Characterization of the Chicken CTCF Genomic Locus, and Initial Study of the Cell Cycle-regulated Promoter of the Gene*

(Received for publication, May 27, 1998, and in revised form, July 16, 1998)

Elena M. Klenova‡§, Sara Fagerlie§§, Galina N. Filippova§, Leo Kretzner¶, Graham H. Goodwin***, Gilbert Loring§§, Paul E. Neiman‡, and Victor V. Lobanenkov‡‡

CTCF is a multifunctional transcription factor encoded by a novel candidate tumor suppressor gene (Filippova, G. N., Lindblom, A., Meinke, L. J., Klenova, E. M., Neiman, P. E., Collins, S. J., Doggett, N. D., and Lobanenkov, V. V. (1998) Genes Chromosomes Cancer 22, 26–36). We characterized genomic organization of the chicken CTCF (chCTCF) gene, and studied the chCTCF promoter. Genomic locus of chCTCF contains a GC-rich untranslated exon separated from seven coding exons by a long intron. The 2-kilobase pair region upstream of the major transcription start site contains a CpG island marked by a “Not-knot” that includes sequence motifs characteristic of a TATA-less promoter of housekeeping genes. When fused upstream of a reporter chloramphenicol acetyltransferase gene, it acts as a strong transcriptional promoter in transient transfection experiments. The minimal 180-base pair chCTCF promoter region that is fully sufficient to confer high level transcriptional activity to the reporter contains high affinity binding element for the transcription factor YY1. This element is strictly conserved in chicken, mouse, and human CTCF genes. Mutations in the core nucleotides of the YY1 element reduce transcriptional activity of the minimal chCTCF promoter, indicating that the conserved YY1-binding sequence is critical for transcriptional regulation of vertebrate CTCF genes. We also noted in the chCTCF promoter several elements previously characterized in cell cycle-regulated genes, including the “cell cycle-dependent element” and “cell cycle gene homology region” motifs shown to be important for S/G2-specific up-regulation of cdk2/Cdc2, cyclin A, and Plk (polo-like kinase) gene promoters. Presence of the cell cycle-dependent element/cell cycle gene homology region element suggested that chCTCF expression may be cell cycle-regulated. We show that both levels of the endogenous chCTCF mRNA, and the activity of the stably transfected chCTCF promoter constructs, increase in S/G2 cells.

CTCF1 is an 11-zinc-finger transcriptional factor with unusual multiple DNA sequence binding specificity (1–3). It is an exceptionally highly conserved protein displaying 93% overall identity and 100% identity in the 11-zinc-finger DNA-binding domain between avian and mammalian amino acid sequences (1). It binds specifically to a number of different target DNA sequences in promoters of chicken, mouse, and human c-myc proto-oncogenes (1, 4, 5). CTCF is ubiquitously expressed (4, 6), and in addition to specific target DNA sequences in vertebrate c-myc genes, a number of other CTCF-target sites have been identified in regulatory regions of several genes including the chicken lysozyme gene transcriptional silencer (2), amyloid precursor gene promoter (3), minimal promoter regions of Pim-1 and Polo-like kinase (PLK) oncogenes, and a silencer element in the upstream non-coding region of the human decay accelerating factor (DAF) gene.

One critical binding site for CTCF is the P2-proximal region of the human c-myc promoter where pausing of polymerase II transcription complexes is regulated in accord with positive and negative expression signals for c-myc (1). CTCF contains transcriptional repressor domains (1), and induction of exogenous CTCF expression in a conditional tetracycline-regulated system results in marked down-regulation of the endogenous c-myc gene transcription and cell growth suppression.

The human CTCF gene maps within one of the smallest regions of overlap for chromosome deletions, which are displayed by a variety of different tumors including breast and prostate cancers (6), and tumor-specific rearrangements of the human CTCF gene have been detected in some primary breast cancer patients (6). Thus, CTCF is potentially an important tumor suppressor gene involved in the pathogenesis of a number of human cancers.

To better understand a mechanism by which CTCF gene expression is regulated, and may contribute to cell proliferation and transformation, we isolated the entire genomic locus and characterized exon-intron organization of the chicken CTCF (chCTCF) gene. We also studied the 5‘ non-coding region of the gene, mapped transcription start sites, and delineated the chCTCF minimal promoter driven by the strictly conserved initiator (Inr)-like element identical to several previously characterized high affinity binding site for the YY1 transcriptional

---

1 The abbreviations used are: CTCF, CTCF-binding factor; PLK, Polo-like kinase; CTCF/chicken CTCF, YY1, Yin Yang 1 transcription factor; CDE, cell cycle-dependent element; CHR, cell cycle genes homology region; UTR, untranslated region; Inr, initiator; CAT, chloramphenicol acetyltransferase; bp, base pair(s); kb, kilobase pair(s).

2 G. N. Filippova, M. Macbeth, Y. J. Hu, and V. V. Lobanenkov, unpublished results.

3 G. N. Filippova, Y. J. Hu, S. Collins, P. E. Neiman, and V. V. Lobanenkov, unpublished results.

* This work was supported by National Institutes of Health NCI RO1 Grants CA68360 and CA71732 (to V. V. L.) and CA20068 (to P. E. N.) and by Cancer Research Campaign grants (to G. H. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** Present address: Dept. of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201.

†† To whom correspondence should be addressed: Fred Hutchinson Cancer Research Center, Suite C2-423, 1100 Fairview Ave. N., Seattle, WA 98109-1024. Tel.: 206-667-4850; Fax: 206-667-6523; E-mail: vlobanen@fred.hccrc.org.

