CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the *DM1* locus

Galina N. Filippova[1]*, Cortlandt P. Thienes[1]*, Bennett H. Penn[1], Diane H. Cho[1], Ying Jia Hu[1], James M. Moore[1], Todd R. Klesert[1,2], Victor V. Lobanenkov[3] & Stephen J. Tapscott[1,2,4]

Published online: 23 July 2001, DOI: 10.1038/ng570

An expansion of a CTG repeat at the *DM1* locus causes myotonic dystrophy (DM) by altering the expression of the two adjacent genes, *DMPK* and *SIX5*, and through a toxic effect of the repeat-containing RNA. Here we identify two CTCF-binding sites that flank the CTG repeat and form an insulator element between *DMPK* and *SIX5*. Methylation of these sites prevents binding of CTCF, indicating that the *DM1* locus methylation in congenital DM would disrupt insulator function. Furthermore, CTCF-binding sites are associated with CTG/CAG repeats at several other loci. We suggest a general role for CTG/CAG repeats as components of insulator elements at multiple sites in the human genome.

Introduction

Myotonic dystrophy (DM) is the most prevalent form of adult muscular dystrophy, with an incidence of approximately 1 in 8,000 individuals worldwide. DM is a dominantly inherited disease with a number of symptoms, including a myotonic myopathy, cataracts and cardiac-conduction defects. The most common form of DM is caused by the expansion of a CTG repeat in the 3′ noncoding region of *DMPK* at the *DM1* locus on chromosome 19q13.3 (refs. 1–4). Unaffected individuals have fewer than 38 CTG repeats, whereas individuals with DM generally have more than 100 repeats. Larger repeat sizes are associated with more severe disease. The most severe congenital forms of DM have both very large expansions and aberrant methylation at the *DM1* locus[5,6].

CTG expansion causes the complex DM phenotype by several possible mechanisms[7]. *DMPK* transcripts with an expanded CUG repeat are aberrantly processed and result in decreased steady-state levels of DMPK[8–11]. Because DMPK-deficient mice have a cardiac conduction defect similar to the cardiac conduction defect in DM[12], it is probable that deficiency of DMPK contributes to the cardiac disease in humans. Moreover, the CTG repeat expansion results in decreased expression of the adjacent *SIX5* (refs. 13,14) and Six5 deficiency in mice causes cataracts[15,16]. Therefore, the cataracts in DM might be caused by SIX5 deficiency. Finally, a toxic effect of the expanded CUG repeat in the RNA might cause the myotonic myopathy, as mice expressing a transgene with an expanded CUG repeat develop a myopathy that is very similar to DM[17].

The existence of a congenital form of DM indicates that additional molecular mechanisms influence the disease phenotype. Congenital DM is associated with large CTG expansions (usually thousands of repeats[5]) and has at least two molecular features that qualitatively distinguish it from the adult disease. CpGs in the region of the CTG repeat are aberrantly methylated in congenital DM, whereas this region is not methylated in adult-onset disease[6]. Moreover, increased levels of skeletal muscle *DMPK* transcripts have been reported in congenital DM, whereas DMPK levels are decreased in adult DM[9,18]. Therefore, the distinct features of congenital DM might be caused by the effect of methylation on regional gene expression.

Several genes have been identified within a 200-kb region surrounding the *DM1* locus[19]

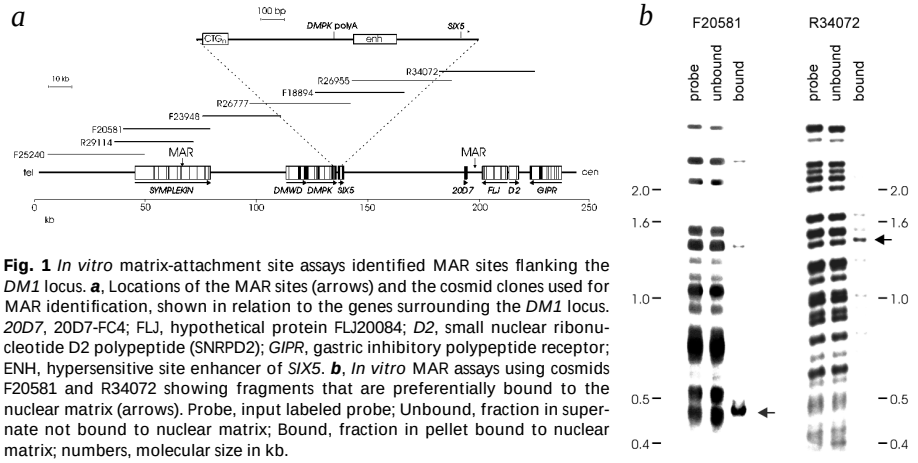


Fig. 1 *In vitro* matrix-attachment site assays identified MAR sites flanking the *DM1* locus. ***a***, Locations of the MAR sites (arrows) and the cosmid clones used for MAR identification, shown in relation to the genes surrounding the *DM1* locus. *20D7*, 20D7-FC4; FLJ, hypothetical protein FLJ20084; *D2*, small nuclear ribonucleotide D2 polypeptide (SNRPD2); *GIPR*, gastric inhibitory polypeptide receptor; ENH, hypersensitive site enhancer of *SIX5*. ***b***, *In vitro* MAR assays using cosmids F20581 and R34072 showing fragments that are preferentially bound to the nuclear matrix (arrows). Probe, input labeled probe; Unbound, fraction in supernate not bound to nuclear matrix; Bound, fraction in pellet bound to nuclear matrix; numbers, molecular size in kb.

*These authors contributed equally to this work. [1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [2]Department of Pathology, University of Washington Medical School, Seattle, Washington 98109, USA [3]Molecular Pathology Section, Laboratory of Immunopathology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]Department of Neurology, University of Washington Medical School, Seattle, Washington 98109, USA. Correspondence should be addressed to S.J.T. (email: stapscot@fhcrc.org) or V.V.L. (e-mail: vlobanenkov@niaid.nih.gov).

