

# An Exceptionally Conserved Transcriptional Repressor, CTCF, Employs Different Combinations of Zinc Fingers To Bind Diverged Promoter Sequences of Avian and Mammalian *c-myc* Oncogenes

GALINA N. FILIPPOVA,<sup>1</sup> SARA FAGERLIE,<sup>1</sup> ELENA M. KLENOVA,<sup>2†</sup> CENA MYERS,<sup>1</sup>  
YVONNE DEHNER,<sup>1</sup> GRAHAM GOODWIN,<sup>2</sup> PAUL E. NEIMAN,<sup>1</sup>  
STEVE J. COLLINS,<sup>1</sup> AND VICTOR V. LOBANENKOV<sup>1\*</sup>

*Fred Hutchinson Cancer Research Center, Seattle, Washington 98104,<sup>1</sup> and Chester Beatty Laboratories, Institute of Cancer Research, London SW3 6JB, United Kingdom<sup>2</sup>*

Received 19 September 1995/Returned for modification 20 December 1995/Accepted 6 March 1996

**We have isolated and analyzed human CTCF cDNA clones and show here that the ubiquitously expressed 11-zinc-finger factor CTCF is an exceptionally highly conserved protein displaying 93% identity between avian and human amino acid sequences. It binds specifically to regulatory sequences in the promoter-proximal regions of chicken, mouse, and human *c-myc* oncogenes. CTCF contains two transcription repressor domains transferable to a heterologous DNA binding domain. One CTCF binding site, conserved in mouse and human *c-myc* genes, is found immediately downstream of the major P2 promoter at a sequence which maps precisely within the region of RNA polymerase II pausing and release. Gel shift assays of nuclear extracts from mouse and human cells show that CTCF is the predominant factor binding to this sequence. Mutational analysis of the P2-proximal CTCF binding site and transient-cotransfection experiments demonstrate that CTCF is a transcriptional repressor of the human *c-myc* gene. Although there is 100% sequence identity in the DNA binding domains of the avian and human CTCF proteins, the regulatory sequences recognized by CTCF in chicken and human *c-myc* promoters are clearly diverged. Mutating the contact nucleotides confirms that CTCF binding to the human *c-myc* P2 promoter requires a number of unique contact DNA bases that are absent in the chicken *c-myc* CTCF binding site. Moreover, proteolytic-protection assays indicate that several more CTCF Zn fingers are involved in contacting the human CTCF binding site than the chicken site. Gel shift assays utilizing successively deleted Zn finger domains indicate that CTCF Zn fingers 2 to 7 are involved in binding to the chicken *c-myc* promoter, while fingers 3 to 11 mediate CTCF binding to the human promoter. This flexibility in Zn finger usage reveals CTCF to be a unique “multivalent” transcriptional factor and provides the first feasible explanation of how certain homologous genes (i.e., *c-myc*) of different vertebrate species are regulated by the same factor and maintain similar expression patterns despite significant promoter sequence divergence.**

The *c-myc* proto-oncogene encodes a nuclear phosphoprotein with leucine zipper and helix-loop-helix structural motifs which is involved in regulating important cellular functions, including cell cycle progression, differentiation, and apoptosis (27, 31, 40). In several types of human and animal cancers, *myc* is deregulated. Maintenance of the level of the *c-myc* mRNA is achieved by regulation of both transcription initiation and transcriptional elongation (for reviews, see references 27 and 40). A wide variety of signals can influence both initiation and elongation of the *c-myc* mRNA. Most of these signals work through *cis*-acting regulatory DNA sequences near or within the *c-myc* gene, and some of these sequences bind specifically a number of nuclear factors (for details, see reference 27). Identification and study of such factors may provide insight into molecular mechanisms of normal and aberrant *c-myc* expression.

Analysis of factors binding to the 5'-flanking noncoding

DNA sequences of the chicken *c-myc* gene (25) resulted in identification (22, 23), purification (24), and molecular cloning of the chicken 11-Zn-finger transcription factor termed CCCTC-binding factor (CTCF) (15).

We have now isolated and analyzed human CTCF cDNA clones and show here that the ubiquitously expressed CTCF factor is (i) an exceptionally highly conserved protein displaying 93% identity between avian and human amino acid sequences and (ii) a “multivalent” factor which utilizes different combinations of individual Zn fingers to specifically bind to diverged regulatory DNA sequences within the promoter-proximal regions of chicken, mouse, and human *c-myc* genes.

We also demonstrate that CTCF contains two strong transcriptional repressor domains transferable to the GAL4 DNA binding domain and that it can repress transcription from reporter constructs containing the P2 promoter of the human *c-myc* gene. The P2-proximal CTCF binding sequence maps precisely within the important regulatory region where pausing of polymerase II transcription complexes is regulated, and CTCF is a predominant nuclear factor binding to this region. We show that elimination of CTCF binding by specific mutation of the P2 promoter-proximal sequence results in increased transcription from stably transfected human *c-myc* reporter constructs. Since chicken CTCF is a repressor for the chicken

\* Corresponding author. Mailing address: Fred Hutchinson Cancer Research Center, Suite C2-023, 1124 Columbia St., Seattle, WA 98104. Phone: (206) 667-4419 or (206) 667-4850. Fax: (206) 667-6523. Electronic mail address: vlobanen@fred.fhrc.org.

† Present address: Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom.

*c-myc* gene promoter (13), function of CTCF appears to be also conserved. Therefore, taken together, our results indicate that CTCF is a major, evolutionarily conserved, negative regulator of vertebrate *c-myc* genes.

## MATERIALS AND METHODS

**Isolation of human CTCF cDNAs.** Two primers corresponding to DNA sequences at amino acids (aa) 1 to 6 and 7 to 13 of the chicken CTCF amino-terminal peptide 1 (15) and three primers corresponding to aa 266 to 271, 276 to 282, and 283 to 288 of the first chicken CTCF Zn finger were used in six combinations to PCR amplify a fragment(s) of human CTCF cDNA from purified, size-fractionated double-stranded human muscle cDNA (Quick-clone cDNA; Clontech Laboratories, Inc.). One of six reactions produced three discrete DNA bands of about 600, 800, and 1,100 bp. DNA from these bands was isolated and ligated into the TA cloning vector (Invitrogen, San Diego, Calif.). The insert sequences of 36 independent plasmids were determined by automated sequencing with a *Taq* DyeDeoxy Terminator Cycle sequencing kit (Applied Biosystems, Inc.) and run through the FASTA DNA sequence homology search using the Wisconsin Genetics Computer Group package. Four inserts were found to have about 82% homology with the chicken CTCF cDNA sequence. One plasmid, p800-3, containing the human CTCF cDNA fragment was used to screen a cDNA library in Uni-ZAP XR vector constructed from poly(A)<sup>+</sup> RNA isolated from early-passage human myeloid cell line HL-60 (3) with a ZAP-cDNA synthesis kit (Stratagene, La Jolla, Calif.). Fourteen positive clones were helper-excised from lambda phage into the Bluescript plasmid. The seven longest clones had identical sequences at the ends of about 4.5-kbp inserts. The three longest clones, p7.1, p9.1, and p10.2, were sequenced on both strands with identical consecutive sets of primers.

**In vitro transcription-translation and nuclear extracts.** Full-length human CTCF and the DNA binding domain of CTCF were synthesized from the p7.1 CTCF cDNA (see Fig. 2A) subcloned into the Bluescript vector with the T3 promoter in the sense orientation and from the pCITE/CTCF1 template containing the 11-Zn-finger domain of CTCF under control of the T7 promoter-CITE leader (15), respectively, by using the TnT reticulocyte lysate coupled in vitro transcription-translation system (Promega Co., Madison, Wis.) as described in the manufacturer's manual. Twelve plasmids for the TnT in vitro synthesis of truncated zinc finger forms of the CTCF DNA binding domain were constructed by cloning in frame into pCITE-4a(+) (Novagen, Madison, Wis.) the human CTCF cDNA fragments PCR amplified with pairs of primers designed to cover certain groups of CTCF Zn fingers as follows: (i) p4a-ZF(1-11), encoding the full-length 11-Zn-finger domain from aa 236 to 622; (ii) amino-terminally truncated forms, i.e., p4a-ZF(2-11), fingers 2 to 11 beginning at the middle of the Zn finger 1 at position 275 and ending at aa 622; p4a-ZF(3-11), fingers 3 to 11, aa 307 to 622; p4a-ZF(4-11), fingers 4 to 11, aa 332 to 622; p4a-ZF(5-11), fingers 5 to 11, aa 367 to 622; and p4a-ZF(6-11), fingers 6 to 11, aa 388 to 622; and (iii) carboxy-terminally truncated forms, i.e., p4a-ZF(1-10), fingers 1 to 10, aa 236 to 549; p4a-ZF(1-9), fingers 1 to 9, aa 236 to 520; p4a-ZF(1-8), fingers 1 to 8, aa 236 to 492; p4a-ZF(1-7), fingers 1 to 7, aa 236 to 463; p4a-ZF(1-6), fingers 1 to 6, aa 236 to 433; and p4a-ZF(1-5), fingers 1 to 5, aa 236 to 404. Translation products synthesized in the presence of [<sup>35</sup>S]Met were visualized on sodium dodecyl sulfate (SDS) gels as described previously (36). Nuclear protein extracts were prepared from isolated cell nuclei by using NUN solution containing 0.3 M NaCl, 1 M urea, and 1% nonionic detergent Nonidet P-40 as described elsewhere (19) and protease and phosphatase inhibitors as described previously for purification of chicken CTCF by sequence-specific chromatography (24).

