8-2011

Endogenous gypsy insulators mediate higher order chromatin organization and repress gene expression in Drosophila

Shaofei Zhang
szhang5@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Shaofei Zhang entitled "Endogenous gypsy insulators mediate higher order chromatin organization and repress gene expression in Drosophila." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biochemistry and Cellular and Molecular Biology.

Mariano S. Labrador, Major Professor

We have read this dissertation and recommend its acceptance:

Bruce McKee, Ranjan Ganguly, Jae Park, Naima M. Moustaid

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Endogenous *gypsy* insulators mediate higher order chromatin organization and repress gene expression in *Drosophila*

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Shaofei Zhang

August 2011

ACKNOWLEDGEMENTS

My deepest gratitude goes first and foremost to my advisor, Dr. Mariano Labrador, for his constant support and professional instructions. His patient assistance and friendly encouragement has sustained me through frustration and depression. I have achieved so much that I could not have achieved without his illuminating suggestions and informative discussions. I am confident that the training I received in his lab will not only add greatly to my career future but also will deeply benefit me all my life.

I am greatly indebted to my doctoral committee members: Dr. Bruce McKee, Dr. Ranjan Ganguly, Dr. Jae Park and Dr. Naima Moustaid-Moussa, who have offered me valuable suggestions in my academic studies, which contributed greatly to the completion of this thesis.

I would like to express my heartfelt gratitude to Hyuck-Joon Kang, who is more like an elder brother to me rather than a lab-mate. Joon is always there to help but he never asks for any return. I would also like to express my gratitude to Piedad, Srilalitha, Heather, Cherie, Todd, who have helped me throughout this academic exploration, and brings me so much joy.

I owe a special debt of gratitude to my parents Yaowu Zhang and Meisheng Liu for their continuous encouragement and unswerving support. I also owe much to my wife Jia Guo, who has accompanied me to go through my doctoral studies and shared all my joys and sorrows.

Abstract

Chromatin insulators play an important role in gene transcription regulation by defining chromatin boundaries. Genome-wide studies in Drosophila have shown that a large proportion of insulator sites are found in intergenic DNA sequences, supporting a role for these elements as boundaries. However, approximately 40% of insulator sites are also found in intragenic sequences, where they can potentially perform as yet unidentified functions. Here we show that multiple Su(Hw) insulator sites map within the 110 kb sequence of the *muscleblind* gene (*mbl*), which forms a highly condensed chromatin structure in polytene chromosomes. Chromosome Conformation Capture assays indicate that Su(Hw) insulators mediate the organization of higher-order chromatin structures at the *mbl* locus, resulting in the formation of a barrier for the progression of RNA polymeraseII (PolII), and producing a repressive effect on basal and active transcription. The interference of intragenic insulators in PolII progression suggests a role for insulators in the elongation process. Supporting this interpretation, we found that mutations in su(Hw) and mod(mdg4) also result in changes in the relative abundance of the *mbl*D isoform, by promoting early transcription termination. These results provide experimental evidence for a new role of intragenic Su(Hw) insulators in higher-order chromatin organization, repression of transcription, and RNA processing.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter III

CHAPTER I

Introduction

Chromatin organization and gene regulation in the nucleus

The genome size of eukaryotes has increased dramatically as a result of evolution over millions of years. A question accompanied with this is how the large amount of genes present in the genome can be expressed efficiently with proper order during the development of a particular organism (Anatskaya and Vinogradov, 2007; Deato and Tjian, 2007; Xiao et al., 2006). Growing evidence suggests that nuclear organization may play an essential role in gene regulation by modulating accessibility of transcription machinery and varying the location of genes (Kosak et al., 2002; Muller et al., 2001; Schubeler et al., 2000).

In the eukaryotic cells，chromosomal DNA is packaged into chromatin, which in humans, for example, compacts two meters of DNA into approximately 5-μm-diameter nucleus. The basic unit of chromatin is the nucleosome. Nucleosomes are composed of a core histone octamer (containing two subunits each of four histones: H2A, H2B, H3 and H4), wrapped around DNA. All four histones contain lysine-rich tails at their amino termini, which undergo post-translational modifications including acetylation, methylation, ubiquitinylation

and phosphorylation (Goll and Bestor, 2002; Shindo, 2009). Histone modifications affect interactions of histones with DNA and nuclear proteins and subsequently act in diverse biological process such as DNA repair, mitosis and transcription regulation (Biancotto et al., 2010; Lo et al., 2004). The combination of histone modifications is thus thought to constitute a ‘histone code’ (Imhof and Becker, 2001; Jenuwein and Allis, 2001). The core histone octamer wraps 146 base pairs of DNA. In addition, linker hitone H1 or H5 wraps another 20 base pairs, forming two full turns of DNA around histone proteins (Luger and Hansen, 2005; Wong et al., 2007; Woodcock, 2006). Nucleosomes are regularly spaced along the genome. Repeating nucleosomes together with "linker" DNA form a 10-nm chromatin fiber, referred to as the "beads on a string" structure, which provides the first level of compaction of DNA into the nucleus. A chain of this structure is arranged into higher-order 30-nm fibers, which further coil into 80-100 nm chromatin fibers (Felsenfeld and McGhee, 1986; Ostashevsky and Lange, 1994). The chromatin structure beyond 30 nm fiber is not fully understood, but it is traditionally thought that the 30 nm fiber is assembled into loops along a central protein scaffold to form euchromatin. This arrangement of chromatin is considered to further condense DNA and form heterochromatin. Higher levels of compaction ultimately results in the highly condensed metaphase chromosome.

The terms euchromatin and heterochromatin were originally defined by their

distinct staining properties with a variety of chemical dyes: euchromain is lightly stained while heterochromatin stains darkly due to its highly condensed structure. In general, heterochromatin is frequently located at the nuclear periphery. It can be subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin refers to heterochromatin that remains compact in all cell types and tends to locate at repetitive sequences found in centromeres and telomeres of chromosomes. Facultative heterochromatin is defined as heterochromatin that can become decondensed during cellular development (Grewal and Jia, 2007; Tamaru, 2010).

The non-uniform compaction of the interphase chromosome is thought to be important for genome function. Such non-uniformity is easily observed in the banding pattern of polytene chromosomes in Drosophila. Polytene chromosomes form when certain specialized cells undergo multiple rounds of DNA replication without cell division, and the sister chromatids remain synapsed together to form a giant chromosome. The banding pattern of Polytene chromosomes is caused by alternation of highly condensed chromatin regions, which form bands, with less condensed chromatin regions, which form interbands. Interestingly, actively transcribed genes are mostly associated with decondensed interbands, whereas transcriptionally inactive genes are frequently located in compacted chromatin bands (Weintraub and Groudine, 1976; Zhimulev et al., 2004). It was also noted

that instead of randomly distributed on the chromosomes, genes with similar expression patterns are clustered in higher eukaryotes (Chen and Stein, 2006; Lercher et al., 2002; Mezey et al., 2008; Prieto et al., 2008; Roy et al., 2002). Therefore, it has been suggested that eukaryotic chromosomes are divided into distinct chromatin domains and that this organization of chromatin structure may function as the first level of regulation to control appropriate tissue-specific expression (Kuhn and Geyer, 2003).

Position-effect variegation in *Drosophila*

The phenomenon of position-effect variegation is one of the best known examples reflecting presence of structurally and functionally distinct chromatin states. The proper expression of the *white* gene is essential for normal red pigmentation of the fly's eye. The *white* gene locates in the euchromatin region of the X chromosome in wild type. A chromosomal inversion induced by X-ray places the *white* gene near the boundary between euchromatin and heterochromatin. As a result, the same *white* gene is expressed in some cells, in which is exposed to euchromatin, and is silenced in other cells, in which is exposed to heterochromatin. This patched distribution of pigmented cells finally leads to a phenotype that is known as variegated eyes (Grewal and Elgin, 2002). It has been suggested that a barrier exists in normal chromosomes to prevent spreading of

heterochromatin into euchromatic regions. In the absence of such barrier, genes such as *white*, which are placed close to heterochromatin may be silenced. The candidates to perform the barrier function are insulators or boundary elements.

Chromatin insulators or Boundary elements in the eukaryotic genome

Insulators or boundary elements are specialized DNA sequences bound by proteins that participate in chromatin organization and gene regulation by establishing and delimiting domains of gene expression and thus set up independent territories of gene activity (Labrador and Corces, 2002; Parnell et al., 2003). Insulators have two characteristic effects on gene expression: First, they block the interaction between enhancers and promoters when located between these elements. Second, they protect the expression of transgenes inserted into heterochromatin by preventing the spread of heterochromatin, and therefore separate active and inactive chromatin domains (Chung et al., 1993; Gaszner and Felsenfeld, 2006; Gerasimova and Corces, 2001; Kellum and Schedl, 1991; Roseman et al., 1995).

During the past few years, insulators have been found in a variety of organisms including yeast, *Drosophila*, and vertebrates (Barges et al., 2000; Engel and Bartolomei, 2003; Gerasimova and Corces, 2001; Pryde and Louis, 1999). In

yeast, the subtelomeric anti-silencing regions (STARs) can protect a reporter gene from being repressed by the neighboring silencing elements when located between them, which suggest the presence of insulators elements (Fourel et al., 1999). The first insulator identified in vertebrate is the 5' boundary of the chicken β-globin locus (cHS4), which posses both enhancer blocking and barrier activities in transgenic assays (Chung et al., 1993). The CCCTC-binding factor (CTCF), which associates with the 5'HS4 sequence in vivo, plays an important role in its insulator activity (Bell et al., 1999; Yusufzai and Felsenfeld, 2004). The 82-kDa CTCF protein, originally identified as a transcription factor, is an 11 zinc finger DNA-binding protein which recognizes diverse DNA regulatory sequences using different combinations of zinc fingers (Dunn and Davie, 2003; Parelho et al., 2008; Renda et al., 2007). CTCF has been shown to be evolutionarily conserved from frog, chicken, and rabbit to humans and appears to be the major insulator protein in vertebrates (Burgess-Beusse et al., 2002; Hore et al., 2008; Moon et al., 2005). The enhancer-blocking activity of CTCF has been studied in detail on the imprinted *Igf2/H19* locus (Chen et al., 2009; Engel and Bartolomei, 2003; Szabo et al., 2004; Yang et al., 2003).

Genomic imprinting is an epigenetic phenomenon which ensures that imprinted genes are expressed from only one allele in a parent-of-origin-dependent manner (Bartolomei and Tilghman, 1997; Ferguson-Smith and Surani, 2001; Kacem and

Feil, 2009). In mice and humans, the insulin-like growth factor 2 (*Igf2*), encoding an embryonic mitogen and H19, a putative tumor suppressor are reciprocally imprinted. *Igf2* is located 90 kb upstream from H19 with a boundary element named differentially methylated domain (DMD) [also known as imprinting control region (ICR)] positioned between two genes. Both genes share the same enhancer downstream of H19, while the DMD/ICR contains multiple binding sites for CTCF. In addition, the DMD/ICR is differentially methylated: in the paternal chromosome the DMD/ICR is methylated, preventing the binding of CTCF. In the maternal chromosome however the DMD/ICR is unmethylated, which allows binding of CTCF. CTCF bound DMR/ICR functions as an insulator to block communication between enhancer and *Igf2.* In this case, enhancer is restricted to activate nearby *H19*. In the paternal allele, methylation of DMD/ICR prevents binding of CTCF and allows downstream enchanter to activate *Igf2* gene.

Chromatin insulators in *Drosophila*

In *Drosophila,* at least five types of insulators have been identified. One of the best-characterized insulators is scs (specialized chromatin structure) and scs' flanking *Drosophila* hsp70 heat-shock genes (Udvardy et al., 1985). The scs element interacts with the eight zinc-fingers protein Zeste-white5 (Zw5), which is required for cell proliferation and differentiation. Homozygous mutations in *zw5*

gene are lethal. Male flies with hypomorphic alleles are viable but sterile, and display a series of developmental defects in bristle, eye and wings (Gaszner et al., 1999). The protein component that binds to scs' sequences and is required for their insulator functions is the boundary-element-associated factor 32 (BEAF 32) (Cuvier et al., 1998; Zhao et al., 1995). The *BEAF-32* gene encodes two different protein isoforms, BEAF-32A and BEAF-32B. These two proteins differ by about 80 amino acids at their N-termini but share common C-terminal region which are required for interaction between two BEAF isoforms (Aravind, 2000). Immunostaining on polytene chromosomes of *Drosophila* shows that BEAF binds to hundreds of sites independent of scs' element, and some of these binding sequences also have insulator properties (Cuvier et al., 1998; Zhao et al., 1995), However, the mechanism by which BEAF functions as insulator protein is not known in detail.

Several other insulators have been identified within the *Bithorax* complex (BX-C). The BX-C locus contains a cluster of homeotic genes, including *Ultrabithorax (Ubx), abdominal-A (abd-A) and Abdominal-B (Abd-B)* that regulate Drosophila body segmentation pattern. The regulatory region of BX-C is further subdivided into nine regulatory sub regions, and expression of *Abd-B* is specifically controlled by the subregulatory domains *infraabdominal-5 (iab-5)*, *iab-6*, *iab-7*, *iab-8* and *iab9*. Three boundary elements are found within the *Abd-B* locus: *Miscadastral*

Pigmentation (MCP), *Frontabdominal-7 (Fab-7)* and *Fab-8*. Proper insulator function of *Fab-7* and *Fab-8* is required to prevent interaction between regulatory sequences in adjacent *iab* subdivisions (Barges et al., 2000). *Fab-7* interacts with the DNA-binding protein GAGA factor (GAF), which is also a component of the insulator SF1 found in the *Antennapedia* complex (Belozerov et al., 2003). The *Fab-8* insulator contains binding sites for the *Drosophila* ortholog of vertebrate CTCF protein (dCTCF) (Kyrchanova et al., 2010; Moon et al., 2005). dCTCF colocalizes at several hundred sites on polytene chromosomes with Centrosomal protein 190 (CP190), another Drosophila insulator protein found in different insulators. Mutations in the *CP190* gene have been shown to affect insulator activity of *Fab-8*, suggesting that CP190 plays an important role in Fab-8 insulator function (Gerasimova et al., 2007; Mohan et al., 2007).

Gypsy **insulators in** ***Drosophila*** **genome**

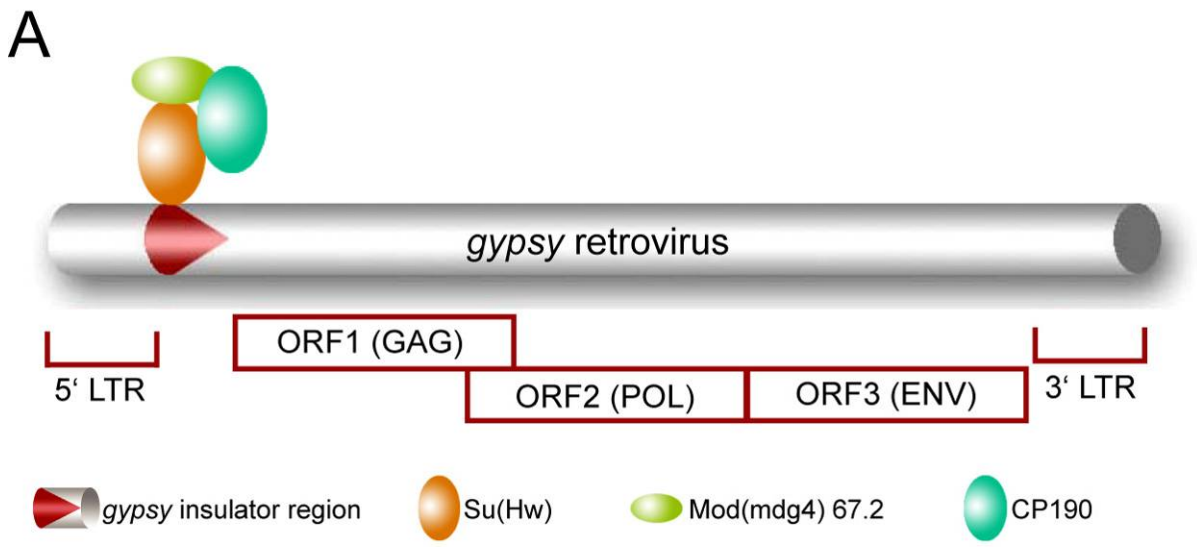
Another insulator that has been studied in detail is a 340-bp DNA fragment located in the 5' untranslated region of the *gypsy* retrotransposon (Geyer and Corces, 1992). The *gypsy* retrotransposon, also known as *mdg4*, belongs to a large class of mobile elements, which are widespread among species (Lee and Langley, 2010; Lorenc and Makalowski, 2003; Pimpinelli et al., 1995; van de Lagemaat et al., 2003). The 7.5 kb *gypsy* retrotransposon is flanked by two long

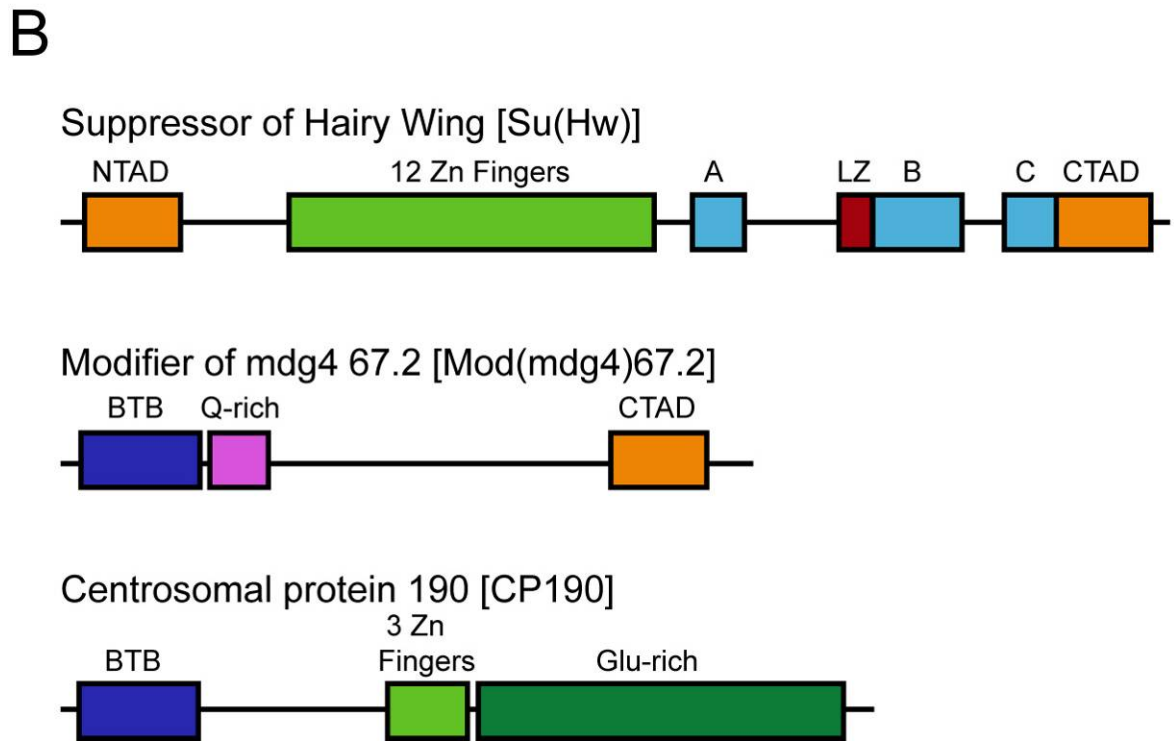
terminal repeats and encodes three retroviral genes: *gag*, *pol* and *env* (Figure 1.1 A). Insertion of *gypsy* into *Drosophila* genes such as *yellow (y)*, *cut*, the BX-C and the *achaete-scute* complex produces a variety of mutations throughout the genome (Corces and Geyer, 1991; Geyer et al., 1988; Mizrokhi et al., 1985). The effect of a *gypsy* insertion on the expression of the *yellow* gene has been studied in detail. *yellow* is regulated by a series of tissue-specific enhancers during development for blackish-brown pigmentation of larval tissues, wing blade, body cuticle, bristles and tarsal claws (Wittkopp et al., 2002). In the y^2 allele, *gypsy* is inserted at -700 bp from the transcription start site. The insertion of *gypsy* prevents interactions of upstream wing and body cuticle enhancers with the promoter, resulting in flies with yellow wing and yellow body phenotype, whereas the downstream enhancer is not affected, producing flies with black bristles (Corces and Geyer, 1991).

Genetic studies have identified a DNA fragment of 430 bp as responsible for *gypsy* mutagenesis. This fragment contains 12 reiterated binding sites for the Suppressor of Hairy wing (Su[Hw]) protein. Presence of this fragment in the original *gypsy* insertion site can reproduce the y^2 phenotype, suggesting that the Su(Hw) binding sequence alone is sufficient to recapitulate *gypsy*-induced phenotypic effects (Geyer and Corces, 1992; Spana and Corces, 1990). Mutations

Figure 1.1 Diagrammatic representations of *gypsy* retrotransposon and protein components of *gypsy* insulators in *Drosophila*

(A). The 7.5-kb *gypsy* retrotransposon has two long terminal repeats of 482 nucleotides and three open reading frames : *gag*, *pol* and *env* which encode gag-specific protease, reverse transcriptase and endonuclease. The region containing 12 Su(Hw) binding sites is shown in red. (B). Su(Hw) has a 12 zinc finger domains necessary for DNA binding. Other domains include: NTAD, amino-terminal acidic domain; LZ, leucine zipper; CTAD, carboxy-terminal acidic domain. The BTB domains found in Mod(mdg4) 67.2 and CP190 are responsible for interacting with Su(Hw) and recruitment to *gypsy* insulator sequences.


A
gypsy retrovirus
ORF1 (GAG)
5‘ LTR
ORF2 (POL)
ORF3 (ENV)
3‘ LTR
gypsy insulator region
Su(Hw)
Mod(mdg4) 67.2
CP190


B
Suppressor of Hairy Wing [Su(Hw)]
NTAD
12 Zn Fingers
A
LZ
B
C
CTAD
Modifier of mdg4 67.2 [Mod(mdg4)67.2]
BTB
Q-rich
CTAD
Centrosomal protein 190 [CP190]
BTB
3 Zn Fingers
Glu-rich

in the *suppressor of Hairy wing* gene [*su(Hw)*] completely rescue *gypsy*-induced mutations, indicating that Su(Hw) is a fundamental component of *gypsy* insulator and is required to elicit *gypsy*-induced mutant phenotypes (Modolell et al., 1983). The 110- kD Su(Hw) is a nuclear protein that is ubiquitously expressed in the nuclei of cells throughout developmental stages (Geyer and Corces, 1992; Roseman et al., 1993; Spana and Corces, 1990). Null mutations in *su(Hw)* are viable but cause female sterility (Harrison et al., 1993). The Su(Hw) protein has two highly acidic domains at both N-terminal and C-terminal regions (Figure 1.1 B). Comparison of the amino acid sequence of Su(Hw) in three different Drosophila species (*D. melanogaster*, *D. ananassae* and *D. virilis*) indicates that the central region (residues from 219 to 623) is highly conserved during evolution (with 80% identity and 95% similarity). This region contains 12 zinc finger motives, through which Su(Hw) binds directly to the octamer motif in the *gypsy* insulator (Parkhurst et al., 1988; Spana et al., 1988) .

Modifier of mdg4 (Mod[mdg4]), is another essential protein component of the *gypsy* insulator complex (Georgiev and Kozycina, 1996; Gerasimova et al., 1995). The *mod(mdg4)* gene encodes at least 29 different isoforms generated by alternative splicing. All the isoforms share the common amino-terminal domain containing a BTB/POZ motif (Gerasimova et al., 1995). Only one of these isoforms, Mod(mdg4)67.2 (also known as Mod[mdg4] 2.2) is involved in the

insulator function. Mod(mdg4)67.2 can interact with each other through its N-terminal BTB domain, while its C-terminal domain mediates interaction with Su(Hw) protein (Figure 1.1 B) (Gause et al., 2001; Ghosh et al., 2001). Mod(mdg4)67.2 does not bind to the *gypsy* insulator DNA directly, instead, it is recruited to insulator sequences through direct protein-protein interaction with Su(Hw) (Gause et al., 2001; Ghosh et al., 2001; Pai et al., 2004). Null mutations in *mod(mdg4)* are lethal, whereas mutations affecting only the Mod(mdg4)67.2 isoform are viable. One of such mutations, $Mod(mdg4)^{u1}$ is caused by the insertion of a *Stalker* retrotransposon into an exon unique to the Mod(mdg4) 67.2 isoform, resulting in the production of a truncated protein that cannot interact with Su(Hw) (Gerasimova et al., 1995; Mongelard et al., 2002).

Centrosomal Protein 190 (CP190) is the third identified component of *gypsy* chromatin insulator in *Drosophila* (Pai et al., 2004). This protein was originally isolated as a microtubule associated protein and has been shown to regulate myosin functions in *Drosophila* embryos (Chodagam et al., 2005; Kellogg et al., 1989). Further studies suggest that CP190 interact with many other insulator proteins and serves as a common protein component of multiple insulator complexes including BEAF insulators, the insulators at the Bithorax complex and the Fab-8 insulator (Gerasimova et al., 2007; Mohan et al., 2007; Negre et al., 2010). Null mutations in *cp190* are lethal, suggesting that CP190 plays an

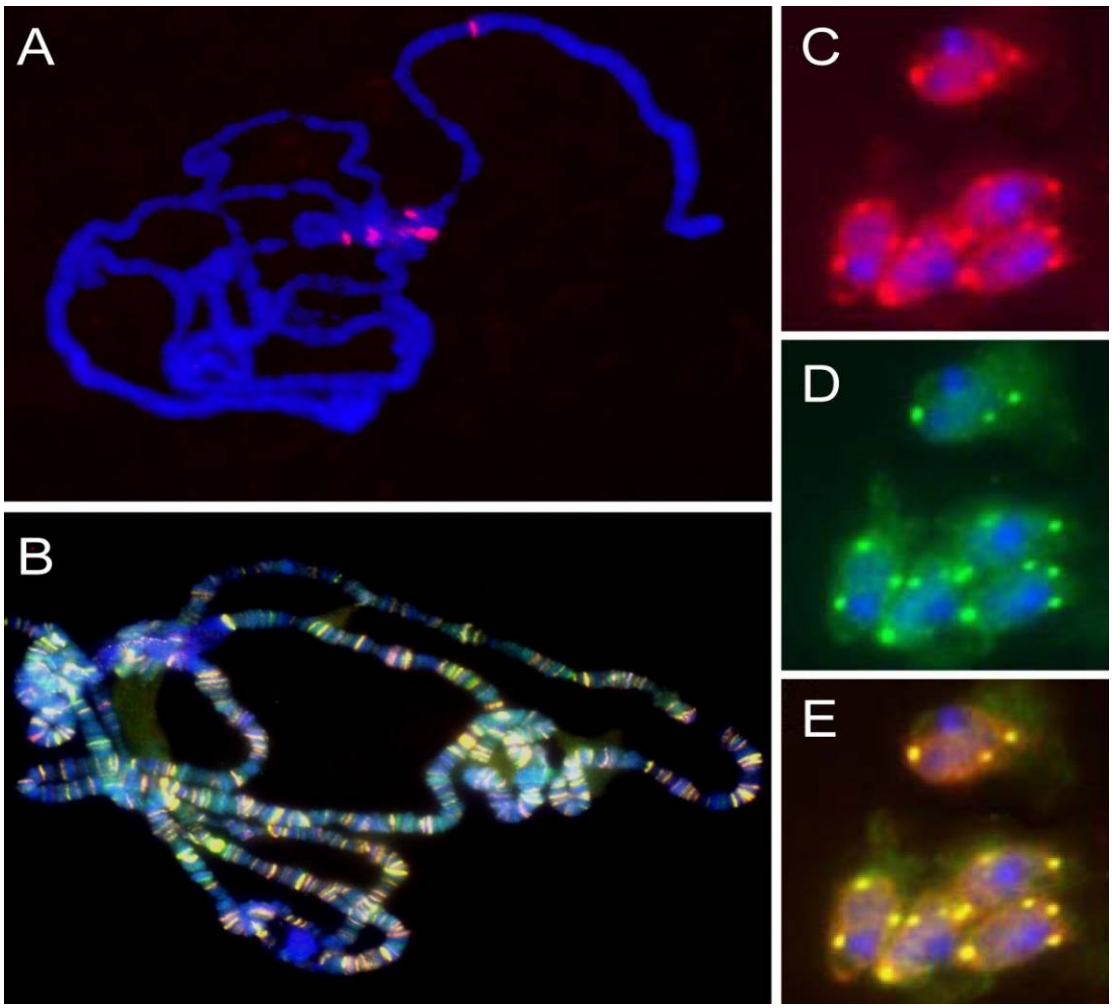
essential role in general insulator function. The *gypsy* insulator function is partially disrupted in a viable *CP190* mutant, which only encodes the first 755 N-termianl amino acids (Oliver et al., 2010; Pai et al., 2004). CP190 contains a BTB domain at its N-terminal end, three copies of C2H2 zinc fingers in the central region, and a Glu-rich motif at C-terminal end (Figure 1.1 B) (Pai et al., 2004). Like Mod (mdg4) 67.2, CP190 is recruited to *gypsy* insulator sequences through interactions with Su(Hw). The BTB domain, which is also found in all Mod(mdg4) isoforms, is necessary for interaction between insulator proteins and has been shown to be required for fly viability (Oliver et al., 2010).

Endogenous *gypsy* insulators and insulator bodies

Immunostaining analysis on the polytene chromosomes of *Drosophila* salivary glands reveals that the *gypsy* insulator proteins Su(Hw) and Mod(mdg4) co-localize at hundreds of sites independent of *Gypsy* sites (Figure 1.2 A, B) (Gerasimova and Corces, 1998, 2001; Pai et al., 2004). These observations

Figure 1.2 Immunostaining on the polytene chromomome and diploid cells showing distribution of *gypsy* retrotransposon, endogenous *gypsy* insulators as well as 'insulator bodies'

(A). Distribution of *gypsy* retrotransposons on polytene chromosomes of *Drosophila* (red). (B). Su(Hw) (red) and Mod(mdg4) 67.2 (green) co-localize at several hundred sites on the polytene chromosomes (Blue: DAPI stained DNA). (C)-(E). Hundreds of binding sites for Su(Hw) (red) and Mod(mdg4) (green) coalesce into large discrete speckles in the nuclei of diploid cells.


A
B
C
D
E

suggest the presence of endogenous Su(Hw) insulators independent of *gypsy* retrotransposon. The first identified endogenous Su(Hw) insulator is a 520 bp fragment resided in the1A-2 cytological location on the X chromosome. The 1A-2 insulator containing two Su(Hw) binding sites was demonstrated to posses enhancer-blocking activity in a transgenic assay (Golovnin et al., 2003; Parnell et al., 2003). Several other Su(Hw) insulators were determined in later experiments (Kuhn-Parnell et al., 2008). However, the function of these in vivo Su(Hw) binding sites remains largely unknown.

In contrast to the distribution pattern of endogenous *gypsy* insulators in polytene chromosomes, Su(Hw) and Mod(mdg4) have been shown to coalesce into around 20-25 large discrete foci in the nuclei of diploid cells, named ‘insulator bodies’. These ‘insulator bodies’ are though to represent higher order nuclear organization structures formed by multiple individual insulator elements that come together and held by interactions between insulator proteins as well as interactions with the nuclear matrix (Gerasimova et al., 2000; Lei and Corces, 2006; Pai et al., 2004).

Predictive models to explain insulating effects of insulators or boundary elements

Up to date, the detailed mechanisms by which chromatin insulators elicit their

functions are unknown. However, several models have been proposed to explain how insulators or boundary elements exert insulation effects (Gaszner and Felsenfeld, 2006; Wallace and Felsenfeld, 2007). The first set of models focus on interactions between insulators and the transcriptional activation machinery. The processive model is consistent with the tracking model of enhancer action. Tracking model suggests that enhancers activate promoters by launching an activating signal, which travels to the promoter and that could be either histone modifications or the RNA polymerase itself (Courey et al., 1986). Activating signals travel from enhancer towards promoter and therefore could be blocked by insulators located between enhancer and promoter. Supporting this model is the fact that insertion of the β-globin 5' HS4 insulator between enhancers and promoters leads to the accumulation of RNA polymerase Ⅱ at the insulator site and prevents spreading of histone H3 and H4 acetylation (Zhao and Dean, 2004). The decoy model is an alternative model that assumes that enhancers activate promoters through direct interaction (Li et al., 1991; Petrascheck et al., 2005; Wu et al., 2004). In this model, the insulation function is explained by the ability of insulators to compete with the promoter for interaction with enhancer. In this view, the enhancer is trapped by a direct interaction with the insulator and thus fails to interact with the promoter (Gaszner and Felsenfeld, 2006).

However, neither the tracking model nor the decoy model can explain the phenomenon observed when two *gypsy* insulators are introduced between an

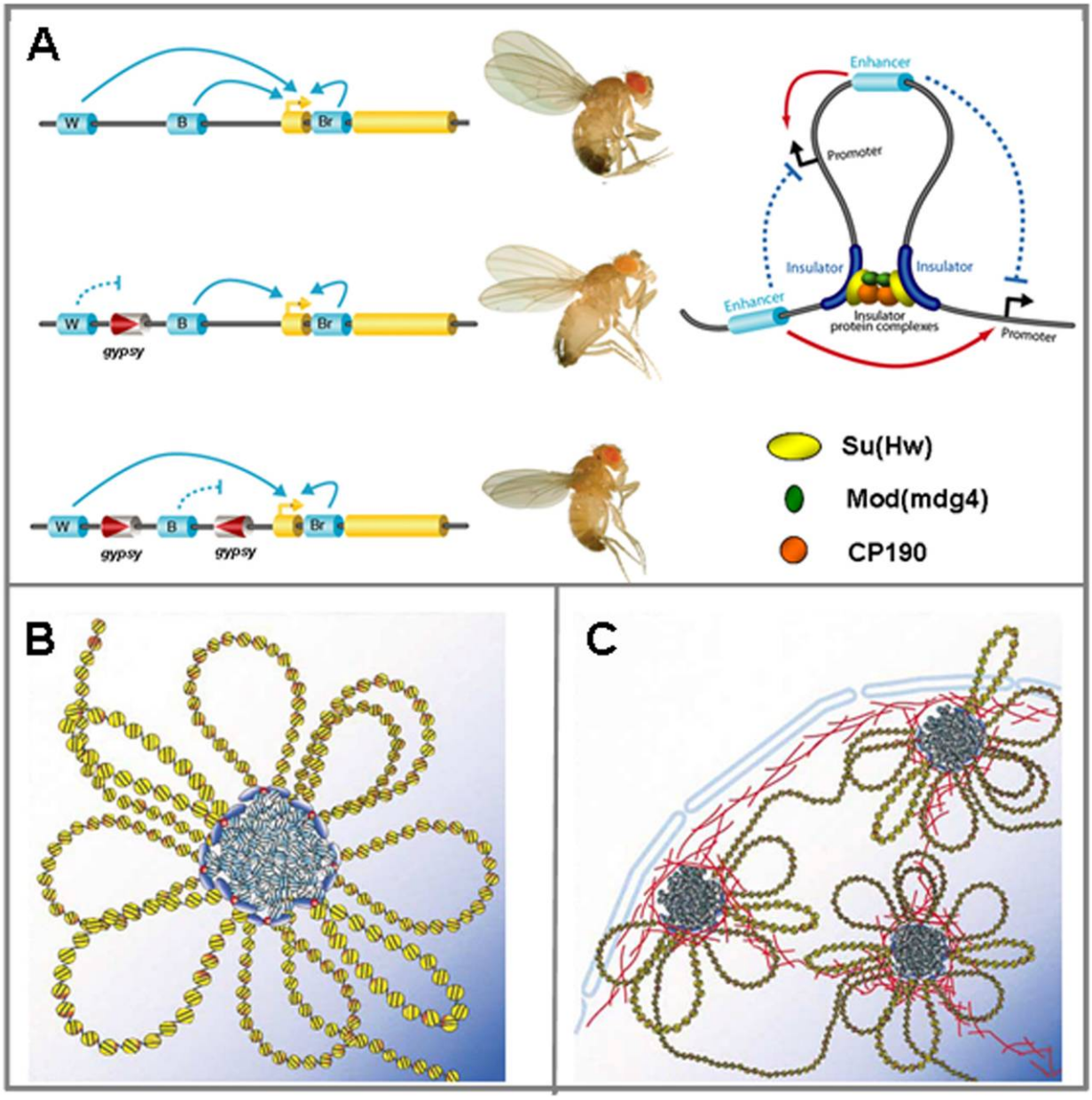
enhancer and a promoter, in which case their insulator activities are neutralized (Cai and Shen, 2001; Muravyova et al., 2001). The *yellow* gene is required for the normal pigmentation of the fly's cuticle structures, and is regulated by a series of tissue-specific enhancers including wing, body and bristle enhancers (Figure 1.3A). Wild type flies have black wings, body and bristles since all enhancers are able to activate the *yellow* promoter (Figure 1.3 A upper panel). A single insertion of *gypsy* insulator can efficiently block communications between upstream wing enhancer and promoter, without affecting downstream body and bristle enhancers. In this case, flies display a yellow wing phenotype (Figure 1.3 A middle panel). However, the insulator activity of *gypsy* insulators can be overcome by introducing a second insulator between body enhancer and *yellow* promoter. As a result, wing enhancer bypasses the paired *gypsy* insulators to activate the promoter, whereas the body enhancer is still blocked, determining a yellow body phenotype (Figure 1.3A bottom panel). As the BTB domain of CP190 protein, one of the major protein component of *gypsy* insulator, can interact with each other (Oliver et al., 2010), the loss of insulator activity described above has been proposed to result from the formation of a chromatin loop between the two adjacent *gypsy* insulators. The looping model suggests that interactions between *gypsy* insulators create topologically isolated domains, such that communication between enhancers and promoters can only occur within the same chromatin domain. According to the looping model, the formation of a chromatin loop brings wing enhancer and *yellow*

promoter in close proximity. As a result, instead of disrupting enhancer-promoter communication, the chromatin loop may facilitate their interaction. On the other hand, the body enhancer is isolated in a separate *gypsy*-insulator-defined domain, which restricts its interaction with the *yellow* promoter. In this model, insulators function primarily by partitioning promoter and enhancer into distinct chromatin loops and the insulating effects result from higher order chromatin organization mediated by insulators.

