MOLECULAR CHARACTERIZATION OF SF-1, A NOVEL CHROMATIN BOUNDARY FROM THE *DROSOPHILA* ANTENNAPEDIA COMPLEX

by

VLADIMIR BELOZEROV

(Under the Direction of Haini Cai)

ABSTRACT

My work has focused on understanding the biological function and the molecular mechanism of SF-1, a novel chromatin boundary from the Scr-*ftz* region in the *Drosophila* Antennapedia complex. Chromatin boundaries regulate gene activity by modulating enhancer-promoter interactions (insulator activity) and protecting genes from the influences of neighboring chromatin (barrier activity). Previous work showed that SF-1 contains both of these activities but they reside in separate DNA fragments: the insulator activity is associated with SF-1/b, whereas the barrier activity lies mainly in SF-1/c. Our transgenic studies show that SF-1/b but not SF-1/c is capable of blocking various enhancers. In particular, SF-1/b blocks the *ftz* distal enhancer, an element that does not rely on promoter competition for the selection of appropriate target. On the other hand, SF-1/c but not SF-1/b prevents the spread of silent chromatin initiated at the *Scr* PRE as revealed by the mini-*white* reporter assay. Based on these results we propose a model for the dual function of SF-1 in the region: a) it protects the *Scr* promoter from inappropriate activation by nearby *ftz* enhancers and b) it protects a non-homeotic gene *ftz* from the effects of *Scr*-proximal PRE. The molecular mechanism of insulator and barrier activities appears to be distinct, i.e. SF-1/b critically depends on the GAGA factor, whereas SF-1/c does not. To gain

further insights into the insulator mechanism, we identified the minimal sequence, SF-1/b3, necessary for enhancer-blocking. This sequence is highly conserved among four *Drosophila* species separated by approximately 5 million years, suggesting an important biological function. SF-1/b3 was used as a bait in a yeast one-hybrid screen of the *Drosophila* cDNA library. Five candidate trans-factors were isolated in this screen. To test biological relevance of these proteins we devised a novel cell culture-based insulator assay, utilizing the inducible expression of GFP driven by the metallothionin enhancer. The transcripts of the candidate genes were targeted by RNAi. The knock-down of *e(bx)* and GAGA factor were found to attenuate the enhancer-blocking activity of SF-1/b in this assay. We demonstrate the utility of this cell culture system for the genome-wide RNAi-mediated search for the protein components of insulators.

INDEX WORDS: *Drosophila*, transcription, enhancer, promoter, chromatin boundary, insulator, barrier, GAGA factor, *ftz*, *Scr*, yeast one-hybrid screen, *e(bx)*, cell culture, RNAi.

MOLECULAR CHARACTERIZATION OF SF-1, A NOVEL CHROMATIN BOUNDARY FROM THE *DROSOPHILA* ANTENNAPEDIA COMPLEX

by

VLADIMIR BELOZEROV

B.S., Moscow State University, Russia, 1996

M.S., Florida State University, 1999

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

MOLECULAR CHARACTERIZATION OF SF-1, A NOVEL CHROMATIN BOUNDARY FROM THE *DROSOPHILA* ANTENNAPEDIA COMPLEX

by

VLADIMIR BELOZEROV

Major Professor: Haini Cai

Committee: Michael Bender
Claiborne Glover
Edward Kipreos

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2004

TABLE OF CONTENTS

Page

CHAPTER 1

GENERAL INTRODUCTION

Gene Regulation in Higher Eukaryotes

The central problem in understanding gene regulation in eukaryotes is to explain how the expression of specific sets of genes is activated during cell growth, differentiation, transformation or in response to environmental cues [1, 2]. The vast variety of distinct genetic programs that are implemented in the course of development and that ultimately lead to the formation of a highly complex organism are encrypted in a genome of a finite size. On a molecular level, all the information required to establish the precise spatio-temporal expression pattern of a particular gene is contained in a defined set of cis-regulatory elements [3, 4]. These elements, known as promoters and enhancers, are short sequences containing a combination of binding sites (modules) for nuclear regulatory proteins such as transcription factors [5]. Moreover, regulatory proteins themselves are composed of several modules: a DNA binding domain, a protein-protein interaction domain, and an effector domain that can directly influence the rate of transcription. It is the unique combination of modules in each cis-regulatory element along with a modular assembly at the level of trans-factors that gives rise to a number of regulatory units large enough to govern the expression of tens of thousands genes according to their individual genetic blueprints. This view, termed the combinatorial theory of gene expression [6], is in the center of modern studies on transcriptional regulation.

Promoters

A core promoter is defined as a minimal DNA sequence sufficient for accurate transcription initiation by RNA polymerase II [7, 8]. In general, core promoters are approximately 40 bp long, and encompass a transcription start site (conventionally termed the +1 position) and some 35 bp of upstream or downstream sequence [9]. The key sequence elements found in core promoters are TATA box, initiator element (Inr), TFIIB recognition element (BRE), and downstream core promoter element (DPE) [10]. It is important to note that not all core promoters contain each one of these elements. For instance, there is a large class of promoters that lack TATA box sequence (TATA-less promoters). Multiple recognition sites for several DNA-binding proteins are often found upstream of the core promoter between positions -50 and -200. These transcription factors include Sp-1, NF-1 (also known as CTF or CCAAT-binding transcription factor), and NF-Y (CBF or CCAAT-box binding factor). Together, these upstream sites and the core promoter, are often referred to as "promoter" [11].

Enhancers

Enhancers were first identified in eukaryotic genomes and operationally defined as DNA sequences that elevate the levels of transcripts produced by RNA polymerase II [12]. Known enhancers vary in length and are capable of activating transcription in a manner independent of their orientation and distance from the core promoter [13]. Several examples illustrating these properties of enhancers include the mouse immunoglobulin Hμ core enhancer located in the second intron of the gene, the T cell receptor α-chain gene enhancer located approximately 69 kb downstream of the promoter, and the wing margin enhancer of the *Drosophila cut* gene that resides 85 kb upstream of the target promoter. An even more dramatic example of the distance-

independence of enhancers is observed in the phenomenon of transvection [14], when an enhancer activates an allelic promoter on a separate chromosome.

What is the molecular mechanism by which enhancers activate transcription of the target promoters? The general answer to this question is not known. Extensive studies on several model enhancers, such as IFN-β, TCRα, and TNF-α, offer interesting mechanistic insights. A prototypic example illustrating the concept of combinatorial theory of gene expression is the virus-inducible enhancer of the IFN-β gene [15]. The sequence of this enhancer is located between positions -110 and -45 relative to the transcription start site. The element contains three regulatory modules that specifically bind NF-κB, proteins of the IRF family, and the ATF-2/c-Jun heterodimer, forming a complex termed the enhanceosome [16, 17]. Significantly, none of these modules alone is capable of activating the IFN-β promoter, whereas multimerized modules can induce transcription in response to non-specific stimuli. Additionally, the high mobility group protein, HMG I(Y), binds to four sites within the enhancer and causes structural changes in the DNA sequence. The resulting DNA conformation is favorable for the binding of the three regulatory proteins and their interaction with each other. HMG I(Y) also acts as a sensitive molecular switch that triggers both assembly and disassembly of the activating complex. The acetylation of HMG I(Y) by PCAF increases its affinity for the activators and thus stabilizes the enhanceosome [18]. In contrast, the acetylation by the coactivator protein CBP results in the disruption of the complex and termination of transcription. Once the functional IFN-β enhanceosome is assembled it recruits GCN5 and SWI/SNF chromatin-remodeling complexes to the core promoter of the IFN-β gene [19]. The nucleosome covering the TATA box of the promoter is acetylated and removed allowing the binding of the Pol II holoenzyme and initiation of transcription.

The IFN-β enhanceosome provides an attractive model to explain the mechanism of enhancer-mediated gene activation. It is not clear, however, if this mechanism is shared by other enhancer-promoter combinations. Neither does the enhanceosome model address the issue of enhancer-promoter specificity over long distances in the genome. Indeed, it is known that a single enhancer can activate a wide range of structurally distinct promoters located as far as 80 kb away or on a homologous chromosome. Thus, a question of fundamental importance is: How do enhancers select the appropriate target promoters and avoid the interactions with other available promoters in the genome?

Chromatin Boundaries

In the nuclei of higher eukaryotic cells genomic DNA undergoes a 10,000 – 20,000-fold compaction as judged by the ratio of the DNA length to metaphase chromosome length [20]. Inside the nucleus the DNA exists in the form of a nucleoprotein complex referred to as chromatin. The high degree of chromatin compaction creates a state intrinsically unfavorable for transcription. Enhancers are the key elements that recruit the enzymatic machinery necessary for relieving the repressed state around the promoter and initiating transcription.

According to the current view, the eukaryotic chromatin is divided into a number of topologically and functionally separate domains [21]. The regulatory sequences, such as enhancers, act on the genes located within the same domain but cannot reach the genes located outside. Insulators or boundary elements are postulated to define the borders between independent chromatin domains [22-24]. The topology of independent domains is viewed as a series of chromatin loops attached to the nuclear matrix, thus arranging the chromosomal DNA in a higher-order structure [23].

Two approaches have been taken to validate this model [25]. The main goal of the structural approach was to identify and characterize DNA sequences that mediate chromatin binding to the nuclear matrix. Such sequences have been isolated and named MAR/SARs (matrix/scaffold attachment regions) [26]. Some controversy still exists, however, regarding the ability of MAR/SAR sequences to serve as DNA anchors *in vivo* [27]. It is also unclear if all MAR/SARs share the properties of typical boundaries. A parallel functional approach takes advantage of the fact that a change in the transcriptional activity of a genomic locus is accompanied by gross rearrangements in the structure of chromatin. If boundary elements delimit the spread of gene activity they should also prevent structural changes in chromatin from spreading into neighboring domains. This phenomenon was first observed in the polytene chromosomes of heat-shocked flies. Exposure to high temperature resulted in dramatic changes in transcription of *hsp70* locus genes accompanied by visible shape changes of the chromosome region containing the locus [28]. However, these structural changes did not propagate beyond the heat shock locus. Later, two insulators, scs (specialized chromatin structure) and scs' were found to flank the hsp70 locus. These insulators as well as many others have been tested for their ability to block enhancer-promoter interactions in test transgenes [29].

Mechanism of Boundary Function

Boundaries are functionally defined as sequences displaying two assayable characteristics [30]. First, they are capable of abolishing the activating effect of the distal enhancers on a target promoter. This property is clearly distinct from general silencing of the region since the promoter-proximal enhancers and the transcriptional potency of the promoter itself remain unaffected. DNA elements capable of blocking enhancers in a position-dependent fashion are

often referred to as insulators. Second, when two boundaries flank a transgenic construct, the reporter genes within the construct tend to be expressed at comparable levels regardless of the chromosomal integration site. Thus boundaries are capable of protecting transgenes against chromosomal position effects (CPE), a characteristic termed the barrier activity [31].

Formally, the two methods of characterizing insulators could reveal either the same or two distinct cis-regulatory activities, depending on the nature of the DNA sequence being tested, the structure of a test transgene and the chromosomal environment in the integration locus. For instance, if the integration of the test transgene occurred in the gene-dense transcriptionally active region with multiple enhancers located nearby, the protection of the transgene against CPE is functionally identical to the disruption of promoter-enhancer interactions. On the other hand, the transgenes integrated into the heterochromatic regions are expected to be repressed by the local environment of highly condensed chromatin. To relieve such repression, the boundaries flanking the transgene are thought to prevent propagation of condensed state into the transgene, the activity potentially different from disrupting enhancer-promoter communication.

The majority of boundaries tested to date display both the insulator and the CPE barrier activities. In a recent study of the chicken β-globin boundary (5'HS4), the authors showed that these activities are attributable to different sequences within the boundary and may require the recruitment of distinct protein components [32]. A 250-bp 5'HS4 contains five protein-binding sequences (FI-FV) as revealed by DNase I footprinting with the nuclear extracts. Previous results demonstrated that the footprinted region II, a recognition sequence for the nuclear protein CTCF, is necessary and sufficient for the enhancer-blocking activity of 5'HS4. However, the deletion of FII has a relatively mild effect on the CPE barrier activity. Moreover, this activity is differentially affected by the deletion of FI, FIII, FIV, and FV, and various combinations thereof,

suggesting the involvement of a distinct trans-factor(s). Interestingly, two other vertebrate insulators examined, DMD of the mouse Igf2 and H19 locus and the BEAD element of the human T-cell receptor α-δ locus, both contain CTCF binding sites and possess enhancer-blocking activity but lack the CPE barrier activity. Our laboratory demonstrated that SF-1, a boundary from the *Drosophila* Antennapedia complex, displays both the insulator and the barrier activity, but these activities are associated with separable DNA sequences, the first example of the activity separation in a fly boundary (P. Majumder, unpublished results).

Almost all of the experiments aimed at elucidating the molecular mechanism of boundary function have been performed with a small set of well-characterized *Drosophila* boundaries. All of these boundaries have been shown to function in both enhancer-blocking and CPE barrier assays at least at the level of resolution afforded by using full-length sequences. Thus, even if the two boundary activities are mechanistically distinct, the models derived from the available experimental data do not accommodate such differences. Three possible mechanisms have been proposed to explain the observed properties of boundaries [30].

1. The chromatin structure model suggests that boundaries contain sequences that are capable of recruiting nuclear factors responsible for remodeling the structure of the surrounding chromatin at the level of primary fiber or higher. A recent study of two *gypsy*-induced alleles, *yellow*^{2} and *yellow*-800 tested this hypothesis [33]. In these alleles the full-length *gypsy* transposon or a 350 bp boundary region are inserted upstream of the *yellow* promoter, separating it from two tissue-specific enhancers. The third instar larval nuclei (in which *yellow* gene is not tanscriptionally active) were treated with DNase I and MNase resulting in chromatin accessibility maps. Three regions have been tested separately: the sequence upstream of *gypsy*, the sequence between *gypsy* and the transcription start site, and the sequence downstream of the *yellow* promoter. It

was found that in both alleles the accessibility of chromatin to nucleases is dramatically increased within the insulator itself and in the promoter-proximal sequence but not in other regions. This effect is restricted to cells with completely functional *gypsy* boundary, since in the backgrounds deficient for either *Su(Hw)* or *mod(mdg4)* the chromatin accessibility to nucleases in the region of interest is indistinguishable from the one in the wild-type *yellow* allele. Interestingly, the number and position of nucleosomes does not depend on the presence of a functional boundary. Therefore, *gypsy* must cause changes in the higher-level chromatin structure. It is currently unclear why chromatin remodeling occurs in a unidirectional manner, only towards the *yellow* promoter.

An example of a direct effect of the boundary DNA on the nucleosome structure is seen in the chicken β-globin locus [34, 35]. A boundary activity has been previously documented in the DNase I hypersensitive region 4 (HS 4) and is believed to protect the β-globin gene from the repressive effects of a 16 kb heterochromatic region located upstream [36]. Histone H3 is heavily methylated on K9 in this region, a modification that facilitates binding of heterochromatic protein HP1. HP1 can recruit histone methylase Suv39h, which, in turn, will lead to the propagation of condensed state further downstream. HS 4 insulator appears to recruit enzymes capable of acetylating K9 residue of H3, thereby preventing methylation by Suv39h (Figure 1 A). This mechanism provides a biologically sound rationale for the CPE barrier activity of boundaries but does not explain enhancer-blocking.

2. Transcriptional model suggests that the boundaries intercept the transcriptional signals sent by the enhancers to the promoters by mimicking the promoter sequences and/or protein complexes present at the promoters (Figure 1 B). This hypothesis is often referred to as the promoter decoy model [30]. This view provides a clear explanation of the strict directionality observed in the

insulator action - insulators can disrupt enhancer-promoter communication only when located between these elements. On the molecular level, several lines of evidence support this model. First, many boundaries contain sequences similar to the ones found in promoter regions of the genes. For example, *gypsy* boundary has been shown to act as a distance-dependent transcriptional stimulator in heterologous transgenes [37, 38], consistent with the fact that this element is located in the vicinity of the transcription start site in the native *gypsy* transposon. Similarly, scs element is found in the promoter region of the heat shock gene [39]. The promoter region of the *even-skipped* gene also contains insulator activity [40]. It should be noted that these elements are distinct from the core promoters, but rather resemble the regulatory promoters. The function of regulatory promoters is thought to be in collecting the stimulatory signals from different enhancers and passing them on to the core promoter [41]. This mode of action explains the possibility of "saturating" an insulator as a result of increased enhancer strength. Indeed, the *gypsy* insulator has been shown to completely block the yolk protein fat body enhancer FBE1. However, when the strength of FBE1 is increased by the presence of multiple copies, the activating signal is partially passed through *gypsy* [42]. Similarly, weakening the *gypsy* insulator by eliminating some of the *Su(Hw)* binding sites causes partial bypass by a single copy of FBE1. Second, the boundary generated by insulators appears to be effective in disrupting transcriptional interactions only. In a recent study [43], the *gypsy* insulator and scs were tested for their ability to prevent physical contact between two FRT sites placed on either side of the boundary. In flies expressing *flp* recombinase, excision of the FRT-flanked transgene occured regardless of the presence of the boundary between the FRT sites. Third, the *gypsy* boundary has been shown to be active outside of the chromosomal context, on plasmids as small as 5.5 kb [43]. Therefore, higher-order chromatin structures may not be required for the boundary activity. If two

independent domains were to be formed by *gypsy* on a plasmid of this size, their size would be extremely small - 1.1 kb and 4.4 kb, respectively. Even the larger of these domains would contain a maximum number of 22 nucleosomes, corresponding to approximately 3.5 turns of the solenoid in the 30 nm fiber. Such a small structure is unlikely to form a higher order chromatin.

3. The looping model is conceptually different from the two above models in that it does not rely on the ability of insulator sequences to interact with proteins directly involved in the process of transcription (such as chromatin modifying complexes, transcription factors, etc.) [30]. Instead, it is suggested that insulators impose a topological or physical constraint on the chromatin fiber between an enhancer and a target promoter. It is hypothesized that insulator sequences bind protein complexes capable of associating with homologous complexes assembled on a number of other insulators and, thus, divide the chromosome into independent loop domains. Large protein complexes (insulator bodies) that are formed at the base of these loops prevent the efficient propagation of transcription-activating signals from an enhancer located in one loop to the promoter in another (Figure 1 C) [23]. This idea is experimentally supported by the observation that the two known components of *gypsy* insulator, *Su(Hw)* and *mod(mdg4)*, overlap at approximately 200 sites on polytene chromosomes from the *Drosophila* salivary glands. Such a large number of sites should result in the diffuse and general nuclear staining in the interphase diploid cells. Instead, *Su(Hw)* and *mod(mdg4)* co-localize at 20-25 nuclear sites, suggesting that each site is an aggregate of approximately 10 individual complexes [44]. Moreover, a recent study shows that the sub-nuclear location of these aggregates is not random [45]. Roughly 75% of them are found on the nuclear periphery, presumably bound to the nuclear lamina. Another study, however, argues that the periferal localization is not essential for the boundary function [46]. More direct evidence supporting the formation of loop domains comes from the observation

that the two specific DNA sequences located at 4D and 7B on the X chromosome co-localize when both contain nearby *gypsy* insertions [45]. In the absence of *gypsy* these sequences are observed at distinct locations within the nucleus, consistent with a large linear distance between them - approximately 3,000 kb. The biochemical information available on the *Su(Hw)* and *mod(mdg4)* proteins fits this model. *Su(Hw)* is a large multi-domain protein that utilizes zinc finger motifs to bind *gypsy* DNA. It also contains a leucine-zipper region highly homologous to the helix 2-coiled-coil region of the bHLH-Zip proteins, and two acidic domains at the termini [47]. The *mod(mdg4)* protein interacts with the leucine zipper of *Su(Hw)* via a highly acidic C-terminal domain. In addition, *mod(mdg4)* contains a BTB domain capable of forming homo-dimers and oligomers, a property necessary for the formation of insulator bodies [48]. Other, yet unknown domains within either protein may mediate the binding of the insulator bodies to the components of the nuclear lamina. An alternative possibility is that *gypsy* DNA itself forms associations with the nuclear matrix, since it has been shown to contain MAR/SAR activity *in vitro* [49]. This view is consistent with the observation of the functional cancellation of two *gypsy* boundaries located immediately next to one another or separated by a small distance [50]. Indeed, this arrangement would promote the preferential interaction of the *mod(mdg4)* proteins associated with the two boundaries, thereby preventing attachment of either boundary to the insulator body elsewhere in the nucleus. Similar observations have been made with test transgenes containing the *gypsy* insulator placed next to a GAGA-dependent Mcp boundary or a synthetic array of GAGA sites. Like *mod(mdg4)*, GAGA factor protein contains a BTB domain but also a zinc finger domain allowing direct binding of GAGA factor to target DNA. These two heterologous boundaries appear to functionally cancel each other, consistent with the preferential interaction of the nearby BTB domains in *mod(mdg4)* and the GAGA factor [51]. In fact, it has

been shown that the BTB domains of the two proteins not only interact *in vitro* and *in vivo*, but also can substitute for each other [48]. It is conceivable that the BTB domain serves as a universal motif that mediates association of the "*gypsy*-like" boundaries with the insulator bodies and that other BTB domain proteins are involved in the formation of boundaries. Sharing common protein components, however, does not appear to be a prerequisite for the interaction of boundaries in the nucleus. Two boundaries flanking the *hsp70* locus, scs and scs' bind distinct proteins, BEAF-32 [52] and *zw5* [53], respectively. The Chromosome Conformation Capture (CCC) experiments show, however, that these two boundaries interact *in vivo* forming a loop domain that contains the entire heat shock locus. This interaction is presumably mediated by BEAF-32 and *zw5*, which co-immunoprecipitate out of the nuclear extracts and interact in the GST-pulldown assay [54].

Biological Roles of Boundaries

The Bithorax complex.

Segment identity in the posterior two-thirds of the *Drosophila* embryo, from parasegment 5 to parasegment 14, is determined by the level of expression of the three homeotic genes comprising BX-C, *Ultrabithorax* (*Ubx*), *abdominal*-A (*abd*-A), and *Abdominal*-B (*Abd*-B) [55, 56]. The spatial and temporal expression patterns of these genes are regulated by a highly complex regulatory sequence that spans more than 300 kb. This genomic region is subdivided into nine autonomous regulatory domains, *abx/bx*, *bxd/pbx*, and *iab-2* through *iab-8* [57]. The sequences within each regulatory domain are responsible for establishing and maintaining the expression levels of each BX-C homeotic gene appropriate for a particular parasegment. The best understood example of homeotic gene regulation is the expression of *Abd*-B. Four enhancers are

responsible for activating *Abd*-B transcription, *iab-5* trough *iab-8*. The location of these enhancers on the chromosome is co-linear with the arrangement of the corresponding parasegment along the anterior-posterior body axis. For example, *iab-5* drives the expression of *Abd*-B in parasegment 10 whereas *iab-8* defines parasegment 13 [58]. The expression levels of *Abd*-B established in each parasegment are maintained by the sequences termed Polycomb Response Elements (PREs) [59]. These elements serve as sites for the assembly of multi-protein complexes containing several *Polycomb*-group (*Pc*-G) proteins and *trithorax*-group (*trx*-G) proteins [60]. Through an unknown mechanism the repressive activity of *Pc*-G and activating activity of *trx*-G combine to create the chromatin state required for the appropriate expression of *Abd-B* throughout the lifespan. The molecular mechanism determining the specificity of the interaction between the *Abd*-B promoter and the appropriate *iab* enhancer and PRE in a parasegment-dependent fashion is currently unknown. One of the possibilities is that each set of the cis-regulatory elements determining *Abd*-B expression in a particular parasegment is isolated into an independent domain. The sequences within these domains become functional only in the cells of appropriate tissue (parasegment) [61]. Consistent with this view, three insulators, Mcp [62], Fab-7, and Fab-8 [63], were identified at the positions of predicted domain boundaries (Figure 2 A). How do distal enhancers, such as *iab*-5, overcome the boundary created by the more promoter-proximal insulators? A possible mechanism may be offered by the discovery of a novel cis-regulatory element, a promoter-targeting sequence (PTS), upstream of Fab-8 [64]. In transgenic assays the PTS has been shown to relieve the boundary effect created by Fab-8 or *gypsy* and allow the interaction of *iab-7* as well as several heterologous enhancers with a distal promoter. It is likely that elements similar to PTS are present in other locations within the *Abd*-B locus and serve to fine-tune the regulatory interactions in this highly complex region. Several

lines of evidence suggest that the Fab boundaries in the Bithorax complex may have a set of very unique characteristics not shared by other, less specialized, boundaries. When scs or *gypsy* boundaries are inserted into the genomic location of Fab-7 by gene conversion they do not faithfully recapitulate all the aspects of the Fab-7 boundary [65]. Surprisingly, *gypsy* loses its enhancer-blocking activity in the CNS at the Fab-7 locus. Also, a recent study [66] suggested that Fab-7, constitutively active in all examined tissues and developmental stages, consists of several sub-elements with developmentally restricted boundary function. It is possible that other boundaries from homeotic gene complexes will share some aspects of the structural and functional complexity of Fab-7.

The mammalian Igf2/H19 locus

The best understood example of the biological function of chromatin boundaries is in the mammalian Igf2/H19 locus [67, 68]. Epigenetic imprinting results in the expression of H19 only from the maternally transmitted allele, whereas Igf2 is expressed only from the paternal allele. A series of molecular genetic analyses led to the proposal that the cis-regulatory element responsible for the establishment of imprinting (Imprinted Control Region or ICR) lies between the two genes [69]. Further, the ICR was shown to have some of the properties of a chromatin boundary; namely some of the sequences within the ICR block distal enhancers from activating the promoters when positioned between the two in cell culture and transgenic mice [70]. The sequence analysis revealed that the mouse, rat, and human ICRs contain multiple copies of a sequence with strong homology to the CTCF-binding site from the chicken β-globin 5' insulator. There are seven copies of CTCF site in the human ICR and four in the mouse ICR. Biochemical tests confirmed that CTCF binds to these sites and that this binding is required for the enhancer-blocking activity. The position of the ICR boundary between the H19 enhancer and the Igf2

promoter explains why Igf2 is not expressed on the maternal allele. However, the activity of the Igf2 promoter would require that the boundary activity of the ICR be abolished on this allele. Since DNA methylation is the primary mechanism of epigenetic imprinting, it is likely that this modification modulates the state of the ICR boundary on the paternal allele. Indeed, several studies demonstrated that CTCF sites are methylated on the paternal allele, preventing the binding of the protein and establishment of the functional boundary [71, 72]. This proposal is supported by the observation that in mice deficient for DNA methylation the ICR boundary is active on both alleles and Igf2 is not expressed at all [73]. Also, in mice carrying the Igf2/H19 locus with a targeted deletion of the ICR the Igf2 gene is constitutively derepressed on the maternal allele [74]. Interestingly, the maternal allele appears to be capable of derepression at any time during tissue development, since temporally conditional deletion of the ICR boundary results in the activation of Igf2 [75]. This observation suggests that the boundary does not permanently change the status of enhancer-promoter communication, but rather modulates it in a dynamic and reversible fashion to provide a sensitive control of gene activity.

Scr-ftz-Antp **Region of the Antennapedia Complex**

The *Drosophila* Antennapedia complex consists of five homeotic genes determining the identity of the head and thorax of the fly [76]. The order in which the genes are found on the chromosome coincides with the order of their expression domains along the anterior-posterior axis of the body, a phenomenon termed co-linearity [77] (Figure 2 A). A unique feature of the Antennapedia complex is the presence of multiple non-homeotic genes spread among the five homeotic genes. Since the coordinated regulation of the homeotic genes is believed to be required for co-linearity, it is important to understand how non-homeotic genes in the cluster

evade these regulatory influences. A convenient model for addressing these questions is the *Scr-ftz-Antp* region that contains two homeotic genes (*Scr* and *Antp*) and a pair-rule gene *ftz*. Similar to the *Abd*-B expression in the Bithorax complex, the expression of these genes is governed by a complex system of cis-regulatory elements (Figure 2 B). The *Scr* transcript is first detected in parasegment 2 (PS2) during gastrulation [78, 79]. This pattern is specified by the T1 and PS2 enhancers, located more that 30 kb upstream of the *Scr* transcription start site, on the other side of the *ftz* gene unit [80, 81]. In germ-band extended embryos, *Scr* accumulates in the PS2 and lateral PS3 ectoderm. Later in development, *Scr* transcript continues to accumulate in the labial and prothoratic ectoderm, with additional accumulation in the sub-esophageal ganglion of the central nervous system. The VM enhancer directs the expression of *Scr* in the anterior and posterior parts of the visceral mesoderm [78]. The early expression pattern of *Scr* and other homeotic genes is primarily controlled by the gap and segmentation genes. These genes are transiently present during early embryogenesis and begin to disappear at gastrulation. The expression of homeotic genes, however, remains active through the rest of the development. This maintenance of the expression state in appropriate tissues is mediated by the PRE and TRE sequences [82]. Several such elements were characterized in the *Scr* region [83]. Interestingly, at least some of them are located near PS2 and T1 enhancers separated from the *Scr* promoter by the entire *ftz* transcription unit (Figure 2 B).

The *ftz* transcripts are initially generated in a continuous band of nuclei, in the region of the embryo destined to develop into the germ band [84]. At the time of cellularization, the *ftz* transcription is directed in the mesoderm by the promoter-proximal "zebra" (Z) enhancer in a characteristic seven-stripe pattern [85] (Figure 2 B). In general, both the stripes and the inter-band regions are approximately the same number of cells in width, although the seventh, most

posterior stripe is slightly wider than the others. This striped expression continues throughout gastrulation supported by the autoregulatory mechanism through the *ftz* distal enhancer and AE1 [86]. Both elements appear to be responsible for the expression of *ftz* in the embryonic ectoderm. Interestingly, the *ftz* distal enhancer sequence also contains the Matrix/Scaffold Associated Regions (MAR/SARs) [87]. At the time when the germ band is almost fully extended, the striped expression fades. *ftz* then becomes expressed primarily in neural cells driven by the neural (N) enhancer [88]. Still later in embryogenesis, *ftz* transcript is also detected in the hindgut.

Despite their close proximity, the cis-regulatory elements in the *Scr-ftz-Antp* region faithfully interact with the appropriate promoters, resulting in a unique spatio-temporal expression pattern for each gene (Figure 2 B). Several mechanisms may be involved in determining the specificity of the enhancer-promoter interactions in this regulatory region [89]. One of them relies on the intrinsic affinity of enhancers for a particular type of promoters. Several examples of such specific interactions have been recently described. The AE1 enhancer is capable of activating a wide range of TATA-containing and TATA-less promoters, such as *ftz*, *eve*, *white,* and *Scr*, in transgenic constructs. However, when AE1 is placed between a TATA-containing and a TATA-less promoter it specifically interacts with the former, a phenomenon termed promoter competition [89, 90]. However, only a limited set of enhancers may rely on promoter competition for their target selection. Another possible mechanism requires chromatin boundaries to prevent the interactions between a promoter and an inappropriate enhancer [22]. Consistent with this idea, we found a boundary element, SF-1, in the Scr-*ftz* intergenic region [91]. SF-1 is positioned approximately 9 kb upstream of the *Scr* promoter, separating all known upstream enhancers of *ftz* from the *Scr*. Additionally, all known *Scr*-proximal PRE/TRE sequences required for the expression of this gene are separated by SF-1 from the *ftz* domain.

The *ftz* transcription unit is evolutionarily mobile [92, 93]. In some fly species such as *D. littoralis* and *D. hydei* the region encompassing the *ftz* gene and its cognate enhancers (often referred to as the Powell conserved region) is inverted with respect to the *D. melanogaster* sequence. The position of SF-1 coincides with the border of the Powell conserved region. It is conceivable that a putative boundary, SF-2, delimits the other border of this mobile region (M. Li, unpublished). If the two boundaries interact, as proposed by the looping model, the upstream enhancers T1 and PS2 would be brought into close physical proximity to the *Scr* promoter, facilitating their interaction. An alternative model to explain this long-distance interaction has recently been proposed [94]. The DNA region located upstream of the *Scr* transcription start site was shown to contain several sequence motifs that are capable of "tethering" the T1 enhancer to the promoter. This "tethering" element appears to be promoter non-specific, since attaching it to the heterologous *ftz* promoter results in potentiation of the *ftz*-T1 interaction.

It is likely that a combination of the three mechanisms (promoter competition, boundaries, and "tethering" elements) is required to achieve the level of precise transcriptional coordination observed in homeotic clusters.

