Regulatory Factor Interactions and Somatic Silencing of the Germ Cell-specific ALF Gene*

MinJung Kim, Dan Li, Yunxia Cui, Konrad Mueller, William C. Chears, and Jeff DeJong1

From the Department of Molecular and Cell Biology, University of Texas at Dallas, Richardson, Texas 75080

Germ cell-specific genes are active in oocytes and spermatocytes but are silent in all other cell types. To understand the basis for this seemingly simple pattern of regulation, we characterized factors that recognize the promoter-proximal region of the germ cell-specific TFIIAα/β-like factor (ALF) gene. Two of the protein-DNA complexes formed with liver extracts (C4 and C5) are due to the zinc finger proteins Sp1 and Sp3, respectively, whereas another complex (C6) is due to the transcription factor RFX1. Two additional complexes (C1 and C3) are due to the multivalent zinc finger protein CTCF, a factor that plays a role in gene silencing and chromatin insulation. An investigation of CTCF binding revealed a recognition site of only 17 bp that overlaps with the Sp1/Sp3 site. This site is predictive of other genomic CTCF sites and can be aligned to create a functional consensus. Studies on the activity of the ALF promoter in somatic 293 cells revealed mutations that result in increased reporter activity. In addition, RNAi-mediated down-regulation of CTCF is associated with activation of the endogenous ALF gene, and both CTCF and Sp3 repress the promoter in transient transfection assays. Overall, the results suggest a role for several factors, including the multivalent zinc finger chromatin insulator protein CTCF, in mediating somatic repression of the ALF gene. Release of such repression, perhaps in conjunction with other members of the CTCF, RFX, and Sp1 families of transcription factors, could be an important aspect of germ cell gene activation.

An intriguing feature of ALF and other germ cell-specific genes is that the sequences sufficient to exert correct control of expression are very short, on the order of 100 base pairs. For instance, a region of the ALF gene between −133 and +21 relative to the transcription start site is able to recapitulate accurate tissue-specific expression and silencing of a linked reporter in transgenic mice (10). The most notable sequences in this region include a TTTCAA element, a GC-rich region with a number of CpG dinucleotides, and an upstream TC-rich region. Although broadly similar features can be observed in several other germ cell-specific promoters (10), individual sequences of germ cell promoters are not precisely homologous, and the identification of factors responsible for their regulation remains an important goal. Such factors are expected to control the two main behaviors of germ cell-specific genes: selective activation during germ cell differentiation and somatic silencing.

Here we have examined DNA-protein complexes that form on the ALF promoter using mouse liver- and testis-derived nuclear extracts. We show that these complexes are due to the zinc finger proteins, Sp1, Sp3, and CTCF, and to the winged helix transcription factor RFX1. Interestingly, sites for Sp1/Sp3 and CTCF overlap, making this a unique subclass of elements that has dual specificity for both types of factors. The identification of these particular proteins is also of interest because they are all members of larger families of factors that are regulated in complex patterns during gametogenesis (11–18). The role of CTCF is especially interesting, since its effect on gene silencing is mediated at the level of chromatin organization (19, 20). Its identification as a germ cell promoter binding factor, as described in this report, suggests a possible role in protecting germ cell genes from inappropriate activation by adjacent enhancers active in somatic tissues. Functional data are presented that support the notion that the ALF promoter is repressed in somatic cells, and we speculate that release of repression might be an important aspect of how such genes ultimately turned on. Overall, the results provide insights into the on-off regulation of germ cell gene promoters.

MATERIALS AND METHODS

Oligonucleotide Probes and Competitors—Mouse ALF promoter DNA was PCR-amplified from mouse genomic DNA with ALF-up (5′-CCCCCTACACCTTCTTGCCC-3′) and ALF-down (5′-TGCGTGCGACAGAGCGGCC-3′) and purified using a MiniElute Kit (Qiagen). A consensus chicken β-globin FII CTCF site (19) was prepared from FII-up (5′-AGGCGCG- CCCCCAGGATGTAATTACGTCCCTCCCCCGGTTAGG GGGCAGCGCGCGCCT-3′) and a complementary oligonucleotide cFII-down. Other oligonucleotides included an Sp1 consensus binding site (top strand, 5′-ATTCGATCGGGGCG-
GGCGGAG-3’), a polyoma virus site for RFX (PyEP; top strand, 5’-GGCCAGTGGCCTAGCAACTAATAC-3’) (21), and other sequences that are indicated in individual figures. Oligonucleotides (10 pmol) were heated at 100 °C for 15 min and cooled at room temperature for 2 h to make duplexes.

Labeling of DNAs was performed with T4 polynucleotide kinase (New England Biolabs) and [γ-32P]ATP (PerkinElmer Life Sciences) at 37 °C for 30 min. Labeled probes were purified using a NICK column (Amersham Biosciences). DNA was labeled with SssI and S-adenosylmethionine at 37 °C for 3 h. Reactions were inactivated at 65 °C for 20 min. To confirm that methylation was complete, DNAs were digested with methylation-sensitive and methylation-insensitive restriction enzymes HpaII and MspI.

