CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease

Rolf Ohlsson, Rainer Renkawitz and Victor Lobanenkov

CTCF is an evolutionarily conserved zinc finger (ZF) phosphoprotein that binds through combinatorial use of its 11 ZFs to ~50 bp target sites that have remarkable sequence variation. Formation of different CTCF–DNA complexes, some of which are methylation-sensitive, results in distinct functions, including gene activation, repression, silencing and chromatin insulation. Disrupting the spectrum of target specificities by ZF mutations or by abnormal selective methylation of targets is associated with cancer. CTCF emerges, therefore, as a central player in networks linking expression domains with epigenetics and cell growth regulation.

Rolf Ohlsson*
Dept of Genetics and Development, Evolution Biology Centre, Uppsala University, Norbyvägen 18A, S-752 36 Uppsala, Sweden.
*e-mail: Rolf.Ohlsson@ebc.uu.se

Rainer Renkawitz‡
Justus-Liebig-Universitaet, Institute of Genetics, Heinrich-Buff-Ring 58-62, Giessen, D-35392, Germany.
‡e-mail: Rainer.Renkawitz@gen.bio.uni-giessen.de

Victor Lobanenkov§
Laboratory of Immunopathology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
§e-mail: VLOBANENKOV@niaid.nih.gov

Identification of evolutionarily conserved segments of homology within DNA regulatory regions, 'phylogenetic footprints', provides a means for discovering functionally relevant site-specific DNA-binding proteins. This approach has its drawbacks, however, because it depends on the conservation of sequences in pivotal *cis* regulatory elements. In some instances, functionally analogous DNA regions, such as replication origins, centromeres and gene promoters, are similarly regulated in different species despite high sequence divergence. For example, although the noncoding regulatory regions of the avian and mammalian *MYC* genes show little sequence identity, they are nevertheless similarly regulated during cell proliferation and differentiation. One simple solution to this puzzle was suggested by one of us more than ten years ago[1–4]. It assumed that an evolutionarily conserved factor with the ability to recognize highly diverged sequences might exist, giving similar functions to dissimilar sequences.

Identification, cloning and characterization of a nuclear protein, CTCF (for 'CCCTC-binding factor'), that binds to a number of different sequences in the human, mouse and avian *MYC* promoters[3–6], and negatively regulates *MYC* in both mammals and birds[4,6,7] provided evidence for the existence of regulatory factor(s) with multiple DNA sequence specificities. In addition, a silencer protein (NeP1) was discovered independently, which binds to the chicken lysozyme silencer[8,9]. Purification and protein sequencing of NeP1 showed it to be CTCF (Ref. 10) and thus identified a new AT-rich CTCF target site in the lysozyme silencer that is markedly dissimilar with GC-rich target sites in *MYC*. In two subsequent, independent studies, proteins binding to the APPβ site of the human amyloid precursor protein (*APP*) gene promoter[11] and to the FII site within the HS4 enhancer-blocking region of the chicken *β-globin* locus[12] were again purified by their ability to interact specifically with the cognate DNA targets. Once sequenced, they both turned out to be CTCF (Refs 13,14). To highlight the unusual ability of this factor to recognize multiple target sites, we suggested the term 'multivalent' CTCF (Refs 6,10).

In keeping with the multivalent character of CTCF, a growing number of different target sites are now implicated in a variety of regulatory functions, ranging from promoter repression and activation, to the creation of hormone-responsive silencers and enhancer-blocking and/or boundary elements. This article presents the first overview of CTCF structure and function, and discusses recent results that highlight links between CTCF, evolution, epigenetics and disease.

The CTCF protein

The organization of the chicken and human *CTCF* loci is shown in Fig. 1a,b. The exon–intron organization is identical in mice and humans, as are the sequences of the entire promoter regions, 5′- and 3′-untranslated regions (UTRs) and each splice donor–acceptor site (G.N. Filipova *et al.*, unpublished). Compared with the chicken *CTCF* gene[15], the mammalian counterparts have five evolutionarily new introns that have a very high density of different classes of Alu- and LINE-family repeats. These differences do not change the *CTCF* open reading frame (ORF), indicating strict conservation for the estimated 300 million years of evolution since birds and humans separated. Exons E2 to E9 of the mammalian CTCF – containing 11 ZFs and one AT-rich DNA-binding motif (KKRRGRP) – are relatively small. The 11 ZFs are distributed in exons E2 to E8, with several individual fingers being split between neighboring exons (Fig. 1b). Exon 10 harbors the stop codon, and a long 3′-UTR. The latter belongs to the rare group of 'ancient conserved untranslated sequences' (ACUTS; http://pbil.univ-lyon1.fr/acuts/ACUTS), strikingly conserved in vertebrate *CTCF* messages possibly reflecting that the 3′-UTR in *CTCF* has an essential regulatory function in mRNA translation[16]. The longer exons, E1 and E10, contain *trans* repressor

