

Dissertation zur Erlangung des Doktores

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

**Molecular basis for targeting PRC1 to
Polycomb response elements**



vorgelegt von

Felice Louisa Frey, MSc

München

15.10.2014

1. Gutachter: Prof. Dr. Heinrich Leonhardt

2. Gutachter: Prof. Dr. Nicolas Gompel

Betreuer: Dr. Jürg Müller

Tag der mündlichen Prüfung: 10.03.2015

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

- dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
- dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
- dass ich mich mit Erfolg der Doktorprüfung im Hauptfach und in den Nebenfächern bei der Fakultät für der

(Hochschule/Universität)

unterzogen habe.

- dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den.....

(Unterschrift)

*) Nichtzutreffendes streichen

Abstract

Proteins of the Polycomb group (PcG) are a class of transcriptional regulators that are essential for the development of animals and plants. PcG proteins act as repressors that are critical for the long-term silencing of genes in cells where these genes should be kept inactive. PcG repression regulates processes ranging from the control of cell fate determination to X chromosome inactivation in female mammals and vernalization in plants. At the molecular level, PcG proteins function in four distinct multiprotein complexes that contain specific enzymatic and nucleosome-binding activities by which they modify and interact with the chromatin of target genes. Polycomb proteins bind to their target genes at so-called Polycomb response elements (PREs), but the molecular interactions by which they are targeted to and anchored at these elements, is not well understood.

Here, I investigated the underlying mechanism of PcG protein complex targeting in *Drosophila*. I purified the PcG protein Polyhomeotic (Ph), a subunit of Polycomb repressive complex 1 (PRC1), from *Drosophila* embryos. This resulted in the isolation of a protein assembly that comprises the known PRC1 subunits Psc (Posterior sex combs), Sce/ Ring (Sex combs extra), Pc (Polycomb) and Scm (Sex comb on midleg) and in addition the subunits of the DNA-binding PcG complex, Pho (Pleiohomeotic) repressive complex (PhoRC). The direct physical association between PRC1 and PhoRC was confirmed *in vitro* and I discovered that Scm mediates the interaction between Ph and the PhoRC subunit Sfmpt (Scm-like protein with four MBT domains). I found that Scm and Sfmpt interact via two distinct surfaces: an N-terminal portion of Sfmpt binds to a central portion of Scm, and the two proteins also bind to each other via their C-terminally located SAM (sterile alpha motif) domains. Previous studies showed that the Scm-SAM domain also binds to the SAM domain of Ph. Ph- and Scm-SAM domains contain two interaction surfaces, a mid loop surface (ML) and an end helix surface (EH). These two interaction surfaces were previously shown to mediate homo- and heteropolymerisation of Ph- and Scm-SAM domains through head-to-tail interactions. Here, I determined the crystal structure of the Scm-SAM/ Sfmpt-SAM

complex. The structure revealed that the ML surface of Scm binds to the EH surface of Sfmbl and that the Sfmbl-SAM domain contains no functional ML surface. The Sfmbl-SAM domain is thus unable to polymerize, but being DNA-tethered in PhoRC, could act as a polymerization seed for DNA-anchoring and oligomerization of PRC1 assemblies.

The work reported in this thesis thus reveals the structural basis for how two different PcG complexes interact with each other and it contributes to our understanding how the PRC1 complex is tethered to Polycomb response elements (PREs) through interaction with the DNA-binding PhoRC complex.

Zusammenfassung

Proteine der Polycomb Gruppe (PcG) sind eine Art von transkriptionalen Regulatoren die eine Schlüsselrolle in der Entwicklung von Tieren und Pflanzen spielen. PcG Proteine agieren als transkriptionale Repressoren, welche eine entscheidende Rolle in der Langzeit Gen-Stilllegung spielen. Sie reprimieren die Expression von Genen in Zellen, in welchen diese nicht exprimiert werden sollen. PcG Proteine regulieren eine Vielzahl an Prozessen, wie z.B. die Zelldifferenzierung, die Inaktivierung des X-Chromosoms in weiblichen Säugetieren bis hin zu der Blütezeit bei Pflanzen. Auf der molekularen Ebene agieren PcG Proteine in vier bestimmten Multiprotein-Komplexen und enthalten spezifische enzymatische und Nukleosom-bindende Aktivitäten mit denen sie mit dem Chromatin von Zielgenen interagieren und es auch modifizieren können. Polycomb Proteine binden an ihre Ziel-Gene an sogenannten Polycomb Response Elementen (PREs). Trotz intensiver Forschung ist bis heute nicht verstanden, wie genau Polycomb Protein Komplexe zu ihren Ziel-Genen rekrutiert und dort gebunden werden.

In dieser Studie untersuchte ich den Mechanismus von Polycomb Komplex Rekrutierung in *Drosophila*. Dafür reinigte ich das PcG Protein Polyhomeotic (Ph) von *Drosophila* Embryonen auf. Ph ist eine der Untereinheiten des Polycomb Komplexes, „Polycomb repressive complex 1“ (PRC1). Dies führte zur Isolierung aller beschriebener PRC1 Komplex Untereinheiten: Psc (Posterior sex combs), Sce/Ring (Sex combs extra), Pc (Polycomb) und Scm (Sex comb on midleg). Außerdem detektierte ich Sfmbt (Scm-like protein with four MBT domains), eine Untereinheit des DNA-bindenden Polycomb Komplexes „Pleiohomeotic repressive complex“ (PhoRC). Eine direkte physische Assoziierung zwischen dem PRC1 Komplex und dem DNA-bindenden PhoRC Komplex wurde durch *in vitro* Experimente bestätigt und es stellte sich heraus, dass die Interaktion zwischen Ph und der PhoRC Untereinheit Sfmbt über Scm stattfindet. Ich entdeckte zwei individuelle Interaktionsflächen zwischen Scm und Sfmbt: Der N-terminus von Sfmbt interagiert mit einem zentralen Teil des Scm Proteins und die

beiden Proteine können ebenfalls über ihre C-terminale SAM (sterile alpha motif) Domänen interagieren.

Vorige Studien hatten bereits gezeigt, dass die Scm-SAM Domäne an die SAM Domäne von Ph binden kann. Die Ph- und die Scm-SAM Domänen enthalten je zwei Interaktionsflächen, eine „Mid Loop Fläche“ (ML) und ein „End Helix Fläche“ (EH). Die Homopolymerisierung und Heteropolymerisierung von SAM Domänen findet mittels dieser beiden Flächen statt. Hier, bestimmte ich die Kristallstruktur des Scm-SAM/ Sfmbt-SAM Komplexes und es zeigte sich, dass die ML Fläche von Scm-SAM an die EH Fläche von Sfmbt bindet und, dass die Sfmbt-SAM Domäne keine funktionelle ML Fläche besitzt. Deswegen kann die Sfmbt-SAM Domäne keine Polymere bilden. Da die Sfmbt-SAM Domäne jedoch über den PhoRC Komplex an DNA gebunden ist, könnte sie als Polymerisationskeim für die DNA-Verankerung und Oligomerisierung von PRC1 Komplexen dienen.

Diese Studie enthüllt die erste strukturelle Basis für die Interaktion zweier unterschiedlicher PcG Protein Komplexe und erweitert unser Verständnis, wie PRC1 Komplexe durch die Interaktion mit dem DNA-bindenden PhoRC Komplex an Polycomb Response Elemente (PREs) gebunden werden.

Table of Content

1. Introduction	4
1.1 The Polycomb and Trithorax system	4
1.2 Polycomb and Trithorax group protein complexes	6
1.2.1 PRC1 and PRC1-like complexes	7
1.2.2 PRC2 complexes.....	10
1.2.3 PhoRC complexes	12
1.2.4 PR-DUB complex and PcG proteins not associated with complexes.....	13
1.2.5 Trx group protein complexes.....	15
1.2.6 Conservation of PcG proteins in animals and plants	18
1.3 Domain architecture of Polycomb group protein complexes	19
1.4 Biological Importance of the PcG system	23
1.4.1 Role of PcG proteins in stem cell maintenance and differentiation	23
1.4.2 Role of the PcG system in tumorigenesis	25
1.5 Recruitment of PcG protein complexes to target genes	26
1.5.1 Polycomb response elements (PREs).....	26
1.5.2 PcG target genes.....	28
1.5.3 Recruitment of PcG proteins to PREs at target genes.....	30
1.5.4 Different PcG recruitment mechanisms in fly and mammals	32
1.6 The Polycomb group protein Polyhomeotic	34
1.6.1 Molecular Function of Polyhomeotic.....	36
1.6.2 Interacting partners of Polyhomeotic	37
1.7 Aims of this study	38

2. Material and Methods	40
2.1. Materials	40
2.1.1 Oligonucleotides used in this study.....	40
2.1.2 Vectors used in this study.....	45
2.1.3 Fly strains used in this study.....	46
2.1.4 Baculoviruses generated and used in this study	47
2.1.5 Antibodies used in this study	47
2.2 Methods	49
2.2.1 Fly husbandry and microscopic analysis of <i>Drosophila</i> tissues	49
2.2.2 Insect cell culture and Baculoviruses	51
2.2.3 Work with bacteria.....	54
2.2.4 Protein biochemistry	55
2.2.5 DNA techniques and cloning	67
3. Results.....	70
3.1 Ph interacts with PRC1 members and the PhoRC complex.....	70
3.2 PhoRC member Sfm bt and PRC1 members Ph and Scm form a stable trimeric complex	
<i>in vitro</i>	78
3.3 Two independent interaction sites are important for Scm:Sfm bt interaction.....	84
3.3.1 Scm:Sfm bt interaction is not dependent on FCS zinc finger domains.....	86
3.3.2 Putative domains involved in Scm:Sfm bt interaction	89
3.3.3 Scm and Sfm bt also interact via their SAM domains	92
3.3.4 The Scm and Sfm bt SAM domains form a stable complex.....	94
3.3.5 Structure of the Scm-SAM/ Sfm bt-SAM complex	96

4. Discussion	103
4.1 Ph interacts with PRC1 members and the PhoRC complex.....	103
4.2 Interaction mapping of PRC1:PhoRC interaction	105
4.3 Role of SAM-containing PcG proteins in PRC1 targeting.....	107
4.3 Direct PRC1 recruitment via PhoRC complex.....	110
4.4 Conservation of PRC1 recruitment mechanism?.....	113
Conclusion	117
5. References	118
6. Appendix	135
6.1 List of Abbreviations	135
6.2 List of Figures	136
6.3 List of Tables	137
6.4 Supplementary material.....	138
Curriculum Vitae	Error! Bookmark not defined.
Acknowledgements	154

1. Introduction

During the development of multicellular organisms a single cell, the zygote, gives rise to many different tissues that form the embryo and later the adult body. While all cells emerging from the single zygote have the same genotype, they vary significantly in their phenotypes. This specialization process is achieved by differential gene expression patterns that are established early in embryogenesis by transient factors. After the initial cues have decayed the chosen cell fates have to be maintained during subsequent cell divisions by a “cellular memory” system. Such a memory system is crucial to maintain cell identities and failures in this system are associated with cancer. The molecular basis for such a “cellular memory” mechanism was first identified in flies in form of the Polycomb (PcG) and Trithorax group (TrxG) system.

1.1 The Polycomb and Trithorax system

Polycomb (PcG) and Trithorax group (TrxG) proteins are essential in maintaining the transcriptional state of a number of developmental regulator genes during the development of multicellular organism (Akam, 1987; Duncan, 1982; Ingham, 1983; Lewis, 1978) (Figure 1.1). To date about 20 PcG proteins and a similar number of TrxG proteins have been described in *Drosophila* (de Ayala Alonso et al., 2007; Kennison and Tamkun, 1988; Lewis, 1978; Nüsslein-Volhard and Wieschaus, 1980) (Table 1.1 and Table 1.2). The PcG and TrxG system is essential for normal differentiation by maintaining cell-fate decisions during embryogenesis (Schuettengruber et al., 2009; Sparmann and van Lohuizen, 2006). Furthermore, the system is critical for propagating the repressed and active state of target genes during the cell division cycle (Martinez and Cavalli, 2006). Besides its key role in maintaining the body pattern and cell differentiation during development, the PcG and TrxG system is also involved in other cellular memory processes in various organisms including imprinting (Mager et al., 2003; Puschendorf et al., 2008; Terranova et al., 2008), X chromosome inactivation in female mammals (Heard, 2004) as well as vernalization in plants (Sung and Amasino, 2004). More recently, the PcG and TrxG system has been shown to play a role in tumor

onset and progression (Bracken and Helin, 2009; Feinberg et al., 2006; Sparmann and van Lohuizen, 2006).

Polycomb group (PcG) proteins



target genes



Trithorax group (TrxG) proteins

Figure 1.1: PcG and TrxG proteins maintain transcriptional states of developmental control genes.

Scheme of repressive PcG complexes antagonized by activating TrxG complexes. PcG/ TrxG target genes include many developmental control genes.

In *Drosophila*, PcG proteins play a crucial role during embryonic development (Lewis, 1978; Kennison, 1995). The best-known PcG target genes are the homeotic genes (Hox genes). Hox gene expression is established during the first few hours of embryogenesis by a number of transcription factors that are encoded by gap and pair rule genes. Together, these factors delimit Hox gene expression domains along the anterior-posterior axis of the embryo to specify the characteristic identity of each body segment (Akam, 1987; Jurgens, 1985; Simon et al., 1992). The expression patterns of Hox genes are then maintained by PcG and TrxG proteins (Capdevilla and García-Bellido, 1981; Jurgens, 1985; Struhl and Akam, 1985). While PcG proteins maintain the repressed state, TrxG proteins function to antagonize the PcG system; they maintain the active state of Hox genes in body segments where these genes should be expressed (Figure 1.1).

Null mutants of PcG proteins result in a characteristic Polycomb phenotype that displays body-patterning defects, where anterior segments are transformed towards posterior segments (Jurgens, 1985). Accordingly, TrxG mutants display mainly posterior to anterior transformations (Ingham, 1983). In addition to their repression of Hox genes, PcG proteins have been shown to target many other genes involved in development such as components of the Notch, Hedgehog or Wnt signaling pathways (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). Furthermore, both, PcG and TrxG proteins, play a role in dosage compensation in female mammals by participating in the

early maintenance of the inactive state during X chromosome inactivation (Kalantry et al., 2006; Pullirsch et al., 2010; Wang et al., 2001).

1.2 Polycomb and Trithorax group protein complexes

Biochemical purifications in *Drosophila* revealed that PcG proteins are the subunits of at least four distinct multi-subunit complexes: Polycomb-repressive complex 1 (PRC1) (Shao et al., 1999), Polycomb-repressive complex 2 (PRC2) (Tie et al., 2001, Müller et al., 2002), Pho-repressive complex (PhoRC) (Klymenko et al., 2006) and Polycomb-repressive deubiquitinase complex (PR-DUB) (Scheuermann et al., 2010) (Figure 1.2).

Similarly, TrxG proteins are subunits of at least five histone-modifying multi-subunit complexes: three Complex Proteins Associated with Set1 (Mohan et al., 2011), the Trithorax Acetylation Complex 1 (TAC1) (Petruk et al., 2001) and the absent, small, or homeotic discs 1 (ASH1) complex (Bantignies et al., 2000; Tanaka et al., 2007). Next to these five histone modifying TrxG complexes, there exist at least five TrxG protein-containing complexes with ATP-dependent chromatin remodeling activity including Switch/Sucrose Non Fermentable (SWI/SNF) and Imitation Switch (ISWI) (Schuettengruber et al., 2011). Different subunits of the PcG and TrxG complexes have been reported to bind to histone tail modifications or possess catalytic activities to post-translationally modify histone tails. TrxG complexes are associated with active histone marks such as methylation of H3-K4 and H3-K36 and additionally have been shown to demethylate the repressive H3-K27me₃ mark. In contrast to that, PcG complexes confer repressive histone marks, especially methylation of H3-K27, but also mono- and deubiquitination of H2A-K118.

In the following paragraphs the core components of the *Drosophila* PcG complexes will be described in more detail (Figure 1.2). Apart from these core components many other factors were found to be associated with PcG complexes including different PcG isoforms, DNA-binding proteins, transcription factors or chromatin-modifying enzymes such as the histone deacetylase RPD3 (Saurin et al., 2001; Tie et al., 2001).

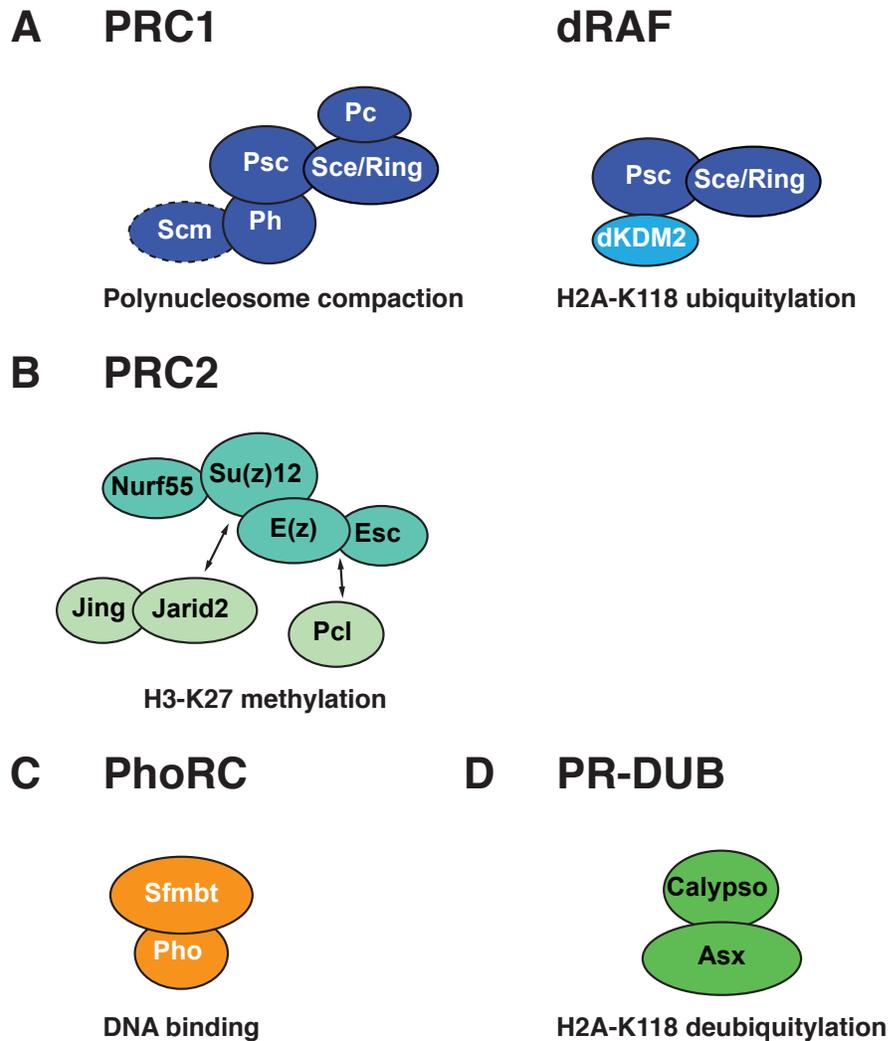


Figure 1.2: PcG protein complexes and their activities in *Drosophila*.

A: PRC1 and PRC1-like complexes (dRAF), B: PRC2 complex with substoichiometric members Pcl and Jing/ Jarid2 implicated in complex recruitment, C: PhoRC complex, D: PR-DUB complex. Modified from (Schwartz and Pirota, 2013; Simon and Kingston, 2013).

1.2.1 PRC1 and PRC1-like complexes

In flies, the canonical PRC1 complex consists of four core members namely Polycomb (Pc), Sex combs extra/Ring (Sce/Ring), Posterior sex combs (Psc) and Polyhomeotic (Ph). In addition, PRC1 contains Sex comb on midleg (Scm) in substoichiometric amounts (Saurin et al., 2001; Shao et al., 1999) (Figure 1.2 A, left panel). PRC1 members contain functional domains for binding and modification of histone tails. For

instance, Pc contains a chromodomain that can bind to tri-methylated histone tails. It has been shown that the Pc chromodomain binds preferentially to H3-K27me₃, a modification that is generated by PRC2 (Fischle et al., 2003; Min et al., 2003). Sce is the catalytic subunit of PRC1 and contains a Ring domain with an E3 ligase activity, which mediates mono-ubiquitination of lysine K118 of histone H2A (H2A-K119 in mammals) (de Napoles et al., 2004; Wang et al., 2004a). Psc is also a Ring finger containing protein. The mouse Psc and Sce homologs, Bmi-1 and Ring1b, respectively, have been shown to interact via their Ring domains and this interaction promotes the E3 ligase activity (Buchenwald et al., 2006; Li et al., 2006). Psc has a close paralogue called Suppressor of zeste 2 (Su(z)2), which can replace Psc in PRC1 (Lo et al., 2009). The function of Ph within the PRC1 complex is not well understood, but it has been proposed to be important for spreading of PcG complexes via its sterile alpha motif (SAM) domain (Kim et al., 2002). Flies contain two closely related Ph paralogs, Ph-proximal and Ph-distal that both co-purify with PRC1 (Shao et al., 1999).

An additional fly PRC1-like complex is the dRing-associated factors (dRAF) complex (Lagarou et al., 2008) (Figure 1.2 A, right panel). This complex also exists in mammals in form of the BCOR complex (Gearhart et al., 2006). The dRAF complex contains the two PRC1 members Sce/Ring and Psc, but lacks Pc, Ph and Scm. Interestingly; the dRAF complex also contains dKDM2, a demethylase, which has been shown to mediate removal of the active mark H3-K36me₂, but also strongly enhances the H2A ubiquitinase activity of Sce/Ring. In fact the dRAF complex has been suggested to mediate bulk levels of H2A mono-ubiquitination (Lagarou et al., 2008). In *Drosophila* the genome wide distribution of Pc largely overlaps with Sce/Ring and Psc distribution indicating that PRC1 and dRAF might have the same target genes (Gutiérrez et al., 2012).

How does PRC1 contribute to gene repression? PRC1 catalyzes mono-ubiquitination of H2A-K118. The role of H2A monoubiquitination in gene repression is somehow controversial. In mammals a number of studies suggest the importance of this modification in target gene repression. An example is that the H2A-K119ub₁ mark has been implicated in restraining RNA Polymerase II activity at bivalent promoters in mouse embryonic stem (mES) cells. Bivalent promoters carry both, active and repressive histone marks, and can change their activation state during differentiation

(Stock et al., 2007). In line with that, H2A-K119ub1 has been shown to prevent eviction of H2A-H2B dimers from nucleosomes by interfering with FACT (Facilitates chromatin transcription complex) recruitment, which is necessary for transcription elongation by RNA Polymerase II (Zhou et al., 2008). Additionally, *in vitro* experiments indicate that presence of H2A-K119ub1 prevents H3-K4 methylation (Nakagawa et al., 2008), which is a mark for active transcription. A more recent study found that H2A-K119ub1 also contributes indirectly to repression by providing a binding surface for recombinant human JARID2-AEBP2 containing PRC2 complexes, which in turn enhances H3-K27me3 of already monoubiquitinated nucleosomes establishing a repressive feedback loop (Kalb et al., 2014). This H2Aub1 dependent PRC2 recruitment mechanism appears to be conserved from fly to men (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). In spite of that, the H2A monoubiquitination does not seem to be crucial for repression of classical PcG target genes such as Hox genes in *Drosophila*. In particular, a catalytic mutant version of *Sce/Ring* (*Sce*^{I48A}) that lacks H2A monoubiquitinase activity could completely rescue silencing of Hox genes in *Drosophila* larvae, but not their viability (Scheuermann et al., 2012). In agreement with that, recent histone genetic assays in our lab have revealed that mutating H2A-K118, the residue that is ubiquitinated by *Sce* in flies, did not result in misexpression of Hox and other classical PcG target genes in imaginal wing disc clones (Pengelly, unpublished). It remains possible that H2A-K118ub1 is crucial for repression of other PcG target genes and that their misexpression is the cause of lethality of animals with a mutated H2A-K118 residue.

Clearly, the H2Aub1 mark is not the only mechanism through which PRC1 represses target genes. In fact, PRC1 induces chromatin compaction even in absence of histone tails *in vitro* (Francis et al., 2004; Grau et al., 2011) and also inhibits ATP-dependent chromatin remodeling by the SWI/SNF complex (Francis et al., 2001; Shao et al., 1999). Psc and its homolog Su(z)2 are crucial for PRC1-mediated chromatin compaction (Francis et al., 2004; King et al., 2005; Lo et al., 2009). Though Ph can also inhibit chromatin remodeling and induce chromatin compaction, it does so less efficiently than Psc (Francis et al., 2001; King et al., 2002). In *Drosophila* Psc and Ph are believed to cooperate in transcriptional repression of a certain class of target genes that do not

require other PRC1 members by altering local, higher-order chromatin structure (Francis et al., 2004; Gutiérrez et al., 2012). A recent study in mouse embryos shed some light on how Ph could contribute to repression. The authors discovered that the ability of Ph to induce and stabilize PRC1 clustering via polymerization of its SAM domain plays a crucial role in gene repression (Isono et al., 2013). Strikingly, the ability of PRC1 complexes to repress target genes by chromatin compaction is conserved to mammals, but might involve different PRC1 subunits, since for example in mouse Ring1B and a Pc homolog (M33) have been reported to contribute to chromatin compaction (Eskeland et al., 2010; Grau et al., 2011).

1.2.2 PRC2 complexes

Drosophila PRC2 contains four core components: Enhancer of zeste (E(z)), Extra sex combs (Esc), Suppressor of zeste 12 (Su(z)12) and nucleosome remodeling factor 55 (Nurf55) (Tie et al., 1998, 2001) (Figure 1.2 B). E(z) catalyzes mono-, di- and trimethylation of H3-K27 via its SET (Suvar3-9, Enhancer-of-zeste, Trithorax) domain (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002). However, E(z) is catalytically inactive *in vitro* unless associated with other PRC2 members. While Su(z)12 and Nurf55 are both necessary for anchoring E(z) to nucleosome substrates, Esc has been shown to boost enzymatic activity of E(z), probably by directly linking it to its substrate, histone H3 (Ketel et al., 2005; Nekrasov et al., 2007; Tie et al., 2007). In addition, Esc and Su(z)12 have been reported to interact not only with target nucleosomes, but also with surrounding nucleosomes in order to regulate the methyltransferase activity of PRC2 (Schmitges et al., 2011; Yuan et al., 2011, 2012).

Next to the four core PRC2 members, there are some additional proteins that have been shown to be associated with PRC2, but that are not essential for catalytic activity. These include Polycomb like protein (Pcl) (Nekrasov et al., 2007), Jumonji, AT rich interactive domain 2 (Jarid2) (Herz et al., 2012a) as well as the zinc finger protein Jing, which is the fly ortholog of AEBP2 (Cao and Zhang, 2004). These additional subunits have been shown to promote complex stability and particularly Pcl and Jarid2 have been reported to play a role in PRC2 binding to specific subsets of target genes (Ciferri et al., 2012; Herz et al., 2012a; Kim et al., 2009; Savla et al., 2008). While Jarid2 and

Jing have been reported to copurify together with core PRC2 members (Herz et al., 2012a). Pcl appears to form a separate PRC2 variant complex (Nekrasov et al., 2007; Schwartz and Pirotta, 2013).

Remarkably, Pcl has a classical PcG phenotype in flies (Duncan, 1982) and co-occupies Polycomb response elements (PREs) with other PcG proteins (Papp and Müller, 2006). Pcl has not only been shown to be essential for PRC2 recruitment in *Drosophila* larvae (Savla et al., 2008), but also promotes high levels of H3-K27 trimethylation at PcG target genes (Nekrasov et al., 2007). While the Tudor domain in mammalian Pcl specifically recognizes H3-K36me_{2/3} and has been suggested to promote binding of PRC2 to partially active PcG target genes (Ballaré et al., 2012; Cai et al., 2013), fly Pcl contains an atypical Tudor domain that cannot bind to methylated H3-K36 (Ballaré et al., 2012; Friberg et al., 2010). The mechanism of how fly Pcl contributes to PRC2 recruitment is still not clear; however, studies in our lab have shown that it can bind to DNA in a non-specific manner (Choi, unpublished).

In contrast to Pcl, Jarid2 mutants do not display classical PcG phenotypes and the gene was not identified in screens for PcG mutants (de Ayala Alonso et al., 2007; Duncan 1982). Regardless, Jarid2 can form a stable complex with PRC2 core members and promotes binding of PRC2 to many target genes (Herz et al., 2012a). Besides a zinc finger and an ARID domain, which are potentially involved in DNA binding, Jarid2 harbors a Jumonji (JmjC) domain, which typically works as a histone demethylase, but has been reported to be catalytically inactive (Li et al., 2010a). Interestingly, Jarid2 has a low occupancy at Hox genes, where enrichment of core PRC2 members is usually very high, implying that it is not involved in recruiting PRC2 complexes to classical PcG target genes. Since Jarid2 mutants only show a slight decrease in global H3-K27me₃ levels, Jarid2 has been suggested to play a role in fine-tuning H3-K27me₃ levels (Herz et al., 2012a). A recent study showed that recombinant human JARID2-AEBP2-containing PRC2 complexes could not only bind to nucleosomes containing H2Aub1, but also exhibit a strongly enhanced trimethylation activity on H2Aub1-containing nucleosomes (Kalb et al., 2014). In contrast to *Drosophila*, mouse embryonic stem (mES) cells require Jarid2 for PRC2 recruitment and proper differentiation (Landeira et al., 2010; Pasini et al., 2010).

Unlike PRC1, PRC2 asserts its major repressive function by modifying histone tails with the H3K27me3 mark (Pengelly et al., 2013). PRC2 has been shown to trimethylate extensive chromatin stretches that generally exceed 10 kilobases. In addition, the genome wide distribution of H3-K27me3 coincides with the distribution of PcG complexes and repressed genes depict high levels of H3-K27me3 in their coding region (Nekrasov et al., 2007; Schwartz et al., 2006; Tolhuis et al., 2006). Furthermore, the histone lysine methylation is a rather stable modification (Bannister et al., 2002). Remarkably, it has been shown that H3-K27me3 prevents acetylation of H3-K27, which is a mark for active genes and promoters (Tie et al., 2009). PRC2 activity is fine-tuned by cues from surrounding chromatin. While the presence of active histone marks such as H3-K4me3, H3-K36me2 and H3-K36me3 decreases catalytic activity (Ketel et al., 2005; Schmitges et al., 2011; Yuan et al., 2011), high nucleosome density and presence of H3-K27me2 and H3-K27me3 have a stimulating effect on PRC2 activity (Margueron et al., 2009). Moreover PRC2 can also bind to the H3-K27me3 modification that it deposits (Hansen et al., 2008; Margueron et al., 2009; Yuan et al., 2012). Hence, positive and negative feedback regulation by surrounding chromatin promotes the maintenance of methylated states.

To date, the molecular mechanisms of repression by the H3-K27me3 mark are not completely resolved. Histone marks usually operate either by recruiting additional regulatory factors or by directly rearranging chromatin structure through mediation of altered interactions to adjacent nucleosomes. In case of H3-K27me3, the former mechanism is supported by studies reporting the specific binding of the PRC1 member Pc via its chromodomain to this histone lysine methylation (Fischle et al. 2003; Min et al., 2003). However, this is possibly not the only way that H3-K27me3 contributes to gene repression.

1.2.3 PhoRC complexes

The PhoRC complex has two core members: Pleiohomeotic (Pho) and Sfmbt (Scm-like protein with four MBT domains) (Klymenko et al., 2006) (Figure 1.2 C). Contrary to other PcG complexes, PhoRC has no known catalytic activity, but its component Pho has a specific DNA binding activity for Polycomb response elements (PREs), which are

essential for PcG recruitment to target genes in flies (Strutt et al., 1997). To date Pho and its homolog Pho like (PhoI) are the only known PcG proteins that can bind to DNA in a sequence specific manner (Brown et al., 1998, 2003; Fritsch et al., 1999). Therefore, PhoRC has been proposed to play a role in recruitment of PcG complexes to target genes (Klymenko et al., 2006; Oktaba et al., 2008). The second PhoRC member, Sfmbt, has four MBT (malignant brain tumor) repeats with which it can bind to mono- and dimethylated lysines in a variety of contexts (Grimm et al., 2009).

Even though homologs of Pho and Sfmbt exist in mammals (Table 1.1) their function seems to have diverged. For instance, YY1, one of the homologs of Pho, shows little overlap with genomic binding profiles of the other PcG proteins (Mendenhall et al., 2010). In agreement with this, YY1 has been shown to perform genome-wide PcG independent functions in mES cells (Vella et al., 2012).

1.2.4 PR-DUB complex and other PcG proteins

The Polycomb repressive deubiquitinase (PR-DUB) complex consists of Calypso, a deubiquitinating enzyme, and Additional sex combs (Asx), which binds Calypso via its conserved N-terminal domain and is required for enzymatic activity (Figure 1.2 D). PR-DUB is conserved to mammals and catalyzes deubiquitination of H2AK118ub1 (H2AK119ub1 in mammals). The PR-DUB complex co-localizes with other PcG complexes at PcG targets and its deubiquitination activity is essential for maintaining PcG repression at Hox genes (Scheuermann et al., 2010).

The puzzle of how H2Aub deubiquitination might contribute to gene repression is still unresolved. PR-DUB strongly synergizes with Sce resulting in a more severe double mutant phenotype indicating that a precise balance between ubiquitination and deubiquitination is important for Hox gene repression (Gutiérrez et al., 2012). Interestingly, even though PR-DUB colocalizes with PRC1 at many target genes, it is only required at a subset of target genes that require Sce for repression (Gutiérrez et al., 2012; Scheuermann et al., 2012). Two alternative mechanisms have been proposed for how H2A deubiquitination could contribute to PcG target gene silencing. On the one hand PR-DUB has been suggested to be important to confine H2A monoubiquitination to specific locations or time points, in chromatin of classical PcG target genes such as

Hox genes. On the other hand it is possible that the major function of PR-DUB is to oppose a promiscuous Sce ubiquitination by deubiquitinating H2Aub1 at target genes, where monoubiquitination is not required. Thus providing sufficient amounts of free ubiquitin for crucial PcG targets such as Hox genes.

Table 1.1: List of known *Drosophila* PcG proteins and their human homologs.

Protein name	Abbr.	Human homologs	Function
Polycomb repressive complex 1 (PRC1)			
Posterior sex combs	Psc	PCGF1 (NSPC1), PCGF2 (Mel18), PCGF3, PCGF4 (BMI1), PCGF5, PCGF6 (MBLR)	H2A-K118 (K119) ubiquitination, chromatin compaction
Suppressor of zeste 2	Su(z)2		Psc paralog
Sex combs extra/Ring	Sce/ dRing	RING1A (RNF1), RING1B (RNF2)	H2A-K118 (K119) ubiquitination
Polycomb	Pc	CBX2 (PC1), CBX4 (PC2), CBX6, CBX7, CBX8 (PC3)	H3-K27me3 binding
Polyhomeotic proximal	Ph-p	PH1, PH2, PH3	Chromatin compaction
Polyhomeotic distal	Ph-d		Ph-p paralog
Sex comb on midleg	Scm	SCMH1, SCMH2, SCML2	H3-K9me1 binding, DNA binding
Polycomb repressive complex 2 (PRC2)			
Enhancer of zeste	E(z)	EZH1, EZH2	H3-K27me1/2/3 methylation
Extra sex combs	ESC	EED	HMTase cofactor, H3-K27me3 binding
Extra sex combs-like	ESCl		Esc paralog
Suppressor of zeste 12	Su(z)12	SUZ12	HMTase cofactor
Polycomb-like	Pcl	PHF1 (PCL1), MTF2 (PCL2), PHF19 (PCL3)	Stimulates H3-K27me3
Pho repressive complex (PhoRC)			
Pleiohomeotic	Pho	YY1, YY2	Sequence specific DNA binding
Pleiohomeotic-like	Phol	YY1, YY2	Pho paralog
Scm-like protein with four MBT domains	dSfmbt	L3MBTL2, MBTD1, hSFMBT1, hSFMBT2	H3-K9me1/2, H4-K20me1/2 binding
Polycomb repressive deubiquitinase (PR-DUB)			
Calypso	-	BAP1	H2A-K118 (K119) deubiquitination
Anterior sex combs	Asx	ASXL1, ASXL2, ASXL3	Deubiquitination co-factor

Protein name	Abbr.	Human homologs	Function
Other PcG proteins			
Super sex combs	Sxc	OGT	O-GlucNAc transferase
Multi sex combs	Mxc	-	-
Cramped	Crm	CRAMP1L	-
Non PcG proteins associated with PcG complexes			
Lysine (K)-specific demethylase 2	Kdm2	KDM2A, KDM2B	H2A-K118 (K119) ubiquitination in dRAF, H3-K36me2 demethylase
Nucleosome remodeling factor 55	Nurf55/ Caf1	RBBP7 (RBAP46), RBBP4 (RBAP48)	Nucleosome binding
Jing	Jing	AEBP2	PRC2 targeting
Jumonji, AT rich interactive domain 2	Jarid2	JARID2	PRC2 targeting

Table was modified from (Di Croce and Helin, 2013; Schwartz and Pirota, 2013; Simon and Kingston, 2013).

Super sex combs (Sxc) is an additional PcG protein that was not found to be co-purified with any of the known PcG complexes. However, a mutant lacking Sxc displays a PcG phenotype in *Drosophila* (Ingham, 1984). More recently it has been shown that Sxc encodes for O-GlcNAc transferase (Ogt), a conserved enzyme catalyzing the addition of O-linked β -N-Acetylglucosamine (O-GlcNAc) sugar moieties to target proteins. Interestingly, Ogt modifies the PcG protein Ph *in vivo* and in this way might contribute to its repressive function (Gambetta et al., 2009). Therefore, Ogt establishes a link between the metabolic state of a cell and gene repression by the PcG system.

1.2.5 Trx group protein complexes

Trx group proteins are best known for counteracting PcG silencing; however, they also play a more general role in gene expression and activation. Similar to PcG proteins they also function in multi-subunit complexes, although they form a more heterogeneous group with diverse functions ranging from histone methylation to ATP-dependent chromatin remodeling and DNA binding activities (Ho and Crabtree, 2010; Schuettengruber et al., 2011).

To date, five Trx complexes harboring histone-modifying activities have been described (Schuettengruber et al., 2011) (Table 1.2). Amongst these five complexes are three *Drosophila* COMPASS-related complexes that incorporate dSET1, TRX and TRR (TRX-related), respectively, as SET domain-containing proteins. All three complexes are H3-K4 methyl-transferases, but, while the dSET1/ COMPASS complex mediates bulk di- and trimethylation of K4, the other two complexes only contribute marginally to global H3-K4 di- and trimethylation levels and have been suggested to mainly function in gene-specific transcriptional regulation (Ardehali et al., 2011; Mohan et al., 2011).

Table 1.2: List of TrxG proteins and their mammalian homologs.

Protein name	Abbrv.	Complex	human Homologs	Function
Histone-modifying complexes				
Absent, small or homeotic discs 2	Ash2	COMPASS, COMPASS-like	ASH2L	H3-K4 methylation
Trithorax	Trx	COMPASS-like, TAC1	MLL1, MLL2	H3-K4 methylation
Trithorax like	Trl	COMPASS-like	MLL3, MLL4	H3-K4 methylation
Ubiquitously transcribed tetratricopeptide repeat, X chromosome	Utx	COMPASS-like	KDM6A, KDM6B, UTY	H3-K27me3 demethylation
Absent, small or homeotic discs 1	Ash1	ASH1	ASH1L	H3-K36 methylation
ATP-dependent chromatin-remodeling complexes				
Brahma	Brm	Brahma complex (SWI/SNF)	BRM, BRG1	Acetylated histone binding
Brahma associated protein 55kD	Bap55		ACTL6A, ACTL6B	
Brahma associated protein 60 kD	Bap60		SMARCD1, SMARCD2	
dalao	dalao		SMARCE1	
enhancer of yellow 3/ SAYP	e(y)3		PHF10	
eyelid/ osa	eld		ARID1A, ARID1B	
moira	mor		SMARCC1, SMARCC2	
Imitation SWI	lswi	NURF (ISWI)	SMARCA1, SMARCA5	Histone binding, reads H3-K4me3
Nucleosome remodeling factor-38kD	NURF 38		PPA1, PPA2	
Enhancer of bithorax/ NURF301	E(bx)		BPTF	
kismet	kis	CHD6, CHD7, CHD8	CHD6, CHD7, CHD8, CHD9	H3-K4me2/3 binding

Protein name	Abbrv.	Complex	human Homologs	Function
other TrxG proteins				
domino	dom	-	SRCAP	H2Av exchange
female sterile (1) homeotic	fs(1)h	-	BRD2, BRD3, BRD4, BRDT	DNA binding
grappa	gpp	-	DOT1L	H3-K79 methylation
lola like	lolal	-	-	DNA binding
zeste	z	-	-	DNA binding

Table was modified from (Kennison and Tamkun, 1988; Schuettengruber et al., 2011).

Specifically, the TRR containing dCOMPASS-like complex was recently reported to be required for genome-wide H3-K4 monomethylation at enhancers (Herz et al., 2012b). Besides its methylation activity the TRR/COMPASS-like complex possesses a demethylation activity of H3-K27me3 mediated by the TrxG protein Utx (Cho et al., 2007; Copur and Muller, 2013; Issaeva et al., 2007; Lee et al., 2007; Mohan et al., 2011). Trx also exists in a COMPASS-like complex (Mohan et al., 2011). In addition, Trx has been reported to be present in a complex called TAC1 that contains cyclic AMP response element-binding protein (dCBP) and was proposed to contribute to gene activation by acetylation of H3-K27 (Petruk et al., 2001; Tie et al., 2009). Moreover, Ash1 has an H3-K36-specific methyltransferase activity (Tanaka et al., 2007; Yuan et al., 2011). To date, no purification of Ash1 complexes has been reported, but the protein has been found to be associated with dCBP (Bantignies et al., 2000).

In addition to these histone-modifying TrxG complexes, *Drosophila* possesses at least five TrxG protein-containing complexes that bind to histone modifications and exert ATP-dependent chromatin-remodeling activities by which they open up chromatin structures (Table 1.2). These include the Brahma complex – it can bind to acetylated histones via its bromodomain (Chalkley et al., 2008; Dingwall et al., 1995; Ho and Crabtree, 2010; Schuettengruber et al., 2011) – and the NURF (nucleosome-remodeling factor) complex – it can bind to histones via its SANT (switching-defective protein 3 (Swi3), adaptor 2 (Ada2), nuclear receptor co-repressor (N-CoR), transcription factor (TFIIIB)) domains and to H3-K4me3 via its PHD (plant homeodomain) finger (Badenhorst et al., 2002; Ho and Crabtree, 2010; Schuettengruber et al., 2011). Furthermore, several chromodomain helicase DNA-binding (CHD) complexes – CHD1

to CHD8 – contain TrxG proteins and are involved in diverse functions such as H3.3 deposition (Konev et al., 2007) and transcriptional elongation (Srinivasan et al., 2008).

1.2.6 Conservation of PcG proteins in animals and plants

PcG family members associated with PRC2 have been identified in multicellular organisms ranging from plants to nematodes to mammals and are consistently involved in differentiation and development. In contrast to that, PRC1 components are absent in nematodes and plants and therefore seem to have appeared later during evolution than PRC2 members (Ringrose and Paro, 2004; Sparmann and van Lohuizen, 2006). While the general function of the PcG system in development is well conserved, individual players have changed and diversified in vertebrates. In particular, during vertebrate evolution, PcG proteins appear to have undergone multiple duplication events. It seems likely that the increasing complexity in body plans demanded for tissue- and developmental-stage specific expression of genes, which is reflected by an increased complexity in chromatin organization and regulation (Ho and Crabtree, 2010). While *Drosophila* contains about 20 PcG proteins mostly in single copies, mammals such as mouse and human possess about twice the amount of known PcG members with multiple homologs for most fly PcG protein (Whitcomb et al., 2007) (Table 1.1 and Table 1.2). Additional complexity is added by the possibility to assemble different subunit homologs combinatorial into distinct PcG complexes (Ho and Crabtree, 2010). Strikingly, the key functional domains of PcG homologs are highly conserved even between evolutionary distant organisms. However, outside the functional domains PcG proteins differ in length and sequence, with novel functional domains emerging in some members (reviewed in (Hennig and Derkacheva, 2009; Whitcomb et al., 2007)).

The diversification of the PcG system resulted in variant complexes with slightly different biochemical activities or functions and specific expression patterns. In addition, target genes of variant complexes have been reported to only partially overlap. Therefore, it is conceivable that complex composition varies in a tissue-specific manner in vertebrates with alternative recruitment mechanisms and specialized functions. Furthermore, they might form at different times during differentiation and development (Bernstein et al., 2006; Gao et al., 2012; Gunster et al., 2001; Kagey et al., 2003,

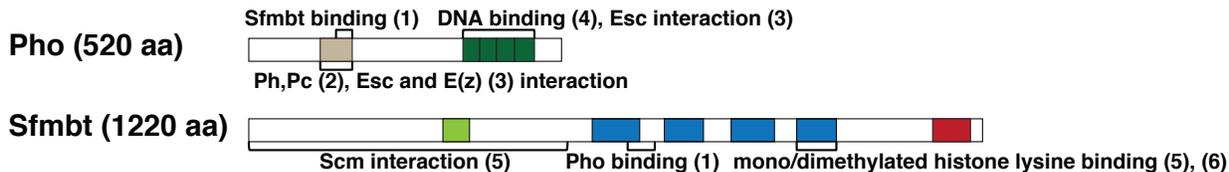
Sparmann and van Lohuizen, 2006). Particularly the PRC1 complex has been shown to exist in at least six different PRC1-like variations in mammals with partially non-redundant functions and composition, but containing RING1B as the conserved unit (Gao et al., 2012). For example in mES cells PRC1 complexes contain only the Pc homolog CBX7, while differentiating cells have been shown to possess CBX2- and CBX4-containing PRC1 complexes (Klauke et al., 2013; Morey et al., 2012; O’Loghlen et al., 2012).