**EMSA, methylation interference, and missing-contact analyses.** Four consecutive fragments of human *c-myc* DNA (A, positions -56 to +111 relative to +1 at the P2 initiation site; B, positions -225 to -38; C, positions -353 to -166; and D, positions -489 to -329) and four consecutive mouse *c-myc* DNA fragments (α, positions -237 to -87; β, positions -157 to +18; γ, positions -49 to +113; and Δ, positions +85 to +254), one by one covering partially overlapping promoter DNA sequences of interest, were PCR amplified and simultaneously end labelled on either strand by using pairs of 15- to 22-bp-long primers, one of which was 5' end labelled with [<sup>32</sup>P]ATP and T4 polynucleotide kinase. A positive-control DNA fragment bearing CTCF binding sequence footprint V of the chicken *c-myc* gene was amplified from the pFpV plasmid (23, 24). These fragments were gel purified by Elutip-D (Schleicher & Schuell, Keene, N.H.) minicolumn chromatography, and basically equal amounts of each fragment were utilized for both electrophoretic mobility shift assay (EMSA) and methylation interference and missing-contact analyses. Each of four DNA fragments containing CTCF binding sites revealed by the EMSA experiments, 5' end labelled on either the top (coding) strand or the bottom (anticoding) strand, was then either partially methylated at guanines with dimethyl sulfate or modified at pyrimidine bases with hydrozine by the C+T reaction of Maxam and Gilbert (28) and incubated with the in vitro-translated DNA binding domain of CTCF. Free DNA probe was separated from the CTCF-bound probe by preparative EMSA, and DNA was isolated from the gel, cleaved at modified bases with piperidine, and analyzed on a sequencing gel as described in detail previously (15, 24). For EMSA with each DNA probe, 1 to 10 μl of the in vitro translation product or nuclear extract was used for reactions in the presence of cold, double-stranded

competitor DNAs [poly(dI-dC) plus poly(dG)-poly(dC) plus oligonucleotide containing strong binding sites for both Sp1 and Egr1 proteins] in a phosphate-buffered saline (PBS)-based buffer containing standard PBS with 5 mM MgCl<sub>2</sub>, 0.1 mM ZnSO<sub>4</sub>, 1 mM dithiothreitol, 0.1% Nonidet P-40, and 10% glycerol. Reaction mixtures were incubated for 30 min at room temperature and then analyzed on 5% polyacrylamide gels run in 0.5× Tris-borate-EDTA buffer.

**Proteolytic-protection analyses.** Two DNA fragments of identical lengths, harboring either the chicken *c-myc* site V sequence or the human *c-myc* site A sequence, were PCR amplified and simultaneously end labelled as described above. These DNA probes were incubated for 30 min at room temperature with 5 μl of the in vitro-translated 11-Zn-finger DNA binding domain of CTCF (CTCF1) in 20 μl of the PBS-based EMSA buffer; then proteinase K (Merck) was added to final concentrations of 0, 0.01, 0.1, 1, 3, 5, and 10 mg/ml, and the samples were incubated for an additional 20 min and loaded on the EMSA gel.

**Reporter and expression constructs and stable and transient transfections.** Expression constructs pGal-CTCF-N and pGal-CTCF-C, containing N- and C-terminal amino acid sequences of chicken CTCF fused in frame to the GAL4 DNA binding domain, respectively, were prepared by ligating the *SmaI-HindIII* (positions 113 to 830) and *MscI-XbaI* (positions 1908 to 2850) fragments of the chicken CTCF cDNA (15) into the pSG424 vector (35). The luciferase reporter construct p5xUAS/TK-Luc contains five GAL4 binding sites (5xUAS [11]) inserted upstream of the minimal herpes simplex virus (HSV) thymidine kinase (TK) promoter of plasmid pTK-Luc. A total of 10<sup>6</sup> Q76 quail fibroblasts growing at about 50% confluence were transfected by the calcium phosphate method (36) with 2.5 μg of the reporter, 2.5 μg of the expression vector, 0.4 μg of the transfection efficiency control plasmid pSVβ-gal, and 5 μg of the pUC18 DNA as a carrier. Cell extracts were prepared 48 h posttransfection, and activity of the reporter constructs was measured by the Luciferase Assay System as described by the manufacturer (Promega). Expression of the GAL-CTCF fusion proteins was monitored by Western blot (immunoblot) analysis using monoclonal antibodies against GAL4 (1-147) protein (a gift from P. Chambon, Institut de Génétique et de Biologie Moléculaire et Cellulaire). Reporter plasmid pAPwtCAT was constructed by ligating the *ApaI-PvuII* fragment of the human *c-myc* 5' noncoding region from positions -121 to +352 relative to the P2 site into the pBLCAT3 promoterless chloramphenicol acetyltransferase (CAT) construct (26). To obtain the pAPacaCAT plasmid, the ACA mutation of the CTCF binding site shown in Fig. 5A was introduced by two-step PCR amplification (with two mutant primers and two flanking normal primers) and religation. The mutated sequence was verified by sequencing. The pCI/CTCF expression construct was made by ligating the insert from the p7.1 full-length human CTCF cDNA plasmid into the pCI expression vector (Promega) under the control of the cytomegalovirus (CMV) immediate-early enhancer-promoter. Each of the two reporter plasmids was cotransfected with the pCMV/β-gal plasmid and with the pSV2neo plasmid into mouse NIH 3T3 fibroblasts by the lipofection method. The molar ratio of CAT reporter plasmid to β-galactosidase (β-gal)-expressing plasmid and *neo*-expressing plasmid was about 10:1:1. Polyclonal stably transfected cell lines were established by pooling all G418-resistant clones from each transfection. Transient-cotransfection experiments were performed with human embryonic kidney 293 cells by using pHIV-LTR/β-gal for normalizing transfection efficiency, the pCI/CTCF expression vector as an effector, and pAPwtCAT and pAPacaCAT as reporter constructs. A number of cotransfection experiments and EMSAs have been initially carried out to ensure that (i) in 293 cells, the pCI/CTCF expression vector is able to produce CTCF, detectable by Western immunoblotting, at levels proportional to the amount of transfected plasmid; (ii) transient transfection into the 293 cell line reproducibly resulted in sufficient signal from our short CAT constructs containing only the P2-proximal *c-myc* promoter region; and (iii) the human immunodeficiency virus (HIV) long terminal repeat (LTR)-driven β-gal construct employed as an internal control for cell transfection efficiency itself neither binds nor responds to CTCF. In both stable-transfection and transient-cotransfection experiments, CAT activity, normalized to the internal copy number of stably integrated reporter constructs or control β-gal activity, was assayed in cell extracts prepared from equal numbers of transfected cells as described previously (37).

**Western, Northern (RNA), and Southern blots.** The Ab1 affinity-purified polyclonal antibody against the N-terminal epitope of CTCF conserved in vertebrates (15) was used at a 1:100 dilution to probe protein gel blots as described by the manufacturer of the enhanced chemiluminescence (ECL) detection kit (Amersham International plc). Northern (RNA) and Southern (DNA) blots were probed with the human CTCF cDNA probe labelled with [α-<sup>32</sup>P]dCTP by nick translation and washed in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C (36).

**Nucleotide sequence accession numbers.** The GenBank/EMBL database accession numbers for the human and mouse CTCF cDNA sequences are U25435 and U51037, respectively.

## RESULTS

**Isolation of the human CTCF cDNA.** The chicken CTCF gene, which encodes a factor binding to the chicken *c-myc* promoter, was cloned previously (15, 22–25). We based our strategy for cloning human CTCF on evidence that at least two

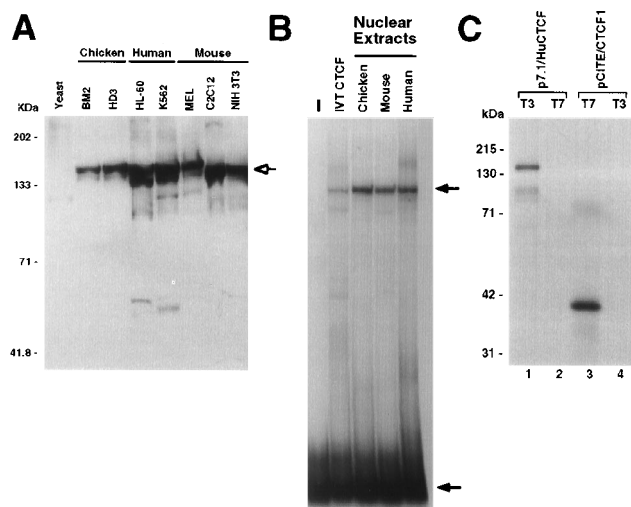


FIG. 1. Characterization of endogenous and cloned CTCF proteins. (A) CTCF proteins are present in cells of different vertebrates. Results of Western immunoblot analysis of *Saccharomyces cerevisiae*, chicken, mouse, and human total cell lysates probed with CTCF antibody Ab1 (15) are shown. BM2 and HD3 are chicken erythroid and myeloid cell lines, respectively; HL-60 and K562 are human myeloid and erythroid cell lines, respectively; and Rauscher MEL, C2C12, and NIH 3T3 are mouse erythroleukemia, myoblast, and fibroblast cell lines, respectively. The major endogenous CTCF protein (arrow), with an apparent molecular mass of about 160 kDa, and positions of the Rainbow molecular mass protein markers (on the left) are indicated. (B) CTCF protein, synthesized from the cDNA template in vitro, and nuclear CTCF proteins of different vertebrates form identical sequence-specific DNA-protein complexes. EMSAs using in vitro-translated human CTCF (IVT CTCF) and nuclear extracts prepared from chicken BM2, mouse NIH 3T3, and human K562 cell lines were performed with <sup>32</sup>P-labelled DNA fragment A from the P2 promoter-proximal region of the human *c-myc* gene (see also Fig. 3 and 4 legends for details). In the EMSA reaction (lane -), a control TnT reticulocyte lysate transcription-translation mixture with no template was used. The positions of the unbound DNA probe and the CTCF-DNA complexes are indicated (arrows). (C) The in vitro-synthesized CTCF polypeptide with a predicted molecular mass of 82 kDa migrates as a 160-kDa protein in SDS-polyacrylamide gels. Full-length CTCF cDNA (lanes 1 and 2) and the 11-Zn-finger domain (lanes 3 and 4) were in vitro transcribed in either the sense (T3 for p7.1/HuCTCF and T7 for pCITE/CTCF1) or the anti-sense orientation and simultaneously translated in the TnT reticulocyte lysate containing [<sup>35</sup>S]methionine. Positions of markers are shown on the left.