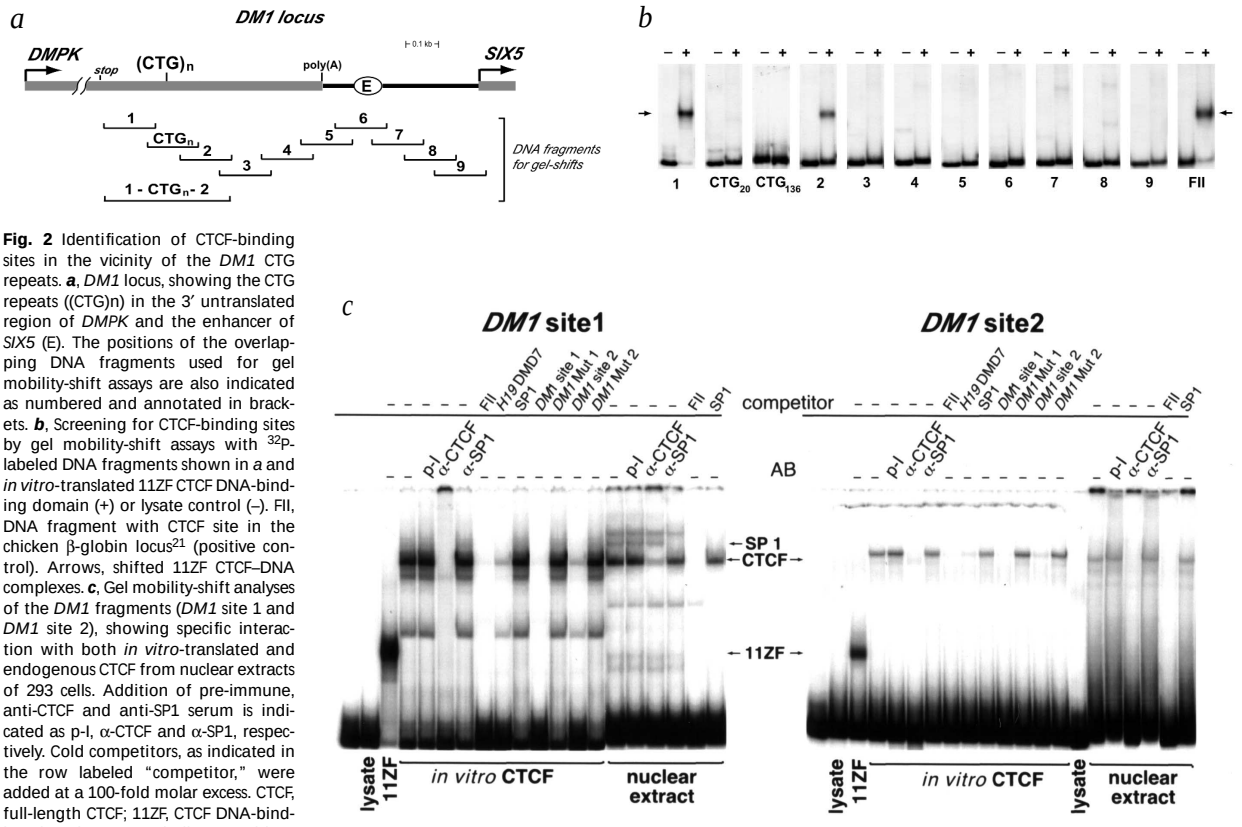


Fig. 2 Identification of CTCF-binding sites in the vicinity of the *DM1* CTG repeats. ***a***, *DM1* locus, showing the CTG repeats ((CTG)n) in the 3′ untranslated region of *DMPK* and the enhancer of *SIX5* (E). The positions of the overlapping DNA fragments used for gel mobility-shift assays are also indicated as numbered and annotated in brackets. ***b***, Screening for CTCF-binding sites by gel mobility-shift assays with ^{32}P-labeled DNA fragments shown in *a* and *in vitro*-translated 11ZF CTCF DNA-binding domain (+) or lysate control (–). FII, DNA fragment with CTCF site in the chicken β-globin locus[21] (positive control). Arrows, shifted 11ZF CTCF–DNA complexes. ***c***, Gel mobility-shift analyses of the *DM1* fragments (*DM1* site 1 and *DM1* site 2), showing specific interaction with both *in vitro*-translated and endogenous CTCF from nuclear extracts of 293 cells. Addition of pre-immune, anti-CTCF and anti-SP1 serum is indicated as p-I, α-CTCF and α-SP1, respectively. Cold competitors, as indicated in the row labeled "competitor," were added at a 100-fold molar excess. CTCF, full-length CTCF; 11ZF, CTCF DNA-binding domain. Arrows indicate positions of shifted protein–DNA complexes.

(Fig 1*a*). Some of these nearly overlap. For example, the polyadenylation site of *DMPK* is less than 300 bp from the transcription start site of *SIX5* and less than 100 bp from the *SIX5* hypersensitive site-enhancer element. Despite their close proximity, these genes have distinct expression patterns, indicating that each can be regulated independently. A principal question is how does the regulatory element for each of these closely spaced genes restrict its activity to the appropriate promoter sequence. One possibility is that boundary elements establish restricted domains of enhancer activity to isolate one gene from another. Two types of boundary elements are thought to delineate regions subject to regulation by *cis*-acting sequences (for review, see ref. 20). Matrix-attachment regions (MARs), also called scaffold-attachment regions, are generally AT-rich sequences that bind the nuclear matrix. These regions form the base of chromatin loops and thereby define domains within which *cis* elements can act. Insulator elements are a second type of boundary element. Insulator elements are identified as sequences that can block an enhancer from regulating a promoter when placed between them, or protect a transgene from silencing by surrounding chromatin. Several well characterized insulators contain binding sites for the zinc-finger protein CTCF and binding of CTCF mediates the inhibition of promoter–enhancer interactions by insulator elements[21–24].

To further investigate the regulation of gene expression in DM, we identified the MAR and insulator elements that flank the *DM1* locus. The MAR elements encompass a domain of approximately 120 kb that contains five identified genes. The CTG repeat is at roughly the center of the domain and is flanked by two CTCF-binding sites. We found that the combination of CTG repeats and CTCF sites at the *DM1* locus forms a functional insulator element located between *DMPK* and *SIX5*. CpG methylation prevents binding of CTCF and might result in loss of insulator activity in congenital DM. Furthermore, we show that CTG repeats at other loci are flanked by CTCF-binding sites. We suggest that CTG/CAG repeats and CTCF-binding sites function together as insulator elements at multiple sites in the human genome.

Results

MARs flank the *DM1* locus and CTCF-binding sites flank the CTG repeat

To identify MARs surrounding the *DM1* locus, we used an *in vitro* MAR assay[25] to screen cosmids covering approximately 200 kb of the region (Fig. 1*a*). We did not identify any MAR elements in the sequences encompassing the CTG repeat or the three adjacent genes (dystrophia myotonica WD motif (*DMWD*), *DMPK* and *SIX5*; data not shown). The nearest MARs were located about 60 kb on either side of the CTG repeat (Fig. 1*a*,*b*); we confirmed this by a second *in vivo* MAR assay[25] (data not shown).

We next sought to determine whether non-MAR insulator elements separate the *SIX5* enhancer from *DMPK*. Recent studies have identified binding sites for CTCF as essential components of vertebrate insulator elements[21–24]. CTCF-binding sites are about 50 bp and variable, possibly because CTCF can use different subsets of its zinc-fingers to recognize diverse DNA sequences[26,27]. We therefore used gel mobility-shift assays to identify CTCF-binding sites. We used 10 overlapping fragments spanning the 1.3-kb region, 150 bp upstream of the CTG repeat to the major transcription start site of *SIX5* (Fig. 2*a*), in gel mobility-shift assays with the *in vitro*-translated DNA-binding domain of CTCF (11ZF, the complete 11 zinc-finger DNA binding domain of CTCF; Fig. 2*b*), which has the same sequence specificity as full-length CTCF[28]. Only two of the fragments interacted with 11ZF CTCF. The two fragments

with CTCF-binding sites were located on either side of the CTG repeat (Fig. 2*b*, probes 1 and 2). Probes containing either a wild-type CTG repeat (CTG_{20}) or an expanded repeat (CTG_{136}), but lacking the flanking CTCF sites, did not bind the 11ZF protein.

To determine whether the two 11ZF-binding regions also bound full-length endogenous CTCF, we used gel mobility-shift assays with nuclear extracts from several CTCF-expressing human cell types, including 293, K562 and primary human fibroblasts (Fig. 2*c* and data not shown). CTCF is ubiquitously expressed[29], and we confirmed that these cell types express CTCF by western blot analysis (data not shown). With both fragments, *DM1* site 1 and *DM1* site 2, the nuclear extracts produced a protein–DNA complex with the same mobility as that generated by *in vitro*-translated, full-length CTCF. These complexes contained CTCF, as they were specifically disrupted by anti-CTCF antiserum but not by pre-immune or anti-Sp1 (SP1 transcription factor) antiserum. The binding of CTCF to each fragment was also specifically diminished by competition with an excess of unlabeled fragments containing CTCF-binding sites previously characterized in sequences from the chicken β-globin gene (FII) and the mouse *H19* gene (DMD7), as well as by each other (*DM1* site 1 and site 2), but not by a competitor fragment containing an