Purpose of the Study

Chromatin boundaries play a fundamental role in maintaining transcriptional independence of individual genes and gene clusters. Boundaries have been identified in eukaryotic organisms ranging from yeast to mammals [22]. In many cases the protein factors that interact with the boundary DNA and possibly mediate their function have been isolated [30]. In *Drosophila* these factors form a diverse group of structurally unrelated proteins, whereas in vertebrates all known boundaries appear to depend on one protein, CTCF. Several models have

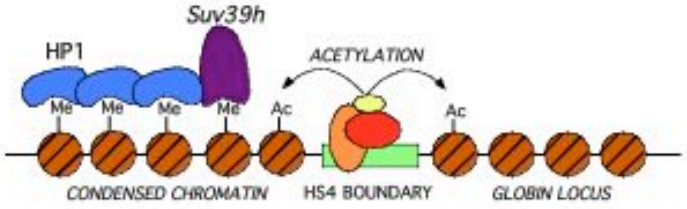
been proposed to explain the mechanism by which boundary elements block enhancers and stop propagation of active or silent chromatin. Despite extensive efforts of many laboratories in this direction a detailed understanding of the molecular mechanism is currently unavailable.

Our laboratory has been investigating gene regulation in the *Drosophila* Antennapedia complex, one of the two homeotic clusters found in the fly. We have been particularly interested in the centromere-distal part of the cluster that contains two homeotic genes, *Sex combs reduced* and *Antennapedia,* and a pair-rule gene *fushi-tarazu*. The architecture of this region is perfectly suited for addressing the question of how regulatory independence of genes is achieved. Despite close proximity of the *Scr* and *ftz* gene promoters and their corresponding cis-regulatory elements, these genes are expressed in drastically different spacio-tempotal patterns. One or more chromatin boundaries may contribute to the regulatory independence of the two genes. Consistent with this hypothesis a novel chromatin boundary, SF-1, was discovered by our laboratory in the *Scr-ftz* intergenic region.

The work presented in this dissertation pursued two main objectives. First, we aimed to obtain evidence supporting the biological role of SF-1 by modeling the behavior of several cis-regulatory sequences from the region in transgenic flies. Second, we sought to gain insights into the molecular mechanism of SF-1 boundary. The sequence elements responsible for the observed insulator and barrier activity of SF-1 were identified and characterized. This allowed us to biochemically isolate candidate SF-1 trans-factors. We tested the relevance of these candidates for the enhancer-blocking function of SF-1 and identified Enhancer of bithorax [e(bx)], as a novel protein component of SF-1.

Figure 1. Three models of the mechanism of chromatin boundaries. (A) According to the chromatin structure model boundary sequences recruit chromatin modifying enzymes. In the example shown the HS4 boundary from the chicken β-globin locus recruits histone H3 K9 acetylase that modifies histones nearby. This modification prevents the K9 methylation by Suv39h. Since the recruitment of HP1 and the spread of henerochromatin are dependent on the K9 methylation, the acetylation of this residue stops propagation of condensed chromatin. **(B)** The "promoter-decoy" model. One of the proposed mechanisms of enhancer function requires the physical interaction between the enhancer-bound transcription activators (colored ovals) and the core promoter. The insulator DNA sequence mimics the structure of the core promoter by recruiting the components of the general transcription machinery. When the insulator is present between the enhancer and a gene promoter, the transcription factors bound to the enhancer become "trapped" by the insulator complex and can not reach the promoter. **(C)** The "looping model" suggests that the proteins bound to the insulator sequences interact with each other in the nucleus and form the "insulator bodies" (green ovals). Such interactions form chromatin loops. If an enhancer and a promoter are located in separate loop domains their interaction is precluded. The loop domains may also be formed by the attachment of the chromatin fiber (blue ribbon) to the nuclear pore complex (light blue oval), nucleolus (green oval) or nuclear matrix components (grey mesh).

A


Suv39h
HP1
ACETYLATION
Me
Me
Me
Me
Ac
Ac
CONDENSED CHROMATIN
HS4 BOUNDARY
GLOBIN LOCUS

B


ENHANCER
PROMOTER
ENHANCER
INSULATOR
PROMOTER

C

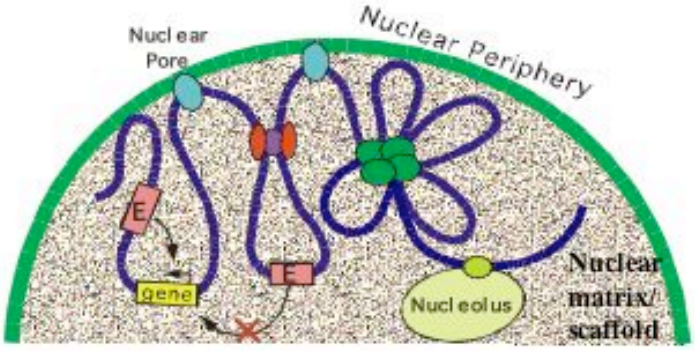
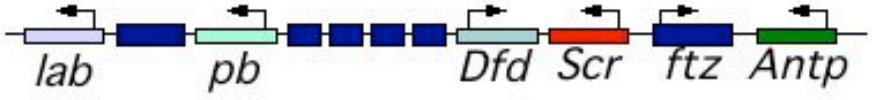
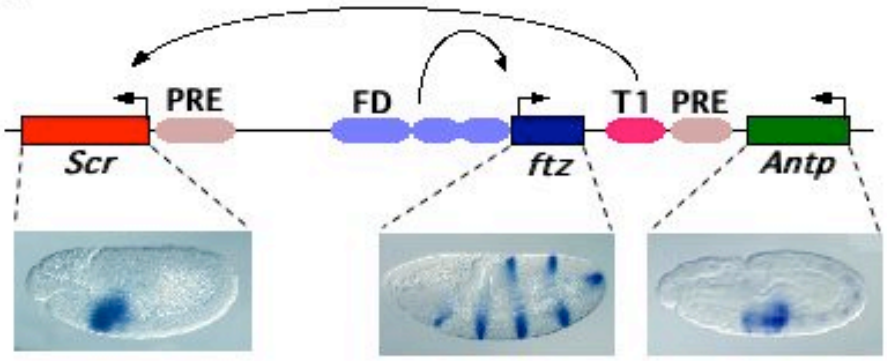

Nuclear Pore
Nuclear Periphery
E
gene
E
Nucleolus
Nuclear matrix/ scaffold

Figure 2. The *Drosophila* Antennapedia complex and gene regulation in the *Scr-ftz-Antp* region. (A) A schematic representation of the genes in the *Drosophila* Antennapedia complex. The homeotic genes are shown as colored boxes; non-homeotic genes, including *ftz*, are shown as blue boxes. The size of the genes and their relative positions are not to scale. **(B)** Some of the cis-regulatory elements required for the expression of *Scr* and *ftz*. Three embryonic enhancers of the *ftz* gene are shown as blue ovals; the *Scr* embryonic enhancer T1 is shown as a red oval; the Polycomb Response Elements required for the maintenance of the homeotic gene expression in late development are shown as purple ovals. The expression patterns of the three genes in the region are shown on the *in situ* images below the diagram.

A



B



CHAPTER 2

MOLECULAR GENETIC CHARACTERIZATION OF SF-1 INSULATOR

Background

The evolutionary conservation of the homeotic genes, both in function and organization, has been attributed to their important roles in animal body patterning, and to their coordinated regulation. Such coordination in the *Drosophila* homeotic complexes often involves extensive regulatory DNA and control elements that function over long distances. In genomic intervals where neighboring genes are closely positioned, long-range enhancers present a challenge for independent gene control. Two complimentary mechanisms are implicated in enhancer-promoter specification within complex genetic loci. The first mechanism, promoter competition, is the result of the preferential interaction between an enhancer and one promoter that reduces or excludes its interaction with other available promoters. An example is the AE1 enhancer of the fushi tarazu (*ftz*) gene in the *Drosophila* Antennapedia complex (ANT-C), which selectively activates the *ftz* promoter but not the neighboring *Sex combs reduced* (*Scr*) promoter (see diagram in Figure 1). This selectivity is due to the preference of AE1 for the TATA-containing *ftz* promoter, over the TATA-less *Scr* promoter, rather than its incompatibility with the *Scr* promoter.

The second mechanism involves the function of chromatin boundaries or insulators. These DNA elements can block transcriptional influences such as enhancer-promoter interactions and chromatin-mediated effects on gene expression. Chromatin insulator function

has been observed in HOM/Hox complexes of several species. In mouse, the functional range of the global hernia and digit enhancers flanking the Hox d10-d13 genes appears to be restricted by chromatin boundary element(s) positioned between the d13 and d11 genes. In the *Drosophila* bithorax homeotic complex (BX-C), multiple boundary elements, including Mcp-1, Fab7, and Fab8, have been identified between the tissue-specific *iab* enhancers in the regulatory region of the *Abdominal B* (*Abd-B*) gene. Although these boundaries have been implicated in modulating the *iab-Abd-B* interactions and maintaining the autonomy between neighboring *iab* enhancers, little is known about the mechanism of their function.

Here we report the presence of a novel chromatin insulator, SF-1, in the Scr-*ftz* region in *Drosophila*, the first such activity identified in the ANT-C. The SF-1 activity persists throughout the animal life cycle, consistent with its role in regulating homeotic genes. Like the Fab-7 insulator from the *Drosophila* BX-C, the highly conserved SF-1 core sequence contains multiple GAGA sites that are essential for its activity. The intergenic position of SF-1 and its ability to restrict promoter access by the promiscuous *ftz* distal enhancer suggest that SF-1 may direct enhancer trafficking in the Scr-*ftz* genomic interval.

Results

A novel enhancer-blocking activity in the Scr-ftz intergenic region

Although intrinsic properties of certain *ftz* enhancers, such as AE1, can account for their exclusive interaction with the cognate promoters, the same mechanism may not apply to all *ftz* enhancers in the region. Furthermore, the Scr distal enhancers, separated from the Scr promoter by the entire *ftz* gene, would have to overcome the interference from a highly competitive *ftz* promoter. To test if chromatin insulators play a role in defining enhancer-promoter interactions

in the Scr-*ftz* region, we examined DNA fragments from the Scr-*ftz* intergenic region for insulator activity. Two tissue-specific enhancers were used in the enhancer-blocking assay, the hairy stripe 1 enhancer (H1) and the rhomboid neuroectoderm enhancer (NEE), which are active in a transverse anterior band and two ventral lateral stripes, respectively. When a neutral 1.4 kb DNA spacer from the λ phage is inserted between the two enhancers, both the *lacZ* and white reporters are expressed in a composite pattern directed by both H1 and NEE, as shown by the whole mount in situ hybridization (construct NλH, Figure 3 A). Insertion of the 2.3 kb EcoRI fragment from the Scr-*ftz* intergenic region reduces the H1-directed white expression and the NEE-directed *lacZ* expression but does not affect the H1-directed *lacZ* expression or the NEE-directed white expression, indicating a selective block of the distal enhancer activities (NFH, Figure 3 B). The enhancer-blocking activity of the element, named SF1 for the Scr-*ftz* boundary, is comparable or even stronger than that of the suHw insulator from the gypsy retrotransposon. In contrast, other DNA fragments of comparable size from the 10 kb region surrounding SF1 exhibited little or no enhancer-blocking activity (H. Cai, personal communication). Importantly, the 16 kb intergenic region contains many closely spaced enhancers that are required for the tissue-specific regulation of Scr and *ftz* genes. The 2.3 kb SF1 region, however, appears to be devoid of any enhancer activities, as assayed in transgenic embryos with several promoters including those from the white, even-skipped (eve) and *ftz* genes. The insulator activity of SF1 is also orientation independent. When the 2.3 kb element is inserted in an inverted orientation between the NEE and H1 enhancers, it blocks the distal enhancers to a comparable level as in the forward orientation (NFrevH, Figure 3 C). Some variation of the enhancer-blocking strength of SF-1 as well as other insulators is observed when a population of embryos is examined. In order to account for this variation we utilized a semi-quantitative method to measure the enhancer-

blocking activity. The insulator activity in each appropriately staged transgenic embryo was described as strong, medium, or weak. The results of this analysis derived from a large number of embryos are shown in Figure 3 D.

The ability of SF-1 to block several other *Drosophila* enhancers was also tested in our lab (H. Cai, P. Shen). A different pair of embryonic enhancers, PE (twist proximal element) and E3 (eve stripe 3 enhancer) was tested first. When the lambda spacer is inserted between the two enhancers, they direct the white and *lacZ* reporter expression in the ventral region and in a mid-embryo stripe, respectively (construct Pλ3, Figure 4 A). Replacing the spacer with SF1 resulted in the block of the E3-mediated expression of the white reporter and the PE-mediated expression of the *lacZ* reporter (PF3, Figure 4 B). Again, SF-1 blocks the distal enhancers more efficiently than the suHw insulator (H. Cai).

SF1 is also active in adult *Drosophila*. Activity of the homeotic selector genes such as Scr is required to maintain the body segment identity throughout the animal life cycle. If SF1 is involved in regulating Scr and *ftz* genes, its boundary activity would be expected to persist to later stages of development. To test this, we examined the enhancer-blocking activity of SF1 in adult tissues with a transgenic yellow gene (H. Cai). The wild type activity of *yellow* is required for the pigmentation of cuticle structures in larval and adult *Drosophila*. (Figure 4 D). The yellow expression is activated in the adult bristles by the bristle-specific enhancer (Br, see diagram in Figure 3) located in the first intron of the gene. A transgenic mini-*yellow* gene that includes the 400 bp upstream sequences and the first intron can produce dark pigmentation in the bristles in a yellow null background (data not shown). Similar dark bristles are observed in flies carrying the transgene with the lambda spacer DNA inserted between the bristle enhancer and the mini-yellow gene promoter (Figure 4 E). When the full-length SF1 is inserted in place of the

spacer DNA, it efficiently blocks the bristle enhancer, reducing the bristle pigmentation to that of the *yellow* $^{-}$ mutant background (Figure 4 F). Again, the enhancer-blocking activity of SF1 is slightly stronger than that of the suHw insulator in a similar assay (Figure 4 G). Thus, the activity of SF1 is present in adult tissues, consistent with its potential role in regulating homeotic genes.

The core insulator sequence of SF-1 is highly conserved

In order to understand the enhancer-blocking mechanism of SF1 and identify its protein components, we sought to define the minimal sequences required for its insulator activity. The 2.3 kb SF1 was divided into three fragments of comparable size (SF1/a, SF-1/b, and SF-1/c) which were then individually tested for the enhancer-blocking activity using the NEE-H1 assay construct (Figure 5 A). The *lacZ* and white reporter expression show that the 700 bp SF-1/a and the 900 bp SF-1/c contain little or no enhancer-blocking activity (Figure 5 B and C). In contrast, SF-1/b strongly blocks the distal enhancers from the downstream reporter genes (Figure 5 B and C). Compared with the full length SF1, SF-1/b is only slightly weaker in blocking the NEE and H1 enhancers (compare Figures 3 D and 5 C). It is important to note that the full-length SF-1 contains seven recognition sites for the GAGA factor, the product of the *Trithorax-like* (*Trl*) gene (shown in Figure 5 B as red squares). A cluster of four of these sites is located within SF-1/b, coincident with the enhancer-blocking activity.

Further truncation of SF-1/b produced three overlapping sub-fragments (SF1/b1, SF-1/b2, and SF-1/b3) of equal length. These DNAs exhibit little enhancer-blocking activity when tested as monomers between NEE and H1 (Figure 6 A). We hypothesized that some of these fragments may still contain the sites necessary and sufficient for the enhancer-blocking activity, but the

truncation reduced their number, so that the activity dropped below the detection limit of the assay. To examine this possibility we tested the insulator activity of SF-1/b fragments in the form of tandem trimers. The multimerized elements showed striking differences in the enhancer-blocking activity. The SF1/b1 and SF-1/b2 fragments did not block distal enhancers (Figures 6 A and B), whereas the SF1/b3 fragment exhibited substantial enhancer-blocking activity (approximately 40% activity of the full-length SF1, Figures 3 D and 6 B). Since SF1/b1-b3 fragments contain significant sequence overlap at their termini (>50 bp), it is unlikely that the enhancer-blocking activity observed in the (SF-1/b3) trimer results from novel junction sequences produced by multimerization. Interestingly, the location of GAGA sites, again, coincides with the enhancer-blocking activity. The 275 bp sequence of SF-1/b3 contains a cluster of three GAGA sites.

Conservation in DNA sequences is often an indication of important biological function. To test if SF1 is evolutionarily conserved, we cloned and sequenced the SF-1/b homologous sequences from three fly species closely related to *D. melanogaster*, namely, *D. mauritiana*, *D. simulans,* and *D. teissieri*. As shown in Figure 7 A and B, the extent of sequence identity within the SF-1/b1 and SF-1/b2 regions decreases significantly with the increase in phylogenetic distance. However, the SF-1/b3 fragment, which contains the strongest insulator activity, remains >97% conserved across all four species separated by 2-5 million years. This degree of conservation is even higher than that of the coding region of the *yellow* gene. The enhancer-blocking activity of the SF1 sub-fragments correlates well with the level of evolutionary conservation of the DNA sequence. The high degree of conservation of the SF1 element in the absence of any detectable accompanying enhancer activity indicates that the insulator may play an important role in gene regulation.

GAGA sites and the 9 bp direct-repeat sequence are essential for the enhancer-blocking activity of the SF-1/b3 minimal insulator

The analysis of the SF1 sequence revealed multiple GAGA factor binding sites. A cluster of three of these sites is found in SF-1/b3 (Figure 8 A, shown in red). GAGA sites are frequently found in regulatory sequences of *Drosophila* homeotic genes and at the heat shock loci. Recently, they were also implicated in insulator/boundary function. It was reported that binding of the GAGA factor to a single GAGAG site in the *Drosophila even-skipped* promoter was essential for the insulator activity. GAGA sites are also required for the insulator function of the Mcp-1 and Fab-7 boundaries from the Bithorax complex. Additionally, we found two copies of a 9 bp sequence separated by 24 bp of DNA (Figure 8 A, shown in green). The arrangement of these direct-repeat sites is suggestive of trans-factor binding, although the TRANSFAC database search did not reveal proteins known to recognize this 9 bp sequence. Finally, we hypothesized that SF-1, an insulator from the homeotic gene complex, may share the molecular mechanism of function with other homeotic insulators. The matrix sequence comparison of SF-1/b3 and Fab-7 identified four 7-8 bp sequences present in both insulators (Figure 8 A, shown in blue).

We tested the functional significance of GAGA sites, direct repeats, and Fab-7 homology sequences in the minimal insulator SF1-b3, using site-directed mutagenesis. Replacement of all three GAGA sites in SF-1/b3 with unrelated sequences abolished its enhancer-blocking activity, making it indistinguishable from that in the lambda DNA control (compare activity charts in Figure 8 B, C, and D). This result indicates that the GAGA sites are essential for the enhancer-blocking activity of SF-1. The presence of a common protein component in several insulators from the *Drosophila* homeotic complexes suggests that they may belong to a conserved family of boundary elements important in regulating homeotic genes. The replacement of the 9 bp direct-

repeat sites with unrelated sequences resulted in the decrease of the number of embryos in the "strong block" category down to about 40% of the wild-type insulator (Figure 8 E). The number of embryos in the "weak block" category concomitantly increased. Interestingly, the impact on the insulator function produced by the direct-repeat site mutation is significantly milder than that of the GAGA site mutation, suggesting that direct-repeat interacting protein may serve as a secondary co-factor in the insulator complex. Lastly, changing the Fab-7 homology sites into unrelated sequences did not significantly reduce the insulator activity of the SF-1/b3 trimer (Figure 8 F), indicating that these sequences are dispensable for insulator activity.

G*AGA factor is involved in the enhancer-blocking activity of SF-1/b3.*

Our molecular genetic analysis indicated that GAGA sites are essential for the activity of the SF-1/b3 minimal insulator. However, a formal possibility remains that another trans-factor, different from the GAGA factor, binds to these sequences. Therefore, we directly tested if the GAGA factor is required for the insulator activity of SF-1. Since the GAGA protein level in early embryos is heavily influenced by the maternal contribution, we examined the SF-1 activity in embryos collected from the *trlR85* heterozygous females mated with the wild type males carrying the SF-1/b3 transgene. The SF-1/b3 mediated enhancer-blocking activity showed a small but consistent decrease in the *trlR85* mutant background (Figure 9), which may be due to the reduced GAGA protein level in these embryos. The reduction of the SF-1/b3 insulator activity appears to be less pronounced than that observed for the GAGA insulator in the *eve* promoter. This difference may be due to the clustering of GAGA sites in SF-1/b3 trimer. A total of nine GAGA sites are present in close proximity to each other in the trimerized fragment, an arrangement that may facilitate cooperative binding of the protein and protect the insulator from

the effect of reduced GAGA protein level. Alternatively, different flanking sequences in the GAGA binding region of SF-1 and *eve* insulators and the resulting differences in their binding affinity may account for the less sensitive response to the GAGA protein in SF-1. Finally, it is possible that GAGA binding proteins other than the GAGA factor are at least partly responsible for the SF1 activity.

G*AGA sites are not sufficient for the enhancer-blocking activity.*

The experiments described above demonstrated that the GAGA recognition sites are required for the enhancer-blocking activity. Their deletion completely abolishes the insulator activity of SF-1/b3. We next asked if a compact cluster of GAGA sites is sufficient to recapitulate enhancer blocking in the embryo. Two types of GAGA sites are present in SF-1, GAGAG and GAGAA. We tested clusters of GAGA sites of each type for their ability to block enhancers in the transgenic assay. We generated tandem arrays of GAGA sites containing eight copies of each sequence and placed them between the NEE and H1 enhancers. In either case the transgenic embryos exhibited a reporter gene expression pattern indicating the lack of insulator activity (Figure 10). We conclude that the GAGA sites alone are not sufficient for enhancer blocking, at least in the form of compact arrays. It is possible that the number of sites and/or the precise spacing between them are critical for them to manifest enhancer-blocking properties. It seems more likely, however, that other sites present in insulator sequences are essential for the activity. Proteins bound to these sites may facilitate recruitment of GAGA factor or interact with it and form complexes that can act as insulators by themselves or in cooperation with other factors that lack DNA-binding activity. The involvement of additional sites is likely since the GAGA binding site is 5 bp long and occurs at high frequency in the genome. The direct-repeat

sites in SF-1/b3 may act in cooperation with GAGA sites to form the insulator. They do not appear to be active in the absence of GAGA sites, since there is no residual insulator activity in the GAGA-mutant SF-1/b3 fragment (Figure 8 D).

SF-1 may direct enhancer trafficking and maintain independence of the Scr and ftz gene regulation.

The location of SF-1 raises the possibility of its role in maintaining the regulatory independence of the *Scr* and *ftz* genes. Although promoter specificity of the *ftz* AE1 enhancer depends on competition from the TATA-containing *ftz* promoter, chromatin insulator function may be necessary to prevent *Scr* enhancers from interfering with *ftz* expression, or to prevent other *ftz* enhancers from influencing *Scr*. We investigated this possibility by testing the promoter preference of the most outlying (and closest to *Scr*) enhancer of *ftz*, the *ftz* distal enhancer. The native *Scr* promoter could not be used for this purpose, since the "core sequence" drives expression of the linked reporter only very weakly and does not faithfully recapitulate the properties of the promoter. Including the sequences upstream and downstream of the core promoter, on the other hand, attenuates expression due to the presence of the PRE activity (data not shown). To circumvent these problems we modeled the *Scr* and *ftz* promoter pair with another pair of well characterized divergently transcribed TATA-containing and TATA-less promoters, *white* and *eve*. The 1.2-kb *ftz* distal enhancer was placed between the two promoters in our double-reporter construct. Multiple transgenic lines carrying this construct were generated. In situ hybridization of transgenic embryos showed that the *ftz* distal enhancer, indeed, strongly activates both the TATA-less *white* promoter and the TATA-containing *eve-lacZ* fusion gene (Figure 11 A). Next we tested the ability of SF-1 to block the interaction of the *ftz* distal

enhancer with the TATA-less *white* promoter by inserting the SF1b element between them. The *white* expression directed by the *ftz* distal enhancer is greatly attenuated (Figure 11 B). Our findings show that different *ftz* enhancers exhibit distinct promoter preferences and may use alternative mechanisms to select their target promoter. The position and the ability of SF-1 to prevent the *ftz* distal enhancer from activating a TATA-less promoter suggest that it may be essential in maintaining the independent gene regulation in the region.

Discussion

Homeotic gene complexes emerge as an excellent model for studying genetic programming of development, and mechanisms of transcriptional regulation. The independent, yet coordinated control of multiple genes by multiple regulatory elements provides a unique opportunity to probe the diverse mechanisms governing the interplay between gene organization and gene regulation. In the Scr-*ftz* region in ANT-C, at least three distinct types of cis-acting elements define the promoter specificity for no less than ten different enhancers. Enhancers such as AE1 distinguish the available promoters based on the core promoter sequence and selectively interact with the TATA-containing *ftz* promoter [89]. The *Scr* distal T1 enhancer appears to depend on a newly identified "promoter tethering element" located near the Scr gene for specific interaction [94].

Here we present evidence that a third type of enhancer-promoter modulation mechanism, the involvement of a chromatin boundary, may be responsible for target promoter specification by the *ftz* distal enhancer. Our findings suggest that the *ftz* distal enhancer does not share the same promoter preferences as AE1 and can equally activate TATA or TATA-less promoters. The intergenic position of the SF-1 chromatin boundary at the junction of the *ftz* transcriptional

unit and the neighboring *Scr* gene, and its ability to block the *ftz* distal enhancer from a TATA-less, *Scr*-like promoter suggest that SF-1 may be essential in maintaining the independent gene regulation in the region. Consistent with this proposed role in regulating the *Scr* homeotic gene, the boundary activity of SF-1 persists through the adult stage of development. Another indication of the functional role of the SF-1 boundary in the genomic interval is the conservation of the DNA sequence of this element during evolution. While the flanking region has diverged significantly (76% identity) in *D. teissieri*, the core insulator sequence remains highly conserved (>97% identity) in this species.

However, it is unclear how SF-1, an insulator positioned within the Scr regulatory region, is circumvented by the *Scr* distal enhancers located downstream of *ftz*. Similar questions exist for the Mcp-1, Fab-7, and Fab-8 boundaries between the *Abd-B* promoter and the distal iab enhancers in BX-C. A specialized DNA element named Promoter Targeting Sequence (PTS) near the *Abd-B* promoter may facilitate the ability of the enhancers to overcome the intervening Fab boundaries. We propose an alternative mechanism based on the recent finding that the suHw insulator activity is abolished by its tandem arrangement. According to our model, SF1 may interact with a putative insulator, SF2, positioned downstream of *ftz* (Figure 5E). The interaction between the two insulators not only neutralizes their ability to block the interactions between the flanking Scr promoter and the enhancers, but also promotes these interactions by "looping-out" the intervening *ftz* domain.

Chromatin boundary function has been shown to be important for gene regulation in the Hox cluster from fly to mouse. However, the protein components involved in the Hox boundary activity, as well as the mechanism of the boundary function, are unknown. We have identified multiple GAGA factor sites that are essential for the enhancer-blocking activity of the SF-1 core

insulator. We have also shown that the *Drosophila* GAGA factor is involved in SF-1 boundary function. Similar findings that GAGA sites are critical for the function of Mcp-1 and Fab-7 boundary elements from the BX-C have been reported recently (P. Schedl, personal communication). These observations suggest that the chromatin insulators from the ANT-C and the BX-C may share common components and mechanism, and belong to a family of conserved boundary elements that regulate enhancer-promoter interactions in the Hox complexes.
It is interesting that the GAGA factor is implicated in the boundary activity in the *Drosophila* Hox clusters. The GAGA factor has been known to regulate transcription by recruiting chromatin remodeling and transcription initiation complexes. However, its role in boundary/insulator activity may not be attributed to its ability to activate transcription but rather to the ability of this protein to forge links among distant DNA elements through its BTB domain. These properties of the GAGA factor are consistent with the looping models proposed for insulator/boundary mechanism.

The hypothesized independent *ftz* transcription domain flanked by boundary elements is also consistent with the observed mobility of *ftz* during evolution. *ftz* is an "accessory" gene unique to the invertebrate homeotic complex. Although it has been found in all major arthropod groups, the protein sequence and function of *ftz* have diverged from the neighboring homeotic genes. Nonetheless, the internal organization of the *ftz* transcription unit including regulatory sequences is highly conserved, possibly due to its important role in segmentation and neural development. The shift in *ftz* function appears to coincide with an increased mobility of the transcription unit as a whole, as the 16 kb genomic region is found inverted in certain *Drosophila* subgenera or missing entirely from the complex in certain insect species. The presence of the

SF1 chromatin boundary at the junction of such an evolutionary mobile unit is consistent with its role in maintaining gene independence during evolution.

Materials and Methods

P-element transformation, whole mount in situ hybridization and visual assessment of reporter gene expression

The $y^1w^{67}c^{23}$ and w^{1118} *Drosophila* strains were used to generate all transgenic lines reported. P-element mediated germ line transformation was carried out as described previously. Three or more independent trangenic lines were obtained and characterized for each test construct. Transgenic embryos were collected and fixed as described previously. Reporter gene expression in the blastoderm stage embryos was detected using the whole mount *in situ* hybridization with the digoxigenin-UTP labeled antisense RNA probes. Expression patterns were visualized by colorimetric reaction following incubation with anti-digoxigenin antibody conjugated to alkaline phosphatase (Genius Kit, Boehringer). The same amount of anti-sense probe was used in each *in situ* hybridization experiment. Fifty to two hundred blastoderm transgenic embryos from multiple random insertion lines were visually inspected by 2-3 individuals in a double-blind format using Zeiss Axioplan 2 microscope. The extent of enhancer-blocking was judged by the level of reporter expression directed by the distal enhancers compared to that of the proximal enhancers. Each embryo was assigned to one the three groups: weak (<30% block), medium (30-70% block) and strong (>70% block). The most frequently observed staining patterns were used to produce the image in the figures. Embryo images were taken using the Olympus DP10 digital camera.

Construction of transgenes and test of enhancer-blocking activity in mutant strains

All P-element constructs used in the embryo enhancer-blocking assays were derivatives of pCaSPeR. The *lacZ* coding region was fused in frame with the *eve* promoter (–42 to +200) and the *eve-lacZ* reporter was inserted into pCaSPeR in the orientation opposite from the *white* promoter. Two pairs of embryonic enhancers, NEE-H1 and PE-E3, were inserted into an *Eco*RI site between the two divergently transcribed reporters. All test DNA fragments were inserted between the enhancers into a unique *Not*I site. Construction of the NLH, PL3, NSH, and PS3 transgenes has been described previously. The 2.3-kb SF1 DNA was sub-cloned from a λ phage genomic clone that hybridized to probes from the Scr region (H. Cai, unpublished). The sub-fragments of SF1 were generated by PCR using the following pairs of primers containing *Not*I ends: 5'GCGGCCGCGAATTCGGTTTTCGAAGCC3' and 5'GCGGCCGCAACTATGGTAGCGCAGAGC3' for SF-1/a; 5'GCGGCCGCAGTGTTGCTGTAAGGACCG3' and 5'GCGGCCGCATTCTGAGCAGCGGAGTCG3' for SF-1/b; 5'GCGGCCGCTCCGCTGCTCAGAATTAGG3' and 5'GCGGCCGCGGATTCCCCATCCTATACC3' for SF-1/c. The PCR products were cloned into pCRII/TOPO vector (Invitrogen). These sequences were subsequently inserted into the *Not* I site of the NEE-H1 assay construct. The sub-fragments of SF-1/b were also generated by PCR. The following primer pairs were used: 5'AGTGTTGCTGTAAGGACCG3' and 5'CTTTAATGGTTTTGCCAGG3' for SF-1/b1; 5'ATATCCTTGCATTGCTTCC3' and 5'ATGGAAATTCACGCGTTCG3' for SF-1/b2; 5'GACAATTCGAATGTCAATGG3' and 5'ATTCTGAGCAGCGGAGTCG3' for SF-1/b3. The PCT products were cloned into the pCRII-TOPO vector, isolated by *Eco*RI digestion, and ligated at high concentration to yield multimers.

Clones containing trimers were selected using PCR analysis. Site-directed mutagenesis of various sites in the SF-1/b3 element was done using the single-stranded DNA method as described previously [95]. The following oligonucleotides were used to perform base substitutions in the SF-1/b3 element: 5'GCTGAAAACAAGCTTCATTGACATT3', 5'GTTTCAAGGCATCGATTGTTTTGTG3' and 5'ATTTCACTGGCTGCAGTTGCACATGT3' for the three GAGA sites; 5'GTGTGCTTCAGCGGCCGCTTCGCCAAAG3' and 5'TTTTTAGCGTCGGCGCGCCGCTTCTCGTG3' for 9 bp direct-repeat sites; 5'TTCTCCATTGGAATTCGAATTGTC3', 5'ATTATTGATTAAGCTTCTGACGCGTCG3', 5'TCGTGTTTTGTGTCTAGAACGTTTCAAG3' and 5'AGGAATTTTCTGGATCCTTTTTTAGCG3' for Fab-7 homology sites. The mini-*yellow* gene in bristle enhancer-blocking constructs was made using the *yellow* genomic region (from –400 to 400 bp downstream of the polyA site) provided by J. Zhou. An 1181 bp *Cla* I fragment in the first intron was deleted, and the *Eco*RV site at +778 was converted to a *Not*I site. The test DNA sequences, such as the λ spacer, SF-1, and su(Hw) insulator were inserted into the *Not*I site. The 1.2 kb *ftz* distal enhancer was provided by L. Pick. It was inserted either by itself or in combination with SF-1/b into the *Eco*RI site of the double-reporter vector to generate the constructs in Fig. 10. The following oligonucleotides were used to generate tandem arrays of the GAGA sites: 5'GGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA3' and 5'GGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC3' for the GAGAG site and 5'GGCCGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAA3' and 5'GGCCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTC3' for the GAGAA site. These arrays were inserted into the *Not*I site of the NEE-H1 vector. The position and orientation of enhancers and insulators were determined by restriction digestions, PCR analyses using P-

element specific primers, and in some cases by DNA sequencing. The integrity of test constructs after integration into the genome was verified by genomic PCR with transgene-specific primers. The SF-1/b sequences from various *Drosophila* species were generated by genomic PCR with *D. melanogaster* SF-1/b primers under low-stringency conditions, cloned into pCRII-TOPO vector (Invitrogen) and sequenced. To test the enhancer-blocking activity of SF-1/b3 in the GAGA mutant background, homozygous males carrying the test transgene were mated with heterozygous *trlR85* females. The reporter expression in embryos was determined by *in situ* hybridization.