Preparation of Tissue Extracts and Recombinant Proteins—Mouse testis and liver were used to prepare nuclear extracts (NE) as described below (22). In brief, razor-minced tissues were homogenized in 20 ml of HB buffer (10 mm HEPES-KOH (pH 7.6), 15 mm KCl, 0.15 mm spermine, 0.5 mm spermidine, 1 mm EDTA, 2.4 mm sucrose, 1% low fat milk, 0.5 mm dithiothreitol, 0.5 mm phenylmethylsulfonyl fluoride, and 1% protease inhibitor) and loaded into an SW28 tube on top of a 10-ml HB cushion. After centrifugation at 24,000 rpm for 60 min at 4 °C, the supernatant was removed, and the nuclear pellet was resuspended with water. The pellet was then resuspended with NL buffer (10 mm HEPES-NaOH (pH 7.6), 100 mm KCl, 0.1 mm EDTA, 10% glycerol, 3 mm MgCl2, 1 mm dithiothreitol, 0.1 mm phenylmethylsulfonyl fluoride, and 1% proteinase inhibitor) and homogenized with a hand-held glass homogenizer. The DNA concentration was measured and diluted to 0.5 mg/ml with NL buffer. One-tenth volume of 4 mm (NH4)2SO4 was added, mixed by inversion, and incubated on ice for 30 min. The sample was then centrifuged at 35,000 rpm for 60 min, and the supernatant was transferred to a new tube. Solid (NH4)2SO4 was added, mixed by inversion, and incubated on ice for 30 min, and centrifuged at 40,000 rpm for 20 min. The pellet was resuspended in 25 mm HEPES (pH 7.6), 0.1 mm EDTA, 40 mm KCl, 10% glycerol, and 1 mm dithiothreitol.

DNA fragments encoding full-length CTCF and an internal zinc domain were amplified from human testis Marathon-ready cDNA (Clontech) with hCTCF-1 (5’-GGCGGAGTGGCCTAGCAACTAATAC-3’) and hCTCF-2 (5’-GGCGGAGTGGCCTAGCAACTAATAC-3’) and with hCTCF-Zn-1 (CGGAGTGGCCTAGCAACTAATAC-3’) and hCTCF-Zn-2 (5’-GGCGGAGTGGCCTAGCAACTAATAC-3’) and with hCTCF-Zn-1 (CGGAGTGGCCTAGCAACTAATAC-3’) and hCTCF-Zn-2 (5’-GGCGGAGTGGCCTAGCAACTAATAC-3’). PCR products were digested with EcoRI and HindIII, purified using a gel extraction kit (Qiagen), and cloned into pGEM-HE (23). Constructs were translated in the TnT Quick coupled transcription and translation kit (Promega). In brief, 2 µg of DNA was mixed with TnT master mix and methionine and incubated at 30 °C for 90 min.

Bandshift Analysis—Binding reactions were typically performed in 1X phosphate-buffered saline, 5 mm MgCl2, 1 mm dithiothreitol, 0.02% Nonidet P-40, 8% glycerol, 0.1 mm ZnSO4, and 80 µg/ml poly(dI-dC). Reactions were incubated at room temperature for 30 min and resolved on 5% native polyacrylamide gels containing 0.5X TBE and 5% glycerol. Some reactions contained antibodies raised against CTCF (C terminus (Upstate Biotechnology, Inc., Lake Placid, NY), CTCF (N terminus) (BD Biosciences), Sp1 (Upstate Biotechnology), Sp3 (Santa Cruz Biotechnology, Santa Cruz, CA), or RFX1 (Santa Cruz Biotechnology).

Promoter Constructs and Expression Vectors—The wild type ALF promoter and a series of eight mutant derivatives were prepared by PCR. For wild type (WT) and mutant 5 to mutant 7 (M5–M7), the reactions contained primer PR0 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PR5 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PR5 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PR6 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), or PR7 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’). For mutant 1 to mutant 4 (M1–M4), the reactions contained primer PR0 as showed above and one of the three primers: PF1 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PF2 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PF3 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), or PF4 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’).

To make expression constructs for CTCF (pFCMV-CTCF), BORIS (pFCMV-BORIS), Sp1 (pFCMV-Sp1), and Sp3 (pFCMV-Sp3), open reading frames for each factor were amplified from human testis RNA using hCTCF-Full-1 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-2 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-3 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-4 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-5 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-6 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-7 (5’-GGGGAAGCTTTCATGGGCC-3’), or hCTCF-Full-8 (5’-GGGGAAGCTTTCATGGGCC-3’). KpnI and NgII sites were added to the primer ends to facilitate cloning, and the amplified fragments were inserted upstream of the luciferase gene in pGL3 Enhancer (Promega).

To make expression constructs for CTCF (pFCMV-CTCF), BORIS (pFCMV-BORIS), Sp1 (pFCMV-Sp1), and Sp3 (pFCMV-Sp3), open reading frames for each factor were amplified from human testis RNA using hCTCF-Full-1 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-2 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-3 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-4 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-5 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-6 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-7 (5’-GGGGAAGCTTTCATGGGCC-3’), or hCTCF-Full-8 (5’-GGGGAAGCTTTCATGGGCC-3’). KpnI and NgII sites were added to the primer ends to facilitate cloning, and the amplified fragments were inserted upstream of the luciferase gene in pGL3 Enhancer (Promega).

Regulation of the Germ Cell-specific ALF Gene

The wild type ALF promoter and a series of eight mutant derivatives were prepared by PCR. For wild type (WT) and mutant 5 to mutant 7 (M5–M7), the reactions contained primer PR0 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PR5 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PR6 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), or PR7 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’). For mutant 1 to mutant 4 (M1–M4), the reactions contained primer PR0 as showed above and one of the three primers: PF1 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PF2 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), PF3 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’), or PF4 (5’-TGTGATCATCCTTGGCCCGGTCCTGCG-3’).

To make expression constructs for CTCF (pFCMV-CTCF), BORIS (pFCMV-BORIS), Sp1 (pFCMV-Sp1), and Sp3 (pFCMV-Sp3), open reading frames for each factor were amplified from human testis RNA using hCTCF-Full-1 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-2 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-3 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-4 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-5 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-6 (5’-GGGGAAGCTTTCATGGGCC-3’), hCTCF-Full-7 (5’-GGGGAAGCTTTCATGGGCC-3’), or hCTCF-Full-8 (5’-GGGGAAGCTTTCATGGGCC-3’). KpnI and NgII sites were added to the primer ends to facilitate cloning, and the amplified fragments were inserted upstream of the luciferase gene in pGL3 Enhancer (Promega).