By extension, the structural model suggests that the chromatin loops formed by inter- or intra- chromosomal interactions may further interact with the nuclear matrix, forming in turn super insulator complexes called ‘insulator bodies’ (Figure 1.3 B). Experimental evidence supporting this assumption comes from the observation that Su(Hw) and Mod(mdg4) co-localize at ~500 sites on polytene chromosomes, but coalesce just into approximately 25 large speckles in diploid cells (Brasset and Vaury, 2005; Gerasimova and Corces, 1998; Spana et al., 1988). Insulator bodies may be the consequence of a particular type of nuclear organization that leads to the partition of the genome into topologically independent domains, and play an essential role in global gene regulation (Figure 1.3 C) (Gurudatta and Corces, 2009; Labrador and Corces, 2002).

Figure 1.3 Looping model showing function of *gypsy* insulators in *yellow* gene expression and nuclear organization

(A) Left: Schematic representation of regulatory region of *yellow* gene and positions of *gypsy* insertion. (W: wing enhancer, B: body enhancer, Br: bristle enhancer, golden bar with arrow: *yellow* promoter). In the wild type, all the three enhancers are able to activate *yellow* promoter (upper panel). A single *gypsy* insertion blocks wing enhancer from activating *yellow* promoter (middle panel). The second *gypsy* insertion neutralizes the insulator activities and allows interaction between wing enhancer and *yellow* promoter. Right: Proposed model to explain *gypsy* insulator function. Chromatin loop formed by paired *gypsy* insulators facilitates interactions between wing enhancer and promoter while looping out body enhancer. (B)-(C). Predictive interactions between insulators and nuclear matrix that form topological chromatin domains. In this model, insulators (blue) organize chromatin fiber (golden) into independent loops through attachment to nuclear lamina. Communication between enhancer and promoter can only occur when they are located in the same loop.


A
W
B
Br
gypsy
Enhancer
Promoter
Insulator
Insulator
Enhancer
Insulator
protein complexes
Promoter
Su(Hw)
Mod(mdg4)
CP190
B
C

General transcriptional machinery in Eukaryotes and assembly and recruitment of RNA polymerase (Pol) II to promoters

In general, the transcription process in eukaryotes can be divided into three major steps: initiation, elongation, and termination (Dynlacht, 1997; Krajewska, 1992). Transcriptional regulation can occur at each of three steps, which involve direct interactions between transcription machinery and a large number of regulatory factors (Kadonaga, 2004; Ptashne, 2005). This dissertation mainly focuses on the regulation of gene transcription during the initiation and elongation steps. Three different forms of RNA polymerase (Pol I, Pol II and Pol III) exist in eukaryotic cells. Pol II is primarily involved in transcription of protein-coding genes. The other two RNA polymerases, Pol I and Pol III, are responsible for transcription of non-coding RNAs such as ribosomal RNAs and tRNAs. Initiation of transcription by Pol II requires assembly of general transcription machinery, composed of general transcription factors (GTFs), mediator and Pol II. GTFs include a group of protein components TFII A, TFII B, TFII D, TFII E, TFII F and TFII H, which function corporately in the transcription initiation step. For example, the existing model for the recruitment of the general machinery to promoter DNA suggests that binding of TBP (TATA box-binding protein), a subunit of TFIID to the TATA box plays a crucial role in the reorganization of the core promoter. This process is facilitated by TFII A and TFII B (Kobayashi et al., 1995; Roberts et al., 1993; Wu et al., 1996)

and followed by recruitment of TFII F together with Pol II. Binding of remaining TFII E and TFII H leads to formation of the preinitiation complex (Langelier et al., 2001; Robert et al., 1996). Transcription is then initiated in the presence of ATP, which turns the preinitiation complex from a closed state into an open state (Jiang et al., 1993).

Transcriptional regulation at the initiation step

Regulation of transcription at the initiation step can be achieved via multiple mechanisms. Much of the regulation is dependent on function of a group of small modular proteins termed activators and repressors, which are required for the correct assembly of the transcriptional machinery on the core promoters (Biggar and Crabtree, 2000; Jiang et al., 1997). The precise mechanism by which activators can increase levels of gene transcription is not fully understood. However, it is generally accepted that activators function by anchoring to regulatory sequences called enhancers by their DNA binding domain, while interacting with components of the general transcriptional machinery on the promoter using their activation domain (Klemm et al., 1995; Stringer et al., 1990; Wu et al., 1996; Xiao et al., 1994). Binding of activators stabilizes the interaction between the transcriptional machinery and the promoter. Upon binding, activators increase the rate and extent of transcription machinery assembly and initiate

various subsequent events, together with additional positive factors named coactivators, which ultimately results in gene activation (Aoyagi and Archer, 2008; Paal et al., 1997; Vorobyeva et al., 2009).

Since DNA in eukaryotic cells is wrapped around histone to form nucleosomes, a structure unfavorable for transcription initiation, chromatin remodeling of the promoter is frequently required and precedes gene activation (Kadonaga, 1998; Roux-Rouquie et al., 1999). Transcriptional activators also recruit various histone acetyltransferases (HATs) and ATP-dependent remodeling enzymes. The combination of histone acetylation plus chromatin remodeling destabilizes local higher-order chromatin structures and permits binding of the transcription machinery on the promoter (Brodolin et al., 2005; Dilworth and Chambon, 2001; Sheldon et al., 1999). Thus transcriptional activators may facilitate transcription initiation by evicting nucleosomes from the target chromatin and thus counteracting repressive chromatin structures associated with the promoter region.

Transcriptional regulation during elongation

After initiation, transcription forwards into the elongation stage. Before entering productive elongation, Pol II complex transcribes only a short DNA sequence until

reaching an intrinsic pausing site, between +20 and +50 relative to transcription start site, where Pol II is halted by two factors: negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). The promoter-proximal pausing of Pol II may serve as an essential checkpoint during the early stage of elongation. Stalled Pol II is ultimately released upon recruitment of the positive transcription elongation factor pTEF-b, which phosphorylates NELF, DSIF and the C-terminal domain (CTD) of the Rpb1 subunit of Pol II. The phosphorylation of negative elongation factors as well as Pol II CTD at Ser2 residues facilitates transition of transcription from stalling of early elongation state to productive elongation state.

Although much emphasis has been laid on gene activation and repression during the transcription initiation step during the last decades, recent studies suggest that regulation of transcription may be particularly prevalent at the elongation stage (Anderson et al., 2011; D'Orso and Frankel, 2010; Gilchrist et al., 2009; Min et al., 2011). The process of transcription elongation is surprisingly complicated, and could be regulated at multiple levels. Since promoter-proximal pausing of RNA Pol II is a rate-limiting step during early elongation, one might predict that transcription elongation may be regulated by controlling transition of RNA Pol II from a non-processive state to a processive state. Indeed, it was reported that binding of heat shock factor (HSF) to heat shock elements stimulates the promoter-proximal paused polymerase release in *Drosophila* (Giardina et al.,

1992; Tang et al., 2000). Some other factors such as HIV Tat, VP16 and E1a have also been suggested to exert their effects by stimulating the processivity of elongating RNA polymerase (Kao et al., 1987; Marciniak and Sharp, 1991; Yankulov et al., 1994).

While these factors participate in transcriptional regulation by acting on the paused RNA Pol II itself, other factors may regulate transcription elongation by modulation of the chromatin environment. Nucleosomes are one of the major obstacles to RNA polymerase elongation *in vivo* when compared with naked DNA. It has been shown that the transcription rate is higher in the body of the gene compared with the promoter region, even though the former region contains a higher density of nucleosomes (Workman, 2006). This observation suggests that nucleosomes are evicted as the RNA polymerase moves forward. It also implies that factors responsible for the removal and replacement of histones can affect transcription rates during elongation. Indeed, it has been reported that nucleosome disassembly/reassembly factors such as FACT, Spt6 and Asf1 promote transcription both *in vitro* and *in vivo* by removing histones from nucleosomes (Adkins et al., 2004; Endoh et al., 2004; Kaplan et al., 2000; Pavri et al., 2006; Schwabish and Struhl, 2006).

Transcriptional regulation occurring at the elongation step appears to be much

more efficient than that occurring in the initiation step, since the formation of the preinitiation complex step is skipped, and therefore may provide a much faster response to the activation signal. In support of this idea, Chip-on-chip studies have revealed that RNA Pol II is frequently located within inactive genes (Radonjic et al., 2005). Occupation of silenced genes by RNA Pol II is suggested to provide potential advantages for maintaining developmentally important genes ready for activation (Kim et al., 2005; Lee et al., 2006; Radonjic et al., 2005).

Research questions addressed in this dissertation

Recently, Chip-on-chip experiments that have determined the *in vivo* binding sites of insulator proteins Su(Hw), dCTCF, BEAF and CP190 indicate that these insulator proteins are frequently located within the transcribed region of genes (Bushey et al., 2009). In addition, the previous immunostaining analysis of the 56F–58A cytogenetic location in our laboratory has shown that most of the strong Su(Hw) immunostaining signals correspond to Su(Hw) binding sites within long genes, which in turn are frequenty located in highly condensed chromatin regions. Interestingly, all the genes with a size larger than 20kb contain at least one Su(Hw) binding site, although the function of these endogenous Su(Hw) insulators is poorly understood.

In this dissertation, I have investigated the role of intragenic Su(Hw) insulators in the organization of higher order chromatin structure and in the regulation of gene expression.

CHAPTER II

Materials and methods

Drosophila stocks and crosses

Flies were raised on standard cornmeal medium and maintained at 25℃. All the fly stocks described in text were obtained from the *Drosophila* Bloomington Stock Center at Indiana University unless otherwise indicated. The lines bearing *mod(mdg4)*^{u1} mutations were obtained from V. Corces. *Drosophila* strains w^{1118}; PBac{RB}*su(Hw)*^{e04061}/TM6B, Tb^1 and y^2ct^6;*mod(mdg4)*^{u1} /TM6B, Tb^{1}were used for studying the effects of loss of Su(Hw) and Mod(mdg4) 67.2 insulator proteins on gene expression. Driver lines: y^1 w*; P{Act5C-GAL4}17bFO1/TM6B, Tb^1 were used to ectopically activate transcription of *mbl* and *Sdc* gene by crossing with y^1w^{67c23};p{$w^{+mc}y^{+mDint2}$=EPgy^2}*mbl*^{EY04602} and y^1w^{67c23};p{$w^{+mc}y^{+mDint2}$=EPgy^2}*Sdc*^{EY04602} respectively . Another driver line gal4hsp/*CyO*; MKRS/TM6 Tb was used to specifically drive expression of *mbl* and

Sdc at third instar larval stage.

In situ hybridization combined with immunostaining on polytene chromosomes

Salivary glands were dissected from third-instar larvae in 0.7% Saline Solution and fixed in 1:2:3 mixture of lactic acid:water:glacial acetic. Chromosomes were squashed and slides were submerged in liquid nitrogen until bubbling was over. After removal of coverslip, slides were transferd into chilled ethanol and slowly warmed up to room temperature (2-3 hours). Slides were air dried and kept at 4°C before hybridization. Probes with an average size of 1 kb were amplified by primers flanking the predictive Su(Hw) binding sites and labeled using Biotin High-Prime random priming kit (Roche). Labeled DNA was ethanol precipitated and resuspended in hybridization buffer to a final concentration of 4ng/μl. Prior to hybridization, slides were washed in 2×SSC at 65°C for 30 minutes, dehydrated in a series of ethanol at 65°C and denatured in 0.1 M NaOH. For hybridization, denatured probes were applied to pretreated slides and incubated at 37°C overnight in a moist chamber. Immunostaining was performed after in situ hybridization. Slides were washed three times in 2×SSC，1× PBS and one time in 1× blocking solution. Antibodies used for immunostaining were rabbit-anti Su(Hw) primary antibody and FITC-conjugated goat anti-rabbit IgG (Jackson

Laboratories). Following incubation, the slides were washed 5 minutes in PBS+0.1% Igepal and stained 25 seconds in 0.1µg/ml 4′,6-diamidino-2-phenylindole (DAPI). Stained polytene chromosome spreads were examined with a Leica DM6000B fluorescence microscope.

Chromatin immunoprecipitation

12-16 hour old *Drosophila* melanogaster embryos were collected from fly population cages supplied with grape-juice agar plates. The embryos were dechorionated with 50% bleach for 90 seconds at room temperature and immediately washed with 1 liter of Embryo Wash Solution (0.12 M NaCl, 0.04% Triton-X100). Embryos were fixed with 1.8% formaldehyde and homogenized in buffer A1 (60 mM KCl, 15 mM NaCl, 4 mM MgCl2, 15 mM HEPES (pH 7.6), 0.5% Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, and 1× EDTA-free protease inhibitor cocktail (Roche). A final concentration of 225mM of glycine was then added to terminate the cross-link reaction. The homogenate was washed three times in washing buffer and resuspended in lysis buffer (140 mM NaCl , 15 mM HEPES pH 7.6, 1 mM EDTA, 0.5mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, 0.05% SDS, 10 mM sodium butyrate, protease inhibitors). Chromatin was sonicated using sonifier (10 s continuous pulses with 20 s intervals on ice) yielding DNA fragments mostly between 500 and 700 bp.

Sonicated chromatin were then centrifuged at maximum speed for 10 min at room temperature. The supernatant was collected and incubated with slurry of Protein A-Sepharose beads to block non-specific binding. After removal of the Sepharose beads, a control sample was set aside for the ‘input’. The rest of the sample was immunoprecipitated with either Su(Hw) antibody or normal rabbit IgG. 50 ul of Protein A-Sepharose beads were added and incubated on a head to head rotating wheel overnight. The next day, Sepharose beads were washed four times in lysis buffer followed by two times wash in TE before elution. Reverse cross-link was performed at 65°C for 6 hours followed by addition of Proteinase K solution and incubation at 50°C for another 2 hours. After phenol/chloroform extraction, the DNA was ethanol-precipitated and used for real-time PCR with appropriate primers.

Immunostaining of polytene chromosomes

Salivary glands were dissected out from wandering third-instar larvae in 0.7% NaCl, passed to a drop of solution I (3.7% formaldehyde in 45% acetic acid). Chromosomes were squashed after 2 to 4 minutes incubation, slides were immersed in liquid nitrogen and the cover slips were removed. The chromosomes were incubated with primary antibody diluted in PBS containing 0.1% Igepal and 1% milk overnight at 4°C in a humid chamber. The slides were washed 5 minutes

with PBS containing 0.1% Igepal and incubated with secondary antibody for two hours at room temperature. After washing in PBS+0.1% Igepal, slides were stained with 0.5μg/ml of DAPI for 25 seconds and mounted in Vectashield medium. Slides were observed under Leica DM6000B fluorescence microscope. The antibodies used for immunostaining were: rabbit anti-Su(Hw), mouse anti-H14 (RNA Pol II) (Covance (Princeton, New Jersey)), FITC-conjugated goat anti-rabbit IgG, Texas red donkey anti-mouse IgM (Jackson Laboratory (Bar Harbor, ME)).

Over expression of Su(Hw), total RNA isolation and quantitative RT-PCR

To over express *Su(Hw)*, full length Su(Hw) coding sequence was cloned into the pBS-actTAP vector. The expression vector was then co-transfected into S2 cells with pBS-PURO which contains a puromycin resistant gene. Stable cell lines were established after three-week selection with 10μg/ml puromycin. TRIzol reagent (Invitrogen) was used to extract total RNA from S2 cells or third-instar larvae. One microgram of total RNA was reverse-transcribed to cDNA in a final volume of 20μl by using the iScript Select cDNA Synthesis Kit (Bio-Rad). Three independent biological RNA samples were prepared for each specific genotype and three parallel technical replicates were performed for each RNA sample. The enrichment of each individual was generated from the mean of nine total

replicates. The transcriptional level of rp49, which encodes a ribosomal protein, was used as an internal control. All the primers used in real-time PCR were optimized before use. Real-time PCR was performed by using iQ SYBR Green Supermix and BioRad iQ5 Multicolor Real-Time PCR Detection System. Melting curve was monitored to ensure the specificity of PCR product. The transcript abundance was determined using relative quantitative method (ΔΔCt value).

Quantitative chromatin conformation capture assay

10^8 *Drosophila* S2 cells were cross-linked with formaldehyde (1%) for 10 minutes at room temperature. The reaction was terminated with addition of 2.5ml glycine (2.5M). Nuclei were then isolated from cross-linked cells and resuspended in 0.5ml of EcoR I digestion buffer containing 1% SDS. After 10 minutes incubation at 65 °C, 10% Triton X-100 was added and mixed gently. EcoR I digestion was carried out at 37 °C for 16 hours. Following digestion, 10% SDS was added to reaction and enzymes were heat inactivated at 65°C for 30 minutes. Digested samples were diluted with T4 DNA ligase buffer containing 1% triton X-100 and incubated with 4,000 unites T4 DNA ligase for two hours at 16 °C. Proteinase K(10mg/ml) was added to ligated samples and incubated overnight at 65°C to reverse cross-linking. DNA was purified by phenol/chloroform extraction and

resolved in TE buffer. RNA was removed by incubation samples with DNAse-free RNase A (10mg/ml) at 37°C for 15 minutes. Enrichment of 3C products was then analyzed by quantitative real-time PCR with primer pairs across the restriction sites. Primer sequences are available upon request. Two minimally overlapping bacterial artificial chromosome (BAC) clones spanning region of interest (RP98-48A11 and RP98-28012) were obtained from Children's Hospital Oakland Research Insitute(CHORI) and used to generate a control library. Interaction frequency between different genome sites were determined by the ratio of the amount of PCR product obtained from 3C library and the amount of PCR product obtained from control library.

DNase I sensitivity assays

Third instar *Drosophila* larvae were collected, frozen in liquid nitrogen and stored at -80°C before use. Isolation of nuclei was performed according to (ElGIN, 1998). Isolated nuclei were resuspended in 1 ml of DNase I digestion buffer (60 mM KCl, 15mM NaCl, 15 mM Tris- HCl, pH 7.4, 0.25 M sucrose, 3 mM $Mgcl_2$, 0.5 mM DTT). Four aliquots of 250µl of nuclear suspension were incubated with 0,1,2,4 units of DNase I (Sigma) respectively on ice for three minutes and digestion was terminated with 5µl of 0.4 M EDTA. Digested nuclei were collected and resuspended in 250µl of sarcosyl lysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM

EDTA 0.5% (w/v) sodium laurylsarcosine), followed by incubation with 3μl of proteinase K at 37°C overnight. DNA was extracted from nuclei with phenol : chloroform : isoamyl alcohol (24:12:1) and treated with RNase A for one hour at 37 °C. After ethanol precipitation, the pellets were resuspended in 100μl of TE buffer. Quantification of remaining DNA after DNase I digestion was carried out by real-time quantitative PCR. DNase I sensitivity of each site tested was determined by plotting the percentage of copies of remaining DNA against units of DNaseI used for nuclei digestion. Three independent biological replicates were prepared for DNase I treatment and three technical replicates were performed in real-time PCR for each DNase I treament. The final digestion profiles were generated form a total of nine replicates.

Oligonucleotides

To perform in situ hybridization assays, the following oligonucleotides were synthesized:

Probe 1:

Sense: AATCTGCAGCGATGCCCGAAGTT

Antisense: GCCTGTGTTAGCTCCAGCGAATT

Probe 2:

Sense: AATCTGCAGCGATGCCCGAAGTT

Antisense: GCCTGTGTTAGCTCCAGCGAATT

Probe 3:

Sense: AATCTGCAGCGATGCCCGAAGTT

Antisense: GCCTGTGTTAGCTCCAGCGAATT

Probe 4:

Sense: AATCTGCAGCGATGCCCGAAGTT

Antisense: GCCTGTGTTAGCTCCAGCGAATT

For the chromatin immunoprecipitation experiment, the following oligonucleotides were synthesized:

Mbl12859494:

Sense: AATAAACACGGCAATGCAAGC

Antisense: CACGAAGCGGGCAAACTGGAG

Mbl12854990:

Sense: GCACATGCCGCACACAGAATCG

Antisense: CAGTTATTGCGCAGCTGTCTGATGT

Mbl12826430:

Sense: CGACGAACTTCCCGAAACCTAT

Antisense: CTGCGGTCAGGCTTCGTATT

Mbl12812914:

Sense: TATCCCTTTACTTCCTGCCACGCCT

Antisense: AGTAGTCGGAAAATGCTGTTGTTGC

For the transcription analysis, the following oligonucleotides were synthesized:

CG18469:

Sense: GTCATTTGGGTCAAGTTGCGTG

Antisense: CTGCCTCAGAGCCAGATTCATT

CG12699:

Sense: GGCGAAGCAGTCTAAGCCCT

Antisense: GCTGCATCTTCGACTTGGGG

mbl:

Sense: GTTCCAGCGCAACAAATGCT

Antisense: CACGGCGGTTTATCACGATT

Sip1:

Sense: GAATGAAGTCCGCTTGCTGC

Antisense: TCTCCTGTTTGGTGCCCTCG

CG6568:

Sense: CAAACATACCCCACTGAACTACCG

Antisense: ACTTATAGCTGCCGCTGCCC

cnk:

Sense: TGGAGCCGTAGGCAATGGAG

Antisense: GGTATCGGGGCCGAGTTCTT

Sdc:

Sense: GAACGTGAATAGCCAGCCCTCC

Antisense: CGCCAATGACAGCAGCCAGAA

Sara:

Sense: CCAACGGCATAGATGTCCCA

Antisense: TGCTGTCCAGTTCGGTGTCC

Fkbp13:

Sense: AGGTGTGCGAACAAAAGTCCAAG

Antisense: CGCTTCTCACCCACGCACAT

EfSec:

Sense: CTGCTGGAGCGTGGAATCAT

Antisense: GGAAGTTGTGCCGTCCGTAT

Acox57D-d:

Sense: GGAAATGTATGCCCAGACGG

Antisense: ATAGATGTTGCTCAAGTTGGCGG

mbl-RA:

Sense: AATACCAATTCCAAGGGGCAAGTAC

Antisense: TAGTGGTTAGAGCCGATACAGATGC

mbl-RB:

Sense: CACAATCACAGCACGAGCACA

Antisense: AGCGTCAGAGGCGAAAGCAA

mbl-RC:

Sense: TGACCTGATATTAAATCTTGT

Antisense: TTAGTTTGCTACACTTCTTCT

mbl-RD:

Sense: CCAGGCGACGCCGTTCATCT

Antisense: CCCATCTGCGTCTGGTAGT

Su(Hw):

Sense: ATTCACATACGCACTCATACGG

Antisense: TGTGACGATTGAGCACTTCCTTTAC

Treh:

Sense: CGTTTAGCAAGGACAGGCAC

Antisense: CTTGCTCAGCCACTCGATG

For the DNase I sensitivity assays, the following oligonucleotides were synthesized:

Gap:

TGCCAATATCAAATCCCAGTTCC

GTATGCGTCACAATCGTCAGAGT

Mbl promo:

CGGAAAAGTGAATTATACCAGAGCC

GATTATTTCTGTTTGCGGTGACTGA

Mbl nonSu:

TTTCCACTTGATGTGCGGCAACTTT

TTTACAACTCGTCTGTCGCTCA

mbl12854990:

GCACATGCCGCACACAGAATCG

CAGTTATTGCGCAGCTGTCTGATGT

Mbl12859494:

AATAAACACGGCAATGCAAGC

CACGAAGCGGGCAAACTGGAG

For the Chromatin conformation capture assays, the following oligonucleotides were synthesized:

Anchor A:

TTAATTTTACGCTGGCCAAACGGGA

Anchor B:

AAATGTCTGTTTGGGACTCTGCGGA

Other primers except anchor A and B used (From left to right in both figure 3.15 A and 2.22 B)

primer1 (13051995):

CCATTCCCTGTTAGTGTTTGTGCTT

primer2 (13088545):

TTGTTTTGCTTGGGGAATCTGCTGG

primer3 (13145839):

ATGCTGTGCACGGTGAAAATTCACA

primer4 (13173110):

CACAAATTCCAATGAACTGGCAACT

primer5 (13192748):

TATGTGTTTTTGCTGGCCGTCCGAC

primer6 (13203359):

GTAAGCGAGATGACAAATAAGGGGA

primer7 (13216672):

AGATTGATGATGCTAGTTTCTGGCG

primer8 (13220748):

TGGGCCTGGGAAAATAATAGAAAAG

primer9 (13233056):

AAATGTGGCCTAGAATATGGCGAAC

primer10 (13239296):

ATTTACGTTTGGATGGGGATAGGCC

primer11 (13244438):

GAGGAAGGGATAAAATCAGCTAACA

primer12 (13271445):

GCGTTCTCAGCCGCAATGGAATAAA

primer13 (13294005):

CTTGATGAACACAGCTTGGCGCACG

CHAPTER III

Results

Intragenic Su(Hw) insulators reduce basal expression levels of mature *mbl* and *Sdc* mRNAs

Previous work in our laboratory established a correlation between the distribution of endogenous Su(Hw) insulators and the structure of polytene chromosomes in *Drosophila* (Wallace et al., 2010). Remarkably, it was found that endogenous Su(Hw) insulators mapping to intragenic regions of long genes were generally found to form condensed polytene chromosome bands. One exception is the *Trehalase* gene (*Treh*), which also contains intragenic Su(Hw) insulators, but is found in an open chromatin structure associated to an interband (Wallace et al., 2010).

To gain insight into the role that intragenic Su(Hw) endogenous insulators play in

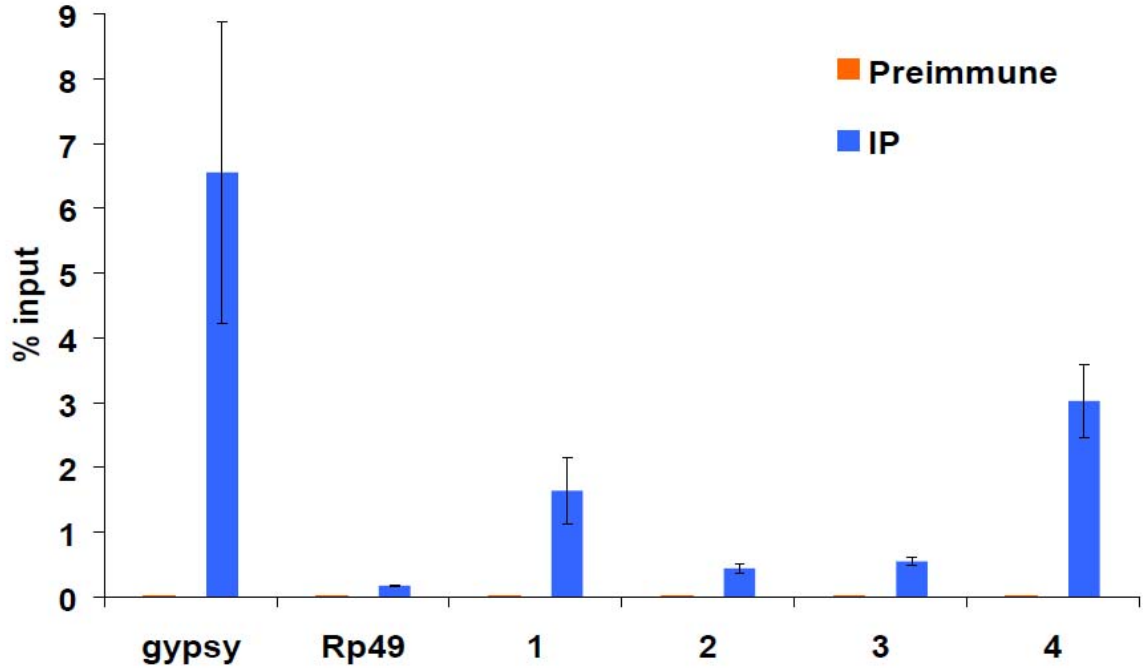
these genes, I have analyzed the transcription levels of *Syndecan (Sdc)*, *muscleblind (mbl)* and *Treh* in *su(Hw)* ^{e04061} and *mod(mdg4)*^{u1} mutant backgrounds. All three genes are long genes (approximately, *mbl* is 110kb, *Sdc* is 90kb and *Treh* is 15kb) and are decorated with intragenic insulator sites (Bushey et al., 2009; Negre et al., 2010). The *Drosophila mbl* gene encodes an alternative splicing factor whose proper function is required for terminal muscle and neural differentiation (Goers et al., 2008; Vicente-Crespo et al., 2008). In contrast to human and mice, which have three muscleblind-like homologs (MBNL 1-3), the *Drosophila* genome contains only one *mbl* gene, encoding four different protein isoforms (*mbl* A-D) generated by alternative splicing (Fernandez-Costa et al., 2011; Holt et al., 2009). Recent studies show that expression of the four isoforms is developmentally regulated and that they are not functionally redundant (Vicente et al., 2007). *Syndecan* belongs to a conserved family of type-I transmembrane proteins, which functions in various biological processes including lipid metabolism and regulation of growth factor pathways (Rapraeger, 2002; Williams, 2001). In *Drosophila*, Syndecan has been shown to play an important role in the regulation of the Slit signaling pathway during muscle pattern formation. Axons and muscle fibers are found to cross the ventral midline of the embryo after loss of *Sdc* function, a typical phenotype induced by failure of Slit signaling (Johnson et al., 2004; Steigemann et al., 2004).

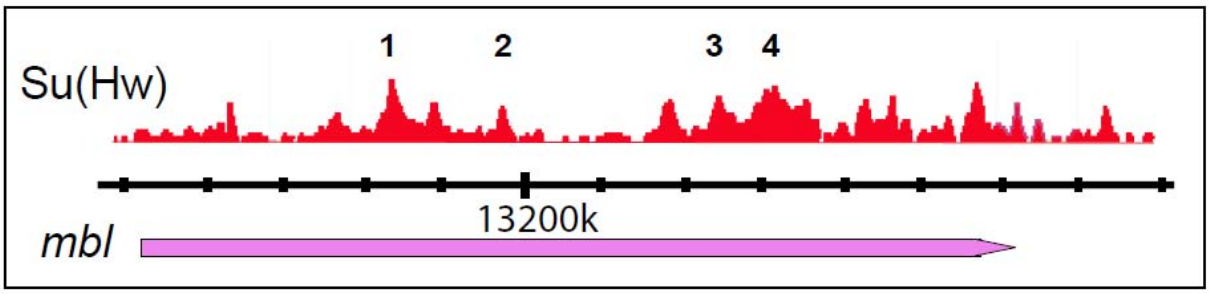
Before testing the effects of Su(Hw) insulators on gene expression, we first

examined the binding of Su(Hw) protein to the targeting sequences *in vivo*. We have picked four out of fourteen Su(Hw) binding sites determined in the chip-chip assay and performed Chromatin Immunoprecipitation assays (Figure 3.1). Chromatin was isolated from 12-16 hour old *Drosophila* embryos and immunoprecipitated with Su(Hw) antibody. The level of DNA enrichment after immunoprecipitation was determined by real-time PCR using primers flanking a subset of predictive Su(Hw) binding sites. Primers designed to target *gypsy* insulator and coding region of *Rp49* were used as positive and negative controls respectively. ChIP assay indicated that all of these sequences are positively bound by Su(Hw) proteins compared with the negative control even though they do not have the same affinity for Su(Hw) binding. The strength of Su(Hw) binding in our Chip assay agrees well with the previous Chip-chip data (Bushey et al., 2009). The $su(Hw)^{e04061}$ allele originated by the insertion of a *piggyBac* transposon in the second exon of *su(Hw)* (see materials and methods). *Su(Hw)* mRNA and protein levels are severely reduced in $su(Hw)^{e04061}$ homozygous flies, and *gypsy* induced mutations y^2 and ct^6 are completely rescued in the $su(Hw)^{e04061}$ background (Figure 3.2). The mutation $Mod\ (mdg4)^{U1}$ is caused by insertion of *Stalker* transposon into the exon unique to Mod(mdg4)-67.2 isoform, producing a truncated protein, which does not interact with Su(Hw) (Gause et al., 2001; Ghosh et al., 2001). The $(mdg4)^{U1}$ mutation has been shown to reduce

Figure 3.1 Chromatin immunoprecipitation assays, coupled to detection by real-time RT-PCR, confirmed binding of Su(HW) proteins to endogenous *gypsy* insulators within *mbl*.

ChIP assays were performed with chromatin isolated from 0-12 hour Drosophila embryos using antibody against Su(Hw) protein. The Y axis represents the amount of chromatin immunoprecipitated with Su(Hw) antibody plotted as a percentage of chromatin from total cell lysate. Primer pairs designed to target *gypsy* retrotransposon and RpL32 gene (Rp49) were used as positive and negative controls respectively. The results represent average of two independent experiments. Quantification of precipitated DNA as performed in triplicate for each chip ssay. Error bars represent the standard error of the mean. Insulator sites 1 to 4 correspond to the following genomic positions (in kb) 1: 13185476; 2: 13199029; 3: 13227557; 4: 13232041.