Since PcG proteins in mammals have been implicated in more dynamic processes such as actively setting up silenced states during differentiation (Ballaré et al., 2012; Brien et al., 2012; Cai et al., 2013), it has been proposed that their function in long-term memory of repressed states in flies has been partly replaced by DNA methylation in vertebrates. Instead the PcG system in mammals has been suggested to play its major role in dynamic silencing of genes and in short-term memory of silencing (Schuettengruber et al., 2009; Sparmann and van Lohuizen, 2006).

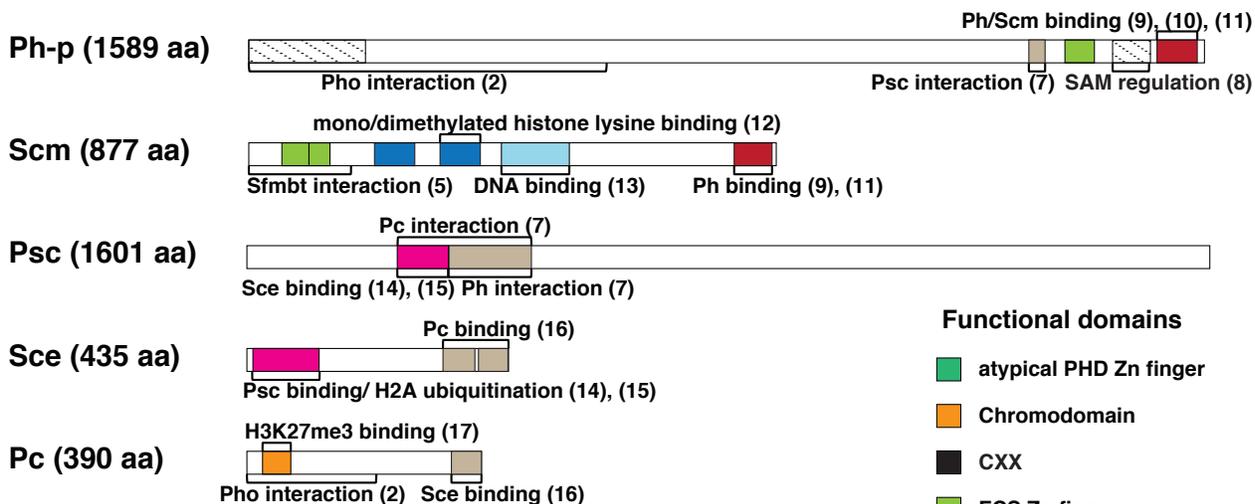
1.3 Domain architecture of Polycomb group protein complexes

PcG proteins have been extensively characterized at the molecular level and contain many described functional domains common to chromatin interacting complexes such as chromodomains and SANT domains (Boyer et al., 2004; Paro and Hogness, 1991; Taverna et al., 2007). In addition, they contain catalytical domains that can modify histone tails such as the SET domain (Müller et al., 2002) and the Ring finger (Buchwald et al., 2006; Gorfinkiel et al., 2004), which are controlled by regulatory domains. A detailed summary of all the described functional domains present in the major PcG complexes in flies, their allocated functions and reported interactions is illustrated in Figure 1.3.

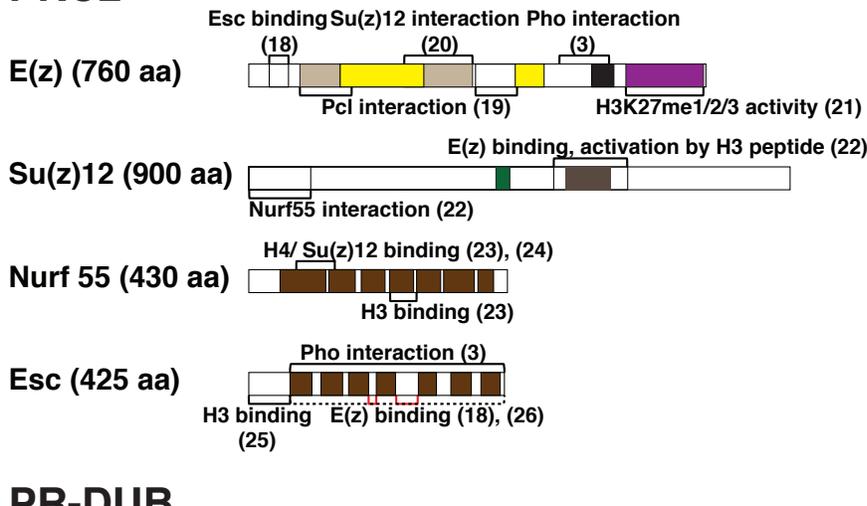
PhoRC



PRC1



PRC2



PR-DUB

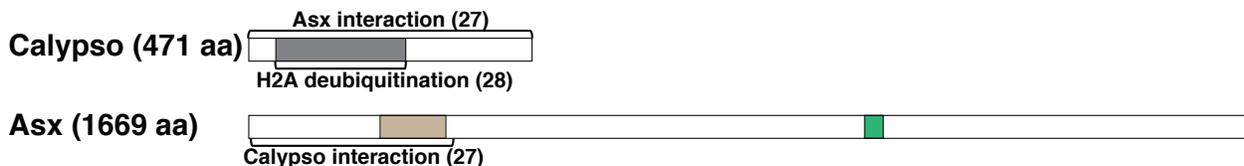


Figure 1.3: Functional domains and interactions of fly PcG complexes.

Schematic representation of major PcG proteins with their known functional domains and mapped interaction surfaces with either other PcG proteins or substrates. Only one paralog of Ph, Psc, Esc and Pho is depicted. PHD: plant homeodomain; Zn finger: Zinc finger; CXX: conserved carboxy domain; FCS: phenylalanine (F)–cysteine (C)–serine (S) sequence motif; HD: homology domain; MBT: malignant brain tumor; Ring: really interesting new gene; SAM: sterile alpha motif; SANT: switching-defective protein 3 (Swi3), adaptor 2 (Ada2), nuclear receptor co-repressor (N-CoR), transcription factor (TFIIIB); SET: Suvar3-9, Enhancer-of-zeste, Trithorax; SLED: Scm-like embedded domain; UCH: ubiquitin carboxy-terminal hydrolase; VEFS: VRN2-EMF2-FIS2-Su(z)12; WD40: tryptophan (W)-aspartic acid (D) 40.

References: (1) (Alfieri et al., 2013), (2) (Mohd-Sarip et al., 2002), (3) (Wang et al., 2004b), (4) (Brown et al., 1998) (5) (Grimm et al., 2009), (6) (Klymenko et al., 2006), (7) (Kyba and Brock, 1998a), (8) (Robinson et al., 2012), (9) (Peterson et al., 1997), (10) (Kim et al., 2002), (11) (Kim et al., 2005), (12) (Grimm et al., 2007), (13) (Bezsonova, 2014), (14) (Buchenwald et al., 2006), (15) (Li et al., 2006), (16) (Gorfinkiel et al., 2004), (17) (Fischle et al., 2003), (18) (Tie et al., 1998), (19) (O’Connell et al., 2001), (20) (Ketel et al., 2005), (21) (Müller et al., 2002), (22) (Yuan et al., 2012), (23) (Nowak et al., 2011), (24) (Schmitges et al., 2011), (25) (Tie et al., 2007), (26) (Jones et al., 1998), (27) (Scheuermann et al. 2010), (28) (Nijman et al., 2005).

To date, we have only very little structural information on the architecture of PcG protein assemblies. In order to understand the molecular mechanisms underlying gene silencing by PcG complexes, it is crucial to gain structural information on PcG members on the atomic level. Due to the large size of many of the *Drosophila* PcG proteins most of the structural information known today derives from single functional domains that were crystallized alone, in complex with histone substrates or in complex with other small functional domains (summarized in Table 1.4). For example, in the case of PRC2, atomic resolution structures of the Esc and Nurf55 subunits in complex with short peptides of the other PRC2 subunits have been determined (Han et al., 2007; Murzina et al., 2008; Schmitges et al., 2011; Song et al., 2008). Nevertheless, there are numerous studies that reported a diversity of different physical associations between individual subunits and with subunits in other PcG protein complexes as well as interactions with histone tails (summarized with according references in Figure 1.3).

Table 1.3: PcG protein structures with protein database accession (PDB) numbers.

Solved structure	Reference	PDB ID
PRC1 complex		
Pc-H3-K27me3 (Pc chromodomain with H3-K27me3 peptide)	(Fischle et al., 2003, Min et al., 2003)	1PDQ 1PFB
Ph-Scm SAM domain heteropolymer	(Kim et al., 2005)	1PK1, 1PK3
Ring1B/Bmi1 complex (two Ring finger domains)	(Buchwald et al., 2006)	2CKL
	(Li et al., 2006)	2H0D
Ring1b Cterm domain (CBX interaction domain)	(Bezsonova et al., 2009)	3H8H
Scm MBT-H3-K9me1/me2 (MBT repeats of Scm with monomethyl-lysine and dimethyl-lysine of H3K9)	(Grimm et al., 2007)	2R5A, 2R5B, and 2R5M
ScmL2-H3-K9me1/H4-K20me2	(Santiveri et al., 2008)	2VYT
ScmL2 SLED domain (NMR structure DNA-binding domain)	(Bezsonova, 2014)	-
PRC2 complex		
Eed-Ezh2 (complex of Eed with N-terminal peptide of Ezh2)	(Han et al., 2007)	2QXV
Nurf55-H4	(Song et al., 2008)	3C99
RbAp46-H4 (Nurf55 or RbAp46 with the first helix of H4)	(Murzina et al., 2008)	3CFV, 3CFS
Pcl tudor domain (NMR structure of Pcl tudor domain)	(Friberg et al., 2010)	2XK0
Nurf55 (1-418)-Su(z)12 (73-143)	(Schmitges et al., 2011)	2YB8
Nurf55-H3 (1-19)		2YBA
Nurf55-H4 peptide (26-45)	(Nowak et al., 2011)	2XYI
Ezh2 SET Domain	(Antonyamy et al., 2013)	4MI5
PhoRC complex		
Sfmbt 4MBT-H4-K20me1	(Grimm et al., 2009)	3H6Z
Pho-Sfmbt 4MBT (complex of Pho spacer and MBT repeat 1 to 4 of Sfmbt)	(Alfieri et al., 2013)	4C5E, 4C5G, and 4C5H
human YY1:MBTD1 4MBT		4C5I
human YY1:adeno-associated virus P5 element (DNA)	(Houbavity et al., 1996)	1UBD
other PcG structures		
hL3MBTL2 FCS ZNF (NMR structure of FCS ZNF domain)	(Lechtenberg et al., 2009)	-
hL3MBTL1-H4K20me2 complexes	(Min et al., 2007)	2PQW, 2RJE and 2RJF

Solved structure	Reference	PDB ID
hL3MBTL1 apo-structure/ L3MBTL1–MES complex		2RJD, 2RJC
hL3MBTL1 197–526-H1.523–27K27me2	(Li et al., 2007)	2RHI
hL3MBTL1-H3.328–34 chimera-Kme2		2RHU
hL3MBTL1 197–526-Kme2/me1		2RHX, 2RHY
hL3MBTL1 206–519(D355N/A)		2RHZ, 2RI2
hL3MBTL1 206–519(N358Q/A)		2RI3, 2RI5

Table was modified from (Müller and Verrijzer, 2009).

Recently a 21 Å EM structure of the human PRC2 core in complex with AEBP2 in combination with chemical-cross-linking expanded our knowledge of the PRC2 architecture and how subunits interact with each other to form a functional holoenzyme that can interact with and modify chromatin (Ciferri et al., 2012). Together with the crystal structures of NURF55, EED and several solved functional domains including the SANT, SET and Zn finger domains a comprehensive picture of how the PRC2 complex is build on an atomic resolution level starts to emerge. The next challenge will be to understand in more molecular detail how these individual complexes interact with each other, are recruited to target genes and combine their individual functions at the chromatin of their target genes in order to achieve repression.

1.4 Biological Importance of the PcG system

In addition to their crucial role in maintaining body patterning during development, PcG proteins play an important role in differentiation and stem cell maintenance. Deregulation of these processes by overexpressing or deleting single PcG proteins has been linked to cancer formation.

1.4.1 Role of PcG proteins in stem cell maintenance and differentiation

In *Drosophila* PcG genes have been shown to play a role in antagonizing stem cell self-renewal and cell differentiation. For instance, Psc and Su(z)2 have been reported to be required for proper differentiation of follicle stem cells into follicular cells in *Drosophila* ovaries. In absence of the two PcG proteins follicle stem cells are incapable of differentiation and develop into tumors due to deregulation of the Wnt signaling pathway (Li et al., 2010).

Similarly, PcG proteins in mammals have been implicated in maintaining both embryonic and adult stem cell identity. For example the human PcG proteins BMI-1, MEL-18 and mouse Ph1 have been reported to be crucial for the self-renewal of various adult stem cell types (Akasaka et al. 1997; Lessard and Sauvageau, 2003; Molofsky et al., 2003, Ohta et al., 2002). Genome-wide studies in mouse and human ES cells have illustrated that most PcG targets are transcription factors and other regulators of developmental pathways (Boyer et al., 2006; Lee et al., 2006). While many of the regulatory cell fate genes are repressed in ES cells, a certain subset of these genes becomes activated during cell fate commitment as demonstrated by studies reporting selective de-repression of PcG targets during differentiation of muscle, nerve and germ cells (Bracken et al., 2006; Caretti et al., 2004; Chen et al., 2005). Therefore, PcG proteins have been suggested to play a role in maintaining the pluripotent state of stem cells by preventing haphazard differentiation. However, the exact role that PcG complexes play in maintaining the pluripotent state of ES cells is somehow controversial. While PRC2 has been reported to be dispensable for ES cell self-renewal (Chamberlain et al., 2008; Pasini et al., 2004; Shen et al., 2008), the PRC1 members Ring1a/ Ring1b have been shown to be crucial for maintaining pluripotency (Endoh et al., 2008). Furthermore PcG proteins are not only involved in preventing differentiation by repressing certain genes, but have been implicated in driving and modulating the differentiation process in response to specific stimuli by repressing pluripotency genes (Pasini et al., 2007).

Interestingly, genes activated during differentiation of ES cells display a loss of PcG protein binding, whereas the genes that are switched off to become permanently repressed during differentiation already have high levels of PcG proteins bound in the non-differentiated state (Bracken et al., 2006). The identification of so called 'bivalent domains' that carry both repressive and active histone marks in ES cells could explain how developmental genes are kept repressed while staying poised for activation during differentiation (Bernstein et al., 2006).

1.4.2 Role of the PcG system in tumorigenesis

In *Drosophila*, deletion or mutation of the PRC1 members Psc/ Su(z)2 and Ph have been shown to result in overproliferation and tumor formation in imaginal disc clones (Classen et al., 2009; Feng et al., 2011; Martinez et al., 2009; Oktaba et al., 2008). Studies analyzing the molecular mechanism of tumor formation identified several PcG target genes that are deregulated in tumor cells including cell cycle regulators such as CycB (Oktaba et al., 2008), components of the Notch signaling pathway (Martinez et al., 2009) as well as Unpaired (Upd) family ligands, which are activators of the JAK/STAT signaling pathway (Classen et al., 2009) suggesting that at least some PcG proteins function as tumor suppressors in *Drosophila*.

Consistent with this, altered levels of PcG and TrxG proteins in mammalian cells are linked to cancer and tumor progression (reviewed in (Bracken and Helin, 2009; Feinberg et al., 2006; Sparmann and van Lohuizen, 2006). One of the first PcG proteins that was associated with human cancer is the Psc homolog, BMI1, which has been reported to cooperate with the proto-oncogene MYC to promote B- and T-cell lymphomas (Haupt et al., 1991; van Lohuizen et al., 1991). However, not only the overexpression or overactivity of PRC1 and PRC2 members such as BMI1, EZH2 and SUZ12 (McCabe et al., 2012; Simon and Lange, 2008; Sneeringer et al., 1999), but also mutations in other PcG components including BAP1 and ASXL1 can lead to oncogenesis in mammals (Shih et al., 2012). Subsequent studies showed that, similar to the situation in flies, malignant transformations could also be caused by inappropriate repression of tumor suppressors by the PcG system. One prominent example for deregulated PcG targets in cancer are the *CDKN2B* and *CDKN2A* loci (Gargiulo et al., 2013; Jacobs et al., 1999; van Lohuizen et al., 1999). These two loci encode for three tumor suppressors, INK4A, INK4B and ARF, which play an important role in restricting cellular proliferation and have been reported to be amongst the most commonly deregulated loci in human cancer. INK4A and INK4B are part of the RB pathway, whereas ARF induces the p53 pathway (Bracken and Helin, 2009; Gargiulo et al., 2013; Sherr, 2001). However, this is not the only crucial PcG target involved in cell-cycle regulation. For instance human PcG homologs of Esc (EED) and Psc (BMI1) have been

reported to regulate cell-cycle progression in an INK4A/ARF-independent manner (Bruggeman et al., 2007; Lessard et al., 1999).

Strikingly, PcG targets include many components of signaling pathways required for gastrulation, differentiation as well as proliferation and maintenance of stem cells pluripotency. These pathways include transforming growth factor- β (TGF β), Notch, Hedgehog, Wnt and fibroblast growth factor (FGF) (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006) and their deregulation has been associated with cancer development (Bieri and Moses, 2006; Reya and Clevers, 2005). In line with that PcG proteins have been reported to interact with DNA methyl-transferases (Mohammad et al., 2009; Viré et al., 2006) and many PcG target genes have been shown to be silenced by DNA methylation in human tumors (Bracken et al., 2006). This suggests that during tumor progression PcG proteins collaborate with DNA methylating enzymes in aberrantly repressing PcG target genes important for differentiation and control of cell proliferation.

1.5 Recruitment of PcG protein complexes to target genes

PcG proteins play an important role in maintaining the repressive chromatin state of many developmental control genes that regulate signaling pathways in development, differentiation and cell proliferation. One key question in understanding the underlying mechanism of the PcG system is how they are recruited to their target genes, since target gene selection is thought to determine the function of the recruited PcG complex. Therefore, the following paragraphs will summarize our current knowledge of how PcG complexes are targeted in flies and how this mechanism has evolved from flies to mammals.

1.5.1 Polycomb response elements (PREs)

PcG proteins are highly enriched at Polycomb response elements (PREs) (Orlando et al., 1998; Strutt et al., 1997; Strutt and Paro, 1997) and have been shown to be constitutively bound to PREs independently of the transcriptional state of the target gene (Nègre et al., 2006; Papp and Müller, 2006; Schwartz et al., 2006; Tolhuis et al., 2006). Conversely, other studies in *Drosophila* tissue culture cells have reported the

absence of PcG proteins from certain PREs in the active state of the target gene (Beisel et al., 2007; Schwartz et al., 2010). Strikingly, Pho and Phol are the only PcG proteins that are known to have specific DNA-binding activity for sequence motifs within the PREs (Brown et al., 1998, 2003).

Polycomb response elements (PREs) are cis-regulatory DNA sequences of several hundred base pair length that are responsible for recruitment of PcG and TrxG proteins in *Drosophila* (Chan et al., 1994; Horard et al., 2000; Poux et al., 1996; Schwartz et al., 2006; Simon et al., 1993). PREs display high nucleosome turnover rates (Deal et al., 2010) and are nucleosome-depleted (Mishra et al., 2001; Mohd-Sarip et al., 2006). In most cases PREs have been identified at the promoter proximal regions to their target genes, but some have been discovered in introns or in distant regions tens of kilobases up- or downstream of their target promoter (Kharchenko et al., 2010). Interestingly, PREs do not contain a common consensus sequence. However, they contain many conserved short sequence motifs, which are recognized by several DNA binding proteins (Ringrose and Paro, 2004; Ringrose et al., 2003). In particular PREs almost always contain Pho binding sites, whereas Phol is only present at a subset of PREs (Klymenko, 2006; Mihaly et al., 1998; Müller and Kassis, 2006; Oktaba et al., 2008; Schuettengruber et al., 2009). In line with that, mutation of Pho binding sites in PRE reporter gene assays resulted in severely impaired silencing capacity (Fritsch et al., 2003), suggesting that PhoRC plays a crucial role in PcG recruitment to target genes.

Nevertheless, Pho binding sites appear to be insufficient for PcG repression (Brown et al., 2005; Déjardin et al., 2005; Mohd-Sarip et al., 2005). In addition, several other sequence-specific DNA binding proteins have been reported to bind to PRE elements and have been implicated in PcG recruitment and silencing. These include the Trl/ GAF (Trithorax-like/ GAGA factor) (Busturia et al., 2001; Hodgson et al., 2001; Mahmoudi et al., 2003; Schwendemann and Lehmann, 2002), Psq (Pipsqueak) (Hodgson et al., 2001, Huang et al., 2002; Schwendemann and Lehmann, 2002), Zeste (Mahmoudi et al., 2003; Mulholland et al., 2003; Ringrose and Paro, 2004), Grh/ NTF-1 (Grainyhead/ neuronal transcription factor 1) (Blastyák et al., 2006), Sp1/ KLF (specificity protein/ Krueppel-like factor) (Brown et al., 2005) and Dsp1 (Dorsal switch protein 1) (Déjardin et al., 2005). However, the exact role that these transcription factors

play in PcG repression is not clear, since, besides Pho mutants, non of the other transcription factors exhibit a clear PcG phenotype (Goldberg et al., 1989; Rappailles et al., 2012) and non of these proteins appear to be sufficient for PcG targeting (Müller and Kassis, 2006). It is possible that some of these transcription factors are only involved in PcG recruitment at specific PREs (Blastyák et al., 2006; Brown et al., 2005) or act redundantly and therefore have a less severe phenotype. Surprisingly, all putative PcG recruiting factors including Pho and Phol do not only localize to PcG-bound regions, but also to promoter regions of active genes, indicating that these DNA-binding factors are not limited to gene silencing, but might also play a role in gene activation (Schuettengruber et al., 2009). Although PREs are necessary and sufficient for PcG recruitment in *Drosophila*, there is only very limited evidence for such elements in vertebrates or plants (Sing et al., 2009; Woo et al., 2010).

1.5.2 PcG target genes

The most studied PcG target genes that are often referred to as classical PcG target genes are the *Drosophila* Homeotic (Hox) genes (Busturia and Morata, 1988; Dura and Ingham, 1988). The Hox genes are present in two clusters, the Bithorax Complex (BX-C) and the Antennapedia complex (Ant-C), and encode transcription factors that are important for controlling the activity of many downstream targets. They start a specific developmental program through a regulated cascade and give identity to each segment of the body (Maeda and Karch, 2006). Misexpression of Hox genes results in homeotic transformations, the characteristic PcG mutant phenotypes (Jurgens, 1985).

The BX-C contains three homeotic genes: *Ultrabithorax (Ubx)*, *Abdominal-A (Abd-A)* and *Abdominal-B (Abd-B)* that are important for the identity of segments in the abdomen and in the posterior part of the thorax (Lewis, 1978; Maeda and Karch, 2006), whereas the Ant-C contains five homeotic genes: *labial (lab)*, *proboscipedia (pb)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)* and *Antennapedia (Antp)* that are crucial for segmental identity of the anterior body segments (Kaufman et al., 1980; Zink et al., 1991; Gindhart and Kaufman, 1995; Strutt et al., 1997). The structure and function of the PRE elements in the BX-C cluster has been studied extensively. For instance *Ubx* is controlled by the two PREs bx and bxd (Chan et al., 1994; Chian et al., 1995), whereas three cis-

regulatory regions including *iab-2* (Shimell et al., 2000), *iab-3* and *iab-4* (Steffen and Ringrose, 2014) control *Abd-A*. The *Abd-B* regulatory region on the other hand has been shown to contain the PREs Mcp (Orlando et al., 1998), Fab-7 (Mihaly et al., 1997) and Fab-8 (Barges et al., 2000).

Although Hox genes are an important group of PcG target genes, early on PcG proteins were shown to bind to more than 100 sites in polytene chromosomes from larval salivary glands (DeCamillis et al., 1992; Franke et al., 1992; Lonie et al., 1994; Zink and Paro, 1989), indicating that they control many more target genes besides the Hox gene cluster. Other PcG target genes include *even-skipped* (Dura and Ingham, 1988), *engrailed* (Americo et al., 2002), *invected* (Strutt and Paro, 1997), *hedgehog* (Maurange and Paro, 2002) and *cyclin A* (Martinez and Cavalli, 2006). Moreover two PREs have been identified at the gene locus of the PcG proteins Ph-p and Ph-d (Bloyer et al., 2003) suggesting an auto-regulatory feedback loop. An algorithmic approach relying on sequence motifs identified in PREs of known target genes predicted about 167 PRE sequences in the *Drosophila* genome (Ringrose et al., 2003).

In 2006 the first genome wide mapping studies of PcG proteins were published and reported between 200-400 PcG controlled genes in the *Drosophila* genome. Interestingly, all three studies found that transcription factors controlling major signaling pathways in development, differentiation and regulation of cell proliferation are highly enriched amongst PcG targets emphasizing the extensive involvement of PcG proteins in coordination of development and cell fate (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). These genome wide studies addressed chromosomal PcG distribution in different developmental stages ranging from embryo- to pupa- to adult-derived cells. Although they reported that the PcG distribution changes during development and is more dynamic than anticipated, they also detected a substantial number of shared PcG binding sites suggesting that PcG proteins are constitutively bound at many target genes (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). Strikingly, two studies that analyzed PcG distribution in mammals found that many of the major signaling pathways controlled by the PcG system in *Drosophila* are also regulated by PcG proteins in mammals (Bernstein et al., 2006; Lee et al., 2006).

1.5.3 Recruitment of PcG proteins to PREs at target genes

As described above PcG proteins have been shown to co-localize at PRE elements at the promoter proximal regions of their target genes. However, the assembly of different PcG complexes at the PRE elements remains elusive. Since Pho and its homolog Pho-like are the only known PcG proteins with a described specific DNA binding activity, a hierarchical recruitment model was suggested to explain subsequential recruitment of PRC1 and PRC2 complexes (Wang et al., 2004b). In this model, PhoRC binds to PREs via its DNA binding domain and recruits the PRC2 complex via a direct interaction with Esc and E(z) to Pho. Next PRC2 trimethylates H3-K27 at PREs and this mark is recognized by the chromodomain of Pc, which results in PRC1 recruitment and H2A-K118 monoubiquitination (Wang et al., 2004b).

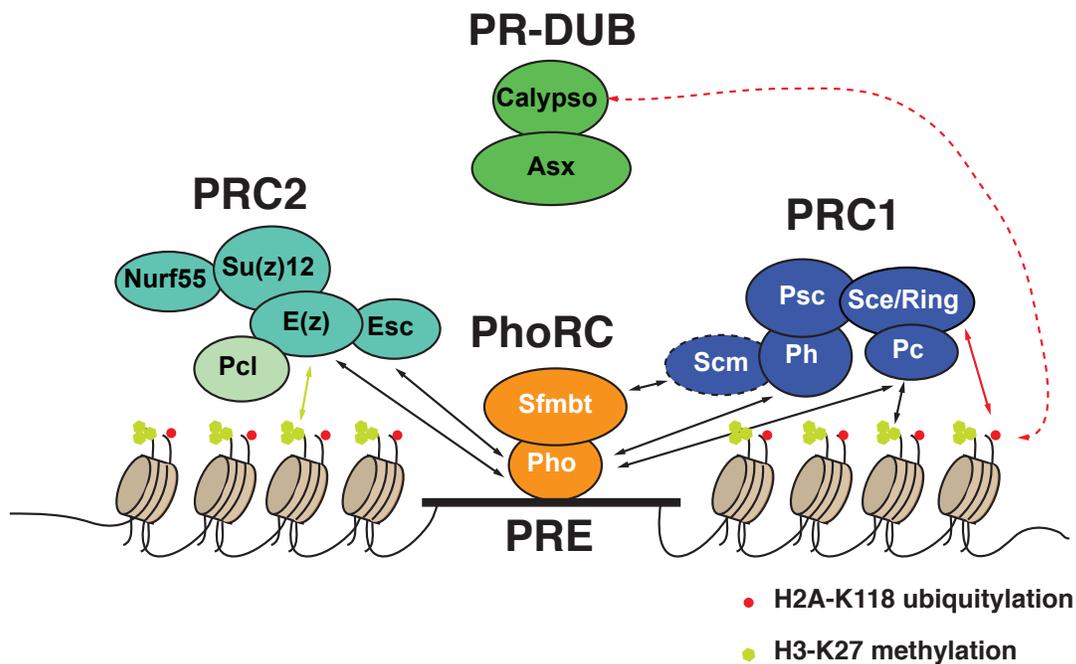


Figure 1.4: Recruitment of PcG complexes to target genes.

Model of how PcG proteins bind at PREs and assert their chromatin modifying activities to repress their target genes. PhoRC binds at PREs via Pho and recruits the other PcG complexes. PRC1 ubiquitinates H2A-K118 via Sce/ Ring and PRC2 mono-, di- and trimethylates histone H3-K27.

However, this hierarchical model has been challenged by several findings that are not compatible with this model. First, genome-wide binding studies in *Drosophila* have shown that PcG proteins form sharp binding peaks at many PREs, whereas the H3K27me3 repressive mark presents broad domains including the regulatory, promoter and coding region of genes. Only the Pc protein itself features a slightly broader distribution trailing with the H3-K27me3 mark in the PRE flanking regions (Kahn et al., 2006; Schuettengruber et al., 2009; Schwartz et al., 2006; Tolhuis et al., 2006). Second, PREs have been shown to be nucleosome-depleted due to high nucleosome turn-over (Deal et al., 2010; Mishra et al., 2001; Mohd-Sarip et al., 2006; Papp and Müller, 2006) making it unlikely that nucleosome modifications are crucial for recruitment to PREs. Third, Pho has not only been reported to directly interact with PRC2 subunits Esc and E(z) (Wang et al., 2004b), but also with the PRC1 components Ph and Pc (Mohd-Sarip et al., 2002). Furthermore, Pho and PRC1 can be co-assembled *in vitro* on naked PRE DNA templates in absence of nucleosomes (Mohd-Sarip et al., 2005).

So how are PcG proteins targeted to PREs? The emerging picture is that PcG proteins are targeted by protein-protein interactions with multiple DNA-binding recruiters such as Trl/ GAF, Psq and Zeste (reviewed in (Müller and Kassis, 2006)). These DNA-binding proteins could provide a platform for assembly of the other PcG complexes by multiple protein-protein interactions. For example the DNA-binding PcG complex PhoRC has been shown to interact with several PRC1 members, including Ph, Pc and Scm (Mohd-Sarip et al., 2002; Grimm et al., 2009). Additionally, the substoichiometric PRC2 members Jarid2 and Pcl have been implicated in PRC2 recruitment (Savla et al., 2008; Herz et al., 2012a). This suggests that interaction and modification of PRE-bound PcG proteins with flanking nucleosomes are not crucial for their recruitment, but instead play an important role in maintaining a repressed chromatin state at promoters and coding regions of target genes (Müller and Kassis, 2006). This does not exclude that histone modifications deposited by PRC1 and PRC2 are important for fine-tuning chromatin repression by for example stabilizing the bound complexes (Ringrose and Paro, 2004). In addition, the PRC1-H3-K27m3 interaction has been proposed to mediate looping to facilitate contacts between PRE-bound PcG proteins and flanking nucleosomes in the promoter region that have to be modified (Müller and Kassis, 2006).

1.5.4 Different PcG recruitment mechanisms in fly and mammals

In contrast to the 200-400 PREs described in *Drosophila*, only two PRE-like elements have been identified in mammals (Sing et al., 2009; Woo et al., 2010). The first element – PRE-kr – was identified in the mouse Kreisler gene and was reported to recruit mainly PRC1 complexes (Sing et al., 2009). The second PRE-like element was identified in the human homeobox D (HOXD) cluster and has been shown to recruit both, PRC1 and PRC2 complexes (Woo et al., 2010). Strikingly, most of the specific DNA binding factors implicated in recruitment of PcG proteins in *Drosophila* – with the exception of YY1 (Pho) (Brown et al., 1998), HMGB2 (DSP1) (Déjardin et al., 2005) and GAF (GAGA) (Nègre et al., 2006) – are not conserved in mammals suggesting that the PcG recruitment mechanisms in mammals and flies differ. Furthermore, genome wide studies in mammals did not detect a clear overlap between YY1 binding sites and PcG target genes (Xi et al., 2007). However, other sequence specific DNA-binding factors might still play a role in recruitment of PcG proteins in mammals, for example the DNA binding proteins REST, RUNX1/CBF β and SNAIL have been reported to partially colocalize with PcG sites and also have been implicated in PcG targeting (Arnold et al., 2013; Dietrich et al., 2012; Yu et al., 2012).

Interestingly, genome wide analysis of PcG targets in ES cells revealed that PcG proteins bind to clusters of unmethylated CpG islands (CGIs) that are depleted of activating factors (Ku et al., 2008). This was further confirmed by two studies reporting that GC rich elements derived from bacterial origin can recruit PRC2 complexes (Mendenhall et al., 2010) and that *de novo* PRC2 recruitment occurs at normally methylated CpG-rich elements in DNA methyltransferase-deficient ES cells (Lynch et al., 2012), respectively. CGIs are short (1-2 kilobase), highly conserved regions of DNA concentrated near the transcription start sites of genes (Blackledge and Klose, 2011; Deaton and Bird, 2011; Sharif et al., 2010).

The situation in mammals seems to be more complex than in flies and it has not been completely resolved how the PcG complexes are recruited to CGIs. PRC1 recruitment in mammals is thought to depend on a combination of CpG Island binding proteins that interact with and target PRC1 subunits and binding to the H3K27me3 mark deposited

by PRC2. Non-canonical PRC1 complexes lack CBX proteins and are thought to rely on sequence specific DNA-binding proteins for recruitment to target sites. For example KDM2B is a H3-K36 demethylase and a CpG binding protein. Consistently, KDM2B has been reported to be part of one PRC1 variant complex – PRC1.1 (dRAF in flies) – and is involved in recruitment of this complex to a subset of CpG islands (Blackledge et al., 2010; Farcas et al., 2012; Wu et al., 2013). Another example of a CpG binding protein involved in PRC1 targeting is E2F6, which is part of the PRC1.6 complex (PhoRC-L complex in flies) (Alfieri et al., 2013; Gao et al., 2012; Trojer et al., 2011). Surprisingly, only two out of six reported PRC1 variant complexes contain a version of the Pc homolog CBX, which could mediate H3-K27me3 binding via its chromodomain (Gao et al., 2012). Furthermore, a study addressing the role of the H3-K27me3 binding CBX chromodomains in dynamics and distribution of the CBX protein reported that deletion of the chromodomain had only minor effects on CBX distribution (Ren et al., 2008), indicating that there exist PRC2 independent recruitment mechanisms even for CBX containing PRC1 complexes.

Remarkably, several PRC2 subunits have been reported to interact with chromatin. The combination of these relatively weak interaction could be sufficient for PRC2 recruitment to CGIs. PRC2-chromatin interactions include binding of the histone chaperone Nurf55 homologs RBAP46 and RBAP48 to histones H3 and H4 (Song et al., 2008) and binding of EED, which is the mammalian Esc homolog, to the H3-K27me3 mark (Margueron et al., 2009). In addition, PRC2 members that are not part of the core complex have been shown to be important for recruitment of other PRC2 components to subsets of target genes. JARID2 and AEBP2 (Jing in flies), for example, have been reported to play a role in PRC2 recruitment by direct interaction with CpG islands. Both proteins harbor zinc finger domains that have been implicated in binding of GC-rich DNA sequences (Kim et al., 2009; Li et al., 2010a; Peng et al., 2009). Furthermore, the three mammalian Pcl homologs (PHF1, MTF2 and PHF19) can bind to H3-K36me_{2/3} via their respective Tudor domains and thus assist in PRC2 recruitment. Since H3-K36me_{2/3} is a mark for active chromatin the Pcl homologs have been implicated in the initiation of PRC2 spreading into active chromatin (Cai et al., 2013), which plays a role during the

differentiation process, where genes have to be switched from the active to the repressed state (Ballaré et al., 2012; Brien et al., 2012).

Furthermore, PRC1 and PRC2 might contribute to each other's recruitment by positive feedback loops driven by their histone modifying activities. Both, the PRC2 mediated H3-K27me3 mark (Cao et al., 2002; Wang et al., 2004b) and the PRC1 mediated H2A-K119ub mark (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014); have been shown to recruit the other PcG complex, respectively. In addition, non-coding RNAs have been implicated in targeting of both PcG complexes, PRC1 and PRC2, in mammals (Pandey et al., 2008; Rinn et al., 2007; Yap et al., 2010).

1.6 The Polycomb group protein Polyhomeotic

The PcG protein Polyhomeotic (Ph) is a core component of the PRC1 complex (Francis et al., 2001; Shao et al., 1999). The *ph* locus encodes two proteins with nearly identical amino acid sequences – Ph-proximal (Ph-p) and Ph-distal (Ph-d) – and is located on the X-chromosome (Dura et al., 1987). The Ph-d protein only differs from the Ph-p by the absence of about 200 amino acids in the N-terminal part and a small region of sequence divergence close to the C-terminus. Both Ph proteins possess three described functional domains that are conserved in mammalian Polyhomeotic homologs: a homology domain (HD), a FCS-Zn finger as well as a SAM domain (Hodgson et al., 1997) (Figure 1.5 A). In *Drosophila* removal of only one of the *ph* homologues results in wild-type-like embryos and viable adults with mild homeotic transformations, whereas removal of both *ph* homologues results in early arrest of development (12 h post fertilization) accompanied by embryonic epidermal defects in the thoracic and abdominal segments and absence of dorsal closure (Dura et al., 1987) (Figure 1.5 B). Furthermore, Ph null mutant clones in imaginal discs exhibit misexpression of several PcG target genes (*Abd-B*, *Ubx* and *cad*) and tumor formation (Beuchle et al., 2001). Interestingly, a Ph-dependent tumor phenotype has also been reported following overexpression of the Ph-p protein (González et al., 2009) suggesting that Ph can act as both, a tumor suppressor as well as an oncogene.

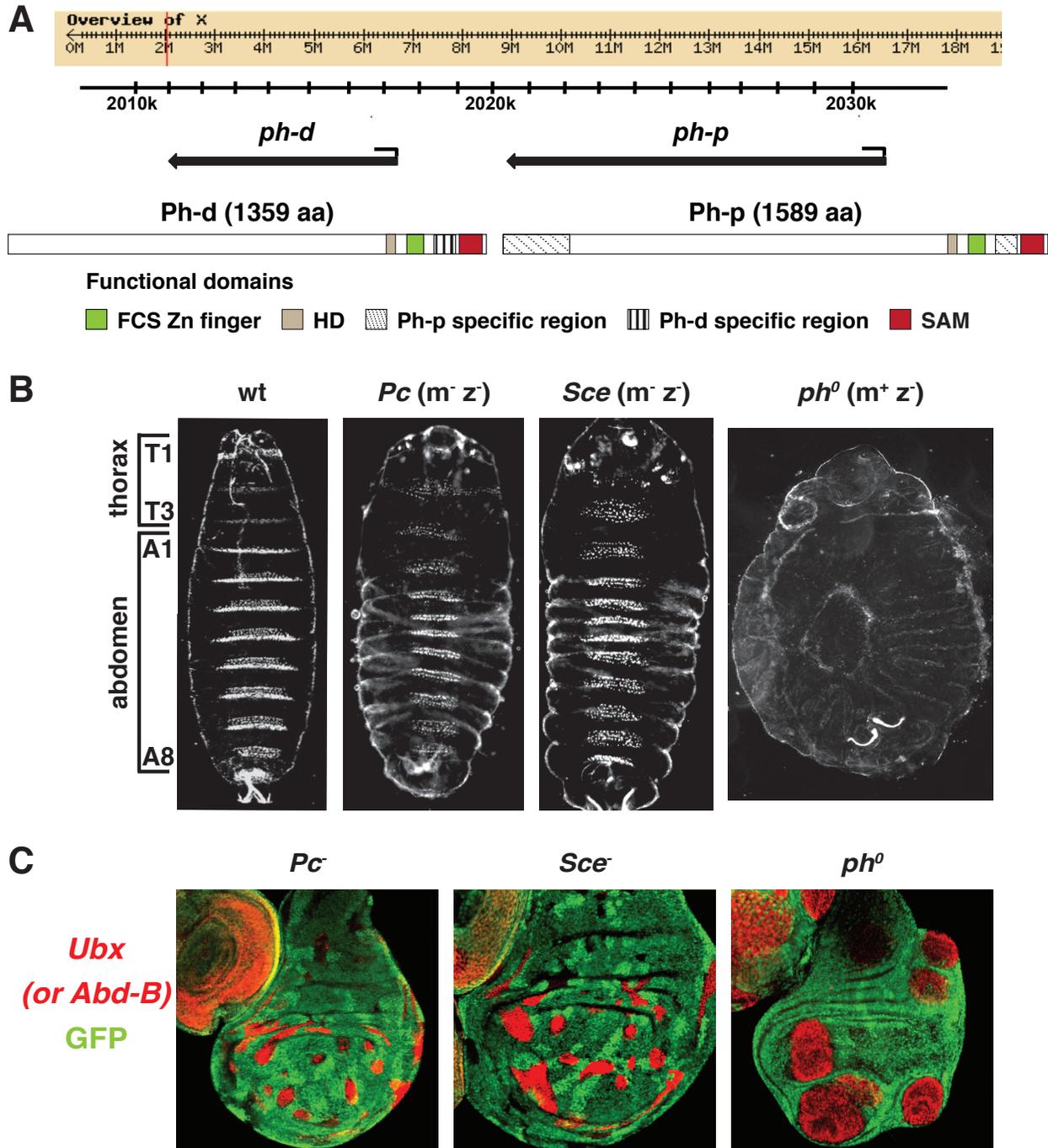


Figure 1.5: Particular role of Polyhomeotic (Ph) genes and proteins in *Drosophila*.

A: *Polyhomeotic* is located on the X-chromosome and encodes two proteins Polyhomeotic-proximal and -distal. Both proteins share three functional domains: FCS Zn finger: phenylalanine (F)–cysteine (C)–serine (S) sequence motif zinc finger; HD: homology domain; SAM: sterile alpha motif. B: Fixed cuticles of wt (wild-type), *Pc* ($m^- z^-$), *Sce* ($m^- z^-$) and *Ph*⁰ ($m^+ z^-$) embryos. Note that embryos lacking Ph display a

more severe phenotype with absence of dorsal closure compared to embryos lacking Pc or Sce/ Ring. Pc and Sce/ Ring mutants show homeotic transformation in their posterior parts. m^- : no maternal contribution, m^+ : maternal contribution, z^- : no zygotic contribution. Pictures adapted from (Gutiérrez et al., 2012). C: Imaginal wing disc clones homozygous for Pc^- , $Sce/ Ring^-$ or Ph^0 (ph^{504} allele) (GFP negative cells) 96 h after induction. Clones were stained for *Ubx* or *Abd-B* (in red) expression, which is repressed in wild-type cells. Note that all three mutant clones show misexpression of PcG target genes, but only Ph^0 clones display a tumor phenotype. Pictures adapted from (Beuchle et al., 2001).

1.6.1 Molecular Function of Polyhomeotic

Besides its role in maintaining segmental identity as a member of the PRC1 complex, Ph has been shown to be crucial for epidermal development in *Drosophila* embryos and Ph mutants display the most severe embryonic phenotypes amongst all PcG mutants (Figure 1.5 B). However, it is not clear if the role of Ph in epidermal development is linked to its role in body patterning or if the Ph protein has two independent functions (Dura et al., 1987). Intriguingly, both Ph-p and Ph-d have PRE elements in their promoter-proximal regions indicating that their expression levels have to be tightly regulated (Fauvarque et al., 1995; Oktaba et al., 2008). Even within the PRC1 complex, Ph appears to play a unique role. Ph and Psc, for instance, are essential for the transcriptional repression of a group of target genes that do not require other PRC1 subunits (Gutiérrez et al., 2012). In addition, Ph and Psc have been reported to mediate PRC1's ability to inhibit chromatin remodeling and induce chromatin compaction (Francis et al., 2001; King et al., 2002). In mouse, Ph plays a crucial role in gene repression by inducing and stabilizing PRC1 clustering via polymerization of its SAM domain (Isono et al., 2013). Interestingly, Ph has been shown to be O-GlcNAc modified by the PcG member Sxc/ Ogt (Gambetta et al., 2009). Studies in our lab suggest that this post-translational modification contributes to Ph function by preventing SAM domain-mediated aggregation of Ph (Gambetta, unpublished).

Despite the fact that the *ph-p* and *ph-d* genes function redundantly and their products can almost completely substitute for each other, they have been speculated to have distinct functions, acting as alternatives in different tissues, developmental stages or even at different target genes in the same cell. In line with that, a study examining

developmental regulation of the Ph proteins found that *ph-d* and *ph-p* are differentially regulated transcriptionally as well as post-transcriptionally during embryogenesis (Hodgson et al., 1997). However, distinct functions for Ph-p and Ph-d have not been clearly allocated and their redundancy is discussed controversially in the literature. A study investigating the role of different PcG proteins in mitosis, for example, reported that *ph-p* is necessary for normal mitosis whereas *ph-d* deletion did not have an effect on mitosis (O'Dor et al., 2006).