RNA Interference—NIH3T3 cells were trypsinized and distributed into 6-well plates without antibiotics at 2 × 105 cells/well. Cells were transfected with CTCF-specific (Dharmacon)
Regulation of the Germ Cell-specific ALF Gene

**RESULTS**

Protein-DNA Interactions on the ALF Promoter—A fragment of the mouse ALF promoter from −117 to −5 relative to the major transcription start site (Fig. 1A) was used as a probe in bandshift reactions that contained nuclear extracts from either mouse liver or testis. The probe spanned a region of the promoter previously shown to drive correct regulation in transgenic mice (10). Six major complexes could be resolved using liver extracts (Fig. 1B; lane 1) and were denoted as C1–C6. Three of these complexes were also abundant in testis extracts (C3, C4, and C5) (lane 2). Additional rapidly migrating complexes were less intense or variably present and were not given separate designations. The results depicted are representative of results obtained with several nuclear extract preparations.

**Zinc Finger Transcription Factors Sp1 and Sp3 Recognize the ALF Promoter**—The ALF promoter sequence is relatively GC-rich and contains a GGGCGG element located in reverse orientation relative to the direction of transcription at positions −84 to −79 (Fig. 1A). To test whether members of the Sp1 family of transcription factors might recognize this element and generate some of the observed complexes, we evaluated the ability of an Sp1 antibody specific to Sp3 selectively diminished formation of the C4 complex in both liver (Fig. 1B; lane 1) and testis (lanes 3 and 4). Additional rapidly migrating complexes were less intense or variably present and were not given separate designations. The results depicted are representative of results obtained with several nuclear extract preparations.

FIGURE 1. The zinc finger transcription factors Sp1 and Sp3 recognize mouse ALF promoter DNA. A, the sequence of the mouse ALF promoter between −133 and +25 is shown. The shaded regions highlight the TC-rich area, the TTCCAAA sequence, and sites for factors identified in this report. CG dinucleotides are indicated in red. The arrows show known transcription start sites, and the dashed line shows the region of the ALF promoter used in initial DNA-protein interaction assays. B, bandshift analysis using ALF promoter DNA with liver (lane 1) and testis (lane 2) extracts reveals six major complexes in liver (C1–C6), and three major complexes in testis (C3–C5). C, liver (lanes 1 and 2) and testis nuclear extracts (lanes 3 and 4) were incubated with the ALF promoter in the absence (lane 1 and 3) or presence (lane 2 and 4) of an Sp1 binding site competitor. Complexes C4 and C5 were sensitive to the addition of competitor, whereas complexes C1, C2, C3, and C6 were unaffected. D, liver and testis nuclear extracts were incubated with an ALF probe in the absence (lanes 1 and 3) or presence (lanes 2 and 4) of an Sp1-specific antibody (lanes 2 and 4). The results show that both the C4 and C5 complexes are specifically eliminated, whereas the other complexes are unaffected. Lanes 5 and 10 are controls showing that the antibodies themselves do not have intrinsic binding activity.

or negative control siRNAs (Ambion) using Lipofectamine 2000 (Invitrogen). The concentration of siRNA that gave the least cell deterioration was 0.58 μg/well, with each well containing 2 ml of Opti-MEM I reduced serum medium (Invitrogen). Cells were cultured for 8 h and then switched to media containing 10% fetal bovine serum and 1% antibiotic solution.

Transfections—293T cells were grown to 50–80% confluence on 6-well plates and transfected with 1 μg of DNA using the FuGENE 6 reagent (Roche Applied Science). PGL3 Control and pGL3 Enhancer served as positive and negative controls, respectively. Whole cell extracts were prepared after 48 h, and luciferase activity was assayed using the luciferase assay system (Promega) and a Turner TD-20e luminometer.

For cotransfection experiments with pmALFpro200 and individual expression vectors, 293T cells were plated at 4 × 10⁵ cells into 6-well plates. After overnight incubation, cells were transfected with expression constructs together with 1 μg of mALFpro200-Luc construct using the FuGENE6 transfection reagent (Roche Applied Science). The pCMV Sport-β-gal vector (100 ng ATCC) was used to monitor transfection efficiency, and the pGL3 Control and pGL3 Enhancer vectors were used as positive and negative controls. After 48 h, luciferase and β-galactosidase assays were performed using the luciferase assay system and the β-galactosidase assay kit (Invitrogen). Luciferase activity was normalized to β-galactosidase activities. All transfections were performed between three and six times each.
6–9). These results demonstrate that the ALF promoter is recognized by the zinc finger transcription factors Sp1 and Sp3.

The Insulator/Silencer Protein CTCF Recognizes the ALF Promoter—Some zinc finger transcription factors, especially those with multiple fingers, have relatively broad site specificity for GC-rich sequences. An example of such a factor is the chromatin insulator and silencer protein CTCF (11, 13). Because it seemed plausible that insulator-like activities might be important for the regulation of germ cell genes in somatic cells, we asked whether CTCF would interact with ALF promoter DNA.

To address this question we first synthesized full-length CTCF protein and a zinc finger DNA binding subdomain (CTCF-Zn) using a coupled transcription-translation system. Control experiments showed that programmed lysates generated strong shifts when incubated with a control CTCF site from the chicken β-globin gene (19) (data not shown). When lysates programmed with the full-length CTCF were used in reactions with a labeled ALF promoter, two complexes were generated (Fig. 2A, lane 1). Competition experiments with unlabelled oligonucleotides corresponding to ALF, a consensus cFII site, and an Sp1 site were consistent with the idea that the CTCF-programmed reticulocyte lysate was specifically recognizing an element within the ALF promoter (lanes 2–4). Lysates programmed with the smaller CTCF-Zn fragment generated a strong, more rapidly migrating complex whose properties in competition experiments matched those of its full-length counterpart (lanes 5–8). In addition, a human ALF promoter probe would also be recognized by the CTCF-Zn protein, albeit more weakly than the mouse probe (Fig. 2B, lanes 1–4).

