Chromatin Domains, Insulators, and the Regulation of Gene Expression

Rodolfo Ghirlando, Keith Giles, Humaira Gowher, Tiaojiang Xiao, Zhixiong Xu, Hongjie Yao, and Gary Felsenfeld

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. 9000 Rockville Pike, Bethesda, MD 20892-0540

Abstract

The DNA sequence elements called insulators have two basic kinds of properties. Barrier elements block the propagation of heterochromatic structures into adjacent euchromatin. Enhancer blocking elements interfere with interaction between an enhancer and promoter when placed between them. We have dissected a compound insulator element found at the 5′ end of the chicken β-globin locus, which possesses both activities. Barrier insulation is mediated by two kinds of DNA binding proteins: USF1/USF2, a heterodimer which recruits multiple enzyme complexes capable of marking histone on adjacent nucleosomes with ‘activating’ marks, and Vezf1, which protects against DNA methylation. We have found that the heterochromatic region upstream of the insulator element is maintained in its silent state by a dicer-dependent mechanism, suggesting a mechanism for Vezf1 function in the insulator.

Enhancer blocking function in the β-globin insulator element is conferred by a binding site for CTCF. Consistent with this property, CTCF binding was found some years ago to be essential for imprinted expression at the Igf2/H19 locus. Work in many laboratories has since demonstrated that CTCF helps stabilize long-range interactions in the nucleus. We have recently shown that in the case of the human insulin locus such an interaction, over a distance of ~300kb, can result in stimulation of a target gene which itself is important for insulin secretion.

Keywords

CTCF; Vezf1; insulation; heterochromatin

Introduction

Early work on the relationship between chromatin structure and gene expression focused on a few relatively simple systems. The chicken β-globin locus was an attractive choice, both because chicken erythrocyte nuclei were relatively free of proteases and nucleases, and because vertebrate β-globin loci were among the first in which local and more distant regulatory elements had been analyzed. The chicken β-globin family consists of four developmentally regulated genes, regulated by local promoters and by elements of the β-globin locus control region (LCR) (Fig. 1). Further upstream we identified an erythroid-specific folate receptor (FR) gene [1]; it is separated from the globin gene cluster by a ~16
kb long condensed chromatin domain. We have shown that the chromatin of the FR gene and the globin cluster carry ‘activating’ histone modifications (histone H3 and H4 acetylation, H3K4 methylation), whereas the 16kb heterochromatic region carries marks associated with silenced chromatin (H3K9 and K27 methylation) [2, 3].

The abrupt transition between the chromatin structure of the β-globin locus and that of the 16 kb heterochromatic region suggested that there might be a DNA sequence that, together with its associated proteins, served as a boundary between the two domains. We had identified a constitutive hypersensitive site (5’HS4) at just that point [4], and asked whether it had any of the properties expected from a boundary element. In particular we asked [5] whether it could block the action of a distal enhancer when placed between an enhancer and promoter, an activity that had earlier been demonstrated [6] for the Drosophila insulator element, gypsy. We found that a 1.2 kb sequence element had this property, and that an ~250 bp ‘core’ sequence within it was also effective. We used this ‘insulator’ fragment to carry out DNase I footprinting experiments with nuclear extracts, and identified five footprinted regions reflecting five distinct protein binding domains. These formed the basis for further investigations, which led to identification of the proteins associated with each domain, as well as their function. During this work it became clear that the 5’HS4 element could not only block distal enhancer action, but could also prevent the spread of heterochromatin. Within this element, at least, these two activities are associated with distinct protein components.

**Blocking the spread of heterochromatin**

The most obvious question raised by the genomic organization shown in Figure 1 is whether the 5’HS4 element can prevent the spread of the immediately adjacent heterochromatin. To examine this possibility we devised an assay in which a reporter driven by an erythroid specific promoter and enhancer was stably integrated into the avian erythroid cell line, 6C2 (Fig. 2). Typically expression from the reporter is silenced within 100 days of cell culture, as frequently happens with transgenes because of the high probability of integrating near a heterochromatic region. However when it was surrounded by two copies of the 250 bp core insulator element, the reporter was protected from silencing [7]. The DNA sequences responsible for this protective effect were identified by repeating the experiments in Fig. 2 with mutated core elements from which one or another of the footprinted elements had been deleted [8]. This revealed that deletion of any one of the elements except footprint 2 was sufficient to abolish protection against silencing.