Acknowledgements

I thank Haini Cai for identifying SF-1 and performing initial analysis of the full-length boundary in the NEE-H1 and PE-E3 double reporter constructs (these data are presented in Figure 2 A and B and Figure 3 A and B). I also thank Ping Shen and Haini Cai for testing SF-1 in the bristle enhancer constructs shown in Figure 3 E-G. I thank Alyssa Ingmundson for technical assistance with embryo counting in the semi-quantitative enhancer-blocking assay; Susan Schweinsberg, Paul Schedl, Krishna Bhat, John McDonald, and Wyatt Anderson for providing fly stocks; and Michael Levine, Pamela Geyer, and Paul Schedl for useful discussion. Most of the work presented in this chapter was published in The EMBO Journal [96]. I am grateful to the editors of The EMBO Journal for allowing me to include parts of the article in this dissertation. This work was supported by the NIH.

Figure 3. Enhancer-blocking activity of SF-1 in transgenic *Drosophila* embryos. (A-C) The reporter expression (*white* and *eve*/*lacZ* fusion gene) in blastoderm stage transgenic embryos visualized by the whole-mount *in situ* hybridization (see methods). Embryos are shown anterior to the left and dorsal side up. Each test transgene is shown below the embryo image. **(A)** The NλH embryos show a composite pattern consisting of comparable levels of NEE-directed ventral lateral expression and the anterior H1 stripe on the *white* and *lacZ* reporters. **(B)** NFH embryos show the reporter expression activated only by the proximal enhancers: NEE-directed ventrolateral stripes detected from *white* and H1-specific expression from the *lacZ* reporter. **(C)** NFrevH embryos, which contain the SF-1 element in reverse orientation, exhibit reporter expression by the proximal enhancers only: NEE from *white*, and H1 from *lacZ*. **(D)** Quantitative assessment of the enhancer-blocking activity of each transgene. Fifty to two hundred transgenic embryos from multiple lines were visually inspected for enhancer-blocking activity, which was categorized into weak, moderate or strong groups according to the level of reporter expression (see Materials and Methods for details). The most frequently observed staining pattern is used in the figure.

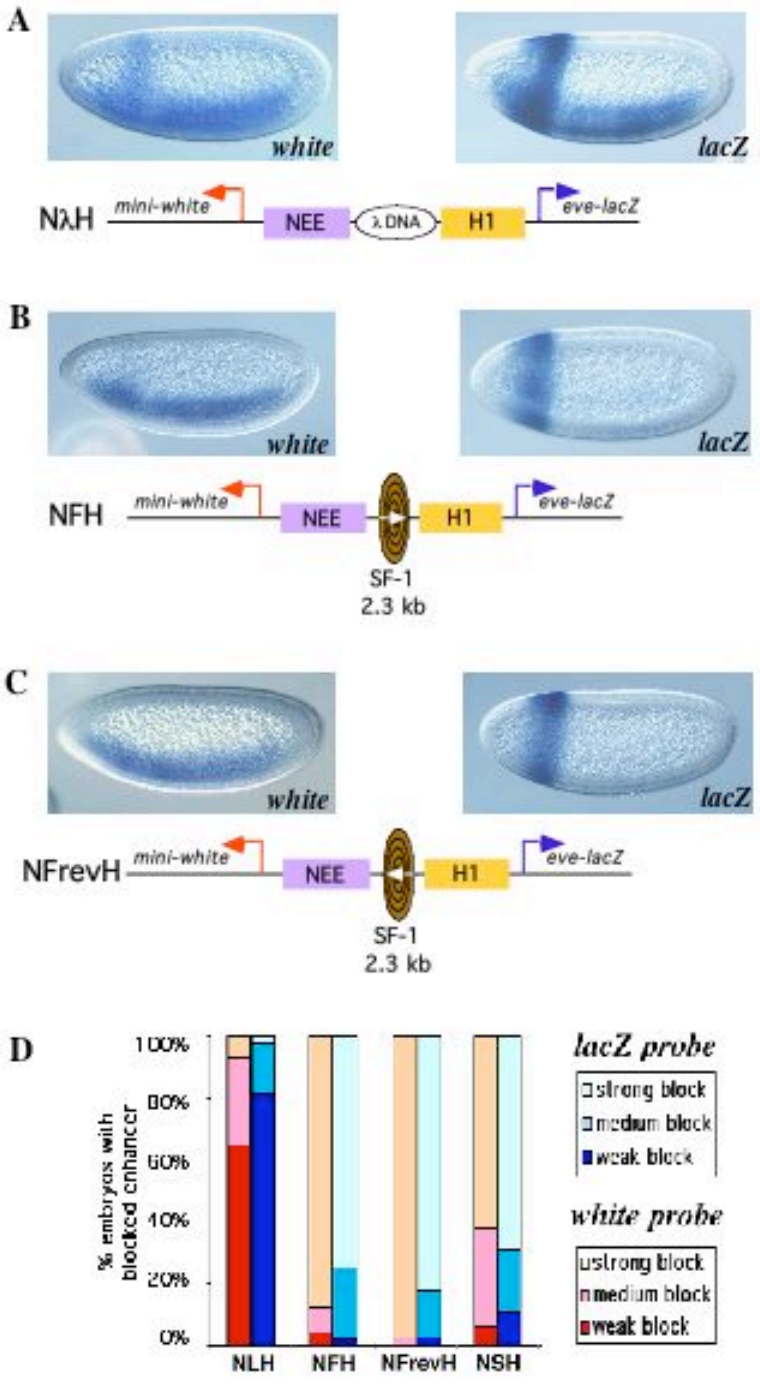

A
white
lacZ
NλH
mini-white
NEE
λ DNA
H1
eve-lacZ
B
white
lacZ
NFH
mini-white
NEE
H1
eve-lacZ
SF-1
2.3 kb
C
white
lacZ
NFrevH
mini-white
NEE
H1
eve-lacZ
SF-1
2.3 kb
D
% embryos with blocked enhancer
100%
80%
60%
40%
20%
0%
NLH
NFH
NFrevH
NSH
lacZ probe
strong block
medium block
weak block
white probe
strong block
medium block
weak block

Figure 4. SF-1 insulator is enhancer-independent and active in adult Drosophila. (A-C) SF-1 blocks a different pair of embryonic enhancers, PE (*twist* mesoderm enhancer) and E3 (*even-skipped* stripe 3 enhancer). The reporter expression (*white* and *eve*/*lacZ* fusion gene) in blastoderm stage transgenic embryos visualized by the whole-mount *in situ* hybridization (see methods). Embryos are shown anterior to the left and dorsal side up. Each test transgene is shown below the embryo image. **(A)** PL3 embryos show a composite pattern consisting of PE-directed ventral expression and E3-directed mid-embryo stripe on the *white* and *lacZ* reporter genes. **(B)** PF3 embryos show that only the proximal enhancer can activate reporter expression: PE-directed ventral stain detected from *white* and E3-specific expression from the *lacZ* reporter. **(C)** Quantitative assessment of the enhancer-blocking activity of each transgene. Fifty to two hundred transgenic embryos from multiple lines were visually inspected for enhancer-blocking activity, which was categorized into weak, moderate or strong groups according to the level of reporter expression (see Materials and Methods for details). The most frequently observed staining pattern is used in the figure. **(D-G)** SF-1 boundary activity in adult *Drosophila*. **(D)** The notum of a Canton-S adult female. Arrows indicate the macrochete bristles on the notum cuticle, both of which exhibit dark pigmentation. The notum of a $yellow^1$ adult female. Note the yellow-colored cuticle and bristles (arrows). **(E)** The notum of the adult female carrying the mini-*yellow* spacer transgene in a *yellow* mutant background, showing the restored pigmentation in the bristles due to the activity of the **Br** enhancer. **(F)** The notum of the adult carrying the mini-*yellow* SF-1 transgene. The bristles are yellow, indicating the lack of *yellow* expression due to the block of the **Br** enhancer by SF-1**. (G)** The notum of the adult containing the mini-*yellow Su(Hw)* transgene. Similar yellow bristles are seen, indicating the *Su(Hw)*-mediated block of the **Br** enhancer**.**

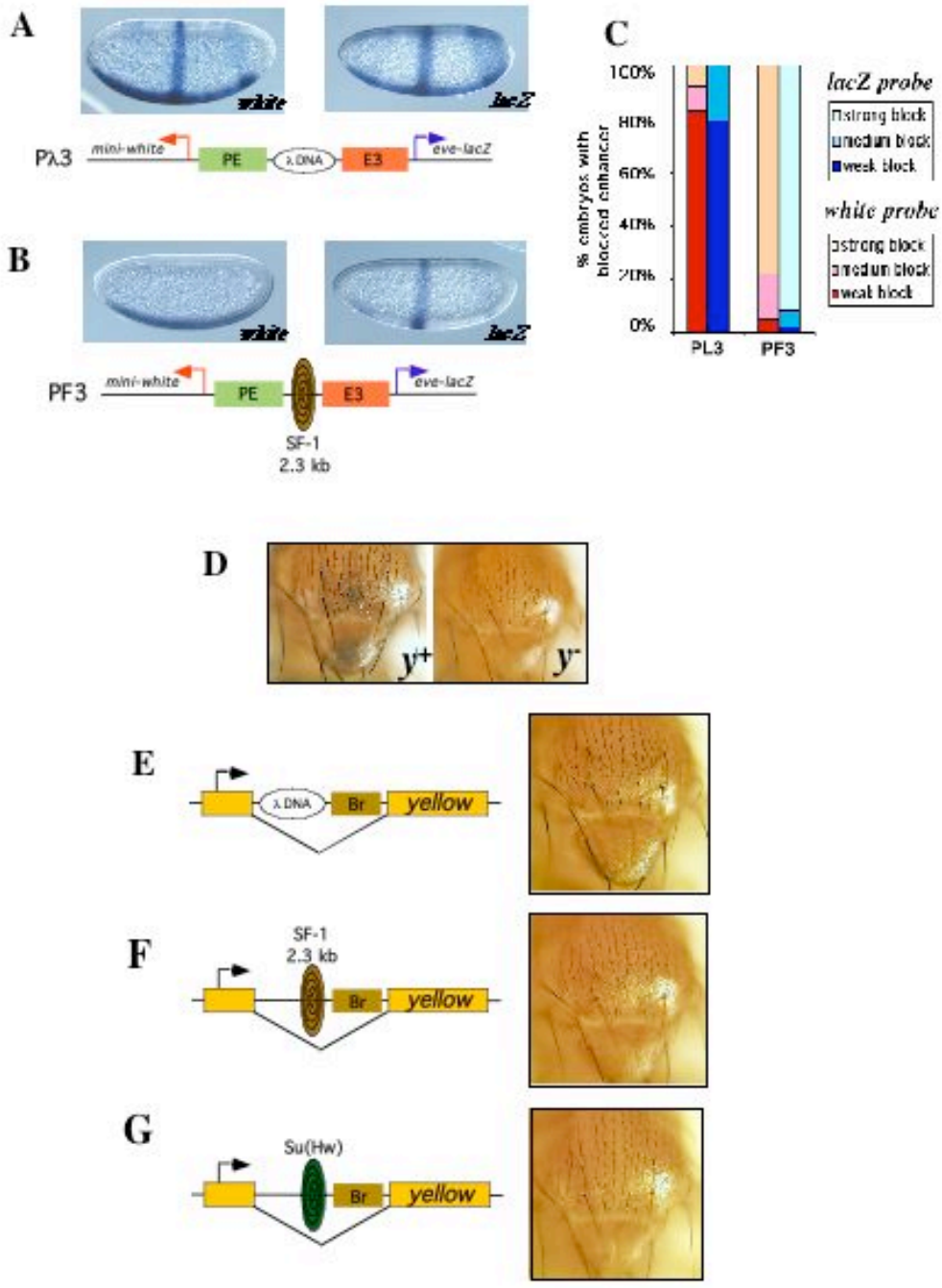
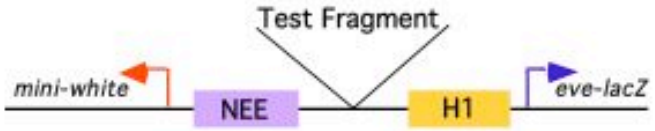
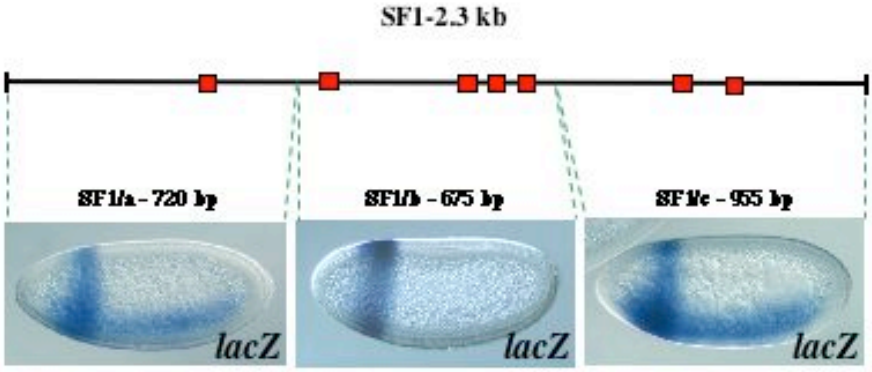

A
white
lacZ
Pλ3
mini-white
PE
λ DNA
E3
eve-lacZ
B
white
lacZ
PF3
mini-white
PE
E3
eve-lacZ
SF-1
2.3 kb
C
100%
80%
60%
40%
20%
0%
% embryos with blocked enhancer
PL3
PF3
lacZ probe
strong block
medium block
weak block
white probe
strong block
medium block
weak block
D
y+
y-
E
λ DNA
Br
yellow
F
SF-1
2.3 kb
Br
yellow
G
Su(Hw)
Br
yellow

Figure 5. The enhancer-blocking activity of SF-1 is located in the SF-1/b fragment. (A) A schematic of the NEE-H1 transgenic construct used to test the enhancer-blocking activity in SF-1 fragments. **(B)** The enhancer-blocking assay (only *eve-lacZ* reporter activity shown) using the NEE-H1 vector was performed with the SF-1 sub-fragments inserted between the NEE and the H1 enhancers. The whole mount *in situ* images of the embryos carrying the SF-1/a, SF-1/b, and SF-1/c fragments are shown. The diagram above the embryos indicates the position and the size of the SF-1/a, b, and c fragments within the context of the full-length insulator. Red boxes indicate GAGA sites. **(C)** Quantitative assessment of enhancer-blocking activity in the above transgenes with both *eve-lacZ* and *white* reporter activity (see Figure 3 and Materials and Methods for details).

A


Test Fragment
mini-white
NEE
H1
eve-lacZ

B


SF1-2.3 kb
SF1a - 720 bp
SF1b - 675 bp
SF1c - 955 bp
lacZ
lacZ
lacZ

C

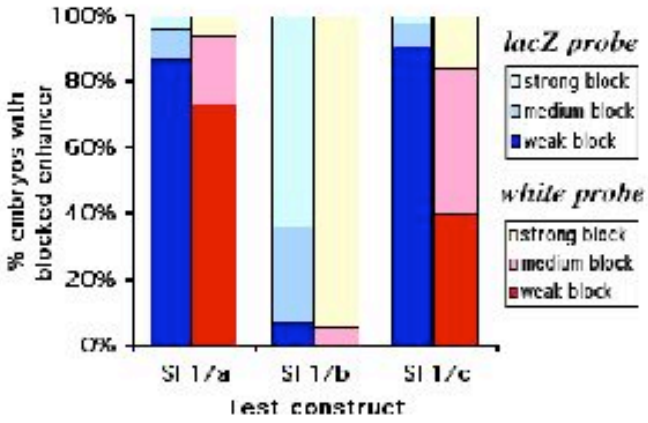
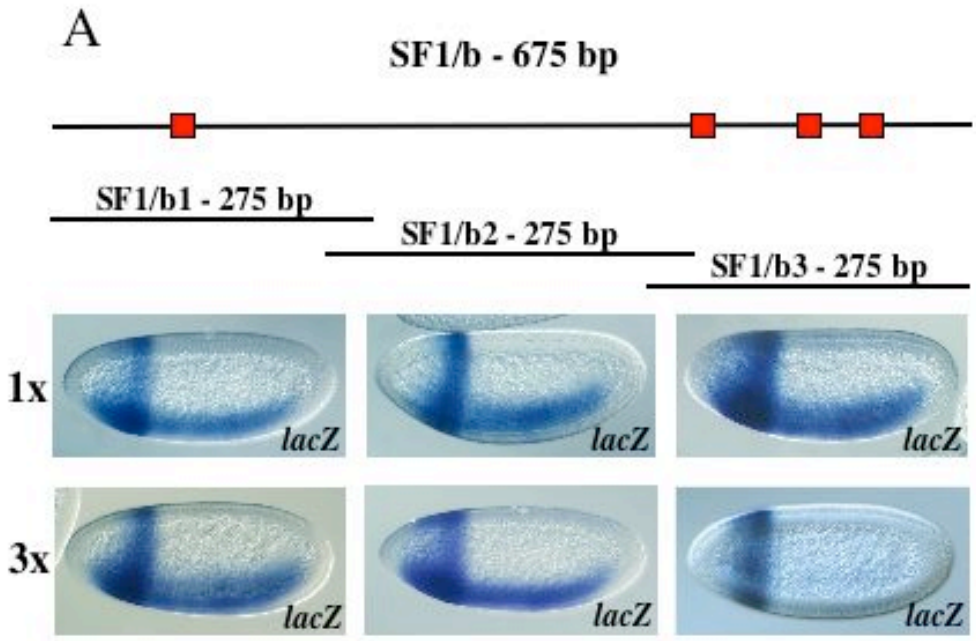

100%
80%
60%
40%
20%
0%
% embryos with blocked enhancer
SF1/a
SF1/b
SF1/c
Test construct
lacZ probe
strong block
medium block
weak block
white probe
strong block
medium block
weak block

Figure 6. The enhancer-blocking activity of SF-1/b resides in the SF-1/b3 sub-fragment. (A) Second round of enhancer-blocking tests using a single copy (top row) of SF-1/b1, SF-1/b2, and SF-1/b3 sub-fragments. All insulator tests were conducted using the NEE-H1 transgenic assay. Only the results of the *lacZ in situ* are shown. The diagram above the embryos indicates the position (orientation same as in Figure 5) and the size of the SF-1/b1, b2 and b3 sub-fragments within context of SF-1/b. Note the terminal overlap between neighboring fragments. Red boxes indicate GAGA sites. The enhancer-blocking activity of the trimerized sub-fragments is shown in the bottom row of embryo images. **(B)** Quantitative assessment of the enhancer-blocking activity of the above transgenes using the *eve-lacZ* reporter activity (see Figure 3 and Materials and Methods for details).


A
SF1/b - 675 bp
SF1/b1 - 275 bp
SF1/b2 - 275 bp
SF1/b3 - 275 bp
1x
3x
lacZ
lacZ
lacZ
lacZ
lacZ
lacZ

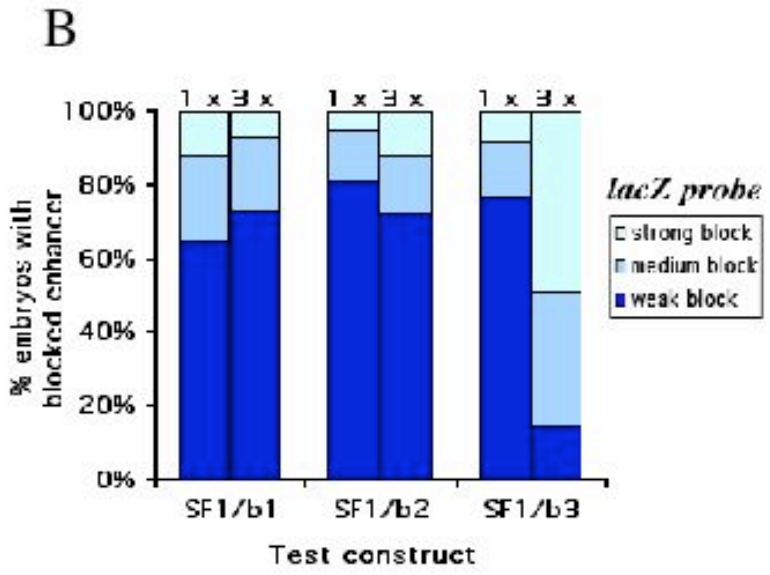
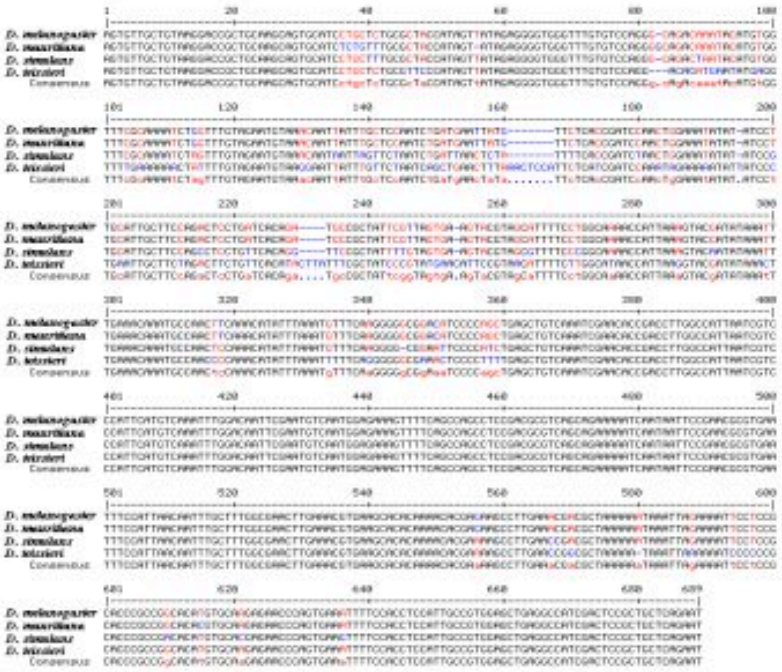

B
1 x 3 x
1 x 3 x
1 x 3 x
100%
80%
60%
40%
20%
0%
% embryos with blocked enhancer
SF1/b1
SF1/b2
SF1/b3
Test construct
lacZ probe
strong block
medium block
weak block

Figure 7. Evolutionarily conservation of the SF-1 enhancer-blocking sequence. (A) Multiple sequence alignment of the SF-1/b insulator sequences from four *Drosophila* species separated by approximately 5 million years. The identical bases are shown in black, bases identical to the consensus are shown in red and bases different from consensus are shown in blue. The enhancer-blocking SF-1/b3 sequence (located at the 3' end of SF-1/b) shows significantly higher degree of conservation among the species than SF-1/b1 and SF-1/b2. **(B)** Sequence identity comparison of the SF-1/b3 sub-fragment between the four *Drosophila* species. Numbers represent the percent nucleotide identity between SF-1/b1, SF-1/b2 and SF-1/b3 in the three species indicated and respective sequences in *D. melanogaster*. The conservation of the *yellow* gene coding region is given as an indication of evolutionary distance between the four fly species.

A



B

Nucleotide identity / DNA region / Species name	SF1/b1	SF1/b2	SF1/b3	yellow coding
D. mauritiana	99.1	100	99.6	98.3
D. simulans	91.5	94.8	98.1	98.1
D. teissieri	76.4	85.7	97.1	92.3

Figure 8. Identification of sequence elements required for the enhancer-blocking activity of SF-1/b3. (A) The DNA sequence of SF-1/b3 is shown with the following sequence elements indicated in color: GAGA sites (red), 9 bp direct repeat sequences (green), and Fab-7 homology sequences (blue). **(B-F)** The sequence elements required for the enhancer-blocking activity of SF-1/b3 were identified by site-directed mutagenesis. All tests were conducted using the NEE-H1 transgene. Only the results of *lacZ in situ* are shown for each test as the embryo image and quantitation chart. The sites mutagenized in each test are shown in the same color as in **(A)**. **(B)** The NλH transgene is shown as a control to illustrate the results of the assay with a neutral DNA element. **(C)** The enhancer-blocking assay with the three copies of the wild-type SF-1/b3. **(D)** The enhancer-blocking assay with the three copies of the SF-1/b3 sequence with the three GAGA sites changed into unrelated sequences in each copy. The insulator activity of SF-1/b3 is lost (compare to **(B)** and **(C)**). **(E)** The enhancer-blocking assay with the three copies of the SF-1/b3 sequence with the two 9 bp direct repeat sites changed into unrelated sequences in each copy. The loss of insulator activity is partial. **(F)** The enhancer-blocking assay with the three copies of the SF-1/b3 sequence with the four Fab-7 homology sites changed into unrelated sequences in each copy. The insulator activity is not affected (compare to **(C)**).

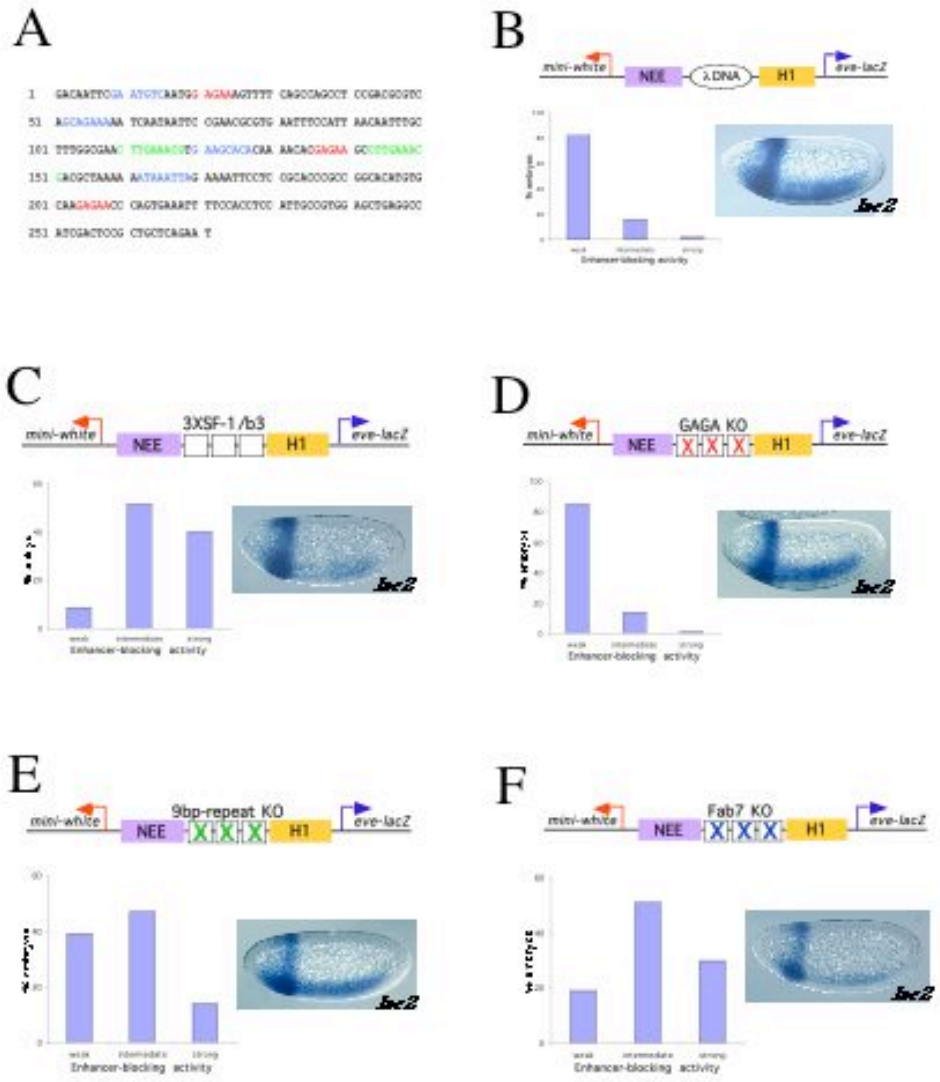

A
B
mini-white
NEE
λ DNA
H1
eve-lacZ
lacZ
C
3XSF-1/b3
D
GAGA KO
E
9bp-repeat KO
F
Fab7 KO
Enhancer-blocking activity

Figure 9. The GAGA factor is required for the insulator activity of SF-1/b3. (A) The *lacZ* expression in the wild-type embryo (top) carrying the transgene diagrammed above the *in situ* images. The reporter is expressed only in the domain of H1 enhancer, indicating a strong insulator activity of the SF-1/b3 trimer. The *lacZ* expression in the embryo (bottom) collected from the cross of the homozygous males carrying the NEE(SF1/b3)$_3$H1 transgene and the heterozygous trl^{R85} females. The reporter expression in the embryos was visualized by the whole mount *in situ* hybridization. Compared to the wild-type, a larger fraction of the mutant embryos exhibit the NEE-induced *lacZ* expression, suggesting that the enhancer-blocking activity of SF-1/b3 depends on the concentration of the GAGA factor. **(B)** Quantitative assessment of the enhancer-blocking activity of the above transgenes using the *eve-lacZ* reporter activity (see Figure 3 and Materials and Methods for details).

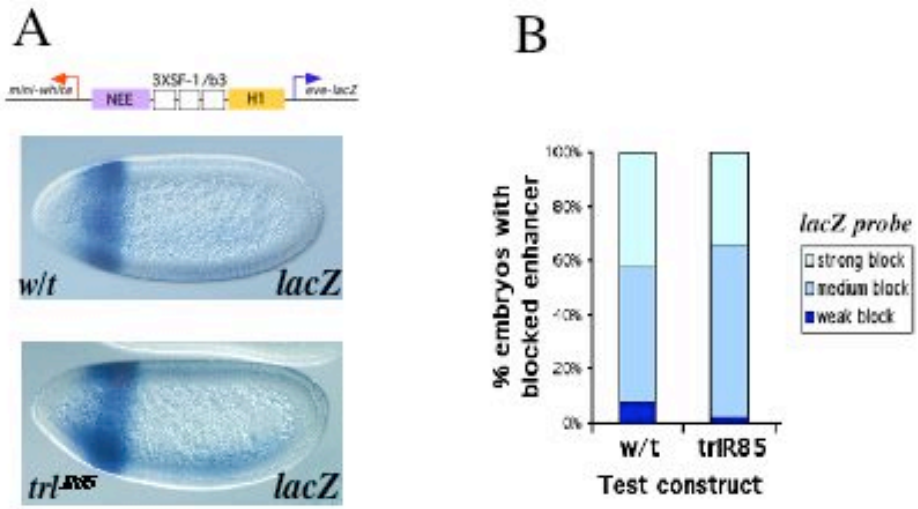
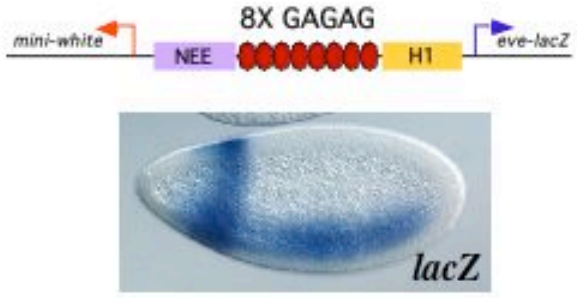

A
mini-white
3XSF-1 /b3
NEE
H1
eve-lacZ
w/t
lacZ
trl^R85
lacZ
B
% embryos with blocked enhancer
100%
80%
60%
40%
20%
0%
w/t
trlR85
Test construct
lacZ probe
strong block
medium block
weak block

Figure 10. GAGA sites are not sufficient for the enhancer-blocking activity. (A) A tandem array of eight GAGAG sites was placed between the NEE and H1 enhancers in the test transgene. The *lacZ* gene expression was analyzed in the transgenic embryos by whole mount *in situ* hybridization. The picture of a typical embryo is shown below the schematic of the transgene. No enhancer-blocking activity was detected. **(B)** The same assay was used to test the insulator activity of the GAGAA sites also known to bind the GAGA factor. These sites are not sufficient for the insulator activity either.