Endogenous CTCF Protein in Liver and Testis Extracts Binds to the ALF Promoter—To directly show that CTCF was responsible for complex formation, we performed supershift reactions using commercially available CTCF-specific antibodies raised against either the N terminus (CTCF N-Ab) or the C terminus (CTCF C-Ab) of CTCF. The N-terminus-specific antibody eliminated formation of both complexes seen with CTCF-programmed lysates, generating an antibody-dependent supershift (Fig. 2C, lanes 1 and 2). The C-terminus-specific antibody eliminated the upper complex but not the lower complex (lanes 5 and 6), indicating that the lower of the two complexes lacks a complete C terminus. Neither antibody had an effect on the complexes formed with CTCF-Zn (Fig. 2C, lanes 3, 4, 7, and 8).

We next asked whether complexes formed on the ALF promoter using liver and testis extracts were actually due to endogenous CTCF. This seemed likely, since the large and small complexes observed with recombinant CTCF comigrated with extract-derived complexes C1 and C3 (Fig. 2D, lanes 1 and 2). When the N-terminus-specific CTCF antibody was added to the reaction, the C3 and C1 complexes were supershifted, but no other complex was affected (compare lanes 2 and 3). Similarly, when these antibodies were added to reactions that contained testis extracts, the C3 complex was supershifted (Fig. 2D, lanes 4 and 5). In addition, the C-terminus-specific antibody was also able to abolish formation of the C3 complex in testis extracts (lanes 7 and 8). Since the antibodies used in these experiments do not recognize the male-specific CTCF-like factor BORIS (12), the C3 complex observed in whole testis extracts is due to CTCF. We did not observe complexes that were shifted by both the Sp1/Sp3 and CTCF antibodies, sug-
suggesting that the probe was not bound by more than one factor. Overall, these results lead to the interesting conclusion that the insulator/silence factor CTCF interacts with an element within the germ cell-specific ALF gene.

Localization of Regulatory Factor Binding Sites—To locate the positions at which Sp1, Sp3, and CTCF were bound, we prepared a series of oligonucleotides that contained mutations along the length of the sequence (mut 5, mut 3-1, mut 3-2, and mut 3-3), or mutations in selected C or G residues (mut CG to AG, mut CG to CT, mut C to A, and mut G to T) (Fig. 3A). These oligonucleotides, together with a wild type control, were then used as competitors for complexes formed with liver NE (Fig. 3B).

The results make several important points about regulatory factor interactions along the ALF promoter. First, the mut Sp1 competitor, which contains a mutated GC box, is the only oligonucleotide that could not compete for the C4/C5 complexes (Fig. 3B, lane 7). Thus, we conclude that this site is critical for the binding of both Sp1 and Sp3. Second, the mut 5 and mut 3-2 oligonucleotides compete for CTCF binding (lanes 3 and 5), whereas mut Sp1 and mut 3-1 do not compete (lanes 4 and 7). These results thus define a 17-nucleotide CTCF recognition motif that overlaps with the Sp1/Sp3 GC box site at the 5'-side (−84) and extends to position −68 at the 3'-side. The competition results also give a clue to the location of the C6 complex. In particular, the inability of mut 3-2, mut 3-3, mut CG to AG, and mut G to T oligonucleotides to compete for this complex (lanes 5, 6, 8, and 11) places the binding site for this factor just downstream of CTCF. The competition properties of C2 and smaller complexes were either indeterminate or similar to CTCF.

The Winged Helix Transcription Factor RFX1 Binds to the Mouse ALF Promoter—Inspection of sequences where complex C6 was formed indicated possible binding sites for C/EBP or RFX factors. Supershift experiments using anti-C/EBP antibodies failed to produce any change (data not shown). On the other hand, a well characterized RFX site from the polyoma virus enhancer was able to eliminate the C6 complex without affecting any of the other complexes (Fig. 3C, lanes 1–3). An antibody-alone control is shown in lane 6. D, an alignment of the mouse ALF RFX site together with RFX sites from several other genes shows strong similarity to the RFX consensus.

FIGURE 3. Localization of factor binding sites and identification of an RFX site. A, the sequence of the wild type ALF promoter DNA fragment is shown at the top (WT). DNAs below the WT sequence contain mutations at specific nucleotides, shown by the dark boxes. The name of each oligonucleotide is shown to the right. B, ALF promoter-dependent complexes formed with liver extracts (lane 1) are competed with each of the oligonucleotides listed in A (lanes 2–11). The results reveal overlapping binding sites for Sp1/Sp3 and CTCF and identify an additional downstream binding site for complex C6. C, reactions containing liver NE and the ALF promoter (lane 1) were incubated with a polyoma virus EP RFX site oligonucleotide (lane 2) or with an oligonucleotide that contained mALF promoter sequences downstream of mutation 3-1. Both oligonucleotides competed for C6 binding but did not affect any of the other complexes. In addition, the presence of an RFX-1-specific antibody resulted in the selective supershift of the C6 complex (compare lanes 4 and 5). An antibody-alone control is shown in lane 6. D, an alignment of the mouse ALF RFX site together with RFX sites from several other genes shows strong similarity to the RFX consensus.
Alignment of the RFX site from the ALF promoter with other RFX sites shows a match to a consensus (25) (GTNRCCNNR-NNGCCGCTAGGGGGNGG-3) at seven of eight positions for which a particular nucleotide is specified and at all three positions where a purine (R) or a pyrimidine (Y) is specified (Fig. 3D).