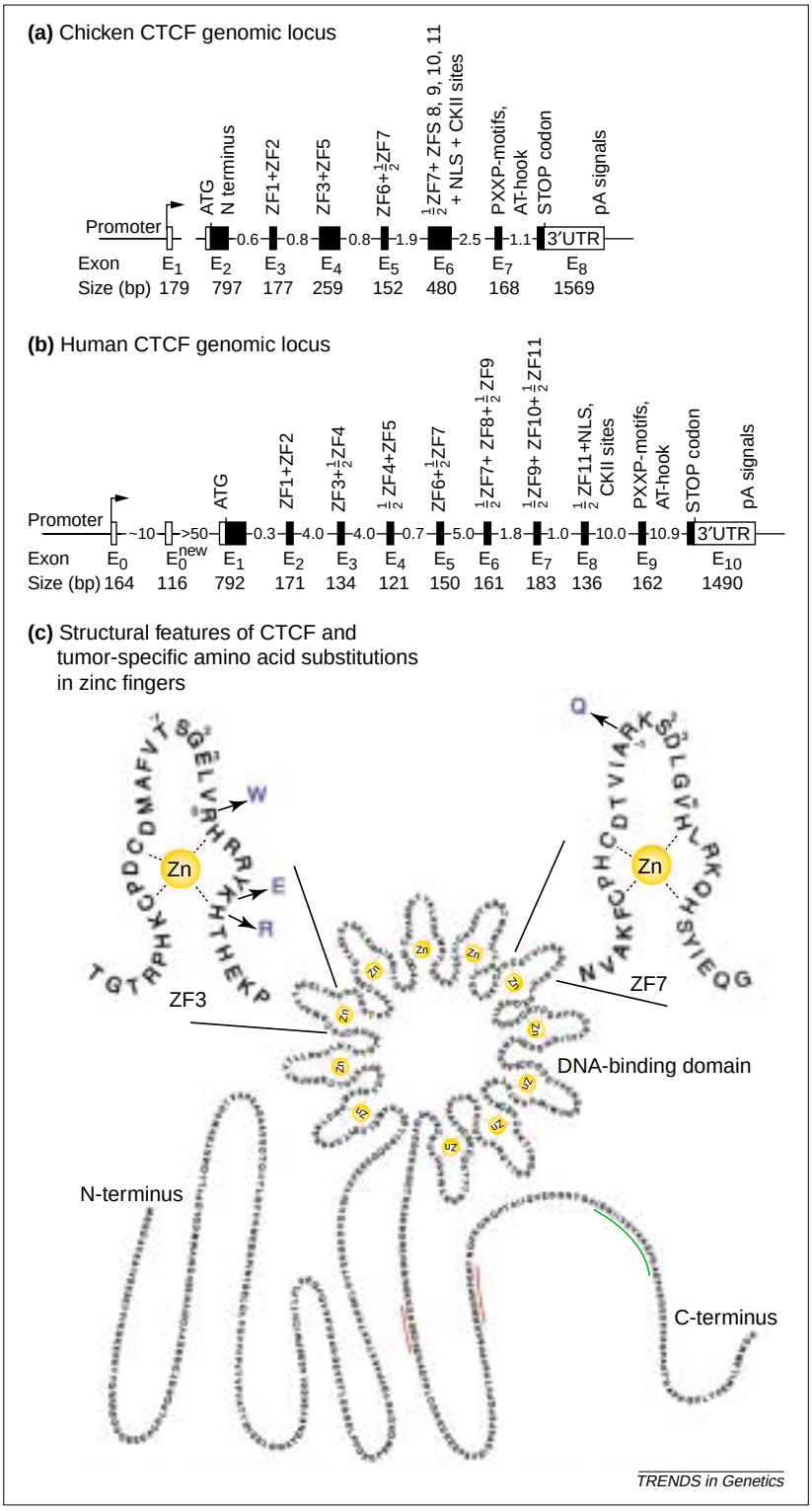

Fig. 1. *CTCF* and its protein product. Genomic organization of the chicken (a) and human (b) *CTCF* genes. Filled boxes, protein coding exons; open boxes, untranslated exons; arrow, transcription start sites. Estimated sizes of introns are in kilobases. The eleven ZFs of mammalian *CTCF* are distributed in exons E2 to E8, with several ZFs being split across neighboring exons[15]. Structure of the avian *CTCF* gene is shown according to Klenova *et al.*[15]. (c) Structural features of CTCF and tumor-specific amino acid substitutions in ZFs. The complete amino acid sequence of the wild-type human CTCF protein shows the DNA-binding domain, which is composed of ten C_2H_2-class ZFs (ZFs 1–10) and one C_2HC-class ZF (C-terminal ZF11). Red, functionally significant sites for CKII phosphorylation[7] and the RGRP-type AT-hook motif[30]; green, pol II-interacting domain. Major base-contacting residues in ZF3 and ZF7 defined by studies of co-crystal structures of multi-ZF factors bound to DNA and tumor-specific missense mutations characterized by G. Filippova and co-workers (unpublished) are indicated on the enlargements.

domains. These domains are individual protein modules that, when tested as fusion proteins with a heterologous DNA-binding domain, confer transcriptional repression. Inhibition of transcription is mediated by the central ZF domain and by sequences both N-terminal and C-terminal to the ZF region[6,17]. Repression by the N-terminal domain is regulated in a cell-type-specific manner[17].

Between avian and mammalian CTCF proteins, 93% of amino acids are identical. However, the identity rises to 100% for the region containing the 11 ZFs[6]. The first ten ZFs are typical units of ~30 residues containing a pair of cysteine residues invariantly separated by 12 amino acids from the pair of histidines. These four residues are coordinated through zinc to form a compact structure with a DNA-recognition α-helix. This type of ZF inserts into the major groove of DNA to make specific contacts with nucleotides by amino acid side chains at positions –1, +2, +3 and +6, marked in Fig. 1c (reviewed in Refs 18,19). The 11th C-terminal C_2HC-type ZF is structurally similar to the C_2HC-type ZFs of the Friend of GATA-1 (FOG) proteins that bind GATA proteins using this type of ZF[20].

Combinatorial use of CTCF zinc fingers

Sequential deletion of each of the CTCF ZFs from either end generated a panel of mutant CTCF proteins for band-shift experiments to assess ZF utilization[6,21,22]. This approach suggested that recognition by native CTCF of different DNA sequences is mediated by varying contributions of individual ZFs[6,10,21–23]. Thus, certain sets of ZFs appear to be necessary for CTCF binding to one target sequence, but are dispensable for binding to another (Fig. 2). However, due to the possible interdependence of DNA-binding properties of the individual ZFs, and/or to additional structural features that could be added to the C_2H_2 ZF fold by the inter-finger linkers (reviewed in Refs 24,25), the 'missing finger' experiments probably provided an incomplete picture of the contribution to target specificity of individual ZFs that usually act in the context of the complete 11 ZF array. Nevertheless, these results suggest that each CTCF ZF might be selectively involved in binding to some targets and dispensible for binding to others.

Co-crystal structures of several transcription factors with multiple ZFs bound to DNA have helped to understand the positioning and nature of amino acids responsible for folding and stability of the C_2H_2 class ZFs, as well as of amino acids establishing DNA contacts[19,26,27]. Recently, extended DNA site recognition by multiple ZFs was shown in the co-crystal structure of the six ZFs of TFIIIA bound to 31 bp of target DNA[27]. Not all of the ZFs that bound to DNA behaved alike. Some ZFs were positioned in the major groove to contact base pairs, whereas other ZFs traversed the DNA minor groove making few or no contacts with the DNA backbone[27].