domains of CTCF were highly conserved. Polyclonal antibodies against an amino-terminal peptide of the chicken CTCF detected nuclear CTCF protein in cells of a variety of vertebrate species, from frogs to humans (Fig. 1A) (15). In addition, EMSAs detected the presence of CTCF DNA binding activity in nuclear extracts from cells of a variety of vertebrates tested (Fig. 1B and data not shown). These observations indicate conserved domains of CTCF polypeptide in both amino-terminal and DNA binding regions. Therefore, primers corresponding to DNA sequences at these regions were used to amplify and subclone a fragment of human CTCF cDNA that we then used to isolate and analyze several lambda clones of human CTCF cDNA as described in Materials and Methods. Figure 2A shows the complete human CTCF cDNA sequence and its longest open reading frame (ORF). In comparing chicken and human coding sequences, we observed about 20% divergence, primarily at the third DNA base pair of CTCF codons (6). In addition to highly conserved coding regions, the 5' non-coding regions of chicken, mouse, and human CTCF cDNAs and their 1.2-kbp-long 3' untranslated regions have multiple domains of 100% homology (6), indicating putative important conserved sequences that might be involved in control of CTCF mRNA turnover, cellular compartmentalization, or translation efficiency.

**Chicken and human CTCF proteins are strictly conserved.** Comparison of human and chicken CTCF amino acid se-

quences (Fig. 2B) shows that the two proteins are practically identical, with homology extending well outside the completely conserved 11-Zn-finger domain. Analysis of the human CTCF sequence reveals the same structural domains previously noted in the amino acid sequence of chicken CTCF (15), namely, 10 Zn fingers of the C2H2 type and 1 Zn finger of the C2HC class, two highly positive domains flanking the 11-Zn-finger domain, three acidic regions in the carboxy-terminal part of the sequence, and putative serine phosphorylation sites adjacent to a potential nuclear localization signal. Unlike most of the Krüppel-GLI class of factors (12), not all of CTCF's 11 Zn fingers are separated by the highly conserved 7-aa linkers (H-C links) of the form (T/S)GE(K/R)P(F/Y)X (5), suggesting that the CTCF DNA binding domain may be formed by discrete groups of fingers.

Probing a "zoo" DNA blot with labelled human CTCF cDNA fragments displays single-copy CTCF genes in frog, chicken, mouse, and human genomes (6). CTCF expression is not restricted to a particular cell type, since Northern blot analysis of total RNA from a variety of chicken, mouse, and human cell lines and tissues detects comparable levels of expression of ~4-kb CTCF mRNA in all cells tested (6, 14).

It was previously observed that the apparent mobility in SDS gels of CTCF protein purified from chicken cells by sequence-specific chromatography suggested a protein of 130 or 160 kDa (23, 24). Moreover, the major CTCF form detected by Western immunoblotting is also about 160 kDa (Fig. 1A). However, the practically identical ORFs of chicken and human CTCF cDNAs (Fig. 2B) predict a protein of 82 kDa. We therefore screened different chicken, mouse, and human cDNA libraries but could not isolate any CTCF cDNAs containing a longer ORF (6, 13). However, when the human cDNA was transcribed and translated in vitro, we noted a single protein with a mobility of about 160 kDa in SDS gels (Fig. 1C, lane 1). Such anomalous electrophoretic migration of proteins is uncommon but has been observed with other translation products (2, 34), including zinc finger proteins (7). It appears that the amino acid sequence responsible for the aberrant migration of CTCF is located outside the DNA binding region, since the in vitro-translated 11-Zn-finger domain of CTCF migrates in accord with its predicted size of about 40 kDa (Fig. 1C, lane 3). The in vitro-translated CTCF product and endogenous CTCF from nuclear extracts comigrate when loaded on the same gel and assayed by immunoblotting (13) and also generate in EMSA retarded complexes with similar mobilities (Fig. 1B). Therefore, both chicken (15) and human (this report) ~4-kb cDNAs represent full-length copies of the mature polyadenylated CTCF mRNA and encode a protein identical to the endogenous CTCF.

**CTCF protein binds specifically to the promoter-proximal regions of avian, mouse, and human *c-myc* genes.** Using the in vitro-translated DNA binding domain of CTCF for gel shift experiments and methylation interference and missing-contact assays, we have determined CTCF binding sequences in the promoter-proximal region of mouse and human *c-myc* genes. Figure 3A demonstrates a schematic outline of our approach. Four consecutive DNA fragments representing DNA sequences of the promoter-proximal regions of mouse *c-myc* (fragments  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\Delta$ ) and human *c-myc* (fragments A, B, C, and D) were synthesized by PCR amplification with pairs of primers (one of which was <sup>32</sup>P end labelled) in order to obtain DNA probes suitable for both EMSA and methylation interference experiments. As a positive control for CTCF binding, a DNA fragment spanning the footprint V region of the chicken *c-myc* promoter (24) was also included in these exper-



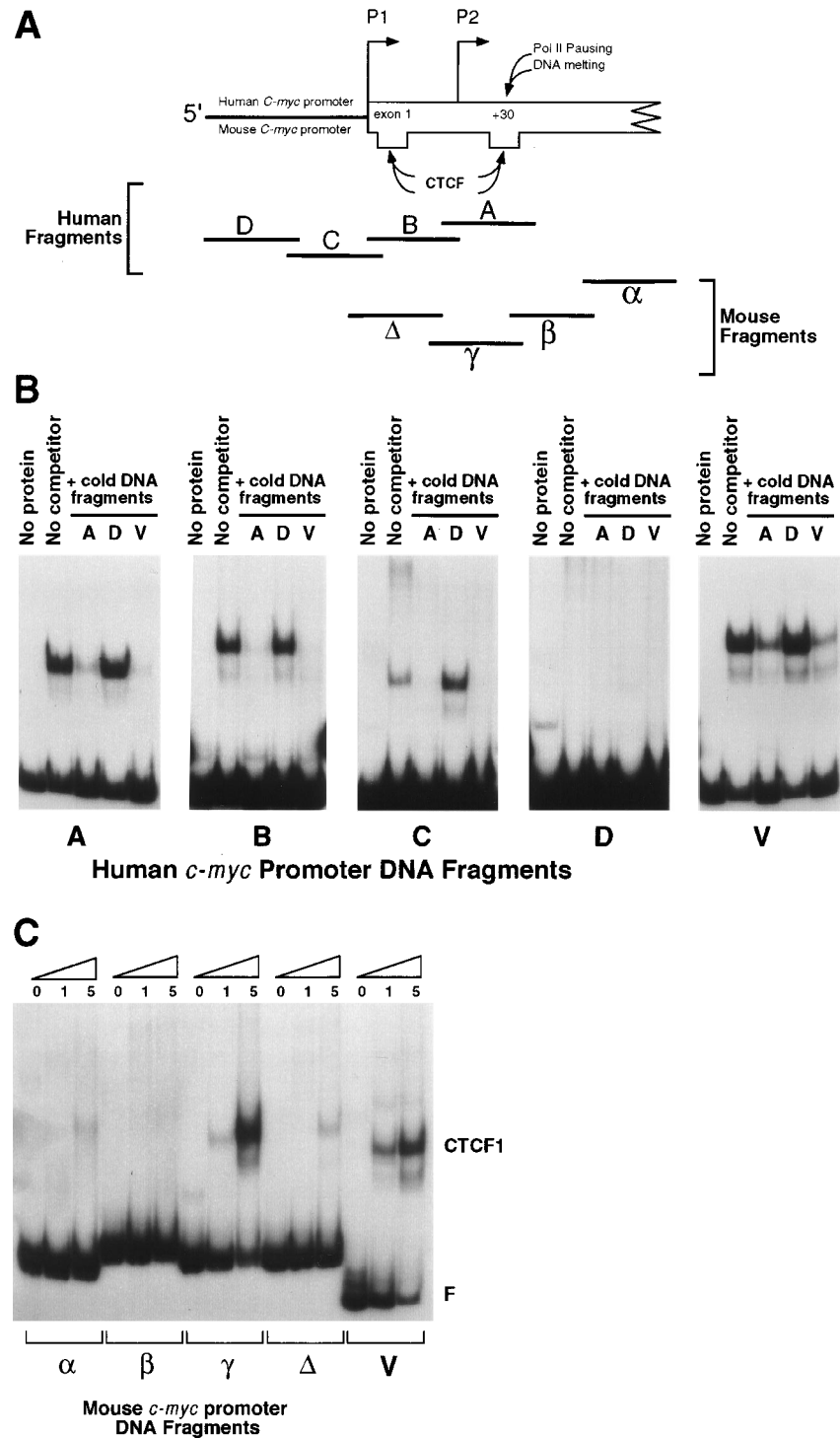


FIG. 3. Sequence-specific CTCF binding to DNA fragments from the promoter-proximal regions of mouse and human *c-myc* genes. (A) Schematic outline of the approach to screen the promoter regions of human and mouse *c-myc* genes for CTCF binding sites. We utilized the indicated DNA fragments from human *c-myc* and mouse *c-myc* P1-P2 promoter regions in EMSA analysis. The +30 region of polymerase II (Pol II) pausing and promoter melting (18) and summary of CTCF binding sites in the human *c-myc* promoter are also shown. (B) EMSA analysis of CTCF binding to human *c-myc* DNA fragments. DNA fragments shown in panel A were tested by EMSA for specific binding to the in vitro-translated 11-Zn-finger domain of CTCF (CTCF1). DNA fragment V harboring the previously defined CTCF binding site of the chicken *c-myc* promoter (24) was also included in these experiments as a positive control. With each DNA probe, EMSA analysis was carried out with either 5  $\mu$ l of control (no template) TnT reticulocyte lysate or 5  $\mu$ l of the lysate containing the CTCF1 DNA binding domain synthesized from the pCITE/CTCF1 template (15). To challenge the specificity of CTCF binding to the human *c-myc* fragments, cross-competition EMSA reactions were also performed by including in the EMSA incubation mixture a 500-fold excess of unlabelled DNA fragments A, D, and V. Note that in these assays, the relatively different positions of shifted bands in EMSA with CTCF1 and different DNA fragments (e.g., with fragments A and V) are due to the different lengths of the DNA probes employed. If fragment V is synthesized to match precisely the length of fragment A, then the positions of CTCF-shifted bands are identical with the two probes (6). (C) EMSA reactions with four mouse *c-myc* promoter DNA fragments with increasing amounts (0 to 5  $\mu$ l) of in vitro-translated CTCF1 were carried out as described above (but excluding cross-competition experiments).