**A Minimal CTCF Site and Prediction of Additional Genomic Sites**—The results in Fig. 5B showed that mutations on either side of a 17-bp region had no effect on CTCF binding. This was surprising, since the binding site is thought to be relatively long (13, 24). To investigate this point, we synthesized two additional sets of oligonucleotides for use as binding site competitors. Members of one set contained the 17-base pair core but differed in the length of their flanking sequences (Fig. 4A). The second set was composed of natural genomic DNA sequences from mice, cows, pigs, and dogs that match the 17-bp core sequence. All of the sequences contain a GC box as part of their flanking sequences are efficient competitors for both Sp1/Sp3 (C4/C5) and CTCF (C1/C3) binding but not for RFX binding (C6) (lanes 2–9). A mouse ALF self-competitor control competes for the same complexes but also abolishes the RFX (C6) complex (lane 10), whereas a mutant Sp1 site competitor only competes for the RFX (C6) complex (lane 11).

When oligonucleotides from the first set were added to bandshift reactions containing the ALF promoter, we observed a binding region-dependent competition for CTCF binding (Fig. 4C). In particular, WT29, with 6 bp of flanking DNA, and WT17, with no flanking DNA, were unable to compete for binding (lanes 4 and 5), whereas the WT and WT40 competitors could compete. WT17 was the only oligonucleotide that could not compete for Sp1/Sp3 binding (lane 5). Overall, these results indicate that although the 17-bp core provides recognition specificity, additional flanking DNA is necessary for stable interactions to occur.

We then asked whether the core site would predict CTCF sites in genomic DNA and whether the surrounding sequences would have an effect on the strength of binding. The experiment was performed under conditions in which the competitor was not in vast excess, so that any differences due to flanking DNA would be observed. The results show that all of these natural genomic sequences could compete for binding (Fig. 4D, lanes 2–9). Because there is a GC box within the CTCF site, all of the oligonucleotides were able to compete for Sp1/Sp3 binding. However, none were able to compete for the C6 complex, since their flanking regions did not contain an RFX site. Control reactions showed that the wild type mouse ALF self-competitor competed with similar efficiency as the test set (lane 10), whereas a mutant GC oligonucleotide was unable to compete for either Sp1/Sp3 or CTCF binding (lane 11).

**Alignment of CTCF Sites**—The ability of CTCF to recognize seemingly dissimilar binding sites, in part through the use of different combinations of zinc fingers, makes identification of CTCF sites a challenging proposition (13). Still, alignments of such sites may be helpful in identifying common sequence features. The alignment shown in Fig. 5A was optimized visually by placing sequences around the distinct stretches of G and C nucleotides in register as well as possible. To help simplify the alignment, the decision was made to allow no gaps, and the use of “top” or “bottom” strands was based on which fit the shaded area (Fig. 5A). The consensus is 5’-CCCNG-GGCCGCTAGGGGGGCGGC-3’. If the stringency is increased to 60% identity at each position, the consensus is 5’-CCN-NNGCCGCTAGGGGGGNGG-3’. If the stringency is increased to 70%, the consensus is 5’-GCNGGCTAGGGGGGG-3’.

An important test of the alignment is whether it could or could not generate a functional CTCF site. To test this point, we synthesized oligonucleotide competitor pools that matched the 50, 60, and 70% stringency thresholds (Fig. 5B). To eliminate
Regulation of the Germ Cell-specific ALF Gene

possible effects of flanking DNA, we synthesized these oligonucleotides with either the natural mouse ALF flanking DNA, or with random AGTC residues. As shown in Fig. 5C, all of the sequences were able to compete for formation of the Sp1/Sp3 and CTCF complexes to varying degrees. At the 70% threshold, however, both types of oligonucleotides became less effective as competitors for CTCF binding as well as for Sp1/Sp3 (lanes 1–6). As expected, only oligonucleotides with the ALF flanking regions were able to compete for the C6 (RFX) complex (lanes 2–4). The data show that a functional binding site had been identified by this analysis and that the determinants for binding lie within a relatively small region. It is clear, however, that many sequence variations are consistent with binding, and further studies will be required to address this specificity issue.

DNA Methylation Status Affects CTCF Binding to the ALF Promoter—Because methylation status is associated with ALF expression (26), we asked whether any of the complexes formed on the ALF promoter would be affected by methylation. To address this point, we prepared methylated and unmethylated ALF probes using SssI methylase. The reaction was monitored with the restriction enzyme HpaII and its methylation-insensitive isoschizomerMspI. Digestion of the unmethylated probe generated two fragments regardless of which enzyme was used, whereas the SssI-methylated probe could only be digested with MspI (Fig. 6A, lanes 1–6).

The results of interaction assays showed that a control ALF probe that had been used in earlier experiments and a probe that had gone through a mock methylation reaction but which was unmethylated were both able to bind CTCF and CTCF-Zn polypeptides (Fig. 6B, lanes 1, 2, 4, and 5). In contrast, a methylated ALF probe was unable to form complexes with either of the recombinant proteins (lanes 3 and 6). In addition, both the endogenous C1 and C3 complexes were absent in reactions that contained liver NE and the methylated ALF probe (Fig. 6C, lanes 1 and 2). Other complexes, most notably C4 and C5, were not affected. The results show that CTCF would be specific for an unmethylated or undermethylated promoter, whereas Sp1 and Sp3 would bind independent of methylation status.

Functional Analysis Suggests That the ALF Promoter Is Repressed in Somatic Cells—The identification of the silencing factor CTCF and the fact that the endogenous ALF gene is normally off in somatic cells suggested the possibility that the promoter might be maintained in a repressed state. To investigate this question, we prepared a wild-type ALF promoter construct spanning nucleotides −105 to +8 linked to a downstream luciferase reporter in the pGL-3 Enhancer vector. This construct was used as a parent to prepare a series of 10-bp mutations spaced at 4-bp intervals along the length of the promoter, beginning at position −99 and ending at position −6.
The strategy for mt1, mt4, mt5, and mt6 was that A and G were changed to T, whereas T and C were changed to G. For mt2, mt3, and mt7 A and G were changed to C, and T and C were changed to A.

