic CTCF as determined by both fluorescent in situ hybridization and somatic cell hybrid methods (11a). Previously noticed multiple CTCF forms in Western blot analyses (16) are likely to be produced by specific proteolytic cleavage (20a). Therefore, the name NeP1 will be replaced by CTCF, irrespective of whether the protein is isolated from chicken or from mammalian sources.

A remarkable flexibility in binding DNA sequence selection by CTCF is evident by comparing the CTCF binding sites in the c-*myc* gene from chickens and humans with the lysozyme silencer site. Pairwise comparison of site A to sites FPV and F1 shows that from a total of 46 positions only 17 (in F1) or 18 (in FPV) positions contain identical nucleotides (Fig. 7). Moreover, homologies with the human c-*myc* site A are restricted to one-half of the chicken c-*myc* site (FPV), whereas the homologous positions in the lysozyme F1 site are distributed over the length of the sequence, thus indicating that there is no common consensus sequence for these three CTCF binding sites. Accordingly, binding of these sequences by CTCF is mediated by making contacts at many different nucleotides as determined by DNA-modification interference assays (Fig. 5 and 7)

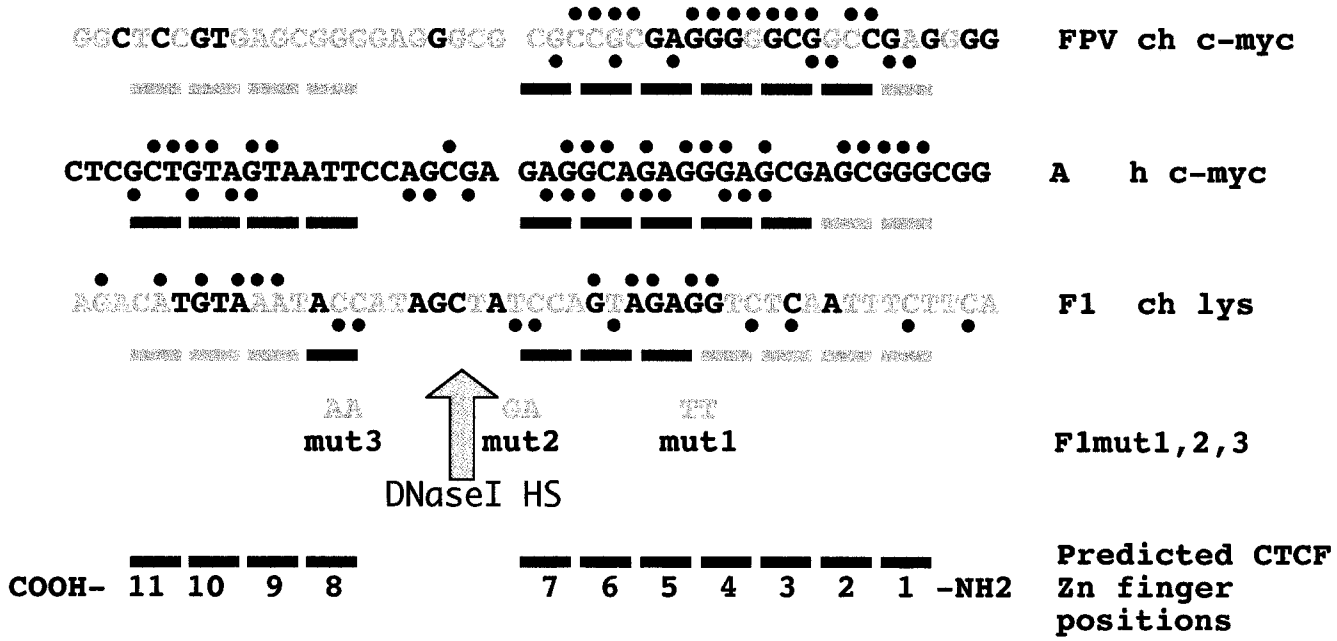

FIG. 7. Comparison of three different CTCF binding sites and the finger disposition model. The CTCF binding sequences FPV from the chicken c-*myc* gene, site A from the human c-*myc* gene, and F1 from the chicken lysozyme silencer are aligned. Identical nucleotides at the same positions relative to the site A sequence are shown in boldface. Indicated are the nucleotides substituted in the three F1 mutated elements. Nucleotides interfering with CTCF binding after DMS or DEPC treatment are indicated by the dots above the sequence (upper strand) or below the sequence (lower strand), and the CTCF-induced DNase I-hypersensitive site is indicated by an arrow. Individual CTCF Zn fingers indispensable for DNA binding are shown by black bars. Fingers dispensable for binding to a particular DNA sequence are printed in gray. For further details, see the text.