In addition to its role in mitosis, Ph has been shown to play a role in control of cell cycle regulation and proliferation. In absence of Ph, imaginal wing disc cells overproliferate and exhibit a shift in cell cycle phasing towards the G2/M phase (Beuchle et al., 2001; Oktaba et al., 2008). Analysis of PcG target genes in wing discs showed that several cell cycle regulator genes are amongst the PcG repressed targets including *Rbf*, *E2F* and *CycB* and these are partly deregulated in Ph mutant clones (Oktaba et al., 2008). Several other studies have examined the overproliferation due to altered Ph levels and have identified derepression of Notch and JAK/STAT signaling pathway members as a possible cause for loss of cell cycle control in mutant cells (Classen et al., 2009; Feng et al., 2011; González et al., 2009; Martinez et al., 2009). Therefore the Ph protein might play a particular role in inactivation of the Notch and JAK/STAT pathway during development. Strikingly, the PcG protein Psc, which together with its paralog Su(z)2 features a similar tumor phenotype as Ph (Beuchle et al., 2001; Oktaba et al., 2008), has been reported to play a direct role in cell cycle progression by mediating proteasome degradation of Cyc-B in mitosis (Mohd-Sarip et al., 2012). However, such a direct role in cell cycle control has not been identified for Ph proteins.

1.6.2 Interacting partners of Polyhomeotic

Polyhomeotic is part of the Polycomb Repressive complex 1 (PRC1). Within the PRC1 complex it has been reported to directly interact with Psc via its homology domain (Kyba and Brock, 1998a) and with the substoichiometric PRC1 member Scm via its SAM domain (Peterson et al., 1997; Kyba and Brock, 1998b; Kim et al., 2005). In addition, the SAM domain of Ph can also mediate homopolymerization of Ph molecules (Kyba

and Brock, 1998b; Kim et al., 2002). Furthermore Ph has been reported to directly interact with the PhoRC member Pho (Mohd-Sarip et al., 2002).

Moreover, Ph has been published to interact with non-PcG proteins. A study in *Drosophila* SL-2 cells, for example, reported that Ph physically interacts with the two chromatin-condensation proteins Topoisomerase II (TOPOII) and Barren (BARR), which also co-localize with PcG proteins to PRE elements in the bithorax cluster (Lupo et al., 2001). In addition to that, a study in *Drosophila* Kc1 cells found Ph associated with the chaperones Hsc4 and Droj2 (Wang and Brock, 2003).

1.7 Aims of this study

Even though the biochemical mechanisms of the PcG system are a field of intense study, the molecular functions of many PcG members are still elusive. The general aim of this study was to get a deeper insight into the biological role of the Polycomb group protein Polyhomeotic (Ph). Ph is part of the Polycomb Repressive complex 1 (PRC1); however, its role within the PRC1 complex is not well understood. Ph null mutants show very severe phenotypes that are not shared by mutants lacking other PRC1 subunits, indicating that Ph either plays a crucial role within PRC1 or might have an additional important function outside of PRC1. This study focused on gaining a better insight into the molecular interactions and mechanisms of the *Drosophila* Ph protein by conducting Tandem Affinity purifications (TAP) with the two Ph paralogs, Ph-p and Ph-d. The specific aims of the purification were (1) to identify novel Ph interacting partners associated with the PRC1 complex, (2) to determine if Ph is present in other complexes outside PRC1 and (3) to analyze if the two paralogs Ph-p and Ph-d associate preferentially with different interactors and thus might play distinct roles in the fly.

In addition to the previously described PRC1 core subunits Psc, Pc and Sce, Scm and Sfmbt were identified as two major Ph interactors. Intriguingly, these two PcG proteins form a molecular link between the PRC1 complex and the DNA binding PhoRC complex suggesting that Ph plays a crucial role in recruitment of PRC1 to Polycomb target genes. Therefore, the next focus of this study was to map the domains important for interaction of PRC1 and PhoRC complex to gain a better insight into PcG recruitment to target genes. These studies revealed that the SAM domains of Ph, Scm and Sfmbt

mediate the interaction between these three proteins. SAM domains are known for their polymerization properties, can form homo- and heteropolymers and have been linked to transcriptional repression. Using X-ray crystallography, I could determine the structure of the complex formed by the SAM domains of Sfm1t and Scm. Together; these studies thus advance our understanding of the molecular interactions that permit assembly of PcG protein complexes at their target genes.

2. Material and Methods

2.1. Materials

2.1.1 Oligonucleotides used in this study

Table 2.1: Sequencing primers used in this study.

Primer Name	Start Site	5'-Sequence-3'	T _m (°C)	Length (bp)
Sfmbt fw1 356	356	ATAACGATGTGGAGAACGGC	60.0	20
Sfmbt fw2 757	757	GAAAACACATCCAGCAAGCA	59.8	20
Sfmbt fw3 1151	1151	GTTCCGGAGTCAGCCGATTT	60.2	19
Sfmbt fw4 1557	1557	GGAGAAGATAATCAGCATTCCGG	60.1	22
Sfmbt fw5 1950	1950	CGAGCACAAGTACAAGGACTG	58.6	21
Sfmbt fw6 2357	2357	TCGACGAATACTACAGTGACGG	60.2	22
Sfmbt fw7 2750	2750	ATCTCATGGATCCCCGACTC	61.2	20
Sfmbt fw8 3152	3152	ATCAGCCCCGAGGAGGAAG	60.3	18
Sfmbt fw9 3559	3559	GTGGGCCCCAGCACTGAAAAT	64.5	20
Sfmbt rv1 440	440	TTGTCCCTCTGACCGTACACT	59.6	21
Sfmbt rv2 840	840	ACTGCAGGTGGCGGTGGT	65.2	18
Sfmbt rv3 1240	1240	TCGCATTCTTAATGTGCGCC	64.3	20
Sfmbt rv4 1648	1648	CCGCATTGAAGTTCTCCTTG	60.8	20
Sfmbt rv5 2050	2050	GGAGGCTGTCGTTGATTTTG	60.6	20
Sfmbt rv6 2443	2443	TGGATGAAAGGTTGAGTGGA	59.1	20
Sfmbt rv7 2844	2844	GTACTIONCATCCGTCCATCCGT	59.8	20
Sfmbt rv8 3250	3250	ACGACTGTGTCGTTGACTGC	59.9	20
Sfmbt rv9 3645	3645	GTTGGTCTTGTGGGATCTGG	60.4	20
Scm f1 356-375	356	AGTTCTGCTCGATCATGTGC	59.0	20
Scm f2 823-842	757	GATTCTACGGAGATCCATGCT	58.2	21
Scm f3 1159-1178	1159	CGCATGGACTCCAGCTCTA	60.1	19
Scm f4 1552-1572	1552	GCGAATCTCATTTCCTTGGT	59.1	20
Scm f5 1960-1980	1960	ACGAATAGCTCTGCGACAAA	58.7	20
Scm f6 2359-2377	2359	GTGAGCACGCCCACTTCT	60.0	18
Scm r1 445-427	445	CCAGACCCTTGCCCGATG	65.0	18
Scm r2 841-823	841	ATGCGTTCATGCGGAATC	60.6	18
Scm r3 1242-1224	1242	AGCGGGAACCATAGCCTC	60.2	18

Primer Name	Start Site	5'-Sequence-3'	Tm (°C)	Length (bp)
Scm r4 1642-1624	1642	GCGGTGTGGCTGACTGAG	61.7	18
Scm r5 2042-2023	2042	GTTGACGAGCTGCTGCAAT	60.2	19
Scm r6 2440-2419	2440	ACTGGATCACCTCTTCGATGG	61.4	21
Ph-d f1	648	CTACGCCATCCAGGTGAAGC	66.9	20
Ph-d f2	1368	CAAGGTGGGAGTGGATGC	64.5	18
Ph-d f3	2035	ACACAGACGGCCCAGAACC	67.6	19
Ph-d f4	2718	GCAGGCCCAAGCCCAAGT	69.9	18
Ph-d f5	3378	CCCGACCAAGGAAACACC	65.4	18
Ph-d f6	4077	CAGCGATATGGTGGCCTGC	69.0	19
Ph-d r1	679	TGGGAAATTCCTGCTTCACC	66.1	20
Ph-d r2	1385	GCATCCACTCCGACCTTGG	68.1	19
Ph-d r3	2040	CTGTGTGGACACGGAGCTGG	69.5	20
Ph-d r4	2726	TGTGTCTGGGCCTGCTGC	69.2	18
Ph-d r5	3398	CTGGGTGTTTCCTTGGTCG	65.5	19
Ph-d r6	4098	CTCGCAGGCCACCATATCG	68.9	19
Ph-p f1	648	CTACGCCATTGAGGTGAAGC	65.0	20
Ph-p f2	1368	CAAGGTGGGAGTGGATGC	64.8	18
Ph-p f3	2035	ACACAGACTGCCCAGAACC	63.4	19
Ph-p f4	2718	GCAGGCACAAGTTCAAGC	62.8	18
Ph-p f5	3378	GCCCACCAAAGAGACACC	63.8	18
Ph-p f6	4077	CTCGGATATGGTTGCTTGC	63.9	19
Ph-p r1	679	TGGGAAACTCTTGCTTCACC	64.5	20
Ph-p r2	1385	GCATCCACTCCCACCTTGG	68.3	19
Ph-p r3	2040	CTGTGTGGATACGGAAGTGG	63.5	20
Ph-p r4	2726	TGTGCCTGCGCCTGTTGC	72.7	18
Ph-p r5	3398	GAAGGTGTCTCTTTGGTGG	60.3	19
Ph-p r6	4098	CTCGCAAGCAACCATATCC	63.6	19
CaSpeR f2	-	TCGTGTTTACTGTTTATTGC	55.4	20
CaSpeR r1	-	CGGTGCCTATATAAAGCAGC	61.2	20
pFasBac fw	-	TGTTGCGCCAGGACTCTAGC	67.0	20
pFasBac rev	-	CAAACCACAAGTAAATGCAGTG	66.7	23
T7 promoter fw	-	TAATACGACTCACTATAGGG	50.9	20
T7 terminator rv	-	CTAGTTATTGCTCAGCGGT	57.1	19

Table 2.2: PCR/ Cloning primers used in this study.

Cloning Project	Primer Name	5'-Sequence-3'	T _m (°C)	Length (bp)
CaSpeR vector				
Ph-p in CaSpeR N-TAP via EcoRI/NotI	EcoRI-Ph-p fw	cgGAATTCatggatcgtcgtgcattg	76.7	26
	NotI-Ph-p rev	tcatcGCGGCCGCtactcgcctcctggatcc	88.8	32
Ph-p in CaSpeR C-TAP via EcoRI/NotI	EcoRI-Ph-p fw	cgGAATTCatggatcgtcgtgcattg	76.7	26
	NotI-Ph-p rev no st	tcatcGCGGCCGCctcgcctcctggatccttg	90.7	32
Ph-d in CaSpeR C-TAP via EcoRI/NotI	Eco-RI-Ph-d fw	ccGAATTCatgccccacggcttcggagc	84.3	28
	NotI Ph-d rev2 no st	tcaGCGGCCGCcatccttcacgtcgcgggtggcac	95.0	36
pFasBac vector				
Sfmbt530-1220 in pFB vector via EcoRI/NotI	EcoRI-Sfmbt530-1220 fw plus start	tcatcGAATTCatgtatgatcccacacactcct	76.0	33
	NotI-Sfmbt530-1220 rev	tcatcGCGGCCGCtactataaaaacggtgatt	79.4	33
Sfmbt530-1220 in pFB-Flag vector via EcoRI/NotI	EcoRI-Sfmbt530-1220 fw no start	tcatcGAATTCtatgatcccacacactcct	73.1	30
	NotI-Sfmbt530-1220 rev	tcatcGCGGCCGCtactataaaaacggtgatt	79.4	33
Sfmbt 530-1136 in pFB vector via EcoRI/NotI	EcoRI-Sfmbt530-1136 fw plus start	tcatcGAATTCatgtatgatcccacacactcct	76.0	33
	NotI-Sfmbt530-1136 rev	tcatcGCGGCCGCctacaccagctccaagtgag	86.5	33
Sfmbt530-1136 in pFB-Flag vector via EcoRI/NotI	EcoRI-Sfmbt530-1136 fw no start	tcatcGAATTCtatgatcccacacactcct	73.1	30
	NotI-Sfmbt530-1136 rev	tcatcGCGGCCGCctacaccagctccaagtgag	86.5	33
Sfmbt1-530 in pFB-HA vector via EcoRI/NotI	EcoRI-Sfmbt1-530 fw no start	tcatcGAATTCatgaacctccgagctgcgca	85.0	33
	NotI-Sfmbt1-530 rev	tcatcGCGGCCGCtactacatgcgccgccaatgc	91.7	36

Cloning Project	Primer Name	5'-Sequence-3'	Tm (°C)	Length (bp)
Ph1298-1589 in pFB-HA vector via EcoRI/NotI	EcoRI-Ph1298-1589 fw no start	tcatcGAATTCaaggcgatgattaagcc	73.9	28
	NotI-Ph1298-1589 rev	tcatcGCGGCCGCctactagggggaggtgt	84.0	30
pEC vectors (LIC cloning)				
Sfmbt-SAM aa1137-1201 in pEC-k-3c-GST (ColE1) vector	Sfmbt-SAM AA1137-3C fw	ccaggggcccgactcgatgCCGATACATGGAACGTGTACGACGTATCC	91.5	49
	Sfmbt-SAM AA1201-3C rv	cagaccgccaccgactgcttaCTTAAGCTGTGC AATAAGGTCCGAAATTTTCAGTGCTGG	88.9	60
Sfmbt-SAM aa1137-1220 in pEC-k-3c-GST (ColE1) vector	Sfmbt-SAM AA1137-3C fw	ccaggggcccgactcgatgCCGATACATGGAACGTGTACGACGTATCC	91.5	49
	Sfmbt-SAM AA1220-3C rv	cagaccgccaccgactgcttaTAAAAACGGTGA TTTGTTGGTCTTGTGGGATCTGGCTC	89.8	59
Scm-SAM aa803-870 in pEC-S-CDF-3c-His vector	Scm-SAM AA803-3C fw	ccaggggcccgactcgatgCCGATCGACTGGA CCATCGAAGAGGTG	93.8	46
	Scm-SAM AA870-3C rv	cagaccgccaccgactgcttaACCATTGACCTTA TTGACTAAATTGCAGATTTTTAAGGCTG GTCC	87.6	66
Scm-SAM aa803-877 in pEC-S-CDF-3c-His vector	Scm-SAM AA803-3C fw	ccaggggcccgactcgatgCCGATCGACTGGA CCATCGAAGAGGTG	93.8	46
	Scm-SAM AA877-3C rv	cagaccgccaccgactgcttaGAGTGCCAGATT ATTCGACGACCATTGACCTTATTGAC	88.7	60
Sfmbt-ZNF aa323-365 in pEC-k-3c-GST (ColE1)/ pEC-S-CDF-3c-His vector	Sfmbt-ZNF aa323-365 3C LIC fw	ccaggggcccgactcgatgATACAGAAAGATG GCATGGCATGGCTGTC	88.7	43
	Sfmbt-ZNF aa323-365 3C LIC rv	cagaccgccaccgactgcttaAAGTACCAACGA ATATAGTTCCGCC	83.4	45
Scm-ZNF1+2 aa55-135 in pEC-k-3c-GST (ColE1)/ pEC-S-CDF-3c-His vector	Scm-ZNF1+2 aa55-135 3C LIC fw	ccaggggcccgactcgatgGGTAGACCGGCCA AACGAG	90.4	38
	Scm-ZNF1+2 aa55-135 3C LIC rv	cagaccgccaccgactgcttaCGAGCAGTTCTT CTTCTGGTG	86.2	42
Ph-p-ZNF aa1357-1406 in pEC-S-CDF-3c-His vector	Ph-p ZNF aa1357-1406 3C LIC fw	ccaggggcccgactcgatgGCTCCAGGCTCGG ATATGG	90.7	38
	Ph-p ZNF aa1357-1406 3C LIC rv	cagaccgccaccgactgcttaTCCACCGATGCC GTTCTTTG	88.7	41

Table 2.3: PCR/ Mutagenesis primers used in this study.

Mutagenesis Project	Primer Name	5'-Sequence-3'	T _m (°C)	Length (bp)
pFasBac vector				
Introduction of HA tag into pFB vector flanked by BamHI and EcoRI	5' BamHI-HA tag	GATCCatgtatccatgatgatgtccagattatgctG	73.9	36
	3' EcoRI-HA tag with start codon	AATTCagcataatctggaacatcatatggatacatG	72.0	36
Deletion of both ZNF domains (aa55-135) Scm1-877	ScmΔZNF aa55-135 fw	pho-ACCAGGCACAGCGGCGGATCGGCA T	84.2	25
	ScmΔZNF aa55-135 rv	pho-TCGTTGCCGTTGCGTGGACGCGGG	85.8	24
Deletion of ZNF domain (aa323-365) Sfmbt1-1220	SfmbtΔZNF aa323-365 fw	pho-AACACTAAGATGGAGGGAGACCAGG C	71.0	26
	SfmbtΔZNF aa323-365 rv	pho-GGGGATCACCGAAGGGTCAATGTTG	75.4	25
pEC vectors				
Scm-SAM: end-helix surface mutant Scm L859R	Scm-SAM L859R fw	pho-CTAGGACCAGCCCGAAAAATCTGCA AT	73.4	27
	Scm-SAM L859R rv	pho-CTTCAGGCCCATGTACTTCATCATC	68.6	25
Scm-SAM: end-helix surface mutant Scm L855E, L859E	Scm-SAM L855E, L859E fw	pho-TGGGCCTGAAGGAAGGACCAGCCG AAAAAATCTGCA	85.2	36
	Scm-SAM L855E, L859E rv	pho-TGTACTTCATCATCATCTCTGAATTT AGCAATAAA	68.0	36

2.1.2 Vectors used in this study

Three types of vectors were used in this study:

- 1) **pCaSpeR-*tub* vector** is a P-element containing Drosophila transformation vector with a white selectable marker. The insertion of target gene and white selectable marker is random. The CaSpeR vector drives ectopic expression under *α-tubulin1* promoter; it contains a 2.6 kb fragment of the *α-tubulin 1* gene, including promoter and 5'-untranslated region sequences (Struhl and Basler, 1993).
- 2) **pFastBac 1 (pFB)** vector is used for the Bac-to-Bac system to generate viruses. The vector is transformed into Dh10αBac, a specialized bacmid-containing *E. coli* strain and subsequent transposition allows for virus generation. pFB vectors contained either a Flag-tag, a haemagglutinin (HA)-tag or no tag. The original plasmid was purchased from *Invitrogen*.
- 3) **pEC-vector** is based on a low-copy cloning plasmid pUC 19 for expression of proteins in *E. coli* strains. It contains a T7 promoter and T7 terminator, a tag (either GST-His or His) and a 3C-PreScission site. In addition, it has an antibiotic resistance for selection and a Cro1 or CDF origin of replication. In this study the pEC-3c-Cro1-GST-His vector and pEC-3c-CDF-S-His vector were used.

Table 2.4: Vectors used in this study.

Vector	Reference	Vector	Reference
pCaSpeR- <i>tub</i> -TAP-Ph-p	(Frey, unpublished)	pFB-Flag-Scm ₁₋₈₇₇	(Grimm et al., 2009)
pCaSpeR- <i>tub</i> -Ph-p-TAP		pFB-Scm ₁₋₈₇₇	
pCaSpeR- <i>tub</i> -TAP-Ph-d		pFB-Flag-Scm ₁₋₈₇₇ ^{ΔZNF1+2}	(Frey, unpublished)
pCaSpeR- <i>tub</i> -Ph-d-TAP		pFB-Scm ₁₋₈₇₇ ^{ΔZNF1+2}	
pFB-Flag-Ph-p ₁₋₁₅₈₉	(Gambetta, unpublished)	pFB-Flag-Scm ₁₋₄₃₅	(Grimm et al., 2009)
pFB-Ph-p ₁₋₁₅₈₉		pFB-Scm ₁₋₄₃₅	
pFB-Flag-Ph-p ₁₂₈₉₋₁₅₈₉		pFB-Flag-Scm ₁₋₁₇₇	
pFB-Ph-p ₁₂₈₉₋₁₅₈₉		pFB-Scm ₁₋₁₇₇	
pFB-Flag-Sfmbt ₁₋₁₂₂₀	(Grimm et al., 2009)	pEC-S-CDF-3c-His-Ph-p ₁₃₅₇₋₁₄₀₆	(Frey, unpublished)
pFB-Sfmbt ₁₋₁₂₂₀		pEC-k-3c-GST-His-Sfmbt ₃₂₃₋₃₆₅	
pFB-Flag-Sfmbt ₁₋₅₃₀		pEC-S-CDF-3c-His-Sfmbt ₃₂₃₋₃₆₅	

Vector	Reference	Vector	Reference
pFB-Sfmbt ₁₋₁₂₂₀ ^{ΔZNF}	(Frey, unpublished)	pEC-k-3c-GST-His-Sfmbt ₁₁₃₇₋₁₂₂₀	(Frey, unpublished)
pFB-HA-Sfmbt ₁₋₅₃₀		pEC-k-3c-GST-His-Scm ₅₅₋₁₃₅	
pFB-Flag-Sfmbt ₅₃₀₋₁₂₂₀		pEC-S-CDF-3c-His-Scm ₅₅₋₁₃₅	
pFB-Flag-Sfmbt ₅₃₀₋₁₁₃₆		pEC-S-CDF-3c-His-Scm ₈₀₃₋₈₇₇ ^{L859R}	
pFB-Sfmbt ₅₃₀₋₁₂₂₀		pEC-S-CDF-3c-His-Scm ₈₀₃₋₈₇₇ ^{L855E, L859E}	
pFB-Sfmbt ₅₃₀₋₁₁₃₆			

2.1.3 Fly strains used in this study

Table 2.5: List of fly strains used in this study.

Strain	Used for	Reference
<i>w</i> ; <i>TAP-Ph-p/TM6C</i>	TAP purification	(Frey, unpublished)
<i>yw</i> ; <i>TAP-Ph-d/TM6C</i>	TAP-purification	(Frey, unpublished)
<i>w</i> ; <i>Ph-p-TAP/TM6C</i>	-	(Frey, unpublished)
<i>w</i> ; <i>Ph-d-TAP/TM6C</i>	-	(Frey, unpublished)
<i>w ph^{del}/FM7c</i>	Recombination crosses	(Feng et al., 2011)
<i>w ph^{del} FRT101/FM7c-twi::EGFP</i>	Rescue experiments	(Frey, unpublished)
<i>w ph^{del} FRT101/FM7c</i>	Rescue experiments	(Frey, unpublished)
<i>w ph⁵⁰⁴ FRT101/FM7c-twi::EGFP</i>	Rescue experiments	(Dura et al., 1987)
<i>w ph⁵⁰⁴ FRT101/FM7c</i>	Rescue experiments	(Dura et al., 1987)
<i>ph-d⁴⁰¹ w¹</i>	Rescue experiments	(Dura et al., 1987)
<i>ln(1)ph410, ph-p⁴¹⁰ w¹</i>	Rescue experiments	(Dura et al., 1987)
<i>w hs-nGFP hs-flp FRT101</i>	Rescue experiments	(Beuchle et al., 2001)
<i>w hs-nGFP hs-flp FRT101; TM3/TM6B</i>	Rescue experiments	(Frey, unpublished)
<i>w; TM3/TM6B</i>	Establishing transgene stocks	-
<i>yw; Dr/TM6C</i>	Establishing transgene stocks	-
<i>w; If/CyO</i>	Establishing transgene stocks	-

2.1.4 Baculoviruses generated and used in this study

Table 2.6: List of baculoviruses generated for this study.

Sfmbt viruses	Scm viruses	Ph viruses
Flag-Sfmbt ₁₋₁₂₂₀	Flag-Scm ₁₋₈₇₇	Flag-Ph-p ₁₋₁₅₈₉
Sfmbt ₁₋₁₂₂₀	Scm ₁₋₈₇₇	Ph-p ₁₋₁₅₈₉
Flag-Sfmbt ^{ΔZNF (323-365)}	Flag-Scm ^{ΔZNF1+2 (55-135)}	Flag-Ph-p ₁₂₉₈₋₁₅₈₉
Sfmbt ^{ΔZNF (323-365)}	Scm ^{ΔZNF1+2 (55-135)}	HA-Ph-p ₁₂₉₈₋₁₅₈₉
Flag-Sfmbt ₁₋₅₃₀	Flag-Scm ₁₋₄₃₅	-
HA-Sfmbt ₁₋₅₃₀	Scm ₁₋₄₃₅	-
Flag-Sfmbt ₅₃₀₋₁₂₂₀	Flag-Scm ₁₋₁₇₀	-
Sfmbt ₅₃₀₋₁₂₂₀	Scm ₁₋₁₇₀	-
Flag-Sfmbt ₅₃₀₋₁₁₃₆	-	-
Sfmbt ₅₃₀₋₁₁₃₆	-	-

2.1.5 Antibodies used in this study

Table 2.7: Primary antibodies used in this study.

Antibody	Antigen	Source	Reference	Use
anti-Ph (Ph2)	Ph-p aa766-984	rabbit polyclonal, affinity purified	(Oktaba et al, 2008)	1:10,000 in Wb
anti-Ph (Ph5)	Ph-p aa86-520	rabbit polyclonal, affinity purified	(Gambetta et al., 2009)	1:1000 in Wb
anti-Scm (MBT/3BWJ)	Scm aa174-435	rabbit polyclonal, crude serum	(Gambetta et al., 2009)	1:15,000 in Wb
anti-Psc (Psc7a)	Psc aa122-519	rabbit polyclonal, affinity purified	(Oktaba, unpublished)	1:250 in Wb
anti-Pc (PCCD2)	Pc	rabbit polyclonal, crude serum	(Papp and Müller, 2006)	1:50,000 in Wb
anti-Sce/ Ring	GST-Sce aa1-274	rabbit polyclonal, affinity purified	(Gorfinkiel et al., 2004)	1:10,000 in Wb
anti-Ogt	human Ogt aa1-300	rabbit polyclonal, affinity purified	Santa Cruz Biotechnology (sc-32921)	1:3000 in WB
anti-E(z) 3TAF	full length E(z)	rabbit polyclonal, crude serum	(Gambetta et al., 2009)	1:5000 in Wb

Antibody	Antigen	Source	Reference	Use
anti-Nurf 55	full length Nurf55	rabbit polyclonal, crude serum	(Gambetta et al., 2009)	1:50,000 in Wb
anti-dSfmbt (MBT)	Sfmbt aa531-980	rabbit polyclonal, affinity purified	(Klymenko et al., 2006)	1:1000 in Wb
anti-Pho (ZF)	Pho aa324-520	rabbit polyclonal, crude serum	(Papp and Müller, 2006)	1:10,000 in Wb
anti-dINO80	dINO80 aa1261-1510	rabbit polyclonal, crude serum	(Klymenko et al., 2006)	1:5000 in Wb
anti-Spt5	Spt5 aa732-1054	guinea pig polyclonal	(Andrulis et al., 2000)	1:50,000 in Wb
anti-HA	HA aa98-106	rabbit polyclonal, affinity purified	Sigma-Aldrich (H6908)	1:6000 in Wb
anti-Flag	Flag peptide	rabbit polyclonal, affinity purified	Sigma-Aldrich (F7425)	1:5000 in Wb
anti-lamin	nuclear lamin	mouse monoclonal	from D. Arndt Jovin, MPI Bioph Chem, Goettingen	1:300 in IF
anti-PAP (peroxidase anti-peroxidase)	HRP	rabbit polyclonal, affinity purified	Sigma-Aldrich (P1291)	1:1000 in Wb
anti-Abd-B	homeodomain fusion protein <i>trpE/Add-B</i>	mouse monoclonal	1A2E9, (Celniker et al., 1989)	1:300 in IF

Wb: Western blot, IF: Immunofluorescence staining, HRP: horse radish peroxidase.

Table 2.8: Secondary antibodies used in this study.

Antibody	Antigen	Source	Reference	Use
HRP anti-rabbit IgG	rabbit IgG	donkey	Amersham Biosciences (NA934)	1:5000 in Wb
HRP anti-mouse IgG	mouse IgG	sheep	Amersham Biosciences (NA931V)	1:5000 in Wb
Cy2 anti-mouse	mouse IgG	goat	Jackson, ImmunoResearch	1:500 in IF
Cy2 anti-rabbit	rabbit IgG	goat	Jackson, ImmunoResearch	1:500 in IF
Cy3 anti-mouse	mouse IgG	goat	Jackson, ImmunoResearch	1:500 in IF
Cy3 anti-rabbit	rabbit IgG	goat	Jackson, ImmunoResearch	1:500 in IF
Cy5 anti-mouse	mouse IgG	goat	Jackson, ImmunoResearch	1:500 in IF
Cy5 anti-rabbit	rabbit IgG	goat	Jackson, ImmunoResearch	1:500 in IF

Wb: Western blot, IF: Immunofluorescence staining, HRP: horse radish peroxidase

2.2 Methods

2.2.1 Fly husbandry and microscopic analysis of *Drosophila* tissues

Flies were grown on standard cornmeal-molasses-yeast medium supplemented with apple juice and methyl paraben as mold inhibitor at 25 °C with 65 % rH (relative humidity). Crosses and their offspring were kept at 25 °C or at 18 °C depending on how fast adult progeny needed to be obtained. Confocal imaging was performed at the Imaging Facility of Max Planck Institute of Biochemistry, on a ZEISS LSM780 confocal laser-scanning microscope equipped with a ZEISS Plan-APO 25x/NA0.8 oil immersion and a ZEISS Plan-APO 63x/NA1.46 oil immersion objective. Pictures of embryonic cuticles were taken with Axio Scope.A1 equipped with an EC Plan-Neo Fluor 10x/NA0.3 objective, DIC setting.

2.2.1.1 Generation of transgenic *Drosophila* stocks and mapping of transgene inserts

The four transgenes α -*tub1-TAP-ph-p*, α -*tub1-TAP-ph-d*, α -*tub1-ph-p-TAP* and α -*tub1-ph-d-TAP* were injected into a *w* host fly line by random insertion with a P-element. The flies developing from the injected flies were collected and crossed to *w; If/CyO* flies. The progeny was screened for red eye color (*w+*). Single male flies with red eyes were crossed to *yw; Dr/TM6C Sb Tb* female flies. The next generation of flies was collected and visible markers such as “Curly” (wings curve away from the body) or “stubble” (shorter and thicker bristles) were documented to map the insertion site of the transgene. Only a small percentage of the transgenic lines could be established as homozygous stocks, all the other fly lines were kept in a heterozygous state over either *CyO* balancer if the transgene was on the second chromosome or *TM6C Sb Tb* if the transgene was inserted on the third chromosome.

2.2.1.2 Generation of clones by mitotic recombination in *Drosophila* larvae

In a first step crosses with the corresponding fly strains, one containing an *wflpGN20F101* chromosome, the other one a F101 site combined with the allele to study, either ph^{del} (Feng et al., 2011) or ph^{504} (Dura et al., 1987) were set up. After egg lay for 24 h the parents were removed from the vial and F1 larvae were heat-shocked for 1 h at 37 °C in a programmable water bath to induce mitotic recombination. Next

larvae were kept at 25 °C for either a 72 or 96 h-period depending on the desired number of cell cycles after recombination. Prior to dissection, larvae were subjected to a second heat shock for 1 h at 37 °C to induce hs-GFP expression. For Ph rescue assays the *wflpGN20F101* chromosome was combined with the rescue transgene on the third chromosome in a first cross. Then *wflp GN20F101; TAP-Ph/ TM6C* males were collected from the next generation and crossed with *ph^{del} FRT101 (w+)/ FM7C-twi::EGFP* females. The progeny from this cross was used to induce mitotic clones with the *ph⁰* allele plus the rescue transgene.

2.2.1.3 Immunostaining of *Drosophila* larval discs

Third instar larvae were dissected in ice-cold PBS with fine forceps. First larvae were split in half; the carcasses were inverted and fat body, digestive system and salivary glands were removed carefully. Subsequently, the carcasses with attached disc tissues were fixed for 20 min in 4 % formaldehyde in PBS-T (0.1% Tween, PBS) and blocked by six 5 min washes with BBT (1% BSA, 0.1% Triton-X-100, PBS). Staining was performed at 4 °C overnight with the appropriate primary antibody (Table 2.7) diluted in BBT. After incubation with primary antibody, the carcasses were washed six times for 5 min at RT, before adding fluorescent-labeled secondary antibody (Table 2.8) and Hoechst dye diluted in BBT for a period of at least 2 h up to overnight. Next two washes with BBT and four washes with PBT were performed within 2 h. Finally, the wing discs were dissected of the carcasses and mounted in Fluoromount-G mounting medium (Fluoromount-G™, SouthernBiotech) before taking confocal images.

2.2.1.4 Preparation of *Drosophila* embryonic cuticles

Flies were crossed at 25 °C in vials and allowed to mate for 24-48 h. Next flies were transferred to a small cage and 12 h egg lays were collected on apple-agar plates at 25 °C. Embryos and non-hatched larvae were collected 24-48 h after laying and dechorionated with 3 % sodium-hypochlorite solution for 3 min on a small filter membrane (Nitrocellulose Membrane with grid, Type AABG, 0.8 µM, 37 mm diameter) and mounted on a vacuum device. After dechorionating, the embryos were washed extensively with PBS-T. If possible the vitelline membrane was carefully removed with a fine capillary. In a next step embryos were transferred from the filter to a slide and

embedded in Hoyer's medium containing lactic acid (1:1). A coverslip was carefully mounted on top and the slide was incubated at 65 °C for several hours until the cuticles appeared clear.

2.2.2 Insect cell culture and Baculovirus production for protein expression

Two different lines of insect cells were used for experiments in this study. Sf21 (IPLB-Sf21 AE, ovarian tissue, *Spodoptera frugiperda*, Invitrogen, Cat no.12682-019) cells were used for transfection and virus generation. High Five cells (BTI-TN-5B1-4 ovarian tissue, *Trichoplusia ni*, Invitrogen, High Five Frozen cells, P/N 51-4005) were used for protein expression.

2.2.2.1 Cultivation of insect cells

Insect cells were cultured in SF-900 III SFM (1x) serum free complete medium (Gibco, Life Technologies) or Express Five SFM (1x) serum free medium (Gibco, Life Technologies) supplemented with 18 mM L-Glutamine (200 mM, 25030-123, Gibco) in 0.5 l medium in 3 l disposable Erlenmeyer flasks with vent cap (431252, Corning) at 27 °C with gentle agitation (90 rpm). Cells were passaged three times per week and split to a density of 0.7-1 x10⁶ cells/ml for Sf21 cells and 0.4-0.8 x10⁶ cells/ml for High Five cells. Cell counts and viability were determined with a Vi-cell XR cell viability analyzer (Beckman coulter).

2.2.2.2 Bac-to-Bac procedure

The Bac-to-Bac procedure was used to generate viruses expressing the proteins of interest in high amounts in insect cell culture.

2.2.2.2.1 Transposition of cDNA from pFastBac into bacmid

The gene of interest was cloned into desired pFB vectors and 2 µl Miniprep DNA or 1 µl Maxiprep DNA were transformed into 100 µl of thawed DH10Bac *E. Coli* cells. After addition of DNA, cells were kept on ice for 30 min, heat-shocked at 42 °C for 45 sec and recovered on ice for 2 min. Then 900 µl of SOC medium was added and cells were shaken for 4 h at 37 °C. Subsequently, cells were spun down at 1500 x g for 4 min in a tabletop centrifuge and resuspended in 200 µl LB by gentle vortexing. For each transformation two selective plates (LB with 50 µg/ml kanamycin, 10 µg/ml tetracycline,

7 µg/ml gentamicin, 100 µg/ml Bluo-Gal, 40 µg/ml IPTG), one with 20 µl and the second one with 160 µl, were plated, sealed with parafilm and incubated for 48 h at 37 °C protected from light. Four to six single white colonies (positive for transposition) were picked and re-streaked on a new selective plate.

2.2.2.2.2 Bacmid purification

Confirmed white colonies were cultured in 3 ml preps (LB medium with 50µg/ml kanamycin, 10 µg/ml tetracycline, 7 µg/ml gentamycin for 20 h. Bacmid DNA was prepared from four independent clones for each construct using buffers from a plasmid Midi kit (QIAGEN). For processing 2 ml of the culture was transferred to a 2 ml-Eppendorf tube and spun for 3 min at 19,000 xg at RT in a tabletop centrifuge. After centrifugation the supernatant was aspirated and the pellet was resuspended in 300 µl P1 (QIAGEN resuspension buffer). Then 300 µl of buffer P2 (QIAGEN lysis buffer) was added and mixed by inverting the tube six times. After 5 min incubation 300 µl of buffer P3 (QIAGEN neutralization buffer) was added and mixed by inverting the tube six times. The lysate was incubated for 5 min on ice before centrifuging at 19,000 x g for 10 min at 4 °C. The supernatant was transferred to a new 2 ml-Eppendorf tube and spun again for 5 min at 19,000 x g at 4 °C. Again the supernatant was transferred to a new 2 ml-Eppendorf tube and supplemented with 800 µl isopropanol. The tube was inverted to mix and incubated for 10 min on ice. Subsequently, the sample was spun at 19,000 x g for 20 min at 4 °C and supernatant was discarded. The pellet was washed in 500 µl 70 % ethanol before spinning at 19,000 x g for 10 min at RT. Next the liquid was removed and the pellet was air dried at 37 °C. Finally, the DNA pellet was resuspended in 100 µl sterile ddH₂O.

2.2.2.2.3 Transfection of Sf21 cells

For transfection Sf21 cells were diluted in SF-900 III medium to 0.4 x 10⁶ cells/ml and 2 ml/well cell suspension was seeded into a 6-well plate (Tissue culture plate, 6 well, flat bottom, BD Falcon). Each transfection included two negative controls: 1) no bacmid DNA, but transfection agent (Cellfectin II, 10362100, Invitrogen), 2) neither bacmid DNA nor transfection agent. Then cells were left for 30-60 min at RT to settle down and

attach to the well surface. The transfection mix was prepared in two parts: First, bacmid DNA was prepared by adding 100 μ l of SF-900 III medium to 1-3 μ l of the appropriate bacmid DNA in 1.5 ml-Eppendorf tubes. Second, masters mix containing 8 μ l of Cellfectin II and 92 μ l of SF-900 III medium for each transfection reaction was prepared. Next 100 μ l of the master mix was added to each bacmid DNA, flicked gently for mixing and incubated for 15-30 min at RT. In the next step 210 μ l of transfection mix was added dropwise to each respective well. Then the cells were incubated with the transfection mix for 3-5 h before exchanging it with 2 ml/well of fresh SF-900 III medium. The transfection plates were sealed with parafilm and incubated for 96 h at 27 °C. Cells were inspected by visual control to monitor the increase in size, which is observed upon transfection. After 96 h the *passage 1* virus (P1) was harvested and transferred to 15 ml-Falcon tubes.

2.2.2.3 Virus amplification

The *passage 1* (P1) virus was used in a ratio of 1:50 to infect 50 ml of 0.4×10^6 cell/ml Sf21 cells. The cells were incubated for 72 h at 27 °C at 90 rpm before spinning them down at 335 x g for 15 min. The supernatant containing the *passage 2* virus (P2) was harvested. The most active P2 virus for each construct was determined by visual control of pellet size and used to infect 250-500 ml of 0.4×10^6 cells/ml Sf21 cells in a ratio of 1:500. The cells were incubated for 72 h and the *passage 3* virus (P3) was harvested by spinning cells in 500 ml-centrifuge tubes (Corning) at 335 x g for 15 min and collecting the supernatant. The P3 virus was used to infect High Five cells for protein expression. A P3 screen determined the amount of virus necessary for expression. Viruses were kept at 4 °C and frozen following the TIPS (Titerless Infected-Cells Preservation and Scale-Up) protocol (Wasiliko and Lee, 2006) at (-80 °C) for long time storage.

2.2.2.4 Protein expression

For expression 0.4×10^6 cells/ml of High Five cells were infected with one to three viruses. The cells were infected in 500 ml aliquots and cultured in disposable 3 l Erlenmeyer with vent cap (Corning, 431252). If several viruses were used, the viruses were screened before use in a co-expression screen to get comparable expression levels for all proteins. After infection cells were incubated for 50-55 h at 27 °C shaking at

90 rpm. Cells were harvested by transferring cultures to 500 ml-buckets (500 ml centrifuge tubes, Corning) and centrifuging for 12 min at 931 x g. Subsequently, supernatant was decanted and pellets were either frozen in liquid Nitrogen and stored at (-80 °C) or processed immediately.

2.2.3 Work with bacteria

2.2.3.1 Transformation in bacteria

Competent cells (Dh5 α (NEB) for cloning, Rosetta (DE3) or Rosetta 2 (Novagen) for protein expression) were thawed for 10 min on ice before adding 1-2 μ l of plasmid DNA (1-200 ng) and flicking the tube carefully to mix cells and DNA. Next the mixture was incubated on ice for 30 min and subsequently heat shocked at 42 °C for 45 sec. Then the cells were kept 5 min on ice to recover before adding 950 μ l RT SOC. Cells were placed at 37 °C for 1 h shaking vigorously (250 rpm). Then several 10-fold serial dilutions were performed and 100-200 μ l of each dilution was plated onto the appropriate selective plate and incubated overnight at 37 °C.

2.2.3.2 Amplifying vectors for Mini, Midi or Maxi Prep

For vector amplification single colonies were picked from transformation plates and grown in an overnight culture at 37 °C in selective medium shaking vigorously (225 rpm). The culture volume was 3 ml for Miniprep, 50-100 ml for Midiprep and 200-300 ml for Maxiprep. The cultures were spun down at 6000 x g for 5-15 min and pellets were processed according to instruction manual of the Qiagen kits (QIAprep Spin Miniprep Kit #27104, Qiagen Plasmid Midi kit #12143, Qiagen Plasmid Maxi kit #12163).

2.2.3.3 Expression of proteins in bacteria

The pEC-vector was used for expression of proteins in bacteria. The vector was transformed into *E.coli* expression strains, either Rosetta (DE3) or Rosetta 2 (Novagen), plated on selective plates and incubated at 37 °C overnight. The transformed plates were used up to two weeks. In the morning a 50-100 ml pre-culture with LB medium and the selective antibiotics was set up and incubated for 4-5 h at 37 °C shaking vigorously at 225 rpm. After the pre-culture was grown dense it was used to inoculated the

expression culture, usually between 2 and 9 l of TB medium plus 10 % phosphor-solution and the selective antibiotics. 500 ml medium was distributed to five air flasks (5 l volume) and inoculated with the pre-culture. The culture was shaken at 37 °C at 225 rpm until the OD₆₀₀ reached 0.6. At this point the temperature of the shaker was switched down to 18 °C and the culture was induced with 0.5 mM IPTG as soon as the OD₆₀₀ had reached a value of 1.5. The cultures were incubated overnight (12-14 h) at 18 °C, 225 rpm and harvested the next morning in 1 l-centrifuge buckets by spinning at 6000 rpm for 15 min at 4 °C in the JA-16 rotor (Beckman Coulter). The pellet was either directly used for cell lysis or frozen in liquid Nitrogen and stored at (-80 °C).

2.2.4 Protein biochemistry

2.2.4.1 Preparation of protein extracts

In this study protein extracts were prepared from (1) *Drosophila* embryos for the initial TAP purification experiments, (2) insect cell culture to reconstitute the observed interactions *in vitro* or (3) bacterial expression strains Rosetta (DE3) or Rosetta 2 to express single protein domains for GST pull-down experiments and crystallization trials. All protein extracts were prepared at 4 °C.

2.2.4.1.1 *Drosophila* embryonic nuclear extracts

For nuclear extract preparation from *Drosophila* embryos, collected embryos (0-14 h) were washed through three consecutive sieves with decreasing pore sizes: 1) 0.75 µm, 2) 0.375 µm and 3) 0.125 µm. Subsequently, embryos were transferred to a beaker in 200 ml embryo wash (0.7 % NaCl, 0.04 % Triton-X100) and dechorionated by adding 60 ml of bleach while stirring for 3 min. Following that, embryos were washed for 5-10 min with deionized water in a 0.125 µm sieve and transferred to a 150 mm Millipore filter. The embryos were dried using a vacuum device and their weight was determined. Then nuclear extracts were prepared according to (Klymenko et al., 2006). The nuclear extract was dialyzed (Spectra/Por Dialysis Membrane MWCO: 6-8,000, Spectrum Laboratories) into a lower salt buffer (15 mM HEPES pH 7.6, 200 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 7.9, 20 % glycerol, 1 mM DTT) overnight with one buffer change after 1 h. The extract quality was estimated by measuring the total protein

concentration and by testing for selected PcG proteins in western blot. The nuclear extract was frozen in liquid Nitrogen in aliquots, and stored at (-80 °C).

2.2.4.1.2 Drosophila embryonic total extract

12-16 h old embryo collections were transferred with a brush to a small filter membrane (Nitrocellulose Membrane with grid, Type AABG, 0.8 µM, 37 mm diameter) mounted on a vacuum device, washed with PBS-T, dechorionated for 3 min with 5 % sodium hypochlorite and washed again. Subsequently, 100-200 embryos were transferred to a 2ml-eppendorf tube with a needle and resuspended in 2x LDS buffer in a ratio of 1 µl per embryo. Next the suspension was treated for 1 min with a Branson sonicator (MS73-tip, 0.5 on, 0.5 off, 33 % Amplitude) to disintegrate the embryos. Then the sample was boiled for 5 min at 95 °C and centrifuged at 19,000 x g for 10 min. Extracts were analyzed by western blot. Typically 10 embryos/lane gave a good signal for the PcG proteins probed for.