Deletion of footprint 4 led to rapid loss of histone acetylation over the reporter gene. We identified the proteins binding to this site as a heterodimer of the regulatory factors USF1 and USF2. In recent studies we have shown that USF1/USF2 can recruit a variety of histone modifying complexes [9]. One of these contains the histone H3 methyltransferase Set1 complex, as well as histone acetyl transferases. Another complex contains among other components the histone H4 arginine 3 methyltransferase PRMT1 [10]. We have found that PRMT1 plays an important role in establishing active chromatin domains. The ultimate effect of USF1/USF2 binding is to deposit on adjacent nucleosomes, through its interactions with these enzymes, a high level of activating histone modifications, effectively preventing the propagation of silencing modifications into the downstream chromatin. Such propagation mechanisms have been proposed for the histone modifications associated with silencing: H3K9 and K27 methylation. In the case of K9 methylation, the modification can be recognized by the heterochromatin protein HP1, which recruits the methylating enzyme that acts on H3K9, and which in turn can methylate the adjacent nucleosome [11], [12]. Plausible evidence exists for an equivalent, polycomb complex-dependent, mechanism for propagation of K27 methylation [13]. Recently, West and his collaborators showed that
USF1/USF2 also recruits the E3 ubiquitin ligase RNF20, which monoubiquitinates histone H2B [14]. Depletion of RNF20 results over time in extension of silencing modifications from the 16 kb heterochromatic domain into the β-globin locus. Heterochromatin formation also extends from the 5’ end of the 16 kb domain, where the authors identify another USF1/USF2 binding site. This results in silencing of the upstream folate receptor gene in the avian erythroid cell line, 6C2. The function of the USF1/USF2 associated complexes is similar to that proposed some years ago for a boundary element in S. cerevisiae; in that case, the barrier was generated by establishing high levels of histone acetylation alone [15].

Role of BGP1/Vezf1 in boundary function

Our recent work has focused on the single protein, BGP1/Vezf1, which occupies footprints 1, 3 and 5 in the 250 bp core insulator element [16]. Deletion of any one of these binding sites results in loss of barrier insulator function [8], [16], and is accompanied by increases in DNA methylation over the promoter region of the reporter gene. Vezf1 binding sites at CpG islands can confer resistance against DNA methylation [16]. We first identified BGP1 many years ago as a protein that binds to G-rich DNA sequences upstream of the chicken adult β-globin gene [17]. More recently the highly homologous gene Vezf1 was cloned from the mouse genome in the laboratory of H. Stuhlmann, and a Vezf1-/− mouse ES cell line was generated [18]. We used this line to study the effects of Vezf1 knockout. Loss of Vezf1 is accompanied by decreases in DNA methylation at many sites genome-wide, including LINE elements and minor satellite repeats as well as near some genes [19]. This led us to measure the abundance of the DNA methyltransferases in the Vezf1 knockout cell line. As shown in Fig. 3 the level of the RNA splice variant coding for the de novo methyltransferase Dnmt3b was markedly decreased, whereas the levels of RNAs coding for the other methyltransferases remained unchanged or increased slightly. Ectopic expression of Vezf1 partly rescued Vezf1 expression [19]. Recent results (Gowher et al in press) show that bound Vezf1 slows the propagation of elongating RNA polymerase II (pol II), and that this results in a change in the relative abundance of splice variants.