This paper is available on line at http://www.jbc.org

Vol. 273, No. 41, Issue of October 9, pp. 26571–26579, 1998
Printed in U.S.A.
factor. A computer-assisted inspection of the chCTCF promoter
DNA sequence revealed the “cell cycle-dependent element” (CDE) and “cell cycle gene homology region” (CHR) motifs
shared by promoters of several cell cycle-regulated genes and
found to be important for specific up-regulation at late S and G2
stages of the cell cycle (7–10), suggesting that chCTCF expression
may be cell cycle-regulated. We show here that both abun-
dance of the endogenous chCTCF mRNA, and the activity of the
stably transfected reporter constructs containing chCTCF promoter,
are increased in S/G2 cells.

EXPERIMENTAL PROCEDURES

Isolation of Chicken CTCF Genomic Clones—A chicken genomic DNA
library, constructed by cloning in Lambda DASH™ II of genomic DNA fragments obtained by Sau3A partial digest, was obtained from A.
Begue and V. Laudet (11). The library was screened with the probes either harboring the most 5’-distal 60-bp Norl-Norl fragment, or con-
taining the open reading frame and 3’-untranslated sequence (3’-UTR)
of the chCTCF cDNA (4). The probes were 32P-labeled according to the
protocol provided by the manufacturer of the random priming kit
(Stratagene). Four overlapping positive λ clones, L2, NN2, L18, and L11
(Fig. 1), were isolated by separate screening with the two probes. Phage
DNA was purified as described by Sambrook (12) and totally sequenced by the M13 subclone method.

Restriction Mapping, Subcloning, and Sequencing Genomic Clones—
Phage DNAs were cut with XbaI, EcoRI, KpnI, BamHI, XhoI, and BglII;
separated on 0.8% agarose gel; transferred onto the Hybond nylon
membrane (Amersham Pharmacia Biotech); and analyzed by consecu-
tive restriction mapping of the DNA fragments corresponding to the 5’-un-
translated, coding, and 3’-untranslated regions of the chCTCF cDNA (4).

The XbaI-produced DNA fragments from all positive phase clones
were subcloned into Bluescript II SK+, and one subclone of each was
amplified DNA products spanning introns. Partial sequencing of the
amplified DNA products using universal M13 sequencing primers was
performed to ensure reliability of results by better assessing experimen-
tal deviations, and 48 h after transfection the CAT activity normalized to the β-galactosidase activity was determined by a
phase-extraction CAT assay as described earlier (20).

To obtain stably transfected NIH3T3 cells, the CAT reporters and the
pEQ222 construct were co-transfected with the pSV2Neo as described above, and selection of G418-resistant clones and establishing of poly-
clonal mass cultures performed as described previously (1).

Analysis of the chCTCF mRNA Levels at Different Stages of the Cell Cycle—Fluorescence-activated cell sorting analysis of the DNA content in
chick myeloid BM2 cells (14) or in a transformed T-cell line MSB-1
(21), which were size-fractionated by the elutriation technique, and
Northern blot analysis of the relative abundance of the chCTCF mRNA
and the histone H2B mRNA relative to the actin mRNA level in cell
fractions with different DNA content, were performed as described by
Thompson et al. (22). RNA blots were consecutively probed with the
p900 chCTCF cDNA insert (4); with the chicken H2b histone cDNA
from the pH2Bdelta4 plasmid (23) to mark the S phase content; and
with the chicken β-actin cDNA probe from the pA2 plasmid (24) as an
internal control to normalize levels of chCTCF and H2b mRNAs.

RESULTS

Genomic Organization of the chCTCF Gene—Four overlap-
pling λ clones containing the entire chCTCF genomic locus were iso-
lated from a screen of ~1010 recombinants by hybridization to the
two regions of the chCTCF cDNA as described under “Experimen-
tal Procedures.” Southern blot hybridization mapping with probes derived from different regions of the chCTCF cDNA (data not shown) demonstrated that two clones (L18 and L11; Fig. 1C) contain all coding exons and the 3’-flanking
region, while two other clones harbor the 5’-non-coding region of the
gene (L2 and NN2; Fig. 1C). Four clones together repre-
sent approximately 35 kb of the chCTCF genomic locus. Re-
striction fragment mapping of each phage and various sub-
cloned, combined with sequencing genomic DNAs in both
directions from the known DNA sequence of the cDNA revealed
junctions at each exon-intron boundary. Finally, measuring of the
intron sizes allowed us to define the complete organization of the
chCTCF gene (Fig. 1A), and to obtain a reasonably detailed restriction map of the locus (Fig. 1B). It contains 8
chCTCF exons. Exon 1, and exons 3 to 7 are relatively small (each less than 500 bp), whereas exons 2 and 8 are relatively
large (797 and 1569 bp, respectively). The long first intron of approximately 15 kb separates coding exons from the first
non-coding GC-rich exon. Exon 2 harbors the translation start
Met residue, and encodes most of chCTCF amino-terminal do-
main upstream of the first zinc-finger. Exon 8 contains a small
carboxyl-terminal portion of the open reading frame (167 bp) and
the entire 3’-UTR of the chCTCF cDNA (1413 bp) with a
polyadenylation signal(s). Table I shows that 11 zinc-fingers of
CFCF are encoded by exons 3 to 6; exon 3 contains fingers 1 and 2,
exon 4 contains fingers 3–5, exon 5 contains finger 6 and a
part of the finger 7, and exon 6 contains the rest of finger 7 and
fingers 8–11. Nuclear localization signal, CKII target phospho-
ylation sites, proline-rich sequence with the P