(11). For example, while the FPV site shows contact positions restricted to about one-half of the sequence, the *myc* site A and the F1 sequences display these contacts over a much longer stretch and contain contact points absent in the site FPV sequence. Since one Zn finger can make specific contacts with only three to four nucleotides (7, 10), the different number and arrangement of contact points in these three CTCF binding sequences suggest that Zn fingers required for binding are different for all three response elements (Fig. 7). Indeed, proteolytic protection analysis has shown that the longer CTCF binding sequence with more contact points involves more CTCF fingers in binding (11). Testing the CTCF Zn finger domain with serially truncated fingers for binding to FPV, to site A (11), and to site F1 has revealed that, upon one-by-one finger deletion, different sets of individual fingers remain sufficient to mediate binding to each DNA sequence (Fig. 6 and 7).

Although a general code for predicting the DNA-Zn finger recognition based on amino acid composition of a given finger is still very speculative, some essential rules have been established for predicting, at least partially, DNA subsite sequence specificity (29). According to these rules, the 11 Zn fingers of CTCF could be aligned with the DNA sequences of three binding sites as depicted in Fig. 7. Taking into account obvious limitations of such predictions, this model puts together reasonably well most of the contact bases determined by interference assays with the particular groups of individual fingers which were found to be sufficient for binding of CTCF to individual target sites.

For example, the C-terminal set of four Zn fingers, which could be deleted without losing binding to the FPV sequence (11), is positioned by this alignment to the region of the FPV sequence devoid of contact points. In the case of the lysozyme sequence F1, three C-terminal fingers are placed by the model to several contact nucleotides at the 5′ end of the F1 sequence. Therefore, removing these fingers is expected to eliminate binding. However, they could be deleted without losing binding (Fig. 6A). This apparent discrepancy with the theoretical model may be explained if the F1 DNA binding by the domain without fingers 1 to 4 is stabilized by fingers 9 to 11, whereas deletion of fingers 9 to 11 could be compensated in DNA binding by the presence of fingers 1 to 4. This appears to be the case, since despite the requirement of fingers from 5 to 8 for binding to the F1 site (Fig. 6A), a domain composed solely of these fingers is not sufficient in binding the F1 DNA (12). Additional support for the finger disposition model presented in Fig. 7 is provided by mutational analysis of CTCF binding sites. Each of the three different point mutations within the F1 site, mut1 to mut3, shown in Fig. 7, which reduce CTCF binding (Fig. 5B) and abolish synergistic transcriptional repression (Fig. 5C), is predicted by the model to alter a recognition subsite for exactly those fingers which cannot be deleted without losing binding (these fingers are shown by black bars in Fig. 7). Similarly, mutations eliminating CTCF binding to two c-*myc* sites, the Nsi mutation in the chicken c-*myc* site FPV (16) and the ACA mutation in the human c-*myc* site A (11), are within subsites predicted by the model to interact with fingers 5 to 6 (essential for binding to the FPV sequence [Fig. 7]) and fingers 10 to 11 (essential for binding to site A [Fig. 7]).