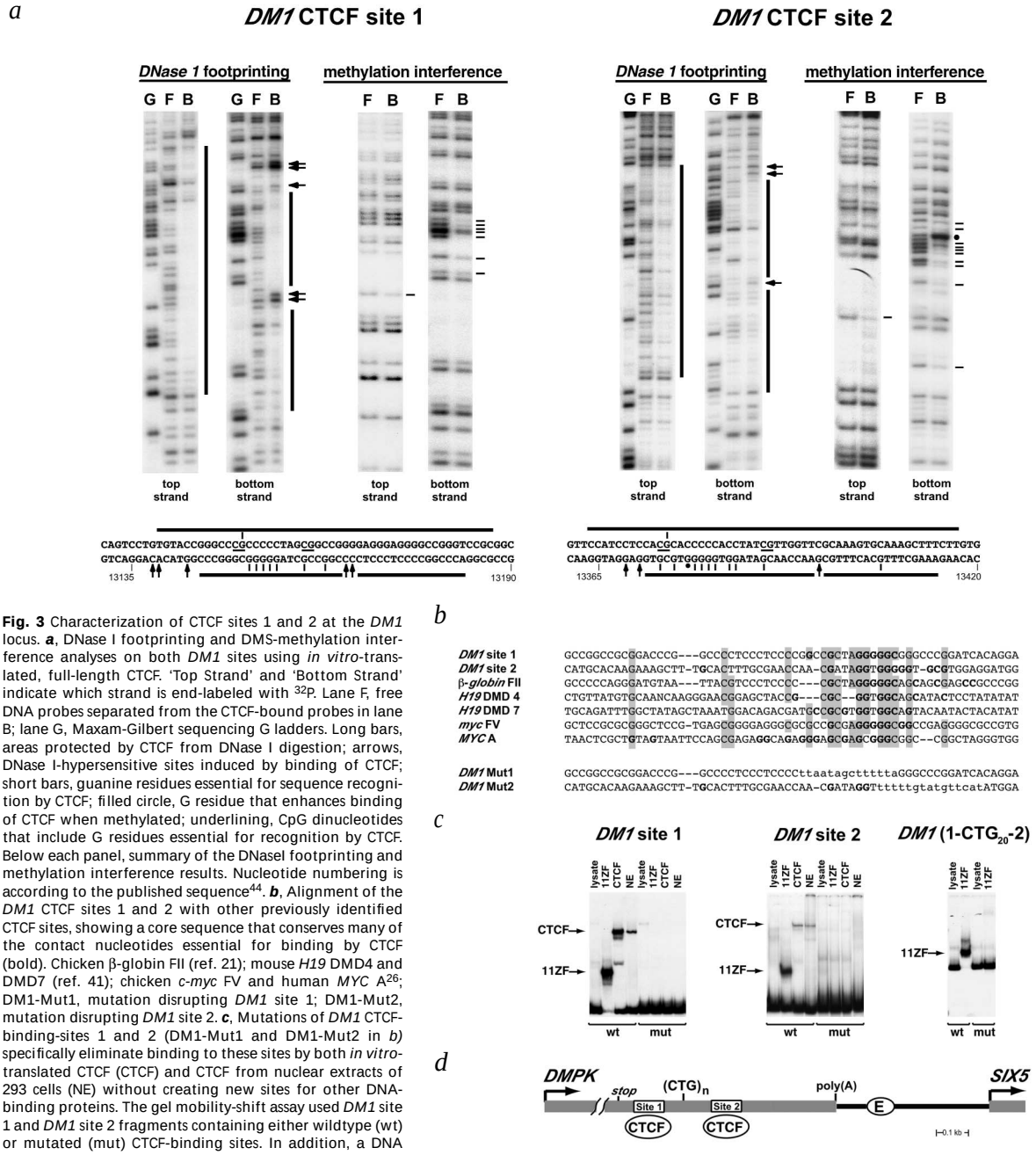


Fig. 3 Characterization of CTCF sites 1 and 2 at the *DM1* locus. ***a***, DNase I footprinting and DMS-methylation interference analyses on both *DM1* sites using *in vitro*-translated, full-length CTCF. 'Top Strand' and 'Bottom Strand' indicate which strand is end-labeled with ^{32}P. Lane F, free DNA probes separated from the CTCF-bound probes in lane B; lane G, Maxam-Gilbert sequencing G ladders. Long bars, areas protected by CTCF from DNase I digestion; arrows, DNase I-hypersensitive sites induced by binding of CTCF; short bars, guanine residues essential for sequence recognition by CTCF; filled circle, G residue that enhances binding of CTCF when methylated; underlining, CpG dinucleotides that include G residues essential for recognition by CTCF. Below each panel, summary of the DNaseI footprinting and methylation interference results. Nucleotide numbering is according to the published sequence[44]. ***b***, Alignment of the *DM1* CTCF sites 1 and 2 with other previously identified CTCF sites, showing a core sequence that conserves many of the contact nucleotides essential for binding by CTCF (bold). Chicken β-globin FII (ref. 21); mouse *H19* DMD4 and DMD7 (ref. 41); chicken *c-myc* FV and human *MYC* A[26]; DM1-Mut1, mutation disrupting *DM1* site 1; DM1-Mut2, mutation disrupting *DM1* site 2. ***c***, Mutations of *DM1* CTCF-binding-sites 1 and 2 (DM1-Mut1 and DM1-Mut2 in *b)* specifically eliminate binding to these sites by both *in vitro*-translated CTCF (CTCF) and CTCF from nuclear extracts of 293 cells (NE) without creating new sites for other DNA-binding proteins. The gel mobility-shift assay used *DM1* site 1 and *DM1* site 2 fragments containing either wildtype (wt) or mutated (mut) CTCF-binding sites. In addition, a DNA fragment ~400 bp long of the *DM1* locus encompassing the CTG repeats and the two CTCF-binding sites (DM1(1-CTG_{20}-2)), either both wild type (wt) or both mutated (mut), shows that mutation of the two identified CTCF-binding sites eliminates binding of CTCF to the entire region surrounding the CTG repeat. ***d***, *DM1* locus, showing positions of CTCF sites 1 and 2 flanking the CTG repeats and separated by approximately 176 bp.

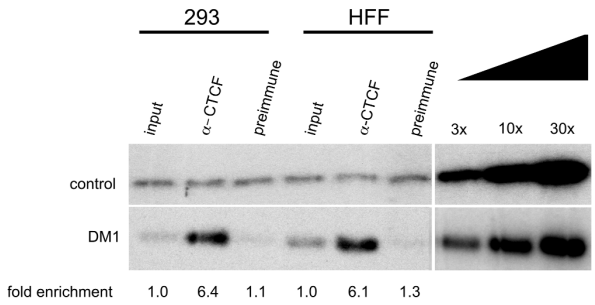


Fig. 4 Binding of CTCF occurs at the *DM1* locus *in vivo*. After treatment of 293 cells and human fibroblasts (HFF) with formaldehyde, lysates were subjected to chromatin immunoprecipitation using either anti-CTCF or pre-immune sera. Co-immunoprecipitating sequences were detected using quantitative PCR, and the 'fold enrichment' of co-precipitating *DM1* sequence was determined relative to a control region (Control) of the *MYC* locus lacking CTCF sites. The 'fold enrichment' represents the average of triplicate PCR reactions. Immunoprecipitations were done twice in each cell type with similar results. The linearity of the PCR was confirmed by titrating increasing amounts (3x, 10x and 30x) of input template into the reaction.

Sp1-binding site. Thus, two regions on either side of the CTG repeat specifically interacted with endogenous CTCF in nuclear extracts. Moreover, CTCF was the most abundant nuclear protein binding to these regions in our gel mobility-shift assays.

Both DNaseI footprinting and methylation interference assays identified a single CTCF-binding site within each fragment. DNaseI footprinting showed that CTCF protects a region of about 50 bp on both strands of each fragment (Fig. 3*a*). In addition, several nuclease-hypersensitive sites were induced by binding of CTCF. Methylation interference identified 'contact' guanine nucleotides whose methylation prevents binding of CTCF. Consistent with descriptions of some other CTCF-binding sites[23,24], most of the CTCF-contact nucleotides and CTCF-induced hypersensitive sites were located on a single DNA strand. Comparison of CTCF-binding sites in each fragment showed a conserved core sequence (Fig. 3*b*). Mutations within this core sequence eliminated binding of CTCF to each fragment (*DM1* site 1 and *DM1* site 2; Fig. 3*c*) and to a larger fragment that encompassed both sites and the intervening CTG repeats (*DM1*(1-CTG_{20}-2), confirming the identity of the two CTCF-binding sites.

In vivo binding of CTCF and nucleosome positioning

In vitro gel mobility-shift assays demonstrated that the CTCF insulator protein was able to bind to sequences adjacent to the CTG repeat at the *DM1* locus. We next used chromatin immunoprecipitation to assess binding of CTCF to these sites in native chromatin *in vivo*. After cross-linking DNA–protein complexes *in vivo* with formaldehyde, we immunoprecipitated sheared chromatin with either anti-CTCF antiserum or control preimmune serum. Quantitative PCR analysis demonstrated that the region of the *DM1* locus containing the CTCF-binding sites was greatly enriched in anti-CTCF immunoprecipitates (Fig. 4). In contrast, a region of *MYC* lacking CTCF-binding sites was not enriched, and we found no enrichment of *DM1* with preimmune serum. The enrichment of the *DM1* site in the anti-CTCF immunoprecipitates was similar in magnitude to the enrichment of CTCF-binding regions in the *H19* and *MYC* genes (data not shown). Therefore, CTCF is bound adjacent to the CTG repeat at the *DM1* locus in human cells *in vivo*.