9
8
7
6
5
4
3
2
1
0
% input
Preimmune
IP
gypsy
Rp49
1
2
3
4

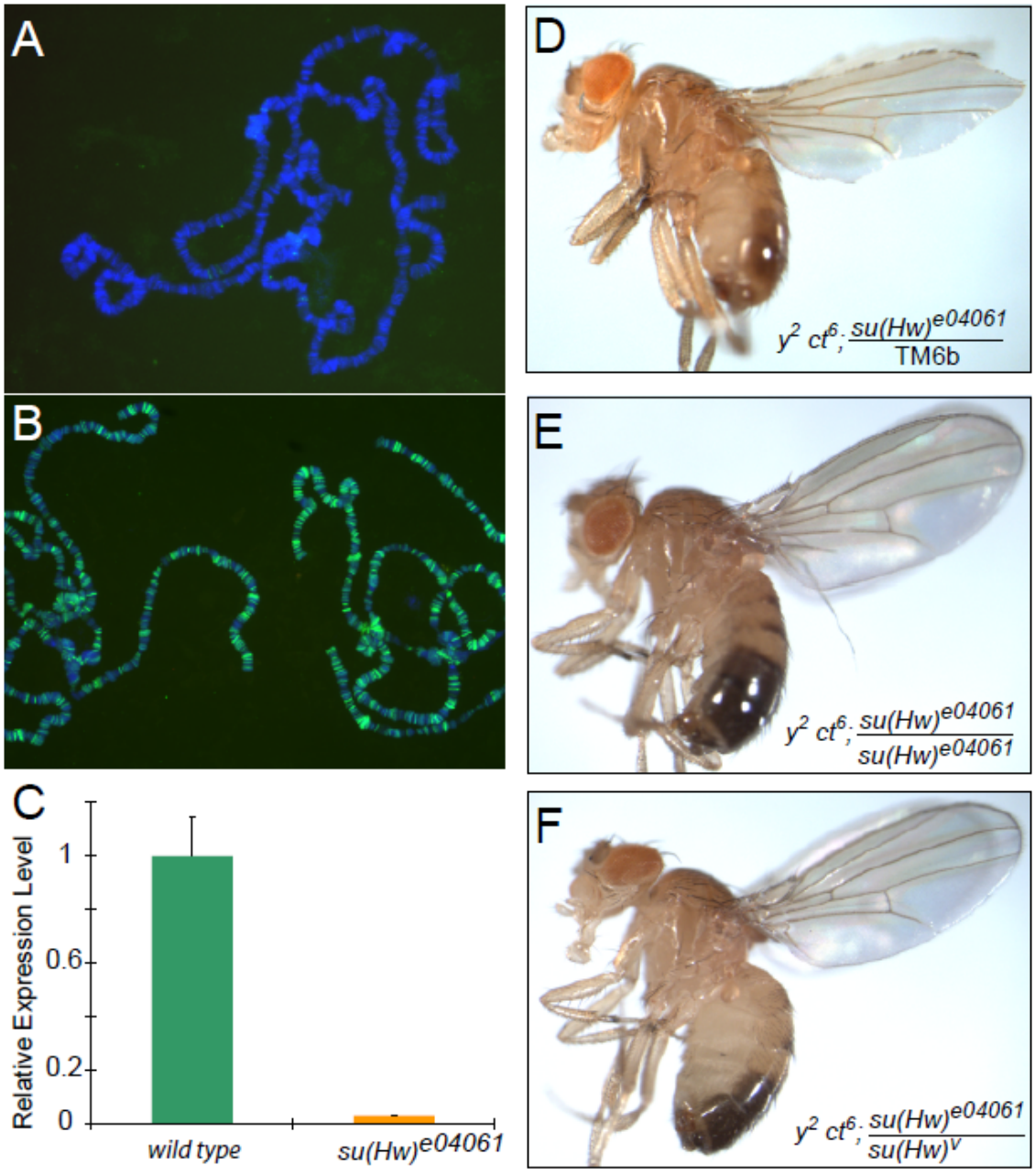

1
2
3
4
Su(Hw)
13200k
mbl

enhancer-blocking activities of Su(Hw) insulators (Georgiev and Kozycina, 1996) and induce bidirectional silencing in some other cases (Gdula and Corces, 1997). However, the mechanisms of Mod(mdg4) function in the Su(Hw) complex remains unclear. Since Mod(mdg4) 67.2 is required for proper function of Su(Hw) insulator, we also examined the binding of Su(Hw) to insulator sites in *Mod(mdg4)^{U1}* background. Immunostaining experiments using Su(Hw) antibody show that Su(Hw) proteins are capable of interacting with DNA in the absence of Mod(mdg4) 67.2 (Figure 3.3), reinforcing the idea that binding of Su(Hw) to DNA is not dependent on Mod(mdg4). Nevertheless, enrichment of Su(Hw) on the polytene chromosomes is dramatically decreased in the *Mod(mdg4)^{U1}* mutant, indicating that the presence of Mod(mdg4) 67.2 facilitates the interaction of Su(Hw) protein with its target sites.

We next extracted total RNA from both wild type third-instar larvae (*su(Hw)^{e04061}* / Tb *su(Hw)$^{+}$* and *mod(md4)^{u1} / Tb mod(mdg4)$^{+}$*) as well as *su(Hw)^{e04061}* and *mod(mdg4)^{u1}* homozygous mutant larvae. To ensure that we were analyzing mature mRNA, total RNA was reverse-transcribed into cDNA with oligo(dT) using the iScript select cDNA synthesis kit (Bio-Rad). Using real-time PCR, we analyzed the transcription levels of *mbl, Sdc and Treh* in the above mentioned genotypes. Enrichment of transcripts was normalized to ribosomal protein 49 (rp49) mRNA levels, which displays no change in *su(Hw)^{e04061}* or *mod(mdg4)^{u1}* mutations

Figure 3.2 *su(Hw)^{e04061}* is a null allele of *su(Hw)*

A-**B**. Immunostaining of polytene chromosomes using anti-Su(Hw) antibodies in *su(Hw)^{e04061}* homozygous (A) and in heterozygous (**B**) *su(Hw)^{e04061}* / TM6b larvae . **C**. mRNA expression levels of *su(Hw)* in wild-type larvae (*su(Hw)^{e04061}* / TM6b) compared with homozygous mutant larvae (*su(Hw)^{e04061}*) using quantitative real time PCR. D-E. The *su(Hw)^{e04061}* allele rescues the *gypsy* induced phenotypes of y^2 and ct^6. Heterozygous flies *su(Hw)^{e04061}* / TM6b carrying y^2 and ct^6 mutations show y^2 and ct^6 phenotypes (**D**), whereas homozygous *su(Hw)^{e04061}* (**E**) and heterozygous (**F**) *su(Hw)^{e04061}/ su(Hw)^{v}*, carrying y^2 and ct^6 mutations have wild-type phenotype.


A
B
C
Relative Expression Level
1
0.6
0.2
0
wild type
su(Hw)e04061
D
y2 ct6; su(Hw)e04061 / TM6b
E
y2 ct6; su(Hw)e04061 / su(Hw)e04061
F
y2 ct6; su(Hw)e04061 / su(Hw)v

compared with wild type (Figure 3.4). Primers used to target cDNAs were described in the materials and methods section. Analysis of real-time PCR data revealed that the transcriptional levels of *mbl* and *Sdc* significantly increased by more than 1.5-fold in the absence of Su(Hw) or Mod(mdg4) proteins, whereas the levels of *Treh* were unchanged (Figure 3.5). The most significant effect of the loss of *Su(Hw)* was found in *mbl*, which displayed a 2.53-fold increase in $Su(Hw)^{e04061}$ ($p = 0.0034$), and 2.2-fold in $mod(mdg4)^{u1}$ ($p = 0.0068$).

The increase in transcription levels of *mbl* and *Sdc* after loss of Su(Hw) protein suggests that intragenic Su(Hw) sites have a repressive effect on transcription. If this is true, over-expressing Su(Hw), should result in an even higher level of transcriptional repression. To further determine the function of intragenic Su(Hw) insulators, we analyzed expression of *mbl*, *Sdc* and *Treh* in Schneider S2 Drosophila tissue culture cells overexpressing Su(Hw). To overexpress Su(Hw), we transfected S2 cells with pBS-actSu(Hw)TAP-tag plasmid in which full length *su(Hw)* cDNA was placed under control of an Actin promoter. The plasmid was stably transfected into S2 cells, and expression of Su(Hw)TAP-tag plasmid was monitored by western blot and immunostaining using specific anti-Su(Hw) and anti-Calmodulin antibodies, which recognize Su(Hw) and the calmodulin peptide fused to Su(Hw) in the overexpressed Su(Hw)TAP-tag protein, respectively. Results show that the overexpressed

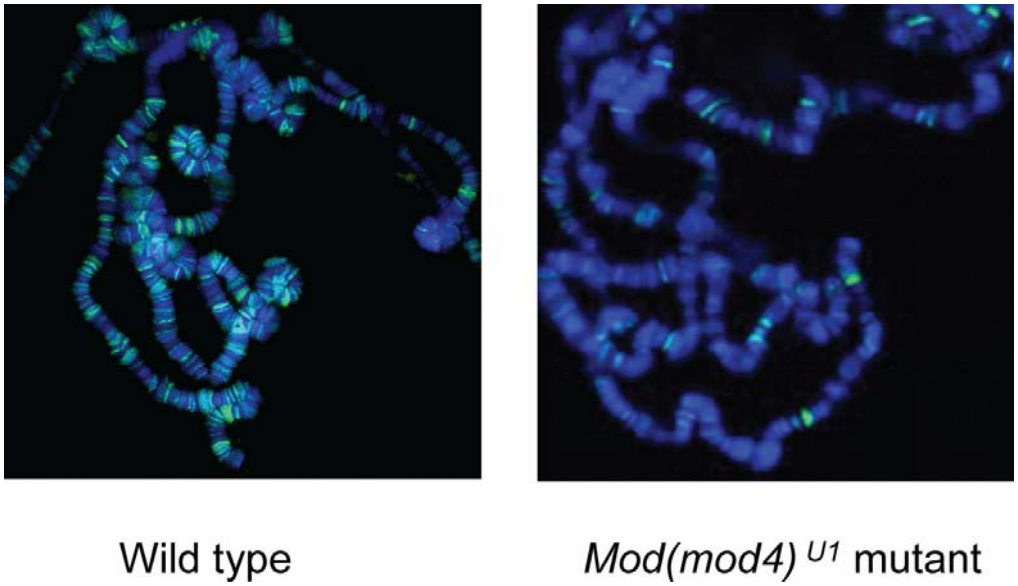
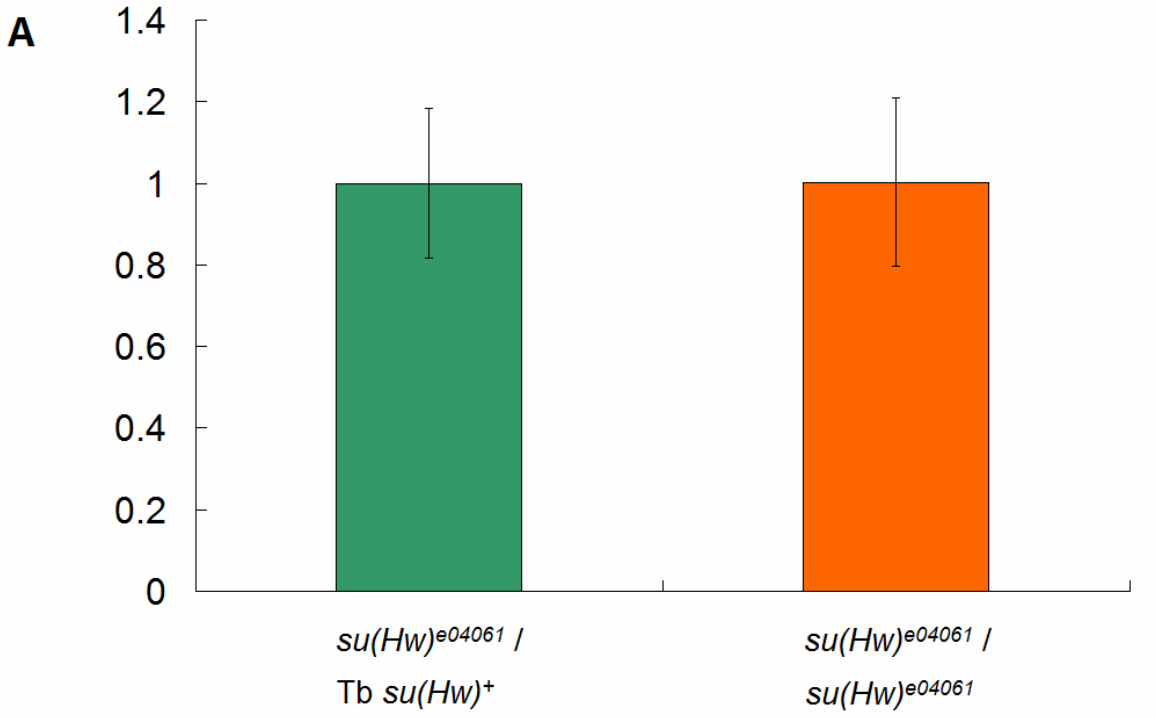


Figure 3.3 Presence of Mod(mdg4) 67.2 facilitates the interaction of Su(Hw) protein and its target sites on chromosomes.

Immunostaining was performed using polytene chromosomes dissected from heterozygous *mod(mdg4)*^{U1} / TM6b larvae (A) and *mod(mdg4)*^{U1} / *mod(mdg4)*^{U1} (B) larvae. Green color represents Su(Hw) binding sites. Blue color represents DAPI stained DNA.

Figure 3.4 Transcriptional level of rp49 is not changed in the absence of either Su(Hw) (A) or Mod (mdg4) (B) protein compared with wild type.

Total RNA was extracted from third-instar larvae. The concentration of RNA was measured with NanoDrop machine and 1μg of total RNA was reverse-transcribed into cDNA. Same amount of cDNA was used from each genotype. The results represent average of three independent experiments. Error bars represent the standard error of the mean.


A
1.4
1.2
1
0.8
0.6
0.4
0.2
0
su(Hw)e04061 /
Tb su(Hw)+
su(Hw)e04061 /
su(Hw)e04061

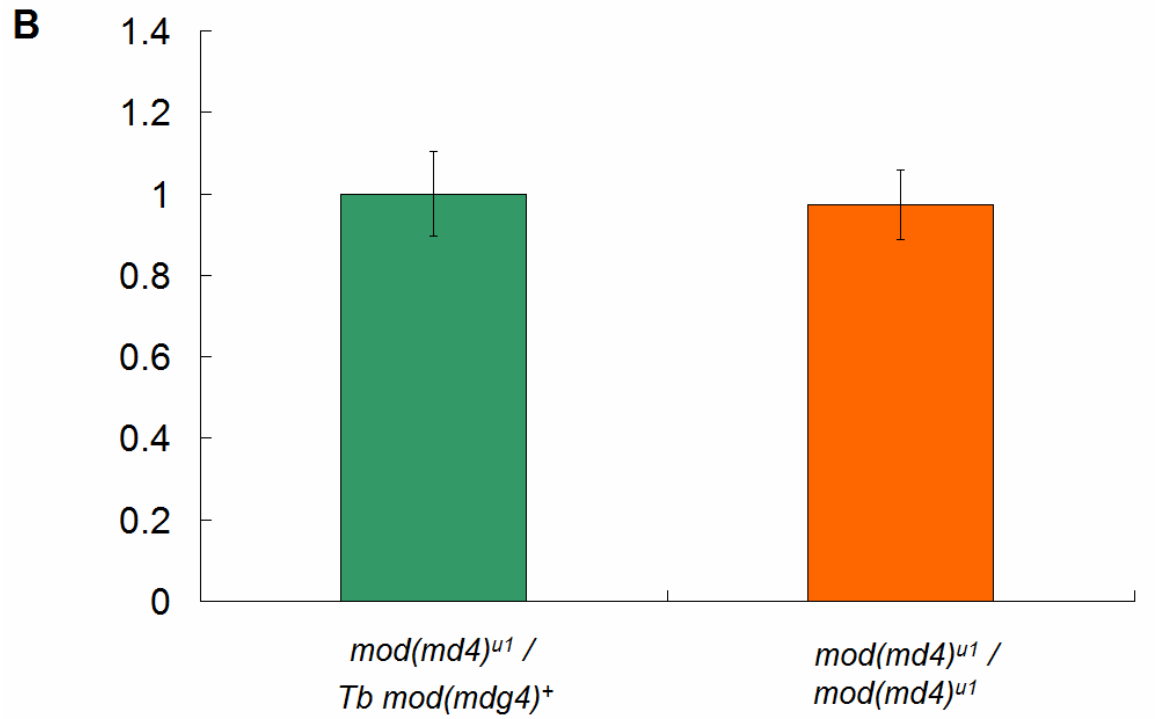
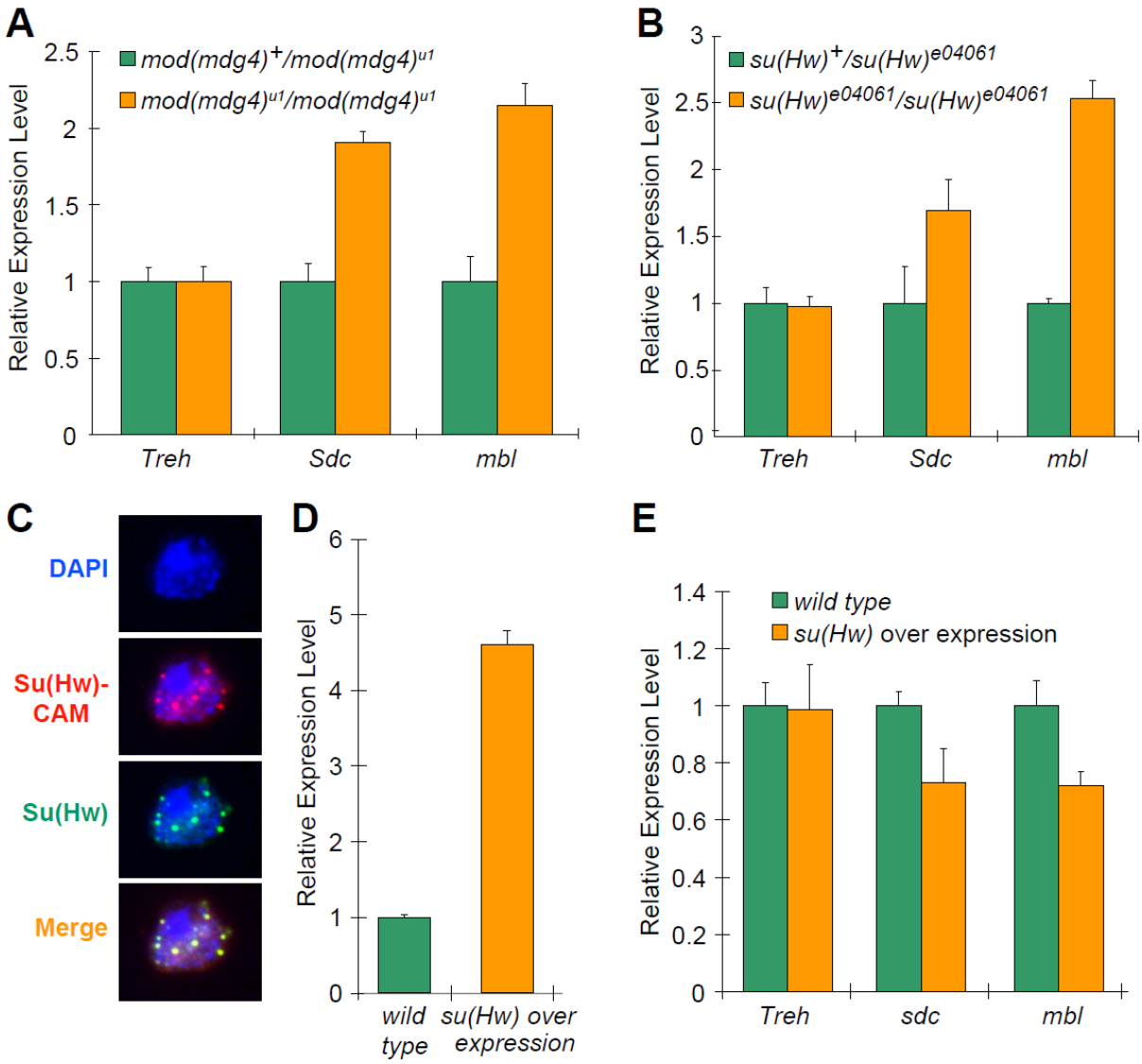

B
1.4
1.2
1
0.8
0.6
0.4
0.2
0
mod(md4)u1 /
Tb mod(mdg4)+
mod(md4)u1 /
mod(md4)u1

Figure 3.5 Loss of insulator proteins increases basal transcription levels of *mbl* and *Sdc*

Changes in transcriptional level of *Treh*, *Sdc* and *mbl* in third instar larval stage in *mod(mdg4)*^{u1} mutant (A) and *su(Hw)*^{e04061} mutant (**B**) are compared side by side. (**C**)-(**E**). Overexpression of Su(Hw) in Drosophila S2 cells suppress transcription of *mbl* and *Sdc*. Overexpressed Su(Hw) protein colocalizes with endogenous Su(Hw) foci in the nucleus (**C**). Immunostaining was performed using anti-Calmodulin (red) and anti-Su(Hw) (green) and DNA was stained with DAPI. Transcription levels of Su(Hw) increased by 4.6-fold in Su(Hw) Overexpression cell lines (**E**). Transcription level of *mbl* and *Sdc* was reduced by approximately 28 and 27 percent, respectively, due to Overexpression of Su(Hw). Note that no obvious change was detected in the expression of *Treh*


A
Relative Expression Level
2.5
2
1.5
1
0.5
0
mod(mdg4)+/mod(mdg4)u1
mod(mdg4)u1/mod(mdg4)u1
Treh
Sdc
mbl
B
Relative Expression Level
3
2.5
2
1.5
1
0.5
0
su(Hw)+/su(Hw)e04061
su(Hw)e04061/su(Hw)e04061
Treh
Sdc
mbl
C
DAPI
Su(Hw)-CAM
Su(Hw)
Merge
D
Relative Expression Level
6
5
4
3
2
1
0
wild type
su(Hw) over expression
E
Relative Expression Level
1.4
1.2
1
0.8
0.6
0.4
0.2
0
wild type
su(Hw) over expression
Treh
sdc
mbl

Su(Hw) protein colocalizes in the cell with endogenous Su(Hw) protein, forming distinctive foci, which are characteristic of insulator bodies (Figure 3.5 C). To test the expression levels of genes in cells overexpressing Su(Hw), RNA was purified from harvested cells and reverse-transcribed into cDNA. Real-time PCR analysis of mRNA levels indicated that overall transcription of *su(Hw)* was increased by more than four times in this cell line (Figure 3.5 D). As predicted, real time PCR data showed that expression of *mbl* and *Sdc* declined by 30% in *Su(Hw)* overexpressing cells, whereas expression of *Treh* did not change significantly (Figure 3.5 E). This result further supports the previous observation that Su(Hw) insulators suppress expression of *mbl* and *Sdc*, which form condensed chromatin bands and that possess endogenous insulators in their intragenic sequences. However, loss of insulator proteins did not influence basal expression of *Treh*, which unlike *mbl* and *Sdc*, is an inducible gene that is located in a decondensed interband region of the chromosome (Wallace et al., 2010). Taken together, these data suggest that intragenic Su(Hw) insulators might play a repressive role in regulating basal gene expression in *mbl* and *Sdc* and that this repressive effect might be mediated by higher order organization of chromatin into condensed chomatin.

In addition to basal mRNA expression levels, intragenic Su(Hw) insulators also reduce active transcription levels of *mbl* and *Sdc*

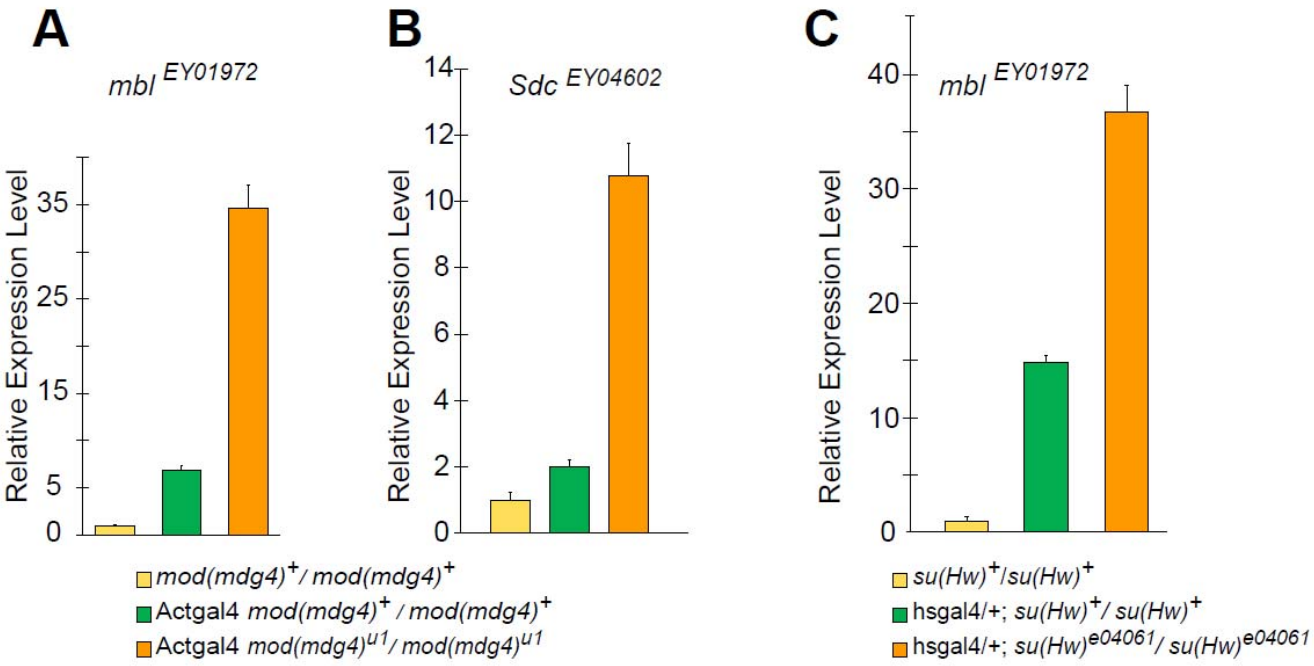
To further confirm that intragenic insulators directly influence active transcription levels in *mbl* and *Sdc*, we took advantage of two existing EP lines that carry insertions of a P element in the promoter region of both *mbl* and *Sdc*, which allow ectopic activation of these genes using the UAS-GAL4 binary system (Bellen et al., 2004) . Transgenic lines $mbl^{EY01972}$ and $Sdc^{EY04602}$ carry a P element with an upstream activating sequence (UAS) followed by an hsp70 promoter inserted to the 5' end of *mbl* and *Sdc*, respectively (Berkeley Drosophila Genome Project (BDGP)). Overexpression of *mbl* and *Sdc* was separately induced by crossing each transgenic line with a GAL4 expression line carrying an Act5C-GAL4 transgene (actin-gal4), which drives expression of GAL4 ubiquitously throughout all developmental stages. We first compared the change in *mbl* and *Sdc* transcriptional levels after GAL4 ectopic activation in wild type with that in the $mod(mdg4)^{u1}$ mutant background. Figure 3.6 A shows that transcription activation by GAL4 induces an approximately 7-fold ($p<0.01$) increase in transcription of *mbl* in wild type. Loss of Mod(mdg4)67.2 protein renders *mbl* more sensitive to GAL4 activation, given that a 35-fold ($p<0.01$) increase in transcription levels was observed in $mod(mdg4)^{u1}$ compared to wild type (Figure 3.6 A). A similar result was obtained for *Sdc*, in which an approximately 5-fold ($p<0.01$) increase in

transcription levels was observed in *mod(mdg4)*^{u1} compared to wild type (Figure 3.6 B). Differences in normal transcriptional levels in *mbl* compared to *Sdc* may explain why there is a 7-fold increase observed in the transcription of *mbl* after ectopic activation, and only a slight increase for *Sdc* transcription after ectopic activation in wild type. The normal transcription level of *Sdc* is in fact much higher than *mbl* (Figure 3.7), making *Sdc* less sensitive to GAL4 activation than *mbl.* However, the loss of Mod(mdg4)67.2 induced an approximately 5-fold ($p<0.01$) increase in the expression level of both *mbl* and *Sdc* after GAL4 activation, suggesting that Mod(mdg4) 67.2 plays a repressive role also in active transcription of both genes.

We next examined expression levels of *mbl* and *Sdc* in a *su(Hw)*^{e04061} homozygous mutant background in combination with ectopic activation of *mbl*. Here, we used a heat-shock GAL4 (hsgal4) activator to drive expression of *mbl* specifically in the third instar larval stage. Heat-shock was carried out at 37℃ for 30 minutes followed by one hour recovery. Real-time PCR data indicates that transcription of *mbl* induced by hsgal4 is more efficient than the actin-gal4 driver, since hsgal4 activation in *su(Hw)* wild type larvae increases *mbl* transcription by 15-fold ($p<0.01$), compared with only a 7-fold ($p<0.01$) increase obtained by actin-gal4 activation (Figure 3.6 A compared to 3.6 C). In addition, results show that hsgal4 activation in homozygous *su(Hw)*^{e04061} is only 2.5-fold ($p<0.01$) that of

Figure 3.6 Loss of insulator proteins increases active transcription levels of *mbl* and *Sdc*

A. Transcriptional level of *mbl* in wild type (yellow), after ectopic activation in wild type (green), and after ectopic activation in *Mod(mdg4)*^{u1} mutant background (orange). **B**. Transcriptional level of *Sdc* in wild type (yellow), after ectopic activation in wild type (green) and after ectopic activation in *Mod(mdg4)*^{u1} mutant background (orange). **C**. Transcriptional level of *mbl* in wild type (yellow), after ectopic activation in wild type (green) and after ectopic activation in *su(Hw)*^{e04061} mutant background (orange). Transcriptional level of *mbl* and *Sdc* in wild type was set as 1.


A
mbl EY01972
Relative Expression Level
0
5
15
25
35
B
Sdc EY04602
Relative Expression Level
0
2
4
6
8
10
12
14
C
mbl EY01972
Relative Expression Level
0
10
20
30
40
mod(mdg4)+/ mod(mdg4)+
Actgal4 mod(mdg4)+ / mod(mdg4)+
Actgal4 mod(mdg4)u1/ mod(mdg4)u1
su(Hw)+/su(Hw)+
hsgal4/+; su(Hw)+/ su(Hw)+
hsgal4/+; su(Hw)e04061/ su(Hw)e04061

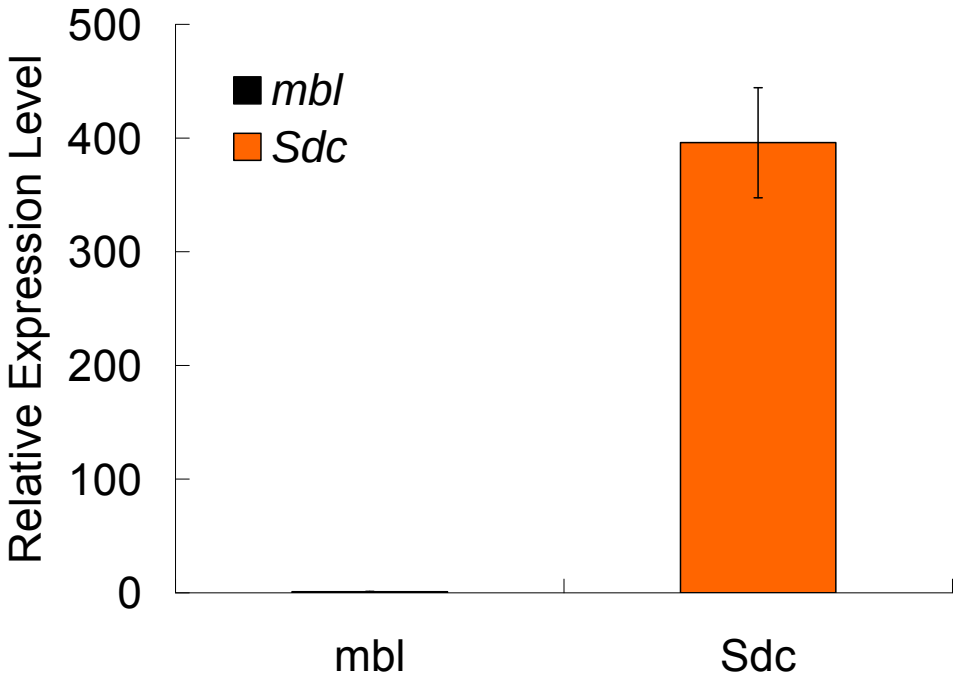


Figure 3.7 Expression level of Sdc is around 400-fold higher than that of *mbl* in wild type

Total RNA was extracted from third-instar larvae and reverse-transcribed into cDNA. Transcriptional level of each gene was normalized to that of *rp49*. The relative transcriptional level of *mbl* after correction was arbitrarily defined to be 1. The results represent average of three independent experiments. Error bars represent the standard error of the mean.

wild type (Figure 3.6 C), whereas the *mod(mdg4)*^{u1} mutation induces a 7-fold ($p<0.01$) increase in transcription of *mbl* after its activation (Figure 3.6 A).Therefore, even though the fold increase in *su(Hw)*^{e04061} compared to wild type is different from the fold increase detected in activated *mbl* in a *mod(mdg4)*^{u1} background, the overall level of transcription in activated *mbl*, compared to non-activated *mbl*, is around 35-fold ($p<0.01$) in both cases. Together, these results suggest that intragenic insulator function reduces both basal transcription and transcriptional activation in *mbl* and *Sdc* by a yet unknown mechanism that is likely independent of the transcriptional activation mechanism.