The reporter constructs were transfected with the Lipofectin reagent
(Life Technologies, Inc.) according to protocol A of Felgner and Holm
(18) into QT6 quail fibroblast cells or into human embryonic kidney 293
cells grown to 30–50% confluence on 9-cm Petri dishes using 1 μg of
CAT reporter plasmids and 1 μg of the internal transfection efficiency
control plasmid expressing β-galactosidase from the human immuno-
deficiency virus type-1 long terminal repeat promoter fused to the
lacZ gene in the pEQU22 construct (19). Four identical transfections for
each reporter were performed to ensure reliability of results by better as-

Transcription Initiation Sites—To map the (chCTCF gene
transcription site(s), we employed the primer extension assays.
Total RNA prepared from chicken cell lines BM2 and HD3
was hybridized at different annealing temperatures with the
excess of 5’-labeled primer-1 (Fig. 3), and incubated with avian
myeloblastosis reverse transcriptase. Extension products were
analyzed on sequencing gels (Fig. 2, lanes 1–6). To determine
Fig. 1. Genomic structure of the chCTCF gene. A, diagram of the exon-intron map of the chCTCF gene. Black and open boxes represent protein coding sequences and untranslated sequences, respectively. An arrow shows transcription initiation and direction. B, in-scale partial restriction map of the chCTCF genomic locus. A portion of the first intron is shown below the main line in the same scale. Restriction sites shown are: B, BamHI; A, Apal; S, Smal; Bg, BglII; K, Kpnl; X, XbaI; E, EcoRI; X, XhoI; Nh, NheI; N, NotI. C, alignment of four overlapping λ clones, isolated from the chicken genomic library.

Table I

Exon-intron structure and splice sites of the chicken CTCF gene

Amino acid positions are numbered according to Klenova et al. (4). Nucleotides shown underlined match to the consensus sequences yyyyyyyyyyyncaG (where // indicates the exon-intron boundary, and y is T or C), for splice acceptor site, and (A/C)AG//gt(a/g)agt for splice donor site consensus (25). ZF, zinc-finger.

| Exon no. | CTCF domain | Exon size | 5’-splice donor | Intron size | 3’-splice acceptor |
|----------|-------------|-----------|----------------|-------------|-------------------|
| 1        | 5’-UTR      | bp        | kb             |            | ttttctccagCAGGAAAGTGAAAATG |
|          |             | 179       | 0.652          |            | M                 |
| 2        | N terminus  | 797       | 0.838          | 3.992      | ccacatcagGTACTCGCCCTCAC |
|          |             |           |                |            | V                  |
| 3        | ZF1 + ZF2   | 177       | 0.798          | 2.526      | ttcacagt-AGAGAGACCATAT |
|          |             |           |                |            | E                  |
| 4        | ZFs 3–5     | 259       | 0.954          | 1.954      | ttgctcctagGTGTCATTTCGGA |
|          |             |           |                |            | V                  |
| 5        | ZF6 and half-ZF7 | 152       | 1.138          | 1.138      | cccctccagCTCGACAAATCATT |
|          |             |           |                |            | A                  |
| 6        | Half-ZF7, ZFs 8–11, NLS, CKII sites | 480       | 1.252          | 1.256      | ttcctctagAGAGAAATGCTGAA |
|          |             |           |                |            | E                  |
| 7        | Pro-rich, AT-hook | 168       | 1.137          | 1.137      | cccctccagCTCGACAAATCATT |
|          |             |           |                |            | A                  |
| 8        | Acidic, Stop codon, 3’-UTR | 1,569     | 1.252          | 1.256      | ATGGACCCGCTGA... |
|          |             |           |                |            | M                  |

accurately nucleotides at the extension product ends, the same 5’-labeled primer served for four extension and dideoxy chain termination reactions with a genomic DNA subclone as a template to produce cognate sequence ladders (Fig. 2, lanes A, T, G, and C). Fig. 2 shows that major specific extension products generated by reverse transcriptase with RNA from two different cell lineages are presented by a doublet band corresponding to the CC dinucleotide between NheI and NotI sites. Since avian myeloblastosis virus reverse transcriptase can often erroneously produce cDNA with one “extra” nucleotide at the 5’-end of mRNA template (26), we assigned the major transcription initiation site of the chCTCF to the G residue designated +1 at the coding strand (first G in the NotI site) as depicted in Fig. 3. In BM2 and HD3 cells, initiation at this site generates the bulk of chCTCF mRNAs with 125-nucleotide-long 5’-UTR leader sequence. However, we previously detected (see Ref. 4) longer chCTCF cDNAs with 5’ ends produced by initiation close to the initiator (Inr) element consensus sequence at –60 position (Fig. 3). Therefore, other transcription start site(s) upstream of the major site mapped by the primer-1 extension with RNA from HD3 and BM2 cells (Fig. 2) may also be present, and perhaps be more efficiently employed in different cell types. Indeed, with primers 2 and 3 complementary to mRNA sequences upstream of the CAG (Fig. 3) used for the primer extension assays, other distal transcription start sites could be detected (data not shown). Only the +1 site identified with the primer 1 (Fig. 2) lies close to the transcription initiation site predicted by the TSSG computer algorithm (shown as “predicted TSS” in Fig. 3) that fairly accurately determines potential transcription start positions in a candidate promoter sequence based on the density calculations for transcription factor binding sites (27); therefore, we assigned the major +1 position as shown in Fig. 3.