In vitro studies have shown that CTG repeats can efficiently position nucleosomes[30–32]. Because approximately 176 bp separates the two *CTCF* footprints (based on an allele with 12 CTG repeats), we sought to determine whether the *CTCF* sites flanked a nucleosome. Standard micrococcal nuclease studies showed a region of approximately 150 nucleotides encompassing the CTG repeats that was protected from digestion in human fibroblasts (Fig. 5 and data not shown), consistent with the existence of a nucleosome centered on the wildtype repeat. A nucleosome in this position would place the two CTCF-binding sites in an inter-nucleosomal region and potentially bring them in proximity, perhaps allowing the interaction of CTCF proteins bound to the two sites.

DM1 insulator function depends on CTCF-binding sites that are methylation-sensitive

Binding of CTCF is a feature of insulator elements at many loci, including the chicken β-globin and human *H19* loci[21–24]. To determine whether the *DM1* region, encompassing the two CTCF-binding sites and the CTG repeat, could function as an insulator element, we compared the insulator activity of this sequence to a well characterized CTCF–insulator element, HS4 (hypersensitive site 4) from the chicken β-globin locus[21]. We used the same constructs and assays as were used before to identify and characterize the HS4 insulator[21,33], which rely on positioning the putative insulator sequence between an enhancer and γ-neomycin phosphotransferase (*neo*; Fig. 6*a*). Sequences with insulator activity result in a decreased number of G418-resistant colonies after transfection into K562 cells because of lower expression of *neo*. When we placed the *DM1* sequence containing

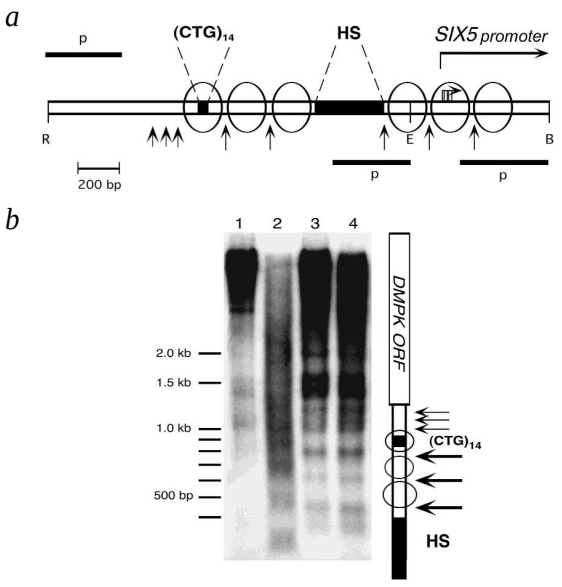


Fig. 5 Nucleosome positioning at the *DM1* locus CTG repeat. ***a***, Schematic representation of results from several micrococcal nuclease (MNase) experiments that revealed a protection consistent with the existence of a nucleosome covering the wildtype CTG repeat. Intact nuclei from wildtype human fibroblasts were digested with varying concentrations of MNase and the DNA was analyzed by digestion with restriction enzymes followed by Southern blot with hybridization to the indicated probes (p). Arrows, regions of nuclease access; spheres, inferred position of nucleosomes. ***b***, Southern blot analysis from one experiment using a probe adjacent to the *Eco*RI site. Lane 1 (control), no MNase added; lane 2 (control), digested with MNase but not *Eco*RI; lanes 3 and 4, experimental lanes showing protection consistent with a nucleosome positioned at the wildtype CTG repeat. HS, hypersensitive site; p, probe; B, *Bam*HI; R, *Rsa*I; E, *Eco*RI; ORF, open reading frame.

the CTCF sites flanking 20 CTG repeats between the mouse HS2 enhancer and *neo* (pJC DM1), the number of G418-resistant colonies was decreased approximately three-fold compared with that of the parent vector (Fig. 6*b*). The *DM1* insulator activity was equal to that of the HS4 insulator element and was independent of orientation relative to the promoter and enhancer. Replacement of the *DM1* sequence with a fragment from λ phage (pJC-λ) did not have a similar effect, even though it was nearly twice the length of the *DM1* sequence. Mutation of both CTCF-binding sites in the *DM1* sequence resulted in almost complete loss of insulator activity (Fig. 6*b*). Individual CTCF-binding-site mutations showed that CTCF-binding-site 1 was sufficient to maintain full insulator activity, whereas CTCF-binding-site 2 only partially contributed to insulator activity (Fig. 6*c*), consistent with *in vitro* binding assays that showed stronger binding of CTCF to site 1 (Fig. 2*c* and data not shown).

These data demonstrate that a region of the *DM1* locus that separates the *SIX5* enhancer from *DMPK* and *DMWD* was able to function as an insulator element, and that an intact CTCF-binding site was necessary for its insulator activity. In congenital DM, the region of the *DM1* locus encompassing the CTCF-binding sites is methylated at CpG dinucleotides[6]. Therefore, we tested the binding of CTCF to *DM1* locus probes methylated *in vitro* at CpG dinucleotides (Fig. 7). CpG methylation substantially reduced binding of CTCF, whereas binding of Sp1 to the *DM1* site 1 probe was not altered by methylation. Therefore, interference with the binding of CTCF by CpG methylation at the *DM1* locus might contribute to the distinct phenotype of congenital DM.

The CTG repeat contributes to *DM1* insulator activity

Unstable CTG/CAG repeats have been associated with many human diseases, yet their sequence has been preserved in human evolution, indicating a possible function for these repeats in the genome. To determine whether the CTG repeat contributed to the insulator activity at the *DM1* locus, we replaced the CTG sequence with a fragment of λ phage DNA of equal size, leaving the flanking sequence, including the CTCF-binding sites, intact. Replacement of the CTG repeats resulted in a small but significant reduction of insulator activity (Fig. 6*d*). Pooling the results from 18 independent transfections yielded an average of 50.2±2.9 colonies with the wildtype *DM1* compared with 64.6±5.1 colonies with the CTG replacement ($P<0.018$). Therefore, the CTG repeats could contribute 25–30% more insulator function in the context of paired CTCF-binding sites. In contrast, expansion of the CTG repeat from 20 to 149 repeats did not diminish insulator function (Fig. 6*c*).

CTG/CAG repeats are associated with CTCF sites at multiple loci

To determine whether CTCF sites are associated with CTG/CAG repeats at other loci in the genome, we screened seven additional repeat containing loci for flanking CTCF sites. Gel mobility-shift

