CRITICAL DNA BINDING INTERACTIONS OF THE INSULATOR PROTEIN CTCF: A SMALL NUMBER OF ZINC FINGERS MEDIATE STRONG BINDING, AND A SINGLE FINGER-DNA INTERACTION CONTROLS BINDING AT IMPRINTED LOCI

Mario Renda1, Ilaria Baglivo1, Bonnie Burgess-Beusse2, Sabrina Esposito1, Roberto Fattorusso1, Gary Felsenfeld2* and Paolo V. Pedone 1*

1Dipartimento di Scienze Ambientali, Seconda Università degli studi di Napoli -via Vivaldi 43, 81100 Caserta, Italy. 2Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD, 20892-0540

Running head: CTCF DNA binding domain

Keywords: CTCF, zinc finger, DNA binding, imprinting, protein-DNA interaction, insulator

*To whom correspondence should be addressed: Gary Felsenfeld, NIDDK, NIH Bldg. 5 Room212, Bethesda, MD 20892-0540; Phone: 301 496-4173; Fax: 301 496-0201; E-mail: gary.felsenfeld@nih.gov and Paolo V. Pedone, Dipartimento di Scienze Ambientali, Seconda Università degli studi di Napoli, Via Vivaldi 43, Caserta, 81100 Italy; Phone: 39-0823-274598; Fax: 39-0823-274605; E-mail: paolov.pedone@unina2.it

The DNA binding protein CTCF mediates enhancer blocking insulation at sites throughout the genome, and plays an important role in regulating allele-specific expression at the Igf2/H19 locus and at other imprinted loci. Evidence is also accumulating that CTCF is involved in large scale organization of genomic chromatin. Although CTCF has eleven zinc fingers, we show here that only four of these are essential to strong binding, and that they recognize a core 12 bp DNA sequence common to most CTCF sites. By deleting individual fingers and mutating individual sites, we determine the orientation of binding. Furthermore, we are able to identify the specific finger and its point of DNA interaction that are responsible for the loss of CTCF binding when CpG residues are methylated in the imprinted Igf2/H19 locus. This single interaction appears to be critical for allele-specific binding and insulation by CTCF.

The CCCTC binding factor (CTCF) is an evolutionarily conserved transcription factor that is involved in various aspects of gene regulation. It was originally identified for its ability to specifically bind regulatory sequences in the promoter-proximal region of the myc oncogene (1, 2) and in the silencer element of the chicken lysozyme gene (3, 4). It is composed of 727 amino acids and contains 11 zinc finger (ZF) domains, the first 10 of the classical Cys2His2 type (5, 6), while the last has an unusual Cys3HisCys sequence. Through combinatorial use of its 11 ZF motifs, CTCF has been reported to bind to a variety of DNA target sites that perform distinct functions, including promoter activation (7, 8) or repression (1, 2), hormone-responsive gene silencing (3, 4), methylation-dependent chromatin insulation and genomic imprinting (9-11).

Of these activities, the ability of CTCF to mediate insulator function has attracted the greatest interest (11). CTCF–dependent insulators have been particularly characterised in the chicken β-globin locus and in the imprinted Igf2/H19 locus in mouse and human. The region including the β-globin genes and the erythrocyte-specific enhancers is flanked on both sides by CTCF insulators (12-14); in the 5’ boundary, Bell et al. identified a 42 bp CTCF binding site (FII) both necessary and sufficient for enhancer blocking activity (12). In the Igf2/H19 locus, the insulator in the differentially methylated domain (DMD, also called the imprinted control region or ICR) located at the 5’ of the H19 gene prevents the downstream enhancers from acting on the upstream Igf2 gene on the maternal allele (15, 16). The insulator activity in mouse is mediated through four repeats (R1 to R4) that are binding sites for CTCF (15-19). The methylation of the H19 DMD insulator on the paternal allele prevents CTCF binding, thus allowing the enhancers to activate the Igf2 gene promoter (16, 17).

Earlier studies of the interactions between CTCF and its binding sites attempted to determine critical interactions between protein and DNA by mutations in the full eleven zinc finger binding domain that either deleted individual fingers from within the cluster, or sequentially deleted fingers from one end or the other (20, 21). Recent genome-wide surveys of CTCF binding sites in vivo (22-24) have led to a much more precise definition of the consensus DNA binding
sequence. Here we have been able to demonstrate that a much smaller ZF construct as well as a much smaller DNA binding site are sufficient to provide binding interactions similar in strength to those observed with the full eleven ZFs on the previously identified FII DNA site. By using various mutants, we have defined in this paper the minimal CTCF domain that binds the chicken β-globin insulator FII site, identifying the zinc fingers that are responsible for this interaction. We show that other sites such as those at the imprinted Igf2/H19 locus behave quite similarly, and we have identified a 12 bp core DNA fragment both necessary and sufficient for CTCF binding to both the FII and Igf2 R3 insulator sites. These results allow us to establish the orientation of CTCF binding at its site, and to identify the interaction – both the individual finger and the particular methylated CpG - that is disrupted when the DNA of the DMD/ICR at the Igf2/H19 locus is methylated on the paternal allele, an event that leads in vivo to monoallelic expression of Igf2.