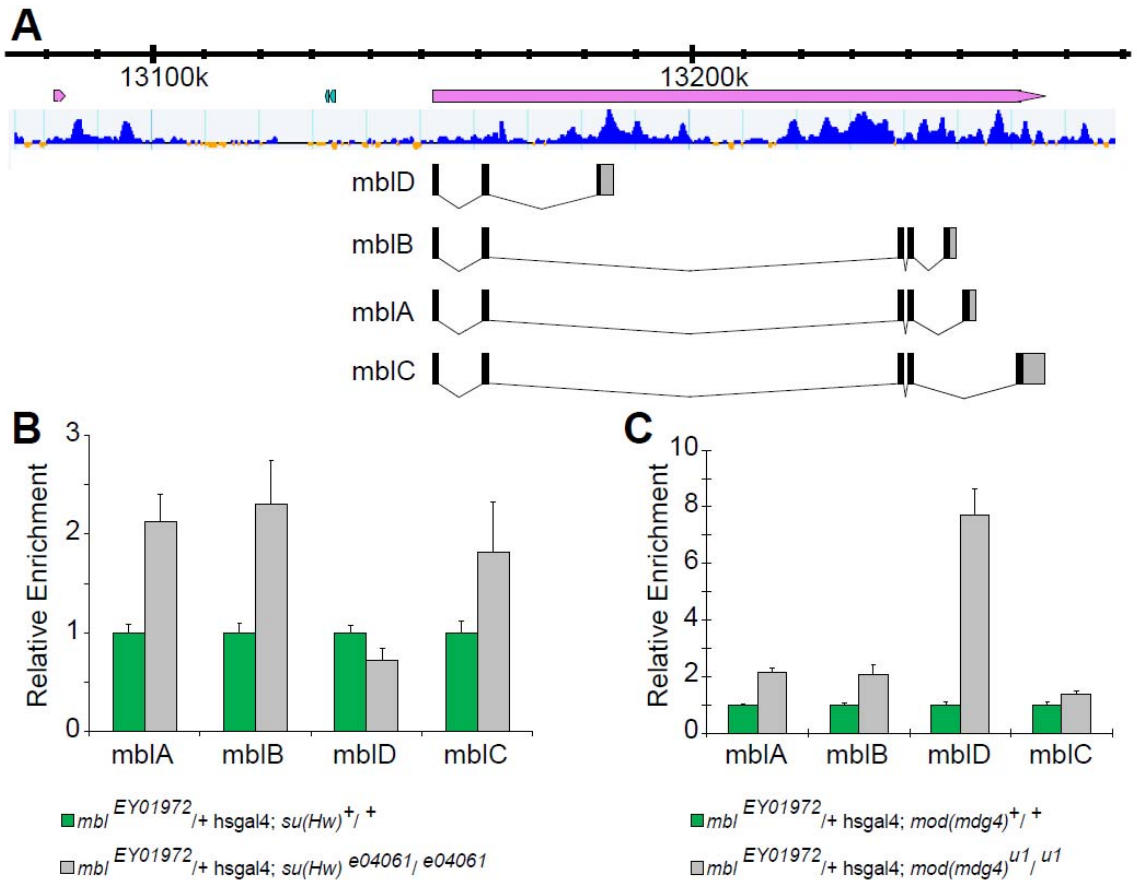
Loss of insulator proteins leads to changes in *mbl* RNA processing in *Drosophila*

In addition to transcriptional levels, intragenic insulators may influence mRNA processing in a manner similar to that of other chromatin proteins found downstream of Pol II start sites (Wada et al., 2009). The *mbl* gene encodes at least four protein isoforms (*mblA, mblB, mblC and mblD*) generated by alternative splicing (Figure 3.8 A), each with a distinct function (Begemann et al., 1997). We asked whether the relative frequency of these *mbl* splice variants increased equally in *su(Hw)*^{e04061} and *mod(mdg4)*^{u1} mutant backgrounds. To answer this question, transcription of *mbl*^{EY01972} was ectopically activated as mentioned earlier,

and the abundance of transcripts corresponding to each specific isoform was determined using isoform-specific primers in *su(Hw)*^{e04061} and *mod(mdg4)*^{u1} mutants, as well as in wild type (see primers in the materials and methods sections). In order to focus on the role of Su(Hw) and Mod(mdg4) 67.2 proteins, we arbitrarily defined the enrichment of each transcript after ectopic activation of *mbl*^{EY01972} in the presence of insulator proteins to be 1, and compared transcript changes caused by loss of Su(Hw) and Mod(mdg4)67.2 proteins. Data show that loss of Mod(mdg4)67.2 protein led to an overall increase in the enrichment of all *mbl* isoforms (Figure 3.8 C). Specifically, *mblA*, *mblB* and *mblC* transcripts displayed 2.1 (p=0.017)-, 2.0 (p=0.057)- and 1.4 (p=0.3)-fold increases, respectively, in the absence of the Mod(mdg4)67.2 protein. Interestingly, results also show that the abundance of *mbl*D increased by almost 8-fold ($p<0.01$) due to loss of Mod(mdg4)67.2, which is significantly more than fold of increase detected for all other *mbl* isoforms. In *su(Hw)*^{e04061}, enrichment of *mblA*, *mblB* and *mblC* transcripts displayed a pattern very similar to that observed in *mod(mdg4)*^{u1} (Figure 3.8 B compared to 3.8 C). However, a different pattern was observed for *mbl*D, for which a slight decrease was observed in *su(Hw)*^{e04061} mutant compared with wild type (Figure 3.8 B). Taken together, these results show that the ratio of *mblD* to other *mbl* isoforms was increased significantly in *mod(mdg4)*^{u1} but decreased in *su(Hw)*^{e04061}. Interestingly, we noticed that the exon unique to *mbl*D contains a

Figure 3.8 Loss of insulator proteins changes the ratio between *mbl* isoforms after induction of transcription at the *mbl* promoter

A. Schematic representation of the *mbl* gene showing its genomic position and four splicing variants. Su(Hw) binding sites within this region are shown as blue peaks based on modENCODE ChIP-on-chip data. **B**. Relative enrichment of *mbl* isoforms after ectopic activation in wild type (green bars) and in *su(Hw)*^{e04061} mutant (gray bars). **C**. Relative enrichment of *mbl* isoforms after ectopic activation in wild type (green bars) and in *mod(mdg4)*^{u1} background (gray bars).


A
13100k
13200k
mblD
mblB
mblA
mblC
B
Relative Enrichment
0
1
2
3
mblA
mblB
mblD
mblC
mbl EY01972/+ hsgal4; su(Hw)+/ +
mbl EY01972/+ hsgal4; su(Hw) e04061/ e04061
C
Relative Enrichment
0
2
4
6
8
10
mblA
mblB
mblD
mblC
mbl EY01972/+ hsgal4; mod(mdg4)+/ +
mbl EY01972/+ hsgal4; mod(mdg4)u1/ u1

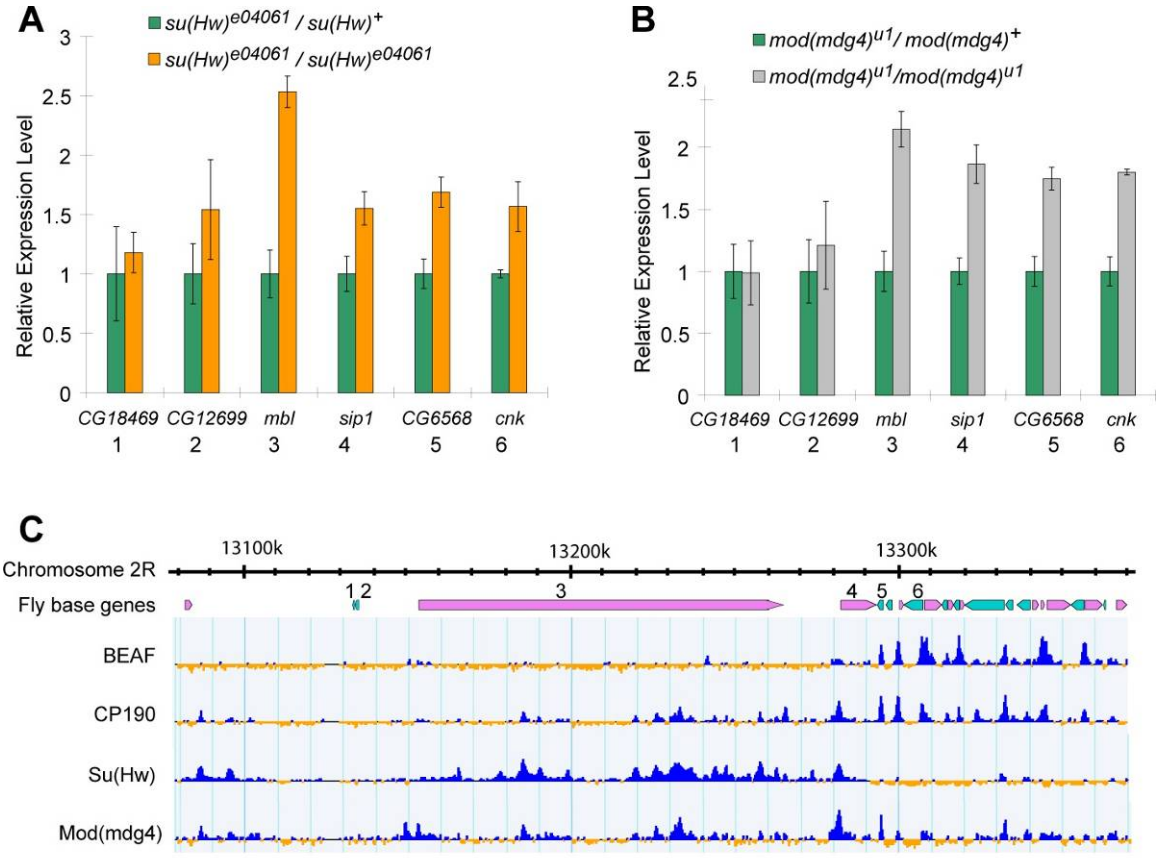
strong Su(Hw) binding site close to the alternative termination site necessary to produce *mblD*. Such sites are absent in all other isoform-specific exons, suggesting that endogenous *gypsy* insulators may play a role in regulating the alternative splicing or termination sites in *Drosophila mbl* pre-mRNA (Figure 3.8 A).

Su(Hw) insulators may help define chromatin domains of gene expression

The current paradigm on insulator function, predicts that insulator disruption should lead to alterations in transcription patterns along contiguous chromatin domains due to malfunction of their boundaries. A detailed look at the distribution of chromatin insulators at the *mbl* locus shows that *mbl*, as well as a large region upstream of the *mbl* promoter, contains a low abundance of BEAF and CP190 insulator proteins, whereas it is very rich in Su(Hw) and Mod(mdg4) proteins (Figure 3.9 C). Conversely, the relative abundance of insulator proteins is inverted downstream of the *mbl* transcription termination site. Downstream of *mbl*, the last significant peak of Su(Hw) appears upstream of the promoter of the *Sip1* gene (Figure 3.9 C). Downstream of *Sip1*, a large number of genes associate to abundant BEAF and CP190 proteins, without a significant presence of Su(Hw) proteins (Figure 3.9 C). This can also be observed in polytene chromosomes, where CP190 bands are found clearly flanking the outer side of the Su(Hw) signals that flank the *mbl* locus (see figure 3.17D). This distribution of insulator

Figure 3.9 Real-time RT-PCR analysis of expression of *mbl* and neighboring genes in third-instar larvae in (A) Su(Hw) mutant and (B) *Mod(mdg4)^{U1}* mutant compared with wild type

Transcriptional level of each gene in wild type was set as 1. (C). Schematic representation of mbl locus on the gene map as well as distribution of four different insulator proteins in this region. The number above each gene indicates the genes analyzed in real-time RT PCR. From let to right, 1:*CG18469*, 2:*CG12699*, 3: *mbl*, 4: *Sip1* and 5: *cnk.*


A
su(Hw)e04061 / su(Hw)+
su(Hw)e04061 / su(Hw)e04061
Relative Expression Level
3
2.5
2
1.5
1
0.5
0
CG18469 CG12699 mbl sip1 CG6568 cnk
1 2 3 4 5 6
B
mod(mdg4)u1/ mod(mdg4)+
mod(mdg4)u1/mod(mdg4)u1
Relative Expression Level
2.5
2
1.5
1
0.5
0
CG18469 CG12699 mbl sip1 CG6568 cnk
1 2 3 4 5 6
C
Chromosome 2R
13100k
13200k
13300k
Fly base genes
1 2
3
4 5 6
BEAF
CP190
Su(Hw)
Mod(mdg4)

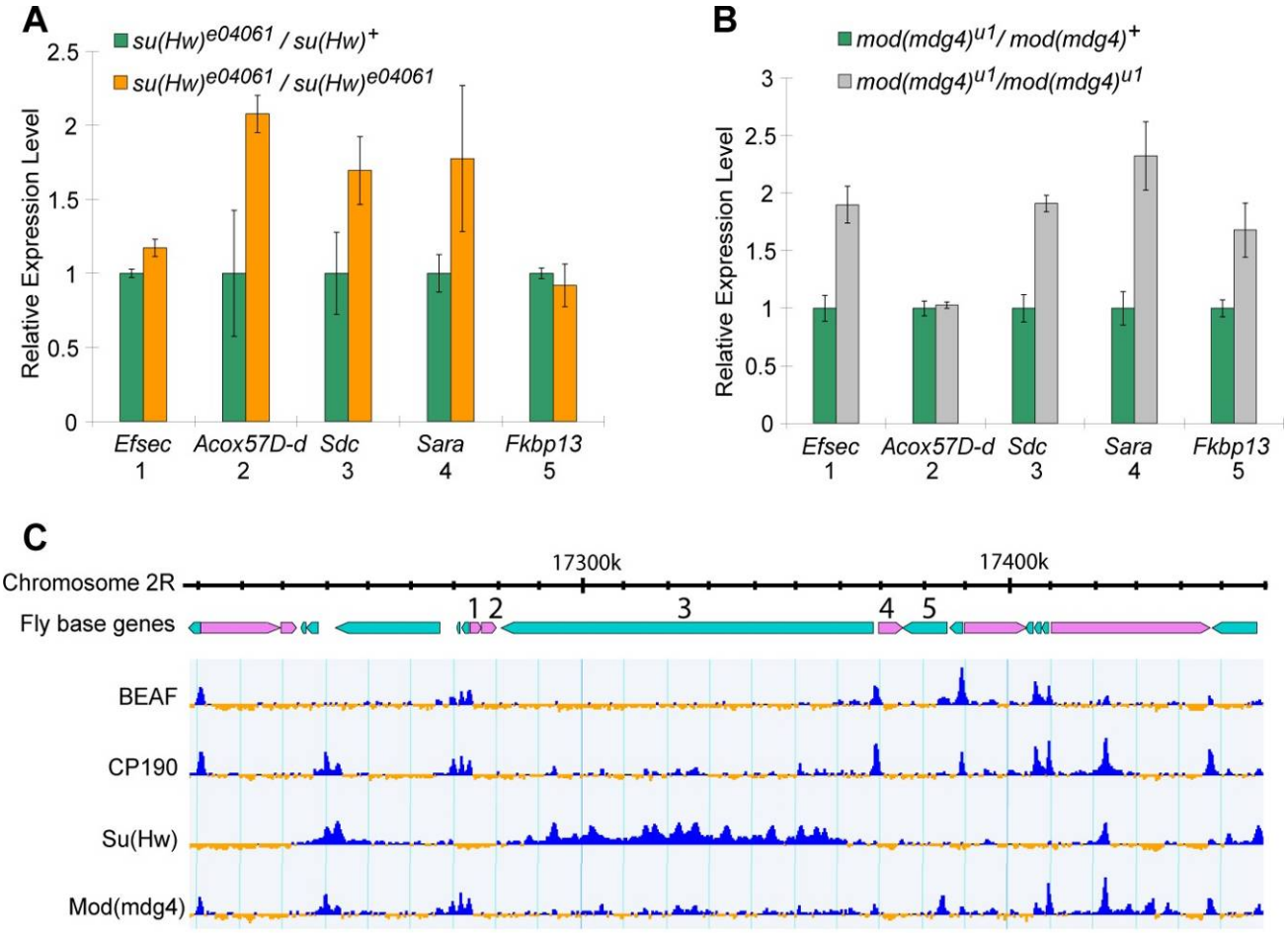
proteins suggests that Su(Hw) insulator sequences in *mbl* and upstream *mbl* define a chromatin domain that is differentiated from the adjacent chromatin domain downstream the *Sip1* promoter, which is characterized by high levels of BEAF and Cp190.

We previously suggested that this type of discontinuity in the distribution of insulator proteins could represent the landmarks for distinct chromatin domains (Wallace et al., 2010). Here we have tested whether transcription of genes adjacent to *mbl*, and lacking Su(Hw) insulator sites, would be also influenced by mutations in insulator proteins, likely resulting from changes in boundary function. Results indicate that genes upstream the *mbl* promoter do not show a significant change in transcription levels in the genetic background of *mod(mg4)*^{u1} or *su(Hw)*^{e04061} homozygous mutants (Figure 3.9 A-B). However, genes downstream *mbl*, including *Sip1,* appear to undergo significant changes of approximately 2-fold in their basal transcription levels, which parallel changes observed in *mbl* (Figure 3.9 A-B). Since none of these genes, with the exception of *Sip1*, have a significant presence of Su(Hw) binding sites, these results suggest that derrepression of transcription after loss of insulator proteins may be a consequence of changes in the genome architecture mediated by Su(Hw) activity.

We next asked whether changes in transcriptional activity of neighboring genes

Figure 3.10 Real-time RT-PCR analysis of expression of *Sdc* and neighboring genes in whole larvae in (A) Su(Hw) mutant and (B) *Mod(mdg4)*^{U1} mutant compared with wild type

Transcriptional level of each gene in wild type was set as 1. (C) Schematic representation of *mbl* locus on the gene map as well as distribution of four different insulator proteins in this region. The number above each gene indicates the genes analyzed in real-time RT PCR. From let to right, 1: *Efsec*, 2: *Acox57D-d*, 3: *sdc*, 4: *Sara* and 5: *Fkbp13.*


A
su(Hw)e04061 / su(Hw)+
su(Hw)e04061 / su(Hw)e04061
Relative Expression Level
2.5
2
1.5
1
0.5
0
Efsec 1
Acox57D-d 2
Sdc 3
Sara 4
Fkbp13 5
B
mod(mdg4)u1/ mod(mdg4)+
mod(mdg4)u1/mod(mdg4)u1
Relative Expression Level
3
2.5
2
1.5
1
0.5
0
Efsec 1
Acox57D-d 2
Sdc 3
Sara 4
Fkbp13 5
C
Chromosome 2R
17300k
17400k
Fly base genes
1 2
3
4 5
BEAF
CP190
Su(Hw)
Mod(mdg4)

could also be observed in a chromosome region where differences in the proposed landmarks of chromatin domains are not as well defined as in *mbl*. To this end, we analyzed the expression of genes neighboring *Sdc*, in which insulator proteins are distributed in a more uniform manner along the chromatin fiber (Figure 3.10 C). Result show that changes in transcriptional activity are still significant in genes both downstream and upstream of *Sdc* (Figure 3.10 A and 3.10 B). Results show that changes in genes adjacent to *Sdc* are more unpredictable and do not respond equally to mutations in *su(Hw)* or *mod(mdg4)*, as they did in *mbl*. For example, genes *Efsec1* and *Fkbp13* are not significantly activated after mutations in *su(Hw)*, but transcription increased almost 2-fold in the background of *mod(mdg4)*^{u1}, whereas the opposite is true for the *Acox57D-d* gene (Figure 3.10 A and 3.10 B). The molecular mechanism responsible for changes in transcription levels of genes neighboring *mdl* and *Sdc* is still unknown and it cannot be ruled out that they respond to pleiotropic effects induced by changes in genes elsewhere in the genome. However, taken together these results support the notion that chromatin insulators function by creating chromatin domains and that disruption of insulator function may change levels of transcription of genes within the domains.

Su (Hw) insulators regulate gene expression differently in brain and Salivary gland tissues

Comparison of gene expression profiles in *Su(Hw)* ^{e04061} and *mod(mdg4)*^{u1} mutants background with the wild type in whole larvae provides an overview of insulator function in gene regulation. However, regulation of many genes occurs in a tissue-specific manner, and analyzing gene expression in whole larvae may not reveal effects of the genetic background. To test whether Su(Hw) insulators function differently in distinct *Drosophila* tissues, we performed real-time RT PCR using RNA extracted from brain and salivary glands. Real time PCR results show that transcription of *mbl* increased in both tissues in the absence of Su(Hw) proteins compared with the wild type (Figure 3.11 and Figure 3.12), which is consistent with the results observed in the whole larvae (Figure 3.9 A-B). However, the change in *mbl* transcription appears to be more significant in salivary glands than that is in brain (Figure 3.11 and 3.12). These differences are probably due to the lower transcriptional level of *mbl* in salivary glands in wild type compared with that of the brain (data not shown). This observation is also true for the gene *CG18469*, whose transcription was obviously increased more in brain than in salivary glands. Interestingly, transcription of the gene CG12699 was reduced in brain while slightly increased in salivary glands. These observations suggest that *Su(Hw)* affects transcription of these genes differently in different tissues.

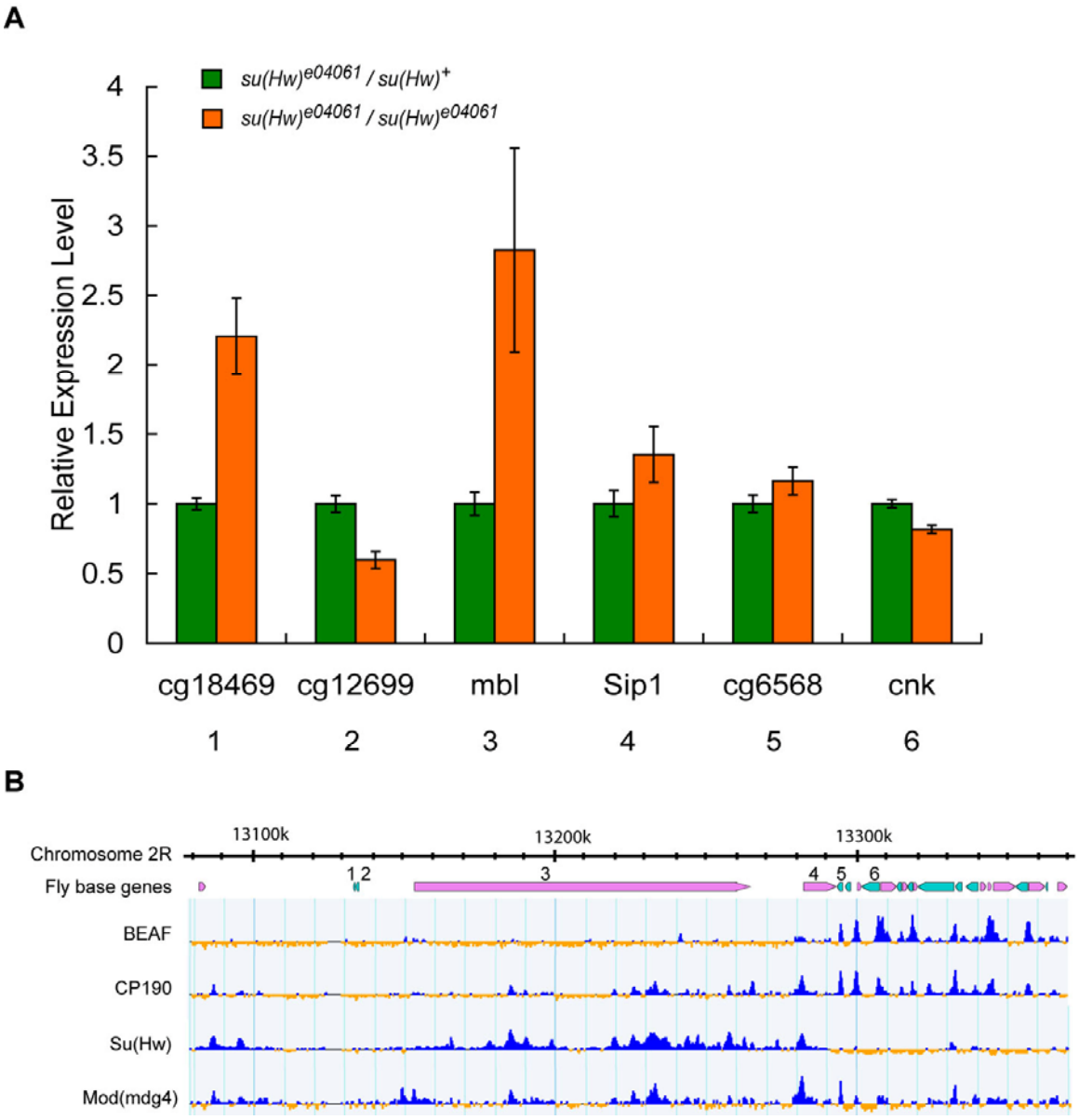


Figure 3.11 Real-time RT-PCR analysis of expression of *mbl* and neighboring genes in brain in *Su(Hw)*^{e04061} mutant background compared with wild type

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *mbl* locus

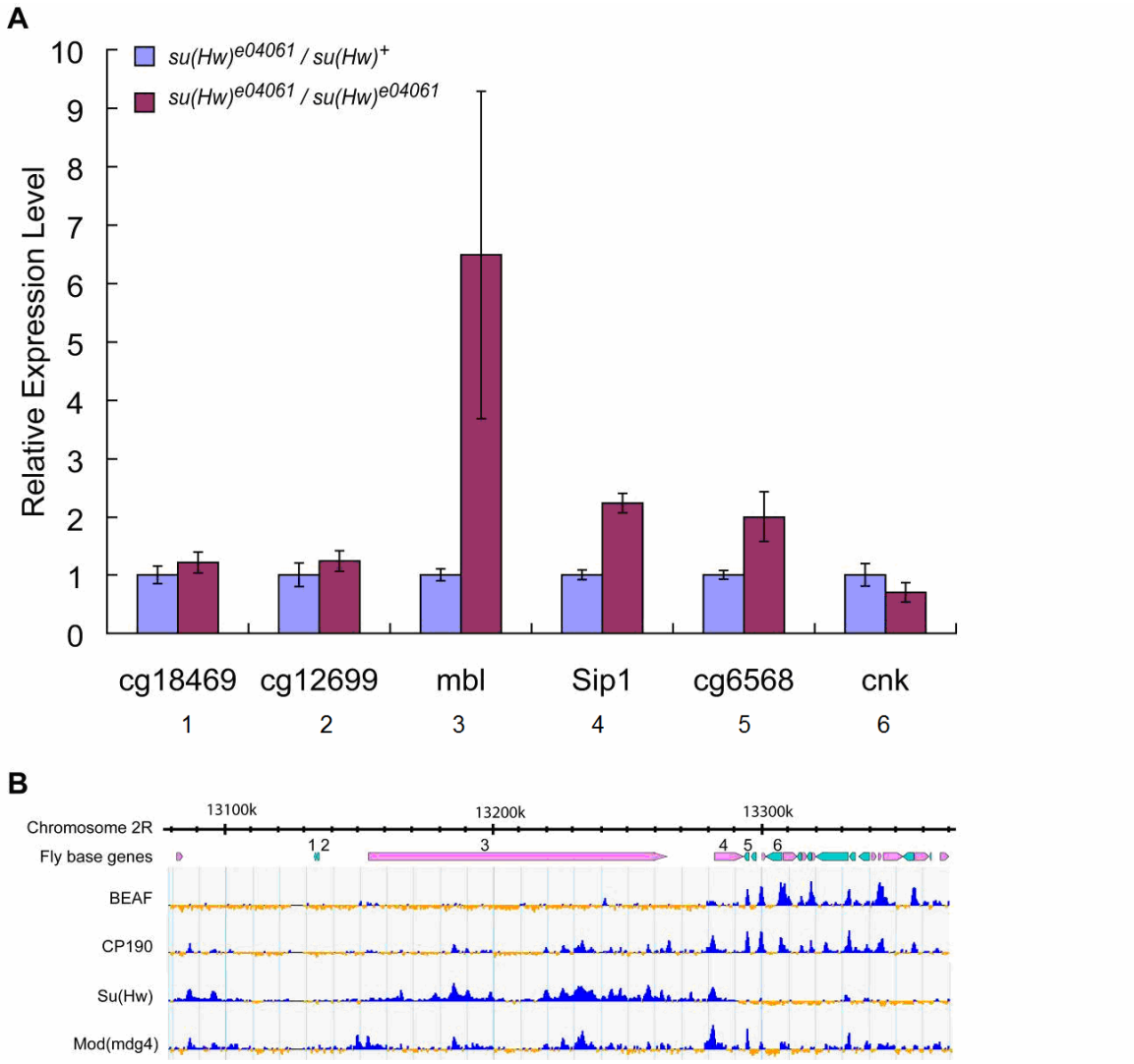


Figure 3 .12 Real-time RT-PCR analysis of expression of *mbl* and neighboring genes in salivary glands in *Su(Hw)^{e04061}* mutant compared with wild type.

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *mbl* locus

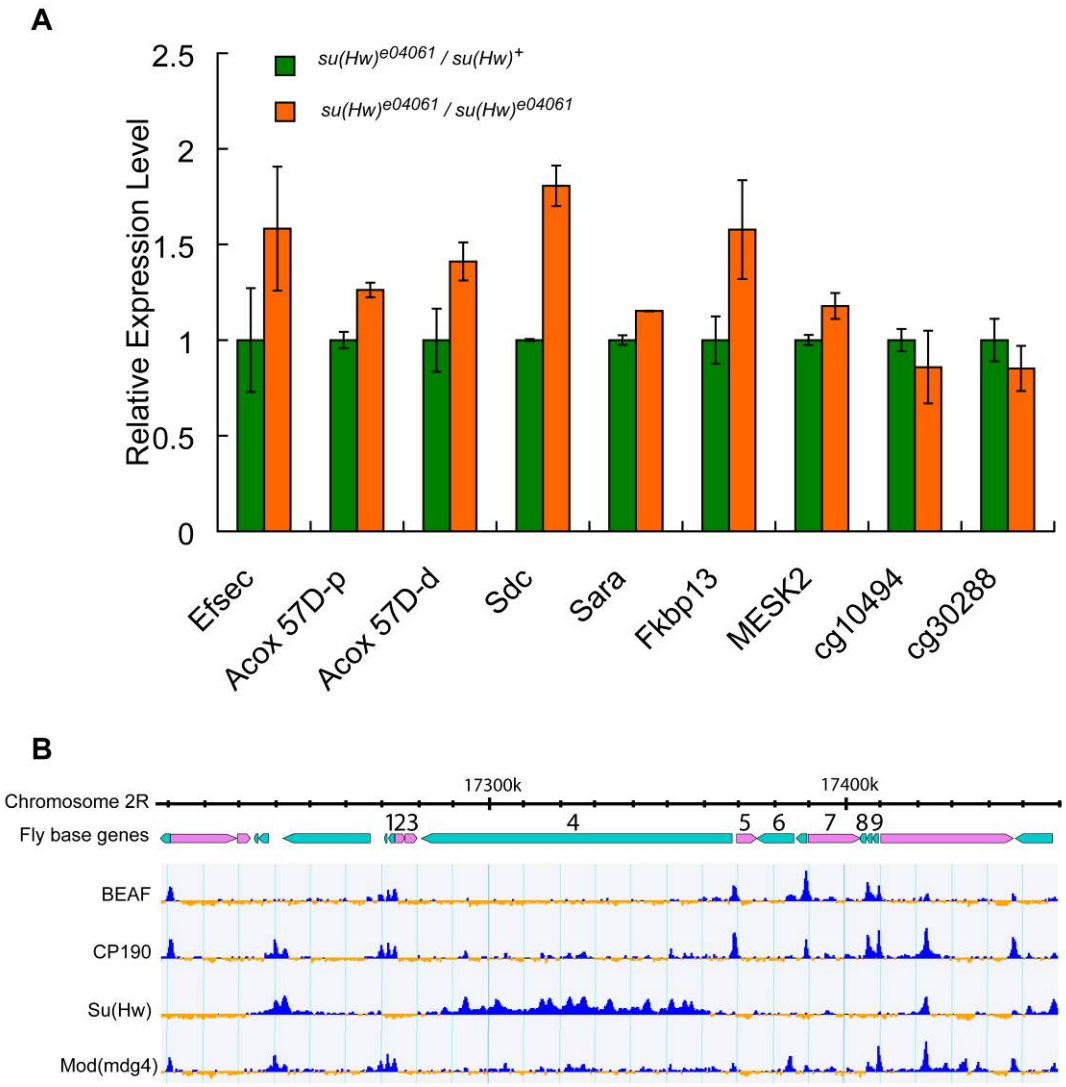


Figure 3. 13 Real-time RT-PCR analysis of expression of *Sdc* and neighboring genes in brain in *Su(Hw)^{e04061}* mutant background compared with wild type.

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *Sdc* locus

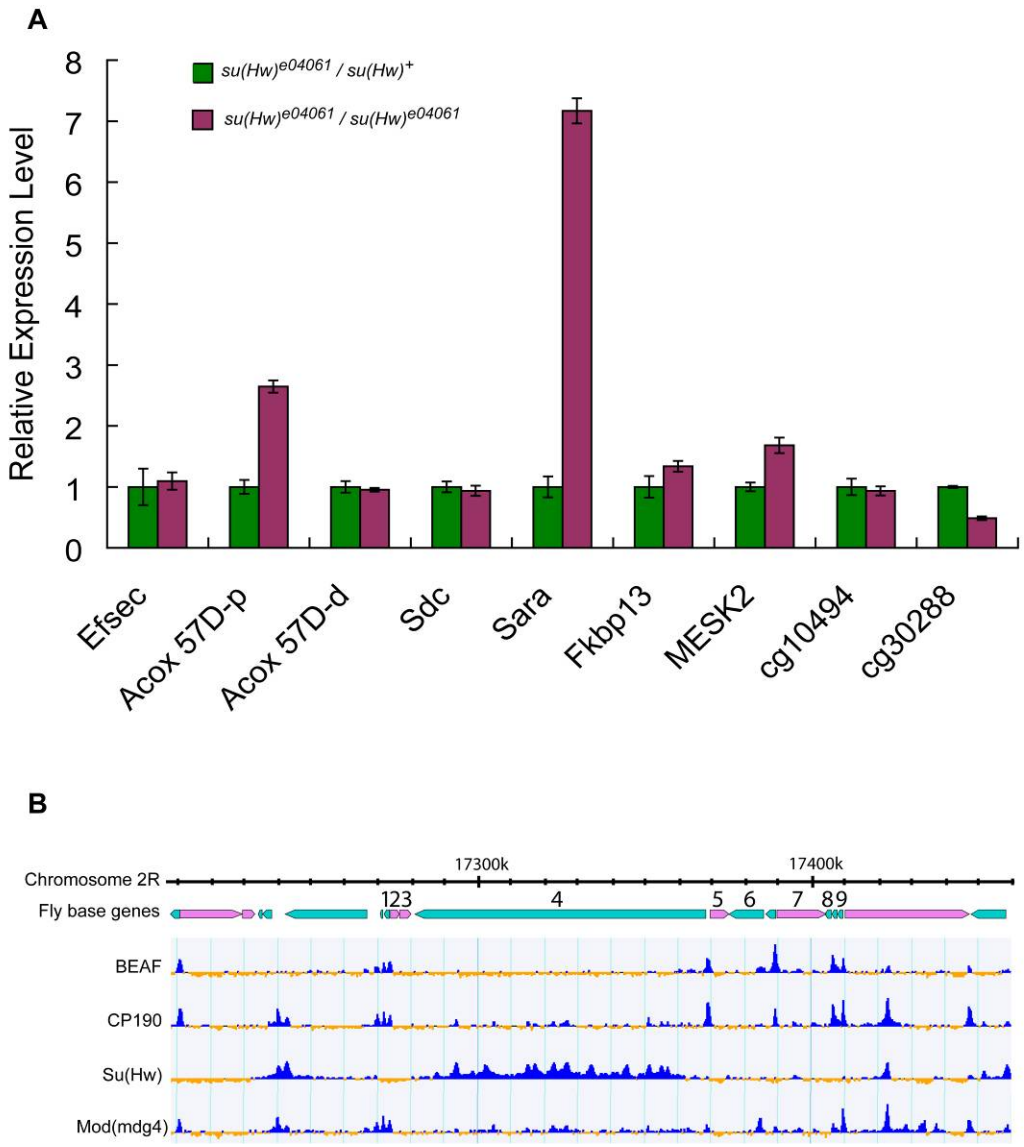


Figure 3. 14 Real-time RT-PCR analysis of expression of *Sdc* and neighboring genes in salivary glands in *Su(Hw)^{e04061}* mutant background compared with wild type.

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *Sdc* locus

However, in striking contrast to genes upstream *mbl*, genes located downstream of *mbl* displayed a very similar expression pattern in brain and salivary glands in the absence of Su(Hw) proteins, reinforcing the idea that insulator proteins may define a chromatin domain in *mbl,* and that the region comprising *mbl* and the genes upstream of *mbl* is differentiated from the region downstream the termination site of *mbl* (Figure 3.11 B).

We next performed the same analysis with genes neighboring *Sdc*. Results show that in brain, transcription of most of the genes adjacent to *Sdc* displayed less than 2-fold increase with the exception of two predicted genes *(CG10494* and *CG30288),* which showed a slight decrease in the *Su(Hw)* ^{e04061} mutant background compared with wild type (Figure 3.13). Again, a different expression pattern was obtained in salivary glands. The most apparent difference between the above two tissues is observed in the expression of the gene *Sara,* which displayed a 7-fold (p=0.032) increase in salivary glands and only slight increase in brain due to absence of Su(Hw) (Figure 3.13 and 3.14). Transcription of the gene *Acox57D-d* was also increased 2.7- fold (p=0.033) in salivary glands, twice as much as the increase observed in brain. Although the mechanisms that cause these differences remains to be understood, these observation suggests that regulation of gene expression by Su(Hw) insulators occurs in a tissue and gene-specific manner.