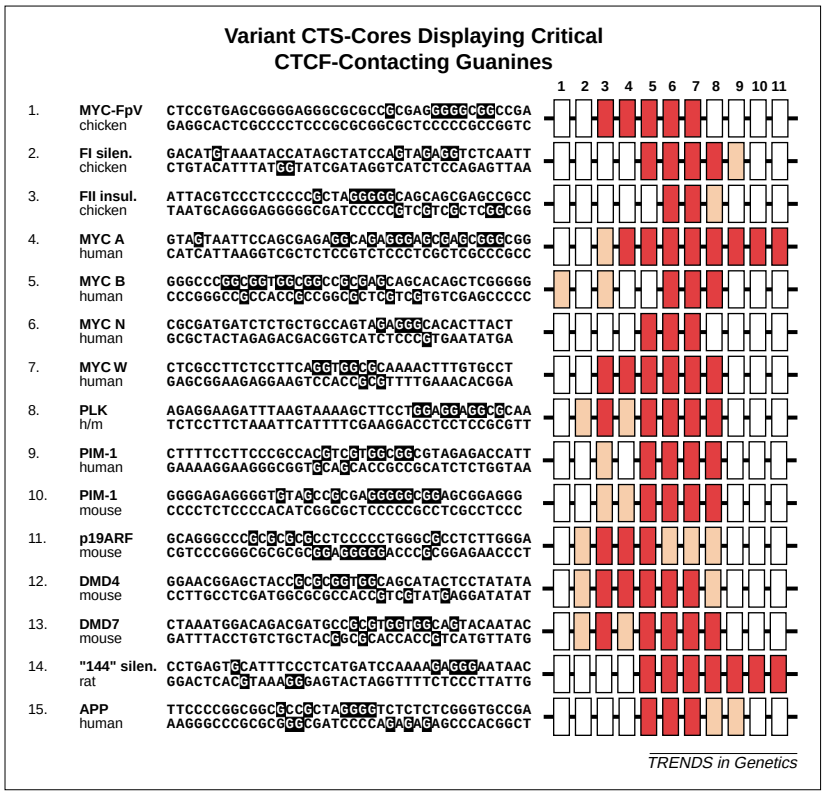

Fig. 2. The wide range of dissimilar CTCF DNA target sites and their recognition by the ZF domain. The left panel shows a series of targets and CTCF-contacting guanines determined by dG-methylation interference within CTCF-bound DNA regions (footprints) defined by protection against DNase I attack. CTCF footprints are usually of ~50–60 bp, and therefore not shown here in full length. The right panel shows the ZFs, with each box representing an individual ZF, and a summary of how different sets of ZFs interact differentially with each target site. Filled boxes show CTCF ZFs that can be deleted from the 11 ZF domain without significantly losing binding to the given DNA. Lighter boxes indicate incomplete loss of binding. Pairwise target comparisons clearly indicate that there is a correlation between different patterns of spatial distribution of contact bases and ZF usage. For example, combinations of CTCF ZF sets required to bind the DMD4 and DMD7 sites from the *H19* ICR are similar to each other, but different from the combinations required for binding to the *β-globin* insulator site FII, or to the *c-myc* site V. FI silencer depicts the silencer element positioned in the chicken *lysozyme* gene. CTCF target sites in the Polo-like kinase (PLK), PIM-1 oncogene (PIM1), the thyroid hormone responsive element 144 and amyloid precursor protein (APP) are also depicted.

The missense ZF mutations found in tumors (G.N. Filipova *et al.*, unpublished) (Fig. 1c), which selectively eliminate binding to some but not to all target sites, suggest additional flexibility for ZF function, showing that the same ZF can behave differently when CTCF is bound to distinct target sequences. It is possible, therefore, that during formation of a CTCF–DNA complex, both DNA and CTCF polypeptide allosterically 'customize' their conformation to engage different ZFs, either for making base contacts or to make a target-specific surface that determines interactions with other nuclear proteins. The sensitivity of CTCF ZF domain to proteinase attack varies upon binding to different targets[6], lending support to this hypothesis.

The ZFs of CTCF are capable of binding either DNA or protein; for example, with several multifunctional factors such as YB-1 (Ref. 28), YY1 and RNP-K proteins (V. Lobanenkov, unpublished). Work is in progress to test whether, in addition to alterations in a spectrum of DNA sequence specificities, any of the cancer-associated mutations might affect interaction of CTCF with other proteins. This bifunctional and combinatorial use of ZFs is unlike other multi-ZF proteins, where the ZFs bind only DNA or only protein. These other factors engage separate groups of ZFs for binding different DNA sequences by each ZF-group (e.g. the MZF1 protein[29]) or for binding DNA by one ZF group independently from binding protein by another group (e.g. the OAZ protein[30]).

Other important motifs in CTCF

Extensive searches for sequence homologies in the N- and C-terminal regions flanking the ZF domain, which together account for approximately two-thirds of the entire protein (Fig. 1c), showed no significant similarities to any previously described protein modules, except for three very short motifs. One of these motifs is a perfect KRRGRP-type AT-hook that might have a role in both DNA binding[31] and protein–protein interactions in chromatin[32,33].

Another motif is the strictly conserved SKKEDSSDSE motif, found in the C-terminal *trans* repressor region[6], where the protein is phosphorylated on the four serines *in vivo* and *in vitro*[7]. Specific mutations of the C-terminal serines abrogated phosphorylation of CTCF *in vitro* and *in vivo,* and phosphorylation by casein kinase II (CKII) *in vitro*[7]. Thus, kinase modifications at CKII sites might be required *in vivo* for attenuation of CTCF activities that are mediated by a conformational change, homo-dimerization, by an interacting protein partner(s), or by combination of these processes. Indeed, completely preventing phosphorylation by substitution of all serines within this site potentiated repression of target promoters, but did not alter CTCF nuclear localization or *in vitro* DNA-binding characteristics. The growth inhibition induced by wild-type CTCF was markedly relieved as a consequence of these mutations[7]. Regulation of CKII by the p38α mitogen-activiated protein kinase (MAPK)[34] indicates a possible pathway that mediates both *MYC* regulation and cell growth control by CTCF (Ref. 7). Because mutation at the very similar CKII-motif in p53 also results in the loss of growth inhibitory activity[35], and because selective phosphorylation of this motif is regulated in response to genotoxic shock by a complex formed by CKII and the components of chromatin transcriptional elongation factor FACT (Ref. 36), the same mechanism operating through the CTCF CKII site would provide a coordination of the events regulated by CTCF and p53. Our preliminary observations with confocal microscopy show that such events might co-localize in cell nuclei, providing further support for possible indirect[21] or direct[37] CTCF–p53 interactions.