Six-fold replicate transfection assays were performed with each construct in somatic 293 cells. Interestingly, results with the mt3, mt4, and mt5 constructs all resulted in an increase in activity relative to the wild-type promoter (Fig. 7B). The fact that three different 10-bp mutations spanned a region of nearly 40 nucleotides argued against the idea that a particular mutation had resulted in the accidental creation of an activator element. Instead, the results seemed to indicate that mutations along this region resulted in a partial release from repression. The position of mt3 mutations coincides with one side of the CTCF binding site, the position of mt4 falls within the predicted site for RFX, and mt5 removes the conserved TTCAAA element. The mt2 resulted in a minor reduction in activity, whereas a construct in which all but one CpG dinucleotide had been mutated to GT showed very low activity. Mutations mt6 and mt7 show only modest reductions in activity, consistent with the fact that these are downstream of the identified transcription start sites. The results support the idea that elements within the ALF promoter may serve to maintain low activity in somatic cells.

We next wanted to test whether some of the factors we had identified had repressive effects on ALF promoter activity in transient cotransfection assays using 293T cells (Fig. 7C). Experiments were performed using constant amounts of pmALF200, a pGL3-based luciferase reporter driven by 200 bp from the mALF promoter. The results showed several interesting trends. First, increasing amounts of pFCMV-CTCF resulted in decreased ALF reporter expression. In contrast, cotransfection with pFCMV-BORIS, which encodes the germ cell-specific counterpart of CTCF (12), resulted in a modest dose-dependent increase in expression. These results are consistent with the notion that CTCF acts to down-regulate expression and with recent reports indicating that CTCF and BORIS may have opposing effects on genome-wide patterns of gene expression in tumor cell lines (27, 28). Additional experiments with
Regulation of the Germ Cell-specific ALF Gene

pFCMV-Sp1 and pFCMV-Sp3 showed that whereas Sp1 did not affect the reporter, Sp3 resulted in a substantial repressive effect even at low concentrations. This is in agreement with reports of Sp3-dependent repression (29). The data demonstrate that two of the factors identified have the ability to repress ALF expression in somatic cells.

If CTCF is a bona fide regulator of the ALF gene, we would predict that a decline in the levels of endogenous CTCF might result in ALF gene activation. To test this possibility, CTCF siRNAs were transfected into NIH3T3 cells, and the level of the endogenous ALF mRNA was measured at various days thereafter. Although it was somewhat difficult to balance siRNA transfection without having adverse effects on the cells, we were able to observe that a decrease in the level of CTCF mRNA (Fig. 7D, lanes 1–4) was associated with the appearance of ALF expression (lanes 9–12). Western analysis showed that treatment was associated with a decline in CTCF protein. ALF expression rose to maximal levels by the second day after transfection (lane 11) and was not observed in untreated control cells or with control RNA samples that had not been treated with reverse transcriptase (lanes 14–17). In addition, an actin control used to normalize the amount of RNA used was unaffected by treatment (lanes 5–8). Collectively, functional assays and binding data indicate that CTCF and Sp3 are repressive factors that could play a role in the somatic cell-specific down-regulation of the ALF gene.

DISCUSSION

The aim of this report was to characterize mechanisms important for the proper expression of germ cell-specific genes. This is an interesting issue, because such genes are regulated in a strict on-off manner and because their promoters are typically very short. Germ cell transcription is also peculiar, because it often involves a switch from somatic to germ cell sites of initiation (30) and, in some cases, is controlled by regulatory elements within retrotransposons (31). Further complexity is suggested by the observation that there are subgroups of germ cell genes whose temporal activation correlates with progression through spermatogenesis (32). Although experiments in transgenic mice show that germ cell promoter modules provide sufficient cues for activation and silencing, many basic problems, including the identification of the relevant regulatory factors, remain to be understood.

Some of the results reported here are summarized in Fig. 8A. One conclusion is that the mouse ALF promoter is recognized by members of the Sp1 family of zinc finger transcription factors. Sp1 not only has a wide role in transcriptional regulation of housekeeping genes but is also reported to control germ cell-specific genes such as PIA5x, H1t, PGK2, LDH-C, cyclin A1, and PDHA-2 (33–39). Moreover, expression of the Sp1 gene is characterized by several unique features in germ cells, including up-regulation during the pachytene stage of meiotic prophase I and the generation of alternate mRNA transcripts (15, 16). Additional members of this family include Sp2, Sp3, and others (14), but the roles of these as gametogenic factors are less well established. Our studies suggest that Sp3 may act as a somatic repressor of ALF expression, a result consistent with other reports of Sp3-dependent repression (29).

Binding data also indicate that members of the winged helix RFX family of transcription factors can recognize the ALF promoter. The particular species identified in this report is the somatic factor RFX1, although other members of this family, including RFX2 and RFX4, are testis-specific (17, 18). RFX sites are present in the germ cell-specific H1t gene, and RFX2 activates H1t expression in GC-2spd germinal cells (25, 40). In this report, we did not further evaluate the role of this factor for ALF expression.

The most surprising finding here was that the ALF promoter is recognized by the chromatin insulator and transcriptional silencer CTCF. This factor was first identified as a protein that bound within a DNase I hypersensitive region of the chicken β-globin gene and exhibited enhancer blocking activity (19). We observed two endogenous CTCF-dependent complexes with an ALF promoter probe, C1 and C3, one of which is due to a truncated form of CTCF that lacks a complete C terminus (C1). Although we do not know if this species has a distinct physiological role, its abundance relative to the full-length C3 complex is notably reduced in extracts of whole testis.
An unusual feature of CTCF is that its binding site specificity is broad or adaptable through the use of alternate combinations of zinc fingers (13, 24). Our studies on the binding of CTCF to the ALF promoter revealed several important points. First, the mouse ALF CTCF site is 17 bp long and could potentially be even smaller. Second, this site has dual specificity for both Sp1/Sp3 and CTCF. Third, this site can be used directly or by alignment to predict other CTCF sites in genomic DNA. Finally, additional flanking DNA is required to provide a stable platform for binding.