iments. Figure 3B and C show that in addition to control fragment V, three of eight DNA fragments efficiently bind the 11-Zn-finger domain of CTCF protein, namely, DNA fragments A and B from the human *c-myc* gene and fragment  $\gamma$  from the mouse gene. Comparison of the proportions of each DNA probe bound by an equal amount of CTCF indicated that binding to fragments A, B, and  $\gamma$  is comparable to that for chicken fragment V. Binding to fragment C was weaker and not characterized further. Unlabelled DNA fragments A, V, and D were also used as competitors in a cross-competition EMSA experiment (Fig. 3B). Fragment A efficiently competed for CTCF binding to itself and to fragments V and B and fragment V competed for binding to itself and to fragments A and B, whereas fragment D, which did not bind CTCF, did not compete for CTCF binding. Seven other 120- to 220-bp-long GC-rich DNA fragments containing multiple CCTC motifs (from the HIV LTR and from the chicken *c-myc* promoter-proximal regions upstream and downstream of site V) were also tested for CTCF binding and found to be negative (6).

To determine which nucleotides are recognized by CTCF in human and mouse fragments A, B, and  $\gamma$  and to compare them with the recognition sequence in chicken fragment V, we carried out missing-contact analysis (for C plus T bases) and methylation interference (for G bases) assays for both strands of each DNA fragment. DNA bases which on removal or modification reduced binding of CTCF resulted in sequencing gel bands of decreased intensity in lanes of CTCF-bound DNA (Fig. 4, lanes B) compared with the free-DNA lanes (lanes F). Inspection of bases required for CTCF binding to four DNA sequences (Fig. 4) reveals the following. (i) CTCF binds to a DNA sequence from positions +5 to +45 immediately downstream of the P2 initiation site of both human (fragment A) and mouse (fragment  $\gamma$ ) *c-myc* promoters. (ii) This P2-proximal CTCF binding sequence is well conserved in the two mammalian *c-myc* genes; moreover, most of the CTCF-contacting nucleotides within the human and the mouse sites are identical. (iii) CTCF also binds to a different GC-rich sequence (in fragment B) immediately downstream of the P1 initiation site of the human *c-myc* promoter. (iv) The P2-proximal CTCF-binding sequence shared by the human and mouse *c-myc* genes (Fig. 4A and C) and the P1-proximal CTCF binding sequence of the human gene (Fig. 4B) are significantly different from one another and from CTCF binding sequence V in the chicken *c-myc* gene (Fig. 4D).

**Different combinations of CTCF Zn fingers bind to divergent sequences in the chicken and human *c-myc* promoters.** As noted above, the amino acid sequence of the CTCF 11-Zn-finger DNA binding domain is 100% conserved between chicken and human CTCF proteins (Fig. 2B), yet visual inspection of the nucleotide contact points of CTCF in the human (fragment A; Fig. 4A) and chicken (fragment V; Fig. 4D) *c-myc* promoters indicates that these CTCF target sequences are clearly divergent. How do the identical CTCF 11-Zn-finger DNA binding domains contact clearly divergent DNA regulatory sequences? A pairwise comparison of CTCF contact points in the chicken and human fragments (Fig. 5A) indicates that CTCF contacts bases within a GC-rich core common to the chicken and human fragments (Fig. 5A, subregion A2). However, in human fragment A, CTCF also contacts sequences at least 12 nucleotides upstream of this GC-rich core (Fig. 5A, subregion A1), and such contact points are absent in chicken fragment V. To determine whether these more upstream sequences are critical for CTCF binding to human fragment A, we selectively mutated three nucleotides within this region by changing TGT to ACA, as noted in Fig. 5A. EMSA with the in vitro-translated DNA binding domain shows that this ACA

mutation knocks out CTCF binding (Fig. 5B). Therefore, the contact bases critical for recognition by CTCF are clearly different in human fragment A and chicken fragment V.

Since human fragment A harbors more CTCF binding nucleotides than chicken fragment V (Fig. 5A), one would predict that more CTCF Zn fingers may be involved in binding to fragment A than to fragment V. To confirm this hypothesis, we first performed proteolytic-protection assays with CTCF target fragment complexes. Specific regions of DNA-binding proteins that make direct contact with the corresponding DNA target fragments are selectively protected from proteolytic degradation (1, 38). We treated the in vitro-synthesized 11-zinc-finger CTCF domain prebound to either DNA fragment A or DNA fragment V with increasing amounts of proteinase K and analyzed the resulting DNA-protein complexes by gel shift assays. Figure 6 shows that the proteinase-resistant complexes formed with both DNA fragments migrated faster than the complexes formed with the untreated, full-length 11-Zn-finger domain. Moreover, proteinase-treated complexes formed with fragment A (Fig. 6B) migrated more slowly than complexes formed with fragment V (Fig. 6A). Since the two DNA fragments were exactly the same length, this result indicates that not all 11 fingers are absolutely required for binding to both fragments and that the site A DNA sequence protects a significantly larger part of the 11-zinc-finger domain than the site V sequence does. The difference in relative mobility was about 20%, suggesting that at least two fingers more are involved in binding to human site A than to chicken site V.

To directly determine which CTCF Zn fingers might be involved in binding to human fragment A versus chicken fragment V, we utilized these fragments as probes in gel shift assays together with serially truncated in vitro-translated CTCF products. As detailed in Materials and Methods, we engineered five amino-terminally (Fig. 7D) and six carboxy-terminally (Fig. 7A) truncated in vitro-translated products of the 11-Zn-finger domain of CTCF. Gel shift assays of these different forms of the CTCF DNA binding domain using the two DNA fragments containing either chicken *c-myc* CTCF binding site V (Fig. 7B and E) or human *c-myc* CTCF binding site A (Fig. 7C and F) demonstrate that N-terminal fingers 1 and 2 are dispensable for binding to site A (Fig. 7F) but that finger 2 is required for binding to the site V sequence (Fig. 7E). On the other hand, C-terminal fingers, including finger 11, are absolutely required for binding to the P2-proximal site A of human *c-myc* (Fig. 7C), but fingers 11 to 8 are dispensable for binding to site V of chicken *c-myc* (Fig. 7B).

Taken together, these data indicate that the group of six zinc fingers, fingers 2 to 7, is sufficient for CTCF binding to chicken *c-myc* site V, while another group of nine fingers, from 3 to 11, mediates CTCF binding to human *c-myc* site A. Because of its ability to recognize and bind to different DNA sequences by employing different groups of Zn fingers, we propose to call CTCF a multivalent factor.

**CTCF contains two transcriptional repressor domains and negatively regulates the human *c-myc* P2 promoter.** Taken together, the strict evolutionary conservation of CTCF and its unusual ability to bind specifically to a number of diverged DNA sequences in the promoter-proximal regions of human, mouse, and chicken *c-myc* genes suggest that CTCF plays an important role in regulation of *c-myc* genes in vertebrate species. To determine whether CTCF might be a positive or negative transcriptional regulator, both amino- and carboxy-terminal CTCF protein domains flanking the 11-Zn-finger region were individually fused to the GAL4 DNA binding domain to produce the pGal-CTCF-N and pGal-CTCF-C expression vectors, respectively, and cotransfected with the reporter pro-