A third motif of potential importance, located between the AT-hook and phosphorylation sites (Fig. 1c), contains two repeats of the PXXP-signature characteristic of SH3-domain binding proteins[38]. A search of the limited number of known nuclear

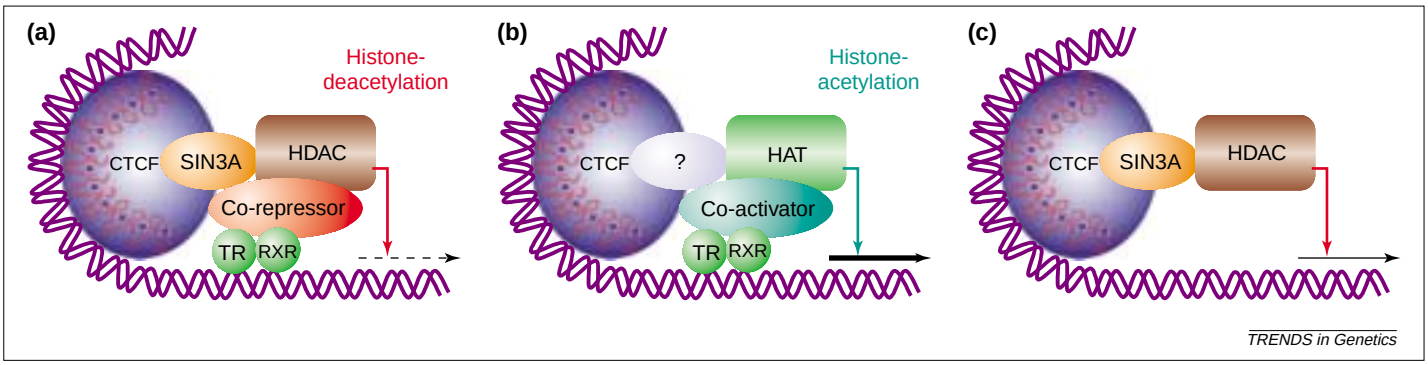

Fig. 3. Synergistic cooperation between hormone receptors and CTCF. (a) Binding of co-repressors N-CoR, SMRT or Alien to the thyroid hormone receptor (TR) allows the interaction with Sin3A and histone deacetylases (see Ref. 63 for more details). On positive TR responsive elements (TREs), this occurs in the absence of ligand, whereas on negative TREs ligand is required. Direct interaction of CTCF with Sin3A mediates binding to histone deacetylases as well[17]. Simultaneous contacts of TR and CTCF to the corepressor/Sin3A complex(es) might allow cooperative binding of histone deacetylases and thereby mediate synergistic repression. (b) Co-activators that recruit histone acetyltransferases (HAT) bind to TR/RXR in the presence of ligand (positive TRE) or without ligand (negative TRE)[63] and might also be recruited by CTCF. (c) Single CTCF-binding sites confer weak gene transcription only[8,41].

SH3-containing proteins, found a candidate tumor suppressor, MYC-binding protein BIN1 (Ref. 39), that associates with CTCF through the PXXP-motif. Preliminary results from the Lobanenkov and the Renkawitz laboratories show that mutation of the crucial prolines in this motif eliminate binding to BIN1 *in vitro* and affect *trans*-repressing activity of CTCF C-terminal region fused to a GAL4 DNA-binding domain. Thus, CTCF–BIN1 interaction might illustrate the functional importance of the conserved PXXP-motifs.

CTCF is a versatile factor

As discussed above, the combinatorial use of the 11 ZFs allows CTCF to be uniquely versatile.

CTCF is a transcription factor

CTCF was one of the first factors shown to bind metazoan silencing elements[4,5,8,9]. Transcription of the chicken lysozyme gene[8,10] and the chicken and human *MYC* genes[4,6] is repressed by CTCF interacting with promoter elements and upstream silencer elements of the gene. In fact, analysis of the lysozyme silencer revealed, for the first time, that a vertebrate silencer shares many features with enhancer elements: they function independently of position and orientation, they comprise functional modules that synergize in transcriptional control, and they might act directly on a minimal promoter[8].

Subsequent analysis of other CTCF target sites identified transcriptional response elements involved not only in gene repression, but also in activation[4,10,13,16,17,21,40–44]. Interestingly, the thyroid hormone receptor (T3R) modulates the effects of CTCF on transcription. In the case of the chicken lysozyme silencer, synergistic repression of the adjacent CTCF and T3R binding sites occurs in the absence of thyroid hormone, and addition of hormone leads to a synergistic activation[8] (Fig. 3). In the context of another negative thyroid hormone response element, the thyroid hormone response is reversed such that the CTCF target site (CTS) and thyroid hormone synergistically repress an adjacent gene[21]. T3R and CTCF also interact directly *in vitro*[42].

The fact that CTCF contains autonomous silencing domains that mediate transcriptional repression led to the identification of CTCF-interacting co-repressors that recruit histone deacetylase (HDAC) activity[17]. Functional implications for these interactions within the context of the different regulatory mechanisms remain to be elucidated. Silencing in general, as tested with minimal promoters, might act directly on the transcriptional start site. Interference with the function of the transcription initiation complex, or modification of the promoter nucleosome(s) may inhibit transcription. The latter mechanism is suggested by the fact that CTCF is able to recruit HDACs[17].

CTCF is a chromatin insulator protein

A recent, exciting development suggests that CTCF is a central player in many chromatin insulators. These operate by blocking the communication between pivotal *cis* regulatory elements, typically gene promoters, and enhancers or silencers[45] (Box 1). Characteristically, insulators function only when positioned between these distinct *cis* elements. Such blocking functions are, as a rule, neutral with respect to enhancers/silencers and promoters (i.e. insulation takes place in the absence of direct effects on the promoters and enhancers), suggesting that the insulator prevents propagation of signals along the chromatin fiber without continuously engaging enhancer or silencer factors.