A


8X GAGAG
mini-white
NEE
H1
eve-lacZ
lacZ

B

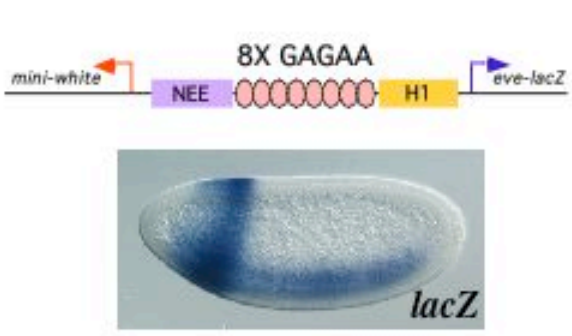
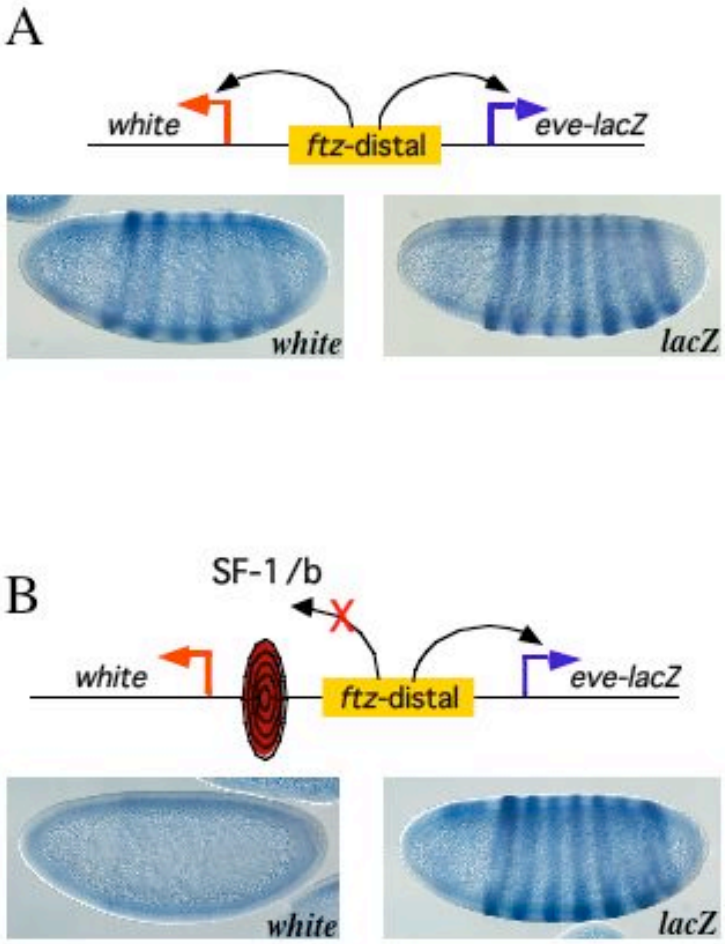

8X GAGAA
mini-white
NEE
H1
eve-lacZ
lacZ

Figure 11. Evidence for the potential role of SF-1 in defining the range of the *ftz*-distal enhancer. (A) The expression of the *white* and *lacZ* reporters was visualized by the whole mount *in situ* hybridization of the embryos carrying the transgene shown above each figure. In the 2-4 h old embryos the *ftz*-distal enhancer activates the expression of the TATA-less *white* promoter and the TATA-containing *eve* promoters at comparable levels. **(B)** The insertion of SF-1/b between the *ftz*-distal enhancer and the *white* promoter redirects the enhancer trafficking in the test transgene. In the majority of examined embryos the *white* reporter is not expressed, whereas the *lacZ* reporter is still activated by the *ftz*-distal enhancer.


A
white
ftz-distal
eve-lacZ
white
lacZ
B
SF-1 /b
white
ftz-distal
eve-lacZ
white
lacZ

CHAPTER 3

BARRIER ACTIVITY OF SF-1

Background

Chromatin organization affects gene activity in diverse organisms [for review see [97-100]]. In *Drosophila*, transgenes such as mini-*white* are influenced by the chromatin environment of the chromosomal insertion site, giving rise to a wide spectrum of eye colors in independent lines, a phenomenon termed Chromosomal Position Effect (CPE) [101, 102]. Similarly, integrated transgenes in mammalian cultured cells are often progressively silenced by the neighboring genome in an insertion site- and copy number-dependent fashion [103, 104]. Endogenous genes such as *white* and *yellow* can also be affected by chromosomal rearrangements that alter their proximity to heterochromatin, causing variegated expression in a field of cells known as Position Effect Variegation (PEV) [105-107]. Further, changes in chromatin structure facilitate coordinated gene regulation in complex genomic loci. In *Drosophila*, formation of silent chromatin domains, mediated by Polycomb Response Elements (PREs), is responsible for keeping the posterior homeotic genes off in the anterior body segments [108, 109].

The mechanisms of these chromatin-mediated effects are not fully understood. Some of them require distinct protein components. HP1 and other enhancer and suppressor of variegation proteins [*E(var)* and *Su(var)*] are involved in PEV, whereas *Polycomb* and *trithorax* group proteins (*Pc*-G and *Trx*-G) regulate homeotic gene expression [60, 110]. However, several

characteristics of these phenomena suggest that they may share a common mechanism: they are often gene non-specific; they require changes in chromatin organization such as nucleosome remodeling and histone modification; and, in many cases, these effects can spread along the chromosomes. The latter characteristic raises the question of how these influences are limited in the genome. Chromatin boundaries are DNA-protein complexes positioned at the ends of chromatin domains [111]. The roles of boundary elements in gene regulation are demonstrated by their ability to protect transgenes from positive or negative influences of neighboring chromatin (barrier function), and to block enhancers or silencers from downstream promoters (insulator function). Interestingly, most chromatin boundaries studied can function both as insulators and as barriers. Studies of the *Drosophila gypsy* boundary showed that these two functions require a set of simple DNA motifs and limited protein components [44, 112]. These results suggested that the two activities share the same underlying mechanism. However, recent studies of the chromatin boundary in the chicken β-globin locus provided evidence that the insulator and barrier functions can be mediated by distinct cis- and trans-factors [32].

How can the barrier activity of a DNA fragment of interest be determined? One assay commonly used in *Drosophila* relies on the sensitivity of the mini-*white* reporter to the transcriptional influences of surrounding chromatin [101]. The mini-*white* reporter was constructed by removing the major intron and most of the 5' regulatory sequences from the *white* gene [102, 113]. The sequences remaining in the promoter, 5'UTR, and possibly in several small introns are sufficient to activate a low level of transcription. The *white* cDNA encodes a transmembrane protein required for the transport of the red pigment precursor in the fly eye [114]. Depending on the amount of the *white* protein expressed the eye color can vary from light yellow to dark red. When a mini-*white* carrying transgene is inserted into various genomic

locations the surrounding chromatin influences the level of the reporter expression which results in fly lines with various eye colors (Figure 12 A, P. Majumder, unpublished). Random insertion of the transgene in the genome produces a comparable number of independent lines in various eye color categories. If two copies of the DNA fragment carrying a barrier activity flank the mini-*white* reporter transgene, the majority of transgenic lines have a uniform yellow eye color. The results of this assay for an SF-1-protected mini-*white* transgene are shown in Figure 12 B (P. Majumder, unpublished). This observation is consistent with the ability of a barrier to prevent spread of the activating or repressive influences of the genomic insertion site into the mini-*white* domain. Thus, the observed level of *white* expression is largely due to the default transcriptional state of the mini-*white* reporter [115].

The primary focus of the research described in Chapter II was to characterize the insulator activity of SF-1, a novel chromatin boundary from the Scr and *ftz* intergenic region in the *Drosophila* Antennapedia homeotic gene complex. A parallel study conducted in our laboratory aimed to understand the barrier activity of SF-1 (P. Majumder, D. Bosu, S. Roy). It was found that SF-1 contains a potent chromatin barrier that shields the mini-*white* transgene from the CPE. The barrier is located in a distinct DNA region (SF-1/c) from the previously reported insulator (SF-1/b) [96] (Figure 12 C). The barrier activity of several boundaries, including scs and *gypsy*, has been tested in the assay described above. Little is known, however, about the mechanism of the barrier function in higher eukaryotes. We previously determined that the GAGA factor is required for the insulator activity of SF-1. Thus, we asked if the same trans-factor is required for the barrier activity of SF-1.

Results

The barrier activity of SF-1 does not require GAGA binding sites

The role of the GAGA factor in the enhancer-blocking activity of SF-1 is consistent with the "looping" model of insulator function. The GAGA factor molecules bound to SF-1/b may interact with other similar complexes in the nucleus mediated by the homologous or heterologous oligomerization of the BTB domain. Chromatin loop domains formed by these interactions are refractory to the influences of enhancers located outside the loops. If the two barrier elements surrounding the mini-*white* reporter were to interact in a similar fashion, the barrier activity could perhaps also be explained by the formation of the loop domains. This hypothesis prompted us to test if the GAGA factor is involved in the barrier activity of SF-1/c. The SF-1/c sequence contains two GAGA recognition sites. We replaced both of them with random sequences by site-directed mutagenesis. The mutant barrier elements were placed on each side of the mini-white in a transgenic construct (Figure 13 A). Forty-three independent transgenic lines were generated with this construct by P-element-mediated germline transformation. The flies in each independent line were visually examined, and an eye color score was assigned to them on the scale of 2.0 to 5.9 (pale yellow to dark red). This eye color range was divided into four categories: 2.0 to 2.9 (pale yellow), 3.0 to 3.9 (dark yellow), 4.0 to 4.9 (orange), and 5.0 to 5.9 (red). The number of transgenic lines in each of these categories was counted, and the distribution for the GAGA-mutant transgene was compared with that for the wild type transgene (Figure 13 B).

In contrast to the insulator activity of SF-1/b, mutations in the GAGA sites do not compromise the barrier activity of SF-1/c. Instead, the proportion of transgenic lines with medium to light eye colors is increased above that seen for the wild type transgene, whereas

fewer lines with the dark eye color were observed. A possible explanation for this is that, in addition to its proposed role in forming chromatin loops, the GAGA factor serves as a transcription activator at many promoter sequences [116]. Therefore, the loss of GAGA binding would lead to attenuated transcription of mini-*white*. In fact, the SF-1/c element containing GAGA site mutations showed a significantly stronger barrier activity, judged by the standard deviation in the distribution of eye colors, than the wild type SF-1/c, or SF-1/a, suggesting that the presence of the activator binding sites partly masked its barrier activity. In summary, these results indicate that the SF-1/c barrier and SF-1/b insulator depend on distinct cis and trans components.

S*F-1 barrier, but not insulator, blocks variegated silencing by an Scr PRE.*

In the majority of cases, the CPE influences on mini-*white* are positive in nature, as the eye colors produced by the unprotected mini-*white* transgene are usually darker than that observed in flies carrying mini-*white* flanked by barriers. However, it has been well documented that blocks of condensed chromatin can propagate along the chromosome and repress the activity of nearby genes. Examples of such domains include the constitutive heterochromatin of centromeres and telomeres [117] and the transcriptionally silent chromatin regions organized by the *Polycomb* group proteins in the homeotic gene complexes [118]. Proximity of the mini-*white* to such chromatin domains results in the stochastic repression of the gene. The repressed state is maintained in a clonal fashion and leads to variegated or gradient eye color [106, 119].

To investigate whether SF-1 barrier is effective in buffering not only positive but also negative influences of chromatin, we tested its ability to block PRE-mediated silencing of mini-*white*. The 2.9 kb DNA region upstream of the *Scr* promoter displays several characteristics of a

PRE. When placed next to the mini-*white* it represses the expression of the reporter, resulting in extremely light, variegated or gradient eye color in 3 out of 4 independent transgenic lines Figure 14 B). This behavior is drastically different from the control transgene without the 2.9 kb element for which none of the six examined lines show eye color phenotypes characteristic of the PRE-mediated repression (Figure 14 A). In fact, hundreds of independent lines have been generated in our laboratory for the barrier activity assays and none of them displayed mosaic or patterned eyes, confirming that our observations on the PRE lines are significant. In addition, paring of transgenes containing the 2.9 kb PRE enhances the silencing of mini-*white* expression, a typical characteristic of Pc-G dependent repression. Finally, the repressive complexes are believed to be assembled at the PRE sequences during later stages of embryogenesis [108]. Therefore, a *bona fide* PRE can be distinguished from a general silencer by the absence of its activity during early embryogenesis. To test for the early activity of the 2.9 kb fragment we inserted the *ftz*-CAT cassette activated by the *ftz* distal enhancer into each of the PRE test constructs on the other side on mini-*white* (these cassettes are omitted in Figure 14 for simplicity). The pre-gastrulation embryos (3-4 h) were collected from the flies carrying the transgenes with or without the PRE and analyzed by whole mount *in situ* hybridization using the CAT probe. In contrast to the late mini-*white* expression, activation of the neighboring CAT reporter by the early *ftz* distal enhancer is not affected by the 2.9 kb *Scr* element, indicating that the silencing of mini-*white* is a canonical PRE-mediated effect.

We tested whether the SF-1/b insulator can block the spread of the PRE-mediated silent chromatin by inserting this fragment between the PRE and mini-*white*. Transgenic lines containing SF-1/b showed no reduction of PRE-mediated silencing, as seen by the low and variegated eye color in all three examined transgenic lines (Figure 14 C). By contrast, a

transgene containing the SF-1/c barrier between the PRE and the mini-*white* resulted in an almost complete block of silencing by the PRE, shown by the near normal levels of eye pigmentation and the absence of variegation (Figure 14 D). These results indicate that the SF-1/c chromatin barrier is distinct from the SF-1/b insulator in its unique ability to attenuate both the positive and the negative influences of chromatin organization on gene expression.

Discussion

We have shown that the 2.3 kb SF-1 boundary in the *Drosophila* Antennapedia complex contains two distinct activities: an insulator activity that blocks enhancer-promoter interactions, and a barrier activity that blocks the transcriptional influences from both positive and negative chromatin effects. Interestingly, these two activities reside in different DNA regions and require different protein factors for their function. Insulator and barrier activities have been reported for several chromatin boundaries including the *Drosophila gypsy*, scs, Fab-7 and the vertebrate β-globin 5'HS4 elements. In some of these boundaries the two activities appear to be inseparable. In particular, the su(Hw) boundary requires simple DNA sequence motifs and two well-characterized factors: the DNA-binding protein Su(Hw) and BTB-domain protein Mod(mdg4). Antibody staining of polytene chromosomes showed that Mod(mdg4) is found at all Su(Hw) interaction sites [47], suggesting that a single protein complex, and possibly a shared mechanism, mediate both insulator and barrier activities.

Separable insulator and barrier activities within the chicken β-globin 5'HS4 boundary have been reported [32]. In mammalian cell culture, tandemly integrated reporters are usually progressively silenced by neighboring chromatin, reminiscent of the CPE and the PEV phenomena in *Drosophila*. The CTCF binding site in 5'HS4, which is necessary and sufficient

for enhancer-blocking activity, is unable to block this silencing. Other sequences in the 5'HS4 are responsible for the barrier activity. Sequences outside of the CTCF site in the 5'HS4 can also block a heterologous MoMLV LTR silencer [120]. Here we describe the first study in *Drosophila* that demonstrates the independence of insulator and barrier activities. This result suggests that the selective and combinatorial use of insulator and barrier function may be a conserved strategy in chromosomal organization and gene regulation in both vertebrates and invertebrates. Although these two activities often co-exist in chromatin boundaries, our study indicates that they do not obligatorily associate with each other and may have different molecular mechanisms. The range of CPE, PEV, and PRE-mediated silencing, despite the different genetic contexts in which they were characterized, are limited only by chromatin barriers, but not by insulators. Our results suggest that these chromatin-mediated effects as a group may utilize common mechanisms that differ fundamentally from the mechanisms of action employed by enhancers/silencers. It should be noted that transcription regulation by enhancers and silencers often requires histone modifications or nucleosome remodeling, which change local chromatin structure [121]. However, chromatin structures that can sustain and propagate over long distances in the genome, such as those seen in PEV and PRE-mediated silencing, may require a distinct molecular mechanism. It is possible that these long-range chromatin effects involve relocating genes to distinct nuclear compartments, whereas enhancers/silencers act through local loop formation.

Both enhancer-promoter interactions and modification of chromatin are at play in the *Scr-ftz* genomic interval in the Antennapedia complex. The latter differs from other *Drosophila* or vertebrate Hox clusters in that it contains both homeotic and non-homeotic genes. Proper expression of these genes requires modulation of enhancer traffic between neighboring

promoters. In addition, PRE-mediated silencing, critical for maintaining the correct spatial pattern of the homeotic genes, should be blocked from spreading into the non-homeotic gene domains. The intergenic position of the SF1 boundary fulfills both requirements: the SF-1/b insulator restricts the long-range enhancers of *ftz* from interfering with the *Scr* promoter, and the SF-1/c barrier protects the *ftz* gene from the silencing effect of the *Scr* PRE (Figure 15 A). The existence of another boundary, SF-2, downstream of the *ftz* transcription unit would explain why the *ftz* enhancers do not influence *Antp*, and *Scr*-distal enhancers and PREs do not influence *ftz*. We are currently searching for the second boundary in the region. SF-1 and the putative SF-2, like other boundaries in the genome, may interact and "loop-out" the intervening *ftz* domain (Figure 15 B). Such interaction would not only ensure functional independence of *ftz*, but also facilitate the communication between the *Scr* promoter and its distal cis-regulatory elements. We propose that the separation and selective association of barriers and insulators could determine the regulatory role of compound boundaries and provide flexibility in their function.

Materials and Methods

P-element transformation, whole-mount in situ hybridization, and visual assessment of reporter gene expression.

The $y^1w^{67}c^{23}$ and w^{1118} *Drosophila* strains were used to generate all transgenic lines reported. P-element mediated germ line transformation was carried out as described previously [122] with the following modifications. The embryos were not dechorionated after collection and were dehydrated for 10 minutes in a 17°C injection room with ambient humidity. To enhance the efficiency of genomic integration of large PRE-containing transgenes the micro-injection was performed using *w*;;Δ2-3*Sb*/TM6*Tb* embryos that constitutively express P-element transposase

[123]. Fourty-three independent lines were generated for the CPE-blocking test with the GAGA-mutant SF-1/c barrier. Three or more independent transgenic lines were obtained for the *Scr* PRE-containing transgenes. The *ftz*-CAT transgenic embryos were collected, fixed, and analyzed by whole-mount *in situ* hybridization as described previously [89].

Assessment of CPE and PRE effects in mini-white transgenic flies

Eighteen or more independent lines were examined for each mini-*white* transgene used in the CPE-blocking assay (P. Majumder, D. Bosu). The eye color of 5 to 7 day old heterozygous females was visually assessed under 10x magnification using a Leica MZ6 dissecting microscope and intermediate illumination with an NCL150 cold light source. Each independent line was assigned an eye color score according to the standard scale. Each grade of eye color intensity in the chart was further defined by optical density at 480 nm (OD480). Eye pigment was extracted from 20 7-day old flies of indicated eye color as described previously [83]. Flies were homogenized in 100 µl AEA buffer (30% EtOH, 0.1% concentrated HCl) and brought to 1 ml by adding 900 µl AEA. The samples were then vortexed for 30 minutes and spun for 10 minutes in a microcentrifuge. Twenty microliters of 0.5% hydrogen peroxide was added to the supernatant to oxidize the extracted pigment. The samples were mixed, spun, and measured for absorbance at OD480 using a Genova Life Analyzer spectrophotometer. Each OD480 reading was repeated tree times, and the mean value was used. To determine if the *Scr* PRE-containing transgenes exhibit pairing sensitivity, heterozygous transgenic parents were mated and the eye color phenotypes of the age-matched progeny were inspected. Since the mini-*white* expression is dosage-sensitive, homozygous progeny are expected to have darker eyes than the heterozygous

parents. Therefore, significant deviations from the Mendelian ratio in the eye color of progeny showing fewer than 25% of dark-eyed flies is consistent with the pairing-sensitive repression.

Construction of transgenes

Full-length SF-1 and its sub-fragments SF-1/a, b, and c were generated by PCR using primers containing *Not*I site and cloned into pCRII/TOPO vector (Invitrogen). The resulting constructs were digested with *Not*I or *Nsi*I, and the DNA inserts were gel extracted, purified, and ligated into the respective sites flanking the mini-*white* reporter in the pCaSpeR transformation vector. Site-directed mutagenesis of the two GAGA sites in SF-1/c was performed using the single-stranded DNA method as described previously [95]. The base substitution in the GAGA sites was performed using the following oligonucleotides:
5'ACAATGAACAGGATCCTGATGAATTA 3' and
5'GTTGTGATGCAGATCTGCTTACTTAG 3'. Transgenes used in the NEE/H1 enhancer-blocking assays were described previously. The transgenes used in the PRE-blocking assays were constructed as follows. The *Scr* promoter region (including the first intron and 2.9 kb of upstream sequence) was PCR amplified and inserted into the *Not*I- and *Bam*HI-digested pCAeb vector, replacing the *eve* promoter. The *ftz*-CAT reporter was cloned into the *Not*I site of this construct. The *ftz*-distal enhancer and either SF-1/b or SF-1/c pairs were assempled in a separate pCRII vector and then placed between the CAT and *lacZ* reporters into a unique *Asc*I site.

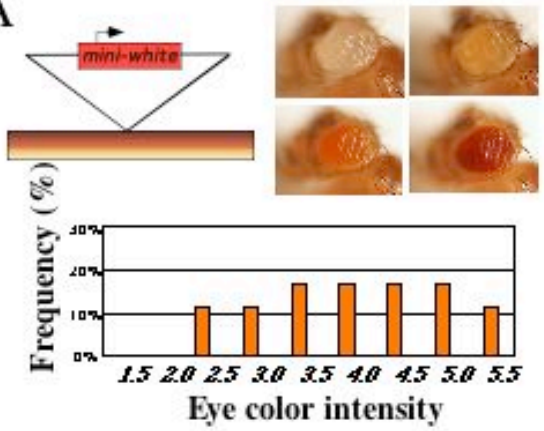
Acknowledgements

I wish to thank the many people whose collaboration I enjoyed on this project. Parimal Majumder Dimple Bosu discovered the separation of the enhancer-blocking and barrier activities

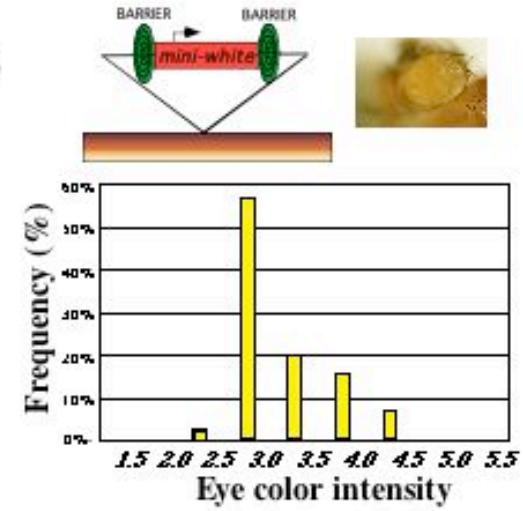
in the SF-1 boundary. I am grateful to Sharmila Roy for technical assistance with embryo microinjection and fly husbandry and to Haini Cai for invaluable guidance and discussion on the barrier project. This work was supported by the NIH.

Figure 12. CPE-blocking activity of the SF-1 chromatin boundary. (A) Examples of the eye color variation caused by CPE in transgenic flies containing the mini-*white* reporter (red box). Heads of a w^{1118} and three transgenic females show increasing level of expression of the mini-*white* reporter. The eye color intensity in each independent transgenic line examined was given a score of 1.5 (lightest) to 5.5 (darkest). The distribution of eye color phenotypes is shown on the bar graph for the lines carrying unprotected mini-*white*. **(B)** Schematic representation of the SF1-flanked mini-*white* transgene randomly integrated into the chromosome (brown bar). Arrow represents the mini-*white* promoter; ovals represent the SF-1 boundary. A typical eye color observed in these transgenic lines is shown on the photograph. The eye color distribution for the SF-1 protected transgenic lines is shown on the bar graph. **(C)** Testing the SF-1 fragments in the CPE protection assay revealed that only SF-1/c contains significant barrier activity (P. Majumder and D. Bosu, unpublished data). The SF-1/b fragment that contains most of the enhancer-blocking activity does not protect the mini-*white* from CPE.

A



B

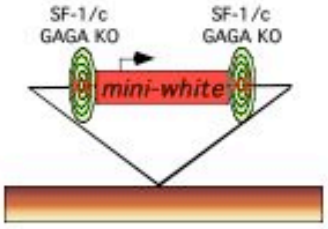


C

	SF-1	SF-1/b	SF-1/c
Insulator	Yes	Yes	No
Barrier	Yes	No	Yes

Figure 13. The barrier and insulator activities of SF-1 require distinct protein factors. (A) A schematic representation of the mini-*white* transgene (red box) flanked by the SF-1/c barrier sequences in which two GAGA sites were changed into unrelated sequences by site-directed mutagenesis. The brown bar represents a chromosome. **(B)** Forty-three independent transgenic lines carrying the construct shown in **(A)** were generated. Their eye color distribution is shown on the bar graph compared to the one for the wild-type SF-1/c transgene. Each segment of the bar (yellow, dark yellow, orange, and red) represents the number of lines in a particular eye color range (shown on top of the graph) relative to the total number of lines. The two distributions are very similar, indicating that the GAGA sites are not necessary for the barrier activity of SF-1/c. A slight decrease in the number of lines with dark eye color for the mutant barrier transgene probably indicates the loss of transcription activation by the GAGA factor.

A


SF-1/c
GAGA KO
SF-1/c
GAGA KO
mini-white

B

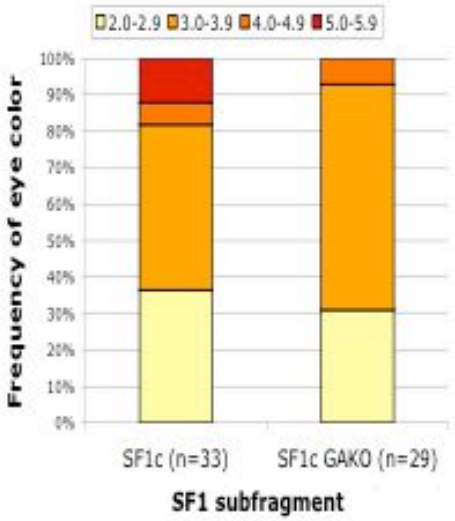

2.0-2.9
3.0-3.9
4.0-4.9
5.0-5.9
Frequency of eye color
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
SF1c (n=33)
SF1c GAKO (n=29)
SF1 subfragment

Figure 14. Chromatin barrier, not insulator, limits PRE-mediated silencing in the *Scr-ftz* region. (A-D) SF-1/c prevents silencing of the mini-*white* by the *Scr* proximal PRE. Representative eye colors of flies from independent lines are shown. Transgenic constructs are diagrammed below the head photographs. The mini-*white* expression is normal in transgenic flies carrying the control transgene without the PRE **(A)**, but silenced in flies carrying the *Scr*-proximal PRE as seen by the low and/or variegated eye color **(B)**. **(C)** The silencing effect is not attenuated by the SF-1/b element placed between the PRE and the mini-*white*, judged by the persistence of the variegated and/or gradient eye color phenotype in all three transgenic lines examined. **(D)** The PRE-mediated silencing appears to be completely blocked by the SF-1/c barrier placed between the PRE and the mini-*white*. All three transgenic lines examined have a normal eye color phenotype.

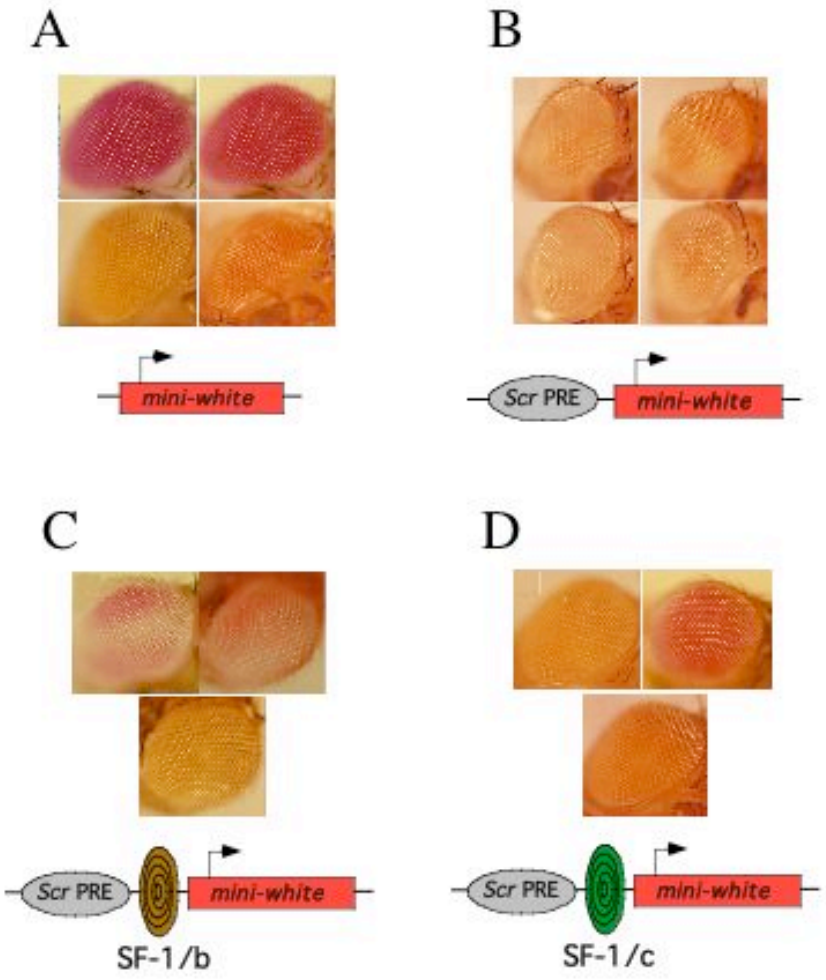
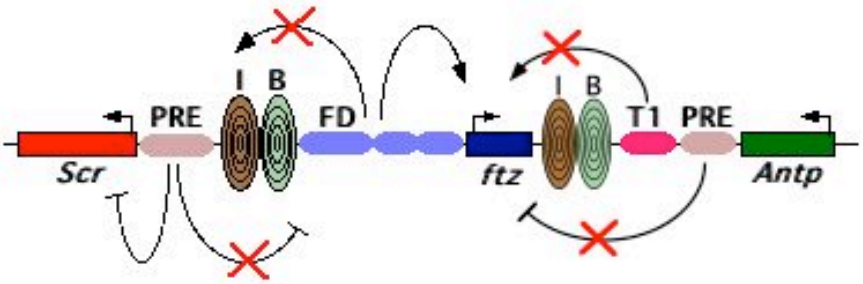
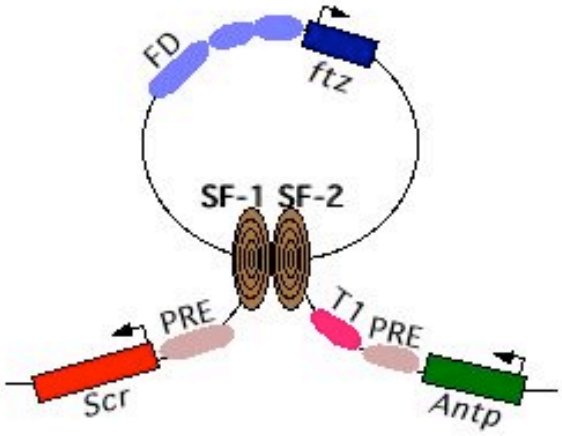

A
mini-white
B
Scr PRE
mini-white
C
Scr PRE
mini-white
SF-1/b
D
Scr PRE
mini-white
SF-1/c

Figure 15. The proposed biological function of the SF-1 insulator and barrier activity. (A) A model that illustrates the possible role of SF-1 in maintaining the transctiptional autonomy of the *ftz* and *Scr* genes. The embryonic enhancers of *ftz*, including the *ftz* distal enhancer (FD), are shown as blue ovals. The SF-1/b insulator (brown oval) prevents the activation of the *Scr* promoter by the *ftz* enhancers. The SF-1/c barrier (green oval) protects *ftz* from the silencing initiated by the *Scr* promoter-proximal PRE. Together these elements form a bipartite compound boundary. Similarly, a putative boundary (SF-2) downstream of *ftz* protects *ftz* from the influence of the *Scr* distal enhancer T1 and the *Antp*-proximal PRE. **(B)** As proposed by the "looping" model of the insulator function, SF-1 may interact with SF-2 and form a chromatin loop encompassing the entire *ftz* gene. Such interaction would ensure the transcriptional independence of the *ftz* domain and bring the *Scr* distal cis-regulatory elements closer to the *Scr* promoter, facilitating their interaction.