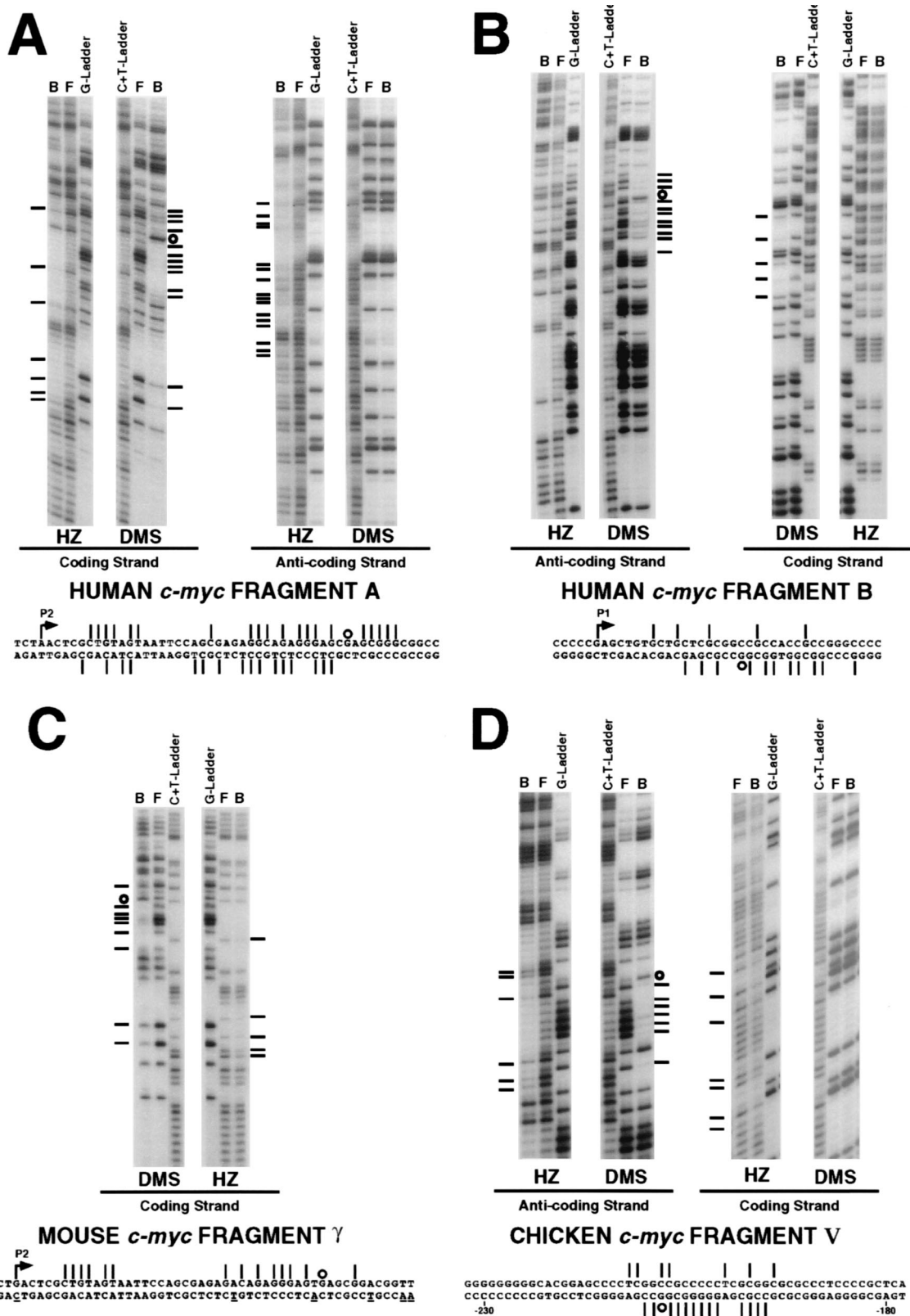


FIG. 4. Identification of variant DNA sequences specifically recognized by CTCF in the promoter-proximal regions of human, mouse, and chicken *c-myc* genes. The results of experiments to determine all DNA bases required for recognition of vertebrate *c-myc* promoters by CTCF are shown. Each of four DNA fragments containing CTCF binding sites revealed by the EMSA experiments (Fig. 3) was subjected to methylation interference analysis (with DNA probes partially methylated at guanines with dimethyl sulfate [DMS]) or missing-nucleoside analysis (with DNA probes modified at pyrimidine bases with hydrozine [HZ]). Lanes F, free DNA probes separated from the CTCF1-bound probes (lanes B). DNA bases which, when missing from the labelled strand or modified, reduce binding of CTCF (bars) and particular methylated G residues preferentially found in CTCF-bound DNA molecules (circles) are indicated. In each panel, the G ladder and the C+T ladder lanes show sequencing reactions run in parallel to facilitate reading of the nucleotide sequence within each CTCF binding site. For each DNA fragment, both coding and noncoding DNA strands (except for mouse fragment  $\gamma$ , which is homologous to human fragment A) were analyzed by both methylation interference and missing-contact analyses, resulting in the summary of guanine and pyrimidine residues required for sequence recognition by CTCF shown below each panel. DNA bases which are different within CTCF binding sites in mouse and human P2-proximal sequences are underlined in panel C.

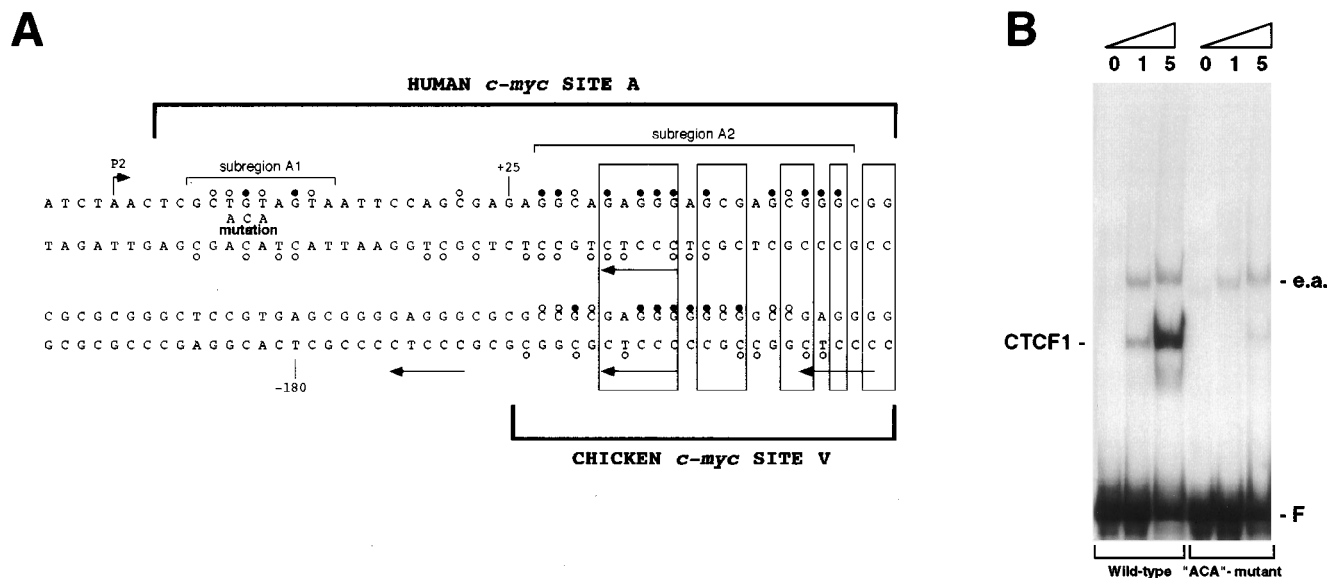


FIG. 5. Selective mutation of nucleotides which distinguish the human *c-myc* P2-proximal CTCF binding sequence from the chicken *c-myc* promoter CTCF binding site V eliminates specific recognition. (A) Comparison of the primary sequence and DNA bases required for CTCF binding to human P2-proximal site A and to chicken promoter site V. CTCF-contacting purine (filled circles) and pyrimidine (open circles) bases determined in two sequences by methylation interference and missing-contact assays (Fig. 4) are indicated. CCGC motifs formerly implicated in CTCF binding (24) (underline arrows) are indicated. The TGT-to-ACA substitution within subregion A1 of human *c-myc* CTCF binding site A is also shown. (B) Two DNA fragments of identical length harboring the P2-proximal DNA sequence of human *c-myc*, with and without the ACA mutation shown in panel A, were synthesized and end labeled by PCR amplification and used for EMSA with 0, 1, and 5  $\mu$ l of the in vitro-translated DNA binding domain of CTCF (CTCF1). Free DNA probes (F) and the presence of an endogenous reticulocyte lysate activity (e.a.) binding outside the CTCF binding sequence are also indicated.

motor containing GAL4 binding sites. The reporter gene consisted of the luciferase gene with either five GAL4 binding sites (11) upstream of a minimal HSV TK promoter (p5xUAS/TK-Luc) or just a TK promoter (pTK-Luc). When cotransfected with the pSG424 vector expressing only the GAL4 DNA binding domain, the two reporter constructs had similar levels of basal transcription (Fig. 8A). Cotransfection with the Gal-CTCF fusion expression vectors, pGal-CTCF-N and pGal-CTCF-C, results in 20- and >100-fold repression of the

p5xUAS/TK-Luc reporter activity, respectively, while activity of the pTK-Luc reporter with no binding sites for the fusion proteins is not inhibited (Fig. 8A). Western immunoblot analysis of total transfected-cell lysates with anti-GAL4 monoclonal antibodies showed production of approximately equal amounts of two fusion proteins (14). Therefore, transcriptional repression was specifically mediated by binding of the Gal-CTCF fusion proteins to the reporter promoter. Figure 8A also indicates that in QT6 fibroblasts, the C-terminal CTCF domain appears to be a stronger repressor than the N-terminal domain. These results indicate that CTCF harbors at least two transcriptional repressor domains.