Organization of the 16 kb heterochromatin domain

Complete understanding of how barrier insulation works requires knowledge of the heterochromatic structures that are being blocked. A physicochemical study of the 16 kb heterochromatic domain was made possible by the observation that cytosines at CpG sites within the domain are highly methylated, whereas those outside are not. Treatment of nuclei from erythroid cells with the restriction enzyme HpaII resulted in release of the intact domain (Fig. 4). The released fraction was sedimented on a sucrose gradient in which the sedimentation coefficient of bulk chromatin in each fraction was calibrated in the analytical ultracentrifuge (Fig. 4), allowing precise determination of the sedimentation coefficient of the 16 kb fragment. In combination with buoyant density measurements in CsCl this permitted determination of the fragment’s frictional coefficient. Its measured value was consistent with that expected for a canonical ‘30 nm fiber’ [20], [21]. This experiment was repeated, replacing HpaII with MspI, which cuts the same sites (CCGG) but is insensitive to the methylation state of the internal C residue. Although the protein-free DNA of the 16 kb domain is cut into small fragments by MspI, the DNA packaged as chromatin is remarkably resistant to attack (Fig. 5A), reflecting its highly compact structure. This property was used in the following experiments as an assay for compactness.

We first asked whether the compact structure of the 16 kb domain could be disrupted by altering the pattern of histone modifications across the region. As with other heterochromatic regions (see below) low levels of transcript could be detected across the domain [22]. When 6C2 cells were treated with the histone deacetylase inhibitor
Trichostatin A (TSA), the transcript abundance rose (Fig. 5B) and histone H4 acetylation across the domain increased from its characteristically low levels (Fig. 5C). Histone H3 acetylation levels also increased [22]. Furthermore these changes were accompanied by an increase in the accessibility of the domain to digestion by MspI (Fig. 5A).

Although work in several laboratories has provided a detailed description of the mechanism of heterochromatin formation in the fission yeast S. pombe, much less is known about this process in vertebrates. We asked whether the roles of the enzymes Dicer and Argonaute, established in S. pombe, were at least partly conserved in 6C2 cells and involved in the establishment or maintenance of the 16kb domain [22]. Consistent with a role for dicer, we detected small RNAs homologous to sites in the 16 kb domain. When siRNA was used to knock down dicer there was an increase in histone acetylation across the region similar to what had been seen after Trichostatin A treatment. Histone H3K9 methylation, a silencing mark, decreased, transcript levels increased (Fig. 5D) and the entire region became more accessible to MspI digestion [22]. Finally, ChIP studies of Argonaute 2 (Ago2) binding detected Ago2 bound at sites within the 16 kb domain; binding was abolished by Dicer knockdown.

How could these properties of Vezf1 be related to its boundary insulator function in the β-globin enhancer? It is possible that invasion of a euchromatic region by an adjacent heterochromatic domain involves extension into that region of the low level transcription observed in heterochromatin. One role of Vezf1 could be to block this extension by interfering with the advance of RNA polymerase, as described above.

**Blocking the effects of a distal enhancer**

As discussed above, positional enhancer blocking was the first insulator activity of the 5’HS4 element to be detected [5]. A single binding site (footprint 2, see Fig 2B) within the 250bp core insulator element is associated with this activity; we showed that it binds the protein CTCF [23]. Not long after this discovery, we and others reported that regulation of imprinting at the mouse and human Igf2/H19 locus was mediated by CTCF binding to the imprinted control region (ICR) of the maternally transmitted allele [24] [25], [26]. The effect was to block access of a distal enhancer to the Igf2 promoter, silencing its expression on this allele. In contrast, the DNA of the CTCF binding sites within the ICR is methylated on the paternal allele, preventing CTCF binding, and allowing the distal enhancer to activate Igf2 expression.

At the time of these studies, the only other well characterized insulator was the gypsy element of Drosophila [6]. It was shown by Corces and his collaborators that this element binds the protein suppressor of hairy wing and subsequently shown that other co-factors are required for its activity. Most importantly, they showed that the gypsy element and its associated proteins organized DNA into looped domains, and they suggested that if enhancer and promoter were sequestered on separate loops they could not interact, giving rise to insulation [27]. We subsequently proposed a similar mechanism for CTCF mediated insulation. This was based in part on the results of an early attempt to identify CTCF co-factors, which detected the nucleolar protein nucleophosmin in co-immunoprecipitation (co-IP) studies [28]. DNA sequences corresponding to CTCF binding sites were localized on the nucleolar surface in a CTCF-dependent fashion; such an arrangement would generate loops in which each end of the loop was tethered to the nucleolar surface, though not necessarily at the same point.

