in situ genetic dissection of the boundary element *Miscadestral pigmentation* in the Bithorax Complex of *Drosophila melanogaster*

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Mario Alessandro Metzler
aus Zürich, Zürich

Basel, 2017

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Markus Affolter
Prof. Dr. François Karch

Basel, den 13. Dezember 2016

Prof. Dr. Jörg Schibler
Dekan

Abstract

In the past, dissection of *cis*-regulatory elements (enhancers, boundary elements, Polycomb Response Elements, etc.) was mostly done using transgenic assays. The study of such elements at the endogenous genomic location was, in many cases, not possible due to technical restrictions. Nowadays, the CRISPR/Cas9 method allows the precise modification, and hence dissection of *cis*-regulatory elements in their natural genomic environment, allowing an unbiased observation of their role in the regulation of their target gene. In this thesis I will describe the dissection of a well studied *cis*-regulatory element at its endogenous genomic location: the *Miscadestral pigmentation* (*Mcp*) boundary in the *Drosophila melanogaster* Bithorax Hox Complex (BX-C).

In *Drosophila melanogaster* the *Miscadestral pigmentation* boundary element and the adjacent Polycomb Response Element divide the *abdominal-A* (*abd-A*) from the *Abdominal-B* (*Abd-B*) gene. Deletion of this elements leads to spurious interaction of *abdominal-A* *cis*-regulatory elements with the *Abdominal-B* gene. Such incorrect interactions of enhancers have to be prevented not only in the Hox cluster but throughout the genome to allow precise gene regulation. The dissection of such an element as a model system can give insights on how the genome is organized in independent gene regions.

I used the CRISPR/Cas9 method in combination with both Non-Homologous-End-Joining and Homologous-Recombination DNA repair mechanisms to induce deletions in the *Mcp* region, and to establish a phiC31-integrase dependent landing site in the *Mcp* genomic location. The established landing site can be used to bring back into

the *Mcp* locus modified genetic material, for example *Mcp* sequences with small deletions or completely unrelated boundary sequences from other genomic loci or species.

The endogenous *Mcp* boundary and the Polycomb Response Element are dissected in great detail, and the obtained results demonstrate that the function of the boundary depends on a single CCCTC-binding factor (CTCF, a known boundary associated factor) site and on four Pleiohomeotic (Pho, part of the Polycomb Protein Group) sites in the Polycomb Response Element. Those results are further refined using classical boundary element assays in transgenic flies. Furthermore, I study the role of the orientation of those regulatory DNA sequences in the endogenous genomic locus.

Surprisingly, some deletions in the *Mcp* region give rise to phenotypes not associated previously with this boundary, suggesting the presence of additional *cis*-regulatory elements other than the already known boundary element and PRE.

This study further demonstrates the versatility of the CRISPR/Cas9 method in studying *cis*-regulatory elements in their endogenous genomic position, a possibility that will enable more precise and clean investigation of gene regulation in the future.

Acknowledgements

"If I have seen further, it is by standing on the shoulders of giants."
Sir Isaac Newton, 1676

In what times I've had the privilege to do studies on genetic regulation! Being in contact with the very people that did so much work on the topic in the past while being able to use the tools a molecular biologist has in his tool set nowadays, CRISPR/Cas9 in the first place.

The first person I have to thank is Dr. Martin Müller. He gave the start impulse to my project, and kept an eye over my chaotic (at least in his eyes) thoughts and experiments. He taught me the strict rules of scientific experimentation and *Drosophila* genetics. Furthermore, he was a big help during the experimentations as well, thanks to his unpaired skills in fly work. Martin was indispensable. It was an honor for me to work with him, a scientist involved in the discovery of the structure of the homeodomain and boundary element long distance interactions. Martin's quote after the establishment of the *Mcp* landing site: "I've waited for those flies since twenty years".

The second person I have to thank is Prof. Dr. François Karch. François was more than just a member of my thesis committee. He was at the very origin of my project, with the cloning of the *Mcp* boundary, and his many efforts and successes deciphering the regulatory logic in the BX-C. François was an invaluable mentor, through his thoughts via e-mail and in person, and in the way he introduced me to other researchers at the *Drosophila* conferences in the States. François's quote after the establishment of the *Mcp* landing site: "Congratulations, I'm jealous". What bigger compliment could I get?

The third person I have to thank is Prof. Dr. Markus Affolter. I have to thank him for having me in his lab, and giving me the freedom to pursue a project that is far from his present main scientific interests. I admire his ability to focus complex knowledge towards simple questions, that are testable in an experimental way. Furthermore, he was able to hire the right people to create an agreeable and stimulating lab environment. Those two qualities were very instructive to me. Markus's quote after I've told him I would go to a non scientific workshop: "How can we motivate PhD students to stay in academia if they are so well informed?".

There is a couple more senior scientists I would like to thank, that together with the three before mentioned ones really gave me the feeling to stand on the shoulders of giants. A thank goes to Prof. Dr. Welcome Bender for the few discussions we had about my work at the American *Drosophila* conferences. Another thank goes to the late Prof. Dr. Walter Gehring for his presence on the second floor and the several lunches, dinners, and discussions we had in Roscoff. Also, I would like to thank Prof. Dr. Michael Levine for accepting my invitation to give a talk at a symposium I've organized and the Postdoc offer he made me (in my second year of Ph.D. studies). Thanks to Prof. Dr. Erik van Nimwegen for the rotation in his lab and the couple of discussions after that when I had some questions about statistical analysis. Last but not least a thank to Dr. Heinz-Georg Belting for discussions about science and life and his challenging questions during lab meetings.

The biggest thank should go to my biggest supporter and critic Beatrice Schibler, thanks for being here for me during all my scientific endeavors and moments away from science. Through good and bad times.

The work at the bench and at the fly station would have been much harder without the support from our floor manager Vaclav Mandak, and the people in the media kitchen Bernadette Bruno, Gina Evora

and Karin Mauro. Therefore, many thanks to them! Further, a big thank to Helen Preiss for remembering our birthdays and taking care of the lab administration (and the coffee machine!).

Thanks to Dr. Dimi Bieli for being a great friend and mentor in the lab, and for helping guiding an Italian through the unknown wildernesses in California and Nevada, watching Arnie driving a tank while sharing a bed in a Californian Motel close to a train track. Also, a big thanks for the corrections made to this thesis. Thanks to Dr. Ilaria Alborelli for being a friend with an open ear in the lab, correcting part of this thesis, and for not driving too much in California. It has been fun having you in the lab and as a friend. Another thanks goes to Dr. Stefan Harmansa for the discussions we had on the travels to the fly meetings in Washington D.C. and Chicago, for corrections on parts of this thesis, and of course for the burgers we shared (and deep dish pizza, but mainly the burgers). Thanks to Dr. Shinya Matsuda for being Shinya.

As I can not go through all lab members of the Affolter lab, I would use the opportunity to thank them all for the big support and the very pleasant atmosphere at work.

Furthermore, I would like to thank all those people that made my years of Ph.D. variated and fun, giving me opportunities to reach further than just my fly station and bench. A thank is due to Angie Klarer for the organization of all the nice trips around the world with the Werner von Siemens Foundation, to take care of the Ph.D. students at the Biozentrum, and the nice chats. Thanks to my Basel carnival clique for the nice time together, a necessary bubble away from work. Moreover, thanks to Dr. Bettina Volz-Tobler and Caroline Mattingley-Scott for the insights they gave me into the university administration due to the photography jobs I could do for them. Thanks to Dr. Sina Heinrichs for the opportunities delivered by the University of Basel Transferable Skills courses. Another thank to Dr. Mario Kaiser and Dr. Corinna Virchow for the involvement into their

start-up magazine project. Thanks also to Dr. Susanne Kudielka for the organization of my art exhibition at the Ramada Plaza Hotel in Basel.

Thanks to the LaTeX community for the great typesetting tools and to Harish Bhanderi and Jakob Suckale for the thesis template.

Last but not least, a big thanks to my parents Barbara and Carlo, and my sisters Selina and Anna for the continuous support. A thank is also due to Beatrice's parents for pleasant moments in Bottmingen and Hasliberg.

Contents

List of Figures **viii**

1 Introduction **1**

1.1 The role of homeotic genes in *Drosophila* development 1

1.1.1 Homeotic genes in *Drosophila melanogaster* 1

1.1.2 Regulation of Hox genes by gap, pair-rule, and segment-polarity genes 3

1.1.3 Gene regulation by Hox proteins in *Drosophila melanogaster* 7

1.1.4 About parasegments and segments in *Drosophila melanogaster* 11

1.1.5 From histoblast nests to adult abdominal segments 12

1.2 Cis-regulatory elements in the Bithorax complex (BX-C) 15

1.2.1 The organization of the Bithorax complex 15

1.2.2 Initiator elements 18

1.2.3 Maintenance elements and Polycomb group proteins 23

1.2.4 Boundary elements 28

1.2.5 Promoter targeting sequences 32

1.2.6 Tissue specific enhancers 32

1.3 The nature of boundary elements 33

1.3.1 Known boundary elements in *Drosophila* 33

1.3.2 Enhancer blocking, barrier function, and long distance interaction 34

1.3.3 The gipsy insulator 35

1.3.4 The scs and scs' boundaries 37

1.4 Boundaries in the BX-C 38

1.4.1 *Fab-7*, the model boundary in the BX-C 38
1.4.2 Other boundaries described in the BX-C 41
1.5 The *Mcp* boundary in the BX-C 42
1.5.1 Deletions and their phenotypes in the *Mcp* region 42
1.5.2 Dissection of the *Mcp* PRE in transgenes 44
1.5.3 Long distance interaction of *Mcp* boundary pairs 45
1.5.4 Directionality of Mcp boundary interactions 47
1.6 Current model of *Abd-B* gene regulation 50
1.7 Methods 52
1.7.1 Use of phiC31 integrase in *Drosophila* 52
1.7.2 The CRISPR/Cas9 method for genome engineering 54

2 Aims of the project 57

3 Results 58
3.1 PAPER MANUSCRIPT: From Blackbox to regulatory logic: *in situ* dissection of the boundary element *Miscadestral pigmentation* in the Bithorax complex of *Drosophila melanogaster* 58
3.1.1 Introduction 58
3.1.2 Materials and Methods 63
3.1.3 Results 66
3.1.4 Discussion 83

4 Supplementary results and discussion 93
4.1 The *MM8* deletion 93
4.2 Transcription through the *Mcp* boundary and PRE pair 94
4.3 Establishment of a second landing site at the *Mcp* genomic location in the BX-C 97

5 Summary and outlook 101

6 Materials & methods 104
6.1 Abdominal cuticle preparations 104
6.1.1 Reagents 104
6.1.2 Protocol 104

6.2 Embryo fixation with Abd-B and Engrailed or Abd-A and Engrailed antibody staining 105
6.2.1 Reagents 105
6.2.2 Protocol 106
6.3 Preparation of electro-competent *E. coli* cells 108
6.3.1 Reagents 108
6.3.2 Protocol 108

7 Copyright disclaimer **109**

Bibliography **110**

List of Figures

1.1 Hox genes and their expression in *Drosophila melanogaster* 3
1.2 Upstream genetic regulation of Hox genes 7
1.3 Role of EXD and HTH in gene regulation 10
1.4 Parasegments and segments 13
1.5 Histoblast nests and their fate in the adult abdomen 16
1.6 Open for business in the BX-C 18
1.7 *iab-6* deletion and replacement 22
1.8 Enhancer from the BX-C and their expression in enhancer reporter experiments 23
1.9 The role of PREs in memory function 26
1.10 Polycomb Group Proteins 29
1.11 Gain- and loss-of-function phenotypes according to the "open for business model" 31
1.12 Boundary element transgenic assays 36
1.13 Dissection of the *Fab-7* boundary 39
1.14 *Mcp* phenotype 43
1.15 Classical *Mcp* deletions 44
1.16 Dissection of the *Mcp* PRE 46
1.17 Pairing sensitivity as a readout for long distance interaction . . . 48
1.18 Orientation dependency of *Mcp* boundary element pairing *in cis* . 49
1.19 Topologically associated domains as an analogy for BX-C DNA loops 52
1.20 CRISPR/Cas9 method representation and applications 56

3.1 Dissection of the endogenous *Mcp* locus using CRISPR/Cas9 with non-homologous-end-joining 68

3.2 Establishment of the phiC31 integrase dependent landing site in the endogenous *Mcp* locus 71
3.3 Deletions in the endogenous *Mcp* boundary and PRE 74
3.4 Dissection of *Mcp* function in the *apterous* locus 78
3.5 Conservation and orientation dependency of the *Mcp* element . . 82
3.6 Working model of *cis*-regulatory interplay at the *Mcp* locus 86
3.7 Breakpoint of deletions in re-entry constructs 88
3.8 Conservation of *Mcp* sequences from differen *Drosophilides* 90

4.1 The *MM8* deletion 95
4.2 Pigmentation phenotypes in flies carrying an inducible promoter in the endogenous *Mcp* genomic location 98
4.3 Establishment of a second landing site in the *Mcp* locus 100

LIST OF FIGURES

1

Introduction

1.1 The role of homeotic genes in *Drosophila* development

1.1.1 Homeotic genes in *Drosophila melanogaster*

"The segmentation pattern of the fly provides a model system for studying how genes control development." E.B. Lewis, 1978

The fruit fly *Drosophila melanogaster* shows a stereotypical segmentation: the body plan is consisting of a head (cephalic segments 1 to 3, or C1-C3), a thorax (thoracic segments 1 to 3, or T1-T3), and an abdomen (abdominal segments 1 to 8, or A1-A8). The segments of those main body structures are established early in embryogenesis and acquire a specific identity.

What defines the identity of a segment? Each segment contains specific structures (such as legs, wings, eyes, antennae) or characteristics (such as pigmentation or hairs), and the identity of a segment is characterized by the arrangement of those structures and characteristics on that specific segment. A first insight about how these segments are specified came from studying mutants in which one segment is transformed into the identity of another one. Such mutations are referred to as homeotic mutations. For example, in wild type *Drosophila* only segment T2 develops wings, while segment T3 develops halteres (structures responsible for balance while flying). Mutant flies were discovered where halteres are transformed into wings (Lewis, 1978).

What causes these particular phenotypes? Genes linked to such transformations were shown to belong to one class of genes which are clustered on the genome. These genes were called Homeotic genes (later shortened to Hox genes) because of their role in such homeotic transformations (Lewis, 1978).

Indeed we have gone a long way since E.B. Lewis published his seminal paper in 1978. Today we understand how these Hox genes are arranged on the chromosomes of flies, and how they are expressed along the body axis in the developing embryo. We also have insight on how the Hox proteins interact with other transcription factors and orchestrate downstream gene activity (see **Section 1.1.3**).

One property of Hox genes is that these genes are arranged on the chromosome in the same order as they are expressed along the anterior/posterior (A/P) body axis. This phenomenon is called collinearity (Deschamps and van Nes, 2005; Kmita and Duboule, 2003). In *Drosophila melanogaster* such collinearity is nicely visible in the developing embryo as illustrated with RNA *in situ* hybridization experiments (**Figure 1.1 A**) and the arrangement of genes on the chromosome (**Figure 1.1 B**). Collinearity does not mean that the expression patterns of Hox genes do not overlap, they indeed do. This is visible in **Figure 1.1 A** where the expression of several Hox genes overlap (note for example how *abd-A* is expressed in regions where either *Abd-B* or *Ubx* are expressed). Collinearity becomes most evident in *Drosophila* embryos when the anterior expression border of each gene is compared to the anterior expression border of the other genes in the Hox cluster. In *Drosophila melanogaster*, there are also exceptions as some former Hox genes have acquired a different function but still reside in the genomic Hox cluster position (*bcd*, *zen* and *ftz*, **Figure 1.1 B**, Lemons and McGinnis (2006)).

What is the role of Hox genes? One of the conserved functions of different Hox genes is to define segmental identity along the anterior-posterior axis (Pearson et al., 2005). Indeed, when one *Drosophila* Hox gene is expressed ectopically, structures are generated that resemble those of the tissues where the protein is usually expressed. One example is the transformation of antennae into legs (Richard and David, 1990; Schneuwly et al., 1987) by ectopic expression of either *Antennapedia* or *Ultrabithorax* in head tissues.

McGinnis and colleagues published an article where they describe that the DNA binding domain (Homeobox) contained in all *Drosophila* Hox genes (Laughon

and Scott, 1984; McGinnis et al., 1984b; Scott and Weiner, 1984) is also present in a whole range of other animals (McGinnis et al., 1984a). This early discovery in the Hox gene field opened a whole new research area: evo-devo. Evo-devo stands for the comparative study of developmental processes and their molecular background throughout different species and families in the animal kingdom.

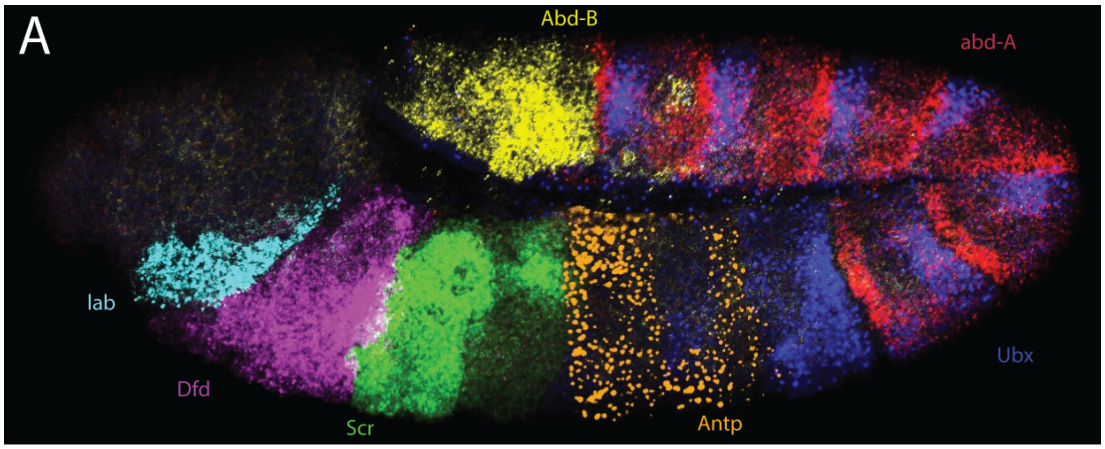
Compared to *Drosophila*, mammals and fish have undergone several rounds of genome and gene duplications, thus they carry multiple Hox clusters (e.g. 4 in mouse, 7 in zebrafish. See **Figure 1.1 B**, Hurley et al. (2005); Maconochie et al. (2003)).

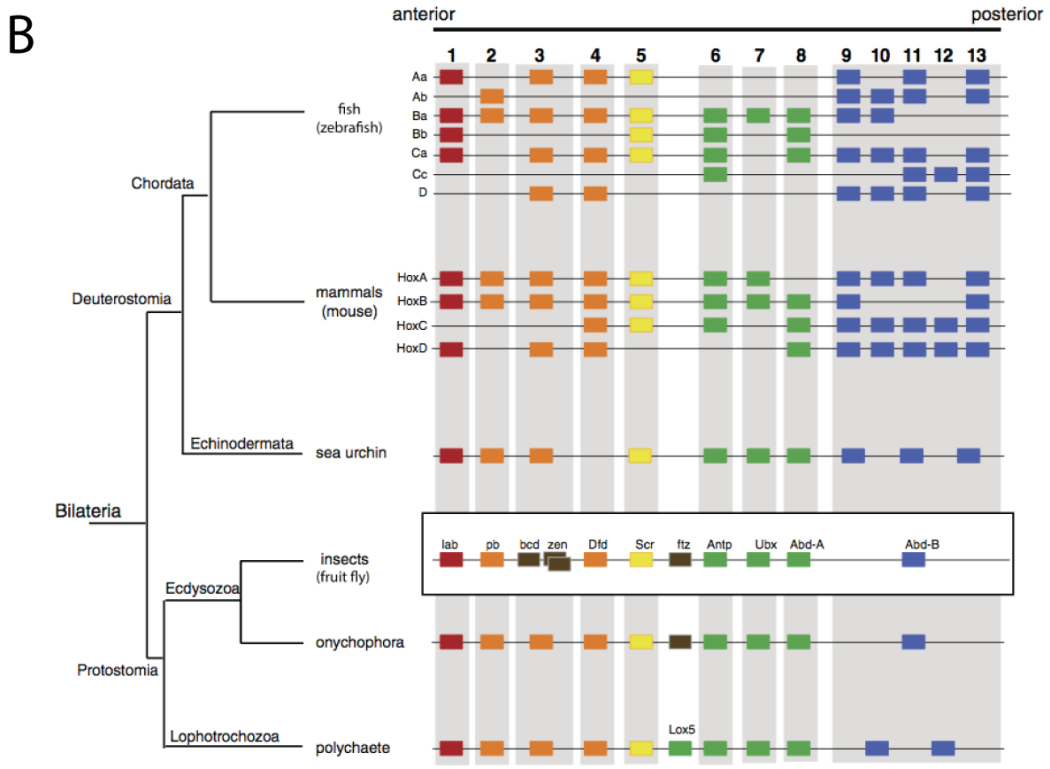
1.1.2 Regulation of Hox genes by gap, pair-rule, and segment-polarity genes

This section will shortly describe the cascade of events that will ultimately lead to the patterning of the *Drosophila* embryo along the anterior/posterior axis.

In *Drosophila melanogaster*, the mother deposits several mRNA molecules (from maternal genes) in the unfertilized eggs in a polarized fashion. For example, bicoid (*bcd)* mRNA is accumulated at the anterior end of the egg (Driever and Nüsslein-Volhard, 1988; Lehmann and Nusslein-Volhard, 1991; Lehmann and

Figure 1.1 *(following page)*: Hox genes and their expression in *Drosophila melanogaster* - A: Expression of Hox genes in *Drosophila melanogaster* embryo (as the embryo is folded on itself during germ band extensions, anterior is on the left and the most posterior tissue is in the middle on top). *lab*: *labial*, *Dfd*: *Deformed*, *Scr*: *Sex combs reduced*, *Antp*: *Antennapedia*, *Ubx*: *Ultrabithorax*, *abd-A*: *abdominal-A*, *Abd-B*: *Abdominal-B*. Modified from (Kosman et al., 2004). B: Organization of Hox clusters in Bilateria. The *Drosophila* Hox gene names and arrangement are marked with a white box. *Drosophila* gene names same as panel (A) plus *proboscipedia* (*pb*) which is not expressed in the embryo and *bicoid* (*bcd*), *zerknullt* (*zen*), and *fushi tarazu* (*ftz*) that are not proper Hox genes as they are not expressed in a collinear way. Same colors are used to mark orthologs, meaning genes that have the same origin. Therefore, *HoxA5* in mammals has the same origin as *Scr* in *Drosophila*. Paralogs are genes that share a common origin an were duplicated during evolution. Therefore, *HoxB9* and *HoxB13* are paralogs in mammals (Koonin, 2005). Modified from (Pick, 2016).


A
Abd-B
abd-A
lab
Dfd
Scr
Antp
Ubx


B
anterior
posterior
1 2 3 4 5 6 7 8 9 10 11 12 13
Aa
Ab
Ba
Bb
Ca
Cc
D
HoxA
HoxB
HoxC
HoxD
lab pb bcd zen Dfd Scr ftz Antp Ubx Abd-A Abd-B
Lox5
fish (zebrafish)
mammals (mouse)
sea urchin
insects (fruit fly)
onychophora
polychaete
Chordata
Deuterostomia
Echinodermata
Bilateria
Ecdysozoa
Protostomia
Lophotrochozoa

Nüsslein-Volhard, 1986; Wang and Lehmann, 1991; Wharton and Struhl, 1991). After fertilization, those mRNAs are translated and protein gradients are established. Those first protein gradients are responsible to regulate a second cohort of factors: the gap genes. Gap genes get their name from mutations that lead to missing segments (or gaps) in the embryo. *cis*-regulatory elements of gap genes read out the concentration of the maternal factors (and other gap gene products), and subsequently the transcription of gap genes is activated only in well defined areas along the anterior/posterior axis of the developing embryo (Akam, 1987). The maternal and gap genes are the framework used by a third generation of regulated gene transcription factors at the end of the blastoderm stage: the pair rule genes (Carroll and Vavra, 1989; Frasch et al., 1988). Pair rule genes are so named because they are expressed in alternate stripes and their deletion leads to embryos with lesions in every other segment. For example, *fushi tarazu* (*ftz*) is expressed where *evenskipped* (*eve*) is not. The classical example of pair rule gene regulation by gap genes is the regulation of the second most anterior *eve* stripe. In the upstream region in the *eve* gene, there are defined enhancer regions, each responsible for a specific stripe. It was shown that several gap gene transcription factors bind to these DNA sequences and either activate or repress that specific enhancer and thus the transcription of *eve*. For *eve* stripe 2, the expression is activated by Bicoid and Hunchback and repressed by Giant and Krüppel (Small et al., 1992).

Pair rule genes are responsible to segment the developing embryo into 14 parasegments (described in more detail in the section about parasegments and segments in this chapter, see **Section 1.1.4**). The pair rule genes as a third tier of regulated genes are responsible for the regulation of a fourth set of genes: the segment polarity genes. The activity of segment polarity genes divides each segment in an anterior and a posterior compartment (DiNardo and O'Farrell, 1987). This step happens during cellularization in the embryo. The first three tiers of transcription factors can still freely diffuse as nuclei are present in a syncytium in the embryo.

Through this cascade of events, each cell in the developing embryo has a defined fate. Is it rather in the anterior or posterior part of the body? Maternal factors give an approximation, and gap genes a more precise spatial information.

Is it in an odd or even numbered segment? Pair rule gene expression will tell. And even more precisely, cells will know if they are in the anterior or posterior part of a single segment due to the expression of segment polarity genes.

As we are speaking about a biological system, it is clear that reality is far from the linearity depicted in the previous paragraphs. Every class of genes described has complex regulatory interactions not only with other gene classes, but also among themselves.

At this point in development, the animal body is segmented and enough information is present for the next class of selector genes to act: the Hox genes. Hox genes are directly regulated by the activity of gap genes (**Figure 1.2 A**, Casares and Sánchez-Herrero (1995); Harding and Levine (1988); Shimell et al. (2000)), pair rule genes (**Figure 1.2 B**, Müller and Bienz (1992)), as well as segment polarity genes (**Figure 1.2 C**, Macías et al. (1994); Mann (1994)), which bind to Hox *cis*-regulatory elements and cooperatively regulate those Hox genes.

In **Figure 1.2** I use data from three publications to illustrate in an exemplary way the inputs into Hox genes of gap genes, pair rule genes, and segment polarity genes. In **Figure 1.2 A**, the input of gap genes on one enhancer of *abd-A* is dissected. This enhancer (*iab-2*) contains binding sites for Hunchback, Giant, and Krüppel proteins. Deletions of the Hunchback sites lead to an expansion of the enhancer activity, while a deletion of the Giant sites has the opposite effect. Deletion of all known gap gene binding sites leads to an expansion of the enhancer activity (Shimell et al., 2000). In **Figure 1.2 B** an enhancer from the *Ultrabithorax* region (*PBX/ABX*) is used to dissect pair rule gene input. In a *ftz* mutant background, only half of the *engrailed* (segment polarity gene) stripes are visible as half of the parasegments are not formed. Furthermore, the enhancer reporter from the *Ubx* region looses all activity, a sign of *ftz* input on *Ubx (Müller and Bienz, 1992).* In **Figure 1.2 C** anti-Ubx staining is used to dissect pair rule input onto *Ubx* gene regulation. When *engrailed* is present *Ubx* is repressed in stripes in the embryo. This repression is not visible anymore in *engrailed* mutant embryos (Mann, 1994).

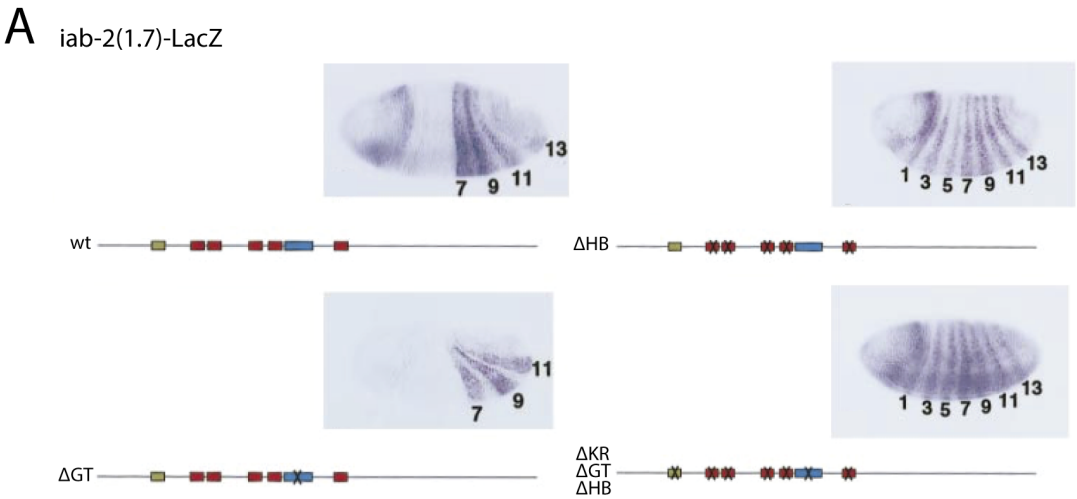
1.1.3 Gene regulation by Hox proteins in *Drosophila melanogaster*

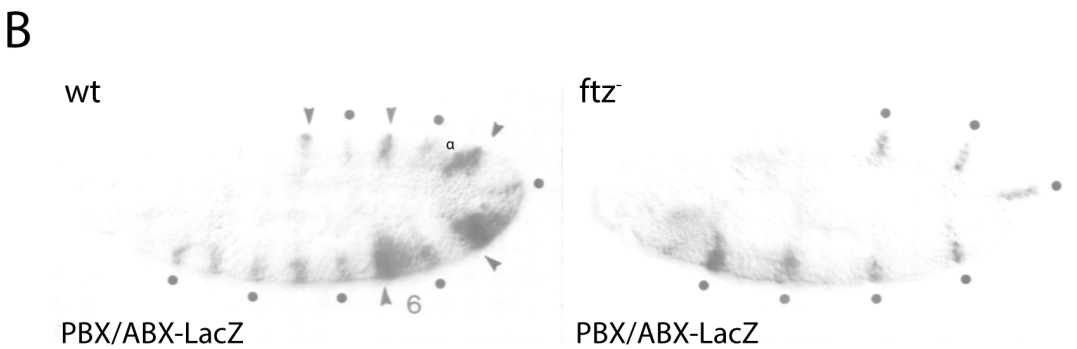
Hox genes are so called selector genes. Selector genes are "master switches" in development. They compute the inputs of previous expression patterns and are responsible, as transcription factors, for the orchestration of the subsequent layers of gene regulation. These subsequent layers of gene regulation are responsible for the definition of tissue identity. Hox regulatory regions process the upstream inputs (described in the previous chapter) and this way regulate the expression of Hox genes.

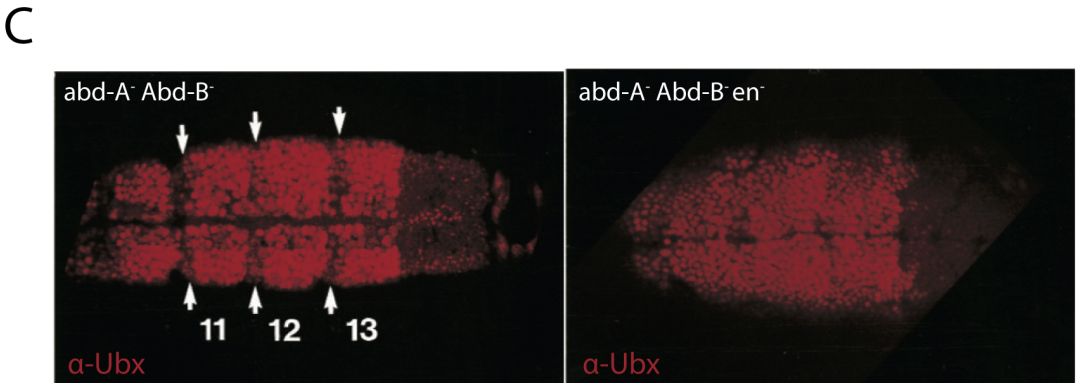
All the genes in the Hox class share a similar DNA binding domain: the homeodomain, consisting of three α-helices connected by loops and an unstructured N-terminal arm. The N-terminal arm interacts with the minor groove on the DNA while the third α-helix interacts with the major groove (Gehring et al., 1994). It is not straightforward to determine which gene is regulated by which Hox gene as all homeodomain proteins recognize virtually the same sequences on the DNA (TAAT) (Berger et al., 2008; Noyes et al., 2008). This was shown *in vitro* and *in vivo* (Affolter et al., 1990; Ekker et al., 1994, 1991; Vachon et al., 1992).

The very similar binding affinity of Hox proteins is puzzling: how would Hox proteins correctly bind to enhancers in tissues where more than one Hox gene is

Figure 1.2 *(following page)*: Upstream genetic regulation of Hox genes - A: Expression pattern of an *abd-A* enhancer fragment (*iab-2(1.7)-LacZ*). In early gastrulation embryonic stages, when binding sites for gap genes are deleted on the reporter the expression is either reduced (in the case of Giant) or expanded (Hunchback and Krüppel). Blue box for Giant (GT), red box Hunchback (HB), yellow box for Krüppel (KR). Modified from (Shimell et al., 2000). B: Expression pattern of the two *Ubx* enhancer regions *PBX* and *ABX* detected with a X-gal staining. In *fushi tarazu* mutants, the expression of this construct is lost (no dark patches in the posterior part of the embryo, only Engrailed signal is visible). Arrowheads mark the presence of X-gal signal, while dots mark Engrailed protein staining. Modified from (Müller and Bienz, 1992). C: *Ultrabithorax* antibody staining in double (*abd-A*, *Abd-B*) and triple (*abd-A*, *Abd-B*, and *en*) mutants. White arrows mark the regions where *Ubx* is repressed by engrailed. In *en* mutant background, this repression ist lost. Modified from (Mann, 1994).


A
iab-2(1.7)-LacZ
7 9 11 13
1 3 5 7 9 11 13
wt
ΔHB
7 9 11
1 3 5 7 9 11 13
ΔGT
ΔKR
ΔGT
ΔHB


B
wt
ftz-
α
6
PBX/ABX-LacZ
α-en
PBX/ABX-LacZ
α-en


C
abd-A- Abd-B-
11 12 13
α-Ubx
abd-A- Abd-B- en-
α-Ubx

expressed? A study on two Hox proteins (Ubx and Antp) led to the discovery of their interaction with Extradenticle (Exd) (Chan et al., 1994). The authors of this study describe how Exd interacts and increases the affinity to a target DNA sequence of Ubx but not Antp protein. Thus, the activity of defined Hox genes depends on the interaction with other transcription factors (Chan et al., 1994). Further studies showed that Exd activity depends on the translocation of the protein to the nucleus. Such a nuclear localization of Exd is dependent on the presence of another homeodomain protein encoded by the *homothorax* (*hth*) gene. Exd/Hth cooperatively bind to DNA with a Hox protein (**Figure 1.3 A,B**, Pai et al. (1998); Rieckhof et al. (1997); Ryoo et al. (1999)). A reduction of Hth activity in abdominal segments 2, 3, and 4 leads to a gain of dark pigmentation on tergites A2 to A4. In contrast, overexpression in abdominal segments 5 and 6 leads to a loss of pigmentation on tergites A5 and A6. Thus, variations in expression of Hth (and consequently Exd nuclear localization) partially phenocopy alleles of the Hox gene *Abd-B*: as seen in following chapters, modulation of *Abd-B* expression levels and expression domain leads to an expanded or reduced pigmentation on male abdomens (Ryoo et al., 1999) (**Figure 1.3 C**).

There is evidence that Hox genes regulate themselves and other Hox genes. One example of auto-activation is the *labial* gene. This activity can be postulated as an enhancer from the *labial* gene (*lab550*) contains binding sites for the Labial protein (Grieder et al., 1997; Ryoo et al., 1999). An example of cross-regulation between Hox genes is the repression of *abd-A* by *Abd-B*. Absence of *Abd-B* will lead to ectopic expression of *abd-A* in more posterior segments (Karch et al., 1990). It was later shown that this specific repression is particular to the epidermis, and only due to the repressor isoform of *Abd-B* (there are two main splicing isoform of *Abd-B*, one containing a transcriptional repressor domain that is lacking in the other isoform). No derepression was observed in the ventral nerve cord of embryos when *Abd-B* is missing (Gummalla et al., 2012).

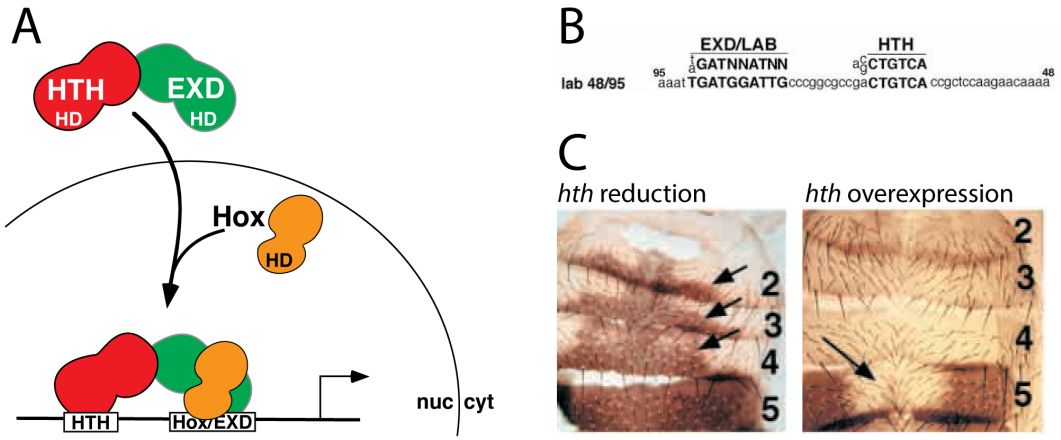


Figure 1.3: Role of EXD and HTH in gene regulation - A: Illustration of HTH/EXD interaction and their role in cooperative gene regulation with Hox proteins. EXD is only transported to the nucleus when bound to HTH. Together they recruit a Hox protein to the correct DNA site. HD: Homeodomain. Modified from (Ryoo et al., 1999). B: Arrangement of EXD, HTH and LAB (the protein coded by the *labial* Hox gene) binding site on a *labial* enhancer. Modified from (Ryoo et al., 1999). C: Modification of HTH levels induce homeotic transformations of segments. Cuticular preparations of male abdomens are shown. In wild types, only segments A5 and A6 are completely pigmented (not shown). *Abd-B* gain-of-function phenotypes in the abdomen (gain of dark pigmentation in A2 to A4) are induced by clonal experiment reducing *hth* expression levels. *Abd-B* loss-of-functions phenotypes (loss of pigmentation in A5 and A6) are achieved by *hth* over-expression. Modified from (Ryoo et al., 1999).

1.1.4 About parasegments and segments in *Drosophila melanogaster*

When observing adult flies, segments are well visible (especially in the abdomen, composed of eight abdominal segments). Lewis, in his seminal paper (Lewis, 1978) postulated that segmental borders are landmarks for the action of the elements in the Hox clusters. Thus, the expression and the action of Hox genes have to be limited by a regulatory system respecting segmental boundaries. Later anatomical and molecular findings led to a refinement of the Lewis-model.

Morata's and Kerridge's observation was that the Hox gene *Ubx* was not only required in the thoracic segment 3 (T3), but that the posterior part of T2 was also affected by *Ubx* alleles (the anterior part of T2 was not affected). Thus, the anterior expression delimitation of *Ubx* is not in T3 but in the middle of T2 (**Figure 1.4 A**). Therefore, it was concluded that Hox gene expression is limited not by segmental boundaries, but by a new metameric unit which was named parasegment (Morata and Kerridge, 1981; Martinez-Arias and Lawrence, 1985).

Parasegments are first molecularly defined by the expression of two pair rule genes, *evenskipped* (*eve*) and *fushi tarazu* (*ftz*). Their sharp anterior expression limits demarcate the parasegmental borders. Later on, this function is taken over by the segment polarity gene *engrailed* (*en*, see **Figure 1.4 A**).