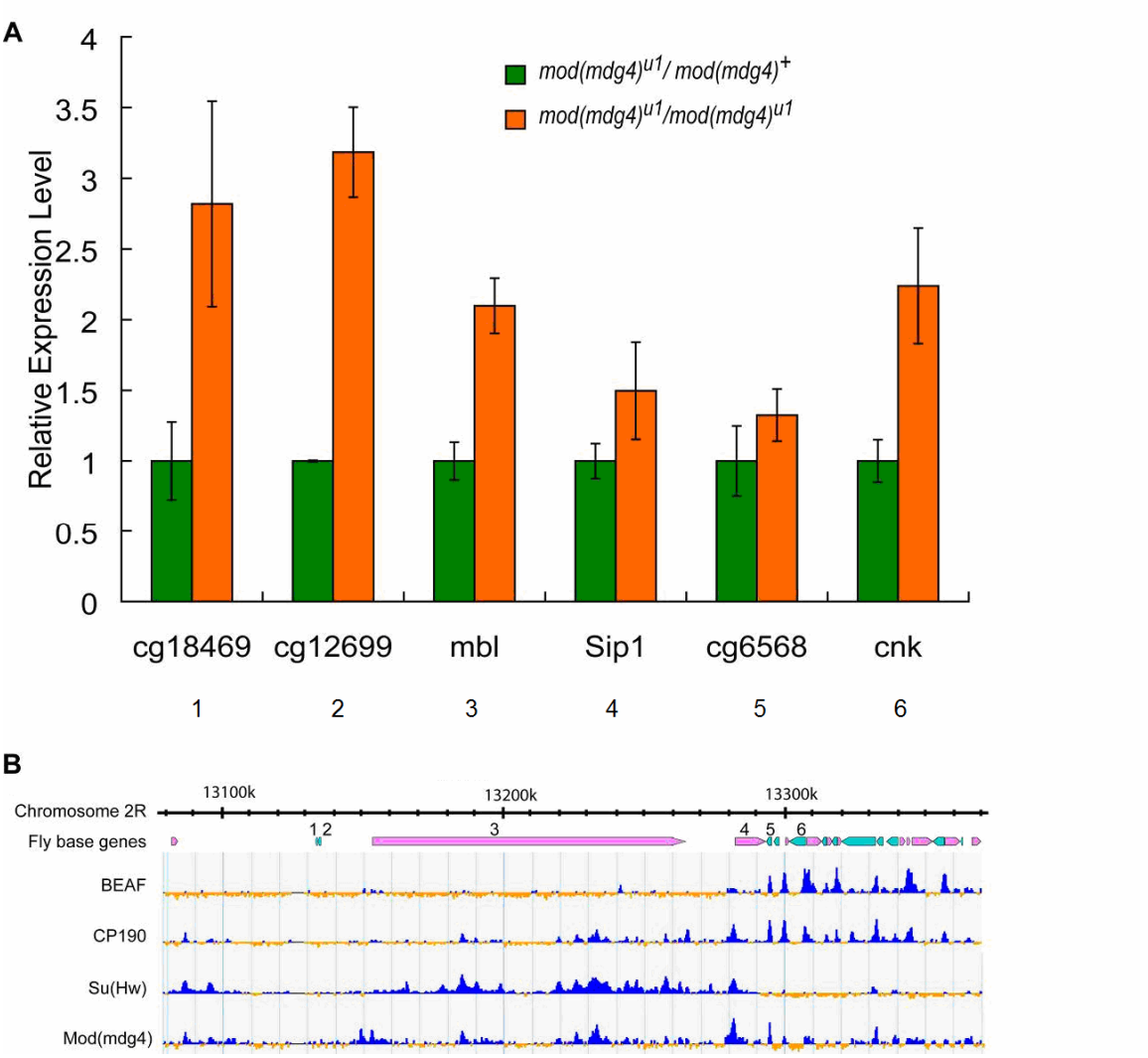


Figure 3. 15 Real-time RT-PCR analysis of expression of *mbl* and neighboring genes in brain in *mod(mdg4)^{U1}* mutant background compared with wild type.

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *mbl* locus

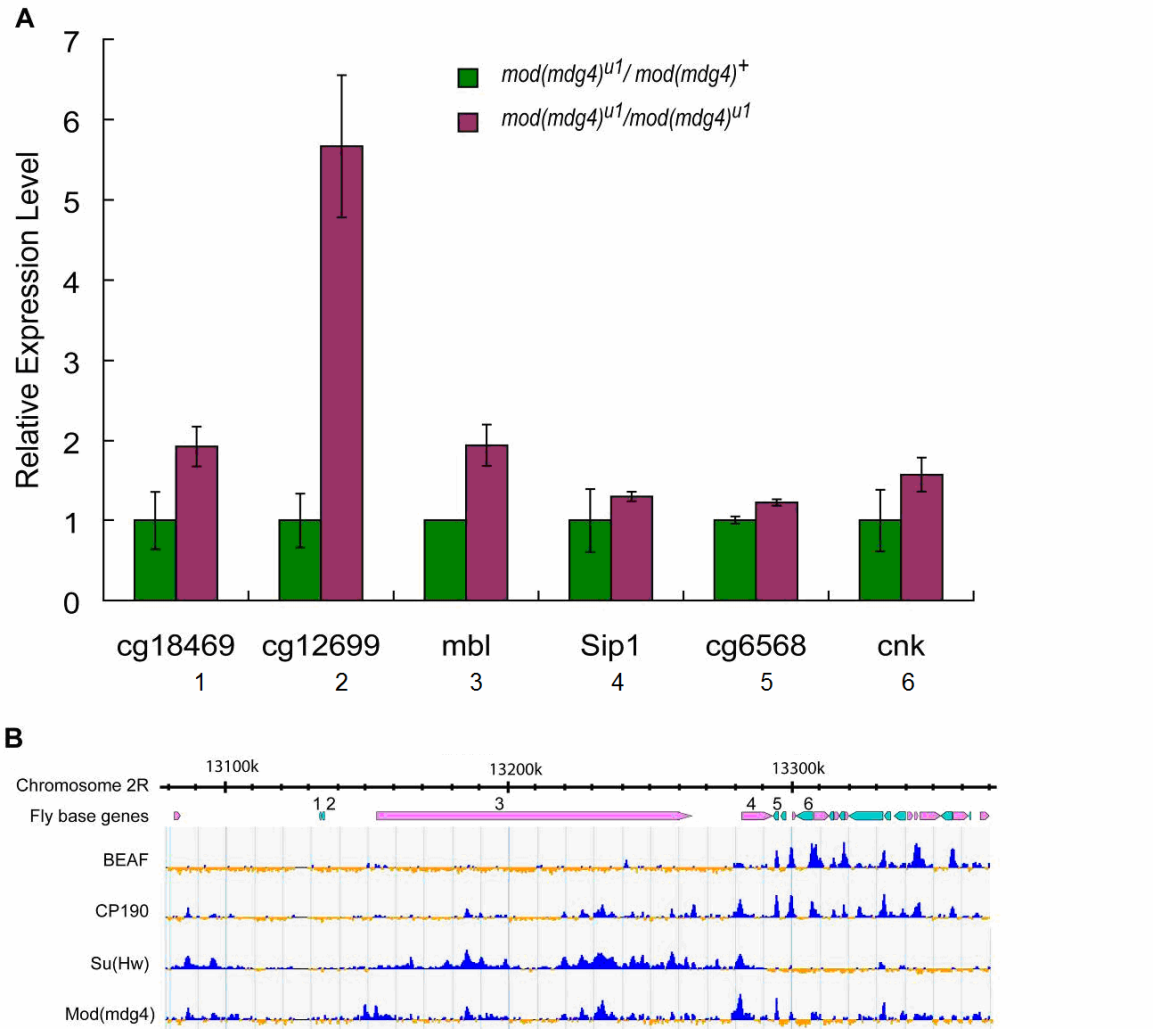


Figure 3. 16 Real-time RT-PCR analysis of expression of *mbl* and neighboring genes in salivary glands in *mod(mdg4)^{U1}* mutant background compared with wild type.

(A). Transcriptional level of each gene in wild type was set as 1. (B). Schematic representation of distribution of genes and insulator proteins near *mbl* locus

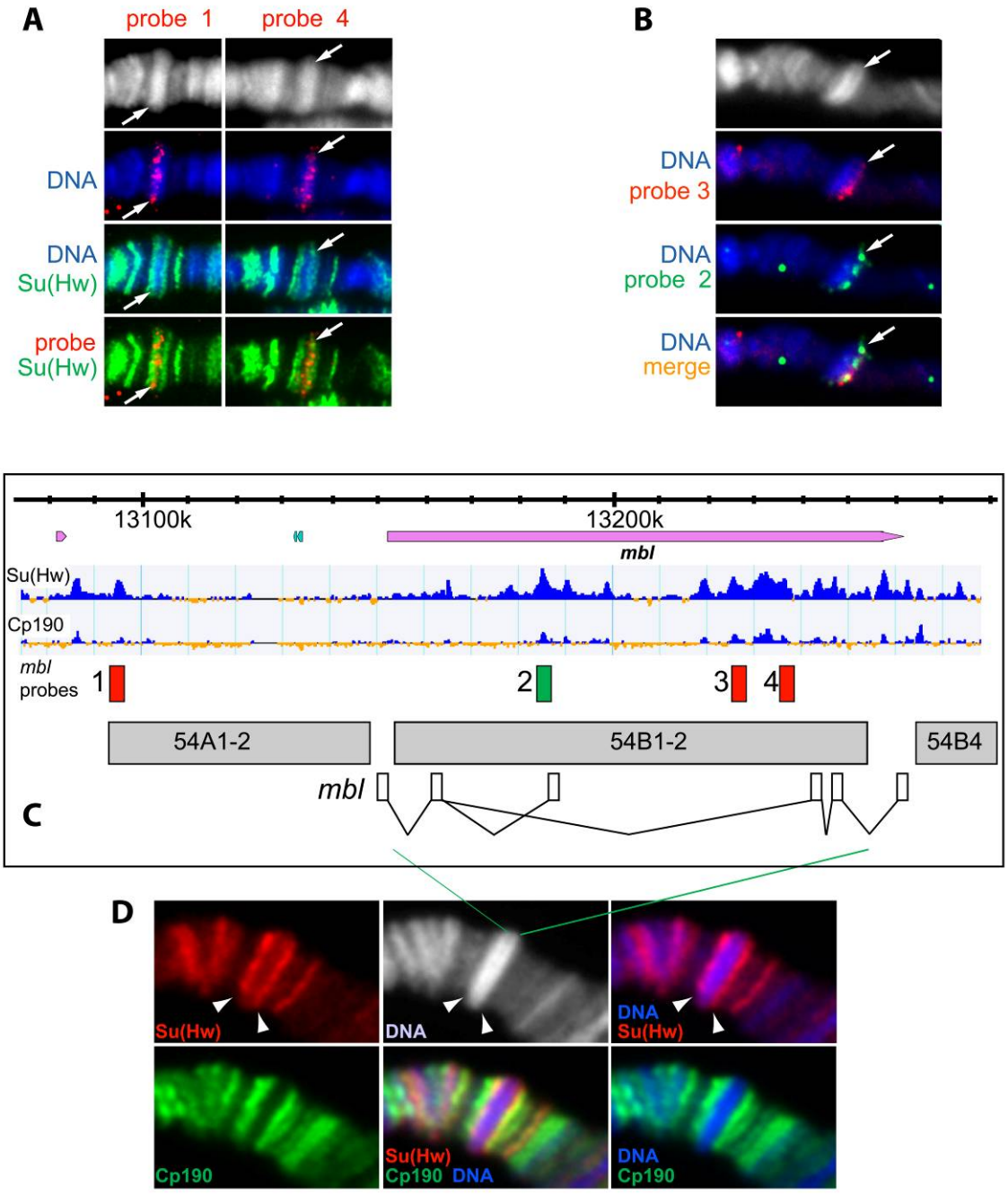
Since changes in gene expression due to mutations in *su(Hw)* and *mod(mdg4)* showed different patterns in the third-instar larvae, we next asked whether these differences are also reflected in specific tissues. Comparison of expression profiles of *mbl* and neighboring genes in *Su(Hw)* ^{e04061} and *mod(mdg4)*^{u1} mutant backgrounds in brain and salivary glands also suggest a different role of Su(Hw) and Mod(mdg4) proteins in the regulation of several genes. For example, in brain, transcription of *CG12699* and *cnk* was increased in *Su(Hw)*^{e04061} mutant background (Figure 3.15 A), and was reduced in *mod(mdg4)*^{u1} mutant background (Figure 3.11 A). In salivary glands, although transcriptional level of *mbl* gene was increased in both *Su(Hw)*^{e04061} and *mod(mdg4)*^{u1} mutants, mutation of *Su(Hw)* has an obviously more dramatic effect in expression of *mbl* gene than *mod(mdg4)*^{u1} mutation (compare figure 3.12 A and 3.16 A). The different effect of Su(Hw) and Mod(mdg4) 67.2 proteins on gene regulation may result from their different role in the insulator complex. Because Su(Hw) mediates the binding of insulator complex to DNA, the loss of Su(Hw) results in the dissociation of the entire insulator complex. On the other hand, loss of Mod(mdg4) 67.2 would likely only disrupt interactions between insulators, leaving other insulator proteins such as Su(Hw) and CP190 still bound to DNA. These observations also underline the complexity of Su(Hw) insulators in the regulation of gene expression.

The *mbl* locus forms a condensed chromatin structure associated to Su(Hw) intragenic insulators in polytene chromosomes

We next asked how chromatin is organized at the *mbl* locus. In previous work in our laboratory, it has been shown that *Sdc* intragenic insulators map to a highly condensed chromatin band in polytene chromosomes, and suggested that insulators could contribute to the organization of chromatin by their ability to establish long-range interactions along the chromatin fiber (Wallace et al., 2010). Here, we have focused our attention on the *mbl* locus, given that previous reports using electron microscopy (EM) and *in situ* hybridization have shown that most intragenic sequences of *mbl* lie within the large 54B1-2 band in polytene chromosome arm 2R (Semeshin et al., 1998). We used fluorescence in situ hybridization (FISH) combined with immunostaining to determine the position of *mbl* in relation to endogenous intragenic insulators (Figure 3.17). In order to identify the cytological location of these Su(Hw) insulator sequences, we designed FISH probes targeted to Su(Hw) binding sites within *mbl* as previously identified (Bushey et al., 2009; Negre et al., 2010). We first used two probes containing Su(Hw) binding sites that we predicted will flank the highly condensed chromatin associated with *mbl* (Figure 3.17 A probe1 and probe 4). FISH combined with immunostaining using antibodies against Su(Hw) revealed that both probes containing insulator sequences colocalized with insulator signals at the edge of

Figure 3.17 The *mbl* locus is organized into a highly condensed chromatin band

(A). Mapping of endogenous Su(Hw) insulators within *mbl* by FISH combined with immunostaining on polytene chromosomes of *Drosophila* third instar larvae. A Su(Hw) binding site within gene *mbl* (probe 4) localized within band 54B1-2, whereas a second binding site near gene *cg10950* (probe 1) localized to the edge of a highly condensed chromatin band 54A1-2. (B). FISH probes 2 and 3 map within *mbl* and contain Su(Hw) binding sites located more than 30kb apart show colocalization on polytene chromosomes. (C). Distribution of Su(Hw) and Cp190 proteins in the *mbl* locus, showing a schematic representation of *mbl*, with its position in the chromosome in kb as well as structure of the gene and the relative position of the FISH probes used in (A) and (B). (D). Immunostaining of polytene chromosomes showing the distribution of Su(Hw) and CP190 proteins near the *mbl* locus.


A
probe 1
probe 4
DNA
DNA
Su(Hw)
probe
Su(Hw)
B
DNA
probe 3
DNA
probe 2
DNA
merge
13100k
13200k
mbl
Su(Hw)
Cp190
mbl
probes
1
2
3
4
54A1-2
54B1-2
54B4
mbl
C
D
Su(Hw)
DNA
DNA
Su(Hw)
Cp190
Su(Hw)
Cp190 DNA
DNA
Cp190

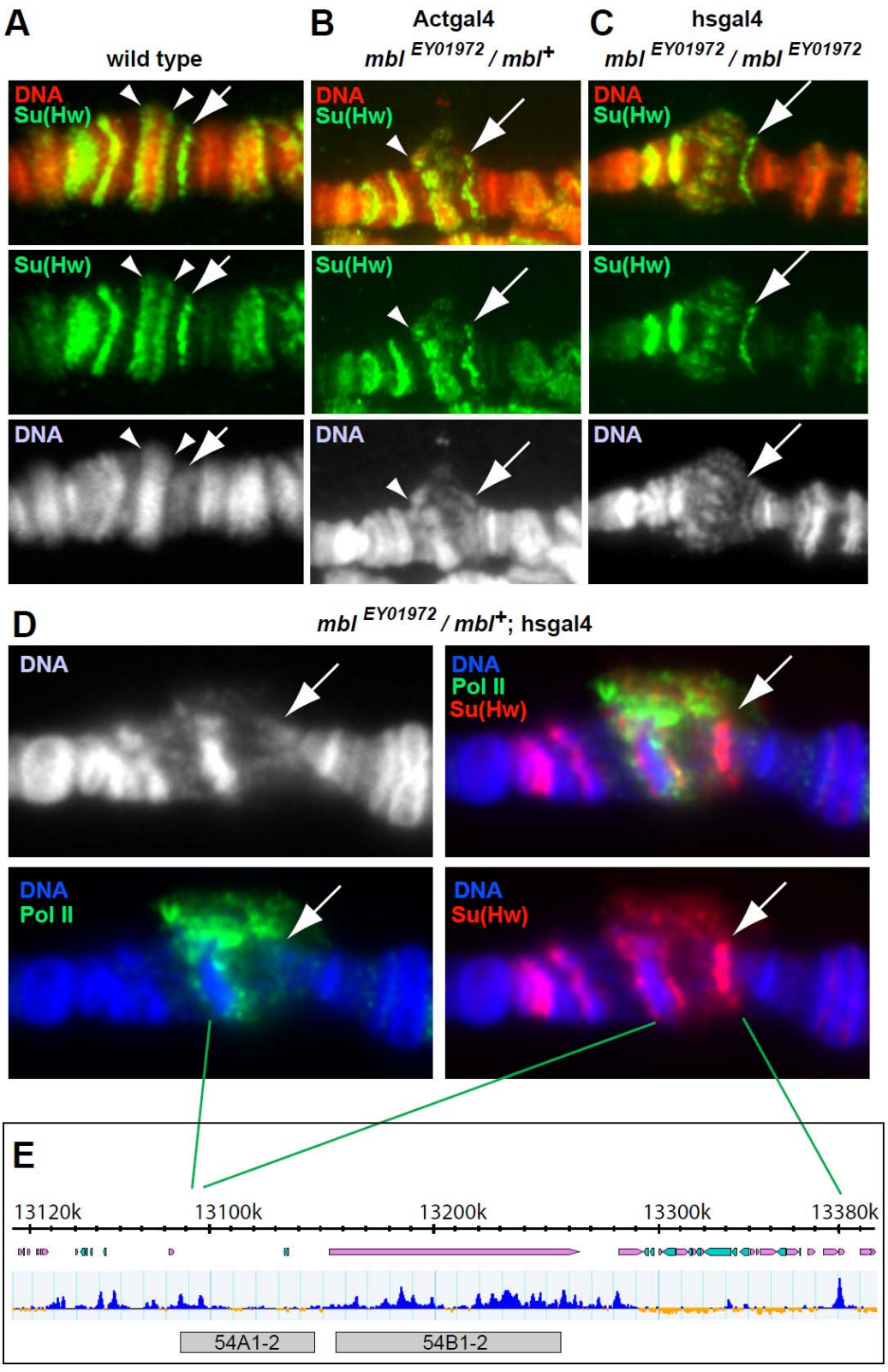
the highly condensed DAPI bands, corresponding to cytological regions 54A1-2 and 54B1-2 (Figure 3.17 A)**.** We next used FISH to combine two probes (probe 2 and probe 3) containing Su(Hw) binding sites mapping within the intragenic sequences of *mbl* (see Figure 3.17 C). As expected, results showed that these sites localized within the 54B1-2 band (Figure 3.17 C and D). Together, these results confirm that *mbl* is organized into a condensed chromosome band and that this band is associated to Su(Hw) insulator sequences.

Ectopic activation of *mbl* disrupts chromatin organization

A general correlation between transcriptional activity and banding pattern in polytene chromosomes has been established for many years. Thus, areas with low transcriptional activity are associated with highly compact bands, whereas actively transcribed genes are localized to decondensed interbands (Zhimulev et al., 2004). Genes such as *mbl* are embedded in bands of condensed chromatin. It is likely that packing of chromatin into bands is mediated by mechanisms involving the formation of higher-order chromatin structures to ensure low basal transcriptional activity in tissues where the gene is off. To directly observe under the fluorescence microscope the changes in chromatin organization that accompany transcriptional activation in genes embedded in condensed chromatin, we took advantage of the $mbl^{EY01972}$ line (Figure 3.18). Our results so far have

Figure 3.18 Condensed chromatin at the *mbl* locus is disorganized following ectopic activation of *mbl* in polytene chromosomes

(A). In wild type, *mbl* appears as highly condensed chromatin flanked by Su(Hw). (B). Following ectopic activation of heterozygous $mbl^{EY01972}$ by Act5C-GAL4 activator, one side of the chromatin becomes noticeably disrupted. (C). Bands 54A1-2 and 54B1-2 were strongly disorganized after transcription was induced at *mbl* on both chromosomes. (A-C). DNA was stained with DAPI (Red) and Su(Hw) is shown in green. (D). *mbl* was ectopically activated using a heat shock GAL4 driver at third instar larval stage. Chromosomes were immunostained with antibodies against Pol II (green) and Su(Hw) (red). PolII greatly accumulated at *mbl* locus and neighboring regions. (E). Distribution of Su(Hw) proteins in the *mbl* locus, showing a schematic representation of *mbl* and neighboring regions.

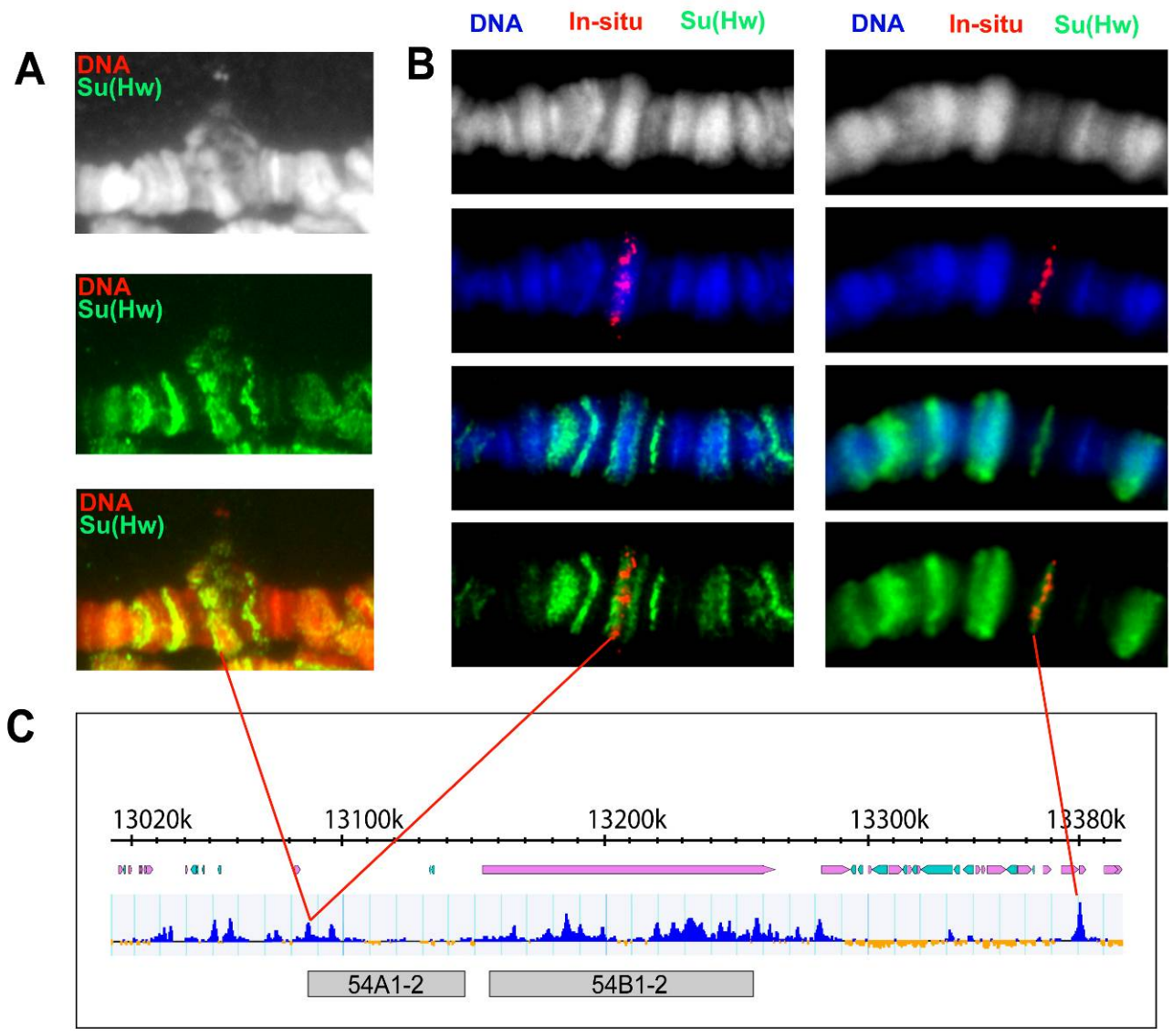

A
wild type
B
Actgal4
mbl EY01972 / mbl+
C
hsgal4
mbl EY01972 / mbl EY01972
DNA
Su(Hw)
Su(Hw)
DNA
D
mbl EY01972 / mbl+; hsgal4
DNA
DNA
Pol II
Su(Hw)
DNA
Pol II
DNA
Su(Hw)
E
13120k
13100k
13200k
13300k
13380k
54A1-2
54B1-2

shown that a large fraction of the *mbl* locus is organized as highly condensed chromatin in a compact DAPI band flanked by two strong Su(Hw) signals (Figure 3.17 A-B and Figure 3.18 A). After activation of transcription at the $mbl^{EY01972}$ promoter by actin-gal4 in $mbl^{EY01972}$ heterozygous flies, half of the *mbl* condensed 54B1-2 band, as well as the adjacent 54A1-2 band, appear disrupted, and Su(Hw) immunostaining signals are broken into disorganized fragments (Figure 3.18 B). However, since we used an *actin* driven GAL4 transgene, and the *actin* promoter stays active throughout development, it is possible that the chromatin disorganization observed is not the result of transcription activation of *mbl* at the condensed *mbl* chromatin. Instead, chromatin disorganization might be the consequence of a failure to form condensed chromatin early during development, when the actin promoter is already producing GAL4 and activating transcription at *mbl*.

To determine whether activation of transcription at a condensed band formed by a large gene leads to chromatin disorganization, we used an hsgal4 transgene to activate transcription of $mbl^{EY01972}$ only after the complete development of polytene chromosomes has taken place. Third- instar larvae carrying both $mbl^{EY01972}$ and hsgal4 activator were collected, and heat shock was carried out at 37 °C for 30 minutes, followed by 1 hour recovery at room temperature. Results show that activation of transcription in third-instar larvae had an even more

Figure 3.19 Ectopic activation of mbl gene leads to disorganization of chromatin structure spanning around 300 bp region

(A). The chromatin of *mbl* region is significantly disrupted after induction of *mbl* transcription. (B). The region of disorganization is roughly located between two strong Su(Hw) binding sites. By using FISH combined with immunostaining, the upstream Su(Hw) binding site is mapped near gene *CG10950*, and the downstream Su(Hw) binding is found near gene *CG30134*. C. Schematic representation of *mbl* and neighboring regions with its position in the chromosome in kb.


A
DNA
Su(Hw)
DNA
Su(Hw)
DNA
Su(Hw)
B
DNA
In-situ
Su(Hw)
DNA
In-situ
Su(Hw)
C
13020k
13100k
13200k
13300k
13380k
54A1-2
54B1-2

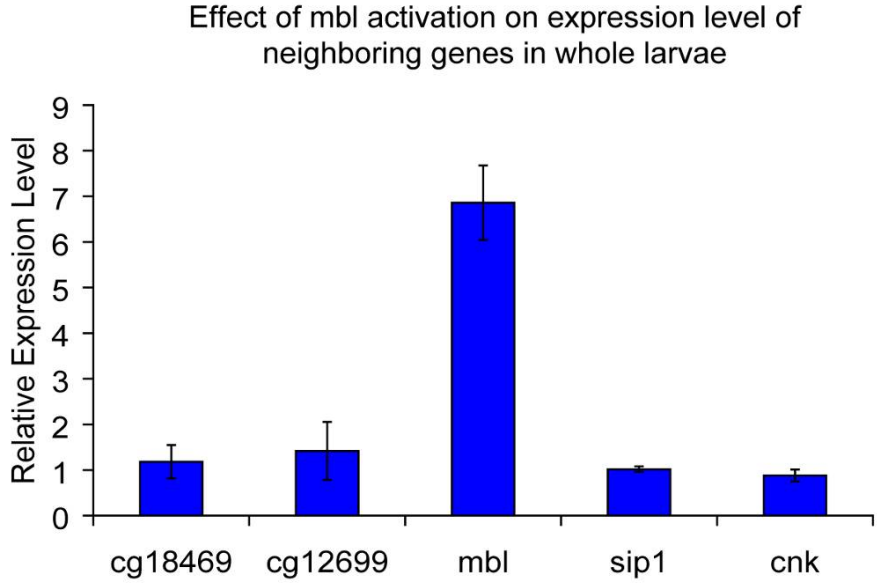
dramatic effect on chromatin organization than in the previous experiments using actin-gal4. In hsgal4 *mbl*^{EY01972} homozygous heat shocked larvae, the 54B1-2 and 54A1-2 bands, as well as adjacent regions, appear largely disrupted (Figure 3.18 C). Interestingly, both DAPI bands appear fragmented into DAPI dots, and only a certain residual amount of Su(Hw) protein remains associated to these DAPI signals. These results show that *mbl* can be activated after chromatin has been condensed to form bands in polytene chromosomes. Activation of *mbl* induces important effects on chromosome structure, which are not limited to the *mbl* locus and spread to adjacent regions, suggesting that activation of genes, especially large genes, could potentially influence transcriptional activity of neighboring genes.

We have used in situ hybridization data and distribution data of Su(Hw) insulators to map the extent of chromatin disorganization induced by ectopic activation of *mbl* in polytene chromosomes (Figure 3.19). Results show that immunostaining signals of Su(Hw) and the structure of the chromosome appear disorganized in a region spanning approximately 300 kb, which is much larger than the 110kb *mbl* DNA that is actively transcribed (Figure 3.18 B-C). To determine whether the extent of chromatin disorganization is encompassed by Pol II activity, we performed immunostaining using a phosphoserine 5-specific H14 antibody (Covance), which specifically recognizes the active form of RNA Polymerase II

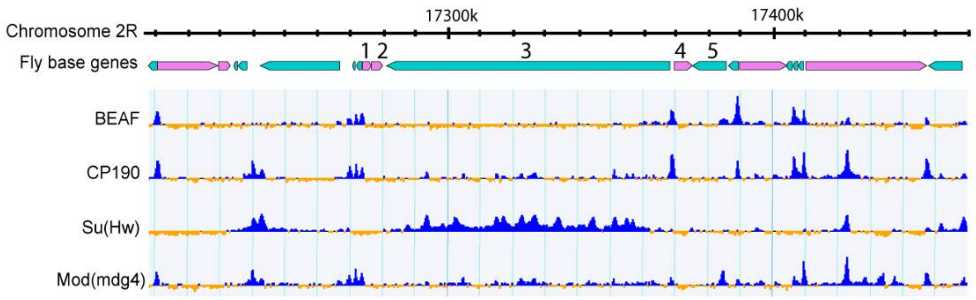
Figure 3.20 Real-time RT-PCR analysis of *mbl* and neighboring genes' expression after ectopic activation of *mbl* in the third-instar larvae

(A) Transcriptional level of each gene before *mbl* activation was set as 1. Enrichment of mRNA was normalized to rp49 internal control. (B) Schematic representation of mbl locus on the gene map as well as distribution of four different insulator proteins in this region. The number above each gene indicates the genes analyzed in real-time RT PCR. From let to right, 1:*CG18469*, 2:*CG12699*, 3: *mbl*, 4: *Sip1* and 5: *cnk.*

A


Effect of mbl activation on expression level of neighboring genes in whole larvae
Relative Expression Level
0
1
2
3
4
5
6
7
8
9
cg18469
cg12699
mbl
sip1
cnk

B


Chromosome 2R
17300k
17400k
Fly base genes
1 2
3
4 5
BEAF
CP190
Su(Hw)
Mod(mdg4)

(Pol II). Results show that a large amount of Pol II accumulated at one half of the *mbl* locus where the DAPI band was broken as a consequence of GAL4 activation, indicating occurrence of robust transcription (Figure 3.18 D).