**EXPERIMENTAL PROCEDURES**

**Cloning and Purification of the Proteins.** DNA fragments encoding the different segments of the human CTCF protein to be expressed as maltose binding protein (MBP) fusion were generated by PCR using the plasmid p4B 7.1 (12) as a DNA template. Oligonucleotides were synthesized on the basis of the published sequence (NM_006565). The following oligonucleotides were used as primers: primer 1, 5′-CGGAATTCCGCTCTGGTAAAGACATTCCA GTG-3′, and primer 2, 5′-AACATGGTCACTCAGCCATCTGGGCACA GCACAATTATCC-3′, for CTCF ZF 1-11; primer 3, 5′- CGGAATTCCGAGAAGCATTCAAGTTC-3′, and primer 4, 5′-AACATGGTCACTCAGCCATCTGGGCACA GCACAATTATCC-3′, for CTCF ZF 4-8; primer 3 and primer 5, 5′-AACATGGTCACTCAGCCATCTGGGCACA GCACAATTATCC-3′, for CTCF ZF 4-7; primer 3 and primer 6, 5′-AACATGGTCACTCAGCCATCTGGGCACA GCACAATTATCC-3′, for CTCF ZF 4-6; primer 3 and primer 7, 5′-AACATGGTCACTCAGCCATCTGGGCACA GCACAATTATCC-3′, for CTCF ZF 4-5; primer 8, 5′- CGGAATTCCGAGAAGCATTCAAGTTC-3′, and primer 4 for CTCF ZF 5-8; primer 8 and primer 5 for CTCF ZF 5-7; primer 9, 5′-CGGAATTCCGAAAGGCCTTTATGGAATGTAT ATTTCG-3′, and primer 4 for CTCF ZF 6-8. All the PCR products were digested with the restriction enzymes EcoRI and Sall and cloned into EcoRI– and Sall– digested pMal C2G (New England BioLabs) bacterial expression vector. All the plasmids obtained were sequenced to confirm that there were no mutations in the coding sequences. The fusion proteins were expressed in the Escherichia coli BL21 host strain. The transformed cells were grown in rich medium plus 0.2% glucose (according to the New England BioLabs’ protocol) at 37°C until the optical density at 600 nm was 0.6 to 0.7, at which time the medium was supplemented with 200 µM ZnSO4 (CTCF ZF 1-11) or 100 µM ZnSO4 (all the others) and protein expression was induced with 0.3 mM IPTG and was allowed to proceed for 2 h. The cells were then harvested, resuspended in 1x PBS (pH 7.4) (25), 1 mM PMSF, 1 µM leupeptin, 1 µM aprotinin and 10 µg/mL lysozyme, sonicated and centrifuged for 30 min at 27500 rcf. The supernatant was then loaded on amylose resin (New England BioLabs) according to the manufacture’s protocol. Following washes with 1x PBS, purified fractions were eluted in maltose elution buffer [10 mM maltose, 100 mM Tris (pH 8.0), and 100 mM NaCl]. The CTCF ZF 4-7 and CTCF ZF 5-8 proteins were expressed also as GST fusions. The coding sequences were generated by PCR as described for the MBP fusion proteins. The DNA fragments were then digested with the restriction enzymes EcoRI and Sall and cloned into EcoRI– and Sall– digested pGex 6P1 (Amersham Biosciences) bacterial expression vector. The fusion proteins were expressed in the Escherichia coli BL21 host strains. The cells were grown in LB medium at 37°C until the optical density at 600 nm was 0.6 to 0.7 at which time the medium was supplemented with 100 µM ZnSO4 and protein expression was induced for 2h with 1 mM of IPTG at 28°C. The cells were then harvested, resuspended in 1x PBS (pH 7.4) (25), 1 mM PMSF, 1 µM leupeptin, 1 µM aprotinin and 10 µg/mL lysozyme, sonicated and centrifuged for 30 min at 27500 rcf. The supernatant was then loaded on a glutathione-sepharose resin (Amersham Biosciences) according to the manufacturer’s protocol. Following washes with 1x PBS, purified fractions were eluted in glutathione elution buffer. [10 mM glutathione, 100 mM Tris (pH 8.0), and 100 mM NaCl]. The CTCF ZF 4-7 and CTCF ZF 5-8 proteins were expressed also as GST fusions. The coding sequences were generated by PCR as described for the MBP fusion proteins. The DNA fragments were then digested with the restriction enzymes EcoRI and Sall and cloned into EcoRI– and Sall– digested pGex 6P1 (Amersham Biosciences) bacterial expression vector. The fusion proteins were expressed in the Escherichia coli BL21 host strains. The cells were grown in LB medium at 37°C until the optical density at 600 nm was 0.6 to 0.7 at which time the medium was supplemented with 100 µM ZnSO4 and protein expression was induced for 2h with 1 mM of IPTG at 28°C. The cells were then harvested, resuspended in 1x PBS (pH 7.4) (25), 1 mM PMSF, 1 µM leupeptin, 1 µM aprotinin and 10 µg/mL lysozyme, sonicated and centrifuged for 30 min at 27500 rcf. The supernatant was then loaded on a glutathione-sepharose resin (Amersham Biosciences) according to the manufacturer’s protocol. Following washes with 1x PBS, purified fractions were eluted in glutathione elution buffer [10 mM glutathione, 100 mM Tris (pH 8.0), and 100 mM NaCl]. The CTCF ZF 4-7 and CTCF ZF 5-8 proteins were expressed also as GST fusions. The coding sequences were generated by PCR as described for the MBP fusion proteins. The DNA fragments were then digested with the restriction enzymes EcoRI and Sall and cloned into EcoRI– and Sall– digested pGex 6P1 (Amersham Biosciences) bacterial expression vector. The fusion proteins were expressed in the Escherichia coli BL21 host strains. The cells were grown in LB medium at 37°C until the optical density at 600 nm was 0.6 to 0.7 at which time the medium was supplemented with 100 µM ZnSO4 and protein expression was induced for 2h with 1 mM of IPTG at 28°C. The cells were then harvested, resuspended in 1x PBS (pH 7.4) (25), 1 mM PMSF, 1 µM leupeptin, 1 µM aprotinin and 10 µg/mL lysozyme, sonicated and centrifuged for 30 min at 27500 rcf. The supernatant was then loaded on a glutathione-sepharose resin (Amersham Biosciences) according to the manufacturer’s protocol. Following washes with 1x PBS, purified fractions were eluted in glutathione elution buffer.
buffer [10 mM glutathione, 100 mM Tris (pH 8.0), and 100 mM NaCl]. To remove the GST from the CTCF ZF 4–7 and CTCF ZF 5–8 fusion proteins, according to the manufacturer’s advice, a cleavage reaction mix containing 50 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, and 1 unit/100 µg of GST fusion protein of Prescission Protease (Amersham Biosciences) was loaded on the resin where the fusion proteins were bound, incubated for 12 h at 4 °C, centrifuged for 5 min at 500 rcf, and the supernatant containing the protein without the GST tag was then recovered.