The chCTCF Gene Promoter Sequence Analyses—We sequenced the 5’-flanking genomic chCTCF DNA region, and inspected this DNA sequence for CpG content and for presence
of potential regulatory elements by the Transcription Element Search Software (TESS) (28) that locates and displays binding sites for transcriptional factors presented in the TransFac database version 3.2. Fig. 3 demonstrates that neither a TATA box nor a CAAT box is present at expected characteristic positions relative to the +1, although farther upstream potential binding sites for CCAAT-binding protein(s) (at −1383) and several TATA elements (for example, at −560) were noted. Lack of the TATA- and CAAT-box elements is a common feature for many GC-rich promoters, which direct transcription via the Inr element that conforms to the consensus sequence Py-Py-A1-N-(T/A)-Py-Py (29), and usually contain multiple sequences with the CCGGCC core for binding the Sp1 transcription factor (reviewed in Ref. 30). To direct specific initiation, functional Inr elements usually encompass a +1 start site within a short range between −5 and +5 (31, 32). However, the Inr-like sequence at the chCTCF major transcription start does not conform well to the consensus, while another Inr element, at −60, shows a perfect match to the consensus, and is positioned at the end of the longest chCTCF cDNAs (4). Moreover, Table II shows that the −60 Inr site is identical to several previously characterized functionally important Inr elements and/or YY1 binding sequences, and that this site is conserved in promoters of vertebrate CTCF genes.

Calculating with the Wisconsin GCG package (version 8, 1994) the G+C content and observed versus expected frequency of CpG methylation-target dinucleotides along the sequence shown in Fig. 3 revealed that the chCTCF 5′-flanking region around transcription start site, especially the region from +110 to −410, has on average of >70% G+C content and values of >0.6 for observed versus expected density of CpG dinucleotides, thus, fulfilling the criteria for the presence of a true CpG island (33). Presence of a CpG island with no TATA or CAAT elements in this region is consistent with the structure of other promoter regions for housekeeping genes (34). Thus, the overall DNA sequence composition of the chCTCF promoter is in good accord with the ubiquitous expression pattern of vertebrate CTCF genes (6), a pattern expected for any gene with a fundamental housekeeping function.

**FIG. 2.** Primer extension mapping of the chCTCF mRNA cap site. [23P]-Labeled primer-1 shown in Fig. 3 was annealed at 45 °C (lanes 1 and 4), 55 °C (lanes 2–4), or 65 °C (lane 6) to 20 μg of total RNA from BM2 (lanes 1 and 2) or HD3 (lanes 4–6) cells, or control yeast tRNA (lane 3), and extended with reverse transcriptase. The extension products were resolved by sequencing gels in parallel with four sequencing reactions as described under “Experimental Procedures.” The double arrow demarcates the major extended products.

**FIG. 3.** DNA sequence of the 5′-flanking region of the chCTCF gene. Restriction sites are underlined, and those sites that were used in preparation of the CAT reporter constructs schematically depicted in Fig. 4 are shown in bold. Sequence motifs with perfect match to previously characterized protein binding sites, and the CDE/CHR motif are boxed. The G residue at the 5′-end of the longest cDNA is double underlined. Transcription start site at +1, mapped by the primer extension, and one predicted by the TSSG algorithm (pTSS), are indicated. Strictly conserved box (SCB) with 100% identity in chicken, mouse, and human CTCF genes is also shown. Presence of a protein binding consensus sequence in the coding or non-coding strand is indicated by 1 or 2, respectively. Note that multiple Sp1, Ap2, and several other binding sites with highly diverged consensus are not included in the figure. See “Results” for other details.
Additionally, the chCTCF proximal promoter region contains multiple Inr-like and Sp1 core sequences. These are too numerous to be shown in the Fig. 3. Many motifs that fit other rather loosely defined consensus sequences for Ap-2, GCF, TCF-1, and numerous “half-sites” for nuclear receptors, were not included in the map shown in Fig. 3. Among the GC-rich sequences proximal to the transcription start site (Fig. 3), there are also several regions somewhat homologous to the GC-rich CTCF-binding sequences detected previously in 5′-flanking regions of vertebrate c-myc genes (1) and in the amyloid protein precursor gene promoter (3). However, only those putative regulatory motifs that, like the Inr/YY1 site, perfectly match previously characterized high affinity binding sites for one or another transcription factor, are shown in Fig. 3. These include consensus binding sequences for NF-κB, GATA family, NF-1, NF-Y, Myb, Ap1, Ap2, SRF, Octa family, PU.1/Ets family, ATF/CREB, E2F, and for Myc/Max heterodimers. Interestingly, two ATF/CREB sites at −245 and −265 positions, E2F site at −210 position, Sp1/144F1 site at −193 position, and ETS box at −160 position of the chCTCF promoter are strictly conserved at identical positions in mouse and human CTCF gene promoters (not shown). Another strictly evolutionarily conserved sequence, called SCB (Fig. 3), is 100% identical within the 5′-non-coding region of chicken, mouse, and human CTCF genes (data not shown). The SCB 20-bp sequence has no obvious similarity to any of known binding sites for transcriptional factors presented in the TransFac data base. The combination of YY1, Ap2, and Myc/Max binding sites within the chCTCF promoter suggests that it may be regulated by Myc (35–37). The presence of multiple binding sites for the Ets family of transcription factors, and of an unusual triplicate GATA-binding site would also predict specific regulation of CTCF gene expression in hematopoietic cells.