2.2.4.1.3 Insect cell cytosolic, nuclear and chromatin extracts

Insect cells were resuspended in lysis buffer (25 mM HEPES pH 7.9, 150mM NaCl, 10 % glycerol, 20µM ZnCl₂, 0.5 % NP-40, 0.5 mM DTT, 0.1 mM AEBSF, 1x complete protease inhibitor cocktail) in a ratio of 1ml/g of insect cell pellet. In order to lyse the cells, pellets were stirred for 1 h at 4 °C on a magnetic stirrer at 200 rpm. The lysate was centrifuged at 12,000 x g for 30 min in a precooled JA-25.50 rotor (Beckman Coulter) to pellet nuclei. The supernatant corresponds to the cytosolic fraction and was decanted and spun again at 12,000 x g for 30 min in order to remove remaining nuclei. After the second centrifugation step the cytosolic fraction was filtered with a syringe driven filter device (Millex-SV 5.00 µm filter unit). The nuclei were resuspended in 0.5 ml Low salt buffer (15 mM HEPES pH 7.6, 20 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 7.9, 20 % glycerol, 1 mM DTT, 1 mM AEBSF, 1x complete protease inhibitor cocktail) per gram of insect cell pellet weight and incubated for 10 min. Next an equal amount of High salt buffer (15 mM HEPES pH 7.6, 800 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 7.9, 20 % glycerol, 1 mM DTT, 1 mM AEBSF, 1x complete protease inhibitor cocktail) was added and nuclei were lysed on a rotating wheel for 30 min at 4 °C. Then

the suspension was spun in an ultracentrifuge for 1 h at 256,136 x g at 4 °C (SW40ti rotor, Beckman Coulter). The soluble nuclear fraction was decanted carefully and filtered with a syringe driven filter device 0.45µm (Millex-SV 5.00 µm filter unit). The chromatin pellet was solubilized by adding 1ml chromatin buffer 1 (50 mM Tris pH 7.9, 5 mM MgCl₂, 0.5 mM EDTA, 25 % glycerol, 20 µM ZnCl₂, 2 mM DTT, 0.1 mM Pefablok-SC, 1x complete protease inhibitor) per 4.17 x 10⁸ cells. The solution was homogenized with a dounce homogenizer (1500 rpm, 50 strokes) and 1/10 volume of 3M-ammonium sulphate was added. Next the sample was sonicated on dry ice (MS73, 0.5/0.5 pulser, 32 % Amplitude) for 50 min and spun for 1 h at 256,136 x g at 4 °C (SW40ti rotor, Beckman Coulter). Then the supernatant was collected and the ammonium sulphate was diluted to a final concentration of 0.1 M with the chromatin buffer. After diluting the ammonium sulphate concentration, the sample was spun for a second time for 1 h at 256,136 x g at 4 °C (SW40ti rotor, Beckman Coulter). The supernatant was collected and the ammonium sulphate concentration was increased to 40 % saturation by gradually adding solid ammonium sulphate powder while stirring on a magnetic stirrer at 4 °C to avoid local high concentrations of ammonium sulphate. Subsequently, the 40 % saturated ammonium sulphate solution was stirred for 30 min for equilibration. Then the solution was centrifuged for a third time for 1 h at 256,136 x g at 4 °C (SW40ti rotor, Beckman Coulter). The supernatant was discarded and the pellet with the precipitated chromatin proteins was resuspended in chromatin buffer 2 (15 mM HEPES pH 7.9, 0.2 mM EDTA, 20 % glycerol, 20 µM ZnCl₂, 2 mM DTT, 0.1 mM Pefablok-SC, 1x complete protease inhibitor). The chromatin extract was dialyzed (Spectra/Por Dialysis Membrane MWCO: 6-8,000, Spectrum Laboratories) into a low salt buffer (15 mM HEPES pH 7.9, 150 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 7.9, 20 % glycerol, 1 mM DTT) overnight with one buffer change after 1 h.

2.2.4.1.4 Bacterial total extracts

The bacterial pellets were resuspended in 25 ml lysis buffer (50 mM phosphate buffer pH 7.5, 250 mM NaCl, 10mM MgCl₂, 25 mM imidazole, 10 % glycerol, 0.1 % Triton-X100, 1 mM Pefablok-CS, 2 mM β-Mercaptoethanol, 1x complete protease inhibitor) per pellet from 1 l expression culture. The pellets were stirred on a magnetic stirrer for

10-15 min to completely resuspend them. Next the suspension was sonificated for 3x 5 min (VS70 tip, 80 % Amplitude, 5x) and between each sonification step the sample was stirred for 5 min on a magnetic stirrer. After sonification, the lysate was transferred to centrifuge tubes and spun for 30 min in a JA-25.50 rotor, 15,000 x g at 4 °C. The supernatant was filtered (5 µm filters) and directly used for the next purification step.

2.2.4.2 Protein purifications

2.2.4.2.1 Tandem-affinity purification of Drosophila nuclear extracts

Before purification IgG sepharose beads (Amersham) were cross-linked with Dimethylpimelimidate. Then 200 µl of cross-linked beads were transferred into an empty column (Bio-Rad 731-1550). To remove non-cross linked IgG chains, the beads were washed sequentially with 1 ml 0.5 M acetic acid (ph 3.5), 5 ml buffer PA (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 % NP40, 2mM MgCl₂, 0.1 mM EDTA, 0.5 mM DTT), again 1 ml 0.5 M acetic acid and finally 5 ml buffer PA. In a next step, beads were equilibrated for 30 min in 10 ml buffer PA on a rotating wheel at 4 °C. Subsequently, nuclear extract from *Drosophila* embryos was transferred to the resin. The extract volume added was at least 10 ml and the extract concentration of different samples was adjusted to approximately the same concentration (between 7-10 mg/ml). Tandem affinity purification (TAP) was conducted as reported in (Klymenko et al., 2006). For elution beads were transferred to a 1.5 ml-Eppendorf tube and eluted three times for 30 min at 4 °C in a thermo mixer/ shaker by adding 200 µl of buffer CE (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 % NP40, 1 mM MgCl₂, 2 mM EGTA, 1 mM imidazole pH 8.0, 10 mM β-Mercaptoethanol).

2.2.4.2.2 Flag-affinity purification of recombinant proteins from insect cell culture

In a first step, the cytosolic, nuclear and chromatin extracts from insect cells were analyzed by SDS-PAGE followed by Coomassie or western blot. The insect cell extracts were combined for purification to get the maximal amount of recombinant proteins in a volume of 12-15 ml and incubated for 10 min on a rotating wheel, e.g. 9 ml of nuclear extract and 3 ml of cytosolic extract. An aliquot of the insect cell extract was removed and kept as an Input control. Agarose beads (Sigma) and Anti-Flag M2 agarose beads

(Sigma) were pre-washed with 10 ml PBS for 5 min and two times 30 min with 10 ml lysis buffer (25 mM HEPES pH 7.9, 150 mM NaCl, 0.05 % NP40, 20 μ m ZnCl₂, 10 % glycerol, 0.5 mM DTT, 0.1 % Pefabloc-SC, 1x complete protease inhibitor). The insect cell extracts were pre-cleared in 15 ml-Falcons by incubating them with 150 μ l agarose beads for 45 min, 4 °C on the rotating wheel. Then the beads were spun down for 5 min at 1500 x g and the supernatant transferred to 300 μ l of equilibrated Anti-Flag-M2 agarose resin. The extracts were allowed to bind to the resin over night at 4 °C on a rotating wheel. The next day the samples were transferred to disposable columns (Biorad) and passed through the column by gravity flow. The beads were extensively washed with increasing concentrations of KCl up to 1.2 M with buffer BC (20 mM HEPES pH 7.9, 0.2 mM EDTA, 20 % glycerol, 2 mM DTT, 0.1 mM Pefabloc-SC, 1x complete protease inhibitor): (1) BC300N (0.3 M KCl, 0.2 % NP40), (2) BC 600N (0.6 M KCl, 0.2 % NP40), (3) BC 1200N (1.2 M KCl, 0.2 % NP40), (4) BC1200 (1.2 M KCl), (5) BC 600 (0.6 M KCl) and (6) BC 300 (0.3 M KCl). In the last step beads were transferred to micro-centrifuge columns and eluted in 300 μ l BC300 buffer supplemented with 0.4 mg/ml Flag-peptide (Sigma) for 1 h at 4 °C in a thermomixer. A second elution step was performed for 2-12 h. Finally, the presence and purity of desired proteins was analyzed on a Coomassie stained SDS-gel and western blot.

2.2.4.2.3 Sucrose gradient sedimentation analysis of Flag-purified complexes

The Flag-purified protein complexes were loaded on a 10-60 % sucrose gradient (20 mM HEPES pH 7.9, 300 mM KCl, 0.2 mM EDTA, 10 % sucrose/ 60 % sucrose, 1 mM DTT, 0.1 mM Pefabloc-SC, 1x complete protease inhibitor). Gradients were prepared with the gradient station (Gradient Station ip, Biocomp) in Seton ultracentrifuge tube at RT and subsequently cooled down to 4 °C. The Seton tube was half filled with light buffer and subsequently heavy buffer was injected underneath the light buffer with a 50 ml Luer-Lock plastic syringe mounted with a cannula. The appropriate gradient mixing program was run on the gradient mixer depending on nature of gradient, gradient composition and tube size. After the gradients were mixed, they were cooled down for 1-2 h at 4 °C.

As size markers for the gradient the gel-filtration kit for protein Molecular Weights 29,000-70,000 Da (Sigma Aldrich) was used. 50 µg of each protein Thyroglobulin, Appoferretin, β-Amylase, Alcohol dehydrogenase, BSA and Carbonic anhydrase were mixed in a final volume of 210 µl sample buffer (20 mM HEPES pH 7.9, 300 mM KCl, 0.2 mM EDTA, 1 mM DTT, 0.1 mM Pefabloc-SC, 1x complete protease inhibitor). Flag-purified samples were concentrated 10x with Amicon Ultra centrifugal filters (10 kDA cut off, Millipore) and subsequently 3x diluted with sample buffer to have a final glycerol concentration of less than 10 %. 100 µl of sample was loaded on each gradient by carefully layering the sample on top of the balanced gradients.

Gradients were run in an ultracentrifuge for 18 h at 256,136 x g at 4 °C (SW40ti rotor, Beckman Coulter) with slow start and slow break to avoid disturbance. After the ultracentrifugation step, the gradients were fractionated one by one in 24 fractions with the gradient fractionator (Gradient Station ip, Biocomp). The following settings were used for fractionation: total distance fractionated 80 mm, distance/fraction 3.33 mm, volume/fraction 0.48 ml and fractionation speed 0.1 mm/sec. After fractionation the remaining 1.5 ml of gradient were transferred to a tube and taken as fraction 25 or void fraction.

In order to concentrate the proteins, trichloroacetic acid precipitation was performed on each fraction. Therefore, one volume of TCA 100 % (w/v) was added to three volumes of protein sample and the solution was supplemented with 4 mg/ml sodium deoxycholate. The sample was vortexed and kept on ice for 30 min. In the following step the sample was spun at 20,000 x g for 10 min at 4 °C in a tabletop centrifuge. The supernatant was aspirated and the pellet was washed twice with three volumes of cold acetone. After each wash the tube was flicked to mix, incubated for 5 min on ice and subsequently spun at 20,000 x g for 10 min at 4 °C in a tabletop centrifuge. After the last washing step the pellet was dried at 95 °C for 5 min and finally resuspended in 55 µl of 2x LDS buffer by shaking it at 95 °C for 5 min in a thermomixer. Subsequently, 15 µl of each concentrated fraction was loaded on 8 % Tris-Glycine Midi gels (NuPAGE Novex, Life Technologies, 8x13 cm) and analyzed by western blotting with the appropriate antibodies. The marker protein fractions were run on a 4-12 % Bis-Tris

gradient Midi gel (NuPAGE Novex, Life Technologies, 8x13 cm) and stained by Coomassie.

2.2.4.2.4 Purification of Scm- and Sfmbt-SAM domains from bacteria for crystallography

The Scm-SAM/ Sfmbt-SAM complex was obtained by co-expressing the two His-fusion constructs (pEC-k-3c-GST-His-Sfmbt₁₁₃₇₋₁₂₂₀ and pEC-CDF-3c-His-Scm₈₀₃₋₈₇₇^{L859R} or pEC-k-3c-GST-His-Sfmbt₁₁₃₇₋₁₂₂₀ and pEC-CDF-3c-His-Scm₈₀₃₋₈₇₇^{L855E, L859E}) in *E. coli* (Rosetta) and copurifying it via Ni²⁺ affinity, Glutathione Sepharose (GST) affinity followed by tag cleavage, cation exchange and size-exclusion chromatography. After the last purification step, samples were concentrated with Amicon Ultra centrifugal filters (3 kDA cut off, Millipore) and submitted to the crystallization facility (MPIB Crystallization facility) for crystallization screening.

2.2.4.2.4.1 His-purification of Scm- and Sfmbt-SAM domains

For His-purification 3 ml of Ni-NTA Agarose beads resin (QIAGEN) was used for extract from 4 l of expression culture. After equilibration the filtered protein extract was applied to the beads and passed through the resin by gravity flow twice. Then the beads were washed in three sequential steps with wash buffers (50 mM phosphate buffer pH 7.5, 250 mM NaCl, 10 mM MgCl₂, 10 % glycerol, 2 mM β-Mercaptoethanol, 1 mM Pefablok-SC, 1x complete protease inhibitor) containing increasing imidazole concentration: (1) 100 ml wash buffer 1 (25 mM imidazole, 0.1 % NP40), (2) 50ml wash buffer 2 (50 mM imidazole) and (3) 50 ml wash buffer 3 (75 mM imidazole). After the washing steps the protein was eluted from the resin by gravity flow with 10 ml elution buffer (250 mM imidazole, 5 % glycerol) and 1 ml fractions were collected. Eluted fractions were tested on a SDS gel for protein amount and purity and pooled for the next step.

2.2.4.2.4.2 GST purification of Scm- and Sfmbt-SAM domains

Pooled fractions from the His-purification step were diluted with GST binding buffer (50 mM phosphate buffer pH 7.5, 200 mM NaCl, 10 mM MgCl₂, 5 % glycerol, 4 mM DTT, 1 mM Pefablok-SC, 1x complete protease inhibitor) to a volume of 25 ml. Prior to purification 6 ml GST resin (Glutathione SepharoseTM 4 Fast Flow, GE Healthcare) per

4 l of expression culture was equilibrated. The sample was passed through the GST-resin by gravity flow and re-applied six times. Then the beads were washed three times with 25 ml GST binding buffer, before adding 12 ml GST binding buffer supplemented with 0.7 ml GST-PreScission protease (core facility, MPI). The column was sealed and incubated with the protease over night on a rotating wheel in order to cut GST-His/ His-tags of the proteins. The next day the flow-through of the column containing the SAM domains without tags was collected. The resin was washed with 6 ml of GST binding buffer and the two fractions were pooled for the next purification step.

2.2.4.2.4.3 Ion exchange Chromatography of Scm- and Sfmbt-SAM domains

Before the next purification step, the sample was dialyzed to change the pH of the sample from pH 7.5 to pH 6.0 for cation exchange chromatography with a MonoS column (5/50, 1 ml, GE Healthcare) and Aekta purifier (GE healthcare). Dialysis was conducted with SnakeSkin Dialysis Tubing membrane (3.5 kDA, ThermoScientific) for 2-3 h against low salt buffer (50 mM phosphate buffer pH 6.0, 75 mM NaCl, 20 % glycerol, 1 mM DTT, 0.05 mM Pefablok-SC, 1x complete protease inhibitor) with one buffer change after 1 h. The sample was collected and the volume adjusted to 20 ml with low salt buffer containing 2.5 % glycerol. Then the sample was loaded in two sequential purification steps with a 10 ml superloop on the MonoS column that had been equilibrated with low salt buffer (2.5 % glycerol). After applying the sample, the column was washed with five column volumes of low salt buffer, before a linear NaCl gradient was started from 75 mM up to 1 M NaCl over 20 column volumes by gradually adding more high salt buffer (50 mM phosphate buffer pH 6.0, 2 M NaCl, 2.5 % glycerol, 2 mM DTT, 0.1 mM Pefablok-SC, 1x complete protease inhibitor) to the low salt buffer. The fraction collector collected 0.5 ml fractions in 96-well plates and the fractions showing the highest UV signal were run on a SDS gel. The three fractions with the highest protein concentrations of both runs were pooled to get a final volume of 3 ml for the Scm-SAM^{L859R}/ Sfmbt-SAM complex. The Scm-SAM^{L855E/ L859E}/ Sfmbt-SAM complex did not bind to the MonoS column and was processed by concentrating the flow-through to a final volume of 3 ml.

2.2.4.2.4.4 Gel-filtration of Scm- and Sfmibt-SAM domains

For gel-filtration 2 ml of concentrated sample was loaded on the column (Superdex75, 16/60g Hilo, GE-Healthcare) that had been equilibrated with size-exclusion buffer (25 mM phosphate buffer pH 6.0, 150 mM NaCl, 2.5 % glycerol, 2 mM DTT). After applying the sample, 1.5 column volumes were collected in 0.5 ml fractions and the fractions showing the highest UV signal were further analyzed on a SDS gel for purity. The fractions with the highest amount of pure protein were pooled and further processed by concentration.

2.2.4.2.5 Crystallization

Crystals of the Scm-SAM/ Sfmibt-SAM complex grew at 4 °C in 0.05 M Tris pH 8.0, 4 % MPD, 0.2 M Ammonium acetate and 40 % PEG3350. They contain two complexes per asymmetric unit (space group P1 21 1). Crystallization screens, data collection, structure determination and refinement were performed with help of Christian Benda, Department of Structural Cell Biology, MPI Biochemistry.

2.2.4.2.6 Data Collection, Structure Determination and Refinement

A 1.975 Å data set of a Scm-SAM/ Sfmibt-SAM crystal was collected at beamline PXII at the Swiss Light Source (SLS) synchrotron facility. The structure was solved by molecular replacement with Ph-SAM/ Scm-SAM (PDB ID: 1PK1) as search model. Data collection and refinement statistics are summarized in Table 2.9. Model building was completed in Coot. The structure was refined in PHENIX to yield a final model. Interaction surfaces of the solved structure were analyzed by PISA (Protein interfaces, surfaces and assemblies).

Table 2.9: X-Ray data Collection and Refinement Statistics

Data collection

Beamline	SLS PXII
Space group	P1 21 1
Unit cell a, b, c (Å)	49.956, 53.97, 61.455, 90, 109.22, 90
Wavelength (Å)	1.0
Resolution range (Å)	47.17 – 1.975

Total reflections	193632 (12564)
Unique reflections	21180 (1797)
Completeness (%)	96.93 (83.18) %
I/σ	11.21 (1.91)
R_{meas} (%)	0.1447
Wilson plot B (Å)	23.5
Refinement	
R_{work} (%)	0.2546
R_{free} (%)	0.2772
Ramachandran values	
Favored region (%)	98
Allowed region (%)	100
Rotamer outliers (%)	0
Overall average B factor (Å)	29.9
R.m.s.d bond lengths (Å)	0.019
R.m.s.d bond angles (°)	1.36

2.2.4.3 Detection and analysis of Proteins

2.2.4.3.1 Mass spectrometry, capillary LC-MS/MS analysis and protein identification

Gabriele Stoehr performed mass spectrometric analysis in Matthias Mann lab, Department of Proteomics and Signal Transduction, MPI Biochemistry. For Mass spectrometry proteins were run on a SDS gel and stained with Coomassie. Then each lane was cut in five equally large pieces that were trypsin digested and loaded on a HPLC system coupled to a MALDI-TOF device for analysis.

2.2.4.3.2 SDS-polyacrylamide-gel-electrophoresis (SDS-Page)

Protein samples were separated according to their size on 6 %, 8 % or 10 % SDS-polyacrylamide gels by gel electrophoresis. Gels were casted following standard protocols for Tris-glycine polyacrylamide gel preparation and run between 80 and 180 V in 1x Laemmli SDS running buffer (10x stock: 0.25 M Tris, 1.92 M Glycine, 1 % SDS, pH 8.3). For very small proteins or samples with a wide range of protein sizes 4-12% Bis-Tris pre-cast gradient gels (NuPAGE Novex Bis-Tris gels, Life Technologies, 8x8 cm) were used. For sucrose gradients and gel purification samples either pre-cast

midi 4-12 % Bis-Tris gradient gels or pre-cast midi 8 % Tris-Glycine gels (NuPAGE Novex, Life Technologies, 8x13 cm) were used depending on protein size. Before applying the sample to the gel, it was boiled for 5 min at 95 °C in 1x LDS sample buffer (NuPAGE LDS Sample Buffer (4x), Life Technologies).

2.2.4.3.3 Silver staining of protein gels

For silver staining SDS gels were fixed with fixing solution (40 % Methanol, 10 % Acetic acid) for 20-30 min and subsequently rinsed with several gel volumes of water over a period of 3 h. In a next step gels were sensitized with 0.02 % (w/v) sodium thiosulfate for 1-2 min and rinsed twice for 1 min with water. Then gels were incubated in a cold 0.1 % (w/v) silver nitrate solution for 20-40 min at 4 °C. Following the incubation, gels were rinsed with two changes of water for 1 min and immediately developed by adding developing solution (0.04 % (v/v) formaldehyde, 2 % (w/v) sodium carbonate). Once sufficient staining was achieved the development was stopped by adding 1 % acetic acid solution. Silver stained gels were stored in 1 % (v/v) acetic acid solution at 4 °C.

2.2.4.3.4 Coomassie Brilliant Blue staining of protein gels

For staining the PageBlue Protein Staining solution (Thermo Scientific) for endpoint staining was used. Prior to staining the SDS gels were rinsed 3x 5 min with fresh changes of ddH₂O. After that 20 ml of PageBlue staining was added to the gel and incubated with gentle agitation for 1 h up to overnight. Finally, PageBlue solution was removed and the SDS gel was rinsed with several changes of ddH₂O for 10-20 min. Gels were stored in ddH₂O at RT.

2.2.4.3.5 Western blotting of protein gels

Protein samples were run on a SDS gel and then transferred to a membrane (Amersham Hybond ECL Nitrocellulose Blotting Membrane, 0.45 µM, GE Healthcare Life Sciences). Therefore, six leaves of Whatman paper (Chromatography paper, 3 mm, Whatman) and one membrane were cut to a slightly larger size than the gel and pre-soaked in transfer buffer (25 mM Tris, 192 mM glycine, 0.05 % SDS). To prepare the transfer sandwich, a blotting cassette was open and assembled with the following

layers: (1) a thick transfer swamp, (2) three Whatman papers, (3) the SDS-gel, (4) the membrane, (5) three layers of Whatman papers and (6) a thin swamp. It was checked that no air bubbles were draped between the gel and the membrane and the cassette was inserted into the blotting chamber. Protein transfer was conducted between 1.5-2 h at 90 V at 4 °C with stirring to ensure stable transfer conditions. Next the membrane was rinsed in PBS-T and blocked in PBS-T-milk 5 % (w/v) for 2 h at RT or up to overnight at 4 °C with gentle agitation. Primary antibody was diluted in PBS-T 5% milk (Table 2.7) and incubated with the membrane over night at 4 °C with gentle agitation. After the primary antibody incubation, the membrane was washed three times fast and 3x 10 min with PBS-T. The secondary antibody, which was horseradish peroxidase conjugated, was diluted in PBS-T-milk 5% 1:5000 (Table 2.8) and incubated for 1-2 h at 4 °C with gentle agitation. Subsequently, the membrane was washed three times quickly and 3x 10 min in PBS-T. For detection the membrane was transferred to a plastic foil and residual liquid was removed before adding the two components of ECL (Enhanced chemiluminescence) solution in a 1:1 ratio. The final volume of ECL solution for a standard mini membrane (8x6 cm) was 1 ml and for a midi membrane (8x15 cm) 2 ml. Different types of ECL solution were used: standard ECL (Amersham ECL Western Blotting Detection Reagent, GE Healthcare), a more sensitive ECL solution (Immobilion Western Chemiluminescent HRP substrate, Millipore) or very sensitive ECL solutions (Lumi-Light Western Blotting Substrate, Roche) and (ECL Prime Western Blotting Detection Reagent, GE Healthcare). ECL solution was spread equally over the membrane and incubated for 1 min before removing excess liquid. Subsequently, the membrane was covered with plastic foil and transferred to a light proof western blotting cassette. In the dark room, films (Amersham Hyperfilm ECL, GE Healthcare) were put on the membrane and exposed for different time periods. Typically, films were exposed for 5 sec, 10 sec, 30 sec, 1 min and 5 min depending on the signal strength. In a last step the film was developed in a developing machine and marker bands were copied from the membrane to the film. Films were either scanned or the membrane was developed with the digital imaging system LAS (Luminiscent Image analyzer, ImageQuant LAS4000, GE Healthcare) to get digital files of the western blots.

2.2.5 DNA techniques and cloning

2.2.5.1 Purification and analysis of DNA fragments

DNA fragments were analyzed by agarose gel electrophoresis or sequencing. For purification DNA fragments were either subjected to gel extraction or PCR purification using QIAGEN kits.

2.2.5.1.1 Agarose gel electrophoresis

Size-dependent electrophoretic separation of DNA in agarose gels was performed for analysis of size and concentration of DNA fragments as well as isolation of specific DNA fragments from DNA mixtures. Therefore, 1.0-2.5 % agarose (Sigma) was added to TBE Buffer (89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA) and boiled until agarose was completely melted. The solution was cooled down and 0.01 % SYBR Safe DNA gel stain (life technologies) was added before pouring the gel with appropriate size. Next DNA samples were supplemented with 6x DNA loading dye (Fermentas) and samples were loaded. Electrophoresis was performed at 90 V for ~1 h with TBE buffer. DNA markers were run in parallel to estimate size and concentration of DNA fragments: (1) GeneRuler 1kb Plus DNA ladder (Fermentas), (2) GeneRuler 100bp Plus DNA ladder (Fermentas) and (3) MassRuler Express DNA Ladder, Reverse (Fermentas). DNA fragments were visualized in a UV detector (Pepqlab).

2.2.5.1.2 Agarose gel Purification of DNA fragments

Depending on DNA fragment size the DNA samples were separated on 1-2.5 % agarose gels containing 0.01 % SYBR Safe DNA gel stain. The fragments were excised from the agarose gel under weak UV light (Safe Imager 2.0, Invitrogen), solubilized and further processed following the QIAquick gel extraction kit (Qiagen).

2.2.5.1.3 PCR product purification

PCR products were either purified by gel extraction (see above) or according to the QIAquick PCR purification kit (Qiagen).

2.2.5.1.4 DNA sequencing

For DNA sequencing samples were sent either to the in house sequencing facility (core facility, MPI), MWG sequencing by Eurofins Genomics or GATC Biotech AG. Sequencing primers used in this study are listed in Table 2.1.

2.2.5.2 Enzymatic processing of DNA

DNA was amplified by Phusion DNA polymerase (Finnzymes), digested by high fidelity restriction enzymes (New England Biolabs) or ligated by T4 ligases (Roche).

2.2.5.2.1 PCR reaction to amplify DNA templates

PCR reaction was generally done in 50 µl reaction volume with 0.02 U/µl Phusion polymerase and 50 pg DNA template. Phusion polymerase (Finnzymes) ready 2x Mastermix containing buffer and NTPs were used. 5-20 ng of template DNA and 0.5 µmol of forward and reverse primer were supplemented and topped up to 50 µl with ddH₂O.

PCR reaction program:

Temperature	Time	No. of cycles
98 °C	30 sec	1
98 °C 65-72 °C (depending on T _m of primer) 72 °C	10 sec 30 sec 30 sec per kb template DNA	25
72 °C	10 min	1

2.2.5.2.2 Enzymatic restriction digest of DNA

For analytical restriction digests the FastDigest system (Thermo Scientific) was used. In a digestion reaction 2 µl of vector Miniprep DNA was digested with 0.5 µl FastDigest restriction enzyme (1FDU, Thermo Scientific) in 1x universal FastDigest Green buffer in a volume of 10 µl at 37 °C for 5 min. After digest the restriction digest was analyzed by agarose gel electrophoresis. For preparative scale restriction digest 2 µg of vector DNA or PCR product was digested with 1 µl of each required high fidelity restriction enzyme (New England Biolabs, 20,000 U/ml) in a total volume of 50 µl. The restriction buffers and supplements such as BSA were used as recommended by NEB. Usually double

restriction enzyme reactions with two restriction endonucleases (2x 1 µl) were set up and incubated for 3 to 12 h at 37 °C.

2.2.5.2.3 DNA ligation

For ligation with overhangs (sticky ends) 2.5 U T4 ligase (bacteriophage DNA ligase, 5 U/µl, Roche) was used to ligate 50 ng of vector DNA and 150 ng of insert DNA (1:3 ratio) in a volume of 10 µl over night at 16 °C. For blunt end ligation the Rapid DNA ligation kit was used. 2.5 U of T4 DNA ligase (5 U/µl) was used to ligate 25 ng of vector DNA in a total volume of 10.5 µl. The reaction was incubated for 5 min at RT.

2.2.5.3 Cloning strategies

Different cloning strategies were used depending on the target vector (see section 2.1.2 Vectors used in this study). DNA oligos used for cloning and mutagenesis are listed in Table 2.2 and 2.3, respectively.

3. Results

3.1 Ph interacts with PRC1 members and the PhoRC complex

The first aim of this study was to biochemically characterize the Polycomb group protein Polyhomeotic by identifying novel interactors employing a tandem affinity purification (TAP) strategy. Therefore, I cloned a transgene expressing a TAP-tagged Ph fusion protein (TAP-Ph) under an *α-tubulin-1* promoter. The tandem affinity purification (TAP) tag consists of three modules (1) a protein A moiety, (2) a TEV protease cleavage site and (3) the Calmodulin binding domain of the Calmodulin binding protein (CBP) (Figure 3.1 A) and allows for rapid purification under native conditions. Since a Flag-tagged Ph transgene under control of the endogenous promoter gave only partial rescue to Ph null embryos (Shao et al., 1999), I decided to express TAP-Ph transgenes using an *α-tubulin-1* promoter to get higher expression levels. Next, I generated transgenic fly lines expressing either *TAP-ph-p* or *TAP-ph-d* in the wild-type background to perform tandem affinity purifications on embryonic material. With this approach I wanted to detect and compare proteins associated with either Ph-p or Ph-d.

First, I tested if the transgenes are sufficiently expressed in embryos (12-16h) and still functional *in vivo*. Both transgenes, *TAP-ph-p* and *TAP-ph-d*, were expressed in the transgenic fly lines (Figure 3.1 B) and were able to give partial rescue to the *ph⁰* phenotype in embryos as well as in larval tissues indicating that they are indeed functional (Figure 3.2).

Embryos lacking both *ph* genes arrest development 12 h after fertilization and display lack of dorsal closure (Figure 3.2 A, panel 2). Rescue experiments with *ph⁰* embryos demonstrated that a single copy of *ph* transgene (*TAP-ph-p* or *TAP-ph-d*) leads to rescue of dorsal closure and formation of a fully developed embryonic cuticle that lacks severe homeotic transformations that are characteristic of PcG mutants and are caused by the widespread misexpression of Hox genes. Nevertheless, I found that the rescued embryos display mild homeotic transformations in the posterior abdominal segments and that their head involution is defective, suggesting that HOX gene repression is likely not fully rescued (Figure 3.2 A, panel 3 and 4).

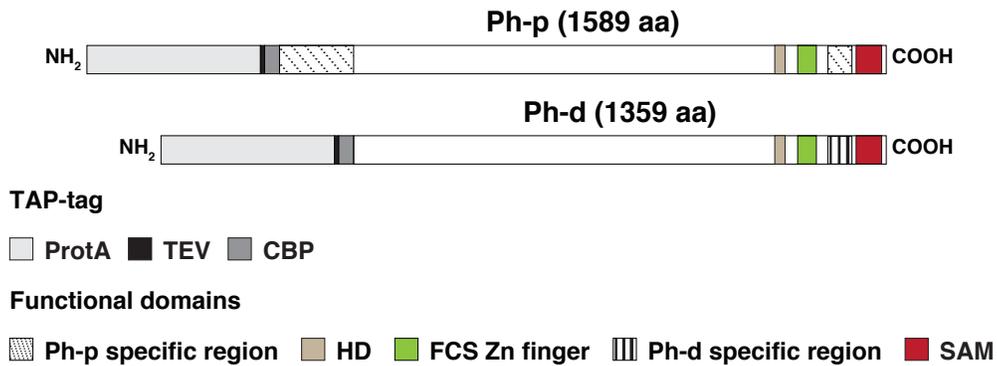
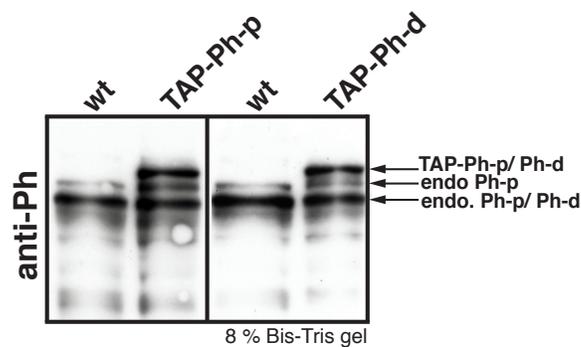
A**B**

Figure 3.1: Expression of *TAP-ph-p/-d* transgenes in embryonic nuclear extract.

(A) Schematic representation of the TAP-Ph proteins. Functional domains conserved in Ph proteins include a zinc finger domain (ZF), a homology domain (HD) and a sterile alpha motif (SAM). In addition, the regions unique to Ph-p, a 194 aa N-terminal part and a 60 aa stretch close to C-terminus where Ph-p and Ph-d diverge in sequence are highlighted (striped pattern). The tandem affinity purification (TAP) tag consists of three modules (1) a protein A moiety, (2) a TEV protease cleavage site and (3) the Calmodulin binding domain of the Calmodulin binding protein (CBP).

(B) Western blot analysis of nuclear extract prepared from wild-type (wt) and transgenic TAP-*ph-p* and TAP-*ph-d* 0-12 h old embryos using an antibody against Ph. Note that the antibody detects three bands of endogenous Ph, which correspond to two isoforms of Ph-p (Ph-p-RA, 1589 aa and Ph-p-RB, 1346 aa) and one isoform of Ph-d (Ph-d-PB, 1359 aa). There is a fourth band visible in the transgenic material, which runs above the three endogenous proteins and corresponds to TAP-Ph-p and TAP-Ph-d, respectively.

In addition, the rescue capacity of a single *ph* transgene was also tested in imaginal wing discs of third instar larvae using clonal analysis. The Hox gene *Abd-B*, which is normally repressed in wing disc cells, is strongly misexpressed in all cells of *ph*⁰ mutant cell clones (Figure 3.2 B, panel 1). A single copy of either the *TAP-ph-p* or the *TAP-ph-d* transgene gives rise to partial rescue and *Abd-B* remains repressed in at least a fraction of clone cells in each disc analyzed (Figure 3.2 B, pane 2 and 3).

I next investigated to what extent the *TAP-ph-p* and the *TAP-ph-d* transgene rescue the distorted morphology and enlarged size of nuclei in *ph*⁰ mutant cells and the formation of tumor tissue by these cells. To analyze nuclear morphology, I stained discs with an antibody against nuclear lamin. I found that both nuclear morphology and clone shape and size are rescued by the presence of the TAP-Ph-p or TAP-Ph-d protein (Figure 3.2 C). Taken together, these data illustrate that both TAP-Ph proteins are at least partially functional and able to emulate the function of the native Ph protein.

The transgenic fly lines were expanded in cages and 0-14 h old embryos were collected to prepare soluble nuclear extract for Tandem Affinity Purification (TAP). The purified material was run on a SDS gel and copurifying proteins were visualized by silver staining (Figure 3.3 A). In contrast to purified material from the control wild-type embryos (left lane), the purified material from TAP-ph-p and TAP-ph-d embryos displayed numerous additional bands (right lane).

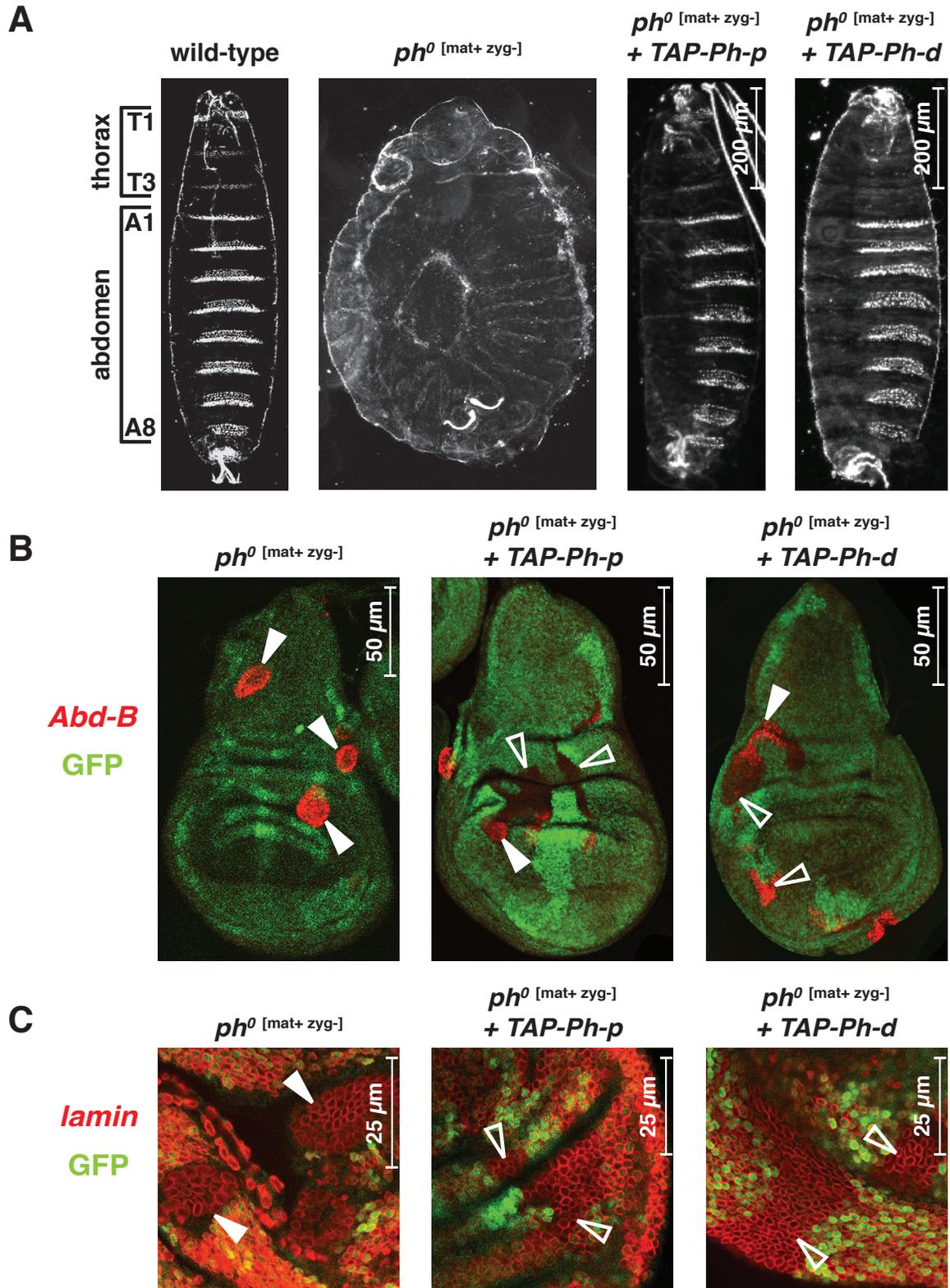


Figure 3.2: TAP-*ph* transgenes partially rescue *ph*⁰ mutants.

(A) Embryonic cuticles of wild-type (wt), ph^0 ^[mat+ zyg-] zygotic mutant embryos without a transgene (no TG) or carrying a *TAP-ph-p* or *TAP-ph-d* transgene. Single copy of the TAP-ph transgene was crossed into the mutant background (ph^{504} allele) and resulted in rescue of dorsal closure and extensive rescue of homeotic transformations. Note that rescued embryos still exhibit mild homeotic transformations (A7 to A8, arrows) and display defective head involution. mat+: maternal contribution, zyg-: no zygotic contribution. Pictures of embryonic cuticles of wt and ph^0 ^[mat+ zyg-] adapted from (Gutiérrez et al., 2012).

(B) Analysis of wing imaginal disc clones homozygous for ph^0 (ph^{del} allele) (GFP negative) 96 h after induction without transgene (ph^0 ; no TG) or with transgene (ph^0 ; *tub-TAP-ph-p*) or (ph^0 ; *tub-TAP-ph-d*). Clones were stained for *Abd-B* (in red) expression, which is repressed in wild-type cells. Repression of the Hox gene *Abd-B* (filled arrow heads) was rescued in at least a fraction of clone cells in presence of the transgenes (empty arrow heads).

(C) Analysis of wing imaginal disc clones homozygous for ph^0 (ph^{del} allele) (GFP negative) 72 h after induction without transgene (ph^0 ; no TG) or with transgene (ph^0 ; *tub-TAP-ph-p*) or (ph^0 ; *tub-TAP-ph-d*). Clones were stained for *lamin B* (in red) to determine the nuclear shape and size. Note that ph^0 clones display a tumor phenotype with enlarged nuclei that form cavities that can be distinguished from the surrounding wild-type tissue (filled arrow heads). In contrast, ph^0 clones rescued with a single *ph* transgene appear wild-type-like (empty arrow heads). Wild-type nuclei of surrounding tissue are GFP-positive (green and red).

The purified material from transgenic embryos was analyzed by Mass spectrometry and was compared to a mock purification performed with extract from wild-type embryos. The score list for interacting proteins was generated by calculating the ratio of enrichment in the transgenic material versus the wild-type material (Table 3.1, for peptide list see 6.4 Supplementary material, Table 6.1). Mass spectrometry and western blotting confirmed that both TAP-fusion proteins, TAP-Ph-p and TAP-Ph-d, interact with all the previously described PRC1 subunits including Psc (Posterior sex combs) and its paralog Su(z)2 (Suppressor of zeste 2), Sce/Ring (Sex combs extra), Pc (Polycomb), Ph-d and Ph-p, as well as Scm (Sex comb on midleg), a substoichiometric member of the PRC1 complex. In addition, the two PcG proteins Sfmbt (Scm-related gene containing four mbt domains) and Pho (Pleiohomeotic), which together form the PhoRC complex, were identified as interacting partners of Ph-p and Ph-d (Table 3.1, Figure 3.4). Interestingly, previous purifications of Ph did not identify Sfmbt as an

interacting partner (Shao et al., 1999; Wang and Brock, 2003); however, a recent study copurified the PhoRC complex with PRC1 by using the Pc subunit as bait (Strübbe et al., 2011). In the analysis by mass spectrometry only very few peptides of Scm were detected in both Ph-d purifications and Scm was undetectable in the Ph-p purification (Table 6.1). However, the western blot analysis showed a clear enrichment of Scm with both, TAP-Ph-p and TAP-Ph-d (Figure 3.4).

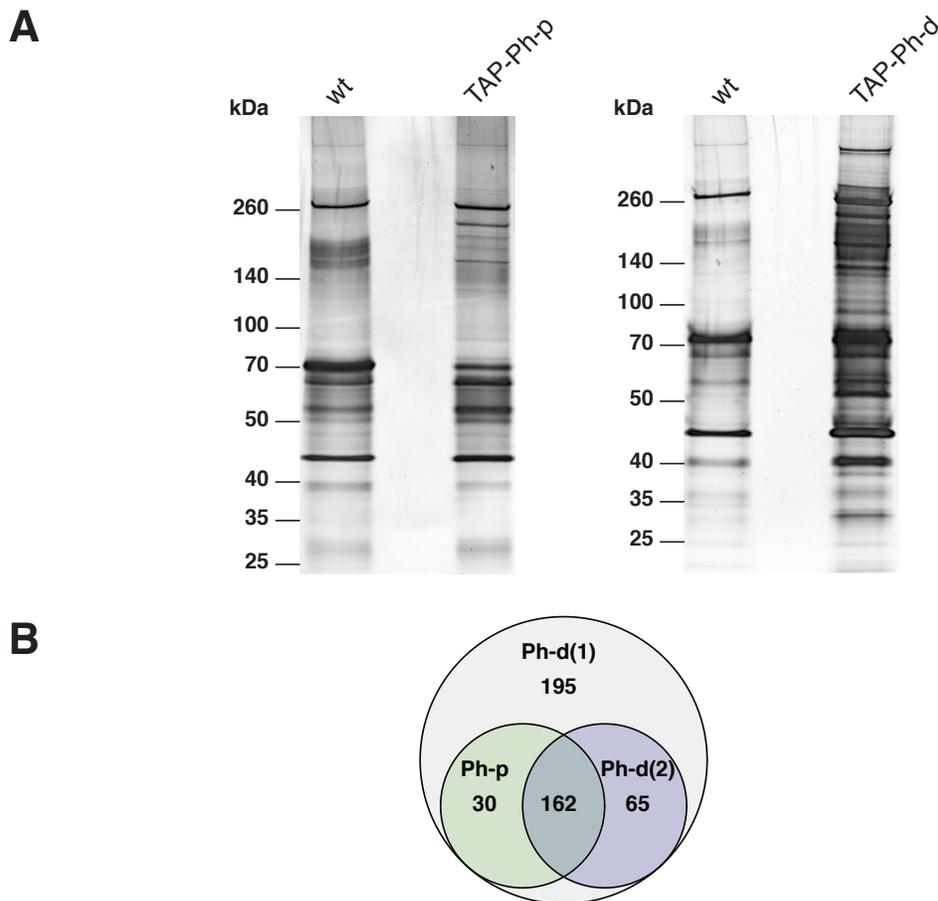


Figure 3.3: Analysis of TAP-purified material by Mass spectrometry.