During the period following these studies there has been a flood of papers addressing questions of CTCF function. A recent review [29] summarizes these results, which show that the principal role of CTCF is to stabilize loop domains throughout the genome.
consequences of this activity depend on the local geometry. For example, as discussed above, when enhancer and promoter are on separate loops they will be insulated from one another. But in other configurations loop formation can bring enhancer and promoter close together, leading to activation (see [29], [30], [31]).

The detailed molecular mechanisms that stabilize CTCF interactions are still under investigation. An important discovery was of the role of cohesin in mediating CTCF loop formation: most CTCF binding sites are also sites of cohesin localization, which is dependent on CTCF binding [32], [33], [34]. Cohesin is essential for loop formation and for insulating activities. We recently have examined the cohesin-CTCF interaction in greater detail. The cohesin complex is composed of four protein components: The long coiled-coil molecules Smc1 and Smc3 which form an open-ended heterodimer, the protein Scc1 (Rad21) which bridges the open end, and SA1/2, which interacts with Scc1 and is external to the Smc1/Smc3/Scc1 trimer. Immunoprecipitation of CTCF co-precipitates the cohesin complex. Further experiments were designed to determine which members of that complex were directly involved in the interaction [35]. They revealed that only SA2 (or SA1) makes direct contact with CTCF; the other members of the cohesin complex are attached to CTCF only indirectly through their interaction with SA2, which in turn makes contact with a site on the C terminal tail of CTCF (Fig. 6A).

A number of other proteins have been shown to interact with CTCF, and in some cases it has been found that they play a role in CTCF function. It is not clear whether all CTCF binding sites, or only a subset, are occupied by any one of these proteins. Recently we used immunoprecipitation /mass spectrometry to search of additional interacting factors [36], and identified the DEAD box RNA helicase p68 as an important co-factor. It forms a heterodimer with a very similar protein, p72. The p68 protein also binds an RNA component, SRA (steroid receptor RNA activator). In a genome-wide ChIP-Seq study, 22% of p68 sites were found associated with CTCF sites, and 7% of CTCF sites bound p68. The p68/SRA complex interacts both with CTCF and the cohesin complex, and appears to stabilize the binding of cohesin. Consistent with these observations, depletion of p68 or SRA results in loss of bound cohesin, but not of bound CTCF (Fig. 6B).

**Long range interactions mediated by CTCF**

As noted above, work in many laboratories has shown that CTCF and its associated cofactors, most notably cohesin, are important in establishing long range contacts within the nucleus [29]. Recent work in our laboratory taking advantage of these results has focused on the organization of chromatin at the human insulin locus [37]. The insulin gene is located immediately upstream of the Igf2/H19 imprinted region (Fig. 7A). It was of interest to determine whether long range interactions within the nucleus between the insulin (INS) gene and distant sites on human chromosome 11 could have regulatory effects. Such contacts could be detected by 4C [38] experiments on human pancreatic islets [39]. From among the strong contacts revealed by this procedure, we chose to focus on the interaction with SYT8, in part because there is a strong CTCF binding site near the SYT8 promoter, as well as near the INS promoter. It seemed likely that interactions between these two sites might be stabilized by bound CTCF. As summarized in Figure 7B, the interaction was confirmed by direct 3C analysis, which also showed that the strength of the contacts increased when islets were treated with glucose. Furthermore, glucose treatment resulted in increased expression of SYT8. To demonstrate further the connection between insulin gene activity and SYT8 expression, the INS gene was inactivated by siRNA targeting of its promoter. There was a marked decrease in SYT8 expression (Fig. 7B). Finally, knocking down CTCF expression in the islets resulted as expected in a decrease in contacts between INS and SYT8. This led to a decrease in SYT8 expression, but, most importantly, INS expression was unaffected.
These results show that the level of insulin expression can affect the level of SYT8 expression through a direct physical contact between the two genes, which are separated by nearly 300 kb. The fact that insulin expression is unaffected by loss of contact accompanying CTCF depletion, whereas SYT8 is downregulated, shows that the INS promoter drives SYT8 expression when it is in close contact, a form of transvection. Consistent with this idea, the INS promoter can stimulate SYT8 expression when it is placed upstream of the SYT8 gene in a plasmid construct [39]. Although SYT8 is a member of the synaptotagmin family of genes that are involved in protein secretion, no clear role has been identified for SYT8 protein, which lacks a calcium binding domain common to other members of the family [39]. To clarify this question siRNA was used to knock down SYT8 expression, which led to a large decrease in secretion of insulin (Fig. 8). Thus the system of physical contacts between INS and SYT8 represents a regulatory network in which increased levels of INS expression result in increased levels of SYT8 expression, which in turn leads to an increase in insulin secretion.