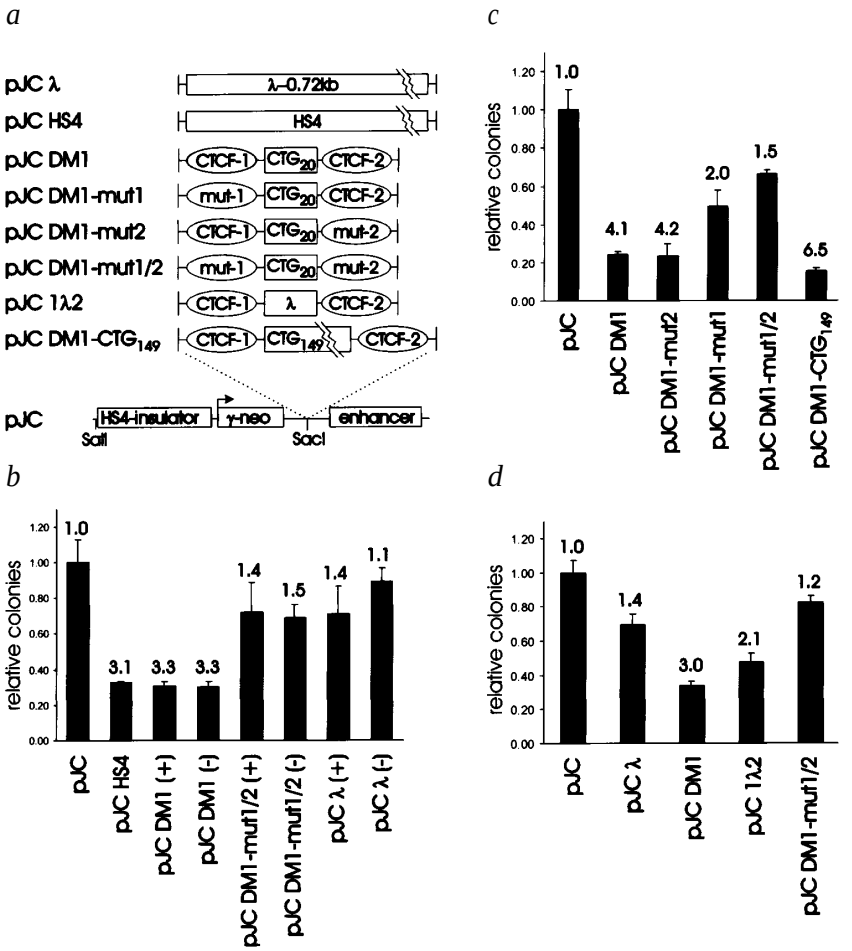


Fig. 6 The CTG repeats and surrounding CTCF-binding sites form a functional insulator. ***a***, Sequences to be tested for insulator activity were cloned into a *Sac*I site in the pJC vector. The pJC vector has been used before to characterize insulator activity[21,33] and contains the mouse hemoglobin beta chain complex (*Hbb*) HS2 enhancer and *neo* driven by the mouse γ-globin promoter with an adjacent insulator element from chicken β-globin gene. The indicated fragments were cloned into the *Sac*I site between the enhancer and promoter to test for insulator activity in a standard colony formation assay and consist of: HS4, the 1.2 kb chicken β-globin insulator element (same as pJC5-4[33]); λ, a 720-bp fragment from λ phage to represent a neutral 'stuffer' sequence; DM1, a 404-bp fragment from the *DM1* locus encompassing both CTCF-binding sites and the CTG_{20} repeat; DM1-mut1/2, like *DM1* but with the mut1 and mut2 mutations shown in Fig. 3*b*; DM1-mut1, like *DM1* but with mut1 mutation; DM1-mut2, same as *DM1* but with a mut2 mutation; 1-λ-2, replacement of the 20 CTG repeats with a 60-bp fragment of the λ 'stuffer' sequence; CTG149, identical to *DM1* but with 149 CTG repeats. ***b***, Colony assays of K562 cells selected in G418 demonstrate that the *DM1* fragment has insulator activity similar to that of the HS4 insulator and that mutation of the CTCF-binding sites eliminates *DM1* insulator activity. The DM1, DM1-mut1/2 and λ fragments were tested in forward (+) and reverse (–) orientations (fragments in *a* are shown in the forward orientation). ***c***, Colony assays showing that mutation of CTCF-binding-site 1 (DM1-mut1) diminishes insulator activity, whereas mutation of site 2 (DM1-mut2) does not diminish insulator activity. In addition, a CTG_{149} expansion does not reduce *DM1* insulator activity. ***d***, Replacement of the CTG repeats with λ DNA demonstrates a slight reduction in insulator activity. Each bar represents the average of three or more independent transfections. Numbers represent fold suppression relative to pJC, and bars represent standard error.

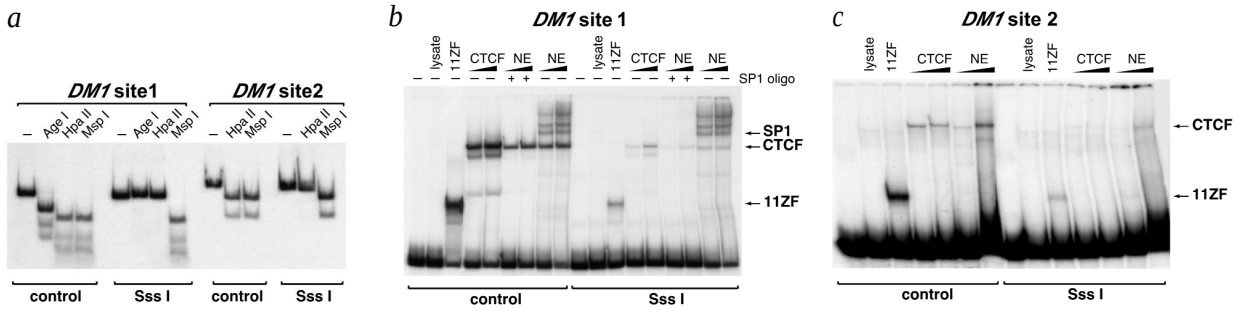


Fig. 7 CpG methylation prevents binding of CTCF to *DM1* site 1 and *DM1* site 2. ***a***, *Sss*I was used to methylate cytosine residues in CpG dinucleotides on the *DM1* site 1 and site 2 probes. Digestion of control or *Sss*I-methylated probe with methylation-sensitive nucleases *Age*I and *Hpa*II or methylation-insensitive *Msp*I shows the complete methylation of the probe after *Sss*I treatment. ***b***,***c***, Control and *Sss*I-methylated DM1 probes were used in gel mobility-shift assays with *in vitro*-translated 11ZF CTCF and with increasing amounts of full-length CTCF or endogenous CTCF from nuclear extracts (NE). The locations of complexes containing 11ZF, full-length CTCF and Sp1 are shown (Fig. 2*c*, confirmation of band identity). CpG methylation substantially diminishes the interaction of CTCF with either the site 1 or the site 2 probe, whereas SP1 binding remains unaffected on the methylated site 1 probe. Quantitation with the phosphorimager shows a 10- to 15-fold reduction in the binding of CTCF to each methylated probe compared with that of the unmethylated probes. Sp1 competitor oligonucleotide: +, present; –, absent.

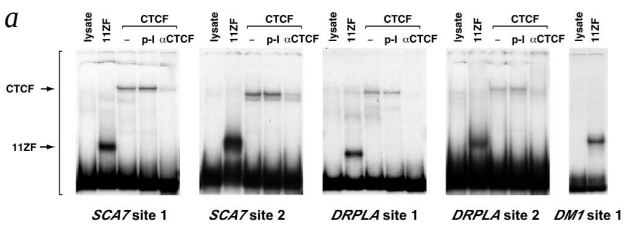
assays identified two CTCF-binding sites flanking the CTG/CAG repeat at the spinocerebellar ataxia 7 (*SCA7*) and dentatorubral-pallidoluysian atrophy (*DRPLA*) loci (Fig. 8*a*), a single CTCF site adjacent to the repeat at the *SCA2* and huntingtin (Huntington disease; *HD*) loci (data not shown) and no binding of CTCF adjacent to the repeat at the *SCA1*, *SCA6* or androgen receptor (*AR*) loci. Therefore, five of the eight repeat-containing loci examined, including *DM1*, had at least one CTG-associated CTCF-binding site. All five loci had a binding site positioned upstream of the CTG sequence, which was the location of the *DM1* site 1 CTCF-binding site (Fig . 8*b*). Because the single *DM1* site 1 CTCF-binding site together with the CTG/CAG repeat was sufficient for full insulator activity (Fig. 6*c*), it is likely that the repeat loci associated with a single upstream site might also function as insulator elements.

Discussion

CTCF was originally identified as a zinc-finger protein that binds the chicken and mammalian *MYC* promoter[26,34–36]. CTCF binds to diverse DNA sequences by using different combinations of its individual zinc-fingers[26,27,37] and mediates both transcriptional repression and activation[26–28,37]. CTCF is also involved in the enhancer-blocking function of vertebrate insulators[21,22,24,38,39]. Here we have described several previously unrecognized features of boundary elements and CTG repeats at the *DM1* locus: (1) The *DM1* locus was positioned roughly at the center of a domain of about 120 kb flanked by MARs. (2) The CTG repeat at the *DM1* locus was flanked by two CTCF-binding sites. (3) CTCF bound to the *DM1* locus *in vivo*. (4) Micrococcal nuclease studies indicated that a nucleosome was positioned at the CTG repeat *in vivo*, placing the two CTCF-binding sites in inter-nucleosomal DNA. (5) The combination of CTCF-binding sites and CTG repeats formed an efficient insulator unit. (6) The hypermethylation at the *DM1* locus in congenital DM was incompatible with binding of CTCF and CTCF-mediated insulator activity. (7) CTCF sites were associated with CAG/CTG repeats at multiple loci, indicating that CAG/CTG repeats might have been selected for their association with CTCF-binding sites in establishing an efficient insulator element at multiple sites in the human genome.