A


I
B
PRE
FD
I
B
T1
PRE
Scr
ftz
Antp

B


FD
ftz
SF-1
SF-2
PRE
T1
PRE
Scr
Antp

CHAPTER 4

IDENTIFICATION OF CANDIDATE SF-1 TRANS-FACTORS

Background

Chromatin boundaries and their protein components

A large number of sequences with enhancer-blocking properties have been identified in the genomes of higher eukaryotes [30]. Although the majority of them were found in *Drosophila*, there are several examples of insulators in vertebrates. For many of these sequences their precise molecular functions have been determined by extensive genetic and molecular genetic analysis. Some examples of the *in vivo* functions boundaries play in their native loci are shown in Table 1.

Relatively little information is available on the protein factors involved in the formation of boundaries. Only five proteins that bind to boundary sequences in higher eukaryotes have been identified so far. Four of them are found in *Drosophila* and include s*uppressor of hairy wing* [*su(Hw)*] - a DNA-binding component of the *gypsy* insulator [44]; Boundary Element Associated Factors (BEAF-32A and BEAF-32B) that interact with scs' [52]; *zeste-white 5* (*zw5*) that binds to scs sequences [53]; and the GAGA factor - a component of the *eve* promoter insulator [40], SF-1 [91], Fab-7 (P. Schedl. Personal communication), and MCP [124]. The fifth insulator-binding protein, CTCF, is found in vertebtates and appears to be a common component of all studied vertebrate enhancer-blockers [69]. Additionally, several putative boundary proteins have been identified in yeast [125]. However, due to the small genome size long-range enhancer-promoter interactions are absent in lower eukaryotes. Therefore, the yeast boundaries serve

primarily as barriers to delimit the borders between transcriptionally silent and active gene domains. Therefore, most of the yeast boundary proteins are transcription activators, such as Rap1 [126], Tbf1 [127], and Reb1 [128] that antagonize the spread of repressive signals, such as DNA condensation and histone modifications, from the silent domains. The yeast boundaries may not share a common mechanism with the elements from higher eukaryotes.

Interestingly, the five metazoan boundary proteins share little overall sequence homology but utilize a common type of DNA-interaction domain, a C2H2 zinc-finger. Different number of zinc-finger motifs are found in these proteins, ranging from twelve in *su(Hw)* to one in the GAGA factor (Table 1). There appear to be no other shared sequence motifs among these proteins. However, it is important to note that the *su(Hw)-mod(mdg4)* complex could be viewed as functionally similar to the GAGA factor in that it utilizes a zinc-finger motif for DNA binding and the BTB domain present in *mod(mdg4)* for interacting with other homologous and heterologous BTB domain proteins as proposed by the "looping" model [50]. This intriguing similarity may be an example of convergent evolution in the mechanism of *Drosophila* boundaries.

The first genes involved in the function of a boundary, *su(Hw)* and *mod(mdg4),* were identified using forward genetics [129, 130]. More recently, a variety of biochemical techniques have proven increasingly useful for isolating novel protein components of insulators. CTCF and BEAF-32 were isolated by a sequence of conventional chromatography steps and DNA-affinity chromatography using concatemerized binding site sequences [52, 131]. *Zw5* cDNA was isolated by screening a *Drosophila* embryonic cDNA expression library with a probe containing the footprinted sequence of the binding site [53].

Identification of DNA-binding proteins

The methods used to identify transcription factors and other DNA-binding proteins can be divided into two general groups, depending on whether the protein is isolated from a complex mixture or a corresponding cDNA is identified by screening. In the first class of methods the DNA-binding proteins are fractionated from crude nuclear extracts using a wide range of biochemical purification techniques, often including a sequence of conventional chromatography steps followed by DNA-affinity chromatography [132-134]. The progress of purification is monitored by a biochemical assay (gel shift, filter binding, Southwestern blotting) [135] or a functional assay (*in vitro* transcription, etc.) [136]. The identity of the purified protein is determined by N-terminal sequencing or mass spectrometry. The second class of methods utilizes cDNA libraries to express candidate DNA-binding proteins in individual cells. These library displays are subsequently screened for the interactions of the proteins with target DNA sequences *in vitro* (expression library screen) [137] or in living cells (yeast one-hybrid screen) [138, 139]. The candidate genes are identified by sequencing the positive cDNA clones followed by database searches.

Yeast one-hybrid technology

The yeast two-hybrid screening methodology has proven effective in the identification of novel binding partners for hundreds of proteins and defining the interaction domains of proteins known to form complexes *in vivo*. The yeast two-hybrid screen was successfully used to establish the genome-wide protein interaction networks [140]. A modification of this methodology for the isolation of new DNA-binding proteins is referred to as the yeast one-hybrid screen [141, 142]. In a two-hybrid experiment one of the interacting proteins has to be known.

Similarly, a prerequisite of the successful one-hybrid screen is a knowledge of the minimal sequence that specifically binds the protein of interest, as shown by an independent biochemical method. In a typical yeast one-hybrid experiment the minimal binding site (or several copies of the site in the form of a tandem array) is placed upstream of a reporter driven by a yeast promoter. The binding site sequence is often called the ”bait”. The bait-containing construct is introduced into *Saccharamyces cerevisiae* as either a plasmid or a genomically integrated copy. Next, a cDNA library fused to a strong transcription activation domain (often Gal4 AD) is transformed into the “bait”-containing strain. It is important that the cDNA library be derived from a tissue or developmental stage that is most likely to express the DNA-binding protein of interest. The synthesis of the protein with binding specificity for the bait sequence results in the recruitment of the activator domain to the bait transgene and the elevation of the reporter expression level. The two most frequently used reporters are yeast His3, allowing survival-based selection, and bacterial *lac*Z suitable for colorimetric detection. Simultaneous use of both selection methods has been reported to greatly reduce the rate of false-positives identified in the screen [143]. The yeast one-hybrid methodology has proven successful in identifying novel DNA-binding proteins from many organisms, including yeast [144], *C. elegans* [145], *Arabidopsis* [146], *Drosophila* [147], and mammals [148], [149].

Results

Putative SF-1 trans-factors isolated in the yeast one-hybrid screen

As described in Chapter II the 9 bp direct-repeat sequences are partially required for the enhancer-blocking activity of SF-1/b3. We sought to identify the protein(s) that specifically binds to this sequence and potentially participates in the formation of the insulator complex. We

generated a yeast strain carrying four copies of the 9 bp site upstream of the His3 reporter stably integrated into the genome. We reasoned that other, yet unknown, sites in the SF-1/b3 sequence may also be required for the insulator activity. Therefore, we made a second recombinant yeast strain carrying the entire SF-1/b3 element upstream of the His3 reporter. Both "bait" sequences caused a low-level expression of His3, which was successfully suppressed by the addition of 45 mM 3-AT to the growth medium. Since the SF-1 insulator appears to be active in all tissues of the embryo, and other known insulator trans-factors have heavy maternal input, we chose to use a *Drosophila* embryonic cDNA library for the screen. The library was transformed into the two "bait" yeast strains, so that each experiment allowed screening $3x10^6$ independent clones. The transformants were grown on selective medium. After 84 hours of incubation at 30°C, surviving colonies were re-streaked on fresh selective plates. A total of 175 colonies were collected for the 9 bp repeat "bait" and 142 colonies for the SF-1/b3 "bait". Only clones that yielded large, healthy colonies after growth on individual plates were further analyzed (39 for the 9 bp repeat "bait" and 31 for the SF-1/b3 "bait"). These clones were grown in liquid selective medium and used for plasmid DNA extraction. The isolated plasmids were transformed into *E. coli* to obtain larger amounts of uncontaminated DNA. Plasmid preparations were analyzed by PCR with pACT2-specific primers and subsequent digestion of the products with *Mse*I, *Msp*I, and *Rsa*I to eliminate duplicate clones. Clones that appeared unique were sequenced.

All clones yielded sequence information sufficient for unambiguous determination of the cDNA identity using the *Drosophila* genome database and the BLAST alignment tool. The genes encoding proteins with well known function and clearly irrelevant for transcription or nuclear function as well as duplicate clones were considered false-positives and discarded. The resulting list of potential true positives is shown in Figure 16 A and B. Interestingly, despite the fact that

the 9 bp repeat “bait” is much shorter than the SF-1/b3 bait, it yielded almost twice as many candidate interactors. This is most likely due to the fact that the SF-1/b3 bait does not contain high-affinity artificial arrays so that only proteins with highest affinity would bind to it. The advantages of using longer native baits were discussed previously [147].

To eliminate additional false-positives we performed database and literature searches to obtain comprehensive information about the remaining candidates (summarized in Figure 17). Candidates with at least one of the following characteristics were considered unlikely to be sequence-specific DNA-binding proteins and were discarded: (i) protein is well characterized and does not contain any DNA-binding domains; (ii) protein is extremely short (<50 aa) and contains no predicted DNA-binding motifs; (iii) protein is predicted by PSORT sequence analysis tool to be cytoplasmic and contains no predicted DNA-binding domains.

Since our transgenic analysis indicated that SF-1 is active during various stages of embryonic development as well as in adult flies we expected the SF-1 trans-factors to be expressed throughout embryogenesis. Thus, we discarded candidates for which a reliable microarray expression profile was available and indicated the absence of detectable levels of transcript during one of these developmental stages. A total of six candidates for both “baits” remained after applying these elimination criteria. These candidates are CG14648, CG1244, CG12822, and CG15812 for the 9 bp repeat “bait” and CG11329 and CG7022 for the SF-1/b3 “bait”. The available embryonic microarray data for five of these candidates are shown in Figure 18. Although the microarray for CG7022 was not available, this gene is known to be expressed during embryonic development.

Verifying the DNA-binding specificity of candidate proteins

To independently test the DNA-binding specificity of the remaining candidates we generated yeast strains carrying the 9 bp repeat and SF-1/b3 "baits" upstream of the *lac*Z reporter. The colorimetric *lac*Z assay is known to be more stringent that the His3 survival-based assay and should allow elimination of non-specific interactions. An additional yeast strain carrying a tandem array of p53 sites upstream of *lac*Z was also generated and used as a positive and a negative control in the assay. These three "bait" strains were transformed with cDNA clones of six candidate trans-factors, p53 cDNA, and Troponin C cDNA.

Since the 9 bp repeat sites are present in the SF-1/b3 sequence, the candidates isolated with the former "bait" would be expected to interact with the latter "bait" to some extent as well. However, none of the candidates exhibiting DNA-binding specificity should interact with the p53 "bait", and conversely p53 protein should not interact with either of the SF-1 "baits" if the assay is specific. The expression of *lacZ* in the transformed yeast cells was visualized by colorimetric reaction with X-gal. The results of the assay are shown in Figure 19. Indeed, p53 protein interacted exclusively with the p53 sites and drove high levels of *lacZ* expression evident from intense blue staining of the corresponding colony. The expression level of *lacZ* in the p53-transformed SF-1 "bait" strains is almost undetectable, indicating that the background expression of *lacZ* is significantly lower than that of His3. All of the 9 bp repeat candidates activated the respective "bait" and, as expected, interacted with the SF-1/b3 bait as well, but not the p53 "bait", confirming sequence specificity of the interactions. The CG11329 protein appeared to interact strongly with both SF-1 baits but interacted equally well with the p53 bait, suggesting that this protein has little sequence specificity. The SF-1/b3 "bait" candidate, CG7022, interacted strongly with the respective "bait" but did not interact with the 9 bp repeat "bait", suggesting that

this protein binds to sequences distinct from the 9 bp repeat. Finally, we tested Troponin C cDNA in the *lacZ* assay. Troponin C is a cytoplasmic protein involved in calcium homeostasis. It was isolated multiple times in the screens with both "baits", but, due to the absence of DNA binding activity, it was eliminated from the list of candidates. Interestingly, the presence of Troponin C in strains with all three baits resulted in overexpression of *lacZ*, a non-specific effect whose mechanism is unclear. In summary, the *lacZ* assay confirmed the specific DNA binding of five candidates and eliminated one candidate as non-specific.

Spatial distribution of candidate trans-factor mRNA in the embryo

Our transgenic studies showed that SF-1 insulator is active in various tissues of the embryo. It is capable of blocking the NEE, H1, PE, E3, *ftz* distal, and AE1 (D. Bosu, unpublished) enhancers. Therefore, we expected that the SF-1 binding protein(s) would be expressed ubiquitously. To confirm that the remaining five candidates conform to this requirement, we examined the distribution of the transcripts of the five genes by whole mount *in situ* hybridization with anti-sense RNA probes. The results of *in situ* analyses are shown in Figure 20. All five genes appear to be expressed ubiquitously throughout the embryo. Moreover, consistent with the microarray data, the mRNA of the candidates is both maternally deposited and zygotically expressed. Some of them, e.g. CG14648, have a very heavy maternal input and a weaker zygotic expression, similar to the pattern observed for *su(Hw)* (H. Cai, personal communication). As a negative control for background staining we performed a parallel *in situ* experiment using a *white* probe. Control embryos showed an undetectable level of staining, consistent with the fact that *white* is not expressed in the embryo.

Discussion

To investigate the molecular mechanism of SF-1 boundary function, we sought to identify protein factors that specifically bind to the insulator sequences. As described in Chapter II the 2.3 kb full-length SF-1 was dissected into elements of smaller size which were individually tested for enhancer-blocking activity *in vivo*. This analysis identified a 275 bp fragment, SF-1/b3, that retains sequence motifs sufficient for enhancer-blocking. The small size of SF-1/b3 made it suitable for use in biochemical experiments aimed at identifying putative trans-factors.

Several methods have been used successfully to isolate novel DNA-binding proteins [133]. Among these the yeast one-hybrid screen appeared best suited for our purpose. Some of the advantages of the yeast one-hybrid methodology are listed below.

1) The candidate DNA-binding factors are identified by DNA sequencing rather than N-terminal protein sequencing or mass spectrometry. 2) The cDNAs encoding the DNA-binding proteins are immediately available for cloning and further characterization. 3) Every gene expressed in a particular tissue and/or at a specific time in development is represented in a high-quality cDNA library. This eliminates the possibility of losing the protein of interest in the process of nuclear extract preparation or chromatographic enrichment. The latter steps are necessary for such approaches as affinity chromatography or Southwestern blotting. 4) The assay is conducted in a eukaryotic cell, making it more likely that the DNA-binding protein will be efficiently expressed and correctly folded and/or post-translationally modified. These advantages are not available when the cDNA library is expressed in *E. coli* for the expression screening. 5) The yeast one-hybrid assay does not absolutely require that the target DNA be an extremely short sequence precisely corresponding to the protein binding site. Instead, larger elements (up to 300-400 bp)

can be used in this assay [147]. This is particularly important for sequences with undefined trans-factor binding sites, such as SF-1/b3.

Over 300 clones were isolated in the primary His3 survival-based screen. A large fraction of them was efficiently eliminated by applying several rational selection criteria. For example, proteins that are cytoplasmic, have no DNA-binding capacity, or are absent from the early embryo were not considered further. The remaining six candidates were tested in a secondary yeast one-hybrid assay with the *lacZ* reporter. This stringent assay eliminated one of the candidates due to its lack of sequence specificity. Since SF-1 is active in all examined tissues in the early embryo, we reasoned that the protein factors required for its enhancer-blocking activity should be expressed ubiquitously. Indeed, every one of the five candidate genes shows ubiquitous distribution as judged by the whole mount RNA *in situ* hybridization.

It is interesting to note that some of the candidate SF-1 trans-factors share common protein domains with other known *Drosophila* insulator proteins (Figure 21 A and B). Namely, CG1244 contains six C2H2 zinc finger motifs, similar to *su(Hw)*, *zw5*, and the GAGA factor. CG15812 contains a BTB domain homologous to those in *mod(mdg4)* and the GAGA factor. The BTB domain in the latter two proteins is proposed to mediate the formation of chromatin loops, one of the possible mechanisms of insulator function [23]. CG14648 and CG12822 do not contain any DNA binding motifs that can be recognized by the available protein sequence analysis software. Given that these candidates consistently behave as sequence-specific DNA-binding proteins in two independent yeast one-hybrid assays, it is possible that they contain novel DNA-binding motifs.

Four of the five candidates identified in the screen have unknown nuclear function. CG7022, however, encodes a previously characterized protein *enhancer of bithorax*, *e(bx)*. This

protein is a crucial component of the Nucleosome Remodeling Factor NURF [150]. The other NURF components are NURF-38, NURF-55, and ISWI ATPase. NURF is believed to associate with promoter sequences at active loci and facilitate transcription by mobilizing nucleosomes. e*(bx)* is a large protein containing multiple domains shared by other factors involved in transcriptional regulation. Its only known DNA-binding motif is the AT-hook, also present in HMGI/Y group proteins [151]. Since the SF-1/b3 sequence is unusually AT-rich (ca. 60% AT) it is possible that *e(bx)* uses these bases for specific binding.

In summary, the yeast one-hybrid screen allowed us to identify five putative SF-1 interacting proteins, whose domain structure, nuclear localization, and expression profile are consistent with possible function in the formation of the insulator complex. The next chapter describes the direct test of the relevance of these candidates for enhancer-blocking.

Materials and Methods

Plasmid construction.

Two oligonucleotides were annealed (5'AATTCTTGAAACGCTTGAAACGCTTGAAACGCTTGAAACG 3' and 5' CTAGCGTTTCAAGCGTTTCAAGCGTTTCAAGCGTTTCAAG 3') to generate a four-copy 9 bp repeat "bait" flanked by *Eco*RI and *Xba*I sites. The SF-1/b3 fragment was prepared by digesting the respective TOPO clone (see Chapter II) with *Eco*RI and *Xba*I. Each "bait" sequence was inserted into pHISi vector (MATCHMAKER One-Hybrid System, BD Biosciences) digested with *Eco*RI and *Xba*I. These "baits" were also inserted into the p*LacZ*i vector using *Eco*RI and *Xho*I sites. These final constructs were confirmed by sequencing. The control bait containing p53 sites was obtained from BD Biosciences.

Yeast one-hybrid screen.

Five micrograms of "bait"-containing pHISi and p*LacZ*i vectors were linearized with *Xho*I and *Nco*I, respectively. The digested plasmids were purified by phenol extraction followed by ethanol precipitation. Three micrograms of purified DNAs were used to transform the YM4271 yeast strain using the LiAc-PEG transformation protocol [152]. Stably integrated transformants were selected on SD/-His or SD/-Ura plates for pHISi and p*LacZ*i strains, respectively. The background expression of *His3* driven by the "bait" sequences was neutralized in the presence of 45 mM 3-AT. The Gal4-AD fusion cDNA library from 0-20h *Drosophila* embryos (BD Biosciences) was amplified once to obtain sufficient amounts of DNA for the screen. The representation of amplified library was tested by PCR amplification of 6 different cDNAs expected to be present in the library. Five micrograms of the library DNA was introduced into the pHISi strains using the LiAc-PEG transformation protocol to yield approximately $3x10^6$ independent clones. The plates were incubated at 30°C for 84 hours. The surviving colonies were re-streaked on fresh SD/-His-Leu+45mM 3-AT plates and grown for 48 hours. Clones that yielded large colonies were grown in liquid SD/-Leu medium. Plasmid DNA was extracted using a previously described protocol [152] and transformed into *E.coli* strain TOP10 (Invitrogen). Bacterial plasmids were isolated using Plasmid Miniprep Kit (Qiagen) and sequenced with the pACT-specific primer. pACT2 plasmids containing candidate cDNAs were introduced into the respective pLacZi bait strains using the LiAc-PEG protocol. The transformants were selected on the SD/-Ura-Leu medium. The colorimetric lacZ assay was performed as described previously [152].

Whole mount RNA in situ hybridization.

Exon fragments of the five candidate genes were amplified using *Drosophila* genomic DNA and the following primers: 5'AGCACATCAAGGTGAACATCG3' and 5'GAACGACGCTCATCAGTGC3' for CG14648; 5'ATGACTGAAGTTGATGTCG3' and 5'CATCTTCAATCACGACGC3' for CG1244; 5'AGCAATTATACCAATGGTGG3' and 5'AATGTCCACGATGCACTGC3' for CG12822; 5'ATCATCTTAAGTGGATGGG3' and 5'CTTGCAATCGGACAATCG3' for CG15812; 5'CGAGTCGGAGTATCACTACG3' and 5'ACCAGCATGTAAAGTTCTCC3' for CG7022. The PCR products were cloned into pCRII-TOPO vector (Invitrogen), and orientation of the exons was determined by restriction analysis. The anti-sense RNA probes were synthesized using the DIG RNA Labeling Kit (Roche). Whole mount RNA *in situ* hybridization was performed as previously described [91].

Bioinformatics and database searches.

The identity of candidate genes was determined by aligning the cDNA sequence with the annotated *Drosophila* genome sequence (release 3) using the NCBI BLAST algorithm. Protein domains were identified using the Pfam program (http://www.sanger.ac.uk/Software/Pfam/). Protein intracellular localization was predicted using PSORT program (http://psort.nibb.ac.jp/). Embryonic microarray and *in situ* hybridization data were obtained from the Berkeley *Drosophila* Genome Project website (http://www.fruitfly.org/cgi-bin/ex/insitu.pl).

Acknowledgements

I wish to thank Claiborne Glover and Walter Schmidt for helpful advise on yeast genetics, Michael Terns for providing me with 3-AT, and Thao Le for technical assistance with

cDNA library amplification. I am grateful to Haini Cai for sharing her expertise in yeast handling and high-efficiency transformation. This work was supported by the NIH.

Table 1. Known *Drosophila* and vertebrate boundaries and their protein components

Insulator	Location	Proposed function	Trans-factor(s)	Trans-factor identification
		DROSOPHILA		
gypsy	the promoter region of the *gypsy* retrotransposon	protects the expression of transposon genes from chromosomal position effects	*Su(Hw)/ mod(mdg4)*	genetic screen for modifiers of gypsy-induced alleles
scs	5' boundary of the 87A7 hsp70 locus	delimit the domain of altered gene expression upon heat shock	*zw5*	screening of the embryonic cDNA expression library
scs'	3' boundary of the 87A7 hsp70 locus		BEAF-32A BEAF-32B	DNA affinity chromatography
Fab-7	flank the borders of the *iab* enhancers at the *Abd-B* gene locus in the Bithorax complex	maintains functional independence of the *iab* enhancers at the *Abd-B* locus	GAGA factor	transgenic analysis of the mutant insulator sequence
Fab-8			dCTCF	reverse genetics
Mcp			-	
SF-1	the *Scr-ftz* intergenic region	maintains regulatory independence of the *ftz* domain	GAGA factor *E(bx)*	Transgenic assay; yeast one-hybrid screen
eve	-30 to -1 region of the *even-skipped* gene	facilitates the activation by distant enhancers	GAGA factor	transgenic assay
		VERTEBRATES		
HS4	the 5' and 3' borders of the chicken β-globin gene cluster	prevents the β-globin gene from repressive effects of the neighboring heterochromatin	CTCF	Chromatographic purification followed by gel mobility shift assays
3'HS		prevents inadequate communication between β-globin gene promoter and the enhancers of the olfactory receptor gene	CTCF	gel mobility shift assay, SouthWestern blotting
Lys 5' A	5' boundary of the lysozyme locus in macrophages, has MAR/SAR activity	ensures high-level of lysozyme gene activity in activated macrophages	CTCF	identified in enhancer-blocking assays

Figure 16. Potential SF-1 trans-factors isolated in the yeast one-hybrid screen. (A) A "bait" construct (shown on top) carrying four copies of the 9 bp direct-repeat site from SF-1/b3 upstream of the His3 reporter was integrated into the yeast genome. This strain was transformed with the *Drosophila* embryonic fusion cDNA library. A total of $3x10^6$ independent clones were screened. Ten candidate SF-1 interacting proteins isolated with this "bait" are listed in the table in green. **(B)** A "bait" construct (shown on top) carrying the entire SF-1/b3 element upstream of the His3 reporter was integrated into the yeast genome. The "bait" strain was transformed with the *Drosophila* embryonic fusion cDNA library. Screening a total of $3x10^6$ independent clones allowed isolation of six candidate trans-factors. They are listed in the table in brown.

A

"9 bp repeat" bait
HDAC1
CG6416
aristaless
Df31
CG14648
CG1244
Abd-A
CG14646
CG12822
piefke
Total: 10

B

SF-1/b3 bait
CG5961
CG6038
HDAC3
CG1153
CG11329
E(bx)
Total: 6

Figure 17. Characteristics of the candidate trans-factor genes. (A, B) The information on the putative trans-factor genes used in the elimination of unlikely candidates based on rational criteria (see text). The genes isolated using the 9 bp repeat "bait" are shown in green **(A)**; the genes isolated with the SF-1/b3 "bait" are shown in brown **(B)**. The information on the function of the genes was gained from searching FlyBase (http://flybase.bio.indiana.edu/). The domain structure of the candidate proteins was predicted using Pfam algorithm (http://www.sanger.ac.uk/Software/Pfam/). The microarray data on the embryonic expression and the in situ expression profiles were found at the Berkeley *Drosophila* Genome Project website (http://www.fruitfly.org/DGC/index.html). The subcellular localization of the candidate proteins was predicted using the PSORT program (http://psort.nibb.ac.jp/).

A

N	CG# or name	Function	Protein Domains	Embryo expression / *in situ*	Subcellular localization
1	CG7471/ HDAC1	Histone deacetylase	No DNA binding motifs	+/Not available	70.6% nuc
2	CG6416	unknown	PDZ domain (signaling molecules, receptors)	Absent/ Not available	94.1% nuc
3	CG3935/ *aristaless*	Transcription factor	homeobox, OAR (protein-protein interactions)	+/ tissue-specific expression	94.1% nuc
4	CG2207	Decondensation factor 31	No DNA binding motifs	-	21.7% nuc
5	CG14648	unknown	possible enzyme domain	+/Not available	55.5% nuc
6	CG1244	unknown	6 C2H2 Zn fingers	+/Not available	94.1% nuc
7	CG10325/ *Abd-A*	Transcription factor	Homeodomain	+/ tissue-specific expression	94.1% nuc
8	CG14646	unknown	RING finger (protein-protein interactions)	+/Not available	76.7% cyt
9	CG12822	unknown	-	+/Not available	55.5% nuc
10	CG15812/ *piefke*	unknown	BTB, PSQ	+/ubiquitous	89% nuc

B

N	CG# or name	Function	Domains	Embryo expression / *in situ*	Subcellular localization
1	CG5961	unknown	F box (ubiquitination)	+/ tissue-specific expression	70.6% cyt
2	CG6038	unknown	-	+/Not available	76.7% cyt
3	HDAC3	Histone deacetylase	No DNA binding motifs	-	94.1% cyt
4	CG1153	unknown	-	absent till after 10h	70.6% cyt
5	CG11329	unknown	PHD, RING finger	+/Not available	55.5% nuc
6	CG7022/ *E(bx)*	Chromatin remodeling machinery	AT-hook, WAC, WAKZ, PHD, bromodomain	+/heavy ubiquitous	94.1%

Figure 18. The most likely trans-factor candidates are expressed throughout embryogenesis. After the elimination of unlikely candidates based on a set of rational criteria, a total of six candidate genes from the screens with both "baits" were remaining. All of these genes are expressed during early and late embryogenesis as judged by the respective microarray profiles. The microarray dada was obtained from the Berkeley *Drosophila* Genome Project website (http://www.fruitfly.org/DGC/index.html). The presence of the transcripts is shown by green bars; the absence of the transcripts is shown by red bars. Each bar corresponds to a one-hour interval during first twelve hours of development. The microarray profile for CG7022 was not available. However, the presence of this transcript in the embryo was evident from the literature.

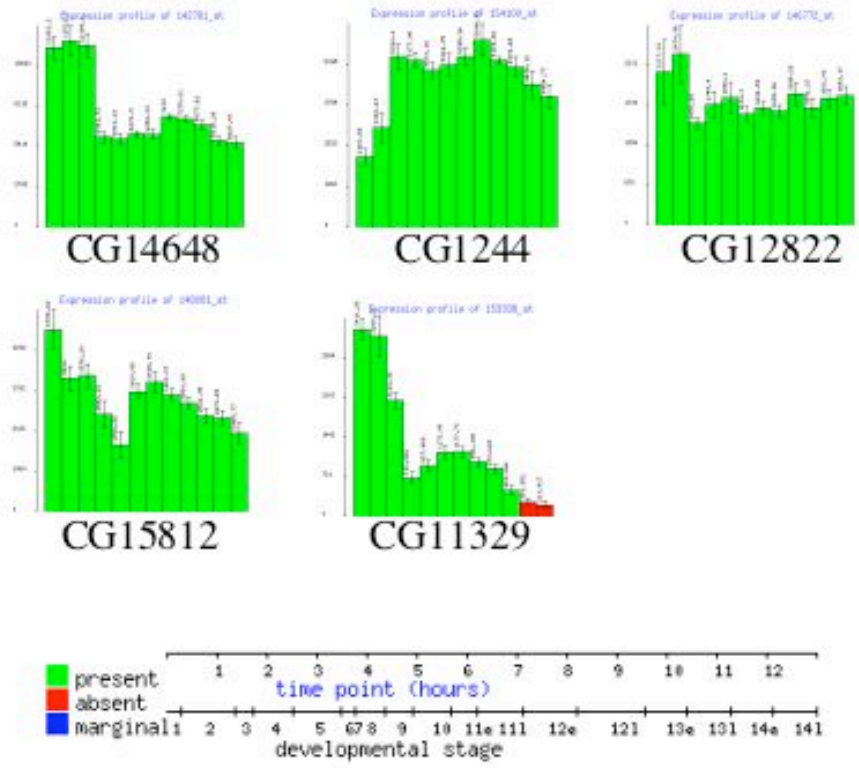

CG14648
CG1244
CG12822
CG15812
CG11329
present
absent
marginal
1 2 3 4 5 6 7 8 9 10 11 12
time point (hours)
1 2 3 4 5 6 7 8 9 10 11e 11l 12e 12l 13e 13l 14e 14l
developmental stage

Figure 19. Testing the sequence specificity of candidate genes in an independent lacZ assay. Three "bait" sequences (four copies of the 9 bp direct repeat, SF-1/b3, and five copies of the p53 sites) were placed upstream of the *lacZ* reporter. These "bait" constructs were integrated into the yeast genome. Each of the "bait" strains was transformed with the Gal4 AD-cDNA fusion plasmids containing the "prey" genes listed on the left. The transformants were selected on the SD/-Ura-Leu medium and assayed for the level of *lacZ* expression. p53 protein (top row) interacts only with the respective "bait", indicating low background level of *lacZ* expression in the assay. CG14648, CG1244, CG12822, CG15812, and CG7022 demonstrate sequence specific binding to the SF-1 derived "baits". CG11329 appears to bind non-specifically to all three "bait" sequences. Similarly, the troponin C gene (bottom row) causes non-specific over-expression of *lacZ* in this assay.

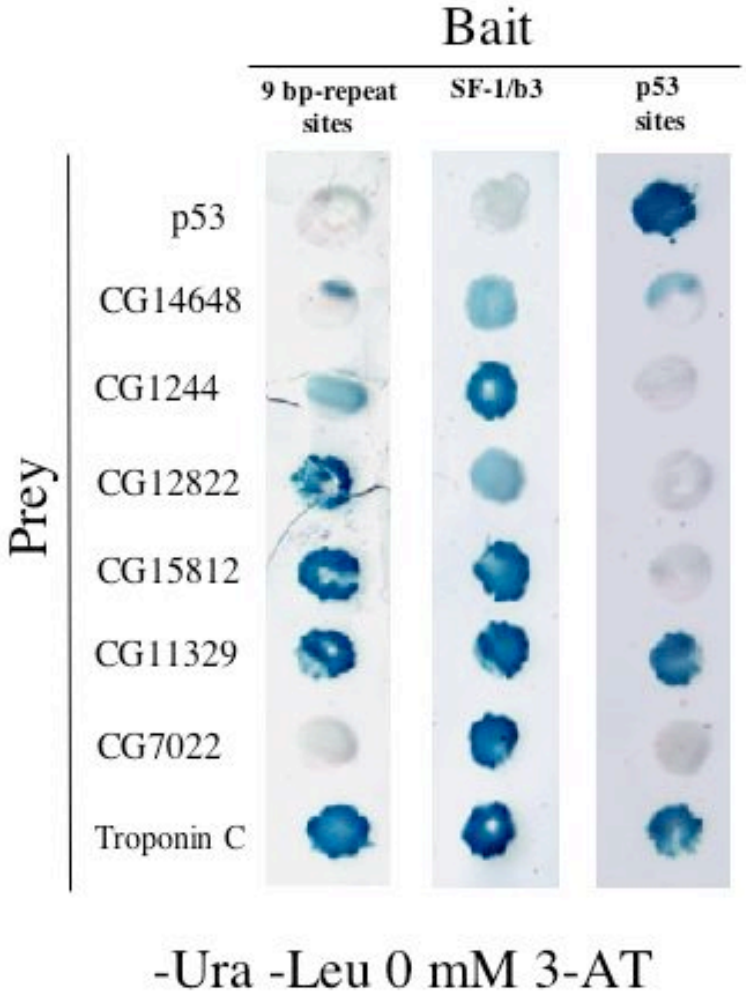

Bait
9 bp-repeat sites
SF-1/b3
p53 sites
Prey
p53
CG14648
CG1244
CG12822
CG15812
CG11329
CG7022
Troponin C
-Ura -Leu 0 mM 3-AT

Figure 20. The candidate trans-factors are ubiquitously expressed in the embryo. The 0-6 hour old w^{1118} *Drosophila* embryos were collected and analyzed by whole mount *in situ* hybridization as described in the Materials and Methods. The appr. 1 kb anti-sense RNA probes corresponding to the four candidate genes (listed on the left side) were prepared as described in the Materials and Methods. For each candidate gene the embryos corresponding to three developmental stages are shown: syncytial blastoderm (0-1h); cellular blastoderm (2-4h); and gastrulation (4-6 h). For CG7022 the embryonic in situ data were available at the Berkeley *Drosophila* Genome Project website (http://www.fruitfly.org/DGC/index.html). The control embryo (lower right) from a parallel *in situ* with the *white* probe shows no detectable background signal.