**Gel Mobility Shift Analysis.** Unless otherwise specified, 0.9 pmol of each of the purified proteins were incubated for 10 min on ice with 55 fmol of the specified labelled duplex oligonucleotide in the presence of 25 mM Hepes (pH 7.9), 50 mM KCl, 6.25 mM MgCl₂, 1% NP-40, 5% glycerol, and 200 ng of double-stranded poly(dI/dC-dI/dC) (Roche). After incubation, the mixture was loaded on a 5% polyacrylamide gel (29:1 acrylamide:bisacrylamide ratio) and run in 0.5x TBE (25) at 4 °C (200 V for 2h, 15 min). As a non-specific competitor for competition experiments, the oligonucleotide NS, 5’-TGGCCAGGGCCGCGCCGTGGCGGGGCGCGGGGT-3’, was used. Protein concentration was determined by a modified version of Bradford’s procedure (26) (Bio-Rad protein assay). In the case of the CTCF ZF 1-11 protein an apparent higher protein concentration (6 pmol) was required likely because not all the protein sample was correctly folded. The affinity of CTCF ZF 1-11 and CTCF ZF 4-8 for FII 45 bp and of CTCF ZF 4-8 for FII core oligonucleotides were measured by a gel mobility shift assay by performing a titration of the proteins with the oligonucleotide. In the case of CTCF ZF 1-11, in a volume of 20 µL, 3 pmol of the protein was incubated with 0.08, 0.12, 0.14, 0.16 and 0.18 pmol of the duplex FII 45 bp; in the case of CTCF ZF 4-8, in a volume of 20 µL, 0.9 pmol of the protein was incubated with 0.08, 0.12, 0.14, 0.16, 0.17 and 0.18 pmol of the duplex FII 45 bp and 0.08, 0.12, 0.14, 0.16, 0.18 and 0.20 pmol of the duplex FII core, respectively. Scatchard analysis of the gel shift binding data was performed to obtain the Kₐ values (27). All numerical values were obtained by computer quantification of the image using an Amersham Biosciences Typhoon Trio+ apparatus.

**RESULTS**

**Identification of a minimal set of fingers for strong binding.** In proteins containing multiple fingers it is clear that adequate specificity and affinity can be provided by a few critical fingers and that different fingers can have very different roles in a polydactyl complex (6, 28). To define the minimal region of the CTCF protein able to bind the CTCF-binding site FII within the chicken β-globin insulator, a site which is an essential component of the enhancer-blocking activity of that element (12), fragments encoding all 11 ZFs of CTCF (CTCF ZF 1-11), and various deletion mutants were cloned into the pMal expression vector. The proteins were expressed in *Escherichia coli* as a maltose binding protein (MBP) fusion and purified. A gel mobility shift analysis was performed on the purified proteins to determine their ability to bind DNA.

As expected, the CTCF ZF 1-11 protein binds a 45 base pair long oligonucleotide duplex which includes the FII site (FII 45bp, for the sequence see Fig. 1A), producing a single complex (Fig. 1B, lane 1). This result confirms that the N-terminal and C-terminal regions of the protein which flank the ZF domains are not required for DNA binding. The binding specificity of the purified CTCF ZF 1-11 protein was demonstrated by competition experiments with unlabeled oligonucleotides; the complex is competed by addition of a 100 fold-excess of unlabeled FII 45bp oligonucleotide (Fig. 1B, lane 2), but not by the same amount of an unrelated oligonucleotide sequence (NS, Fig. 1B, lane 3). The affinity of the CTCF ZF 1-11 protein for the FII 45bp oligonucleotides was measured using a gel mobility shift assay. Titration of the protein with the oligonucleotide is shown in Figure 1C; Scatchard analysis (27) of these data leads to an apparent dissociation constant of 3.3 ± 1.0 x 10⁻¹⁰ M.

We compared this result with the DNA binding capability of proteins with a reduced number of fingers. Interestingly, similar high affinity binding to the FII site, with an apparent Kₐ of 1.0 ± 0.2 x 10⁻¹⁰ M, can also be obtained with a protein that includes only the ZF domains from 4 to 8 (CTCF ZF 4-8, Fig. 2A). This result indicates that ZFs outside the region that includes the fingers from 4 to 8 do not contribute significantly to DNA binding affinity to the FII site. Starting from the construct CTCF ZF 4-8
we made different C-terminal (Fig. 2B) and N-terminal (Fig. 2C) deletion mutants to identify the minimal DNA binding domain. The protein including ZFs 4 to 7 (CTCF ZF 4-7) still binds the FII site with high affinity (Fig. 2B, lane 2), indicating that when ZF 4, 5, 6 and 7 are present, a ZF 8 contribution is not essential for FII DNA binding. The protein including ZF 4 to 6 (CTCF ZF 4-6, Fig. 2B, lane 3) is able to bind the FII site, but with lower affinity compared to the CTCF ZF 4-7 protein (Fig. 2B, compare lane 2 and lane 3), indicating that ZF 7 plays an important role in stabilising the interaction with the FII DNA site. However, a further deletion of ZF 6 produces the protein CTCF ZF 4-5, which does not bind DNA (Fig. 2B, lane 4), indicating that ZF 6 is involved in FII DNA interaction.