Ectopic activation of *mbl* affects transcription of adjacent genes in a tissue-and gene-specific manner

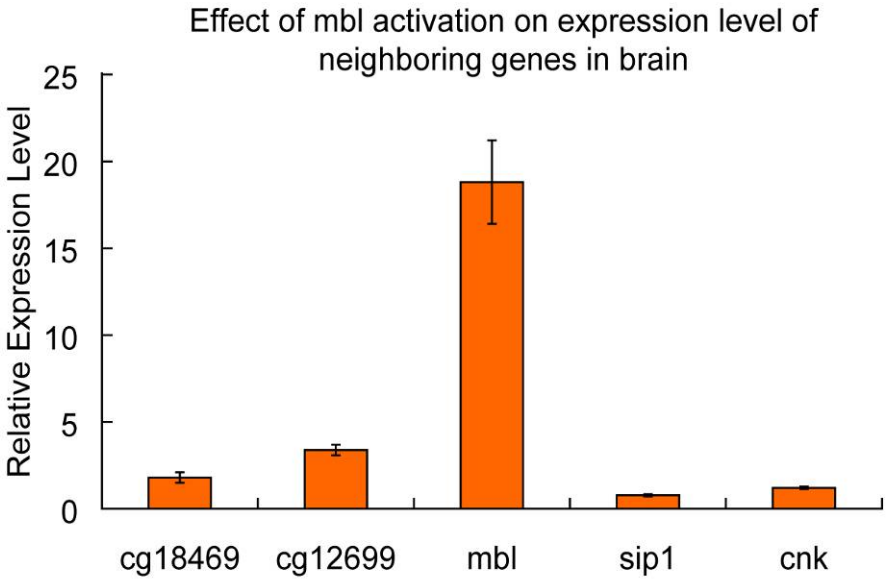
Since in the previous experiments, Pol II signals as well as disrupted chromatin structures appeared to expand beyond the boundaries of *mbl* (Figure 3.18 B-C), we asked whether overexpression of a large gene such as *mbl* could potentially influence chromatin organization and possibly the transcriptional activity of adjacent genes, and whether chromatin insulators may have a role preventing changes in chromatin structure from having an effect on transcription of adjacent genes. FISH combined with immunostaining experiments show that *mbl* ectopic activation has noticeable effects in a region flanked between two strong Su(Hw) binding sites. An upstream Su(Hw) binding sites is mapped near gene *CG10950*, whereas a downstream site is located near gene *CG30134* (Figure 3.19). In order to test whether induced decondensation of chromatin at *mbl* would result in higher transcriptional activity of neighboring genes, we extracted total RNA from third-instar larvae carrying both $mbl^{EY01972}$ and actin-gal4 activator. RNA obtained from $mbl^{EY01972}$ heterozygous flies without actin-gal4 activator was used as control. A total of four *mbl* neighboring genes located between genes *CG10950* and

CG30134 were selected and their transcriptional levels were determined using real-time RT PCR before and after *mbl* activation respectively. Two out of four genes (*CG18469* and *CG12699*) are located upstream of *mbl* and the other two genes (*Sip1* and *cnk*) are located downstream of *mbl* (Figure 3.20 B). Our real time PCR analysis reveals that *mbl* gene displayed about seven-fold ($p<0.01$) increase in the expression level (Figure 3.20 A). A slight increase was also observed in the transcription of two genes *CG18469* and *CG12699* located upstream of *mbl*. In contrast, no obvious changes in transcription were detected for *mbl* downstream neighboring genes. However, the RNA samples used in our real-time RT PCR experiments were obtained from whole larvae, which do not allow discrimination of transcriptional changes in different tissues. For this reason, we next extracted RNA from both brain and salivary glands before and after *mbl* activation respectively, and examined the transcriptional level of five genes mentioned above. Real-time PCR results show that transcriptional levels of *mbl* displayed a much more significant increase in both tissues than the average transcription level observed in the whole larvae (Figure 3.21). Specifically, transcription of *mbl* was boosted by 18.8-fold ($p<0.01$) in brain and 41.4-fold ($p<0.001$) in salivary glands. The dramatic increase in the transcription of *mbl* detected in salivary glands is also consistent with the abundant accumulation of RNA Pol II on the polytene chromosomes at the *mbl* locus observed using immunosgaining experiments. The upstream gene *CG18469* displayed 1.8-fold

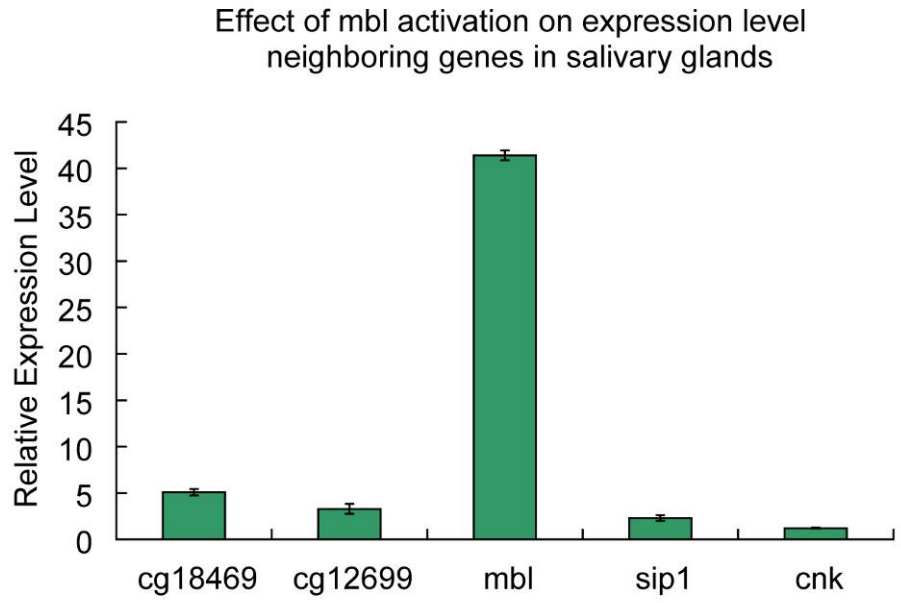
Figure 3.21 Real-time RT-PCR analysis of *mbl* and neighboring genes' expression after ectopic activation of *mbl* in brain

RNA was extracted from brain (A) and salibary glands (B). Transcriptional level of each gene before *mbl* activation was set as 1. Enrichment of mRNA was normalized to rp49 internal control. (C) Schematic representation of *mbl* locus on the gene map as well as distribution of four different insulator proteins in this region.

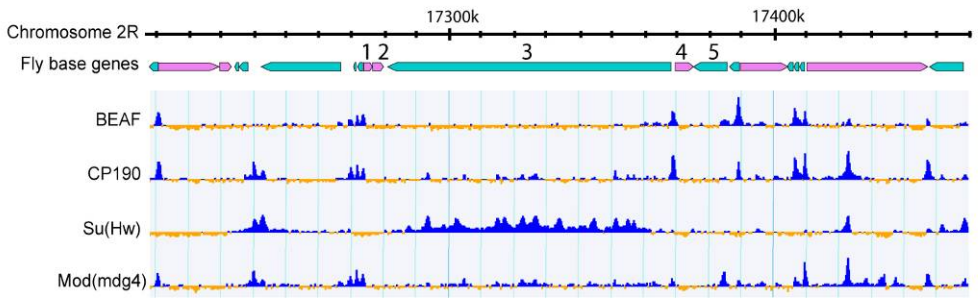
A


Effect of mbl activation on expression level of neighboring genes in brain
Relative Expression Level
25
20
15
10
5
0
cg18469
cg12699
mbl
sip1
cnk

B


Effect of mbl activation on expression level neighboring genes in salivary glands
Relative Expression Level
45
40
35
30
25
20
15
10
5
0
cg18469
cg12699
mbl
sip1
cnk

C


Chromosome 2R
17300k
17400k
Fly base genes
1 2
3
4 5
BEAF
CP190
Su(Hw)
Mod(mdg4)

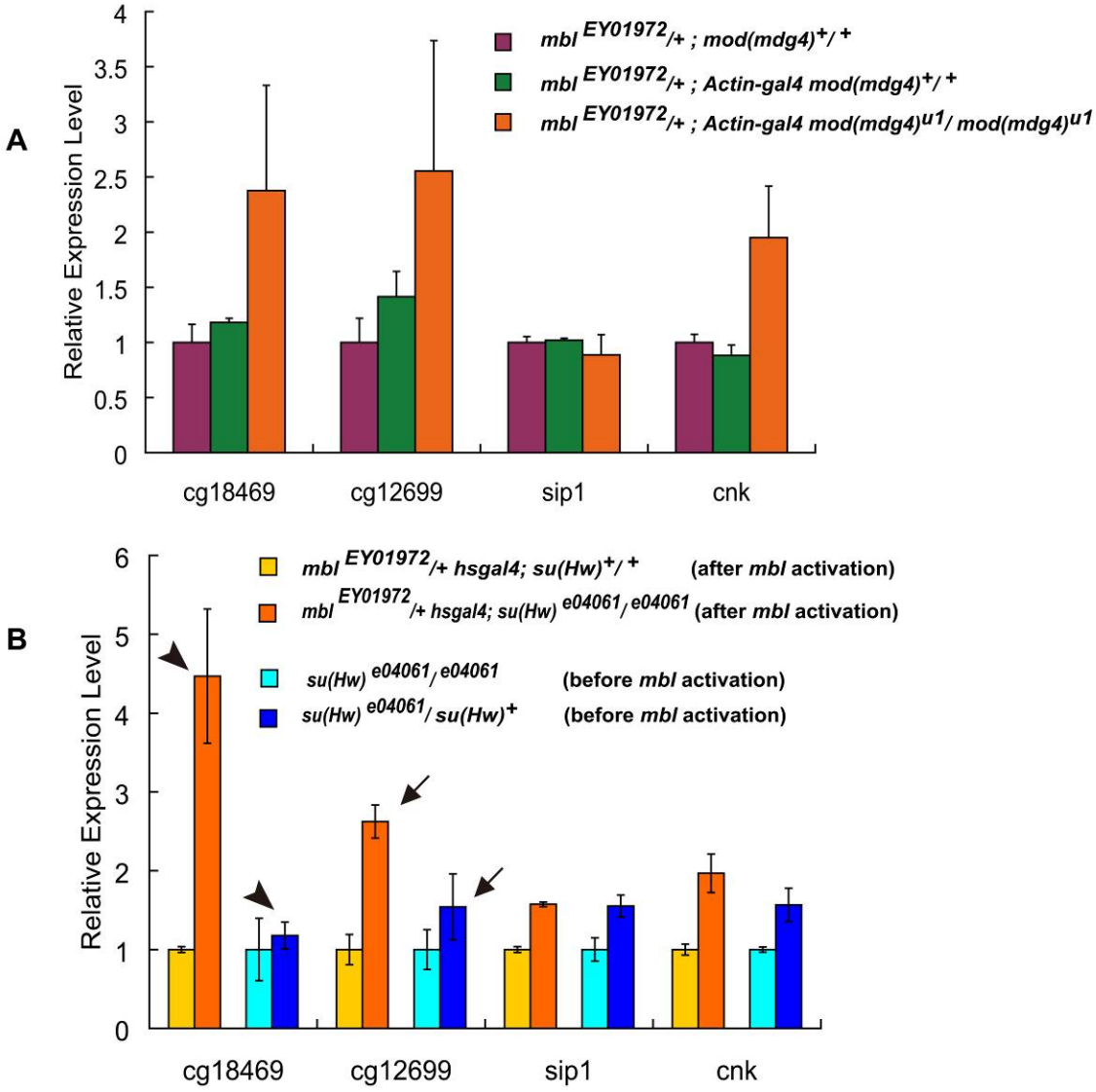
(p=0.07) increase in brain and 5-fold (p=0.051) increase in salivary glands. Transcription of gene *CG12699*, located upstream *mbl*, was also increased by 3.4-fold (p=0.004) and 3.3 –fold (p=0.072) in brain and salivary glands respectively. In contrast to the significant increase observed in the genes upstream of *mbl*, the downstream genes show much less significant changes. For example, only about 1.2- fold (p=0.14) increase in transcription is detected for the gene *cnk* in brain and the transcription of *Sip1* was even decreased by approximate 20% in brain (p=0.11). This observation is somehow unexpected considering the fact that an 'open' chromatin conformation and a significant accumulation of RNA Pol II caused by *mbl* activation were also found both upstream and downstream *mbl*. However, this result is consistent with previous observations that several regions with rather low transcriptional activities are located at interbands (Demakov et al., 2004), suggesting that the decondensed state of chromatin and transcriptional activity could be uncoupled. Taken together, here we show that although the maintenance of a condensed chromatin structure is concomitant with a continuous low expression of associated genes, decondensation of chromatin structure does not necessarily lead to high gene expression levels. In addition, our results also suggest that insulators may have a role preventing the spreading of changes in local chromatin structure to adjacent genes.

Endogenous Su(Hw) insulators have a stronger repressive effect on transcription of *mbl* and its upstream neighboring genes after *mbl* activation than before activation

We next asked why genes downstream of *mbl* are less affected by *mbl* activation than those located upstream of *mbl*. Considering that there is a strong Su(Hw) binding site right upstream of *Sip1* promoter, one may expect that this insulator could function as a boundary to protect downstream genes form ectopic activation. In order to test this possibility, we compared the expression of genes after induction of *mbl* transcription in both presence and absence of insulator proteins in the third-instar larvae. As mention above (Figure 3.20 A), when we activated *mbl* transcription using actin-gal4 activator, no obvious change was detected in the transcription of *Sip1* gene. We found that removal of Mod(mdg4) 67.2 protein in a *mod(mdg4)*^{u1} genetic background did not lead to any significant change in *Sip1* transcription (Figure 3.22 A). We did observed about two-fold (p=0.063) increase in the transcription of *cnk* in the absence of Mod(mdg4) 67.2 protein compared with wild type***.*** However, the two-fold increase is not likely caused by *mbl* activation, since removal of Mod(mdg4) 67.2 protein itself could almost double the transcription of the *Sip1* gene (Figure 3.9 B). These results do not support a role of Su(Hw) insulators upstream of *Sip1* in isolating downstream genes from *mbl* activation. Interestingly, we found that transcription of *CG18469* and *CG12699* was significantly increased after *mbl* activation without Mod(mdg4)

Figure 3.22 Real-time RT-PCR analysis of effects of *mbl* gene activation on the transcription of neighboring genes in the absence insulator proteins

(A). Expression of *mbl* was ectopically activated with actin-gal4 activator, transcriptional level of *mbl* adjacent genes were analyzed using gene specific primers. Purple and green colors represent transcriptional levels of these genes before and after *mbl* activation respectively in the presence of Mod(mdg4) 67.2 protein, orange color represents gene expression after *mbl* activation without Mod(mdg4) 67.2 protein. (B) Gene *mbl* was activated with heat shock activator. Transcriptional levels of the same set of genes were determined. Purple: before heat shock treatment in the presence of Su(Hw) protein. Green: after heat shock treatment in the presence of Su(Hw) protein. Orange: after heat shock treatment in the absence of Su(Hw) protein. (C). Schematic representation of genes tested on the gene map and distribution of insulator proteins in this region. 1: *CG18469*, 2: *CG12699*, 3: *Sip1* 4: *cnk*


A
Relative Expression Level
4
3.5
3
2.5
2
1.5
1
0.5
0
mbl EY01972/+ ; mod(mdg4)+/+
mbl EY01972/+ ; Actin-gal4 mod(mdg4)+/+
mbl EY01972/+ ; Actin-gal4 mod(mdg4)u1/ mod(mdg4)u1
cg18469
cg12699
sip1
cnk
B
Relative Expression Level
6
5
4
3
2
1
0
mbl EY01972/+ hsgal4; su(Hw)+/+ (after mbl activation)
mbl EY01972/+ hsgal4; su(Hw) e04061/e04061 (after mbl activation)
su(Hw) e04061/e04061 (before mbl activation)
su(Hw) e04061/ su(Hw)+ (before mbl activation)
cg18469
cg12699
sip1
cnk
C
Chromosome 2R
13100k
13200k
13300k
Fly base genes
1 2
3
4
BEAF
CP190
Su(Hw)
Mod(mdg4)

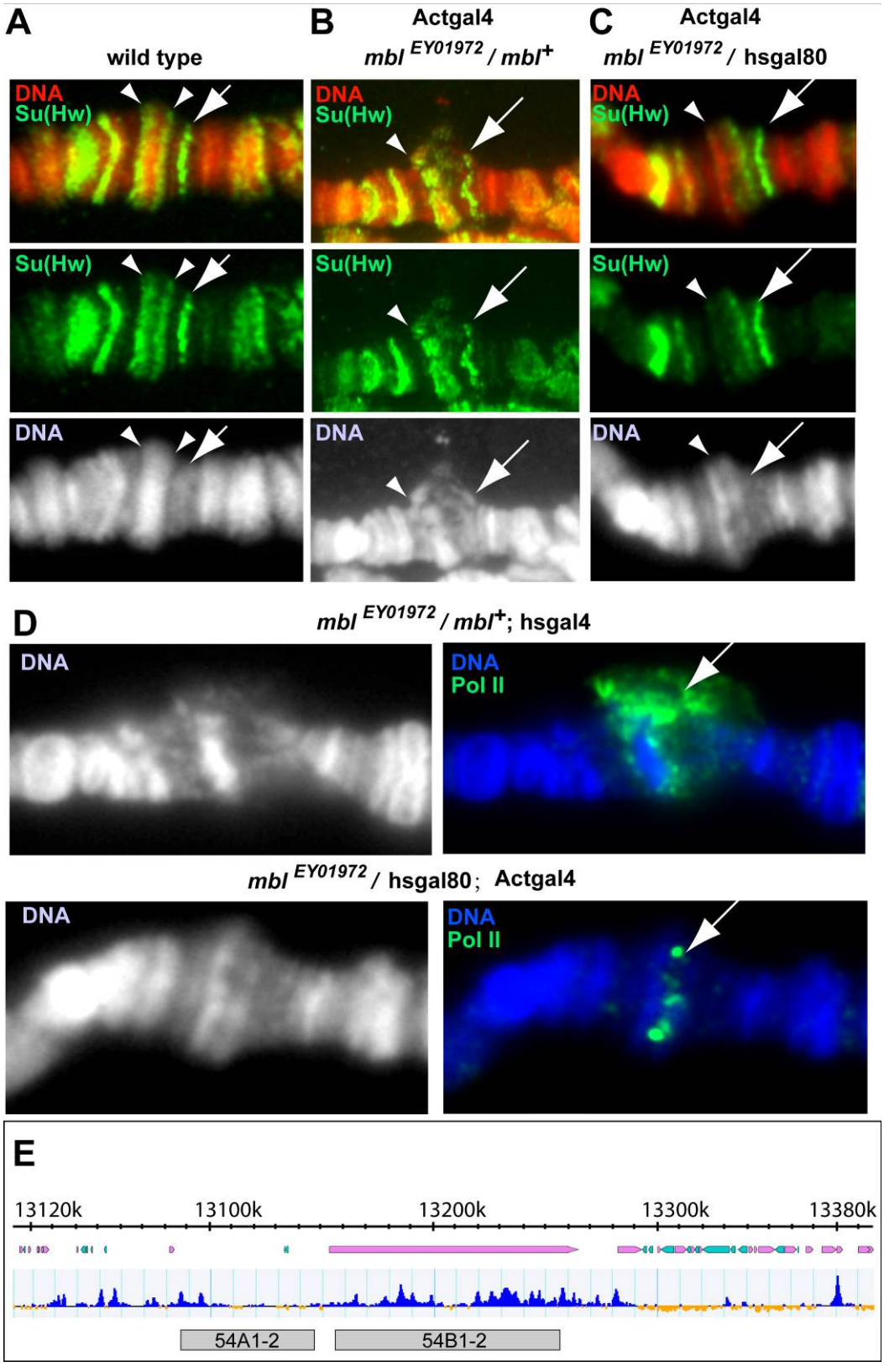
67.2. It appears that these increases are not caused by absence of Mod(mdg4) 67.2 protein since our real-time PCR shows that *Mod(mdg4)*^{u1} mutation did not induce obvious change in transcription of *CG12699* and *CG18469* (Figure 3.9B). A very similar result was also observed in the transcription of *mbl* as described above. *mbl* activation in a *Mod(mdg4)*^{u1} mutant background induces about 5-fold ($P<0.01$) increase in *mbl* transcription, compared with only 2.2-fold ($p<0.01$) increase before *mbl* activation (compare figure 3.6 A with figure 3.5 A).

In order to further validate this result, we also activated *mbl* transcription using heat shock Gal4 as mentioned above. We arbitrarily defined the transcriptional level of each gene after *mbl* activation to be 1 (yellow bar in figure 3.22 B) and compared the change in gene transcription due to lack of Su(Hw) protein (orange bar in figure 3.22 B). The relative transcriptional level before *mbl* activation in wild type (light blue bar in figure 3.22 B) or *su(Hw)* ^{e04061} mutant (dark blue in figure 3.22 B) is listed in figure 3.16. Results show that *Su(Hw)* has little boundary effect in the transcription of genes downstream *mbl* (*Sip1* and *cnk*) after ectopic activation of *mbl* (compare the blue bar with the orange bar). This result further suggests that the Su(Hw) insulator site upstream the *Sip1* promoter does not function to prevent changes in chromatin structure from spreading from *mbl* into downstream genes. On the other hand, lack of Su(Hw) proteins caused a significant increase in the transcription of genes *CG18469* and *CG12699*

(upstream *mbl*) after ectopic activation of mbl. Specifically, the gene *CG12699* underwent a 1.5-fold (p=0.098) increase in transcription in a *su(Hw)* ^{e04061} mutant background compared with wild type, but the same gene produced a 2.6-fold (p=0.011) increase in transcription levels after *mbl* activation in the same *su(Hw)* ^{e04061} mutant background (arrow in figure 3.22 B). A larger difference was observed in the transcription of gene *CG18469*, which displayed 1.2 (0.078) -and 4.5-fold (p=0.033) increase respectively before and after *mbl* activation in *su(Hw)* ^{e04061} mutants. One possibility to explain these results is that Su(Hw) functions as the only insulator in a chromatin domain formed by *mbl* plus the region upstream *mbl* promoter, and therefore lack of Su(Hw) protein allows spreading of changes in chromatin structure such as those taking place during *mbl* activation. The region downstream the *mbl* termination site constitutes a different chromatin domain where other insulator proteins such as CP190, BEAF or CTCF are more abundant than Su(Hw), and likely function to prevent spreading of changes in chromatin structure. Under this assumption, the effect of the lack of Su(Hw) in this domain is negligible, since the barrier function is provided by the other insulator proteins.

Figure 3.23 Establishment of highly condensed or less condensed chromatin confirmation at the *mbl* locus is a dynamic process

(A). In wild type, *mbl* locus appears as highly condensed chromatin band flanked by Su(Hw) insulator signals. (B). Following ectopic activation of heterozygous $mbl^{EY01972}$ by Act5C-GAL4 activator, one side of the chromatin becomes noticeably disrupted. C. The chromatin structure near mbl locus is largely recovered upon expression of GAL80. (D). Accumulation of Pol Ⅱ at *mbl* locus is greatly reduced due to effect of GAL80. E. Distribution of Su(Hw) proteins in the *mbl* locus, showing a schematic representation of *mbl* and neighboring regions.


A
wild type
B
Actgal4
mbl EY01972 / mbl+
C
Actgal4
mbl EY01972 / hsgal80
DNA
Su(Hw)
Su(Hw)
DNA
D
mbl EY01972 / mbl+; hsgal4
DNA
DNA
Pol II
mbl EY01972 / hsgal80; Actgal4
DNA
DNA
Pol II
E
13120k
13100k
13200k
13300k
13380k
54A1-2
54B1-2

Establishment of banding pattern of polytene chromosomes is a dynamic process

Since ectopic activation of *mbl* transcription is able to disrupt the highly condensed chromatin band, we questioned whether the condensed chromatin structure found in bands could be rebuilt once transcription activation is turned off. To test this possibility, we utilized a transgenic line carrying heat shock GAL80 transgene to further control the expression of the *mbl* allele driven by GAL4/UAS system. GAL80 has been shown to be a transcriptional repressor. Upon expression, GAL80 directly acts on GAL4 by binding to its transactivation domain, which blocks the activity of GAL4 (Lue et al., 1987; Suster et al., 2004). To activate expression of GAL80,third-instar larvae carrying mbl^{EY1972} ,actin-gal4 as well as hs-gal80 were incubated at 37℃ for 30 minutes followed by 1 hour recovery. Flies carrying mbl^{EY1972}; actin-gal4 or mbl^{EY1972}/hsgal4 without hs-gal80 were used as control. Immunostaining with rabbit anti-Su(Hw) antibodies combined with mouse anti-H14 (RNA Pol II) antibodies (Covance; Princeton, New Jersey) shows that, expression of GAL80 largely repressed expression *mbl*, as evidenced by significantly reducing accumulation of RNA Pol II at *mbl*, compared with control (Figure 3.23D). As a result, the disrupted condensed chromatin band at the *mbl* locus as well as two strong Su(Hw) signal bands flanking the DAPI stained DNA band was largely recovered (Figure 3.23 A-C). These observations

suggest that formation of highly condensed chromatin in bands is a dynamic process, which may depend on the transcriptional activity of associated genes.

Loss of endogenous Su(Hw) insulator function reduces DNase I accessibility along the *mbl* locus

Immunostaining of polytene chromosomes using antibodies against Pol II showed that the DAPI band 54B1-2 (*mbl* DNA) had a more open structure and contained a distinct level of transcriptionally active Pol II staining when compared with the adjacent band 54A1-2, which appears significantly more condensed and completely lacks Pol II staining (Figure 3.24). These differences at the polytene chromosome level led us to ask whether intragenic insulators at the *mbl* locus function by causing changes in chromatin structure that influence DNA accessibility, and therefore determine the levels of accessibility and initiation of transcription by Pol II. Since further analysis using polytene chromosomes from *su(Hw)* and *mod(mdg4)* mutant larvae did not reveal any significant differences (data not shown), we performed a DNase I-sensitivity assay. For this assay, nuclei were first isolated from third-instar larvae and then digested with increasing concentrations of DNase I. A 107 bp fragment located upstream of the *yellow* gene, which shows no detectable digestion under similar conditions (Chen and Corces, 2001), served as an internal control for equal loading of DNA in different

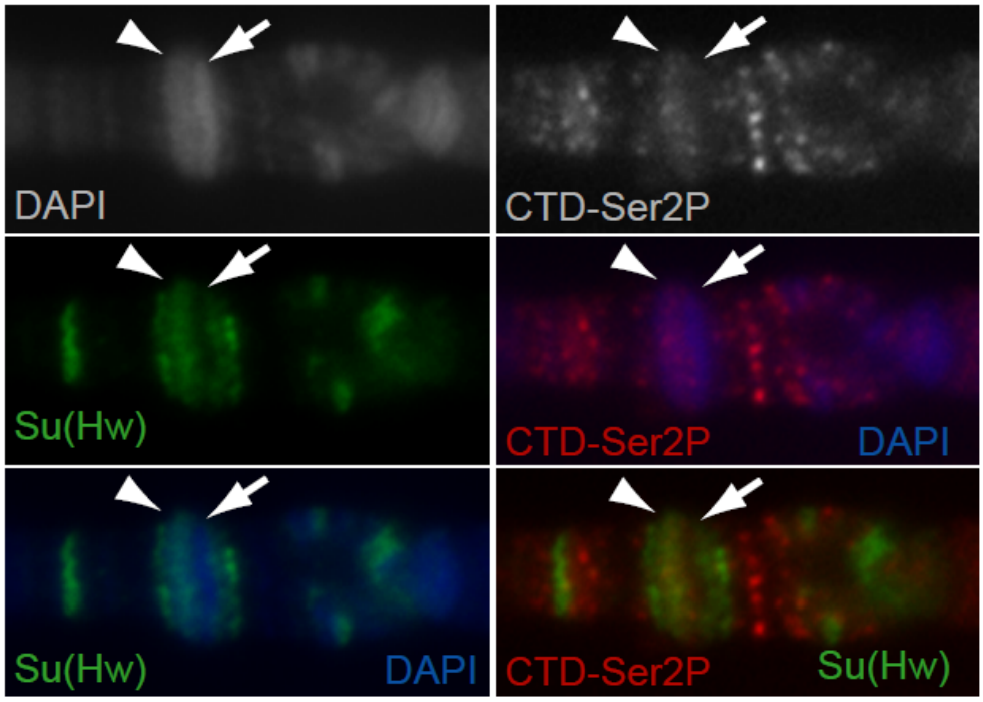
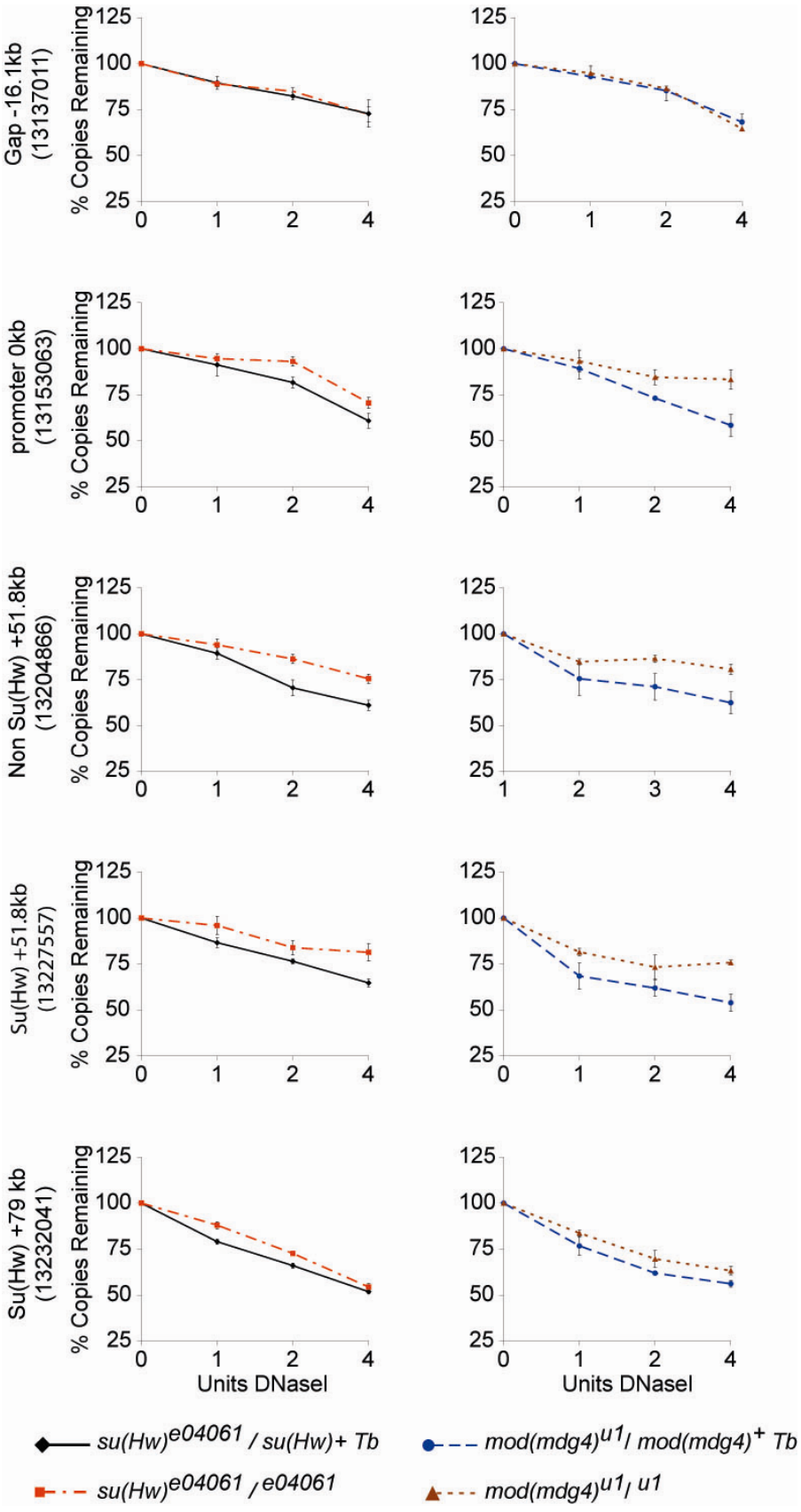


Figure 3.24 Phosphorylated RNA polymerase II is found associated to *mbl* in the 54B1-2 band

Immunostaining of polytene chromosomes using antibodies anti-Su(Hw) (Green) and anti-CTD-Ser2P (red). Arrowheads point to the 54B1-2, showing a different level of DAPI staining when compared with the 54A1-2 band (arrows), which mostly contains intergenic DNA. PolII and Su(Hw) associate with 54B1-2 and are missing from 54A1-2.

Figure 3.25 Removal of *gypsy* insulator proteins reduces chromatin accessibility to DNase I digestion in *mbl* chromatin.

DNase I accessibility assays were performed with nuclei isolated from wild type, *su(Hw)*^{e04061} and *mod(mdg4)*^{u1} backgrounds. Graphs were generated by plotting the percentage of remaining DNA copies, corrected for DNA content, against the number of units of DNase I used in the reaction. The digestion profiles were based on three independent experiments each performed in triplicate. The position in the *mbl* locus for every site analyzed is indicated to the left of the graph (parenthesis indicate the position in kbs in the chromosome).


Gap -16.1kb (13137011)
promoter 0kb (13153063)
Non Su(Hw) +51.8kb (13204866)
Su(Hw) +51.8kb (13227557)
Su(Hw) +79 kb (13232041)
% Copies Remaining
125
100
75
50
25
0
1
2
3
4
Units DNaseI
su(Hw)e04061 / su(Hw)+ Tb
su(Hw)e04061 / e04061
mod(mdg4)u1/ mod(mdg4)+ Tb
mod(mdg4)u1/ u1

reactions. The degree of digestion was analyzed by real-time PCR as described previously (McArthur et al., 2001). Before testing the effect of endogenous *gypsy* insulators on chromatin structure within *mbl* gene, we first examined the chromatin accessibility to DNase I of a gene-free region located about 16 kb upstream of *mbl*, which is also embedded in a condensed band but lacking endogenous *gypsy* insulators. As expected, no obvious change in chromatin accessibility was observed in mutant larvae, after removal of either Su(Hw) or Mod(mdg4)67.2 proteins, compared with wild type (Figure 3.25). Then we determined accessibility to DNase I digestion at four regions downstream of the promoter of *mbl*: promoter region, two Su(Hw) binding sites and one non-Su(Hw) binding site. Results indicate that DNase I sensitivity of all these regions declines significantly in the absence of Su(Hw) protein (Figure 3.25). A similar pattern was also observed in the *mod(mdg4)*^{u1} mutant. This result is consistent with previous observations indicating that the presence of the *gypsy* insulator is able to increase accessibility of chromatin to DNase I digestion (Chen and Corces, 2001). To compare these changes among different sites along *mbl* locus, we employed the method of McArthur and Bibb (McArthur and Bibb, 2006), which estimates the DNaseI sensitivity at a given site by the percent loss of DNA copies after nuclei digestion with 4 units of DNase I. In this way, higher DNase I sensitivity of chromatin is reflected by more loss of DNA at a specific site after DNase I digestion. Figure 3.26 shows that gypsy insulator proteins have little effect on the

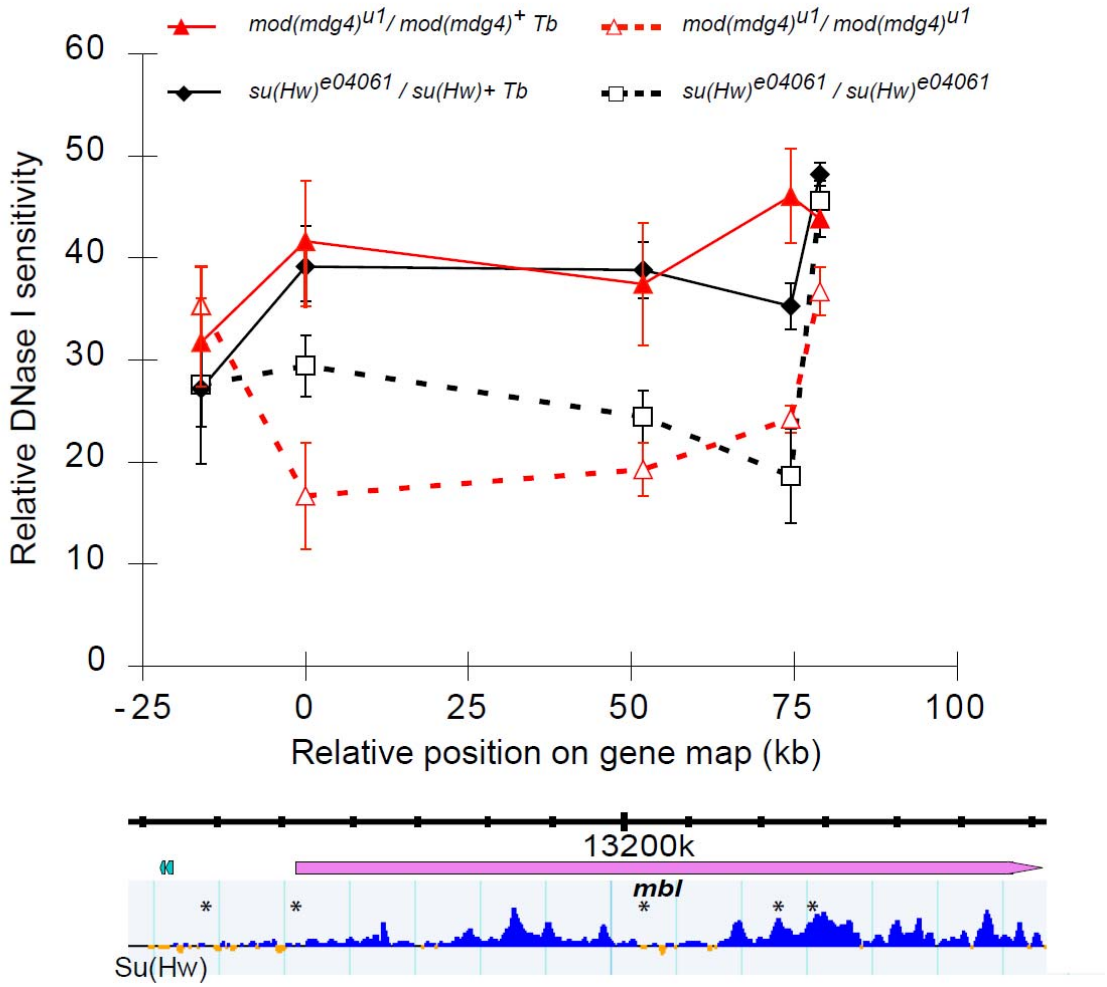


Figure 3.26 Relative DNAseI sensitivity in *mbl* significantly decreases in *su(Hw)* and *mod(mdg4)* mutants

DNAseI sensitivity for each site was estimated by the percent loss of DNA copies (Y axis) after digestion of chromatin with 4 units DNase I for 3 minutes on ice. Numbers on X axis represent genomic distance (in kb) of each site tested to *mbl* transcription start site (position 0).