Many diseases are associated with instability of CTG/CAG repeats[40]. A fundamental question in human biology is why have CTG/CAG repetitive elements evolved and been maintained in our population. In many cases, such as at the *DRPLA* or *HD* loci, the CAG repeat encodes a polyglutamine tract in the protein product. However, glutamine can be encoded by either CAG or CAA and, therefore, selection for a sequence of glutamines would not necessitate a perfect CAG repeat. Furthermore, as at the *DM1* locus, CTG repeats are found in noncoding regions. Therefore, if there is evolutionary pressure to establish and maintain these repeats, it must not depend simply on codon usage.

Our study supports the conclusion that the CTG/CAG repeats form a functional component of an insulator element, as replacing the repeats with a non-repeat containing sequence of similar size resulted in a decrease in insulator activity. *In vitro* studies have shown that CTG/CAG repeats act as strong nucleosome positioning elements[30–32]. Indeed, our micrococcal nuclease studies are consistent with a nucleosome positioned on the wild-type repeats at the *DM1* locus in human fibroblasts *in vivo* and it is possible, but remains to be proven, that the CTG/CAG repeat might have a similar nucleosome positioning capacity *in vivo*. Nucleosomes at the insulator element in the insulin-like growth factor 2 (*IGF2*)/*H19* locus are phased to position the CTCF-binding sites in inter-nucleosomal regions[24,41], indicating that the nucleosome positioning might affect the binding of CTCF and,



b

	site 1 upstream		site 2 downstream
DM1	+	CTG_n	+
DRPLA	+	CTG_n	+
SCA7	+	CTG_n	+
HD	+	CTG_n	-
SCA2	+	CTG_n	-

Fig . 8 CTCF-binding sites are associated with the CTG/CAG repeat at other loci. ***a***, Gel retardation assays with probes (see Web Table A) from regions flanking the CTG/CAG repeat at the *SCA7* and *DRPLA* loci using *in vitro*-translated 11ZF or full-length CTCF with either no antisera (–), pre-immune sera (p-I) or anti-CTCF antisera (α-CTCF). *DM1* site 1, positive control. ***b***, Location of the CTCF-binding sites relative to the CTG repeat.

ultimately, insulator function. Therefore, one function of the CTG repeat at the *DM1* locus might be to establish a local nucleosome positioning that would place the CTCF-binding sites in an accessible position; however, additional studies will be needed to test this. Although replacement of the CTG repeat with a portion of λ DNA of equal size demonstrated that the CTG repeat contributed to insulator activity, this experiment did not address mechanism because the 'neutral' sequence might not preclude appropriate nucleosome positioning. Accordingly, it will be useful to determine whether a sequence that prevents nucleosome positioning, such as CCG[42], further diminishes insulator activity, or whether the spacing between the repeat and the CTCF site is an important variable.

Loss of binding of CTCF at the *DM1* locus might contribute to the distinct phenotype of congenital myotonic dystrophy. Methylation-sensitive restriction endonucleases were previously used to demonstrate that the *DM1* locus is methylated in congenital DM, but not in adult-onset DM or in unaffected individuals[6]. Individuals with congenital DM had complete CpG methylation of the assayable sites in the regions that we have now identified as the CTCF-binding sites. Previous studies of binding of CTCF to the insulator elements at the *IGF2/H19* locus showed that CpG methylation inhibits binding[22–24]. Here we have demonstrated that CpG methylation similarly inhibited binding of CTCF to sites at the *DM1* locus. Further, methylation at the *DM1* locus resulted in the loss of an *in vivo* protein footprint at the sequence we have now identified as a CTCF-binding site (*DM1* site 1)[6]. Although the site was originally identified as encompassing a consensus Sp1-binding site, our study has shown that CTCF was the most abundant nuclear factor binding to that site, and that binding of CTCF but not Sp1 was inhibited by CpG methylation at this site. Therefore, the loss of the *in vivo* protein footprint in congenital DM demonstrated before[6] was most likely because of the inhibition of the binding of CTCF to the CpG methylated site, which would result in a loss of insulator function.

The molecular and developmental consequences of interfering with the binding of CTCF at the *DM1* locus remain to be determined. The insulator is located between *DMPK* and the *SIX5* enhancer, and there are several other genes within the larger region bounded by MAR elements. One simple prediction is that elimination of binding of CTCF at the *DM1* locus alters the regulation of local gene expression, possibly permitting the *SIX5* enhancer to interact with the *DMPK* promoter. Although a formal test of this prediction will need to be done in a mouse model, it is important to reconsider the limited amount of published information on gene expression in congenital DM. One finding was an increased amount of *DMPK* RNA in skeletal muscle in congenital DM[18], whereas many studies in adult onset DM found either normal or reduced amounts of *DMPK* RNA[8,9,11]. Given our results, loss of insulator activity between the *SIX5* enhancer and the *DMPK* promoter because of the methylation of the CTCF-binding site in congenital DM may result in higher levels of *DMPK* expression. Higher levels of *DMPK* expression might be expected to cause a more severe disease, as part of the pathology of DM is caused by an accumulation of the CUG-containing RNA in the nucleus.

Methods

Matrix-attachment assays. We obtained cosmid clones from Laurence Livermore National Laboratory. We purified DNA for probes first by maxiprep (Qiagen), then by a cesium chloride gradient. We did matrix-attachment assays essentially as described before[25]. We washed 2.5×10^8 Raji cells at 900*g* for 10 min at RT three times in 50-ml volumes of 5 mM Tris, pH 7.5, 20 mM KCl, 0.125 mM spermidine, 0.05 mM spermine, 0.5mM EDTA, 1% thioglycol, 0.7 μg/ml pepstatin and 0.2 mM phenylmethyl sulfonyl fluoride. We lysed cells at 4 ºC in 12 ml of the same buffer containing 0.1% digitonin (Fluka) using 15 strokes of a 15-ml Dounce homogenizer with a B type pestle. We washed the nuclei at 900*g* for 10 min at 4 ºC twice in 25-ml volumes of the buffer described above plus digitonin and stored them for 1 month at –20 ºC in the same buffer containing digitonin and 50% glycerol. We washed the nuclei for 30 s at 2,000*g* twice in 1-ml volumes of the buffer described above plus digitonin in a Capsulefuge (Tomy), and stabilized them for 20 min at 37 ºC in 50 μl of the same buffer. For *in vitro* MAR assays, we extracted the nuclei at RT with 1 ml of 5 mM HEPES–NaOH, pH 7.4, 2 mM EDTA–KOH, 2 mM KCl, 0.25 mM spermidine, 2 M NaCl, 0.1 % digitonin, 0.7 μg/ml pepstatin and 0.2 mM phenylmethyl sulfonyl fluoride. We washed the extracted nuclei six times in a microfuge with digestion buffer (20 mM Tris, pH 7.4, 70 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, 0.125 mM spermidine, 0.05 mM spermine, 0.1% digitonin, 0.7 μg/ml pepstatin and 0.2 mM phenylmethyl sulfonyl fluoride). We slowly rotated the washed scaffolds with 200 units each of *Eco*RI, *Hind*III and *Xba*I in digestion buffer. After 6–8 h, we added 40 mM EDTA and 1 fmol end-labeled probe and continued the incubation for 10–12 h. We pelleted matrix-associated DNA in a microfuge for 1 min. We accomplished digestion of supernatant and pellet fractions for 3 h at 50 ºC by adding 0.3 M NaCl, 0.5% SDS and 1 mg/ml freshly made protease K, then extracted them with phenol and precipitated them with ethanol. We separated matrix-associated and unassociated DNA by electrophoresis through an agarose gel. *In vivo* assays were like the *in vitro* assays except that we used 25 mM 3,5-diiodosalicylic acid lithium salt (Sigma) instead of NaCl in the extraction buffer, continued digestion for 12–16 h and did not add exogenous DNA. We separated equal amounts of matrix-associated and unassociated DNA by electrophoresis and then carried out Southern blot analysis with probes for the region of interest.