We next determined the effect of deleting fingers from the N-terminus of the cluster CTCF ZF 4-8 (Fig. 2C). When ZF 4 is removed, the resulting protein CTCF ZF 5-8 is also capable of FII DNA binding, but with reduced affinity compared to CTCF ZF 4-8 (Fig. 2C, lanes 1 and 2), indicating an important role for ZF 4 in FII DNA recognition. The protein CTCF ZF 6-8, obtained by further deleting ZF 5, does not bind the FII site (Fig. 2C, lane 3), indicating that ZF 5 is essential for FII high affinity recognition.

The above results show that the proteins CTCF ZF 4-7 and ZF 5-8, each containing 4 ZF domains, are both able to bind the FII site, although CTCF ZF 4-7 shows higher DNA binding affinity (compare CTCF ZF 4-7, Fig. 2B lane 2 with CTCF ZF 5-8, Fig. 2C lane 2). Inasmuch as these two proteins share the ZFs 5, 6 and 7, we tested the CTCF ZF 5-7 construct for its DNA binding affinity to the FII site. Interestingly, the CTCF ZF 5-7 protein does not bind the FII 45bp oligonucleotide (Fig. 2D, lane 2); this result indicates that ZF 8, which we have shown to be dispensable for CTCF high affinity binding to the FII site when ZF 4 is present (compare CTCF ZF 4-8 and CTCF ZF 4-7, Fig. 2B lanes 1 and 2), becomes essential when ZF 4 is missing (compare CTCF ZF 5-8 to CTCF ZF 5-7, Fig. 2D lanes 1 and 2). On the other hand ZF 4, which we have shown to be important in stabilising the CTCF ZF 4-8 DNA binding on the FII site (compare CTCF ZF 4-8 to CTCF ZF 5-8, Fig. 2C, lanes 1 and 2), becomes essential when ZF 8 is missing (compare CTCF ZF 4-7 Fig. 2B, lane 2 to CTCF ZF 5-7, Fig. 2D, lane 2). The results indicate that, in order to obtain high affinity FII DNA binding, four zinc fingers are necessary, with the protein CTCF ZF 4-7 showing the highest DNA affinity. The only three zinc finger containing protein which shows appreciable DNA binding to the FII site in the binding condition we have tested is the construct CTCF ZF 4-6. Considering that CTCF ZF 4-6 and CTCF ZF 5-7 share ZF 5 and 6, we can argue that ZF 4 makes a stronger contribution to DNA binding affinity at this site compared to ZF 7. Similarly, the observation that CTCF ZF 4-7 binds more strongly than CTCF ZF 5-8 to the FII DNA site suggests that in this DNA interaction ZF 4 makes a more significant contribution than ZF 8.

In order to exclude the possibility that the MBP moiety might influence the DNA capability of the fusion proteins, we expressed both the CTCF ZF 4-7 and CTCF ZF 5-8 proteins as GST fusions and tested their ability to bind the FII 45bp oligonucleotide; both proteins are able to bind the probe as GST fusions (data not shown) and after the GST moiety was removed by proteolytic cleavage (see Experimental procedures) (Fig. 2E, CTCF ZF 4-7 cut, lane 1 and Fig. 2F, CTCF ZF 5-8 cut, lane 1). The binding specificity of the purified proteins without the tag was demonstrated by competition experiments with unlabeled oligonucleotides; the complexes are competed by addition of a 100 fold-excess of unlabeled FII 45bp oligonucleotide (Fig. 2E, lane 2 and Fig. 2F, lane 2), but not by the same amount of an unrelated oligonucleotide sequence (NS, Fig. 2E, lane 3 and Fig. 2F, lane 3). Addition of 50 mM EDTA to the binding reactions (Fig. 2E, lane 4 and Fig. 2F, lane 4) abolishes the DNA binding activity, confirming that binding of the proteins to DNA is metal dependent.

Other CTCF binding sites. In order to assess if the CTCF ZF 4-8 region could also be involved in the recognition of other biologically significant CTCF DNA target sites, we first tested the different protein deletion mutants for their capability to bind one of the CTCF sites responsible for the chromatin insulator function of the mouse Igf2/H19 imprinting control region (site R3 according to Engel et al., 2004 ref. 19; see Fig. 1A). Essentially, the DNA binding results obtained with the different CTCF deletion mutants on a 31 base pair long probe including the R3 sequence (R3 31bp, Fig. 3A) are similar to those obtained with the FII site, with the most significant difference being that the CTCF ZF 4-6 protein, which is still able to bind the FII

"Downloaded from http://www.jbc.org/ by guest on November 7, 2017"
sequence (Fig. 2B, lane 3), is not able to bind the R3 site (Fig. 3A, lane 7). Interestingly, the CTCF ZF 4-8 DNA binding domain is also able to bind an oligonucleotide including the R4 CTCF site of the Igf2/H19 imprinting control region (19) (R4 31bp, Fig. 3B) and one derived from the CTCF recognised sequence present in the promoter of the human amyloid β-protein precursor (APP) gene (7) (APP 45bp, Fig. 3C; for the sequences see Fig. 1A). The alignment of all of these sites reveals that 8 bases in a 12-bp long sequence are shared among all of them (Fig. 1A); we defined this 12-bp sequence as the core site.