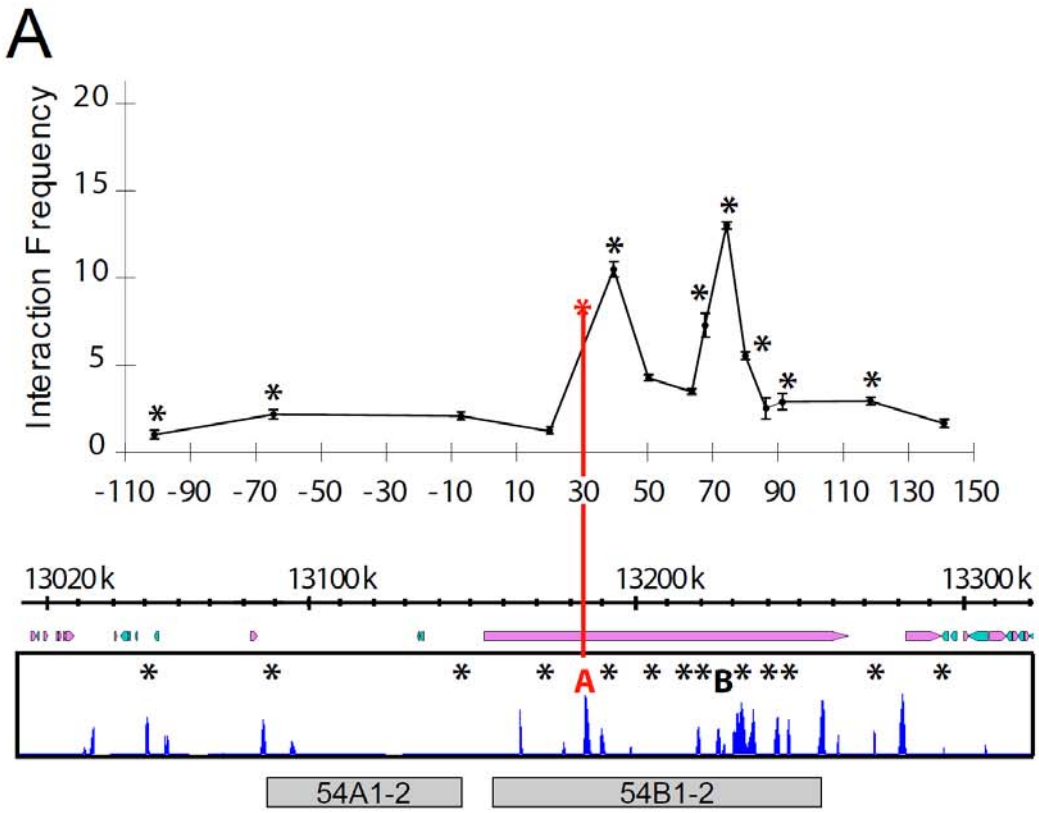
DNase I sensitivity upstream of *mbl* gene: 4 units of DNase I digestion induces about 30 percent loss of DNA in the presence or absence of gypsy insulator proteins. In contrast, DNase I sensitivity within the intragenic region of *mbl* was clearly reduced due to loss of Su(Hw) or Mod(mdg4)67.2 proteins. For example, loss of DNA was reduced by 9.7, 14.4, 16.7 percent respectively, at the three sites located at the 5' end of *mbl*, including the promoter region, in $su(Hw)^{e04061}$ compared with wild type. A more dramatic effect was detected in $mod(mdg4)^{u1}$ mutants, where loss of DNA declined by 24.9, 18.2 and 21.8 respectively. Surprisingly, at the most 3' end site, where the strongest Su(Hw) binding was found, reduction of DNA loss was only 2.6 percent in $su(Hw)^{e04061}$and 7.1 percent in $mod(mdg4)^{u1}$. We draw the following conclusion from the above results: first, the removal of *gypsy* insulator proteins makes chromatin at the *mbl* locus less accessible to DNase I, but has minor effects upstream of the transcription start site where insulators are absent. Second, the decrease in the DNase I accessibility of at the *mbl* locus in $su(Hw)^{e04061}$ and $mod(mdg4)^{u1}$ is not limited to insulator protein binding sites.

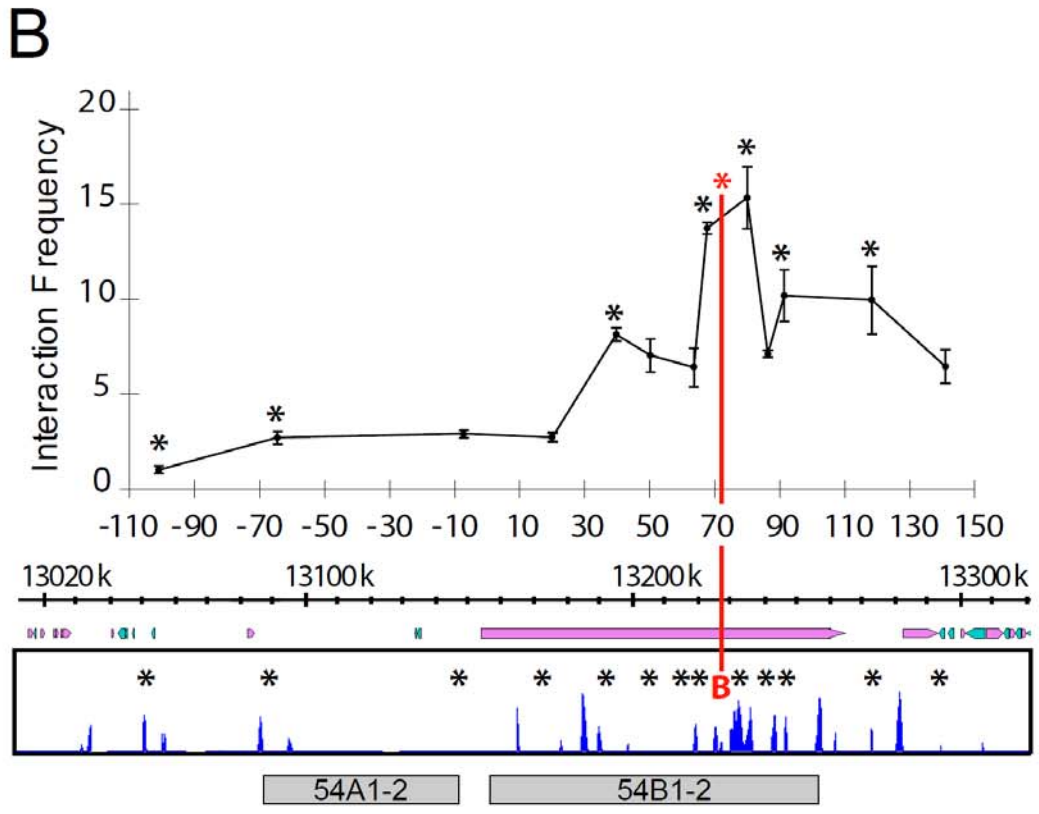
Chromosome Conformation Capture assays show that Su(Hw) insulators mediate the formation of chromatin loops in intragenic sequences of mbl

Some of the phenomena described here, such as formation of condensed chromatin in bands, broad changes in chromatin organization after induction of

Figure 3.27 Intragenic *mbl* Su(Hw) insulators form higher-order chromatin structures

A schematic representation of the 280 kb genomic region including *mbl* is shown at the bottom of each panel. Su(Hw) binding sites from Kc cells (Bushey et al., 2009) are shown as blue peaks. X axis indicates relative distance (in kb) of each site to transcription start site of *mbl* (position 0). The Y axis represents interaction frequencies between two *Eco*RI fragments obtained from the 3C library relative to the interaction frequencies obtained from BAC clones. Red asterisks indicate positions of anchor primer regions in each experiment (A and B). Black asterisks in the ChIP-on-chip profile indicate all sites tested for potential interactions. The two anchor fragments are 40 kb apart. Asterisks in the curve represent elements that coincide with Su(Hw) insulator peaks. In each graph, the lowest cross-linking frequency was defined to be 1. Error bars represent standard errors after three real-time PCR measurements for each point.


A
Interaction Frequency
20
15
10
5
0
-110 -90 -70 -50 -30 -10 10 30 50 70 90 110 130 150
13020k
13100k
13200k
13300k
A
B
54A1-2
54B1-2


B
Interaction Frequency
20
15
10
5
0
-110 -90 -70 -50 -30 -10 10 30 50 70 90 110 130 150
13020k
13100k
13200k
13300k
B
54A1-2
54B1-2

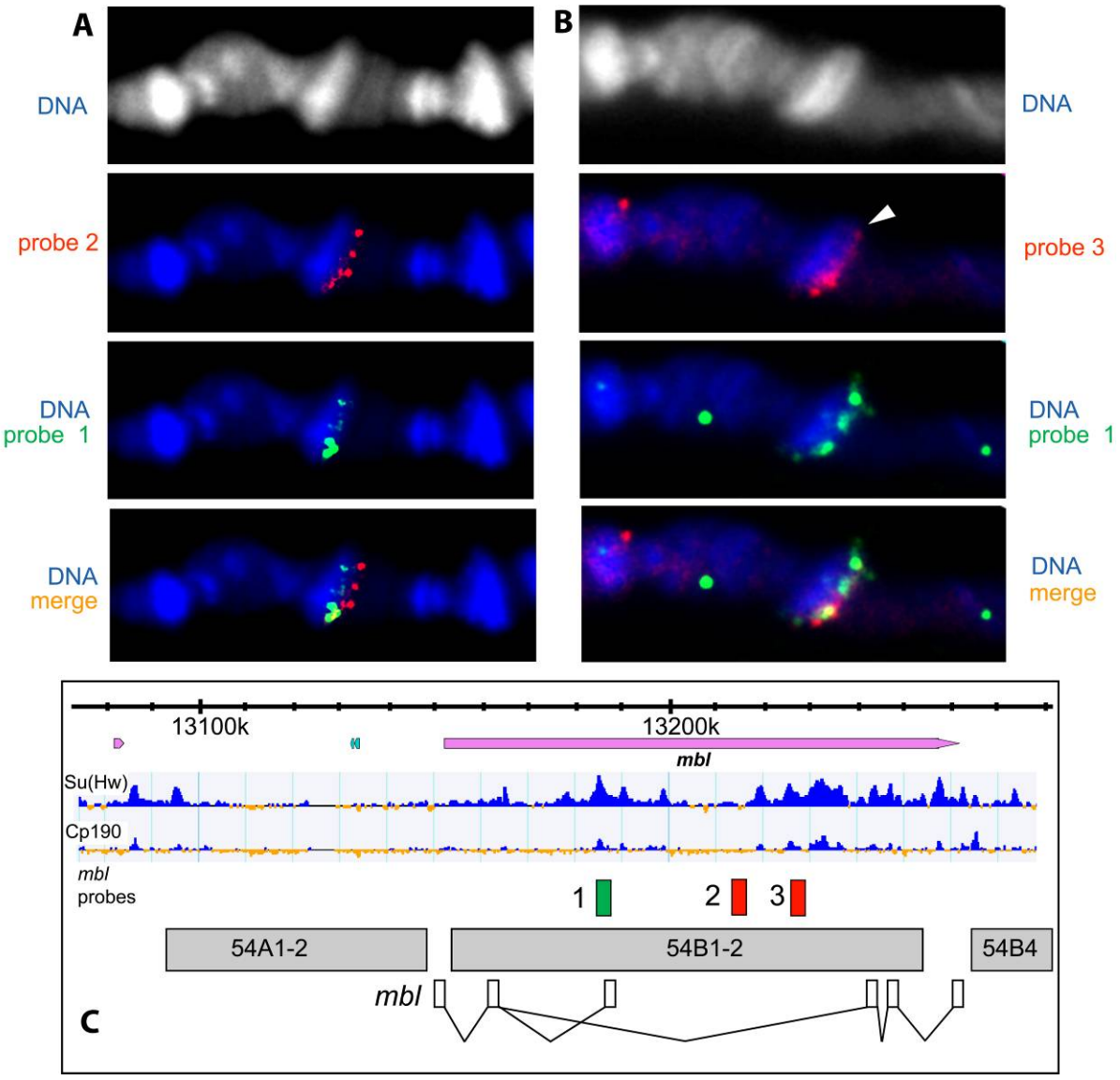
transcription, and DNAseI accessibility cannot be simply explained by classic mechanisms of gene transcription regulation in *cis*-. Such phenomena could be the response to mechanisms involving higher-order chromatin structure mediated by long-range interactions within the chromatin fiber (Labrador and Corces, 2002; West et al., 2002). To address the possibility that endogenous Su(Hw) insulators are directly involved in the three-dimensional organization of *mbl*, a high-resolution chromosome conformation capture (3C) assay was performed. 3C assays provide a powerful tool to detect interactions between two distant chromosomal regions (Kruithof et al., 2009; Wong et al., 2007). To investigate the potential long-range interactions between Su(Hw) binding sites at *mbl*, we have generated a 3C library from *Drosophila* S2 cells prepared from ligation products of an *Eco*RI digestion performed in nuclei. We first used a primer located 30 kb downstream of the promoter region of *mbl*, which contains a strong Su(Hw) binding site, as an ‘anchor’ and assessed interactions of this anchor with other DNA segments throughout the *mbl* locus (Figure 3.27 A). To control for the efficiency of different primer sets, we have also used a library generated using two minimally overlapping bacterial artificial chromosome (BAC) clones (RP98-48A11 and RP98-28012) spanning *mbl* plus neighboring loci, and mixed in equimolar amounts. The interaction frequency between two genomic sites was calculated by dividing the amount of PCR product obtained from the 3C library by that obtained from the control library. As shown in Figure 3.27 A, we have detected frequent

interactions between the anchor region and five other genomic locations (located at 39.7, 67.7 80.0, 91.4 and 118.4 kb downstream of the 5' end of *mbl*), as evidenced by local peaks in interaction frequencies. Four out of five sites analyzed are found within *mbl*, and one of them is located between *mbl* and *Sip1*. Interestingly, all five regions contain Su(Hw) binding sites. In contrast, genomic sites located at 50.3, 60.6 and 86.4 kb downstream of *mbl* promoter, do not contain Su(Hw) binding sites, and clearly show significantly lower interactions with the anchor region (Figure 3.27 A).

To further validate this result, we selected a region, located at 70 kb downstream of *mbl* promoter, which also contains a Su(Hw) binding site, as a second anchor. A strong interaction with this anchor was observed at a downstream Su(Hw) binding site located at 6.9 kb. The high interaction frequency between these two regions cannot be simply explained by the short distance between them (6.9kb) , since a site with similar distance located upstream of the anchor region, which does not contain a Su(Hw) binding site, displayed a much lower interaction frequency (Figure 3.27 B). Strong interaction with this anchor was also found at three other downstream regions containing a cluster of Su(Hw) binding sites (Figure 3.27 B). Taken together, these results show that, even though relative crosslinking efficiencies between anchor regions and other chromosomal regions are inversely proportional to their distances, local peaks

Figure 3.28 Su(Hw) insulators mediate physical interaction between distinct genomic sites, while looping out intervening DNA sequences

(A). Florescence in-situ hybridization using different combinations of probes with *mbl* gene. Left figure: Dig-labeled probe 1 contains a strong Su(Hw) binding site located near 5' end of *mbl* gene. Biotin-labeled probe 2 is located approximately 30.9kb downstream of proble 1, which contains no Su(Hw) binding site. Probe 1 and probe 2 were detected with anti-dig-fluorescein (green) and anti-avidin-rhodamin (red) respectively. Right figure: the same probe 1 is used as in left figure. Biotin-labeled probe 3 containing a strong Su(Hw) binding site is located 40.9kb downstream of probe 1. The probes are detected with same set of secondary antibody as in left figure. Note that the in-situ signals of probe 1and probe 3 colocalize very well (right), whereas those signals of probe 1 and probe 2 separates from each other (left). (B). Schematic representation of *mbl* locus on gene map and distribution of insulator proteins in this region. The number below insulator protein distribution map indicates relative position of probes on the gene map used in figure A.


A
B
DNA
DNA
probe 2
probe 3
DNA
probe 1
DNA
probe 1
DNA
merge
DNA
merge
13100k
13200k
mbl
Su(Hw)
Cp190
mbl
probes
1
2
3
54A1-2
54B1-2
54B4
mbl
C

can always be detected at Su(Hw) binding sites. These results strongly support the hypothesis that Su(Hw) insulators regulate higher-order chromatin organization through formation of multiple chromatin loops at the *mbl* locus.

If the above model is correct, we expect that the interacting Su(Hw) binding sites which serve as 'nodes' in the chromatin loops would colocalize on the polytene chromosomes. In contrast, genomic sites located in the middle of the chromatin loop which do not contain endogenous Su(Hw) insulators could potentially be distinguishable from those Su(Hw) binding sites. In order to test this possibility, we performed fluorescence in-situ hybridization assays with 3 different probes (probe 1-3, figure 3.28 C) containing the *Eco*RI reorganization sites used in the 3C experiments. Both probe 1 and probe 3 contain strong Su(Hw) binding sites. In the 3C assay these two sites were used as anchor sequences, and showed the highest interaction frequency (Figure 3.27 A). Probe 1 is labeled with digoxin and detected by anti-dig-fluorescein (green). Probe 3 is labeled with biotin and detected by anti-avidin-rhodamin (red). As shown in figure 3.28, even though the linear distance between probe 1 and probe 3 is more than 40 kb, their in-situ hybridization signals colocalized very well on the polytene chromosomes. The biotin-labeled probe 2, on the other hand, does not contain any Su(Hw) binding sites but has a shorter linear genomic distance (approximately 30kb) with probe 1. 3C results show that probe1 and probe 2 displayed an apparently lower

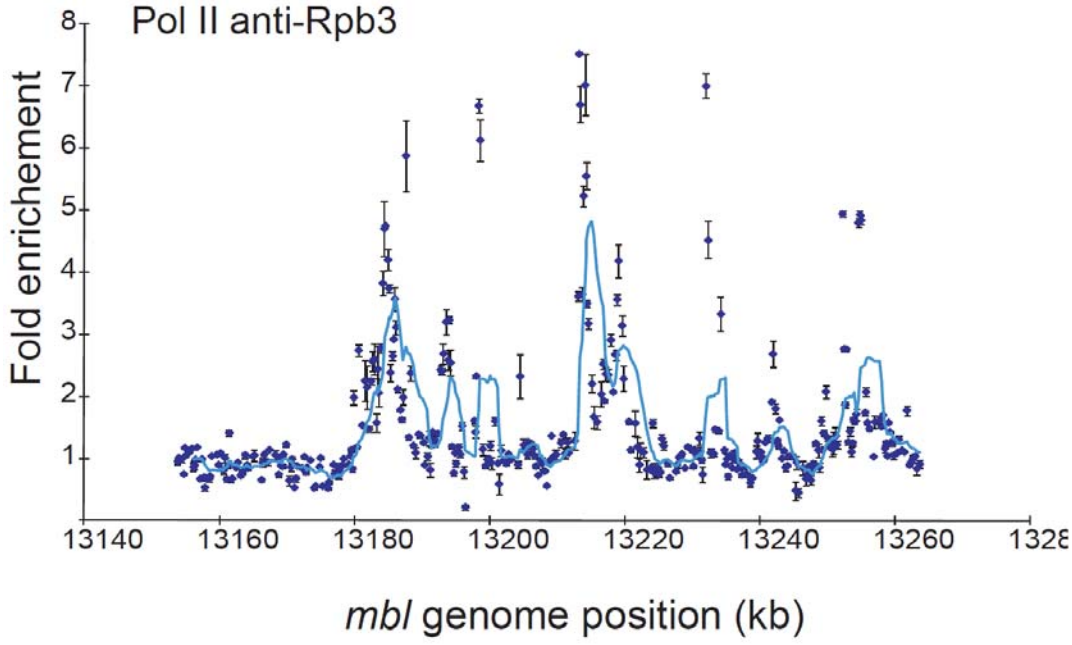
interaction frequency. Correspondingly, the in-situ signal of probe 2 does not overlap with probe 1 as much as probe 3. This result provides an additional experimental evidence that Su(Hw) insulators mediate physical interaction between distinct genomic sites, while looping out intervening DNA sequences.

Pol Ⅱ is enriched at intragenic insulator sites in *mbl*

Together, above data suggest that insulators may form higher-order chromatin structures that prevent the free passage of PolII, therefore decreasing the transcription levels of *mbl*. To test this hypothesis we analyzed previously published wide-genome ChIp-on-Chip data (GEO accession code GSE6714) containing the position of both, serine 2 phosphorylated PolII and Rbp3, a subunit of PolII , from Drosophila S2 cells (Walker and Sikorska, 1987a). This analysis shows that independently of the antibody used (anti-CTD-Ser2P or anti-Rpb3) PolII is significantly enriched at Su(Hw) insulator sites along *mbl* (Figure 3.29 and figure 3.30 A). Interestingly, it does not appear that the any of the insulator sites contributes more importantly as a barrier against Pol Ⅱ passage, suggesting that Pol Ⅱ is only temporarily stall at each insulator site. These sites are the same that are involved in long-range interactions in 3C experiments shown in Figure 3.27, and suggest a model in which the coalescence of insulators is necessary in order to prevent free passage of elongating polymerase (Figure 3.30 B).

Figure 3.29 Distribution of PolII in the *mbl* gene

Accumulation of PolII at intragenic Su(Hw) insulator sites is observed with antibodies anti- phosphorylated Pol II and with antibodies anti-Rpb3. Data points correspond to IP values representing the relative enrichment log2 IP/WCE obtained from ChIP-on-Chip data GEO accession GSE6714 (Muse et al. 2007).Contiguous lines represent the moving average calculated using overlapping sets of 10 data point intervals.


Pol II anti-Rpb3
Fold enrichement
8
7
6
5
4
3
2
1
13140
13160
13180
13200
13220
13240
13260
mbl genome position (kb)

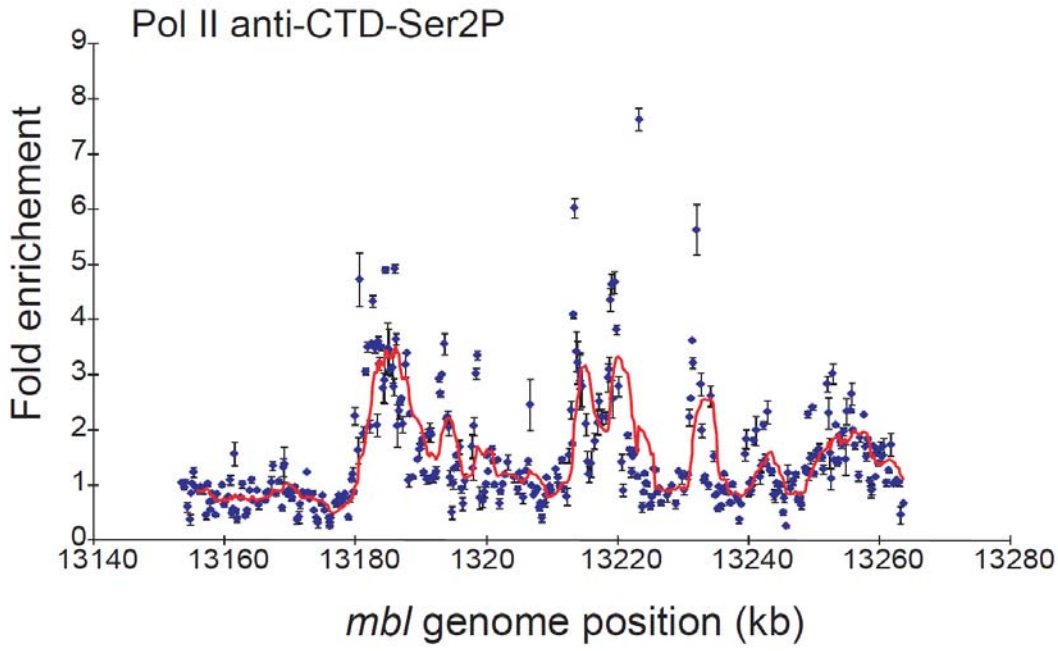
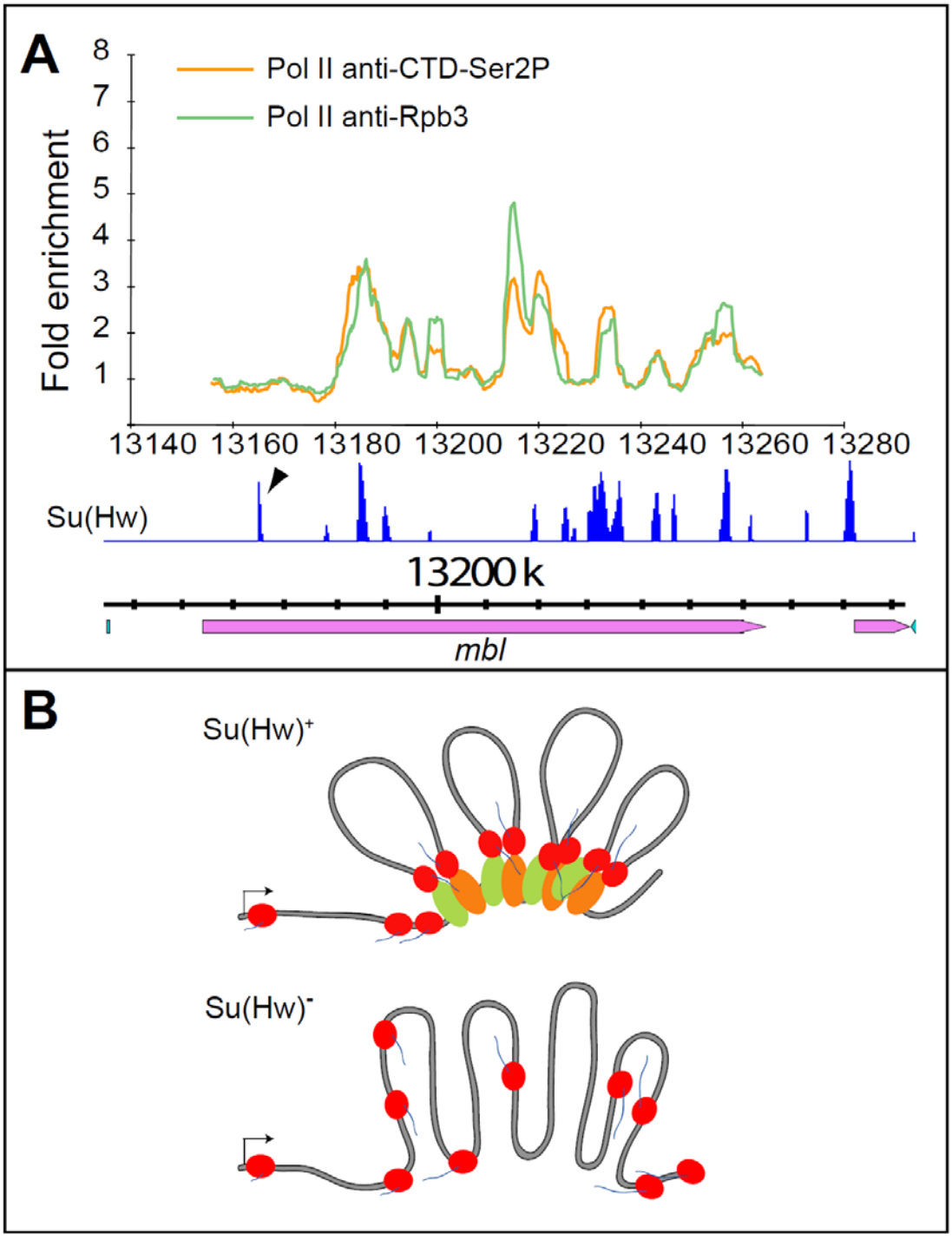

Pol II anti-CTD-Ser2P
Fold enrichement
9
8
7
6
5
4
3
2
1
0
13140
13160
13180
1320
13220
13240
13260
13280
mbl genome position (kb)

Figure 3.30 Intragenic Su(Hw) insulator sites are enriched with RNA polymerase II

A. Ten points moving average of the distribution of Pol II, anti-Rpb3 (green line) and anti-CTD-Ser2P (orange line), using ChIP-chip data (GEO accession GSE6714). RNA PolII peaks at all intragenic insulator sites, except the most 5' site (arrow head). **B**. An interpretation of the data suggesting that insulators form higher-order chromatin structures mediated by protein interactions that prevent free passage of PolII , and reducing the amount of basal transcription. Mutations in insulator proteins $(Su(Hw)^-)$ eliminate chromatin loops and allow a higher level of transcription. Green and orange ovals are insulator proteins. Red ovals represent RNA pol II.


A
Pol II anti-CTD-Ser2P
Pol II anti-Rpb3
Fold enrichment
8
7
6
5
4
3
2
1
13140 13160 13180 13200 13220 13240 13260 13280
Su(Hw)
13200 k
mbl
B
Su(Hw)+
Su(Hw)-

Chapter IV

Discussion

Chromatin insulators are traditionally viewed as elements that function by blocking enhancer activity and spreading of heterochromatin. However, evidence directly addressing endogenous insulator function is very limited, and derives mostly from transgenic assays, leaving open the possibility that insulators could perform functions that are not directly predicted from the well- established properties mentioned above. In addition, whereas most evidence supporting that endogenous insulators function as enhancer-blockers and heterochromatin barriers comes from studies of CTCF in vertebrates, evidence from Drosophila, where as many as 6 different insulator proteins are known, is lacking. Although it has been shown that loss of insulator proteins in Drosophila has important positive and negative effects on transcription, the specific function and role of particular insulator sites in the Drosophila genome remains speculative (Muller et al., 2001; Parnell et al., 2006; Soshnev et al., 2008).

In this work, we have analyzed the effect of the loss of insulator proteins on the transcription levels of long Drosophila genes containing intragenic insulator sites. The following conclusions can be drawn from our results: 1) Su(Hw)

intragenic insulators repress basal and active transcription levels of *mbl* and *Sdc*, two long Drosophila genes that contain more than 10 intragenic insulator sites, but do not affect transcription levels of *Treh*, an inducible gene that contains two Su(Hw) intragenic insulators. Repression of transcription in *mbl* and *Sdc* by Su(Hw) should be independent of the enhancer-blocking activity of the insulators, since it is observed also when *mbl* transcription is induced ectopically using GAL4-UAS binary system. 2) Su(Hw) intragenic insulators could play a role in pre-mRNA processing during transcription, given that the *mblD* isoform resulting from an early termination site is enriched up to 8-fold after loss of Mod(mdg4)67.2. The same isoform is depleted by more than 20% after loss of Su(Hw). 3) *mbl* DNA is organized into a condensed chromatin domain that is defined along the chromatin fiber with differentially distributed insulator proteins. 4) Intragenic Su(Hw) insulators help maintain high accessibility to chromatin in the *mbl* locus, given that loss of insulator proteins systematically reduces accessibility to DNAase I in *mbl* intragenic sequences. Interestingly, accessibility correlates negatively with transcriptional activity of *mbl*. 5) Intragenic Su(Hw) insulators participate in long range interactions, mediating the formation of higher-order chromatin structures both within and outside of the *mbl* gene. 6) Serine2 phosphorylated PolII accumulates at insulator sites, suggesting that insulators function as barriers that interrupt passage of RNA polymerase and slow down transcriptional elongation.

Repression of transcription and higher-order chromatin structure.

Insulators can potentially perform different functions, depending on their location relative to genes. Here we have shown that intragenic insulators have a repressive effect on transcription of *mbl* and *Sdc*, but have no effect on *Treh*. These differences likely correspond to structural and regulatory peculiarities of these genes: for example, whereas *mbl* and *Sdc* are extremely long, highly regulated tissue-specific genes that form condensed bands in tissues where their expression is off, *Treh* is an inducible gene with shorter introns that is located in an interband . In addition, *Treh* has multiple transcription initiation sites that are separated by Su(Hw) insulator sites, which unlike in *mbl* are also occupied by Cp190, and CTCF proteins, suggesting that these insulator sites may have an alternative role in promoter selection or mRNA processing with no consequences on transcription levels.

Several possibilities may explain the repressive effects of Su(Hw) insulators on the expression of *mbl* and *Sdc*, but perhaps the most likely derives from the insulator boundary activity. Insulators can potentially interfere with the progression of PolII and therefore obstruct elongation, reducing the overall transcription rate of a sequence. For example, in the B-Globin locus, insulators

have the ability to block PolII when it progresses from the LCR to the promoters of the globin genes (Zhao and Dean, 2004). In addition, it has been shown that PolII is systematically stalled at intragenic CTCF sites in transcribing inducible long mammalian genes (Wada et al., 2009). More interestingly, a recent finding describing the molecular mechanisms regulating expression of PUMA, a gene that triggers apoptosis in mammals and is activated by p53, shows that intragenic CTCF insulators block PolII elongation, preventing high levels of basal transcription in the absence of p53 activation. Our results suggest that *mbl* is likely regulated by a related mechanism, in which intragenic Su(Hw) insulators would modulate PolII elongation, repressing basal transcription by approximately 2-fold, and reducing active transcription by a factor of up to 5-fold.

mbl, as well as *Sdc*, are found within condensed chromatin bands in polytene chromosomes, in which transcription is highly repressed. Remarkably, our results show that, contrary to what it is normally assumed, the mechanism of transcriptional repression at *mbl* by Su(Hw) insulators appears to be independent of accessibility to DNA. DNAseI accessibility experiments performed in this work have shown that chromatin at *mbl* is actually more accessible to DNAseI in wild type, when transcription is strongly repressed, than in the background of *su(Hw)* and *mod(mdg4)67.2* mutants, when transcription can be up to 5-fold stronger. This result suggests that in *mbl* accessibility to DNA is uncoupled from

transcriptional activity, and also points to a mechanism consisting of interference of PolII elongation, rather than a mechanism preventing accessibility and PolII initiation events. In PUMA, the first 6 kb of the gene is under continuous transcriptional activity, and chromatin does not appear to oppose particular resistance to transcription initiation in the absence of p53. Transcription in PUMA is actually suppressed by an intragenic CTCF insulator that prevents elongation (Gomes and Espinosa, 2010). Our results suggest that intragenic Su(Hw) insulators in *mbl* and *Sdc* could play a role in Drosophila that may be analogous to that of CTCF in PUMA.