Promoter Activity of the chCTCF Gene 5′-Flanking Region in Transfection Experiments—To test whether DNA sequence upstream of the chCTCF transcription start can function as a promoter, and to assess the importance of various transcription factor target sites, we engineered five CAT reporter constructs containing from 180 to 2180 bp of the chCTCF 5′-flanking DNA extending in 5′-direction from the NheI site at the +1 position (Fig. 3) fused to the CAT-encoding region of the pBLCAT3 plasmid (38) as schematically shown in Fig. 4A. The reporter constructs were transfected into QT6 quail fibroblast cells, and 48 h after transfection their normalized CAT activity was compared one to another and to the pSV2CAT construct. The promoterless plasmid pBLCAT3 was employed to determine, and to subtract, the nonspecific background CAT signal.

Fig. 4B shows that 1) the genomic chCTCF sequence between PstI and NheI sites shown in Fig. 2 can efficiently drive transcription of a reporter gene, 2) consecutive truncations of chCTCF promoter from −2 kb to −0.2 kb do not result in any marked change in CAT activity levels, and 3) significant chCTCF promoter activity, comparable to activity of a strong SV40 early promoter of the pSV2CAT construct, is retained with even the shortest pSN/CAT construct that harbors only the 180-bp fragment defined by PstI and NheI sites (Figs. 3 and 4). Essentially similar results were obtained with the five chCTCF promoter-CAT constructs transiently transfected into human embryonic kidney 293 cell line (data not shown).

Activity of the Minimal chCTCF Promoter Depends on the Conserved Inr-like Element—The minimal chCTCF promoter contains at −60 the GCCATTTT-motif identical to one of the most common high affinity binding site for the ubiquitous transcription factor YY1 (38, 45), which is reported to play an important role in activity of a number of promoters (see Refs. 40–45 and Table II). This motif is 100% conserved in promoters of chicken, mouse, and human CTCF genes (Table II), suggesting a critically important contribution of this site in transcriptional regulation of vertebrate CTCF genes. Since major core nucleotides required for function of the identical Inr element and YY1 binding have previously been well characterized (29, 46), we were able to design and introduce specific mutation into...
the −60 Inr-like element GCCATTT-motif of the minimal chCTCF promoter to create a "non-Inr" sequence atggTTT as shown in Fig. 3, and to test whether this mutation would affect activity of the promoter. Fig. 4B shows that this mutation reduces activity of the chCTCF minimal promoter to a level close to the background CAT signal produced by the promoterless pBLCAT3 construct, indicating that the conserved Inr-like element at −60 position is a critical determinant of the transcriptional strength of this promoter in transient transfection assays.

Increase of the chCTCF mRNA Levels and of the Promoter Activity in S/G2 Cells—We searched for additional sequence homologies between the chCTCF promoter and the GenBank Eucaryotic Promoter Database to test whether any of the other previously characterized regulatory elements are shared. One 18-bp-long chCTCF promoter region from −216 to −233 (Fig. 3) displayed only one nucleotide difference to the CCCAGCGG-CGGTTTGGAA motif that is 100% conserved in promoters of mouse and human PLK genes (Table III), and that is shown by mutational analyses to be essential for activation of the PLK promoter at S/G2 phase of the cell cycle (10). Moreover, as shown in Table III, very similar sequence segments, called the "R box" (47) or the CDE combined with the CHR motifs (7–9), are present in promoters of several additional cell-cycle-regulated genes, such as cdc25c, cdc2, and cyclin A (7).

Since the CDE/CHR motifs were found to be critically important for the S/G2-specific transcriptional up-regulation (reviewed in Ref. 8), we wondered whether chCTCF expression may also be cell cycle-regulated in the similar fashion. We have initially tested chCTCF expression during cell cycle by extracting RNA from logarithmically growing cell populations fractionated by elutriation into cell cycle compartments as described previously (22). Fig. 5 shows that chCTCF mRNA concentration normalized to that of the actin message (which is not cell cycle-regulated; Ref. 22), is approximately 5-fold increased in cell fractions enriched in S phase and G2 DNA content relative to cell fractions with G1 DNA content. Similar results were obtained with two chicken cell lines, BM2 (Fig. 5) and MSB-1 (data not shown). Additionally, preliminary results on primary chicken fibroblasts induced to proliferate after growth arrest by serum starvation also support conclusion that chCTCF mRNA is up-regulated with an increase in S/G2 phase of the cell cycle (data not shown).

To test whether S/G2-increased transcriptional activity of the promoter may regulate chCTCF mRNA abundance during the cell cycle, we have stably transfected into NIH3T3 cells two chCTCF promoter-based reporter constructs, the longest one, pPN/CAT, and the shortest one, pSN/CAT (see Figs. 3 and 4 for details), and promoterless pBLCAT3 construct to determine background CAT values. Polyclonal mass cell cultures with each stably transfected reporter were established, and employed to assay for normalized CAT activity during one cycle of synchronous progression from resting state (achieved by serum starvation) to proliferation induced by addition of growth factors (10% fetal serum). Fig. 6 demonstrates that activity of the longest promoter-CAT construct pPN/CAT is increased within the first hour after serum induction, decreases later to the level of uninduced resting cells, and then significantly increases again at a 15–20-h interval after induction when most of cells are expected to be in the S/G2 phase as determined by the correlation with the levels of the H2b histone mRNA expression (data not shown). Compared with the pSN/CAT reporter, the shorter promoter construct pSN/CAT which does not include the CDE/CHR motif or the E box (Fig. 3), shows weaker general activity in stably transfected cells, and less stimulation at 15–20 h after induction (Fig. 6). Therefore, the S/G2 up-regulation of chCTCF mRNA levels appears to correlate with up-regulation of the stably transfected reporter construct containing −2000 bp of the chCTCF gene promoter.