Length of the minimum DNA binding site. It has been proposed that CTCF binding sites would not be detected by gel shift analyses with 20-30 bp double stranded oligonucleotides because these probes would be too short to accommodate the CTCF protein and guarantee efficient CTCF binding in vitro (1, 29). Having demonstrated that 4 ZFs are sufficient for CTCF high affinity DNA binding we reasoned that we could reduce the size of the oligonucleotide and still get high affinity binding. Firstly, we derived from the FII 45bp probe a 23bp oligo (FII 23bp), which included the core site, and used it as a probe in a gel shift experiment (Fig. 4A). Interestingly, the binding of the different CTCF deletion mutants to the FII 23bp oligonucleotide is similar to that observed with the FII 45bp (compare Fig. 4A and Fig. 2B to 2D). In order to test if the core site, present in the FII 23bp probe, was indeed responsible for the recognition of the target site by the CTCF proteins, we designed a new oligonucleotide (FII core, Fig. 4B) in which the core was preserved and the flanking sequences were changed with respect to the wild type FII 23bp sequence. Interestingly, all the CTCF deletion mutants bind the FII core (Fig. 4B) just as they do the FII 23bp site (Fig. 4A). Moreover, we determined by gel mobility shift assay the affinity of the CTCF ZF 4-8 protein for the FII core oligonucleotide (Fig. 4C) and found a $K_d$ value of $1.7 \pm 0.3 \times 10^{-10}$ M very similar to the one obtained for the interaction of the same protein with the FII 45 bp oligonucleotide; these results clearly indicate that all the specific DNA contacts that the CTCF ZF 4-8 protein makes with the FII target site can be mapped within the core sequence.

CTCF orientation on its binding site. In order to define the orientation of the CTCF ZF DNA binding domain on the FII core site, and to map which of the fingers was able to interact with the individual bases, we designed oligonucleotides bearing various mutations in the core site and used them in gel mobility shift assays performed with the different CTCF deletion mutants. First, we tested the FII core mut 1 oligonucleotide, mutated in the DNA triplet located at the 3’ end of the core (Fig. 5); interestingly, this mutation does not affect the DNA binding of the CTCF ZF 5-8 protein (Fig. 5D), which lacks ZF 4, while it strongly reduces the DNA binding affinity of the CTCF ZF 4-8, CTCF ZF 4-7 and CTCF ZF 4-6 proteins (Fig. 5A, 5B and 5C, respectively), which all have the ZF 4; this results strongly suggests that ZF 4 is recognising the 3’ end site of the FII core. Having determined that ZF 4 recognises the 3’ end of the core, we reasoned that the other ZFs could read the DNA sequence in a 3’→5’ direction. We have previously shown that the CTCF ZF 4-6 protein, which is still able to bind to the FII sequence, is not able to bind the R3 site of the mouse Igf2/H19 imprinted control region (Fig. 3A, lane 7). The sequence comparison between the FII and the R3 site (Fig.1A) shows that there are 3 differences in the core sequence, the most 3’ of which is a G to T transversion at position 9. When the same G to T mutation is introduced in the FII core site (FII core mutT), CTCF ZF 4-6 DNA binding is strongly inhibited (Fig. 5E); this result suggests that a single base mutation in the 3’ half of the core sequence can be responsible for the different binding of the CTCF ZF 4-6 to the FII and R3 DNA sites, confirming that CTCF ZF 4-6 recognises the 3’ of the FII core and the CTCF protein recognises the R3 core site with a similar orientation, in spite of the differences in the DNA sequence.

We then tested the binding of various CTCF finger deletion mutants, using as a probe the FII core mut 2 oligonucleotide, in which the first 3 bases at the 5’ end of the core were mutated; this mutation strongly reduces the DNA binding affinity of the CTCF ZF 4-8, CTCF ZF 4-7 and CTCF ZF 5-8 proteins (Fig. 6A, 6B and 6C, respectively), but does not affect the DNA binding of the CTCF ZF 4-6 protein (Fig. 6D). Considering that all the proteins whose DNA binding is inhibited by the mutation of the first 3 bases of the core have ZF 7, while only the DNA binding of the CTCF ZF 4-6 protein, which lacks ZF 7, is not affected by this mutation, we can conclude that ZF 7 is the one recognising the most 5’ triplet of the FII core. In order to understand if all the three bases mutated in the
FII core mut 2 oligonucleotide were important for CTCF DNA binding, we used as probes in similar gel shift experiments point mutants of the FII core oligonucleotide in which each of the first 3 bases of the core was singly mutated (FII core mut 3, FII core mut 4 and FII core mut 5, Fig. 7); interestingly, the mutation of the cytosine in position 2 (FII core mut 4) has the strongest inhibitory effect on CTCF ZF 4-8, CTCF ZF 4-7 and CTCF ZF 5-8 DNA binding (Fig. 7A, 7B and 7C), suggesting that the contact of ZF 7 with this base, or its complementary guanine, is essential for high affinity DNA binding. We note that mutation of position 3 of the core (FII core mut 5, T replacing G) does not prevent the binding, consistent with the recently published consensus (22) in which A, T or G can be found at this position.

**Effect of DNA methylation: Disruption of a critical finger-DNA interaction.** CTCF DNA binding to different target sequences, including the FII site, has been demonstrated to be sensitive to DNA methylation (16, 17). Interestingly, Bell and Felsenfeld (16) already demonstrated that methylation on both strands of the first CpG in the FII core sequence alone significantly reduced CTCF binding. This CpG dinucleotide includes on the upper strand the cytosine in the second position of the core, which if mutated together with its complementary guanosine inhibits CTCF binding (see FII core mut 4, Fig. 7). In order to test if the methylation of this base, like the mutation, would affect CTCF DNA binding, we produced an oligonucleotide, FII core met, in which each of the first 3 bases of the core was methylated on both strands of the core site in the FII oligonucleotide (R3 core met). This oligonucleotide, which is 23 bp long, includes in the R3 core sequence, only the methylation of the cytosine of the first CpG outside the core sequence does not affect CTCF DNA binding (Fig. 9A, oligo R3 for met 1, lane 3); of the three cytosines in the CpGs located in the R3 core sequence, only the methylation of the first (Fig. 9B, R3 for met 2, lane 4) and second (Fig. 9B, R3 for met 3, lane 5) cytosines inhibits the protein binding, with the first being the one with the strongest effect, while the methylation of the third CpG in the core had no effect (Fig. 9B, R3 for met 4, lane 6). We then tested if also in the case of the R3 site, the CTCF DNA binding domain would be able to bind the R3 core site just as it does the entire wild type 31 bp sequence. Indeed, the protein CTCF ZF 4-7 is able to bind with high affinity the R3 core oligonucleotide, which is 23 bp long and has the R3 core flanked by scrambled DNA sequences (Fig. 9C, lane 1); as expected, methylation of the first CpG of the core site in the R3 core oligonucleotide (R3 core met) inhibits CTCF ZF 4-7 DNA binding (Fig. 9C, lane 2); this results is similar to the one obtained with the methylation of the first CpG of the FII core (see Fig. 8C). Taken together these results
further confirm that the DNA binding modality on the R3 site is similar to the one described on the FII site, thus suggesting that the methylation of the first CpG of the R3 core is preventing the interaction of ZF 7.