(A) Silver staining of tandem affinity purified material from nuclear extracts of wild-type, TAP-ph-p and TAP-Ph-d transgenic embryos.

(B) Venn-diagram to visualize overlap of Mass Spectrometry hits identified in three independent Ph purifications. Two purifications were conducted with material from embryos expressing TAP-Ph-d and one with embryos expressing TAP-Ph-p protein. The total number of hits identified varied between the three different purifications: Ph-d purification 1 (n=452), Ph-d purification 2 (n=227) and Ph-p purification (n=192).

Besides the above-mentioned PcG proteins, 11 additional proteins were identified in all three independent Ph purifications: Ge-1, Sop2, Nup107, Myo61F, Pyd, Edc3, Gnf1, Osp, Lamin, Imitation SWI and Hcf. These 11 proteins can be classified into three groups: (1) proteins involved in RNA processing, (2) proteins involved in nuclear structure and (3) proteins involved in chromatin binding and remodeling. The first group includes Ge-1 and Edc3, which are both involved in RNA decapping (Eulalio et al., 2008; Fan et al., 2011) and have been linked to silencing through miRNAs (Eulalio et al., 2007). The second group consists of proteins involved in nuclear structure such as the actin binding protein Sop2, which is a constituent of the cytoskeleton (Zallen et al., 2002), the nucleocytoplasmic transporter protein Nup107 (Chen and Xu, 2010) as well as the actin binding Myo61F, which participates in directing movement of organelles along actin filaments (Hegan et al., 2007; Morgan et al., 1995). In addition, lamin, a component of the nuclear lamina that can interact with chromatin is also part of the nuclear structure (Gruenbaum et al., 1988).

Table 3.1: Ph interacting proteins identified by Mass spectrometric analysis.

Proteins identified	Score (Ph-d 1)	Score (Ph-d 2)	Score (Ph-p)
Ph-d (Polyhomeotic distal)	13.54	13.26	13.32
Psc (Posterior sex combs)	8.71	9.33	7.45
Sce/Ring (Sex comb extra)	8.59	10.74	8.59
Pc (Polycomb)	8.50	9.03	5.78
Su(z) 2 (Suppressor of zeste 2)	6.53	-	-
Ge-1 (Enhancer of mRNA-decapping protein 4 homolog)	4.79	8.46	4.17
Sop2 (Suppressor of profilin 2)	4.57	8.30	4.83
Nup107 (Nucleoprin 107kD)	4.56	6.40	5.21
Sfmbt (Scm-related gene containing four mbt domains)	4.36	5.36	7.06
Myo61F (Myosin 61F)	4.32	8.45	5.29
Ph-p (Polyhomeotic proximal)	3.96	5.73	12.13
Pho (Pleiohomeotic)	3.84	-	-
Pyd (Polychaetoid)	3.64	5.63	5.71
Edc3 (Enhancer of decapping 3)	3.37	9.55	5.72

Proteins identified	Score (Ph-d 1)	Score (Ph-d 2)	Score (Ph-p)
Gnf1 (Germline transcription factor 1)	2.98	2.69	4.17
Osp (Outspread)	2.96	4.16	5.08
Lam (Lamin Dm0)	2.92	6.35	5.06
Iswi (Imitation Switch)	2.42	6.35	4.85
HCF (Host cell factor)	0.79	5.66	4.38

Major hits detected with high score in all three independent purifications with exception of Su(z)2 and Pho that were only significantly enriched in Ph-d purification 1. Scm was detected by mass spec analysis in both Ph-d purifications, but due to poor “sequence coverage” is not represented amongst the major hits. PRC1 subunits and PhoRC subunits are highlighted in blue and orange, respectively.

The third group comprises chromatin binding and remodeling proteins and includes Iswi and HCF. While Iswi is part of several chromatin-remodeling complexes such as NURF (nucleosome-remodeling factor), ACF (ATP-utilizing chromatin assembly and remodeling factor), and CHRAC (chromatin accessibility complex) (Badenhorst et al., 2002; Eberharter et al., 2001; Ito et al., 1997; Tsukiyama et al, 1995, 1995a), HCF has not only been linked to chromatin remodeling, but has also been shown to possess histone acetylation activity (Guelman et al., 2006; Suganuma et al., 2008). Gnf1 has been speculated to be involved in DNA replication (FlyBase Curators et al., 2004). The remaining protein, Pyd is of unknown function.

Note that there were no striking differences detected between proteins copurifying with either Ph-d or Ph-p (Figure 3.3 B). While the number of hits varied significantly between different purifications, the overlap of all three purifications accounted for 162 hits and the major hits were recovered reproducibly (Table 3.1). In the first Ph-d (1) purification I identified about twice the amount of proteins as in the subsequent Ph-d (2) and Ph-p purifications, which were conducted in parallel.

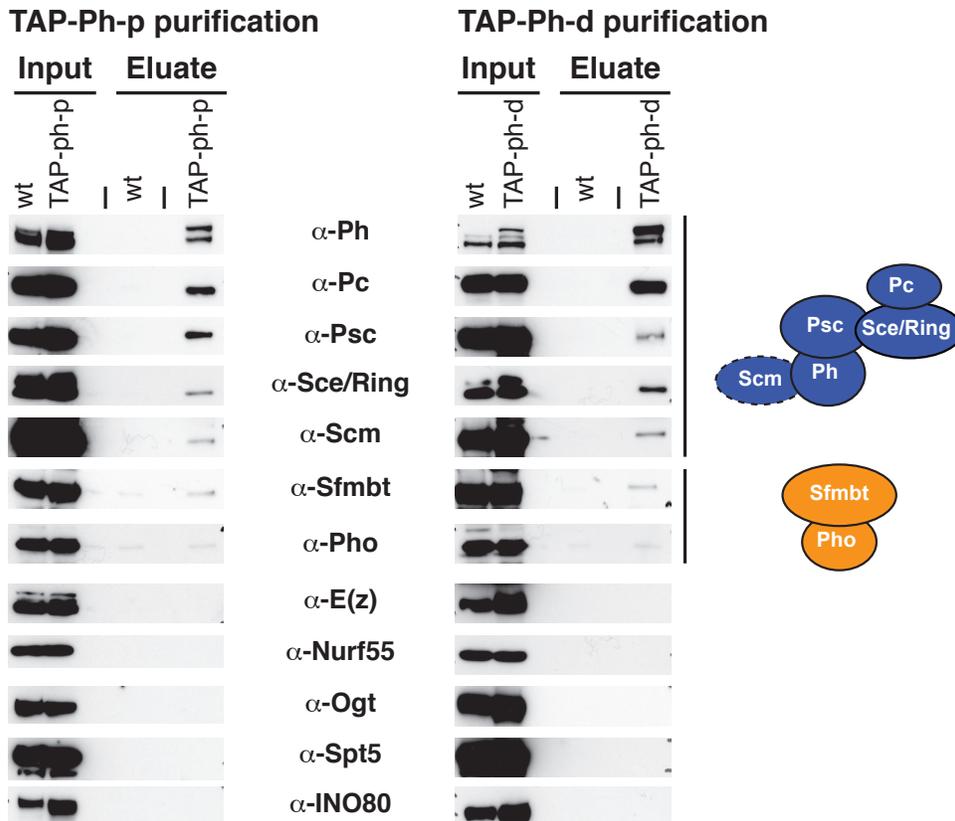


Figure 3.4: TAP-Ph purifications identify PRC1 members and PhoRC members as major Ph interactors.

Western blotting of TAP-purified material Ph, Pc, Psc, Sce/ Ring, Scm, Sfmbt and Pho are detected in *TAP-ph-p* and *TAP-ph-d* eluates. E(z), Nurf55, Ogt, Spt5 and INO80 were undetectable in this material. Note that Pho is also detected in eluates from the control purification from wild-type embryos.

3.2 PhoRC member Sfmbt and PRC1 members Ph and Scm form a stable trimeric complex *in vitro*

Since I found Sfmbt in all three Ph purifications with a high score and could also confirm its association by western blotting (Figure 3.4, row 6, left and right panel), I decided to focus on the question, whether Ph or one of the other PRC1 members can directly interact with the PhoRC complex via Sfmbt. This interaction could form a link between the specific DNA binding activity of PhoRC and the PRC1 complex that lacks specific DNA binding activity. On one hand earlier studies reported that Ph can interact with Pho

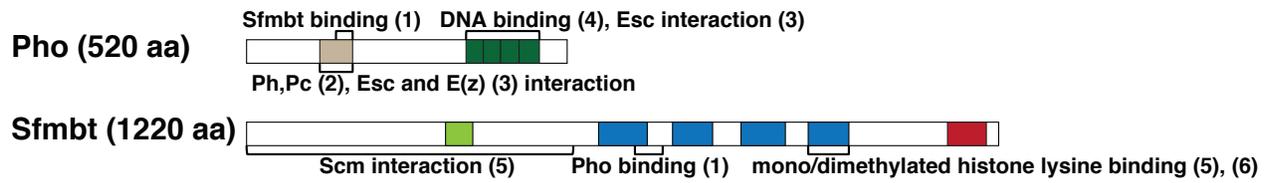
via its N-terminal part in *in vitro* assays (Mohd-Sarip et al., 2002). On the other hand, previous TAP purifications using TAP-Pho or TAP-Sfmbt as bait have failed to identify PRC1 subunits as copurifying components (Alfieri et al., 2013; Klymenko et al., 2006). Nevertheless, Scm, a substoichiometric PRC1 subunit, has been shown to interact with both, Ph and Sfmbt. Specifically, Scm directly interacts with the N-terminal part of Sfmbt (Grimm et al., 2009) and Scm has also been reported to form hetero-polymers with Ph via their respective C-terminal SAM domains (Kim et al., 2005) (Figure 3.5 A). Therefore, it was conceivable that Scm mediates the observed interaction.

To test for a direct physical interaction between Ph and Sfmbt, both proteins were co-expressed in insect cells using the baculovirus expression system and the overexpressed proteins were isolated by affinity-purification using a Flag epitope tag on one of the two proteins. In parallel, Ph, Sfmbt and Scm were co-expressed and purified to account for the possibility of indirect interaction mediated by Scm. Intriguingly, the interaction between Ph and Sfmbt does not appear to be direct, but depends on Scm (Figure 3.5 B). Since full length Ph was insoluble, all the interaction assays were performed with an N-terminal truncation of Ph (Ph₁₂₉₈₋₁₅₈₉). The truncated version of Ph contains all the functional domains.

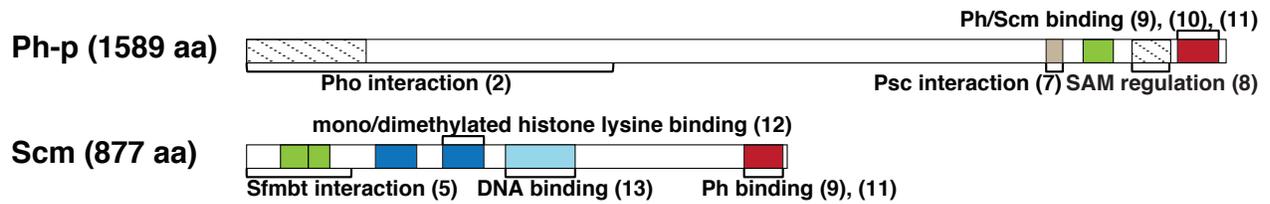
In a next step, I investigated whether the observed interaction between Sfmbt and the two PRC1 subunits Scm and Ph is stable or rather a transient interaction. Therefore, I purified the trimeric complex Sfmbt:Scm:Ph₁₂₉₈₋₁₅₈₉ from insect cells either by pulling on Ph or on Sfmbt and subsequently run the purified complex on an analytical Superdex 200 column, which has a resolution of 10-600 kDa to test whether the assembly stays intact. In parallel, each single complex component was purified (Flag-Sfmbt, Flag-Scm and Flag-Ph₁₂₉₈₋₁₅₈₉) (Figure 3.6) and was run on the Superdex 200 column in order to monitor shifts in protein migration upon complex formation. While Flag-Sfmbt was only present in the void fractions, Flag-Scm and Flag-Ph₁₂₉₈₋₁₅₈₉ were trailing through all fractions with peaks at about 600 kDa and about 400 kDa, respectively (data not shown).

A

PhoRC



PRC1



Functional domains

■ FCS Zn finger
 ■ HD
 ■ MBT
 Ph-p specific regions
 ■ SAM
 ■ SLED
 ■ Zn finger

B

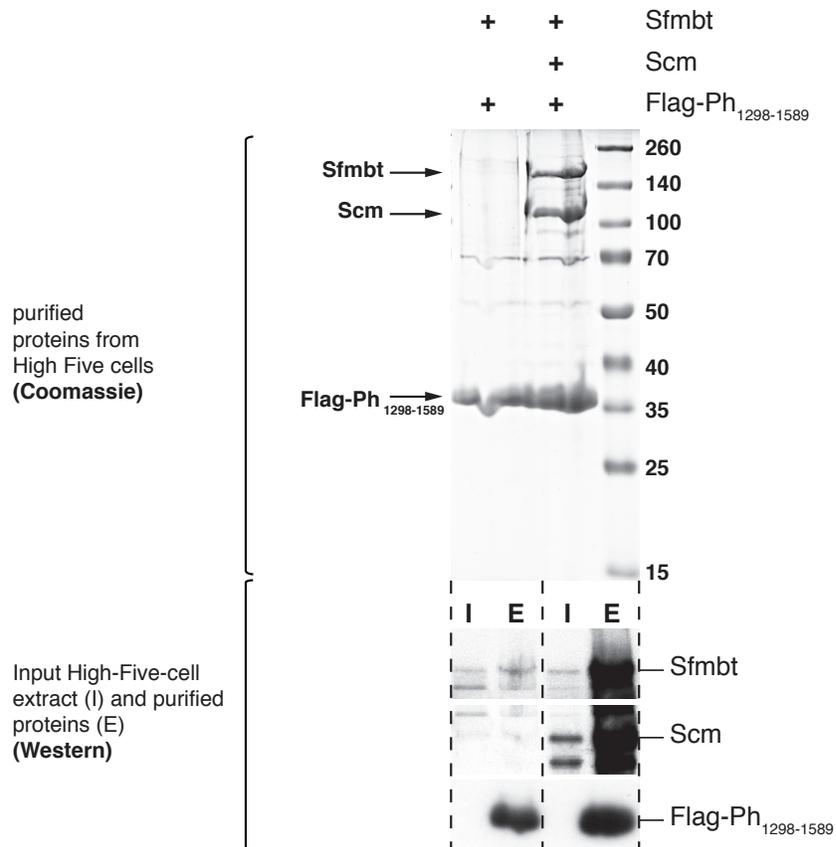


Figure 3.5: Scm mediates Ph interaction with Sfmbt.

(A) Schematic drawing of PRC1 and PhoRC complex members involved in Ph:Sfmbt

interaction. Known functional domains are indicated including zinc finger domains (ZF), homology domains (HD), sterile alpha motifs (SAM) and malignant brain tumor repeats (MBT). In addition, the regions unique to Ph-p, a 194 aa N-terminal part and a 60 aa stretch close to C-terminus are highlighted (striped pattern).

References: (1) (Alfieri et al., 2013), (2) (Mohd-Sarip et al., 2002), (3) (Wang et al., 2004b), (4) (Brown et al., 1998) (5) (Grimm et al., 2009), (6) (Klymenko et al., 2006), (7) (Kyba and Brock, 1998a), (8) (Robinson et al., 2012), (9) (Peterson et al., 1997), (10) (Kim et al., 2002), (11) (Kim et al., 2005), (12) (Grimm et al., 2007), (13) (Bezsonova, 2014).

(B) Proteins were enriched by Flag-affinity purification from High Five cell extracts containing either only Flag-Ph₁₂₉₈₋₁₅₈₉ and Sfm_{bt} (lane 1) or Flag-Ph₁₂₉₈₋₁₅₈₉, Scm and Sfm_{bt} (lane 2). Purified proteins were visualized by Coomassie staining. Input material as well as purified material was probed by western blot to validate comparable starting material in different cell extracts. Note that an N-terminal truncation of Ph was used for the interaction assays, since full length Ph was insoluble. The truncated version of Ph contains all the functional domains.

In the case where the complex was purified by pulling on Flag-Ph₁₂₉₈₋₁₅₈₉, it behaved very similarly to the single proteins, with a slight shift of Scm towards the void fractions (data not shown). However, the complex purified by pulling on Flag-Sfm_{bt} displayed a completely altered migration pattern. All three proteins were shifted to the void fraction with Flag-Sfm_{bt} and trailing of HA-Ph₁₂₉₈₋₁₅₈₉ and Scm through lower molecular weight fractions was strongly reduced (data not shown) indicating that the three proteins are associated in a high molecular weight assembly.

The apparent size of the complex was much bigger than expected from the sum of the molecular weights of the single proteins, since the purified complex was mainly found in the void fractions of the analytical S200 column. Therefore, I decided to run the complex on a 15-60 % sucrose gradient, which can resolve higher molecular weight assemblies (Figure 3.7). Comparable to runs on the analytical S200 column, Sfm_{bt} was only present in higher molecular weight fractions and did not trail through the gradient (Figure 3.7, up; upper panel), while single Flag-Scm and single Flag-Ph₁₂₉₈₋₁₅₈₉ proteins trailed all through the gradient. This running behavior is in agreement with their ability to form homopolymers. A peak was observed at the fractions corresponding approximately to the molecular weight of the single proteins (Figure 3.7 up; middle and lower panel).

summary, the trimeric complex purified by pulling on Flag-Sfmbt is stable on a sucrose gradient.

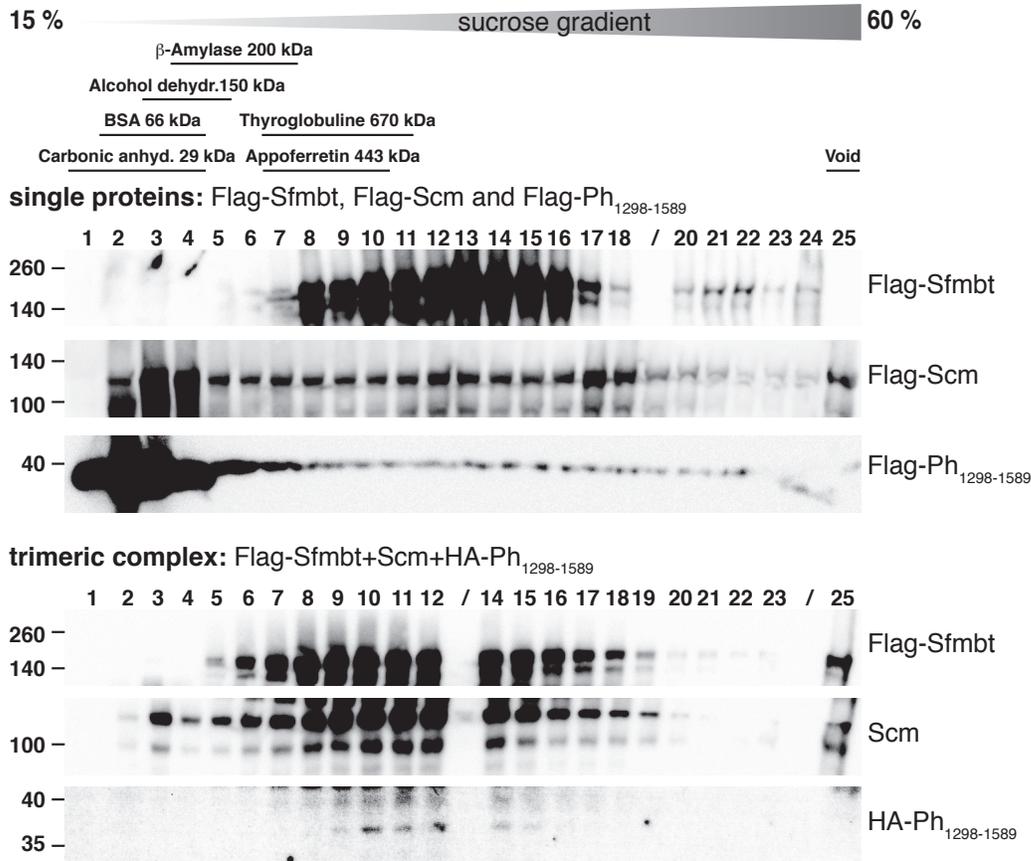


Figure 3.7: Sfmbt, Scm and Ph form a stable trimeric complex *in vitro*.

Flag-purified single proteins and trimeric complex were ~5x concentrated and run on a 15-60 % sucrose gradient. Marker proteins with known molecular weight were run in parallel on a 15-60 % sucrose gradient. The distribution of marker proteins is indicated on top of the panels with lines. The gradient was fractionated in 24x 0.5 ml plus one void fraction (25) of about ~1 ml. The fractions were TCA precipitated, resuspended in 60 μ l 2x LDS loading buffer and run on 8 % Tris-Glycine Midi gels. Western blots were probed for anti-Flag (up, all three panels; down, upper panel), anti-Scm (down, middle panel) or anti-HA (down, lower panel).

3.3 Two independent interaction sites are important for Scm:Sfmbt interaction

Having established that the PhoRC member Sfmbt forms a stable trimeric complex with the two PRC1 members Ph and Scm, I next wanted to investigate how the trimeric complex is assembled. Therefore, I first focused on the conserved functional domains, which likely play a role in the interaction. Scm and Ph have been shown to form homo- as well as heteropolymeric structures via their SAM domains (Peterson et al., 1997), this interaction was consolidated by a crystal structure of the copolymer of Ph- and Scm-SAM (Kim et al., 2005). Therefore, I proceeded by investigating the interaction surfaces of Scm and Sfmbt. Scm and Sfmbt form a complex *in vitro* and the interaction had previously been mapped to the N-terminal parts of both proteins (Grimm et al., 2009) (Figure 3.8 A). In agreement with that, I could isolate a stable complex between Scm and Sfmbt (Figure 3.8 B).

Next, I tested C-terminally truncated versions of both proteins for interaction using the same constructs that were used in earlier studies (Grimm et al., 2009). The N-terminal part of Sfmbt (Sfmbt₁₋₅₃₀) interacts with full length Scm (Figure 3.8 C, lane 1); however, in contrast to Grimm et al. I was unable to detect a stable interaction between the Sfmbt₁₋₅₃₀ protein and the C-terminally truncated Scm₁₋₄₃₅ or Scm₁₋₁₇₀ proteins (Figure 3.8 C, lanes 2 and 3). In summary, the interaction mapping showed that the N-terminal part of Sfmbt can interact with full length Scm, but not with C-terminally truncated versions of Scm that contain either only the FCS zinc finger domains (Scm₁₋₁₇₀) or the FCS zinc finger domains and the MBT repeat domain (Scm₁₋₄₃₅).

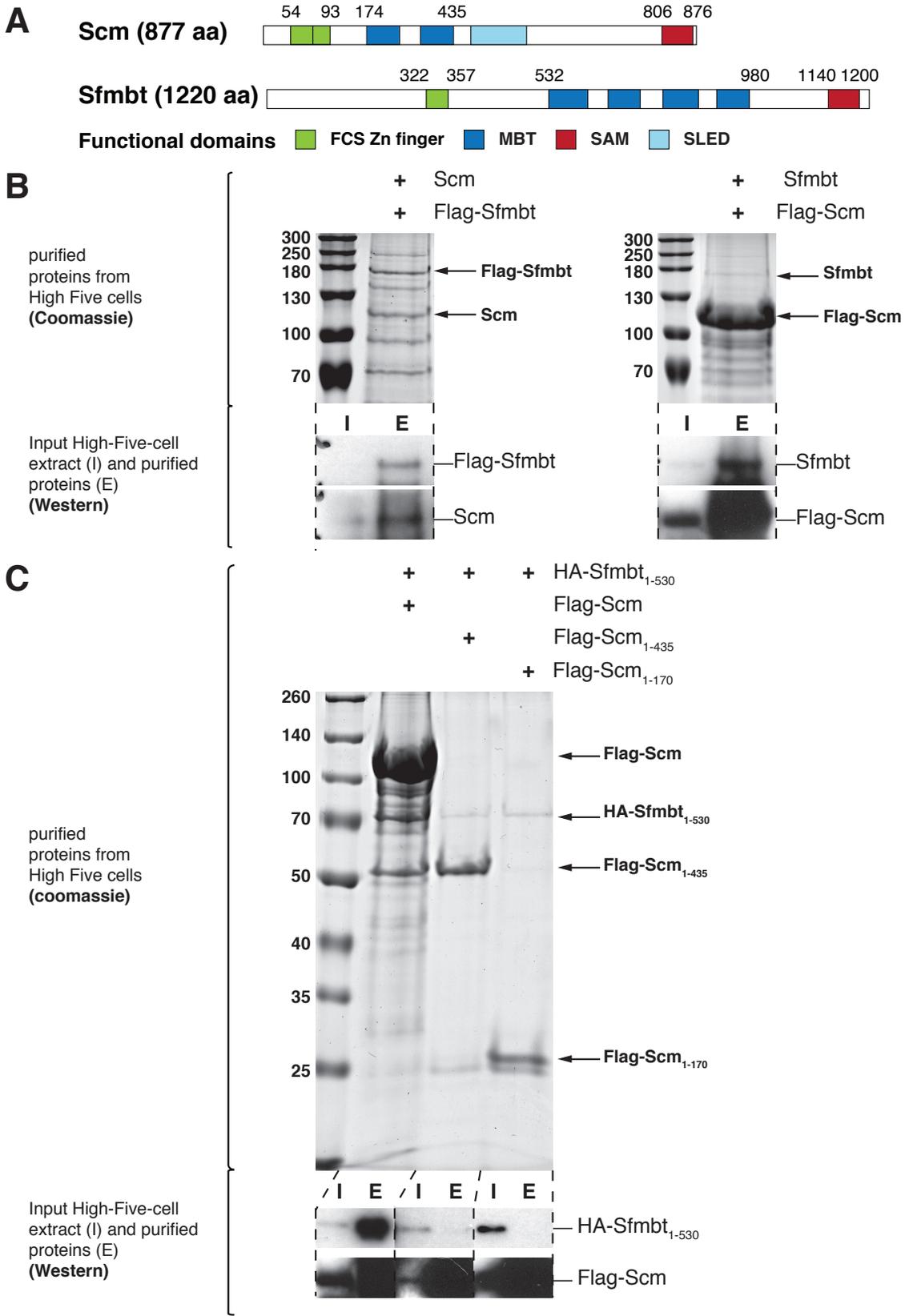


Figure 3.8: An N-terminal portion of Sfmbt interacts with Scm.

(A) Schematic drawing of Scm and Sfmbt with functional domains including zinc finger domains, MBT (malignant brain tumor) repeats and SAM (sterile alpha motif) domains.

(B) Proteins were enriched by Flag-affinity purification from a High Five cell extract containing either Flag-Sfmbt and Scm (panel 1) or Flag-Scm and Sfmbt (panel 2) co-expressed. Purified proteins were visualized by Coomassie staining. Input material as well as purified material was probed by western blot to validate comparable starting material in different cell extracts. Note that Flag-Scm is much stronger expressed than Flag-Sfmbt and has various degradation bands.

(C) Proteins were enriched by Flag-affinity purification from High Five cell extract containing either Flag-Scm and HA-Sfmbt₁₋₅₃₀ (panel 1), Flag-Scm₁₋₄₃₅ and HA-Sfmbt₁₋₅₃₀ (panel 2) or Flag-Scm₁₋₁₇₀ and HA-Sfmbt₁₋₅₃₀ (panel 3) co-expressed. Purified proteins were visualized by Coomassie staining. Input material as well as purified material was probed by western blot. HA-Sfmbt₁₋₅₃₀ forms a complex with Scm, but not with truncated Scm versions Flag-Scm₁₋₄₃₅ or Flag-Scm₁₋₁₇₀. Note that there is a faint band visible in Coomassie staining at the expected molecular weight of HA-Sfmbt₁₋₅₃₀ in lanes 2 and 3; however, no enrichment of HA-Sfmbt₁₋₅₃₀ was detectable by western blotting.

3.3.1 Scm:Sfmbt interaction is not dependent on FCS zinc finger domains

The N-terminal portion of Sfmbt (Sfmbt₁₋₅₃₀), which is still able to mediate interaction to Scm, contains the FCS zinc finger domain as the major recognizable protein domain. Zinc finger domains are small, independent modules that are folded around a central zinc ion, which stabilizes their structure. These domains have been proposed to mediate diverse interactions including protein-DNA, protein-RNA as well as protein-protein interactions (Gamsjaeger et al., 2007). A number of PcG proteins, namely Ph, Scm, Sfmbt as well as dL(3)MBTL2 contain an atypical C2C2 zinc finger domain termed FCS zinc finger due to its characteristic phenylalanine-cysteine-serine sequence motif. The FCS zinc fingers of PcG proteins have particularly been implicated in unspecific binding of regulatory RNA molecules (Wang et al., 2011); however, their exact role is still not clear. Therefore, I wanted to address whether the FCS zinc finger domains play a role in the Scm:Sfmbt interaction.

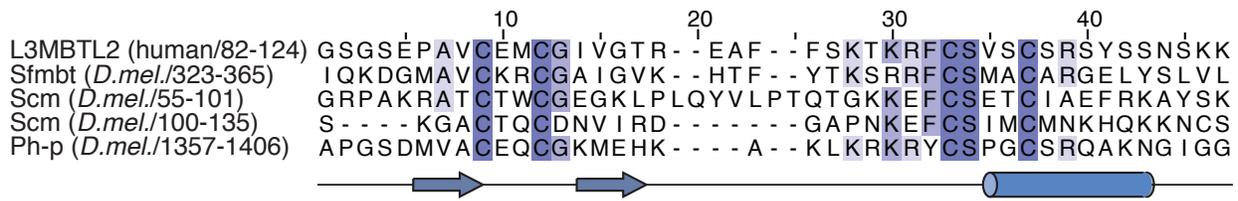
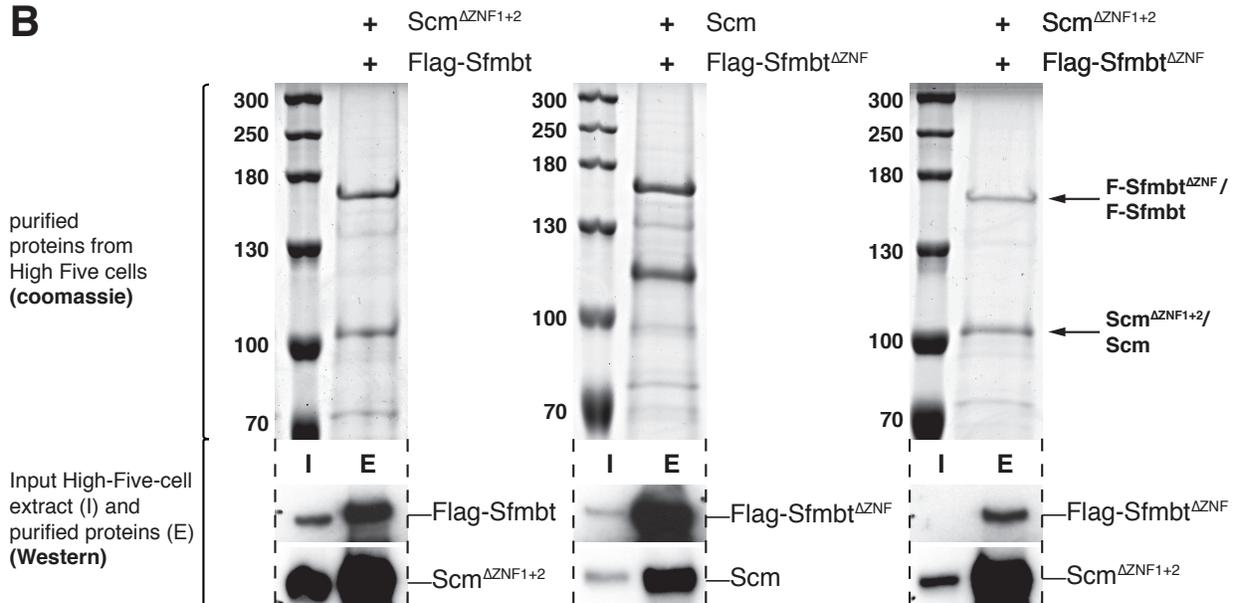
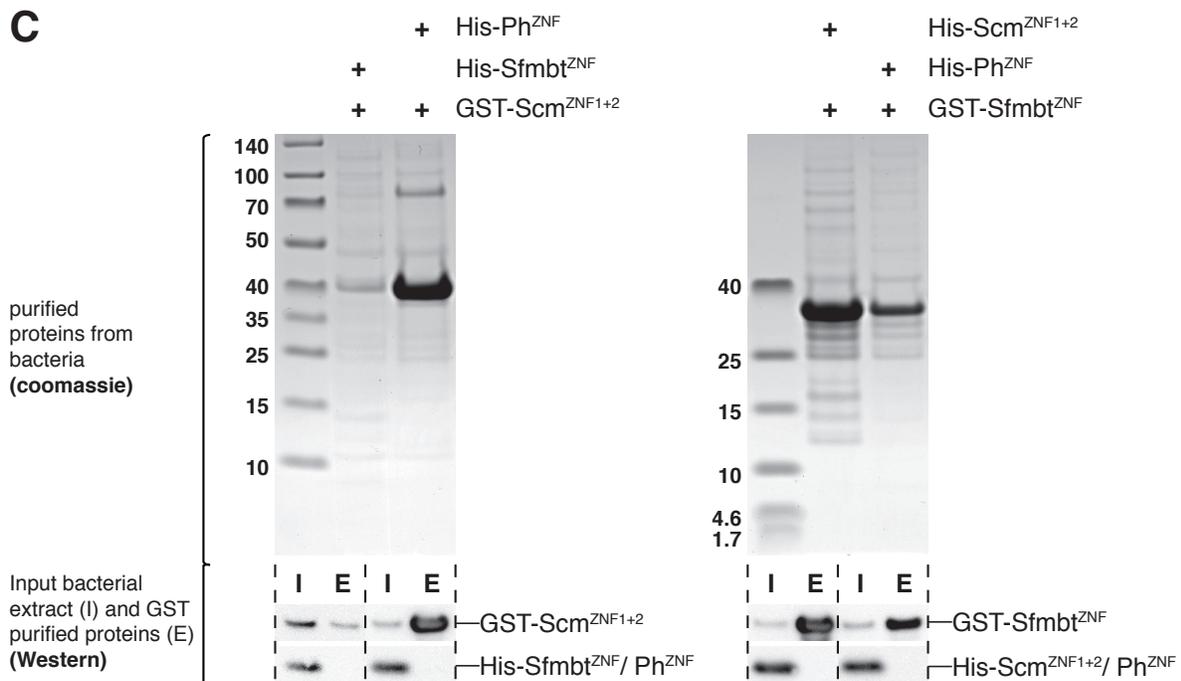
A**B****C**

Figure 3.9: The Scm:Sfmtb interaction is not dependent on FCS zinc finger domains.

(A) Sequence alignment of FCS zinc finger domains of human L3MBTL2 and the *Drosophila* proteins Sfmbt, Scm and Ph-p. The blue colored secondary structure illustrated below is based on the solution structure of human L3MBTL2 FCS zinc finger (Lechtenberg et al., 2009). The alignment was assembled by T-coffee and edited with Jalview. The color pattern is according to sequence identity.

(B) Proteins were enriched by Flag-affinity purification from High Five cell extract containing different zinc finger deletion mutants either Flag-Sfmbt and Scm^{ΔZNF1+2} (panel 1), Flag-Sfmbt^{ΔZNF} and Scm (panel 2) or Flag-Sfmbt^{ΔZNF} and Scm^{ΔZNF1+2} (panel 3) coexpressed. Purified proteins were visualized by Coomassie staining. Input material as well as purified material was probed by western blot to validate comparable starting material in different cell extracts. Zinc finger deletion constructs were cloned by deleting aa55-135 of Scm and aa323-365 of Sfmbt. These are the same borders shown in the alignment and used for expressing zinc finger domains in bacteria.

(C) *Drosophila* zinc finger sequences shown in A were cloned into pEC vectors with either His-GST-tag or His-tag. Proteins were enriched by GST-Pulldown from bacterial lysates expressing either GST-Scm^{ZNF1+2} and His-Sfmbt^{ZNF}, GST-Scm^{ZNF1+2} and His-Ph^{ZNF}, GST-Sfmbt^{ZNF} and His-Scm^{ZNF1+2} or GST-Sfmbt^{ZNF} and His-Ph^{ZNF}. 30 % of purified proteins were run on a 4-12 % gradient gel and visualized by Coomassie staining. Input material (0.0075 %) and purified material (5 %) were probed by western blot.

To test this hypothesis, I conducted pull-down experiments upon deletion of the FCS zinc finger domains in the full-length context as well as direct interaction studies with the expressed zinc finger domains. In order to define borders for the *Drosophila* PcG FCS zinc fingers I took advantage of the human L3MBTL2 FCS zinc finger solution structure, which was solved previously, and consists of two β-strands that form an anti-parallel sheet, followed by a 13-residue loop and an α-helix towards the end of the sequence (Lechtenberg et al., 2009). By sequence alignment of human L3MBTL2 to *Drosophila* Sfmbt, Ph and Scm zinc fingers I determined the residues involved in the zinc finger structure (Figure 3.9 A).

In a first approach I deleted the zinc finger domains from the full-length Scm and Sfmbt proteins and subsequently checked with Flag pull-down with Baculo expressed proteins whether the interaction was lost. Scm and Sfmbt proteins can still interact with each other in the absence of (1) the Sfmbt zinc finger, (2) both Scm zinc fingers or (3) the Sfmbt and Scm zinc fingers (Figure 3.9 B, panel 1, 2 and 3, respectively).

In conclusion, the FCS zinc finger domains are not essential for formation of the Scm:Sfmbt complex. To extend this analysis and exclude that a contribution of the FCS ZNF for complex formation would be masked by other interaction surfaces, I tested whether the isolated Scm and Sfmbt zinc finger domains interact. I coexpressed the FCS zinc finger domains of Scm and Sfmbt in bacteria with or without GST-tag in order to perform GST pull-down assays. I could not detect any enrichment of Sfmbt zinc finger (aa323-365) with GST-Scm zinc fingers (aa55-135) or vice versa (Figure 3.9 C, panel 1 and 2, respectively). As a control for specificity the Ph zinc finger (aa1357-1406) was expressed, but did not show any unspecific binding to either of the GST-tagged zinc fingers (Figure 3.9 C). To summarize, the FCS zinc fingers of Scm and Sfmbt cannot directly interact with each other and are apparently not necessary for Scm:Sfmbt interaction.

3.3.2 Putative domains involved in Scm:Sfmbt interaction

After excluding the FCS zinc fingers as possible domains mediating interaction between Scm and the N-terminal portion of Sfmbt, it appears that none of the previously described functional domains are involved. Therefore, I focused on identifying possible undescribed domains by having a closer look at the secondary structure prediction of Scm and Sfmbt proteins.

The first 1-530 amino acids of Sfmbt are still able to interact with Scm although somehow less efficiently than the full-length protein. Analysis of the Sfmbt N-terminus by the secondary structure prediction tool PSIPRED revealed that there is secondary structure predicted primarily for the N-terminal 100 amino acids of Sfmbt (Figure 3.10 A). In contrast to the other described functional domains, which are highly conserved throughout evolution, this part of the protein is not conserved in vertebrates. By blast searches of the first 300 amino acids in NCBI only insect orthologs were identified. The identity ranged from 91 % in *Drosophila yakuba* (6-15 myr divergence), to 67 % in *Drosophila virilis* (40 myr divergence) and 51 % in *Ceratitis capitata* (~100 myr divergence). Even though this part of the protein was mainly conserved in drosophilid species, it also showed significant conservation in *Ceratitis capitata* (~100 myr divergence). However, more distant insects such as *Anopheles gambiae*

(~260 myr divergence) or *Bombus impatiens* (~290 myr divergence) do not contain the first 100 amino acids in their Sfmbt orthologs (Figure 3.10 A).

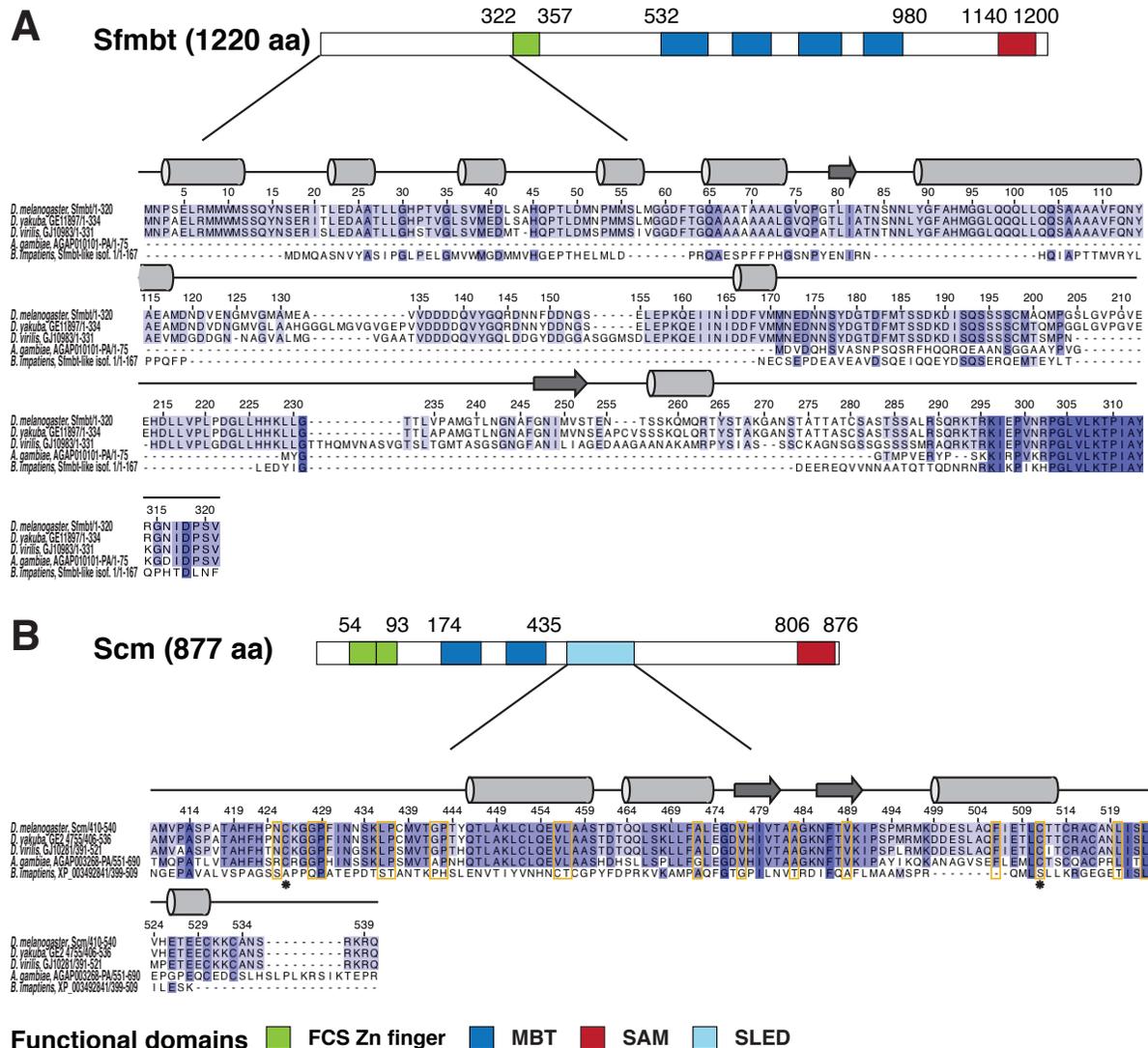


Figure 3.10: Putative domains involved in Scm:Sfmbt interaction.

(A) Schematic drawing of Sfmbt with functional domains including FCS zinc finger domain, MBT (malignant brain tumor) repeats and SAM (sterile alpha motif) domain. Shown below is a sequence alignment of Sfmbt N-terminus (aa 1-320) of insect orthologs including *D. melanogaster*, *D. yakuba*, *D. virilis*, *A. gambiae* and *B. impatiens*. The alignment was assembled by T-coffee and edited with Jalview. The color pattern is according to sequence identity. The secondary structure prediction shown on top was conducted for Sfmbt using PSIPRED.

(B) SLED domain as putative interaction surface in Scm. Schematic drawing of Scm

with functional domains including FCS zinc finger domains, MBT (malignant brain tumor) repeats and SAM (sterile alpha motif) domain. Shown below is a sequence alignment of the MBT adjacent part of Scm (aa 410-540) of insect orthologs including *D. melanogaster*, *D. yakuba*, *D. virilis*, *A. gambiae* and *B. impatiens*. The alignment was assembled by T-coffee and edited with Jalview. The color pattern is according to sequence identity. The secondary structure prediction shown on top was conducted for Scm using PSIPRED. The residues highlighted in orange in the alignment are conserved up to humans. Two single point mutations (C425Y and C511Y) within the SLED domain have been reported to result in partial loss of function of Scm in flies (Borneman et al., 1998) and are marked with an asterisk. The first mutation (C425Y) is located in a putative binding surface, while the second mutation (C511Y) was mapped to the DNA binding site of the SLED domain (Bezsonova, 2014).