**DISCUSSION**

Our search for factors specifically binding to the 5′-flanking non-coding DNA sequences of the chicken, mouse, and human c-myc genes resulted in identification, purification, and molecular cloning of the evolutionarily conserved 11-zinc-finger transcription factor CTCF (1, 4–6, 48, 49). Besides CTCF, there is no other example of a "universal" factor that binds to the regulatory regions of all vertebrate c-myc oncogenes, e.g. to the avian and human c-myc promoters despite their sequence divergence.

**CTCF gene is ubiquitously expressed (6), and in addition to**

**CTCF-binding sequences in vertebrate c-myc genes (1, 4) sev-
CTCF is related to its potential role as an important tumor suppressor gene involved in the pathogenesis of a number of different human cancers. We recently demonstrated that human CTCF maps to chromosome 16q22.1 segment within one of the smallest regions of overlap for chromosome deletions displayed by a variety of different tumors including breast and prostate cancers, and we have observed genomic rearrangements at the CTCF locus in several breast cancer samples (6). Thus, studies of CTCF gene structure and regulation may provide important insights into fundamental mechanisms regulating cell proliferation in normal and transformed cells.

We isolated overlapping genomic λ clones (Fig. 1C), and mapped genomic region containing chCTCF with several restriction enzymes (Fig. 1B). We also tested whether multiple CTCF-related loci might be present in the genome. Southern blot hybridization of genomic DNA digested with several restriction enzymes with DNA probes containing different chCTCF exons showed no other DNA fragments besides those that correspond to the genomic map shown in Fig. 1B (Ref. 6, and data not shown), indicating that chCTCF gene is a single copy gene with no CTCF-related pseudogenes or close homologues. The same conclusion was reached for mouse and human CTCF genes (6). Consistent with these results, fluorescence in situ hybridization with human metaphase chromosomes showed one single chromosomal locus containing CTCF (6). Therefore, in addition to being exceptionally conserved (1), CTCF also appears to be a non-redundant gene.

We identified the exon-intron structure of the chCTCF gene (Fig. 1A), and determined which parts of the CTCF protein are encoded by each exon (Table I). In many zinc-finger factor genes, individual fingers are either encoded by one separate exon (for example, in the HP.10 gene all 11 zinc-fingers are in one domain within the last 3’-exon (50), or each separate zinc-finger is encoded in a separate exon (for example, four fingers in WTI gene (51), GATA1 and GATA3 genes (52)). In contrast, the chCTCF zinc-fingers are distributed over exons 3 to 6 with the finger 7 being “torn apart” between exons 5 and 6 (Table I). The evolution of this particular organization of the 11-zinc-finger chCTCF DNA-binding domain is unclear, but comparative examination of the chCTCF intron-exon structure with that of mammalian CTCF genes may provide new insights.

The first exon of chCTCF is separated from the exon 2 by a relatively long intron of ~15 kb with multiple NoI sites close to the 3’-1st exon to 5’-1st intron junction (Fig. 1). The 75% GC-rich first exon of the chCTCF gene is not translated. In the genomic context this type of non-coding exons may have various functions. It may contain downstream transcriptional signals as exemplified by the mdr-16 gene (53), or harbor elements enhancing an upstream promoter activity as shown for the O-6-methylguanine-DNA-methyltransferase gene (54), or it may be alternatively spliced to create different types of mRNAs as described for the aromatic l-amino acid decarboxylase gene (55). In the mRNA context, the GC-rich untranslated exons frequently result in 5’-UTR mRNA sequences which are involved in regulation of the mRNA turnover, transport, and cellular compartmentalization, and translational efficiency (reviewed in Ref. 56). To support an essential role for certain non-coding exons, “experiments of nature” revealed that mutations in the untranslated exon can frequently be associated with tumorigenesis, as reported for the BCL-6 oncogene (57). We would predict that the chCTCF first non-coding exon likely has an important function because several regions of 100% identity in the first non-coding exons of the chicken (4), mouse, and human (1) CTCF genes have been maintained without a change throughout an estimated 300 million years of evolution from birds to humans.

To be able to study transcriptional regulation of the chCTCF gene, we determined and analyzed DNA sequence upstream of the first exon (Fig. 3). As shown in Fig. 3, the transcription initiation site predicted within this sequence by the TSSG computer algorithm of the Gene-Finder computer tools for analysis of human and model organisms genome sequences (27) is very close to the +1 initiation site identified experimentally by the primer extension assay (Fig. 2). Mapping the major transcription start site at the +1 position shown in Fig. 3 suggested that the DNA region extending upstream of this site should serve as the chCTCF promoter. The ~2-kb genomic region upstream of the major transcription start site has no TATA box but contains a CpG island with multiple Sp1 binding sites and other sequence motifs characteristic of a promoter of housekeeping genes. Indeed, when fused to the reporter CAT gene, it is found to direct efficient transcription in transiently transfected cells (Fig. 4). Moreover, regulatory elements of the chCTCF promoter that are sufficient in transient transfection experiments to confer high levels of transcriptional activity comparable to the activity of the well characterized strong SV40 early promoter, were delineated within the minimal 180-bp region immediately upstream of the transcription start site (Fig. 4).
The minimal chCTCF promoter region with strong basal activity contains an Inr-like element that is identical to a number of previously characterized high affinity binding sites for the transcription factor YY1 and that is strictly conserved in promoters of chicken, mouse, and human CTCF genes (Table II). Mutations of the core nucleotides of this element which are reported to eliminate YY1 binding (see Ref. 29, and references therein) severely reduced transcriptional activity of the minimal chCTCF promoter in transient transfection experiments (Fig. 4), indicating that the conserved Inr-like YY1-binding element is critical for transcriptional regulation of vertebrate CTCF genes.