**DISCUSSION**

In this manuscript we have defined the CTCF DNA binding domain responsible for the interaction of this protein with various biologically relevant DNA binding sites; this domain includes ZFs from 4 to 8. Interestingly, ZF 8, which we have shown to be dispensable for CTCF high affinity binding to the FII site when ZF 4 is present, becomes essential when ZF 4 is missing, suggesting that the contribution of this finger to CTCF DNA binding affinity could vary depending on the particular DNA site. Moreover, we have identified in the FII CTCF site of the chicken β-globin insulator a 12 bp core site, conserved in various other CTCF DNA binding sites, which is recognised by the CTCF DNA binding domain (CTCF ZF 4-8) with high affinity, comparable to that measured with an oligonucleotide bearing a wild type 45 base pairs sequence. We demonstrated that ZF 4 recognizes the 3’ end of the identified core site, while ZF 7 recognizes the 5’ end; it seems likely that ZF 8 interacts with the DNA sequence outside the core, providing additional but non-sequence-specific DNA contacts. These results allowed us to orient the protein on its target DNA and obtain the first results revealing how individual CTCF ZFs are contributing to the interaction with the target sites. Considering that the data in the literature indicate that most of the Cys2His2 zinc fingers bind 3bp units (28), the 12 bp core identified in this study could be specifically recognized by ZF 4 to ZF 7 oriented on the DNA sequence in the 3’ to 5’ direction.

We note that earlier investigations that made use of sequential deletions from either end of the eleven ZF full binding domain have reported a variety of critical contacts, in some cases involving many more fingers (21). We believe that binding can be partially stabilized by non-sequence-specific interactions between peripheral fingers and outlying DNA sequences. As we have shown, however, the critical interactions for strong binding are confined to a very small DNA sequence and a 5 finger central region of CTCF. Our results should also be compared with those of Quittschke et al. (20), who made use of long constructs in which individual fingers were deleted from the interior to show that the central fingers 5-7, and more weakly fingers 8-10, could contribute to binding to the APP gene promoter sequence; moreover, in agreement with the CTCF DNA binding data on the APP gene promoter (20), our results provide direct confirmation that the N-terminal end of the zinc finger domain is aligned toward the transcriptional start site.

We also found that methylation of the cytosine in position 2 of the core site inhibits ZF 7 DNA binding on both the FII and R3 sites, strongly affecting CTCF DNA binding affinity. Having tested the effect on CTCF DNA binding of the methylations of all the 4 CpGs present in the R3 site, one of which is located outside the core, we have demonstrated that only the methylations of the cytosines on the upper strand of the first and second CpGs in the core inhibit CTCF binding, with the first having a stronger effect. Considering that the CTCF- binding sites present in the Igf2/H19 insulator are very well conserved, it is likely that we have identified the most important methylation signals which affect CTCF DNA binding in these sequences, and we can affirm that ZF 7 DNA binding is inhibited by these methyl groups.

Recent studies on genome-wide localization of CTCF binding sites have provided a redefinition of the CTCF consensus sequence (22-24). Kim et al. have identified a 20-basepair motif which is present in over 75% of the experimentally identified vertebrate CTCF binding sites (22); according to the authors, CTCF binding in vivo is mediated mostly by this single consensus motif, which is also highly conserved evolutionarily as demonstrated by comparing potential CTCF binding sites in different vertebrate genomes. A similar CTCF binding motif was demonstrated by Xie et al. to be significantly enriched in conserved noncoding regions of the human genome (24) and these repeated regulatory motifs have been proposed to be involved in insulator function to limit the spread of gene activation; a comparable CTCF binding site consensus sequence has also been proposed for the Drosophila CTCF protein (23). Interestingly, the 12 bp core sequence we have identified as essential for FII and R3 sites CTCF DNA recognition is entirely consistent with this CTCF consensus motif; in fact, CTCF binding sites randomly selected from the genome (22) have sequences with strong homology to those
Interestingly, it has been reported that different proteins, including CHD8, Sin3A and YB-1 (30-32) bind to the zinc finger cluster of CTCF, suggesting that the zinc finger domain mediates not only DNA binding but also protein-protein interactions; on the basis of our results, the zinc finger domains outside the region from ZF 4 to ZF 8 are good candidates to be responsible for these interactions essential for CTCF activity.

Co-crystal structures of several transcription factors with multiple ZFs bound to DNA have helped in understanding the positioning and nature of amino acids responsible for DNA contacts (6, 28); in these polydactyl DNA complexes not all the ZFs that bind DNA behave alike: some are positioned in the major groove to contact base pairs, whereas others traverse the DNA minor groove either making few stabilising contacts with the DNA phosphodiester backbone or not contacting the nucleic acid at all. Only the resolution of the co-crystal structure of the identified CTCF minimal DNA binding domain with its target site will clarify the contribution of each CTCF ZF to the interaction with DNA.