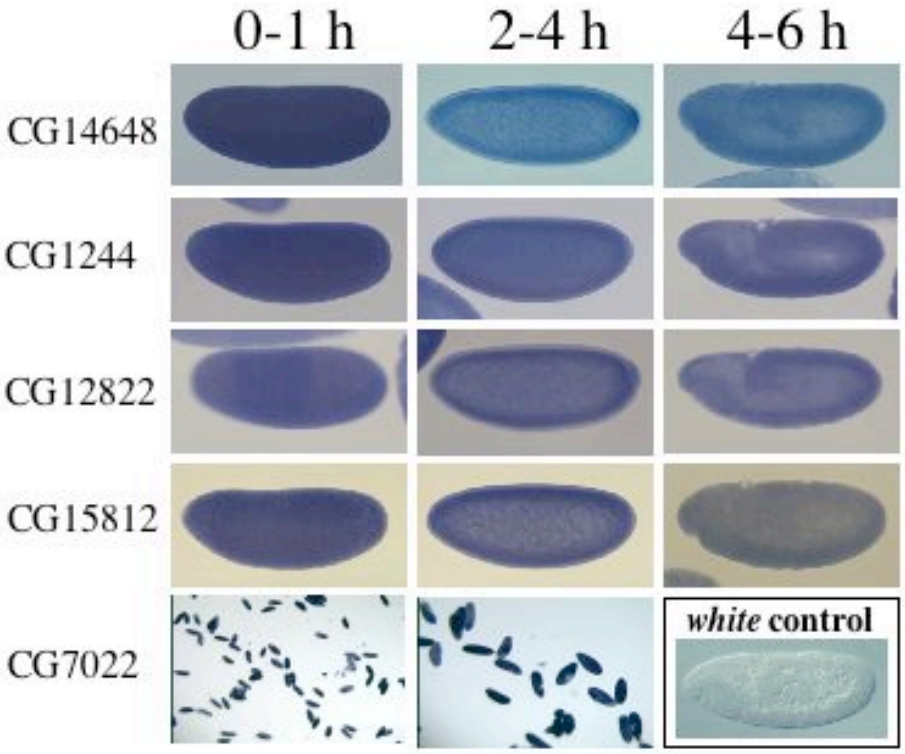
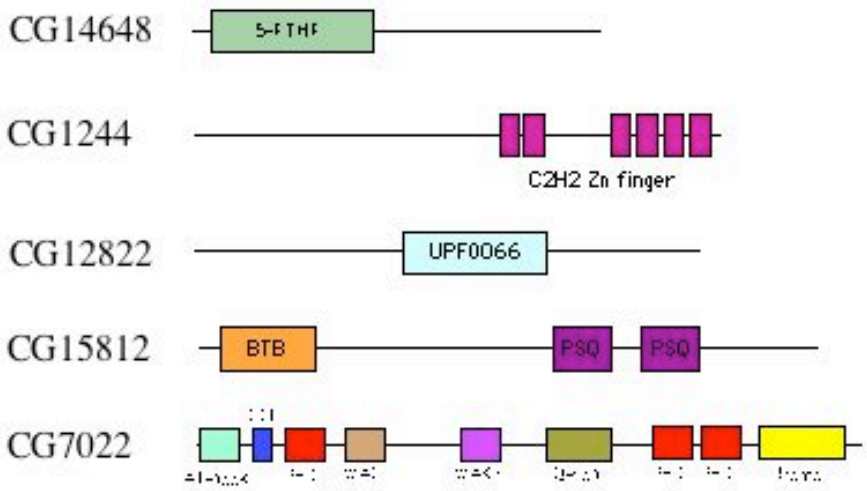

0-1 h
2-4 h
4-6 h
CG14648
CG1244
CG12822
CG15812
CG7022
white control

Figure 21. The final list of candidate SF-1 interacting proteins. (A) A compiled list of the putative SF-1 trans-factors after the unlikely candidates were eliminated using a set of stringent assays. The candidates isolated with the 9 bp direct repeat "bait" are shown in green; the only remaining candidate isolated with the SF-1/b3 "bait" is shown in brown. For all five genes the protein domain structure, subcellular localization, mRNA distribution, and protein length (GenBank, http://www.ncbi.nlm.nih.gov/entrez) are consistent with or suggestive of the putative trans-factor function. **(B)** The domain structure of the five genes isolated in the yeast one-hybrid screen. Three of the candidates, CG1244, CG15812, and CG7022 have previously characterized DNA-binding domains.

A

CG# / gene name	Function	Domains	mRNA present / spatial distribution	Subcellular localization	Protein length, aa
CG14648	unknown	possible enzyme domain	+/ ubiquitous	55.5% nuc	545
CG1244	unknown	6 C2H2 Zn fingers	+/ ubiquitous	94.1% nuc	1152
CG12822	unknown	none predicted	+/ ubiquitous	55.5% nuc	412
CG15812 (*pfk*)	unknown	BTB, PSQ	+/ubiquitous	89% nuc	518
CG7022 *E(bx)*	chromatin remodeling	AT-hook, WAC, WAKZ, PHD, bromodomain	+/heavy ubiquitous	94.1% nuc	2669

B



CHAPTER 5

CELL CULTURE-BASED INSULATOR ASSAY

Background

RNAi in Drosophila cell culture

Forward genetic analysis remains a powerful tool for studying gene function in *Drosophila*. However, performing genetic studies is not always possible because of the lack of mutations for particular genes, lethality or sterility associated with some mutations, and the complexity of phenotypes observed with certain hypomorphic alleles. The availability of the complete annotated *Drosophila* genome sequence led to the rapid development of reverse genetic techniques for gene function analyses. Among these the RNA interference (RNAi) – based methods proved most useful. Originally developed in *C. elegans*, RNAi was shown highly effective in *Drosophila* embryos [153, 154]. Early experiments used micro-injection of the *in vitro* synthesized double-stranded RNA (dsRNA) into pre-cellular blastoderm embryos. Later developments showed that RNAi can be achieved by the expression of gene-specific dsRNA from a transgenic vector [155]. Moreover, RNAi can be induced in a wide variety of spacial and temporal patterns using the UAS-Gal4 binary system [156].

Despite the elegance and utility of RNAi methodology in live flies, this technology is not amenable for genome-wide reverse genetic screens. The *Drosophila* cell culture system (particularly Schneider S2 cell line [157]) has become a valuable tool in analyzing biological function [158]. S2 cells appear to be derivatives of embryonic hematopoietic cells or their

developmental precursors [159] and are suitable for studies of many biological processes, such as signal transduction [160], apoptosis [161], mRNA processing [162], and transcription regulation [163]. The genes of interest can be efficiently "knocked out" in S2 cells by simple incubation in a medium containing specific dsRNA. Extensive studies of RNAi in this system showed that short (approximately 1 kb) dsRNAs corresponding to a portion of the gene of interest are sufficient to induce RNAi. dsRNAs introduced into S2 cells are cleaved in the cytoplasm into short 21-23 siRNAs by a ribonuclease III, *Dicer*. The *Drosophila* genome contains two *Dicer* genes, *Dcr-1* and *Dcr-2*, whereas human and *C. elegans* genomes each have one copy of the gene [164]. Recent studies showed that the two *Drosophila Dicer* genes have distinct activities: *Dcr-1* is primarily involved in micro RNA (miRNA)-mediated gene silencing, whereas *Dcr-2* is a crucial component of the RNAi pathway [165]. siRNA molecules associate with the RNA-Induced Silencing Complex (RISC) and load it onto target transcripts via a homology-dependent process followed by the degradation of the target by RISC. Interestingly, the role of *Dicer* may extend beyond the initial cleavage of dsRNA, since siRNAs fail to induce RNAi when co-transfected with *Dicer*-specific siRNA in mammalian cells [166]. In *C. elegans*, *Drosophila,* and mouse RNAi appears to have an amplification mechanism, either via catalysis, synthesis, or possibly both. The experiments in these systems demonstrate that the effect of RNAi persists for many rounds of cell division, remaining active over a 50-100 fold increase in cell mass. These features of RNAi combined with the ease of dsRNA introduction into *Drosophila* S2 cells make this system extremely useful for studies of gene function. It becomes particularly instrumental in the study of the genes crucial for oogenesis and cell viability when the RNAi induction in the fly results in lethality. Moreover, the RNAi response can be easily optimized in cell culture by varying the amount of introduced dsRNA, whereas this factor can not be easily changed in

transgenic flies. These features make RNAi in cells particularly attractive for reverse-genetics and genome-wide functional screens [167]. Several examples of such studies have been recently published [168, 169].

Drosophila cell culture model for the analysis of insulator function

The advantages described above make the cell culture system ideally suited for the study of chromatin insulators and their protein components. Namely, the functional relevance of putative SF-1 trans-factors identified biochemically can be conveniently tested by RNAi. Moreover, the additional protein components of the insulator complex can be identified in a genome-wide RNAi screen. These applications require a robust and sensitive enhancer-blocking assay in cultured *Drosophila* cells.

The only published enhancer-blocking assay in *Drosophila* cell culture utilized the *hsp*27 promoter fused to the chloramphenicol acetyl-transferase (CAT) reporter gene [52]. The *hsp*27 promoter contains an Ecdysone Response Element (ERE) between positions –544 and –579, and three heat shock response elements between positions –270 and –370 [170]. The test insulator sequences were placed at position –129, separating the enhancers from the promoter. The assays were performed in transiently or stably transformed *Drosophila* Kc cells following heat-shock or ecdysone treatments. None of the tested scs' insulator elements showed any appreciable enhancer-blocking activity in the transient assay with the ecdysone induction. In mixed stable cell lines the enhancer-blocking activity was observed with the ecdysone-induction protocol, but was up to ten-fold less pronounced with the heat-shock treatment, perhaps, because of the weaker activating potential of the heat-shock elements. It is also important to note that this study used stable cell lines containing large tandem arrays of integrated plasmids, necessitating the use

of an additional insulator downstream of CAT to prevent cross-activation by ERE from the neighboring copy of the construct. These drawbacks, combined with the inconvenience of the CAT activity assay, inability to observe reporter expression levels in individual cells, and high cost of ecdysone prompted us to develop a new and improved enhancer-blocking assay in *Drosophila* cells.

Results

Enhancer-blocking assay in Drosophila S2 cells

Since the previously reported enhancer-blocking assay in *Drosophila* cultured cells suffered from several drawbacks, we sought to develop a new and improved assay. We chose fluorescent reporters (GFP and RFP) to monitor the gene expression activity, since these reporters allow visualization in live cells and have a fairly large linear range of detection. We chose to use the *Drosophila* metallothionein (Mtn) enhancer, which has a potent (over 100-fold) transcriptional response upon induction of cells with heavy metal ions [171] and has been used extensively for high-level production of recombinant proteins in *Drosophila* cell cultures [172, 173]. The induction of expression driven by the metallothionein promoter is due to the binding of the transcription factor MTF-1 to multiple binding sites in the promoter that are collectively termed the Metal Response Element, MRE [174-176]. Since we lacked the information on the potency of the Mtn "core" promoter, we chose to use the MT enhancer in combination with a well characterized "core" promoter. We tested two TATA-containing promoters, *eve* and SV40 minimal. When the MT enhancer was placed immediately upstream of the *eve*-GFP fusion the transiently transfected cells displayed extremely bright green fluorescence after a 24 hour induction with 1 mM Cu^{2+} (Figure 22 A). GFP expression was easily detectable even in

uninduced cells. By contrast, the GFP expression was undetectable in cells transfected with the MT-SV40 construct before or after induction. Next, we tested the inducible expression of a different fluorescent reporter, RFP, in the same assay. The recently developed fast-maturing variant of RFP [177, 178] was potently induced by 1 mM Cu^{2+} (Figure 22 B).

SF-1, SF-1/b, and su(Hw) insulators are active in transiently transfected S2 cells

To prove the suitability of the MT enhancer – *eve* promoter combination for the enhancer-blocking assay in cultured cells, we first tested the distance sensitivity of the MT enhancer. Placing the 1.8 kb fragment from the *white* coding region between the MT enhancer and the *eve* promoter did not significantly weaken the activity of the enhancer judged by fluorescence microscopy (data not shown). In order to measure the enhancer activity in a more quantitative way, the transfected cells were analyzed by flow cytometry. Comparison of the flow cytometry data derived from the cells carrying constructs with and without the spacer confirmed that the MT enhancer functions over distance (Figure 23 A and B). To test whether the MT enhancer can be blocked by a previously characterized insulator, we placed the *Su(Hw)* insulator, SF-1, and SF-1/b into the assay construct in place of the spacer. All three insulators efficiently blocked the MT enhancer as judged by a severe reduction of the fluorescence intensity in transfected cells after induction (Figure 23 C, D, E).

Since these tests were performed using different DNA constructs, we wanted to account for possible variations in transfection and induction efficiency from well to well. Thus, we co-transfected the same test constructs with the identical amounts of the MT-*eve*RFP plasmid to serve as an internal control. In cells with the MT enhancer immediately next to the *eve*GFP reporter (Figure 24 A) or separated by a 1.8 kb spacer (Figure 24 B) a significant increase in both

green and red fluorescence intensity was observed upon induction with 1 mM Cu^{2+}. By contrast, in cells carrying the insulator-containing constructs the induction caused the increase in red fluorescence only (Figure 24 C, D, E). In these samples the proportion of cells exhibiting bright green fluorescence did not increase significantly compared with the uninduced control.

Double-reporter construct

As was described above, co-transfection of the GFP insulator test construct with the no-insulator RFP control allowed us to eliminate variations in transfection efficiency and induction conditions between individual experiments. However, the ratio between the number of GFP and RFP plasmids varies somewhat from cell to cell. Moreover, if the enhancer-blocking assays were to be performed in cells with genomically integrated copies of the construct, the RFP control reporter located on a separate plasmid would not reflect the influences on the genomic environment of the GFP reporter. To circumvent these problems we designed a double-reporter construct in which divergently transcribed *eve*GFP and *eve*RFP genes share one MT enhancer positioned in between (Figure 25 A). Since both reporters are driven by identical promoters their transcripts could be expected to accumulate at comparable levels upon induction by the MT enhancer. Indeed, when the double-reporter construct was transiently transfected into S2 cells, we observed a consistent correlation between the levels of GFP and RFP in most cells, both induced and uninduced (Figure 25 B). In order to obtain an average ratio of the green to red fluorescence for a large population we analyzed the transfected cells by flow cytometry. Twenty thousand cells were counted in each experiment. The resulting data were subjected to software compensation to eliminate the contribution of green fluorescence to the red channel and vice versa. The results of this analysis suggest that there is a linear correlation between the level of

GFP and RFP in the majority of cells (Figure 26 B). The insertion of the 1.8 kb spacer from the *white* coding region slightly shifts the distribution of cells into the region of green fluorescence (Figure 26 C). The reasons for this shift are unclear. The insertion of *Su(Hw)*, SF-1 or SF-1/b between the MT enhancer and the *eve*GFP reporter results in each case in significant reduction of GFP expression, leaving RFP expression unaffected (Figure 26 D, E, F). To quantify the enhancer-blocking activity in these constructs we divided the flow chart for the parental double-reporter construct in two halves by a diagonal (Figure 26 B). Next we approximated the average RFP to GFP ratio in the population of transfected cells by calculating the ratio between the number of cells in the top and bottom half of the chart. This ratio was normalized to be 100% for the parental double reporter. Using the same diagonal gating method the RFP/GFP ratios were calculated for the spacer- and insulator-containing constructs (Figure 26 G). The 350 bp *Su(Hw)* insulator has the strongest enhancer-blocking activity, followed by SF-1 and SF-1/b.

RNAi abolishes the insulator activity of su(Hw)

We have shown the ability of the *Su(Hw)* insulator to block the MT enhancer in transiently transfected S2 cells. Both genes required for the activity of this boundary in flies, *su(Hw)* and *mod(mdg4)*, are also expressed in S2 cells (P. Majumder, unpublished, and data herein). Therefore, it seems likely that these two proteins are involved in the formation of the insulator complex in both systems. We tested this supposition through the use of RNA interference. S2 cells were transiently transfected with the *Su(Hw)* insulator-containing double reporter construct shown in Figure 27 A. Then, cells were treated with the increasing amounts of double-stranded RNAs corresponding to the exon regions of *su(Hw)* and *mod(mdg4)* genes. Cells were also treated with the GAGA factor dsRNA as a negative control (Figure 27 B). Most

protocols for RNAi in *Drosophila* cells recommend 72 hour incubation as an optimal period of time that permits degradation of the protein synthesized prior to dsRNA treatment while keeping the level of the target transcript low. After the 72-hour incubation the cells were induced by the addition of 1 mM Cu^{2+}. Twenty-four hours after induction the cells were analyzed by fluorescent microscopy and flow cytometry (Figure 27 C and D). Since the duration of the RNAi experiment is long enough to allow cells to divide multiple times, the percentage of transiently transfected cells is substantially lower than in our previous experiments. To accommodate this difference we increased the number of counted cells to 100,000. Comparison of the compensated flow charts for the induced cells treated with the *su(Hw)* dsRNA and the control cells revealed a significant increase in the amount of green fluorescence (lower right quadrant) in the dsRNA-treated sample (Figure 27 D). The recovery of GFP expression depended on the dosage of dsRNA and was highest (over 3 fold increase) in cells treated with 15 μg of dsRNA (Figure 27 E). The recovery of GFP fluorescence was less pronounced in the *mod(mdg4)* dsRNA-treated cells (less than 2-fold increase). Surprisingly, the increase in fluorescence did not correlate with the amount of the *mod(mdg4)* dsRNA added, probably because of the detrimental effect of the depletion of this protein on global gene transcription. The fluorescent images of the dsRNA-treated cells demonstrate that the level of GFP expression increases in the *su(Hw)* and *mod(mdg4)* RNAi samples while the level of RFP control remains the same (Figure 27 C). The addition of the GAGA factor dsRNA to the cells did not result in any significant change in GFP fluorescence (Figure 27 E). This is an expected result since the GAGA factor is not known to be involved in the activity of the *Su(Hw)* insulator.

To demonstrate that the observed changes in GFP expression were a direct consequence of the altered transcript level of the respective gene we performed a semi-quantitative RT-PCR

analysis of the cells subjected to RNAi. The primer pairs were designed in such a way that only the sequences of mature spliced transcripts would be amplified and give products of the expected size. All reactions also included a pair of primers designed to detect the actin 88F transcript that served as an internal control. The results of RT-PCR analysis are shown in Figure 27 F. Both *su(Hw)*, *mod(mdg4)*, and GAGA factor transcripts appear to be knocked-down by RNAi in a dsRNA dosage-sensitive fashion. The amount of actin 88F transcript is approximately the same across all samples indicating that the observed reduction of transcript levels is not due to unequal input of total RNA. Notably, even in the cells treated with the highest amount of dsRNA residual transcripts were observed, necessitating further optimization of the RNAi protocol.

e(bx) affects the activity of SF-1/b

By performing the "proof-of-concept" experiments described above we demonstrated that the sensitivity of our cell culture assay permits the detection of the insulator activity loss as a result of the RNAi-mediated depletion of an insulator protein component. These results suggest that this assay can be used for the fast and efficient validation of candidate insulator components identified by indirect biochemical approaches. We sought to test if the candidate genes identified in the yeast one-hybrid screen are involved in the function of SF-1 using the RNAi assay. To this end, dsRNAs corresponding to the exon regions of the five candidate genes were generated (Figure 28 B) and introduced into the cells transiently transfected with the SF-1/b-containing double-reporter construct (Figure 28 A). As was shown in Chapter II, the enhancer-blocking activity of SF-1 depends on the GAGA factor. Thus, we used the GAGA dsRNA treatment of cells as a positive control for this RNAi experiment. The *su(Hw)* and *mod(mdg4)* dsRNAs were used as negative controls, since SF-1 insulator activity does not require these proteins. The

dsRNA treatment and Cu^{2+} induction were performed as described above. Twenty-four hours after induction the cells were examined by fluorescence microscopy (Figure 28 C) and flow cytometry. The degree of recovery of GFP fluorescence was determined by the analysis of flow cytometry data as described in the Materials and Methods. As expected the GAGA factor RNAi cells exhibited higher green fluorescence than the mock RNAi cells or cells treated with the *su(Hw)* or *mod(mdg4)* dsRNAs (Figure 28 D). Among the genes identified in the yeast one-hybrid screen only *e(bx)* had a significant effect on the GFP fluorescence. In fact, in the *e(bx)* dsRNA-treated cells the degree of GFP expression recovery was even higher than that in the GAGA RNAi cells. The other four candidate genes appeared to have no effect on the insulator activity (Figure 28 D).

We verified the effect of RNAi on the level of respective transcripts by semi-quantitative RT-PCR. Total RNA was extracted from the dsRNA-treated and control cells, reverse transcribed, and analyzed by PCR with gene-specific primers. The actin 88F primers were added to each reaction as an internal control. Transcripts corresponding to *su(Hw)*, *mod(mdg4)*, GAGA factor, CG1244, CG12822, and *e(bx)* were "knocked-down" in a dsRNA dosage-sensitive fashion (Figure 28 E and F). Transcripts of CG14648 and CG15812 were undetectable by our analysis even in untreated control cells, indicating that these transcripts are either extremely rare or that these genes are not expressed in S2 cells.

Discussion

The only enhancer-blocking assay in cultured *Drosophila* cells published to date [52] is sub-optimal for several reasons, e.g., because it lacks sensitivity and does not allow monitoring the insulator activity in live cells. Here we describe the development and validation of a new

cell-culture insulator assay. In this system two divergently transcribed fluorescent reporters, GFP and RFP, each fused to the *eve* minimal promoter, are activated by a single copper-inducible metallothionin (MT) enhancer. Since the promoters driving GFP and RFP are the same, the MT enhancer appears to be shared, resulting in comparable expression levels of the two reporters. Insertion of an insulator between the MT enhancer and the *eve*GFP reporter results in an attenuated GFP signal upon induction. On the other hand, the RFP signal remains unchanged, serving as an *in situ* control for general silencing and induction. The inability of the MT enhancer to communicate with the eve promoter over the insulator sequence is not simply due to the distance sensitivity of the enhancer. Placing a neutral DNA spacer between the MT enhancer and eveGFP does not compromise their interaction. Three insulators (Su(Hw), SF-1, and SF-1/b) were tested and shown to function in this assay.

Further, we demonstrated that the cell culture enhancer-blocking system is suitable for the study of insulator protein components. Reducing the mRNA levels of *su(Hw)* and *mod(mdg4)* genes in S2 cells resulted in significant reduction of the *Su(Hw)* insulator activity. This effect is specific since the reduction of the GAGA factor mRNA level did not influence the activity of the *Su(Hw)* insulator. In the reciprocal test the activity of SF-1/b was attenuated by the GAGA RNAi but not the *su(Hw)* or *mod(mdg4)* RNAi.

These results allowed us to test the genes identified in the yeast one-hybrid screen as putative components of the SF-1 insulator. Remarkably, the *e(bx)* RNAi led to a partial loss of the activity of SF-1/b, suggesting the involvement of this gene in the insulator function. Other candidates seemed to have no effect on SF-1/b. In fact, two of them, CG14648 and CG15812, appeared not to be expressed in S2 cells.

What is the possible role of *e(bx)* in the formation of the insulator complex? It is conceivable that this protein binds to the SF-1 sequence in the form of a quaternary complex NURF. NURF is a large complex, containing four subunits: ISWI ATPase; NURF-55, a WD-40 repeat protein; NURF-38, an inorganic pyrophosphatase; and NURF-301, recently identified as a product of the *e(bx)* gene [179]. Two other ATP-dependent ISWI chromatin remodeling complexes have been identified in *Drosophila*: ACF and CHRAC. These two complexes are involved in nucleosome sliding that results in the formation of regular ordered arrays [180, 181]. In contrast, NURF disrupts nucleosomal arrangement and activates transcription *in vitro* [150, 182]. It is possible that NURF-mediated nucleosome sliding facilitates the recruitment of other factors or generates a transcriptionally favorable state in the SF-1 sequence. The latter possibility would argue in favor of the "promoter decoy" model of insulator function.

Several lines of evidence are consistent with our finding that *e(bx)* is involved in the SF-1 insulator function. First, *e(bx)* contains two AT hook domains, also present in HMGI/Y proteins, that are known to bind to the minor groove of AT-rich sequences [151]. The sequence of SF-1/b3 is over 60% AT and possibly contains sequence motifs particularly favorable for the AT-hook binding. Second, *e(bx)* has been found to associate with the GAGA factor in the embryonic nuclear extract. The two proteins interact specifically *in vitro*, and their interaction is mediated by the N-terminal region of *e(bx)* and the zinc finger region of the GAGA factor [183]. It is conceivable that the two proteins bind to their respective sequence motifs in SF-1 and interact with each other forming the insulator complex. Finally, *e(bx)* mutations have been shown to affect the expression of the homeotic genes, *Ubx* and *engrailed*, consistent with a potential function of SF-1 in regulating homeotic genes in the Antennapedia complex.

In summary, we developed a novel cell culture-based insulator assay that can serve as a convenient means of testing the function of the putative insulator proteins isolated in biochemical or genetic screens. Several advantages of conducting RNAi in cultured cells versus living animals are immediately apparent. (i) The RNAi conditions can be easily manipulated to achieve the optimal response. (ii) The assay can be conducted with the genes that are essential for oogenesis or development. For example, the requirement of *e(bx)* for oogenesis in *Drosophila* [179] would preclude "knocking-out" this gene by RNAi with a maternal driver. (iii) The ease of specific RNAi induction in *Drosophila* S2 cells and the availability of the large collection of dsRNAs (*Drosophila* RNAi Screening Center at the Harvard Medical School) makes it possible to conduct genome-wide reverse genetic screens for additional protein components of insulators.

Materials and Methods

Generation of insulator test constructs.

All test constructs described were made in the pCasPeR vector to allow the establishment of stable lines with single-copy insertions upon co-transfection with the source of transposase. EGFP and RFP open reading frames were amplified by PCR using *Pfu* Turbo polymerase (Stratagene) using pEGFP-N3 (Clontech) and pRed H-Stinger (provided by S. Barolo) as templates. The EGFP or RFP reporters were inserted into *Bam*HI and *Pst*I sites replacing *lac*Z in the previously described pCAeb-*lac*Z construct [184]. To make the double-reporter construct (pCA2xReporter) the *eve*-RFP fusion was assembled in pCRII and then placed into the *Nsi*I and *EcoR*V sites of the pCA*eve*-GFP vector, replacing the mini-*white* gene. The MT enhancer was PCR amplified from *Drosophila* genomic DNA using the primers

5'GTTGCAGGACAGGATGTGG3' and 5'AACGCGGCTTTACACACGG3' and cloned into pCR-TOPO vector (Invitrogen). The MT enhancer was placed into pCA*eve*-GFP upstream of the promoter or between divergently transcribed reporters in pCA2xReporter. In each case, a *Not*I site was engineered between the MT enhancer and the *eve*-GFP. The 1.8 kb *Sma*I fragment from the mini-*white* gene, SF-1, SF-1/b, and *Su(Hw)* were isolated from respective clones in pCRII vector as *Not*I fragments and placed into the *Not*I site of the single- and double-reporter constructs. Clones with the SF-1 and SF-1/b insulators in the "forward" orientation relative to the *eve*-GFP were identified by PCR.

Cell culture, transfections, and flow cytometry.

In all experiments *Drosophila Schneider* cells (S2) were maintained in HyQ SFX-Insect serum-free medium (HyClone) at 28°C. For transfection, plasmid DNAs were prepared using the Plasmid Mini Kit (Qiagen) and dissolved at 1 μg/μl in sterile TE buffer. Transfections were performed using Cellfectin reagent (Life Technologies). Briefly, $9x10^5$ cells were aliquoted into each well of a 6-well plate. After cells had attached to plastic, the medium was removed and replaced with 1 ml of transfestion cocktail (2-3 μg of test plasmid and 5 μl Cellfectin in 1 ml SFX medium). In co-transfection experiments equal amounts of the GFP and RFP plasmids were mixed with Cellfectin. After 5 hours of incubation the transfection mix was replaced with 2 ml of medium. Cells were induced by the addition of 1 mM $CuSO_4$ 24 hours after transfection. The induction was allowed to proceed for another 24 hours. The induction of cells was monitored by fluorescence microscopy (Zeiss Axioplan 2). For flow cytometry cells were washed off the plastic, spun down at 100xg for 5 minutes and resuspended in sterile PBS. The analysis was performed using the FACSCalibur flow cytometer (Becton Dickinson Immunocytometry

Systems) at the UGA CTEGD Flow Cytometry Facility. Twenty thousand cells were counted in a typical experiment. Fluorescence was excited with the argon laser at 488 nm. Green fluorescence was detected with FL1 530/30 BP filter; red fluorescence was detected with FL2 585/42 BP filter. Photomultiplier voltages were set at 430V for FL1 and 420V for FL2. No hardware compensation was used. The optimal software compensation was determined to be FL1-52% and FL2-40%. FloJo and Summit software packages were used for data analysis.