Chromatin immunoprecipitation. We added formaldehyde fixation buffer (10× buffer: 50 mM HEPES, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 11 mM NaCl and 11% formaldehyde) to the media of cultured human fibroblasts or 293 cells for 10 min and quenched the reaction with 125 mM glycine. We washed cells sequentially in PBS, pH 7.4, buffer 1 (10 mM Tris, pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100 and 10 mM sodium butyrate) and buffer 2 (buffer 1 without Triton X-100 and with 0.2 M NaCl). We pelleted cells, lysed them (in 40 mM Tris-HCl pH 8.0, 1% Triton X-100, 4 mM EDTA, 300 mM NaCl, 10 mM sodium butyrate and 'mini-protease inhibitor cocktail' (Boehringer Mannheim)), and then sonicated them for 2 min. After centrifugation, we pre-cleared the supernate with protein A Dynabeads (Dynal) and mixed 750 μg protein with 5 μl CTCF antiserum (Upstate Biotechnology), incubated this overnight at 4 ºC with anti-CTCF (UBI), washed it five times following the manufacturer's protocol and eluted it with 1% SDS and 0.1 M $NaHCO_3$. We heated samples to 65 ºC for 4 h, extracted them with phenol:chloroform twice, ethanol-precipitated them and used 2 ng DNA in PCR reactions with [α-^{32}P]dCTP. We separated PCR products by PAGE and quantified them using a phosphorimager. We used the following primers: DM1-1A, 5′–CTGCCAGTTCACAACCGCTCCGAG–3′; DM1-1B-, 5′–GCA GCATTCCCGGCTACAAGGACCCTTC–3′; MYC+ (control), 5′–GCTGCC TCCACAAGCTCTCCACTTGCC–3′; MYC– (control), 5′–GCTGGAGGC CTGTGGTTAGCCCTGAGATGTGTC–3′.

Micrococcal nuclease assays. We did micrococcal nuclease experiments as described before[43]. We isolated nuclei from confluent human fibroblast cultures as follows: We collected cells with 0.1% trypsin and 2 mM EDTA and centrifuged them at 200*g* for 5 min. We washed pellets in ice-cold reticulocyte suspension buffer (RSB: 10 mM Tris, pH 7.4, 10 mM NaCl and 5 mM $MgCl_2$), resuspended them gently in ice-cold lysis buffer (RSB with 0.1% Nonidet-P40) and incubated them on ice for 10 min. We pelleted nuclei at 200*g* for 3 min, resuspended them again in lysis buffer and transferred them to 1.5-ml microfuge tubes (3×10^7 nuclei/tube). We pelleted nuclei once more at 200*g* for 3 min, resuspended them in 200 μl ice-cold RSB supplemented with 0.1 mM $CaCl_2$ and 5–50 units/ml micrococcal nuclease (Pharmacia), then incubated them for 15 min at 37 ºC to allow digestion. We terminated digestions by adding 250 μl STOP solution (25 mM EDTA, 10 mM Tris-HCl pH 8.0, 1% SDS and 2 mg/ml proteinase K) and incubated them overnight at 37 ºC. We purified genomic DNA, digested it to completion with various restriction enzymes (NEB), size-fractionated it by electrophoresis through 1.6% agarose TAE gels and transferred to it Hybond-N membranes (Amersham). After cross-linking the samples with ultraviolet radiation, we incubated them with gel-purified, random-primed ^{32}P-labeled DNA

probes overnight at 65 °C in FBI buffer (10% PEG, 7% SDS and 1.5X SSC) containing 100 μg/ml salmon sperm DNA. We used 0.1× SSC and 0.1% SDS at 60 °C for washes.

Enhancer-blocking constructs. Plasmid pJC HS4 (pJC5-4) was a gift from G. Felsenfeld. We derived pJC from pJC HS4 by deletion of the *Sac*I HS4 element. For pJC DM1, we generated a 404-bp fragment by PCR amplification of cosmid F18894 with oligonucleotides DM1-1A and DM1-2B and inserted it into the *Sac*I site of pJC. For pJC-λ, we inserted a 720-bp fragment of λ-CI857 from positions 25,159–25,880 of GenBank accession number NC_001416 into the *Sac*I site of pJC. We generated mutations of the *DM1* CTCF sites 1 and 2 with the QuikChange protocol (Stratagene) using complementary oligonucleotides from bases 13,136–13,179 and 13,347–13,406 (GenBank accession number L08835), respectively, incorporating the mutated sequences shown in Fig. 3*b*. We constructed pJC CTG_{149} by the replacing a CTG_{20} fragment (positions 13229-13262) in pJC DM1 with a *Bsa*I–*Bbs*I CTG_{149} fragment from pACT (obtained from C. Thornton) with blunt ends created with T4 polymerase. We constructed pJC 1-λ-2 using a fragment of λ-CI857 from positions 25,159–25,211 inserted into the same position as in pJC CTG_{149}. We confirmed the identity of all clones by sequencing.

Tissue culture and colony assays. We maintained K562 cells at a density between 0.4×10^6 and 1×10^6 cells/ml in Iscove's modified Dulbecco's medium with 10% calf serum (Hyclone). We isolated enhancer-blocking constructs by the Endofree Maxi procedure (Qiagen) except for the experiments in Fig. 6*c*, for which we used minipreps (Qiagen). We cleaved constructs with *Sal*I, then extracted them with phenol and precipitated them. We did colony assays essentially as described before[21,33]. We washed K562 cells at an initial density of 0.5×10^6–0.9×10^6 cells/ml in 50 ml cold PBS and resuspended them at a density of 2×10^7 cells/ml in PBS. We added 0.2 μg linearized constructs to 0.5 ml washed cells and incubated them on ice for 10 min. We electroporated the cells at 200 V and 960 μF with a Gene Pulser (BioRad), incubated them on ice for 10 min and cultured them in 20 ml Iscove's media. After 20–24 h of culture, we plated 10 ml of cells in 15-cm dishes with 0.3% agar (Difco) and 570 μg/ml G418, and counted colonies after 17 d of culture.

In vitro translation and nuclear extracts. We translated full-length human CTCF and the 11ZF CTCF DNA-binding domain *in vitro* from pCITE-7.1 and pCITE-11ZF, respectively[37], using the TnT reticulocyte lysate-coupled *in vitro* transcription–translation system (Promega). We prepared nuclear extracts from the human cells according to a published protocol[26].

Gel mobility-shift assay. We amplified 10 overlapping DNA fragments, spanning the 1.3-kb region of the *DM1* locus (Fig. 2*a*), by PCR using primers labeled at the 5′ end with [γ-^{32}P]ATP. Positions of fragments 1–4 and CTGn on a described DNA sequence[44] (GenBank accession number L08835) were: 1, 13,087–13,233; CTGn, 13,179–13,341; 2, 13,282–13,485; 3, 13,409–13,634; and 4, 13,561–13,773. Positions of fragments 5–9 on a described DNA sequence[45] (GenBank accession number X84813) are: 5, 236–387; 6, 316–498; 7, 421–622; 8, 544–775; and 9, 688–895. We amplified fragments containing sequences adjacent to the CAG/CTG repeats at other genomic loci from human genomic DNA using primers for *DRPLA, SCA7, SCA1, SCA6, HD, AR,* and *SCA2* (Web Table A). We PCR amplified DNA fragments with the previously identified CTCF-binding site FII at the chicken β-globin locus[21] and the DMD7 site at the mouse *H19* locus[41] from the genomic DNA. We gel-purified the 5′-end-labeled DNA fragments and used them for gel mobility-shift assays with 1–10 μl of the *in vitro*-translated proteins or nuclear extracts as described before[26]. For binding reactions, we used buffer containing standard PBS with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM DTT, 0.1% NP40 and 10% glycerol in the presence poly(dI-dC). As Sp1-like and 'G-string'-binding factors can bind to short GC-rich segments within the extended CTCF site and can obscure the binding of CTCF[35], we did gel mobility-shifts with nuclear extracts in the presence of unlabeled double-stranded oligonucleotides containing an Sp1-binding site, unless otherwise specified. For 'supershift' assays, we added to the binding reaction an antibody against the C-terminal domain of CTCF (Upstate Biotechnology) or against Sp1 (Santa Cruz Biotechnology). We incubated reaction mixtures with a final volume of 20 μl for 30 min at RT and then analyzed them by 5% nondenaturing PAGE run in 0.5xTris–borate–EDTA buffer as described before[26].