CTCF is the major protein in nuclear extracts binding to the P2-proximal DNA sequence A of the human *c-myc* gene under our EMSA conditions (Fig. 1B). To determine whether CTCF might act as a transcriptional repressor when bound to the human *c-myc* promoter, we analyzed the functional contributions of both endogenous and exogenous CTCF binding to the P2-proximal site of the human *c-myc* gene. The site is situated within the region between 121 bp upstream (*Apa*I site) and 352 bp downstream (*Pvu*II site) of the P2 promoter (Fig. 8B), which excludes the P1 upstream sequences. This 473-bp sequence around the P2 promoter is sufficient to correctly initiate RNA transcription from stably transfected constructs (21) and is fully responsible for the suppression of *c-myc* transcription upon induced cell differentiation (10). We prepared CAT reporter constructs harboring this P2-proximal human *c-myc* DNA sequence (Fig. 8B), and in one of these constructs we engineered the ACA mutation that specifically eliminates CTCF binding to this P2-proximal sequence (Fig. 5). To generate stable transfectants with these two reporter constructs, we cotransfected them with the pSV/neo plasmid along with the pCMV/ $\beta$ -gal plasmid into mouse NIH 3T3 fibroblasts. G418-resistant clones from each transfection were pooled, and CAT

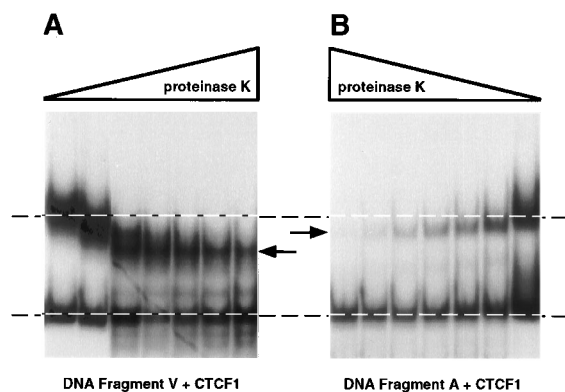


FIG. 6. Two different CTCF binding DNA sequences protect different numbers of zinc fingers from proteolytic attack. Two DNA fragments of identical length, harboring either chicken *c-myc* site V sequence (A) or human *c-myc* site A sequence (B), were preincubated with the in vitro-translated 11-Zn-finger DNA binding domain of CTCF (CTCF1), then treated with increasing amounts of proteinase K as described in Materials and Methods, and analyzed by EMSA on the same gel. The identical mobilities of two free DNA probes and of two untreated DNA-CTCF1 complexes (lower and upper dashed lines, respectively) and the positions of complexes retaining DNA-protein binding during proteinase treatment (arrows) are indicated.



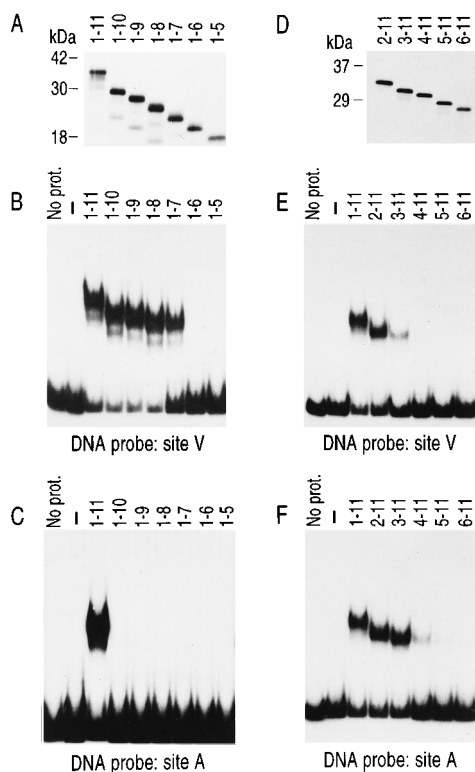


FIG. 7. Different combinations of CTCF zinc fingers are required to bind human and chicken *c-myc* promoters. Each lane is labelled by two numbers indicating the first and the last zinc finger of each truncated form of the 11-finger CTCF DNA binding domain. Full-length 11-Zn-finger polypeptide (panel A, lane 1-11) and six carboxy-terminal deletion (panel A, lanes 1-10 to 1-5) and five amino-terminal deletion (panel D, lanes 2-11 to 6-11) forms of the DNA binding domain were synthesized in vitro as described in Materials and Methods, and 1- $\mu$ l aliquots of each translation product were analyzed by SDS gel electrophoresis. (A and D) Basically equal amounts of each truncated form were synthesized. Positions of the molecular mass markers are indicated on the left. (B and C) EMSA analysis of C-terminally truncated forms binding to chicken *c-myc* fragment V and human *c-myc* fragment A (Fig. 3), respectively. (E and F) Similar analysis for N-terminally truncated forms. Gel shift reactions included equal amounts (5  $\mu$ l) of each in vitro translation product with the control reticulocyte lysate mixture (lanes -) and with no protein (lanes No prot.).

activity, normalized to the reporter copy number or  $\beta$ -gal activity, was assayed in extracts from equal numbers of these cells. We measured CAT activity in cells grown under three different conditions: normal growth, when cells were passaged every third day and did not reach confluence; growth arrest, when confluent cells were kept in serum-deprived medium for 2.5 days; and serum response, when confluent cells were serum starved for 2 days and then transferred to a fresh serum-containing medium for 12 h prior to being harvested. Under all three cell growth conditions, the ACA mutation results in a three- to sixfold increase in reporter gene activity (Fig. 8C). Moreover, the repressing effect of CTCF binding to the P2-proximal site was most profound in growth-arrested cells, i.e., under conditions in which transcription from the *c-myc* promoter has been reported to be inhibited (see reference 27 for a review). Thus, mutational analysis of the P2-proximal CTCF-binding site strongly suggests that CTCF is a repressor of transcription from the major human *c-myc* gene promoter.

To examine the ability of exogenously supplied CTCF to repress the *c-myc* P2 promoter, we performed transient-cotransfection experiments with a CMV promoter-driven CTCF expression vector and the two *c-myc* promoter-CAT reporter

constructs described above (Fig. 8B). These transient-cotransfection experiments are potentially complicated by endogenous CTCF present in target cells which might repress reporter constructs and mask any effect of the exogenous CTCF. Therefore, to assess any effect of exogenous CTCF produced by the transfected expression vector, conditions in which endogenous CTCF was limiting with respect to the transfected target constructs were established (i.e., binding of endogenous CTCF was saturated). Under such conditions, an excess of target constructs free of bound endogenous CTCF should respond to exogenous CTCF produced by the cotransfected expression vector. Figure 8D (bars for 0  $\mu$ g of CTCF) shows that with an input of 1  $\mu$ g of *c-myc* promoter-CAT constructs per transfection, the target constructs appeared to be in excess, since there was little difference in CAT reporter activity between the wild-type and mutated constructs. Under these conditions, introduction of as little as 0.2  $\mu$ g of CTCF expression vector resulted in repression of the wild-type but not the ACA-mutated promoter, indicating that the sequence-specific interaction of exogenously expressed CTCF with the P2-proximal DNA region can specifically repress the promoter. At a higher input of exogenous CTCF (2.0 and 10  $\mu$ g of expression vector), a stronger repression was achieved. However, some of this stronger repressing effect does not require binding of CTCF to the P2-proximal site because the ACA-mutated promoter also becomes repressed (Fig. 8D, two rightmost bars). This finding indicates that at a high input level, CTCF can either bind to low-affinity sites in the mutated promoter or interact with other transcription factors involved in transcription from the P2 promoter of the human *c-myc* gene. This may be quite specific for the P2 *c-myc* promoter, since in cotransfection experiments with several other promoters, including HIV LTR, murine leukemia virus LTR, simian virus 40, and HSV TK, we noted no promoter suppression by even high levels of exogenously expressed CTCF (data not shown).

Taken together, the presence of two strong CTCF repressor domains (Fig. 8A), our observation that mutation of a CTCF binding site within the *c-myc* promoter results in increased reporter gene activity (Fig. 8C), and the suppression of the *c-myc* promoter activity by exogenous CTCF (Fig. 8D) indicate that CTCF is a major physiological repressor of the human *c-myc* P2 promoter.

## DISCUSSION

We have cloned the human *c-myc* promoter-binding protein CTCF and noted >93% identity with the chicken CTCF protein. While more than 90% amino acid identity between avian and mammalian nuclear proteins has been described for some structural DNA- and RNA-binding factors (such as histones and SR proteins), it is not common for sequence-specific DNA-binding transcription factors. Among them, only a few examples of such extreme conservation have been found: *oct-1* (33), *gata-3* (16), *ets-1* (44), and *max* (39). The fact that chicken and human CTCF amino acid sequences did not noticeably diverge (Fig. 2B) during the estimated 200 to 300 million years of evolution (20) is suggestive of a vital CTCF function conserved in all vertebrates. Moreover, no significant amino acid sequence alterations, either inside or outside the CTCF DNA binding domain, were tolerated to maintain this conserved function. Here, we present data showing that at least one such conserved function of CTCF involves binding and repressing *c-myc* gene promoters.

We found that CTCF is able to bind specifically to a number of diverged DNA sequences in the promoter-proximal regions of chicken, mouse, and human *c-myc* genes (Fig. 3 and 4).