The sensitivity of CTCF to methylation status raises the question of whether this factor would bind the endogenous promoter. Inspection of earlier data shows that of 13 ALF promoter-derived PCR products, only two were methylated at any of the three CpGs located within the CTCF site (26). The remaining 11 sequences were not modified at any of the sites, although most were modified at one or more sites located upstream or downstream. Thus, the reported methylation pattern of CpG dinucleotides in and around the first exon of the ALF gene in somatic cells appears not to preclude CTCF binding. Of course, it is possible that binding is normally transient, for instance working in a “touch-and-go” manner to establish a repressive state at specific times in development or when the DNA becomes undermethylated. In addition, the fact that CTCF, Sp1, and Sp3 bind to overlapping sites and have differential sensitivity to methylation status suggests that the relevant interaction(s) and functional consequences would depend on DNA interaction affinities, the extent of promoter methylation, and relative factor concentrations in a particular tissue.

Several recent studies have shown that CTCF levels are reduced and BORIS levels are increased in some cancer cell lines and that this is associated with the derepression of some testis-specific genes (27, 28). This observation is consistent with results presented here showing a direct interaction between CTCF and a germ cell promoter and with the observation that CTCF and BORIS had opposing effects on the expression of an ALF promoter-reporter construct (Fig. 7C). CTCF is normally down-regulated in male germ cells and replaced with BORIS (12). We speculate that this exchange could result in a chromatin state that is permissive for interactions with the basal transcription machinery and site-specific regulators, resulting in gene activation. Once activated, ALF, together with other germ cell transcription factors, is likely to mediate downstream effects on gene expression (Fig. 8A). The observation that CTCF is up-regulated later in spermatogenesis (12) suggests a possible role in the reestablishment of a silent, somatic state.

The demonstration that there are direct interactions between a germ cell-specific promoter and an insulator such as CTCF provides a potential mechanism by which such genes might protect themselves from inappropriate activation by enhancers of adjacent genes (Fig. 8B). Silencer elements and repressive factors have been suggested for other germ cell-specific genes, including PGK-2, H1t, SP-10, and LDHC (41–44). Although in many cases the factors responsible have not been identified, several studies have shown that E2F6 is important to maintain the somatic silencing of germ cell genes such as α-tubulin, Tex12, SMC1β, and STAG3 (45, 46), whereas repression of the SP-10 gene involves TDP-43 (47). In addition to the role of specific factors, higher levels of DNA methylation are sometimes associated with silent germ cell genes (48). Overall, the results point to the idea that release of somatic silencing may be critical for the activation of germ cell genes.

We had demonstrated earlier that TBP was capable of recognizing the TATA-like TTCAAA element (26), and in the current study we identify four additional transcription factors as candidate regulators of the ALF gene. Three of these factors bind to a recognition element that has dual specificity for members of the Sp1 and CTCF families. These of these factors on ALF promoter activity support the idea that repression and derepression are an important aspect of germ cell gene regulation. In future studies, we hope to address the physiological role of each factor and the role of chromatin packaging in germ and somatic tissues.

REFERENCES

1. DeJong, J. (2006) Gene (Amst.) 366, 39–50
2. Veenstra, G. J. C., and Wolffe, A. P. (2001) Trends Biochem. Sci. 26, 665–671
3. Hochheimer, A., and Tjian, R. (2003) Genes Dev. 17, 1309–1320
4. Upadhyaya, A. B., Lee, S. H., and DeJong, J. (1999) J. Biol. Chem. 274, 18040–18048
5. Ozer, I., Moore, P. A., and Lieberman, P. M. (2000) J. Biol. Chem. 275, 122–128
6. Han, S. Y., Zhou, L. P., Upadhyaya, A. B., Lee, S. H., Parker, K. L., and DeJong, J. (2001) Biol. Reprod. 64, 507–517
7. Upadhyaya, A. B., Khan, M., Mou, T.-C., Junker, M., Gray, D. M., and DeJong, J. (2002) J. Biol. Chem. 277, 34208–34216
8. Han, S. Y., Xie, W., Hammes, S. R., and DeJong, J. (2003) J. Biol. Chem. 278, 45586–45593
9. Huang, M., Wang, H., Li, J., Zhou, Z., Du, Y., Lin, M., and Sha, J. (2006) Int. J. Mol. Med. 17, 599–604
10. Han, S. Y., Xie, W., Kim, S. H., Yue, L., and DeJong, J. (2004) Biol. Reprod. 71, 933–941
11. Klenova, E. M., Morse, H. C., Ohlisson, R., and Lobanenkov, V. V. (2002) Semin. Cancer Biol. 12, 399–414
12. Loukinov, D. I., Pugacheva, E., Vatolin, S., Pack, S. D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A. C., Cui, H., Niemitz, E. L., Rasko, J. E., Ducquier, F. M., Kistler, M., Breen, J. J., Zhuang, Z., Quitschke, W. W., Renkwitz, R., Klenova, E. M., Feinberg, A. P., Ohlisson, R., Morse, H. C., and Lobanenkov, V. V. (2002) Proc. Natl. Acad. Sci. 99, 6806–6811
13. Ohlisson, R., Renkwitz, R., and Lobanenkov, V. (2001) Trends Genet. 17, 520–527
14. Suske, G. (1999) Gene (Amst.) 238, 291–300
15. Persengiev, S. P., Raval, P. J., Rabinovitch, S., Millette, C. F., and Kilpatrick, D. L. (1996) Endocrinology 137, 638–646
16. Thomas, K., Sung, D. Y., Yang, J., Johnson, K., Thompson, W., Millette, C., McCarrey, J., Breitbart, A., Gibbs, R., and Walker, W. (2005) Biol. Reprod. 72, 898–907
17. Reith, W., Ucla, C., Barras, E., Gaud, A., Durand, B., Herrero-Sanchez, C., Kobr, M., and Mach, B. (1994) Mol. Cell. Biol. 14, 1230–1244
18. Morottoni-Yano, K., Yano, K.,–L, Saito, H., Sun, Z., Iwama, A., and Miki, Y. (2002) J. Biol. Chem. 277, 836–842
19. Bell, A. C., West, A. G., and Felsenfeld, G. (1999) Cell 98, 387–396
20. Bell, A. C., West, A. G., and Felsenfeld, G. (2001) Science 291, 447–450
21. Nakayama, A., Murakami, H., Maeyama, N., Yamashiro, N., Sakakibara, A., Mori, N., and Takahashi, M. (2003) J. Biol. Chem. 278, 233–240
22. Sierra, F. (1990) J. Biol. Chem. 265, 1309–1320
23. Sierra, F. (1990) J. Biol. Chem. 265, 1309–1320
24. Filippova, G. N., Qi, C.-F., Ulmer, J. E., Moore, J. M., Ward, M. D., Hu, Y.J., Loukinov, D. I., Pugacheva, E. M., Klenova, E. M., Grundy, P. E., Feinberg, A. P., Ohlisson, R., Morse, H. C., and Lobanenkov, V. V. (2002) Proc. Natl. Acad. Sci. 99, 6806–6811
Regulation of the Germ Cell-specific ALF Gene