Although analyzing specific composition and arrangement of known target sites for transcription factors in the chCTCF promoter (Fig. 3) has provided us with initial clues on possibly since presence of CDE/CHR elements were shown to be essential for transcriptional regulation of vertebrate genes. We also showed that, together with PLK, cdc25C, cdc2, and cyclin A genes, CTCF gene most probably belongs to a particular group of important genes that are transcriptionally activated in S and G2 phases of the cell cycle, and that the chCTCF promoter and promoters of all other genes of the S/G2 up-regulated family share common CDE/CHR-like consensus sequence.

We do not yet know whether the net amount of CTCF protein, its DNA-binding activity, or nuclear accumulation are cell cycle-regulated coordinately with CTCF mRNA. Furthermore, interaction of CTCF with cell cycle regulatory apparatus is likely to be very complex because CTCF interacts with and regulates promoters of genes that are up-regulated at different stages of the cell cycle: c-myc (at G1/S), PLK (at G2/S), and other target promoters that are not known to be cell cycle-regulated. Here, we showed that CTCF itself may also be cell cycle-regulated. Moreover, our preliminary results indicate that chCTCF expression is, in turn, regulated by Myc and by CTCF itself, suggesting the possibility of a regulatory network involved in cell proliferation control based on multiple feedback loops between CTCF and its target genes.

Acknowledgments—We thank Drs. A. Begue and V. Laudet for generously procuring the chicken genomic library. We thank Degan Collins and John Rasko for critical reading of the manuscript and helpful suggestions, and Tim Knight and Tariq Awad for help with preparing figures.

REFERENCES

1. Filippova, G. N., Fagerlie, S., Klenova, E., Myers, C., Dehner, Y., Goodwin, G., Neiman, P., Collins, S., and Lobanenkov, V. V. (1996) Mol. Cell. Biol. 16, 2802–2813.
2. Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Burde, J., Lottspeich, F., Filippova, G. N., Lobanenkov, V., and Renkawitz, R. (1997) Mol. Cell. Biol. 17, 1281–1288.
3. Vostrov, A. A., and Quitschke, W. W. (1997) J. Biol. Chem. 272, 33533–33539.
4. Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., Neiman, P. E., and Lobanenkov, V. V. (1995) Mol. Cell. Biol. 13, 7612–7624.
5. Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, M., Klenova, E. M., Polatskaya, A. V., and Goodwin, G. H. (1990) Oncogene 5, 1743–1753.
6. Filippova, G. N., Lindblom, A., Meinke, L. J., Klenova, E. M., Neiman, P. E., Collins, S. J., Duggett, N. D., and Lobanenkov, V. V. (1998) Genes Chromosomes Cancer 22, 26–36.
7. Zwick, J., Lucibello, F. C., Wolfram, L. A., Gross, C., Truss, M., Engeland, K., Muller, R. (1995) EMBO J. 14, 4514–4522.
8. Zwick, J., and Muller, R. (1997) Trends Genet. 13, 3–6.
9. Liu, N., Lucibello, F. C., Kerner, K., Wolfram, L. A., Zwick, J., and Muller, R. (1997) Nucleic Acids Res. 25, 4915–4920.
10. Uchiumi, T., Longo, D. L., and Ferris, D. K. (1997) J. Biol. Chem. 272, 9166–9174.
11. Amouyel, P., Laudet, V., Martin, P., Li, R., Quatannens, B., Staehelin, D., and Saule S. (1989) J. Virol. 63, 3382–3388.
12. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13. Sanger, F., Nicklen, S., and Coulson, A. R. (1986) Bio/Technology 24, 104–108.
14. Moscovici, C., Zeller, N., and Moscovici, M. G. (1982) In Expression of Differentiated Function in Cancer Cells (Revoltella, R. F., Pontier, G. M., et al.), pp. 345–449, Raven Press, New York.
15. Beug, H., Doedelin, G., Freudenstein, C., and Graf, T. (1982) J. Cell. Physiol. Suppl. 53, 195–207.
16. Chomczynski, P., and Sacchi, N. (1986) Anal. Biochem. 162, 156–159.
17. Luckow, V., and Schutz, G. (1987) Nucleic Acids Res. 15, 3490.
18. Seligman, P. L., Gadek, T. R., Holm, M., Roman, R., Chen, H. W., Wenz, M., Northrop, J. P., Singold, G. M., and Danielsen, M. (1967) Proc. Natl. Acad. Sci. U. S. A. 64, 7413–7417.
19. Biegalke, B. J., and Gehalle, A. P. (1991) Virology 183, 381–385.
20. Seed, B., and Sheen, J.-Y. (1988) EMBO J. 7, 271–277.
21. Nazarian, K., and Witty, R. L. (1975) J. Natl. Cancer Inst. 55, 453–458.
22. Thompson, C. B., Challoner, P. B., Neiman, P. E., and Groudine, M. (1985) Nature 314, 363–366.
23. Challoner, P. B., Moss, S. B., and Groudine, M. (1989) Mol. Cell. Biol. 9, 592–513.
24. Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J., and Kirschner, M. W. (1989) Cell 20, 95–105.
25. Smyth, P., Shapiro, M. B., and Harris, N. L. (1990) Methods Enzymol. 183, 252–278.
26. Gupta, K. C., and Kingsbury, D. W. (1984) Nucleic Acids Res. 12, 3829–3841.
27. Solovyev, V., and Salamov, A. (1997) ISMB 5, 294–302.
28. Schag, J., and Overton, G. C. (1997) TESS: Transcription Element Search
4 G. N. Filippova, S. Fagerlie, D. Dimitrov, I. Chernukhin, Y. J. Hu, E. M. Klenova, J. Ulmer, L. Kretzner, S. Collins, P. E. Neiman, and V. V. Lobanenkov, submitted for publication.
chCTCF Genomic Locus and Promoter