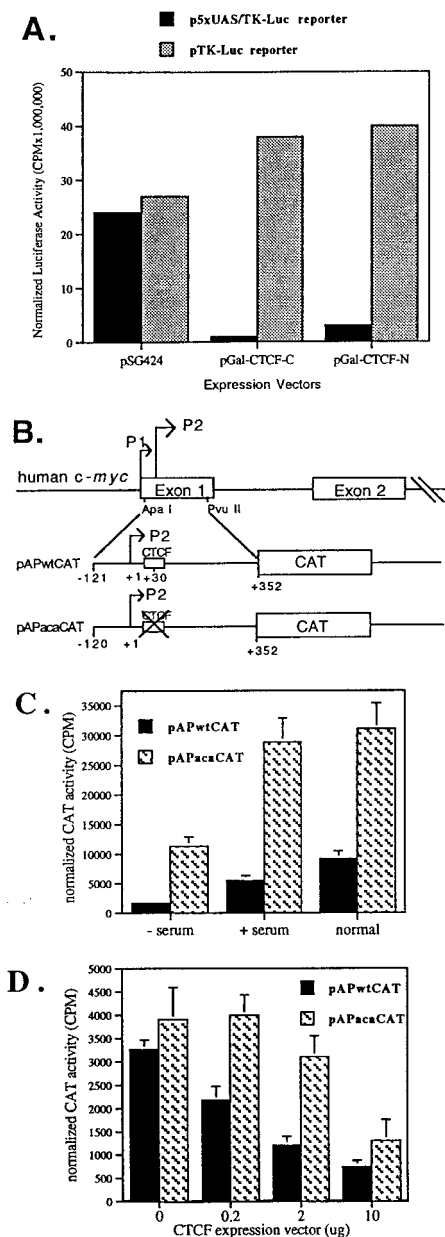


FIG. 8. CTFCF contains two transcriptional repressor domains and negatively regulates the P2 promoter of the human *c-myc* gene. (A) Expression vectors producing the GAL4 (1-147) DNA binding domain alone (pSG424) or fused to the C-terminal (pGal-CTCF-C) or N-terminal (pGal-CTCF-N) CTFCF domain flanking the 11-Zn-finger region were cotransfected into QT6 fibroblasts along with the TK promoter-based reporter plasmids containing five (p5xUAS/TK-Luc) or no (pTK-Luc) GAL4 binding sites, and the activity of the reporter, normalized to the expression from the cotransfected pSV/ $\beta$ -gal construct, was measured. Representative results from one of five independent experiments are shown. The standard deviation calculated from five experiments was <10%. (B) Scheme for the reporter *c-myc*-CAT constructs. See text for details. (C) Transcriptional activities of the wild-type P2 promoter-CAT construct (pAPwtCAT) and of the CTFCF binding site-mutated construct (pAPacaCAT) stably transfected into NIH 3T3 fibroblasts and assayed under three different cell growth conditions as described in the text. Standard error was calculated by measuring normalized CAT activity in four separate plates of each stably transfected mass culture. Results of these experiments were identical when CAT activity was normalized to either  $\beta$ -gal activity from cotransfected pCMV/ $\beta$ -gal plasmid or the copy number of stably integrated reporter constructs estimated by Southern blotting. (D) Transcriptional repression assayed by measuring CAT activity from the pAPwtCAT and pAPacaCAT reporter constructs transiently cotransfected along with increasing amounts of the pGL/CTFCF expression vector into 293 cells. Error bars represent the standard deviations of the means of four transfections for each combination of the reporter and effector constructs.

Although there is absolute (100%) identity between the chicken and human CTFCF proteins within the 11-zinc-finger domain (Fig. 2B), the specific sequences to which CTFCF binds in the respective *c-myc* promoters are clearly divergent (Fig. 4 and 5A). How do identical CTFCF DNA binding domains contact these different DNA regulatory sequences? We have now determined that CTFCF utilizes different combinations of Zn fingers to bind to these diverged chicken and human *c-myc* promoters. First, proteinase protection assays indicated that more CTFCF Zn fingers are involved in binding to the P2-proximal human promoter sequence (site A) than in binding to chicken site V (Fig. 6). Second, a gel shift analysis using serially truncated Zn fingers reveals that fingers 2 to 7 are involved in binding to the chicken *c-myc* promoter (site V), while fingers 3 to 11 appear to be required for binding to the human *c-myc* promoter (site A) (Fig. 7). Therefore, in binding to the chicken *c-myc* promoter, only 6 of the 11 CTFCF Zn fingers are utilized, while human CTFCF utilizes no more than 9 fingers in binding to the diverged human *c-myc* promoter. In so concluding, however, we acknowledge certain limitations of our gel shift DNA binding studies with terminally deleted CTFCF forms possibly arising from an apparent lack of rigorous quantitative analyses of binding affinities of different groups of CTFCF Zn fingers to different DNA sequences and also from possible complex interdependence of individual Zn fingers and other regions of the full-length CTFCF protein. A definitive demonstration of exactly how different individual CTFCF Zn fingers recognize different nucleotide sequences will perhaps require a crystal structure analysis of CTFCF complexes with a number of its DNA binding sites. However, such analysis may be a problem because of the size of the components involved. At present, it is worthwhile to note that the two different base-specific contact patterns in chicken and human *c-myc* promoters determined by the methylation interference experiments (Fig. 4) are consistent with DNA subsite sequences predicted for CTFCF fingers 2 to 7 and for fingers 3 to 11 by the set of rules proposed for the Zn finger DNA recognition code based on the cocrystal structure of several Zn finger domains and their cognate DNA binding sites (43).

There appears to be considerable evolutionary conservation in the patterns of *c-myc* expression, with *myc* expression enhanced in mitogen-stimulated cells and repressed during terminal differentiation in multiple vertebrate species. It is perhaps not surprising that CTFCF, which appears to be an important transcriptional repressor of *c-myc*, also displays marked evolutionary conservation. What is quite surprising, however, is the considerable evolutionary divergence of the CTFCF target sequences in the human and chicken *c-myc* promoters that requires different combinations of CTFCF Zn fingers to bind to these regulatory sequences. Despite this strict evolutionary conservation, there appears to be considerable flexibility inherent in the CTFCF DNA binding domain to enable it to bind to diverged sequences within the *c-myc* promoters of different species.

Two other relatively large Zn finger proteins, Evi-1 and MZF1, which can bind different DNA sequences have been previously reported. MZF1 protein is a 13-Zn-finger protein which appears to harbor two independent DNA binding domains, a nine-Zn-finger domain separated from an additional four-Zn-finger domain by a glycine-proline-rich region (9). The *Evi-1* oncogene protein also contains two domains of Zn fingers, an amino-terminal domain of seven fingers and a carboxy-terminal domain of three fingers (29). In both MZF1 (30) and the *Evi-1* protein (8, 32), each of these domains binds independently to distinct, diverged target DNA sequences. In contrast, CTFCF also binds to diverged sequences but does so

by utilizing different combinations of Zn fingers within a single DNA binding domain. In this respect, we propose to define CTCF as a multivalent factor.

We have accumulated several lines of evidence indicating that CTCF is a negative regulator of the human *c-myc* promoter. Transient-transfection assays utilizing chimeric GAL4 DNA binding domain-CTCF fusion products indicate that CTCF harbors at least two transcriptional repressor domains (Fig. 8A). Moreover, in stable-transfection experiments utilizing *c-myc* promoter-CAT constructs, we have observed that selectively mutating the CTCF binding site within the human P2-proximal promoter region (site A, Fig. 5A) results in increased reporter gene activity (Fig. 8C). Finally, we have observed in transient-cotransfection assays that CTCF specifically represses human *c-myc* promoter activity through both DNA binding-dependent and -independent pathways (Fig. 8D). Taken together, these observations provide strong evidence that CTCF is a negative regulator of *c-myc* transcription.

Several previous observations indicate that the P2-proximal region to which CTCF binds and which it negatively regulates appears to be critical for *c-myc* transcriptional regulation: (i) the level of mRNAs initiated at the P2 promoter is usually more than 80% of steady-state *c-myc* RNA levels in normal cells (40); (ii) activity of transcription from the P2 promoter is regulated by the rate of pausing and release of polymerase II at the sequence immediately downstream of P2 (4, 41, 42); (iii) pausing of polymerase II and promoter melting demonstrated by *in vivo* footprinting analysis occur at around position +30 (18), at a sequence that maps precisely within the CTCF binding site (Fig. 4A); and (iv) transcription of P2-initiated RNA is blocked at the same site when *c-myc* is down-regulated in cells induced to differentiate (17). Note that EMSA analysis of both mouse and human nuclear extracts with human fragment A containing these P2-proximal sequences indicates that CTCF is the predominant nuclear protein interacting with the DNA region downstream of the P2 initiation site (Fig. 1B). Thus, CTCF binding to the P2-proximal site likely plays a critical role in the complex regulatory events mediated through this site. Since we did not attempt to distinguish the effect of CTCF on initiation versus elongation, the precise molecular mechanism of this repression remains to be determined.

Taken together, the marked evolutionary conservation of CTCF repressor domains and the ability to bind promoters of avian and mammalian *c-myc* genes suggest that the *c-myc*-repressing function of CTCF may be conserved also in vertebrates. Although previous mutational analysis of the CTCF binding site in the chicken *c-myc* promoter indicated that the Nsi mutation designed to specifically knock out CTCF binding results in a decrease of transcription (15), we have recently found that besides CTCF, this mutation also eliminates an overlapping strong binding site for the Egr1 family of transcription activators (6). We have also found that CTCF and two Sp1-like and Egr1 proteins bind to the site V sequence of the chicken *c-myc* promoter in a pairwise mutually exclusive fashion and demonstrated by cotransfection experiments that chicken CTCF represses the chicken *c-myc* promoter (13). This suggests that chicken CTCF might actively repress the chicken *c-myc* promoter by a combination of two mechanisms, by displacing positive Sp1 and Egr1 family factors from the site V sequence and bringing its own transcription repressor domains to the promoter.

How can CTCF be a negative regulator of endogenous *c-myc* gene transcription if it is ubiquitously expressed in different cells, including proliferating cells with active *c-myc* expression? Perhaps there is a specific posttranslational modification of CTCF that regulates this repressor activity. Indeed, we have

recently documented (14) that CTCF is phosphorylated *in vivo* and that the negative effect of CTCF on transcription strongly depends on the site-specific reversible phosphorylation of its C-terminal *trans*-repressor domain. Moreover, in comparison with the wild-type CTCF protein, transient expression of the constitutively hypophosphorylated form of CTCF resulted in much stronger repression of target promoters (14). An additional reason to believe that the repression of *c-myc* genes by CTCF is regulated at the level of specific posttranslational modifications comes from our recent observation that in rapidly growing erythroid precursor HD3 cells CTCF is highly phosphorylated. However, upon induction of terminal differentiation of the cells, both *c-myc* expression and CTCF phosphorylation are extinguished (14).

In conclusion, we again note the 100% evolutionary conservation within the entire CTCF 11-Zn-finger domain, including those fingers that are not directly involved in binding to the *c-myc* promoters. This suggests that these fingers are involved in another conserved biological function(s), which might include zinc finger protein-protein or protein-RNA interactions or interactions with regulatory elements of other specific target genes. It is therefore likely that the strict conservation of CTCF is driven by another preserved function(s) beyond binding to *c-myc* promoters. We are now attempting to identify such a function(s).