For Scm analysis, I focused on the region between the MBT repeats and the SAM domain, since this part seems to be essential for Sfmfbt interaction. I showed that Scm could not interact with Sfmfbt via its N-terminal 1-435 amino acids. In addition, Grimm et al. demonstrated that an Scm-SAM deletion construct, which lacks the SAM domain, could still interact with the N-terminal portion of Sfmfbt (Grimm et al., 2009). Furthermore, neither the FCS zinc fingers nor the MBT domains of Scm are involved in interaction to Sfmfbt (Figure 3.9 and Grimm, unpublished).

Interestingly, the region downstream of the MBT domains (aa 410-540), contains a so-called Scm-Like Embedded Domain (SLED). The SLED domain was only very recently identified in human Scml2 (Sex Comb on Midleg-Like 2) and its structure was determined (Bezsonova, 2014). The SLED domain is conserved from metazoans to humans and is found exclusively within the Scm family of proteins, where it is located directly adjacent to the MBT repeats (Bezsonova, 2014). The degree of conservation of this domain (aa 410-540) ranges from 100 % identity in *Drosophila yakuba* to 92 % identity in *Drosophila virilis*. More distant insects such as *Anopheles gambiae* and *Bombus impatiens* exhibit 63 % and 27 % identity, respectively (Figure 3.10 B). A simple Blast search with the *Drosophila melanogaster* SLED domain did not pick up on any mammalian homologs. However, some of the highly conserved residues identified in the human SLED containing proteins (Bezsonova, 2014) are indeed present in the *Drosophila* sequence (Figure 3.10, orange boxes).

Bezsonova et al., who solved the in-solution structure of the human Scml2 SLED domain, reported that this domain binds DNA with μM affinity and that it may also engage in binding additional ligands via a loop region located on the opposite face to the identified DNA-binding site (Bezsonova, 2014). Intriguingly, a hypomorphic Scm allele was assigned to a single point mutation in *Drosophila* Scm (C425Y, allele k2), which is located in this particular loop region (Borneman et al., 1998).

To summarize, it is possible that the first 100 amino acids of Sfmbt and the SLED domain of Scm are putative interaction sites for Scm/Sfmbt complex formation. However, further analysis would be needed to assess this idea.

3.3.3 Scm and Sfmbt also interact via their SAM domains

Strikingly, all three PcG proteins – Sfmbt, Scm and Ph – involved in PhoRC:PRC1 interaction contain a highly conserved C-terminal SAM domain. SAM domains are small ~70 amino acid long helical domains which can mediate protein-protein interactions and are involved in various biological processes ranging from signal transduction to transcriptional and translational regulation (Qiao et al., 2005). In addition, some SAM domains have been implicated in RNA binding (Aviv et al., 2003; Green et al., 2003). While the SAM domains of Ph and Scm are known to form homo- as well as heteropolymers and have been studied extensively (Kim et al., 2002, 2005; Peterson et al., 1997; Robinson et al., 2012), little is known about the role of the Sfmbt-SAM domain. Even though the Sfmbt-SAM domain does not appear to be essential for the interaction with Scm (Figure 3.8 C), it could still play a role in Scm:Sfmbt interaction, possibly constituting a second interaction surface.

In order to test this hypothesis, I conducted pull-down experiments with (1) Sfmbt₁₋₅₃₀, a C-terminally truncated Sfmbt version that contains amino acids 1-530 and thus lacks the MBT repeats and SAM domain, (2) Sfmbt₅₃₀₋₁₂₂₀, an N-terminally truncated Sfmbt version, that contains the MBT repeats and SAM domain, but lacks the N-terminus and (3) Sfmbt₅₃₀₋₁₁₃₆, an N-terminally truncated version of Sfmbt that lacks the SAM domain (Figure 3.11 A). Not only the N-terminal part, but also the C-terminal part of Sfmbt can interact with Scm independently, indicating that the Scm:Sfmbt interaction is mediated by multiple surfaces and not just a single interaction site (Figure 3.11 B, lane 1 and 2).

Remarkably, deletion of the SAM domain in the N-terminally truncated Sfmbt version results in severely reduced Scm enrichment (Figure 3.11 B, lane 3). To conclude, I identified the SAM domain of Sfmbt as a second interaction site for binding to Scm.

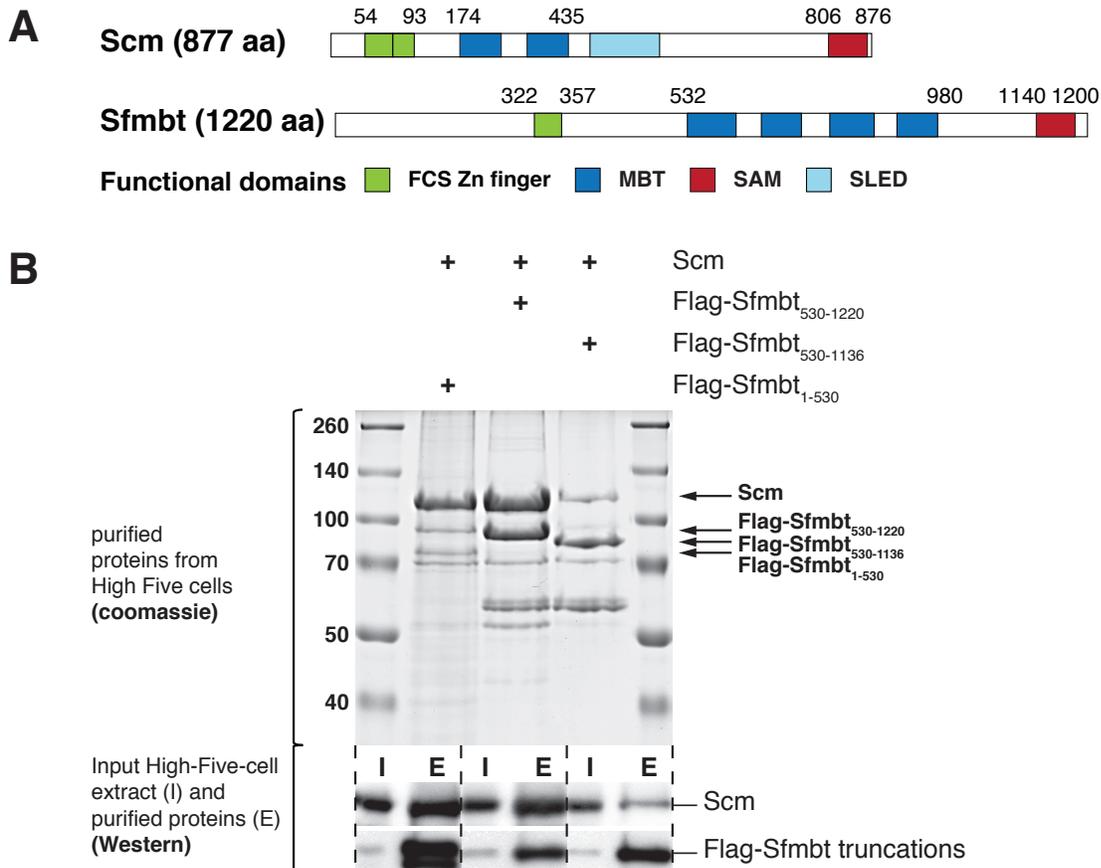


Figure 3.11: Scm and Sfmbt also interact via their SAM domains.

(A) Schematic drawing of Sfmbt and Scm with functional domains including zinc finger domains, MBT (malignant brain tumor) repeats and SAM (sterile alpha motif) domain. (B) Proteins were enriched by Flag-affinity purification from High Five cell extract containing Flag-Sfmbt₁₋₅₃₀ and Scm (lane 1), Flag-Sfmbt₅₃₀₋₁₂₂₀ and Scm (lane 2) or Flag-Sfmbt₅₃₀₋₁₁₃₆ and Scm (lane 3) coexpressed. Purified proteins were visualized by Coomassie staining. Input material as well as purified material was probed by western blot to validate comparable starting material in different cell extracts.

3.3.4 The Scm and Sfm bt SAM domains form a stable complex

Since the SAM domain of Sfm bt is important for interaction with Scm, it seems likely that it directly interacts with the SAM domain of Scm. To verify this assumption, I first had a closer look at the interaction mode of Ph- and Scm-SAM domains. Ph- and Scm-SAM each contain two interaction surfaces, the mid loop surface (α 3-4, ML) and the end helix surface (α 4-5, EH) (Figure 3.12 A, ML (green), EH (red)). These two surfaces constitute the interface in homomeric as well as heteromeric SAM:SAM interactions (Kim et al., 2002, 2005) (Figure 4.2). Kim et al. found that SAM polymers assemble preferentially in a head-to-tail fashion. Furthermore, in the Ph-SAM/Scm-SAM copolymer the ML surface of Ph strongly favors binding to the EH surface of Scm, hence supporting the model of individual Scm-SAM and Ph-SAM blocks that are joined by a single junction (Kim et al., 2002, 2005).

Surprisingly, sequence alignment of SAM domains present in PcG proteins showed that Sfm bt-SAM does not contain a functional ML surface. In the Sfm bt-SAM domain, most of the residues forming the hydrophobic core of the ML surface present in Ph- and Scm-SAM domains are replaced by polar residues (R1170^{Sfm bt}, Q1173^{Sfm bt}, D1178^{Sfm bt}) rendering self-association of Sfm bt-SAM domains unlikely (Figure 3.12 A). In particular, an A to R amino acid substitution in one of the residues essential for forming the ML surface (R1170^{Sfm bt}) would result in steric hindrance of the interface (Figure 3.12 A, black asterisk). In contrast, the EH surface of Sfm bt-SAM is more conserved, suggesting that it could engage in heterodimer formation with the Scm-SAM domain.

Therefore, I expressed GST-His-Sfm bt-SAM and His-Scm-SAM constructs in bacteria, purified the recombinant protein by Ni-affinity and gel-filtration chromatography and conducted GST pull-down assays with the purified SAM domains. Since I wanted to prevent Scm from forming homopolymers, I mutated one of the residues of the EH surface (L859R^{Scm}), which is needed for Scm homopolymer formation, but would still allow for heteromeric interaction with Sfm bt-SAM via the Scm-SAM ML surface. In gel-filtration analysis the Sfm bt-SAM domain migrated at the size of a monomer, confirming that it cannot form homopolymers (data not shown).

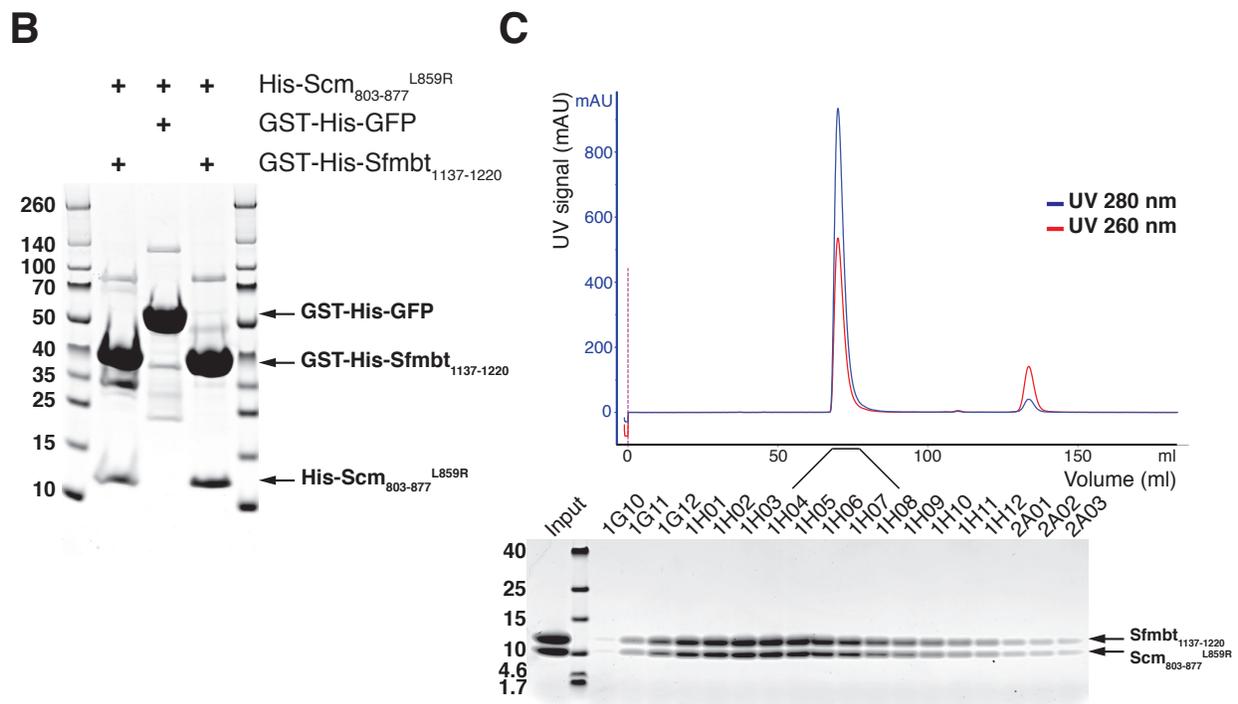
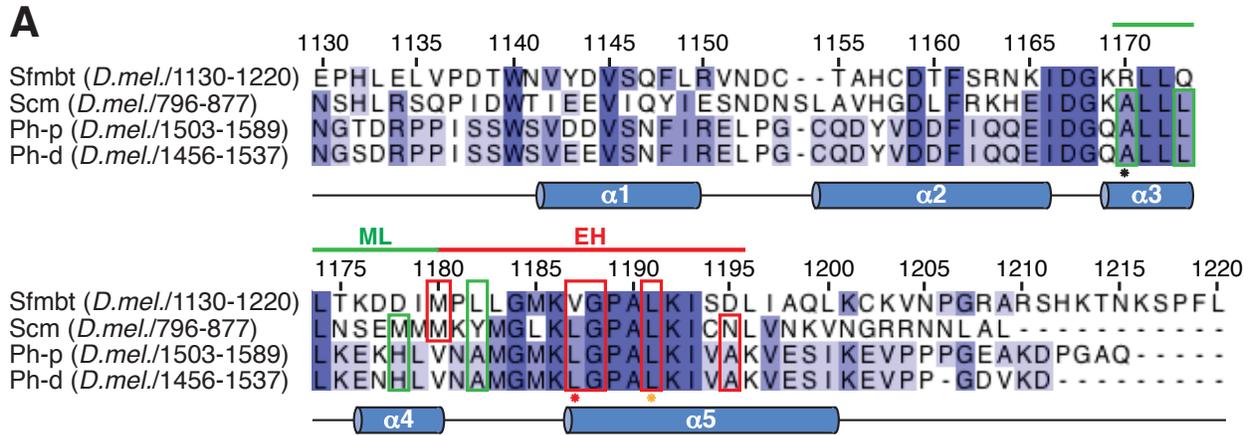


Figure 3.12: The Scm and Sfmbt SAM domains form a stable complex.

(A) Sequence alignment of SAM domains of the *Drosophila* PcG proteins Sfmbt, Scm, Ph-p and Ph-d. The color pattern is according to sequence identity. The hydrophobic residues that form the core of the two interaction surfaces, mid-loop surface (ML) and end-helix (EH) surface, are marked in green and red, respectively (Kim et al., 2002). The five helices present in the Ph-SAM structure are illustrated below the alignment as blue cylinders (Kim et al., 2002). Note that Sfmbt-SAM lacks a functional mid loop surface rendering it unlikely that Sfmbt can form homopolymers. For interaction studies either an L859R^{Scm} point mutation (orange asterisk) or an L855E/L859E^{Scm} double mutation (red and orange asterisk) were introduced into Scm EH surface to prevent Scm-SAM self-association. The sequence alignment was assembled by T-coffee and

edited with Jalview.

(B) GST pull-down assay of purified Sfbmt- and Scm-SAM^{L859R} domains. Sfbmt-SAM (1137-1220) and Scm-SAM (803-877^{L859R}) were either purified separately (lane 1) or coexpressed and purified together (lane 3). As a control GST-GFP and His-Scm-SAM^{L859R} were employed (lane 2). Purified proteins were run on a 4-12 % gradient gel and visualized by Coomassie staining.

(C) Gel-filtration of purified Sfbmt-SAM and Scm-SAM^{L859R} domains with a Superdex75 chromatography column showed that these domains comigrate during gel-filtration and elute in a single peak. The peak fractions 1G10-2A03 (1 % of total volume) were run on a 4-12 % gradient gel and visualized by Coomassie staining.

Nevertheless, Sfbmt-SAM can form a heterodimer with Scm-SAM (Figure 3.12 B). The SAM domains were more stable when coexpressed and copurified (Figure 3.12 B, lane 3) compared to being separately expressed, purified and only mixed upon GST-binding (Figure 3.12 B, lane 1). His-Scm-SAM did not display any background binding to the GST control GST-GFP (Figure 3.12 B, lane 2).

In addition to that, untagged Sfbmt- and Scm-SAM comigrate in the same fractions during gel-filtration chromatography demonstrating that these two SAM domains form indeed a stable complex *in vitro* (Figure 3.12 C). To summarize, SAM domain interaction is not only important for Ph:Scm interaction, but also plays a role in Scm:Sfbmt interaction establishing a second interaction site between these two proteins.

3.3.5 Structure of the Scm-SAM/ Sfbmt-SAM complex

To gain structural insight into how the PRC1 complex is tethered to Polycomb response elements (PREs) at PcG target genes via the DNA binding PhoRC complex, Christian Benda (Department of Structural Cell Biology, MPI Biochemistry) and I determined the crystal structure of the Scm-SAM/ Sfbmt-SAM complex. Therefore, point mutations were introduced to disrupt one of the polymer interfaces of Scm to be able to work with largely monomeric and soluble domains. Based on the known structure of Scm (Kim et al., 2005), I introduced either a single point mutation (L859R^{Scm}, Figure 3.12 A, – orange asterisk) or two point mutations (L855E^{Scm}, L859E^{Scm} Figure 3.12 A, – red and orange asterisk) in the end helix (EH) surface of Scm to abolish self-association. While the complex containing Scm-SAM^{L859R} only yielded crystals containing oligomeric

Scm-SAM^{L859R}, the Scm-SAM^{L855E/L859E}/Sfmbt-SAM complex yielded crystals containing the Scm-SAM^{L855E/L859E}/Sfmbt-SAM dimer.

3.3.5.1 Overall Structure of the Scm-SAM/ Sfmbt-SAM complex

The structure of the Scm-SAM/ Sfmbt-SAM complex was solved by molecular replacement with Ph-SAM/ Scm-SAM (PDB ID: 1PK1) as search model and refined to 1.975 Å resolution. The asymmetric unit contains two Scm/ Sfmbt heterodimers that interact laterally to form a tetrameric ensemble (Figure 3.13 A). One Scm-SAM/ Sfmbt-SAM complex is displayed in Figure 3.13 B in the front and top view. The Sfmbt-SAM domain is a helical bundle containing five α-helices similar to the Ph- and Scm-SAM domain. Analysis of the interfaces in the tetramer showed that the major interaction is between the ML surface of Scm-SAM and the EH surface of Sfmbt-SAM in both dimers present in the asymmetric unit. The buried area in these interfaces is between 520 and 580 Å² and there are three salt bridges and two hydrogen bonds present. These two interaction surfaces have previously been shown to be crucial for SAM:SAM interactions (Kim et al., 2001, 2002, 2005). While the two Scm-SAM domains and the two Sfmbt-SAM domains present in one tetramer showed little contact in the crystal structure, there are several interactions observed between the Scm-SAM domain of dimer one and the Sfmbt-SAM domain of dimer two, and between the Sfmbt-SAM domain of dimer one and the Scm-SAM domain of dimer two, respectively (Figure 3.13 A). However, the buried area in these interfaces is considerably smaller (~370 Å²) and there are fewer interactions present. These interactions include one hydrogen bond (L877^{Scm} – K1176^{Sfmbt}) and one salt bridge (E859^{Scm} – K1201^{Sfmbt}) and are most likely due to the close packing of the two dimers in the crystal lattice. The salt bridge occurs via one of the mutated residues (E859^{Scm}) introduced into the Scm EH surface and cannot form in the Scm-SAM wt domain. Since the Scm-SAM/ Sfmbt-SAM complex behaves as a heterodimer in gel-filtration chromatography, even at elevated concentrations, it seems likely that the observed lateral interactions are an artifact of crystallography.

3.3.5.2 Structure of the Scm-SAM/ Sfmbt-SAM Interface

The interaction between Scm-SAM and Sfmbt-SAM is comparable to the Ph-SAM/ Scm-SAM interface and the interfaces observed in Ph-SAM or Scm-SAM homopolymers (Kim et al., 2002, 2005). The interface of an Scm-SAM/ Sfmbt-SAM dimer is formed by the ML surface of Scm (α 3-4) and the EH surface Sfmbt (α 4-5) (Figure 3.13 C). Consistent with previously published SAM domain structures, the interface comprises a hydrophobic core region, which is enlaced by a polar region. The hydrophobic side chains of A838^{Scm}, L841^{Scm}, L842^{Scm}, M846^{Scm} and the aromatic moiety of Y850^{Scm} of the Scm-SAM ML surface interact with M1180^{Sfmbt}, V1187^{Sfmbt}, G1188^{Sfmbt} and L1191^{Sfmbt} forming the hydrophobic core region on the Sfmbt-SAM EH surface (Figure 3.13 C, left panel). Additionally, there are multiple hydrogen bonds and salt bridges stabilizing the complex. A polar patch of the Scm-SAM ML surface mediates three salt bridges (E833^{Scm} – K1186^{Sfmbt}, D835^{Scm} – K1192^{Sfmbt} and K849^{Scm} - D1177^{Sfmbt}) and two additional hydrogen bonds to the polar residues of the Sfmbt-SAM EH surface. The hydrogen bonds are mediated by the backbone carbonyl oxygen of H832^{Scm} (H832O) to the backbone Nitrogen of G1188^{Sfmbt} (G1188N) and by Y850^{Scm} to D1177^{Sfmbt} (Figure 3.13 C, middle panel (front view) and right panel (top view)). Additionally, the interface is stabilized by two water molecules bridging residues Y850^{Scm} and D1177^{Sfmbt}, respectively residues E833^{Scm} and K1192^{Sfmbt} (not shown). Furthermore, the Scm-SAM C-terminal tail was not present in the previously reported Scm-SAM structure (Kim et al., 2005), but the Scm structure reported here revealed that the Scm C-terminus is packed against α 5 of the Scm-SAM domain (Figure 3.13 B).

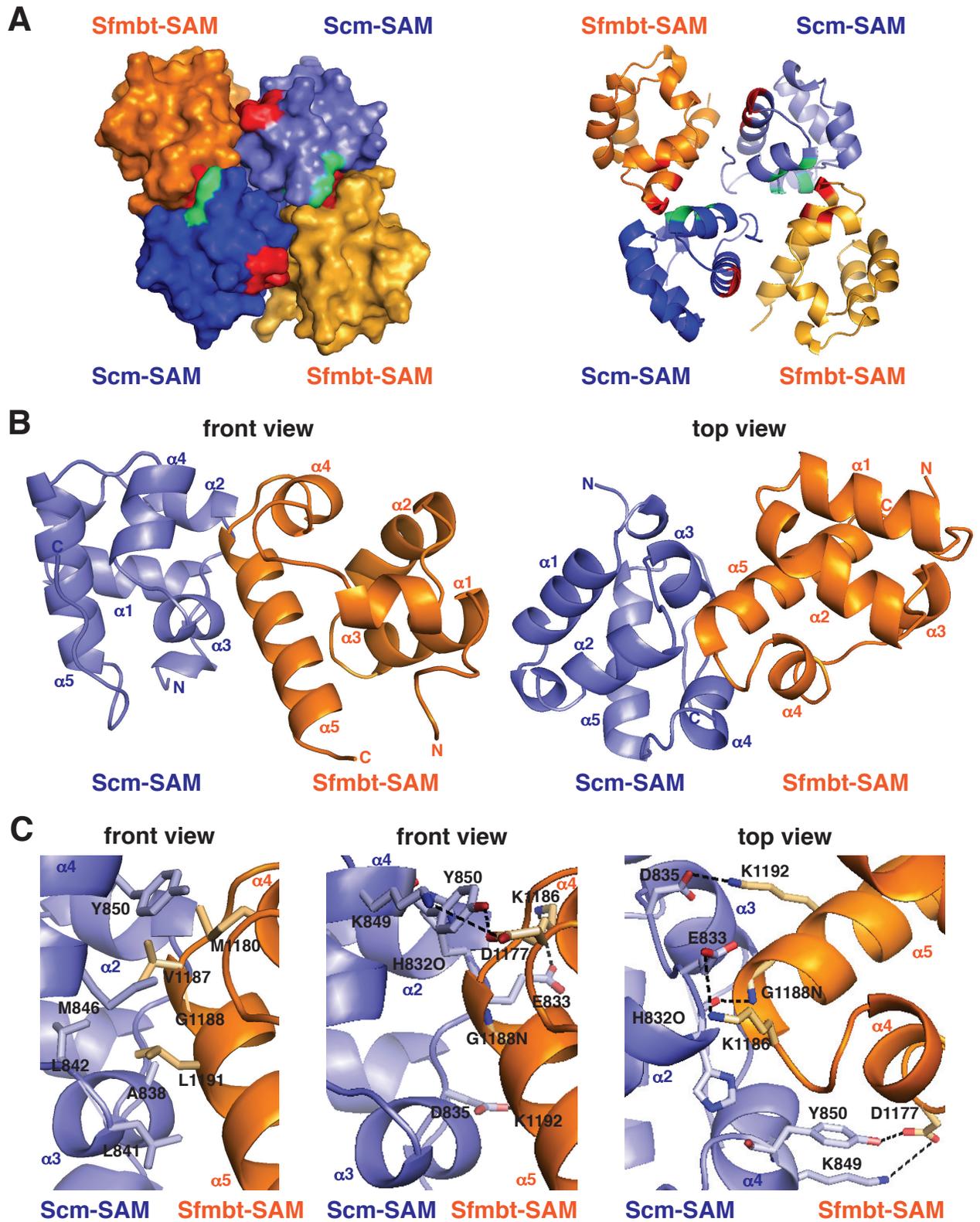


Figure 3.13: Structure of the Scm-SAM/ Sfmbt-SAM complex.

A: The asymmetric unit contains two Scm-SAM/ Sfmibt-SAM heterodimers that interact laterally to form a tetrameric ensemble. Surface presentation (right) and ribbon diagram (left) of tetrameric ensemble. Scm-SAM is depicted in blue Sfmibt-SAM in orange. One Scm-SAM/ Sfmibt-SAM dimer is shown in dark color and the second one in slightly lighter color. The hydrophobic core of the ML surface (green) and EH surface (red) is highlighted.

B: An Scm-SAM/ Sfmibt-SAM dimer is shown in two orientations. The front view (left) and the top view (right). Scm-SAM is depicted in blue, Sfmibt-SAM in orange.

C: Close-up view of the Scm-SAM/ Sfmibt-SAM interface. The left panel highlights the hydrophobic contacts shown in the front view. The numbering of residues corresponds to the full-length protein. The residues comprising the polar patch are illustrated in the middle panel (front view) and in the right panel (top view). Atoms are colored in red for oxygen and blue for nitrogen. Hydrogen bonds and salt bridges are indicated with dashed lines. No backbone atom designations indicate that the side chains mediate the interaction.

3.3.5.2 Comparison of SAM/ SAM Interfaces of Ph, Scm and Sfmibt

A schematic illustration comparing the most conserved hydrophobic and polar interactions involved in the SAM/ SAM interfaces of (1) Ph-SAM/ Ph-SAM, (2) Scm-SAM/ Scm-SAM, (3) Ph-SAM/ Scm-SAM and (4) Scm-SAM/ Sfmibt-SAM is shown in Figure 3.14. While very similar residues are involved in forming the polar interface of Scm-SAM/ Sfmibt-SAM, the specific interactions vary somehow compared to the other hetero- and homo-SAM interfaces.

Compared to the polar interactions in the Scm-SAM homopolymer, which involve two hydrogen bonds ($H832^{Scm} - L855N^{Scm}$, $H832O^{Scm} - G856N^{Scm}$) and one salt bridge ($D835^{Scm} - K860^{Scm}$), the polar interactions of the ML surface of Scm to Sfmibt-SAM EH surface include two hydrogen bonds ($H832O^{Scm} - G1188N^{Sfmibt}$, $Y850^{Scm} - D1177^{Sfmibt}$) and three salt bridges ($E833^{Scm} - K1186^{Sfmibt}$, $D835^{Scm} - K1192^{Sfmibt}$ and $K849^{Scm} - K1177^{Sfmibt}$). Only the salt bridge ($D835^{Scm} - K860^{Scm}$, respectively $D835^{Scm} - K1192^{Sfmibt}$) is conserved. This salt bridge appears to be one of the most conserved polar interactions, since it can be found in all four interfaces. The other interactions differ, even though they involve similar residues.

Compared to the four residues of the EH surface mediating the polar interactions of Scm ($K854^{Scm}$, $L855^{Scm}$, $G856N^{Scm}$ and $K860^{Scm}$) and Ph ($K1560^{Ph}$, $L1561^{Ph}$, $G1562^{Ph}$

and K1566^{Ph}), the polar interactions of the Sfmibt EH surface are mediated by three conserved residues (K1186^{Sfmibt}, G1188^{Sfmibt} and K1192^{Sfmibt}) and one residue (D1177^{Sfmibt}) that is located more N-terminal in $\alpha 4$. The corresponding residues in Ph and Scm are already part of the ML surface. As mentioned in the previous paragraph, there are two additional salt bridges present in the Scm-SAM/ Sfmibt-SAM interface (E833^{Scm} – K1186^{Sfmibt} and K849^{Scm} – K1177^{Sfmibt}), which are absent in the other three interfaces.

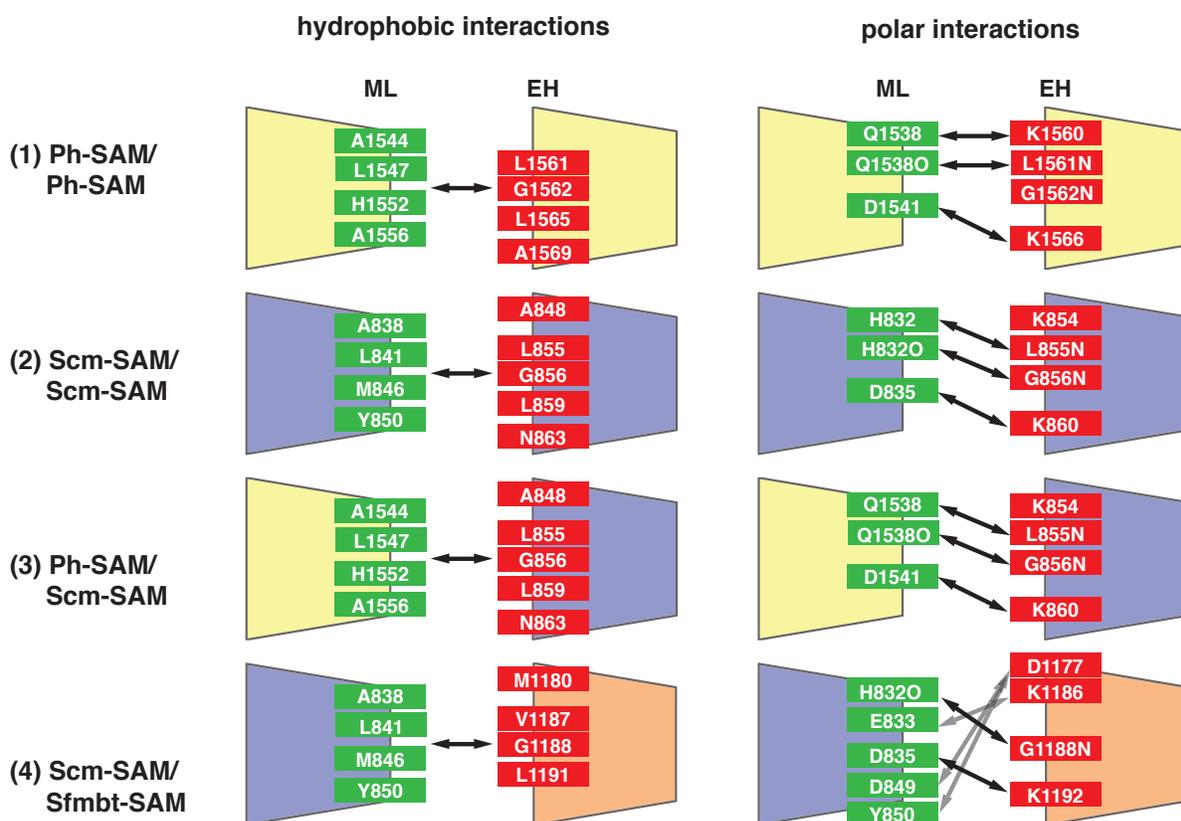


Figure 3.14: Conserved interactions between SAM/ SAM interfaces.

Schematic illustration of the most conserved hydrophobic and polar interactions present between different SAM domain complexes: (1) Ph-SAM/ Ph-SAM (yellow), (2) Scm-SAM/ Scm-SAM (blue), (3) Ph-SAM (yellow)/ Scm-SAM (blue) and (4) Scm-SAM (blue)/ Sfmibt-SAM (orange). The numbering of the residues corresponds to the full-length proteins. Arrows indicate an interaction observed in the crystal structure. No backbone atom designations indicate that the side chains mediate the interaction.

Figure adapted from Figure 3 C (Kim et al., 2005).

In summary, I solved the structure of the Scm-SAM/ Sfmbt-SAM complex and this structure revealed that the complex employs the same interfaces for complex formation as previously described hetero- and homo-SAM complexes, matching a model where SAM:SAM interactions would contribute to targeting PRC1 to PREs via the DNA binding PhoRC complex.

4. Discussion

In this study biochemical purification of PRC1 from *Drosophila* embryos revealed an interaction between PRC1 and the DNA binding PhoRC complex. This suggests a mechanism for how PRC1, which does not contain a specific DNA binding activity for Polycomb response elements (PREs), could be anchored at target genes. Subsequent *in vitro* reconstitution studies confirmed the interaction between subunits of the two complexes. The interaction is mediated by the PRC1 subunit Scm, which interacts with both, the core PRC1 subunit Ph and the PhoRC subunit Sfmt. Further mapping identified the SAM domains of the three proteins as important interaction modules and I determined the structure of a complex containing the SAM domain of Scm and Sfmt.

4.1 Ph interacts with PRC1 members and the PhoRC complex

PcG proteins exist in distinct multimeric complexes *in vivo*. These complexes co-bind at PREs in the genome, but how the different complexes assemble at these genomic sites is largely unknown.

Here, I used a tandem affinity purification (TAP) strategy to investigate the interactions of the PRC1 subunit Ph. This analysis identified the core PRC1 subunits Psc, Sce/ Ring, Pc and the substoichiometric PRC1 subunit Scm as major Ph-interacting proteins. In addition, subunits of the PhoRC complex were identified (Figure 3.4 and Table 3.1). While Sfmt was identified in all three purifications with a high score, Pho was only recovered in one of the three purifications, indicating that Sfmt mediates the Ph interaction. Interestingly, previous purifications of Ph did not identify Sfmt as an interacting partner (Shao et al., 1999, Wang and Brock, 2003); however, a recent study copurified the PhoRC subunits with PRC1 when using the Pc subunit as bait (Strübbe et al., 2011). A possible reason why the PhoRC complex was not discovered as a Ph interactor in previous purifications could be technical differences in the purification schemes used. While the two earlier studies (Shao et al., 1999; Wang and Brock, 2003) had multi-step purification procedures, Strübbe et al. and this study relied on a rapid

one- or two-step purification procedure. It is likely that this permitted the identification of weaker interactions due to almost physiological conditions.

Nevertheless, previous tandem affinity purifications conducted in our lab, using the Sfmbt protein as bait, identified neither Scm nor Ph as interacting partners (Alfieri et al., 2013). This could be due to the fact that the interactions with PRC1 were perturbed since Sfmbt was C-terminally tagged in these experiments (Alfieri et al., 2013), possibly interfering with SAM domain interactions to Scm that I uncovered here. In addition, none of the major interacting partners of Sfmbt that are forming the PhoRC-L complex including HDAC1/Rpd3, HP1b, NAP1 and Mga (Alfieri et al., 2013) were pulled down in this purification. First, it is possible that the interaction is indirect and could not be detected with the tandem mass spectrometry approach carried out in this study. In line with that, the second PhoRC core member Pho was only recovered in one out of three Ph purifications. Second, it has not been clarified if Sfmbt exists in one or in two different assemblies *in vivo*. One assembly being the PhoRC complex, which only contains the Pho and Sfmbt subunits and the second one being the PhoRC-L complex, which contains additional Sfmbt interactors including HDAC1/Rpd3 and HP1b. Therefore, it is possible that only the PhoRC complex and not the PhoRC-L complex interacts with PRC1.

Other identified Ph interactors include 11 proteins that were detected repeatedly in three independent Ph purifications namely Ge-1, Sop2, Nup107, Myo61F, Pyd, Edc3, Gnf1, Osp, Lamin, Imitation SWI and Hcf. According to their functions these 11 proteins were assigned to three groups: (1) proteins involved in RNA processing, (2) proteins involved in nuclear structure and (3) proteins involved in chromatin binding and remodeling. The first group includes Ge-1 and Edc3. The second group contains SOPs, Nup107, Myo61F and lamin and the third group comprises Iswi and HCF. Interestingly, the latter two proteins have also been shown to copurify with TAP-tagged Sfmbt (Vidal Matos, 2009), but did not show any association with TAP-tagged Calypso (Scheuermann et al., 2010) indicating that they could be specific interactors. The remaining two Ph interactors Gnf1 and Pyd could not be assigned to any of these three groups. It is not clear, if these additional interactors associate with PRC1 or PhoRC in a substoichiometric manner or if they are part of a distinct Ph containing entity that is less

abundant than PRC1. There are several arguments supporting the idea that Ph also exists in complexes other than PRC1. First, a study on the dRAF complex found that a significant fraction of Psc and Ph protein is not associated with either PRC1 or dRAF, suggesting that they are either present as free molecules or as part of other, still unknown complexes (Lagarou et al., 2008). Moreover, genetic interaction studies using different *ph* alleles suggest that the Ph protein may perform different functions in combination with different subsets of PcG proteins (Cheng et al., 1994). However, the Ph purifications reported here and elsewhere (Shao et al., 1999; Wang and Brock, 2003) did not provide any evidence that Ph would be present in other complexes besides PRC1.

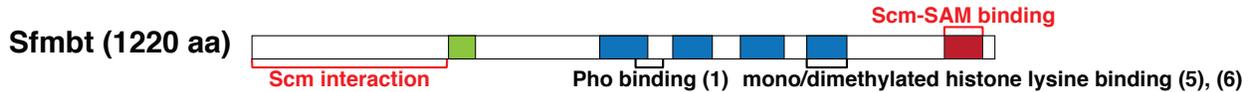
Surprisingly, Ph-d and Ph-p gave very similar molecular interaction profiles (Figure 3.3 B). On the one hand this was expected since they are very similar in sequence and can functionally almost completely substitute each other. On the other hand they have been suggested to act as alternatives in different tissues, developmental stages or even at different target genes in the same cell. However, the embryonic nuclear extract comprising 0-14 h old embryos could be too heterogeneous to identify subtle differences between proteins associated with Ph-p and Ph-d.

4.2 Interaction mapping of PRC1:PhoRC interaction

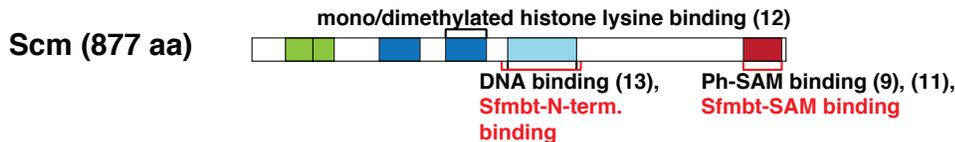
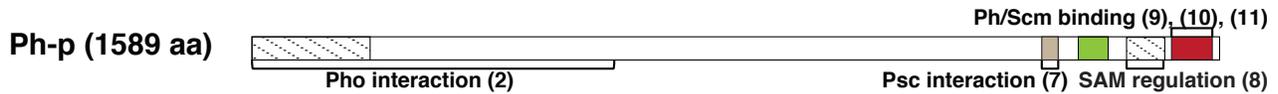
In vitro reconstitution of PRC1 and PhoRC interplay revealed that the interaction is mediated by PRC1 subunits Ph and Scm with the PhoRC subunit Sfm bt. While Ph and Sfm bt do not interact with each other directly, they form a stable trimeric complex with Scm. In agreement with that, Scm, a substoichiometric PRC1 member, has been linked to both, Ph and Sfm bt. Scm has been reported to directly interact with the N-terminal part of Sfm bt (Grimm et al., 2009) and in addition forms heteropolymers with Ph via its C-terminal SAM domain (Peterson et al., 1997). Since the interaction between Scm and Ph is well described and consolidated by a crystal structure of the copolymer of Ph- and Scm-SAM (Kim et al., 2005), I focused on defining the interaction surfaces of Scm and Sfm bt in depth. A detailed interaction mapping showed that Scm and Sfm bt have two interaction surfaces: one mediated by the N-terminal part of Sfm bt with a central portion

of Scm containing the SLED domain, and a second newly-discovered interaction mediated by their SAM domains (Figure 4.1).

PhoRC



PRC1



Functional domains



Figure 4.1: Novel identified interactions linking PRC1 and PhoRC complex.

Schematic drawing of PRC1 and PhoRC complexes that mediate interaction with previously known (black color) and novel identified interaction sites (in red color). References: (1) (Alfieri et al., 2013), (2) (Mohd-Sarip et al., 2002), (3) (Wang et al., 2004b), (4) (Brown et al., 1998) (5) (Grimm et al., 2009), (6) (Klymenko et al., 2006), (7) (Kyba and Brock, 1998a), (8) (Robinson et al., 2012), (9) (Peterson et al., 1997), (10) (Kim et al., 2002), (11) (Kim et al., 2005), (12) (Grimm et al., 2007), (13) (Bezsonova, 2014).

The first interaction site, which is mediated by the N-terminal part of Sfmbt to Scm, has been described previously (Grimm et al., 2009) and has been speculated to be mediated by the FCS zinc fingers present in the N-terminal portion of both proteins. However, pull-down experiments indicated that the FCS zinc fingers of Sfmbt and Scm can neither directly interact with each other nor are they necessary for Scm:Sfmbt interaction eliminating the FCS zinc fingers as putative interaction domains. In addition, pull down experiments with C-terminally truncated versions of Sfmbt and Scm showed that the N-terminal part of Sfmbt (Sfmbt₁₋₅₃₀) interacts with full length Scm, but not with C-terminally truncated Scm versions (Scm₁₋₄₃₅ or Scm₁₋₁₇₀) (Figure 3.8 C). This result is in conflict with previous results that implicated the N-terminal portion of Scm in Sfmbt

interaction (Grimm et al., 2009). Instead we propose two candidate regions, the first 100 amino acids of Sfm bt and the SLED domain of Scm, which is located between the MBT repeats and the SAM domain, as a putative interaction site of Scm and Sfm bt. However, further analysis is necessary to verify these two domains as interaction surfaces. This could include a GST-pull-down assay with the two proposed domains to test for direct interaction or an indirect interaction test with deletion constructs.

The second interaction site is mediated by the SAM domain of Sfm bt that binds directly to the SAM-domain of Scm. Scm-SAM and Sfm bt-SAM form a stable complex *in vitro* comparable to the Ph-SAM/ Scm-SAM complex. Interestingly, Sfm bt-SAM carries an amino acid substitution in one of the surfaces necessary for self-association and therefore cannot form homopolymers such as Ph-SAM or Scm-SAM. However, interaction studies with purified SAM domains showed that the end helix surface (EH) of Sfm bt-SAM is able to interact with the mid-loop surface (ML) of Scm-SAM. Structural analysis of the Scm-SAM/ Sfm bt-SAM complex showed that these two domains indeed interact via the same interface as other hetero- and homo-SAM complexes.

4.3 Role of SAM-containing PcG proteins in PRC1 targeting

Intriguingly, the three SAM domain containing PcG proteins Ph, Scm and Sfm bt can interact via their SAM domains and thus form a link between the PRC1 and PhoRC complex. The SAM domains of Ph and Scm are essential for the function of these proteins *in vivo* (Peterson et al., 2004; Robinson et al., 2012; Roseman et al., 2001). Rescue assays in fly embryos with different Ph constructs showed that a Ph construct lacking the SAM domain fails to rescue any features of the Ph null mutant phenotype demonstrating its importance for Ph function (Gambetta, unpublished). In particular, Ph-SAM polymerization is required for the gene silencing function of Ph in flies (Robinson et al., 2012). Similarly, Scm proteins lacking the SAM domain or carrying point mutations in the SAM domain that disrupt polymerization have been reported to be incapable of rescuing the lethality of Scm null mutant embryos (Peterson et al., 2004). Furthermore, the SAM domain is crucial for long-range repression mediated by Scm (Roseman et al., 2001). SAM domains in these proteins thus play essential roles in PcG silencing.