Thus, the parasegments are defined by molecular and not anatomical landmarks. Nevertheless they have anatomical consequences. In scanning electron images of germ band extended embryos (**Figure 1.4 B**) there are anatomical landmarks that indicate the boundaries of segments and parasegments: parasegment boundaries are marked by small grooves on the surface of the embryo, while tracheal pits mark the segmental boundaries (Lawrence, 1988).

The best way to define parasegments and segments is by the expression of *engrailed* (Kornberg et al., 1985; Morata and Lawrence, 1975). *engrailed* separates each segment into an anterior domain (absence of *engrailed* expression) and a posterior domain (*engrailed* expression). At the same time, the anterior expression boundary of each *engrailed* stripe marks the parasegmental boundary, while the posterior expression limit marks the segmental boundary (**Figure 1.4 A**, Martinez-Arias and Lawrence (1985)).

As *engrailed* is expressed throughout development, the parasegmental and segmental boundaries can be tracked in all stages of development: in larvae the most evident anatomical landmark is the bristle pattern of denticle belts on the ventral cuticle (**Figure 1.4 C**). Surprisingly, they do not demarcate either the parasegmental nor the segmental boundaries. The posterior *en* expression boundary is positioned within a denticle belt (between the first and second row of bristles), while the anterior *en* expression boundary is positioned in the naked cuticle between two denticle belts, without any visible morphological effects (**Figure 1.4 D**, Payre (2004)). Therefore, on the larval cuticle the parasegmental boundary does not have anatomical consequences.

In other larval structures, like in wing and haltere imaginal discs, the parasegmental boundary is very important. As those imaginal disc structures consist of cells of a single segment (T2 for wing, T3 for haltere), the parasegmental boundary subdivides these discs into anterior and posterior compartments (**Figure 1.4 E**). In adult flies, especially in the abdomen, the posterior domain is less visible in the external cuticle. In the abdomen, the tissue positive for *engrailed* expression are limited to the intertegal region. Therefore, there is neither a segmental nor a parasegmental boundary on the visible tergites (**Figure 1.4 F**, see next section for anatomical details of fly abdomens).

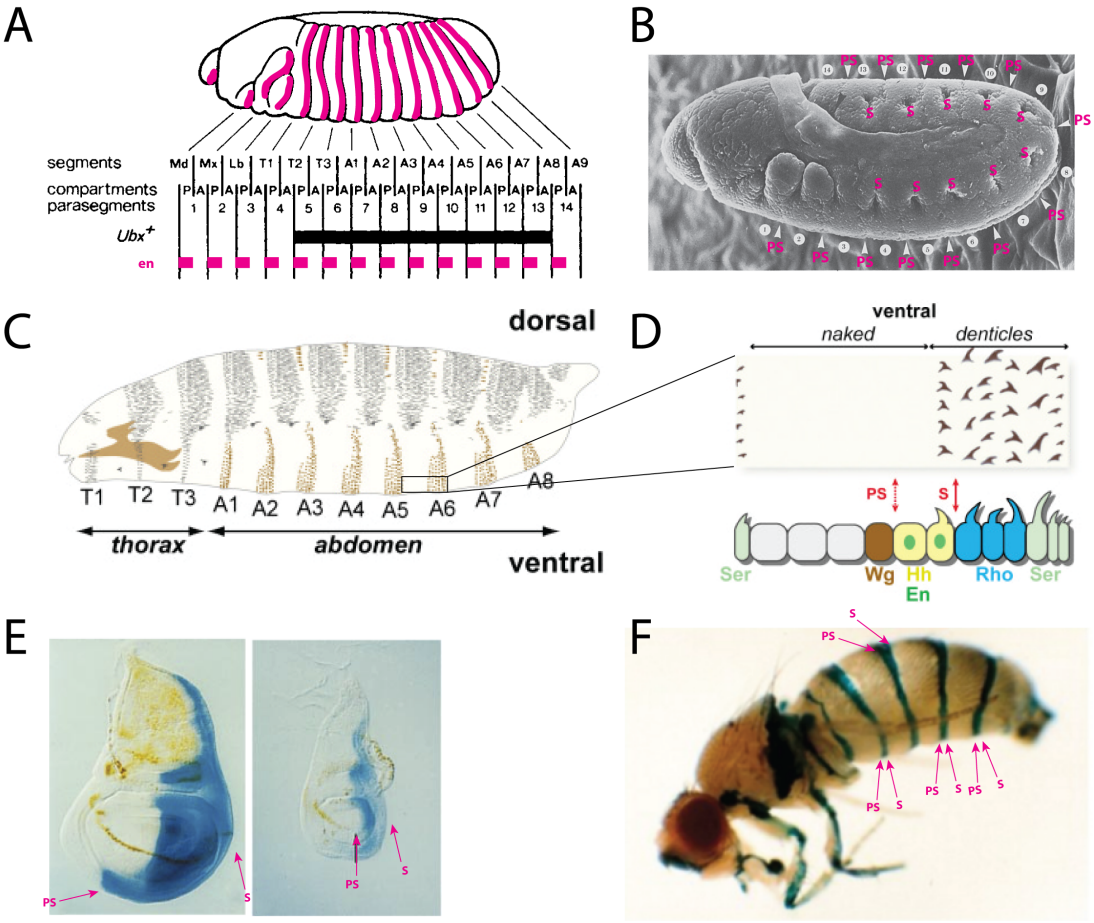
In conclusion, the expression of *en* is a very good marker for the parasegmental and segmental subdivision of different tissues throughout development. This knowledge is important to grasp the nature of Hox phenotypes in those tissues, as the parasegmental boundaries can be followed even in tissues where there is no clear anatomical landmark. For the aim of this study, we can summarize that in the adult fly abdomen parasegments can be translated directly into segments as the external visible part (the tergite) does not contain either segmental nor parasegmental boundary.

1.1.5 From histoblast nests to adult abdominal segments

To understand adult phenotypes of abdominal segments, a short primer on the development of this tissue is needed.

During embryonic development, nests of cells with the same cellular fate are established along the dorsal/ventral axis of each segment that during pupal metamorphosis will give rise to the cuticle of the adult abdomen. Those cells are called histoblasts. Unlike imaginal discs, which are isolated structures within the larvae, histoblast are imaginal precursors that are embedded into larval tissues. Histoblast are established during late embryogenesis but remain quiescent until metamorphosis (Guerra et al., 1973). During metamorphosis those cells start to proliferate (**Figure 1.5 A**) and give rise to the adult tissue, substituting the larval cells (Curtiss and Heilig, 1995).

Figure 1.4 *(following page)*: Parasegments and segments - A: Schematic representation of segmental and parasegmental relationship. Segments, parasegments and compartments are marked along the A/P axis. The posterior compartment of segments is marked by the *engrailed* expression, depicted as pink stripes. Those cells that are positive for *engrailed* expression are positioned in the anterior part of parasegments. The expression domain of *Ubx* is depicted as a black bar, note how the boundaries of expression of this Hox gene are defined by parasegmental and not segmental boundaries. Modified from (Martinez-Arias and Lawrence, 1985). B: Electron scanning microscopy image of a germ band extended *Drosophila* embryo. White arrows indicate small furrows that mark parasegment boundaries. Numbers in white circles indicate the parasegment number. Pink "s" mark the tracheal pits, which are positioned at segmental boundaries. Modified from (Lawrence and Martinez-Arias, 1985). C: Illustration of a *Drosophila* larva. Segments are marked along the A/P axis and divided into thorax and abdomen. Note that on the ventral side three thoracic and 8 abdominal arrays of hairs are present. They are called denticle belts. Modified from (Payre, 2004). D: Ventral topology of bristle hairs within a denticle belt. "Hairy" areas are alternated with "naked" ones. The segmental boundary is positioned according to the *engrailed* expression and thus is not at the border between "hairy" and "naked" areas, but more posterior right after the first row of bristles. The parasegmental boundary does not have any distinctive morphological structures on the cuticula of larvae, and is positioned in the "naked" region. Wg: Wingless, Hh: Hedgehog, En: Engrailed, Rho: Rhomboid, Ser: Serrate. Modified from (Payre, 2004). E: *Lac-Z* expression driven by *engrailed* enhancer in wing and haltere discs. Note how in this case the parasegmental boundary is the one separating anterior and posterior compartments of the discs. Modified from (Shashidhara et al., 1999). F: *Lac-Z* expression driven by *engrailed* enhancer in an adult fly. Note how in the abdomen the *engrailed* expression is restricted to the intertegal region. Thus the biggest part of visible adult abdominal structures (dorsally the tergites) do not contain segmental or parasegmental boundaries. Modified from (Hama et al., 1990).


A
segments
Md Mx Lb T1 T2 T3 A1 A2 A3 A4 A5 A6 A7 A8 A9
compartments
parasegments
Ubx+
en
B
PS
C
dorsal
T1 T2 T3 A1 A2 A3 A4 A5 A6 A7 A8
thorax
abdomen
ventral
D
ventral
naked
denticles
PS
S
Ser
Wg
Hh
En
Rho
Ser
E
PS
S
F
S
PS

There are eight histoblast nests per segment (4 per hemisegment). Their progeny will generate the adult abdomen cuticle (**Figure 1.5 B,C**): two pairs of dorsal nests (anterior and posterior) giving rise to the tergite structures and consisting of around sixteen cells each, a spiracular pair giving rise to the adult spiracles (around three cells), and one ventral pair that will establish the sternite and pleurite tissue (around five cells) (Curtiss and Heilig, 1995).

The majority of the visible part of adult tergites is constituted from cells deriving from the anterior pair of the dorsal histoblast nests. *engrailed* is not expressed in those anterior dorsal nests (Kopp et al., 1997; Struhl et al., 1997). Thus, the histoblast cells that do not see an *engrailed* signal will be the ones mostly visible on the cuticle of adult animals (**Figure 1.5**). Removal of Engrailed activity (in double mutants with the sister gene *invected*) can transform posterior nest cells into anterior nest identity (Lawrence et al., 1999).

As the adult phenotypes that are described in this thesis are limited to tergites, we have to consider that the effects we are observing might be limited to the cells deriving from the anterior dorsal histoblast nest.

1.2 Cis-regulatory elements in the Bithorax complex (BX-C)

1.2.1 The organization of the Bithorax complex

The Hox genes are aligned on the genome in the same order as they are expressed along the body axis. Therefore, mutations along the chromosome will affect single segments (or better: parasegments) along the A/P axis of the fly. Hox genes in *Drosophila melanogaster* are divided in two clusters: the Antennapedia complex (Ant-C) and the Bithorax complex (BX-C). Gene products from the Ant-C are responsible for the identity of the anterior segments until thoracic segment 2 (T2), while BX-C genes are responsible for the segmental identity of segments T3 and all abdominal segments (A1 - A8).

Approximately, each Ant-C gene is responsible for a single parasegment: *labial*, *proboscipedia*, *Deformed*, *Sex combs reduced* and *Antennapedia* are responsible for the identity of parasegments 1 to 5, respectively. In contrast, the genes in

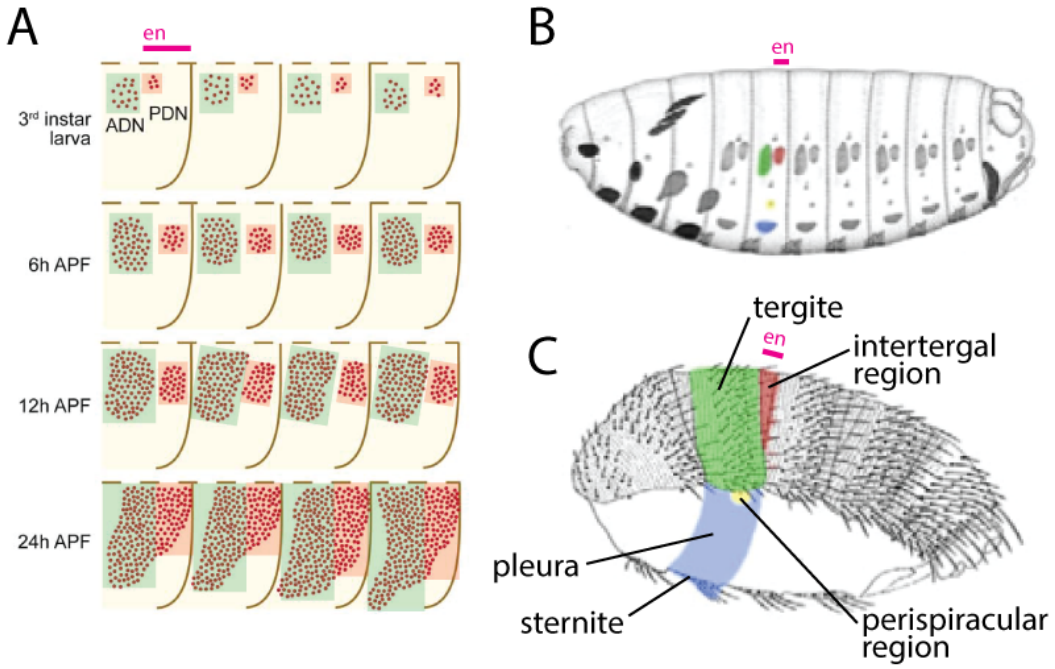


Figure 1.5: Histoblast nests and their fate in the adult abdomen - A: Proliferation of histoblast nests in abdominal segments during pupal development. Red dots represent histoblast cells. APF: After Puparium Formation. ADN: Anterior Dorsal Nest. PDN: Posterior Dorsal Nest. A pink bar marks the *engrailed* expression region in one segment. Note how the parasegmental boundary (anterior en expression boundary) divides the anterior and the posterior dorsal nest. Green and red highlight corresponds to the color code in panel (B). Modified from (Salvaing et al., 2008). B: Anlage of adult abdominal tissue: in late embryogenesis histoblasts are established in each abdominal parasegment. In green: dorsal anterior histoblast nest, in red: dorsal posterior histoblast nest, in yellow: spiracular histoblast nest, in blue: ventral histoblast nest. A pink bar marks the engrailed expression region in one segment. Modified from (Ninov et al., 2007). C: Anatomical names of adult abdominal structures. Colors correspond to the ones in panel (B) and indicate the fate map of histoblast nests. A pink bar marks the *engrailed* expression region in one segment. Note how the *engrailed* domain corresponds to the intertegal region. Thus, there is no segmental or parasegmental boundary going through tergites. Modified from (Ninov et al., 2007).

the BX-C are responsible for 9 parasegments (6 to 14). This led to the following prediction by Lewis:

"The wild type allele of each BX-C gene will be assumed to code for a BX-C substance which controls one or more components of an intersegmental transformation. The various BX-C substances are presumed to act indirectly by repressing or activating other sets of genes which then directly determine the specific structures and functions that characterize a given segment." E.B. Lewis, 1978

Therefore, for each parasegment one pseudo-allele was predicted. In **Figure 1.6 C** those predicted alleles are named *abx/bx*, *bxd/pbx*, *iab-2*, *iab-3*, *iab-4*, *iab-5*, *iab-6*, *iab-7* and *iab-8*. Mutations in each of those regions will affect a very specific segment of the adult fly abdomen (compare color code in **Figure 1.6 A,B,C**).

Surprisingly, in 1985 a study revealed that the BX-C does only consist of three genetic complementation groups (Sánchez-Herrero et al., 1985). Therefore, it was postulated that in the BX-C there are only three protein coding genes with the names *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*).

The regions that previously where named pseudo-alleles by Lewis are regulatory regions of those three genes. The ones of *Ubx* are named *abx/bx* and *bxd/pbx*, the ones of *abd-A* are named *infraabdominal-2* (*iab-2*), *iab-3* and *iab-4*, *Abd-B* regulatory regions are called *iab-5*, *iab-6*, *iab-7* and *iab-8* (**Figure 1.6 C**).

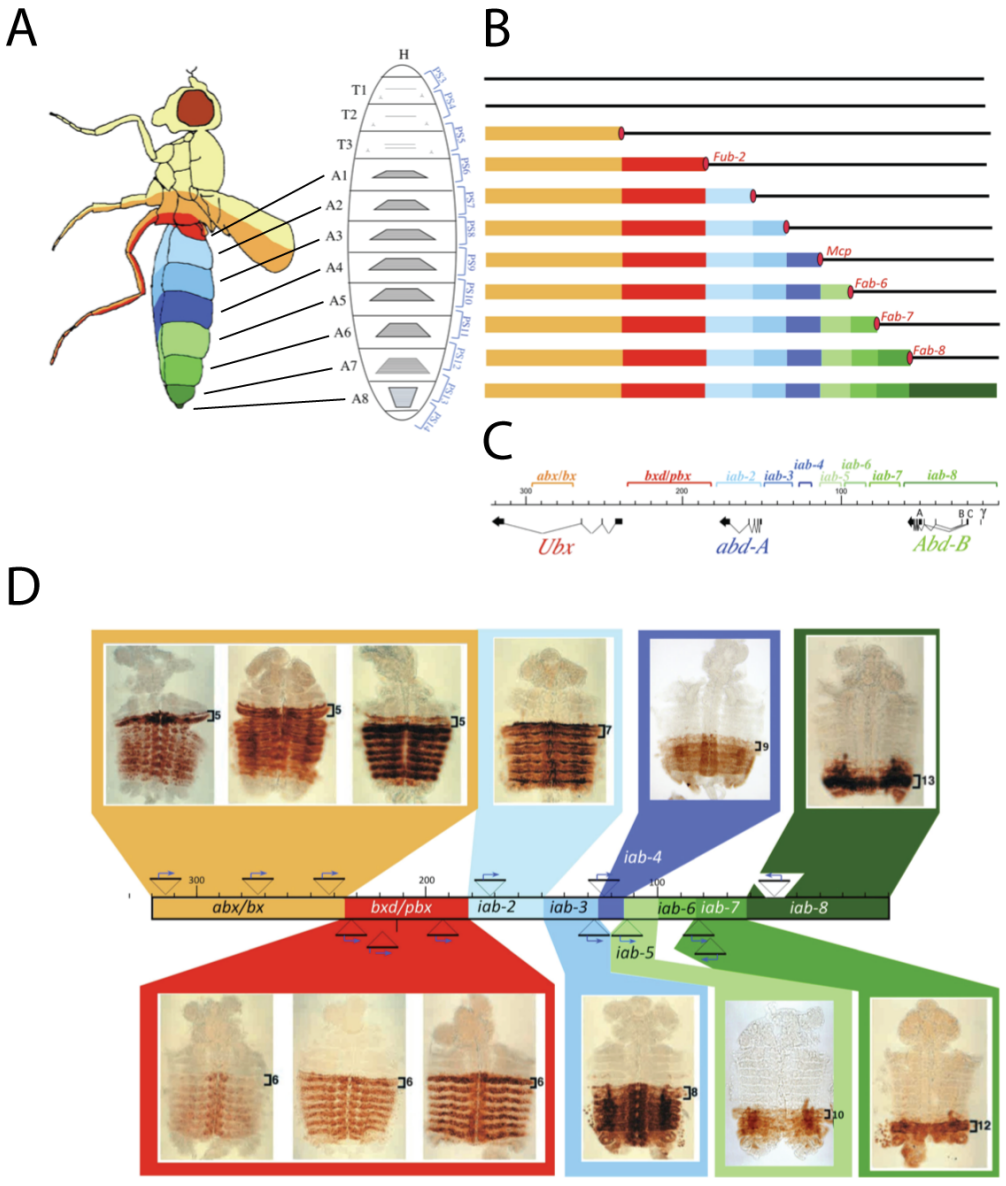
Soon after the molecular technologies had been developed, many rearrangement breakpoints that lead to the original Lewis-model were molecularly mapped (Bender et al., 1983; Karch et al., 1985). These studies substantiated the genetic analysis of the BX-C that Ed Lewis had developed since the early 1950's. Therefore, the segment-specific *cis*-regulatory elements responsible for *Ubx*, *abd-A* or *Abd-B* regulations are assumed to be localized in the region of these breakpoints. For example, mutations in the orange region in **Figure 1.6 B** (*abd/bx*) will affect the tissues in orange shown in **Figure 1.6 A** (posterior compartment of the wing, anterior compartment of the haltere). The same is true also for the other colors in **Figure 1.6**.

1.2.2 Initiator elements

Indications about specific enhancers in the BX-C region came from enhancer trap insertions in the region (**Figure 1.6 D**) and reporter constructs (**Figure 1.8**). In enhancer trap experiments, a transposable element is inserted in a genomic region and piggybacks the activity of enhancers in that region and the reporter gene on the transposon is activated accordingly (O'Kane and Gehring, 1987). The enhancer traps in the BX-C region showed a particular pattern: insertions that were at several kilobases from each other could result in the same parasegmental boundaries, while insertions that were very close to each other could give rise to a different pattern. In **Figure 1.6 D** one can observe how insertions in differently colored genomic regions give different expression patterns in the embryo (the more proximal in the genome, the more anterior the signal). Importantly, insertions in the same regulatory region (represented by the same color) have the same anterior expression boundary.

The patterns received with such experiments look as follows: when moving from the *Ubx* region towards the *Abd-B* region, for each *cis*-regulatory element, the anterior border of expression moves stepwise (parasegment by parasegment)

Figure 1.6 *(following page)*: Open for business in the BX-C - A: Relation between the adult segments and the larval segments is depicted. Furthermore, parasegments and their relation to segments are also illustrated. Modified from (Maeda and Karch, 2015). B: Genomic regulatory regions that are active in the different segments of the animal are marked with colored boxes, the black line represents inactive regulatory sequences in that specific animal segment. Boundaries (*Fub-2*, *Mcp*, *Fab-6*, *Fab-7*, *Fab-8*) that separate active from inactive regulatory regions are marked by a red oval. Positions on the genome correspond to illustration of the region in panel (C). Modified from (Maeda and Karch, 2015). C: Illustration of the BX-C genomic region. Coding sequences are marked by arrows and their respective names (*Ubx*, *abd-A*, *Abd-B*). Above the black line (numbers represent the coordinates from the original sequencing clone) are the regulatory regions with their names (*abx/bx*, *bxd/pbx*, *iab-2* to *iab-8*). Modified from (Maeda and Karch, 2015). D: Enhancer trap insertions and their respective expression in the BX-C. Insertions in the same *iab* gives the same expression pattern. Note how insertions in different genomic locations result in distinct patterns. The more proximal insertions show a more anterior expression boundary. Modified from (Maeda and Karch, 2015).


A
H
T1
T2
T3
A1
A2
A3
A4
A5
A6
A7
A8
B
Fub-2
Mcp
Fab-6
Fab-7
Fab-8
C
abx/bx
bxd/pbx
iab-2
iab-3
iab-4
iab-5
iab-6
iab-7
iab-8
Ubx
abd-A
Abd-B
D
abx/bx
bxd/pbx
iab-2
iab-3
iab-4
iab-5
iab-6
iab-7
iab-8

more and more posteriorly. Thus, the regulatory regions of the *Ubx* gene are active in more parasegments than the ones from the *abd-A* or *Abd-B* genes. This led to the hypothesis that the BX-C is separated into distinct regulatory regions that are responsible for one single parasegment (Bender et al., 1983).

This sequential activation of regulatory regions along the chromosome (as depicted in **Figure 1.6 B**) led to the formulation of the "open-for-business" model: starting from the most proximal genomic region of the BX-C, the regulatory regions are activated one by one according to their genomic location, parasegment by parasegment (**Figure 1.6 B**, Maeda and Karch (2015); Peifer et al. (1987)). An important aspect of the "open-for-business" model was that the stepwise transcriptional activation of the BX-C depended on boundary elements separating two neighboring *cis*-regulatory regions. In **Figure 1.6 B**, these boundaries are indicated by red ovals.

But how are those regions activated? That's where initiator elements come into the play. Those elements are enhancers that integrate the previously described gap, pair-rule, and segment polarity genes inputs. Each *iab* region contains at least one initiator element. Those initiator elements are defined as DNA sequences from the BX-C region whose expression is restricted to a subset of parasegments when analyzed in reporter constructs (**Figure 1.8**).

A first report about the anterior expression restriction of initiators was the analysis of four different DNA sequences from the *Ubx* and *abd-A* gene regions. The expression in reporter constructs of those DNA fragments (coordinates and expression patterns in **Figure 1.8 A**) shows a clear anterior expression boundary. This expression boundary moves posteriorly stepwise, parasegment by parasegment, the more distal the DNA was originally isolated from the BX-C (Simon et al., 1990). For the *Abd-B* region, the four initiators responsible for segmental gene expression were localized with similar LacZ-reporter experiments (Mihaly et al., 2006) (**Figure 1.8 B**).

Comparing the position of initiators on the genome with the embryonic expression patterns there is a clear collinearity: genetic elements show activities along the anterior/posterior body axis in the same order as they are arranged on the genome. The initiators which are most proximal (closer to the centromere)

show a more anterior expression. Deletions of initiator elements result in a loss-of-function homeotic transformation: the segment with the most anterior expression of the deleted initiator element, acquires the identity of its anterior neighboring segment. One example is the deletion of the *iab-6* initiator. *iab-6* is usually responsible for the identity of PS11 (A6 in adult flies). In flies carrying this mutation the abdominal segment A6 is transformed into abdominal segment A5 (Iampietro et al., 2010). See **Figure 1.7** for a molecular explanation for such a phenotype in accordance to the "open for business" model.

Another experiment by Iampietro and colleagues demonstrated that initiator elements read out the position information from gap, pair rule and segment polarity genes independently from their genomic position in the *Abd-B* gene. When the *iab-5* initiator element is inserted in place of the *iab-6* initiator the *iab-6* region is "opened" already in PS10 where the *iab-5* initiator is active. As a consequence, PS10 acquires the identity of PS11 in the embryo. In adults the A5 sternite is transformed into an A6 sternite (no bristles on the cuticle, **Figure 1.7 D**).

Combining the finding that 1) initiators are active from a defined anterior border to the posterior end of the embryo (**Figure 1.6 D** and **Figure 1.8**) with the finding that 2) deletions of initiators lead to a loss-of-function phenotype, the relation between homeotic transformation and initiator elements can be reconstructed: cells in a certain segment are defined by the one initiator with the most posterior activity. For example cells in PS11 are defined by *iab-6*, although other more anterior initiators are active as well in that one segment. Therefore, every segment is controlled by only one *iab* region and its initiator element. When an initiator is deleted, the cells in the respective segment will not see the activity of the defining initiator. Therefore, cells now react to the next more anterior initiator as their identifying signal, as this specific initiator is active in this segment. Therefore, they will acquire the identity of the adjacent anterior segment (**Figure 1.7**). For example, cells in PS11 (A6 in adults) are defined by the initiator in *iab-6*. When the *iab-6* initiator is deleted, the activity left in this segment is the one from *iab-5* and thus PS11 (A6 in adults) will look the same as PS10 (A5).

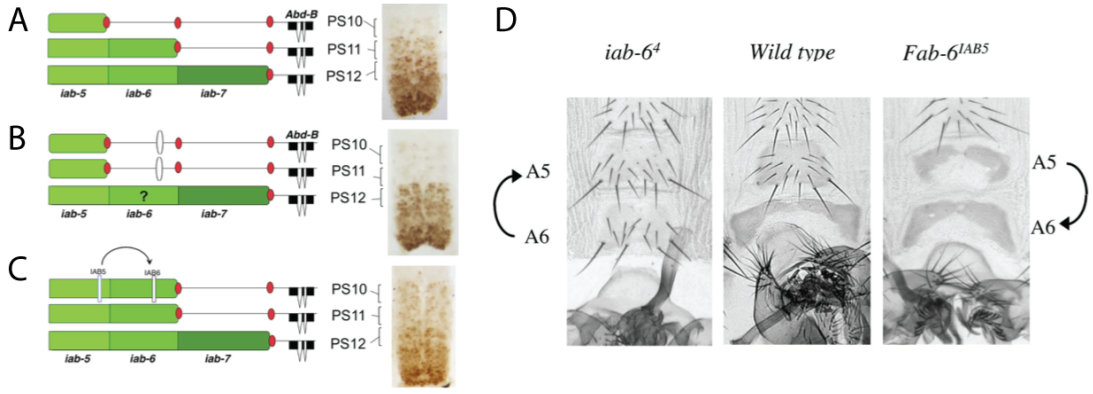


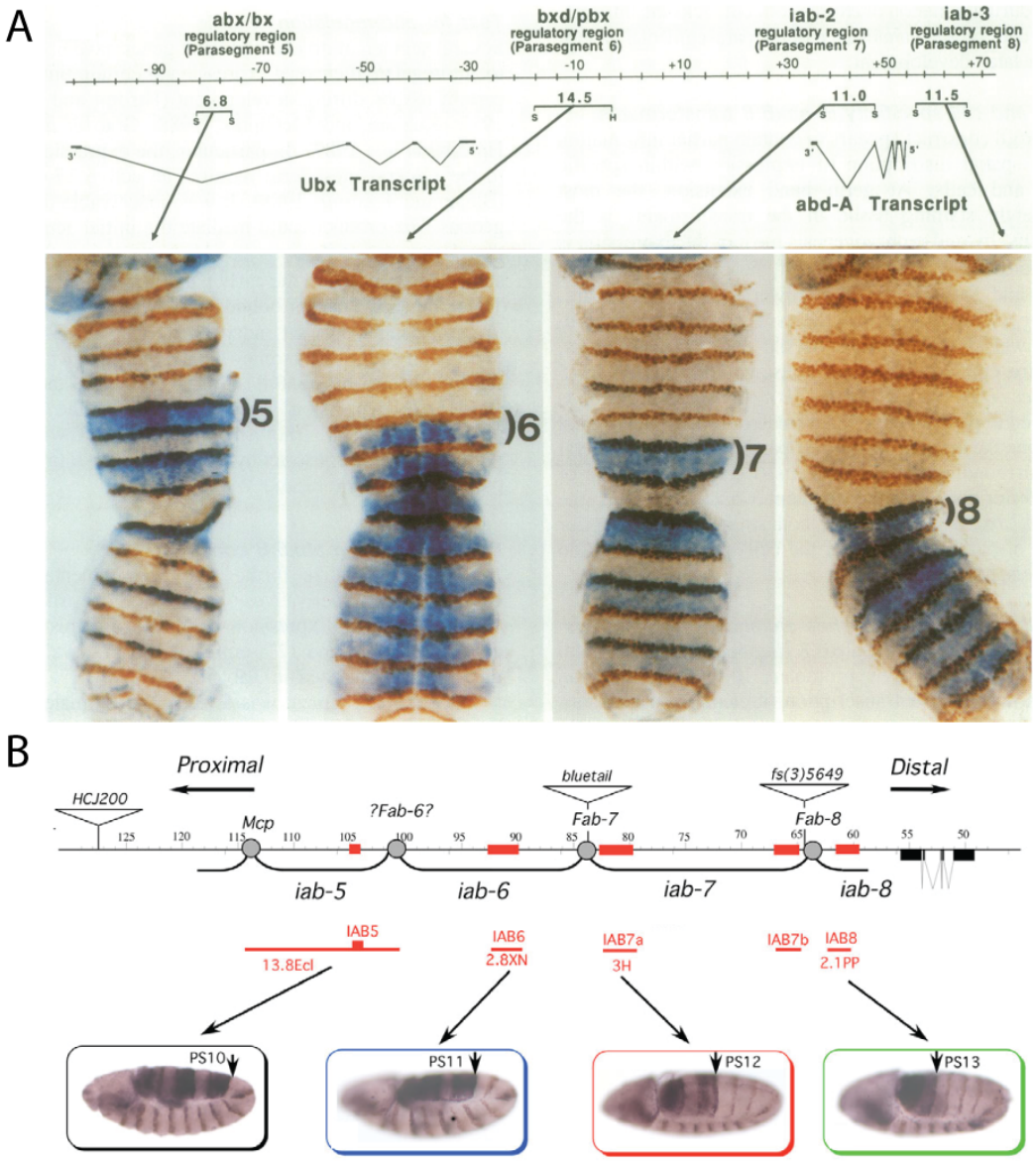
Figure 1.7: *iab-6* deletion and replacement - A: Wild type situation. Red ovals represent the boundary elements, green rectangles indicate the active genomic regions in each parasegment. The thin lines (not in green boxes) represent inactive genomic regions. In PS10 *iab-5* is active, in PS11 *iab-5* and *iab-6* are active, in PS12 *iab-5*, *iab-6* and *iab-7* are active. This is reflected by the *Abd-B* expression in the embryonic ventral nerve cord (on the right): PS10 shows the weakest anti-Abd-B staining, while PS11 and PS12 show stronger expression. Modified from (Maeda and Karch, 2015). B: Deletion of the *iab-6* initiator prevents the activation of *iab-6* in PS11. Therefore, PS10 and PS11 both experience only the activity of the *iab-5* initiator and therefore show the same low Abd-B protein levels (with levels associated to PS10 in wild type). Modified from (Maeda and Karch, 2015). C: When the *iab-5* initiator is inserted in place of the *iab-6* initiator the *iab-6* region is activated already in PS10. Therefore, the levels in PS10 and PS11 are again the same, but PS10 acquires the levels of PS11. Modified from (Maeda and Karch, 2015). D: Phenotypes of the *iab-6* initiator element deletion and replacement in adult abdomens. Note that PS10 corresponds to segment A5 where *iab-5* is active and PS11 to A6 where *iab-6* is active. In wild type males, the bristle patterns on ventral abdominal segments A5 and A6 are different: about 15 bristles grow in A5; A6 is devoided of bristles. When the *iab-6* initiator is deleted A6 is transformed into A5 and shows ectopic bristles on A6. When the *iab-6* initiator deletion is rescued with a *iab-5* initiator the opposite situation happens (A5 looses the bristles and is transformed into A6). Modified from (Maeda and Karch, 2015).

Comparing the findings of transgenic reporter construct experiments (**Figure 1.8**) with enhancer trap experiments in the endogenous locus (**Figure 1.6 D**) we notice that the expression in the endogenous locus is much more stable and consistent than in the transgene (for example we can see some pair rule like pattern in reporter constructs, as seen in **Figure 1.8**. These effects are not visible in enhancer trap experiments). Thus, we can reason that initiators are embedded in a more complex regulatory environment in the endogenous locus, that gives them a more controlled and precise expression pattern.

1.2.3 Maintenance elements and Polycomb group proteins

Hox genes have to be regulated for a long period during development (at least until third instar larva, Foronda et al. (2006)) but their regulatory factors (mainly gap and pair rule genes) are present only early in embryogenesis. So, how do initiators and thus Hox genes keep their expression constant when their regulatory input changes? There must be an additional mechanism. Reporter constructs containing initiator elements have been observed to expand their region of activity in later embryogenesis (Busturia et al., 2001; Maeda and Karch, 2006). This decay of the anterior expression limit might derive from Hox auto-regulation

Figure 1.8 *(following page)*: Enhancer from the BX-C and their expression in enhancer reporter experiments - A: Expression of enhancer reporter constructs. Transcripts of *Ubx* and *abd-A* are shown. Four fragments (marked as horizontal square brackets) are depicted, together with their size and the restriction sites used for cloning (S for SalI, H for HindIII). The corresponding expression patterns from these four constructs in 6h old embryos are shown below and connected with an arrow to the corresponding DNA fragment. Notice how these fragments differ in their anterior expression pattern. For the *abx/bx* and *iab-2* initiator region a pair rule input is possible, as the reporter shows a pair rule like pattern. Modified from (Simon et al., 1990). B: The position of boundaries (gray circles), *Abd-B* transcript (black boxes), initiator elements (red boxes), and transposable insertions (triangles) are indicated on the black line representing the *Abd-B* genomic region. Reporter constructs containing the initiators (red lines) show restricted expression patterns in embryos with extended germ band. Note how, similar to panel (A), the anterior expression boundary moves posteriorly segment by segment when more posterior initiators are used. Modified from (Mihaly et al., 2006).


A
abx/bx
regulatory region
(Parasegment 5)
bxd/pbx
regulatory region
(Parasegment 6)
iab-2
regulatory region
(Parasegment 7)
iab-3
regulatory region
(Parasegment 8)
- 90
-70
-50
-30
-10
+10
+30
+50
+70
6.8
14.5
11.0
11.5
Ubx Transcript
abd-A Transcript
)5
)6
)7
)8
B
Proximal
Distal
HCJ200
Mcp
?Fab-6?
bluetail
Fab-7
fs(3)5649
Fab-8
iab-5
iab-6
iab-7
iab-8
IAB5
13.8Ecl
IAB6
2.8XN
IAB7a
3H
IAB7b
IAB8
2.1PP
PS10
PS11
PS12
PS13

(Kuziora and McGinnis, 1988) or activation by unspecific binding of other Hox proteins expressed in other segments to the unspecific homeobox binding site. Furthermore, cell-type specific factors may also lead to such an expansion of activity. Surprisingly, some DNA fragments isolated from *iab* regions were able to keep the expression pattern in check even in later stages of development and did not show an expansion of the expression domain. Those fragments contain elements called maintenance elements or Polycomb Response Elements (PRE). PREs contain binding sites for Polycomb and Trithorax group proteins. Deletions of such binding sites reverse this memory function (Chiang et al., 1995; Busturia et al., 2001).

Figure 1.9 shows an illustration of an experiment where the *iab-6* initiator element is flanked by a PRE and drives a *LacZ* reporter gene. In this configuration the *LacZ* expression shows a clear restriction to PS6 and more posterior parasegments both in early and late embryogenesis. Deletion of the PRE on the same transgene leads to a loss of the clear *LacZ* expression boundary later in embryogenesis (Maeda and Karch, 2009). This transgenic experimental set-up clearly points to a role for the PRE in the maintenance of Hox gene expression after the early regulatory inputs are lost. Other PREs were studied and characterized with analogous experiments (see also **Section 1.5.2**).

Maintenance elements belong to the basic layout of every *iab* region. They contain binding sites both for Polycomb group and Trithorax group proteins. Therefore, they are also referred to as Trithorax Response Elements (TRE). Polycomb group proteins (PcG) are involved in silencing of *iab* regions where initiators are not active. On the other hand, Trithorax group proteins maintain the active state established by an active initiator element in *iab* regions (Kennison, 1993; Paro, 1990).

Biochemical studies have shown that the substrate of Polycomb and Trithorax group proteins is the chromatin surrounding the PRE/TRE (Simon, 1995). Its activity is regulated by specific covalent modifications of histones and/or by nucleosomal remodeling (Pirrotta, 1997).

The correct interplay between initiator and maintenance elements is crucial, as the positional information that is read out by the initiator elements is fading

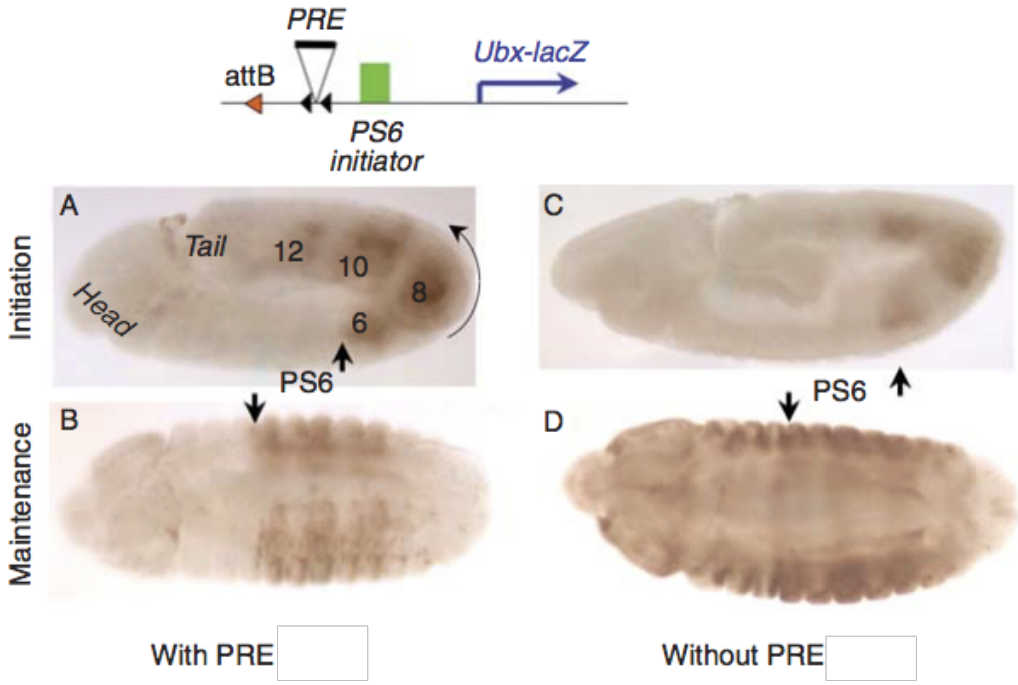


Figure 1.9: The role of PREs in memory function - Lack of PRE leads to an expansion of initiator element activity. On top the architecture of the transgene. The transgene contains a PS6 initiator and a PRE (from the *bxd/pbx* region). The initiator drives the transcription of a *LacZ* reporter under the control of an *Ubx* promoter. In early embryos (germ band expanded, panels A an C) the expression of *LacZ* is limited anteriorly by the PS9 anterior boundary independently of the presence of the PRE (A versus C). In later stages (panels B and D) the signal is expanded both anteriorly and posteriorly in the absence of the PRE (panel D). Black arrows mark the PS6 anterior boundary. Modified from (Maeda and Karch, 2009).

during early embryogenesis, as gap and pair rule genes are not expressed anymore (described in the previous section).

In *Drosophila*, PREs and TREs DNA sequences can be predicted according to the binding of a couple of proteins (GAGA binding factor, Pleiohomeotic, Engrailed, and Zeste). DNA binding can be predicted *in silico* using the binding motifs of those proteins (Ringrose et al., 2003). PcG proteins are targeted to PREs by a small subset of DNA binding members of the PcG, e.g. Pho, GAF or Dorsal switch protein 1 (DSP-1, **Figure 1.10**). Indeed, when the binding sites for Pho and GAF are deleted in initiator reporters with an associated memory function (or PRE) the maintenance of an expression boundary is lost (Busturia et al., 2001).

But how do Polycomb proteins lock a certain state to a stretch of DNA on the genome? PcG proteins were first described in *Drosophila* (Slifer, 1942). Mutations could be isolated which phenocopied homeotic gain of function phenotypes. For example, alleles of *Polycomb* (*Pc*, a member of PcG) are often associated with patches of black pigment on the fourth abdominal segment of male flies, extra sex comb teeth on the second and third pair of legs, or antenna to leg transformations. These phenotypes are reminiscent of those found for *Abd-B*, *Scr* or *Antp* gain-of-function alleles, respectively (Busturia and Morata, 1988). These observations suggest that it is the role of PcG proteins to delimit expression of homeotic genes to certain territories.

Three multi-protein complexes have so far been characterized which are recruited to PREs. Their names are PRC1 (Polycomb-repressive complex 1), PRC2, and PhoRC (Pleiohomeotic repressive complex). Those complexes are composed of several proteins with enzymatic activity. The substrate of PcG proteins is the chromatin surrounding the PRE, which specialized proteins modify by covalently linking methyl- or ubiquitin-groups (PRC2 and PRC1, respectively. **Figure 1.10**) to specific amino acids in the N-terminal arms of the histone proteins (Schwartz and Pirrotta, 2007).

The sequence of events occurring at the PRE is as follows: the PRE senses the activity of the surrounding chromatin. If it is poised for transcription, PRCs do not exert their function (Papp and Müller, 2006). If the chromatin close to the PRE is inactive, the first complex that is recruited to the DNA is PRC2. Its

main role is tri-methylation of lysine 27 on histone 3 (Nekrasov et al., 2007). This histone modification is catalyzed by the methyl-transferase protein Enhancer of Zeste (E(z)) (Viré et al., 2006). Subsequently, this epigenetic mark is recognized by PRC1. PRC1 reinforces repression by ubiquitylation of lysin 119 on histone 2A (Wang et al., 2004).

But how is a PRE switched on or off? This is probably the biggest question about PREs nowadays. How does this element decide if it is in the off state (PRE, repressive) or on state (TRE, maintenance)? There are findings that link such activities to transcription running through the PRE from one direction or the other (Cavalli and Paro, 1998; Herzog et al., 2014), and thus acting as a switch, but other studies refute such ideas (Erokhin et al., 2015).

1.2.4 Boundary elements

The "open for business" model proposes that each *iab* region is opened sequentially. This was finally demonstrated on a molecular level by combining genetics and biochemistry in the work of Bowman and colleagues (Bowman et al., 2014). They analyzed Polycomb modifications of histones in the BX-C. The elegance of their work was that they were able to isolate cells from single parasegments and thus achieved an unprecedented resolution for the analysis of chromatin modifications within each *iab* along the A/P axis.

The boundary between the Polycomb silenced regions and the not silenced regions lays precisely at the genomic positions previously identified by genetic studies. In the next paragraphs, I will go through the evidence that lead to the definition of boundary elements in the BX-C.

For the *Abd-B* region extensive studies were performed on the sequences that separate the single *iab* regions from each other. Those sequences were called boundary elements. The boundary separating the *abd-A* from the *Abd-B* region (thus between *iab-4* and *iab-5*) is called *Miscadestral pigmentation* (*Mcp*). The boundaries separating *iab-5*, *iab-6*, *iab-7* and *iab-8* are called *Fab-6*, *Fab-7* and *Fab-8*, respectively (see **Figure 1.6**).

Deletions of single boundary regions will lead to fused *iab* regions. Such a fusion of two previously independent *cis*-regulatory regions will lead to a gain-

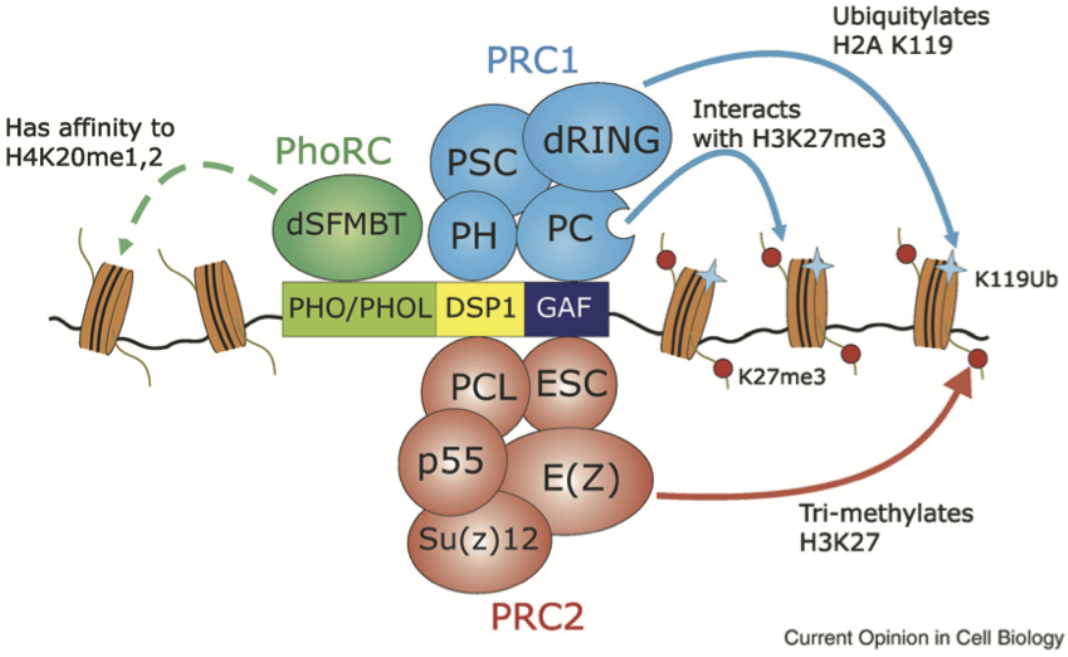


Figure 1.10: Polycomb Group Proteins - PcG proteins bind to specific sequences called Polycomb Response Elements (PRE, green, yellow and blue boxes on the thin black line representing DNA). The first complex that is recruited to the DNA is PRC2. Its main role is tri-methylation of lysine 27 on histone 3 (H3K27m3). H3K27m3 is catalyzed by the methyl-transferase protein Enhancer of Zeste (E(z)). This epigenetic mark is recognized by PRC1 which contains Polycomb. PRC1 reinforces repression by ubiquitylation of lysine 119 on histone 2A. Modified from (Schwartz and Pirrotta, 2008).

of-function phenotype. In the newly fused regulatory region the more posterior of the two *iabs* determines the target gene and expression level, while both the anterior and posterior *iabs* determine the segments in which the newly formed regulatory region is active. Therefore, the segmental identity of the more anterior segment acquires the identity of the posterior segment. For example, in male adult flies the segment A5 is pigmented, while A4 is not. Deletion of the *Mcp* boundary leads to ectopic pigmentation in A4 (Celniker et al., 1990). The reason for this is that the regulatory region *iab-4* is not separated anymore from *iab-5*. The two regions now act as one, and *Abd-B* is activated in PS9 in the embryo (usually the expression boundary is in PS10, Galloni et al. (1993)).

Another example is the deletion of the *Fab-7* boundary that separates the *iab-6* from the *iab-7* region (Gyurkovics et al., 1990). When this boundary is deleted, PS11 will acquire the identity of PS12 as the two regions *iab-6* and *iab-7* are not able to work independently, as shown in **Figure 1.11**. Therefore, the *Abd-B* expression levels are increased in PS11 to the levels found in PS12. **Figure 1.11** is also a good representation of the "open for business" model: the three lines with green ovals represent the situation of the *iab-5*, *iab-6* and *iab-7* regions in the three parasegments PS10, PS11 and PS12. In **Figure 1.11 A** the wild type situation is depicted: in PS10 only *iab-5* is active and therefore "open", in PS11 *iab-5* and *iab-6* are open, and in PS12 *iab-5*, *iab-6* and *iab-7* are open. A deletion of the *iab-7* initiator element (**Figure 1.11 B**) leads to a loss of function phenotype in PS12. In contrast, a deletion of a boundary element (**Figure 1.11 C**) leads to a gain of function phenotype in PS11, as *iab-6* and *iab-7* are fused and therefore "opened" together.

For the sake of precision, it is worth noting that the phenotypes of *Mcp* and *Fab-7* described above are achieved by deletions (Mcp^1 and $Fab\text{-}7^1$) that remove both the boundary elements and associated PREs. A dissection of the individual roles of boundary element and PRE was achieved for *Fab-7* in the endogenous genomic location (see **Section 1.4.1**). This was not the case for *Mcp*.

In summary, two initiator elements will end up in the same domain when a boundary is removed, and therefore the interaction of both initiators with the promoter is spurious leading to ectopic expression.

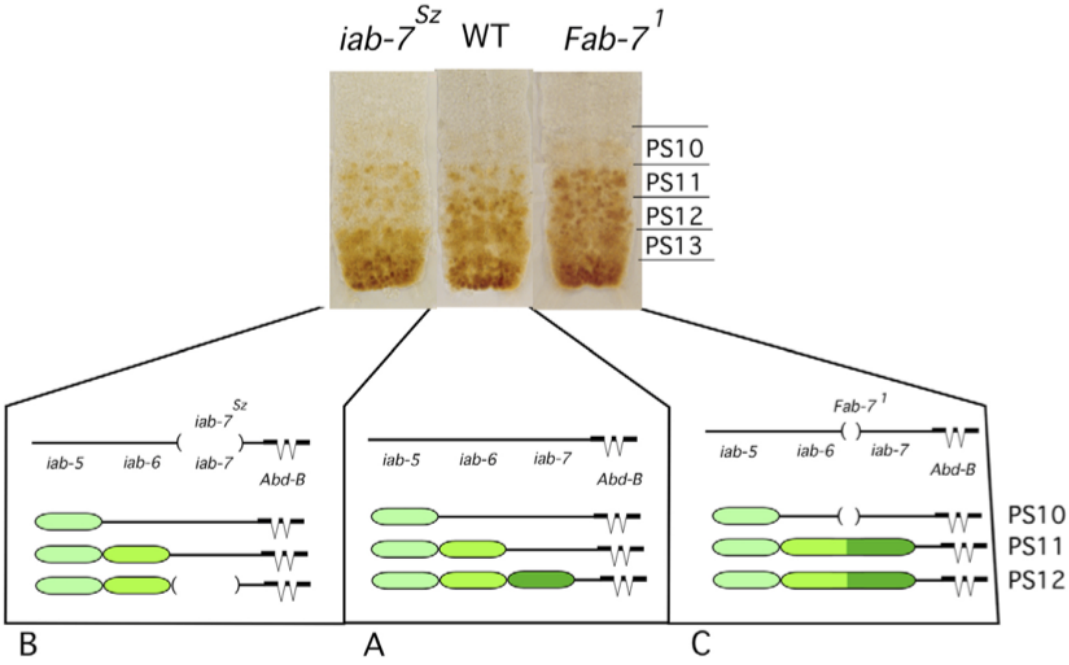


Figure 1.11: Gain- and loss-of-function phenotypes according to the "open for business model" - Expression of *Abd-B* in embryonic ventral nerve cord in different mutant backgrounds. *iab-7*^{Sz} removes the *iab-7* initiator element. *Fab-7*^{1} removes the *Fab-7* boundary that separates *iab-6* from *iab-7*. WT: wild type control. A: In wild type the *iab* regions *iab-5*, *iab-6* and *iab-7* are activated according to the "open for business" model (see **Figure 1.6 B** and **Figure 1.7**). B: In the case of a deleted initiator element the result is a loss of function phenotype: *Abd-B* levels in the relative parasegments are the same as the anterior adjacent parasegment (compare levels of *Abd-B* in PS12 and PS11 in *iab-7*^{Sz} vs wild type). C: In the case of a boundary deletion the two *iab* regions are not able to act separately. Thus the *Abd-B* expression is again the same in PS11 and PS12 (like in panel B) but instead of a loss of function there is a gain-of-function phenotype (higher level of *Abd-B* in PS11). Modified from (Maeda and Karch, 2006).

1.2.5 Promoter targeting sequences

In a general genome context, boundaries are thought of as *cis*-regulatory elements preventing enhancers to contact promoters in a genomic region on the other side of the boundary, therefore creating independent regulatory domains. If this is the case, how do enhancers in the different *iabs* contact the *Abd-B* promoter, having to surpass several boundaries? Kyrchanova and colleagues have called this problem the "boundary paradox in the BX-C" (Kyrchanova et al., 2015).

To allow correct interaction between the enhancer elements on different *iabs* with the promoter of coding sequences, it was therefore proposed that each *iab* region contains a promoter targeting sequence (PTS). Indeed, in transgenic assays, the PTS from the *iab-7* region was able to promote enhancer/promoter interaction (Zhou and Levine, 1999). A second PTS was predicted in the *iab-6* region with transgenic assays using those sequences probing PTS activity, strengthening the evidence for such *cis*-regulatory elements (Chen et al., 2005). However, a clean deletion of the *iab-7* PTS element only weakly impaired *iab-7* function, while no phenotype is visible for a *iab-6* PTS deletion in the endogenous locus (Mihaly et al., 2006).

Recent experiments that replaced the *Fab-7* boundary in the endogenous locus with *Fab-8* sequences showed that only *Fab-8* fragments containing the PTS were able to allow correct insulator bypass of the boundary (Kyrchanova et al., 2016). In this study, the activity of the *iab-7* PTS was confirmed and further narrowed to 83bp.

The evidence accumulated in the field suggests that PTS elements are indeed present in the *Abd-B* gene, however further work is necessary to define those elusive *cis*-regulatory elements and their interaction with the *Abd-B* promoter.

1.2.6 Tissue specific enhancers

Further investigations found DNA sequences originating from the BX-C that give an expression signal in reporter assays, but behave differently from the previously described initiator elements. The activity domain of those enhancers is not restricted to a certain subset of parasagments as it was the case for initiator elements, but instead they are active in specific tissues (like epidermis (Simon

et al., 1990) or individual imaginal discs (Pirrotta et al., 1995)). Therefore, those enhancers were named tissue specific enhancers.

Interestingly, transposable element insertions into the endogenous BX-C (enhancer trap experiments, O'Kane and Gehring (1987)) do never show such tissue specific expression pattern and are always restricted to a parasegmental boundary (**Figure 1.6 D**, Bender et al. (1983)). Thus, it is probable that those tissue specific enhancers are restricted on their activity in the BX-C both on a tissue level (by their own activity) and on a segmental level (by PRE activity in the endogenous BX-C, according to the "open for business" model).

1.3 The nature of boundary elements

1.3.1 Known boundary elements in *Drosophila*

The classification of a DNA sequence as an insulator or boundary element is based on the phenotype when the sequence is deleted and/or the activity of the DNA in transgenes. In the BX-C, deletion of a boundary gave rise to a gain of function phenotype instead of a loss of function observed when an initiator or enhancer was deleted. Transgenic assays were used to test if a boundary contained the following characteristics: enhancer blocking function, barrier function, and long distance interaction (**Figure 1.12**, reviewed in Kuhn and Geyer (2003); Matzat and Lei (2014)).

Several boundary element classes were identified in *Drosophila* based on genetic and biochemical criteria. Their classification depends mainly on their associated proteins. Boundaries can be broadly grouped in two classes: class I) the Supressor of Hairy wing dependent boundaries and class II) the Centrosomal protein 190 (CP190), Boundary element associated factor 32 (BEAF-32), and CCCTC-binding factor (CTCF) dependent boundaries (Nègre et al., 2010). Class I contains virtually only the Suppressor of Hairy wing (Su(Hw)) bound insulator (Su(Hw) is a protein discovered because of its binding to a natural transposable element), while class II contains a diverse array of independent boundaries. Examples of class II boundaries are the boundaries in the Hox clusters. The boundaries in the BX-C, and more specifically in the *Abd-B* region, are the best studied.

1.3.2 Enhancer blocking, barrier function, and long distance interaction

The definition of an insulator (or boundary element) is based on the phenotypes such DNA sequences exhibited in transgenic constructs. The functions that are associated with such elements are enhancer blocking (**Figure 1.12 B,C**), chromatin barrier function (**Figure 1.12 D,E**), and long distance interaction (**Figure 1.12 F,G**). Those functions lead to the hypothesis that the main role of boundary elements may be to protect a given promoter from spurious activation by enhancers that usually are not responsible for that specific promoter and therefore enable the correct enhancer-promoter interaction. Furthermore, these elements are implicated with the blocking of heterochromatin spreading (reviewed in Kuhn and Geyer (2003); Matzat and Lei (2014)).

Enhancer blocking function can be tested with a simple transgenic assay: a sequence thought to be a boundary element is inserted between an enhancer and a promoter of a reporter gene (**Figure 1.12 A,B**). If a gene is not expressed, this might be an indication that the sequence inserted is indeed an insulator, as it prevents correct enhancer/promoter interaction. In such a case, it has still to be ruled out that the repression of the reporter gene is not due to silencing. To test this, the putative insulator is cloned on the other side of the enhancer. Thus, the insulator DNA is not between the enhancer and the promoter anymore (**Figure 1.12 C**). In that situation, the insulator function should not block the enhancer/promoter interaction anymore (Kellum and Schedl, 1991).

Chromatin barrier function can be tested with a *mini-white* reporter. Such reporters show a big variety of eye pigmentation reflecting the genomic location they are inserted into by transposable elements. This phenomenon is named chromosomal position effect (Levis et al., 1985). This effect can be used to test the capacity of insulators to block the spread of heterochromatin (Kellum and Schedl, 1991): when a *mini-white* reporter is inserted close to heterochromatin the eye color is either weak or variegated (some ommatidia colored and others lacking color). The insertion of a boundary element blocks the spreading of heterochromatin and the eye color becomes darker and more constant throughout ommatidia.

Long distance interaction can either be tested *in cis* or *in trans*: it was shown that boundary elements can pair physically on the same transgene, but also over very long distances when inserted on the sister chromosome, and even (in some cases) on different chromosomes. In transgenes, the hypothesis of loop generation was put forward after a surprising observation: compared to enhancer blocking assays, in which a single boundary is inserted between an enhancer and a reporter, Kuhn and colleagues inserted two *gypsy* insulators (**Figure 1.12 F**). One insulator is enough to prevent enhancer-promoter interaction, an additional insulator reverts this phenotype and the reporter is expressed (Kuhn et al., 2003). The enhancer is probably brought in vicinity of the promoter by the physical interaction of the two insulators, looping out the DNA laying between them (**Figure 1.12 G**).

1.3.3 The gipsy insulator

Most of the classical *Drosophila melanogaster* alleles were isolated as spontaneous mutagenic events. Much later, it was demonstrated that many of them were due to insertions of mobile genetic elements (transposons) into gene regions. This ectopic genetic material disturbs the function of the gene and thus gives rise to a phenotype. One of this transposable elements is the *gipsy* transposon. Relevant examples for the topic of this thesis are *gipsy*-dependent *abx* and *bx* alleles of the *Ubx* gene region. Such alleles induce haltere to wing homeotic transformations (Peifer and Bender, 1986). As it is the case for *abx* and *bx* alleles, many of the *gipsy* insertions leading to a phenotype were not localized in coding regions, but in regulatory regions of a gene.

How can an insertion of a *gypsy* element in regulatory regions disrupt gene function? It was found that the *gypsy* transposable element contains a 430bp long DNA fragment containing 12 copies of a repeated motif. When tested in an enhancer blocking assay this repeat motif was sufficient to inhibit correct reporter expression (Geyer and Corces, 1992; Geyer et al., 1988). The repeat region was shown to be bound by a zinc finger protein called Suppressor of Hairy wing (Su(Hw)). These studies suggested that *gypsy* element insertions act as boundary (or insulator) elements preventing proactive enhancer-promoter interactions. In

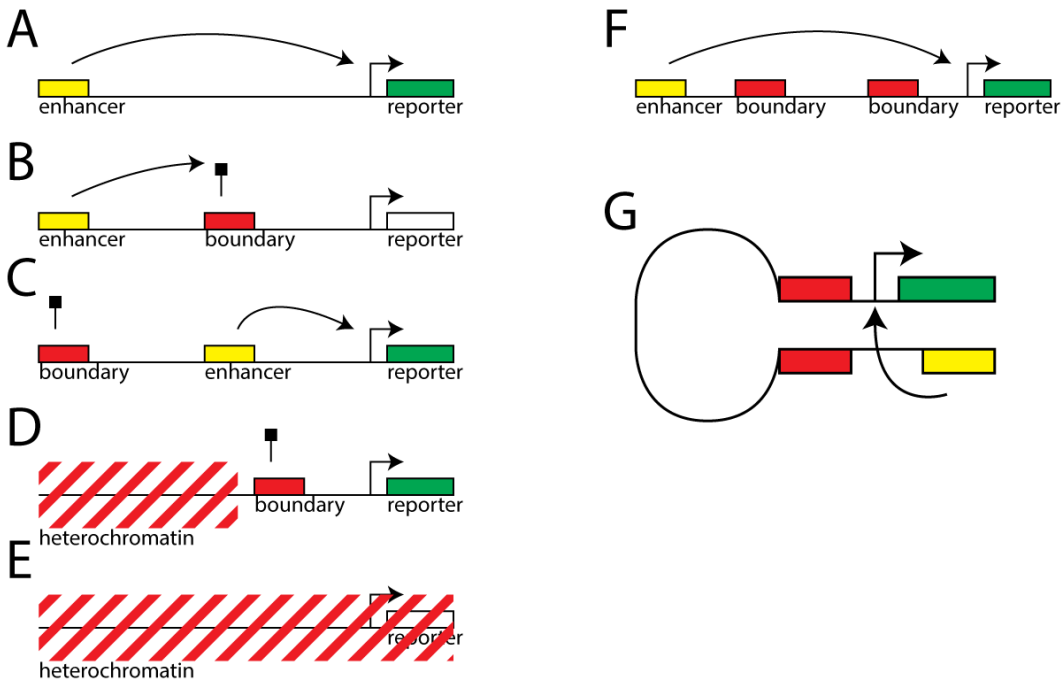


Figure 1.12: Boundary element transgenic assays - A: In a control transgenic construct a chosen enhancer (yellow box) is able to activate a reporter gene (green box for active transcription, promoter as a black arrow) posed on the same transgene. B,C: Enhancer blocking assay. When a boundary element (red box) is positioned in between the enhancer and the reporter gene from panel (A), the interaction between enhancer and promoter is disrupted (the reporter gene the box is white and not green anymore). In controls experiments, the boundary is positioned away from the enhancer and promoter (C). If the boundary is a boundary and not a silencer, this should reverse the effect seen in (B). D,E: Barrier function assay. Inserting a boundary element close to an active reporter gene, the activity of the reporter gene is protected by the boundary element from heterochromatin spreading (red dashed lines). Removal of the boundary leads to an inactivation of the gene due to heterochromatin spreading over the gene region. F,G: Boundary element skipping and long distance interaction. When an additional boundary element is inserted in the situation described in panel (B) the enhancer blocking activity of the boundary is reverted and the reporter gene is activated. The most likely mechanism for such a phenomenon is illustrated in panel (G): The two boundary elements are able to establish a DNA-DNA long distance interaction, and therefore the enhancer is brought in proximity of the promoter of the reporter gene.

fact, such insertions are often between an enhancer and its target promoter (Geyer and Corces, 1992). Indeed, when studying *gypsy*-induced mutant alleles in *su(Hw)* mutant background, their phenotypes can be reverted (Modolell et al., 1983). Not only did Su(Hw)-dependent insulators show enhancer blocking, but when flanking transgenic insertions they could shield reporter genes from chromosomal position effects (Roseman et al., 1993). This shielding effect may depend on the barrier function or on insulator/insulator looping (Comet et al., 2006; Sigrist and Pirrotta, 1997).

A recent study analyzed the Su(Hw) binding sites in the *Drosophila* genome using chromatin immunoprecipitation (ChIP). In contrast to the clustered Su(Hw)-binding sites on the *gypsy* transposon, endogenous genomic Su(Hw) binding sites frequently consist of a single Su(Hw) binding motif. The sites analyzed show a constant Su(Hw) binding throughout development and in diverse tissues. The authors also analyzed the transcriptional changes in a *su(Hw)* mutant background. They observed a widespread and general change in gene expression. The genes that were differentially expressed (compared to wild type) were in most cases not in proximity of *gypsy* insertions (and thus in proximity of clustered Su(Hw) sites) but close to single Su(Hw) sites. These results support a model that Su(Hw) is responsible to maintain a constant genomic architecture (Adryan et al., 2007).

1.3.4 The scs and scs' boundaries

scs and scs' (Specialized Chromatin Structure domain boundary elements) were the first boundaries described in *Drosophila* (Kellum and Schedl, 1991). Those two boundaries are flanking a 14kb region containing five genes (two of them coding for heat shock 70 proteins). They interact with each other and thus probably create a chromatin loop (Kyrchanova et al., 2013). This insulator/ insulator interaction is mediated by the two insulator-associated proteins BEAF-32 and Zw5 (Gaszner et al., 1999; Zhao et al., 1995).

1.4 Boundaries in the BX-C

1.4.1 *Fab-7*, the model boundary in the BX-C

The best studied boundary in the BX-C is *Fab-7*. This boundary separates the *iab-6* from the *iab-7* region. *iab-7* is responsible for the identity of segment A7 which is not formed in wild type adult male flies. *iab-6* is responsible for the identity of segment A6, which is darkly pigmented in adult male flies. Deletion of the *Fab-7* boundary leads to a well visible phenotype: A6 is transformed into A7 (Gyurkovics et al., 1990). Therefore, only one darkly pigmented tergite (A5) is visible in mutant male flies, and not two (A5 and A6 in wild type flies, see **Figure 1.13 A**). The first description of the *Fab-7* boundary restricted its functional DNA sequences to a 4kb region between the *iab-6* and *iab-7* regions.

An insertion of a transposable element in the region (called "bluetail", see **Figure 1.13 B**, Galloni et al. (1993)) was instrumental to genetically dissect this boundary with a good resolution at the endogenous locus. Deletions caused by imprecise excision of the "bluetail" transposon enabled the determination of three different phenotypical classes (I,II, and III) (Mihaly et al., 1997).

The three phenotypes correlate with the presence of three DNase hypersensitive regions in the *Fab-7* boundary (HS1, HS2, and HS3. **Figure 1.13 B,C**). HS stands for DNase I hypersensitive site, and are DNA sequences that can be digested by the DNase I enzyme due to open chromatin. This open chromatin state is a sign of transcription factor binding. Deletion of all three regions together leads to a class I phenotype, meaning a complete transformation of A6 into A7 as observed in *Fab-7*^{1}. Deletion of HS1 and HS2 leads to a class II phenotype: a partial transformation of A6 into A7. Such a phenotype can be modulated by Polycomb group or Trithorax group alleles. The last class, class III (removal of HS3) shows no discernible phenotype (**Figure 1.13 D**).

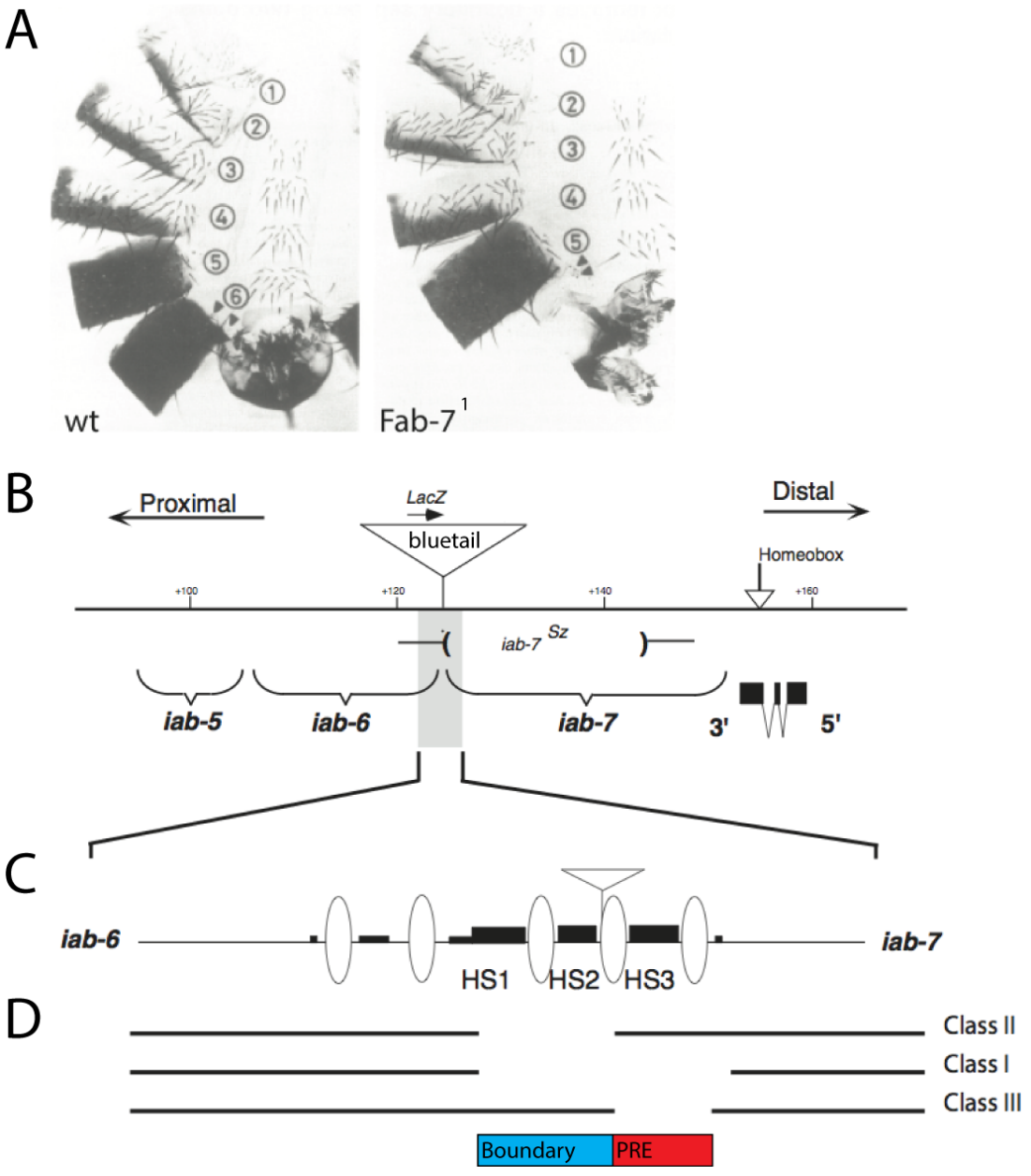
The sequences in HS3 are necessary to see a modulation of *Abd-B* phenotypes due to Polycomb and Trithorax alleles. Thus, it was postulated that HS3 contains a PRE. As gain-of-function phenotypes were only visible when HS1 and HS2 are

deleted, it was postulated that those sequences contain the boundary element (Mihaly et al., 1997).

Recently, a more precise dissection of the Fab-7 boundary allowed a higher resolution on the functional units. The insulator function associated previously to HS1 and HS2 could be restricted to HS1 (Wolle et al., 2015). The function of the boundary depends on the binding of GAF. Removal of those binding sites destroys the enhancer blocking function of the boundary (Schweinsberg et al., 2004).

Furthermore, *Fab-7* facilitates interaction between regulatory regions and gene promoters: it was shown by the use of the DamID method that this boundary

Figure 1.13 *(following page)*: Dissection of the *Fab-7* boundary - A: Classical *Fab-7*^{1} phenotype. A spreaded preparation of an adult male abdominal cuticle is shown. wild type males have 6 visible abdominal segments (A1 to A6). The tergites are differentially pigmented. A1 remains unpigmented. A2-A4 develop a stripe of black pigment along the posterior border of each tergite. A5 and A6 tergites are completely black. On the ventral side, male sternites A1 and A6 are devoid of sternite bristles. They are only present in A2 to A5. In *Fab-7*^{1} males these morphological features remain essentially unchanged, except for the absence of A6 tergites. Arrowheads point to the 6th and 7th tracheal openings. Modified from (Gyurkovics et al., 1990). B: Representation of the *Fab-7* region with the insertion of the bluetail transposon. In brackets the different *cis*-regulatory regions are indicated (*iab-5* to *iab-7*). The *Abd-B* gene is indicated by one of its transcripts. The position of the homeobox is also indicated by a triangle. The location of the bluetail transposable element is indicated by the triangle on top of the black line, the orientation of the *LacZ* reporter gene is shown by an arrow. Modified from (Mihaly et al., 1997). C: Characterization of the *Fab-7* element. The bluetail transposon is marked again as a triangle, the black boxes mark DNAse hypersensitive sites (smaller boxes mark regions that showed weaker sensitivity to DNase) while the ovals represent the putative positions of the nucleosomes. Modified from (Mihaly et al., 1997). D: Genetic dissection of *Fab-7*. The three phenotypical classes of deletions are marked. Class I deletes both the PRE and insulator region of *Fab-7*. Class II deletes the insulator region on the left side of the bluetail insertion. Class III deletions (the deletions are localized on the right of the bluetail insertion) do not show the typical *Fab-7* phenotype, but can enhance the phenotype of Class I and II in trans. Therefore, class III was predicted to contain the *Fab-7* PRE, while Class I contains the proper boundary element. Blue box symbolizes the predicted location of boundary element, red box the location of PRE. Modified from (Mihaly et al., 1997).


A
wt
Fab-7 1
B
Proximal
LacZ
bluetail
Distal
Homeobox
+100
+120
+140
+160
iab-7 Sz
iab-5
iab-6
iab-7
3'
5'
C
iab-6
iab-7
HS1
HS2
HS3
D
Class II
Class I
Class III
Boundary
PRE

interacts with the *Abd-B* promoter. The synthetic binding of a DNA methylase enzyme (DamID method) to the *Fab-7* region methylates DNA close to the *Abd-B* promoter (Cléard et al., 2006). This experiment proves that a boundary in the BX-C indeed is able to form loops in the endogenous genomic location, confirming transgenic assays that attributed the long distance interaction function to boundary elements.

Given the modularity and similarity of *iab* regions, the *Fab-7* boundary can be replaced by the *Fab-8* boundary. To observe a complete rescue, both the *iab-7* PTS and *Fab-8* sequences have to be inserted in the *Fab-7* locus, replacing the endogenous *Fab-7* boundary element. These experiments demonstrate the functional similarity of the boundaries in the *Abd-B* region, even though their sequence similarity is very scarce (Kyrchanova et al., 2016). Indeed, the opposite experiment works as well: the *Fab-7* boundary can replace the *Fab-8* boundary, even though *Fab-7* does not contain any CTCF binding sites (in opposition to the other boundaries in the *Abd-B* gene) (Iampietro et al., 2008). Other boundaries, not originally from the BX-C (like *scs* or *gypsy*) are not able to substitute the *Fab-7* boundary: replacement of the *Fab-7* boundary with *scs* or *gypsy* leads to a loss of function (segment A6 transformed into A5 in adult males) (Hogga et al., 2001). In conclusion, the boundaries in the BX-C seem to be an independent subclass of boundaries, that are similar in function to each other.

1.4.2 Other boundaries described in the BX-C

Further boundaries were isolated from the BX-C. The ones that were experimentally described are *Fub* (separating the *Ubx bxd/pbx* region from the *abd-A iab-2* region, Bender and Lucas (2013)), *Mcp* (separating *abd-A iab-4* from *Abd-B iab-5*, Karch et al. (1994)), *Fab-6* (separating *iab-5* from *iab-6, Mihaly et al. (2006)*), and *Fab-8* (separating *iab-7* from *iab-8,* Barges et al. (2000)).

The interaction between *Mcp*, *Fab-6*, *Fab-7*, *Fab-8*, and the *Abd-B* promoter was studied in transgenic assays that tested for the ability of boundary element pairs to show boundary element bypass as described in **Figure 1.12 F** (Kyrchanova et al., 2011). All those boundaries interact with each other in transgenic constructs, and enable enhancer/promoter interactions in line with the DNA loop

model. Furthermore, in similar set-ups all boundaries except *Mcp* interact with the *Abd-B* promoter. Those interactions where dependent on the action of CTCF (but not exclusively, as *Fab-7* does not bind CTCF), as deletion of CTCF binding sites on the boundary elements studied reduced the interactions between the boundaries themselves and between the boundaries and the promoter. The importance of CTCF binding sites was shown also in other studies, e.g. it confers enhancer blocking activity to the *Fab-8* boundary (Kyrchanova et al., 2016). Interestingly, CTCF is the only known insulator protein conserved from *Drosophila melanogaster* to *Homo sapiens* (Moon et al., 2005).

1.5 The *Mcp* boundary in the BX-C

1.5.1 Deletions and their phenotypes in the *Mcp* region

The first *Mcp* allele that was isolated (Mcp^1) is dominant and was found to carry a 3.5kb deletion between the *iab-4* and *iab-5* regions (Karch et al., 1990). It has a striking phenotype in adult males: in wild type flies only A5 and A6 are pigmented. In *Mcp* mutants, A4 is pigmented as well (**Figure 1.14 A**). Thus, *Mcp* induces a homeotic transformation in which A4 acquires A5 identity (Celniker et al., 1990). This transformation is triggered by an anterior expansion of *Abd-B* expression. This change in expression is best visible in the ventral nerve cord of germ band retracted embryos. Antibody staining against Abd-B in mutant embryos shows an anterior *Abd-B* expression border in PS9 compared to PS10 in wild type (**Figure 1.14 B**, Crosby et al. (1993)).

Further endogenous deletions in the *Mcp* locus helped to narrow down the boundary region. Mcp^{B116} is a smaller deletion (around 800bp) that removes a sequence that is contained in the Mcp^1 deletion (**Figure 1.15**). This deletion shows a similar phenotype as Mcp^1. Combined with the Mcp^{H27} deletion the boundary can be restricted to a region of around 400bp (Karch et al., 1994). As will be discussed below, this 400bp interval is similarly organized as the *Fab-7* region: on the *iab-4* side, there is a boundary element and on the *iab-5* side, there is a PRE.

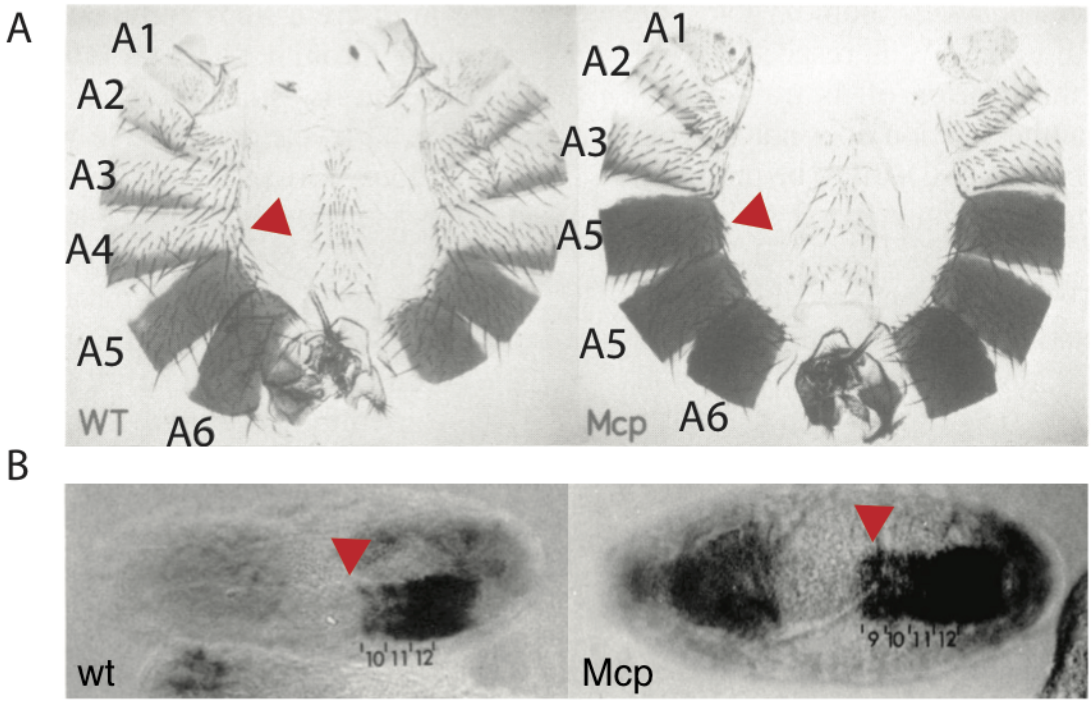


Figure 1.14: *Mcp* phenotype - A: Adult abdominal cuticle spreads from wild type and *Mcp* males are depicted. Note the change in phenotype in the A4 tergite. Arrowheads mark the same segment, in wild type this segment has an A4 identity, in mutants an A5 identity. Modified from (Celniker et al., 1990). B: Ventral views of whole mount embryos hybridized with a probe against *Abd-B* transcripts are shown (anterior to the right). In wild type animals the gene is transcribed only until parasegment 10. In *Mcp* mutants the expression of *Abd-B* is broader, as it is visible in the embryonic *in situs*: *Abd-B* is expressed from posterior until parasegment 9. Arrowheads mark the anterior expression boundary of *Abd-B*. In adult animals this leads to an abnormal pigmentation seen in panel A of this figure. Modified from (Crosby et al., 1993).

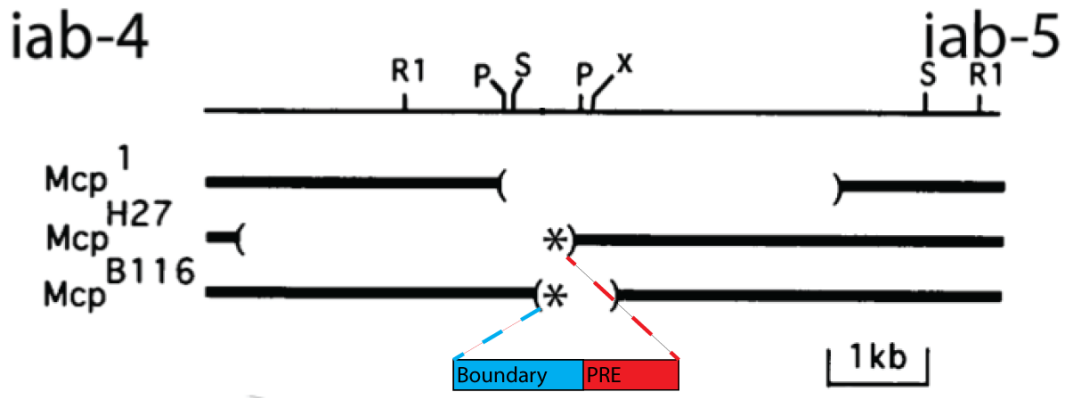


Figure 1.15: Classical *Mcp* deletions - Genetic characterization of *Mcp*. *Mcp* is located between *iab-4* and *iab-5*. The line at the top symbolize the DNA. Relevant restriction sites are marked (R1: EcoRI; P: PstI; S:SalI; X:XbaI). Below the DNA, the extensions of three *Mcp* deletions are shown. The minimal deletion interval defined by the Mcp^{H27} and Mcp^{B116} breakpoints (marked with an asterisk) is 415bp in size. The blue and red box symbolize the boundary element and PRE in the *Mcp* region. Both those elements are deleted in all alleles. Modified from (Karch et al., 1994).

1.5.2 Dissection of the *Mcp* PRE in transgenes

Early studies associated an enhancer silencing activity to the *Mcp* boundary (Busturia and Bienz, 1993). In those experiments, a DNA fragment containing *Mcp* was tested for its ability of silencing the *bxd* enhancer from the *Ubx* locus in transgenes. Indeed, *Mcp* was able to restrict the expression pattern of the *bxd* enhancer. This experimental setup is analogous to the one introduced in **Section 1.2.3**, which describes a widely used assay for the identification of PRE containing DNA fragments. Based on the observed maintenance activity of *Mcp*, it could be concluded that the *Mcp* fragment used by Busturia and Bienz contains a PRE.

Later on, this silencing activity was assigned to the four Pho binding sites on the distal side of the *Mcp* element (the side facing the *iab-5* region) (Busturia et al., 2001). In this experiment a *pbx* enhancer (specific for PS6, therefore active in the posterior half of the haltere disc) was flanked by a *Mcp* sequence on a *LacZ* reporter plasmid (**Figure 1.16 A**). *Mcp* is inserted between the *pbx* enhancer and the *Ubx-LacZ* reporter. As *Mcp* does indeed contain a PRE, the

expression in the posterior compartment of the haltere disc is maintained during development. This construct was inserted randomly into the genome using P transposable element transgenesis. As the genomic background varies in such insertions, various different insertions with the same construct were analyzed and scored, as genomic background effects may interfere with the enhancer and/or PRE activity. Flies that show a restricted *LacZ* expression in the posterior haltere disc compartment were scored as "silenced". On the opposite, flies showing an expanded *LacZ* expression in the anterior haltere compartment and in the wing disc were scored as "derepressed".

When a 822bp SalI-XbaI *Mcp* sequence is used (see restriction sites in **Figure 1.15**), the *pbx* signal domain is mainly not expanded (15 out of 23 "silenced" in **Figure 1.16 B**). Instead, when all Pho binding sites are removed from the PRE, such a silencing does more often not happen and the enhancer signal is expanded (8 out of 9 "derepressed" in **Figure 1.16 B**).

However, a 102bp fragment (called MCP1 in **Figure 1.16 B,C**) containing the four Pho binding sites is not sufficient to restrict the *pbx* pattern: 17/18 lines show anterior expansion (see **Figure 1.16 C**). Robust *Mcp* mediated maintenance requires two GAF binding sites (compare MCP7 and MCP7* in **Figure 1.16 C**). Hence, the *Mcp* PRE is defined by the 137bp contained in MCP7.

In conclusion, *Mcp* contains a PRE and its activity is dependent on the Pho and GAF binding site in transgenic assays.

1.5.3 Long distance interaction of *Mcp* boundary pairs

One additional feature of the *Mcp* boundary element is its capacity to interact with a second *Mcp* boundary element over very large genomic distances. Transgenic insertions at different genomic locations have been shown to physically and genetically interact with each other, sometimes even when the two insertions are on different chromosome arms or on different chromosomes (Müller et al., 1999; Vazquez et al., 2006) (**Figure 1.17**).

Mcp boundary mediated interactions could only be monitored as long as the *Mcp* PRE (or a heterologous PRE, see Li et al. (2011)) is present on the same transgene together with a *mini-white* reporter. *mini-white* is a modified *white*

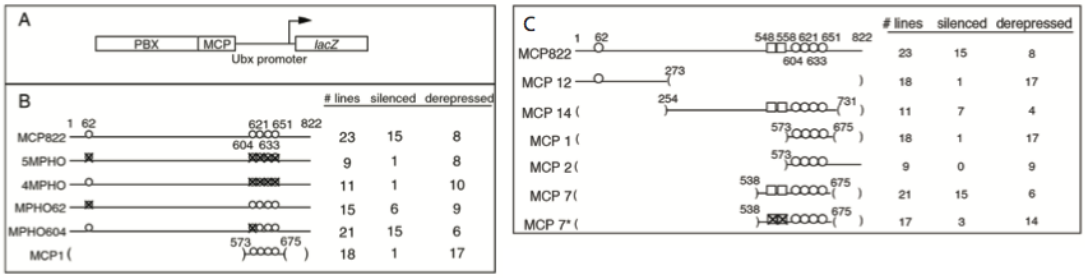


Figure 1.16: Dissection of the *Mcp* PRE - Pho and GAF are responsible for the silencing activity of *Mcp*. Through a transgenic assay the silencing activity of *Mcp* was tested. In different constructs the single Pho and GAF sites were mutated, and the minimal region for silencing was identified. A: The general design of the constructs is illustrated: the *pbx* enhancer was fused to the different *Mcp* elements (with mutated sites, or shorter versions). The *LacZ* gene monitors the activity of the *pbx* enhancer in haltere discs. B,C: With the full length *Mcp* the expression is restricted to the posterior part of the haltere disc ("silenced" in the figure). With a mutated and non functional PRE, the *LacZ* expression will be all over the haltere and wing disc (this is the meaning of "derepressed" in the figure). Round signs on the DNA line represent PHO sites, squares represent GAF sites. Modified from (Busturia et al., 2001).

gene lacking all the tissue specific enhancers. Nevertheless, *mini-white* is able to rescue *white* mutant flies in heterozygous conditions, with variations due to genomic position effects.

But what is the role of the *Mcp* PRE in monitoring long distance interaction? When a transgene containing *mini-white* is observed in *white* mutant background, the intensity of the eye color is gene dosage dependent: heterozygous flies have less eye color than homozygous flies. Adding a PRE to the same transgene usually has two consequences: 1) heterozygous flies often display less *mini-white* activity and 2) the gene dosage dependency is very often lost. Indeed, heterozygous flies accumulate more eye pigment than their homozygous siblings (**Figure 1.17 A**). Therefore, it is postulated that PREs work best when paired (reviewed in Kassis (2002)).

When two *Mcp* boundaries (containing the PRE) are paired, an adjacent *mini-white* cassette is silenced. Through this assay, one can assess if two *Mcp* boundaries are interacting *in trans* even when inserted at different genomic loci (Müller et al., 1999; Vazquez et al., 2006). Such interaction is orchestrated by

the insulator subunit of the *Mcp* boundary, the *Mcp* PRE is dispensable for such a long distance function (Li et al., 2011).

The fact that two *Mcp* boundaries can interact was later verified *in cis* (using a boundary element bypass assay, described in **Figure 1.12 F,G**) (Gruzdeva et al., 2005). A 340bp long fragment was sufficient to induce boundary pairing. Such a fragment contained the CTCF binding site, while the PRE sequences were absent (**Figure 1.17 C**).

Taken together, those experiments demonstrate that the *Mcp* boundary is able to pair *in cis* and *in trans*, even over long distances, and that this interaction is dependent on the insulator unit of the boundary, and not on the PRE.

1.5.4 Directionality of Mcp boundary interactions

In the previous section I've gone through the evidence that the *Mcp* boundaries establish interactions with themselves when multiple *Mcp* boundaries are distributed on chromosomes. A further study raised an additional aspect: *Mcp* boundaries have a directionality. When a pair of *Mcp* elements is inserted in the same transgenic construct, their orientation determines how the established loop would look like, and if the enhancer can contact the promoter in an optimal way (see **Figure 1.18**, Kyrchanova et al. (2007)).

This experiment further demonstrated that the PRE is not necessary for chromatin/chromatin interactions established by the *Mcp* boundary, as Kyrchanova and colleagues are able to induce looping in their assays with *Mcp* boundaries not containing any PRE. Furthermore, the directionality raise an interesting point because in mammals it was postulated that the orientation of CTCF binding sites might have a function (Guo et al., 2015).

However, the question remains whether this directionality of *Mcp* interactions plays a role in the endogenous location, or if it is a feature of transgenic assays.

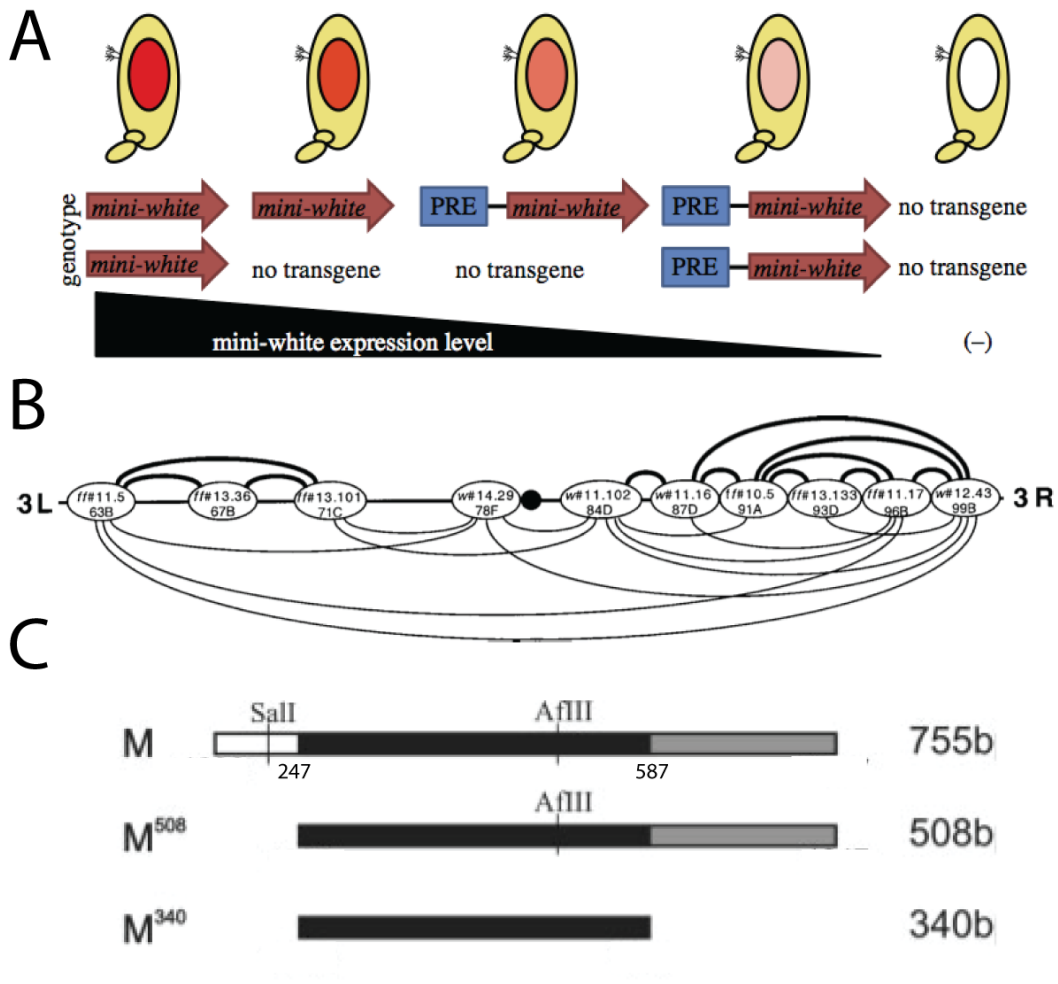


Figure 1.17: Pairing sensitivity as a readout for long distance interaction - A: Illustration of pairing sensitivity. The eye color of flies is represented on top. Below an illustration of the genotypes, the presence of *mini-white* (red arrow) and PRE (blue box) on transgenes on both sister chromosomes. At the bottom the level of *mini-white* expression. Modified from (McElroy et al., 2014). B: The *Mcp* element can establish long distance interactions. Ten P{*Mcp*, *miniwhite*} insertions and their genomic location on chromosome 3 are shown. Four of them are on chromosome arm 3L, six on 3R. The black dot in the center symbolizes the centromere. Pairs of P{*Mcp*, *miniwhite*} inserts are connected by a thick line, a thin line or no line at all. They represent the following scores respectively: strong interaction (weaker eye color than both parents), weak interaction (weaker eye color than one of the two parents) or no interaction (same or darker eye color as parents). Modified from (Müller et al., 1999). C: PstI-PstI *Mcp* fragment. The putative insulator is marked in black, while the PRE is marked in grey on the 755bp long *Mcp* fragment used in the Gruzdeva et al. study. Numbers mark the distance from the proximal PstI site. Both the 508bp and 340bp derivative constructs were able to mediate long distance interaction. Modified from (Gruzdeva et al., 2005).

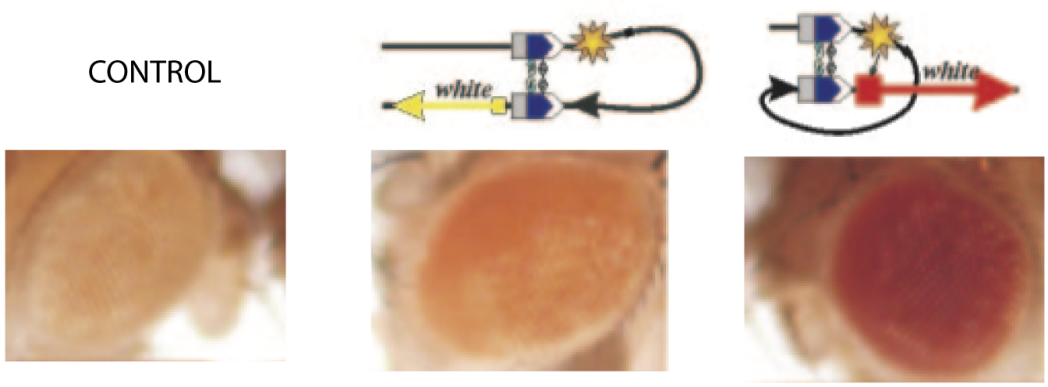


Figure 1.18: Orientation dependency of *Mcp* boundary element pairing *in cis* - *Mcp* is orientation dependent *in cis*. On the left a control eye color. In the middle the inserted transgene contains two *Mcp* boundaries facing each other flanking an enhancer. The promoter and the coding sequence of the *white* reporter gene is laying outside the putative boundary loop. In this case the eye color is not substantially darker than the control. When the *Mcp* boundaries are in the same orientation on the transgene (on the right) the enhancer is positioned in a way that it can contact the *white* promoter, leading to clearly darker eye colors. Yellow and red arrows represent the *mini-white* reporter cassette with respective activity (yellow weak, red strong). Star represents the enhancer. Blue arrow box represents the boundary element *Mcp* and its orientation. Modified from (Kyrchanova et al., 2007).

1.6 Current model of *Abd-B* gene regulation

Combining the "open for business" concept with the chromatin/chromatin interaction capacity of boundaries and the knowledge of other *cis*-regulatory elements described beforehand, we can nowadays postulate a model on how the gene *Abd-B* is regulated.

According to the "open for business" model, genomic regions are opened according to their activities (reviewed in Maeda and Karch (2015)). Indeed, Bowman and colleagues showed that BX-C chromatin is more open (less regions silenced by H3K27m3) the more posterior the cells are positioned along the A/P axis. This suggests that chromatin is opened up along the collinearity rule of the BX-C. When the initiator in one region (in the *Abd-B* gene *iab-5* to *iab-8*) is activated this will probably lead to an inhibition of PRE activity on this region (Bowman et al., 2014). Thus, initiators are activated in a linear way from anterior to posterior corresponding to the location of the cell in the A/P body axis. This activation will lead to a less and less repressed BX-C chromatin the more posterior the cell is localized in the embryo.

But what does separate the different regions? How is it controlled that Polycomb silences only blocks of DNA? The results of Bowman and colleagues show very sharp boundaries of H3K27me3 at the location of chromatin boundaries.

The domains in the BX-C remind of a general genomic feature: it is known that genomic regions can be physically separated from each other. Chromosomes by themselves occupy discrete regions in the nucleus (**Figure 1.19 A**). Single chromosomes are themselves divided into domains named topological associated domains (TADs). Those TADs can either be associated to the nuclear lamina (and repressed) or not (**Figure 1.19 B**). TADs were defined using chromosome capture methods, where DNA-DNA interactions can be biochemically assessed. Inside TADs DNA-DNA interactions are very common, while across TAD boundaries those interactions are greatly reduced (**Figure 1.16 C**, reviewed in Matharu and Ahituv (2015)).

One difference between general TADs and the independent regions in the BX-C is that the DNA between most of the *Abd-B* boundaries does not contain coding sequences but just *cis*-regulatory elements (**Figure 1.19 D**). The current model

of the BX-C would suggest that the different boundaries are all in close physical proximity to each other in the nucleus. This interaction might be lost when the initiator in one of the *iabs* is activated. This would allow the initiator and other *cis*-regulatory elements in the *iab* region to interact with the *Abd-B* promoter. The loops that would not be opened are silenced by polycomb proteins (see green circles in **Figure 1.19 D**).

Those mechanisms (activation according to initiator activity and silencing of inactivated regions) would lead to a different DNA conformation according to the segment in which the cell is positioned, as this topological information is given by the initiator itself (**Figure 1.19 D**): in segment A4 the whole *Abd-B* gene is silenced and all boundary elements are in contact. Moving posteriorly, boundary elements are released one-by-one segmentally from contacting the other *Abd-B* boundaries. In A5, *Mcp* is released and the regulatory elements in the *iab-5* region can interact with the *Abd-B* promoter. In A6, both *Mcp* and *Fab-6* are released, and the regulatory elements in *iab-6* can contact the *Abd-B* promoter.

As described in previous chapters, those long distance DNA-DNA interactions are thought to be mediated by insulator protein pairing (reviewed in Bushey et al. (2009)). In the case of the *Abd-B* regions it is likely that those loops are created by CTCF pairing (aided by CP190). Nevertheless, there is the exception of the *Fab-7* boundary that does not bind CTCF, therefore it might be that different insulator proteins interact with each other through common binding of CP190 (Bushey et al., 2009).

Still, nowadays we have no definitive data that would completely validate this sequence of events in *Abd-B*. Nevertheless, interactions between boundaries, sharp limits in polycomb modifications, and the rich world of genetic analysis data collected in the last decades in the BX-C would indicate this way. Further development of biochemical methods and tissue isolation may clarify the situation even better in the future.

1.7 Methods

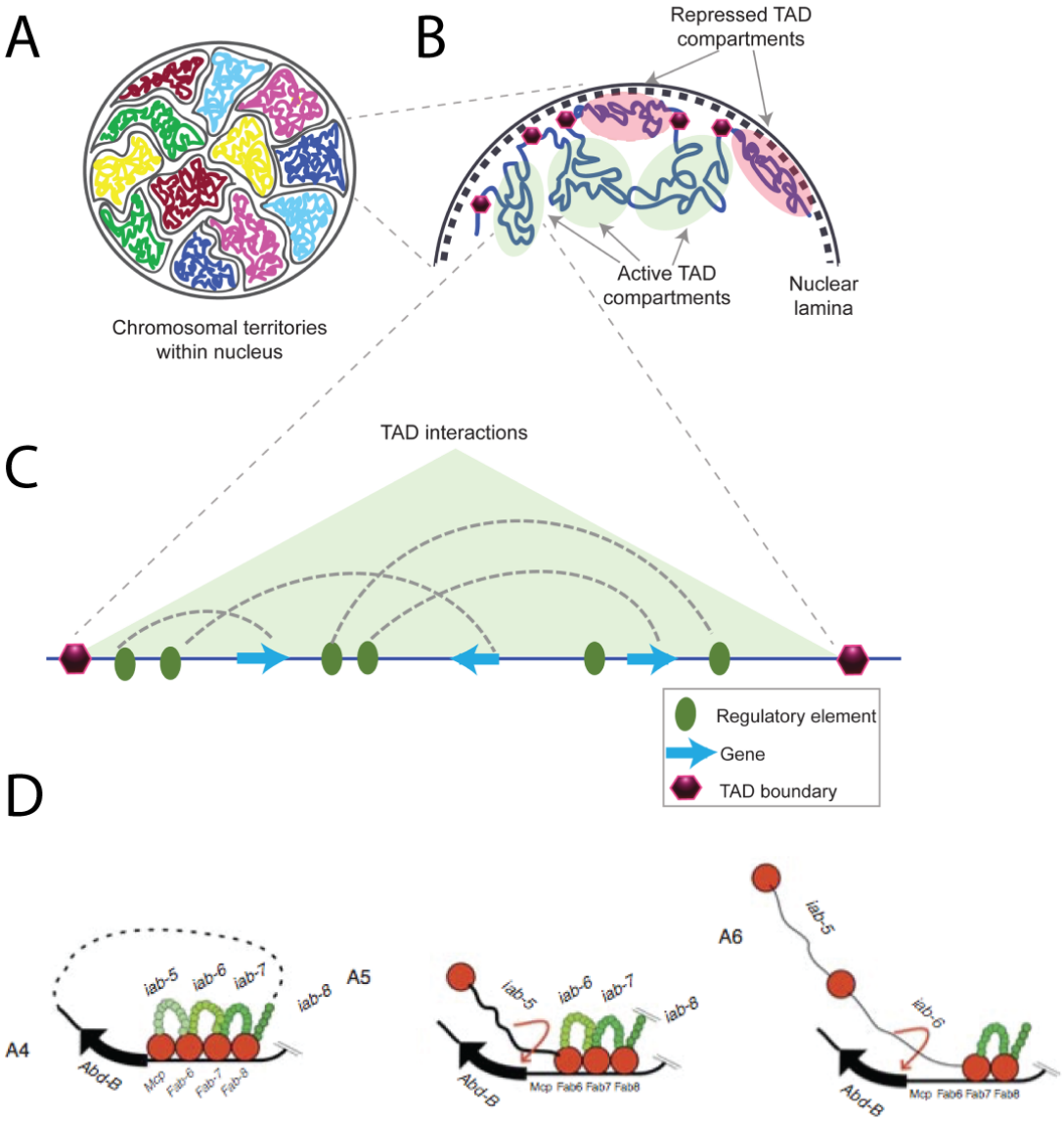
1.7.1 Use of phiC31 integrase in *Drosophila*

The integration of modified DNA sequences into the genome of an animal is crucial for modern genetic studies. The first transgenesis methods were based on transposable elements, in *Drosophila* most commonly based on P-Elements (Spradling and Rubin, 1982). Those systems have one crucial disadvantage: the location of integration can not be influenced. For this reason, in *Drosophila* the phiC31-based integration system has been established for precise integration of DNA in specific genomic loci (Bischof et al., 2007). phiC31-integrase encodes an integrase responsible for the recombination of two short DNA sequences, bacterial attachment site (attB) and phage attachement site (attP).

Previously introduced attP or attB sites in the genome can be used to re-integrate genetic material at that one position in a precise and efficient way. This allows studies of multiple modified sequences at the same location and thus the conditions are normalized for position effects.

In this thesis, the Recombination Mediated Cassette Exchange method (RMCE) is used (Bateman et al., 2006). In the case of RMCE, two attachment sites are used (i.e. two attB sites on the genome and two attP on the plasmid). Thereby,

Figure 1.19 *(following page)*: Topologically associated domains as an analogy for BX-C DNA loops - A: Chromosomes in interphase usually occupy defined regions inside the nucleus (different colors for different chromosomes). Modified from (Matharu and Ahituv, 2015). B: Each chromosome is predicted to be subdivided into multiple topological associated domains (TADs). Usually, lamina associated DNA is silenced, while DNA reaching in the middle of the nucleus is active. Modified from (Matharu and Ahituv, 2015). C: Each TAD is composed of two delimiting boundaries and can contain multiple coding sequences and multiple *cis*-regulatory elements. Inside TADs there can be DNA-DNA interaction between sequences that are not close to each other. Modified from (Matharu and Ahituv, 2015). D: Topology of the *Abd-B* genome region (*Abd-B* coding sequence as black arrow) in different abdominal segments (A4 to A6). In A4 there is no *Abd-B* expression in wild type flies. Therefore all the inter-boundary regions (boundaries as red circles) are silenced by polycomb induced modifications (green circles). Modified from (Maeda and Karch, 2009).


A
B
Repressed TAD compartments
Active TAD compartments
Nuclear lamina
Chromosomal territories within nucleus
C
TAD interactions
Regulatory element
Gene
TAD boundary
D
A4
A5
A6
iab-5
iab-6
iab-7
iab-8
Abd-B
Mcp
Fab6
Fab7
Fab8

instead of inserting the entire injected plasmid into the genome, only the genetic material between the attachment sites is inserted.

1.7.2 The CRISPR/Cas9 method for genome engineering

The story on how CRISPR/Cas9 has gone from being first described in 1987 (Ishino et al., 1987) to being the poster child of next generation biology and medicine starting from 2012 onwards (when CRISPR/Cas9 was described as a RNA-guided DNA nuclease, Jinek et al. (2012)) is a tale that illustrates the long road from basic research to application, and how the nature of those applications can not be predicted (reviewed in Doudna and Charpentier (2014)).

After the discovery of a series of short direct repeats interspaced with short sequences in the genome of *Escherichia coli* (Ishino et al., 1987), almost twenty years passed during which such sequences were found in many other bacteria and archaea, but no insight into the role of such sequences was gained. In 2005, several publications suggested that the Clustered Regularly Interspaced Short Palindrome Repeats (CRISPR) contained spacers that had extrachromosomal origin. Those sequences originated from phages that infected those bacteria (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). The fact that bacteria integrate phage DNA into specific regions of their chromosome indicated that these repeated sequences might be part of a bacterial immune system (Makarova et al., 2006). Indeed, such an acquired bacterial immunity was demonstrated with a phage infection experiment in *Streptococcus thermophilus* (Barrangou et al., 2007).

The finding that only one *cas* gene (coding for the Cas9 protein, a endonuclease able to induce double strand breaks) is needed for such bacterial defense function (Sapranauskas et al., 2011) opened the door for the development of a method that would allow directed DNA cutting *in vitro* (Jinek et al., 2012) and *in vivo* (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). In these studies, the two RNA components of the system (crRNA and tracrRNA, two small RNA molecules that cooperatively are needed to bring Cas9 to the genome) were combined to one molecule (guide RNA, or gRNA. **Figure 1.20 A,B**).

The method developed is based on the Cas9 protein and a gRNA. gRNA contains a guide sequence that is designed individually with a specific locus in mind: the RNA sequence is complementary to the target DNA. A restriction in designing gRNAs is the presence of a PAM sequence (Protospacer Adjacent Motif, see **Figure 1.20 A**) on the genomic DNA. This PAM sequence is necessary for the Cas9 activity. Interestingly, the PAM sequence changes according to the Cas9 donor species. For the commonly used *Streptococcus aureus* Cas9, the PAM consists of 5'-NGG-3' nucleotides (Kleinstiver et al., 2015).

In *Drosophila*, CRISPR/Cas9 can be used to create precise clean deletions using one pair of gRNA flanking the target region (**Figure 1.20 C**) or the additional usage of a plasmid containing homology arms to integrate genetic material at the location of choice (**Figure 1.20 D**, Gratz et al. (2013a)).

Those two different genome modification mechanisms rely on two DNA repair mechanism: non-homologous-end-joining (NHEJ) for clean deletions and homologous recombination (HR) for integration of foreign DNA (Gratz et al., 2013a).

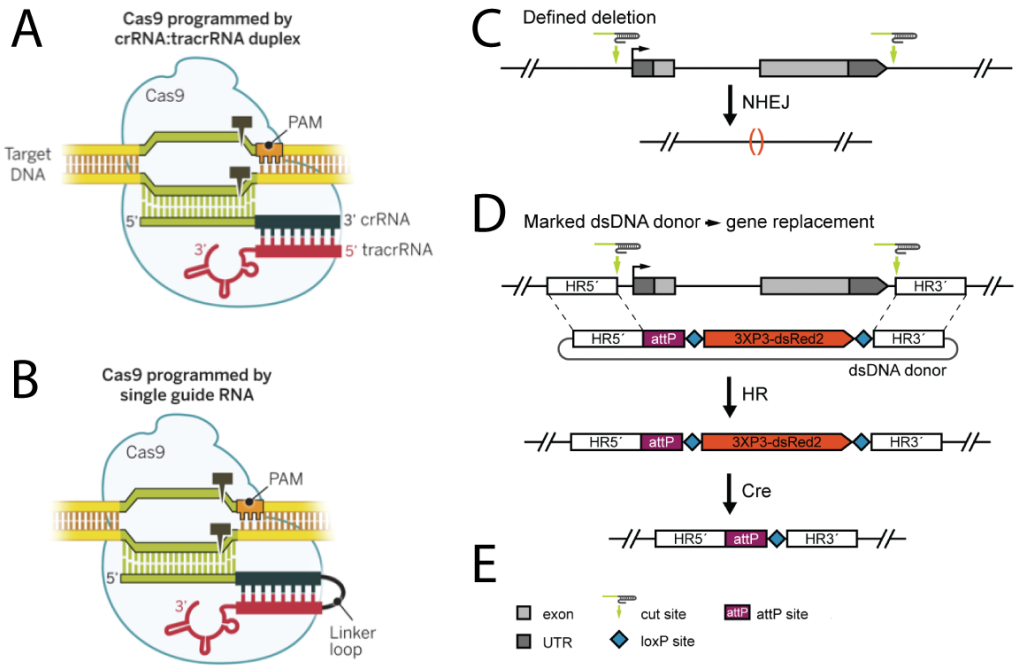


Figure 1.20: CRISPR/Cas9 method representation and applications - A: Interaction between the Cas9 protein, the two involved RNA molecules (crRNA and tracrRNA, together they guide the Cas9 enzyme to the correct genomic location, where a PAM sequence is present) and the target DNA. The Cas9 protein opens the double stranded DNA helix and induces breaks in both strands (dark grey arrowhead and box). Modified from (Doudna and Charpentier, 2014). B: Engineered RNA molecules made scientific usage more convenient: fusing the crRNA and tracrRNA molecules depicted in panel (A) with a linker made the CRISPR/Cas9 method usable with just two components (the Cas9 protein and the chimeric RNA molecule). Modified from (Doudna and Charpentier, 2014). C: Use of the CRISPR/Cas9 method to induce specific deletions in a genomic locus of choice. Grey boxes show the structure of a hypothetical transcript (the promoter shown as a black arrow). Green arrows show the position of the gRNA/Cas9 binding locations. The cutting of the two DNA strands at those locations will activate the Non-Homologous-End-Joining (NHEJ) DNA repair mechanism. In some cases this will create a chromosome where the DNA sequences in between the gRNA/Cas9 binding sites are deleted (illustrated by red brackets). Modified from (Gratz et al., 2013b). D: Another DNA repair mechanism that can be used in combination with the CRISPR/Cas9 method is Homologous Recombination (HR). Providing a plasmid containing homologous sequences to the target DNA outside the gRNA/Cas9 binding sites will integrate the DNA sequences in between the homology arms on the plasmid at that specific genomic location, substituting the sequences in between the gRNA/Cas9 binding on the genome. This way a positive selection is feasible (in the case illustrated, transformands would have red fluorescent eyes). Modified from (Gratz et al., 2013b). E: Legend for panels (C) and (D). Modified from (Gratz et al., 2013b).

2

Aims of the project

DNA fragments can have different characteristics when studied at their natural position or in transgenes. Many studies analyzed *Mcp* sequences in transgenes. Before my studies no precise *Mcp* analysis was possible at the endogenous locus in the BX-C due to technical restrictions. We therefore aimed at studying the boundary element *Mcp* in the endogenous location.

To allow such a precise analysis, I established two phiC31 integrase dependent landing sites into the endogenous *Mcp* genomic location. The DNA homologous recombination repair mechanism used to generate these landing sites induced two separate deletions in the endogenous *Mcp* genomic region. The landing sites can be used to reintroduce modified boundary sequences which can be studied in the background of two different *Mcp* region deletions. My re-entry experiments allowed the precise dissection of known regulatory elements (like the insulator and PRE function) and the discovery of yet unknown elements in the region of the *Mcp* boundary.

Furthermore, bigger and clean DNA deletions allow more convenient genetic studies in *Drosophila*. Therefore, I established fly strains containing several deletions in the region, making use of the CRISPR/Cas9 technology to induce DNA double strand breaks and the cellular NHEJ DNA repair mechanism. Those deletions were used to screen for phenotypes caused by deletions in the wider *Mcp* genomic region in the BX-C.

3

Results

3.1 PAPER MANUSCRIPT: From Blackbox to regulatory logic: *in situ* dissection of the boundary element *Miscadestral pigmentation* in the Bithorax complex of *Drosophila melanogaster*

Mario A. Metzler, Dimitri Bieli, Stefan Harmansa, Markus Affolter, Martin Müller

Author contributions at the time of thesis submission: MAM and MM planned and carried out the experiments. MAM and MM wrote the manuscript.

3.1.1 Introduction

Hox genes in *Drosophila melanogaster* pattern the body along the anterior/ posterior axis and are subdivided into two complexes: 1) the Antennapedia complex is responsible to establish the identity of the most anterior segments until thoracic segment 2 (T2) and 2) the Bithorax complex (BX-C) is responsible to assign the identity of segments T3 to abdominal segment 8 (A8). The transcriptional regulation of Hox genes in *Drosophila melanogaster* has been studied extensively and is known to be highly complex (Duncan, 2002; Lewis, 1978). Indeed, the contribution of many different *cis*-regulatory elements is required for the correct expression pattern of Hox gene transcripts throughout development.

3.1 PAPER MANUSCRIPT: From Blackbox to regulatory logic: *in situ* dissection of the boundary element *Miscadestral pigmentation* in the Bithorax complex of *Drosophila melanogaster*

In the fly abdomen the segmental identity is controlled by three genes belonging to the Bithorax Complex (BX-C): *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*). *Ubx* is responsible for the properties of the first abdominal segment (A1, and thoracic segment 3), *abd-A* for A2 to A4, and *Abd-B* for A5 to A8. It was shown that mutations in those three genes can lead to transformations of one segment into another. Such homeotic mutations led to the characterization of several distinctive *cis*-regulatory elements: 1) initiator elements, 2) tissue-specific enhancers, 3) boundary elements, and 4) Polycomb Response Elements (PRE) (Maeda and Karch, 2006). To understand how gene transcription is controlled in each abdominal segment, the analysis of the interplay between different *cis*-regulatory elements is crucial.

The transcription of *Abd-B* in the posterior-most abdominal segments of the adult fly is associated with regulatory regions *iab-5* to *iab-8* (**Figure 3.1 A**). Each of those regions contains an initiator element: an early enhancer that integrates the input of gap genes (Qian et al., 1991; Shimell et al., 2000), pair-rule genes (Müller and Bienz, 1992), and segment-polarity genes (Macías et al., 1994; Mann, 1994). Each initiator element has a sharp anterior activity limit which corresponds to a parasegmental (PS) border. For example, *iab-5* initiator element activity starts in PS10 and extends all the way to the posterior end. The next *iab* regulatory element on the chromosome is *iab-6*. It becomes active in PS11. As revealed by anti-Abd-B antibody staining in late embryos, *Abd-B* expression levels in PS10 and PS11 are clearly different (Busturia and Bienz, 1993; Simon et al., 1990). Thus, through this delimitation of initiator element activity, the expression levels of *Abd-B* are precisely controlled parasegment by parasegment.

Limiting gene regulation to just those early enhancers rise an important problem: a single gene is controlled by multiple initiator elements active in different parasegments and cross-talk between those initiator elements has to be prevented. It is known now that boundary elements are responsible to keep the *iab* regions insulated from each other. Within the *Abd-B* domain the corresponding boundary elements are called *Mcp*, *Fab-6*, *Fab-7* and *Fab-8* (**Figure 3.1 A**). *Mcp* separates *iab-4* from *iab-5*, *Fab-6* separates *iab-5* from *iab-6*, and so forth. Evidence for the existence of boundary elements was obtained from the molecular analysis of a peculiar class of *Abd-B* alleles. These were *Abd-B* gain-of-function alleles which

expanded the activity of the *Abd-B* gene by exactly one parasegment. Those alleles turned out to represent small deletions between two neighboring *iab* regions. As a result, the deletion of a *Abd-B* boundary leads to the fusion of two *iab* regions. In the new fused region, both parasegments acquire the identity of the posterior parasegment (Barges et al., 2000; Gyurkovics et al., 1990; Karch et al., 1994; Mihaly et al., 1997, 1998). The correct arrangement of initiator elements and boundaries is therefore crucial for precise developmental gene regulation.

Over the last twenty years, several studies have shown that within the realms of the *Abd-B* gene, boundary elements are always flanked by a Polycomb Response Element (PRE) (Maeda and Karch, 2015). PREs are DNA sequences that enable binding of Polycomb protein complexes (Müller and Kassis, 2006). Polycomb proteins are responsible to mediate specific histone modifications that silence chromatin (Ringrose and Paro, 2004). A very elegant study recently demonstrated that silenced chromatin is established according to the position along the anterior/posterior axis of the animal: more anterior tissues have more extended silenced chromatin in the BX-C region compared to more posterior tissues. The regions that are not silenced coincide with active initiators in the observed tissues (Bowman et al., 2014). These findings corroborate the predictions made by the "open for business" model formulated almost thirty years ago by Peifer et al. (1987). The model states that *iab* regulatory regions are becoming active one after the other from parasegment to parasegment along the anterior/posterior axis of the animal. Importantly, the model also posits that the demarcation between active and silenced chromatin always coincides with the position of a boundary element (Peifer and Bender, 1986; Maeda and Karch, 2015).

The interplay between boundary, PRE and the associated initiator element has best been studied in the *Frontabdominal-7* (*Fab-7*) boundary in the BX-C region. The Fab-7 boundary is flanked by a PRE in close proximity in the *iab-7* region (Mishra et al., 2001) and is separating the *iab-6* (PS11) and *iab-7* (PS12) regions. In wild-type conditions *Abd-B* expression is lower in PS11 compared to PS12. Removal of the *Fab-7* boundary element/PRE pair in $Fab\text{-}7^{1}$ flies (referred to as class I *Fab-7* alleles) leads to a fusion of the *iab-6* (responsible for PS11) and *iab-7* regions. This fusion leads to a uniform expression of Abd-B in PS11 and PS12. Phenotypically this leads to the transformation of adult

segment A6 into A7 (Mihaly et al., 1997). More precise deletions in the *Fab-7* region lead to two other classes of alleles: class II alleles correspond to a deletion of the *Fab-7* boundary element only. This leads to a partial A6 to A7 transformation of tergites (an *Abd-B* gain-of-function phenotype) and ectopic sternite bristles in A6 (transformation of A6 into A5, an *Abd-B* loss-of-function phenotype). This phenotype can be either pushed towards a stronger gain-of-function phenotype in a Polycomb mutant background or towards a stronger loss-of-function in a Trithorax (proteins responsible for epigenetic modifications that lead to an activated chromatin state) mutant background. Class III deletions, which remove only the PRE, do not give any altered phenotype (Mihaly et al., 1997).

In further studies, the *Fab-7* boundary was shown to interact with the *Abd-B* promoter both at the endogenous locus and in transgenic assays (Cléard et al., 2006; Kyrchanova et al., 2011). This suggests that boundaries physically interact with the promoter and therefore subdivide the chromatin into loops.

For interactions between boundaries and initiator elements, studies on the *iab-6* initiator and *Fab-6* boundary were informative. A deletion of the *iab-6* initiator leads to an *Abd-B* loss-of-function phenotype (sternite A6 transformed into sternite A5). Instead, deletion of the *Fab-6* boundary gives rise to an *Abd-B* gain-of-function phenotype (sternite A5 transformed into sternite A6). When both the initiator and the boundary are deleted, this leads to a complete loss of function (visible in tergites and sternites) in A5 and A6, since both of the segments are transformed towards A4 identity. This represents a somehow surprising result, as the *iab-5* initiator element is not deleted, and therefore the phenotype in the adult cuticle is not in line with the "open for business" model. In the embryonic ventral nerve cord the model still holds: activity in PS10 is not altered, while PS11 acquires a PS10 identity because of the lack of the *iab-6* initiator element. The authors explain this discrepancy with a possible deletion in the *iab-5* region of an adult cuticle enhancer (Iampietro et al., 2010).

These elegant *in situ* studies on *Fab-7* and *Fab-6* gave insight into the nature of boundaries and PREs in the *Abd-B* region and how they separate *cis*-regulatory elements of the same gene. *Miscadestral pigmentation* (*Mcp*), the boundary that separates the *abd-A* and *Abd-B* genes, can give insights into how entire gene loci

are insulated from each other. *Mcp* separates the initiator element regions *iab-4* and *iab-5* and thus the regulatory regions of the genes *abd-A* and *Abd-B* (Celniker et al., 1990).

Classical *Mcp* deletion alleles were able to restrict the Mcp region to around 400 base pairs (bp) (Karch et al., 1994). Furthermore, transgenic assays have revealed that *Mcp* is similarly organized as the *Fab-6* and *Fab-7* regions: a boundary element abuts the *iab-4* regulatory region and is closely linked to a PRE on its *iab-5* side. Some of the hallmarks attributed to the *Mcp* region are described below: 1) Two *Mcp* elements can pair *in trans* over large distances, meaning that distant genetic loci can be brought into physical vicinity (Müller et al., 1999; Vazquez et al., 2006). 2) When interposed between an enhancer and its promoter, *Mcp* can act as an enhancer blocker (Gruzdeva et al., 2005; Gohl et al., 2008). Gruzdeva et al. found that this function is contained within the boundary element of *Mcp*. 3) *Mcp* can control the stable expression of BX-C enhancers throughout development via its PRE (Busturia et al., 2001). This activity is contained within a 140bp fragment that contains GAF and Pho binding sites.

Further transgenic studies have indicated that the orientation of the *Mcp* boundary is important for correct chromatin-chromatin interactions (Kyrchanova et al., 2007) and that this long-distance interaction function is mediated by the boundary and not the PRE (Li et al., 2011). The establishment of such long-distance interactions between boundaries is thought to facilitate the interaction between distant enhancers and promoters (Gruzdeva et al., 2005). All these results indicate the importance of the boundary element and the PRE for the correct function of the *Mcp* region in regulating *Abd-B* expression. ChIP experiments have shown that the boundary part of *Mcp* is a target of CTCF, a known boundary element protein conserved up to humans (Holohan et al., 2007; Smith et al., 2009).

Further indications that CTCF and Polycomb group proteins play an important role in the function of the boundary element and PRE comes from phenotypes of CTCF and Polycomb group protein alleles. Animals heterozygous for several Polycomb Group Protein alleles show a partial transformation of the fourth abdominal segment into the fifth, mimicking a *Mcp* deletion (Bornemann

et al., 1998). Furthermore, combinations of *Pho* alleles give mixed *Mcp*/*Fab-7* phenotypes (A4 to A5, and A6 to A7 transformations) (Girton and Jeon, 1994). Similarly, CTCF alleles show black pigmentation in A4 male tergites, indication of a partial A4 to A5 transformation (Mohan et al., 2007).

Until now, *in situ* studies in the *Mcp* genomic locus similar to the ones carried out in the *Fab-7* region have not been undertaken. Furthermore, all current endogenous deletions in the *Mcp* region remove both the boundary and the PRE, thus it has not been possible to separately analyze these *cis*-regulatory elements. To understand the logic of *cis*-regulatory elements at gene boundaries, we established several tools to study the function of the *Mcp* boundary in great detail at the endogenous locus.

In this work, we introduced a phiC31-integrase dependent landing site at the *Mcp* locus that was used to dissect the locus and identify the different *cis*-regulatory units. Furthermore, we made use of the CRISPR/Cas9 technology to create new deletions in the locus.

Our analysis of the *Mcp* boundary region confirms that the boundary region and the PRE contribute to *Mcp* function. But to our surprise, the phenotype associated with the boundary/PRE deletion differs from the one described for the classical *Mcp* alleles. Those alleles show a transformation of A4 into A5. With our tailored deletions, we also observe a change of identity in A3. This segment acquires black pigment and thus A5 identity.

Finally, we suggest a model of how the different *cis*-regulatory elements interact at the boundary between the *abd-A* and *Abd-B* gene. This model is able to explain the phenotypes that were described in our studies.

3.1.2 Materials and Methods

Stocks used in this study

Flies were grown on standard cornmeal agar. Mcp^{B116} and *y w;DfP9/DpP5,Sb* flies were a gift from François Karch. *y w vas-Cas9* (Gratz et al., 2014), w^{1118}, *w; Xa/(CyO;TM3),* and *y w* flies were obtained from Bloomington Drosophila Stock Center. *y w;S110501/TM6C* flies were obtained from the Szeged stock

center. *Mcp[1]/TM6C* flies were a gift from Ana Busturia. *y w M{vas-int.Dm}zh-2A* , a stock producing phiC31-integrase under the control of the vasa promoter, and insertion platform M{3xP3-RFP. attP}zh-86Fb were obtained from Johannes Bischof (Bischof et al., 2007). $ap^{c1.4b}$ is described in (Bieli et al., 2015a).

Fly DNA injections were carried out according to standard protocols (Spradling and Rubin, 1982; Bateman et al., 2006; Bischof et al., 2007).

Generation of CRISPR/Cas9 induced deletions

gRNAs were cloned as described in (Gratz et al., 2014). We injected pairs of gRNA binding to flanking sequences to the DNA that was aimed to be deleted into *vas-Cas9* embryos. Survivors were crossed with w^{1118} flies and the progeny screened for the dominant change in abdominal pigmentation associated with *Mcp* deletions in adult males. The isolated transformands were crossed with *w;* ap^{Xa}*/(CyO;TM3)* to establish balanced stocks.

For the generation of the Mcp^{MM1} following gRNA guide sequences were used: gctggcttttacagcatttc and gctttgttacccctgaaaat. For Mcp^{MM3}: gaaagtcgggtctgcaaata and gctttgttacccctgaaaat. For Mcp^{MM4}: gctggcttttacagcatttc and gaatggggccatttgtgtat. For Mcp^{MM5}: gctttgttacccctgaaaat and gcaccgtgggcccagtaatt.

Deletions were verified by PCR and sequencing.

Generation of CRISPR/Cas9 induced homologous recombination and establishment of a Mcp phiC31-integrase dependent landing site

The backbone for the recombination plasmid was designed in silico containing following elements: MCS5-attP-3xP3-EGFP-SV40poly(A)-attP-FRT-MCS3. This construct was synthesised and inserted in a pUC57 vector at Genewiz. The two multiple cloning sites (MCS5 and MCS3) were used to clone homology arms into the vector.

Homology arms were cloned from *y w* flies genomic DNA using the following primers: CCTGCCGACTGAACGAATGC and ACGCCCTGATCCCGATACACATAC for the proximal arm (*iab-4* region), GCGTTTGTGTGGGTAG-

TAAATGTATC and AAAGGCCAACAAAGAACACATGGACG for the distal arm (*iab-5* region).

The recombination plasmid was injected into *vas-Cas9* flies with two gRNAs containing the following guides: gctggcttttacagcatttc and gctttgttacccctgaaaat. These gRNAs are identical to the ones used for the creation of Mcp^{MM1}. Concentrations used according to (Gratz et al., 2014).

The described strategy leads to a small duplication on the *iab-5* side when re-entry plasmids are injected (see **Figure 3.7**). However, this duplication does not lead to a phenotype when the wild-type *m0* construct is injected. The duplication is removed with the *m14* deletion construct and does not lead to a phenotype either.

Successful homologous recombination events were isolated by screening for EGFP expression in larvae. The insertion was confirmed by PCR and sequencing. It is referred to as Mcp^{attP}. A balanced stock (*y w M{vas-int.Dm}zh-2A;*Mcp^{attP}*/TM3,Sb*) was established using a *y w M{vas-int.Dm}zh-2A;Dr Mio/TM3,Sb* stock. Because of poor survival rates in injection experiments, the Mcp^{attP} chromosome was combined with *DpP5,Sb* and stock *y w M{vas-int.Dm}zh-2A;*Mcp^{attP}*/DpP5,Sb* was established. It was used for the injection of the re-entry plasmids described below.

Cloning and injections of re-entry plasmids

Constructs for phiC31-integrase mediated transgenesis were generated based on plasmid piB-LLFY(BI) (Bieli et al., 2015a). This plasmid contains a *mini-yellow* marker which is used to select transformands. Re-entry DNA fragments were either generated by PCR mutagenesis or DNA synthesis (ATG:biosynthetics and Genewiz).

Transgenic lines for the various re-entry plasmids were obtained for docking sites $ap^{c1.4b}$, $ap^{c2.73c}$ and Mcp^{attP}. The *mini-yellow* marker can be deleted from Mcp^{attP} transgene insertions by Flp treatment.

Immunostaining

Embryos were prepared and stained using standard procedures. The primary antibodies used were: 1:10 anti-ABD-B (mouse, 1A2E9) from Developmental Studies Hybridoma Bank and 1:500 anti-engrailed (rabbit, d-300, sc-28640) from Santa Cruz Biotechnology.

Wing disc staining were done according to (Bieli et al., 2015b).

Abdominal cuticle preparation

Flies were fixed overnight in 70% EtOH/30% glycerol. Subsequently, abdomens were removed from the carcass and cut along the dorsal midline with a razor blade. The cut abdomens were incubated for 20 minutes in 10% KOH. Afterwards, the abdominal cuticles were positioned on a slide outside up and covered with a cover slip. The prep was incubated at 60°C for four hours. After incubation the cover slip was washed away in a beaker glass filled with distilled water. Abdominal cuticle were mounted in Hoyer's mounting medium.

3.1.3 Results

CRISPR/Cas9 induced deletions in the *Mcp* locus reveal an unknown function distally from the *Mcp* boundary

The *Mcp* boundary was previously delimited in the endogenous locus by mainly two deletions: the Mcp^{1} deletion removes around 3400bp, while the Mcp^{B116} deletion removes a region of around 850bp (Karch et al. (1994), **Figure 3.1 B**). Both of these deletions remove the boundary element and the PRE.

In order to delimit the *Mcp* boundary element/PRE pair in a more precise way than previously possible, we made use of the CRISPR/Cas9 system to create precise deletions. First, we induced a deletion called Mcp^{MM4} that removes a similar stretch of DNA as the Mcp^{1} deletion. Mcp^{MM4} shows the same phenotype as Mcp^{1} (A4 to A5 transformation) and therefore phenocopies this classical deletion (**Figure 3.1 B,C**).

Next, we established two deletions that remove only the core element of the boundary/PRE pair as previously defined in transgenic studies. The relevant

sequences are localized between the SalI and XbaI restriction sites. In Mcp^{MM1}, the entire 0.8kb region between the SalI and XbaI sites is deleted. Mcp^{MM3} shares the same distal break point with Mcp^{MM1} but its proximal break is located just to the left of the boundary element and deletes 525bp (**Figure 3.1 B**). Inspection of the abdomen of homozygous male adults revealed an unexpected result. Not only was there black pigment in A4, but also most of the A3 tergite was black (**Figure 3.1 C**). This phenotype was not expected as, according to the current models, a deletion of a boundary should give a gain of function in only one more anterior segment (as seen previously with deletion Mcp^{MM4}). A3 to A5 tergite transformations were previously described for $Abd\text{-}B^{Sab}$ alleles. They are associated with a base pair change in a Krüppel binding site of the *iab-5* initiator element (Ho et al., 2009). We found no such mutation in any of our new *Mcp* alleles (data not shown).

It has previously been noticed that hemizygosity for the BX-C (*DfP9/+*) can produce in males visible subtle haplo-insufficiency effects. These include white spots in A5, occasional bristles on the sixth sternite and patchy pigmentation on the A4 tergite (Karch et al. (1985), see also **Figure 3.1 C**). The latter phenotype suggests that the activity of the *Mcp* boundary region could be pairing dependent. Therefore, we wished to analyze the phenotypes of various *Mcp* alleles in hemizygous conditions. The phenotypes of Mcp^{1}*/DfP9* and Mcp^{MM4}*/DfP9* males remained similar to those of homozygous siblings (compare top and bottom pictures in **Figure 3.1 C**). Significant changes were detected for Mcp^{MM1}*/DfP9* and Mcp^{MM3}*/DfP9*, where the patchy black pigmentation in A3 was enhanced to an almost complete transformation of A3 into A5. Interestingly, conspicuous patches of black pigment were also detected on the tergite A3 of Mcp^{B116}*/DfP9* males.

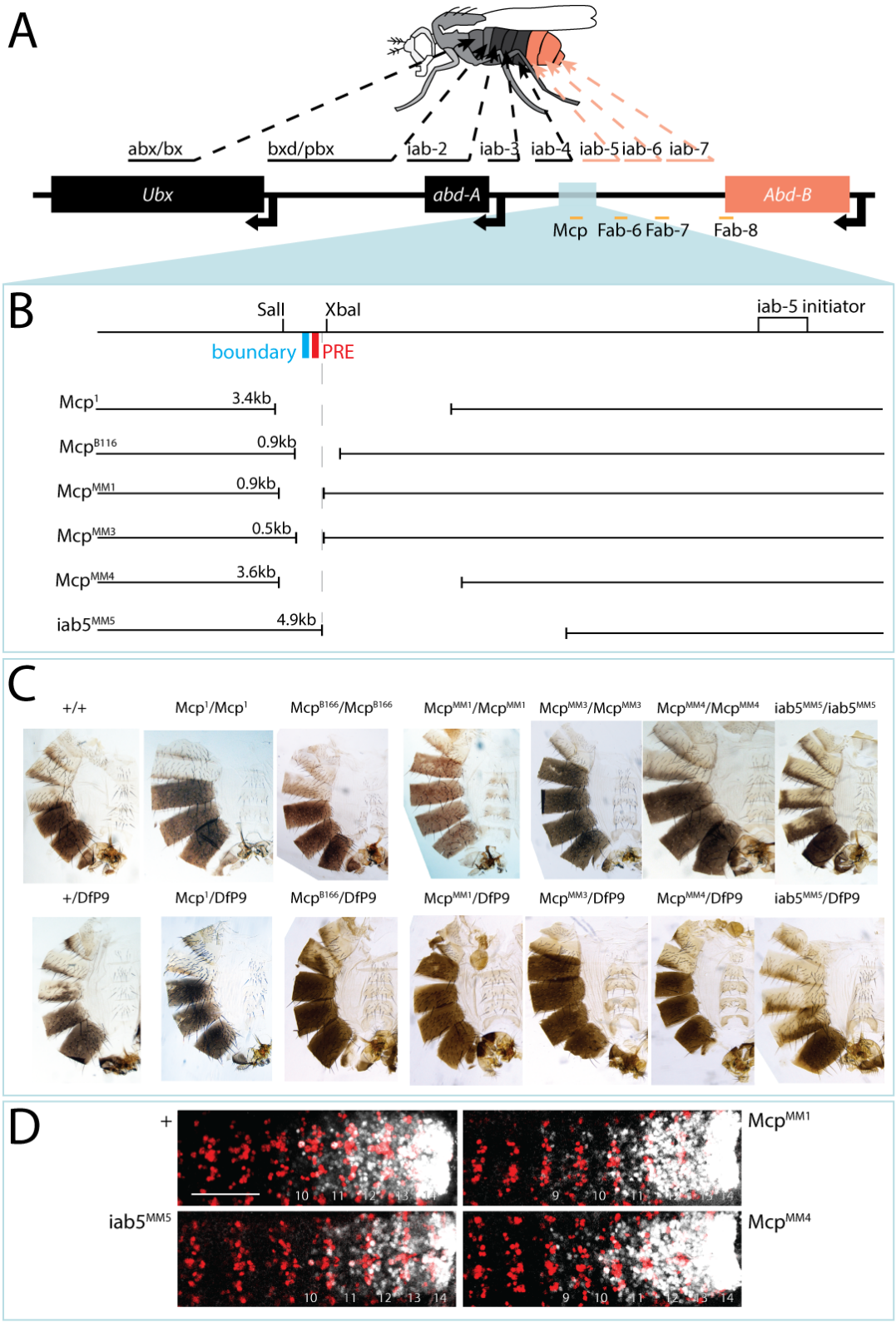
Comparison of the breakpoints of Mcp^{MM1} and Mcp^{1} shows that the two deletions differ distally to the PRE, while their proximal break points are virtually equal. Therefore, we established a deletion that started at the distal breakpoint of Mcp^{MM1} and extended for 4.9kb into the *iab-5* region. Importantly, this deletion does not affect the *iab-5* initiator element (see $iab5^{MM5}$ in **Figure 3.1 B**). Homozygous and hemizygous males display an *iab-5* loss-of-function phenotype: unpigmented patches are scattered across the A5 tergite. Thus, we refer to this

allele as $iab5^{MM5}$ (**Figure 3.1 C**). This result suggests that the Mcp^{1} deletion affects more regulatory elements than just the core boundary/PRE pair. The loss of function associated with the $iab5^{MM5}$ deletion leads to the speculation that this allele may uncover a tissue specific enhancer active in tissues giving rise to the A5 tergite.

All alleles discussed so far should be caused by a differential expression of *Abd-B*. Therefore, we visualized *Abd-B* expression by anti-Abd-B antibody staining and imaged the ventral nerve cord. In wild-type embryos, *Abd-B* expression was very strong in the most posterior segment (PS14) and became fainter stepwise parasegment by parasegment until PS10. Homozygous Mcp^{MM1} and Mcp^{MM4} both show an expanded *Abd-B* expression until PS9, while $iab5^{MM5}$ does not show a discernible difference from wild-type. Interestingly, Mcp^{MM1} does not show additional signal in PS8. PS8 corresponds to A3 in adult males, and tergite

Figure 3.1 *(following page)*: Dissection of the endogenous *Mcp* locus using CRIPSR/Cas9 with non homologous end joining - A: Illustration of the BX-C genomic locus in *Drosophila melanogaster*. Black boxes represent the coding regions of *Ubx* and *abd-A*. *Abd-B* coding sequences are marked with a light red box. Promoters are marked with black arrows. Above the black line the regulatory regions and their corresponding adult segments are connected by dotted lines. Segments controlled by *Abd-B* are highlighted in light red. Below the black line representing the BX-C, boundaries from the *Abd-B* region are marked with orange lines. B: The black line at the top represents a chromosomal stretch of the *iab-4/iab-5* region. On its left, the SalI and XbaI sites demarcate a fragment containing the *Mcp* boundary (blue box) and the *Mcp* PRE (red box). The location of the *iab-5* initiator is illustrated as a black outlined box. Below, black interrupted lines indicate the location and size (indicated in kilobases above the proximal breakpoint) of deletions in the *Mcp* locus. Mcp^{1} and Mcp^{B116} are two classical *Mcp* deletions while the other deletions are newly CRISPR/Cas9 induced deletions (Mcp^{MM1}, Mcp^{MM3}, Mcp^{MM4} and $iab5^{MM5}$). C: Cuticle preps of homozygous (first row) or hemizygous (second row, we used *DfP9* as a deficiency that removes all three BX-C genes) male adult flies corresponding to the deletions depicted in panel (B). D: *Abd-B* expression in embryonic ventral nerve cord. *Abd-B* expression in white, *engrailed* expression in red. The anterior *engrailed* expression boundary delimits parasegmental borders. Parasegments showing *Abd-B* expression are marked with a number. Scale bar equals 50μm.

3.1 PAPER MANUSCRIPT: From Blackbox to regulatory logic: *in situ* dissection of the boundary element *Miscadestral pigmentation* in the Bithorax complex of *Drosophila melanogaster*



A3 is clearly transformed in adult Mcp^{MM1} males (**Figure 3.1 D**).

Establishment of a phiC31 integrase dependent landing site in the Mcp locus

In order to dissect the endogenous Mcp boundary in greater detail, we engineered the Mcp locus using CRISPR/Cas9-induced homologous recombination. We wished to introduce a Mcp^{MM1}-like deletion and replace it with two phiC31-integrase dependent landing sites for Recombination-Mediated-Cassette-Exchange (RMCE, (Bateman et al., 2006)). Towards that end, we designed a plasmid containing homology arms from the Mcp region (around 4kb in length) flanking an attP cassette containing an EGFP coding sequence driven by an 3x3P enhancer. Upon integration of the RMCE-target cassette, a 789bp *Mcp* sequence was deleted. This sequence corresponds to the SalI-XbaI fragment and contains the boundary element/PRE pair. To remove the *mini-yellow* injection marker of the re-entry constructs, an FRT site was inserted on the *iab-5* side of the RMCE cassette (**Figure 3.2 A,B**).

We planned to identify flies with a homologous recombination event using the activity of 3x3P-EGFP reporter. Thus, successfully transformed flies should have EGFP expression in their eyes. No such flies could be found. However, we were able to isolate correct insertions thanks to EGFP expression in the abdominal segments of larvae (**Figure 3.2 C**). We conclude that the EGFP cassette acts as an enhancer trap in the BX-C, but is silenced in eye tissues. Integration was verified using primers p1, p2, p3 and p4 (**Figure 3.2 B**) by PCR (**Figure 3.2 D**) and sequencing (data not shown).

After integration of re-entry constructs into the Mcp^{attP} landing site, the *yellow* injection marker can be removed by Flipase treatment (**Figure 3.2 E**). Furthermore, the re-entry *Mcp* fragment is flanked by LoxP sites and can thus be removed by Cre-recombinase treatment.

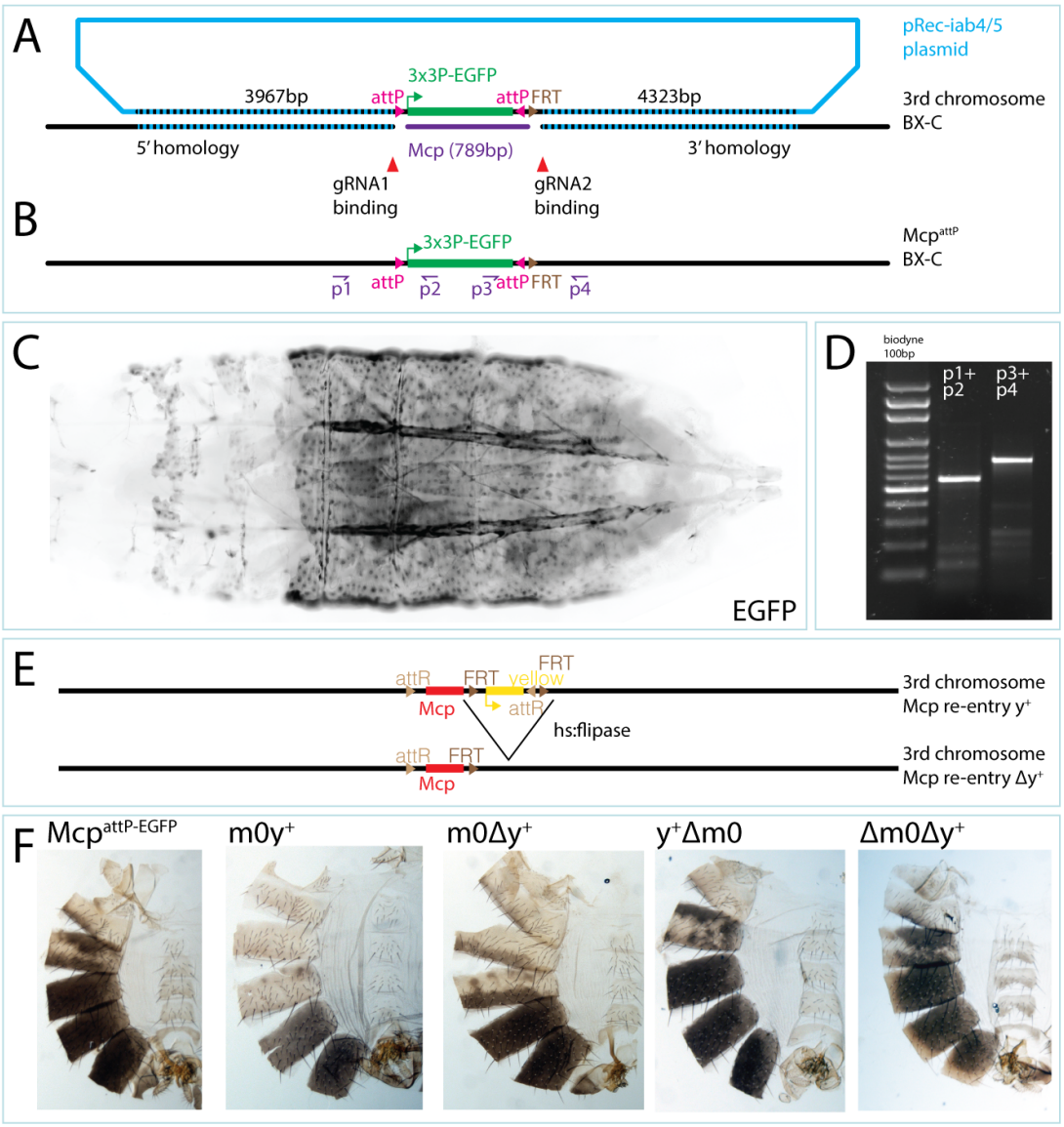
As expected, homozygous Mcp^{attP} males (landing site illustrated in **Figure 3.2 B**) give a strong *Abd-B* gain-of-function phenotype reminiscent of that described for Mcp^{MM1}: tergite A4 is completely and A3 incompletely pigmented (**Figure 3.2 F**). Re-insertion of the wild type *Mcp* sequence gives a complete

rescue of this phenotype. Derivatives of Mcp^{m0y+} were obtained by step-wise Flp and Cre treatments. Removal of the *mini-yellow* reporter (referred to as $Mcp^{m0\Delta y+}$) does not affect the abdominal pigmentation pattern of homozygous males: it remains as in wild type flies. In contrast, deletion of the 0.8kb SalI-XbaI fragment ($Mcp^{\Delta m0}$) again reconstitutes the strong *Abd-B* gain-of-function phenotype, irrespective of the presence of the *mini-yellow* transformation marker ($y+\Delta m0$ and $\Delta m0\Delta y+$ in **Figure 3.1 F**).

Dissection of the *Mcp* boundary and PRE at the endogenous locus in the BX-C

We wished to identify relevant DNA sequences within the SalI-XbaI fragment of the *Mcp* region. Therefore, a collection of overlapping small deletions within that fragment was constructed. The mutated DNA fragments were brought back into the piB-LLFY plasmid for RMCE at the Mcp^{attP} docking site. Design of the deletions was aided by *in silico* detection of protein binding sites within

Figure 3.2 *(following page)*: Establishment of the phiC31 integrase dependent landing site in the endogenous *Mcp* locus - A: CRISPR/Cas9 induced homologous recombination at the *Mcp* locus. In blue a representation of the homology donor plasmid, in black the *Mcp* region on the third chromosome. Homologous sequences are marked as dotted lines. The cutting sites of gRNA-Cas9 are indicated by red triangles. The deleted *Mcp* sequence is illustrated in purple. An EGFP cassette controlled by the 3x3P enhancer is inserted instead. The EGFP cassette is flanked by attP sites. On the right, the attP site is abutted by an FRT site. B: Illustration of the *Mcp* region on the third chromosome after homologous recombination. Primers used for verification by PCR are indicated. C: Expression of EGFP in third instar larvae carrying the landing site. D: Genotyping of landing site carrying flies using primers indicated in (B). E: Re-entry in the *Mcp* landing site. The 3x3P-EGFP is replaced with a cassette containing different *Mcp* fragments (flanked by LoxP sites for excision, not shown in figure). The *mini-yellow* reporter is used as an injection marker. The final configuration leaves one attR and an FRT site plus the inserted construct. F: Abdominal pigmentation phenotypes of the wild-type (*m0*) re-entry construct. Both the *Mcp* fragment and the *yellow* injection marker can be independently removed. When *m0* is floxed out, presence or absence of *mini-yellow* does not significantly modify these two phenotypes.


A
pRec-iab4/5
plasmid
3x3P-EGFP
3967bp
attP
attP FRT
4323bp
3rd chromosome
BX-C
5' homology
Mcp (789bp)
3' homology
gRNA1
binding
gRNA2
binding
B
Mcp^attP
BX-C
3x3P-EGFP
p1 attP p2 p3 attP FRT p4
C
EGFP
D
biodyne
100bp
p1+
p2
p3+
p4
E
attR FRT yellow FRT
Mcp
attR
hs:flipase
3rd chromosome
Mcp re-entry y+
attR FRT
Mcp
3rd chromosome
Mcp re-entry Δy+
F
Mcp^attP-EGFP
m0y+
m0Δy+
y+Δm0
Δm0Δy+

the 0.8kb SalI-XbaI fragment. First, previously mapped binding sites for GAF and Pho (Busturia et al., 2001) as well as for Pita (Zolotarev et al., 2016) are indicated in **Figure 3.3 B**. The GAF and Pho sites constitute the relevant DNA elements for the PRE function in transgenic assays. Pita is a boundary associated protein described by (Zolotarev et al., 2016). Second, JASPAR database and tools (Mathelier et al., 2016) were used to map binding sites for insulator binding proteins of *Drosophila melanogaster*: Zeste, BEAF-32 , Su(Hw), Trl, CTCF. Zw5 and DspI have no weight matrixes for DNA binding affinity in the JASPAR database.

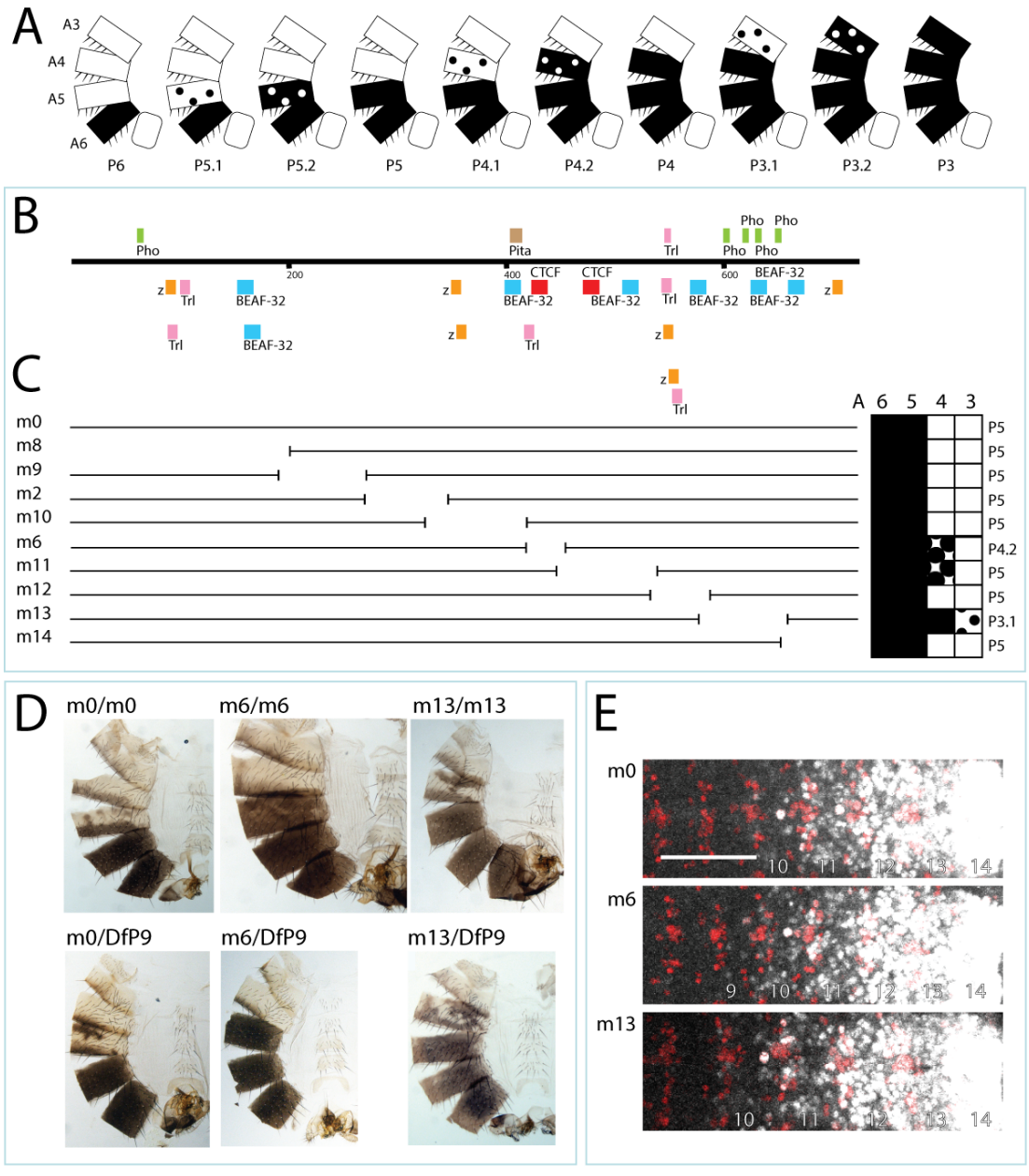
Re-entry events were isolated with the help of the *mini-yellow* marker for nine overlapping deletion plasmids. Correct orientation of the RMCE insert was verified by PCR. The *mini-yellow* marker was removed by Flp-treatment and homozygous stocks were established. As described above, the wild-type SalI-XbaI fragment (*m0*) suppresses the gain-of-function phenotype observed in Mcp^{attP} males and essentially restores wild-type pigmentation. Similar observations were obtained for Mcp^{m8}, Mcp^{m9}, Mcp^{m2}, Mcp^{m10}, Mcp^{m12} and Mcp^{m14} stocks. The *Abd-B* gain of function phenotype was retained only in two of the nine deletions. Mcp^{m6} and Mcp^{m11} males display patchy pigmentation on tergite A4 but not A3. This phenotype is enhanced in hemizygous males in which A4 is now completely pigmented. The gain-of-function phenotype is even stronger in Mcp^{m13} males. There, homozygous males have a complete A4 to A5 transformation, and black spots can also be seen on A3. This phenotype is somewhat enhanced in hemzygous males.

These observations highlight the importance of a 41bp interval containing a putative CTCF binding site (deleted in Mcp^{m6}), a 93bp interval deleted by Mcp^{m11}, and of a 82bp interval containing four Pho binding sites (deleted in Mcp^{m13}) for proper *Mcp* function at the endogenous locus. Deletions Mcp^{m6} and Mcp^{m11} are located in the boundary element, deletion Mcp^{m13} removes part of the PRE. It therefore appeared conceivable that a 253bp fragment extending from the left of Mcp^{m6} to the right of Mcp^{m13} could be sufficient to substitute for the entire 0.8kb SalI-XbaI fragment. A corresponding transgenic line was established and the *mini-yellow* marker deleted. Its phenotype was analyzed in adult homozygous males. Mcp^{m253} completely rescues the *Mcp* phenotype (data

not shown, see **Figure 3.7** for breakpoints). This observation reinforces the conclusion that all information for *Mcp* function is contained in a short stretch of DNA separating *iab-4* from *iab-5*.

The phenotypes observed in Mcp^{m6} and Mcp^{m13} are an indication of incorrect *Abd-B* expression. Therefore, we stained late homozygous embryos carrying the *m0*, *m6* and *m13* re-entry constructs with *Abd-B* and *engrailed* (*en*, used to mark parasegmental boundaries. The anterior *en* expression boundary corresponds to the anterior parasegment boundary) antibodies and imaged their ventral nerve cord (VNC). As expected, Mcp^{m0}/Mcp^{m0} embryos show *Abd-B* expression in PS10 to PS14, indistinguishable from wild-type (compare **Figure 3.1 D** and **Figure 3.3 E**). Mcp^{m6}/Mcp^{m6} embryos show a gain of function phenotype: in addition to Abd-B expression from PS10 to PS14, they show also weak *Abd-B* expression in PS9. This weak expression resembles the phenotype observed in complete *Mcp* deletions (like Mcp^{MM1}, compare **Figure 3.1 D** and **Figure 3.3 E**). Surprisingly, deletion of the PRE (Mcp^{m13}/Mcp^{m13}) does not show any evident gain of function phenotype as there is no signal in PS9 (**Figure 3.3 D**).

Figure 3.3 *(following page)*: Deletions in the endogenous *Mcp* boundary and PRE - A: Classification for male adult abdominal pigmentation. Wild-type flies have a class P5 abdominal pigmentation. B: The solid black line depicts the *Mcp* boundary/PRE region, above the line are the five Pho and one Trl (GAGA factor) binding site analyzed in (Busturia et al. 2001) and the predicted binding for Pita according to (Zolotarev et al., 2016). Below the black line the positions of JASPAR hits for insulator proteins are depicted (Mathelier et al. 2016). C: The interrupted black lines illustrate the position and size of re-entry deletion constructs (compared to panel B), to the left the name of the deletion and to the right the phenotype in pigmentation pattern and abdominal pigmentation class introduced in panel (A). Mcp^{m0} is the wild-type re-entry. D: Cuticle prep of re-entry deletions showing relevant phenotypes, Mcp^{m6} removes one of the two predicted CTCF sites, and Mcp^{m13} is a deletion in the PRE region removing 4 Pho sites. E: Staining against Abd-B and En in embryonal ventral nerve cord of homozygous embryos. Abd-B antibody signal in white, Engrailed antibody signal in red. En signal was used to delimit parasegments. Parasegments are numbered. Parasegments showing *Abd-B* expression are marked with a number. Scale bar 50μm.


A
A3
A4
A5
A6
P6
P5.1
P5.2
P5
P4.1
P4.2
P4
P3.1
P3.2
P3
B
Pho
Pita
Trl
Pho Pho
Pho Pho
200
400
600
CTCF
CTCF
BEAF-32
Z
Trl
C
m0
m8
m9
m2
m10
m6
m11
m12
m13
m14
A 6 5 4 3
P5
P5
P5
P5
P5
P4.2
P5
P5
P3.1
P5
D
m0/m0
m6/m6
m13/m13
m0/DfP9
m6/DfP9
m13/DfP9
E
m0
m6
m13
9
10
11
12
13
14

Long distance interaction assays in the *apterous* region

Numerous studies have provided evidence that the protein CTCF plays an important role in genome organization (Herold et al., 2012). It is thought that distant DNA fragments containing CTCF binding sites can be brought into close proximity, and thereby forming looped structures. Such chromatin looping might for example juxtapose distant tissue-specific enhancers to their target promoter.

In the previous section, we have presented genetic evidence that CTCF could also play a role for *Mcp* function: the Mcp^{m6} allele deletes part of the *Mcp* boundary including a putative CTCF binding site and causes an *Abd-B* gain-of-function phenotype. A previous study has reported that the boundary part of *Mcp*, and not the PRE, is mediating physical interaction between *Mcp* elements placed on distant transposable elements (Li et al., 2011). We wished to re-evaluate these findings with our set of overlapping deletions. The genetic assay system to test their potential to interact *in trans* with a distant *Mcp*-containing transgene is described below.

The long distance interaction assay consists of two transgene components (see **Figure 3.4 B**): 1) A phiC31-integrase docking site for RMCE was previously established in the *apterous* (*ap*) locus (see **Figure 3.4 A**, Bieli et al. (2015a)). It can be used to establish transgenes containing our set of *Mcp* deletions. Insertions are identified through the *mini-yellow* marker. A *mini-white* reporter ($ap^{MM\text{-}Mcp}$) inserted at the identical genomic position was previously shown to be dominantly silenced, presumably due to the neighboring *ap* PRE (Gohl et al., 2008). 2) The $ap^{MM\text{-}Mcp}$ transgene was exposed to a P-element transposase source and several new insertion were isolated on chromosome 2. One of them, 17.1^{A*}, is inserted at position 2R: 6,220,525 (R6.13), about 500kb distal to the *ap* locus. Heterozygous $17.1^{A*}/+$ males have uniform orange-brown eye color, their homozygous siblings develop a yellowish and variegating eye pigmentation. Hence, the gene dosage-dependence normally seen for *mini-white* reporters is not observed for the 17.1^{A*} insert. This phenomenon has been reported for many PRE and *mini-white* containing transgene constructs and is known as PRE-mediated pairing-dependent silencing of *mini-white* (reviewed by Kassis (2002)). It is this unusual behavior of *mini-white* which is used in our assay to score boundary-boundary interactions.

For PREs, pairing-dependent *mini-white* silencing is only seen for homozygous inserts (T1/T1) but not for *trans*-heterozygots (T1 +/+ T2). In the latter case, *mini-white* expression is again dosage dependent. The situation can change when T1 and T2 also contain a boundary element. In this case, boundary elements on T1 and T2 can interact and juxtapose the two PREs, thereby again inducing PRE-dependent *mini-white* silencing (see **Figure 3.4 B**). Since our 0.8kb *Mcp* fragment contains a boundary element and a PRE, the long-distance interaction test can be done with just this fragment.

For the long-distance interaction assay, homozygous 17.1^{A^*} virgins were crossed with *yw* males (control) or with males carrying our set of *Mcp* deletions in the *ap* locus (see **Figure 3.4 C**). From the progeny of these crosses, $17.1^{A^*}/+$ and $17.1^{A^*}/ap^{m}$ flies were collected and aged for seven days before their eye color were compared and scored. The observations are interpreted in the following way: 1) ap^{m0} contains the wild-type 0.8kb *Mcp* fragment. $17.1^{A^*}/ap^{m0}$ flies have clearly lighter eye color than $17.1^{A^*}/+$ controls (yellow, variegating vs. uniform orange brown; see **Figure 3.4 D**). Lighter eye color in *trans*-heterozygous flies indicates that the two *Mcp* elements on ap^{m0} and 17.1^{A^*} interact and induce PRE-mediated *mini-white* silencing. 2) If $17.1^{A^*}/ap^{m}$ flies have similar eye color as $17.1^{A^*}/+$ controls, it can be concluded that interaction between the two *Mcp* elements is lost and as a consequence also the PRE-pairing dependent silencing of the *mini-white* reporter on 17.1^{A^*}.

The data for the various *Mcp* deletions is summarized in **Figure 3.4 C** and examples are depicted in **Figure 3.4 D**. Three classes of interactions could be found: 1) *Mcp* fragments *m8*, *m13* and *m14* did not interfere with long-distance interaction, hence their deletions do not remove protein binding sites important for the *Mcp*-*Mcp* interaction (data for *m8* and *m14* not shown). 2) *Mcp* fragments *m9*, *m2* and *m12 in trans* to 17.1^{A^*} produced eye pigmentation intermediate between $17.1^{A^*}/ap^{m0}$ and $17.1^{A^*}/+$ (data not shown). This suggests that their deletions partially reduce *Mcp*-*Mcp* interaction. 3) *Mcp* fragments *m10*, *m6* and *m11* basically abolish silencing of the *mini-white* on 17.1^{A^*}. It can thus be assumed that the interaction capability is missing on these *Mcp* deletions. It is conceivable that these deletions remove binding sites for proteins involved in *Mcp*-mediated long-distance interaction.

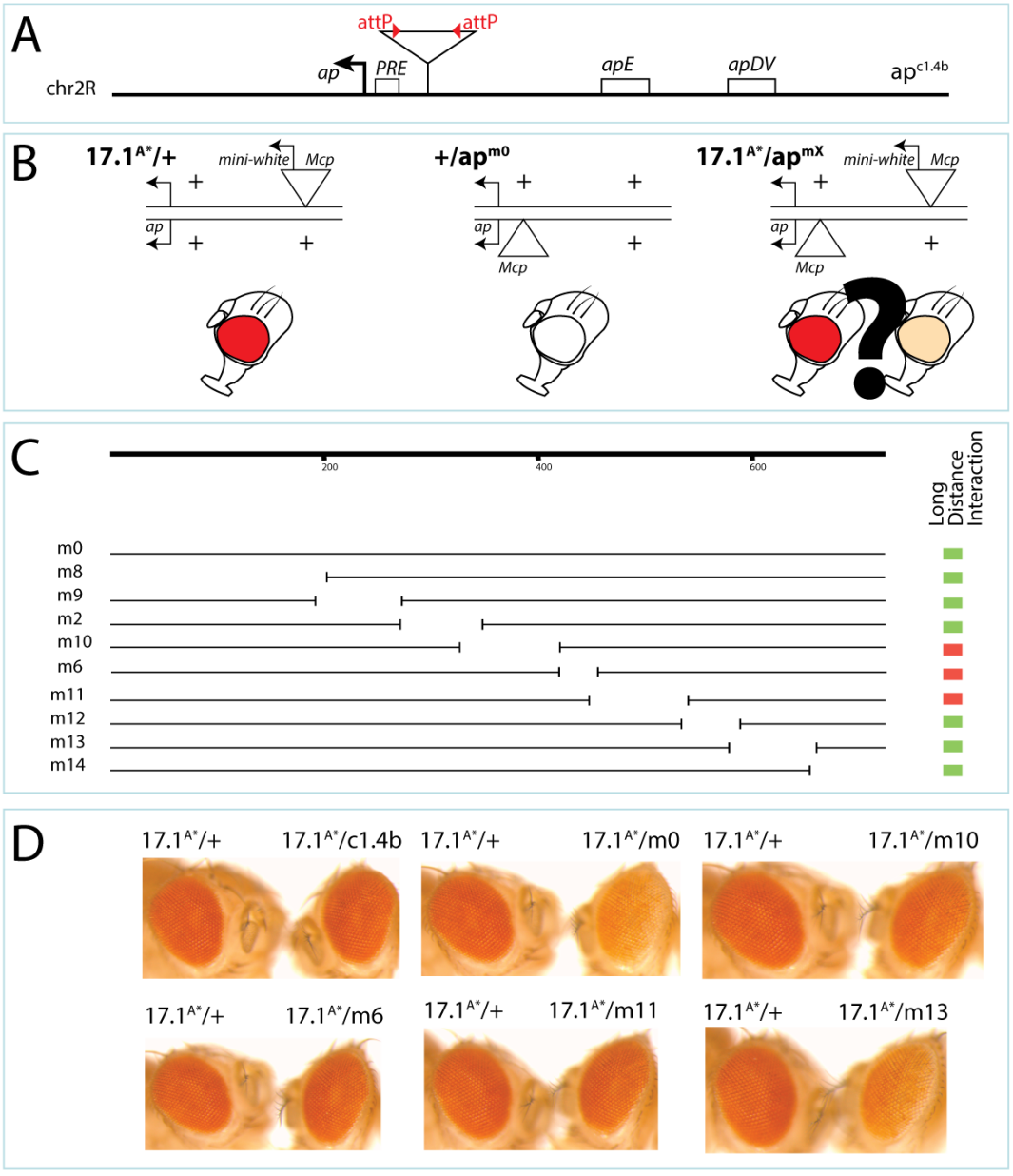
From the genetic data presented in **Figure 3.4 C**, it is clear that DNA sequences important for long-distance interaction are spread over the boundary part of *Mcp* (with the one exception of *m12* showing weak interaction), including the putative CTCF sites (see **Figure 3.3 B**). In contrast, deletion of most of the PRE (ap^{m13}) has no effect on *Mcp*-mediated long-distance interaction. These observations corroborate the findings of (Li et al., 2011) that it is the boundary part of *Mcp* that mediates *Mcp-Mcp* interactions. This activity might also be of relevance for intra-BX-C interactions between other boundary elements and/or the *Abd-B* promoter.

Conservation of *Mcp* boundary between *Drosophilides*

To test the conservation of the boundary function during *Drosophila* evolution, we made use of the established landing site in the endogenous *Mcp* genomic location. We re-inserted the orthologous sequences from other *Drosophila* species:

Figure 3.4 *(following page)*: Dissection of *Mcp* long-distance interaction function in the *apterous* locus - A: Insertion $ap^{c1.4b}$ in the apterous gene is marked with a triangle above the black line representing the apterous genomic region. On the black triangle the two attP sites used for RMCE are marked with two red arrowheads. The landing site is inserted in close proximity to the *ap* PRE. B: Illustration of the long distance interaction assay. The two parallel black lines represent the two sister chromosomes. The *ap* region is represented by its promoter (black arrow on chromosome). Transgenic insertions are marked by a black triangle. On the left always the *ap* insertion, on the right the insertions in 17.1^{A*}. $17.1^{A*}/+$ flies have orange brown eye color, as no PRE-pairing dependent silencing is established. $+/ap^{m0}$ flies do not show any eye color, as *mini-white* is not present on the transgene. $17.1^{A*}/ap^{mX}$ is the experimental situation, where the different *Mcp* deletion constructs are assayed for their eye color, and therefore interaction function. C: Representation of *Mcp* DNA as a black line with base pairs coordinates, for predicted binding sites see **Figure 3.3 B**. *Mcp* deletion constructs analogous to **Figure 3.3 C** with activities from the long distance interaction assay illustrated as green (deletion construct does displays the activity) and red squares (deletion construct does not display the activity). D: Relevant results from the long distance interaction assay. In each picture on the left the control eye color (flies without *Mcp* insert in *ap*, only *Mcp* and *mini-white* insert 17.1*), on the right flies with both the inserts (*Mcp* insert in *ap*, *Mcp* and *mini-white* insert 17.1*).

3.1 PAPER MANUSCRIPT: From Blackbox to regulatory logic: *in situ* dissection of the boundary element *Miscadestral pigmentation* in the Bithorax complex of *Drosophila melanogaster*



Drosophila erecta, *Drosophila simulans*, and *Drosophila pseudoobscura*. The first two are closely related to *D. melanogaster*, the latter belongs to the *obscura* group which diverged from the *melanogaster* group about 25-30 million years ago. For all three species, a DNA fragment corresponding to the 0.8kb *Mcp* fragment of *D. melanogaster* was chosen. Over this stretch of DNA, the conservations are as follows: 83.4% for *D. simulans*, 70.7% for *D. erecta* and 34.6% for *D. pseudoobscura* (see **Figure 3.8**). Short conserved sequence stretches are scattered all over the 0.8kb fragment. In addition, there are two longer conserved DNA elements. One comprises about 160bp next to the SalI site. Deletion of this region in Mcp^{m8} did not have an effect on *Mcp* function (see **Figure 3.3 C**). The other lies at the other end of the fragment, where the *Mcp* boundary element/PRE pair is located.

Homozygous stocks could be established for Mcp^{sim}, Mcp^{ere} and Mcp^{pse}. As could be expected, Mcp^{sim} and Mcp^{ere} yield a complete rescue. Homozygous and hemizygous males show the typical black pigmentation in A5 and A6, but not in A4. *Mcp* function is impaired in Mcp^{pse}: patches of black pigment can be detected on A4 but not A3. This indicates that Mcp^{pse} partially rescues the gain-of-function phenotype of Mcp^{attP} or Mcp^{MM1}. When looking for an explanation for this difference, one has to look at the CTCF and Pho binding sites on these different constructs. 1) The CTCF site (CTGGCGCCCCCTATT) is conserved in Mcp^{sim} and Mcp^{ere}, while Mcp^{pse} has five base changes (CTGGCGCtCtCTgac). 2) for the Pho sites, one out of four is completely conserved in all three constructs (the second most distal), while the other three have variations from the consensus (ATGGC). Those sites in Mcp^{pse} are (from the proximal site to the most distal one): Agaaga, gcGGC, ATGGC and gTGtg. As a following step, it would be interesting to replace those divergent CTCF and Pho binding sites in the *D. pseudoobscura* fragment with the *D. melanogaster* ones. As a prediction, this should give a better (if not complete) rescue.

Sequences from related *Drosophila* species seem to work well in the *D. melanogaster* *Mcp* locus. Therefore we used the information about conservation to design a synthetic *Mcp* element. We designed a synthetic 0.8kb SalI/XbaI fragment that includes all the conserved DNA stretches and called it Mcp^{cons}. Sequences that are not conserved were replaced by random DNA sequences. Care was taken not

to insert binding sites for known boundary protein binding sites. Somewhat surprisingly, Mcp^{cons} flies are not able to rescue the *Mcp* phenotype of Mcp^{attP} at all: they show an *Abd-B* gain-of-function phenotype in A4 and A3 (**Figure 3.5 A**). We can conclude that 1) the PRE and CTCF sites alone are not sufficient to rescue the phenotype and that 2) Even though sequences from other *Drosophilides* are fairly different from the *D. melanogaster* sequences they give a better rescue than our synthetic element that is based on the conservation to those species.

Effect of boundary orientation

The boundary regions of the *Abd-B* domain are all organized in the same way: each boundary element is abutted by a PRE on its distal side. While the boundary separates neighbouring *iab* regulatory regions, the PRE is part of the regulatory landscape of the distal *iab*. By changing the order of boundary element and PRE by flipping the 0.8kb SalI-XbaI fragment, the *iab-5* PRE is repositioned into *iab-4* in Mcp^{rev1} flies. One would predict that this *Mcp* allele could lead to an *Abd-B* gain-of-function phenotype because the *iab-5* domain loses the *iab-5* PRE. In fact, segment A4 is transformed into A5 identity.

The apparent loss-of-function in Mcp^{rev1} flies could also be explained by other reasons. Current models on *Abd-B* regulation propose that the boundaries of the *Abd-B* domain are contacting each other and/or the promoter. It has been shown in transgene assays that the pairing between boundaries is orientation dependent (Kyrchanova et al., 2007). Furthermore, in mammals, CTCF binding site orientation was reported to play a role in genome organization (Zlotorynski, 2015). We therefore decided to investigate the orientation dependence of three sequences contained within the boundary element/PRE pair of *Mcp*. Towards that end, the following fragments were flipped in the context of the 0.8kb SalI-XbaI interval: 1) The 197bp boundary element in Mcp^{rev2}, 2) the 41bp CTCF site Mcp^{rev3}, and 3) the 132bp PRE in Mcp^{rev4} (see **Figure 3.5 B,C** for results and **Figure 3.7** for coordinates).

Our observations suggest that these DNA fragments are not (or only slightly) orientation dependent. The abdominal cuticle of Mcp^{rev3} male is indistinguishable from wild-type. Black patches in the A4 tergite of Mcp^{rev2} and Mcp^{rev4} indicated

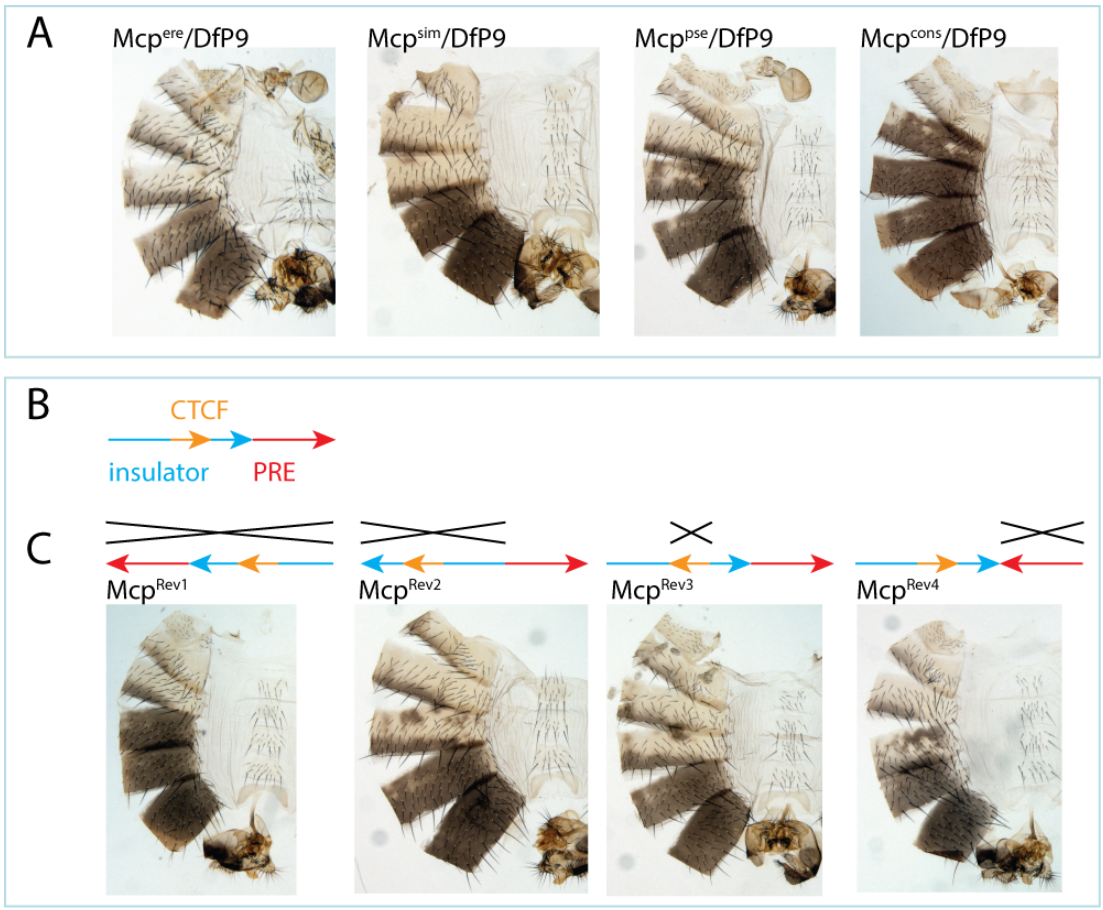


Figure 3.5: Conservation and orientation dependency of the *Mcp* element - A: Re-entry constructs in the endogenous *Mcp* locus of conserved sequences from other *Drosophilides*. *Mcp* sequences from *Drosophila erecta* and *Drosophila simulans* show a good rescue of the *Mcp* phenotype. *Mcp* like sequences from *Drosophila pseudoobscura* do show an *Abd-B* gain of function phenotype (nevertheless, not a complete A4 to A5 transformation). When the conserved *Drosophila melanogaster Mcp* sequences are interspaced with random DNA sequences with the same length of the replaced not conserved original *Mcp* sequences no rescue is visible (strong A4 and A3 transformation to A5 identity). All constructs are observed over *DfP9*. B: Illustration of the functional units of *Mcp* and their orientation in unmodified sequences. C: Adult abdominal phenotypes of re-entry constructs with reversed elements, all constructs are observed over *DfP9*. The only way to get a complete A4 transformation to A5 is when the whole boundary element/PRE pair is inverted (Mcp^{rev1}). Mcp^{rev2} reverses the DNA containing the boundary function (included the CTCF binding site), phenotypically those flies do not show a strong *Abd-B* gain of function. When only the CTCF site is reversed (orange arrow, Mcp^{rev3}) we still do not observe a strong *Abd-B* gain of function phenotype. When the PRE is inverted (red arrow, Mcp^{rev4}) we observe a slight stronger *Abd-B* gain of function phenotype.

that the orientation of the respective DNA fragment plays a minor role. However, as those cuticles derive from hemizygous flies, those patches in A4 and the rare loss of function in A5 might derive from hemizygousity. Furthermore, it is conceivable that flipping of these DNA fragments could alter the distance between proteins bound to them and this might create a mild phenotype as observed in Mcp^{rev2} and Mcp^{rev4}.

3.1.4 Discussion

Dissection of the *Mcp* boundary element and PRE in the endogenous *Mcp* locus in the BX-C

The study of the boundary element *Miscadestral pigmentation* (*Mcp*) in its endogenous genomic location gave us novel insight into gene regulation at the *abd-A*/*Abd-B* boundary.

The previous knowledge about boundaries in the BX-C postulates a gain of function phenotype in the next adjacent anterior segment upon boundary element deletion. In the case of *Mcp* this would correspond to a transformation of segment A4 into the identity of segment A5. Deletion of the core *Mcp* region (boundary + PRE) gives a stronger phenotype than postulated (A5 and A4 transformation). In contrast, when we deleted just the boundary (and not the PRE) we observed the postulated phenotype (A5 to A4 transformation, **Figure 3.3 D**).

The study of the *Fab-7* boundary showed that an additional deletion of the PRE to the boundary will make the gain-of-function phenotype associated with a boundary element deletion more stable: instead of a mixed gain/loss of function one would get a full gain-of-function phenotype (Mihaly et al., 1997). In the endogenous *Mcp* region we find that a deletion of the boundary/PRE pair not only gives a stronger A4 to A5 transformation, but an additional A3 to A5 gain-of-function phenotype as well (**Figure 3.1 C**). This result is in contradiction with the phenotype associated with the Mcp^{1} deletion (Celniker et al., 1990), which does not show any A3 to A5 gain of function phenotype, even thought the boundary and the PRE are both deleted. We therefore hypothesized that there are additional *cis*-regulatory elements in the region that is deleted in the Mcp^{1} but not in the Mcp^{MM1} allele. The existence of such an element is necessary to

explain the difference in phenotype between the classical Mcp^{1} deletion and our new deletions (Mcp^{MM1} and Mcp^{MM3}) in the core *Mcp* region (**Figure 3.1 B,C**).

Indeed, when we deleted a region distal to the *Mcp* PRE, which removes sequences deleted in the Mcp^{1} deletion but not in our Mcp^{MM1} deletion ($iab5^{MM5}$), we get a novel phenotype for the *Mcp* region. The phenotype associated with the $iab5^{MM5}$ deletion is an almost complete loss of function phenotype in segment A5 (**Figure 3.1 C**). This result strongly suggests the presence of a previously not described regulatory element.

Furthermore, we aimed at assigning the gain of function in A3 and the gain of function in A4 to short definite DNA sequences. For this purpose, we established an attP RMCE landing site replacing the endogenous *Mcp* locus, leaving a deletion similar to Mcp^{MM1}. *Mcp* re-entry constructs containing different *Mcp* fragments were established and inserted into this landing site. Most of these constructs rescue the *Mcp* phenotype. Two deletion constructs give two distinct phenotypical classes: deletion of the CTCF site gives an A4 gain of function phenotype (**Figure 3.3 D**), while deletion of the PRE gives an A4 and a partial A3 gain of function phenotype (**Figure 3.3 D**). We conclude that the additional A3 pigmentation derives from missing PRE/TRE activity. Experiments by (Busturia et al., 2001) have shown that a missing PRE leads to an expanded expression of enhancers from the BX-C.

Thus, combining the finding that a deletion of the PRE leads to an A3 gain of function (not predictable by previous models) with the discovery of an additional regulatory element on the distal side of the *Mcp* boundary we postulate the following working model: the newly discovered distal regulatory element is an enhancer (late *iab-5* enhancer, or $iab\text{-}5^{L5}$) that is controlled by the PRE during development. The absence of PRE activity leads to an expansion of $iab\text{-}5^{L5}$. Therefore, the dark pigmentation in A4 and A3 in alleles similar to Mcp^{MM1} is not due to a faulty insulator action, but lack of a correct memory function by the PRE.

In contrast, inhibiting insulator function by the deletion of the CTCF binding site will lead to a phenotype that is in line with current BX-C models: removal of the boundary between *iab-4* and *iab-5* in re-entry construct *m6* leads to a

fusion of such regions and thus results in a one segment gain of function (dark pigmentation in A4).

Taken together, those results lead to our model on how the *Mcp* locus is regulated (**Figure 3.6**). In a wild type situation the boundary element (with the CTCF site) is responsible to prevent influence of *iab-4 cis*-regulatory elements to the *Abd-B* region. The putative enhancer *iab-5*^{L5} is responsible for correct *Abd-B* expression in tissues giving rise to the adult A5 tergite, while kept in check by the *Mcp* PRE (**Figure 3.6 A**).

When the putative enhancer *iab-5*^{L5} is removed (*iab-5*^{MM5}) this leads to an *Abd-B* loss-of-function phenotype in tergite A5 as *Abd-B* is not expressed in those tissues anymore (**Figure 3.6 B**).

Removal of the *Mcp* PRE will lead to an extension of the *iab-5*^{MM5} enhancer activity to tissues giving rise to tergites A4 and A3. Therefore, there is pigmentation in both of these segments (**Figure 3.6 C**). However, this raises questions about the phenotype of *Mcp*^{rev} flies. When the PRE is brought to the *iab-4 region* no ectopic A3 pigmentation is seen. Thinking along the lines of the "open for business" model, the PRE should not be able to interact with *iab-5* regulatory elements. Nevertheless, the physical vicinity might still be sufficient for the PRE to interact with *iab-5*^{L5}.

Removal of the proper *Mcp* boundary element (by removal of the CTCF binding, re-entry constructs *m6* and *m11*) will give rise to the classical *Mcp* phenotype (*Abd-B* gain-of-function in A4 but not in A3). This result is in line with the predicted result, as the removal of the boundary will likely lead to a fusion of the regions *iab-4* and *iab-5*. Therefore, the enhancer elements present in *iab-4* will now be able to interact with the *Abd-B* region and this leads to an ectopic *Abd-B* expression (**Figure 3.6 D**).

Removal of the whole *Mcp* boundary element/PRE pair (*Mcp*^{MM1}) leads to the expansion of the *iab-5*^{L5} enhancer activity, additionally the *iab-4* and *iab-5* regions are fused by the lack of a proper boundary element. As a result one gets a consistent A4 and A3 *Abd-B* gain-of-function phenotype (**Figure 3.6 E**).

The classical *Mcp*^{1} deletion would remove all the *cis*-regulatory elements in the *Mcp* region (boundary element, PRE, and enhancer). Therefore, as the putative enhancer would also be removed, the interference caused to the current *Abd-B*

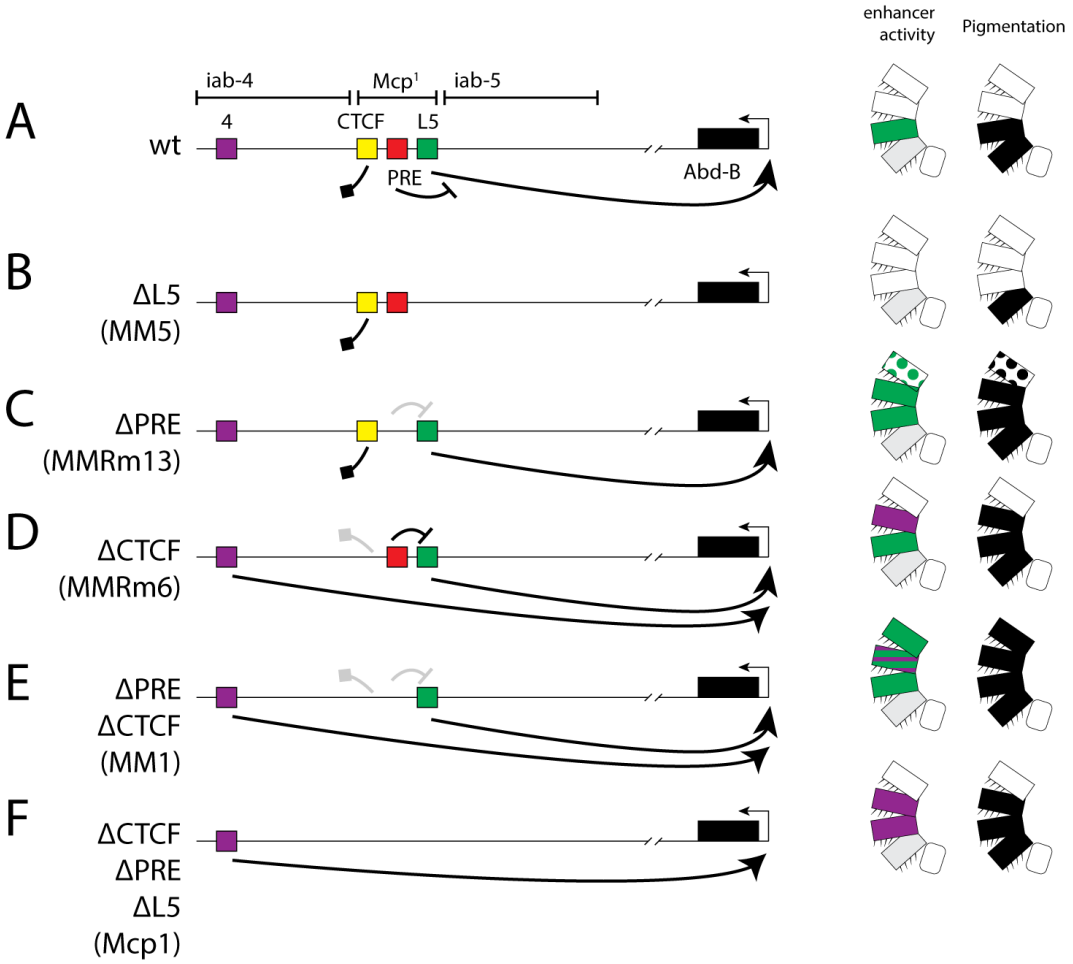
"open for business" model is removed. The regions *iab-4* and *iab-5* are fused, and the predicted phenotype (A4 *Abd-B* gain-of-function phenotype) corresponds to the observed phenotype (**Figure 3.6 F**).

Functional dissection of the *Mcp* boundary element/PRE pair in the *apterous* locus

Using the same *Mcp* SalI-XbaI deletion constructs as in the dissection of the *Mcp* element in the endogenous locus, we dissected the long-distance interaction function contained in the 0.8kb SalI-XbaI *Mcp* fragment.

We were able to localize the long distance interaction function in the region of two predicted CTCF binding sites (deletion constructs *m10*, *m6* and *m11*). As CTCF is known to enable such long distance chromatin-chromatin interactions

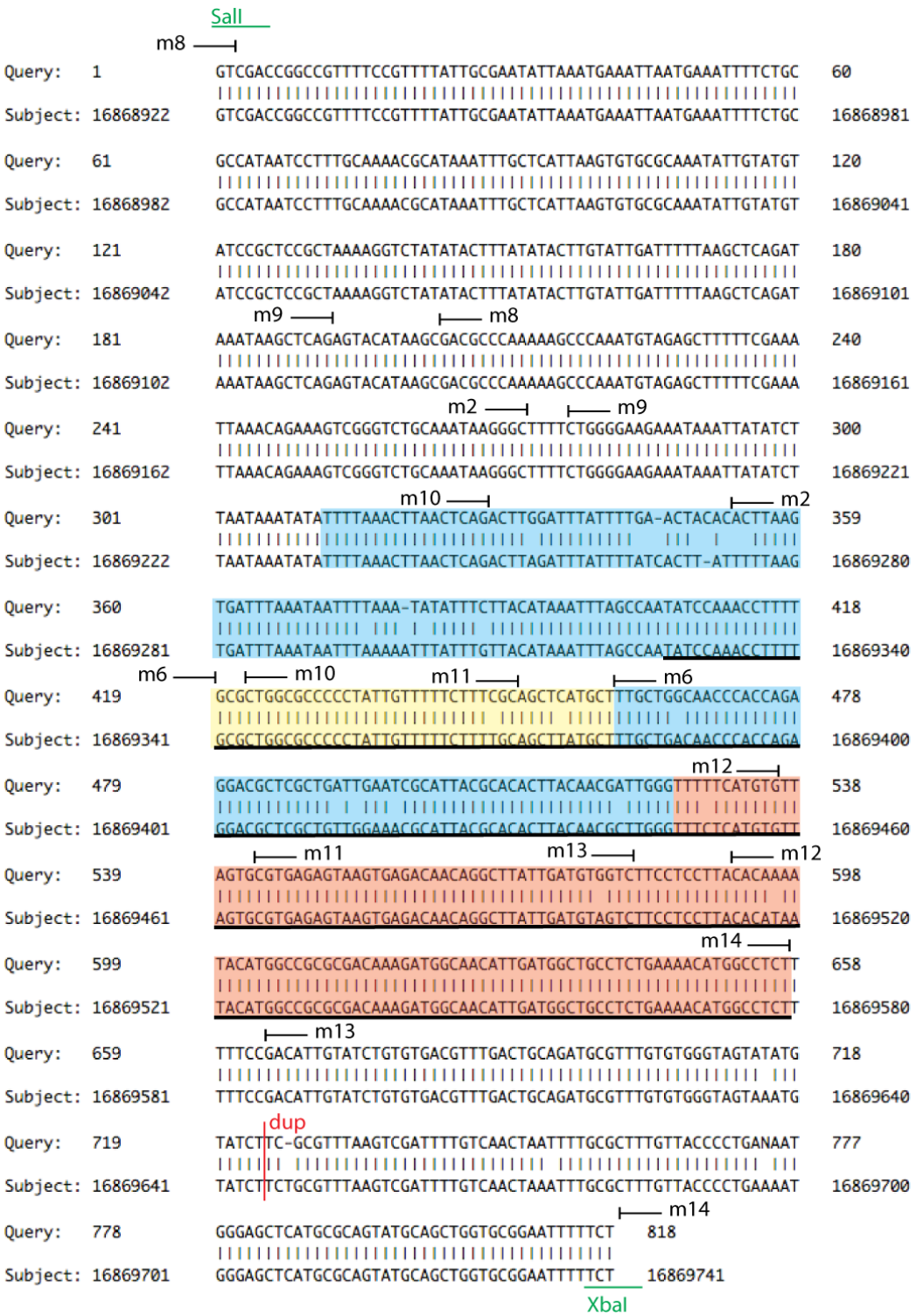
Figure 3.6 *(following page)*: Working model of *cis*-regulatory interplay at the *Mcp* locus - Left the description of the alleles, in the middle an illustration of the *Mcp* region, on the right an illustration of the activity of different regulatory elements on the male adult abdomen and the resulting pigmentation phenotype. A: Illustration of the wild type condition. *iab-4*, *iab-5*, and the span of the region missing in the Mcp^1 deletion are depicted by black lines. *Abd-B* protein coding sequence is depicted as a black box and the promoter as a a thin black arrow. Regulatory regions are depicted as colored squares. In purple an unknown *iab-4* late enhancer, in yellow the insulator function of the boundary containing the CTCF binding site, in red the PRE and in green the predicted late enhancer in *iab-5*. In wild type conditions the insulator will block interaction between the *Abd-B* promoter and *iab-4* regulatory elements. The late enhancer in *iab-5* can interact with the *Abd-B* promoter and this signal is kept stable through the *Mcp* PRE. B: lack of the *iab-5* late enhancer results in the loss of *Abd-B* activity in A5 and thus lack of pigmentation. C: Loss of the PRE leads to an expansion of the *iab-5* late enhancer activity and thus pigmentation in A3. D: Loss of the insulator function leads to input on the *Abd-B* promoter from enhancers in the *iab-4* region and thus pigmentation in A4. E: Loss of both the insulator and PRE functions leads to a more consistent A4 and A3 gain of function. F: Situation of the Mcp^1 deletion. When the insulator, PRE, and enhancer are missing we get a comparable phenotype to (D). As there is no enhancer to keep in check the PRE function is not needed, and the enhancers in *iab-4* take over. Thus observe a gain of function phenotype in A4.


enhancer activity
Pigmentation
A
wt
iab-4
4
Mcp1
CTCF
L5
iab-5
PRE
Abd-B
B
ΔL5
(MM5)
C
ΔPRE
(MMRm13)
D
ΔCTCF
(MMRm6)
E
ΔPRE
ΔCTCF
(MM1)
F
ΔCTCF
ΔPRE
ΔL5
(Mcp1)

this came as no surprise.

Our work makes a strong argument towards the fact that findings in transgene construct are useful, but might not show the real effects in the endogenous conditions of a *cis*-regulatory element. Furthermore, our result are an illustration on how predicted binding sites might be used differently according to the chromosomal environment as the CTCF site in *m11* might play a role in *apterous* but not in the endogenous locus. We can, therefore, confirm that the long distance interaction is linked to the CTCF binding sites (and not to the PRE) in line with the findings of (Vazquez et al., 2006) and (Li et al., 2011).

Figure 3.7 *(following page)*: Breakpoint of deletions in re-entry constructs - BLAST result from Flybase (Attrill et al., 2016) . The query sequence is the *Mcp* sequence used in our studies. Subject sequence is genome release R6.12. The two restriction sites used are marked by green lines and names. Deletions breakpoints are marked as interrupted lines. Proximal and distal breakpoints of each deletion are labeled with the respective name. The red line labelled with "dup" is the proximal start of the sequence duplicated after re-integration in the *Mcp* landing site. Highlighted colors correspond to the inverted sequences in **Figure 3.5 B**. The thick line represents the re-entry construct Mcp^{253} that is able to rescue the Mcp^{attP} phenotype.


SalI
m8
Query: 1 GTCGACCGGCCGTTTTCCGTTTTATTGCGAATATTAAATGAAATTAATGAAATTTTCTGC 60
Subject: 16868922 GTCGACCGGCCGTTTTCCGTTTTATTGCGAATATTAAATGAAATTAATGAAATTTTCTGC 16868981
Query: 61 GCCATAATCCTTTGCAAAACGCATAAATTTGCTCATTAAGTGTGCGCAAATATTGTATGT 120
Subject: 16868982 GCCATAATCCTTTGCAAAACGCATAAATTTGCTCATTAAGTGTGCGCAAATATTGTATGT 16869041
Query: 121 ATCCGCTCCGCTAAAAGGTCTATATACTTTATATACTTGTATTGATTTTTAAGCTCAGAT 180
Subject: 16869042 ATCCGCTCCGCTAAAAGGTCTATATACTTTATATACTTGTATTGATTTTTAAGCTCAGAT 16869101
m9
m8
Query: 181 AAATAAGCTCAGAGTACATAAGCGACGCCCAAAAAGCCCAAATGTAGAGCTTTTTCGAAA 240
Subject: 16869102 AAATAAGCTCAGAGTACATAAGCGACGCCCAAAAAGCCCAAATGTAGAGCTTTTTCGAAA 16869161
m2
m9
Query: 241 TTAAACAGAAAGTCGGGTCTGCAAATAAGGGCTTTTCTGGGGAAGAAATAAATTATATCT 300
Subject: 16869162 TTAAACAGAAAGTCGGGTCTGCAAATAAGGGCTTTTCTGGGGAAGAAATAAATTATATCT 16869221
m10
m2
Query: 301 TAATAAATATATTTTAAACTTAACTCAGACTTGGATTTATTTTGA-ACTACACACTTAAG 359
Subject: 16869222 TAATAAATATATTTTAAACTTAACTCAGACTTAGATTTATTTTATCACTT-ATTTTTAAG 16869280
Query: 360 TGATTTAAATAATTTTAAA-TATATTTCTTACATAAATTTAGCCAATATCCAAACCTTTT 418
Subject: 16869281 TGATTTAAATAATTTAAAAATTTATTTGTTACATAAATTTAGCCAATATCCAAACCTTTT 16869340
m6
m10
m11
m6
Query: 419 GCGCTGGCGCCCCCTATTGTTTTTCTTTCGCAGCTCATGCTTTGCTGGCAACCCACCAGA 478
Subject: 16869341 GCGCTGGCGCCCCCTATTGTTTTTCTTTTGCAGCTTATGCTTTGCTGACAACCCACCAGA 16869400
m12
Query: 479 GGACGCTCGCTGATTGAATCGCATTACGCACACTTACAACGATTGGGTTTTTCATGTGTT 538
Subject: 16869401 GGACGCTCGCTGTTGGAAACGCATTACGCACACTTACAACGCTTGGGTTTCTCATGTGTT 16869460
m11
m13
m12
Query: 539 AGTGCGTGAGAGTAAGTGAGACAACAGGCTTATTGATGTGGTCTTCCTCCTTACACAAAA 598
Subject: 16869461 AGTGCGTGAGAGTAAGTGAGACAACAGGCTTATTGATGTAGTCTTCCTCCTTACACATAA 16869520
m14
Query: 599 TACATGGCCGCGCGACAAAGATGGCAACATTGATGGCTGCCTCTGAAAACATGGCCTCTT 658
Subject: 16869521 TACATGGCCGCGCGACAAAGATGGCAACATTGATGGCTGCCTCTGAAAACATGGCCTCTT 16869580
m13
Query: 659 TTTCCGACATTGTATCTGTGTGACGTTTGACTGCAGATGCGTTTGTGTGGGTAGTATATG 718
Subject: 16869581 TTTCCGACATTGTATCTGTGTGACGTTTGACTGCAGATGCGTTTGTGTGGGTAGTAAATG 16869640
dup
Query: 719 TATCTTC-GCGTTTAAGTCGATTTTGTCAACTAATTTTGCGCTTTGTTACCCCTGANAAT 777
Subject: 16869641 TATCTTCTGCGTTTAAGTCGATTTTGTCAACTAAATTTGCGCTTTGTTACCCCTGAAAAT 16869700
m14
Query: 778 GGGAGCTCATGCGCAGTATGCAGCTGGTGCGGAATTTTTCT 818
Subject: 16869701 GGGAGCTCATGCGCAGTATGCAGCTGGTGCGGAATTTTTCT 16869741
XbaI

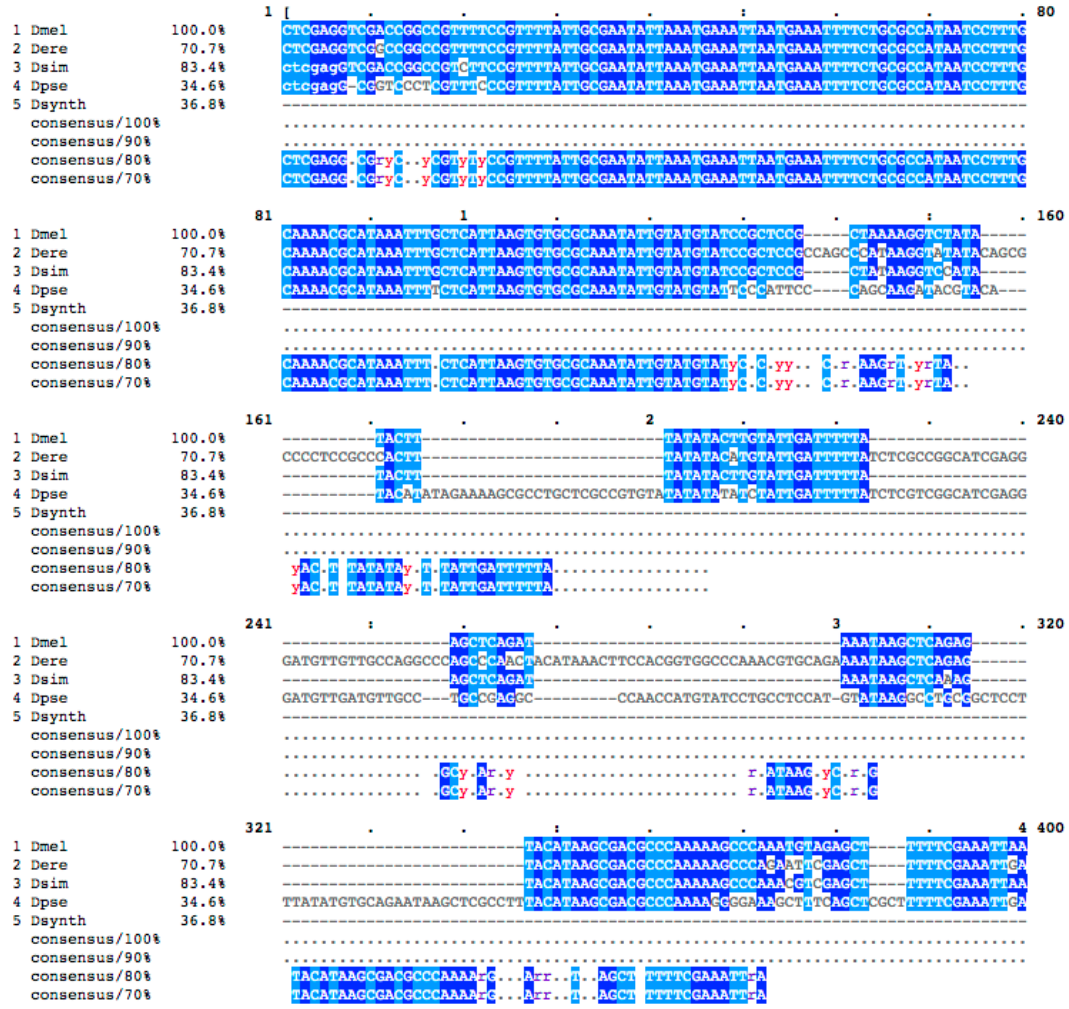


Figure 3.8: Conservation of *Mcp* sequences from differen *Drosophilides* - Alignment of *D. melanogaster*, *D. erecta*, *D. simulans*, and *D. pseudoobscura*. Furthermore, the synthetic boundary element/PRE pair is aligned as well as "Dsynth". For the alignment we used Kalign on the EBI (European Bioinformatics Institute) server (using standard settings) and MView on the EBI server for the visualization (using standard settings). (Figure continues on next page)

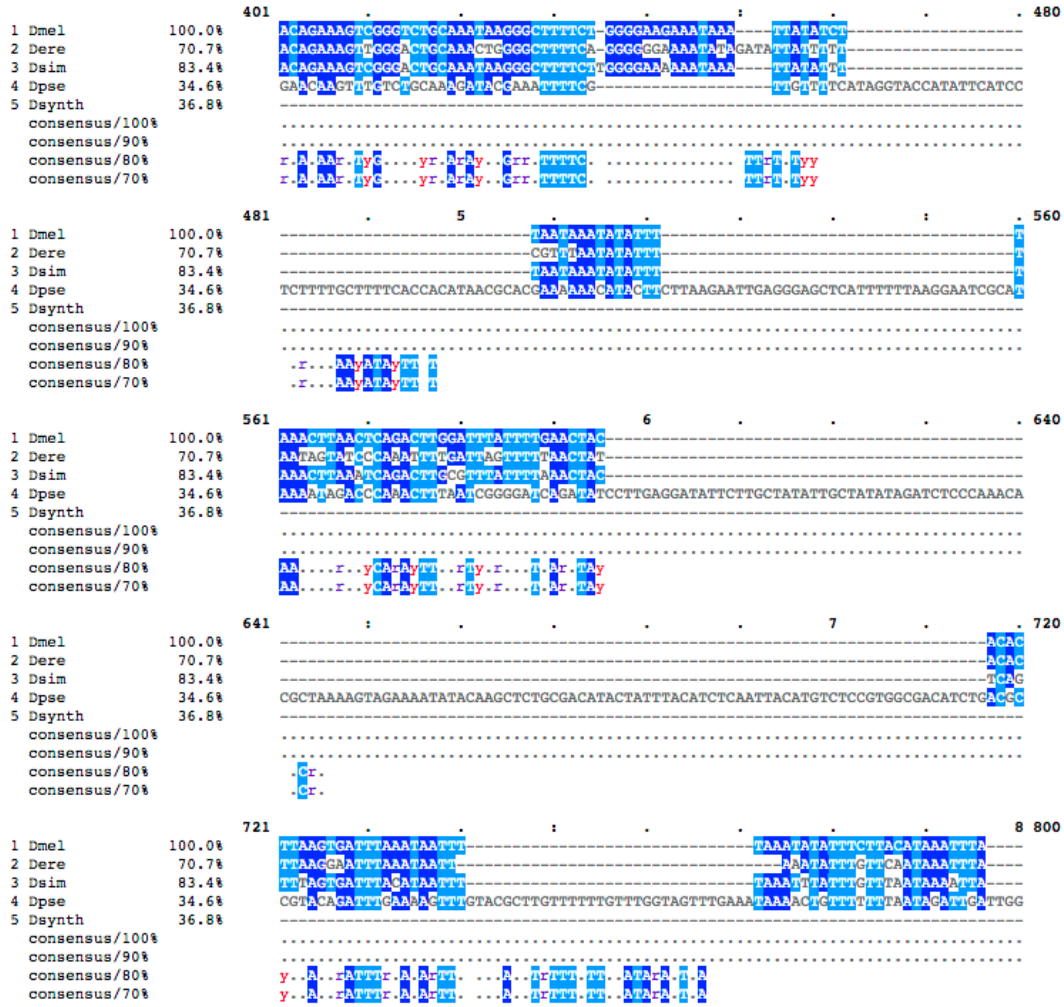


Figure 3.8 (cont.)**:** (Figure continues on next page)

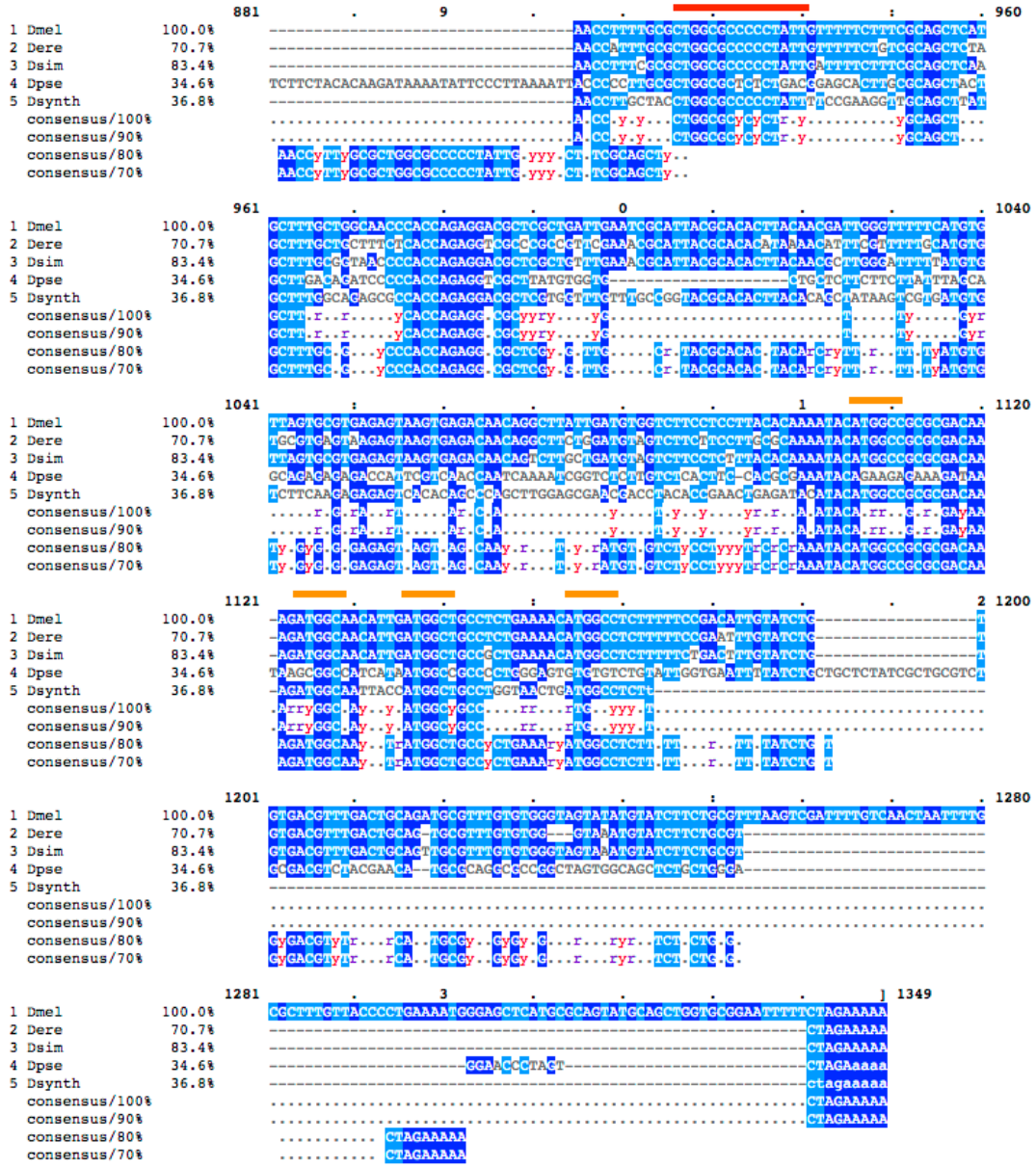


Figure 3.8 (cont.)**:** Red line marks the span of the CTCF site, orange lines mark the span of the four Pho sites.

4

Supplementary results and discussion

4.1 The *MM8* deletion

I want to introduce a further deletion that is peculiar, and had an important role in my project. The deletion *MM8* (details in **Figure 4.1 B,C,D**). This deletion was isolated by chance during the establishment of deletion Mcp^{MM1} using the same gRNA pair. The heterozygous phenotype of such a deletion gives very weak *Abd-B* gain- or loss-of-function phenotypes in A4 or A5 (data not shown). We initially thought, that it would be a small deletion inside *Mcp*. Deletion screening by PCR and DNA sequencing revealed the real nature of the *MM8* deletion: removal of the *Mcp* boundary element and PRE pair and additional DNA on both sides of *Mcp* resulting in a deletion of around 15kb. Most likely the *MM8* deletion removes all the *iab-4* sequences, the *Mcp* boundary element/ PRE pair, and the putative enhancer $iab\text{-}5^{L5}$.

The hemizygous phenotype of this deletion is striking (**Figure 4.1 A**): we observe an *Abd-B* loss-of-function phenotype on tergite A5. Furthermore, this phenotype is copied to tergites A4 and A3. Unfortunately we were not able to obtain *MM8* homozygous flies.

How can we explain such a phenotype? We can, with a certain confidence, argue that all *iab-4* regulatory elements are deleted. Work by Bender and Hudson (2000); Fitzgerald and Bender (2001) has restricted the putative *Fab-4* boundary

to a small region using transposable enhancer traps (*HCJ200*, giving an *iab-3* specific signal and *pHFiab-4*, giving an *iab-4* specific signal). Our deletion removes most DNA in the *iab-4* region reaching this putative boundary (but not deleting it, see **Figure 4.1 B,D**). Furthermore, we can explain the *Abd-B* loss-of-function phenotype by the deleted putative *iab-5*^{L5} enhancer.

Even though the *iab-4* initiator element is most probably deleted, we get a partial *Abd-B* gain-of-function phenotype in A4. This would still be quite in line with the "open for business" model, as region *iab-4* (although almost not existing anymore) and *iab-5* are fused (*iab-5* initiator element is not affected by the *MM8* deletion).

In contrast, we can not explain the resulting A3 *Abd-B* gain-of-function phenotype. The PRE-dependent phenotype described in our manuscript does not hold up, as *MM8* should have deleted the late enhancer and therefore its signal can not be expanded.

In summary, the deletion *MM8* was very helpful to delimitate the DNA sequences removed in *iab5*^{MM5} (which was induced at a later time point, despite the name), but it raises more questions than it answers them and can, for the moment, be seen as a curiosity. Hopefully, we will understand the reason for such a phenotype in the future.

4.2 Transcription through the *Mcp* boundary and PRE pair

If the DNA loop model of boundary action holds true, one has to ask how those loops can be opened to permit changes in genome topology (see **Figure 1.19 D** for a loop model of the *Abd-B* gene).

One hypothesis is that non-coding transcription might play a role. Transcription through a boundary element may evict proteins that are bound to this DNA and therefore "break" the link to other boundary elements. Previous experiments have shown that a phenotype can be induced when transcription is sent through different BX-C regions (Bender and Fitzgerald, 2002; Hogga and Karch, 2002).

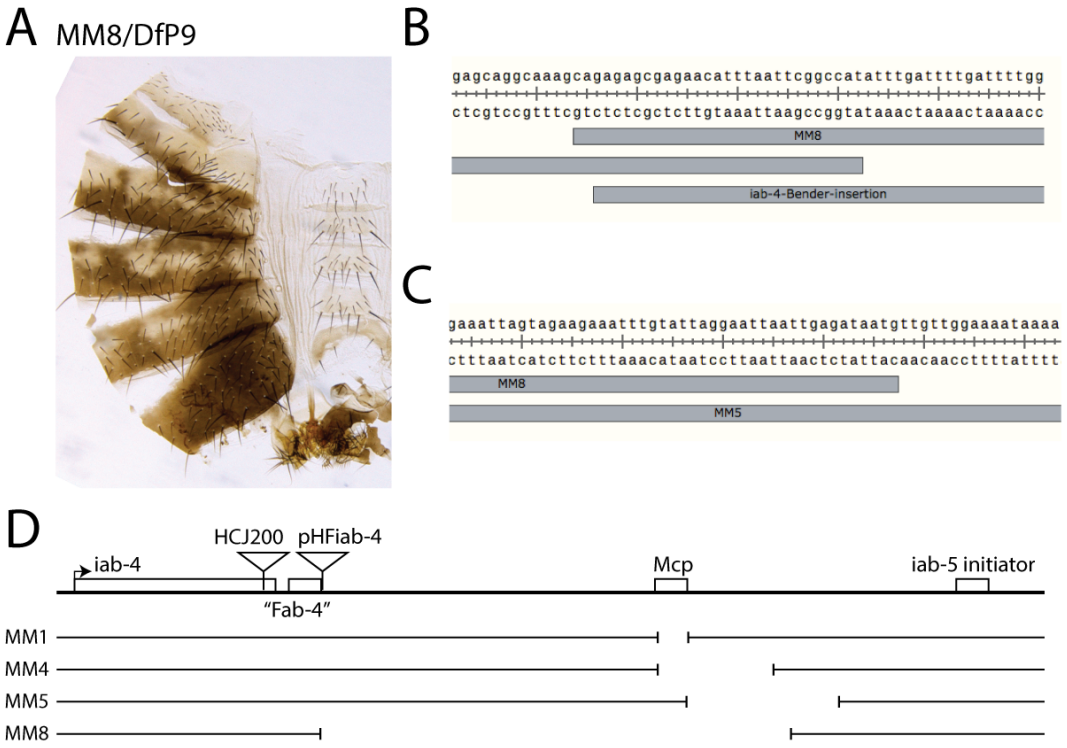


Figure 4.1: The *MM8* deletion - A: Phenotype of *MM8* deletion over *DfP9*. Note how there is an *Abd-B* loss-of-function phenotype on tergite A5. This mixed identity of tergite A5 is then reproduced both in A4 and A3. B: Proximal breakpoint of deletion *MM8*, note how the breakpoint (end of gray box named *MM8*) coincides with the insertion of the enhancer trap transposable element isolate by (Fitzgerald and Bender, 2001). C: Distal breakpoint of deletion *MM8*. D: Genomic topography of the *MM8* deletion. On top a line representing a stretch of the BX-C. The box with the arrow represents the non-coding *iab-4* RNA (promoter depicted as black arrow). The *Mcp* boundary, the putative *Fab-4* boundary, and the *iab-5* initiator are marked as black boxes. Insertions of enhancer trap transposable elements (from Bender and Hudson (2000); Fitzgerald and Bender (2001)) around the *Fab-4* putative boundary are marked as triangles. *HCJ200* gives a enhancer trap pattern related to *iab-3* activity (anterior expression border in PS8), while *pHFiab-4* an enhancer trap pattern related to *iab-4* activity (anterior expression border in PS9). Below, the three deletions Mcp^{MM1}, Mcp^{MM4} and $iab5^{MM5}$ (described already in the paper manuscript) are represented as interrupted lines. At the bottom, the span of the deletion *MM8* is depicted as an interrupted line. Note how the proximal breakpoint coincides with the "*iab-4* like" enhancer trap insertion.

Furthermore, it was shown that the activity of a PRE can be modulated according to transcriptional red-through direction (Herzog et al., 2014).

To test this hypothesis in the case of the *Mcp* boundary, we established three different constructs in the *Mcp* genomic locus. 1) A Gal4 inducible promoter (minimal heat shock promoter) driving transcription trough the *Mcp* boundary (wild type re-entry in the *Mcp* landing site) 2) A Gal4 inducible promoter driving transcription away from the Mcp boundary (in the *iab-4* region) 3) A control construct containing only the Gal4 binding sites (6xUAS), but no promoter (**Figure 4.2**). All of these lines do not show any phenotypes when no Gal4 is expressed (data not shown).

We used two different Gal4 drivers to induce transcription in those lines: 1) A *snail* driver that is active in histoblasts, and therefore in those cells that later will give rise to the adult tergite 2) An *actin* driver that is active throughout development in all tissues.

The results can be summarized as follows: the *Abd-B* gain-of-function phenotypes induced by transcription are the strongest when transcription is driven through the *Mcp* boundary element and PRE and when the *actin* driver is used. The combination of *actin* Gal4 driver and the inducible promoter inducing transcription through *Mcp* gives an A4 and A3 transformation to A5, similar to a *Mcp* null allele. The histoblast driver does not give such a strong phenotype, as A4 is partially pigmented and no pigmentation is visible in A3. Transcription away from *Mcp* does not give any additional pigmentation in the case of the histoblast driver, and no flies were viable with the *actin* driver. We can not explain this mortality, but probably transcription through the *abd-A* regulatory regions might cause this effect. To test if this mortality is due to *abd-A* over-expression we combined this construct with a GMR driver, that activates transcription in the eye. No eye phenotype was observed in such a case. Ectopic expression of a *UAS-abd-A* transgene in the eye leads to visible rough eye phenotypes (data not shown).

For the control construct (6xUAS, no minimal heat shock promoter) no ectopic pigmentation was visible with the histoblast driver, and some pigmentation was visible with the *actin* driver in A4 (but definitively less than with the construct driving transcription trough *Mcp*).

In conclusion: transcription trough the Mcp boundary element and PRE is able to induce an *Abd-B* gain-of-function phenotype. This phenotype might be linked to some functional regulation of the boundary and/or PRE by transcription, or simply be a result of boundary and/or PRE proteins being displaced from the DNA. Further studies are needed to understand this behavior. Nevertheless, it is interesting that with a completely different assay we are able to induce ectopic pigmentation on the tergite A3, and is therefore an ulterior sign that the *Mcp* region can be linked to such a phenotype.

For further studies, the following lines have already been established but not yet tested: *iab-4-Mcp-UAS-iab-5*, *iab-4-Mcp-UAS-hsp->-iab-5* and *iab-4-Mcp-<-hsp-UAS-iab-5*. The first line has the UAS sites (without a minimal *heath shock* promoter) inserted on the *iab-5* side of the *Mcp* boundary element/PRE pair, while the second and third line have the possibility (through the inducible minimal *heath shock* promoter) to induce transcription into the *iab-5* region (away from the boundary) or through *Mcp* into the *iab-4* region. Those lines will be instrumental as controls to the previously mentioned experiments, as it might be that it is not transcription through *Mcp* that induces the phenotype, but transcription through *iab-5*.

4.3 Establishment of a second landing site at the *Mcp* genomic location in the BX-C

The discovery of an additional cis-regulatory element in the *iab-5* region posed us in front of a problem: the established landing site Mcp^{attP} can not be used to dissect this elusive element. I decided to establish a new landing site that enables the dissection of the region deleted in Mcp^{MM4}. This new landing site was successfully established (Mcp^{attP2} in **Figure 4.3**). I've used the CRISPR/Cas9 toolset to induce DNA double strand breaks (using the following gRNAs: gctggcttttacagcatttc and gaatggggccatttgtgtat) and homologous recombination. The recombination vector used for the Mcp^{attP} landing site (see paper manuscript for details) was modified by removing all *iab-5* sequences until the KflI restriction site (see **Figure 4.3**). Therefore the right recombination arm was reduced to

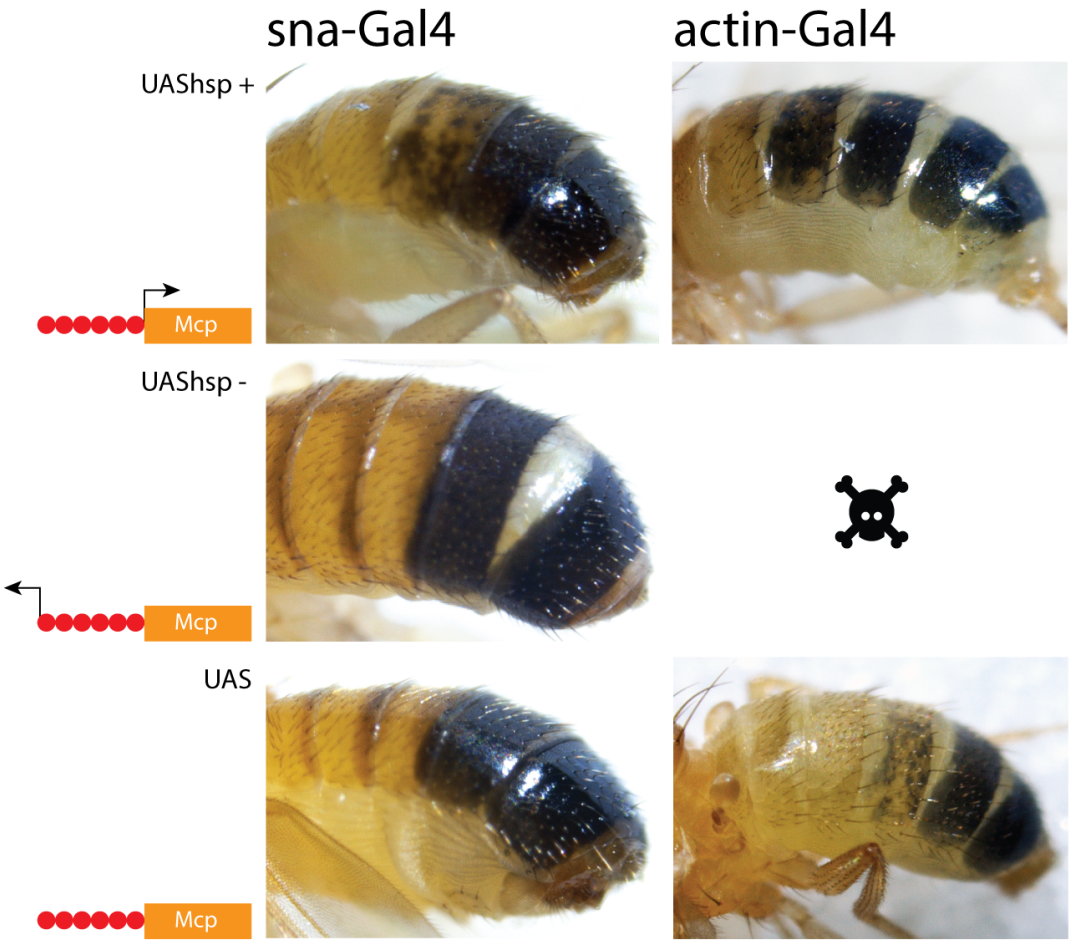


Figure 4.2: Pigmentation phenotypes in flies carrying an inducible promoter in the endogenous *Mcp* genomic location - Gal-4 was expressed using two drivers: *snail* (expressed in histoblasts) and *actin* (expressed in all tissues). In the *Mcp* endogenous genomic locations we inserted three UAS/promoter configurations proximal to the *Mcp* boundary element (therefore, the inducible promoters are in the *iab-4* region): 1) A Gal4 inducible promoter initiating transcription through the boundary element, 2) A Gal4 inducible promoter initiating transcription away from the boundary element, 3) UAS sites (Gal4 binding) without a promoter.

1.4kb instead of 4.3kb. Establishment of the new landing site Mcp^{attP2} was tested by PCR and DNA sequencing. The new landing site has therefore the same proximal break point as Mcp^{attP} and the distal break point at the KflI restriction side in *iab-5* mentioned before.

This new landing site shows no additional pigmentation in A3, but as predicted, a complete transformation of A4 into A5. Furthermore, as a curiosity, the integration of 3x3P-EGFP in Mcp^{attP2} as an injection marker gives expression in the eyes and the abdomen. This was not the case for Mcp^{attP} (only signal in abdomen, not eyes), indicating a role in eye tissue silencing of the unknown cis-regulatory element in *iab-5*. Furthermore, the EGFP levels in the larval cuticle are much higher for the newly established Mcp^{attP2} compared to the original landing site (data not shown).

Re-entry of constructs in the new Mcp^{attP2} landing site are on the way, and will hopefully help the deciphering of the *cis*-regulatory element $iab\text{-}5^{L5}$ and to explain in a more comprehensive way the difference in phenotype between Mcp^{MM1} (A4 and A3 to A5 transformation) and Mcp^{MM4} (only A4 to A5 transformation, no transformation of A3).

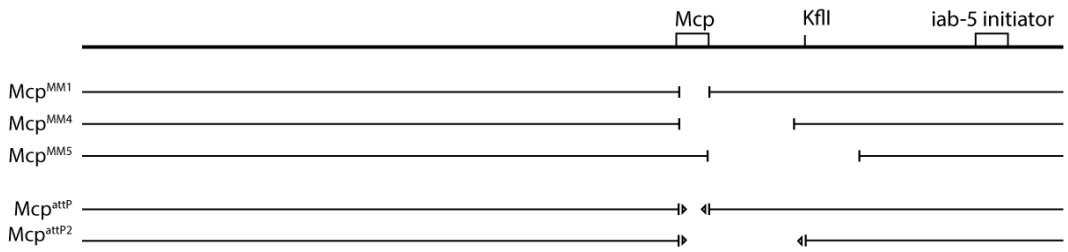


Figure 4.3: Establishment of a second landing site in the *Mcp* locus - Thick line on top represents the endogenous *Mcp* genomic locus. Boxes indicate the positions of the core *Mcp* boundary element/PRE pair and the *iab-5* initiator element. The KflI restriction site used to shorten the homology arm of the recombination plasmid used for Mcp^{attP} is indicated. Deletions Mcp^{MM1}, Mcp^{MM4} and $iab5^{MM5}$ are described in the paper manuscript. Mcp^{attP} is the landing site used in the experiments described in the paper manuscript (black arrowheads representing attP sites). The new landing site is marked as Mcp^{attP2}. This landing site removes more DNA on the distal site of the *Mcp* boundary compared to the Mcp^{attP} landing site. The induced deletion by landing site integration is comparable to the Mcp^{MM4} deletion.

5

Summary and outlook

In my work I have established tools that allow an unprecedented view at the boundary element and the Polycomb Response Element at the endogenous *Miscadestral pigmentation* locus in the Bithorax Hox Complex.

The CRISPR/Cas9 method enables study of *cis*-regulatory elements at the endogenous locus with a previously inconceivable ease (in terms of time, money, and tools). In the time of my doctoral studies I was in the right position, and at the right time, to use its potential to the fullest. Future studies of *cis*-regulatory elements in *Drosophila* will lack rigor when not investigated at their endogenous genomic location. Every laboratory can use the new inexpensive CRISPR/Cas9 tool set without much prior experience.

I have established two phiC31-integrase-dependent landing sites that can be used for further work in studying the effect of mutagenized DNA sequences at the endogenous boundary between the *abd-A* and *Abd-B* gene in the BX-C.

My studies have dissected the *Mcp* region using two strategies: 1) establishment of precise deletions induced by CRISPR/Cas9 DNA double strand breaks and non-homologous-end-joining DNA repair mechanism. 2) Use of the established phiC31-integrase dependent landing site. Both methods were extremely important to dissect the sequence requirements of the *Mcp* boundary element and PRE in great detail.

My studies document that the *Mcp* region defined by the classical Mcp^1 deletion contains at least three functional elements. Two of them, the boundary element and the PRE, have been inferred from transgenic studies. Both of them

are crucially involved in keeping in check the anterior expression limit of the *Abd-B* gene. A third *cis*-regulatory element is located immediately distal to the boundary/PRE pair. My data suggests that this element could be an enhancer active later in development ($iab5^{L5}$). Deletion of $iab5^{L5}$ leads to an *Abd-B* loss-of-function phenotype: in adult males, the identity of segment A5 is changed to A4. In contrast, late embryos that carry the deletion of the putative enhancer do not show an *Abd-B* phenotype. I therefore argue that the change in *Abd-B* expression happens only in later stages of development.

An unexpected phenotype was observed when clean deletions of the *Mcp* boundary/PRE pair were created. Current models for *Abd-B* function predicted that these alleles would cause a gain-of-function phenotype that transforms segment A4 into A5. However, apart from this expected phenotype, also segment A3 was partially transformed into A5 identity. My deletion analysis indicates that the *Mcp* boundary/PRE pair differentially contributes to these phenotypes: 1) Deletion of the CTCF binding site in the boundary element gives a A4 tergite transformation to A5, therefore this construct rescues the A3 phenotype but not the A4 phenotype. 2) Deletion of the PRE leads to strong A4 to A5 tergite transformation, but spots of pigmentation are also visible in A3. Therefore, the additional *Abd-B* gain-of-function observed in *Mcp* boundary element/PRE pair deletions are due to missing PRE sequences.

My model on *Mcp* function proposes that the *Mcp* PRE keeps the putative $iab5^{L5}$ enhancer in check. Ongoing work is testing the possible enhancer activity of $iab5^{L5}$ with reporter constructs. Furthermore, the Mcp^{attP2} landing site should be a powerful tool to investigate the nature of this *cis*-regulatory element *in situ*. Alternatively, $iab5^{L5}$ could also act as a PTS. Such elements are supposed to mediate interaction between the different *iabs* and the *Abd-B* promoter. As this interaction is important, disturbing this mechanism might lead to unexpected phenotypes. Such PTS elements were identified in the *iab-6* and *iab-7* regions but not in *iab-5* yet. Furthermore, as the results of my work include the finding that DNA sequences thought to contain certain *cis*-regulatory elements might contain more elements than predicted, I do not want to rule out that the $iab\text{-}5^{L5}$ region contains more than just one additional *cis*-regulatory element.

From a biochemical point of view, it would be of big interest to know how the chromatin of the Abd-B domain is organized in some of the newly created *Mcp* alleles. I would be very curious to know how some of those deletions affect the topography of chromatin loops proposed to exist between *Mcp* and flanking boundary elements and/or the *Abd-B* promoter. A possible technique that would allow such insights is the circular chromosome conformation capture (4C) method (Zhao et al., 2006). This chromosome capture method allows to investigate the interaction of one specific locus (in this case *Mcp*) with the rest of the genome.

To conclude, the work presented in this thesis highlights the importance and feasibility of doing *cis*-regulatory element research at the endogenous genomic locus. Even elements characterized long ago by deletions in the endogenous genomic location might be mis-understood when not looked at with the genetic resolution that is possible nowadays.

6

Materials & methods

6.1 Abdominal cuticle preparations

6.1.1 Reagents

- Fixing solution: 75% Ethanol /25% Glycerol
- Tissue cleaning solution: 10% KOH
- Hoyer's medium: dissolve 15g of gum arabic in 25mL of water. Heat to 60C°and stir overnight. Add 100g of chloral hydrate, and after it has dissolved, add 10g of glycerol. Centrifuge at 10'000g for half an hour and filter through glass wool. Recover supernatant and store at room temperature. All steps should be performed unter the fume hood.

6.1.2 Protocol

1. Fix overnight in fixing solution
2. Separate abdomen from thorax using two pair of forceps on a glass slide in fixing solution
3. Fix abdomens with one pair of forceps with the dorsal side looking upwards and cut with a razor blade along the dorsal midline (it helps when one side of the blade is rested on the coverslip). Make sure that the genitals are cut, as this will help with following steps

4. Transfer the cut abdomens into tissue cleaning solution for 15 minutes
5. Transfer the abdomens on a drop of cleaning solution on a glass slide and flatten them with the tips of forceps
6. Add a coverslip. If the flattening was done correctly, the half tergites should be flatten out nicely on both sides of the pleural tissue
7. Incubate at 55°C for 4 hours
8. Incubate the glass slide vertically in a beaker glass full of water for around two minutes. After that, when removing the slides from the beaker the cover slip will slide away
9. Transfer the abdomens in a drop of Hoyer's medium, arrange them by flattening them, and add a cover slip
10. Incubate over night at 55°C
11. Images were taken with a Leica microscopy camera on a Leica binocular

6.2 Embryo fixation with Abd-B and Engrailed or Abd-A and Engrailed antibody staining

6.2.1 Reagents

- 4% sodium hypochlorite (bleach)
- Fixing solution: 1 Volume 4% V/V paraformaldehyde in PBS : 1 Volume Heptane
- 1X PBS buffer: 8.0g/L NaCl, 0.2g/L KCl, 1.42g/L Na_2HPO_4, 0.24g/L KH_2PO_4
- PBT buffer: 1X PBS, 0.3% V/V Triton X-100
- PBTN: 2% V/V Normal goat serum in PBT

- Abd-B/en primary staining solution: 1:10 anti-ABD-B (mouse, 1A2E9) from Developmental Studies Hybridoma Bank, 1:500 engrailed Antibody (rabbit, d-300, sc-28640) from Santa Cruz Biotechnology in PBTN
- abd-A/en primary staining solution: 1:100 AbdA Antibody (mouse, C-11, sc-390990) from Santa Cruz Biotechnology, 1:500 engrailed Antibody (rabbit, d-300, sc-28640) from Santa Cruz Biotechnology in PBTN
- Secondary staining solution: 1:1000 goat anti-mouse Alexa Fluor 488 from life technologies, 1:1000 goat anti rabbit Cy5 from Jackson ImmunoResearch
- Mounting medium: Vectashield H-1000 from Vector Laboratories

6.2.2 Protocol

Fixation

1. Prepare the egg laying plates by dropping liquid yeast on grape juice plates. Dry in front of blowing heater.
2. Cages on grape juice plates were filled with around 200 adult flies
3. Keep the cages running for a couple of day before collections. Best results are achieved if cage is substituted with a clean one every day, and grape juice plates changed in the morning and in the evening
4. Collect embryos after around 14 hours
5. Pour 4% bleach onto the grape juice plates containing the embryos, incubate the embryos for two minutes
6. Collect embryos by pouring the bleach solution through nylon nets
7. Transfer embryos to fixing solution and shake for 20 minutes
8. Remove lower phase
9. Add 1 Volume of methanol
10. Vortex for 1 minute

11. Remove top phase (embryos will fall to the bottom of eppendorf tube)
12. Wash directly 5 times with methanol
13. Embryos can be stored in methanol at -20°C

Staining

1. Wash with 1mL PBT, 3x directly and 6x10 minutes at room temperature
2. Incubate in PBTN for 30 minutes at room temperature
3. Incubate with primary staining solution overnight at 4°C
4. Wash with 1mL PBT, 3x directly and 6x10 minutes at room temperature
5. Incubate with secondary staining solution for 2 hours at room temperature
6. Wash with 1mL PBT, 3x directly and 6x10 minutes at room temperature
7. Remove all the liquids and add three drops of mounting medium
8. Prepare the glass slides by applying double sticking tape around the edge, the size should be similar to cover slips used
9. Transfer the embryos in the mounting medium to the prepared glass slide. Distribute the embryos on the slide with a plastic pipette tip
10. Add the cover slip. It should stick on the tape wall and not press the embryos
11. Add nail polish around the coverslip edge
12. Images were taken with a Leica SP5 confocal microscope

6.3 Preparation of electro-competent *E. coli* cells

6.3.1 Reagents

- Low salt LB: 10g tryptone, 5g yeast extract, fill up to 1L with distilled water.
- 10% glycerol (sterile)

6.3.2 Protocol

1. Inoculate 10ml low salt LB with your bacteria strain. Let grow overnight at 37°C. Add antibiotics according to resistance of bacteria.
2. In the morning dilute culture 1:100 in 1L of low salt LB with no antibiotics. Let culture grow in two 3L Erlenmeyer flasks with baffles.
3. Grow until OD 0.6 to 0.8 (3 to 4 hours). When this threshold is reached, immediately cool in ice slurry and never warm up before the final freezing.
4. Distribute the 1L culture to four centrifuge bottles fitting the SLA-3000 rotor. Pre-cool the bottles and centrifuge to 4°C. Centrifuge for 10 minutes at 3300 rpm.
5. Resuspend the pellet in each bottle with 5ml sterile and cooled 10% glycerol with a glass pipette.
6. Fill 2ml Eppendorf tubes with the suspended bacteria glycerol solution.
7. Centrifuge at 8000rpm in cooled bench-top centrifuge.
8. Remove supernatant and resuspend in 1ml 10% glycerol.
9. Prepare 50μl aliquots and flash-freeze in liquid nitrogen.
10. Store at -80°C.

7

Copyright disclaimer

All foreign materials in the introduction of this thesis (mainly panels in the figures) are described in the figure legends and the source is cited. Figures with no citations are either new figures by myself or own figures taken from my master thesis (Metzler, 2012). Furthermore, some text from my master thesis (like figure legends and single paragraphs in the introduction) was reused and edited for this doctoral thesis.

Bibliography

Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., Russell, S., and White, R. (2007). Genomic mapping of suppressor of hairy-wing binding sites in drosophila. *Genome biology*, 8(88):R167. 37

Affolter, M., Percival-Smith, A., Müller, M., Leupin, W., and Gehring, W. J. (1990). Dna binding properties of the purified antennapedia homeodomain. *Proceedings of the National Academy of Sciences*, 87(1111):4093–4097. 7

Akam, M. (1987). The molecular basis for metameric pattern in the drosophila embryo. *Development (Cambridge, England)*, 101(11):1–22. 5

Attrill, H., Falls, K., Goodman, J. L., Millburn, G. H., Antonazzo, G., Rey, A. J., Marygold, S. J., and Consortium, F. (2016). Flybase: establishing a gene group resource for drosophila melanogaster. *Nucleic acids research*, 44:NaN–NaN. 88

Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Müller, 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 (Cambridge, England)*, 127(44):779–790. 41, 60

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). Crispr provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(58195819):1709–1712. 54

Bateman, J. R., Lee, A. M., and Wu, C. t. (2006). Site-specific transformation of drosophila via ϕc31 integrase-mediated cassette exchange. *Genetics*, 173(22):769–777. 52, 64, 70

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B., and Hogness, D. S. (1983). Molecular genetics of the bithorax complex in drosophila melanogaster. *Science (New York, N.Y.)*, 221(46054605):23–29. 17, 20, 33

Bender, W. and Fitzgerald, D. P. (2002). Transcription activates repressed domains in the drosophila bithorax complex. *Development (Cambridge, England)*, 129(21):4923–4930. 94

Bender, W. and Hudson, A. (2000). P element homing to the drosophila bithorax complex. *Development (Cambridge, England)*, 127(1818):3981–3992. 93, 95

Bender, W. and Lucas, M. (2013). The border between the ultrabithorax and abdominal-a regulatory domains in the drosophila bithorax complex. *Genetics*, 193(44):1135–1147. 41

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T., and et al. (2008). Variation in homeodomain dna binding revealed by high-resolution analysis of sequence preferences. *Cell*, 133(77):1266–1276. 7

Bieli, D., Kanca, O., Gohl, D., Denes, A., Schedl, P., Affolter, M., and Müller, M. (2015a). The drosophila melanogaster mutants apblot and apxasta affect an essential apterous wing enhancer. *G3 (Bethesda, Md.)*, 5(66):1129–1143. 64, 65, 76

Bieli, D., Kanca, O., Requena, D., Hamaratoglu, F., Gohl, D., Schedl, P., Affolter, M., Slattery, M., Müller, M., and Estella, C. (2015b). Establishment of a developmental compartment requires interactions between three synergistic cis-regulatory modules. *PLoS genetics*, 11(1010):e1005376. 66

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for drosophila using germ-line-specific phic31 integrases. *Proceedings of the National Academy of Sciences*, 104(99):3312–3317. 52, 64

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (crisprs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, 151:2551–2561. 54

Bornemann, D., Miller, E., and Simon, J. (1998). Expression and properties of wild-type and mutant forms of the drosophila sex comb on midleg (scm) repressor protein. *Genetics*, 150(22):675–686. 62

Bowman, S. K., Deaton, A. M., Domingues, H., Wang, P. I., Sadreyev, R. I., Kingston, R. E., Bender, W., and Reinberg, D. (2014). H3k27 modifications define segmental regulatory domains in the drosophila bithorax complex. *eLife*, 3:e02833. 28, 50, 60

Bushey, A. M., Ramos, E., and Corces, V. G. (2009). Three subclasses of a drosophila insulator show distinct and cell type-specific genomic distributions. 23(1111):1338–1350. 51

Busturia, A. and Bienz, M. (1993). Silencers in abdominal-b, a homeotic drosophila gene. *The EMBO journal*, 12(44):1415–1425. 44, 59

Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The mcp silencer of the drosophila abd-b gene requires both pleiohomeotic and gaga factor for the maintenance of repression. *Development (Cambridge, England)*, 128(1111):2163–2173. 23, 25, 27, 44, 46, 62, 73, 84

Busturia, A. and Morata, G. (1988). Ectopic expression of homeotic genes caused by the elimination of the polycomb gene in drosophila imaginal epidermis. *Development (Cambridge, England)*, 104(44):713–720. 27

Carroll, S. B. and Vavra, S. H. (1989). The zygotic control of drosophila pair-rule gene expression. ii. spatial repression by gap and pair-rule gene products. *Development (Cambridge, England)*, 107(33):673–683. 5

Casares, F. and Sánchez-Herrero, E. (1995). Regulation of the infraabdominal regions of the bithorax complex of drosophila by gap genes. *Development (Cambridge, England)*, 121(66):1855–1866. 6

Cavalli, G. and Paro, R. (1998). The drosophila fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell*, 93(44):505–518. 28

BIBLIOGRAPHY

Celniker, S. E., Sharma, S., Keelan, D. J., and Lewis, E. B. (1990). The molecular genetics of the bithorax complex of drosophila: cis-regulation in the abdominal-b domain. *The EMBO journal*, 9(1313):4277–4286. 30, 42, 43, 62, 83

Chan, S.-K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. (1994). The dna binding specificity of ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, 78(44):603–615. 9

Chen, Q., Lin, L., Smith, S., Lin, Q., and Zhou, J. (2005). Multiple promoter targeting sequences exist in abdominal-b to regulate long-range gene activation. *Developmental biology*, 286(22):629–636. 32

Chiang, A., O'Connor, M. B., Paro, R., Simon, J., and Bender, W. (1995). Discrete polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development (Cambridge, England)*, 121(6):1681–1689. 25

Cléard, F., Moshkin, Y., Karch, F., and Maeda, R. K. (2006). Probing long-distance regulatory interactions in the drosophila melanogaster bithorax complex using dam identification. *Nature genetics*, 38(88):931–935. 41, 61

Comet, I., Savitskaya, E., Schuettengruber, B., Nègre, N., Lavrov, S., Parshikov, A., Juge, F., Gracheva, E., Georgiev, P., and Cavalli, G. (2006). Pre-mediated bypass of two su(hw) insulators targets pcg proteins to a downstream promoter. *Developmental cell*, 11(1):117–124. 37

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and et al. (2013). Multiplex genome engineering using crispr/cas systems. *Science (New York, N.Y.)*, 339(61216121):819–823. 54

Crosby, M. A., Lundquist, E. A., Tautvydas, R. M., and Johnson, J. J. (1993). The 3' regulatory region of the abdominal-b gene: genetic analysis supports a model of reiterated and interchangeable regulatory elements. *Genetics*, 134(33):809–824. 42, 43

Curtiss, J. and Heilig, J. S. (1995). Establishment of drosophila imaginal precursor cells is controlled by the arrowhead gene. *Development (Cambridge, England)*, 121(1111):3819–3828. 13, 15

Deschamps, J. and van Nes, J. (2005). Developmental regulation of the hox genes during axial morphogenesis in the mouse. *Development (Cambridge, England)*, 132(1313):2931–2942. 2

DiNardo, S. and O'Farrell, P. H. (1987). Establishment and refinement of segmental pattern in the drosophila embryo: spatial control of engrailed expression by pair-rule genes. *Genes & development*, 1(1010):1212–1225. 5

Doudna, J. A. and Charpentier, E. (2014). Genome editing. the new frontier of genome engineering with crispr-cas9. *Science (New York, N.Y.)*, 346(62136213):1258096. 54, 56

Driever, W. and Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in drosophila embryos. *Cell*, 54(11):83–93. 3

Duncan, I. W. (2002). Transvection effects in drosophila. *Annual review of genetics*, 36:521–556. 58

Ekker, S. C., Jackson, D. G., Kessler, D. P. v., Sun, B. I., Young, K. E., and Beachy, P. A. (1994). The degree of variation in dna sequence recognition among four drosophila homeotic proteins. *The EMBO journal*, 13(1515):3551. 7

Ekker, S. C., Young, K. E., Kessler, D. P. v., and Beachy, P. A. (1991). Optimal dna sequence recognition by the ultrabithorax homeodomain of drosophila. *The EMBO journal*, 10(55):1179. 7

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of polycomb response elements. 112(4848):201515276. 28

Fitzgerald, D. P. and Bender, W. (2001). Polycomb group repression reduces dna accessibility. *Molecular and cellular biology*, 21(19):6585–6597. 93, 95

Foronda, D., Estrada, B., de Navas, L., and Sánchez-Herrero, E. (2006). Requirement of abdominal-a and abdominal-b in the developing genitalia of drosophila breaks the posterior downregulation rule. *Development (Cambridge, England)*, 133(11):117–127. 23

Frasch, M., Warrior, R., Tugwood, J., and Levine, M. (1988). Molecular analysis of even-skipped mutants in drosophila development. *Genes & development*, 2(1212):1824–1838. 5

Galloni, M., Gyurkovics, H., Schedl, P., and Karch, F. (1993). The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. *The EMBO journal*, 12(33):1087–1097. 30, 38

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 & development*, 13(1616):2098–2107. 37

Gehring, W. J., Affolter, M., and Burglin, T. (1994). Homeodomain proteins. *Annual review of biochemistry*. 7

Geyer, P. K. and Corces, V. G. (1992). Dna position-specific repression of transcription by a drosophila zinc finger protein. *Genes & development*, 6(1010):1865–1873. 35, 37

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. *Proceedings of the National Academy of Sciences*, 85(2222):8593–8597. 35

Girton, J. R. and Jeon, S. H. (1994). Novel embryonic and adult homeotic phenotypes are produced by pleiohomeotic mutations in drosophila. *Developmental biology*, 161(22):393–407. 63

Gohl, D., Müller, M., Pirrotta, V., Affolter, M., and Schedl, P. (2008). Enhancer blocking and transvection at the drosophila apterous locus. *Genetics*, 178(11):127–143. 62, 76

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J., and O'Connor-Giles, K. M. (2013a). Genome engineering of drosophila with the crispr rna-guided cas9 nuclease. *Genetics*, 194(44):1029–1035. 55

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., and O'Connor-Giles, K. M. (2014). Highly specific and efficient crispr/cas9-catalyzed homology-directed repair in drosophila. 196(44):961–971. 63, 64, 65

Gratz, S. J., Wildonger, J., Harrison, M. M., and O'Connor-Giles, K. M. (2013b). Crispr/cas9-mediated genome engineering and the promise of designer flies on demand. *Fly*, 7(44):249–255. 56

Grieder, N. C., Marty, T., Ryoo, H. D., Mann, R. S., and Affolter, M. (1997). Synergistic activation of a drosophila enhancer by hom/exd and dpp signaling. *The EMBO journal*, 16(2424):7402–7410. 9

BIBLIOGRAPHY

Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A., and Georgiev, P. (2005). The mcp element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. *Molecular and cellular biology*, 25(99):3682–3689. 47, 48, 62

Guerra, M., Postlethwait, J. H., and Schneiderman, H. A. (1973). The development of the imaginal abdomen of drosophila melanogaster. *Developmental biology*, 32(22):361–372. 13

Gummalla, M., Maeda, R. K., Castro Alvarez, J. J., Gyurkovics, H., Singari, S., Edwards, K. A., Karch, F., and Bender, W. (2012). abd-a regulation by the iab-8 noncoding rna. *PLoS genetics*, 8(55):e1002720. 9

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., and et al. (2015). Crispr inversion of ctcf sites alters genome topology and enhancer/promoter function. *Cell*, 162(44):900–910. 47

Gyurkovics, H., Gausz, J., Kummer, J., and Karch, F. (1990). A new homeotic mutation in the drosophila bithorax complex removes a boundary separating two domains of regulation. *The EMBO journal*, 9(88):2579–2585. 30, 38, 39, 60

Hama, C., Ali, Z., and Kornberg, T. B. (1990). Region-specific recombination and expression are directed by portions of the drosophila engrailed promoter. *Genes & development*, 4(77):1079–1093. 13

Harding, K. and Levine, M. (1988). Gap genes define the limits of antennapedia and bithorax gene expression during early development in drosophila. *The EMBO journal*, 7(11):205–214. 6

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). Ctcf: insights into insulator function during development. *Development (Cambridge, England)*, 139(6):1045–1057. 76

Herzog, V. A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., Ruf, M., Pospisilik, A., Dimond, A., Senergin, H. B., Vargas, M. L., Simon, J. A., and Ringrose, L. (2014). A strand-specific switch in noncoding transcription switches the function of a polycomb/trithorax response element. *Nature Genetics*, 46(9):973–981. 28, 96

Ho, M. C. W., Johnsen, H., Goetz, S. E., Schiller, B. J., Bae, E., Tran, D. A., Shur, A. S., Allen, J. M., Rau, C., Bender, W., and et al. (2009). Functional evolution of cis-regulatory modules at a homeotic gene in drosophila. *PLoS genetics*, 5(1111):e1000709. 67

Hogga, I. and Karch, F. (2002). Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of polycomb-mediated silencing. *Development (Cambridge, England)*, 129(21):4915–4922. 94

Hogga, I., Mihaly, J., Barges, S., and Karch, F. (2001). Replacement of fab-7 by the gypsy or scs insulator disrupts long-distance regulatory interactions in the abd-b gene of the bithorax complex. *Molecular cell*, 8(55):1145–1151. 41

Holohan, E. E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). Ctcf genomic binding sites in drosophila and the organisation of the bithorax complex. *PLoS genetics*, 3(77):e112. 62

Hurley, I., Hale, M. E., and Prince, V. E. (2005). Duplication events and the evolution of segmental identity. *Evolution & Development*, 7(66):556–567. 3

Iampietro, C., Cléard, F., Gyurkovics, H., Maeda, R. K., and Karch, F. (2008). Boundary swapping in the drosophila bithorax complex. *Development (Cambridge, England)*, 135(2424):3983–3987. 41

Iampietro, C., Gummalla, M., Mutero, A., Karch, F., and Maeda, R. K. (2010). Initiator elements function to determine the activity state of bx-c enhancers. *PLoS genetics*, 6(1212):e1001260. 21, 61

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in escherichia coli, and identification of the gene product. *Journal of bacteriology*, 169(1212):5429–5433. 54

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-rna-guided dna endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(60966096):816–821. 54

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). Rna-programmed genome editing in human cells. *eLife*, 2:e00471. 54

Karch, F., Bender, W., and Weiffenbach, B. (1990). abda expression in drosophila embryos. *Genes & development*, 4(99):1573–1587. 9, 42

Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H., and Schedl, P. (1994). Mcp and fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of drosophila melanogaster. *Nucleic acids research*, 22(1515):3138–3146. 41, 42, 44, 60, 62, 66

Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M., and Lewis, E. B. (1985). The abdominal region of the bithorax complex. *Cell*, 43(11):81–96. 17, 67

Kassis, J. A. (2002). Pairing-sensitive silencing, polycomb group response elements, and transposon homing in drosophila. *Advances in genetics*, 46:421–438. 46, 76

Kellum, R. and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, 64(55):941–950. 34, 37

Kennison, J. A. (1993). Transcriptional activation of drosphila homeotic genes form distant regulatory elements. *Trends in genetics : TIG*, 9(33):75–79. 25

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., and et al. (2015). Engineered crispr-cas9 nucleases with altered pam specificities. *Nature*, 523(75617561):481–485. 55

Kmita, M. and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science (New York, N.Y.)*, 301(56315631):331–333. 2

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics1. *dx.doi.org*. 3

Kopp, A., Muskavitch, M. A., and Duncan, I. (1997). The roles of hedgehog and engrailed in patterning adult abdominal segments of drosophila. *Development (Cambridge, England)*, 124(1919):3703–3714. 15

Kornberg, T., Sidén, I., O'Farrell, P., and Simon, M. (1985). The engrailed locus of drosophila: In situ localization of transcripts reveals compartment-specific expression. *Cell*, 40(11):45–53. 11

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W., and Bier, E. (2004). Multiplex detection of rna expression in drosophila embryos. *Science (New York, N.Y.)*, 305(56855685):846. 3

Kuhn, E. J. and Geyer, P. K. (2003). Genomic insulators: connecting properties to mechanism. *Current Opinion in Cell Biology*, 15(33):259–265. 33, 34

Kuhn, E. J., Viering, M. M., Rhodes, K. M., and Geyer, P. K. (2003). A test of insulator interactions in drosophila. *The EMBO journal*, 22(10):2463–2471. 35

Kuziora, M. A. and McGinnis, W. (1988). Autoregulation of a drosophila homeotic selector gene. *Cell*, 55(33):477–485. 25

Kyrchanova, O., Ivlieva, T., Toshchakov, S., Parshikov, A., Maksimenko, O., and Georgiev, P. (2011). Selective interactions of boundaries with upstream region of abd-b promoter in drosophila bithorax complex and role of dctcf in this process. *Nucleic acids research*, 39(88):3042–3052. 41, 61

Kyrchanova, O., Leman, D., Parshikov, A., Fedotova, A., Studitsky, V., Maksimenko, O., and Georgiev, P. (2013). New properties of drosophila scs and scs' insulators. *PloS one*, 8(44):e62690. 37

Kyrchanova, O., Mogila, V., Wolle, D., Deshpande, G., Parshikov, A., Cléard, F., Karch, F., Schedl, P., and Georgiev, P. (2016). Functional dissection of the blocking and bypass activities of the fab-8 boundary in the drosophila bithorax complex. *PLoS genetics*, 12(77):e1006188. 32, 41, 42

Kyrchanova, O., Mogila, V., Wolle, D., Magbanua, J. P., White, R., Georgiev, P., and Schedl, P. (2015). The boundary paradox in the bithorax complex. *Mechanisms of development*, 138:122–132. 32

Kyrchanova, O., Toshchakov, S., Parshikov, A., and Georgiev, P. (2007). Study of the functional interaction between mcp insulators from the drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication. *Molecular and cellular biology*, 27(88):3035–3043. 47, 49, 62, 81

Laughon, A. and Scott, M. P. (1984). Sequence of a drosophila segmentation gene: protein structure homology with dna-binding proteins. *Nature*, 310(59725972):25–31. 2

Lawrence, P. A. (1988). The present status of the parasegment. *Development (Cambridge, England)*, 104:61–65. 11

Lawrence, P. A., Casal, J., and Struhl, G. (1999). hedgehog and engrailed: pattern formation and polarity in the drosophila abdomen. *Development (Cambridge, England)*, 126(1111):2431–2439. 15

Lawrence, P. A. and Martinez-Arias, A. (1985). The cell lineage of segments and parasegments in drosophila. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 312(11531153):83–90. 13

Lehmann, R. and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in drosophila. *Cell*, 47(11):141–152. 3

Lehmann, R. and Nusslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the drosophila embryo. *Development (Cambridge, England)*, 112(33):679–691. 3

Lemons, D. and McGinnis, W. (2006). Genomic evolution of hox gene clusters. *Science (New York, N.Y.)*, 313(57955795):1918–1922. 2

Levis, R., Hazelrigg, T., and Rubin, G. M. (1985). Effects of genomic position on the expression of transduced copies of the white gene of drosophila. *Science (New York, N.Y.)*, 229(47134713):558–561. 34

Lewis, E. (1978). A gene complex controlling segmentation in drosophila. *Nature*, 276(56885688):565–570. 1, 2, 11, 58

Li, H.-B., Müller, M., Bahechar, I. A., Kyrchanova, O., Ohno, K., Georgiev, P., and Pirrotta, V. (2011). Insulators, not polycomb response elements, are required for long-range interactions between polycomb targets in drosophila melanogaster. *Molecular and cellular biology*, 31(44):616–625. 45, 47, 62, 76, 78, 88

Macías, A., Pelaz, S., and Morata, G. (1994). Genetic factors controlling the expression of the abdominal-a gene of drosophila within its domain. *Mechanisms of development*, 46(11):15–25. 6, 59

Maconochie, M., Nonchev, S., Morrison, A., , , and Krumlauf, R. (2003). Paralogous hox genes:function and regulation. *dx.doi.org*. 3

Maeda, R. K. and Karch, F. (2006). The abc of the bx-c: the bithorax complex explained. *Development (Cambridge, England)*, 133(88):1413–1422. 23, 31, 59

Maeda, R. K. and Karch, F. (2009). The bithorax complex of drosophila an exceptional hox cluster. *Current topics in developmental biology*, 88:1–33. 25, 26, 52

Maeda, R. K. and Karch, F. (2015). The open for business model of the bithorax complex in drosophila. *Chromosoma*, 124(33):293–307. 18, 20, 22, 50, 60

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006). A putative rna-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic rnai, and hypothetical mechanisms of action. *Biology direct*, 1(11):7. 54

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). Rna-guided human genome engineering via cas9. *Science (New York, N.Y.)*, 339(61216121):823–826. 54

Mann, R. S. (1994). Engrailed-mediated repression of ultrabithorax is necessary for the parasegment 6 identity in drosophila. *Development (Cambridge, England)*, 120(1111):3205–3212. 6, 7, 59

Martinez-Arias, A. and Lawrence, P. A. (1985). Parasegments and compartments in the drosophila embryo. *Nature*, 313(60046004):639–642. 11, 13

Matharu, N. and Ahituv, N. (2015). Minor loops in major folds: Enhancer–promoter looping, chromatin restructuring, and their association with transcriptional regulation and disease. *PLoS genetics*, 11(1212):e1005640. 50, 52

Mathelier, A., Fornes, O., Arenillas, D. J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., and et al. (2016). Jaspar 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic acids research*, 44:NaN–NaN. 73

Matzat, L. H. and Lei, E. P. (2014). Surviving an identity crisis: A revised view of chromatin insulators in the genomics era. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(33):203–214. 33, 34

McElroy, K. A., Kang, H., and Kuroda, M. I. (2014). Are we there yet? initial targeting of the male-specific lethal and polycomb group chromatin complexes in drosophila. *Open Biology*, 4(33):140006. 48

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., and Gehring, W. J. (1984a). A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. *Cell*, 37(22):403–408. 3

BIBLIOGRAPHY

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984b). A conserved dna sequence in homoeotic genes of the drosophila antennapedia and bithorax complexes. *Nature*, 308(59585958):428–433. 3

Metzler, M. (2012). Genetic characterization of the insulator function contained in miscadestral pigmentation. 109

Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cléard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch, F. (2006). Dissecting the regulatory landscape of the abd-b gene of the bithorax complex. *Development (Cambridge, England)*, 133(1515):2983–2993. 20, 23, 32, 41

Mihaly, J., Hogga, I., Gausz, J., Gyurkovics, H., and Karch, F. (1997). In situ dissection of the fab-7 region of the bithorax complex into a chromatin domain boundary and a polycomb-response element. *Development (Cambridge, England)*, 124(99):1809–1820. 38, 39, 60, 61, 83

Mihaly, J., Mishra, R. K., and Karch, F. (1998). A conserved sequence motif in polycomb-response elements. *Molecular cell*, 1(77):1065–1066. 60

Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S. E., and Schedl, P. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both gaga and pleiohomeotic for silencing activity. *Molecular and cellular biology*, 21(44):1311–1318. 60

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. *Proceedings of the National Academy of Sciences*, 80(6):1678–1682. 37

Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R. A. H., Renkawitz-Pohl, R., Saumweber, H., and et al. (2007). The drosophila insulator proteins ctcf and cp190 link enhancer blocking to body patterning. *The EMBO journal*, 26(1919):4203–4214. 63

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(22):174–182. 54

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S. T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., and et al. (2005). Ctcf is conserved from drosophila to humans and confers enhancer blocking of the fab-8 insulator. *EMBO reports*, 6(22):165–170. 42

Morata, G. and Kerridge, S. (1981). Sequential functions of the bithorax complex of drosophila. *Nature*, 290(58095809):778–781. 11

Morata, G. and Lawrence, P. A. (1975). Control of compartment development by the engrailed gene in drosophila. *Nature*. 11

Müller, J. and Bienz, M. (1992). Sharp anterior boundary of homeotic gene expression conferred by the fushi tarazu protein. *The EMBO journal*, 11(1010):3653–3661. 6, 7, 59

Müller, J. and Kassis, J. A. (2006). Polycomb response elements and targeting of polycomb group proteins in drosophila. *Current opinion in genetics & development*, 16(55):476–484. 60

Müller, M., Hagstrom, K., Gyurkovics, H., Pirrotta, V., and Schedl, P. (1999). The mcp element from the drosophila melanogaster bithorax complex mediates long-distance regulatory interactions. *Genetics*, 153(33):1333–1356. 45, 46, 48, 62

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russell, S., White, R. A. H., and et al. (2010). A comprehensive map of insulator elements for the drosophila genome. 6(11):e1000814. 33

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H. G., Wilm, M., and Müller, J. (2007). Pcl-prc2 is needed to generate high levels of h3-k27 trimethylation at polycomb target genes. *The EMBO journal*, 26(1818):4078–4088. 28

Ninov, N., Chiarelli, D. A., and Martín-Blanco, E. (2007). Extrinsic and intrinsic mechanisms directing epithelial cell sheet replacement during drosophila metamorphosis. *Development (Cambridge, England)*, 134(22):367–379. 16

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., and Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, 133(77):1277–1289. 7

O'Kane, C. J. and Gehring, W. J. (1987). Detection in situ of genomic regulatory elements in drosophila. *Proceedings of the National Academy of Sciences*, 84(2424):9123–9127. 18, 33

Pai, C.-Y., Kuo, T.-S., Jaw, T. J., Kurant, E., Chen, C.-T., Bessarab, D. A., Salzberg, A., and Sun, Y. H. (1998). The homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in drosophila. *Genes & development*, 12(33):435–446. 9

Papp, B. and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional on and off states by trxg and pcg proteins. *Genes & development*, 20(1515):2041–2054. 27

Paro, R. (1990). Imprinting a determined state into the chromatin of drosophila. *Trends in genetics : TIG*, 6(1212):416–421. 25

Payre, F. (2004). Genetic control of epidermis differentiation in drosophila. *The International journal of developmental biology*, 48(22):207–215. 12, 13

Pearson, J. C., Lemons, D., and McGinnis, W. (2005). Modulating hox gene functions during animal body patterning. *Nature reviews. Genetics*, 6(1212):893–904. 2

Peifer, M. and Bender, W. (1986). The anterobithorax and bithorax mutations of the bithorax complex. *The EMBO journal*, 5(99):2293–2303. 35, 60

Peifer, M., Karch, F., and Bender, W. (1987). The bithorax complex - control of segmental identity. *Genes & development*, 1(99):891–898. 20, 60

Pick, L. (2016). Hox genes, evo-devo, and the case of the ftz gene. *Chromosoma*, 125(33):535–551. 3

Pirrotta, V. (1997). Chromatin-silencing mechanisms in drosophila maintain patterns of gene expression. *Trends in genetics : TIG*, 13(88):314–318. 25

Pirrotta, V., Chan, C. S., McCabe, D., and Qian, S. (1995). Distinct parasegmental and imaginal enhancers and the establishment of the expression pattern of the ubx gene. *Genetics*, 141(44):1439–1450. 33

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). Crispr elements in yersinia pestis acquire new repeats by preferential uptake of bacteriophage dna, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, 151:653–663. 54

Qian, S., Capovilla, M., and Pirrotta, V. (1991). The bx region enhancer, a distant cis-control element of the drosophila ubx gene and its regulation by hunchback and other segmentation genes. *The EMBO journal*, 10(66):1415–1425. 59

Richard, S. and David, S. (1990). Functional dissection of ultrabithorax proteins in d. melanogaster. *Cell*, 60(44):597–610. 2

Rieckhof, G. E., Casares, F., Ryoo, H. D., Abu-Shaar, M., and Mann, R. S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*, 91(22):171–183. 9

Ringrose, L. and Paro, R. (2004). Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annual review of genetics*, 38:413–443. 60

Ringrose, L., Rehmsmeier, M., Dura, J.-M., and Paro, R. (2003). Genome-wide prediction of polycomb/trithorax response elements in drosophila melanogaster. *Developmental cell*, 5(55):759–771. 27

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. *The EMBO journal*, 12(22):435–442. 37

Ryoo, H. D., Marty, T., Casares, F., Affolter, M., and Mann, R. S. (1999). Regulation of hox target genes by a dna bound homothorax/hox/extradenticle complex. *Development (Cambridge, England)*, 126(2222):5137–5148. 9, 10

Salvaing, J., Mouchel-Vielh, E., Bloyer, S., Preiss, A., and Peronnet, F. (2008). Regulation of abd-b expression by cyclin g and corto in the abdominal epithelium of drosophila. *Hereditas*, 145(33):138–146. 16

Sánchez-Herrero, E., Vernós, I., Marco, R., and Morata, G. (1985). Genetic organization of drosophila bithorax complex. *Nature*, 313(59985998):108–113. 17

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The streptococcus thermophilus crispr/cas system provides immunity in escherichia coli. *Nucleic acids research*, 39(2121):9275–9282. 54

Schneuwly, S., Klemenz, R., and Gehring, W. J. (1987). Redesigning the body plan of drosophila by ectopic expression of the homoeotic gene antennapedia. *Nature*, 325(61076107):816–818. 2

Schwartz, Y. B. and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nature reviews. Genetics*, 8(11):9–22. 27

Schwartz, Y. B. and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Current Opinion in Cell Biology*, 20(33):266–273. 29

Schweinsberg, S., Hagstrom, K., Gohl, D., Schedl, P., Kumar, R. P., Mishra, R., and Karch, F. (2004). The enhancer-blocking activity of the fab-7 boundary from the drosophila bithorax complex requires gaga-factor-binding sites. *Genetics*, 168(33):1371–1384. 39

Scott, M. P. and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the antennapedia, ultrabithorax, and fushi tarazu loci of drosophila. *Proceedings of the National Academy of Sciences*, 81(1313):4115–4119. 3

Shashidhara, L. S., Agrawal, N., Bajpai, R., Bharathi, V., and Sinha, P. (1999). Negative regulation of dorsoventral signaling by the homeotic gene ultrabithorax during haltere development in drosophila. *Developmental biology*, 212(22):491–502. 13

Shimell, M. J., Peterson, A. J., Burr, J., Simon, J. A., and O'Connor, M. B. (2000). Functional analysis of repressor binding sites in the iab-2 regulatory region of the abdominal-a homeotic gene. *Developmental biology*, 218(11):38–52. 6, 7, 59

Sigrist, C. J. and Pirrotta, V. (1997). Chromatin insulator elements block the silencing of a target gene by the drosophila polycomb response element (pre) but allow trans interactions between pres on different chromosomes. *Genetics*, 147(1):209–221. 37

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during drosophila development. *Current Opinion in Cell Biology*, 7(33):376–385. 25

Simon, J., Peifer, M., Bender, W., and O'Connor, M. (1990). Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. *The EMBO journal*, 9(1212):3945–3956. 20, 23, 32, 59

Slifer, E. H. (1942). A mutant stock of drosophila with extra sex-combs. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 90(11):31–40. 27

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the drosophila embryo. *The EMBO journal*, 11(1111):4047–4057. 5

Smith, S. T., Wickramasinghe, P., Olson, A., Loukinov, D., Lin, L., Deng, J., Xiong, Y., Rux, J., Sachidanandam, R., Sun, H., and et al. (2009). Genome wide chip-chip analyses reveal important roles for ctcf in drosophila genome organization. *Developmental biology*, 328(22):518–528. 62

Spradling, A. C. and Rubin, G. M. (1982). Transposition of cloned p elements into drosophila germ line chromosomes. *Science (New York, N.Y.)*, 218(45704570):341–347. 52, 64

Struhl, G., Barbash, D. A., and Lawrence, P. A. (1997). Hedgehog organises the pattern and polarity of epidermal cells in the drosophila abdomen. *Development (Cambridge, England)*, 124(1111):2143–2154. 15

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J., and Cohen, S. M. (1992). Homeotic genes of the bithorax complex repress limb development in the abdomen of the drosophila embryo through the target gene distal-less. *Cell*, 71(33):437–450. 7

Vazquez, J., Müller, M., Pirrotta, V., and Sedat, J. W. (2006). The mcp element mediates stable long-range chromosome–chromosome interactions in drosophila. *Molecular Biology of the Cell*, 17(55):2158–2165. 45, 46, 62, 88

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., and et al. (2006). The polycomb group protein ezh2 directly controls dna methylation. *Nature*, 439(70787078):871–874. 28

Wang, C. and Lehmann, R. (1991). Nanos is the localized posterior determinant in drosophila. *Cell*, 66(44):637–647. 5

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., and Zhang, Y. (2004). Role of histone h2a ubiquitination in polycomb silencing. *Nature*, 431(70107010):873–878. 28

Wharton, R. P. and Struhl, G. (1991). Rna regulatory elements mediate control of drosophila body pattern by the posterior morphogen nanos. *Cell*, 67(55):955–967. 5

Wolle, D., Cléard, F., Aoki, T., Deshpande, G., Schedl, P., and Karch, F. (2015). Functional requirements for fab-7 boundary activity in the bithorax complex. *Molecular and cellular biology*, 35(2121):3739–3752. 39

BIBLIOGRAPHY

Zhao, K., Hart, C. M., and Laemmli, U. K. (1995). Visualization of chromosomal domains with boundary element-associated factor beaf-32. *Cell*, 81(66):879–889. 37

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U., and et al. (2006). Circular chromosome conformation capture (4c) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics*, 38(1111):1341–1347. 103

Zhou, J. and Levine, M. (1999). A novel cis-regulatory element, the pts, mediates an anti-insulator activity in the drosophila embryo. *Cell*, 99(66):567–575. 32

Zlotorynski, E. (2015). Chromosome biology: Ctcf-binding site orientation shapes the genome. *Nature Reviews Molecular Cell Biology*, 16(1010):578–579. 81

Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A. A., Lando, A. S., Eliseeva, I. A., Kulakovskiy, I. V., Maksimenko, O., and Georgiev, P. (2016). Architectural proteins pita, zw5,and zipic contain homodimerization domain and support specific long-range interactions in drosophila. *Nucleic acids research*, page gkw371. 73, 74

Mario Metzler

Strassburgerallee 97
4055 Basel
Switzerland

m.metzler@me.com
+41 79 731 73 64

Nationality: Swiss
Date of birth: 17.02.1989

Experience

2013/03 - 2017/03 **Ph.D. Candidate and Research Assistant in molecular genetics**
Prof. Dr. Markus Affolter, Biozentrum University of Basel
- Establishment of CRISPR/Cas9 in the laboratory
- Teaching and supervision of B.Sc. students

2015/10 - present **Advisor in University of Basel Graduate Academy Focus Group**
Vice rectorate for research, University of Basel
- Advising on concept for new graduate career service center

2015/09 **Visiting student Siemens and Novartis NIBR in Shanghai, China**
- Topic: how European companies do business in China

2015/08 - 2017/03 **Collaboration with Prof. Pavel Georgievs lab in Moscow, Russia**
- Technology transfer and instruction of collaborators

2015/10 - present **Advisor, Photographer, Illustrator for a magazine start-up**
Avenue — Magazin für Wissenskultur, www.avenue.jetzt

2014/12 - 2016/12 **Founder and Chair, Ph.D. Association at the Biozentrum of the University of Basel**
- Representing 120 doctoral candidates
- Leading a board of 15 people
- Managing relations with director of the institute and rectorate

2011/06 - 2011/09 **Intern, Columbia University, New York City**
Prof. Dr. Oliver Hobert, Neurogenetics
- Learning molecular biology techniques
- International environment

Education

2011/09 - 2012/12 **M.Sc. in Molecular Biology**, Biozentrum, University of Basel (5.7)
2008/10 - 2011/06 **B.Sc. in Molecular Biology**, University of Basel (4.8)
2004/09 - 2008/06 **Matura**, Liceo Cantonale Lugano 2, Savosa, Switzerland (5.4)

Awards

2013/03 - 2016/03 **Ph.D. Fellowships for Excellence** with support from the Werner von Siemens Foundation in Zug, Switzerland (awarded yearly to 10 out of around 450 applicants)
2015/06 **Best poster award** (out of 40) - Biozentrum Ph.D. retreat
2013/06 **Best talk award** (out of 15), Biozentrum Ph.D. retreat

Languages

Native **German** and **Italian**, highly proficient in spoken and written **English**, good working knowledge of spoken **French**

Computer Skills	**OS X, Windows and Linux**	Expert knowledge (server and desktop **administration**, **graphical user interfaces** and **command line**)
	Excel, Python and R	Management, statistics, and graphing of **big datasets**
	Web programming	Ruby on Rails, HTML, CSS, JavaScript, PHP, MySQL, PostgreSQL (one example: http://drosophila.me)
	Visual communication	**Adobe Illustrator**, **Microsoft PowerPoint**, **Apple Keynote**, and photography
Personal development	**Consulting**	
	2016/04	**McKinsey DIVE**, 3 days workshop in Amsterdam McKinsey & Company, Switzerland
	2016/01	**Strategic Management and Innovation Specialization** Manage 21st Century Business Strategy Copenhagen Business School on Coursera
	2015/08	**Two Speed IT: How Companies Can Surf the Digital Wave, a BCG Perspective** BCG and CentraleSupélec on Coursera
	People and Project management	
	2016/08	**Kommunikation Grundlagen** Bundesamt für Sport, Jugend+Sport Leiterweiterbildung
	2016/03	**Project Management Training Certificate** Nicolas Abbondanza, SPOL AG
	2015/10	**Conflict Management** Dr. Susanne Kress, PIW Training & Beratung GmbH
	Marketing	
	2016/04	**Introduction to Marketing** Wharton, University of Pennsylvania on Coursera
Further interests	2015/03	Solo Roadtrip from Chicago to New Orleans
	2014/03	Roadtrip in California and Nevada with coworkers
	2014/02 - 2014/11	**Solo art exhibition** at the Ramada Plaza Hotel in Basel
	Since 2007	Swiss Youth and Sport **alpine ski teacher** (d/f/i/e)
	Since 2007	Swiss Youth and Sport **orienteering teacher**

Photographer and fine art artist (http://mariometzler.ch)

No military duties left