Conclusions

The relationship between chromatin organization within the nucleus and biological function is under intensive study in many laboratories, both at the largest scales of organization [40] [41] and within individual gene clusters [42]. On the smaller scale it is clear that the eukaryotic genome is organized in such a way that individual regions – genes or gene clusters – may be cordoned off so that they are less responsive to signals from their neighbors. The studies described here began with the identification of the compound insulator element at the 5’ end of the chicken β-globin locus. Within this element one can detect a variety of strategies used by the genome to establish such independent regions. To protect against the encroachment of heterochromatin into a euchromatic region, insulator elements may recruit factors that put ‘activating’ histone modifications on adjacent nucleosomes, blocking the spread of ‘silencing’ histone marks. Interference with silencing by heterochromatin may also involve blocking RNA dependent mechanisms for the formation or extension of heterochromatic structures. Another way of compartmentalizing the genome is to create loop domains in which regulatory elements outside the domain are functionally excluded, or those inside the domain are brought together to increase interaction [31] [43]. Our recent study of the INS locus shows that such contacts can have important regulatory consequences. The 4C contact map generated as part of that study [39] suggests that there may be many more interactions with regulatory consequences between insulin and distant neighbors. Work in many laboratories makes it clear that regulation over great distances, mediated by elements that help stabilize such interactions, is an essential aspect of large scale organization within the nucleus. How the detailed nuclear architecture is generated and changes during development, and how it is maintained through cell division or re-established afterwards, are questions that must now be answered.