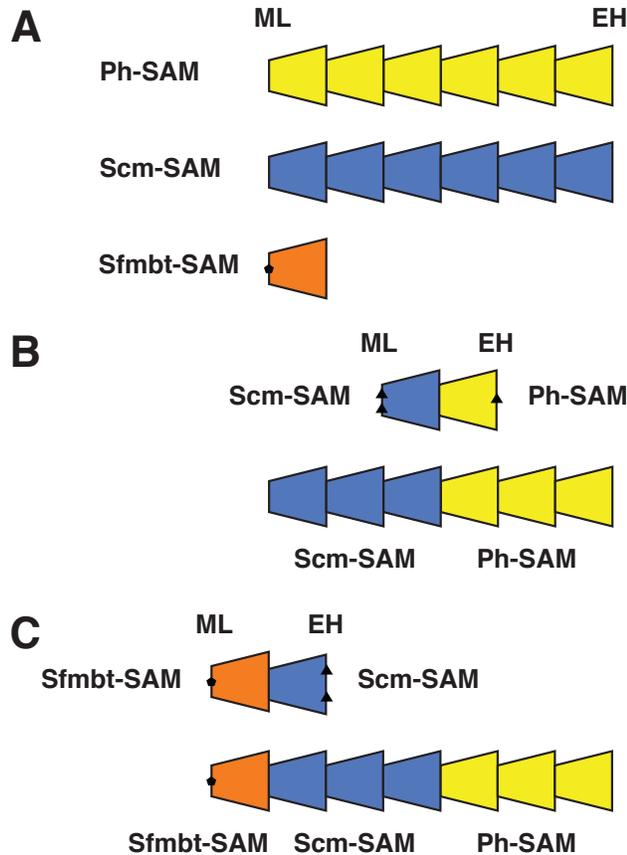


Figure 4.2: Regulation of SAM domain polymerization: A model for Ph-, Scm- and Sfmt-SAM domain interaction.

A: Model of SAM domain self-association of the three PcG proteins Ph (yellow), Scm (blue) and Sfmt (orange). The two surfaces important for interaction are mid-loop (ML) surface pointing towards the left side and end-helix (EH) surface pointing towards the right side. Note that while Ph- and Scm-SAM domains form helical, head-to-tail homo-polymeric structures, Sfmt-SAM is monomeric, because it has a non-functional ML surface.

B: Model of copolymer formation between Ph- and Scm-SAM as proposed by (Kim et al., 2005). The upper panel shows the preferred interface between Ph- and Scm-SAM that was used to obtain the co-crystal structure. Note that Scm-SAM ML and Ph-SAM EH were mutated in order to prevent self-association. The lower panel shows the model for a Ph-/ Scm-SAM copolymer consisting of a single junction between a Ph-SAM ML surface and an Scm-SAM EH surface which is extended by Ph-SAM on the right side and Scm-SAM on the left side.

C: Model of copolymer formation between Ph-, Scm- and Sfmt-SAM. The upper panel exhibits the interface between Scm-SAM and Sfmt-SAM that was used to obtain the co-crystal structure. Note that Scm-SAM EH was mutated to prevent self-association, whereas Sfmt-SAM is monomeric due to its non-functional ML surface.

The lower panel shows the copolymer model between Ph-SAM and Scm-SAM from B extended by the Sfbmt-SAM domain. Association of the Sfbmt-SAM domain to the ML surface of Scm would prevent further extension of the Scm-SAM polymer.

Ph-SAM and Scm-SAM can both form helical, head-to-tail homopolymeric structures (Kim et al., 2002, 2005). In addition, Ph-SAM and Scm-SAM have been suggested to copolymerize in form of a single junction copolymer, since they both preferentially polymerize in one single orientation (Figure 4.2 A). In contrast to that, Sfbmt-SAM is unable to polymerize due to several amino acid exchanges in the ML surface (Figure 3.12 A). However, we could show that Sfbmt-SAM is able to interact with Scm via its EH surface and in this way could contribute to regulation of Scm-SAM polymerization. Since the second interaction surface of Sfbmt-SAM is non-functional, the Sfbmt-SAM domain could cap Scm-SAM polymers and thus prevent excessive Scm-SAM polymerization. In line with that, Sfbmt associates with the Mid-loop surface of Scm, which is the interaction surface preferentially used for Scm polymerization in the Ph-SAM/ Scm-SAM copolymer (Figure 4.2 C).

Interestingly, Yan and 'Modulator of the activity of ETS' (Mae) are two SAM domain containing *Drosophila* transcriptional repressors that also interact via their SAM domains (Qiao et al., 2004). Yan is a target of the receptor tyrosine kinase pathway and belongs to the ETS family of transcriptional repressors. Yan-SAM polymerization is critical for transcriptional repression and Mae has been shown to regulate Yan polymerization by depolymerizing Yan-SAM. Similar to Sfbmt-SAM, Mae-SAM only possesses one functional interaction surface in its SAM domain (in this case a ML surface) that can only bind to one end of the Yan-SAM polymer and thus blocks further extension. Remarkably, Mae-SAM has a much stronger binding affinity to Yan-SAM than Yan-SAM to itself (Qiao et al., 2004). There are some interesting parallels in how these polymeric transcriptional repressors are regulated. Since Ph- and Scm-SAM have similar binding affinities with K_d values in the range of 50-200 nm, competition and depolymerization between these two SAM domains is not likely (Kim et al., 2005); however, it would be interesting to determine the binding affinity of Sfbmt-SAM to Scm-SAM. Preliminary results indicate that the Scm-SAM self-association is stronger than –

Scm-SAM:Sfmbt-SAM interaction, but this should be confirmed with isothermal titration calorimetry (ITC).

How could SAM domain polymerization contribute to PcG repression? Clearly SAM domains play a crucial role in PcG repression. Since Ph-SAM and Scm-SAM form polymers it is tempting to speculate that they contribute to spreading of PcG complexes from the PREs into flanking regions or are needed for the interaction between complexes bound at distinct sites along the chromatin fiber thus playing a role in long-range interaction and repression. However, Ph chromatin immunoprecipitation (ChIP) profiling suggests that Ph is sharply located at PRE elements and is not trailing into flanking regions such as the H3-K27me3-binding Pc protein does. Scm has been shown to colocalize with Sfmbt at the PREs of several PcG target genes (Grimm et al., 2009), but there is no genome wide binding profile available for Scm. It would be interesting to determine to which degree Scm profiles overlap with the profiles of other PcG proteins, particularly with Ph and Sfmbt, and if it is possible to detect spreading of Scm into PRE flanking regions by ChIP analysis.

Another possible mechanism of how SAM domains could contribute to PcG repression could be through chromatin compaction. In this scenario SAM:SAM interaction would lead to Ph and Scm polymerization, which would cluster at PREs. Chromatin interaction domains present in these proteins including MBT repeats that bind methylation marks, SLED domains that can bind to DNA and FCS Zinc finger domains, which potentially bind to chromatin, could help to compact the chromatin in the PRE surrounding areas to achieve a repressive chromatin state of target genes.

4.3 Direct PRC1 recruitment via PhoRC complex

My data, together with recent publications in the field (Tavares et al., 2012), suggest that the hierarchical recruitment model (Wang et al., 2004b), which has dominated the Polycomb field for many years, is outdated. This model claims the sequential recruitment of PcG complexes to PREs, where first the PhoRC complex is bound to PREs by direct protein-DNA interactions, then PRC2 is recruited by direct interaction to the PhoRC complex and deposits its H3-K27me3 mark, which is subsequently bound by

PRC1. In this hierarchical model binding of the H3-K27me3 mark by the chromo domain of Pc is proposed to be the major mechanism of PRC1 recruitment to PREs.

Interestingly, we find that PRC1 can also directly interact with one of the PhoRC complex members, namely Sfmbt. The interaction is due to two binding surfaces between Sfmbt and Scm, one of them mediated via the respective SAM domains, which is supported by structural studies. The direct binding between the PRC1 and the PhoRC complexes mediated by SAM domain containing proteins Ph, Scm and Sfmbt that are polymerization-competent could be a way to increase the residence time of both complexes on chromatin. Next to the specific DNA binding affinity of Pho, the weak DNA binding affinity of Scm via the SLED domain and the H3-K27me3 binding mediated by the chromo domain of Pc could contribute to stabilize the PcG complexes at PREs. Consistent with that a recent study found that in cells lacking the PRC1 core subunits Psc and its homolog Su(z)2 not only PRC1 recruitment to PREs is disrupted, but also PhoRC subunits bind with reduced levels to PREs suggesting a cross-talk between these two complexes (Khan et al., 2014). In addition, feedback loops mediated by deposited histone marks that are in turn bound by other PcG proteins, reinforce PcG complex binding at PREs (Kalb et al., 2014).

Therefore a possible model could be that PcG complex assembly is initiated by PhoRC binding to specific regions at PREs, but this assembly is subsequently stabilized by combinatorial interactions between the different PcG complexes as well as between the complexes and chromatin flanking the PREs. In this scenario the stable assembly of different PcG complexes at PREs is provided by a multitude of relatively weak protein-protein, protein-DNA as well as protein-chromatin interactions.

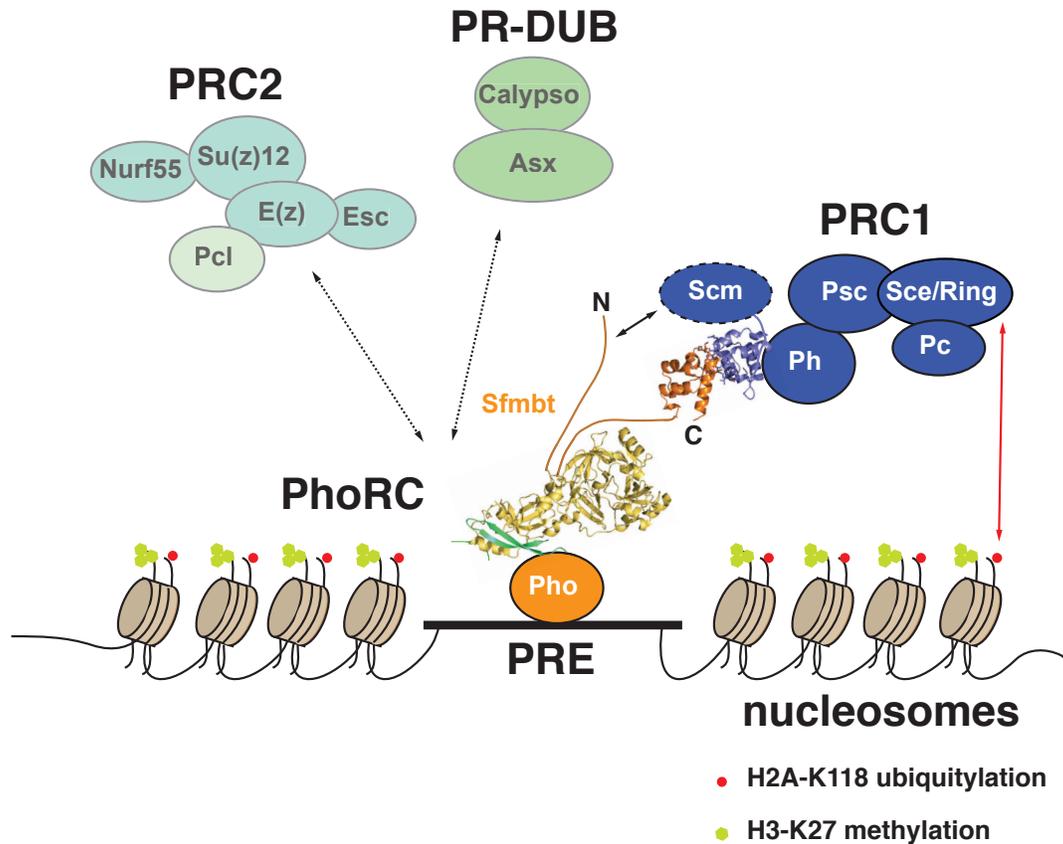


Figure 4.3: PRC1 complex recruitment to PREs via the PhoRC complex.

Interactions between the PRC1 members Ph and Scm with the PhoRC member Sfm link PRC1 to PREs. Pho binds with its DNA-binding domain to Pho-binding sites within PREs. Pho binds to the four MBT repeats of Sfm via a spacer region (PDB ID: 4C5E) (Alfieri et al., 2013). Sfm on the other hand has two interaction surfaces for the PRC1 member Scm. One is mediated via the N-terminal portion of Sfm and the second one is mediated via the SAM domain of Sfm (orange) to the SAM domain of Scm (blue) (this study). Scm-SAM can also interact with the SAM domain of Ph (PDB ID: 1PK1) (Kim et al., 2005).

Our structural and biochemical data link PRC1 directly to the PRE-binding PhoRC complex via the three proteins: Ph, Scm and Sfm. In line with our model, chromatin immunoprecipitation (ChIP) data from transformed embryos carrying either a wild-type fragment of the *bx*d PRE (PRE_D) or a mutated *bx*d PRE fragment (PRE_D *pho mut*) with point mutations in all six Pho protein-binding sites showed that absence of Pho binding sites resulted in severely reduced binding levels of Pho, Sfm and Ph. Interestingly, more recent ChIP assays in imaginal discs with the same reporter lines revealed that

not only the PhoRC members are strongly reduced at the $PRE_{D\ pho\ mut.}$, but also PRC1 members including Scm and Ph as well as PRC2 members (Sheahan, unpublished data). This indicates that direct interactions of PRC1 and PRC2 with PhoRC are essential for their recruitment. Contrary to our findings, Scm has been reported to be recruited to the *bxd* PRE independently from Pho in S2 cells (Wang et al., 2010). However, this study used an RNAi-knockdown strategy to reduce Pho protein levels. The knockdown might not have been efficient in reducing Pho levels, because they could still detect Scm and reduced levels of Pc and Su(z)12 at the *bxd* PRE after knockdown of Pho. Moreover, they did not account for Pho-like (Phol) levels in this study. ChIP assays and genetic experiments indicate that Phol can largely compensate for loss of Pho (Brown et al., 2003, Khan et al., 2014) and therefore Phol might have played a role in recruiting Scm in the absence of Pho.

Another interesting point is the strong genetic interaction between Scm and Sfmbt. Sfmbt-Scm^{D215N} double-mutant cell clones display a tumor-like phenotype with unrestricted cell proliferation (Grimm et al., 2009) reminiscent of cell clones lacking the PRC1 members Ph or Psc-Su(z)2 (Oktaba et al., 2008). This phenotype is not observed in either of the single mutants and appears to be specific to Scm and Sfmbt. Therefore, it is tempting to speculate that this severe phenotype is caused by inability to recruit the PRC1 members Ph and Psc-Su(z)2 in case both, Scm and Sfmbt, are not present. Nevertheless, it is possible that the recruitment mechanism via Scm and Sfmbt is not the only way that Ph and Psc are recruited to target genes, since they have been shown to be important for repression of a subset of target genes that do not require other PRC1 members and their mutant phenotypes are much more severe than mutants lacking the other PRC1 or PhoRC members (Gutiérrez et al., 2012; Oktaba et al., 2008).

4.4 Conservation of PRC1 recruitment mechanism?

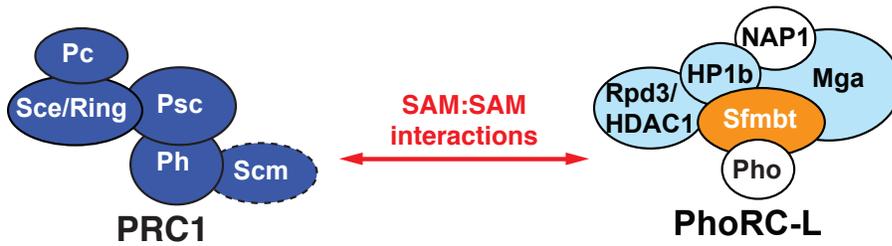
How conserved is the novel PRC1 recruitment mechanism described in this study between flies and vertebrates? At a first glance it seems unlikely that this mechanism is important for PRC1 recruitment in mammals, because the Pho homolog YY1 does not play a major role in targeting PcG complexes. Genome-wide mapping in mouse ES cells has demonstrated that YY1 binding sites do not overlap with other PcG binding

sites, instead YY1 is associated primarily with active genes and exerts PcG-independent functions in ES cells (Mendenhall et al., 2010; Vella et al., 2012).

Nevertheless, SAM mediated gene repression is conserved in mammals. The polymerization capacity of the Ph-SAM domain is important for clustering of PRC1 complexes into microscopically visible foci, so-called Polycomb bodies. Disrupting polymerization by introducing point mutations in the interaction surface leads to Polycomb body disassembly, reduced binding of PRC1 and PRC2 to target genes, chromatin decompaction and moderate target gene derepression, even though mutant Ph still assembles into PRC1 (Isono et al., 2013). In addition, a study on human PHC3 SAM showed that this SAM domain utilizes the same interaction surfaces as fly Ph, but forms larger polymers (Nanyes et al., 2014). Thus, polymerization of Ph-SAM was suggested to be important to capture and/ or retain PRC1 at target genes to stabilize transcriptional repression.

Interestingly, mammals possess the PRC1.6/ E2F.6 complex, which resembles the *Drosophila* PhoRC-L complex containing Pho and Sfmbt (Alfieri et al., 2013; Gao et al., 2012). Is it possible that this mammalian complex is involved in PRC1 recruitment via SAM domain interactions? The PRC1.6/ E2F.6 complex does not contain Pho, but instead includes E2F.6 as a DNA binding factor. In addition the PRC1.6/ E2F6 complex contains L3MBTL2, which is an Sfmbt homolog. L3MBTL2 has been reported to be a close homolog of *Drosophila* Sfmbt, but lacks the C-terminal SAM domain. This thus argues against a scenario where SAM:SAM interaction would provide the means for PRC1 recruitment (Bonasio et al., 2010) (Figure 4.4).

Flies



Humans

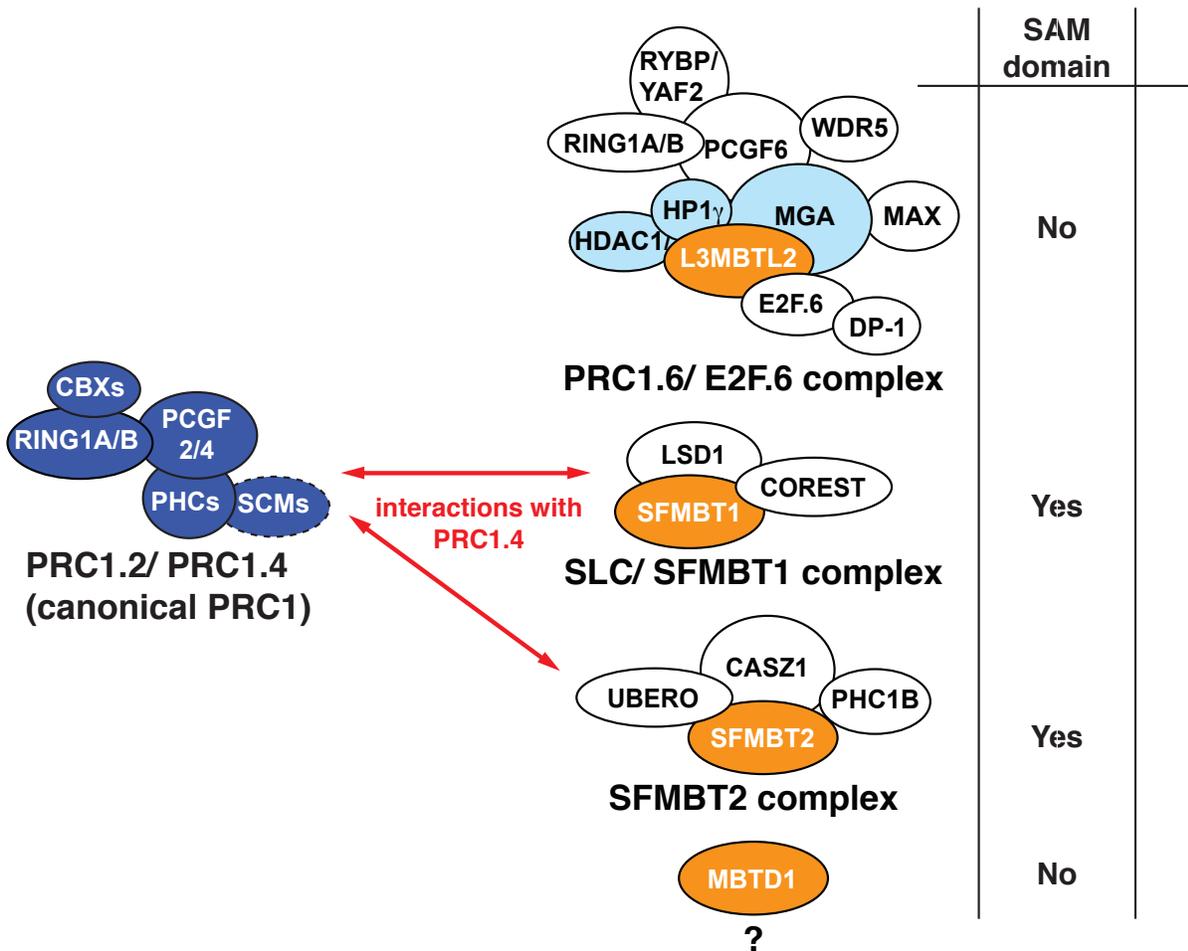


Figure 4.4: A possibly conserved role of SAM domains in PRC1 recruitment.

Schematic illustration of PRC1 and the Sfmbt containing complex PhoRC-L in flies and their mammalian counterparts. Drawing adapted from (Alfieri et al., 2013). The fly PhoRC-L complex was described in (Alfieri et al., 2013), human PRC1.2, PRC1.4 and PRC1.6 in (Gao et al., 2012; Ogawa et al., 2002), the SFMBT1 and 2 complexes in (Zhang et al., 2013). The interactions partners of MBTD1 and its genome-wide

binding profile are unknown. Sfmbt and its human homologs L3MBTL2, SFMBT1, SFMBT2 and MBTD1 proteins are depicted in orange, orthologous subunits identified in both *Drosophila* and human assemblies are labeled in light blue, and other associated proteins are shown in white. PRC1 complexes are colored blue. In this study I identified a SAM:SAM mediated interaction between the fly PRC1 and PhoRC complex (red arrow). Note that the SAM domain containing Sfmbt homologs SFMBT1 and 2 have been reported to interact with PRC1.4 (Zhang et al., 2013). This interaction is possibly mediated via SAM domains, suggesting an evolutionary conserved role for SAM domains in PcG targeting and repression.

A closer look at the human homologs of fly Scm and Sfmbt showed that all three Scm homologs – SCM1, SCML1 and SCML2 – as well as two of the four known Sfmbt homologs – SFMBT1 and SFMBT2 – contain SAM domains with conserved interaction surfaces. The three Scm homologs have been found to associate with the canonical mammalian PRC1 complexes – PRC1.2 and PRC1.4 – that also contain CBX (Pc homologs), RING1B and PHC (Ph homologs) proteins and resemble the fly PRC1 complex (Gao et al., 2012). SFMBT1 and 2 have been reported to be part of two functionally distinct complexes. While SFMBT1 is part of the SLC (SFMBT1/ LSD1/ CoREST) complex, which contains components of the LSD1/ CoREST histone demethylase complex, SFMBT2 is associated with proteins such as castor zinc finger 1 (CASZ1), ubiquitin-conjugated enzyme E2O (UBE2O) and the PHC1B (Zhang et al., 2013). Intriguingly, both SFMBT1 and 2 were found to associate with the PRC1.4 complex, the closest homolog of *Drosophila* PRC1 (Figure 4.4). Moreover, SFMBT1 was found to be able to recruit PRC1.4 to gene targets in a tethering assay (Zhang et al., 2013). Therefore, it is possible that one mechanism of canonical PRC1 recruitment to target genes could be mediated by SAM:SAM interactions between Sfmbt, Scm and Ph. However, this link has to be further confirmed by genome-wide colocalization studies in ChIP assays and interaction studies between the SAM domains of the human Ph, Scm and Sfmbt homologs. Other mechanisms include H3-K27me3-recognition by CBX chromo domain (Gao et al., 2012), recruitment to unmethylated CpG and association with specific DNA-binding factors islands (reviewed in (Di Croce and Helin, 2013; Klose et al., 2013; Simon and Kingston, 2013).

Remarkably, even if the mechanism of PRC1 tethering to PcG target genes via SAM domains is conserved in mammals, it seems to have evolved to a more specialized recruitment option that is applied to only a certain set of targets, which are controlled by either the SLC complex or the SFMBT2 containing complex. These targets would include genes involved in nucleosome and chromatin assembly functions and would control certain processes such as the dynamic regulation of histone loci (Zhang et al., 2013).

Conclusion

In this study, I identified two interaction surfaces between the PRC1 subunit Scm and the PhoRC subunit Sfmbt, one mediated by the N-terminal portion of Sfmbt, presumably with the SLED domain of Scm, and the other one mediated by the SAM domains of both proteins. For the later interaction the work here reports the structural basis. Obtaining structural insights into PcG assembly at target genes is crucial for understanding how these key developmental regulators maintain the repressive chromatin state of target genes throughout cell divisions to maintain determined cell fates. Here, I present the first atomic-level information on how different PcG complexes interact with each other. These interactions shed light on how the PRC1 complex is targeted to PREs in order to execute its chromatin modifying and compaction activities at nucleosomes of target genes. In addition to the recognition of the H3-K27me3 mark by the chromo domain of Pc, SAM domain polymerization involving Ph-SAM, Scm-SAM and Sfmbt-SAM are intriguing mechanism how the PRC1 complex could spread from PREs to flanking nucleosomes or contact other complexes assembled at distinct chromosomal loci to ultimately change the chromatin structure of target genes that have to be repressed.

5. References

- Akam, M. (1987). The molecular basis for metamereric pattern in the *Drosophila* embryo. *Development* *101*, 1–22.
- Akasaka, T., Tsuji, K., Kawahira, H., Kanno, M., Harigaya, K., Hu, L., Ebihara, Y., Nakahata, T., Tetsu, O., Taniguchi, M., et al. (1997). The role of mel-18, a mammalian Polycomb group gene, during IL-7-dependent proliferation of lymphocyte precursors. *Immunity* *7*, 135–146.
- Alfieri, C., Gambetta, M.C., Matos, R., Glatt, S., Sehr, P., Fraterman, S., Wilm, M., Muller, J., and Muller, C.W. (2013). Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. *Genes Dev.* *27*, 2367–2379.
- Americo, J., Whiteley, M., Brown, J.L., Fujioka, M., Jaynes, J.B., and Kassis, J.A. (2002). A complex array of DNA-binding proteins required for pairing-sensitive silencing by a polycomb group response element from the *Drosophila* engrailed gene. *Genetics* *160*, 1561–1571.
- Andrulis, E.D., Guzmán, E., Döring, P., Werner, J., and Lis, J.T. (2000). High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev.* *14*, 2635–2649.
- Antonyamy, S., Condon, B., Druzina, Z., Bonanno, J.B., Gheyi, T., Zhang, F., MacEwan, I., Zhang, A., Ashok, S., Rodgers, L., et al. (2013). Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. *PLoS One* *8*, e84147.
- Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). *Drosophila* Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J.* *30*, 2817–2828.
- Arnold, P., Schöler, A., Pachkov, M., Balwiercz, P.J., Jørgensen, H., Stadler, M.B., van Nimwegen, E., and Schübeler, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res.* *23*, 60–73.
- Aviv, T., Lin, Z., Lau, S., Rendl, L.M., Sicheri, F., and Smibert, C.A. (2003). The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nat. Struct. Biol.* *10*, 614–621.
- De Ayala Alonso, A.G., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007a). A Genetic Screen Identifies Novel Polycomb Group Genes in *Drosophila*. *Genetics* *176*, 2099–2108.
- De Ayala Alonso, A.G., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007b). A Genetic Screen Identifies Novel Polycomb Group Genes in *Drosophila*. *Genetics* *176*, 2099–2108.
- Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* *16*, 3186–3198.
- Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* *19*, 1257–1265.
- Bannister, A.J., Schneider, R., and Kouzarides, T. (2002). Histone Methylation. *Cell* *109*, 801–806.
- Bantignies, F., Goodman, R.H., and Smolik, S.M. (2000). Functional interaction between the coactivator *Drosophila* CREB-binding protein and ASH1, a member of the trithorax group of chromatin modifiers. *Mol. Cell. Biol.* *20*, 9317–9330.
- Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Müller, M., Shanower, G., Schedl, P., Gyurkovics, H., and Karch, F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex *iab-7* domain and insulates *iab-7* from initiation elements and a PRE in the adjacent *iab-8* domain. *Development* *127*, 779–790.

- Beisel, C., Bunes, A., Roustan-Espinosa, I.M., Koch, B., Schmitt, S., Haas, S.A., Hild, M., Katsuyama, T., and Paro, R. (2007). Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 16615–16620.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* *125*, 315–326.
- Beuchle, D., Struhl, G., and Müller, J. (2001). Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* *128*, 993–1004.
- Bezsonova, I. (2014). Solution NMR Structure of the DNA-binding Domain from Scml2 (Sex Comb on Midleg-like 2). *J. Biol. Chem* *289*, 15739–15749.
- Bezsonova, I., Walker, J.R., Bacik, J.P., Duan, S., Dhe-Paganon, S., and Arrowsmith, C.H. (2009). Ring1B contains a ubiquitin-like docking module for interaction with Cbx proteins. *Biochemistry* *48*, 10542–10548.
- Bierie, B., and Moses, H.L. (2006). TGF-beta and cancer. *Cytokine Growth Factor Rev.* *17*, 29–40.
- Blackledge, N.P., and Klose, R. (2011). CpG island chromatin: a platform for gene regulation. *Epigenetics* *6*, 147–152.
- Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* *38*, 179–190.
- Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* *157*, 1–15.
- Blastyák, A., Mishra, R.K., Karch, F., and Gyurkovics, H. (2006). Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol. Cell. Biol.* *26*, 1434–1444.
- Boyer, S., Cavalli, G., Brock, H.W., and Dura, J.-M. (2003). Identification and characterization of polyhomeotic PREs and TREs. *Dev. Biol.* *261*, 426–442.
- Bonasio, R., Lecona, E., and Reinberg, D. (2010). MBT domain proteins in development and disease. *Semin. Cell Dev. Biol.* *21*, 221–230.
- Bornemann, D., Miller, E., and Simon, J. (1998). Expression and Properties of Wild-Type and Mutant Forms of the *Drosophila* Sex Comb on Midleg (SCM) Repressor Protein. *Genetics* *122*, 1621–1630.
- Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). Opinion: The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* *5*, 158–163.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* *441*, 349–353.
- Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* *9*, 773–784.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development* *20*, 1123–1136.
- Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* *19*, 1273–1281.
- Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L., and 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.

- Brown, J.L., Fritsch, C., Mueller, J., and 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*, 285–294.
- Brown, J.L., Grau, D.J., DeVido, S.K., and 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*, 5181–5189.
- Bruggeman, S.W.M., Hulsman, D., Tanger, E., Buckle, T., Blom, M., Zevenhoven, J., van Tellingen, O., and van Lohuizen, M. (2007). *Bmi1* Controls Tumor Development in an *Ink4a/Arf*-Independent Manner in a Mouse Model for Glioma. *Cancer Cell* *12*, 328–341.
- Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins *Bmi1* and *Ring1b*. *EMBO J.* *25*, 2465–2474.
- Busturia, A., and Morata, G. (1988). Ectopic expression of homeotic genes caused by the elimination of the Polycomb gene in *Drosophila* imaginal epidermis. *Development* *104*, 713–720.
- Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the *Drosophila* *Abd-B* gene requires both *Pleiohomeotic* and GAGA factor for the maintenance of repression. *Development* *128*, 2163–2173.
- Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell* *49*, 571–582.
- Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* *15*, 57–67.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* *298*, 1039–1043.
- Capdevila, M.P., and García-Bellido, A. (1981). Genes Involved in the Activation of the Bithorax Complex of *Drosophila*. *Wilhelm Roux Arch. Dev. Biol.* *190*, 339–350.
- Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb *Ezh2* methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes & Development* *18*, 2627–2638.
- Celniker, S.E., Keelan, D.J., and Lewis, E.B. (1989). The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the Abdominal-B domain. *Genes & Development* *3*, 1424–1436.
- Chalkley, G.E., Moshkin, Y.M., Langenberg, K., Bezstarosti, K., Blastyák, A., Gyurkovics, H., Demmers, J.A.A., and Verrijzer, C.P. (2008). The transcriptional coactivator SAYP is a trithorax group signature subunit of the PBAP chromatin remodeling complex. *Mol. Cell. Biol.* *28*, 2920–2929.
- Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency. *STEM CELLS* *26*, 1496–1505.
- Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* *13*, 2553–2564.
- Chen, X., Hiller, M., Sancak, Y., and Fuller, M.T. (2005). Tissue-Specific TAFs Counteract Polycomb to Turn on Terminal Differentiation. *Science* *310*, 869–872.
- Chen, X., and Xu, L. (2010). Specific Nucleoporin Requirement for Smad Nuclear Translocation. *Mol. Cell. Biol.* *30*, 4022–4034.
- Cheng, N.N., Sinclair, D.A., Campbell, R.B., and Brock, H.W. (1994). Interactions of polyhomeotic with Polycomb group genes of *Drosophila melanogaster*. *Genetics* *138*, 1151–1162.

- Chiang, A., O'Connor, M.B., Paro, R., Simon, J., and BENDER, W. (1995). Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development* *121*, 1681–1689.
- Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., et al. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J. Biol. Chem.* *282*, 20395–20406.
- Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. *eLife* *1*, e00005.
- Classen, A.-K., Bunker, B.D., Harvey, K.F., Vaccari, T., and Bilder, D. (2009). A tumor suppressor activity of *Drosophila* Polycomb genes mediated by JAK-STAT signaling. *Nat. Genet.* *41*, 1150–1155.
- Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *CellReports* *7*, 1456–1470.
- Copur, O., and Muller, J. (2013). The histone H3-K27 demethylase Utx regulates HOX gene expression in *Drosophila* in a temporally restricted manner. *Development* *140*, 3478–3485.
- Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* *20*, 1147–1155.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* *111*, 185–196.
- Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* *328*, 1161–1164.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development* *25*, 1010–1022.
- DeCamillis, M., Cheng, N.S., Pierre, D., and Brock, H.W. (1992). The polyhomeotic gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. *Genes & Development* *6*, 223–232.
- Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* *434*, 533–538.
- Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., and Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet.* *8*, e1002494.
- Dingwall, A.K., Beek, S.J., McCallum, C.M., Tamkun, J.W., Kalpana, G.V., Goff, S.P., and Scott, M.P. (1995). The *Drosophila* snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. *Mol. Biol. Cell* *6*, 777–791.
- Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of *Drosophila melanogaster*. *Genetics* *102*, 49–70.
- Dura, J.M., and Ingham, P. (1988). Tissue- and stage-specific control of homeotic and segmentation gene expression in *Drosophila* embryos by the polyhomeotic gene. *Development* *103*, 733–741.
- Dura, J.M., Randsholt, N.B., Deatrck, J., Erk, I., Santamaria, P., Freeman, J.D., Freeman, S.J., Weddell, D., and Brock, H.W. (1987). A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in *D. melanogaster*. *Cell* *51*, 829–839.
- Eberharter, A. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J.* *20*, 3781–3788.
- Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y.-I., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* *135*, 1513–1524.

- Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* 38, 452–464.
- Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.-F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes & Development* 21, 2558–2570.
- Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2008). Deadenylation is a widespread effect of miRNA regulation. *RNA* 15, 21–32.
- Fan, S.-J., Marchand, V., and Ephrussi, A. (2011). *Drosophila* Ge-1 Promotes P Body Formation and oskar mRNA Localization. *PLoS One* 6, e20612.
- Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* 1, e00205.
- Fauvarque, M.O., Zuber, V., and Dura, J.M. (1995). Regulation of polyhomeotic transcription may involve local changes in chromatin activity in *Drosophila*. *Mech. Dev.* 52, 343–355.
- Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7, 21–33.
- Feng, S., Huang, J., and Wang, J. (2011). Loss of the Polycomb group gene polyhomeotic induces non-autonomous cell overproliferation. *EMBO Rep.* 12, 157–163.
- Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development* 17, 1870–1881.
- FlyBase Curators, Swiss-Prot Project Members, and InterPro Project Members (2004). Gene Ontology annotation in FlyBase through association of InterPro records with GO terms.
- Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a functional core polycomb repressive complex. *Mol. Cell* 8, 545–556.
- Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574–1577.
- Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.W., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* 11, 2941–2950.
- Friberg, A., Oddone, A., Klymenko, T., Müller, J., and Sattler, M. (2010). Structure of an atypical Tudor domain in the *Drosophila* Polycomblike protein. *Protein Sci.* 19, 1906–1916.
- Fritsch, C., Beuchle, D., and Müller, J. (2003). Molecular and genetic analysis of the Polycomb group gene *Sex combs extra/Ring* in *Drosophila*. *Mech. Dev.* 120, 949–954.
- Gambetta, M.C., Oktaba, K., and Müller, J. (2009). Essential role of the glycosyltransferase *sxc/Ogt* in polycomb repression. *Science* 325, 93–96.
- Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* 32, 63–70.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* 45, 344–356.
- Gargiulo, G., Cesaroni, M., Serresi, M., de Vries, N., Hulsman, D., Bruggeman, S.W., Lancini, C., and van Lohuizen, M. (2013). In Vivo RNAi Screen for BMI1 Targets Identifies TGF- β /BMP-ER Stress Pathways as Key Regulators of Neural- and Malignant Glioma-Stem Cell Homeostasis. *Cancer Cell* 23, 660–676.

- Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb Group and SCF Ubiquitin Ligases Are Found in a Novel BCOR Complex That Is Recruited to BCL6 Targets. *Mol. Cell. Biol.* **26**, 6880–6889.
- Gindhart, J.G., and Kaufman, T.C. (1995). Identification of Polycomb and trithorax group responsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* **139**, 797–814.
- Goldberg, M.L., Colvin, R.A., and Mellin, A.F. (1989). The *Drosophila* *zeste* locus is nonessential. *Genetics* **123**, 145–155.
- González, I., Simón, R., and Busturia, A. (2009). The Polyhomeotic protein induces hyperplastic tissue overgrowth through the activation of the JAK/STAT pathway. *Cell Cycle* **8**, 4103–4111.
- Gorfinkiel, N., Fanti, L., Melgar, T., García, E., Pimpinelli, S., Guerrero, I., and Vidal, M. (2004). The *Drosophila* Polycomb group gene *Sex combs extra* encodes the ortholog of mammalian Ring1 proteins. *Mech. Dev.* **121**, 449–462.
- Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & Development* **25**, 2210–2221.
- Green, J.B., Gardner, C.D., Wharton, R.P., and Aggarwal, A.K. (2003). RNA recognition via the SAM domain of Smaug. *Mol. Cell* **11**, 1537–1548.
- Grimm, C., de Ayala Alonso, A.G., Rybin, V., Steuerwald, U., Ly-Hartig, N., Fischle, W., Müller, J., and Müller, C.W. (2007). Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein *Sex comb on midleg*. *EMBO Rep.* **8**, 1031–1037.
- Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Müller, J., and Müller, C.W. (2009). Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.* **28**, 1965–1977.
- Gruenbaum, Y., Landesman, Y., Drees, B., Bare, J.W., Saumweber, H., Paddy, M.R., Sedat, J.W., Smith, D.E., Benton, B.M., and Fisher, P.A. (1988). *Drosophila* nuclear lamin precursor Dm0 is translated from either of two developmentally regulated mRNA species apparently encoded by a single gene. *J. Cell Biol.* **106**, 585–596.
- Guelman, S., Suganuma, T., Florens, L., Swanson, S.K., Kiesecker, C.L., Kusch, T., Anderson, S., Yates, J.R., Washburn, M.P., Abmayr, S.M., et al. (2006). Host Cell Factor and an Uncharacterized SANT Domain Protein Are Stable Components of ATAC, a Novel dAda2A/dGcn5-Containing Histone Acetyltransferase Complex in *Drosophila*. *Mol. Cell. Biol.* **26**, 871–882.
- Gunster, M.J., Raaphorst, F.M., Hamer, K.M., den Blaauwen, J.L., Fieret, E., Meijer, C.J., and Otte, A.P. (2001). Differential expression of human Polycomb group proteins in various tissues and cell types. *J. Cell. Biochem.* **81**, 129–143.
- Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Müller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139**, 117–127.
- Han, Z., Xing, X., Hu, M., Zhang, Y., Liu, P., and Chai, J. (2007). Structural Basis of EZH2 Recognition by EED. *Structure* **15**, 1306–1315.
- Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300.
- Haupt, Y., Alexander, W.S., Barri, G., Klinken, S.P., and Adams, J.M. (1991). Novel zinc finger gene implicated as *myc* collaborator by retrovirally accelerated lymphomagenesis in E mu-*myc* transgenic mice. *Cell* **65**, 753–763.
- Heard, E. (2004). Recent advances in X-chromosome inactivation. *Curr. Opin. Cell Biol.* **16**, 247–255.

- Hegan, P.S., Mermall, V., Tilney, L.G., and Mooseker, M.S. (2007). Roles for *Drosophila melanogaster* myosin IB in maintenance of Enterocyte brush-border structure and resistance to the bacterial pathogen *Pseudomonas entomophila*. *Mol. Biol. Cell* 18, 4625–4636.
- Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* 25, 414–423.
- Herz, H.-M., Mohan, M., Garrett, A.S., Miller, C., Casto, D., Zhang, Y., Seidel, C., Haug, J.S., Florens, L., Washburn, M.P., et al. (2012a). Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in *Drosophila*. *Mol. Cell. Biol.* 32, 1683–1693.
- Herz, H.M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y. h, Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012b). Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes & Development* 26, 2604–2620.
- Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. *Nature* 463, 474–484.
- Hodgson, J.W., Cheng, N.N., Sinclair, D.A., Kyba, M., Randsholt, N.B., and Brock, H.W. (1997). The polyhomeotic locus of *Drosophila melanogaster* is transcriptionally and post-transcriptionally regulated during embryogenesis. *Mech. Dev.* 66, 69–81.
- Hodgson, J.W., Argiropoulos, B., and 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, 4528–4543.
- Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a polycomb response element and *in vitro* binding of polycomb group complexes containing GAGA factor. *Mol. Cell. Biol.* 20, 3187–3197.
- Houbaviy, H.B., Usheva, A., Shenk, T., and Burley, S.K. (1996). Cocystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13577–13582.
- Huang, D.-H., Chang, Y.-L., Yang, C.-C., Pan, I.-C., and King, B. (2002). pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol. Cell. Biol.* 22, 6261–6271.
- Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* 306, 591–593.
- Ingham, P.W. (1984). A gene that regulates the bithorax complex differentially in larval and adult cells of *Drosophila*. *Cell* 37, 815–823.
- Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Dev. Cell* 26, 565–577.
- Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C.M., Nakamura, T., Mazo, A., Eisenbach, L., and Canaani, E. (2007). Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* 27, 1889–1903.
- Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-Containing and ATP-Utilizing Chromatin Assembly and Remodeling Factor. *Cell* 90, 145–155.
- Jacobs, J.J., Scheijen, B., Voncken, J.W., Kieboom, K., Berns, A., and van Lohuizen, M. (1999). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes & Development* 13, 2678–2690.
- Jones, C.A., Ng, J., Peterson, A.J., Morgan, K., Simon, J., and Jones, R.S. (1998). The *Drosophila* esc and E(z) Proteins Are Direct Partners in Polycomb Group-Mediated Repression. *Cell. Biol.* 18, 2825–2834.
- Jurgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* 316, 153–155.

- Kagey, M.H., Melhuish, T.A., and Wotton, D. (2003). The polycomb protein Pc2 is a SUMO E3. *Cell* 113, 127–137.
- Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the *Drosophila* *ubx* gene. *J. Biol. Chem.* 281, 29064–29075.
- Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial Interactions Are Required for the Efficient Recruitment of Pho Repressive Complex (PhoRC) to Polycomb Response Elements. *PLoS Genet.* 10, e1004495.
- Kalantry, S., Mills, K.C., Yee, D., Otte, A.P., Panning, B., and Magnuson, T. (2006). The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat. Cell Biol.* 8, 195–202.
- Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* 21, 569–571.
- Kaufman, T.C., Lewis, R., and Wakimoto, B. (1980). Cytogenetic Analysis of Chromosome 3 in *DROSOPHILA MELANOGASTER*: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics* 94, 115–133.
- Kennison, J.A. (1995). The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function. *Annu. Rev. Genet.* 29, 289–303.
- Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 85, 8136–8140.
- Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes. *Mol. Cell. Biol.* 25, 6857–6868.
- Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2010). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471, 480–485.
- Kim, C.A., Phillips, M.L., Kim, W., Gingery, M., Tran, H.H., Robinson, M.A., Faham, S., and Bowie, J.U. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* 20, 4173–4182.
- Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* 9, 453–457.
- Kim, C.A., Sawaya, M.R., Cascio, D., Kim, W., and Bowie, J.U. (2005). Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *J. Biol. Chem.* 280, 27769–27775.
- Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* 37, 2940–2950.
- King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol. Cell. Biol.* 22, 7919–7928.
- King, I.F.G., Emmons, R.B., Francis, N.J., Wild, B., Müller, J., Kingston, R.E., and Wu, C.-T. (2005). Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. *Mol. Cell. Biol.* 25, 6578–6591.
- Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell Biol.* 15, 353–362.
- Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet.* 9, e1003717.