Although it is important to note that direct proof of CTCF involvement in chromatin insulators is still lacking, the accumulated data shows guilt by association. The initial observation showed that CTCF interacts with the core insulator sequences in several previously characterized vertebrate insulators, notably the chicken *β-globin* FII insulator, the *Xenopus* repeat organizer elements and the BEAD-A insulator from the locus for T-cell receptor-γ (Ref. 14). Because CTCF target sites flank the entire *β-globin* gene cluster[46], CTCF presumably protects this domain from the adverse effects of adjacent regulatory elements (Fig. 4).

Box 1. Models for the action of insulators

Our understanding of the insulator mechanism is intimately linked with our understanding of the mode of enhancer function. It is hotly debated, however, whether enhancers operate by DNA looping, tracking–looping or linking modes[a]. In the looping model, the enhancers are postulated to establish communication with target promoters directly by means of a higher-order chromatin structure. Conversely, both the tracking–looping and linking models assume that the enhancer signals progressively spread along the chromatin fibre before encountering promoter regions (Fig. I). Both the linking and tracking–looping models of enhancer action are compatible with an insulator 'roadblock' model, which describes the physical blocking of spreading of enhancer signals (Fig. I). This model, however, cannot readily explain the blocking of enhancers that operate in the looping mode.

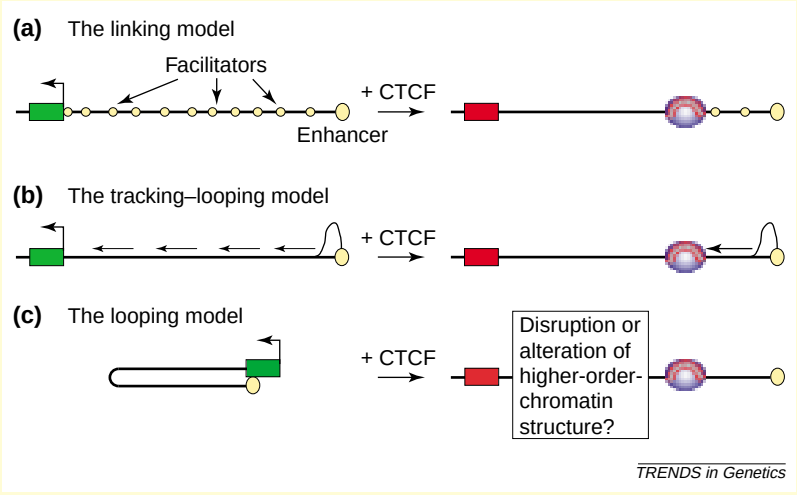

Fig. I. Enhancer and insulator models[b]. Blocking of enhancer function could be explained by a physical obstruction (roadblock) of the propagation of a linear enhancer signal in both the linking (a) and tracking–looping (b) models. The looping model (c) is thought to involve a higher-order chromatin conformation that is altered when CTCF binds to its targets sites. Black line, the chromatin fiber in the *Igf2/H19* domain; small yellow circles, the enhancer facilitators; large yellow circle, the enhancer; green, active loci; red, inactive loci.

References

a Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* 13, 2465–2477

b Bell, A.C. *et al.* (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 291, 447–450

Another more recent example concerns the identification of the *H19* imprinting control region (ICR) as a chromatin insulator. This is situated in the 5′-flank of the *H19* gene and 90 kb downstream of the *Igf2* gene (Box 2)[47,48]. This domain, which is maternally unmethylated and paternally methylated, regulates the repression of the maternal allele of the *Igf2* gene[49], presumably by blocking the communication between the *Igf2* promoters and downstream enhancers[50]. Three reports describe different strategies to document that CTCF target sites in the *H19* ICR (Box 2) are necessary and sufficient for the insulator function[22,51,52]. Importantly, chromatin immunopurification analyses reveal that CTCF is associated only with the maternal allele of the *H19* ICR, strengthening the link between CTCF and allelic control of *Igf2* expression[22]. Interestingly, the *β-globin* FII insulator and the *H19* ICR insulator differ in the use of CTCF ZFs for binding to their

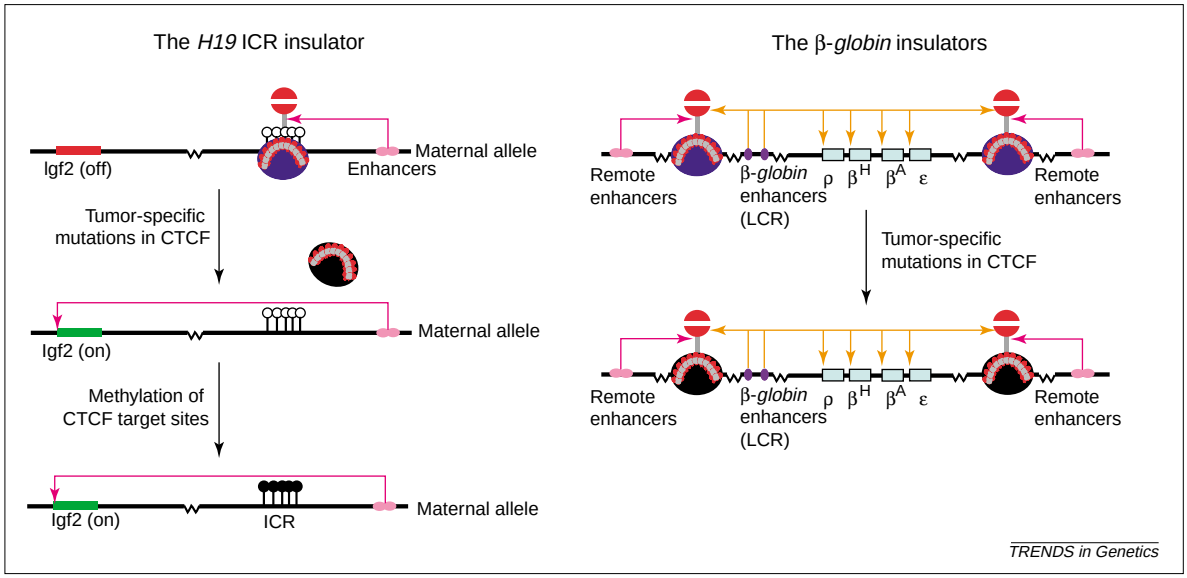

Fig. 4. Cancer-associated CTCF mutations and epimutations of CTCF target sites. Elimination of CTCF binding to target sites in the *H19* ICR because of mutations in tumors is thought to deregulate expression of the maternal *IGF2* allele. Conversely, CTCF mutations do not eliminate binding to CTCF target sites in the *β-globin* locus, for example (see text for details). Wild-type CTCF is represented in blue, and the mutated CTCF is in black.