To fit a recognition code and to adjust with contact nucleotides distributed over very long DNA sequences of sites A and F1, the finger disposition model suggested here requires a gap between N-terminal and C-terminal sets of individual fingers as shown in Fig. 7. However, inspection of the CTCF amino acid sequence (11, 16) shows that in the contiguous 11-Zn-finger CTCF domain these sets of fingers are not separated from one another. Therefore, it is likely that, in order to make contacts with CTCF, the structure of the DNA segment within site A or F1 can locally alter, i.e., by melting and looping out the DNA separating the two contact point clusters. Such structural alteration indeed does appear to take place when CTCF binds the site F1 sequence. It has been noticed previously that binding of CTCF induces a strong DNase I-hypersensitive site within the 50-bp-long F1 footprint (4) at the position that coincides exactly with the DNA region expected to loop out according to the model (Fig. 7). Moreover, the DNA bending center of the F1-CTCF complex has been mapped to this region as well (1). In contrast to CTCF binding sites A and F1, the CTCF contacting nucleotides are grouped in one region of the chicken c-*myc* site FPV (Fig. 7), and the finger disposition model does not require structural changes in DNA in order to align contact points with the set of indispensable fingers. Accordingly, there is no DNase I-hypersensitive site within the FPV CTCF footprint (22). Thus, local DNA structure alterations induced by CTCF may depend on the primary sequence context of its binding sites and may be determined by combinatorial contributions of different sets of individual Zn fingers to binding. Currently, we are testing possible functional differences of different CTCF binding sites.

In conclusion, we have shown that, despite the difference in target DNA binding sequences, the NeP1 protein required for transcriptional regulation of the lysozyme gene silencer by the thyroid receptor is identical to the 11-Zn-finger transcriptional repressor CTCF. This finding suggests a new role for CTCF in structural organization of at least some chromatin regions which contain thyroid-receptor-responsive regulatory elements. Because of the remarkable flexibility in binding site selection by utilizing different sets of Zn fingers exemplified here by comparing CTCF binding sequences from the lysozyme silencer and from c-*myc* genes, we conclude that CTCF is a true multivalent transcriptional factor possessing multiple sequence specificities.

ACKNOWLEDGMENTS

We are grateful to L. Schäfer-Pfeiffer for excellent technical assistance and help with the cell culture and to N. McKenna for critically reading the manuscript.

The work was supported by the Deutsche Forschungsgemeinschaft (SFB 272), by the Fonds der Chemischen Industrie, and, in part, by the NIH/NCI RO1 grants CA68360 to V. V. Lobanenkov and CA20068 to P. E. Neiman.

REFERENCES

1. **Arnold, R., M. Burcin, B. Kaiser, M. Muller, and R. Renkawitz.** 1996. DNA bending by the silencer protein NeP1 is modulated by TR and RXR. Nucleic Acids Res. **24:**2640–2647.
2. **Baniahmad, A., I. Ha, D. Reinberg, S. Tsai, M. J. Tsai, and B. W. O'Malley.** 1993. Interaction of human thyroid hormone receptor beta with transcription factor TFIIB may mediate target gene derepression and activation by thyroid hormone. Proc. Natl. Acad. Sci. USA **90:**8832–8836.
3. **Baniahmad, A., A. C. Köhne, and R. Renkawitz.** 1992. A transferable silencing domain is present in the thyroid hormone receptor, in the v-erbA oncogene product and in the retinoic acid receptor. EMBO J. **11:**1015–1023.
4. **Baniahmad, A., C. Steiner, A. C. Köhne, and R. Renkawitz.** 1990. Modular structure of a chicken lysozyme silencer—involvement of an unusual thyroid-hormone receptor-binding site. Cell **61:**505–514.
5. **Burcin, M., A. C. Köhne, D. Runge, C. Steiner, and R. Renkawitz.** 1994. Factors influencing nuclear receptors in transcriptional repression. Semin. Cancer Biol. **5:**337–346.
6. **Chen, J. D., and R. M. Evans.** 1995. A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature **377:**454–457.
7. **Desjarlais, J. R., and J. M. Berg.** 1992. Towards rules relating zinc finger protein sequences and DNA binding site preference. Proc. Natl. Acad. Sci. USA **89:**7345–7349.
8. **Eckerskorn, C., and F. Lottspeich.** 1989. Internal amino acid sequence analysis of proteins separated by gel electrophoresis after tryptic digestion in polyacrylamide matrix. Chromatographia **28:**92–94.