A. P., Cleton-Jansen, A.-M., Moerland, E. W., Cornelisse, C. J., Suzuki, H., Komiya, A., Lindblom, A., Dorion-Bonnet, F., Neiman, P. E., Morse, H. C., Collins, S. J., and Lobanenkov, V. V. (2002) Cancer Res. 62, 48–52

25. Horvath, G. C., Kistler, W. S., and Kistler, M. K. (2004) Biol. Reprod. 71, 1551–1559

26. Xie, W., Han, S. Y., Khan, M., and DeJong, J. (2002) J. Biol. Chem. 277, 17765–17774

27. Hong, J. A., Kang, Y., Abdullaev, Z., Flanagan, P. T., Pack, S. D., Fischette, M. R., Adnani, M. T., Loukinov, D. I., Vatolin, S., Risinger, J. I., Custer, M., Chen, G. A., Zhao, M., Nguyen, D. M., Barret, J. C., Lobanenkov, V. V., and Schrump, D. S. (2005) Cancer Res. 65, 7763–7774

28. Vatolin, S., Abdullaev, Z., Pack, S. D., Flanagan, P. T., Custer, M., Loukinov, D. I., Pugacheva, E., Hong, J. A., Morse, H., Schrump, D. S., Risinger, J. I., Barret, J. C., and Lobanenkov, V. V. (2005) Cancer Res. 65, 7751–7762

29. De Luca, P., Majello, B., and Lania, L. (1996) J. Biol. Chem. 271, 8533–8536

30. Schmidt, E. E., Ohbayashi, T., Makino, Y., Tamura, T., and Schibler, U. (1997) J. Biol. Chem. 272, 5226–5334

31. Peaston, A. E., Evslikov, A. V., Graber, J. H., deVries, W. N., Holbrook, A. E., Solter, D., and Knowles, B. B. (2004) Dev. Cell 7, 597–606

32. Shima, J. E., McLean, D. J., McCarrey, J. R., and Griswold, M. D. (2004) Biol. Reprod. 71, 319–330

33. Zhang, P. L., Stroud, J., Eddy, C. A., Walter, C. A., and McCarrey, J. R. (1999) Biol. Reprod. 60, 1329–1337

34. Bonny, C., Cooker, L. A., and Goldberg, E. (1998) Biol. Reprod. 58, 754–759

35. Muller, C., Yang, R., Beck-von-Peczoz, L., Idos, G., Verbeek, W., and Koeffler, H. P. (1999) J. Biol. Chem. 274, 11220–11228

36. Clare, S. E., Fantz, D. A., Kistler, W. S., and Kistler, M. K. (1997) J. Biol. Chem. 272, 33028–33036

37. Wilkerson, D. C., Wolfe, S. A., and Grimes, S. A. (2002) Biol. Reprod. 67, 1157–1164

38. Santti, H., Mikkonen, L., Hirvonen-Santti, S., Toppari, J., Janne, O. A., and Palvimo, J. J. (2003) Biochim. Biophys. Res. Commun. 308, 139–147

39. Iannello, R. C., Kola, I., and Dahl, H.-H. M. (1993) J. Biol. Chem. 268, 22581–22590

40. Grimes, S. R., Prado, S., and Wolfe, S. A. (2005) J. Cell. Biochem. 94, 317–326

41. Horvath, G. C., Clare, S. E., Kistler, M. K., and Kistler, W. S. (2001) Biol. Reprod. 65, 1074–1081

42. Jethanandani, P., and Goldberg, E. (2001) J. Biol. Chem. 276, 35414–35421

43. Mizuno, K., Goto, M., Masamune, Y., and Nakanishi, Y. (1992) Gene (Amst.) 119, 293–297

44. Reddi, P. P., Shore, A. N., Shapiro, J. A., Anderson, A., Stoler, M. H., Acharya, K. K. (2003) Dev. Biol. 262, 173–182

45. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R. J., Hoffmann, B., Morkel, M., Birchmeier, C., and Hagemeyer, C. (2005) Curr. Biol. 15, 1051–1057

46. Storre, J., Schafer, A., Reichert, N., Barbero, J. L., Hauser, S., Eilers, M., Gaubatz, S. (2005) J. Biol. Chem. 280, 41380–41386

47. Acharya, K. K., Govind, C. K., Shore, A. N., Stoler, M. H., and Reddi, P. P. (2006) Dev. Biol. 295, 781–790

48. Maclean, J. A., and Wilkinson, M. F. (2005) Curr. Top. Dev. Biol. 71, 131–197