Box 2. A model of genomic imprinting

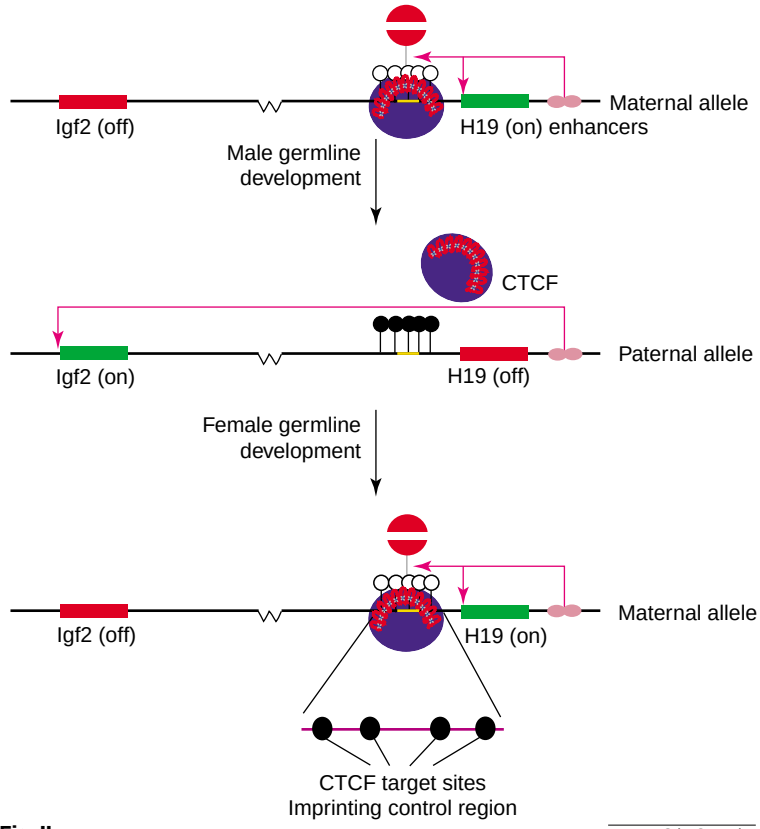

Fig. II

The *Igf2* and *H19* genes are expressed monoallelically from the paternal and maternal alleles, respectively[a–d]. Whereas *Igf2* encodes a growth factor, the *H19* gene encodes a noncoding, polysome-associated transcript that is implicated in the translational control of *Igf2* mRNA[e].

The methylation mark works as a binary switch of chromatin insulator function. The differential methylation patterns are established during gametogenesis (Fig. II). Open circles, unmethylated CpG dinucleotides; filled circles, methylated CpGs in the *H19* ICR; cerise circles, endogenous enhancers 3′ of the *H19*, arches represent enhancer–promoter communications through the looping, tracking–looping or linking modes (Box 1 Fig. I).

References

a Bartolomei, M.S. and Tilghman, S.M. (1997) Genomic imprinting in mammals. *Annu. Rev. Genet.* 31, 493–525

b Ohlsson, R. (ed.) (1999) *Genomic Imprinting, An Interdisciplinary Approach*, Springer-Verlag

c Ohlsson, R. *et al.* (eds) (1995) *Genomic Imprinting: Causes and Consequences*, Cambridge University Press

d Feinberg, A.P. (2001) Genomic imprinting and cancer. In *The Metabolic and Molecular Bases of Inherited Disease* (Scriver, C.R. *et al.*, eds), pp. 527-537, McGraw–Hill

e Li, Y-M. *et al.* (1998) The *H19* transcript is associated with polysomes and may regulate *IGF2* expression in *trans*. *J. Biol. Chem.* 273, 28247–28252

target sites[22]. Whether this means that the mechanism of insulation differs between different CTCF target sites is open to question.

Possible alternative explanations employ CTCF-mediated alteration of higher-order chromatin structures as a common theme[53]. For example, CTCF might translocate or prevent translocation of the *Igf2/H19* domain to different nuclear domains. This could trigger the recruitment of a silencer function that represses only the maternal *Igf2* allele[54]. The recent demonstration that the *gypsy* insulator directs regulatory elements to different nuclear compartments[55], supports this view. This scenario is unlikely to include allele-specific association to nuclear matrix of the *H19* ICR, as defined by resistance to high-salt extraction[56]. To examine these issues further, it might be important to establish whether the insulator function of the *H19* ICR shows a preference for some types of enhancer and whether the *H19* ICR controls localization of regulatory elements to different nuclear compartments in an allele-specific manner.

CTCF, epigenetics and cancer

Methylation

The binding of CTCF to the *Igf2/H19* insulator is methylation-sensitive[22,51,52]. As mentioned above, this is highlighted by the demonstration that CTCF *in vivo* is associated exclusively with the unmethylated, maternally-inherited allele of the *H19* ICR (Ref. 22). The assumption that this leads to a methylation-sensitive insulator was recently confirmed in an episome-based assay[57]. Box 2 summarizes these observations, which indicate that CTCF function marks the *Igf2/H19* expression domain in a parent-of-origin-dependent manner. To examine how methylation affects CTCF binding, an experimentally based prediction of the zinc fingers involved will be required. This will be key in enabling us to understand the general mechanics of the link between CTCF and the methylation status of its varying target sites.

The fact that the CpG-rich CTCF target sites of the *H19* ICR are protected against methylation during early postimplantation development[58] suggests that CTCF prevents *de novo* methylation. However, the fact that the four CTCF target sites comprise a mere 10% of the entire differentially methylated domain of the mouse *H19* ICR hints at the possibility that additional factors organize the methylation-free domain on the maternal allele in concert with CTCF. It will be of obvious interest to determine whether CTCF is involved in the establishment and maintenance of methylation-free regions during

germline development. Given that CTCF target sites are often CpG rich, we do not rule out that CTCF might have a more general function to shape the epigenetic profile of the vertebrate genome, perhaps involving CpG islands, for example.