FOOTNOTES

This work was partially funded by grants PRIN 2006 from MIUR to R.F and P.V.P., FIRB 2003 from MIUR to P.V.P. and L.R. N.5 2003 from Regione Campania to P.V.P. This work was supported by the Intramural Research Program, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

REFERENCES

Fig. 1. (A) Sequence alignment of 4 well characterised CTCF DNA binding sites (7, 12, 19). The FII oligonucleotide includes the CTCF binding site within the insulator located at the 5’ boundary of the chicken ß-globin locus (12); the R3 and R4 sequences are derived from 2 of the 4 repeats present in the mouse ICR at the 5’ of the H19 gene that are binding sites for CTCF and through their insulator activity regulate imprinting at the Igf2/H19 locus (19); the APP oligonucleotide includes the CTCF recognised sequence present in the promoter of the human amyloid ß-protein precursor gene (7). The 12 bp core site is underlined; an asterisk indicates the bases that are identical in all the four DNA sequences. (B) Gel mobility shift analysis of CTCF ZF 1-11 DNA binding to the FII 45 bp oligonucleotide; the DNA binding specificity has been investigated by competition experiments. Purified CTCF ZF 1-11 protein (6 pmol) was incubated with 55 fmol of the labeled FII 45 bp oligonucleotide in the absence (lane 1) or presence of a 100-fold excess of unlabeled specific oligonucleotide FII 45 bp (lane 2) and a 100-fold excess of an unlabeled oligonucleotide with a non specific sequence (NS, lane 3), then subjected to the gel shift analysis. (C) Gel mobility shift titration of CTCF ZF 1-11 with the FII 45 bp oligonucleotide (see Experimental Procedures) (lower panel) and Scatchard analysis of the gel shift binding data (upper panel). The ratio of bound to free DNA is plotted vs the molar concentration of bound DNA in the reaction mixture.
Fig. 2. Identification of the CTCF minimal DNA binding domain. (A) Gel mobility shift titration of CTCF ZF 4-8 with the FII 45 bp oligonucleotide (see Experimental Procedures) (lower panel) and Scatchard analysis of the gel shift binding data (upper panel). The ratio of bound to free DNA is plotted vs the molar concentration of bound DNA in the reaction mixture. (B) Gel mobility shift DNA binding analysis of different CTCF ZF 4-8 C-terminal deletion mutants: CTCF ZF 4-8 (lane 1), CTCF ZF 4-7 (lane 2), CTCF ZF 4-6 (lane 3) and CTCF ZF 4-5 (lane 4). (C) Gel mobility shift DNA binding analysis of different CTCF ZF 4-8 N-terminal deletion mutants: CTCF ZF 4-8 (lane 1), CTCF ZF 5-8 (lane 2) and CTCF ZF 6-8 (lane 3). (D) Gel mobility shift analysis of CTCF ZF 5-7 DNA binding. (E-F) Gel mobility shift DNA binding analysis of CTCF ZF 4-7 and CTCF ZF 5-8 after the proteolytic cleavage of the GST tag. The DNA binding specificity and metal requirement for binding were investigated. The purified CTCF ZF 4-7 cut (E) and CTCF ZF 5-8 cut (F) fragments (1.5 pmol) without the GST tag were incubated with 55 fmol of the labeled FII 45 bp oligonucleotide in the absence (lane 1) or presence of a 100-fold excess of unlabeled specific oligonucleotide FII 45 bp (lane 2), a 100-fold excess of a unlabeled oligonucleotide with a non specific sequence (NS, lane 3) or 50 mM EDTA (lane 4) and then subjected to the gel shift analysis.

Fig. 3. DNA binding analysis of the CTCF deletion mutants on different DNA binding sites. In (A) the R3 31bp oligonucleotide, in (B) the R4 31bp oligonucleotide and in (C) the APP 45bp oligonucleotide were used as probes in gel mobility shift experiments.

Fig. 4. Identification of the minimal core sequence recognised by CTCF in the FII site. Gel mobility shift analysis of CTCF deletion mutants DNA binding to the FII 23bp oligonucleotide (A) and to the FII core oligonucleotide (B); the sequences of the FII 23bp and FII core oligonucleotides are indicated; the core sequence is underlined and the bases which have been mutated are indicated in italics. (C) Gel mobility shift titration of CTCF ZF 4-8 with the FII core oligonucleotide (see Experimental Procedures) (lower panel) and Scatchard analysis of the gel shift binding data (upper panel). The ratio of bound to free DNA is plotted vs the molar concentration of bound DNA in the reaction mixture.

Fig. 5. ZF 4 positioning and CTCF orientation on the FII core site. Comparison of CTCF ZF 4-8 (A), CTCF ZF 4-7 (B), CTCF ZF 4-6 (C) and CTCF ZF 5-8 (D) DNA binding to the FII core (lane 1) and FII core mut 1 (lane 2) oligonucleotides. (E) Comparison of CTCF ZF 4-6 DNA binding to the FII core (lane 1) and FII core mutT (lane 2) oligonucleotides. In gel shift experiments presented in A, B and D 0.2 pmol of the purified proteins were used, while in C and E 0.9 pmol of the purified CTCF ZF 4-6 protein were used. The sequences of the FII core, FII core mut 1 and FII core mutT oligonucleotides are indicated; the core sequence is underlined and the bases which have been mutated are indicated in italics.

Fig. 6. ZF 7 positioning and CTCF orientation on the FII core site. Comparison of CTCF ZF 4-8 (A), CTCF ZF 4-7 (B), CTCF ZF 5-8 (C) and CTCF ZF 4-6 (D) DNA binding to the FII core (lane 1) and FII core mut 2 (lane 2) oligonucleotides. 0.2 pmol of the purified proteins were used in A and B experiments, 0.9 pmol in C and D experiments. The sequences of the FII core and FII core mut 2 are indicated; the core sequence is underlined and the bases which have been mutated are indicated in italics.