9. **Eick, D., F. Kohlhuber, D. A. Wolf, and L. J. Strobl.** 1994. Activation of pausing RNA polymerases by nuclear run-on experiments. Anal. Biochem. **218:**347–351.
10. **El-Baradi, T., and P. Tomas.** 1991. Zinc finger proteins: what we know and what we would like to know. Mech. Dev. **35:**155–169.
11. **Filippova, G. N., S. Fagerlie, E. M. Klenova, C. Myers, Y. Dehner, G. Goodwin, P. E. Neiman, S. J. Collins, and V. V. Lobanenkov.** 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-*myc* oncogenes. Mol. Cell. Biol. **16:**2802–2813.

11a. **Filippova, G. N., N. Doggett, L. Meinke, P. E. Neiman, S. Collins, and V. V. Lobanenkov.** Unpublished data.

11b. **Filippova, G. N., and V. V. Lobanenkov.** Unpublished data.

12. **Filippova, G. N., M. Macbeth, and V. Lobanenkov.** Unpublished data.
13. **Fondell, J. D., F. Brunel, K. Hisatake, and R. G. Roeder.** 1996. Unliganded thyroid hormone receptor alpha can target TATA-binding protein for transcriptional repression. Mol. Cell. Biol. **16:**281–287.
14. **Hanna-Rose, W., and U. Hansen.** 1996. Active repression mechanisms of eukaryotic transcription repressors. Trends Genet. **12:**229–234.
15. **Hörlein, A. J., A. M. Naar, T. Heinzel, J. Torchia, B. Gloss, R. Kurokawa, A. Ryan, Y. Kamel, M. Soderstrom, C. K. Glass, and M. G. Rosenfeld.** 1995. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. Nature **377:**397–404.
16. **Klenova, E. M., R. H. Nicolas, H. F. Paterson, A. F. Carne, C. M. Heath, G. H. Goodwin, P. E. Neiman, and V. V. Lobanenkov.** 1993. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-*myc* gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell. Biol. **13:**7612–7624.
17. **Köhne, A. C., A. Baniahmad, and R. Renkawitz.** 1993. NeP1—a ubiquitous transcription factor synergizes with v-ERBA in transcriptional silencing. J. Mol. Biol. **232:**747–755.
18. **Krumm, A., L. B. Hickey, and M. Groudine.** 1995. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev. **9:**559–572.
19. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82:**488–492.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.

20a. **Lobanenkov, V. V.** Unpublished data.

21. **Lobanenkov, V. V., R. H. Nicolas, V. V. Adler, H. Paterson, E. M. Klenova, A. V. Polotskaja, and G. H. Goodwin.** 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. Oncogene **5:**1743–1753.
22. **Lobanenkov, V. V., R. H. Nicolas, M. A. Plumb, C. A. Wright, and G. H. Goodwin.** 1986. Sequence-specific DNA-binding proteins which interact with (G+C)-rich sequences flanking the chicken c-myc gene. Eur. J. Biochem. **159:**181–188.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Marcu, K. B., S. A. Bossone, and A. J. Pate.** 1992. Myc function and regulation. Annu. Rev. Biochem. **61:**809–860.
25. **Renkawitz, R.** 1990. Transcriptional repression in eukaryotes. Trends Genet. **6:**192–197.
26. **Strobl, L. J., and D. Eick.** 1992. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. EMBO J. **11:**3307–3314.
27. **Strobl, L. J., F. Kohlhuber, J. Mautner, A. Polack, and D. Eick.** 1993. Absence of a paused transcription complex from the c-myc P2 promoter of the translocation chromosome in Burkitt's lymphoma cells: implication for the c-myc P1/P2 promoter shift. Oncogene **8:**1437–1447.
28. **Sturm, R., T. Baumruker, B. R. J. Franza, and W. Herr.** 1987. A 100-kD HeLa cell octamer binding protein (OBP100) interacts differently with two separate octamer-related sequences within the SV40 enhancer. Genes Dev. **1:**1147–1160.
29. **Suzuki, M., and N. Yagi.** 1994. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. Proc. Natl. Acad. Sci. USA **91:**12357–12361.
30. **Wong, J. M., Y. B. Shi, and A. P. Wolffe.** 1995. A role for nucleosome assembly in both silencing and activation of the Xenopus TR beta A gene by the thyroid hormone receptor. Genes Dev. **9:**2696–2711.