RNAi experiments. Exon fragments from the five candidate genes identified in the yeast one-hybrid screen, each approximately 800 bp long, were generated by PCR with the *Drosophila* genomic DNA as template and cloned into pCRII-TOPO vector. The following primer pairs were used: 5'AGCACATCAAGGTGAACATCG3' and 5'GAACGACGCTCATCAGTGC3' for CG14648; 5'ATGACTGAAGTTGATGTCG3' and 5'CATCTTCAATCACGACGC3' for CG1244; 5'AGCAATTATACCAATGGTGG3' and 5'AATGTCCACGATGCACTGC3' for CG12822; 5'ATCATCTTAAGTGGATGGG3' and 5'CTTGCAATCGGACAATCG3' for CG15812; 5'CGAGTCGGAGTATCACTACG3' and 5'ACCAGCATGTAAAGTTCTCC3' for CG32346. An 800 bp fragment of *su(Hw)* cDNA was generated by PCR with the 5'AGGAAAAGAAGGGCAAGCTGC3' and 5'AGCATATGTCCTTCTTCTCC3' primer pair and TOPO-cloned. The pCRII clones of the full-length *mod(mdg4)* and GAGA cDNAs were provided by Haini Cai. Each of the nine TOPO clones was linearized by *Not*I and *Hind*III to prepare templates for run-off *in vitro* transcription reaction. SP6 and T7 MEGAscript transcription kits (Ambion) were used to generate sense and anti-sense RNA strands. The resulting RNAs were quantified by gel electrophoresis and diluted 1:10 in nuclease-free water (Ambion). Equal amounts of the sense and anti-sense strands were mixed in PCR tubes, denatured by heating at 92°C for 10 minutes and annealed by slow cooling (1°C/min) to room

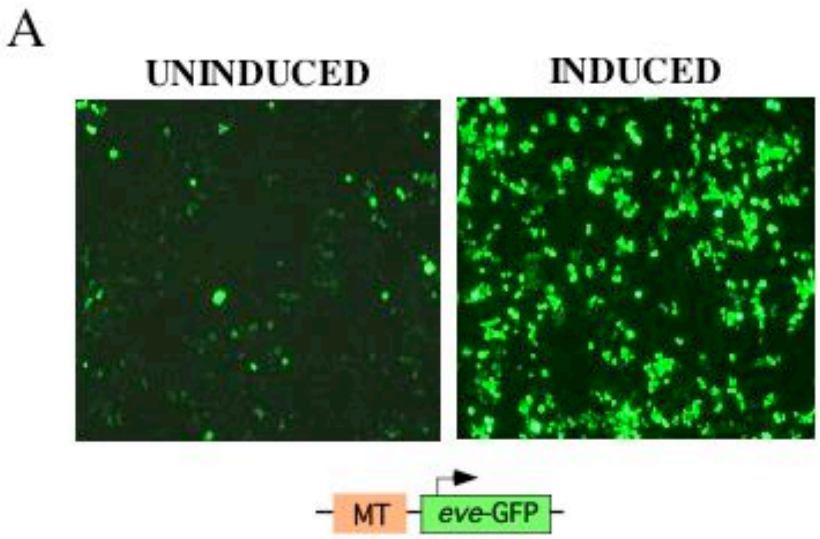
temperature. The concentration and quality of double-stranded RNAs were verified by gel electrophoresis. For the RNAi experiment a large-scale transfection of the pCA2xReporter plasmids containing *Su(Hw)* and SF-1/b insulators was performed. Twenty-four hours after transfection cells were washed off the plates, counted, and aliquoted into 12-well plates at $3x10^5$ cells/well. Five wells were set up for each RNAi test: (i) no RNAi/Cu^{2+}(-) control, (ii) no RNAi/Cu^{2+}(+) control, (iii) 5μg dsRNA/Cu^{2+}(-), (iv) 5μg dsRNA/Cu^{2+}(+), and (v) 15μg dsRNA/Cu^{2+}(+). Cells were allowed to attach to wells and then transfected with the specified amount of dsRNA using the Cellfectin reagent protocol (Invitrogen). Cells were incubated for 72 hours to allow for the turnover of the target protein. 1 mM Cu^{2+} was added into appropriate wells for induction. 24 hours after induction cells were washed off the wells, spun at 100xg for 5 minutes, resuspended in PBS, and analyzed by flow cytometry using BD FACSCalibur. One hundred thousand cells were analyzed for each well. The flow cytometry data were processed with the use of Summit. The remaining cells were spun down, and the total RNA was extracted using the TRIZOL reagent (Invitrogen) per manufacturer's protocol. RNA pellets were re-suspended in 20 μl of nuclease-free water, treated with RNase-free DNase (Ambion), and analyzed by gel electrophoresis to determine the yield of total RNA. The isolated RNA was used as a template for RT-PCR analysis using the OneStep RT-PCR kit (Qiagen). The following primer pairs were used to detect transcripts: 5'GTCAGCTGAGACTGCAACG3' and 5'TTCGTTGAACTCCAAACATTGCG3' for CG14648; 5'CAGTCTGACGATGATGACG3' and 5'ATCTTGCGCTTCTGCAGACG3' for CG1244; 5'ATCGTGGACATTCTGGAGGC3' and 5'TAAGAATGTCCTGTATTCTAGAGC3' for CG12822; 5'ACGCGAATCCAGCAGTTCG3' and 5'CGTATTGTCGCGCTTTCGG3' for CG15812; 5'ATAAGCGACGCCATCTGTCC3' and 5'TTACGTTCTTCCTCATCATCC3' for CG32346;

5’TTATTAGCACCGCGGAATCG3’ and 5’AGAGCGTTTGTTTGTCGACAC3’ for *mod(mdg4)*; 5’GGAAAACACAGCCCGAAACA3’ and 5’CCTCATCCGTCAGCTGCTCT3’ for *su(Hw)*; 5’GTATCGGCAATCCAATTGTTG3’ and 5’CTGATGATGGATGTCCTCCA3’ for the GAGA factor; 5’GATGGTGTCTCCCACACCGT3’ and 5’CGATCGGCAATACCAGGGT3’ for actin 88F. All RT-PCR reactions were performed in a semi-quantitative multiplex format with the primers for actin 88F transcript serving as a control. The optimal ratio between the concentration of the actin 88F primers and the candidate gene primers was empirically determined to be 1:10. The conditions of PCR reactions were: 55°C annealing, 1 min extension, 20 cycles. The RT-PCR products were analyzed by gel electrophoresis.

Acknowledgements

I wish to thank Dr. Michael Adang and the members of his lab for providing me with S2 cell stock and pIZT plasmid, as well as for sharing their expertise on work with *Drosophila* cell culture. I thank Dr. Kelley Moremen for providing pEGFP-N3 plasmid as a source of EGFP ORF, Dr. Scott Barolo for pRed H-Stinger plasmid, a source of fast-maturing RFP ORF. I thank the members of the Fechheimer lab for being helpful and cooperative in sharing the cell culture facility and Julie Nelson and Diana Martin for introducing me to flow cytometry instruments and analysis software.

Figure 22. The inducible expression of GFP and RFP reporters by the metallothionein enhancer. (A) The test construct (shown below the cell images) contained the metallothionein enhancer (MT, pink box) driving the expression of the eve promoter fused to the EGFP open reading frame. *Drosophila Schneider S2* cells were transiently transfected with this construct and photographed 24 hours post-transfection (uninduced cells). The Cu^{2+} was added to a final concentration of 1 mM into the culture medium to induce the activity of the MT enhancer. The cells were photographed 24 hours after induction. A strong inducible expression of GFP is observed. **(B)** An analogous construct was made with the fast-maturing RFP. A consistent induction profile is observed with this reporter. However, the fluorescence signal is somewhat weaker that with GFP.


A
UNINDUCED
INDUCED
MT
eve-GFP

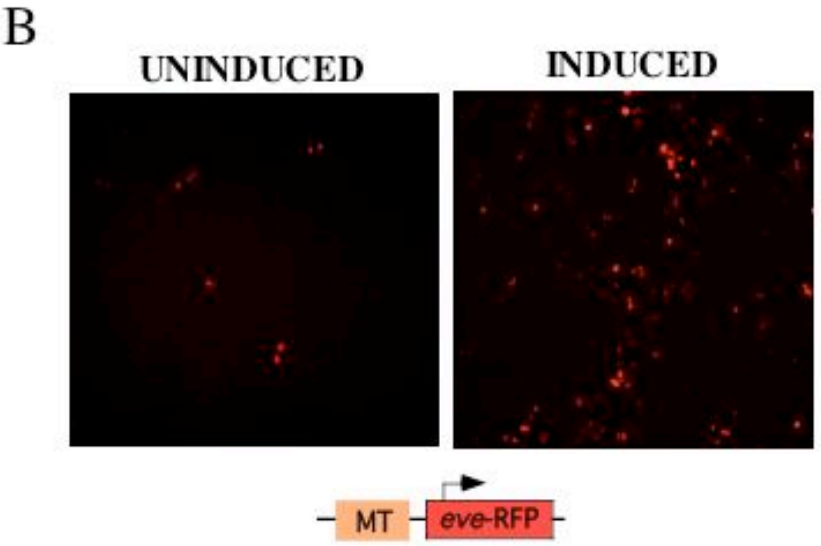

B
UNINDUCED
INDUCED
MT
eve-RFP

Figure 23. The activity of the MT enhancer can be blocked by the insulators. (A-E) The parental MT-*eve*GFP construct (Figure 22 A) and derivative constructs with the spaces and insulator sequences inserted between the enhancer and the promoter are shown on the left. These constructs were transiently transfected into the *Drosophila Schneider* S2 cells. The uninduced and induced cell samples were analyzed by flow cytometry as described in the Materials and Methods. Twenty thousand cells were counted in each experiment. The cells exhibiting GFP fluorescence above the background are highlighted by green shadow on the histograms. The log of fluorescence intensity is shown on x and the number of cells on y. The induction of cells with 1 mM Cu^{2+} results in over-expression of GFP in cells transfected with the parental construct **(A)**. The insertion of a 1.8 kb spacer from the *white* coding region does not compromise the ability of MT to activate GFP transcription **(B)**. However, the enhancer is completely blocked by *su(Hw)* insulator **(C)**, SF-1 **(D)**, and SF-1/b **(E)**.

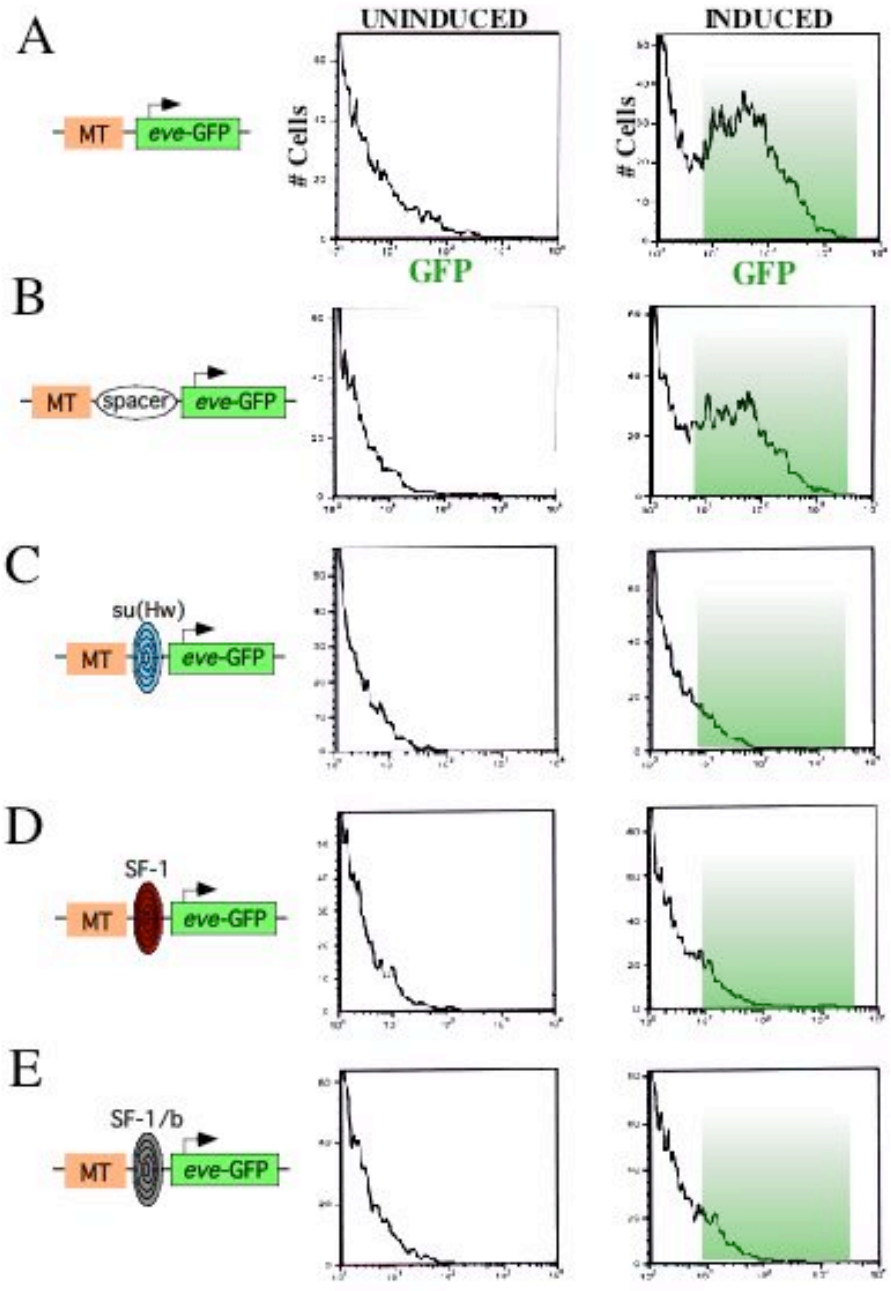

A
UNINDUCED
INDUCED
MT
eve-GFP
Cells
GFP
B
MT
spacer
eve-GFP
C
su(Hw)
MT
eve-GFP
D
SF-1
MT
eve-GFP
E
SF-1/b
MT
eve-GFP

Figure 24. Cell culture insulator assay with a second fluorescent reporter as an internal control for induction. (A-E) The *Drosophila Schneider* S2 cells were transiently co-transfected with equal amounts of the two plasmids: the GFP-containing insulator test construct and the RFP-containing internal control plasmid. The plasmid combinations used are shown on the left side of the figure. The uninduced and induced cell samples were analyzed by flow cytometry as described in the Materials and Methods. Twenty thousand cells were counted in each experiment. The log of green fluorescence intensity is shown on x and the log of the red fluorescence intensity is shown on y. The induction with 1 mM Cu^{2+} results in the concomitant increase in both green and red fluorescence in cells transfected with parental plasmid **(A)** or the spacer-containing derivative **(B)**. In the case of insulator-containing constructs the red fluorescence increases upon induction, whereas the green fluorescence remains unchanged **(C, D, E)**.

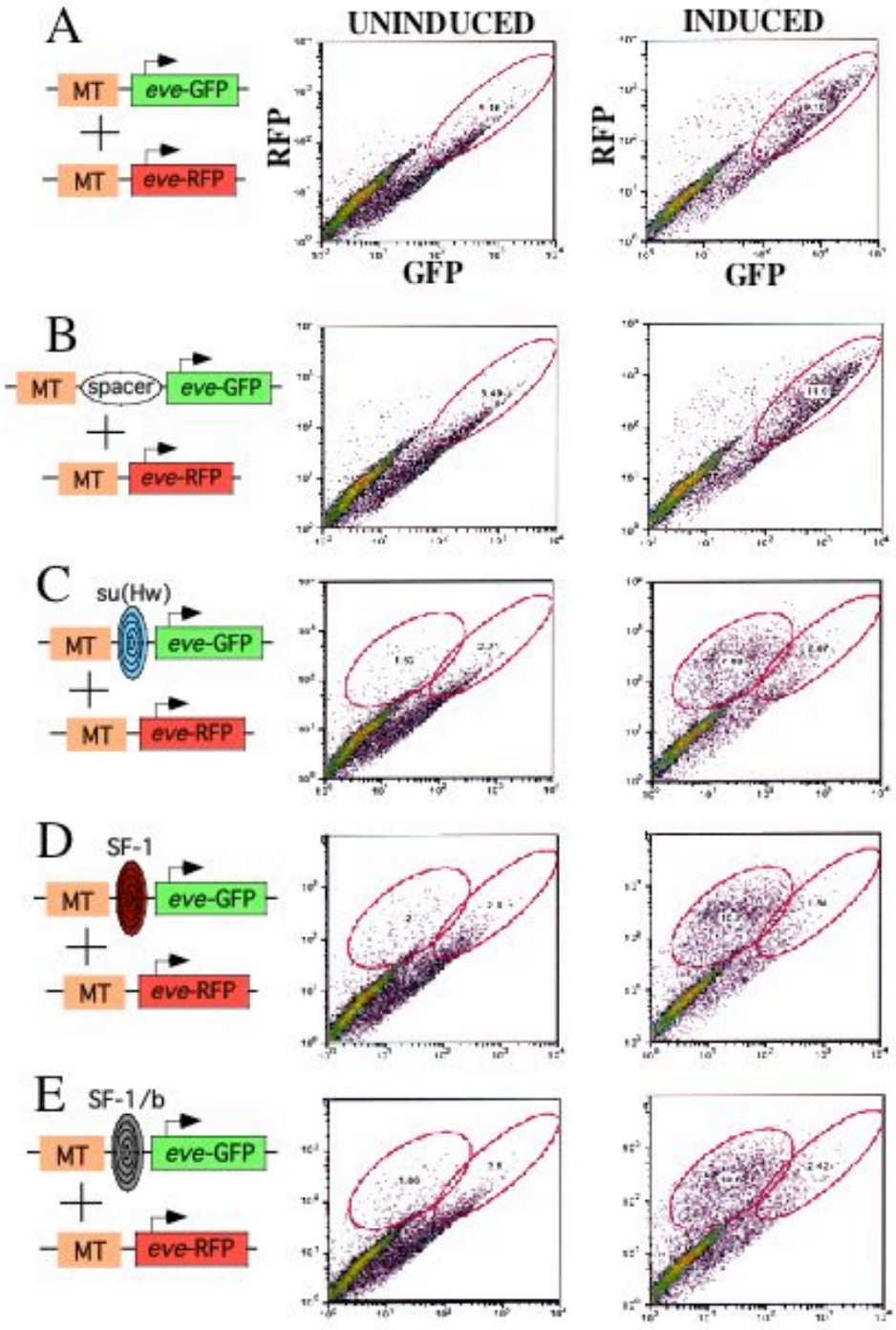

A
UNINDUCED
INDUCED
MT
eve-GFP
MT
eve-RFP
RFP
GFP
B
MT
spacer
eve-GFP
MT
eve-RFP
C
su(Hw)
MT
eve-GFP
MT
eve-RFP
D
SF-1
MT
eve-GFP
MT
eve-RFP
E
SF-1/b
MT
eve-GFP
MT
eve-RFP

Figure 25. Insulator assay with the double-reporter construct. (A) A schematic of the double-reporter construct. The divergently transcribed GFP and RFP reporters are fused to the identical eve promoters. The MT enhancer (pink box) is placed between the reporter genes. **(B)** The *Drosophila Schneider* S2 cells were transiently transfected with the double-reporter construct. The fluorescence photographs were taken before and after induction of cells with 1 mM Cu^{2+}. The Zeiss Axioplan 2 microscope with GFP and PI filter sets was used for imaging cells. Both reporters are efficiently induced by the shared MT enhancer. There is a direct correlation between the levels of GFP and RFP expression in most cells.

A

B

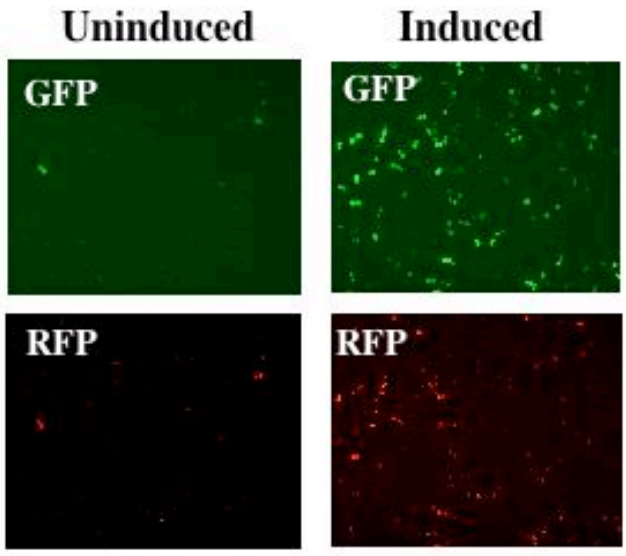


Figure 26. The cell culture enhancer-blocking assay using the double-reporter construct. **(A)** A schematic of the double-reporter construct with the position of insulator insertion is shown. **(B-F)** The *Drosophila Schneider* S2 cells were transiently transfected with the double-reporter constructs containing DNA elements indicated on top of charts. The cells were induced with 1 mM Cu^{2+} and analyzed by flow cytometry as described in the Materials and Methods. Twenty thousand cells were counted in each experiment. A direct correlation between the levels of GFP and RFP expression in most cells is observed for the parental double-reporter construct **(B)**. A similar induction profile is also seen in the spacer-containing control construct **(C)**. The prevalent induction of the red but not green fluorescence is observed in the cells transfected with the constructs containing *Su(Hw)* insulator **(D)**, SF-1 **(E)**, and SF-1/b **(F)**. **(G)** Quantitative assessment of the enhancer-blocking activity in the tested constructs is shown on the bar graph. Refer to the text for the method of calculating RFP/GFP ratio.

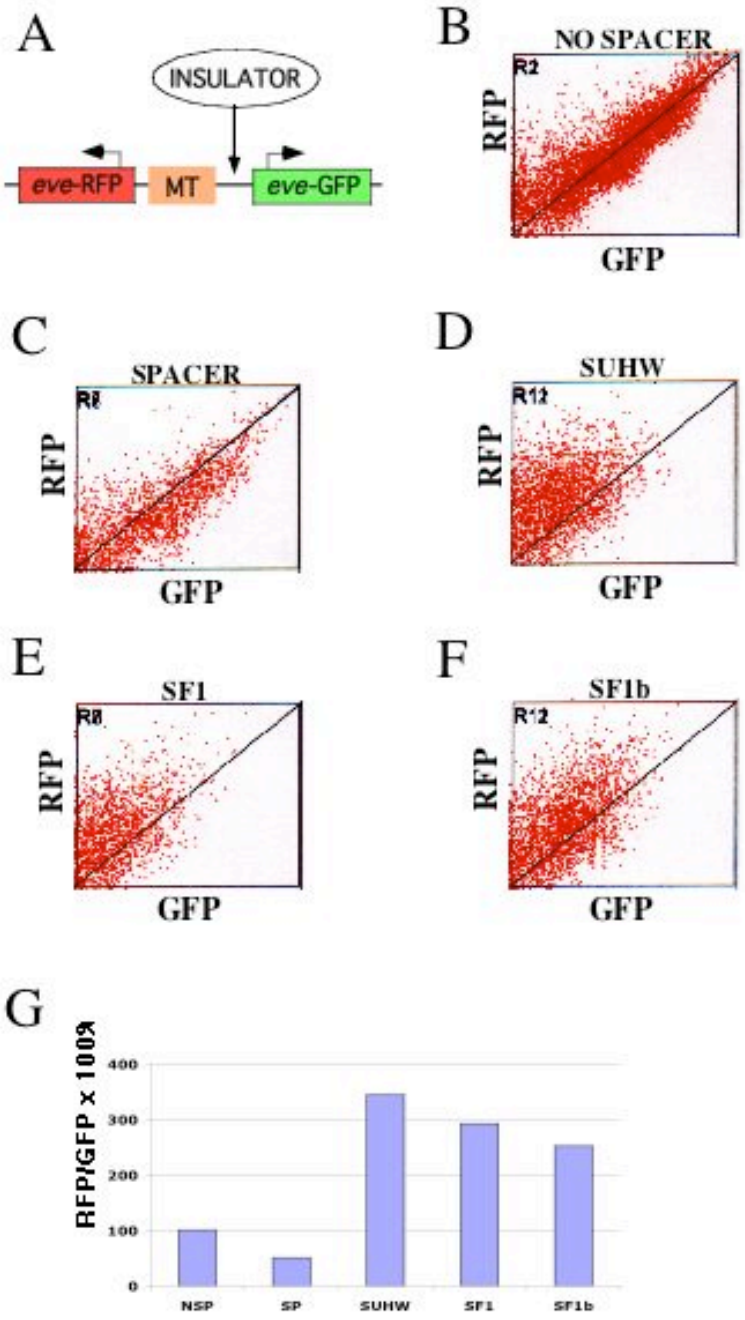

A
INSULATOR
eve-RFP
MT
eve-GFP
B
NO SPACER
R2
RFP
GFP
C
SPACER
RFP
GFP
D
SUHW
RFP
GFP
E
SF1
RFP
GFP
F
SF1b
RFP
GFP
G
RFP/GFP x 100%
400
300
200
100
0
NSP
SP
SUHW
SF1
SF1b

Figure 27. RNAi-mediated knock-out of the insulator function. (A) A schematic of the *Su(Hw)* insulator-containing double-reporter construct used for the experiment. **(B)** The *in vitro* synthesized double-stranded RNAs corresponding to *su(Hw)* gene (lane 1), *mod(mdg4)* (lane 2), and the GAGA factor (lane 3). M – DNA size marker. **(C)** The *Drosophila Schneider* S2 cells were transiently transfected with the *Su(Hw)*-containing double-reporter construct (shown in A), treated with dsRNAs indicated on the left side for 78 hours, induced with 1 mM Cu^{2+} and photographed using Zeiss Axioplan 2 microscope with GFP and PI filter sets. A slight increase in GFP fluorescence is observed in the *su(Hw)* and *mod(mdg4)* dsRNA-treated cells compared to the untreated control. **(D)** An example of the flow cytometry analysis of the *su(Hw)* dsRNA-treated cells and the untreated control is shown. One hundred thousand cells were counted. The log of the green fluorescence intensity is shown on x and the log of the red fluorescence intensity is shown on y. A significant increase in the green fluorescence intensity is observed in the *su(Hw)* RNAi cells. **(E)** The degree of recovery of GFP expression in cells treated with different dsRNAs. Refer to Materials and Methods for the details on quantitative assessment of GFP fluorescence recovery. **(F)** The semi-quantitative multiplex RT-PCR analysis of the dsRNA-treated cells. The genes targeted by RNAi are shown on top. The first lane in each triplet represents an untreated control; the second lane – cells treated with 5 μg respective dsRNA; third lane – 15 μg dsRNA. The top band in each lane represents the actin 88F internal control. For all three genes the dsRNA dosage-sensitive reduction of the transcript level is observed.

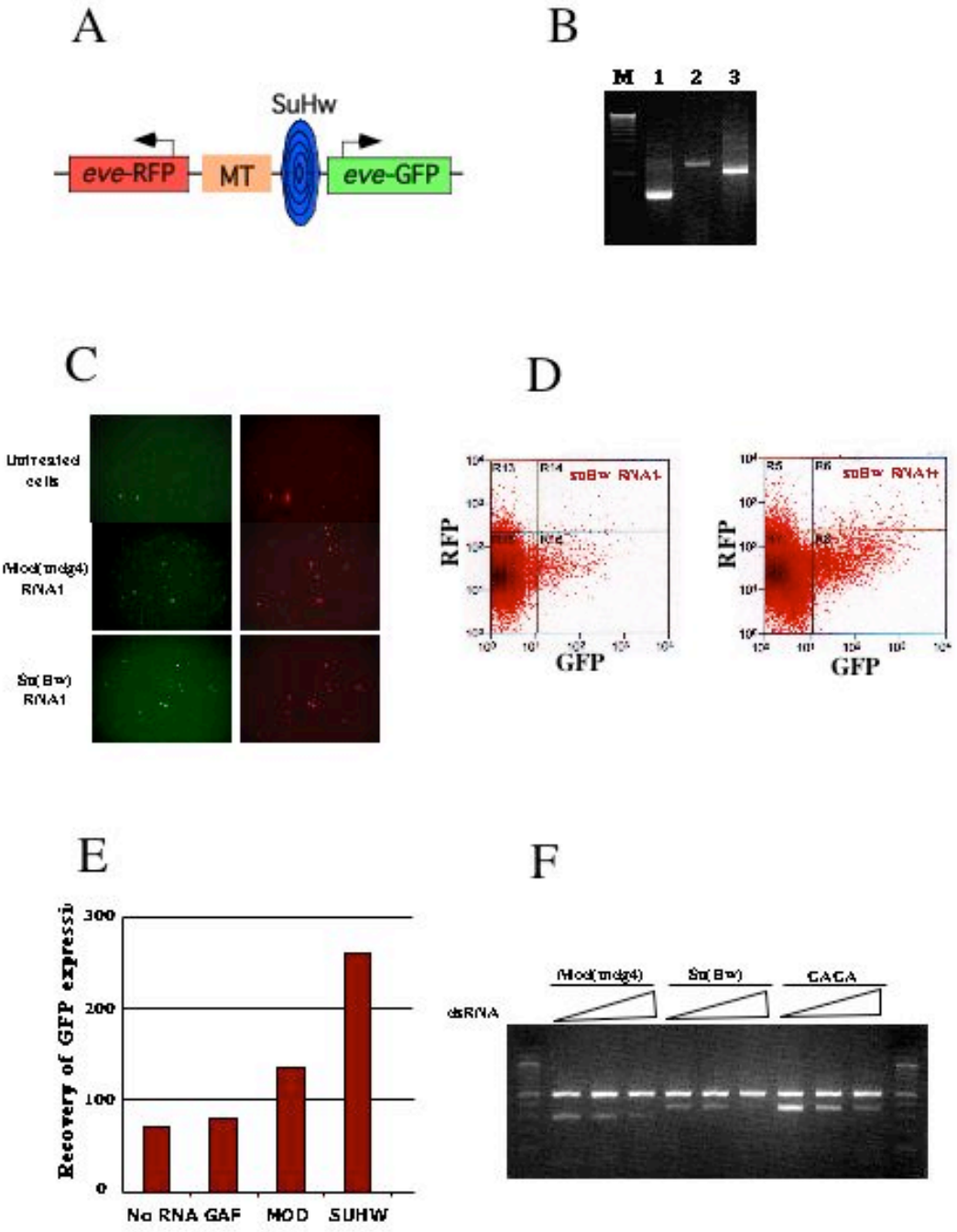
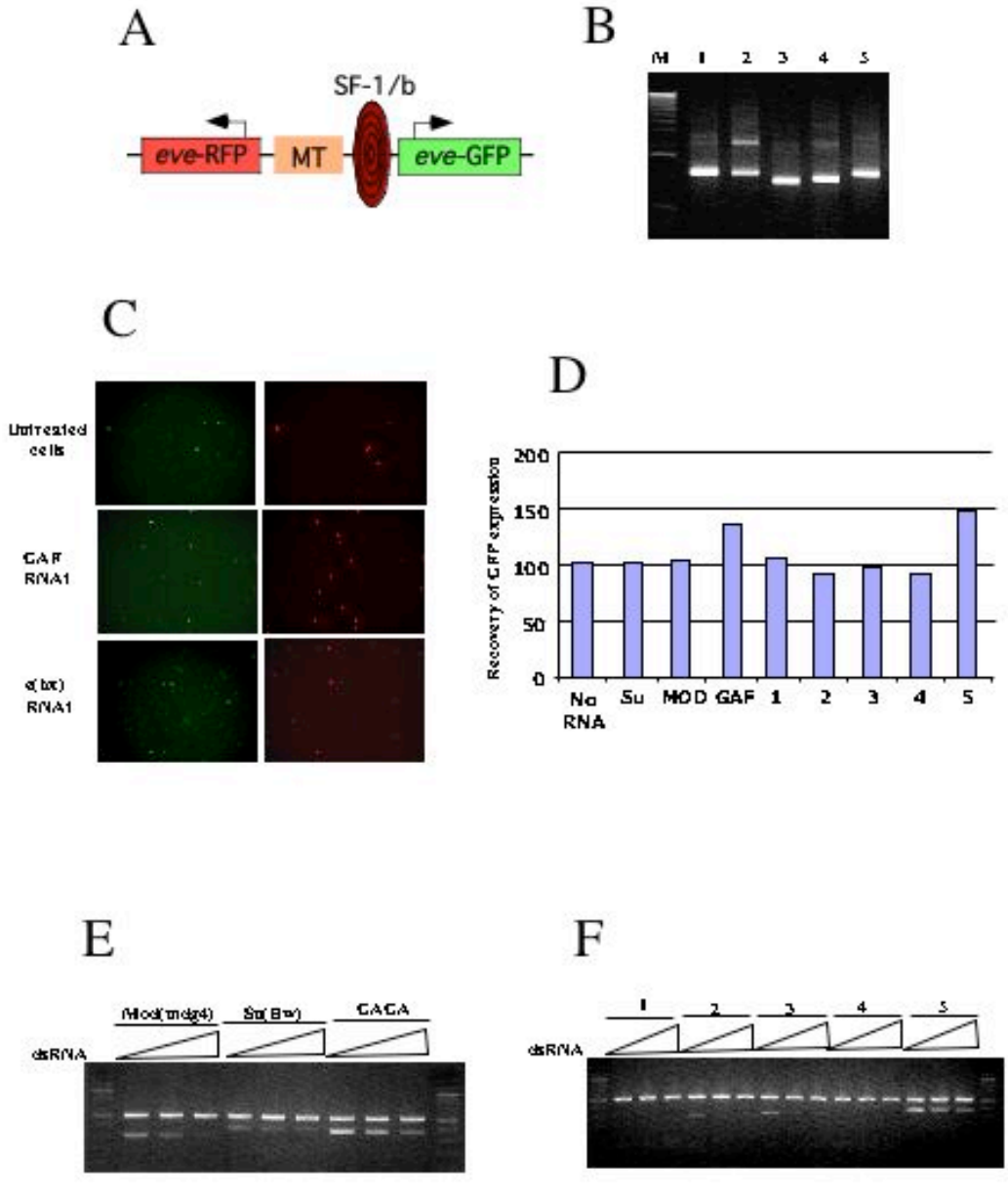

A
SuHw
eve-RFP
MT
eve-GFP
B
M 1 2 3
C
Untreated cells
(Mod(mdg4) RNA1
Su(Hw) RNA1
D
RFP
GFP
E
Recovery of GFP expressi
300
200
100
0
No RNA GAF MOD SUHW
F
(Mod(mdg4)
Su(Hw)
GAGA
dsRNA

Figure 28. RNAi-mediated knock-out of the SF-1 insulator function. (A) A schematic of the SF-1/b insulator-containing double-reporter construct used for the experiment. **(B)** The *in vitro* synthesized double-stranded RNAs corresponding to CG14648 (lane 1), CG1244 (lane 2), CG12822 (lane 3), CG15812 (lane 4), and CG7022 (lane 5). M – DNA size marker. **(C)** The *Drosophila Schneider* S2 cells were transiently transfected with the SF-1/b-containing double-reporter construct (shown in A), treated with dsRNAs indicated on the left side for 78 hours, induced with 1 mM Cu^{2+} and photographed using Zeiss Axioplan 2 microscope with GFP and PI filter sets. A slight increase in GFP fluorescence is observed in the GAGA factor and *e(bx)* dsRNA-treated cells compared to the untreated control. **(D)** The degree of the recovery of GFP expression in cells treated with different dsRNAs. Refer to Materials and Methods for the details on quantitative assessment of GFP fluorescence recovery. Only the GAGA factor and *e(bx)* RNAi cells exhibit consistent increase in the GFP fluorescence. **(E and F)** The semi-quantitative multiplex RT-PCR analysis of the dsRNA-treated cells. The genes targeted by RNAi are shown on top. In **(F)** 1 represents CG14648, 2 – CG1244, 3 – CG 12822, 4 – CG15812, 5 – CG7022. The first lane in each triplet represents an untreated control; the second lane – cells treated with 5 μg respective dsRNA; third lane – 15 μg dsRNA. The top band in each lane represents the actin 88F internal control. For all three genes the dsRNA dosage-sensitive reduction of the transcript level is observed.