Histone acetylation

CTCF interacts with SIN3A to recruit HDACs[17], indicating that CTCF has the potential to influence chromatin organization at or around its target sites (Fig. 3). Whether this means that any CTCF-dependent modification in the histone acetylation status of chromatin at target sites can be stably propagated by division, is open to question. In particular, it is not known whether CTCF–target interactions are the same following DNA replication or whether the random distribution of hypoacetylated histones to the daughter strands facilitates CTCF–DNA interaction once replication is completed. The latter suggests that both methylation and histone acetylation marks on daughter strands might be 'read' by CTCF after replication.

Cancer

The ability of CTCF to read epigenetic marks might also provide a cause-and-effect link to some forms of neoplasia, because epigenetic disturbances are common in human cancer (reviewed in Ref. 59). For example, derepression of the maternal *IGF2* allele is linked with abnormal methylation of the CTCF target sites within the *H19* ICR in a wide range of cancer types (reviewed in Refs 60–62). This could reflect the failure of CTCF to establish the chromatin insulator function on the maternal allele[57] (Box 2). This maternal→paternal switch in epigenotype might result from perturbations in methylation protection during development of the soma, or the female germline, thereby eliminating insulator function.

Interestingly, Filippova *et al.* (unpublished) identified several different tumor-specific mutations that result in substitutions of amino acids at positions critical either for ZF formation or DNA recognition (Fig. 1c), and analyzed them for the effect on CTCF binding to different targets (Fig. 2). Remarkably, these tumor-derived mutations abrogate CTCF binding to the *Igf2/H19* ICR targets, eliminating or reducing CTCF interaction with the promoters of these growth-regulating genes, but have no effect on binding to the targets of proliferation-neutral genes, such as the *β-globin* FII insulator (Fig. 4) or the silencer of the *lysozyme* gene. These data can be viewed as an outcome of experiments of Nature that reveal differential contributions of individual ZFs to recognition of distinct CTCF binding sequences. They also stress the possibility that CTCF mutations leave the maternal *H19* ICR allele and a subset of other CTCF target sites unguarded, resulting in abnormal methylation (Fig. 4). Moreover, CTCF displays major features that characterize tumor suppressor genes. The human gene maps within one of the smallest regions of overlap for common loss of heterozogosity at 16q22.1 observed in many solid tumors (reviewed in Ref. 64), and overexpression of CTCF in a variety of cell lines results in the block of cell cycle progression at several different stages that is not associated with an immediate cell death[65].

Concluding remarks

The information summarized here identifies CTCF as a central factor, capable of assuming a multitude of forms to function as a transcriptional activator, a repressor/silencer or an insulator. Perturbations of these functions by mutation of coding sequences or obstruction of target sites by methylation occurs in certain cancers, including those characterized by loss of heterozygosity at the CTCF locus. These features provisionally define CTCF as a tumor suppressor gene. The role of CTCF in imprinting, however, suggests a functional richness not associated with other tumor suppressor genes. This activity might indicate a crucial role for CTCF in natural selection, because it relates to induction of functional and stable hemizygosity for expression of certain genes. Further studies of CTCF in flies, fish, mice and humans are required to uncover the biological crevices accessible to the 11 zinc fingers of this unique protein.

Acknowledgements
We thank Herbert Morse III, Andrew P. Feinberg, Andras Paldi, Elena Klenova, Galina Filippova, Elena Pugacheva, Chandrasekhar Kanduri and Gary Franklin for valuable advice and assistance during the course of this work. This work was supported by the Natural Science Research Council (NFR), the Swedish Cancer Research Foundation (RmC/CF), the Swedish Pediatric Cancer Foundation (BCF) and the Lundberg Foundation to R.O, the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to R.R. and intramural NIAID NIH Research Funding to V.V.L.

References

1 Lobanenkov, V.V. (1988) Chromatin proteins specifically interacting with certain DNA sequences. *Mol. Biol.* (*Mosk*) 22, 869–891
2 Lobanenkov, V.V. and Goodwin, G.H. (1989) CCCTC-binding protein: a new nuclear protein factor which interaction with 5′-flanking sequence of chicken *c-myc* oncogene correlates with repression of the gene. *Proc. Acad. Sci. USSR* (*Moscow*) 309, 741–745
3 Lobanenkov, V.V. *et al.* (1989) CCCTC-binding factor (CTCF): a novel sequence-specific DNA-binding protein which interacts with the 5′-flanking sequence of the chicken *c-myc* gene. In *Gene Regulation and AIDS: Transcriptional Activation, Retroviruses and Pathogens.* (Papas, T.S., ed.), pp. 45–68, Portfolio Publishing Corp.
4 Lobanenkov, V.V. *et al.* (1990) A novel sequence-specific DNA-binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′- flanking sequence of the chicken *c-myc* gene. *Oncogene* 5, 1743–1753
5 Klenova, E.M. *et al.* (1993) CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken *c-myc* gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* 13, 7612–7624
6 Filippova, G. *et al.* (1996) 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
7 Klenova, E.M. *et al.* (2001) Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* 21, 2221–2234
8 Baniahmad, A. *et al.* (1990) Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* 61, 505–514
9 Kohne, A.C. *et al.* (1993) NeP1. A ubiquitous transcription factor synergizes with v-ERBA in transcriptional silencing. *J. Mol. Biol.* 232, 747–755
10 Burcin, M. *et al.* (1997) 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
11 Quitschke, W.W. *et al.* (1996) The initiator element and proximal upstream sequences affect transcriptional activity and start site selection in the amyloid beta-protein precursor promoter. *J. Biol. Chem.* 271, 22231–22239