#### ACKNOWLEDGMENTS

We are grateful to Paul Goodwin and Tim Knight for assistance with image analysis and help in preparing figures; to Michael Parker for providing help with DNA sequence assembly and analysis; to Mary Kay Dolejsi for automated DNA sequencing and oligonucleotide synthesis; and to LeMoyné Mueller, Gilbert Loring, and Sandra Jo Thomas for technical assistance. We thank V. KewalRamani for the 293 cell line, M. Linal for the QT6 cell line, Anton Krumm and Mary Peretz for mouse and human *c-myc* gene plasmids, and P. Chambon for anti-GAL4 antibodies. Michael Emerman, Philippe Soriano, Mark Groudine, and Stephen Tapscott are thanked for discussions and review of the manuscript.

Cena Myers was supported by a 1994 summer graduate student fellowship. This work was funded by NIH/NCI grants RO1 CA20068 to P. E. Neiman and RO1 CA55397 to S. J. Collins, by the NIH RO3 TW00057 Fogarty Award and by American Cancer Society grant DB54 to P. Neiman and V. Lobanenko, by a Human Frontier Science Program (HFSP) Long-Term Fellowship to E. M. Klenova, by Cancer Research Campaign grants to G. H. Goodwin, and by Pilot Study Seattle Breast Cancer Research Program grant NCI P20 CA66186-01 to V. Lobanenko.

#### REFERENCES

1. Bogenhagen, D. F. 1993. Proteolytic footprinting of transcription factor TFIIIA reveals different tightly binding sites for 5S RNA and 5S DNA. *Mol. Cell. Biol.* **13**:5149–5158.
2. Casaregola, S., A. Jacq, D. Laoudj, G. McGurk, S. Margaron, M. Tempete, V. Norris, and I. B. Holland. 1992. Cloning and analysis of the entire *Escherichia coli* *ams* gene: *ams* is identical to *hmp1* and encodes a 114 kDa protein that migrates as a 180 kDa protein. *J. Mol. Biol.* **228**:30–40.
3. Collins, S. J., R. C. Gallo, and R. E. Gallagher. 1977. Continuous growth and differentiation of human myeloid leukemia cells in suspension culture. *Nature (London)* **270**:347–349.
4. 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.
5. 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.
6. Filippova, G. N., P. E. Neiman, S. J. Collins, and V. V. Lobanenko. Unpublished data.
7. Franklin, A. J., T. L. Jetton, K. D. Shelton, and M. A. Magnuson. 1994. BZP, a novel serum-responsive zinc finger protein that inhibits gene transcription. *Mol. Cell. Biol.* **14**:6773–6788.
8. Funabiki, T., B. L. Kreider, and J. N. Ihle. 1994. The carboxyl domain of zinc fingers of the *Evi-1* myeloid transforming gene binds a consensus sequence of GAAGATGAG. *Oncogene* **9**:1575–1581.

9. Hromas, R., S. J. Collins, D. Hickstein, W. Raskind, L. L. Deaven, P. O'Hara, F. S. Hagen, and K. Kaushansky. 1991. A retinoic acid-responsive human zinc finger gene, MZF1, preferentially expressed in myeloid cells. *J. Biol. Chem.* **266**:14183-14187.
10. Ishida, S., K. Shudo, S. Takada, and K. Koike. 1994. Transcription from the P2 promoter of human protooncogene *myc* is suppressed by retinoic acid through an interaction between the E2F element and its binding proteins. *Cell Growth Differ.* **5**:287-294.
11. Kakidani, H., and M. Ptashne. 1988. Gal4 activates gene expression in mammalian cells. *Cell* **52**:161-167.
12. Kinzler, K. W., J. M. Ruppert, S. H. Bigner, and B. Vogelstein. 1988. The *GLI* gene is a member of the *Kruppel* family of zinc finger proteins. *Nature (London)* **332**:371-374.
13. Klenova, E. M., G. N. Filippova, C. Meyers, S. Fagerlie, P. E. Neiman, G. H. Goodwin, and V. V. Lobanekov. Unpublished data.
14. Klenova, E. M., G. N. Filippova, P. E. Neiman, G. H. Goodwin, and V. V. Lobanekov. Functional regulation of the chicken *c-myc* transcriptional repressor CTCF by phosphorylation. Submitted for publication.
15. Klenova, E. M., R. H. Nicolas, H. F. Paterson, A. F. Carne, C. M. Heath, G. H. Goodwin, P. E. Neiman, and V. V. Lobanekov. 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.
16. Ko, L. J., M. Yamamoto, M. W. Leonard, K. M. George, P. Ting, and J. D. Engel. 1991. Murine and human T-lymphocyte GATA-3 factors mediate transcription through a *cis*-regulatory element within the human T-cell receptor  $\delta$  gene enhancer. *Mol. Cell. Biol.* **11**:2778-2784.
17. Kohlhuber, F., L. J. Strobl, and D. Eick. 1993. Early down-regulation of *c-myc* in dimethylsulfoxide-induced mouse erythroleukemia (MEL) cells is mediated at the P<sub>1</sub>/P<sub>2</sub> promoters. *Oncogene* **8**:1099-1102.
18. Krumm, A., T. Meulia, M. Brunvand, and M. Groudine. 1992. The block to transcriptional elongation within the *c-myc* gene is determined in the promoter-proximal region. *Genes Dev.* **6**:2201-2213.
19. Lavery, D. J., and U. Schibler. 1993. Circadian transcription of the cholesterol 7 $\alpha$  hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.* **7**:1871-1884.
20. Lewin, B. 1990. *Genes IV*, p. 504-506. Oxford University Press, New York.
21. Lipp, M., R. Schilling, S. Wiest, G. Laux, and G. W. Bornkamm. 1987. Target sequences for *cis*-acting regulation within the dual promoter of the human *c-myc* gene. *Mol. Cell. Biol.* **7**:1393-1400.
22. Lobanekov, V., and G. Goodwin. 1989. CCCTC-binding protein: a new nuclear protein factor which interaction with 5'-flanking sequence of chicken *c-myc* oncogene correlates with repression of the gene. *Proc. Acad. USSR (Moscow)* **309**:741-745.
23. Lobanekov, V. V., V. V. Adler, E. M. Klenova, R. H. Nicolas, and G. H. Goodwin. 1989. CCCTC-binding factor (CTCF): a novel sequence-specific DNA binding protein which interacts with the 5'-flanking sequence of the chicken *c-myc* gene, p. 45-68. *In* T. S. Papas (ed.), *Gene regulation and AIDS: transcriptional activation, retroviruses and pathogens*. Portfolio Publishing Corp., Woodlands, Tex.
24. Lobanekov, 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.
25. Lobanekov, 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.
26. Luckow, B., and G. Schutz. 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res.* **15**:5490.
27. Marcu, K. B., S. A. Bossone, and A. J. Pate. 1992. *Myc* function and regulation. *Annu. Rev. Biochem.* **61**:809-860.
28. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
29. Morishita, K., D. S. Parker, M. L. Mucenski, N. A. Jenkins, N. G. Copeland, and J. N. Ihle. 1988. Retroviral activation of a novel gene encoding a zinc-finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* **54**:831-840.
30. Morris, J. F., R. Hromas, and F. J. Rauscher III. 1994. Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol. Cell. Biol.* **14**:1786-1795.
31. Packham, G., and J. L. Cleveland. 1995. *c-Myc* and apoptosis. *Biochim. Biophys. Acta* **1242**:11-28.
32. Perkins, A. S., R. Fishel, N. A. Jenkins, and N. G. Copeland. 1991. *Evi-1*, a murine zinc finger proto-oncogene, encodes a sequence-specific DNA-binding protein. *Mol. Cell. Biol.* **11**:2665-2674.
33. Petryniak, B., L. M. Staudt, C. E. Postema, W. T. McCormack, and C. B. Thompson. 1990. Characterization of chicken octamer-binding proteins demonstrates that POU domain-containing homeobox transcription factors have been highly conserved during vertebrate evolution. *Proc. Natl. Acad. Sci. USA* **87**:1099-1103.
34. Query, C. C., R. C. Bentley, and J. D. Keene. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**:89-101.
35. Sadowski, I., and M. Ptashne. 1989. A vector for expressing Gal (1-147) fusions in mammalian cells. *Nucleic Acids Res.* **17**:7539.
36. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. Seed, B., and J.-Y. Sheen. 1988. A simple phase-extraction assay for chloramphenicol acetyltransferase activity. *Gene* **67**:271-277.
38. Shuman, J. D., C. R. Vinson, and S. L. McKnight. 1990. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* **249**:771-774.
39. Sollenberger, K., T. Kao, and E. Taparowsky. 1994. Structural analysis of the chicken *max* gene. *Oncogene* **9**:661-664.
40. Spencer, C. A., and M. Groudine. 1991. Control of *c-myc* regulation in normal and neoplastic cells. *Adv. Cancer Res.* **56**:1-48.
41. 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.
42. Strobl, L. J., F. Kohlhuber, J. Mautner, A. Polack, and D. Eick. 1993. Absence of a paused transcription complex from the *c-myc* P<sub>2</sub> promoter of the translocation chromosome in Burkitt's lymphoma cells: implication for the *c-myc* P<sub>1</sub>/P<sub>2</sub> promoter shift. *Oncogene* **8**:1437-1447.
43. Suzuki, M., M. Gerstein, and N. Yagi. 1994. Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res.* **22**:3397-3405.
44. Watson, D. K., M. J. McWilliams, P. Lapis, J. A. Lautenberger, C. W. Schweinfest, and T. S. Papas. 1988. Mammalian *ets-1* and *ets-2* genes encode highly conserved proteins. *Proc. Natl. Acad. Sci. USA* **85**:7862-7866.