References


**Highlights**

- Heterochromatin is blocked by elements that maintain positive histone modifications
- Heterochromatin formation in vertebrates can involve dicer dependent mechanisms
- CTCF mediated stabilization of long range contacts affects gene expression
- The human insulin gene regulates the distant SYT8 gene through physical contact
Fig.1.
The chicken $\beta$-globin locus, showing the four globin genes, the upstream locus control region (LCR) that acts as a set of positive regulatory elements for those genes. Upstream is an erythroid-specific folate receptor gene, driven by the HSA regulatory element. The two gene loci are separated by an $\sim$16 kb long heterochromatin domain. In erythroid cells the histones associated with the gene loci are marked by ‘activating’ modifications, and those associated with the heterochromatic region by ‘silencing’ marks. The insulator element, 5’HS4, lies just downstream of the heterochromatin domain. The red triangles summarize two kinds of functions for this insulator: to block inappropriate cross-talk between HSA and the LCR, and to prevent the advance of the heterochromatic region into the globin locus.
Fig. 2.
A. Assay for barrier function. An avian erythroid cell line, 6C2, is stably transformed with a construct coding for a fragment of the IL2 receptor (IL2R), driven by a strong erythroid specific enhancer and promoter. Below is a representative FACS analysis of expression of IL2R as a function of time. Left panel: the uninsulated reporter is silenced after 100 days in culture. Right panel: Surrounding the reporter with core insulator elements protects against silencing [8]. B. DNA sequence of the core element, with footprinted regions representing protein binding sites marked in red [44].
Expression of Dnmt3b in a Vezf1<sup>−/−</sup> mouse ES cell line. Analysis of the splice variants of Dnmt3b; Dnmt3b transcript is decreased in Vezf1<sup>−/−</sup> cells. A) Relative quantitative RT-PCR analysis of DNA MTase transcripts using Taqman expression assay. The Ct values for all MTases are normalized to that of β-actin. B) Map of Dnmt3b gene showing the splicing event leading to the short isoform. Arrows indicate the position of primers for RT-PCR. To analyze alternative splicing, a two fold dilution series of total RNA from WT and Vezf1<sup>−/−</sup> ES cells was reverse transcribed and amplified by PCR for 27 cycles. The bands representing the splicing forms were quantified using Phosphoimager and the ratio of long isoform to the short form is shown in the bar diagram. C) Quantitation of band intensities shown in B). AS=alternatively spliced [19].
Fig. 4.
(A) Excision of the 16 kb heterochromatic domain by HpaII (green triangles) digestion of nuclei. (B) Centrifugation of the digest on a calibrated sucrose gradient, with turkey chromatin added as a carrier. The 16 kb domain is detected by PCR using domain-specific primers, and the y axis shows the abundance of the 16 kb fragment relative to the abundance (from PCR) obtained with fraction 20 taken as zero. Each fraction of the gradient is then run separately in the analytical ultracentrifuge and sedimentation is monitored by UV absorbance, allowing calibration of the sucrose gradient and precise measurement of the domain’s sedimentation coefficient [45, 46].
Fig. 5.
Mechanisms of heterochromatin formation. Effects of raising histone acetylation levels or knocking down Dicer on chromatin structure (A) Sucrose gradient centrifugation of an MspI digest of nuclei from 6C2 cells treated with Trichostatin A (TSA), an inhibitor of histone deactylases, or a control. Arrow shows direction of sedimentation. Experiments were similar to those in Fig. 4, except that Msp I was used to digest nuclei [22]. Msp I cuts the same sites as HpaII, but does not distinguish mcPpG from CpG. There is an increase in smaller (more slowly sedimenting) chromatin in MspI digests after TSA treatment (red) than in the control digest (blue), reflecting increased enzyme accessibility and a more open chromatin structure. (B) Induction of transcript over the 16 kb domain after treatment with TSA. Low levels of transcript from the 16 kb heterochromatic region are detected by qRT-PCR in wild type 6C2 cells, using probes specific for discrete sites in the region. Treatment with 1 μM TSA for 17 hours increases the amounts of this RNA. The abundance of RNA at each probed site is plotted relative to that in untreated cells. (C) Induction of histone H4 acetylation over the domain after TSA treatment for 4 hours (red), compared to control (blue) or 5 days after removing TSA (yellow). Chromatin immunoprecipitation with anti-H4Ac antibody was followed by PCR with probes specific for the region. (D) Induction of transcript following siRNA knockdown of Dicer. A qRT-PCR assay depicts the fold change of RNA levels at 72 h post-transfection. Experiments shown here are typical results, similar to those presented in greater detail in [22].
Fig. 6.
(A) Details of the interaction between CTCF and cohesin [35]. (B) Details of the interaction between the CTCF-cohesin complex and p68 [36].
Fig. 7.
Long range contacts and biological function in the human insulin locus. A) Map of the region around the human insulin gene. Abbreviations: Ins, insulin; TH, tyrosine hydroxylase; Igf2, insulin like growth factor 2 (sense/antisense); ICR, imprinted control region (with binding sites for CTCF). Dotted lines indicate possible interactions between the insulin promoter region and other more distant genes, which were searched for by 4C and 3C methods [39]. B) Summary of results obtained in these experiments with pancreatic islets from multiple human donors, showing the interaction between the human insulin and SYT8 loci, separated by about 300kb in the genome. In islets, addition of glucose results in increase in the contact between the loci, measured by 3C, and stimulation of SYT8 expression.

*Biochim Biophys Acta*. Author manuscript; available in PMC 2013 July 01.
expression (top right). Targeting of the *insulin* promoter with siRNA results in a decrease in *SYT8* expression (bottom left). Knockdown of CTCF results in loss of contact between the genes, downregulation of *SYT8*, but no change in *insulin* expression (bottom middle). To study the role of SYT8 in islet function we knocked down SYT8 and found that this resulted in a decrease in insulin secretion (bottom right, and see following figure) [39].
Fig. 8.
Knockdown of SYT8 expression causes large changes in insulin secretion [39]. Islets were treated either with a control siRNA (blue) or an siRNA that targets the SYT8 promoter (red). Insulin secretion was measured as a function of time after glucose addition (see [39] for detailed procedure).