26579

Software on the WWW; Technical Report CBIL-TR-1997-1001-v0.0 of the Computational Biology and Informatics Laboratory, University of Pennsylvania, Philadelphia

29. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994) Mol. Cell. Biol. 14, 116–127
30. Smale, S. T. (1994) in Transcription: Mechanisms and Regulation (Conaway, R. C., and Conaway, J. W., eds) pp. 63–81, Raven Press, New York
31. O'Shea-Greenfield, A., and Smale, S. T. (1992) J. Biol. Chem. 267, 1391–1402
32. Zenzie-Gregory, B., O'Shea-Greenfield, A., and Smale, S. T. (1992) J. Biol. Chem. 267, 2823–2830
33. Larsen, F. G., Gundersen, R., Lopez, R., and Prydz, H. (1992) Genomics 13, 1695–1107
34. Bird, A. P. (1986) Nature 321, 209–213
35. Grandori, C., and Eisenman, R. N. (1997) Trends Biochem. Sci. 22, 177–181
36. Bouchard, C., Staller, P., and Eders, M. (1998) Trends Cell Biol. 8, 202–206
37. Henriksen, M., and Luscher, B. (1996) Adv. Cancer Res. 68, 109–182
38. Yant, S. R., Zhu, W., Millinoff, D., Slightom, J. L., Goodman, M., and Gumucio, D. L. (1995) Nucleic Acids Res. 23, 4353–4362
39. Meier, V. S., and Groner, B. (1994) Mol. Cell. Biol. 14, 128–137
40. Flanagan, J. R., Becker, K. G., Ennist, D. L., Gleason, S. L., Driggers, P. H., Levi, B. Z., Appella, E., and Orato, K. (1992) Mol. Cell. Biol. 12, 38–44
41. Shi, Y., Seto, K., Chang, L. S., and Shenk, T. (1991) Cell 67, 377–388
42. Chan, P. K., Chan, F. Y., Morris, S. W., and Xie, Z. (1997) Nucleic Acids Res. 15, 1225–1232
43. Riggs, K. J., Saleque, S., Wong, K.-K., Merrell, K. T., Lee, J.-S., Shi, Y., and Calame, K. (1993) Mol. Cell. Biol. 13, 7487–7496
44. Ye, J., Young, H. A., Ortaldo, J. R., and Ghosh, P. (1994) Nucleic Acids Res. 25, 5672–5678
45. Hyde-DeRuyscher, R. P., Jennings, E., and Shenk, T. (1995) Nucleic Acids Res. 23, 4457–4465
46. Usheva, A., and Shenk, T. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 13571–13576
47. Sugarman, J. L., Schonthal, A. H., and Glass, C. K. (1995) Mol. Cell. Biol. 15, 3283–3290
48. Lobanenkov, V. V., Adler, V. V., Klenova, E. M., Nicolas, R. H., and Goodwin, G. H. (1986) Eur. J. Biochem. 159, 181–188
49. Lobanenkov, V. V., Adler, V. V., Klenova, E. M., Nicolas, R. H., and Goodwin, G. H. (1989) in Gene Regulation and AIDS: Transcriptional Activation, Retroviruses and Pathogenesis (Papas, T. S., ed) pp. 45–68, Portfolio Publishing Corp., Woodlands, TX
50. Lanfrancone, L., Pengue, G., Pandolfi, P.-P., Salcini, A. E., Giacomacci, A., Longo, L., Doni, E., De Luca, P., La Mantia, G., Pelicci, P.-G., and Lania, L. (1992) Genomics 12, 708–728
51. Gessler, M., Konig, A., and Bruns, G. A. P. (1992) Genomics 12, 807–813
52. Labastie, M.-C., Bories, D., Cabret, C., Greigou, J.-M., Shretien, S., and Romeo, P.-K. (1994) Genomics 21, 1–6
53. Cohen, D., Pickart, R. L., Hsu, S. I., De Pinho, R. A., Carrasco, H., and Horwitz, S. B. (1991) J. Biol. Chem. 266, 2219–2224
54. Harris, L. C., Potter, P. M., Tano, K., Shiota, S., Mitra, S., and Brent, T. P. (1991) Nucleic Acids Res. 16, 6163–6167
55. Ichinose, H., Sumi-Ichinose, C., Ohye, T., Hagino, Y., Fujita, K., and Nagatsu, T. (1992) Biochemistry 31, 11546–11550
56. Curtis, D., Lehmann, R., and Zamo, P. D. (1995) Cell 81, 171–178
57. Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B. H., Knowles, D. M., Offit, K., Chaganti, R. S., and Dalla-Favera, R. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 12520–12524