A
SF-1/b
eve-RFP
MT
eve-GFP
B
M 1 2 3 4 5
C
Untreated cells
GAF RNA1
e(bx) RNA1
D
Recovery of GFP expression
200
150
100
50
0
No RNA
Su
MOD
GAF
1
2
3
4
5
E
dsRNA
CACA
F
dsRNA
1
2
3
4
5

CHAPTER 6

CONCLUSION

The work described in this dissertation has focused on understanding the biological function and the molecular mechanism of SF-1, a novel chromatin boundary from the Scr-*ftz* region in the *Drosophila* Antennapedia complex. Chromatin boundaries regulate gene expression in two ways. They modulate enhancer-promoter interactions by blocking distal enhancers (insulator activity) and protect genes from the influences of neighboring chromatin (barrier activity). Previous work by our laboratory has shown that SF-1 contains both of these activities. Interestingly, they reside in separate DNA fragments: the insulator activity is associated with the SF-1/b fragment, whereas the barrier activity lies mainly in the SF-1/c fragment. This is the first example of such activity separation in an animal model.

Our transgenic studies showed that SF-1/b but not SF-1/c is capable of blocking various embryonic and adult enhancers. In particular, SF-1/b blocks the *ftz* distal enhancer, an element that does not rely on promoter competition for the selection of appropriate target promoter. On the other hand, SF-1/c but not SF-1/b prevents the spread of silent chromatin initiated at the *Scr* PRE as revealed by the mini-*white* eye color assay. Based on these results we proposed a model for the dual function of SF-1 in the region: a) it protects the *Scr* promoter from inappropriate activation by nearby *ftz* enhancers and b) it protects a non-homeotic gene *ftz* from the effects of *Scr*-proximal PRE. The *ftz* gene domain may be similarly protected by another, putative, boundary, SF-2, located downstream of *ftz*.

The molecular mechanism of insulator and barrier activities appears to be distinct, i.e. SF-1/b critically depends on the GAGA factor, whereas SF-1/c does not. To gain further insights into the mechanism of insulator function, we identified the minimal sequence, SF-1/b3, necessary for the enhancer-blocking activity. This sequence is highly conserved (>98% identity) among four *Drosophila* species separated by approximately 5 million years. This degree of the evolutionary conservation in a non-coding sequence is suggestive of an important biological function of SF-1.

The SF-1/b3 minimal insulator sequence, as well as the 9 bp direct repeat site found in SF-1/b3, were used as baits in a yeast one-hybrid screen of the *Drosophila* embryonic cDNA library. Five putative SF-1 trans-factors were isolated in this screen. To test the functional relevance of these proteins, we developed a novel cell culture-based enhancer-blocking assay. This assay utilizes the inducible expression of GFP and RFP reporters driven by the metallothionin enhancer in transiently transfected S2 cells. The insertion of an insulator sequence between the reporter and the enhancer significantly attenuated the level of fluorescent signal. We demonstrate the utility of this cell culture system for the genome-wide RNAi-mediated search for the protein components of insulators.

We "knocked-down" the expression of the candidate SF-1 trans-factors by RNAi and observed the effect of the reduced transcript level on the SF-1/b insulator activity. The "knock-down" of *enhancer of bithorax* and the GAGA factor were found to attenuate the enhancer-blocking activity of SF-1/b in this assay. It is known that these two proteins interact *in vitro*. Therefore, it is possible that they also interact at the SF-1/b sequence and form the insulator complex. Future work will be needed to elucidate the aspects of the GAGA and *e(bx)* function that result in enhancer-blocking.

REFERENCES

1. Felsenfeld, G., *Quantitative approaches to poblems of eukaryotic gene expression.* Biophys Chem, 2003. **100**(1-3): p. 607-13.
2. Struhl, K., *Fundamentally different logic of gene regulation in eukaryotes and prokaryotes.* Cell, 1999. **98**(1): p. 1-4.
3. Berger, S.L. and G. Felsenfeld, *Chromatin goes global.* Mol Cell, 2001. **8**(2): p. 263-8.
4. Felsenfeld, G., et al., *Chromatin structure and gene expression.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9384-8.
5. Kadonaga, J.T., *Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors.* Cell, 2004. **116**(2): p. 247-57.
6. Collado-Vides, J., *A linguistic representation of the regulation of transcription initiation. I. An ordered array of complex symbols with distinctive features.* Biosystems, 1993. **29**(2-3): p. 87-104.
7. Buratowski, S., *Mechanisms of gene activation.* Science, 1995. **270**(5243): p. 1773-4.
8. Verrijzer, C.P., et al., *Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II.* Cell, 1995. **81**(7): p. 1115-25.
9. Butler, J.E. and J.T. Kadonaga, *The RNA polymerase II core promoter: a key component in the regulation of gene expression.* Genes Dev, 2002. **16**(20): p. 2583-92.
10. Kadonaga, J.T., *The DPE, a core promoter element for transcription by RNA polymerase II.* Exp Mol Med, 2002. **34**(4): p. 259-64.

11. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter.* Annu Rev Biochem, 2003. **72**: p. 449-79.

12. Khoury, G. and P. Gruss, *Enhancer elements.* Cell, 1983. **33**(2): p. 313-4.

13. Blackwood, E.M. and J.T. Kadonaga, *Going the distance: a current view of enhancer action.* Science, 1998. **281**(5373): p. 61-3.

14. Wu, C.T. and J.R. Morris, *Transvection and other homology effects.* Curr Opin Genet Dev, 1999. **9**(2): p. 237-46.

15. Maniatis, T., et al., *Structure and function of the interferon-beta enhanceosome.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 609-20.

16. Carey, M., *The enhanceosome and transcriptional synergy.* Cell, 1998. **92**(1): p. 5-8.

17. Merika, M. and D. Thanos, *Enhanceosomes.* Curr Opin Genet Dev, 2001. **11**(2): p. 205-8.

18. Munshi, N., et al., *Coordination of a transcriptional switch by HMGI(Y) acetylation.* Science, 2001. **293**(5532): p. 1133-6.

19. Yie, J., K. Senger, and D. Thanos, *Mechanism by which the IFN-beta enhanceosome activates transcription.* Proc Natl Acad Sci U S A, 1999. **96**(23): p. 13108-13.

20. Lamond, A.I. and W.C. Earnshaw, *Structure and function in the nucleus.* Science, 1998. **280**(5363): p. 547-53.

21. Zlatanova, J.S. and K.E. van Holde, *Chromatin loops and transcriptional regulation.* Crit Rev Eukaryot Gene Expr, 1992. **2**(3): p. 211-24.

22. Gerasimova, T.I. and V.G. Corces, *Boundary and insulator elements in chromosomes.* Curr Opin Genet Dev, 1996. **6**(2): p. 185-92.

23. Labrador, M. and V.G. Corces, *Setting the boundaries of chromatin domains and nuclear organization.* Cell, 2002. **111**(2): p. 151-4.

24. Bell, A.C. and G. Felsenfeld, *Stopped at the border: boundaries and insulators.* Curr Opin Genet Dev, 1999. **9**(2): p. 191-8.

25. Udvardy, A., *Dividing the empire: boundary chromatin elements delimit the territory of enhancers.* Embo J, 1999. **18**(1): p. 1-8.

26. Mirkovitch, J., M.E. Mirault, and U.K. Laemmli, *Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold.* Cell, 1984. **39**(1): p. 223-32.

27. Hancock, R., *A new look at the nuclear matrix.* Chromosoma, 2000. **109**(4): p. 219-25.

28. Udvardy, A., E. Maine, and P. Schedl, *The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains.* J Mol Biol, 1985. **185**(2): p. 341-58.

29. Kellum, R. and P. Schedl, *A group of scs elements function as domain boundaries in an enhancer-blocking assay.* Mol Cell Biol, 1992. **12**(5): p. 2424-31.

30. West, A.G., M. Gaszner, and G. Felsenfeld, *Insulators: many functions, many mechanisms.* Genes Dev, 2002. **16**(3): p. 271-88.

31. Chung, J.H., M. Whiteley, and G. Felsenfeld, *A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila.* Cell, 1993. **74**(3): p. 505-14.

32. Recillas-Targa, F., et al., *Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6883-8.

33. Chen, S. and V.G. Corces, *The gypsy insulator of Drosophila affects chromatin structure in a directional manner.* Genetics, 2001. **159**(4): p. 1649-58.

34. Bresnick, E.H. and G. Felsenfeld, *Dual promoter activation by the human beta-globin locus control region.* Proc Natl Acad Sci U S A, 1994. **91**(4): p. 1314-7.

35. Chung, J.H., A.C. Bell, and G. Felsenfeld, *Characterization of the chicken beta-globin insulator.* Proc Natl Acad Sci U S A, 1997. **94**(2): p. 575-80.

36. Ghirlando, R., et al., *Physical properties of a genomic condensed chromatin fragment.* J Mol Biol, 2004. **336**(3): p. 597-605.

37. Wei, W. and M.D. Brennan, *The gypsy insulator can act as a promoter-specific transcriptional stimulator.* Mol Cell Biol, 2001. **21**(22): p. 7714-20.

38. Wei, W. and M.D. Brennan, *Polarity of transcriptional enhancement revealed by an insulator element.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14518-23.

39. Kuhn, E.J., C.M. Hart, and P.K. Geyer, *Studies of the role of the Drosophila scs and scs' insulators in defining boundaries of a chromosome puff.* Mol Cell Biol, 2004. **24**(4): p. 1470-80.

40. Ohtsuki, S. and M. Levine, *GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo.* Genes Dev, 1998. **12**(21): p. 3325-30.

41. Rollins, R.A., P. Morcillo, and D. Dorsett, *Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes.* Genetics, 1999. **152**(2): p. 577-93.

42. Scott, K.C., A.D. Taubman, and P.K. Geyer, *Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength.* Genetics, 1999. **153**(2): p. 787-98.

43. Parnell, T.J. and P.K. Geyer, *Differences in insulator properties revealed by enhancer blocking assays on episomes.* Embo J, 2000. **19**(21): p. 5864-74.

44. Gerasimova, T.I., et al., *A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation.* Cell, 1995. **82**(4): p. 587-97.

45. Gerasimova, T.I., K. Byrd, and V.G. Corces, *A chromatin insulator determines the nuclear localization of DNA.* Mol Cell, 2000. **6**(5): p. 1025-35.

46. Xu, Q., et al., *Nuclear location of a chromatin insulator in Drosophila melanogaster.* J Cell Sci, 2004. **117**(Pt 7): p. 1025-32.

47. Ghosh, D., T.I. Gerasimova, and V.G. Corces, *Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function.* Embo J, 2001. **20**(10): p. 2518-27.

48. Read, D., et al., *Functional studies of the BTB domain in the Drosophila GAGA and Mod(mdg4) proteins.* Nucleic Acids Res, 2000. **28**(20): p. 3864-70.

49. Nabirochkin, S., M. Ossokina, and T. Heidmann, *A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence.* J Biol Chem, 1998. **273**(4): p. 2473-9.

50. Cai, H.N. and P. Shen, *Effects of cis arrangement of chromatin insulators on enhancer-blocking activity.* Science, 2001. **291**(5503): p. 493-5.

51. Melnikova, L., et al., *Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14806-11.

52. Zhao, K., C.M. Hart, and U.K. Laemmli, *Visualization of chromosomal domains with boundary element-associated factor BEAF-32.* Cell, 1995. **81**(6): p. 879-89.

53. Gaszner, M., J. Vazquez, and P. Schedl, *The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction.* Genes Dev, 1999. **13**(16): p. 2098-107.

54. Blanton, J., M. Gaszner, and P. Schedl, *Protein:protein interactions and the pairing of boundary elements in vivo.* Genes Dev, 2003. **17**(5): p. 664-75.

55. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-70.

56. Lewis, E.B., *Control of body segment differentiation in Drosophila by the bithorax gene complex.* Prog Clin Biol Res, 1982. **85 Pt A**: p. 269-88.

57. Lewis, E.B., *Regulation of the genes of the bithorax complex in Drosophila.* Cold Spring Harb Symp Quant Biol, 1985. **50**: p. 155-64.

58. Celniker, S.E., et al., *The molecular genetics of the bithorax complex of Drosophila: cis-regulation in the Abdominal-B domain.* Embo J, 1990. **9**(13): p. 4277-86.

59. Simon, J., et al., *Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products.* Dev Biol, 1993. **158**(1): p. 131-44.

60. Simon, J.A. and J.W. Tamkun, *Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes.* Curr Opin Genet Dev, 2002. **12**(2): p. 210-8.

61. Mihaly, J., et al., *Chromatin domain boundaries in the Bithorax complex.* Cell Mol Life Sci, 1998. **54**(1): p. 60-70.

62. Karch, F., et al., *Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of Drosophila melanogaster.* Nucleic Acids Res, 1994. **22**(15): p. 3138-46.

63. Barges, S., et al., *The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain.* Development, 2000. **127**(4): p. 779-90.

64. Zhou, J. and M. Levine, *A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo.* Cell, 1999. **99**(6): p. 567-75.

65. Hogga, I., et al., *Replacement of Fab-7 by the gypsy or scs insulator disrupts long-distance regulatory interactions in the Abd-B gene of the bithorax complex.* Mol Cell, 2001. **8**(5): p. 1145-51.

66. Schweinsberg, S.E. and P. Schedl, *Developmental modulation of Fab-7 boundary function.* Development, 2004. **131**(19): p. 4743-9.

67. Banerjee, S., et al., *Igf2/H19 imprinting control region (ICR): an insulator or a position-dependent silencer?* ScientificWorldJournal, 2001. **1**: p. 218-24.

68. Wolffe, A.P., *Transcriptional control: imprinting insulation.* Curr Biol, 2000. **10**(12): p. R463-5.

69. Lewis, A. and A. Murrell, *Genomic imprinting: CTCF protects the boundaries.* Curr Biol, 2004. **14**(7): p. R284-6.

70. Szabo, P., et al., *Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function.* Curr Biol, 2000. **10**(10): p. 607-10.

71. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.* Nature, 2000. **405**(6785): p. 486-9.

72. Schoenherr, C.J., J.M. Levorse, and S.M. Tilghman, *CTCF maintains differential methylation at the Igf2/H19 locus.* Nat Genet, 2003. **33**(1): p. 66-9.

73. Li, E., C. Beard, and R. Jaenisch, *Role for DNA methylation in genomic imprinting.* Nature, 1993. **366**(6453): p. 362-5.

74. Thorvaldsen, J.L., et al., *Analysis of sequence upstream of the endogenous H19 gene reveals elements both essential and dispensable for imprinting.* Mol Cell Biol, 2002. **22**(8): p. 2450-62.

75. Srivastava, M., et al., *H19 and Igf2 monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19.* Genes Dev, 2000. **14**(10): p. 1186-95.

76. Kaufman, T.C., M.A. Seeger, and G. Olsen, *Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster.* Adv Genet, 1990. **27**: p. 309-62.

77. Duboule, D. and G. Morata, *Colinearity and functional hierarchy among genes of the homeotic complexes.* Trends Genet, 1994. **10**(10): p. 358-64.

78. Riley, P.D., S.B. Carroll, and M.P. Scott, *The expression and regulation of Sex combs reduced protein in Drosophila embryos.* Genes Dev, 1987. **1**(7): p. 716-30.

79. Mahaffey, J.W. and T.C. Kaufman, *Distribution of the Sex combs reduced gene products in Drosophila melanogaster.* Genetics, 1987. **117**(1): p. 51-60.

80. Gindhart, J.G., Jr., A.N. King, and T.C. Kaufman, *Characterization of the cis-regulatory region of the Drosophila homeotic gene Sex combs reduced.* Genetics, 1995. **139**(2): p. 781-95.

81. Gorman, M.J. and T.C. Kaufman, *Genetic analysis of embryonic cis-acting regulatory elements of the Drosophila homeotic gene sex combs reduced.* Genetics, 1995. **140**(2): p. 557-72.

82. Ingham, P.W., *trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective.* Int J Dev Biol, 1998. **42**(3): p. 423-9.

83. Gindhart, J.G., Jr. and T.C. Kaufman, *Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene Sex combs reduced.* Genetics, 1995. **139**(2): p. 797-814.

84. Hafen, E., A. Kuroiwa, and W.J. Gehring, *Spatial distribution of transcripts from the segmentation gene fushi tarazu during Drosophila embryonic development.* Cell, 1984. **37**(3): p. 833-41.

85. Hiromi, Y., A. Kuroiwa, and W.J. Gehring, *Control elements of the Drosophila segmentation gene fushi tarazu.* Cell, 1985. **43**(3 Pt 2): p. 603-13.

86. Pick, L., et al., *Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated.* Genes Dev, 1990. **4**(7): p. 1224-39.

87. Amati, B., et al., *Nuclear scaffold attachment stimulates, but is not essential for ARS activity in Saccharomyces cerevisiae: analysis of the Drosophila ftz SAR.* Embo J, 1990. **9**(12): p. 4007-16.

88. Doe, C.Q., et al., *Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis.* Science, 1988. **239**(4836): p. 170-5.

89. Cai, H.N., et al., *Genomic context modulates insulator activity through promoter competition.* Development, 2001. **128**(21): p. 4339-47.

90. Ohtsuki, S., M. Levine, and H.N. Cai, *Different core promoters possess distinct regulatory activities in the Drosophila embryo.* Genes Dev, 1998. **12**(4): p. 547-56.

91. Belozerov, V.E., et al., *A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila.* Embo J, 2003. **22**(12): p. 3113-21.

92. Maier, D., D. Sperlich, and J.R. Powell, *Conservation and change of the developmentally crucial fushi tarazu gene in Drosophila.* J Mol Evol, 1993. **36**(4): p. 315-26.

93. Maier, D., A. Preiss, and J.R. Powell, *Regulation of the segmentation gene fushi tarazu has been functionally conserved in Drosophila.* Embo J, 1990. **9**(12): p. 3957-66.

94. Calhoun, V.C., A. Stathopoulos, and M. Levine, *Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex.* Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9243-7.

95. Kunkel, T.A., *Rapid and efficient site-specific mutagenesis without phenotypic selection.* Proc Natl Acad Sci U S A, 1985. **82**(2): p. 488-92.

96. Belozerov, V.E., et al., *A novel boundary element may facilitate independent gene regulation in the Antennapedia Complex of Drosophila.* EMBO J., 2003. **22**(12).

97. Bell, A., et al., *The establishment of active chromatin domains.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 509-14.

98. Richards, E.J. and S.C. Elgin, *Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects.* Cell, 2002. **108**(4): p. 489-500.

99. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression.* Science, 2003. **301**(5634): p. 798-802.

100. Ahmad, K. and S. Henikoff, *Epigenetic consequences of nucleosome dynamics.* Cell, 2002. **111**(3): p. 281-4.

101. Levis, R., T. Hazelrigg, and G.M. Rubin, *Effects of genomic position on the expression of transduced copies of the white gene of Drosophila.* Science, 1985. **229**(4713): p. 558-61.

102. Pirrotta, V., H. Steller, and M.P. Bozzetti, *Multiple upstream regulatory elements control the expression of the Drosophila white gene.* Embo J, 1985. **4**(13A): p. 3501-8.

103. Festenstein, R., et al., *Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner.* Nat Genet, 1999. **23**(4): p. 457-61.

104. Ayyanathan, K., et al., *Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation.* Genes Dev, 2003. **17**(15): p. 1855-69.

105. Spofford, J.B., *Parental control of position-effect variegation. II. Effect of sex of parent contributing white-mottled rearrangement in Drosophila melanogaster.* Genetics, 1961. **46**: p. 1151-67.

106. Schotta, G., et al., *Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila.* Semin Cell Dev Biol, 2003. **14**(1): p. 67-75.

107. Sass, G.L. and S. Henikoff, *Comparative analysis of position-effect variegation mutations in Drosophila melanogaster delineates the targets of modifiers.* Genetics, 1998. **148**(2): p. 733-41.

108. Ringrose, L. and R. Paro, *Remembering silence.* Bioessays, 2001. **23**(7): p. 566-70.

109. Hagstrom, K. and P. Schedl, *Remembrance of things past: maintaining gene expression patterns with altered chromatin.* Curr Opin Genet Dev, 1997. **7**(6): p. 814-21.

110. Kennison, J.A., *The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function.* Annu Rev Genet, 1995. **29**: p. 289-303.

111. Burgess-Beusse, B., et al., *The insulation of genes from external enhancers and silencing chromatin.* Proc Natl Acad Sci U S A, 2002. **99 Suppl 4**: p. 16433-7.

112. Spana, C., D.A. Harrison, and V.G. Corces, *The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon.* Genes Dev, 1988. **2**(11): p. 1414-23.

113. O'Hare, K., et al., *DNA sequence of the white locus of Drosophila melanogaster.* J Mol Biol, 1984. **180**(3): p. 437-55.

114. Ewart, G.D., et al., *Mutational analysis of the traffic ATPase (ABC) transporters involved in uptake of eye pigment precursors in Drosophila melanogaster. Implications for structure-function relationships.* J Biol Chem, 1994. **269**(14): p. 10370-7.

115. Roseman, R.R., V. Pirrotta, and P.K. Geyer, *The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects.* Embo J, 1993. **12**(2): p. 435-42.

116. Wilkins, R.C. and J.T. Lis, *Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation.* Nucleic Acids Res, 1997. **25**(20): p. 3963-8.

117. Perrod, S. and S.M. Gasser, *Long-range silencing and position effects at telomeres and centromeres: parallels and differences.* Cell Mol Life Sci, 2003. **60**(11): p. 2303-18.

118. Pirrotta, V., et al., *Assembly of Polycomb complexes and silencing mechanisms.* Genetica, 2003. **117**(2-3): p. 191-7.

119. Pirrotta, V. and L. Rastelli, *White gene expression, repressive chromatin domains and homeotic gene regulation in Drosophila.* Bioessays, 1994. **16**(8): p. 549-56.

120. Yao, S., et al., *Retrovirus silencer blocking by the cHS4 insulator is CTCF independent.* Nucleic Acids Res, 2003. **31**(18): p. 5317-23.

121. Vermaak, D., K. Ahmad, and S. Henikoff, *Maintenance of chromatin states: an open-and-shut case.* Curr Opin Cell Biol, 2003. **15**(3): p. 266-74.

122. Rubin, G.M. and A.C. Spradling, *Genetic transformation of Drosophila with transposable element vectors.* Science, 1982. **218**(4570): p. 348-53.

123. Kaufman, P.D. and D.C. Rio, *Germline transformation of Drosophila melanogaster by purified P element transposase.* Nucleic Acids Res, 1991. **19**(22): p. 6336.

124. Busturia, A., et al., *The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression.* Development, 2001. **128**(11): p. 2163-73.

125. Bi, X. and J.R. Broach, *Chromosomal boundaries in S. cerevisiae.* Curr Opin Genet Dev, 2001. **11**(2): p. 199-204.

126. Bi, X. and J.R. Broach, *UASrpg can function as a heterochromatin boundary element in yeast.* Genes Dev, 1999. **13**(9): p. 1089-101.

127. Fourel, G., et al., *Cohabitation of insulators and silencing elements in yeast subtelomeric regions.* Embo J, 1999. **18**(9): p. 2522-37.

128. Fourel, G., et al., *An activation-independent role of transcription factors in insulator function.* EMBO Rep, 2001. **2**(2): p. 124-32.

129. Harrison, D.A., et al., *The gypsy retrotransposon of Drosophila melanogaster: mechanisms of mutagenesis and interaction with the suppressor of Hairy-wing locus.* Dev Genet, 1989. **10**(3): p. 239-48.

130. Georgiev, P.G. and T.I. Gerasimova, *Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster.* Mol Gen Genet, 1989. **220**(1): p. 121-6.

131. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.* Cell, 1999. **98**(3): p. 387-96.

132. Moxley, R.A., H.W. Jarrett, and S. Mitra, *Methods for transcription factor separation.* J Chromatogr B Analyt Technol Biomed Life Sci, 2003. **797**(1-2): p. 269-88.

133. Gadgil, H., S.A. Oak, and H.W. Jarrett, *Affinity purification of DNA-binding proteins.* J Biochem Biophys Methods, 2001. **49**(1-3): p. 607-24.

134. Gadgil, H., L.A. Jurado, and H.W. Jarrett, *DNA affinity chromatography of transcription factors.* Anal Biochem, 2001. **290**(2): p. 147-78.

135. Silver, S.C. and S.W. Hunt, 3rd, *Techniques for cloning cDNAs encoding interactive transcriptional regulatory proteins.* Mol Biol Rep, 1993. **17**(3): p. 155-65.

136. Kamakaka, R.T. and J.T. Kadonaga, *The soluble nuclear fraction, a highly efficient transcription extract from Drosophila embryos.* Methods Cell Biol, 1994. **44**: p. 225-35.

137. Marini, M.G., et al., *Cloning MafF by recognition site screening with the NFE2 tandem repeat of HS2: analysis of its role in globin and GCSl genes regulation.* Blood Cells Mol Dis, 2002. **29**(2): p. 145-58.

138. Liao, M.X., et al., *Yeast one-hybrid system used to identify the binding proteins for rat glutathione S-transferase P enhancer I.* Biomed Environ Sci, 2002. **15**(1): p. 36-40.

139. Misawa, H. and M. Yamaguchi, *Identification of transcription factor in the promoter region of rat regucalcin gene: binding of nuclear factor I-A1 to TTGGC motif.* J Cell Biochem, 2002. **84**(4): p. 795-802.

140. Giot, L., et al., *A protein interaction map of Drosophila melanogaster.* Science, 2003. **302**(5651): p. 1727-36.

141. Wang, M.M. and R.R. Reed, *Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast.* Nature, 1993. **364**(6433): p. 121-6.

142. Alexander, M.K., B.D. Bourns, and V.A. Zakian, *One-hybrid systems for detecting protein-DNA interactions.* Methods Mol Biol, 2001. **177**: p. 241-59.

143. Clontech, *MATCHMAKER One-Hybrid System User Manual*. 1998: Palo Alto, CA.

144. Mannhaupt, G., et al., *Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast.* FEBS Lett, 1999. **450**(1-2): p. 27-34.

145. Thatcher, J.D., C. Haun, and P.G. Okkema, *The DAF-3 Smad binds DNA and represses gene expression in the Caenorhabditis elegans pharynx.* Development, 1999. **126**(1): p. 97-107.

146. Liu, Q., et al., *Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis.* Plant Cell, 1998. **10**(8): p. 1391-406.

147. Yu, Y., et al., *A double interaction screen identifies positive and negative ftz gene regulators and ftz-interacting proteins.* Mech Dev, 1999. **83**(1-2): p. 95-105.

148. Oda, E., et al., *Cloning and characterization of a GC-box binding protein, G10BP-1, responsible for repression of the rat fibronectin gene.* Mol Cell Biol, 1998. **18**(8): p. 4772-82.

149. Cathomen, T., et al., *A genetic screen identifies a cellular regulator of adeno-associated virus.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14991-6.

150. Mizuguchi, G., et al., *Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin.* Mol Cell, 1997. **1**(1): p. 141-50.

151. Reeves, R. and L. Beckerbauer, *HMGI/Y proteins: flexible regulators of transcription and chromatin structure.* Biochim Biophys Acta, 2001. **1519**(1-2): p. 13-29.

152. Clontech, *Yeast Protocols Handbook*. 2001: Palo Alto, CA.

153. Misquitta, L. and B.M. Paterson, *Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation.* Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1451-6.

154. Kennerdell, J.R. and R.W. Carthew, *Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway.* Cell, 1998. **95**(7): p. 1017-26.

155. Piccin, A., et al., *Efficient and heritable functional knock-out of an adult phenotype in Drosophila using a GAL4-driven hairpin RNA incorporating a heterologous spacer.* Nucleic Acids Res, 2001. **29**(12): p. E55-5.

156. Duffy, J.B., *GAL4 system in Drosophila: a fly geneticist's Swiss army knife.* Genesis, 2002. **34**(1-2): p. 1-15.

157. Schneider, I., and Blumenthal, A.B., *Drosophila cell and tissue culture*, in *The Genetics and Biology of Drosophila*, M.A.a.T.R.F. Wright, Editor. 1978, Academic Press: London. p. 265-315.

158. Echalier, G., *Drosophila Cells in Culture*. 1997, New York: Academic Press.

159. Abrams, J.M., et al., *Macrophages in Drosophila embryos and L2 cells exhibit scavenger receptor-mediated endocytosis.* Proc Natl Acad Sci U S A, 1992. **89**(21): p. 10375-9.

160. Clemens, J.C., et al., *Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6499-503.

161. Ilangovan, R., et al., *Inhibition of apoptosis by Z-VAD-fmk in SMN-depleted S2 cells.* J Biol Chem, 2003. **278**(33): p. 30993-9.

162. Celotto, A.M. and B.R. Graveley, *RNA interference of mRNA processing factors in Drosophila S2 cells.* Methods Mol Biol, 2004. **257**: p. 245-54.

163. Tourmente, S., et al., *Enhancer and silencer elements within the first intron mediate the transcriptional regulation of the beta 3 tubulin gene by 20-hydroxyecdysone in Drosophila Kc cells.* Insect Biochem Mol Biol, 1993. **23**(1): p. 137-43.

164. Tijsterman, M. and R.H. Plasterk, *Dicers at RISC; the mechanism of RNAi.* Cell, 2004. **117**(1): p. 1-3.

165. Lee, Y.S., et al., *Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways.* Cell, 2004. **117**(1): p. 69-81.

166. Carthew, R.W., *Gene silencing by double-stranded RNA.* Curr Opin Cell Biol, 2001. **13**(2): p. 244-8.

167. Kuttenkeuler, D. and M. Boutros, *Genome-wide RNAi as a route to gene function in Drosophila.* Brief Funct Genomic Proteomic, 2004. **3**(2): p. 168-76.

168. Boutros, M., et al., *Genome-wide RNAi analysis of growth and viability in Drosophila cells.* Science, 2004. **303**(5659): p. 832-5.

169. Lum, L., et al., *Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells.* Science, 2003. **299**(5615): p. 2039-45.

170. Beaulieu, J.F., A.P. Arrigo, and R.M. Tanguay, *Interaction of Drosophila 27,000 Mr heat-shock protein with the nucleus of heat-shocked and ecdysone-stimulated culture cells.* J Cell Sci, 1989. **92 (Pt 1)**: p. 29-36.

171. Bunch, T.A., Y. Grinblat, and L.S. Goldstein, *Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells.* Nucleic Acids Res, 1988. **16**(3): p. 1043-61.

172. Angelichio, M.L., et al., *Comparison of several promoters and polyadenylation signals for use in heterologous gene expression in cultured Drosophila cells.* Nucleic Acids Res, 1991. **19**(18): p. 5037-43.

173. Johansen, H., et al., *Regulated expression at high copy number allows production of a growth-inhibitory oncogene product in Drosophila Schneider cells.* Genes Dev, 1989. **3**(6): p. 882-9.

174. Balamurugan, K., et al., *Metal-responsive transcription factor (MTF-1) and heavy metal stress response in Drosophila and mammalian cells: a functional comparison.* Biol Chem, 2004. **385**(7): p. 597-603.

175. Giedroc, D.P., X. Chen, and J.L. Apuy, *Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation.* Antioxid Redox Signal, 2001. **3**(4): p. 577-96.

176. Zhang, B., et al., *The Drosophila homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals.* Mol Cell Biol, 2001. **21**(14): p. 4505-14.

177. Bevis, B.J. and B.S. Glick, *Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed).* Nat Biotechnol, 2002. **20**(1): p. 83-7.

178. Barolo, S., B. Castro, and J.W. Posakony, *New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP.* Biotechniques, 2004. **36**(3): p. 436-40, 442.

179. Badenhorst, P., et al., *Biological functions of the ISWI chromatin remodeling complex NURF.* Genes Dev, 2002. **16**(24): p. 3186-98.

180. Ito, T., et al., *ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor.* Cell, 1997. **90**(1): p. 145-55.

181. Varga-Weisz, P.D., et al., *Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II.* Nature, 1997. **388**(6642): p. 598-602.

182. Tsukiyama, T. and C. Wu, *Purification and properties of an ATP-dependent nucleosome remodeling factor.* Cell, 1995. **83**(6): p. 1011-20.

183. Xiao, H., et al., *Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions.* Mol Cell, 2001. **8**(3): p. 531-43.

184. Zhou, J., et al., *The regulation of enhancer-promoter interactions in the Drosophila embryo.* Cold Spring Harb Symp Quant Biol, 1997. **62**: p. 307-12.