Fig. 7. Identification of the bases of the triplet located at the 5’ end of the FII core site which are important for CTCF DNA binding. Gel mobility shift analysis of CTCF ZF 4-8 (A), CTCF ZF 4-7 (B) and CTCF ZF 5-8 (C) DNA binding to the FII core (lane 1), FII core mut 2 (lane 2), FII core mut 3 (lane 3), FII core mut 4 (lane 4) and FII core mut 5 (lane 5)

10
oligonucleotides. 0.2 pmol of the purified proteins were used in A and B, while 0.9 pmol in C. The sequences of the FII core, FII core mut 2, FII core mut 3, FII core mut 4 and FII core mut 5 are indicated; the core sequence is underlined and the bases which have been mutated are indicated in italics.

Fig. 8. Analysis of the sensitivity to the cytosine methylation of the FII core site of the different CTCF deletion mutants DNA binding. Gel mobility shift analysis of CTCF ZF 1-11 (A), CTCF ZF 4-8 (B), CTCF ZF 4-7 (C), CTCF ZF 5-8 (D) and CTCF ZF 4-6 (E) DNA binding to the FII core (lane 1) and FII core met (lane 2) oligonucleotides. 6 pmol of the purified protein were used in A, 0.2 pmol in B and C and 0.9 pmol in D and E. The sequences of the FII core and FII core met are indicated; the core sequence is underlined; the cytosine which has been methylated is indicated. The FII core met oligonucleotide is methylated only on the upper strand.

Fig. 9. Definition of the methylation signal which affects CTCF binding to the R3 site and identification of the core site recognised by the protein. (A) Gel mobility shift analysis of CTCF ZF 4-7 DNA binding to the R3 31 bp (lane 1), R3 fully met (lane 2), R3 for met (lane 3) and R3 rev met (lane 4) oligonucleotides. The 4 CpG present in the R3 31 bp oligonucleotide are methylated on both strands in R3 fully met, only on the upper strand in R3 for met and only on the lower strand in R3 rev met. (B) Gel mobility shift analysis of CTCF ZF 4-7 DNA binding to R3 for met 1 (lane 3), R3 for met 2 (lane 4), R3 for met 3 (lane 5) and R3 for met 4 (lane 6) oligonucleotides. (C) Gel mobility shift analysis of CTCF ZF 4-7 DNA binding to the R3 core (lane 1) and R3 core met (lane 2) oligonucleotides. The sequences of the R3 31 bp, R3 for met, R3 for met 1, R3 for met 2, R3 for met 3, R3 for met 4, R3 core and R3 core met oligonucleotides are indicated; the core sequence is underlined, the cytosines which have been methylated are indicated and the bases which have been mutated in R3 core and R3 core met are indicated in italics; in B and C the complementary oligonucleotides used to obtain the double strands probes were not methylated.
Figure 1

A

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FII</td>
<td>CCCCCAGGGGATGAATTACAAGCGTCTCCCTGCCCTAGGAGGCAAGGAGAG</td>
</tr>
<tr>
<td>R3</td>
<td>------------</td>
</tr>
<tr>
<td>R4</td>
<td>------------</td>
</tr>
<tr>
<td>APP</td>
<td>GCAGGAGGGGGCCAGAGACAGGCTCCCGCGGGCCAGCTAGGAGGAGAG</td>
</tr>
</tbody>
</table>

B

Protein: CTCF ZF 1-11
Comp: Fl 45bp 100x NS 100x

C

Kd = 3.3 (± 1.0) x 10^-10 M

Protein CTCF ZF 1-11
DNA [ ]

Figure 1
Figure 2
Figure 4

A

B

C

Protein

CTCF ZF 1-11
CTCF ZF 4-8
CTCF ZF 4-7
CTCF ZF 5-8

Probe
Fill 23bp

1 2 3 4 5

Fill core

DNA[]

Kd = 1.7 (± 0.3) x 10^{-3} M

Protein CTCF ZF 4-8

F II 23bp: CCTCCCGCTAGGAGGCACAG

F II core: AAGAACCCTAGGGGCTACT
Figure 5
Figure 6
Figure 7

A

Protein: CTCF ZF 4 – 8

Probes:

<table>
<thead>
<tr>
<th></th>
<th>Fil core</th>
<th>Fil core mut 2</th>
<th>Fil core mut 3</th>
<th>Fil core mut 4</th>
<th>Fil core mut 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

Protein: CTCF ZF 4 – 7

<table>
<thead>
<tr>
<th></th>
<th>Fil core</th>
<th>Fil core mut 2</th>
<th>Fil core mut 3</th>
<th>Fil core mut 4</th>
<th>Fil core mut 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

Protein: CTCF ZF 5 – 8

<table>
<thead>
<tr>
<th></th>
<th>Fil core</th>
<th>Fil core mut 2</th>
<th>Fil core mut 3</th>
<th>Fil core mut 4</th>
<th>Fil core mut 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F II core: AAGAAACGGCAGGGGCTACT
F II core mut 2: AAGAAACACCTAGGGGCTACT
F II core mut 3: AAGAAACGGCAGGGGCTACT
F II core mut 4: AAGAAACGGCAGGGGCTACT
F II core mut 5: AAGAAACGGCAGGGGCTACT
Figure 8
Figure 9
Critical DNA binding interactions of the insulator protein CTCF: A small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci

Mario Renda, Ilaria Baglivo, Bonnie Burgess-Beusse, Sabrina Esposito, Roberto Fattorusso, Gary Felsenfeld and Paolo V. Pedone

J. Biol. Chem. published online September 7, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M706213200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts