DISSERTATION

SUBMITTED TO THE COMBINED FACULTIES FOR THE NATURAL
SCIENCES AND FOR MATHEMATICS OF THE
RUPERTO-CAROLA UNIVERSITY OF HEIDELBERG, GERMANY
FOR THE DEGREE OF DOCTOR OF NATURAL SCIENCES.

PRESENTED BY
DIPLOMA BIOLOGIST RAQUEL VIDAL MATOS
BORN IN OLIVEIRINHA, AVEIRO, PORTUGAL
ORAL EXAMINATION:

BIOCHEMICAL CHARACTERIZATION OF THE *DROSOPHILA* POLYCOMB PROTEIN DSFMBT

REFEREES:
DR. ANNE EPHRUSSI, EMBL HEIDELBERG
PROF. DR. IRMGARD SINNING, HEIDELBERG UNIVERSITY, BIOCHEMISTRY CENTER (BZH)

This PhD thesis work was carried out under the supervision of **Dr. Jürg Müller** at the **European Molecular Biology Laboratory** (EMBL) in Heidelberg, Germany from **2005 to 2009**.

Dedicated to my mother

Abre os olhos e encara a vida! A sina
Tem que cumprir-se! Alarga os horizontes!
Por sobre lamaçais alteia pontes
Com tuas mãos preciosas de menina.

Nessa Estrada da vida que fascina
Caminha sempre em frente, além dos montes!
Morde os frutos a rir! Bebe nas fonts!
Beija aqueles que a sorte te destina!

Trata por tu a mais longínqua estrela,
Escava com as mãos a própria cova
E depois, a sorrir, deita-te nela!

Que as mãos da terra façam, com amor,
Da graça do teu corpo, esguia e nova,
Surgir à luz a haste de uma flor!...

Florbela Espanca

ACKNOWLEDGEMENTS

Foremost, I would like to thank my supervisor, Dr. Jürg Müller for the opportunity to work in his group, the scientific discussions, his endless enthusiasm for science, and for his constant support, help and guidance throughout this PhD work.

I am extremely thankful to my Thesis Advisory Committee, Dr. Anne Ephrussi, Dr. Asifa Akhtar and Professor Dr. Irmgard Sinning for the valuable contributions to my PhD work and their constant help and support. I also wish to thank Dr. Lars Steinmetz and Professor Dr. Herbert Steinbeisser for being part of my Thesis Defence Committee.

I wish to thank Dr. Andreas Ladurner and Dr. Christoph Müller for the discussions and ideas that they gave me regarding my project. I wish to thank the current and past members of the Ladurner and Akhtar Lab for their comments and input during the Chromatin Group meetings. I would like to thank our collaborators Dr. Sven Fraterman and Dr. Matthias Wilm for the mass spectrometry analysis. I wish to acknowledge Dr. James Kadonaga, Dr. Steven Henikoff and Dr. Jeffrey Simon for reagents. I am grateful to Sandra Müller and Ann-Mari Voie, Annette Gerlinger and Brigitte Leeds, Vladimir Benes and the Genecore, Hüseyin Besir and Anne-Marie Lawrence for their valuable contributions and discussions during my PhD. I wish to thank Corinna Gorny for her kind understanding and support. I would also like to thank the Fly community and the Genome Biology Unit for creating the best work environment.

I wish to express my deep gratitude to the Dean of graduate studies, Dr. Helke Hillebrand as well as to Milanka Stojkovic and Tiziana Novarini for their constant help and support, and for making my life easier.

I warmly thank my previous supervisor and friend, Dr. Olga Amaral for the support that she gave me to accept the EMBL fellowship, for the strength in the less fortunate moments and for teaching me the fundaments of lab routine. I wish as well to express my gratitude to Prof. Dr. Maria João Prata and Prof. Dr. Maria Teresa Borges.

I am indebted to the past and present members of the Müller group who contributed a great deal to the invaluable scientific and non-scientific discussions, and for the pleasant and unforgettable time I spent in the lab. In particular, I would like to express my deep gratitude to Tanya Klymenko for sharing with me her knowledge and savvy, Andrès Gaytan for teaching me *Drosophila* genetics and for his sense of humour that helped so much in the dark moments, Max Nekrasov for teaching me the principles of biochemistry, for the invaluable contributions to my PhD work, and for being a great colleague and friend, Betti Papp for the support and encouragement and Kasia Oktaba

for passing me her scientific expertise in ChIP, for the valuable corrections on my PhD dissertation and for being a dear friend. I would like also to thank Johanna Scheuermann for the translation of the summary of the PhD dissertation to German and for the scientific discussions. I would like to hearty thank Archana Prusty for the scientific discussions and troubleshooting and more importantly, for her love and support, and for her cheerful and caring personality that made my life in the EMBL and ultimately in Heidelberg, much more colourful.

I wish to deeply thank Nadia Dubé and Anne-Marie Glynn for having the time to help me with the corrections of my PhD dissertation, for the advice, and for the nice moments spent together. My hearty thanks to my friends Claudia Chica, Corinne Kox, Andreia Feijão, Hélio Roque, Paulo Cunha, Guillaume Valentin, Aynur Kaya-Çopur, Nadia Dubé, Stella Lamprinaki, Evangelia Petsalaki, Jop Kind, Silvia Santos, Pedro Beltrão, Janina Karres, Alexandra Manaia, Luis Vacs, Carmelo Lopez and the Portuguese community for all the laughter, courage and support during my lows and highs, and for contributing greatly for all the great moments and memories that I will carry forever in my heart.

My heartfelt gratitude to my forever best friends Ana Cristina, Joana, Diana (as daibas) and Gustavo for being there for me unconditionally, for showing me support, love and understanding, for believing in me when I didn't, for listening and reasoning with me, for making me feel that even when I am not there, I always belong and for giving me always the deepest feeling of home wherever! I would like to express my deep and heartfelt gratitude to my mum, Anabela, my father José Carlos, my sisters, Sara, Margarida and Maria for always believing in me, for being my safest harbour, where I always find the greatest love and support and where I always fell like home no matter what. There are no words that can express my love for you! I deeply thank Anibal and Paula for the love and support. I wish to thank my grandparents, Elvira and António for their love, support, and understanding. This PhD work would have never been possible without the love and support of my family.

SUMMARY

Polycomb group (PcG) proteins are transcriptional regulators that maintain the repression of a large set of developmental control genes. PcG proteins form distinct multiprotein complexes: PhoRC (Pho Repressive Complex), PRC1 (Polycomb Repressive Complex 1), PRC2 (Polycomb Repressive Complex 2) and its variant Pcl-PRC2. These complexes repress target genes by modifying their chromatin. PcG chromatin modifications are thought to provide a memory that permits the transcriptional OFF state to be maintained in a heritable manner. PcG protein complexes assemble at specific *cis*-regulatory sequences called Polycomb Response Elements (PREs). Although all PcG complexes are targeted to PREs, only the PhoRC subunit Pho has sequence-specific DNA binding activity. The mechanism by which PRC1 or PRC2 are targeted and tethered at PREs is only poorly understood. During my Ph. D. studies, I performed a biochemical characterization of dSfmbt, the protein that together with Pho forms PhoRC. Using a Tandem Affinity Purification strategy, I purified proteins associated with dSfmbt. These purified complexes contained not only dSfmbt and Pho but also the histone deacetylase Rpd3, the histone chaperone NAP1, the chromatin binding protein HP1b, and an uncharacterized protein, CG3363. This is further supported by the observation that in addition to the Pho and dSfmbt also Rpd3 is bound to PREs of PcG target genes. dSfmbt forms a stable complex with the PRC1 subunit Scm *in vitro* and these two proteins are bound at PREs of PcG target genes. Using genetic interaction assays, I found that Scm and dSfmbt act in a highly synergistic manner to repress PcG target genes *in vivo* during *Drosophila* development. Taken together, these studies thus suggest that the PhoRC complex comprises not only Pho and dSfmbt but also additional chromatin-modifying and chromatin-binding subunits. The molecular and functional interactions between dSfmbt and Scm underscore the central role of dSfmbt as a molecular adaptor between the DNA-binding Pho subunit and PRC1.

ZUSAMMENFASSUNG

Die Proteine der Polycomb-Gruppe (Polycomb group, PcG) sind Transkriptionsregulatoren, die die Repression einer Vielzahl von Zielgenen aufrecht erhalten. Sie bilden separate Multiproteinkomplexe: PhoRC (Pho Repressive Complex), PRC1 (Polycomb Repressive Complex 1), PRC2 (Polycomb Repressive Complex 2) und dessen Variante Pcl-PRC2. Diese Komplexe reprimieren ihre Zielgene, indem sie deren Chromatin modifizieren. Von solchen PcG-Modifikationen wird angenommen, dass sie ein "Gedächtnis" für den inaktiven Transkriptionsstatus (OFF state) erbringen, der so auf vererbbare Weise aufrechterhalten wird. Die PcG-Komplexe assemblieren auf spezifischen *cis*-regulatorischen Sequenzen, den sogenannten PREs (Polycomb Response Elements). Obwohl alle PcG-Komplexe auf den PREs vorhanden sind, bindet nur die Pho-Untereinheit des PhoRC sequenz-spezifisch an DNA. Es ist bisher nicht ausreichend verstanden, durch welchen Mechanismus PRC1 und PRC2 die PREs erkennen und binden. Im Rahmen meiner Doktorarbeit habe ich das PcG-Protein dSfmbt, das zusammen mit der Pho-Untereinheit den PhoRC bildet, biochemisch charakterisiert. Mit Hilfe der Tandem-Affinitätsreinigungstechnik (Tandem Affinity Purification) habe ich Proteine identifiziert, die mit dSfmbt assoziieren. Die gereinigten Komplexe enthielten nicht nur dSfmbt und Pho, sondern auch die Histon-Deacetylase Rpd3, das Histon-Chaperon NAP1, das Chromatin-bindende Protein HP1b und ein uncharakterisiertes Protein, CG3363. Dieses Ergebnis wird weiter dadurch unterstützt, dass zusätzlich zu Pho und dSfmbt auch Rpd3 auf den PREs von PcG-Zielgenen gebunden ist. dSfmbt bildet *in vitro* einen stabilen Komplex mit der PRC1-Komponente Scm und diese beiden Proteine sind ebenfalls auf den PREs von PcG-Zielgenen gebunden. Ich habe mit Hilfe von genetischen Interaktionsexperimenten festgestellt, dass Scm und dSfmbt *in vivo* im Rahmen der *Drosophila*-Entwicklung ihre Zielgene mit stark ausgeprägter Synergie reprimieren. Zusammenfassend legen diese Untersuchungen also nahe, dass PhoRC nicht nur die Untereinheiten Pho und dSfmbt besitzt, sondern zusätzliche chromatin-bindende und -modifizierende Komponenten umfasst. Die molekularen und funktionellen Interaktionen zwischen dSfmbt und Scm unterstreichen die zentrale Rolle von dSfmbt als molekularem Adapter zwischen der DNA-bindenden Pho-Untereinheit und PRC1.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Metazoans develop from a single cell, the zygote, into a complex organism that is formed by different cell types. The zygote undergoes a complex developmental process called embryogenesis, whose major steps are mitosis, growth, migration of cells and differentiation into tissues. The differentiation process occurs in response to different stimuli that changes the expression state of genes within the cells. These differences of gene expression among cells will give rise to different tissues.

The pattern of expression of these genes leads to the formation of segments, where each has a particular developmental fate. The maintenance of this pattern of gene expression throughout many cell generations is crucial for the normal development of the organism.

"Cellular memory" is a process that allows the transmission of a specific transcriptional state from the mother cell to the daughter cells (*Hadorn, 1978*). Genetic and molecular studies have identified two distinct groups of proteins, which are the key factors involved in the "cellular memory" process. These two regulatory groups of proteins are the Polycomb group (PcG) and the trithorax group (trxG) proteins (Figures 1 and 2). These proteins are in charge of maintaining the transcriptional state of genes during the development of metazoans (reviewed in *Francis & Kingston, 2001*; *Ringrose & Paro, 2004*; *Müller & Verrijzer, 2009*).

1. BODY PATTERNING DURING *DROSOPHILA* DEVELOPMENT

In *Drosophila*, the formation of body segments is determined at the blastoderm stage of development (*Chan & Gehring, 1971*). Each of these body segments has a particular pattern of gene expression that will give rise to specific appendages in the adult animal. The identity of these segments differes among themselves and is stable throughout development.

In fertilized *Drosophila*'s eggs, there are maternally deposited factors that give rise to anterior-posterior and dorso-ventral polarity in the embryo. These maternal factors will activate zygotic segmentation genes accordingly to the established polarity of the embryo. The pattern of expression of these segmentation genes through the embryo axis allows the formation of segments. Each segment has a particular pattern of gene expression and gives rise to specific appendages and structures located at specific

anatomical positions of the adult (reviewed in *St Johnston & Nüsslein-Volhard, 1992*; *Pankratz & Jäckle, 1993*). The segmentation genes are composed of two different sets of genes: the gap genes and the pair-rule genes. Both are transcriptional regulators; the gap proteins establish the repressive transcriptional state of genes, and the pair-rule proteins establish the active transcriptional state of genes. These segmentation genes regulate the expression patterns of many genes, such as those from the Antennapedia complex (ANT-C; *Kaufman et al, 1990*) and the Bithorax complex (BX-C; *Duncan, 1989*; *Lewis, 1978*). These two gene clusters determine and maintain the identity of the *Drosophila* segments. ANT-C and BX-C are homeotic (HOX) genes that encode proteins that characteristically have a conserved DNA-binding domain (homeodomain). This domain binds to *cis*-regulatory elements in target genes, regulates their expression by activating or repressing them and gives rise to the different pattern of expression of the segments. HOX genes start being expressed in the blastoderm state of embryogenesis and their expression is maintained throughout the entire course of development. As HOX proteins determine segment identity, segments where these proteins have been removed, will acquire the identity of another segment and the anatomy as well as the survival of the animal will be severely compromised.

Gap and pair-rule regulators are present transiently during the first few hours of embryogenesis, but their initially established expression pattern of HOX genes, has to be maintained throughout the course of the development of the organism. Following the blastoderm stage, the maintenance of the transcriptional state of the HOX genes is under the control of two different transcriptional regulators, the PcG and the trxG proteins. Specifically, PcG proteins maintain the repressed state of HOX genes and trxG proteins maintain the active state of these genes (reviewed in *Francis & Kingston, 2001*; *Ringrose & Paro, 2004*; *Müller & Verrijzer, 2009*). In contrast to gap and pair-rule proteins, the PcG and trxG proteins are present ubiquitously throughout the organism and are needed to maintain the appropriate state of transcription of HOX genes and other cell specification genes (reviewed in *Ringrose and Paro, 2004*; *Oktaba et al 2008*).

PcG and trxG proteins are the regulators that confer “cellular memory” by maintaining the OFF and ON transcription state of genes respectively, which will lead to the formation of the anatomy of the organism. Three distinct and structurally independent PcG protein complexes have been biochemically purified (Figure 1): the Polycomb Repressive Complex 1 (PRC1), the Polycomb Repressive Complex 2 (PRC2), and the Pho Repressive Complex (PhoRC). Other Polycomb proteins have not yet been associated with these complexes or they might belong to still unidentified complexes (Figure 1 and table 1).

In *Drosophila*, the trxG proteins have also been described to be organized into different multimeric complexes with a high molecular weight. Although little is known about these complexes, four have been identified: the Trithorax Acetylation Complex 1 (TAC1), the Brahma (BRM) complex, the Absent, small or homeotic discs 1 (Ash1) complex, and the Ash2 complex (Figure 2). As there are trxG proteins that do not belong to any of these three complexes, it is probable that unknown trxG complexes are awaiting discovery (table 2).

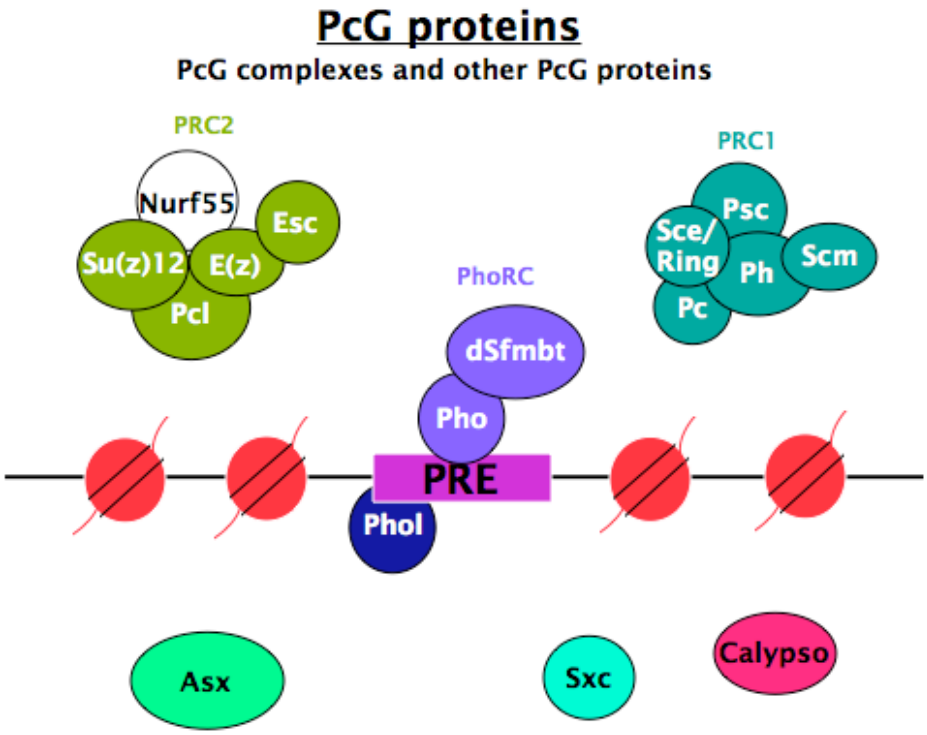

Figure 1 – PcG proteins and complexes in *Drosophila*.
PRC2 is composed of the proteins Extra sex combs (Esc), Enhancer of zeste (E(z)), Polycomb-like (Pcl), Supressor of zeste 12 (Su(z)12), and Nurf55. PRC1 is composed of the proteins Posterior sex combs (Psc), Sex combs extra (Sce/Ring), Polyhomeotic (Ph), Polycomb (Pc), and Sex comb on midleg (Scm). PhoRC is composed of the proteins Pleiohomeotic (Pho) and Drosophila Scm-related gene containing four malignant brain tumor domains (dSfmbt). Other PcG proteins are: Pleiohomeotic-like (Phol), Additional sex combs (Asx), Super Sex combs (Sxc) and Calypso (see text for details). Only Pho and Phol have sequence-specific DNA binding activity, thus binding to PREs.

2. PcG PROTEINS AND COMPLEXES IN *DROSOPHILA*

The PcG genes were first identified in *Drosophila* by the observation of mutant phenotypes that suggested that they were responsible for the repression of several HOX genes in the ANT-C and in the BX-C. The first PcG genes identified were: *extra sex combs* (*esc*) and *Polycomb* (*Pc*), whose absence in the mutant fly lead to the formation of ectopical sex combs on the second and third legs of *Drosophila* males. The sex combs structure is restricted to the first pair of male legs in wild-type *Drosophila*. Therefore, the mutant phenotype is caused by the misexpression of the HOX gene *Sex combs reduced* (*Scr*) in tissues where its expression should be repressed. In these PcG mutants, there is a misexpression of *Scr* in the second and third leg imaginal discs, allowing ectopic expression of this HOX gene and giving rise to sex combs structures in these appendages (*Kennison, 1995*; *2004*). Since the discovery of these two PcG genes in *Drosophila*, many more genes that cause homeotic transformations due to HOX genes misexpression have been cloned (Table 1).

Table 1 – Polycomb group gene families and conserved domains.

Protein name (*Drosophila*)	Gene	*H. sapiens* Homologues	Conserved domains	Cloning references
Polycomb	*Pc*	HPC1, HPC2, HPC3	Chromodomain	*Paro & Hogness, 1991*
Polyhomeotic	*Ph*	HPH1, HPH2	SAM domain	*DeCamilis et al, 1992*
Posterior sex combs	*Psc*	BMI1/ MEL18	RING-finger	*Brunk et al, 1991*
Supressor of zeste 2	*Su(z)2*		RING-finger	*Brunk et al, 1991*
dRING/ Sex combs extra	*Sce*	RING1	RING-finger	*Fritsch et al, 2003*
Enhancer of zeste	*E(z)*	EZH1, EZH2	SET domain SANT domain	*Jones et al, 1990*
Extra sex combs	*Esc*	EED	WD40 repeats	*Frei et al, 1985*
Extra sex combs like	*Escl*		WD40 repeats	*Wang et al, 1985*
Supressor of zeste 12	*Su(z)12*	SU(Z)12	VEFS-box Zn-finger	*Birve et al, 2001*
Polycomb-like	*Pcl*	PHF1/ hPcl1	Tudor domain PHD-finger	*Lonie et al, 1994*
Pleiohomeotic	*pho*	YY1	Zn-finger	*Brown et al, 1998*
Pho-like	*phol*		Zn-finger	*Brown et al, 2003*
Sex combs on midleg	*Scm*	SCML1, SCML2	SAM domain Zn-finger MBT repeats	*Bornemann et al, 1996*
Additional sex combs	*Asx*	ASX1, ASX2, ASX3	PHD-finger ASXH domain	*Sinclair et al, 1998*
dSfmbt	*dSfmbt*	Sfmbt	SAM domain	*Klymenko et al, 2006*

			Zn-finger MBT repeats	
Calypso	*calypso*			*Gaytan de Ayala Alonso et al, 2007*
Super sex combs	*sxc*	Ogt	TPR repeats Glycosyl transferase domain	*Gambetta et al, 2009*

- POLYCOMB REPRESSIVE COMPLEX 2 (PRC2)

PRC2 is a 600 KDa PcG complex that contains the PcG proteins Enhancer of zeste (E(z)), Extra sex combs (Esc), and Supressor of zeste 12 (Su(z)12), as well as the histone-binding protein Nurf55 (*Czermin et al, 2002*; *Müller et al, 2002*). From the core four members of the PRC2 complex, E(z) is a PcG protein that has a SET domain in its structural architecture. This domain has been found to be responsible for transferring methyl residues into lysines present in histone tails (*Rea et al, 2000*). The Histone Methyl transferase (HMTase) activity of E(z) is highly specific for the trimethylation of the lysine residue 27 in the histone H3 (H3K27me3) and it has been shown to be critical for HOX gene silencing *in vivo* (*Müller et al, 2002*). Although E(z) is the HMTase, when isolated its activity is 1000-fold lower than when present in the tetrameric PRC2 complex (*Ketel et al, 2002*). Studies performed to dissect the importance of each component of the PRC2 showed that although E(z) is the enzymatic unit of the complex, its activity is increased when forming a complex with Esc protein (*Müller et al, 2002*; *Nekrasov et al, 2005*).

Su(z)12 and Nurf55 proteins are also crucial units for the activity of the PRC2 complex as they constitute the minimal PRC2 subcomplex that binds to mononucleosomes (*Nekrasov et al, 2005*; *Ketel et al, 2005*). Esc protein also contributes to the high affinity binding of the complex to the mononucleosomes (*Nekrasov et al, 2005*).

A variant of PRC2 has been recently purified, the PRC2-Pcl complex (*Nekrasov et al, 2007*). PRC2-Pcl has the same characteristics as PRC2, but also contains the PcG protein Polycomb-like (Pcl) as a member of the complex. PRC2-Pcl is thought to represent only a subset of the total PRC2 present in cells (*Nekrasov et al, 2007*). The presence of Pcl is thought to be crucial for the high levels of H3K27me3 that are found in PcG target genes (*Nekrasov et al, 2007*). Pcl is also thought to be involved in the correct targeting of PRC2 to PcG target genes, but little is known about the mechanism of how this protein generates high levels of H3K27me3 (*Nekrasov et al, 2007*). The mammalian homologue of PRC2 also displays the same HMTase activity towards H3K27me3 (*Cao et al, 2002*; *Kuzmichev et al, 2002*).

- POLYCOMB REPRESSIVE COMPLEX 1 (PRC1)

PRC1 is a PcG multiprotein complex that is composed of five proteins: Polyhomeotic (Ph), Posterior sex combs (Psc), Polycomb (Pc), Sex comb on midleg (Scm), and Sex combs extra (Sce/Ring) (*Franke et al, 1992*; *Shao et al, 1999*; *Saurin et al, 2001*). In addition to these PcG proteins, several TBP (TATA-binding-protein)-associated factors, part of the general transcription factor TFIID (dTAFII proteins), and the sequence-specific DNA-binding factor Zeste were also found as components of PRC1 (*Saurin et al, 2001*). The homologous PRC1 complex in mammals (hPRC-H) is composed of the mammalian homologs of the five PcG proteins found in the *Drosophila* PRC1 (*Levine et al, 2002*).

PRC1 interacts with chromatin *in vitro* by blocking remodelling by the SWI2/SNF2 complex, and this complex also induces compaction of chromatin in nucleosomal arrays (*Francis et al, 2001*; *2004*). Another characteristic of PRC1 is its binding to methylated lysines in histone tails. In particular the chromodomain of the Pc protein binds specifically to H3K27me3 (*Cao et al, 2002*; *Czermin et al, 2002*; *Fischle et al, 2003*; *Min et al, 2003*).

The component of PRC1 Sce/Ring and its human homologue Ring 1B are catalytic subunits that specifically ubiquitinate lysine 119 of histone H2A (H2AK119ubi) (*Wang et al, 2004*). Although PRC1-mediated ubiquitylation is a post-translational modification of histone tails related to transcriptional repression, the function and mechanism of this modification remains unclear and further studies are needed to unravel its role.

The co-purification of the TATA-box-binding protein (TBP)–associated factors (TAFIIs) as stoichiometric components of PRC1 (*Saurin et al, 2001*) suggests that this complex might use a combination of mechanisms to repress gene expression and not only by chromatin remodelling. For instance, PRC1 may have a role in blocking transcription as it interacts with components of the RNA polymerase II machinery (*Dellino et al, 2004*).

- PHO REPRESSIVE COMPLEX (PHORC)

PhoRC is a PcG protein complex composed of two proteins: Pleiohomeotic (Pho) and "*Drosophila* Scm-related gene containing four Malignant Brain Tumor (MBT) domains" (dSfmbt) (*Klymenko et al, 2006*). Pho, and the functionally redundant Pho-like are the only DNA-binding PcG proteins (*Brown et al, 1998*; *2003*). dSfmbt is the other PcG component of PhoRC and has a particular domain structure with a Zinc (Zn)-finger in the N-terminal part of the protein, four MBT repeats in the central portion of the protein,

and an α-sterile motif (SAM) in the C-terminal (*Klymenko et al, 2006*; figure 11). Its Zn-finger is a Zn-finger type FCS, which has been characterized to bind to RNA in a non-specific manner (*Zhang et al, 2004*). This type of Zn-finger is also present in the PcG proteins Ph and Scm, and also in the transcription factor Lethal (3) malignant brain tumor (L(3)mbt; *Zhang et al, 2004*).

The SAM domain, present in the C-terminal portion of the dSfmbt protein, has been reported to be involved in protein-protein interaction by forming homo-oligomers and hetero-oligomers (*Thanos et al, 1999*; *Kim et al 2002*). This domain is also present in two other PcG proteins, Ph and Scm (Figure 11), and is responsible for the interaction between these two proteins (*Peterson et al, 1997*; *2004*). L(3)mbt also has a SAM domain in the C-terminal portion of the protein.

The MBT repeats of dSfmbt show specific binding to the mono and dimethylated lysine residues, 9 and 27 in histone H3, and 20 in histone H4 (*Klymenko et al, 2006*; *Grimm et al, 2009*). Scm (Figure 11) and L(3)mbt also have MBT repeats. The MBT repeats of Scm bind unspecifically to mono and dimethylated lysine residues in histones H3 and H4 (*Grimm et al, 2007*). Unspecific lysine binding activity has also been reported for the human homologue of L(3)mbt, L3MBTL1 (*Li et al, 2007*; *Min et al, 2007*).

PhoRC is specifically targeted to PREs in a Pho-dependent manner, suggesting that the methyl binding function of dSfmbt is not required for the targeting of the complex to these regions (*Klymenko et al, 2006*). Recent studies have shown that this complex is not only targeted to HOX genes (*Oktaba et al, 2008*). It is in fact also targeted to other genes required for the patterning along the anterioposterior, dorsoventral, and proximodistal axis in imaginal discs, as well as genes involved in cell cycle regulation (*Oktaba et al, 2008*). The involvement of PhoRC in repressing genes whose expression is cell cycle dependent suggests that the mechanism of repression by PcG proteins is more dynamic than previously thought and is not only involved in body patterning, but also in cell growth and proliferation (*Oktaba et al, 2008*).

- OTHER PCG PROTEINS

Apart from the different PcG proteins interacting to form three PcG complexes, there are other PcG proteins that have not yet been associated with any complexes, and whose function is crucial for HOX genes repression. These proteins are Pleiohomeotic-like (Phol; *Brown et al, 2003*), Super sex combs (Sxc; *Ingham, 1984*), Additional sex combs (Asx; *Jürgens, 1985*; *Sinclair et al, 1998*) and Calypso (*Gaytan de Ayala Alonso et al, 2007*). The weak phenotype observed in *pho* mutants, led to the supposition that

there was another protein that acted redundantly to Pho, leading to the identification of Phol (*Brown et al, 2003*). Pho and Phol both repress HOX genes and bind to the same DNA motif (*Brown et al, 2003*). Recent studies showed that *sxc* gene encodes for the Ogt protein, a glycosyltransferase involved in the addition of N-acetylglucosamine (GlcNAc) to proteins (*Gambetta et al, 2009*). *Calypso* was identified in a genetic screen designed to identify new PcG genes (*Gaytan de Ayala Alonso et al, 2007*). Further biochemical studies are required to structurally and functionally characterize these proteins.

3. TRXG PROTEINS AND COMPLEXES OVERVIEW IN *DROSOPHILA*

The trxG genes were discovered in genetic screens that searched for mutations that suppress the PcG phenotype of PcG mutants (*Ingham, 1981*; *Kennison & Russel, 1987*; *Kennison & Tamkun, 1988*; *Shearn, 1989*). The first genes of the trxG to be identified were *trx*, *ash1* and *ash2* (reviewed in *Kennison, 1993*). Other trxG genes are shown in table 2. The maintenance of the transcriptional ON state of HOX genes is expected to involve many regulatory steps such as transcriptional activation, post-translational modification of HOX proteins or expression of required HOX protein cofactors. This suggests that the trxG genes are far more heterogeneous and complex both in function and structure than the PcG genes.

Table 2 – Selected Trithorax Group (trxG) gene families and conserved domains.

Protein name (*Drosophila*)	Gene	*H. sapiens* Homologues	Conserved domains	Cloning references
Trithorax	*trx*	MLL (ALL-1/ HTRX), MLL2, MLL3, MLL5	PHD finger (s) Bromodomain SET domain	*Mazo et al, 1990*
Absent, small or homeotic discs 1	*ash1*	hASH1	PHD finger SET domain	*Tripoulas et al, 1994*
Absent, small or homeotic discs 2	*ash2*	hASH2	PHD finger SPRY domain	*Adamson & Shearn, 1996*
Brahma	*brm*	hBRM	Bromodomain BRK domain ATPase domain	*Tamkun et al, 1992*
Osa	*osa*	OSA1/ hELD1	ARID DNA binding domain EHD1, EHD2 domains	*Vazquez et al, 1999*
Moira	*mor*	BAF155	Chromodomain BRCT and SWIRM protein binding domain SANT domain	*Crosby et al, 2000*
Little imaginal	*lid*	RBP2	ARID DNA binding	*Gildea et al, 2000*

discs 1			domain PHD fingers Leucine zipper motif	
Zeste	*z*		Leucine zipper motif	*Biggin et al, 1988*
GAF (GAGA factor)	*trl*		BTB/ POZ domain Zn-finger	*Farkas et al, 1994*
Kismet	*kis*	CHD7	BRK domain SNF2-related ATPase domain Chromodomain RmlC-like cupin DEAD-DEAH box helicase, N-terminal	*Daubresse et al, 1999*
Tonalli	*tna*	hZimp10	SP-RING Zn-finger	*Gutierrez et al, 2003*
Kohtalo	*kto*	MED12L		*Treisman, 2001*
Skuld	*skd*	THRAP2	TRAP domain	*Treisman, 2001*

The trxG regulators can be divided in two different complexes (Figure 2) that have been purified from *Drosophila*: the TAC1 (*Petruk et al, 2001*) and the BRM complex (*Papoulas et al, 1998*; *Crosby et al, 1999*; reviewed in *Simon & Tamkun, 2002*). Both the Ash1 and Ash2 proteins are not associated with either of these two complexes but there is evidence for their association with yet unknown two distinct protein complexes (*Papoulas et al, 1998*).

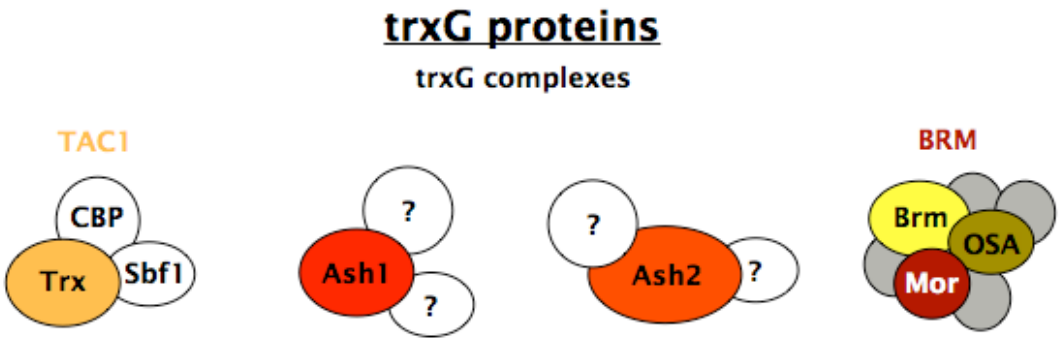

Figure 2 – TrxG proteins and complexes in *Drosophila*.
TAC1 is composed of the Trithorax (Trx) protein, CREB-binding protein (CBP), and SET domain binding protein (Sbf1). BRM is composed of the proteins Brahma (Brm), Moira (Mor), OSA and four additional proteins. The trxG proteins Ash1 and Ash2 belong to two additional protein complexes yet uncharacterized (see text for details).

- TAC1 AND TRX PROTEIN

TAC1 is estimated to weight 1MDa and is composed of the subunits Trithorax (Trx), dCBP (CREB-binding protein), which has histone acetyltransferase activity and the anti-phosphatase Sbf1 (SET domain binding factor; *Petruk et al, 2001*; figure 2). This complex has acetyltransferase activity confered by the dCBP protein and also has HMTase activity conferred by the SET domain of the Trx protein (*Rea et al, 2000*; *Smith et al, 2004*).

Trx is a highly conserved protein with homologues in yeast (Set1) and in humans (MLL). In yeast, Set1 is a member of the protein complex COMPASS (COMplex Proteins ASsociated with Set1; *Miller et al, 2001*) and MLL is part of the MLL complex in humans (human COMPASS; *Yokoyama et al, 2004*). The MLL complex also contains the trxG protein ASH2L, which is the human homologue of the *Drosophila* Ash2 (*Yokoyama et al, 2004*; table 2).

Trx, Set1, and MLL have a SET domain that it is responsible for the methylation of lysine 4 in histone H3 (H3K4me) (*Santos-Rosa et al, 2002*; *Milne et al, 2002*; *Nakamura et al, 2002*; *Smith et al, 2004*). It has also been shown that MLL directly binds to the promoter sequences of a target HOX gene and by this mean, regulates its expression (*Milne et al, 2002*; *Nakamura et al, 2002*). A correlation between the MLL-SET H3K4 methylation and the activation of a target HOX gene has also been reported (*Milne et al, 2002*; *Nakamura et al, 2002*).

- BRM

The BRM complex has a molecular weight of 2MDa and is highly related to the SWI/SNF chromatin-remodelling complex found in *Saccharomyces cerevisiae*. The subunits of this complex are the trxG proteins brahma (Brm), Moira (Mor), and Osa and four additional accessory proteins (*Papoulas et al, 1998*; *Crosby et al, 1999*; reviewed in *Simon & Tamkun, 2002*; figure 2). Brm has a DNA-dependent ATPase domain that is typical for the SWI2/SNF2 protein family that is involved in chromatin remodelling (*Papoulas et al, 1998*; *Collins et al, 1999*; *Kal et al, 2000*). Brm was found to be associated with transcriptionally active chromatin on polytene chromosomes, suggesting that this protein plays a general role in transcription at many loci in addition to at the HOX genes (*Armstrong et al, 2002*).

- ASH1 AND ASH2

Ash1 and Ash2 are two other trxG proteins that belong to as of yet uncharacterized complexes (*Papoulas et al, 1998*). Ash2 is a protein that is involved in the wing development of *Drosophila* (*Adamson & Shearn, 1996*; *Amorós et al, 2002*; *Angulo et al, 2004*). Ash2 has been described to be in a 500 KDa complex that is structurally independent of the Ash1 and the BRM complex (*Papoulas et al, 1998*).

Gel filtration chromatography of embryonic extracts suggested that Ash1 is present in a 2 MDa complex, but its composition is yet to be discovered (*Papoulas et al, 1998*). Ash1 is a SET domain protein and hence has HMTase activity, but the specificity of this enzymatic activity has been controversial. Isolated Ash1 SET domain with its preSET and postSET domains has been reported to methylate lysines 4 and 9 in histone H3 (H3K4me & H3K9me), and lysine 20 in histone H4 (H4K20me; *Beisel et al, 2000*). A different study showed that Ash1 was required only for H3K4 methylation *in vivo* and not for H3K9 or H4K20 methylation (*Byrd & Shearn, 2003*). On the other hand, a recent study showed that recombinant *Drosophila* Ash1 protein and its human homologue ASH1L specifically methylate lysine 36 in histone H3 (H3K36me; *Tanaka et al, 2007*).

The substrate for the SET proteins Ash1 and Trx are lysine residues in histones that correlate with transcriptionally active chromatin (i.e. H3K4 and H3K36). Thus, trxG maintain the active transcriptional state of its target genes (*Yokoyama et al, 2004*; *Steward et al, 2006*; *Tanaka et al, 2007*).

4. POLYCOMB RESPONSE ELEMENT (PRE)

PcG and trxG proteins bind at *cis*-regulatory elements of its target genes, maintaining their transcriptional status throughout many cell generations. The PREs were initially identified as regulatory sequences that prevented inappropriate activation of HOX reporter genes in a PcG-dependent manner in *Drosophila* embryos and larvae (*Müller & Bienz, 1991*; *Simon et al, 1993*; *Chan et al, 1994*; *Christen & Bienz, 1994*). Further functional assays such as transgenic analysis and chromatin immunoprecipitation identified PREs and their regulated genes at five loci, namely at some HOX genes of the BX-C and the ANT-C clusters and at the *engrailed*, *polyhomeotic*, and *hedgehog* genes (*Mihaly et al, 1998*; *Bloyer et al, 2003*; *Brown et al, 1998*; *Maurange & Paro, 2002*; *Orlando & Paro, 1993*; *Strutt & Paro, 1997*; *Fritsch et al, 1999*). The silencing function of PREs is continuously required as its excision from a

silenced HOX reporter gene results in loss of repression independently of the developmental time at which it occurs (*Busturia et al, 1997*; *Sengupta et al, 2004*).

The function of the PREs is not only restricted to the HOX genes. They also work as general transcription silencers with the capacity to silence reporter genes containing enhancer and promoter sequences from genes that are normally not under the control of PcG proteins (*Sengupta et al, 2004*). The PREs are located in close proximity to the promoter of their target genes (*Schwartz et al, 2006*; *Oktaba et al, 2008*).

Pho and Phol DNA binding sites are the main signature of PREs (*Oktaba et al, 2008*). Mutational studies on these binding sites showed that their binding to these sites is essential for silencing of the target genes (*Brown et al, 1998*; *Mihaly et al, 1998*; *Fritsch et al, 1999*; *Shimell et al, 2000*; *Busturia et al, 2001*; *Mishra et al, 2001*; *Brown et al, 2003*; *Ringrose et al, 2003*).

Moreover, a number of proteins have been reported to bind to the PREs and this binding has been proposed to be involved in the recruitment of the PcG proteins to these sites. These proteins include Trithorax-like (Trl; also known as GAGA factor [GAF]; *Horard et al, 2000*; *Busturia et al, 2001*; *Hodgson et al, 2001*; *Mahmoudi et al, 2003*; *Schwendemann & Lehmann, 2002*), Pipsqueak (Psq; *Hodgson et al, 2001*; *Huang et al, 2002*; *Schwendemann & Lehmann, 2002*), Zeste (*Mulholland et al, 2003*; *Mahmoudi et al, 2003*; *Ringrose & Paro, 2004*), Grainyhead (Grh; also known as neuronal transcription factor 1 [NTF1]; *Blastyak et al, 2006*), Dispersed (Dsp1; *Dejardin et al, 2005*) and Sp1/KLF family members (*Brown et al, 2005*).

- PREs AS TARGETING PLATFORMS FOR PcG PROTEIN COMPLEXES

Chromatin immunoprecipitation studies have shown that the three PcG complexes PRC1, PRC2, and PhoRC are bound at the PREs of HOX genes and other target genes (*Strutt et al, 1997*; *Orlando et al, 1997*; *Klymenko et al, 2006*; *Papp & Müller, 2006*; *Oktaba et al, 2008*). However, the mechanism through which the PcG complexes assemble at the PREs is unknown.

Pho, one of the proteins of the PhoRC, is the only PcG protein that binds directly to the PREs (*Klymenko et al, 2006*; *Oktaba et al, 2008*), but the targeting of PRC1 and PRC2 to the PREs remains elusive. E(z), as a PRC2 subunit, methylates H3K27 and this modification is recognized by the chromo domain of the PRC1 subunit Pc (*Cao et al, 2002*; *Czermin et al, 2002*; *Kuzmichev et al, 2002*). Also, the function of E(z) is required for the recruitment of PRC1 components to target genes (*Rastelli et al, 1993*; *Czermin et al, 2002*; *Wang et al, 2004*).

Altogether, this evidence has led to a model where PRC2 is upstream of PRC1 in PcG-mediated silencing, but how PRC2 is targeted to the PREs is still unknown. It has been proposed that the recruitment of PRC2 to the PREs is by direct interactions between the PRC2 components E(z) and Esc and the PhoRC subunit Pho (*Wang et al, 2004*). According to this model, PRE-tethered PRC2 would deposit H3K27me3 histone marks in nucleosomes at the PREs and thereby creating binding sites for the chromo domain of Pc that would result in the targeting of PRC1 to H3K27me3 PREs. However, recent studies do not support this model (reviewed by *Müller & Kassis, 2006*). First the Pho protein can interact directly with the PRC1 components Ph and Pc (*Mohd-Sarip et al, 2002*). Also, Pho and PRC1 can co-assemble on naked PRE DNA templates in the absence of nucleosomes (*Mohd-Sarip et al, 2005*). These reconstituted Pho-PRC1 protein-DNA assemblies showed protein-DNA contacts across an extended stretch of PRE DNA and they appear to adopt a conformation that is difficult to conciliate with the formation of nucleosome core particles (*Mohd-Sarip et al, 2005*; *2006*). Secondly, tandem affinity purification (TAP) studies using Pho as a bait protein did not reveal any component of the PRC1 or PRC2 complexes (*Klymenko et al, 2006*). Neither the TAP purification using the PRC2 components E(z) or Pcl as bait proteins showed any members of PhoRC or PRC2 (*Nekrasov et al, 2007*). Moreover, the purifications of PRC1 complexes have not showed any members of either the PRC2 or PhoRC complexes. Lastly, studies using quantitative ChIP analysis suggested that the PREs present at the *Ubx* gene are depleted of nucleosomes (*Papp & Müller, 2006*; *Mohd-Sarip et al, 2006*). In these studies, it has also been suggested that PREs are a platform of assembly for PhoRC, PRC1, and PRC2, as subunits of these complexes were present at the PREs. Thus, current evidence suggests that not only PhoRC but also PRC1 and PRC2 are targeted to the PREs through interactions with Pho, dSfmbt, or other as of yet unknown protein and that this targeting is not mediated by histone modifications nor by nucleosomal interactions (Figure 3).

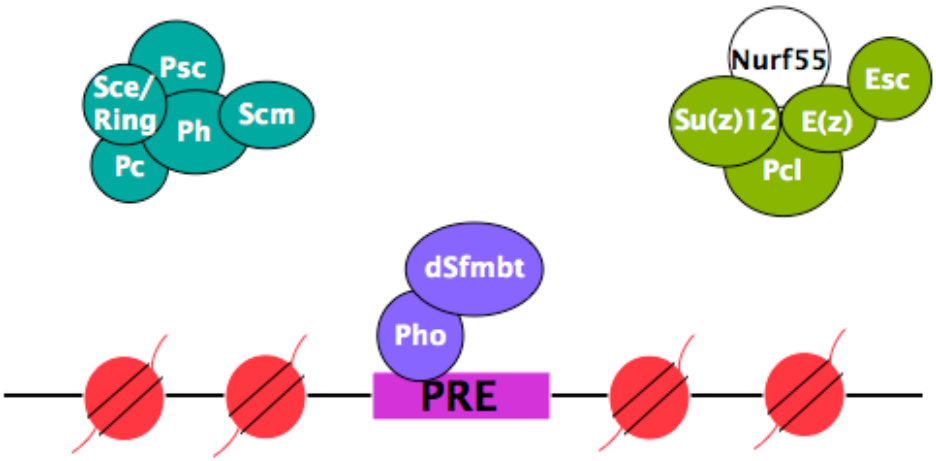

Figure 3 – PcG complexes PRC1, PRC2 and PhoRC at the PREs. Pho, a member of PhoRC, binds to a specific DNA motif called PREs, while how PRC1 and PRC2 are recruited to PREs remains unknown.

AIM

The main goal of my thesis work is to determine whether dSfmbt is the physical link between PhoRC and PRC1 or PRC2 by interacting directly with one of its subunits, or by forming a different complex that could mediate the binding between PhoRC and PRC1 or PRC2. In order to understand how PRC1 and PRC2 are targeted to the PREs, two different approaches will be taken: 1) an *in vivo* biochemical approach using dSfmbt as a bait protein in a Tandem Affinity Purification (TAP) and 2) *in vitro* and *in vivo* interaction studies between the PhoRC component dSfmbt and the PRC1 subunits Scm and Ph. An additional goal of my work is to elucidate the *in vitro* specific function of the trxG protein Ash1.

MATERIALS AND METHODS

1. EMBRYONIC NUCLEAR EXTRACT

Embryonic nuclear extracts were performed as reported in *Klymenko et al, 2006*. Preparation of embryonic nuclear extract from *Drosophila* was done on ice or at 4°C. Dechorionated embryos were taken up in buffer NU1 (15 mM HEPES pH 7.6; 10 mM KCl; 5 mM $MgCl_2$; 0.1 mM EDTA pH 7.9; 0.5 mM EGTA pH 7.9; 350 mM Sucrose; 2 mM DTT; 0.2 mM PMSF), at a ratio of 1 ml buffer NU1 per 1 gram of embryos and were homogenized with a glass dounce homogenizer. Lysate containing nuclei from 30 to 40 grams of embryos was filtered by gravity through a double layer of miracloth mounted on a funnel. After the lysate was mostly drained through the miracloth, the cloth was rinsed with 2-3 volumes of buffer NU1. Nuclei were pelleted by centrifugation in a pre-cooled Superlite GSA rotor at 9000 rpm for 15 minutes. After wiping off the lipid layer (upper layer) and discarding the supernatant (intermediate liquid portion), the nuclei pellet was resuspended in 0.5 ml of low-salt buffer per gram of embryos (Low-salt buffer: 15 mM HEPES; 20% glycerol; 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA pH 7.9; 1 mM DTT; "Complete EDTA-free" protease inhibitor cocktail – Roche). Care was taken so that yolk was not resuspended with the nuclei. Resuspended nuclei were transferred into a 50 ml falcon tube and lysed by addition of 0.5 ml high-salt buffer per gram of embryos (high-salt buffer: 15 mM HEPES; 20% glycerol; 1.5 mM $MgCl_2$; 800 mM KCl; 0.2 mM EDTA pH 7.9; 1 mM DTT). The Falcon containing the solution was put on a rotating wheel and mixed for 20 minutes. After lysis was completed, soluble nuclear material was separated from insoluble chromatin and lipids by centrifugation in a pre-cooled SW40 rotor at 38000 rpm for 1 hour in ultraclear tubes. After centrifugation, a thin lipid layer was removed by suction (upper layer), the soluble nuclear extract was carefully taken using a glass pipet (intermediate liquid portion). Care was taken not to contaminate the sample with the lower debris portion. Soluble nuclear extract was then, dialysed against Dialysis Buffer (15 mM HEPES; 20% glycerol; 1.5 mM $MgCl_2$; 200 mM KCl; 0.2 mM EDTA pH7.9; 1 mM DTT) in Spectra/Por Membrane 1 (cut-off: 6-8000). After dialysis, soluble nuclear extract was quick-frozen in liquid nitrogen and stored at -80°C.

2. TANDEM AFFINITY PURIFICATION (TAP) PROCEDURE

TAP procedure was performed as reported in *Klymenko et al, 2006*. For the TAP purification (Figure 4), 200 μl cross-linked IgG beads were sequencially washed with 1 ml of 0.5 M Acetic acid, 5 ml of PA buffer, 1 ml of 0.5 M Acetic acid, 5 ml of PA buffer (PA buffer: 10 mM Tris-HCl pH 8.0; 150 mM NaCl; 0.1% NP40; 2 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM DTT). All steps next were carried out at 4°C. IgG resin was equilibrated with 10 ml of PA buffer for 30 minutes, in a falcon tube. After equilibration, remove the PA buffer and add 10 ml of nuclear extract. Nuclear extract was thawed on ice. Binding was performed for 2-4 hours at 4°C in a rotating wheel. After the incubation period, discard the unbound material by centrifugation at 1500 rpm for 2 minutes The IgG resin was washed 4 times with 10 ml of PA buffer (each wash for 5-10 minutes in the rotating wheel at 4°C; discard the washing solution by centrifugation at 1500 rpm for 2 minutes). IgG beads were transferred to a 1.5 ml tube for cleavage with TEV protease in 2 ml of PA buffer for 12-16 hours in a rotating wheel at 4°C. After cleavage, eluate was removed and beads were washed with 1.5 ml of CB buffer (CB buffer: 10 mM Tris-HCl pH 8.0; 150 mM NaCl, 0.1% NP40; 1 mM $MgCl_2$; 2 mM $CaCl_2$; 1 mM Imidazole pH 8.0; 10 mM β-mercaptoethanol). Pool the elution from TEV cleavage and the wash. Adjust the volume to 10 ml with CB buffer and the calcium concentration to a final concentration of 3 mM. To this solution, add 200 μl of calmodulin beads (pre-washed with 5 ml of CB buffer) in a fresh 15 ml falcon tube. Binding was performed during 3 hours. After this incubation period, unbound material was discarded by centrifugation at 1500 rpm for 2 minutes. Beads were washed 3 times with 10 ml of CB buffer and, were transferred to a 1.5 ml tube. The bound material was eluted with 300 μl (2-3 elutions were made for a 1 hour each in a rotating wheel at 4°C) of CE buffer (CE buffer: 10 mM Tris-HCl pH 8.0; 150 mM NaCl; 0.1% NP40; 1 mM $MgCl_2$; 2 mM EGTA; 1 mM Imidazole pH 8.0; 10 mM β-mercaptoethanol).

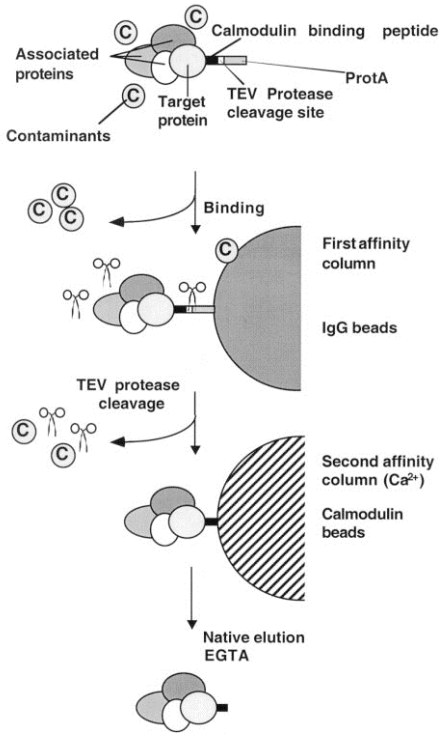

Figure 4 - TAP purification layout (*Rigaut et al, 1999*).

3. SILVER STAINING OF THE PROTEIN GELS

The Silver staining protocol for MS-MS analysis was adapted from *Shevchenko et al, 1996*. All the steps were performed on a shaking table at room temperature except the incubation in silver solution that was performed at 4°C. All solutions must be done fresh each time. Gels were fixed with Fixing solution (40% Methanol; 10% Acetic acid) for 20-30 minutes; then rinsed with water several times, to remove the acid, over a period of 3-4 hours. Sensitize the gels with 0.02% (w/v) sodium thiosulfate solution for 1-2 minutes and rinse them twice with water for 1 minute each. Incubate in (chilled) 0.1% (w/v) silver nitrate solution for 20-40 minutes at 4°C. After the Silver nitrate solution incubation, gels were rinsed with two changes of water (1 minute each) and immediately developed with developing solution (0.04% (v/v) formaldehyde; 2% (w/v) sodium carbonate) on a shaking table. Developing solution is changed for fresh one when it turned yellow; when sufficient staining was obtained, the development was quenched by adding 1% Acetic acid solution. The silver stained gels were stored in 1% (v/v) Acetic acid solution at 4°C.

4. PROTEIN IDENTIFICATION BY MS-MS AND LC-MS/MS

Sven Fraterman, a PhD student in Matthias Wilm group at EMBL, performed the protein identification by MS-MS and by LC-MS/MS of the eluted material from TAP purification (adapted from *Fraterman et al, 2007*).

Table 3 - List of all the peptides obtained in the dSfmbt-TAP, TAP purification.

Protein name (CG number)	Peptides	Peptides
CG 3363	ILLFDPR IQGAVVYYR LNATIIDAAAR VINDSDDRLDAAPSK DAEEQASQTPAGR FLDMLISK GISNEDAFNLR HAAAAAAAAAAAAAAAAASAK IEEHEETEVYR KIETPAEGPTPGKR LAVVMQGR LPGAQGNNLK LQEFTFR NAIGTPPGLNVLK QEPPPANSLFKDESPLK QFPDITEEAR SFTFVIPYLNDR SSNDVADQKPQTQAVETLPLK SSNDVADQKPQTQAVETLPLKTESK VAEPLAAEELSK VINDSDDRLDAAPSK	AHYSFERPPEETSTK APVVKPQSTSVSK DFLHSNWWNVER DQWNAPVTHFNPDLAVR FIVYALSPR GLVNYYK IFNYFR IPSQLSVEIEK KAHMLDIQVPR LLPSGPSMTK LNATIIDAAAR LQFFPNAPYDAFVTPSSK MGIINNEANVYR QLVIDEQDDEQAETDVKQEVER QQLELLK RLNATIIDAAAR SFKEEPNKLQEFTFR TLPNETVQVVK TYELPPIILK
dSfmbt (CG 16975)	DDIMPLLGMK DIYPQDDLPQIPK DNNFDDNGSELEPK EAMIEVHEDDATIELFK FGYMMIR GELYSLVLNTK GNIDPSVIPIQK INDSLQSR ISDLIAQLK LFKDIYPQDDLPQIPK LLQLTKDDIMPLLGMK MNFTFDEYYSDGK MYDPTHSYDWLPR TFTWEGYLR TLPSNFYNK TLPSNFYNKINDSLQSR TNSFVEGMK	DIYPQDDLPQIPKYER DTGAVAAGQHLFHR DWKDFLVGR LADIDSSEPHLELVPDTWNVYDVSQFLR LFKDIYPQDDLPQIPKYER MMWMSSQYNSER RMYDPTHSYDWLPR YKDWKDFLVGR
Pho (CG 17743)	AADNIFSSK FAQSTNLK GKEEFGIDGFTSQQNK GNLSQENNISER NIGYGENQETSK TLSNHTGNTGDLHALPSSVPFR	ERGNLSQENNISER FSLDFNLR FTNAQTLEMPHPISSVQIMDHLIK GKEEFGIDGFTSQQNKEYQK KGDNVINYNIHENDKIK MNEGNHYDLHR
HDAC1/ Rpd3 (CG 7471)	FHSDEYVR LHISPSNMTNQNTSEYLEK SDNDAGATANAGSGSGSGSGAGAK SIRPDNMSEYNK YGEYFPGTGDLR	
HP1b (CG 7041)	ILGATDSSGHLMFLMK KSFLEDDTEEQKK LSTSSTPESIR SFLEDDTEEQK SFLEDDTEEQKK	

	TEYYLK WKGSDHADLVPAK	
NAP-1 (CG 5330)	GIPGFWLTVFR MLPAPVQNR NTAIMSEMVQPHDEPAIR QYLQQMVK YQVQYQPLFDK YQVQYQPLFDKR	

BRM-associated protein 55 (CG6546)	KFYVDTNYVTVPR SPLGGDFLSR VAAQIPTVHYEFPNGYHQDFGSER VGYAQEDSPK	
ISWI ATPase (CG8625)	AALDSLGESSLR FGANQVFSSK KIDEAEPLTEEEIQEKENLLSQGFTAWTK	
Reptin (CG 9750)	ATEVNTEDVK GLGLDDVLEAR IATDTSLR	
Pontin	GLGLDEVGAAVHSAAGLVGQK NQISKDDIEDVHSLFLDAK	
β-tubulin at 56D (CG9277)	ALTVPELTQQMFDAK AVLVDLEPGTMDSVR EVDEQMLNIQNK FPGQLNADLR IMNTYSVVPSPK INVYYNEASGGK IREEYPDR ISEQFTAMFR LAVNMVPFPR YLTVAAIFR	GHYTEGAELVDSVLDVVR GHYTEGAELVDSVLDVVRK IREEYPDRIMNTYSVVPSPK LHFFMPGFAPLTSR MSATFIGNSTAIQELFK MSATFIGNSTAIQELFKR NSSYFVEWIPNNVK SGPFGQIFRPDNFVFGQSGAGNNWAK
α-tubulin at 84B (CG1913)	AVFVDLEPTVVDEVR DVNAAIATIK EDLAALEK EIVDLVLDR FDLMYAK LSVDYGK NLDIERPTYTNLNR TVGGGDDSFNTFFSETGAGK VGINYQPPTVVPGGDLAK	IHFPLVTYAPVISAEK QLFHPEQLITGK QLFHPEQLITGKEDAANNYAR TVGGGDDSFNTFFSETGAGK
α-tubulin at 67C (CG8308)	DVNAAVSAIK EELTASGSSASVGHDTSANDAR ELYHPEQLISGKEDAANNYAR ENIAVLER IGINYEKPAFVPDGDLAK IHFPLVAYAPLMSAER LDFAVYPSPK SIFVDLEPTVIDDVR TFFTETGNGK TKEELTASGSSASVGHDTSANDAR	
Hsp-4a (CG 4264)	ATLDEDNLK ARFEELNADLFR DAGTIAGLNVLR IINEPTAAAIAYGLDKK LLQDLFNGK LVTHFVQEFK STAGDTHLGGEDFDNR	FEELNADLFR HWPFEVVSADGKPK KTFFPEEISSMVLTK LLQDLFNGKELNK LSKEDIER NQVAMNPTQTIFDAK SVIHDIVLVGGSTR WLDANQLADKEEYEHR
EF-1α (CG8280)	EVSSYIK IGGIGTVPVGR LPLQDVYK QTVAVGVIK VETGVLKPGTVVVFAPANITTEVK	
PepB (CG 6143)	AAAALEENER EGIEESYR NQNPPSLLDLPR	
Actin	AGFAGDDAPR	

	GYSFTTTAER IIAPPER	
40S ribosomal protein S3a (CG2168)	IASDYLK LIAEDVQDR VFEVSLADLQK	
Reptin (CG 9750)	ATEVNTEDVK GLGLDDVLEAR IATDTSLR	
Pontin (CG4003)	GLGLDEVGAAVHSAAGLVGQK NQISKDDIEDVHSLFLDAK	
CG 12262	ENVLIGEGAGFK GITFEDVR	
Quaking related 58E-1 (CG3613)	ISYALGEIR LLDDEVEK	
Stonewall (CG3836)	ASISFAGIR QISPTGGASLLQK SGIAPEVQR TSFQDLLQQASQTVQR VYKPGTGQEIAPTSPAGGLDKR	
Host cell factor (CG1710)	ELIERPESETNTR ISASDLNSEHIIQAENHSFANR TNLEPGTAYR WLQDPAAAK YLNDLYILDTR	
Longitudinals lacking protein G (CG12052)	KENTAPDVASTAEIQR LSGAYTLEQTK NLNADEVMR	

5. CHROMATIN IMMUNOPRECIPITATION

Chromatin immunoprecipitation (ChIP) assays were performed as previously described in *Oktaba et al, 2008*. It was used three independent chromatin preparations per specific antibody. For each ChIP reaction, 50-200 μl of chromatin solution was taken up in 1 ml RIPA buffer (140 mM NaCl; 10 mM Tris-HCl pH 8.0; 1 mM EDTA; 1% Triton X-100; 0.1% SDS; 0.1% sodium deoxycholate; 1 mM PMSF). Chromatin was pre-cleared by adding 50% (v/v) beads suspension (0.1 g Protein A Sepharose CL-4B beads - GE healthcare - equilibrated in 1 ml RIPA [with "Complete EDTA-free" protease inhibitor cocktail – Roche], incubated at 4°C for 1 hour, centrifuged at 4°C for 10 minutes at 16000g and beads resuspended in 500 μl RIPA [with "Complete EDTA-free" protease inhibitor cocktail – Roche]) and incubated at 4°C for 1 hour. The chromatin was removed after centrifugation at 4°C for 30 seconds at 16000g and incubated with specific antibodies overnight at 4°C. In parallel, beads were blocked in RIPA containing 1 mg/ml of BSA overnight at 4°C. Antibody-chromatin complexes were recovered by incubation with 40 μl of 50% (v/v) blocked beads suspension at 4°C for 3 hours. At 4°C, 10 minutes 1 ml washes were performed as follows: 5x RIPA, 1x LiCl (250 mM LiCl; 10 mM Tris-HCl pH 8.0; 1 mM EDTA; 0.5% NP40; 0.5% sodium deoxycholate), 2x TE (10 mM Tris-HCl pH 8.0; 1 mM EDTA), removing each time the solution by centrifugation at 4°C for 30

seconds at 16000g. Antibody-chromatin complexes were suspended in 100 μl TE and incubated with 50 μg/ml RNase (Roche) at 37°C for 30 minutes, then adjust to 0.5% SDS and 0.5 mg/ml of proteinase K and incubated at 37°C overnight. For reversal of cross-links, samples were incubated at 65°C for 6 hours, followed by phenol/chlorophorm extraction using Phase-Lock heavy gel tubes (Eppendorf) following the manufacturer's instructions and DNA precipitation. The pellet was resuspended in 30 μl H_2O prior to storage at -20°C. 1-2 μl of eluate was used to check enrichment using specific primers by qPCR.

6. Real-time quantitative PCR – Rt-qPCR

Rt-qPCR was performed as previously reported in *Oktaba et al, 2008*. ChIP eluate was analysed using a real-time PCR instrument (Applied Biosystems 7500) using SYBR Green (Applied Biosystems) and standart settings (Applied Biosystems). PCR was performed in duplicates and serial dilutions of purified input DNA were measured together with the immunoprecipitated DNA samples. This allowed to calculate the amount of target sequence in immunoprecipitated chromatin relative to the amount of target sequence in input chromatin. With each antibody, immunoprecipitation reactions from independently generated batches of chromatin were performed and for each PCR fragment, the amount of DNA in the immunoprecipitated material was expressed as percentage of DNA present in the input material.

Table 4- List of primers used for qPCR analysis in ChIP experiments.

Name	Forward primer (5' to 3')	Reverse primer (5' to 3')	Position to TSS (kb)
Ubx	GCAGCATAAAACCGAAAGGA	CGCCAAACATTCAGAGGATAG	-30
Ubx	ATGATATCTCGTCTGGCACTAC	AGACATCCAGCAAACTGCGATA	+8
Abd-B	CACTTTCGAGCAAGAGCG	AACAGCGATTCACAGACAGC	0
Abd-B	GGAATACCGCACTGTCGTAGG	GCAGCCATCATGGATGTGAA	+72
en	GTTCACTCCCTCTGCGAGTAG	GAAAACGCAGATTGAAACGTC	0
en	GGCGGTGTCAATATTTTGGT	CGCCTTAAGGTGAGATTCAGTT	+3
ap	AGTGTGAGTGAGTGCGATGTC	CATTTGCCACTACGTGAGAGC	-2
ap	GGTAAGCCAAACCGTGAATG	CCCTCGATTTTCTCCCTTAAC	+9
Dll	CCTAGCCACAAAGCGACATT	CCCTGCTGAGAGCAGAAACT	-1
Dll	ATCCGGAGCACCTATCAGC	GGTAGCCGGAGTAGGAGTTCT	+3
eve	TTGTGACTTTGGGTCTGAGG	GAAGAGGAGTGGGGAGTCG	+2
eve	CTCAATCCGCTCCATCAGTT	CCGCAATCACAGTTGTCGT	+9
pnr	GAGCAGGGGTGTTGAGACA	TCTTTCCTTCAGGGACTGTCA	+4
pnr	AGTAGCACCTTGATGGGACAT	GTCTGTCCGTAAAGGGGAAAG	+5
C1	ACACTGCGAGCGCCTCACACGC	CCTAGGTGAATGTGCGGCACAC	
C2	TCAAGCCGAACCCTCTAAAAT	AACGCCAACAAACAGAAAATG	

C3	CCGAACATGAGAGATGGAAAA	AAAGTGCCGACAATGCAGTTA
C4	CAGTTGATGGGATGAATTTGG	TGCCTGTGGTTCTATCCAAAC

7. IMMUNOSTAINING OF *DROSOPHILA* LARVAL DISCS

Immunostainings of *Drosophila* larval discs were made as previously described in *Beuchle et al, 2001*. Third instar larvae were dissected in PBS (137mM NaCl; 2.7 mM KCl; 10 mM Sodium Phosphate; 2 mM Potassium Phosphate; final pH 7.4) by splitting them in half and turning the anterior part inside out. Fat body, salivary glands and gut are discarded and the carcass with the discs attached was fixed for 20 minutes in 4% formaldehyde in PBT (0.1% tween in PBS). Carcasses were then blocked by several washes with BBT (1% Bovine Serum Albumin – BSA – and 0.1% Triton X-100 in PBS). Staining was performed at 4°C overnight with the corresponding primary antibodies diluted in BBT. Primary antibody was washed off with BBT (6 times changed for a period of 1 hour). Carcasses were then stained at 4°C overnight with fluorescently labeled secondary antibody diluted in BBT in tubes protected from light. The secondary antibody was then washed off by washing twice with BBT and four times with PBT for a 1 hour period. Discs were then carefully taken out of the carcasses and mounted in Fluoromount-G on glass slides.

Table 5 – List of antibodies used for *Drosophila* discs immunostaining.

Antibody	Antibody reference or source
Anti-Ubx	Mouse monoclonal FP.3.38. Described in *White & Wilcox, 1984*
Anti-Abd-B	Mouse monoclonal 1A2E9. Described in *Celniker et al, 1990*
Anti-En	Mouse monoclonal 4D9

LIST OF FLY STOCKS USED IN THIS STUDY:

dSfmbt-CTAP construct rescue:

yw hs-flp; hs-nGFP FRT40

w; dSfmbt^{1} FRT40/SM6B

yw hs-flp; dSfmbt^{1} FRT40; α-tubulin-dSfmbt-TAP (w^{+})

dSfmbt-CTAP fly population genotype:

yw hs-flp; +; α-tubulin-dSfmbt-TAP (w^{+})

For functional analysis of dSfmbt and Scm:

yw hs-flp; hs-nGFP FRT 40

yw hs-flp; [hs-nGFP FRT40; $Scm^{Su(z)302}$]/ SM5-TM6B

w; $dSfmbt^{1}$ FRT40/ SM6B

w; FRT82 Scm^{D1}/ TM6C

w; [$dSfmbt^{1}$ FRT40; FRT82 Scm^{D1}]/ SM5-TM6B

yw; FRT40 FRT42D P[y^{+}] $calypso^{2}$/ SM6B

yw hs-flp; FRT40 FRT42D P[y^{+}] $calypso^{2}$; $Scm^{Su(z)302}$]/ SM5-TM6B

yw hs-flp; [FRT42D hs-nGFP; Scm^{D1}]/ SM5-TM6B

FLP expression was induced by 1 hour heat-shock at 37°C, 24-48 hours after egg laying at 25°C; dissection was performed 96 hours after heat-shock. GFP expression was induced by 1 hour heat-shock at 37°C, followed by 1 hour at 25°C prior to dissection.

8. Protein expression and FLAG-purification using baculovirus expression system in Sf9 cells

Baculovirus production was performed with the Bac-to-Bac system (GibcoBRL), using full-length cDNAs inserted into pFastBacDUAL. For this study the following virus were created: FLAG-Scm (cDNA in pFastBacDUAL was a gift from J. Simon), FLAG-ScmΔSAM. Viruses expressing untagged-dSfmbt, untagged-Pho and untagged-Ph were described previously (*Klymenko et al, 2006*; *Francis et al, 2001*). FLAG purification was performed as previously described (*Müller et al, 2002*) with the following modifications. FLAG affinity purification was carried out as following: 50 ml of Sf9 cell culture (cell density 10^{6} cells/ml) was infected with a single virus or viruses for different combinations of complexes components. 48h after infection, cells were pelleted in a 50 ml falcon tube in a Sorvall Megafuse centrifuge at 3000 rpm for 15 minutes. Cells were suspended in 10 ml of Lysis Buffer F (Lysis Buffer F: 20 mM Tris-HCl pH 8.0; 300 mM NaCl; 20% glycerol; 4 mM $MgCl_2$; 0.4 mM EDTA; 2 mM DTT; 0.05% NP40; 10 μM $ZnCl_2$; "Complete EDTA-free" protease inhibitor cocktail – Roche) and then disrupted with ultrasonic cell disruptor (50 strokes 3 times with a 5 minutes interval on ice) at 4°C. Insoluble material was pelleted by ultracentrifugation in a Beckman SW40 rotor at 12000 rpm for 25 minutos at 4°C. After centrifugation, the supernatant was poured into a 15 ml falcon tube. M2 anti-FLAG beads from Sigma were equilibrated accordingly with the

advised from the producer. 300 μl of M2 anti-FLAG beads were added for the 10 ml of whole-cell extracts. Binding was carried out for 2-4 hours at 4°C on a rotating wheel. After incubation, the beads were washed with BC buffer (20 mM HEPES pH 7.9; 0.4 mM EDTA; 20% glycerol; "Complete EDTA-free" protease inhibitor cocktail – Roche) using stepwise KCl salt washes (3 times with 600 [with 0.05% NP40] and 1200 [with 0.05% NP40] mM KCl, twice with 1200, 600 and 300 mM KCl). Beads were transferred to a 1.5 ml tube and the complex was eluted using BC buffer with 300 mM KCl and 0.4 mg/ml of FLAG peptide (DYKDDDDK – Sigma) for 1-2 hours. Purified complexes were stored at -80°C.

9. SITE-DIRECTED MUTAGENESIS PCR

For this assay I used *Pfu* Turbo (Stratagene), which is a DNA polymerase. The PCR reaction goes as follow: 5 μl of 10x *Pfu* Turbo buffer (provided by manufactor); 100 ng of plasmic DNA; 1 μl of primer 1 (at a concentration of 1 pmol/μl); 1 μl of primer 2 (at a concentration of 1 pmol/μl); 1 μl of 10 mM dNTPs; 1 μl *Pfu* Turbo (2.5 enzymatic units) and H2O till a final volume of 50 μl.

The PCR programme is: 95°C for 2 minutes; [95°C for 1 minute; 55°C-68°C for 1 minute; 68°C for 2 minutes/ kb of plasmid] x 18 cycles; 68°C for 2 minutes/kb of plasmid + 5 minutes.

After PCR, I added 1 μl (20 enzymatic units) of restriction enzyme *DpnI* directly on the mixture. Incubate at 37°C for 1h. The *DpnI* restriction enzyme cuts G^{6m} ATC sequence, hence eliminating the template plasmid and leaving only the PCR product. After, 2 μl of the solution was transformed in competent *E. coli* cells using a standard protocol.

Table 6 – Primers used for Site-directed mutagenesis PCR

Use	Primer sequence	Primer name
dSfmbt mutation of STOP codon to a *NotI* site	AAATCACCGTTTTTAGGCGGCCGCGATTTTAGTTTTTTC	**Forward primer (5' to 3')**
	TTTAGTGGCAAAAATCCGCCGGCGCTAAAATCAAAAAAG	**Reverse primer (5' to 3')**
Scm mutation of two aas to 2 STOP codons.	CATCTGCGGTCGCAGTAGTAGGACTGGACCATCG	**Forward primer (5' to 3')**
	CGATGGTCCAGTCCTACTACTGCGACCGCAGATG	**Reverse primer (5' to 3')**

10. MONONUCLEOSOME ASSEMBLY

The mononucleosome assembly protocol was adapted from *Nekrasov et al, 2005*. Recombinat *Xenopus* octamers were prepared as described (*Luger et al, 1999*) and mononucleosomes were assembled onto a (radiolabelled) 201 bp '601' DNA template (*Thåström et al, 1999*). For assembly, 7 μg octamer was mixed with 5 μg DNA template in a volume of 15 μl, which corresponds to a concentration of nucleosomes of 3.3 pmol/μl (2M NaCl, 10 mM Tris pH 8.0, 0.1 mM EDTA, 10 mM β-mercaptoethanol) at room temperature followed by stepwise reduction of the salt concentration by addition of Tris-EDTA to obtain a solution containing 0.17 pmol nucleosomes/μl (100 mM NaCl/Tris-EDTA).

11. HISTONE METHYLTRANSFERASE ASSAYS

The HMTase assays were adapted from *Nekrasov et al, 2005*. Reactions were carried out in a volume of 50 μl and contained 7 pmol of nuclesome, 2 pmol of Ash1 protein, 10 mM HEPES Ph 7.9, 0.25 mM EDTA, 200 mM NaCl, 10% glycerol, 2 mM dithiothreitol (DTT), 2.5 mM $MgCl_2$ and 4 μM S-adenosyl-L-[methyl-^{14}C]methionine. Reactions were incubated for 40 minutes at 30°C and resolved in a 18% SDS-polyacrylamide gel run at 12-15V/cm; after Coomassie staining, the gel was dried and exposed for autoradiography.

RESULTS

1. BIOCHEMICAL CHARACTERIZATION OF A NEW PCG COMPLEX

• BIOCHEMICAL PURIFICATION OF A NEW PCG COMPLEX

In order to understand if the PhoRC subunit dSfmbt is the physical link between the PhoRC and PRC1 and/or PRC2, I performed a TAP-tag purification using dSfmbt as a bait protein (*Rigaut et al, 1999*; figure 4).

I firstly constructed a vector that carried the cDNA of dSfmbt under the α-tubulin promoter, with a TAP-tag at the N-terminus of the protein (TAP-dSfmbt). I then generated *Drosophila* transgenic lines carrying this vector and after a series of crosses, I isolated three transgenic TAP-dSfmbt fly lines with insertions in the third chromosome. In order to check for functionality of the TAP-dSfmbt protein, the transgene was placed into a dSfmbt null background (*dSfmbt* $^{-}$). These homozygous mutant flies die as 3rd instar larvae and misexpress the HOX gene *Ubx* in the wing imaginal disc (*Klymenko et al, 2006*). Therefore, the functionality of the TAP-dSfmbt protein would rescue the viability of the flies and lead to the proper expression of *Ubx*. However, the TAP-dSfmbt protein did not rescue viability of *dSfmbt* $^{-}$ allele nor it rescued the misexpression of *Ubx* in wing imaginal discs.

As the N-terminus TAP-tag could potentially perturb the structure of the protein, I constructed a different vector where the dSfmbt protein had a C-terminus TAP-tag (dSfmbt-TAP; figure 5A). I isolated two dSfmbt-TAP transgenic lines ([*dSfmbt* $^{-}$*; tub-dSfmbt-TAP1*] and [*dSfmbt* $^{-}$*; tub-dSfmbt-TAP2*]) with insertions in the third chromosome, and repeted the procedure described above to verify the functionality of the protein. Although the dSfmbt-TAP protein did not rescue the viability of *dSfmbt* $^{-}$ flies, dSfmbt-TAP rescued *Ubx* misexpression in wing imaginal discs (Figure 5B).

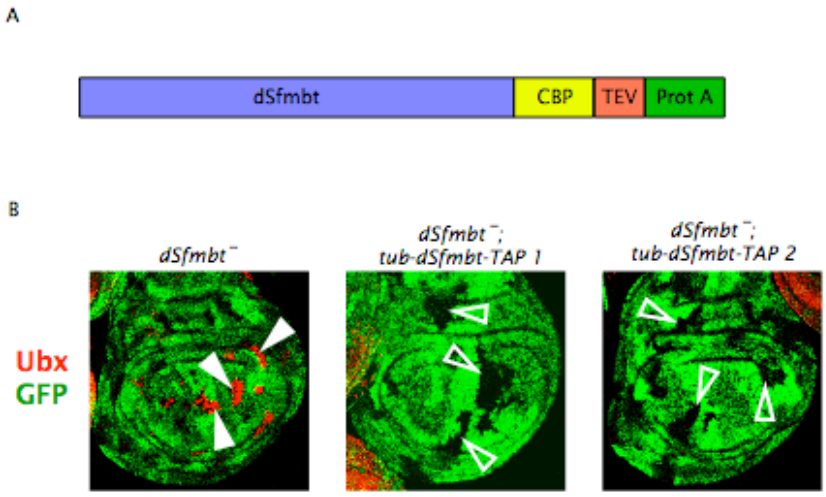

Figure 5 - dSfmbt-TAP transgene rescues *Ubx* misexpression in dSfmbt mutant clones.
(A) Schematic representation of the dSfmbt-TAP protein. **(B)** Analysis of wing imaginal discs after 96h of induction of clones of cells that are homozygous for *dSfmbt*$^{-}$ (marked by the absence of GFP) with ([*dSfmbt*$^{-}$*; tub-dSfmbt-TAP 1*] or [*dSfmbt*$^{-}$*;tub-dSfmbt-TAP 2*]) or without (*dSfmbt*$^{-}$) transgene. The dSfmtb-TAP rescue transgene was stained with an antibody against the HOX protein Ubx (which is not expressed in this tissue in wild-type animals) seen here in red. The misexpression of Ubx seen in *dSfmbt*$^{-}$ (*full arrowhead*; *Klymenko et al, 2006*) is completely rescued in both transgenic lines ([*dSfmbt*$^{-}$*; tub-dSfmbt-TAP 1*] and [*dSfmbt*$^{-}$*; tub-dSfmbt-TAP 2*]), where no misexpression of Ubx is observed (*empty arrowhead*).

I expanded the fly line dSfmbt-TAP 1 in a wild type background and prepared nuclear extracts from embryos collected from 0h to 12h post fertilization. Nuclear extracts from a wild-type fly population were used as a negative control for the dSfmbt purification.

I performed three purifications using independent transgenic and wild-type embryonic nuclear extract preparations. The purified material was separated in a 4-12% Bis-Tris SDS polyacrylamide gel and silver stained. Purified material from wild-type and dSfmbt-TAP purification were separated and analysed by MS-MS for identification of peptides (white rectangles; figure 6A). The material from wild-type and dSfmbt-TAP purification was also analysed by LC-MS/MS for a more thorough analysis. Both MS-MS and LC-MS/MS were performed and analysed by Sven Fraterman, from the laboratory of Matthias Wilm at EMBL.

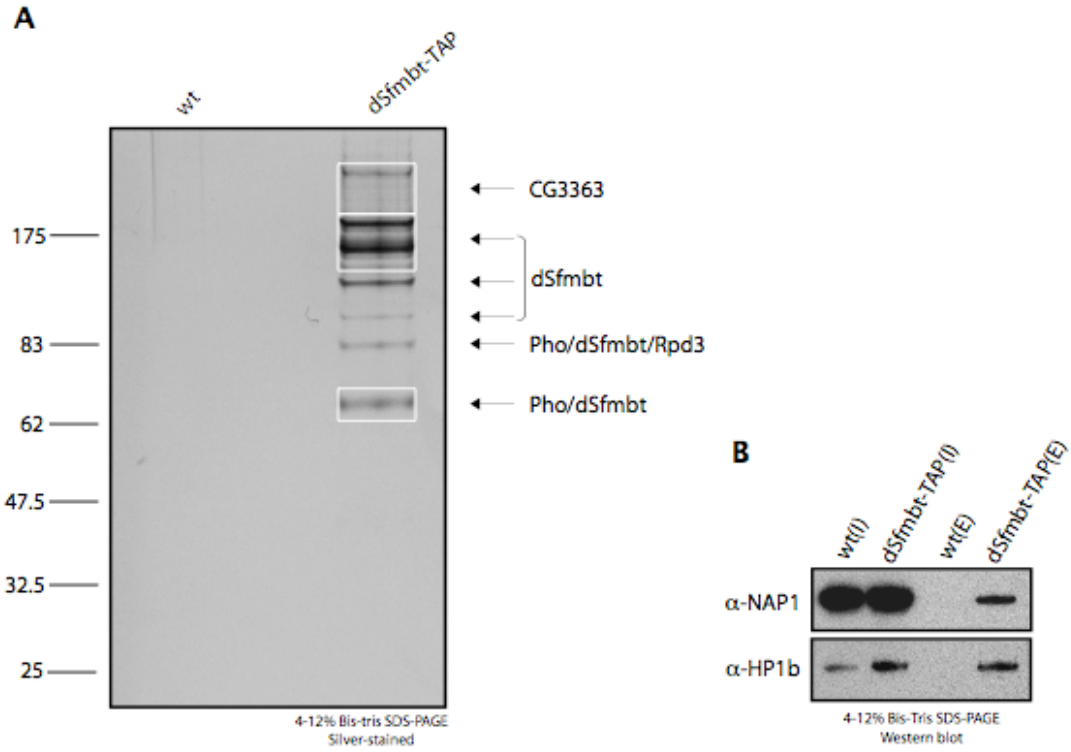

Figure 6 – dSfmbt complex purified from *Drosophila* embryos.
(A) Proteins eluted from the calmodulin affinity column (see materials and methods) were separated on a 4-12% polyacrylamide gel and bands were revealed by silver staining. The molecular weight protein marker can be seen on the left of the gel. **Wild-type (wt)** – contains purified material from 0-12h wt embryonic nuclear extracts; **dSfmbt-TAP** – contains purified material from nuclear extracts of transgenic line *tub-dSfmbt-TAP 1*, prepared from 0-12 hours old embryos. The identification of these proteins was done using MS-MS peptide microsequencing (right side of the gel). White rectangles represent the portions analysed by MS-MS. Single bands were also analysed by MS-MS.
(B) Western Blot analysis of two additional members of dSfmbt complex, NAP1 and HP1b, that were identified by LC-MS/MS. NAP1 is present in the same amounts both in the input material **(I)** of wild-type embryonic nuclear extracts **(wt)** and of *tub-dSfmbt-TAP 1* embryonic nuclear extracts **(dSfmbt-TAP)**. NAP1 protein co-purifies with dSfmbt **(dSfmbt-TAP (E))** and is absent from the eluted material of wild-type embryonic nuclear extracts **(wt (E))**. The same applies for the HP1b protein.

The MS-MS analyses identified peptides from the PcG proteins Pho and dSfmbt, which form the known complex PhoRC (*Klymenko et al, 2006*), and also identified peptides from Rpd3/HDAC1 protein, as well as from an uncharacterized protein, CG3363 (Figure 6A). The LC-MS/MS analyses corroborated the identification of these peptides, and also identified peptides from two chromatin-associated proteins, NAP1 and HP1b (Figure 6B).

Pho is the only PcG protein that binds to specific DNA sequences called PREs (*Brown et al, 1998*). This protein forms a stable complex with dSfmbt forming the PhoRC

complex, which is a PcG protein complex (*Klymenko et al, 2006*). Pho and PhoRC are found at the PREs of many target genes and are required for the tethering of the other PcG complexes to these gene loci (reviewed in *Müller & Verrijzer, 2009*; *Kwong et al, 2008*; *Oktaba et al, 2008*).

Rpd3 is a histone deacetylase enzyme that has been found in several complexes related to transcriptional repression: Rpd3-Sin3 complex in yeast (*Kadosh & Struhl, 1998*; *1998*), Rpd3-Groucho (*Chen et al, 1999*), Rpd3-Even-skipped (*Mannervik & Levine, 1999*), PRC2 complex (*Tie et al, 2001*; *2003*), and PRC1 complex in *Drosophila* (*Chang et al 2001*). In humans it has been reported to interact with the retinoblastoma protein (Rb) in the repression of the E2F target genes (*Zhang et al, 2000*). Rpd3 is *Drosophila*'s homologue of the human Histone deacetylase 1 (HDAC1) and has been involved in the deacetylation of H3K9 and H3K36 (reviewed by *Thiagalingam et al, 2003*; *Lee & Shilatifard, 2007*).

Heterochromatic protein b (HP1b) belongs to a family of chromosomal proteins that are implicated in heterochromatin packaging and gene silencing (reviewed in *Kwon & Workman, 2008*). HP1b is the *Drosophila* homolog of human HP1α. This family of proteins characteristically possesses two major domains: a chromo domain in N-terminal part, and a chromo shadow domain in C-terminal (reviewed in *Kwon & Workman, 2008*). The chromo domain is responsible for the binding of this protein to the trimethylated state of histone H3 in lysine 9 (H3K9me3), and the chromo shadow domain is known to be responsible for protein – protein interaction as it recognizes the specific sequence PxVx[M/L/V], PxVxL being the strongest recognition motif (*Smothers & Henikoff, 2000*; *Bannister et al, 2001*; *Jacobs et al, 2001*; *2002*; *Thiru et al, 2004*). This PxVxL motif is present in the N-terminal part of dSfmbt (Figure 7). The HP1 family is composed three proteins that are present in different states of chromatin, HP1a, b, and c, with HP1b being present in both euchromatin and heterochromatin (*Smothers & Henikoff, 2001*).

Figure 7 – Binding motif of HP1b chromo shadow domain in dSfmbt.
The PxVxL binding motif of the chromo shadow domain (purple) is present in the first 54 aminoacids of the dSfmbt protein. A hydrophobic (leucine) and a slightly polar (threonine) aminoacid at positions -6 and -7 (green) strengthen the interaction between the proteins (*Thiru et al, 2004*).

NAP1 is a Nucleosome Assembly Protein that is conserved from yeast to humans. *Drosophila* NAP1 (dNAP1; *Ito et al, 1996*), human NAP1 (hNAP1; *Ishimi et al, 1984*), and yeast NAP1 (yNAP1; *Ishimi & Kikuchi, 1991*) are functionally conserved proteins (reviewed in *Park & Luger, 2006*). dNAP1 acts as a chaperone that transports newly synthesized histones from the cytoplasm to the nucleus and deposits these histones onto the chromatin by an ATP-facilitated process (*Ito et al, 1996*). NAP-1 interacts mostly with the H2A/H2B dimer and its location inside the cell is dependent on the cell cycle; it is found in the nucleus during S phase and in the cytoplasm when G2 starts (*Ito et al, 1996*; reviewed in *Park & Luger, 2006*). CG3363 is an uncharacterized *Drosophila* protein with no prominent domains on its structure. Being only conserved in Diptera, this protein is not likely to be a PcG protein, as PcG proteins are highly conserved from yeast to humans.

The new complex purified and described in this work containing Pho, dSfmbt, Rpd3, HP1b, NAP1, and CG3363 as subunits will be called from now on the "dSfmbt complex".

- dSFMBT COMPLEX LOCALIZES AT THE PREs

As Pho and dSfmbt both colocalize at the PREs of the HOX genes and other PcG targets (*Klymenko et al, 2006*; *Papp & Müller, 2006*; *Oktaba et al, 2008*), I performed ChIP experiments in order to determine whether Rpd3, NAP1, and HP1b proteins were also tethered at the PREs. Three independent preparations of chromatin were used as input material for the precipitation with antibodies against Pho, dSfmbt, Rpd3, NAP1, and HP1b proteins. The chromatin used for these experiments was prepared by a previous PhD student, Maxim Nekrasov, and originated from imaginal discs and central nervous system (CNS) wild-type larval tissues.

Quantitative real time PCR was performed to quantify the absolute amount of DNA recovered from specific regions of the *Drosophila* genome. The regions checked were PREs from PcG target genes (*Chan et al, 1994*; *Mihaly et al, 1997*; *Oktaba et al, 2008*), a genomic region of each of these genes, and also euchromatic and heterochromatic regions that were chosen randomly in the *Drosophila* genome. The PcG target genes chosen for this experiment were: *Ultrabithorax* (*Ubx*), *Abdominal-B* (*Abd-B*), *engrailed* (*en*), *apterus* (*ap*), *distalless* (*Dll*), *even skipped* (*eve*), and *pannier* (*pnr*).

As previously reported, Pho and dSfmbt colocalized at the tested PREs of PcG target genes and were absent from the corresponding genomic regions, as well as from the euchromatic and heterochromatic regions (*Oktaba et al, 2008*; figure 8, purple and

blue graphs). Rpd3 was specifically localized to all the PREs tested, supporting that dSfmbt and Rpd3 proteins colocalized with each other *in vivo* (Figure 8, yellow graph). NAP1 did not show any preference in binding to the PREs or other genomic regions in this experiment (Figure 8, red graph). As there are no clear negative controls for this experiment, I could not conclude whether this protein was present or absent at the tested genomic regions (Figure 8, red graph). The HP1b antibody was not functional in these experiments, so I could not determine the localization of this protein.

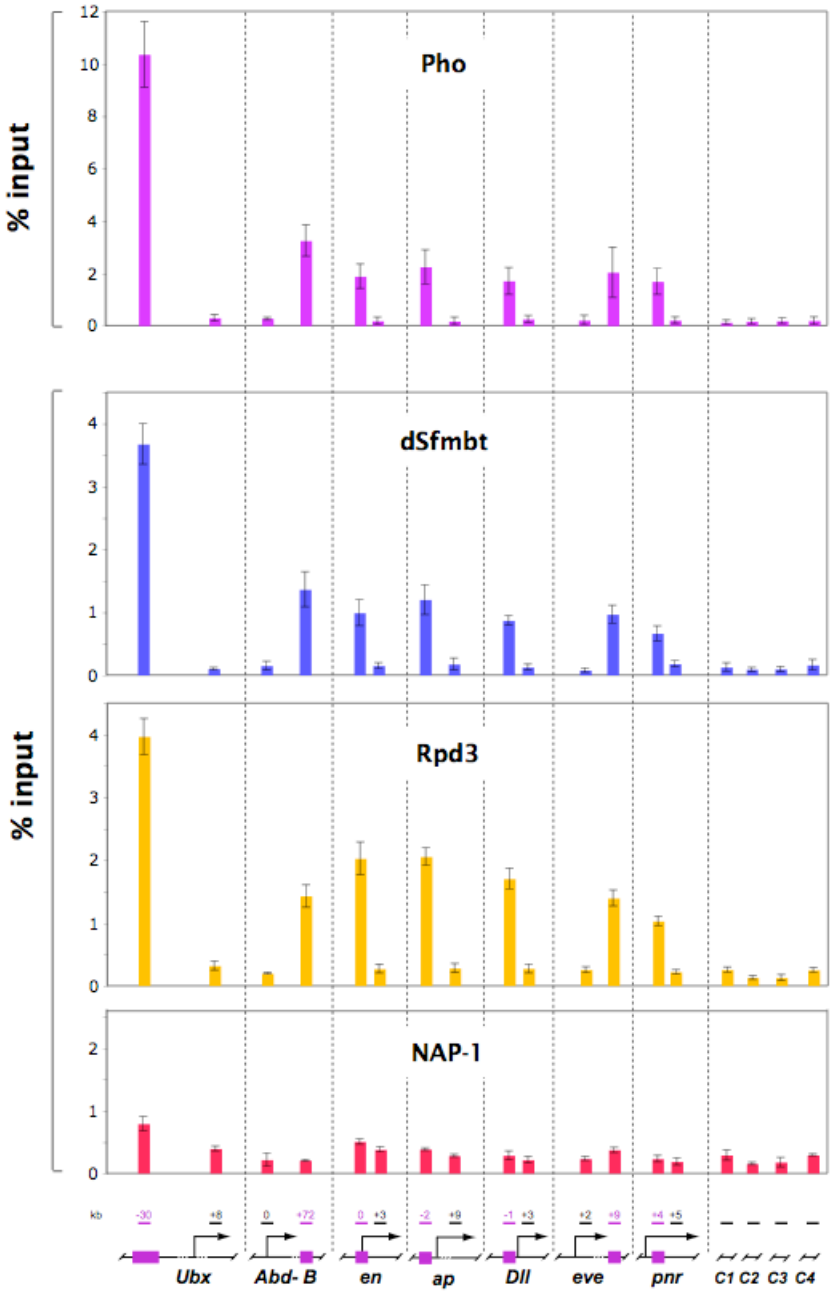

Figure 8 – dSfmbt, Pho and Rpd3 colocalize at the PREs in PcG target genes.
ChIP analysis monitoring dSfmbt (blue graph), Pho (purple graph), Rpd3 (yellow graph), and NAP1 (red graph) binding in imaginal disc/CNS tissues dissected from wild-type larvae. Graphs show the results from three independent immunoprecipitation reactions using different batches of chromatin preparations. ChIP signals were quantified by qPCR and are presented as percentages of the input chromatin precipitated at each region. Error bars correspond to the standard deviations. The location of PREs (purple boxes) and other regions with respect to transcription start sites of *Ubx*, *Abd-B*, *en*, *ap*, *Dll*, *eve*, and *pnr* genes are indicated in kilobases. C1-C4 indicate euchromatic and heterochromatic control regions outside these genes (see **table 4** for qPCR primer sequences). dSfmbt (blue graph), Pho (purple graph), and Rpd3 (yellow graph) proteins are specifically enriched at the PRE of each gene, but not at the analysed intervals in the control regions of the same genes or in the control regions C1-C4. NAP1 (red graph) protein is enriched at the *Ubx* PRE relatively to the analysed coding region of this gene, but there is no difference between the PREs and corresponding coding regions of the other analysed target genes and the control regions C1-C4.

Although Rpd3 is not a PcG protein, it forms a stable complex with PcG proteins with whom it colocalizes at the PREs of PcG target genes. Therefore Rpd3 contributes to the silencing of PcG target genes together with PcG proteins (Figure 8). I concluded that dSfmbt forms a stable complex with Pho (PhoRC; *Klymenko et al, 2006*), Rpd3, NAP1, and HP1b (Figure 6). The function of this complex on gene repression is as of yet unknown and further experiments need to be performed. However, a possible repressional mechanism and the contributions of each protein will be discussed further.

2. CHARACTERIZATION OF THE METHYL-LYSINE BINDING ACTIVITY OF DSFMBT *IN VIVO*

The MBT repeats of dSfmbt bind specifically to mono and dimethylated lysines of histone H3 (H3K9 and H3K27) and histone H4 (H4K20) (*Klymenko et al, 2006*; *Grimm et al, 2009*). The *in vivo* function of these dSfmbt repeats is not known, but it has been proposed that their binding to methylated lysines assists the PRE-tethered PcG complexes with scanning the flanking chromatin for particular lysine modifications (reviewed in *Müller & Kassis, 2006*). In order to determine the *in vivo* importance of the methyl-lysine binding activity of dSfmbt in repression, I generated three different transgenic flies: 1) a transgene that expresses wild-type dSfmbt protein as a control; 2) a transgene that expresses a dSfmbt protein that carries the mutations D917A, E947A, and Y948F (dSfmbt$^{D917A\ E947A\ Y948F}$ - dSfmbt^{AAF}); 3) a transgene that expresses a dSfmbt protein that lacks all four MBT repeats (dSfmbt$^{\Delta MBT}$). The expression of the transgenes was under the control of the α-tubulin promoter (tub-dSfmbt; tub-dSfmbt^{AAF}; tub-dSfmbt$^{\Delta MBT}$). The aminoacids D917, E947, and Y948 are located in the fourth MBT repeat of dSfmbt and are responsible for the binding of this protein to monomethylated H4K20 peptide (*Grimm et al, 2009*). Therefore, the transgenic flies tub-dSfmbt^{AAF} express a dSfmbt protein that is unable to bind to monomethylated H4K20 (*Grimm et al, 2009*).

I inserted the transgene tub-dSfmbt^{AAF} and tub-dSfmbt$^{\Delta MBT}$ in a dSfmbt null background (*dSfmbt*$^{-}$) to see if these non-functional dSfmbt proteins could rescue the misexpression of the HOX gene *Ubx* in wing imaginal discs (Figure 9). The same procedure was done for the transgene tub-dSfmbt as a control (Figure 9).

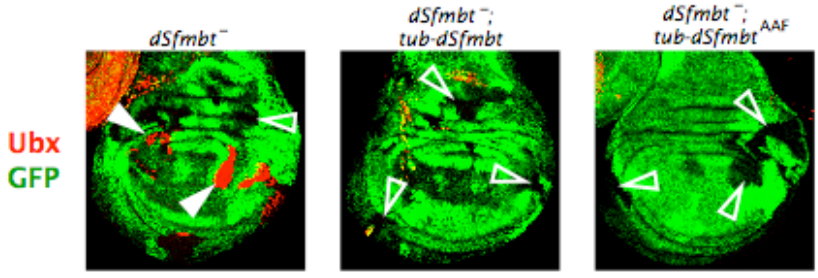

Figure 9 – The Methyl-lysine binding activity of dSfmbt is not involved in the *Ubx* repression in wing imaginal discs.

Analysis of 96h wing imaginal discs after induction of cell clones that are homozygous for *dSfmbt*$^{-}$ (marked by the absence of GFP) by immunostaining against Ubx protein (in red). Left: wing imaginal disc with clones of *dSfmbt*$^{-}$ misexpresses *Ubx* gene in most of the cell clones (*Klymenko et al, 2006*; full arrowheads and empty arrowhead); Middle: wing imaginal disc with

clones of *dSfmbt*$^{-}$ in animals carrying a dSfmbt wild-type transgene under the tubulin promoter rescues the *Ubx* misexpression ([*dSfmbt*$^{-}$*;tub-dSfmbt*]; empty arrowheads); Right: wing imaginal disc with clones of *dSfmbt*$^{-}$ in animals carrying a dSfmbt transgene with no methyl-lysine binding activity ([*dSfmbt*$^{-}$*; tub- dSfmbt*^{AAF}]; *Grimm et al, 2009*) rescues the misexpression of the *Ubx* gene.

Both tub-dSfmbt and tub-dSfmbt^{AAF} transgenes rescue the Ubx misexpression of *dSfmbt*$^{-}$ cell clones in wing imaginal discs. Therefore, the methyl-lysine binding activity of the MBT repeats of dSfmbt alone is not required for the PcG repression of target genes. Unfortunately, a tub-dSfmbt$^{\Delta MBT}$ stable transgenic line could not be established.

3. dSFMBT AND THE PRC1 COMPONENT, SCM

- ### dSFMBT AND SCM COLOCALIZE AT THE PREs

The PRC1 subunit Scm also contains MBT repeats (Figure 11), and these MBT repeats possess a methyl-lysine binding activity. However, the specificity of these MBT repeats is not as high as for the dSfmbt MBT repeats (*Grimm et al, 2007*; *2009*). *In vivo* experiments of the functionality of the MBT repeats of Scm showed, as for the MBT repeats of dSfmbt, that a transgene lacking the MBT repeats of Scm (Scm$^{\Delta MBT}$) and a transgene that abolishes the methyl-lysine binding activity of Scm (Scm^{N324A}) could rescue almost fully the *Ubx* misexpression in wing imaginal discs seen in *Scm* mutant clones (*Grimm et al, 2007*). As these two PcG proteins have the same methyl-lysine binding activity and since abolishing this activity leads to a similar phenotype, it is possible that they act redundantly in *Drosophila*.

In order to know if dSfmbt and Scm colocalized at the PREs of PcG target genes, firstly I performed ChIP experiments. Three independent preparations of chromatin were used as input material for the precipitation with antibodies against Scm and dSfmbt proteins. The chromatin used was prepared by Maxim Nekrasov and originated from imaginal discs and CNS wild-type *Drosophila* larval tissues. The same regions as those used for the colocalization of the dSfmbt complex subunits were tested.

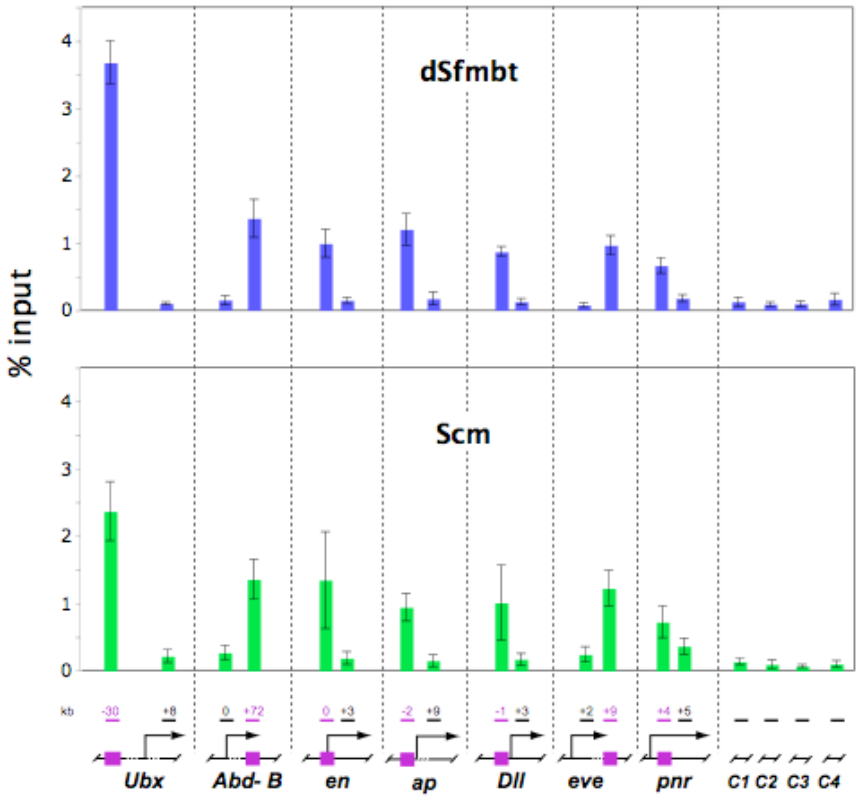

Figure 10 – dSfmbt and Scm colocalize to PREs in PcG target genes.
ChIP analysis monitoring dSfmbt (blue graph) and Scm (green graph) binding in imaginal discs/CNS tissues dissected from wild type *Drosophila* larvae. Graphs show the results from three independent immunoprecipitation reactions from different batches of chromatin preparations. ChIP signals were quantified by Rt-qPCR and are presented as percentages of input chromatin precipitated at each region. Error bars correspond to standard deviations. The location of PREs (purple boxes) and other regions with respect to transcription start sites in the *Ubx*, *Abd-B*, *en*, *ap*, *Dll*, *eve*, and *pnr* genes are indicated in kilobases. C1-C4 indicate euchromatic and heterochromatic control regions outside these genes (see **table 4** for qPCR primer sequences). dSfmbt (blue graph) and Scm (green graph) proteins are specifically enriched at the PRE of each analyzed gene, but not at the analyzed intervals in the coding regions of these same genes or in control regions C1-C4.

ChIP experiments showed that Scm and dSfmbt are specifically bound to all the tested PREs of PcG target genes (Figure 10, blue and green graphs), in accordance with the hypothesis that these two proteins function redundantly in *Drosophila*.

- BIOCHEMICAL CHARACTERIZATION OF DSFMBT AND PRC1 COMPONENTS SCM AND PH

Scm and dSfmbt proteins have a very similar domain structures (Figure 11). They both contain MBT repeats, a zinc finger in the N-terminal part of the protein, and a SAM domain in the C-terminus (Figure 11). The PRC1 subunit Ph also contains a zing finger and a SAM domain in the C-terminal portion of the protein (Figure 11).

SAM domains are responsible for the formation of homomeric and heteromeric protein structures (*Thanos et al, 1999*; *Kim et al, 2002*). It has been described that the PRC1 subunits Scm and Ph interact with each other through these domains (*Peterson et al, 1997*; *2004*).

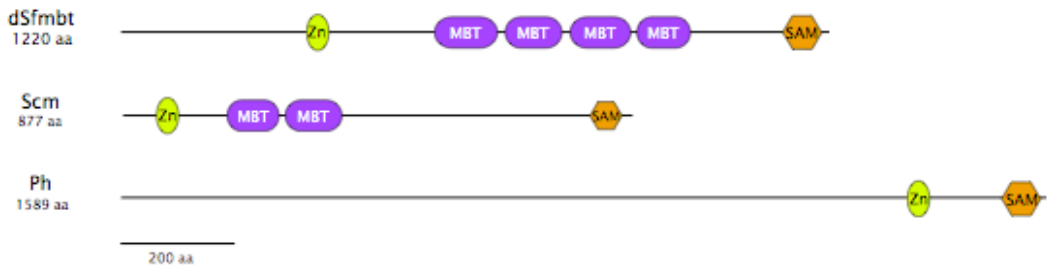

Figure 11 – dSfmbt, Scm, and Ph protein domain structure.
MBT – Malignant Brain Tumor; Zn – Zinc finger; SAM - α-sterile motif; aa- aminoacids.

As dSfmbt, Scm and Ph colocalize at PREs of PcG target genes (Figure 10; *Oktaba et al, 2008*) and as these three proteins contain SAM domains in their C-terminus (Figure 11), I performed interaction assays between these proteins using the baculovirus system in Sf9 cells in order to test whether Scm and/or Ph would form a stable complex with dSfmbt. These interaction assays were performed by infecting Sf9 cells with baculovirus containing a tagged protein (FLAG-Scm) and with baculovirus containing an untagged protein (dSfmbt, Ph or Pho), followed by FLAG affinity purification. The interaction Scm-Ph was used as a positive control (*Peterson et al, 1997*; *2004*) and the co-expression of Scm-Pho proteins as a negative control (*Klymenko et al, 2006*). Purified material was separated by SDS-polyacrilamyde gel and stained with Coomassie. In order to corroborate the results found on the Coomassie stained gel, input and purified material were immunoblotted against the FLAG epitope (which recognizes FLAG-Scm protein), dSfmbt, Ph, and Pho.

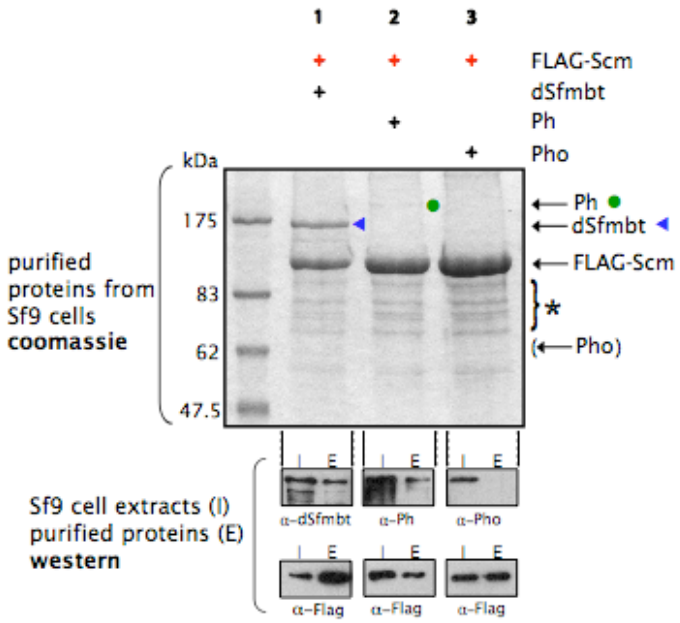

Figure 12 – dSfmbt and Scm proteins interact *in vitro*.
FLAG-tagged Scm and untagged dSfmbt, Ph or Pho proteins were affinity-purified via FLAG-tag, separated by SDS-PAGE, and visualized by Coomassie staining (top). Western blot of corresponding Sf9 total cell extract input prior to purification **(I)** and eluted purified proteins **(E)** reveal relative enrichment of proteins after purification (below). Note that dSfmbt (blue arrowhead) forms a stable complex with FLAG-Scm, while Ph (green dot) co-purifies less efficiently with FLAG-Scm than dSfmbt (*Grimm et al, 2009*). Also note that co-expression of Pho with FLAG-Scm results in the purification of FLAG-Scm only, and Pho is not detected by Coomassie staining or western blot analysis of the eluted material. (Asterisk shows FLAG-Scm degradation products).

FLAG-Scm interacts stably with the Ph protein (Figure 12, lane 2, green dot; *Peterson et al, 1997*; *2004*) and also forms a stable complex with the dSfmbt protein (Figure 12, lane 1 blue arrowhead; *Grimm et al, 2009*). Although in the Coomassie stained gel dSfmbt and Scm appear to form a more stable complex than Scm and Ph, the recovery of dSfmbt and Ph proteins observed by western blot analysis is comparable (Figure 12, lane 1E α-dSfmbt and lane 2E α-Ph). FLAG-Scm does not interact with the Pho protein (Figure 12, lane 3; *Klymenko et al, 2006*). A baculovirus expressing FLAG-Ph was not stable and the baculovirus expressing FLAG-dSfmbt had a low recovery in FLAG affinity purification, so I could not test the interactions between dSfmbt and Ph.

As the Scm – Ph interaction is mediated by their SAM domains (*Peterson et al, 1997*; *2004*), I performed interaction assays to find out whether Scm and dSfmbt might also interact through their SAM domains. The interaction assays were performed as described above with the addition of a set of assays where Sf9 cells were infected with baculovirus for FLAG-ScmΔSAM and baculovirus for untagged dSfmbt or Ph protein. Purified material was separated by SDS-polyacrilamyde gel and stained with Coomassie.

Western blots were performed on input material to show the similar protein expression profile of each baculovirus. The interactions Scm – Ph and Scm – dSfmbt were used as positive controls (Figure 12; *Peterson et al, 1997*; *2004*; *Grimm et al, 2009*) and expression of Scm-Pho proteins were used as negative controls (*Klymenko et al 2006*).

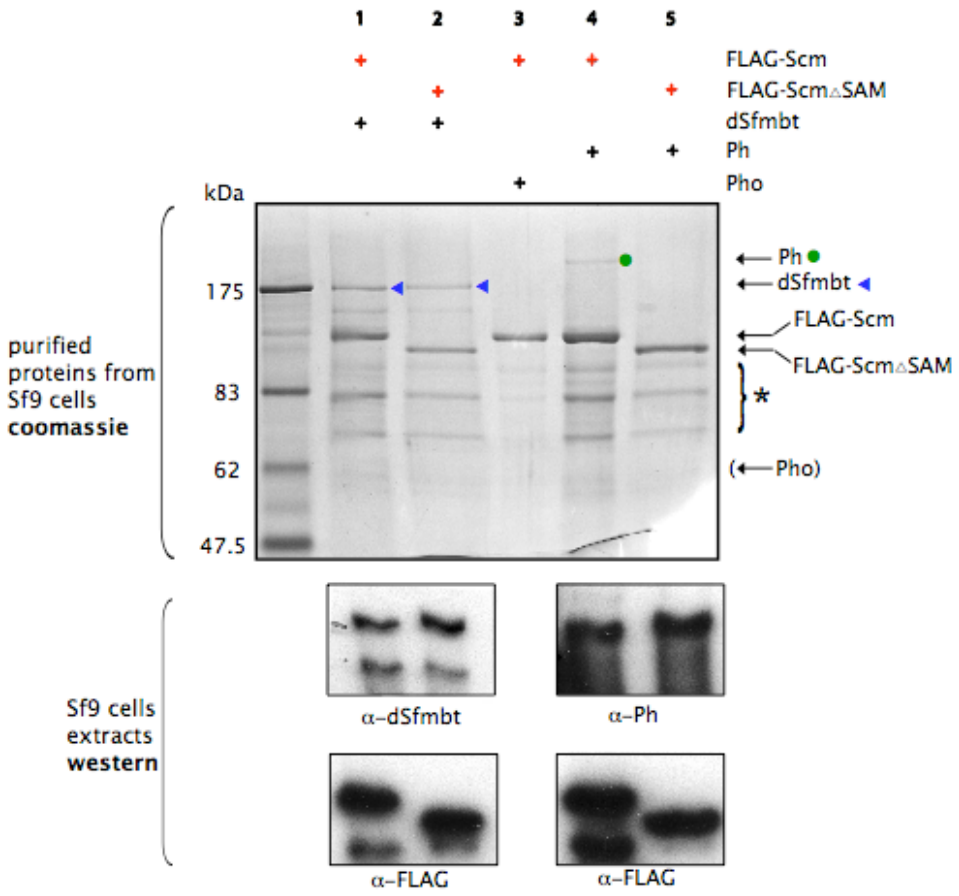

Figure 13 –SAM is not required for dSfmbt-Scm interaction.
FLAG-tagged ScmΔSAM and untagged dSfmbt or Ph full-length proteins were affinity purified via FLAG-tag, separated by SDS-PAGE and visualized by Coomassie staining (top). Immunopurification of FLAG-Scm with dSfmbt (lane 1) or Ph (lane 4) full length are positive controls for the specific interaction of Scm-dSfmbt and Scm-Ph. Immunopurification of FLAG-Scm with Pho full length is a negative control of this interaction (lane 3 and **fig. 12**). Note that dSfmbt (lane 2, blue arrowhead) still forms a stable complex with FLAG-ScmΔSAM, showing that this domain does not mediate the interaction between these two proteins. On the other hand, Ph (lane 5) no longer interacts with FLAG-ScmΔSAM, corroborating previous findings that this domain is responsible for the interaction of these two proteins (*Peterson et al, 1997; 2004*). Western blot analysis shows the expression of dSfmbt, Ph, and the FLAG-tagged Scm constructs in Sf9 cell extracts (below).

The Scm protein forms a complex with Ph (Figure 13, lane 4 green dot; *Peterson et al, 1997*; *2004*), which results from interactions of their respective SAM domains (Figure 13, lane 4 (green dot) and 5; *Peterson et al, 1997*; *2004*). dSfmbt and Scm form

a stable complex (Figure 13, lane 1, blue arrowhead), but the interaction between these two proteins is not mediated by the SAM domain of Scm (Figure 13, lane 2, blue arrowhead). A recent study reported that it is the N-terminal portion (up until the MBT repeats) of dSfmbt and Scm that mediates this interaction (*Grimm et al, 2009*).

- *IN VIVO* CHARACTERIZATION OF THE DSFMBT-SCM COMPLEX IN *DROSOPHILA*

In order to test the function of the dSfmbt – Scm complex and its methyl-lysine binding activity *in vivo*, I removed the *dSfmbt* function in animals that lack wild-type Scm protein and expressed the MBT mutant protein Scm^{D215N} (Scm^{D215N} allele encodes for $Scm^{Su(z)302}$; *Bornemann et al, 1998*). Specifically, clones of *dSfmbt* null mutant cells were induced in Scm^{D215N} mutant *Drosophila* larvae and clones of *dSfmbt* $^{-}Scm^{D215N}$ double mutant cells were analyzed for the misexpression of PcG target genes. In the wing imaginal discs, cell clones lacking *dSfmbt* $^{-}$ showed widespread misexpression of the PcG target gene *Ubx* (Figure 5; *Klymenko et al, 2006*), but they do not showed misexpression of *Abd-B* (Figure 14).

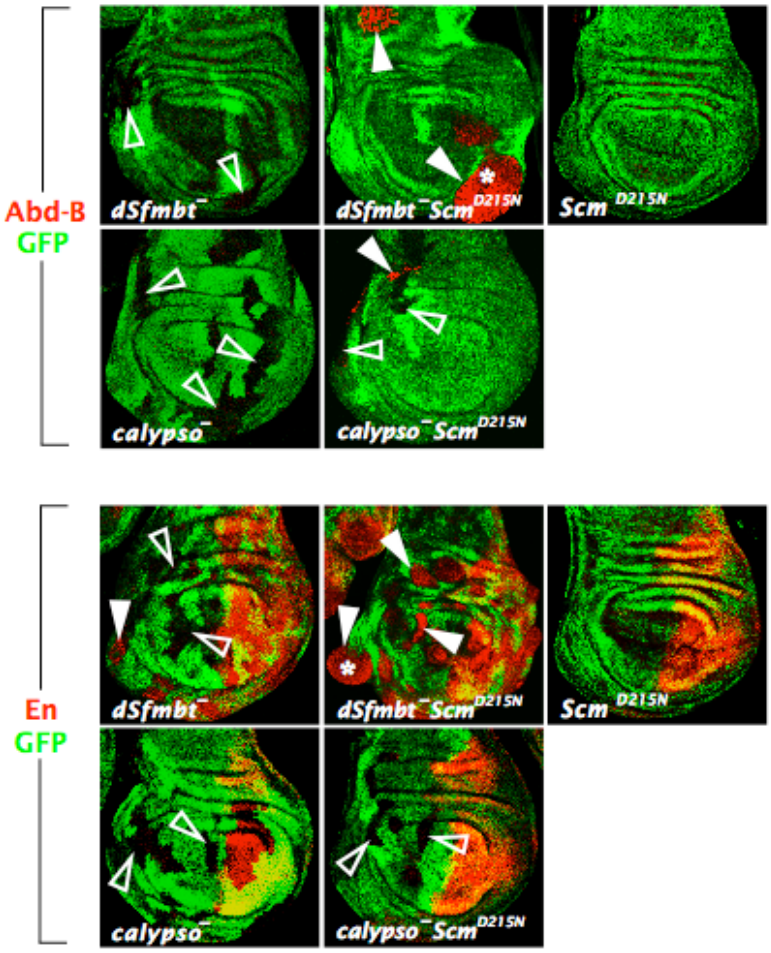

Figure 14 – dSfmbt and Scm interact functionally to maintain the Polycomb repression.
dSfmbt and Scm act redundantly to maintain repression of Polycomb target genes *Abd-B* and *en* in *Drosophila* (*Grimm et al, 2009*). Wing imaginal discs stained with antibodies against Abd-B (red, top) or En protein (red, bottom) as indicated. Left: discs with clones of *dSfmbt* or *calypso* single mutant cells that are marked by the absence of nuclear GFP. Right: discs from Scm^{D215N} mutant larvae; these animals were trans-heterozygous for Scm^{D215N} and the protein null mutation Scm^{D1} (*Bornemann et al, 1998*) and all cells thus express Scm^{D215N} instead of wild-type Scm protein. Nuclear GFP was used here to reveal all nuclei. Middle: Scm^{D215N}/Scm^{D1} mutant discs with clones of *dSfmbt* or *calypso* mutant cells. The *dSfmbt* Scm^{D215N} double mutant and *calypso* Scm^{D215N} double mutant cells are GFP-negative. Abd-B is not expressed in wild-type wing discs and remains repressed in *dSfmbt* or *calypso* single mutant cells (left, empty arrowheads) or in Scm^{D215N} mutant discs (right), but is strongly misexpressed in *dSfmbt* Scm^{D215N} double mutant cells (middle, arrowheads). In clones of *calypso* Scm^{D215N} double mutant cells (middle), Abd-B is misexpressed in a small fraction of clone cells (arrowhead), but remains repressed in the majority of clone cells (empty arrowheads). En expression is confined to posterior compartment cells of wild-type imaginal discs and this pattern is unchanged in Scm^{D215N} mutant discs (right). En remains repressed in *dSfmbt* or in *calypso* single mutant clones in the anterior compartment (left, empty arrowheads) with the exception of some *dSfmbt* mutant clones in the hinge that shows misexpression of En (filled arrowhead). Note that En in strongly misexpressed in almost all *dSfmbt* Scm^{D215N} double mutant clones in the anterior compartment (middle, arrowhead), but remains repressed in *calypso* Scm^{D215N} double mutant clones. Also note that only *dSfmbt* Scm^{D215N} but not *calypso* Scm^{D215N} double mutant clones show the tumour-like phenotype (asterisks).

Similarly, *Abd-B* is not misexpressed in wing imaginal discs of Scm^{D215N} mutant animals (Figure 14). In contrast, *Abd-B* is strongly misexpressed in clones of *dSfmbt* $^{-}$ Scm^{D215N} double mutant cells (Figure 14). A similar high synergy between these two PcG repressor proteins is observed with the *en* gene. In imaginal discs with *dSfmbt* $^{-}$ single mutant clones, *en* is only misexpressed in a subset of clones in specific regions of the disc, but remains repressed in other parts of the disc and it is not misexpressed in Scm^{D215N} single mutants. In contrast, *en* is strongly misexpressed in clones of *dSfmbt* $^{-}$ Scm^{D215N} double mutant cells (Figure 14). In addition, *dSfmbt* $^{-}$ Scm^{D215N} double mutant cells clones show a tumour-like phenotype that is characterized by unrestricted cell proliferation (Figure 14; asterisks). This phenotype is not observed in either of the single mutants (Figure 14) and is characteristic of cell clones lacking the PRC1 components Psc-Su(z)2 or Ph (*Oktaba et al, 2008*).

To test whether this strong genetic interaction between dSfmbt and Scm was specific, the same clonal analysis strategy was used to remove the function of the PcG gene *calypso* (*Gaytan de Ayala Alonso et al, 2007*) in Scm^{D215N} mutant *Drosophila* larvae. As observed for *dSfmbt* $^{-}$, clones of *calypso* $^{-}$ single mutant cells in the wing imaginal disc showed misexpression of *Ubx* (*Gaytan de Ayala Alonso et al, 2007*), but maintained repression of *Abd-B* and *en* (Figure 14). In clones of *calypso* $^{-}$ Scm^{D215N} double mutant cells, *en* remained fully repressed and the clones did not show the tumour-like phenotype observed in *dSfmbt* $^{-}$ Scm^{D215N} mutant clones (Figure 14). *Abd-B* became misexpressed in a fraction of *calypso* $^{-}$ Scm^{D215N} clone cells, but misexpression was much less extensive that in *dSfmbt* $^{-}$ Scm^{D215N} mutant clones (Figure 14). Removal of *dSfmbt* function in Scm^{D215N} mutant animals therefore resulted in much more severe Polycomb phenotypes compared to removing *calypso* in this genetic background. Taken together these results suggest a particularly strong synergy between dSfmbt and the PRC1 component Scm in the repression of target genes and in the control of cell proliferation.

4. CHARACTERIZATION OF THE TRXG PROTEIN, ASH1

The HMTase protein Ash1 has been reported to methylate H3K4, H3K9, H3K36 and H4K20 (*Beisel et al, 2002*; *Byrd & Shearn, 2003*; *Tanaka et al, 2007*). The Ash1 protein is found specifically 8 Kb downstream of the transcription start site of the *Ubx* gene in *Drosophila* larval tissues where this gene is active (*Papp & Müller, 2006*). In tissues where *Ubx* is active, the H3K4me3 modification correlates specifically with the presence of Ash1 and the H3K9me3 modification is absent from the body of the active *Ubx* gene (*Papp & Müller, 2006*). When *ash1* mutants were analysed, the histone modification H3K4me3 was absent and there was an extention of the H3K9me3 modification into the body of the *Ubx* gene in tissues where this gene should be active (*Papp & Müller, 2006*).

In order to understand the specificity of the HMTase protein Ash1, I performed HMTase assays in *Xenopus* wild type mononucleosomes or octamers *in vitro*. I used a portion of the Ash1 protein where only the Pre-SET, SET and Post-SET domains were present ($Ash1^{1032-1619}$; Figure 15A). This protein was expressed with a C-terminal FLAG tag in the Sf9 cells using the baculovirus system, followed by FLAG affinity purification (Figure 15B). The E(z)/Esc/Su(z)12 trimeric PRC2 complex was used as a positive control (*Nekrasov et al, 2005*). $Ash1^{1032-1619}$ or the trimeric complex was incubated with octamers or mononucleosomes and the incubated material was separated using a SDS polyacrylamide gel, and subsequently stained with Coomassie.

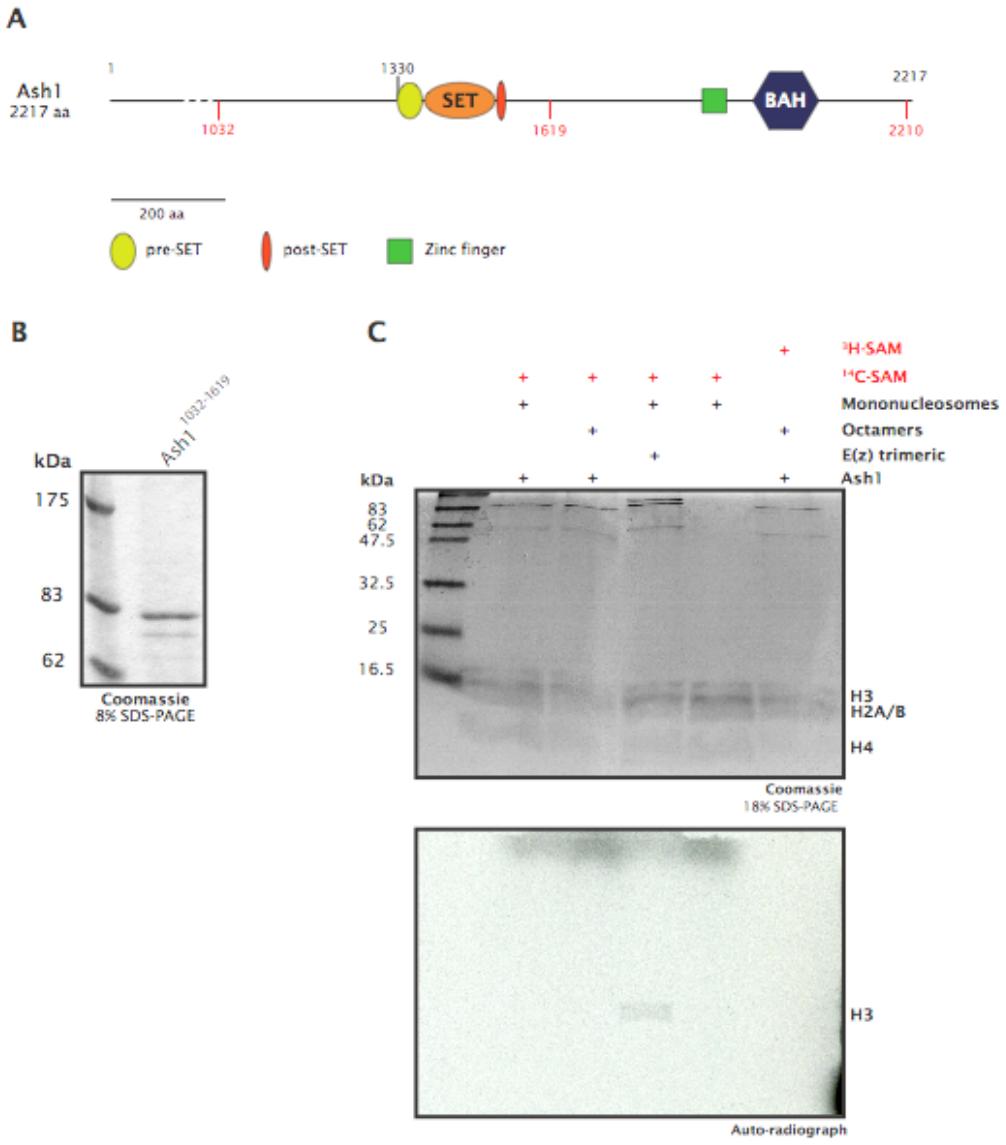

Figure 15 – Ash1 does not methylate nucleosomes or octamers *in vitro*.

A – Scheme of the Ash1 protein and its domains, SET (**S**u(var)3-9, **E**nhancer of zeste, **T**rithorax) domain, pre and post SET domains, Zinc finger and BAH (**B**romo **A**djacent **H**omology) domain. Numbers in red are aminoacids of Ash1 protein and represent the portion of the protein that was expressed for the HMTase assay; Ash1$^{1032\text{-}2210}$ could not be recovered from FLAG purification in sufficient amounts to perform HMTase assays. **B** – Ash1$^{1032\text{-}1619}$ purified protein. **C** – HMTase assay of Ash1$^{1032\text{-}1619}$ using *Xenopus* mononucleosomes or octamers; HMTase assay was performed with ^{3}H and ^{14}C; E(z) trimeric complex was used as a positive control for methylation (*Nekrasov et al, 2005*). Above: SDS-PAGE where the incubated material (above the gel) was separated. Below: Autoradiograph of the SDS-PAGE showing the HMTase activity of the PRC2 trimeric complex, but not of the Ash1 protein.

The Ash1$^{1032-1619}$ protein shows no HMTase activity either in mononucleosomes or in octamers, while the PRC2 trimeric complex shows methylation of histone H3 (Figure 15C; *Nekrasov et al, 2005*). This result does not corroborate with previously published Ash1 protein HMTase activity (*Beisel et al, 2002*; *Byrd & Shearn, 2003*; *Tanaka et al, 2007*). In these studies, several portions of the Ash1 protein were analysed in HMTase assays and a similar construct as the one used for this assay (Figure 15A and B) showed HMTase activity (*Beisel et al, 2002*). However, the type of tag and the system used here for expressing Ash1 protein was different from the previous reports (Sf9 cells in this study; *E. coli* in *Beisel et al, 2002*; HeLa cells in *Tanaka et al, 2007*). This could explain the different results, as the Ash1 protein might need some post-translational modifications and/or co-factors that are not available in cultured Sf9 cells.

DISCUSSION AND CONCLUSION

In this section I will discuss the potential *in vivo* function of the two complexes identified in this thesis work: the dSfmbt-Scm complex and the dSfmbt complex in *Drosophila* embryos. The functional relevance of Rpd3, HP1b, and NAP1 proteins in the dSfmbt complex will also be discussed. A model for the recruitment of the different PcG complexes will be proposed.

1. dSFMBT COMPLEX VERSUS dSFMBT-SCM COMPLEX

dSfmbt and Scm proteins colocalize at the PREs of PcG target genes (Figure 10) and form a stable complex in Sf9 cells (Figure 12). Although the Scm and Ph proteins interact through their SAM domains (Figure 13; *Peterson et al, 1997*; *2004*), these domains are not responsible for the dSfmbt-Scm complex (Figure 13; *Grimm et al, 2009*). dSfmbt and Scm show similar function as methyl-lysine binding proteins *in vitro* (*Klymenko et al, 2006*; *Grimm et al, 2007*; *2009*). *In vivo*, these proteins do not only interact, but also show redundancy in their methyl-lysine binding activity (Figure 14; *Grimm et al, 2009*).

As Scm protein was not found as a subunit of the dSfmbt complex (Table 3), the interaction between these two proteins possibly occurs later in development or occurs transiently. This would explain why Scm did not copurify with dSfmbt in the TAP purification. Previous studies have shown that the Pcl protein is not a member of the PcG complex PRC2 in *Drosophila* larvae (*Savla et al, 2008*), being present in PRC2 only during embryogenesis for the stablishment of high levels of H3K27me3 (*Nekrasov et al, 2007*). In larvae, Pcl is thought to be a member of a different complex from PRC1 and PRC2 (*Savla et al, 2008*). Therefore, the Scm-dSfmbt interaction occurs possibly only later during larval development. The Scm-dSfmbt complex may also occur only transiently *in vivo* for a specific function related to their activity as methyl-lysine binding proteins.

According to the previous model of the PcG complexes in chromatin function (reviewed in *Müller & Verrijzer, 2009*), the PhoRC is tethered to the PREs through the DNA binding function of Pho protein and its binding is crucial for the tethering of PRC1 and PRC2 at the PREs (*Wang et al, 2004*; *Mohd-Sarip et al, 2005*; *Klymenko et al, 2006*). PRC2 methylates H3K27 in the surrounding nucleosomes (*Nekrasov et al, 2007*) and the chromo domain of Pc binds to H3K27me3 (*Cao et al, 2002*; *Czermin et al, 2002*; *Fischle*

et al, 2003; *Min et al, 2003*). The targeting and tethering of PRC1 or PRC2 to the PREs was not explained in that model.

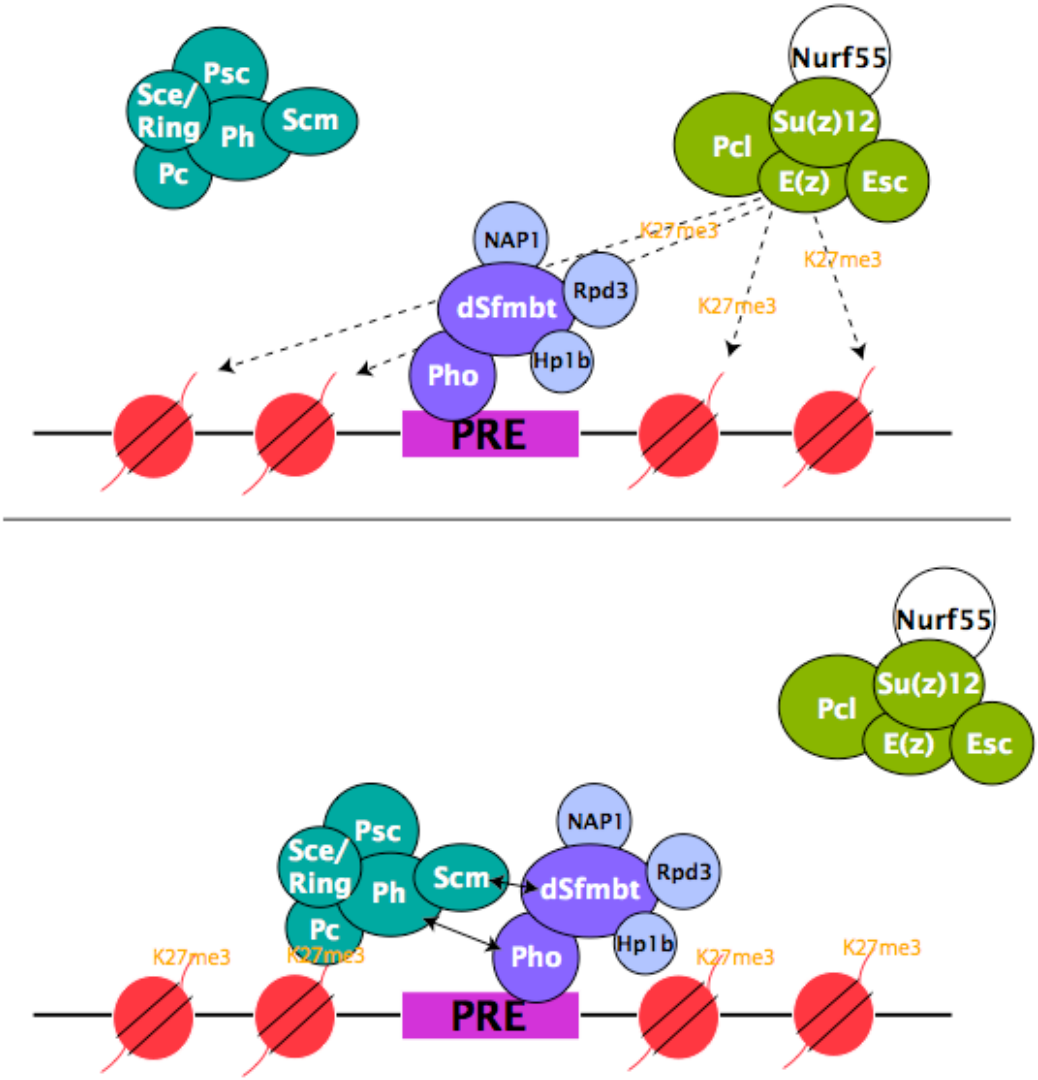

Figure 16 – Tethering of PRC1 to the PREs.
PhoRC and dSfmbt complex are tethered to the PREs through the protein Pho. PRC2 subunit E(z) trimethylates H3K27 (K27me3; up, broken arrows). Through dSfmbt-Scm and/or Pho-Ph interactions (below, double arrows) PRC1is targeted to the PREs where the chromodomain of the Pc protein would bind to H3K27me3 of the PRE surrounding nucleosomes. This binding would assist on the chromatin compaction function of PRC1.

With the identification of the dSfmbt-Scm complex, the tethering of PRC1 to the PREs can be explained by the interaction between these two proteins (*Grimm et al, 2009*) and the interaction between Pho and Ph (*Mohd-Sarip et al, 2005*). The tethering of PRC1 to the PhoRC-PREs would then allow the binding of Pc to H3K27me3 modification

of the surrounding nucleosomes, hence contributing to chromatin compaction (Figure 16; *Francis et al, 2001*; *2004*).

The dSfmbt complex in embryos is constituted by the proteins, Rpd3, NAP1, HP1b, Pho, and CG3363. In larvae, the Scm protein interacts with dSfmbt and this complex is crucial for PcG repression. As well the dSfmbt-Scm complex provides a mechanistic model of recruitment of PRC1 to the PREs (Figure 16).

2. THE NEW PCG COMPLEX – DSFMBT COMPLEX

The dSfmbt complex is composed of the proteins Pho, dSfmbt, Rpd3, HP1b, NAP1, and CG3363 (Figure 6). In this part of the discussion, the occurrence and possible function of Pho, Rpd3, HP1b, and NAP1 in the dSfmbt complex will be discussed.

- PHO

Pho together with dSfmbt form the PhoRC complex, which was purified using Pho as a bait protein in TAP purification (*Klymenko et al, 2006*). This complex was again purified by TAP purification using dSfmbt as a bait protein (Figure 6A), which shows that PhoRC, apart from being very stable, is a complex that in fact occurs *in vivo*, being important for the PcG repression.

- RPD3

The Rpd3 protein has been found to be a member of many protein complexes including as a subunit of PRC2 (*Tie et al, 2001*; *2003*) and PRC1 (*Chang et al, 2001*) in *Drosophila*. Rpd3 has been reported to interact with the Pcl protein in several biochemical assays (*Tie et al, 2001; 2003*), but it was not found when the PRC2 complex was purified using E (z) or Pcl as bait protein in TAP purification (*Nekrasov et al, 2007*). Moreover, in the TAP purification using dSfmbt as a bait protein where Rpd3 was found (Figure 6A and table 3), no subunits of PRC2 were found (Table 3). The use of different experimental procedures may explain the differences between my results and the ones previously reported. The Rpd3-Pcl interaction was found by co-immunoprecipitation with specific antibodies and GST-pulldowns (*Tie et al, 2001; 2003*) whose validation depend on negative controls. Hence, in absence of controls, conclusions must be drawn carefully and additional experiments must be carried to corroborate the interaction.

The TAP purification was designed to purify complexes in their most natural occurrence (*Rigaut et al, 1999*). The purification was made from embryos of transgenic animals using a bait protein that should be assembled in the *in vivo* complex. Two steps of this purification allow the recovery of the most stable complex and the cleaning of the sample from contaminants as well as weak interacting partners (Figure 4). Thus, Rpd3 is not a stable subunit of the PRC2 complex *in vivo*, but is a stable subunit of the newly identified dSfmbt complex. Rpd3 has also been reported to interact with members of the PRC1 complex in S2 cell lines (*Chang et al, 2001*). Although Rpd3 was purified as a member of PRC1, the S2 cell line is a *Drosophila* derivative cell line and it cannot be extrapolated to the whole organism. The PRC1 complex was purified from *Drosophila* embryos using FLAG purification (*Shao et al, 1999; Francis et al, 2001*). In this purification, Rpd3 protein was not found as a member of PRC1 (*Shao et al, 1999; Francis et al, 2001*).

In summary, Rpd3 is not a stable subunit of either PRC2 or the PRC1 complex in *Drosophila*, but is indeed a stable member of the novel dSfmbt complex as it was purified from *Drosophila* embryos using TAP purification (Figure 6 and table 3). As well, it colocalizes together with dSfmbt at the PREs of several PcG target genes (Figure 8).

rpd3 mutants do not show a PcG phenotype (*Mottus et al, 2000*), but Rpd3 has a role in transcriptional repression. In yeast, the role of Rpd3 in transcriptional repression directly depends on its histone deacetylase catalytic activity (*Kadosh & Struhl, 1998*). In *Drosophila*, it has been shown that *Pc*, *Psc*, and *Pcl* mutants show a stronger misexpression of HOX genes when combined with Rpd3 mutation (*Chang et al, 2001; Tie et al, 2001*). Rpd3 has also been shown to interact with the transcriptional repressor Groucho and that its histone deacetylase activity is necessary for the repression by the Groucho protein (*Chen et al, 1999*). The association of Rpd3 with repression is also seen in humans; it has been reported that Rpd3 forms a complex with the Retinoblastoma protein (Rb) to repress E2F target genes (*Zhang et al, 2000*). Hence, Rpd3 is an important component of transcriptional repression in several organisms and, although not a PcG protein, it is a subunit of the dSfmbt complex. Genetic assays will be performed to address the importance of Rpd3 function in the dSfmbt complex and in PcG repression.

- HP1B

HP1b is a *Drosophila* protein found both in euchromatin and in heterochromatin (*Smothers & Henikoff, 2001*). HP1b has a chromo domain and a chromo shadow domain separated by an unstructured portion, the hinge (*Smothers & Henikoff, 2001*). The chromo domain is monomeric and it binds to methylated H3K9 (*Bannister et al, 2001; Jacobs et al, 2001*). The hinge has a nuclear localization signal (NLS), which targets the protein to the nucleus (*Smothers & Henikoff, 2001*). The chromo shadow domain forms a symmetric dimer and is involved in protein-protein interaction (*Brasher et al, 2000*; *Thiru et al, 2004*). Specifically, the chromo shadow domain recognizes a specific motif in proteins, PxVx[M/L/V], being the strongest interaction with the chromo shadow domain using the motif PxVxL (*Thiru et al, 2004*). The third aminoacid of this motif is the most promiscuous of the three as it is the most variable within proteins that interact with the HP1 family (*Thiru et al, 2004*). The PxVxL motif is present in the N terminal portion of the dSfmbt protein (Figure 7). The valine residue (position 0) is required for the specific binding to the HP1b chromo shadow domain, and the residues at positions -2 and +2 confer specificity to the binding (*Thiru et al, 2004*). The interaction between the motif and the chromo shadow domain is strengthened when hydrophobic aminoacids are present at position -6 and -7 (*Thiru et al, 2004*). At these two positions, the dSfmbt protein has a hydrophobic (leucine) and a slightly polar (threonine) residue, which would support the hypothesis of a direct physical interaction between this protein and HP1b (Figures 6B and 7).

A member of the HP1 family in humans, HP1γ interacts with the MBT protein L3MBTL1 (human homologue of the Drosophila L(3)mbt protein) (*Trojer et al, 2007*). L3MBTL1 protein has three variations of the binding motif of the HP1 chromo shadow domain: PxVxH, PxVxG, and VxVxN (*Thiru et al, 2004*). The L3MBTL1-HP1γ interaction is possibly mediated by the VxVxN motif as interaction experiments showed that the minimal portion of the L3MBTL1 protein for the interaction to occur is a portion that contains the VxVxN motif (*Trojer et al, 2007*). So not only HP1b is a stable member of the dSfmbt complex, but also dSfmbt and HP1b possibly interact directly and stably through the PxVxL motif of dSfmbt and the chromo shadow domain of HP1b.

- NAP1

NAP1 is a histone chaperone and it associates preferably to the histone dimer H2A/H2B in *Drosophila melanogaster*, although it has been shown to also interact with the H3/H4 dimer (*Ito et al, 1996*). On the contrary, yNAP1 preferentially binds to the dimer H3/H4 instead of H2A/H2B (*Fujii-Nakata et al, 1992*). Although dNAP1 and yNAP1 have a conserved function of shuttling histones from the cytoplasm to the nucleus, the preferential dimer association is different, which can be explained by the low identity (31%) between these two proteins (*Ito et al, 1996*).

It has been shown that NAP1 is mainly localized at the nucleus during S phase and in the cytoplasm in G2 during cell cycle in the *Drosophila* embryo, in accordance with its function as a histone shuttle (*Ito et al, 1996*). NAP1 has been shown to associate with CAF1, showing also a chromatin assembly function (*Ito et al, 1996*). In the nucleus, NAP1 is also a stable subunit of the dSfmbt complex (Figure 6B and table 3). Therefore, its function may be related to chromatin compaction of silenced genes or may have a yet unknown function in transcriptional repression.

The different associations of NAP1 may explain the result seen on figure 6B, where the relative amount of NAP1 in the input material of dSfmbt-TAP nuclear extract is higher that in the purified material. This would mean that not all the NAP1 present in the input is associated with the dSfmbt complex, which would explain its different functions in the nucleus. As no histones were found in the purification (Table 3) and as there are no nucleosomes at the PREs (*Papp & Müller, 2006*; *Mohd-Sarip et al, 2006*) where the dSfmbt complex assembles (Figure 8), NAP1's function in the dSfmbt complex is possibly not related to its function as a histone chaperone and has a different function in the context of this complex.

The NAP1 ChIP experiments did not show a preferent binding towards the PREs or towards the genomic region (Figure 8, red graph). Assuming that the NAP1 antibody immunoprecipitates NAP1-binding chromatin, then NAP1 protein would be present at all the tested genomic regions (Figure 8, red graph). This would support the hypothesis that NAP1 protein has a different function as a subunit of the dSfmbt complex (presence at the PREs) and as a histone chaperone or a partner of CAF1 (presence at other genomic regions).

NAP1 is an important and crucial protein in *Drosophila* development. Its knock out results in an embryonic lethal phenotype when there is no maternal deposition of NAP1 but, even when there is maternal deposition, very few homozygous flies are fertile and from these most have a low viability and die few days after hatching (*Lankenau et al, 2003*).

In conclusion, NAP1 protein is a stable subunit of dSfmbt complex, but its function in transcriptional repression mediated by PcG has to be further analysed *in vivo*. To this end, experiments for analyzing genetic interaction between dSfmbt and NAP1 by removing the function of dNAP1 in *dSfmbt* mutant clones would be a good way to assess the function of NAP1 in this complex.

3. POSSIBLE FUNCTION OF THE DSFMBT COMPLEX ON CHROMATIN

During *Drosophila* development, the same gene can become active and repressed. Genes like *vestigial* (*Williams et al, 1993*; *Wu & Cohen, 2002*), *teashirt* (*Corè et al, 1997*; *Wu & Cohen, 2002*), and *Homothorax* (*Casares & Mann, 2000*; *Wu & Cohen, 2002*) change their transcriptional state from active to repressed in certain tissues during *Drosophila* development.

The *Ubx* gene shows both trxG and PcG proteins at the PREs and the body of the gene, independently of its active or repressed state (*Papp & Müller, 2006*). The difference between the ON and OFF transcriptional state of the *Ubx* gene is the distribution along the gene of the different histone modifications and the presence of Ash1 and Kismet proteins only in the ON state (*Papp & Müller, 2006*). Therefore, for the transition of a transcriptionally active to a repressed gene to happen, a change in the profile of histone modifications that are present in that gene is necessary. Although for lysine methylation, the positioning of the lysine is important for the transcriptional state that will translate (i.e. H3K9me3, H3K27me3 and H4K20me3 are repressive histone modifications, and H3K4me3 and H3K36me3 are marks that correlate with activation), with acetylation/ deacetylation it is much simpler. Histone acetylation correlates with gene activation and histone deacetylation correlates with transcriptional repression of genes (*Hebbes et al, 1988*; *Braunstein et al, 1993*). Hence, the dSfmbt complex may be important for the switch of a transcriptionally active gene to its repressed state through the histone deacetylase activity of Rpd3 subunit (Figure 17).

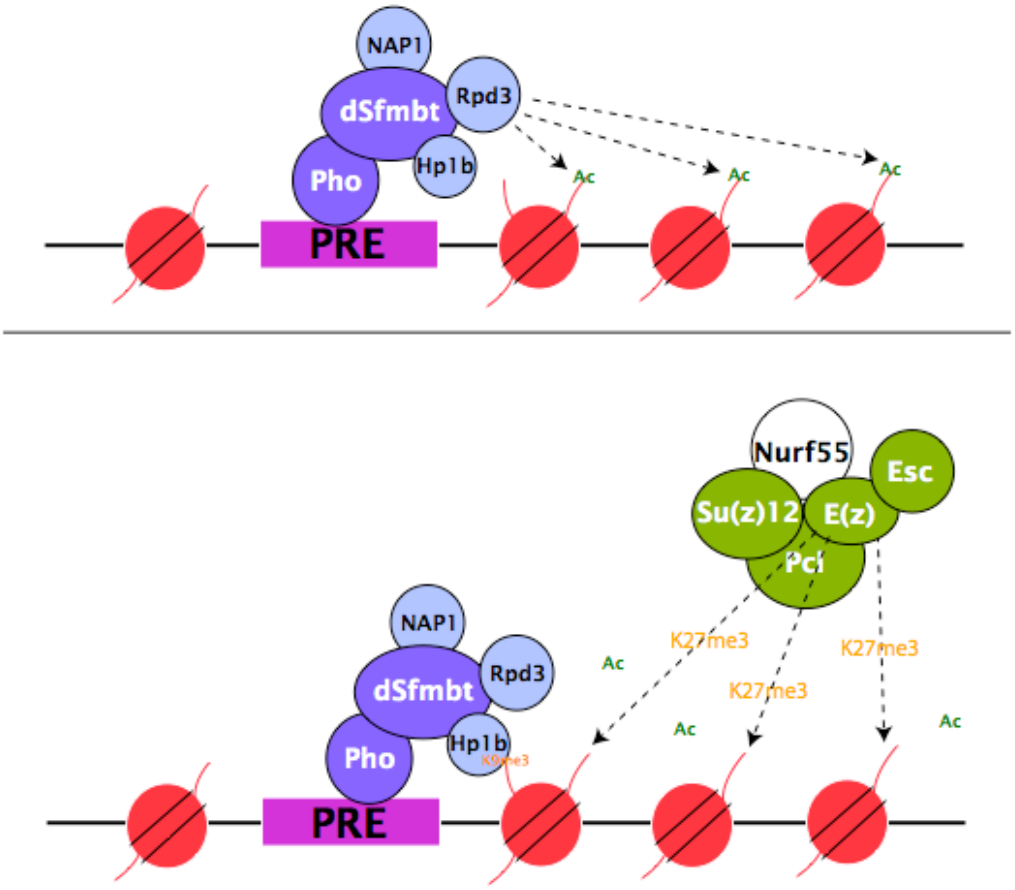

Figure 17 – Model of possible function of the dSfmbt-Rpd3 complex.
The dSfmbt complex binds to the PREs through the protein Pho. The Rpd3 protein deacetylates (acetyl residue (Ac in green)) the histone tails (up, broken arrows). PRC2 subunit E(z) trimethylates H3K27 (K27me3; below, broken arrows) thus, leading to gene repression. Su(var)3-9 or other H3K9 specific HMTases methylate H3K9 and HP1b binds to this modification which may help on chromatin compaction. Finally, through dSfmbt-Scm interactions and/or Pho-Ph interactions, PRC1 is targeted and tethered to the PREs (**Figure 16**).

PhoRC together with the dSfmbt complex is targeted and tethered to the PREs through the DNA binding motif of the Pho protein, both in active and repressed genes (*Klymenko et al, 2006*; *Papp & Müller, 2006*). In active genes, PRE-tethered dSfmbt complex would deacetylate the surrounding nucleosomes by its Rpd3 subunit (Figure 17). The deacetylation would possibly trigger a recruitment of several HMTase enzymes that would deposit histone repressive marks (H3K27me3 by PRC2 complex; H3K9me3 by Su(var)3-9 protein or another H3K9 specific HMTase; reviewed in *Ebert et al, 2006*). Previous studies had shown specific presence of H3K27me3 and H3K9me3 modifications at the repressed HOX gene *Ubx* (*Papp & Müller, 2006*). The chromo domain of HP1b subunit could then bind to the H3K9me modification of surrounding nucleosomes (Figure 17), which would also contribute to chromatin compaction. The

function of NAP1 is yet to be determined, but it could be involved in chromatin compaction together with PRC1 (through the dSfmbt-Scm complex; figure 16) by shuttling histones to the surrounding PRE-chromatin. Further studies need to be performed in order to determine the actual function of this complex *in vivo*.

In conclusion, the dSfmbt complex may be involved in the switch from active genes to their repressional state during *Drosophila* development (Figure 17). The interaction between Scm and the dSfmbt complex would help the tethering of PRC1 to the PREs and the repression of PcG target genes (Figure 16).

REFERENCES

Adamson, A. L., Shearn, A., 1996. Molecular genetic analysis of Drosophila ash2, a member of the trithorax group required for imaginal disc pattern formation. *Genetics*, **144** (2), 621-633.

Amorós, M., Corominas, M., Deák, P., Serras, F., 2002. The ash2 gene is involved in Drosophila wing development. *Int. J. Dev. Biol.*, **46** (3), 321-324.

Angulo, M., Corominas, M., Serras, F., 2004. Activation and repression activities of ash2 in Drosophila wing imaginal discs. *Development*, **131** (20), 4943-4953.

Armstrong, J. A., Papoulas, O., Daubresse, G., Sperling, A. S., Lis, J. T., Scott, M. P., Tamkun, J. W., 2002. The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *EMBO J.*, **21** (19), 5245-5254.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., Kouzarides, T., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410** (6824), 120-124.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., Sauer, F., 2002. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. *Nature*, **419** (6909), 857-862.

Beuchle, D., Struhl, G., Müller, J., 2001. Polycomb group proteins and heritable silencing of Drosophila Hox genes. *Development*, **128** (6), 993-1004.

Biggin, M. D., Bickel,S., Benson, M., Pirrotta, V., Tijan, R., 1988. Zeste encodes a sequence-specific transcription factor that activates the Ultrabithorax promoter in vitro. *Cell*, **53** (5), 713-722.

Birve, A., Sengupta, A. K., Beuchle, D., Larsson, J., Kennison, J. A., Rasmuson-Lestander, A., Müller J., 2001. Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. *Development*, **128**, 3371-3379.

Blastyak, A., Mishra, R., Karch, F., Gyurkovics, H., 2006. Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol. Cell Biol.*, **26** (4), 1434-1444.

Bloyer, S., Cavalli, G., Brock, H. W., Dura, J. M., 2003. Identification and characterization of polyhomeotic PREs and TREs. *Dev. Biol.*, **261** (2), 426-442.

Bornemann, D., Miller, E., Simon, J., 1996. The Drosophila Polycomb group Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development*, **122** (5), 1621-1630.

Bornemann, D., Miller, E., Simon, J., 1998. Expression and properties of wild-type and mutant forms of the Drosophila sex comb on midleg (SCM) repressor protein. *Genetics*, **150** (2), 675-686.

Brasher, S. V., Smith, B. O., Fogh, R. H., Nietlispach, D., Thiru, A., Nielsen, P. R., Broadhurst, R. W., Ball, L. J., Murzina, N. V., Laue, E. D., 2000. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimmer. *EMBO J.*, **19** (7), 1587-1597.

Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D., Broach, J. R., 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes & Dev.*, **7** (4), 592-604.

Brown J. L., Fritsch C., Müller J., Kassis J. A., 2003. The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development*, **130** (2), 285-294.

Brown, J. L., Grau, D. J., DeVido, S. K., Kassis, J. A., 2005. An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. *Nucleic Acids Res.*, **33** (16), 5181-5189.

Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M. L., Kassis, J. A., 1998. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell*, **1**, 1057-1064.

Brunk, B. P., Martin, E. C., Adler, P. N., 1991. Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature*, **353**, 351-353.

Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., Sakonju, S., 2001. The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development*, **128** (11), 2163-2173.

Busturia, A., Wightman, C. D., Sakonju, S., 1997. A silencer is required for maintenance of transcriptional repression throughout Drosophila development. *Development*, **124**, 4343-4350.

Byrd, K. N., Shearn, A., 2003. ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. *Proc. Natl. Acad. Sci. U S A*, **100** (20), 11535-11540.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., Zhang, Y., 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039-1043.

Casares, F., Mann, R. S., 2000. A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identifies in Drosophila. *Development*, **127** (7), 1499-1508.

Celniker, S. E., Sharma, S., Keelan, D. J., Lewis, E. B., 1990. The molecular genetics of the bithorax complex of Drosophila: cis-regulation in the Abdominal-B domain. *EMBO J.*, **9** (13), 4277-4286.

Chan, C. S., Gehring, W. 1971. Determination of blastoderm cells in Drosophila melanogaster. *Proc Natl Acad Sci U S A.* **68** (9), 2217-2221.

Chan, C. S., Rastelli, L., Pirrotta, V., 1994. A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *EMBO J.*, **13** (11), 2553-2564.

Chang, Y. L., Peng, Y. H., Pan, I. C., Sun, D. S., King, B., Huang, D. H., 2001. Essential role of Drosophila Hdac1 in homeotic gene silencing. *Proc. Natl. Acad. Sci. U S A*, **98** (17), 9730-9735.

Chen, G., Fernandez, J., Mische, S., Courey, A. J., 1999. A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development. *Genes & Dev.*, **13** (17), 2218-2230.

Christen, B., Bienz, M., 1994. Imaginal disc silencers form Ultrabithorax: evidence for Polycomb response elements. *Mech. Dev.*, **48** (3), 255-266.

Collins, R. T., Furukawa, T., Tanese, N., Treisman, J. E., 1999. Osa associates with the Brahma chromatin remodelling complex and promotes the activation of some target genes. *EMBO J.*, **18** (24), 7029-7040.

Corè, N., Charroux, B., McCormick, A., Vola, C., Fasano, L., Scott, M. P., Kerridge, S., 1997. Transcriptional regulation of the Drosophila homeotic gene teashirt by the homeodomain protein Fushi tarazu. *Mech. Dev.*, **68** (1-2), 157-172.

Crosby, M. A., Miller, C., Alon, T., Watson, K. L., Verrijzer, C. P., Goldman-Levi, R., Zak, N. B., 1999. The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in Drosophila melanogaster. *Mol. Cell Biol.*, **19** (2), 1159-1170.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., Pirrotta, V., 2002. Drosophila Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell, 111 (2).

Daubresse, G., Deuring, R., Moore, L., Papoulas, O., Zakrajsek, I., Waldrip, W. R., Scott, M. P., Kennison, J. A., Tamkun, J. W., 1999. The Drosophila kismet gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development*, **126** (6), 1175-1187.

DeCamillis, M., Cheng, N. S., Pierre, D., Brock, H. W., 1992. The polyhomeotic gene of Drosophila encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. *Genes Dev.*, **6**, 223-232.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decouville, M., Locker, D., Cavalli, G., 2005. Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. *Nature*, **434** (7032), 533-538.

Dellino, G. I., Schwartz, Y. B., Farkas, G., McCabe, D., Elgin, S. C., Pirrotta, V., 2004. Polycomb silencing blocks transcription initiation. *Mol. Cell*, **13**, 887-893.

Duncan I., 1986. Control of bithorax complex functions by the segmentation gene fushi tarazu of D. melanogaster. *Cell*, **47** (2), 297-309.

Ebert, A., Lein, S., Schotta, G., Reuter, G., 2006. Histone modification and the control of heterochromatic gene silencing in Drosophila. *Chromosome Res.*, **14** (4), 377-392.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., Karch, F., 1994. The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature*, **371** (6500), 806-808.

Fischle, W., Wang, Y., Allis, D., 2003. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.*, **15**, 172-183.

Francis, N. J., Kingston, R. E. 2001. Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell Bio.* **2**, 409-21.

Francis, N. J., Kingston, R. E., Woodcock, C., 2004. Chromatin Compaction by a Polycomb Group Protein Complex. *Science*, **306**, 1574-1577.

Francis, N., Saurin, A., Shao, Z., Kingston, R., 2001. Reconstitution of a functional core polycomb repressive complex. *Mol. Cell*, **8**, 545-556.

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H. W., Paro, R., 1992. Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. *EMBO J.*, **11**, 2941-2950.

Fraterman, S., Zeiger, U., Khurana, T. S., Wilm, M., Rubinstein, N. A., 2007. Quantitative proteomics profiling of sarcomere associated proteins in limb and extraocular muscle allotypes. *Mol. Cell Proteomics*, **6** (4), 728-737.

Frei, E., Bopp, D., Burri, M., Baumgartner, S., Edstrom, J. E., Noll, M., 1985. Isolation and structural analysis of the extra sex combs gene of Drosophila. *Cold Spring Harb. Symp. Quant. Biol.*, **50**, 127-134.

Fritsch, C., Beuchle, D., Kehle, J., Müller, J., 1999. Transcriptional silencing of homeotic genes by Polycomb Group proteins. *Europ. Dros. Res. Conf.*, **16**, 267.

Fritsch, C., Beuchle, D., Müller, J., 2003. Molecular and genetic analysis of the Polycomb group gene Sex combs extra/ Ring in Drosophila. *Mech. Dev.*, **120** (8), 949-954.

Fuhii-Nakata, T., Ishimi, Y., Okuda, A., Kikuchi, A., 1992. Functional analysis of nucleosome assembly protein, NAP-1. The negatively charged COOH-terminal region is not necessary for the intrinsic assembly activity. *J. Biol. Chem.*, **267** (29), 20980-20986.

Gambetta, M. C., Oktaba, K., Müller, J., 2009. Essential Role of the Glycosyltransferase Sxc/Ogt in Polycomb Repression. *Science*, **325** (5936), 95-96.

Gaytán de Ayala Alonso, A., Gutiérrez, L., Fritsch, C., Papp, B., Beuchle, D., Müller, J., 2007. A genetic screen identifies novel polycomb group genes in Drosophila. *Genetics*, **176** (4), 2099-2108.

Gildea, J. J., Lopez, R., Shearn, A., 2000. A screen for new trithorax group gene identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. *Genetics*, **156** (2), 645-663.

Grimm, C., de Ayala Alonso, A. G., Rybin, V., Steuerwald, U., Ly-Hartig, N., Fischle, W., Müller, J., Müller, C. W., 2007. Structural and functional analyses of methyl-lysine binding by the malignant brain tumor repeat protein Sex comb on midleg. *EMBO Rep.*, **8** (11), 1031-1037.

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Müller, J., Müller, C. W., 2009. Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.*, (published online).

Gutiérrez, L., Zurita, M., Kennison, J. A., Vázquez, M., 2003. The Drosophila trithorax group gene tonally (tna) interacts genetically with the Brahma remodelling complex and encodes an SP-RING finger protein. *Development*, **130** (2), 343-354.

Hadorn E. 1978. Transdetermination. In " The Genetics and Biology of Drosophila", Vol 2c, 556-617. Eds. M. Ashburner and T.R.F. Wright. Academic Press

Hebbes, T. R., Thorne, A. W., Crane-Robinson, C., 1988. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.*, **7** (5), 1395-1402.

Hodgson, J. W., Argiropoulos, B., Brock, H. W., 2001. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol. Cell Biol.*, **21** (14), 4528-4543.

Horard, B., Tatout, C., Poux, S., Pirrotta, V., 2000. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Mol. Cell Biol.*, **20** (9), 3187-3197.

Huang, D. H., Chang, Y. L., Yang, C. C., Pan, I. C., King, B., 2002. pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol. Cell Biol.*, **22** (17), 6261-6271.

Ingham, P. W., 1981. Trithorax: a new homeotic mutation of Drosophila melanogaster. *Wilhelm Roux's Archives*, **190**, 365-369.

Ingham, P. W., 1984. A gene that regulates the bithorax complex differentially in larval and adult cells in Drosophila. *Cell*, **37** (3), 815-823.

Ishimi, Y., Hirosumi, J., Sato, W., Sugasawa, K., Yokota, S., Hanaoka, F., Yamada, M., 1984. Purification and initial characterization of a protein which facilitates assembly of nucleosome-like structure from mammalian cells. *Eur. J. Biochem.*, **142** (3), 431-439.

Ishimi, Y., Kikuchi, A., 1991. Identification and molecular cloning of yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. *J. Biol. Chem.*, **266** (11), 7025-7029.

Ito, T., Bulger, M., Kobayashi, R., Kadonaga, J. T., 1996. Drosophila NAP-1 is a core histone chaperone that functions in a ATP-facilitated assembly of regularly spaced nucleosomal arrays. *Mol. Cell Biol.*, **16** (6), 3112-3124.

Jacobs, S. A., Khorasanizadeh, S., 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, **295** (5562), 2080-2083.

Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D., Khorasanizadeh, S., 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.*, **20** (18), 5232-5241.

Jones, R. S., Gelbart, W. M., 1990. Genetic analysis of the Enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. *Genetics*, **126**, 185-199.

Jürgens, G., 1985. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature*, **316**, 153-155.

Kadosh, D., Struhl, K., 1998. Histone deacetylase activity of Rpd3 is important for transcriptional repression in vivo. *Genes & Dev.*, **12** (6), 797-805.

Kadosh, D., Struhl, K., 1998. Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol. Cell Biol.*, **18** (9), 5121-5127.

Kal, A. J., Mahmoudi, T., Zak, N. B., Verrijzer, C. P., 2000. The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. *Genes & Dev.*, **14** (9), 1058-1071.

Kaufman, T. C., Seeger, M. A., Olsen, G., 1990. Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster. *Adv. Genet.* **27**, 309-362.

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

Kennison, J. A., 2004. Introduction to Trx-G and Pc-G genes. *Methods Enzymol.*, **377**, 61-70.

Kennison, J. A., Russel, M. A., 1987. Dosage-Dependent Modifiers of homoeotic Mutations in Drosophila melanogaster. *Genetics*, **116** (1), 75-86.

Kennison, J. A., Tamkun, J. W., 1988. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc. Natl. Acad. Sci. U S A*, **85** (21), 8136-8140.

Kennison, J., 1993. Transcriptional activation of Drosophila homeotic genes from distant regulatory elements. *Trends Genet.*, **9**, 75-79.

Ketel, C. S., Andresen, E. F., Vargas, M. L., Suh, J., Strome, S., Simon, J. A., 2005. Subunit contributions to histone methyltransferase activities of fly and worm Polycomb group complexes. *Mol. Cell Biol.*, **25**, 6857-6868.

Kim, C. A., Gingery, M., Pilpa, R. M., Bowie, J. U., 2002. The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.*, **9** (6), 453-457.

Klymenko, T., Papp, B., Fischle W., Köcher T., Schelder M., Fritsch C., Wild B., Wilm, M., Müller J., 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & Dev.*, **20**, 1110-1122.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., Reinberg, D., 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Dev.*, **16**, 2893-2905.

Kwon, S. H., Workman, J. L., 2008. The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. *Mol. Cells*, **26** (3), 217-227.

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J. R., White, R., 2008. Stability and dynamics of polycomb target sites in Drosophila development. *PLoS Genet.*, **4** (9), e1000178.

Lankenau, S., Barnickel, T., Marhold, J., Lyko, F., Mechler, B. M., Lankenau, D. H., 2003. Knockout targeting of the Drosophila nap1 gene and examination of DNA repair tracts in the recombination products. *Genetics*, **163** (2), 611-623.

Lee, J. S., Shilatifard, A., 2007. A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat. Res.*, **618** (1-2), 130-134.

Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., Kingston, R. E., 2002. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Cell*, **17**, 6070-6078.

Lewis, E. B., 1978. A gene complex controlling segmentation in Drosophila. *Nature*, **276**, 565-570.

Li, H., Fischle, W., Wang, W., Duncan, E. M., Liang, L., Murakami-Ishibe, S., Allis, C. D., Patel, D. J., 2007. Strutural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell*, **28** (4), 677-691.

Lonie, A., D'Andrea, R., Paro, R., Saint, R., 1994. Molecular characterisation of the Polycomblike gene Drosophila melanogaster, a trans-acting negative regulator of homeotic gene expression. *Development*, **120**, 2629-2636.

Luger, K., Rechsteiner, T. J., Richmond, T. J., 1999. Preparations of nucleosome core particle from recombinant histones. *Methods Enzymol.*, **304**, 3-19.

Mahmoudi, T., Zuijderduijn, L. M., Mohd-Sarip, A., Verrijzer, C. P., 2003. GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. *Nucleic Acids Res.*, **31** (14), 4147-4156.

Mannervik, M., Levine, M., 1999. The Rpd3 histone deacetylase is required for segmentation of the Drosophila embryo. *Proc. Natl. Acad. Sci. U S A*, **96** (12), 6797-6801.

Maurange, C., Paro, R., 2002. A cellular memory module conveys epigenetic inheritance of hedgehog expression during Drosophila wing imaginal disc development. *Genes & Dev.*, **16** (20), 2672-2683.

Mazo, A., Huang, D., Mozer, B., Dawid, I., 1990. The trithorax gene, a trans-acting regulator of the Bithorax complex in Drosophila, encodes a protein with zinc-binding domains. *Proc. Natl. Acad. Sci. U S A*, **87**, 2112-2116.

Mihaly, J., Hogga, I., Gausz, J., Gyurkovics, H., 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*, **124** (9), 1809-1820.

Mihaly, J., Mishra, R. K., Karch, F., 1998. A conserved sequence motif in Polycomb-response elements. *Mol. Cell*, **1** (7), 1065-1066.

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., Shilatifard, A., 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U S A*, **98** (23), 12902-12907.

Milne, T. A., Briggs, S. D., Brock H. W., Martin, M. E., Gibbs, D., Allis, C. D., Hess, J. L., 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell*, **10** (5), 1107-1117.

Min, J., Allali-Hassani, A., Nady, N., Qi, C., Ouyang, H., Liu, Y., MacKenzie, F., Vedadi, M., Arrowsmith, C. H., 2007. L3MBTL1 recognition of mono- and dimethylated histones. *Nat. Struct. Mol. Biol.*, **14** (12), 1229-1230.

Min, J., Zhang, Y., Xu, R. M., 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & Dev.*, **17**, 1823-1828.

Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S. E., 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. *Mol. Cell Biol.*, **21** (4), 1311-1318.

Mohd-Sarip, A., Cléard, F., Mishra, R. K., Karch, F., Verrijzer, C. P., 2005. Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes & Dev.*, **19** (15), 1755-1760.

Mohd-Sarip, A., van der Knaap, J. A., Wyman, C., Kanaar, R., Schedi, P., Verrijzer, C. P., 2006. Architecture of a Polycomb nucleoprotein complex. *Mol. Cell*, **24** (1), 91-100.

Mohd-Sarip, A., Venturini, F., Chalkley, G. E., Verrijzer, C. P., 2002. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Mol. Cell Biol.*, **22** (21), 7473-7483.

Mottus, R., Sobel, R. E., Grigliatti, T. A., 2000. Mutational analysis of a histone deacetylase in Drosophila melanogaster: missense mutations suppress gne silencing associated with position effect variegation. *Genetics*, **154** (2), 657-668.

Mulholland, N. M., king, I. F., Kingston, R. E., 2003. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes & Dev.*, **17** (22), 2741-2746.

Müller J., Verrijzer P., 2009. Biochemical mechanism of gene regulation by polycomb group protein complexes. *Curr. Opin Genet. Dev.* **19** (2), 150-158.

Müller, J., Bienz, M., 1991. Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. *EMBO J.*, **10** (11), 3147-3155.

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., Simon, J. A., 2002. Histone methyltransferase activity of a Drosophila polycomb group repressor complex. *Cell*, **111**, 197-208.

Müller, J., Kassis, J. A., 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.*, **16** (5), 476-484.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., Canaani, E., 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell*, **10** (5), 1119-1128.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H. G., Wilm, M., Müller, J., 2007. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.*, **26** (18), 4078-4088.

Nekrasov, M., Wild, B., Müller, J., 2005. Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.*, **6**, 348-353.

Oktaba, K., Guttiérrez, L., Gagneur, J., Girardot, C., Sengupta A., S., Furlong E., E., M., Müller J., 2008. Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in *Drosophila*. *Dev. Cell*, **15** (6), 877-889.

Orlando, V., Paro, R., 1993. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell*, **75** (6), 1187-1198.

Orlando, V., Strutt, H., Paro, R., 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods*. **11** (2), 205-214.

Pankratz, M. J., Jäckle, H., 1993. Blastoderm segmentation. In "The development of Drosophila melanogaster", 467-516, Eds M. Bate and A. Martinez-Arias. Cold Spring Harbor. Cold Spring Harbor Laboratory Press.

Papoulas, O., Beek, S. J., Moseley, S. L., McCallum, C. M., Sarte, M., Shearn, A., Tamkun, J. W., 1998. The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. *Development*, **125** (20), 3955-3966.

Papp, B., Müller, J., 2006. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes & Dev.*, **20** (15), 2041-2054.

Park, Y. J., Luger, K., 2006. Structure and function of nucleosome assembly proteins. *Biochem. Cell. Biol.*, **84** (4), 540-558.

Paro, R., Hogness, D. S., 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. U S A*, **88**, 263-267.

Peterson, A. J., Kyba, M., Bornemann, D., Morgan, K., Brock H. W., Simon, J., 1997. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol. Cell Biol.*, **17** (11), 6683-6692.

Peterson, A. J., Mallin, D. R., Francis, N. J., Ketel, C. S., Stamm, J., Voeller, R. K., Kingston, R. E., Simon, J. A., 2004. Requirement for sex comb on midleg protein interaction in Drosophila polycomb group repression. *Genetics*, **167** (3), 1225-1239.

Petruk, S., Sedkov, Y., Smith, S., Tillib, S., Kraevski, V., Nakamura, T., Canaani, E., Croce, C. M., Mazo, A., 2001. Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. *Science*, **294** (5545), 1331-1334.

Rastelli, L., Chan, C. S., Pirrotta, V., 1993. Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. *EMBO J.*, **12** (4), 1513-1522.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting C. P., Allis, C. D., Jenuwein, T., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, **406** (6796), 593-599.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Séraphin, B., 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, **17** (10), 1030-1032.

Ringrose, L., Paro, R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet.* **38**, 413-443.

Ringrose, L., Rehmsmeier, M., Dura, J. M., Paro, R., 2003. Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. *Dev. Cell*, **5** (5), 759-771.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., Kouzarides, T., 2002. Active genes are tri-methylated at K4 of histone H3. *Nature*, **419** (6905), 407-411.

Saurin, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P., Kingston, R. E., 2001. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature*, **412**, 655-660.

Savla, U., Benes, J., Zhang, J., Jones, R. S., 2008. Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development*, **135** (5), 813-817.

Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X. Y., Bourgon, R., Biggin, M., Pirrotta, V., 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat. Genet.*, **38** (6), 700-705.

Schwendemann, A., Lehmann, M., 2002. Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc. Natl. Acad. Sci. U S A*, **99** (20), 12883-12888.

Sengupta, A. K., Kuhrs, A., Müller, J., 2004. General transcriptional silencing by a Polycomb response element in Drosophila. *Development*, **131** (9), 1959-1965.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., Kingston, R. E., 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*, **98**, 37-46.

Shearn, A., 1989. The ash-1, ash-2 and trithorax genes of Drosophila melanogaster are functionally related. *Genetics*, **121** (3), 517-525.

Shevchenko, A., Wilm, M., Vorm, O., Mann, M., 1996. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.*, **68**, 850-858.

Shimell, M. J., Peterson, A. J., Burr, J., Simon, J. A., O'Connor, M. B., 2000. Functional analysis of repressor binding sites in the iab-2 regulatory region of the abdominal-A homeotic gene. *Dev. Biol.*, **218** (1), 38-52.

Simon, J. A., Tamkun, J. W., 2002. Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr. Opin. Genet. Dev.*, **12** (2), 210-218.

Simon, J., Chiang, A., Bender, W., Shimell, M. J., O'Connor, M., 1993. Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.*, **158** (1), 131-144.

Sinclair, D. A. R., Milne, T. A., Hodgson, J. W., Shellard, J., Salinas, C. A., Kyba, M., Randazzo, F., Brock, H. W., 1998. The *Additional sex combs* gene of *Drosophila* encodes a chromatin protein that binds to shared and unique Polycomb group sites on polytene chromosomes. *Development*, **125**, 1207-1216.

Smith, S. T., Petruk, S., Sedkov, Y., Cho, E., Tillib, S., Canaani, E., Mazo, A., 2004. Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. *Nat. Cell Biol.*, **6** (2), 162-167.

Smothers, J. F., Henikoff, S., 2000. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol.*, **10** (1), 27-30.

Smothers, J. F., Henikoff, S., 2001. The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. *Mol. Cell Biol.*, **21** (7), 2555-2569.

St Johnston, D., Nüsslein-Volhard, C., 1992. The origin of pattern and polarity in the Drosophila embryo. *Cell*, **68**, 201-219.

Steward, M. M., Lee, J. S., O'Donovan, A., Wyatt, M., Bernstein, B. E., Shilatifard, A., 2006. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit MLL complexes. *Nat. Struct. Mol. Biol.*, **13** (9), 852-854.

Strutt, H., Cavalli, G., Paro, R., 1997. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.*, **16** (12), 3621-3632.

Strutt, H., Paro, R., 1997. The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. *Mol. Cell Biol.*, **17** (12), 67773-6783.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., Kennison, J. A., 1992. brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell*, **68** (2), 561-572.

Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D., Kitajima, S., 2007. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*, **397** (1-2), 161-168.

Thanos, C D., Goodwill, K. E., Bowie, J. U., 1999. Oligomeric Structure of the Human EphB2 receptor SAM Domain. *Science*, **283** (5403), 833-836.

Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., Widom, J., 1999. Sequence motifs and free energies of select natural and non-natural nucleosome posiotining DNA sequences. *J. Mol. Biol.*, **288** (2), 213-229.

Thiagalingam, S., Cheng, K. H., Lee, H. J., Mineva, N., Thiagalingam, A., Ponte, J. F., 2003. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann. N. Y. Acad. Sci.*, **983**, 84-100.

Thiru, A., Nietlispach, D., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V., Laue, E. D., 2004. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J.*, **23** (3), 489-499.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., Harte, P. J., 2001. The Drosophila Polycomb Group proteins ESC and E(z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development*, **128** (2), 275-286.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A., Harte, P. J., 2003. A 1-megadalton ESC/E(z) complex from Drosophila that contains polycomblike and RPD3. *Mol. Cell Biol.*, **23** (9), 3352-3362.

Treisman, J., 2001. Drosophila homologues of the transcriptional coactivation complex subunits TRAP240 and TRAP230 are required for identical processes in eye-antennal disc development. *Development*, **128** (4), 603-615.

Tripoulas, N. A., Hersperger, E., La Jeunesse, D., Shearn, A., 1994. Molecular genetic analysis of the Drosophila melanogaster gene absent, small or homeotic discs 1 (ash1). *Genetics*, **137**, 1027-1038.

Trojer, P., Li, G., Sims, R. J. 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S. D., Wang, Y. H., Reinberg, D., 2007. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell*, **129** (5), 915-928.

Vázquez, M., Moore, L., Kennison, J. A., 1999. The trithorax group gene osa encodes an ARID-domain protein that genetically interacts with the brahma chromatin-remodeling factor to regulate transcription. *Development*, **126** (4), 733-742.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., Zhang, Y., 2004. Role of histone H2A ubiquitination in polycomb silencing. *Nature*, **431**, 873-878.

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., Jones, R. S., 2004. Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell*, **14** (5), 637-646.

Wang, L., Jahren N., Vargas M. L., Andersen E. F., Benes J., Zhang J., Miller E. L., Jones R. S., Simon J. A., 2006. Alternative ESC and ESC-like subunits of a Polycomb Group histone methyltransferase complex are differentially deployed during Drosophila development. *Mol. Cell Biol.*, **26**, 2637-2647.

White, R. A., Wilcox, M., 1984. Protein products of the bithorax complex in Drosophila. *Cell*, **39** (1), 163-171.

Williams, J. A., Paddock, S. W., Carroll, S. B., 1993. Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions. *Development*, **117** (2), 571-584.

Wu, J., Cohen, S. M., 2002. Repression of Teashirt marks the initiation of wing development. *Development*, **129** (10), 2411-2418.

Yokoyama, A.,Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., Cleary, M. L., 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell Biol.*, **24** (13), 5639-5649.

Zhang, H. S., Gavin, M., Dahiya, A., Postigo, A. A., Ma, D., Luo, R. X., Harbour, J. W., Dean, D. C., 2000. Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF. *Cell*, **101** (1), 79-89.

Zhang, H., Christoforou, A., Aravind, L., Emmons, S. W., Van den Heuvel, S., Haber, D. A., 2004. The C. elegans Polycomb gene SOP-2 encodes an RNA binding protein. *Mol. Cell*, **14** (6), 841-847.