Our data indicate that Su(Hw) does not have a direct role in the formation of bands of condensed chromatin in polytene chromosomes. In fact, insulators are not required for the maintenance of condensed bands in chromosomes, given that condensation of chromatin in these bands appears normal at the fluorescence microscope level in $su(Hw)^{e04061}$ and $mod(mdg4)^{u1}$ mutant larvae in salivary glands. Since insulators exert a strong repressive effect on transcription of *mbl*, and given our data suggesting that repression is independent of accessibility, it can be argued that the higher-order chromatin structures mediated by insulators are independent of the chromatin condensation at the chromatin band and are responsible for the repression of transcription (Figure 3.17). The lack of correlation between accessibility and transcription, as well as the accumulation of

PolII at intragenic insulator sites, suggest that elongation is the rate limiting step by which insulators repress transcription at *mbl*, and that long-range interactions between insulator sites might be required to prevent Pol II from engaging in productive elongation. Figure 3.30 B illustrates a model in which insulators function as repressors of elongating polymerases. In this model, higher-order structures mediated by insulators block polymerases, reducing basal transcription rate, which is increased after the loss of insulator proteins.

Su(Hw) insulators and mRNA processing

A major question related to the association of insulators to intragenic sequences is the specific role that they play during elongation. Evidence that Su(Hw) can have a role in the regulation of cotranscriptional processes in *mbl* is revealed by our observation in this study that loss of Su(Hw) insulator proteins changed the ratio between different *mbl* isoforms. Remarkably, the frequency of *mblD* increased by 8-fold in the absence of the Mod(mdg4)67.2 protein, but decreased more than 20% after loss of Su(Hw). The mechanism by which Su(Hw) controls the relative frequency of *mbl* isoforms remains unknown, and a mechanism involving mRNA stability cannot be discarded with our data. However, given that factors controlling PolII elongation can potentially increase the frequency of termination of transcription (de la Mata et al., 2003), the simplest mechanism to explain our

observations is that stalling or slowing down of PolII at insulator sites increases the chances of transcription termination at nearby termination sites. This hypothesis is supported by the observation that, in *su(Hw)* mutants, in which the insulator activity is missing, the frequency of early termination events producing *mblD* is significantly reduced. We cannot disregard the possibility, however, that direct or indirect interactions between transcription termination machinery and components of the *gypsy* insulator, such as Mod(mdg4) 67.2, could have inhibitory effects on the process of transciption termination. This possibility is strongly supported by previous results describing genetic interactions between *gypsy* insulators and *suppressor of forked* [*su(f)*] (Depken and Schiessel, 2009; Hoover et al., 1992; Kosak et al., 2002; Routh et al., 2008; Rutledge et al., 1988). *su(f)* is the *Drosophila* homolog of the human Cleavage stimulation Factor-77 (CstF-77), which is required for cleavage and subsequent polyadenylation of pre-mRNA in eukaryotes (Hockert et al.; Luger and Hansen, 2005). Genetic interactions between CstF-77 and Su(Hw) were uncovered after analyzing mutational insertions of the *gypsy* retrotransposon into introns of genes such as *forked* (*f*), which led to the conclusion that insertions were disrupting normal transcription by inducing early termination events at polyA sites located at the LTRs of the retrotransposon (Benoit et al., 2002; Pan et al., 2006; Rutledge et al., 1988). Termination events at these sites were suppressed either by mutations in *su(f)* or *su(Hw)* by a mechanism that is still unclear, but is linked to the presence of

binding sites for the Su(Hw) insulator protein in the *gypsy* insulator (Kepper et al., 2008). Interestingly, a strong Su(Hw) binding site exist near the *mblD* specific early termination site (Figure 3.8 A). This site was independently identified by ModEncode and the Corces laboratory using ChIP-on-chip and was confirmed by us using ChIP (Figure 3.1). Such interactions could explain our observation of a significant enrichment of *mblD* in the absence of Mod(mdg4) 67.2. These mechanisms are not mutually exclusive, and changes in the ratio among *mbl* isoforms may occur in response to changes in the function of Su(Hw) due to a combination of PolII stalling effects, as well as in response to possible interactions between insulator proteins and elements of the transcription termination machinery.

Recent findings that more than 40% of endogenous insulators occur in intragenic sequences (Bushey et al., 2009; Negre et al., 2010), suggests the possibility that insulators can be strongly involved in processes directly related with transcription elongation and/or mRNA maturation. We previously reported that in polytene chromosomes the majority of DAPI bands correspond to long genes containing intragenic Su(Hw) insulator sites, which suggests that Su(Hw) may have a role in the repression of transcription of a large number of genes in the Drosophila genome that share similar characteristics (di Bari et al., 2006). Data presented here provides evidence suggesting that chromatin organization

mediated by insulators in these genes may function by actively repressing and modulating the elongation process. Our data is particularly interesting in light of the increasing evidence revealing the importance of the role of elongation in transcriptional regulation (Dorigo et al., 2004). For example, in humans and in Drosophila, more than 30% of the genes have paused polymerases in the promoter region, and bivalent genes in mouse ES cells are transcriptionally engaged by PolII, which is paused in the middle of the gene (Fan et al., 2004; Mergell et al., 2004; Mozziconacci and Victor, 2003; Walker and Sikorska, 1987a, b). Many of these genes are important developmental genes in which transcriptional activation depends on the activation of elongation, rather than activation of initiation, as seems to be the case in PUMA and perhaps *mbl* and *Sdc*.

Transcription in eukaryotes requires numerous processes, including capping, splicing, cleavage/polyadenylation as well as chromatin remodeling, histone modifications, and mRNA transport, taking place in a cotranscriptional manner that depends on interactions of a multitude of factors with PolII (Woodcock, 1994). Many of these factors correspond to classic chromatin remodeling complexes such as SWI/SNF, which in addition to a role in regulating chromatin accessibility in promoters, is also involved in elongation and regulation of splicing (Baudy and Bram, 1978; Horowitz et al., 1997; McBryant et al., 2008).

Findings in this work provide evidence for the first time connecting chromosome structure, insulators and higher-order chromatin organization, as new intragenic factors that play a role in the regulation of gene transcription and mRNA maturation in Drosophila. This prospect is supported by the recent observation that Enhancer of Yellow 2 (ENY2), which was previously shown to interact with Su(Hw), is a component of the TOH complex, which is in turn involved in transcription elongation, mRNA biogenesis, and mRNA export (Yao et al., 1993). Connections between factors such as chromosome structure, higher-order chromatin organization and insulators with elements of transcription elongation such as histone modifications, nucleosome remodeling, splicing, termination and export are currently very vague, making further analysis of the relationship between all these factors necessary to fully understand the complexity of transcription in eukaryotes.

List of References

Adkins, M.W., Howar, S.R., and Tyler, J.K. (2004). Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. Mol Cell *14*, 657-666.

Anatskaya, O.V., and Vinogradov, A.E. (2007). Genome multiplication as adaptation to tissue survival: evidence from gene expression in mammalian heart and liver. Genomics *89*, 70-80.

Anderson, S.J., Sikes, M.L., Zhang, Y., French, S.L., Salgia, S., Beyer, A.L., Nomura, M., and Schneider, D.A. (2011). The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. J Biol Chem.

Aoyagi, S., and Archer, T.K. (2008). Dynamics of coactivator recruitment and chromatin modifications during nuclear receptor mediated transcription. Mol Cell Endocrinol *280*, 1-5.

Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci *25*, 421-423.

Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Muller, M., Shanower, G., Schedl, P., Gyurkovics, H., and Karch, F. (2000). 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 *127*, 779-790.

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

Baudy, P., and Bram, S. (1978). Chromatin fiber dimensions and nucleosome orientation: a neutron scattering investigation. Nucleic Acids Res *5*, 3697-3714.

Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Mlodzik, M. (1997). muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development *124*, 4321-4331.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Bellen, H.J., Levis, R.W., Liao, G., He, Y., Carlson, J.W., Tsang, G., Evans-Holm, M., Hiesinger, P.R., Schulze, K.L., Rubin, G.M.*, et al.* (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics *167*, 761-781.

Belozerov, V.E., Majumder, P., Shen, P., and Cai, H.N. (2003). A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. EMBO J *22*, 3113-3121.

Benoit, B., Juge, F., Iral, F., Audibert, A., and Simonelig, M. (2002). Chimeric

human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. Proc Natl Acad Sci U S A *99*, 10593-10598.

Biancotto, C., Frige, G., and Minucci, S. (2010). Histone modification therapy of cancer. Adv Genet *70*, 341-386.

Biggar, S.R., and Crabtree, G.R. (2000). Chemically regulated transcription factors reveal the persistence of repressor-resistant transcription after disrupting activator function. J Biol Chem *275*, 25381-25390.

Brasset, E., and Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. Heredity *94*, 571-576.

Brodolin, K., Zenkin, N., and Severinov, K. (2005). Remodeling of the sigma70 subunit non-template DNA strand contacts during the final step of transcription initiation. J Mol Biol *350*, 930-937.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A *99 Suppl 4*, 16433-16437.

Bushey, A.M., Ramos, E., and Corces, V.G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. Genes Dev *23*, 1338-1350.

Cai, H.N., and Shen, P. (2001). Effects of cis arrangement of chromatin insulators

on enhancer-blocking activity. Science *291*, 493-495.

Chen, N., and Stein, L.D. (2006). Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans. Genome Res *16*, 606-617.

Chen, S., and Corces, V.G. (2001). The gypsy insulator of Drosophila affects chromatin structure in a directional manner. Genetics *159*, 1649-1658.

Chen, Y., Dhupelia, A., and Schoenherr, C.J. (2009). The Igf2/H19 imprinting control region exhibits sequence-specific and cell-type-dependent DNA methylation-mediated repression. Nucleic Acids Res *37*, 793-803.

Chodagam, S., Royou, A., Whitfield, W., Karess, R., and Raff, J.W. (2005). The centrosomal protein CP190 regulates myosin function during early Drosophila development. Curr Biol *15*, 1308-1313.

Chung, J.H., Whiteley, M., and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505-514.

Corces, V.G., and Geyer, P.K. (1991). Interactions of retrotransposons with the host genome: the case of the gypsy element of Drosophila. Trends Genet *7*, 86-90.

Courey, A.J., Plon, S.E., and Wang, J.C. (1986). The use of psoralen-modified DNA to probe the mechanism of enhancer action. Cell *45*, 567-574.

Cuvier, O., Hart, C.M., and Laemmli, U.K. (1998). Identification of a class of

chromatin boundary elements. Mol Cell Biol *18*, 7478-7486.

D'Orso, I., and Frankel, A.D. (2010). RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol *17*, 815-821.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

Deato, M.D., and Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. Genes Dev *21*, 2137-2149.

Demakov, S., Gortchakov, A., Schwartz, Y., Semeshin, V., Campuzano, S., Modolell, J., and Zhimulev, I. (2004). Molecular and genetic organization of Drosophila melanogaster polytene chromosomes: evidence for two types of interband regions. Genetica *122*, 311-324.

Depken, M., and Schiessel, H. (2009). Nucleosome shape dictates chromatin fiber structure. Biophys J *96*, 777-784.

di Bari, M.G., Ciuffini, L., Mingardi, M., Testi, R., Soddu, S., and Barila, D. (2006). c-Abl acetylation by histone acetyltransferases regulates its nuclear-cytoplasmic localization. EMBO Rep *7*, 727-733.

Dilworth, F.J., and Chambon, P. (2001). Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. Oncogene *20*, 3047-3054.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Dunn, K.L., and Davie, J.R. (2003). The many roles of the transcriptional regulator CTCF. Biochem Cell Biol *81*, 161-167.

Dynlacht, B.D. (1997). Regulation of transcription by proteins that control the cell cycle. Nature *389*, 149-152.

ElGIN, L.L.W.M.J.S.S.C.R. (1998). Mapping chromatin structure in Drosophila (Oxford, Oxford University Press).

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Engel, N., and Bartolomei, M.S. (2003). Mechanisms of insulator function in gene regulation and genomic imprinting. Int Rev Cytol *232*, 89-127.

Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. Mol Cell *16*, 655-661.

Felsenfeld, G., and McGhee, J.D. (1986). Structure of the 30 nm chromatin fiber. Cell *44*, 375-377.

Ferguson-Smith, A.C., and Surani, M.A. (2001). Imprinting and the epigenetic

asymmetry between parental genomes. Science *293*, 1086-1089.

Fernandez-Costa, J.M., Llamusi, M.B., Garcia-Lopez, A., and Artero, R. (2011). Alternative splicing regulation by Muscleblind proteins: from development to disease. Biol Rev Camb Philos Soc.

Fourel, G., Revardel, E., Koering, C.E., and Gilson, E. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. EMBO J *18*, 2522-2537.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-713.

Gaszner, M., Vazquez, J., and Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. Genes Dev *13*, 2098-2107.

Gause, M., Morcillo, P., and Dorsett, D. (2001). Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. Mol Cell Biol *21*, 4807-4817.

Gdula, D.A., and Corces, V.G. (1997). Characterization of functional domains of the su(Hw) protein that mediate the silencing effect of mod(mdg4) mutations. Genetics *145*, 153-161.

Georgiev, P., and Kozycina, M. (1996). Interaction between mutations in the suppressor of Hairy wing and modifier of mdg4 genes of Drosophila

melanogaster affecting the phenotype of gypsy-induced mutations. Genetics *142*, 425-436.

Gerasimova, T.I., Byrd, K., and Corces, V.G. (2000). A chromatin insulator determines the nuclear localization of DNA. Mol Cell *6*, 1025-1035.

Gerasimova, T.I., and Corces, V.G. (1998). Polycomb and trithorax group proteins mediate the function of a chromatin insulator. Cell *92*, 511-521.

Gerasimova, T.I., and Corces, V.G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annu Rev Genet *35*, 193-208.

Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O., and Corces, V.G. (1995). A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell *82*, 587-597.

Gerasimova, T.I., Lei, E.P., Bushey, A.M., and Corces, V.G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. Mol Cell *28*, 761-772.

Geyer, P.K., and Corces, V.G. (1992). DNA position-specific repression of transcription by a Drosophila zinc finger protein. Genes Dev *6*, 1865-1873.

Geyer, P.K., Green, M.M., and Corces, V.G. (1988). Mutant gene phenotypes mediated by a Drosophila melanogaster retrotransposon require sequences homologous to mammalian enhancers. Proc Natl Acad Sci U S A *85*, 8593-8597.

Ghosh, D., Gerasimova, T.I., and Corces, V.G. (2001). Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. EMBO J *20*, 2518-2527.

Giardina, C., Perez-Riba, M., and Lis, J.T. (1992). Promoter melting and TFIID complexes on Drosophila genes in vivo. Genes Dev *6*, 2190-2200.

Gilchrist, D.A., Fargo, D.C., and Adelman, K. (2009). Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. Methods *48*, 398-408.

Goers, E.S., Voelker, R.B., Gates, D.P., and Berglund, J.A. (2008). RNA binding specificity of Drosophila muscleblind. Biochemistry *47*, 7284-7294.

Goll, M.G., and Bestor, T.H. (2002). Histone modification and replacement in chromatin activation. Genes Dev *16*, 1739-1742.

Golovnin, A., Biryukova, I., Romanova, O., Silicheva, M., Parshikov, A., Savitskaya, E., Pirrotta, V., and Georgiev, P. (2003). An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in Drosophila. Development *130*, 3249-3258.

Gomes, N.P., and Espinosa, J.M. (2010). Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. Genes Dev *24*, 1022-1034.

Grewal, S.I., and Elgin, S.C. (2002). Heterochromatin: new possibilities for the

inheritance of structure. Curr Opin Genet Dev *12*, 178-187.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35-46.

Gurudatta, B.V., and Corces, V.G. (2009). Chromatin insulators: lessons from the fly. Brief Funct Genomic Proteomic *8*, 276-282.

Harrison, D.A., Gdula, D.A., Coyne, R.S., and Corces, V.G. (1993). A leucine zipper domain of the suppressor of Hairy-wing protein mediates its repressive effect on enhancer function. Genes Dev *7*, 1966-1978.

Hockert, J.A., Yeh, H.J., and MacDonald, C.C. The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation. J Biol Chem *285*, 695-704.

Holt, I., Jacquemin, V., Fardaei, M., Sewry, C.A., Butler-Browne, G.S., Furling, D., Brook, J.D., and Morris, G.E. (2009). Muscleblind-like proteins: similarities and differences in normal and myotonic dystrophy muscle. Am J Pathol *174*, 216-227.

Hoover, K.K., Gerasimova, T.I., Chien, A.J., and Corces, V.G. (1992). Dominant effects of suppressor of Hairy-wing mutations on gypsy-induced alleles of forked and cut in Drosophila melanogaster. Genetics *132*, 691-697.

Hore, T.A., Deakin, J.E., and Marshall Graves, J.A. (2008). The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. PLoS Genet *4*, e1000169.

Horowitz, R.A., Koster, A.J., Walz, J., and Woodcock, C.L. (1997). Automated electron microscope tomography of frozen-hydrated chromatin: the irregular three-dimensional zigzag architecture persists in compact, isolated fibers. J Struct Biol *120*, 353-362.

Imhof, A., and Becker, P.B. (2001). Modifications of the histone N-terminal domains. Evidence for an "epigenetic code"? Mol Biotechnol *17*, 1-13.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiang, F., Mannervik, B., and Bergman, B. (1997). Evidence for redox regulation of the transcription factor NtcA, acting both as an activator and a repressor, in the cyanobacterium Anabaena PCC 7120. Biochem J *327 (Pt 2)*, 513-517.

Jiang, Y., Smale, S.T., and Gralla, J.D. (1993). A common ATP requirement for open complex formation and transcription at promoters containing initiator or TATA elements. J Biol Chem *268*, 6535-6540.

Johnson, K.G., Ghose, A., Epstein, E., Lincecum, J., O'Connor, M.B., and Van Vactor, D. (2004). Axonal heparan sulfate proteoglycans regulate the distribution and efficiency of the repellent slit during midline axon guidance. Curr Biol *14*, 499-504.

Kacem, S., and Feil, R. (2009). Chromatin mechanisms in genomic imprinting. Mamm Genome *20*, 544-556.

Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of

transcription factors and chromatin-modifying machines. Cell *92*, 307-313.

Kadonaga, J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247-257.

Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489-493.

Kaplan, C.D., Morris, J.R., Wu, C., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes Dev *14*, 2623-2634.

Kellogg, D.R., Field, C.M., and Alberts, B.M. (1989). Identification of microtubule-associated proteins in the centrosome, spindle, and kinetochore of the early Drosophila embryo. J Cell Biol *109*, 2977-2991.

Kellum, R., and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64*, 941-950.

Kepper, N., Foethke, D., Stehr, R., Wedemann, G., and Rippe, K. (2008). Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. Biophys J *95*, 3692-3705.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Klemm, R.D., Goodrich, J.A., Zhou, S., and Tjian, R. (1995). Molecular cloning

and expression of the 32-kDa subunit of human TFIID reveals interactions with VP16 and TFIIB that mediate transcriptional activation. Proc Natl Acad Sci U S A *92*, 5788-5792.

Kobayashi, N., Boyer, T.G., and Berk, A.J. (1995). A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. Mol Cell Biol *15*, 6465-6473.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science *296*, 158-162.

Krajewska, W.M. (1992). Regulation of transcription in eukaryotes by DNA-binding proteins. Int J Biochem *24*, 1885-1898.

Kruithof, M., Chien, F.T., Routh, A., Logie, C., Rhodes, D., and van Noort, J. (2009). Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. Nat Struct Mol Biol *16*, 534-540.

Kuhn-Parnell, E.J., Helou, C., Marion, D.J., Gilmore, B.L., Parnell, T.J., Wold, M.S., and Geyer, P.K. (2008). Investigation of the properties of non-gypsy suppressor of hairy-wing-binding sites. Genetics *179*, 1263-1273.

Kuhn, E.J., and Geyer, P.K. (2003). Genomic insulators: connecting properties to mechanism. Curr Opin Cell Biol *15*, 259-265.

Kyrchanova, O., Ivlieva, T., Toshchakov, S., Parshikov, A., Maksimenko, O., and Georgiev, P. (2010). Selective interactions of boundaries with upstream

region of Abd-B promoter in Drosophila bithorax complex and role of dCTCF in this process. Nucleic Acids Res.

Labrador, M., and Corces, V.G. (2002). Setting the boundaries of chromatin domains and nuclear organization. Cell *111*, 151-154.

Langelier, M.F., Forget, D., Rojas, A., Porlier, Y., Burton, Z.F., and Coulombe, B. (2001). Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II. J Biol Chem *276*, 38652-38657.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lee, Y.C., and Langley, C.H. (2010). Transposable elements in natural populations of Drosophila melanogaster. Philos Trans R Soc Lond B Biol Sci *365*, 1219-1228.

Lei, E.P., and Corces, V.G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. Nat Genet *38*, 936-941.

Lercher, M.J., Urrutia, A.O., and Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet *31*, 180-183.

Li, R., Knight, J.D., Jackson, S.P., Tjian, R., and Botchan, M.R. (1991). Direct

interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell *65*, 493-505.

Lo, W.S., Henry, K.W., Schwartz, M.F., and Berger, S.L. (2004). Histone modification patterns during gene activation. Methods Enzymol *377*, 130-153.

Lorenc, A., and Makalowski, W. (2003). Transposable elements and vertebrate protein diversity. Genetica *118*, 183-191.

Lue, N.F., Chasman, D.I., Buchman, A.R., and Kornberg, R.D. (1987). Interaction of GAL4 and GAL80 gene regulatory proteins in vitro. Mol Cell Biol *7*, 3446-3451.

Luger, K., and Hansen, J.C. (2005). Nucleosome and chromatin fiber dynamics. Curr Opin Struct Biol *15*, 188-196.

Marciniak, R.A., and Sharp, P.A. (1991). HIV-1 Tat protein promotes formation of more-processive elongation complexes. EMBO J *10*, 4189-4196.

McArthur, M., and Bibb, M. (2006). In vivo DNase I sensitivity of the Streptomyces coelicolor chromosome correlates with gene expression: implications for bacterial chromosome structure. Nucleic Acids Res *34*, 5395-5401.

McArthur, M., Gerum, S., and Stamatoyannopoulos, G. (2001). Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. J Mol Biol *313*, 27-34.

McBryant, S.J., Krause, C., Woodcock, C.L., and Hansen, J.C. (2008). The silent information regulator 3 protein, SIR3p, binds to chromatin fibers and assembles a hypercondensed chromatin architecture in the presence of salt. Mol Cell Biol *28*, 3563-3572.

Mergell, B., Everaers, R., and Schiessel, H. (2004). Nucleosome interactions in chromatin: fiber stiffening and hairpin formation. Phys Rev E Stat Nonlin Soft Matter Phys *70*, 011915.

Mezey, J.G., Nuzhdin, S.V., Ye, F., and Jones, C.D. (2008). Coordinated evolution of co-expressed gene clusters in the Drosophila transcriptome. BMC Evol Biol *8*, 2.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev *25*, 742-754.

Mizrokhi, L.J., Obolenkova, L.A., Priimagi, A.F., Ilyin, Y.V., Gerasimova, T.I., and Georgiev, G.P. (1985). The nature of unstable insertion mutations and reversions in the locus cut of Drosophila melanogaster: molecular mechanism of transposition memory. EMBO J *4*, 3781-3787.

Modolell, J., Bender, W., and Meselson, M. (1983). Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. Proc Natl Acad Sci U S A *80*, 1678-1682.

Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I.,

Leers, J., White, R.A., Renkawitz-Pohl, R., Saumweber, H., *et al.* (2007). The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. EMBO J *26*, 4203-4214.

Mongelard, F., Labrador, M., Baxter, E.M., Gerasimova, T.I., and Corces, V.G. (2002). Trans-splicing as a novel mechanism to explain interallelic complementation in Drosophila. Genetics *160*, 1481-1487.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., *et al.* (2005). CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep *6*, 165-170.

Mozziconacci, J., and Victor, J.M. (2003). Nucleosome gaping supports a functional structure for the 30nm chromatin fiber. J Struct Biol *143*, 72-76.

Muller, W.G., Walker, D., Hager, G.L., and McNally, J.G. (2001). Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. J Cell Biol *154*, 33-48.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V., and Georgiev, P. (2001). Loss of insulator activity by paired Su(Hw) chromatin insulators. Science *291*, 495-498.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A., *et al.* (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet *6*,

e1000814.

Oliver, D., Sheehan, B., South, H., Akbari, O., and Pai, C.Y. (2010). The chromosomal association/dissociation of the chromatin insulator protein Cp190 of Drosophila melanogaster is mediated by the BTB/POZ domain and two acidic regions. BMC Cell Biol *11*, 101.

Ostashevsky, J.Y., and Lange, C.S. (1994). The 30 nm chromatin fiber as a flexible polymer. J Biomol Struct Dyn *11*, 813-820.

Paal, K., Baeuerle, P.A., and Schmitz, M.L. (1997). Basal transcription factors TBP and TFIIB and the viral coactivator E1A 13S bind with distinct affinities and kinetics to the transactivation domain of NF-kappaB p65. Nucleic Acids Res *25*, 1050-1055.

Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell *16*, 737-748.

Pan, Z., Zhang, H., Hague, L.K., Lee, J.Y., Lutz, C.S., and Tian, B. (2006). An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. Gene *366*, 325-334.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., *et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422-433.

Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S., and Corces, V.G. (1988). The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev *2*, 1205-1215.

Parnell, T.J., Kuhn, E.J., Gilmore, B.L., Helou, C., Wold, M.S., and Geyer, P.K. (2006). Identification of genomic sites that bind the Drosophila suppressor of Hairy-wing insulator protein. Mol Cell Biol *26*, 5983-5993.

Parnell, T.J., Viering, M.M., Skjesol, A., Helou, C., Kuhn, E.J., and Geyer, P.K. (2003). An endogenous suppressor of hairy-wing insulator separates regulatory domains in Drosophila. Proc Natl Acad Sci U S A *100*, 13436-13441.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W., and Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res *33*, 3743-3750.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese, C., and Gatti, M. (1995). Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc Natl Acad Sci U S A *92*, 3804-3808.

Prieto, C., Risueno, A., Fontanillo, C., and De las Rivas, J. (2008). Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. PLoS One *3*, e3911.

Pryde, F.E., and Louis, E.J. (1999). Limitations of silencing at native yeast telomeres. EMBO J *18*, 2538-2550.

Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes. Trends Biochem Sci *30*, 275-279.

Radonjic, M., Andrau, J.C., Lijnzaad, P., Kemmeren, P., Kockelkorn, T.T., van Leenen, D., van Berkum, N.L., and Holstege, F.C. (2005). Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon S. cerevisiae stationary phase exit. Mol Cell *18*, 171-183.

Rapraeger, A.C. (2002). Heparan sulfate-growth factor interactions. Methods Cell Biol *69*, 83-109.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem *282*, 33336-33345.

Robert, F., Forget, D., Li, J., Greenblatt, J., and Coulombe, B. (1996). Localization of subunits of transcription factors IIE and IIF immediately upstream of the

transcriptional initiation site of the adenovirus major late promoter. J Biol Chem *271*, 8517-8520.

Roberts, S.G., Ha, I., Maldonado, E., Reinberg, D., and Green, M.R. (1993). Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. Nature *363*, 741-744.

Roseman, R.R., Johnson, E.A., Rodesch, C.K., Bjerke, M., Nagoshi, R.N., and Geyer, P.K. (1995). A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. Genetics *141*, 1061-1074.

Roseman, R.R., Pirrotta, V., and Geyer, P.K. (1993). The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. EMBO J *12*, 435-442.

Routh, A., Sandin, S., and Rhodes, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc Natl Acad Sci U S A *105*, 8872-8877.

Roux-Rouquie, M., Chauvet, M.L., Munnich, A., and Frezal, J. (1999). Human genes involved in chromatin remodeling in transcription initiation, and associated diseases: An overview using the GENATLAS database. Mol Genet Metab *67*, 261-277.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature *418*, 975-979.

Rutledge, B.J., Mortin, M.A., Schwarz, E., Thierry-Mieg, D., and Meselson, M. (1988). Genetic interactions of modifier genes and modifiable alleles in Drosophila melanogaster. Genetics *119*, 391-397.

Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I., and Groudine, M. (2000). Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev *14*, 940-950.

Schwabish, M.A., and Struhl, K. (2006). Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol Cell *22*, 415-422.

Semeshin, V.F., Artero, R., Perez-Alonso, M., and Shloma, V.V. (1998). Electron microscopic in situ hybridization of digoxigenin-dUTP-labelled DNA probes with Drosophila melanogaster polytene chromosomes. Chromosome Res *6*, 405-410.

Sheldon, L.A., Smith, C.L., Bodwell, J.E., Munck, A.U., and Hager, G.L. (1999). A ligand binding domain mutation in the mouse glucocorticoid receptor functionally links chromatin remodeling and transcription initiation. Mol Cell Biol *19*, 8146-8157.

Shindo, N. (2009). [Histone modification: a new era of targeting epigenetics]. Rinsho Ketsueki *50*, 282-288.

Soshnev, A.A., Li, X., Wehling, M.D., and Geyer, P.K. (2008). Context differences reveal insulator and activator functions of a Su(Hw) binding region. PLoS

Genet *4*, e1000159.

Spana, C., and Corces, V.G. (1990). DNA bending is a determinant of binding specificity for a Drosophila zinc finger protein. Genes Dev *4*, 1505-1515.

Spana, C., Harrison, D.A., and Corces, V.G. (1988). The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev *2*, 1414-1423.

Steigemann, P., Molitor, A., Fellert, S., Jackle, H., and Vorbruggen, G. (2004). Heparan sulfate proteoglycan syndecan promotes axonal and myotube guidance by slit/robo signaling. Curr Biol *14*, 225-230.

Stringer, K.F., Ingles, C.J., and Greenblatt, J. (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. Nature *345*, 783-786.

Suster, M.L., Seugnet, L., Bate, M., and Sokolowski, M.B. (2004). Refining GAL4-driven transgene expression in Drosophila with a GAL80 enhancer-trap. Genesis *39*, 240-245.

Szabo, P.E., Tang, S.H., Silva, F.J., Tsark, W.M., and Mann, J.R. (2004). Role of CTCF binding sites in the Igf2/H19 imprinting control region. Mol Cell Biol *24*, 4791-4800.

Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: jumonji crosses the line. Genes Dev *24*, 1465-1478.

Tang, H., Liu, Y., Madabusi, L., and Gilmour, D.S. (2000). Promoter-proximal

pausing on the hsp70 promoter in Drosophila melanogaster depends on the upstream regulator. Mol Cell Biol *20*, 2569-2580.

Udvardy, A., Maine, E., and Schedl, P. (1985). 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 *185*, 341-358.

van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet *19*, 530-536.

Vicente-Crespo, M., Pascual, M., Fernandez-Costa, J.M., Garcia-Lopez, A., Monferrer, L., Miranda, M.E., Zhou, L., and Artero, R.D. (2008). Drosophila muscleblind is involved in troponin T alternative splicing and apoptosis. PLoS One *3*, e1613.

Vicente, M., Monferrer, L., Poulos, M.G., Houseley, J., Monckton, D.G., O'Dell K, M., Swanson, M.S., and Artero, R.D. (2007). Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. Differentiation *75*, 427-440.

Vorobyeva, N.E., Soshnikova, N.V., Nikolenko, J.V., Kuzmina, J.L., Nabirochkina, E.N., Georgieva, S.G., and Shidlovskii, Y.V. (2009). Transcription coactivator SAYP combines chromatin remodeler Brahma and transcription initiation factor TFIID into a single supercomplex. Proc Natl Acad Sci U S A *106*,

11049-11054.

Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A.*, et al.* (2009). A wave of nascent transcription on activated human genes. Proc Natl Acad Sci U S A *106*, 18357-18361.

Walker, P.R., and Sikorska, M. (1987a). Chromatin structure. Evidence that the 30-nm fiber is a helical coil with 12 nucleosomes/turn. J Biol Chem *262*, 12223-12227.

Walker, P.R., and Sikorska, M. (1987b). Chromatin structure. Further evidence against the existence of a beaded subunit for the 30-nm fiber. J Biol Chem *262*, 12218-12222.

Wallace, H.A., Plata, M.P., Kang, H.J., Ross, M., and Labrador, M. Chromatin insulators specifically associate with different levels of higher-order chromatin organization in Drosophila. Chromosoma *119*, 177-194.

Wallace, H.A., Plata, M.P., Kang, H.J., Ross, M., and Labrador, M. (2010). Chromatin insulators specifically associate with different levels of higher-order chromatin organization in Drosophila. Chromosoma *119*, 177-194.

Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev *17*, 400-407.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes

have an altered conformation. Science *193*, 848-856.

West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev *16*, 271-288.

Williams, K.J. (2001). Interactions of lipoproteins with proteoglycans. Methods Mol Biol *171*, 457-477.

Wittkopp, P.J., Vaccaro, K., and Carroll, S.B. (2002). Evolution of yellow gene regulation and pigmentation in Drosophila. Curr Biol *12*, 1547-1556.

Wong, H., Victor, J.M., and Mozziconacci, J. (2007). An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. PLoS One *2*, e877.

Woodcock, C.L. (1994). Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. J Cell Biol *125*, 11-19.

Woodcock, C.L. (2006). Chromatin architecture. Curr Opin Struct Biol *16*, 213-220.

Workman, J.L. (2006). Nucleosome displacement in transcription. Genes Dev *20*, 2009-2017.

Wu, F.Y., Wang, S.E., Chen, H., Wang, L., Hayward, S.D., and Hayward, G.S. (2004). CCAAT/enhancer binding protein alpha binds to the Epstein-Barr virus (EBV) ZTA protein through oligomeric interactions and contributes to cooperative transcriptional activation of the ZTA promoter through direct

binding to the ZII and ZIIIB motifs during induction of the EBV lytic cycle. J Virol *78*, 4847-4865.

Wu, Y., Reece, R.J., and Ptashne, M. (1996). Quantitation of putative activator-target affinities predicts transcriptional activating potentials. EMBO J *15*, 3951-3963.

Xiao, H., Friesen, J.D., and Lis, J.T. (1994). A highly conserved domain of RNA polymerase II shares a functional element with acidic activation domains of upstream transcription factors. Mol Cell Biol *14*, 7507-7516.

Xiao, L., Kim, M., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. Gene Expr Patterns *6*, 409-419.

Yang, Y., Hu, J.F., Ulaner, G.A., Li, T., Yao, X., Vu, T.H., and Hoffman, A.R. (2003). Epigenetic regulation of Igf2/H19 imprinting at CTCF insulator binding sites. J Cell Biochem *90*, 1038-1055.

Yankulov, K., Blau, J., Purton, T., Roberts, S., and Bentley, D.L. (1994). Transcriptional elongation by RNA polymerase II is stimulated by transactivators. Cell *77*, 749-759.

Yao, J., Lowary, P.T., and Widom, J. (1993). Twist constraints on linker DNA in the 30-nm chromatin fiber: implications for nucleosome phasing. Proc Natl Acad Sci U S A *90*, 9364-9368.

Yusufzai, T.M., and Felsenfeld, G. (2004). The 5'-HS4 chicken beta-globin

insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci U S A *101*, 8620-8624.

Zhao, H., and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res *32*, 4903-4919.

Zhao, K., Hart, C.M., and Laemmli, U.K. (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. Cell *81*, 879-889.

Zhimulev, I.F., Belyaeva, E.S., Semeshin, V.F., Koryakov, D.E., Demakov, S.A., Demakova, O.V., Pokholkova, G.V., and Andreyeva, E.N. (2004). Polytene chromosomes: 70 years of genetic research. Int Rev Cytol *241*, 203-275.

VITA

Shaofei Zhang was born in Baotou, Inner Mongolia, China. He entered University of Science and Technology of China (USTC) in 2001 after graduated from high school. Shaofei received his Bachelor's degree with a major of Biochemistry and Molecular Biology in 2005. In August 2005, he enrolled in the Biochemistry, Cellular and Molecular Biology program as a graduate research assistant. Shaofei was awarded the Wright Research Award in May, 2011. He will start his post-doc research in National Institutes of Health (NIH) from September, 2011.