12 Walters, M.C. *et al.* (1999) The chicken β-globin 5′HS4 boundary element blocks enhancer-mediated suppression of silencing. *Mol. Cell. Biol.* 19, 3714–3726
13 Vostrov, A. and Quitschke, W. (1997) The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J. Biol. Chem.* 272, 33353–33359
14 Bell, A.C. *et al.* (1999)The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396
15 Klenova, E.M. *et al.* (1998) Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *J. Biol. Chem.* 273, 26571–26579
16 Klenova, E.M. *et al.* (1997) Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS–PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. *Nucleic Acids Res.* 25, 466–474
17 Lutz, M. *et al.* (2000) Transcriptional Repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* 28, 1703–1713
18 el-Baradi, T. and Pieler, T. (1991) Zinc finger proteins: what we know and what we would like to know. *Mech. Dev.* 35, 155–169
19 Klug, A. and Schwabe, J.W. (1995) Protein motifs. Zinc fingers. *FASEB J.* 9, 597–604
20 Fox, A.H. *et al.* (1999) Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *EMBO J.* 18, 2812–2822
21 Awad, T.A. *et al.* (1999) Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. *J. Biol. Chem.* 274, 27092–27098
22 Kanduri, C. *et al.* (2000) Functional association of CTCF with the insulator upstream of the *H19* gene is a parent of origin-specific and methylation-sensitive. *Curr. Biol.* 10, 853–856
23 Quitschke, W.W. *et al.* (2000) Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. *Nucleic Acids Res.* 28, 3370–3378
24 Wolfe, S.A. *et al.* (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212
25 Pabo, C.O. and Nekludova, L. (2000) Geometric analysis and comparison of protein–DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.* 301, 597–624
26 Choo, Y. and Klug, A. (1997) Physical basis of a protein–DNA recognition code. *Curr. Opin. Struct. Biol.* 7, 117–125
27 Nolte, R.T. *et al.* (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2938–2943
28 Chernukhin, I.V. *et al.* (2000) Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.* 275, 29915–29921
29 Morris, J.F. *et al.* (1994) Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol. Cell. Biol.* 14, 1786–1795
30 Hata, A. *et al.* OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. *Cell* 100, 229–240
31 Aravind, L. and Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* 26, 4413–4421
32 Currie, R.A. (1997) Functional interaction between the DNA binding subunit trimerization domain of NF-Y and the high mobility group protein HMG-I(Y). *J. Biol. Chem.* 272, 30880–30888
33 Cairns, B.R. *et al.* (1999) Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Mol. Cell* 4, 715–723
34 Sayed, M. *et al.* (2000) Stress-induced activation of protein kinase CK2 by direct interaction with p38 mitogen-activated protein kinase. *J. Biol. Chem.* 275, 16569–16573
35 Milne, D.M. *et al.* (1992) Mutation of the casein kinase II phosphorylation site abolishes the anti-proliferative activity of p53. *Nucleic Acids Res.* 20, 5565–5570
36 Keller, D.M. *et al.* (2001) A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol. Cell* 7, 283–292
37 Wadgaonkar, R. *et al.* (1996) Cloning and characterization of proteins that bind to the C-terminal domain of p53. *Proc. Annu. Meet. Am. Assoc. Cancer Res.* 37, A4092
38 Sparks, A.B. *et al.* (1996) Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1540–1544
39 Sakamuro, D. *et al.* (1996) BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat. Genet.* 14, 69–77
40 Yang, Y. *et al.* (1999) CTCF is essential for up-regulating expression from the amyloid precursor protein promoter during differentiation of primary hippocampal neurons. *J. Neurochem.* 73, 2286–2298
41 Lutz, M. *et al.* (2000) Modulation of thyroid hormone receptor silencing function by co-repressors and a synergizing transcription factor. *Biochem. Soc. Trans.* 28, 386–389
42 Perez-Juste, G. *et al.* (2000) An element in the region responsible for premature termination of transcription mediates repression of c-myc gene expression by thyroid hormone in neuroblastoma cells. *J. Biol. Chem.* 275, 1307–1314
43 Arnold, R. *et al.* (1996) DNA bending by the silencer protein NeP1 is modulated by TR and RXR. *Nucleic Acids Res.* 24, 2640–2647
44 Arnold, R. *et al.* (2000) The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(*)0401 gene. *Gene* 253, 209–214
45 Bell, A.C. *et al.* (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 291, 447–450
46 Saitoh, N. *et al.* (2000) Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO J.* 19, 2315–2322
47 Thorvaldsen, J.L. *et al.* (1998) Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* 12, 3693–3702
48 Kanduri, C. *et al.* (2000) The 5′-flank of the murine *H19* gene in an unusual chromatin conformation unidirectionally blocks enhancer–promoter communication. *Curr. Biol.* 10, 449–457
49 Bartolomei, M.S. and Tilghman, S.M. (1997) Genomic imprinting in mammals. *Annu. Rev. Genet.* 31, 493–525
50 Leighton, P. *et al.* (1995) An enhancer deletion affects both *H19* and *Igf2* expression. *Genes Dev.* 9, 2079–2089
51 Hark, A.T. *et al.* (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* 405, 486–489
52 Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482–485
53 Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* 13, 2465–2477
54 Constancia, M. *et al.* (2000) Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat. Genet.* 26, 203–206
55 Gerasimova, T.I. *et al.* (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell* 6, 1025–1035
56 Horsthemke, B. *etal.* (1999) The mechanisms of genomic imprinting. In *Genomic Imprinting, An Interdisciplinary Approach* (Ohlsson, R., ed.), pp. 91–118, Springer-Verlag
57 Holmgren, C. *et. al.* (2001) CpG methylation regulates the *Igf2/H19* insulator. *Curr. Biol.* 11, 1–3
58 Tremblay, K. *et al.* (1997) A 5′ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol. Cell. Biol.* 17, 4322–4329
59 Feinberg, A. (2001) Cancer epigenetics takes center stage. *Proc. Natl. Acad. Sci. U. S. A.* 98, 392–394
60 Allshire, R. and Bickmore, W. (2000) Pausing for thought on the boundaries of imprinting. *Cell* 102, 705–708
61 Tycko, B. (1999) Genomic imprinting and cancer. In *Genomic Imprinting: An Interdisciplinary Approach* (Ohlsson, R., ed.), pp. 133–170, Springer-Verlag
62 Reik, W. *et al.* (2000) Imprinting in development and disease. *Int. J. Dev. Biol.* 44, 145–150
63 Burke, L.J. and Baniahmad, A. (2000) Co-repressors 2000. *FASEB J.* 14, 1876–1888
64 Filippova, G. *et al.* (1998) 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 Chromosomes Cancer* 22, 26–36
65 Rasko, J. *et al.* (2001) Cell growth inhibition by the multivalent multifunctional transcriptional factor CTCF. *Cancer Res.* 16, 6002–6007