DNase I footprinting and methylation interference analyses. We performed DNase I footprinting and methylation interference analyses with *in vitro*-translated, full-length CTCF and the *DM1* fragments 1 and 2 positive for binding of CTCF. We amplified the fragments by PCR and labeled them at the 5′ end on either the top (coding) or the bottom (anti-coding) strand with primers for *DM1* site 1 (DM1-1A, 5′–CTG CCAGTTCACAACCGCTCCGAG–3′ and DM1-1B, 5′–GCAGCATTCC CGGCTACAAGGACCCTTC–3′) and for *DM1* site 2 (DM1-2A, 5′–CTTTCTTTCGGCCAGGCTGAGGCC–3′ and DM1-2B, 5′–AAAGCA AATTTCCCGAGTAAGCAGGC–3′). We then either incubated the purified probes with CTCF and then partially digested them with DNase I, or partially methylated them at guanine residues with dimethyl sulfate and then incubated them with CTCF. We isolated free DNA probes from the CTCF-bound probes by preparative gel mobility-shift assays and gel-purified them, and then either analyzed them on a sequencing gel or first cleaved at modified guanine residues by piperidine and then analyzed them as described in detail before[26,35].

In vitro CpG methylation. We methylated labeled and purified *DM1* fragments 1 and 2 at cytosine residues in CpG dinucleotides with *Sss*I methyltransferase (New England Biolabs) in presence of 0.8 mM S-adenosyl methionone for 16 h at 37 °C followed by heat inactivation. We confirmed the methylation status of the DNA fragments by digesting them with methylation-sensitive nucleases *Age*I and *Hpa*II or methylation-insensitive *Msp*I, an isoschisomer of *Hpa*II. We used the same unmethylated probes as a control in these experiments.

Acknowledgments

This work was supported by National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Disease R01-AR45203 (S.J.T.) and National Institutes of Health R01-CA68360 (G.N.F.). We thank P. Rollini and K. Fournier for help with MAR assays, L. Ashworth at the Lawrence Livermore National Laboratory Genome Center for her help in compiling a contiguous sequence for the DM1 locus, J.D. Brook and M.G. Hamshere for cosmids, A.C. Bell, G. Felsenfeld, and C.A. Thornton for plasmids and P. Neiman, S. Collins, M. Groudine and B. Trask for critical comments on the manuscript.

Received 22 February; accepted 24 May 2001.

1. Brook, J.D.M.C. *et al.* Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **68**, 799–808 (1992).
2. Harley, H.G. *et al.* Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**, 545–546 (1992).
3. Mahadevan, M. *et al.* Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255**, 1253–1255 (1992).
4. Fu, Y.H. *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256–1258 (1992).
5. Tsilfidis, C., MacKenzie, A.E., Mettler, G., Barcelo, J. & Korneluk, R.G. Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nature Genet*. **1**, 192–195 (1992).
6. Steinbach, P., Glaser, D., Walther, V., Wolf, M. & Schwemmle, S. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am. J. Hum. Genet*. **62**, 278–285 (1998).
7. Tapscott, S.J. Deconstructing myotonic dystrophy. *Science* **289**, 1701–1702 (2000).
8. Novelli, G.G. *et al.* Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. *Biochem. Mol. Biol. Int*. **29**, 291–297 (1993).
9. Fu, Y.H. *et al.* Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science* **260**, 235–238 (1993).
10. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. & Singer, R.H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol*. **128**, 995–1002 (1995).
11. Krahe, R. *et al.* Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing. *Genomics* **28**, 1–14 (1995).
12. Berul, C.I. *et al.* DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J. Clin. Invest*. **103**, R1–R7 (1999).
13. Klesert, T.R., Otten, A.D., Bird, T,D. & Tapscott, S.J. Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nature Genet*. **16**, 402–406 (1997).
14. Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C. & Moxley III, R.T. Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nature Genet*. **16**, 407–409 (1997).
15. Klesert, T.R. *et al.* Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nature Genet*. **25**, 105–109 (2000).
16. Sarkar, P.S. *et al.* Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nature Genet*. **25**, 110–114 (2000).
17. Mankodi, A. *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**, 1769–1773 (2000).
18. Sabouri, L.A. *et al.* Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. *Nature Genet*. **4**, 233–238 (1993).

19. Alwazzan, M., Hamshere, M.G., Lennon, G.G. & Brook, J.D. Six transcripts map within 200 kilobases of the myotonic dystrophy expanded repeat. *Mamm. Genome* **9**, 485–487 (1998).
20. Bell, A.C., West, A.G. & Felsenfeld, G. Gene Regulation: insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447–450 (2001).
21. Bell, A.C., West, A.G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98,** 387–396 (1999).
22. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).
23. Hark, A.T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486–489 (2000).
24. Kanduri, C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol*. **10**, 853–856 (2000).
25. Mirkovitch, J., Mirault, M.E. & Laemmli, U.K. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39**, 223–232 (1984).
26. Filippova, G. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol*. **16**, 2802–2813 (1996).
27. Burcin, M. *et al.* Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol. Cell. Biol*. **17**, 1281–1288 (1997).
28. Vostrov, A. & Quitschke, W. The zinc finger protein CTCF binds to the APB-beta domain of the amyloid beta-protein precursor promoter: evidence for a role in transcriptional activation. *J. Biol. Chem*. **272**, 33353–33359 (1997).
29. Filippova, G.N. *et al.* A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chrom. Cancer* **22**, 26–36 (1998).
30. Wang, Y-H., Amirhaeri, S., Kang, S., Wells, R.D. & Griffith, J.D. Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* **265**, 669–671 (1994).
31. Godde, J.S. & Wolffe, A.P. Nucleosome assembly on CTG triplet repeats. *J. Biol. Chem*. **271**, 15222–15229 (1996).
32. Wang, Y. & Griffith, J. Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics* **25**, 570–573 (1995).
33. Chung, J.H., Whiteley, M. & Felsenfeld, G. A 5′ element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**, 505–514 (1993).
34. Lobanenkov, V.V., Nicolas, R.H., Plumb, M.A., Wright, C.A. & Goodwin, G.H. Sequence-specific DNA-binding proteins which interact with (G+C)-rich sequences flanking the chicken c-myc gene. *Eur. J. Biochem*. **159**, 181–188 (1986).
35. Lobanenkov, V.V. *et al.* A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* **5**, 1743–1753 (1990).
36. Klenova, E.M. *et al.* CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken cmyc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol*. **13**, 7612–7624 (1993).
37. Awad, T.A. *et al.* Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel DNA sequence. *J. Biol. Chem*. **274**, 27092–27098 (1999).
38. Wolffe, A.P. Imprinting insulation. *Curr. Biol*. **10**, R463–465 (2000).
39. Reik, W. & Murrell, A. Silence across the border. *Nature* **405**, 408–409 (2000).
40. Reddy, P.S. & Housman, D.E. The complex pathology of trinucleotide repeats. *Curr. Opin. Cell. Biol*. **9**, 364–372 (1997).
41. Kanduri, C. *et al.* The 5′-flank of the murine H19 gene in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. *Curr. Biol*. **10**, 449–457 (2000).
42. Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D. & Griffith, J. Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J. Mol. Biol*. **263**, 511–516 (1996).
43. Gerber, A.N., Klesert, T.R., Bergstrom, D.A. & Tapscott, S.J. Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev*. **11**, 436–450 (1997).
44. Mahadevan, M.S. *et al.* Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. *Hum. Mol. Genet*. **2**, 299–304 (1993).
45. Boucher, C.A. *et al.* A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable (CTG)n repeat. *Hum. Mol. Genet*. **4**, 1919–1925 (1995).