- Klymenko, T. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & Development* 20, 1110–1122.
- Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A., et al. (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* 317, 1087–1090.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4, e1000242.
- Kyba, M., and Brock, H.W. (1998a). The *Drosophila* polycomb group protein Psc contacts ph and Pc through specific conserved domains. *Mol. Cell. Biol.* 18, 2712–2720.
- Kyba, M., and Brock, H.W. (1998b). The SAM domain of polyhomeotic, RAE28, and scm mediates specific interactions through conserved residues. *Dev. Genet.* 22, 74–84.
- Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes & Development* 22, 2799–2810.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.
- Lechtenberg, B.C., Allen, M.D., Rutherford, T.J., Freund, S.M.V., and Bycroft, M. (2009). Solution structure of the FCS zinc finger domain of the human polycomb group protein L(3)mbt-like 2. *Protein Sci.* 18, 657–661.
- Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318, 447–450.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-I., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313.
- Lessard, J., Schumacher, A., Thorsteinsdottir, U., van Lohuizen, M., Magnuson, T., and Sauvageau, G. (1999). Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes & Development* 13, 2691–2703.
- Lessard, J., and Sauvageau, G. (2003). *Bmi-1* determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 423, 255–260.
- Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010a). Jarid2 and PRC2, partners in regulating gene expression. *Genes & Development* 24, 368–380.
- Li, H., Fischle, W., Wang, W., Duncan, E.M., Liang, L., Murakami-Ishibe, S., Allis, C.D., and Patel, D.J. (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* 28, 677–691.
- Li, X., Han, Y., and Xi, R. (2010b). Polycomb group genes Psc and Su(z)2 restrict follicle stem cell self-renewal and extrusion by controlling canonical and noncanonical Wnt signaling. *Genes & Development* 24, 933–946.
- Li, Z., Cao, R., Wang, M., Myers, M.P., Zhang, Y., and Xu, R.-M. (2006). Structure of a *Bmi-1*-Ring1B polycomb group ubiquitin ligase complex. *J. Biol. Chem.* 281, 20643–20649.
- Van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. (1991). Identification of cooperating oncogenes in *Eμ*-myc transgenic mice by provirus tagging. *Cell* 65, 737–752.

- Van Lohuizen, M., Jacobs, J.J.L., Kieboom, K., Marino, S., and DePinho, R.A. (1999). The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* 397, 164–168.
- Lonie, A., D'Andrea, R., Paro, R., and Saint, R. (1994). Molecular characterisation of the Polycomblike gene of *Drosophila melanogaster*, a trans-acting negative regulator of homeotic gene expression. *Development* 120, 2629–2636.
- Lupo, R., Breiling, A., Bianchi, M.E., and Orlando, V. (2001). *Drosophila* chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol. Cell* 7, 127–136.
- Lynch, M.D., Smith, A.J.H., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* 31, 317–329.
- Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. *Development* 133, 1413–1422.
- Mager, J., Montgomery, N.D., de Villena, F.P.-M., and Magnuson, T. (2003). Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nat. Genet.* 33, 502–507.
- Mahmoudi, T., Zuijderduijn, L.M.P., Mohd-Sarip, A., and Verrijzer, C.P. (2003). GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. *Nucleic Acids Res.* 31, 4147–4156.
- Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762–767.
- Martinez, A.-M., and Cavalli, G. (2006). The role of polycomb group proteins in cell cycle regulation during development. *Cell Cycle* 5, 1189–1197.
- Martinez, A.-M., Schuettengruber, B., Sakr, S., Janic, A., Gonzalez, C., and Cavalli, G. (2009). Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling. *Nat. Genet.* 41, 1076–1082.
- Maurange, C., and Paro, R. (2002). A cellular memory module conveys epigenetic inheritance of hedgehog expression during *Drosophila* wing imaginal disc development. *Genes & Development* 16, 2672–2683.
- McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E., et al. (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci.* 109, 2989–2994.
- Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 6, e1001244.
- Mihaly, J., Hogga, I., Gausz, J., Gyurkovics, H., and Karch, F. (1997). In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. *Development* 124, 1809–1820.
- Mihaly, J., Mishra, R.K., and Karch, F. (1998). A conserved sequence motif in Polycomb-response elements. *Mol. Cell* 1, 1065–1066.
- Mishra, R.K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S.E., and Schedl, P. (2001). The *iab-7* polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell. Biol.* 21, 1311–1318.
- Mohammad, H.P., Cai, Y., McGarvey, K.M., Easwaran, H., Van Neste, L., Ohm, J.E., O'Hagan, H.M., and Baylin, S.B. (2009). Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. *Cancer Res.* 69, 6322–6330.

- Mohan, M., Herz, H.-M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., and Shilatifard, A. (2011). The COMPASS Family of H3K4 Methylases in *Drosophila*. *Mol. Cell. Biol.* **31**, 4257–4257.
- Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Mol. Cell. Biol.* **22**, 7473–7483.
- Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes & Development* **19**, 1755–1760.
- Mohd-Sarip, A., van der Knaap, J.A., Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C.P. (2006). Architecture of a polycomb nucleoprotein complex. *Mol. Cell* **24**, 91–100.
- Mohd-Sarip, A., Lagarou, A., Doyen, C.M., van der Knaap, J.A., Aslan, U., Bezstarosti, K., Yassin, Y., Brock, H.W., Demmers, J.A.A., and Verrijzer, C.P. (2012). Transcription-Independent Function of Polycomb Group Protein PSC in Cell Cycle Control. *Science* **336**, 744–747.
- Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.-K., Clarke, M.F., and Morrison, S.J. (2003). Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nat. Cell Biol.* **425**, 962–967.
- Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47–62.
- Morgan, N.S., Heintzelman, M.B., and Mooseker, M.S. (1995). Characterization of Myosin-IA and Myosin-IB, Two Unconventional Myosins Associated with the *Drosophila* Brush Border Cytoskeleton. *Dev. Biol.* **172**, 51–71.
- Mulholland, N.M., King, I.F.G., and Kingston, R.E. (2003). Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes & Development* **17**, 2741–2746.
- Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* **16**, 476–484.
- Müller, J., and Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr. Opin. Genet. Dev.* **19**, 150–158.
- 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., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208.
- Murzina, N.V., Pei, X.Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F., et al. (2008). Structural Basis for the Recognition of Histone H4 by the Histone-Chaperone RbAp46. *Structure* **16**, 1077–1085.
- Nanyes, D.R., Junco, S.E., Taylor, A.B., Robinson, A.K., Patterson, N.L., Shivarajpur, A., Halloran, J., Hale, S.M., Kaur, Y., Hart, P.J., et al. (2014). Multiple polymer architectures of human polyhomeotic homolog 3 (PHC3) SAM. *Proteins* **82**, 2823–2830.
- Nègre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal Distribution of PcG Proteins during *Drosophila* Development. *PLoS Biol.* **4**, e170.
- Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* **26**, 4078–4088.
- Nijman, S.M.B., Luna-Vargas, M.P.A., Velds, A., Brummelkamp, T.R., Dirac, A.M.G., Sixma, T.K., and Bernards, R. (2005). A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773–786.
- Nowak, A.J., Alfieri, C., Stirnimann, C.U., Rybin, V., Baudin, F., Ly-Hartig, N., Lindner, D., and Müller, C.W. (2011). Chromatin-modifying Complex Component Nurf55/p55 Associates with Histones H3 and H4

and Polycomb Repressive Complex 2 Subunit Su(z)12 through Partially Overlapping Binding Sites. *J. Biol. Chem.* **286**, 23388–23396.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801.

O'Connell, S., Wang, L., Robert, S., Jones, C.A., Saint, R., and Jones, R.S. (2001). Polycomb-like PHD fingers mediate conserved interaction with enhancer of zeste protein. *J. Biol. Chem.* **276**, 43065–43073.

O'Dor, E., Beck, S.A., and Brock, H.W. (2006). Polycomb group mutants exhibit mitotic defects in syncytial cell cycles of *Drosophila* embryos. *Dev. Biol.* **290**, 312–322.

O'Loughlen, A., Muñoz-Cabello, A.M., Gaspar-Maia, A., Wu, H.-A., Banito, A., Kunowska, N., Racek, T., Pemberton, H.N., Beolchi, P., Laval, F., et al. (2012). MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33–46.

Ogawa, H., Ishiguro, K.-i., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A Complex with Chromatin Modifiers That Occupies E2F- and Myc-Responsive Genes in G0 Cells. *Science* **296**, 1132–1136.

Ohta, H., Sawada, A., Kim, J.Y., Tokimasa, S., Nishiguchi, S., Humphries, R.K., Hara, J., and Takihara, Y. (2002). Polycomb group gene *rae28* is required for sustaining activity of hematopoietic stem cells. *J. Exp. Med.* **195**, 759–770.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev. Cell* **15**, 877–889.

Orlando, V., Jane, E.P., Chinwalla, V., Harte, P.J., and Paro, R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17**, 5141–5150.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246.

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

Paro, R., and 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.

Pasini, D., Bracken, A.P., Jensen, M.R., Denchi, E.L., and Helin, K. (2004). *Suz12* is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein *Suz12* is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* **27**, 3769–3779.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). *JARID2* regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). *Jarid2/Jumonji* Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell* **139**, 1290–1302.

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**, 698–699.

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

- Peterson, A.J., Mallin, D.R., Francis, N.J., Ketel, C.S., Stamm, J., Voeller, R.K., Kingston, R.E., and Simon, J.A. (2004). Requirement for sex comb on midleg protein interactions in *Drosophila* polycomb group repression. *Genetics* 167, 1225–1239.
- Petruk, S., Sedkov, Y., Smith, S., Tillib, S., Kraevski, V., Nakamura, T., Canaani, E., Croce, C.M., and Mazo, A. (2001). Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. *Science* 294, 1331–1334.
- Poux, S., Kostic, C., and Pirrotta, V. (1996). Hunchback-independent silencing of late Ubx enhancers by a Polycomb Group Response Element. *EMBO J.* 15, 4713–4722.
- Pullirsch, D., Härtel, R., Kishimoto, H., Leeb, M., Steiner, G., and Wutz, A. (2010). The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. *Development* 137, 935–943.
- Puschendorf, M., Terranova, R., Boutsma, E., Mao, X., Isono, K.-I., Brykczynska, U., Kolb, C., Otte, A.P., Koseki, H., Orkin, S.H., et al. (2008). PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat. Genet.* 40, 411–420.
- Qiao, F. (2005). The Many Faces of SAM. *Sci. STKE Signal Transduct. Knowl. Environ.* 286, re7.
- Qiao, F., Song, H., Kim, C.A., Sawaya, M.R., Hunter, J.B., Gingery, M., Rebay, I., Courey, A.J., and Bowie, J.U. (2004). Derepression by depolymerization: Structural insights into the regulation of Yan by Mae. *Cell* 118, 163–173.
- Rappailles, A., Decoville, M., and Locker, D. (2012). DSP1, a *Drosophila* HMG protein, is involved in spatiotemporal expression of the homeotic gene *Sex combs reduced*. *Biol. Cell* 97, 779–785.
- Ren, X., Vincenz, C., and Kerppola, T.K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell. Biol.* 28, 2884–2895.
- Reya, T., and Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nat. Cell Biol.* 434, 843–850.
- Ringrose, L., and 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., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev. Cell* 5, 759–771.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.
- Robinson, A.K., Leal, B.Z., Chadwell, L.V., Wang, R., Ilangovan, U., Kaur, Y., Junco, S.E., Schirf, V., Osmulski, P.A., Gaczynska, M., et al. (2012). The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. *J. Biol. Chem.* 287, 8702–8713.
- Roseman, R.R., Morgan, K., Mallin, D.R., Roberson, R., Parnell, T.J., Bornemann, D.J., Simon, J.A., and Geyer, P.K. (2001). Long-range repression by multiple polycomb group (PcG) proteins targeted by fusion to a defined DNA-binding domain in *Drosophila*. *Genetics* 158, 291–307.
- Santiveri, C.M., Lechtenberg, B.C., Allen, M.D., Sathyamurthy, A., Jaulent, A.M., Freund, S.M.V., and Bycroft, M. (2008). The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J. Mol. Biol.* 382, 1107–1112.
- Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660.
- Savla, U., Benes, J., Zhang, J., and Jones, R.S. (2008). Recruitment of *Drosophila* Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development* 135, 813–817.

- Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* 465, 243–247.
- Scheuermann, J.C., Gutiérrez, L., and Müller, J. (2012). Histone H2A monoubiquitination and Polycomb repression: The missing pieces of the puzzle. *Fly (Austin)* 6, 162-168.
- Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* 42, 330–341.
- Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in *Drosophila* Embryos. *PLoS Biol.* 7, e100003.
- Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* 12, 799–814.
- Schwartz, Y.B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–864.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705.
- Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative Epigenetic Chromatin States of Polycomb Target Genes. *PLoS Genet.* 6, e1000805.
- Schwendemann, A., and 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, 12883–12888.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98, 37–46.
- Sharif, J., Endo, T.A., Toyoda, T., and Koseki, H. (2010). Divergence of CpG island promoters: A consequence or cause of evolution? *Dev. Growth Differ.* 52, 545–554.
- Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502.
- Sherr, C.J. (2001). The INK4a/ARF network in tumour suppression. *Nat. Rev. Mol. Cell Biol.* 2, 731–737.
- Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* 12, 599–612.
- Shimell, M.J., Peterson, A.J., Burr, J., Simon, J.A., and O'Connor, M.B. (2000). Functional analysis of repressor binding sites in the *iab-2* regulatory region of the abdominal-A homeotic gene. *Dev. Biol.* 218, 38–52.
- Simon, J.A., Chiang, A., and Bender, W. (1992). Ten different Polycomb group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. *Development* 114, 493–505.
- Simon, J.A., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.* 158, 131–144.
- Simon, J.A., and Lange, C.A. (2008). Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat. Res.* 647, 21–29.
- Simon, J.A., and Kingston, R.E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* 49, 808–824.
- Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138, 885–897.

- Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci.* *107*, 20980–20985.
- Song, J.-J., Garlick, J.D., and Kingston, R.E. (2008). Structural basis of histone H4 recognition by p55. *Genes & Development* *22*, 1313–1318.
- Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* *6*, 846–856.
- Srinivasan, S., Dorighi, K.M., and Tamkun, J.W. (2008). Drosophila Kismet Regulates Histone H3 Lysine 27 Methylation and Early Elongation by RNA Polymerase II. *PLoS Genet* *4*, e1000217.
- Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* *15*, 340–356.
- Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* *9*, 1428–1435.
- Strübbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. (2011). Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc. Natl. Acad. Sci.* *108*, 5572–5577.
- Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*. *EMBO J.* *4*, 3259.
- Struhl, G., and Basler, K. (1993). Organizing activity of wingless protein in *Drosophila*. *Cell* *72*, 527–540.
- Strutt, H., Cavalli, G., and 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*, 3621–3632.
- Strutt, H., and Paro, R. (1997). The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol. Cell. Biol.* *17*, 6773–6783.
- Suganuma, T., Gutiérrez, J.L., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat. Struct. Mol. Biol.* *15*, 364–372.
- Sung, S., and Amasino, R.M. (2004). Vernalization and epigenetics: how plants remember winter. *Curr. Opin. Plant Biol.* *7*, 4–10.
- Nakagawa, T., Kajiyani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M., et al. (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes & Development* *22*, 37–49.
- Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* *397*, 161–168.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* *148*, 664–678.
- Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* *14*, 1025–1040.
- Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* *15*, 668–679.

- Tie, F., Furuyama, T., and Harte, P.J. (1998). The *Drosophila* Polycomb Group proteins ESC and E(Z) bind directly to each other and co-localize at multiple chromosomal sites. *Development* 125, 3483–3496.
- Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and 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, 275–286.
- Tie, F., Stratton, C.A., Kurzhals, R.L., and Harte, P.J. (2007). The N terminus of *Drosophila* ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. *Mol. Cell. Biol.* 27, 2014–2026.
- Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* 136, 3131–3141.
- Tolhuis, B., de Wit, E., Muijters, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699.
- Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., et al. (2011). L3MBTL2 Protein Acts in Concert with PcG Protein-Mediated Monoubiquitination of H2A to Establish a Repressive Chromatin Structure. *Mol. Cell* 42, 438–450.
- Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83, 1021–1026.
- Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., and Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res.* 40, 3403–3418.
- Vidal Matos, R. (2009) Biochemical Characterization of the *Drosophila* Polycomb protein dSfmbt. Ruperto-Carola University of Heidelberg, Germany, 84 pages; www.ub.uni-heidelberg.de/archiv/10684.
- Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871–874.
- Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat. Genet.* 28, 371–375.
- Wang, Y.J., and Brock, H.W. (2003). Polyhomeotic stably associates with molecular chaperones Hsc4 and Droj2 in *Drosophila* Kc1 cells. *Dev. Biol.* 262, 350–360.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassisi, J.A., and Jones, R.S. (2004b). Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Mol. Cell* 14, 637–646.
- Wang, L., Jahren, N., Miller, E.L., Ketel, C.S., Mallin, D.R., and Simon, J.A. (2010). Comparative analysis of chromatin binding by Sex Comb on Midleg (SCM) and other polycomb group repressors at a *Drosophila* Hox gene. *Mol. Cell. Biol.* 30, 2584–2593.
- Wang, R., Ilangoan, U., Leal, B.Z., Robinson, A.K., Amann, B.T., Tong, C.V., Berg, J.M., Hinck, A.P., and Kim, C.A. (2011). Identification of Nucleic Acid Binding Residues in the FCS Domain of the Polycomb Group Protein Polyhomeotic. *Biochemistry* 50, 4998–5007.
- Wasiliko, D.J., and Lee, S.E. (2006). TIPS: Titerless Infected-Cells Preservation and Scale-Up. *BioProcess J* 5, 29-32.
- Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. *Trends Genet.* 23, 494–502.
- Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140, 99–110.

- Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Mol. Cell.* 49, 1134–1146.
- Xi, H., Yu, Y., Fu, Y., Foley, J., Halees, A., and Weng, Z. (2007). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res.* 17, 798–806.
- Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol. Cell* 38, 662–674.
- Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K.L., Woo, A.J., Chouinard, C.R., Labadorf, A., Akie, T.E., Moran, T.B., et al. (2012). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol. Cell* 45, 330–343.
- Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* 286, 7983–7989.
- Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science* 337, 971–975.
- Zallen, J.A., Cohen, Y., Hudson, A.M., Cooley, L., Wieschaus, E., and Schejter, E.D. (2002). SCAR is a primary regulator of Arp2/3-dependent morphological events in *Drosophila*. *J. Cell Biol.* 156, 689–701.
- Zhang, J., Bonasio, R., Strino, F., Kluger, Y., Holloway, J.K., Modzelewski, A.J., Cohen, P.E., and Reinberg, D. (2013). SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & Development* 27, 749–766.
- Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Mol. Cell* 29, 69–80.
- Zink, B., and Paro, R. (1989). In vivo binding pattern of a trans-regulator of homoeotic genes in *Drosophila melanogaster*. *Nature* 337, 468–471.
- Zink, B., Engström, Y., Gehring, W.J. and Paro, R. (1991). Direct interaction of the Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of *Drosophila melanogaster*. *EMBO J.* 10, 153.

6. Appendix

6.1 List of Abbreviations

aa (amino acid)	PcG Polycomb group
Abd-B (Abdominal-B)	Pcl (Polycomb-like)
<i>Antp</i> (<i>Antennapedia</i>)	PCR (polymerase chain reaction)
Ash (Absent, small and homeotic discs)	Ph (Polyhomeotic)
Asx (Additional sex combs)	Ph-d (Polyhomeotic-distal)
ChIP (chromatin immunoprecipitation)	PHD (plant homeodomain)
<i>D.m.</i> (<i>Drosophila melanogaster</i>)	Ph-p (Polyhomeotic-proximal)
dRAF (<i>Drosophila</i> Ring-associated factors complex)	Pho (Pleiohomeotic)
dSfmbt (<i>Drosophila</i> Scm-like with four MBT domains)	PhoRC (Pho repressive complex)
DTT (Dithiothreitol)	PMSF (phenylmethanesulfonyl fluoride)
Esc (Extra sex combs)	PRC (Polycomb repressive complex)
ESCs (embryonic stem cells)	PRC1 (Polycomb repressor complex 1)
E(z) (Enhancer of zeste)	PRC2 (Polycomb repressor complex 2)
H2A-ub1 (monoubiquitination of histone H2A)	PR-DUB (Polycomb repressive deubiquitinase)
H3-K4me (methylation on lysine 4 of histone H3)	PRE (Polycomb response element)
H3-K9me (methylation on lysine 9 of histone H3)	Psc (Posterior sex combs)
H3-K27me (methylation on lysine 27 of histone H3)	RING (Really interesting new gene)
H3-K36me (methylation on lysine 36 of histone H3)	RYBP (Ring and YY1 Binding Protein)
H4-K20me (methylation on lysine 20 of histone H4)	SAM (sterile alpha motif)
HA (haemagglutinin)	SANT (Swi3, Ada2, N-CoR, TFIIB)
HD (homology domain)	Sce (Sex combs extra)
HMTase (histone methyltransferase)	Scm (Sex combs on midleg)
HOX (Homeobox)	SDS (sodium dodecylsulfate)
IP (immunoprecipitation)	SET (Suvar3-9, Enhancer-of-zeste, Trithorax Spt5 = Suppressor of Ty 5)
kb (kilobases)	Su(z)2 (Suppressor of zeste 2)
kDa (kilo Daltons)	Su(z)12 (Suppressor of zeste 12)
MBT (malignant brain tumor)	Sxc (Super sex combs)
me1 (monomethylation)	trxG (trithorax group)
me2 (dimethylation)	ub (ubiquitin)
me3 (trimethylation)	Ubx (Ultrabithorax)
MW (molecular weight)	UCH (ubiquitin carboxy-terminal hydrolase)
NURF (nucleosome remodeling factor)	UTX (Ubiquitously transcribed tetratricopeptide repeat, X chromosome)
O-GlcNAc (O-linked N-acetylglucosamine)	VEFS (VRN2-EMF2-FIS2-Su(z)12)
Ogt (O-GlcNAc transferase)	wt (wild-type)
PBS (phosphate buffered saline)	Zn (zinc)
Pc (Polycomb)	

6.2 List of Figures

	Page
1.1 PcG and TrxG proteins maintain transcriptional states of developmental control genes.	5
1.2 PcG protein complexes and their activities in <i>Drosophila</i> .	7
1.3 Functional domains and interactions of fly PcG complexes.	20
1.4 Recruitment of PcG complexes to target genes.	30
1.5 Particular role of Polyhomeotic (Ph) genes and proteins in <i>Drosophila</i> .	35
3.1 Expression of <i>TAP-Ph-p/ -d</i> transgenes in embryonic nuclear extract.	71
3.2 <i>TAP-Ph</i> transgenes partially rescue <i>ph⁰</i> mutants.	73
3.3 Analysis of TAP-purified material by Mass spectrometry.	75
3.4 TAP-Ph purifications identify PRC1 members and PhoRC members as major Ph interactors.	78
3.5 Scm mediates Ph interaction with Sfmbt.	80
3.6 Sfmbt, Ph and Scm form a trimeric complex <i>in vitro</i> .	82
3.7 Sfmbt, Ph and Scm form a stable trimeric complex <i>in vitro</i>	83
3.8 An N-terminal portion of Sfmbt interacts with Scm.	85
3.9 The Scm:Sfmbt interaction is not dependent on FCS zinc finger domains.	87
3.10 Putative domains involved in Scm:Sfmbt interaction.	90
3.11 Scm and Sfmbt also interact via their SAM domains.	93
3.12 The Scm and Sfmbt SAM domains form a stable complex.	95
3.13 Structure of the Scm-SAM/ Sfmbt-SAM complex.	99
3.14 Conserved interactions between SAM/ SAM interfaces.	101
4.1 Novel identified interactions linking PRC1 and PhoRC complex.	106
4.2 Regulation of SAM domain polymerization: A model for Ph-, Scm- and Sfmbt-SAM domain interaction.	108
4.3 PRC1 complex recruitment to PREs via the Pho-RC complex.	112
4.4 A possibly conserved role of SAM domains in PRC1 recruitment.	115

6.3 List of Tables

	Page
1.1: List of known <i>Drosophila</i> PcG proteins and human homologs.	14
1.2 List of TrxG proteins and their mammalian homologs.	16
1.3 PcG protein structures with protein database accession (PBD) numbers.	22
2.1 Sequencing primers used in this study	40
2.2 PCR/ Cloning primers used in this study.	42
2.3 PCR/ Mutagenesis primers used in this study	44
2.4 Vectors used in this study.	45
2.5 List of fly strains used in this study.	46
2.6 List of Baculoviruses generated for this study.	47
2.7 Primary antibodies used in this study	47
2.8 Secondary antibodies used in this study.	48
2.9 X-Ray data Collection and Refinement Statistics	63
3.1 Ph interacting proteins identified by Mass spectrometric analysis	76
6.1 Peptides identified in TAP-Ph-p/ -d eluates.	138

6.4 Supplementary material

Table 6.1: Peptides identified in TAP-Ph-p/ -d eluates.

Peptides identified	Ph-d 1	Ph-d 2	Ph-p
Ph-d (Polyhomeotic distal)			
AGISFDEDFAK	+	+	
ENHLVNAMGMK	+	+	+
GPTATLVPIDSPK	+	+	+
GSVGTPSIR	+	+	
HPLLPLSSR	+	+	
QSNAAVQPPSSTIPNSVSGK	+	+	
QSNAAVQPPSSTIPNSVSGKEEPK	+		
RHPLLPLSSR	+	+	
TPPPSPEATTSVK	+	+	
VDPQRPLR	+	+	+
Ph-p (Polyhomeotic proximal)			
AGITFDEK			+
GPTATLVPIGSPK	+	+	+
HLVNAMGMK	+	+	+
HTSLTLEK		+	+
QSNAAVQPPSSTTPNSVSGKEEPK		+	+
QTHTPSTPNRPSAPSTPNTNCNSIAR			+
RADTESDTTTPVSTTASQGISASAILAGGTLPLK			+
YDVASPPHPGIAQQQATSGTGPATGSGSVTPTSHR			+
Ph-d/ Ph-p (Polyhomeotic distal/ proximal)			
ATMQEDIK	+	+	+
CLETLAQK	+	+	+
ELPGCQDYVDDFIQQEIDGQALLLKK	+		
LDEAMAEK	+	+	+
LSGIASAPGSDMVACEQCGK	+	+	+
MVVMSTTGTPITLQNGQTLHAATAAGVDK	+	+	
NGIGGVGSGETNGLGTGGIVGVDAMALVDR	+		
NQPDGTQGMFIQQQPATQTLQTQQNQIIQCNVTQTPTK	+		
PASSVSTQTAQNQSLLK	+	+	
PFQGNPQMLTTTTQNAK	+		
PGAPVMPHNGTQVR	+		
QEFPTHHTSGSGTELK	+	+	+
QQQLQLFQK	+	+	+
TEIGQVAGQNK	+	+	+
VGVDAAQGK	+	+	+
VGVDAAQGKLAQK	+		
VVGHLTTVQQQQATNLQQVVNAAGNK	+	+	+
YADKDVSEPPK	+	+	+

Psc (Posterior sex combs)			
ASRPNPFANIPNDVNR		+	
DRPEEAALATPEQR	+	+	+
DVAATPPTETLK	+	+	
EIVKPLKPEK		+	+
ENQEQLAVEVASSK	+	+	
FEIDAQR		+	
FSIDIMYK	+		
FVYDKFEIDAQR	+	+	
IMSPSGVSTLSPR		+	
LPDQPQDQVQAAK	+	+	+
LSPPLPTVDFK		+	+
LTAATAPQTK	+	+	+
LVPGLYER		+	+
QNSVTIIDMSDPERR		+	
QSPPAAVPGHLLAAK	+		
RYQPILPK	+	+	
SDTTLQAIVYK	+	+	+
SIGGGSVENNSNAAQKPHLYGPK	+		
SPSPLTVPPLTIR	+	+	+
SPVNNYIEIVK	+	+	+
TAAQMGSHTASSENK		+	
TDSEPELVDTLRPR	+	+	
VATPPPPSSPR		+	
VEPVSLPEDQK	+	+	+
VEPVSLPEDQKAEASIK	+		
VGNEVFNDYLQK	+		
VTPLKPVLTPTQVDK	+		
VTSGAFSEDPK	+	+	+
VYESPQPLVKPAPR		+	+
YLQCPAMCR		+	
Sce/Ring (Sex combs extra)			
ADPNFDLLISK	+	+	+
AMSVLTSER		+	+
EEYEAIQEK	+	+	+
FCSDCIVTALR	+	+	+
FNQTQSQQALVNSINEGIK	+	+	
IYPSREEYEAIQEK	+	+	
KPQEVITDSTEIAVSPR	+	+	+
LTLDLGADLPEACR		+	+
MQVDDASNPPSVR	+	+	+
SEESSEDSQMDCR		+	
SLHSELMCPICLDMLK	+		
STPSPVPSNSSSSKPK	+	+	+

TSLDPAPNK		+	+
TTANATVDHLSK	+	+	+
TWELSLYELQR	+		
Pc (Polycomb)			
DNATDDPVDLVYAAEK		+	
IASEAATQLK	+	+	+
LIDIYEQTNK	+	+	+
QPLTPLSPR		+	
VVITDVTVNLETVTIR	+		
Su(z) 2 (Suppressor of zeste 2)			
ILLYDNEQTK	+	+	
INQDIEPEHSVR	+		
LDSTSTSEALNR	+		
Ge-1 (Enhancer of mRNA-decapping protein 4 homolog)			
IQNVAEFK		+	
IQPAALESGLK	+	+	+
LMEQYLK	+	+	
LTEFLAAR	+	+	
MELLIDLK		+	
NLSQLAYK		+	
QLHDAFSVGIK		+	
QLLMAGQINK		+	
QLVESSLHK		+	
SIDSAVLQTIR	+	+	
SLEILLAR	+	+	
SQAATPAPPYDLR		+	
SSQFQCFQVK		+	
TELTDAMLETQR		+	
VCNIATSMR		+	
VLTELYR		+	+
VLTSGGVHTR		+	
Sop2 (Suppressor of profilin 2)/Arpc1			
DIEEPPTTPWGNR	+		
IFQSMDR	+	+	+
KPLGQLMAEFR	+	+	
LADVLNQHDLR	+	+	+
NAYVWTQGDDGK	+		
STVTSLDWHPNVLVLLAGSTDYK	+		
TENTDTVVDSIHQNAITSVR	+	+	
TQIALSPNNHEIHIYSR	+		
WKPALVLLR		+	
Nup107 (Nucleoprin 107kD)			
ATAGVFSGHLGSLK	+		
EVIQQLYALNATLR	+		

FLEQVEQK		+	+
GEQQNPLHAHDR			+
IADASLELLNSK	+	+	
ILLQTDR	+	+	
IVAAYVEALIAR	+		
KPQTSHAASSQDNFTER		+	
LADEISSENR	+	+	
LHEDPNFEQNVSVLHEK	+		
LPIEGNPR	+	+	
Sfmbt (Scm-related gene containing four mbt domains)			
DTGAVAAGQHLFHR	+	+	
GELYSLVLNTK	+	+	
GNIDPSVIPIQK		+	+
INDSLQSR		+	
ISDLIAQLK	+	+	
LVCVATVAR		+	
MNFTFDEYYSDGK		+	
VNDCTAHCDTFSR	+		
Myo61F (Myosin 61F)			
AGVQQLVK		+	
DKGDLILIIPR		+	
DKGHLVIIGTQ		+	
DLVVTAAR		+	+
FILLSNK		+	
GDLILIIPR		+	
GDVVTSPLNQELAIYAR		+	
GGVIDIQTGAEPGVVR	+	+	
GHLVIIGTQ		+	
GIISILDEECLRPGEPTDK		+	
GQCVLISGESGSGK		+	
IDFTLTNHNDLDMVIR		+	
LLGVNASELEAALTHR	+	+	+
LNISLQAK		+	+
METGLHER		+	
NDAPNGFNEEFIANAK		+	
NNDLLFR		+	
NSLEHNVVK		+	
QVQQALTVIDFTK		+	
RPETAITQFR	+	+	
SLIEENR		+	
SNPVLEAFGNAK		+	
TLFDTEDAYQEK		+	+
TYELFLER		+	
VICNLIEEK	+	+	+

VNDFVASTFGSEQLK		+	
YLGLMENLR		+	
Pyd (Polychaetoid)			
AEELATAQFNATK	+	+	
AVVQDDVQAEYITR	+	+	
DSGNTVQLVVK	+		
FSTPLQDDDK	+		
GSAFELYR		+	+
GSSGGAAQEDFYSSR	+	+	
HALDLAGSR	+	+	+
HALLDITPNAVDR	+	+	
LGP NAPDDLK		+	+
LNLVVLRL	+		
NNSVTQADYTK	+		
PVVLFGPVSDLAR	+		
QSKPGDIPLNTSGSSSR	+		
SFGGPNDLNR	+	+	+
VPGYGFGIAVSGGR	+	+	+
Edc3 (Enhancer of decapping 3)			
AGLSLQR		+	+
ASVLAIIDPPPCGINEVAIK		+	
EQDLPSTPR	+	+	+
GASDLAITLLGGAR	+	+	+
GNFNSALSDK		+	
HDENILASK		+	
LGSYVSNTR	+	+	
LLEQNSSSPEISLFK		+	
LPHFSNILGK	+	+	+
PIDIVSNGDGFYK		+	
QDLDGHTAPPPVVKPTPVK		+	
QISAEITIVR		+	
QLASHGLTVLLYVEQAK		+	+
QNAEVVLK		+	
SIDLIEPAK		+	+
TINIGAAATGR	+	+	+
YRHDENILASK		+	
Gnf1 (Germline transcription factor 1)			
AALLSGPPGIGK		+	+
AVLEFENEDIDR	+		
GGMQELIALIK		+	
HVDPTELFGGETK	+		
LAQELHDHTR		+	+
SLSGYFTGQQQAVSR		+	
YLVVGEEAGPK		+	

Osp (Outspread)			
CVENAALEEK		+	
ETDLAQALEK		+	
LAQGIEENEGLYK		+	
LEQSLTQLQQIR		+	
LTMDLEQK		+	
LVLASLSPSSR	+	+	+
MSELLATLQR		+	
SEQLLEQR		+	
YQTEIEQLR		+	
Lam (Lamin Dm0)			
DEIQSLYEDK	+	+	
DQIHSQEINESR	+	+	+
LTIEVQTTR		+	
NLEQETLSR	+	+	+
VELQNLNDR	+	+	+
Iswi (Imitation Switch)			
AALDSLGESSLR	+	+	+
CNTLITLIER		+	
DDIDNIAK		+	
DIDVVNGAGK	+		
ENIELEEK		+	
FGANQVFSSK		+	+
KIDAEPLTEEEIQEK		+	
LDGQTPHEDR		+	
LITESTVEEK		+	
LQNILMQLR		+	
NYTEIEDR		+	
TALELQR		+	+
YLVIDEAHR		+	
HCF (Host cell factor)			
IVPSVTASHSLR		+	
LSAINSCGR		+	
SAEALVLETAEIR	+	+	
SGSALGLGVEATSTVLK		+	
TLISNQSGVK		+	+
TSGEAPLPR	+	+	+
TSGGMTVQTLPK		+	
TYGDSPPPR		+	
VLNPTGPQPR		+	+
YLNDLYILDTR		+	
YSNELYELQATK		+	
Scm (Sex comb on midleg)			
EFCSETCIAEFR		+	

LPLQYVLPTQTGK LVDSTEIHAIGHCEK	+	+	
Pontin/ Tip49 ALLLAGPPGTGK LDPSIFDALQK TISNVVIGLK		+	+
Caf-1/Nurf55 (Chromatin assembly factor 1 subunit) GEFGGFGSVCGK HPSKPEPSGECQPDRL IGEEQSTEDAEDGPPPELLFIHGGHTAK LMIWDTR SDNAAESFDDAVEER TPSSDVLVFDYTK TVALWDLR	+	+	+
Actin-42A CPESLFQPSFLGMEACGIHETTYNSIMK IVAPPER	+	+	+
Actin-5C CPEALFQPSFLGMEACGIHETTYNSIMK IIAPPER IIAPPERK IKIIAPPER IKIIAPPERK	+	+	+
Actin 42-A/ Actin-5C DLYANTVLSGGTTMYPGIADR EITALAPSTMK KDLYANTVLSGGTTMYPGIADR LCYVALDFEQEMATAASSSSLEK QEYDESGPSIVHR YPIEHGIVTNWDDMEK TTGIVLDSGDGVSHTVPIYEGYALPHAILR DSYVGDEAQSK DSYVGDEAQSKR HQQVMVGMGQKDSYVGDEAQSK IWHHTFYNELR AGFAGDDAPR CDVDIRK DLTDYLMK EIVRDIKEK GYSFTTTAER GYSFTTTAEREIVR HQQVMVGMGQK ILTERGYSFTTTAER LDLAGR	+	+	+

LDLAGRDLTDYLMK			
SYELPDGQVITIGNER	+	+	+
VAPEEHPVLLTEAPLNPK	+	+	+
VAPEEHPVLLTEAPLNPKANR	+		
RGILTLK			
AVFPSIVGR	+	+	+
AVFPSIVGRPR	+	+	+
Unc-115			
EGQALVALDR		+	
EYGRPNAEDISR		+	
IYPYHLLITNYR		+	
IYTYSYLTDAPHYLR		+	
LNSGIGSAIAK		+	
SEASVDSITEGDRR		+	
SPGMNNEEPIELSHYPAAK		+	
TPLSPHFHR		+	
TVSPGLILR		+	
VLQAGDNHHFHPTCAR		+	
VLVDAIR		+	
YESPIGASPSR	+	+	
Dd1c-1 (Dynein light chain 1)/ Cdlc1			
NFGSYVTHETR		+	+
YNPTWHCIVGR		+	
Mlc1 (Myosin akali light chain 1)			
ALNLNPTLALIEK		+	+
EKEQGCEYEDFIECLK		+	
EQGCYEDFIECLK		+	
IKLDEFPLPIYSQVK	+		
LDEFPLPIYSQVK	+	+	+
LMSDPVVF		+	
Nop56			
AQAIIDAAK		+	+
IVPDNYMFAK		+	
SAGVAQLGLGHSYSR	+		
YPASTVQILGAEK		+	
Cep97 (CG3980)			
CLPTSLETTLAK		+	
IDELNDVDLK	+		
IDNIDSYLK	+	+	+
IETLSLAR	+	+	
LIEHVTESSAVTIQK	+	+	
QLILDENELQK		+	
TQEYIEQLGK		+	+
TRPSTLALESK	+	+	

VLNLEGNNIK	+	+	
VTVTALNPQLHK	+		
YLSSVCPLVGK		+	
SF2 (SR family splicing factor 2)			
DGSGVVEFLR	+	+	+
DIQDLFHK		+	+
EAGDVCFADTYK	+	+	+
GPPFAFVEFEDAR	+	+	+
GTPTYSPVR	+	+	+
LRVEFPR		+	
SHEGEVAYIR	+	+	+
VMVTGLPASGSWQDLK		+	
Troponin C, isoform 3			
FIVEEDAEAMQK		+	
SSVDEDLTPEQIAVLQK		+	
IMPalpha2 (Importin subunit alpha)/ Pendulin			
EAAWTVSNITAGNQK	+		
ILKPFIDLLDTK	+		
IQAVVDSDAVPR	+	+	+
LETLQQHENEEVYK	+		
LGLLLQHMK		+	
NINDEDLTSPLK	+		
QIQAVIQAGIFQQLR	+		
Nurf-38 (Nucleosome remodeling factor- 38kD)			
FVANCFPHK		+	+
IIAIDVNDPLASK	+	+	+
NADFANTIIAETHK	+	+	+
NSPSYSLYFK		+	
TIYNMVVEVPR		+	
pABPC1 (Polyadenylate-binding protein)			
ALDTMNFDLVR	+		
FSSAGPVLSIR	+		
GAQPQVQGTHAAAAAANMR		+	
HESVFGVNLVVK	+		
IAFSPYGNITSAK		+	
NPPVPQLHQTQPIPQQLQGK		+	+
PLYVALAQR	+	+	
SGVGNVFIK		+	
VEEAVAVLQVHR	+	+	+
lost/ growl			
AAALFANEEEFKK	+		
AESAETATPVQTK	+	+	+
ILPVLQQLK		+	+
IPSFVGADK		+	+

SVYLPSTR		+	
VDVLADATEEQKK		+	
VGIGFNNIFNR	+	+	
YDTPVDIATPTEIIR	+		
Rpl17 (Ribosomal protein L17)			
LVVHHIQVNR		+	+
Arpc5 (Actin-related protein2/3 complex, subunit 5)/ P16-ARC			
DHALNITLR	+	+	+
GFEIPSEGSSGHLLQWHEK	+	+	
GGVGCIVR	+	+	
KIDVDQYNEDNFR	+		
SVEALLSALQNAPLR	+		
Cp110 (CG14617)			
ALDEHGALEASR	+	+	
ILQFEQSGLGK	+	+	
MQAEFDHQQR		+	
B52 protein/ SRp55 (Serine-arginine protein 55)			
LIVENLSSR		+	
VYVGGLPYGVR		+	
cindr (GH03163)			
AATIAVEDTDFDR	+	+	
AATIAVEDTDFDRVER	+		
ASGVTGMFPDNFVR		+	
ASILTDMR		+	
ATRPNSLAIR	+	+	
LEGLVLNQQR		+	
LFAAEQAVSSNSSSITTK	+	+	
LTTAVSHDLADGLAASK	+	+	+
MLSGANQLPESGAAGNVVIR	+	+	
NLDLTKPPGVGASVSIQQR	+	+	
RPPTIGATVTTTSALTSK	+	+	+
SPPPPVLSK	+	+	
TIEELVNALK	+	+	
TPTGGVPMTGNLLGGK	+	+	
TSDDNSTGEEPQLAKPK		+	
VAQLEQR		+	
VDKLEGLVLNQQR		+	
VGVFDPNFVK	+	+	+
VGVFPSNFVQHIEPSPVLASK		+	
gammaTub37C (Tubulin gamma-2 chain at 37C)			
DVFFYQADDNHYIPR		+	
LYNQENVFLSK	+		
NAFLDNFR	+		
VINNIMTSPYSK	+		

VQEEVFDILDR	+	+	
Msp (Mini spindles)			
AAALNCINSFGEK	+	+	
AAINELAAIIEAPEK	+		
AQISTLPQVTLK	+		
AVTPLLADVDPLK	+	+	
DGSASGIGASPQR		+	
EEFTELLR	+		
EITPEELQEK	+	+	
FYDTNPSVLIK	+		
GVQSTNPTVR	+	+	
HQAPIPEEPK		+	
IVSACVAATTLALR		+	
KPATVSGGGATSAPTAALK		+	
LAVGDDSQYNK	+		
LIKPSIGDLAPALHR	+		
LLTSLVK	+		
LQAIISEAR	+		
MVVDNALAQEK	+		
NLFPGFHALGDNK	+		
QAGISQILIR	+		
QLAVEIYR	+	+	
SIIEFGFQLQPK	+		
TLNEPDPTVR		+	+
TVGDVMTGIVQK		+	
VDIAPQITEALLK	+		
VFGYVMEGLK		+	
VLSFAFEQK	+		
YIEEGLAEIER	+		
Det (Deterin)			
ENDTATCFVCGK	+	+	
HAPQCEFAK		+	
KLNLLEQHR	+		+
LNLLEQHR		+	
MAEAGFYWTGTK	+	+	+
NLTVSQFLEILGTVVK		+	
RLDEFTR		+	
SWFPETASCSISK		+	
Tum (tumbleweed)/ AcGAP			
AESTIESLPVIAGNER		+	
ALSALASFDDLRR	+		
DFANAVQNPDTK	+	+	
DTLAFLILHFQR	+		
EATAPPLTPVNAMAPHVVAESGTPQLQHR	+		

EHRPSLPK		+	
FGAVGLR		+	
FYGTTPASAHK	+		
GLTEVGLYR	+	+	
IGDGLSSTPR	+	+	
IIDMEIK	+		
IMAVADLLR	+	+	
LAFHLTLPSSR	+	+	
NDLLSLYATPFK	+		
NQIPSVGNK	+	+	
SLTEPLIPTSQWK	+		
SSVGIGVEQHTVDVGQGAER	+	+	
TPATVIK		+	
VTIPQDGQGVIR	+	+	+
YLLTVSCVPQTGTPTTK		+	
Arpc3A (Actin-related protein 2/3 comple, subunit 3A)			
AQGQQDMYSLAISK	+	+	
PQTAQDADLMR	+	+	+
QQVGNMAILPLR	+	+	+
QYLLQLR		+	
VFNTEDGKPNK	+	+	+
Klp61F (Kinesin-like protein at 61F)			
ALKDELQNK	+		
ALSHLFDLR	+	+	
DIEETLSTLEYAHR	+		
EIQTNLQVIEENNQR	+		
FATIIDSSLQSVEEHAK	+		
FIGHATVATDLVQESNR	+	+	
GSVIIQGLEEIPVHSK	+		
IFSEVSMSLVEK		+	
ISDQHSQAFVAK	+	+	+
LLQESLGGR	+	+	
QQLEAVQEK	+		
THTMVGNETAELK		+	
TLVATSPHQEIVR	+	+	
TSIIATISPGHK	+		
VHHNQVEIICQESK		+	
RpL31 (Ribosomal protein L31)			
RVHNIGFK		+	+
SAINEVVTR	+	+	+
RpL30 (Ribosomal protein L30)			
LALVMK		+	+
LVLIASNTPALR	+	+	+
TEVQHYSGTNIELGTACGK		+	

VCTLSITDPGDSDIIR	+	+	
Belle			
FLVLDEADR		+	+
Droj2 (DnaJ-like-2)			
AISQAYEVLSDADKR	+	+	+
DLIVSTQPGEVIR	+	+	+
FFGAGFGGSGGGR	+	+	+
QVYDEGGEEAAIK	+		
Actn (Alpha-actinin)			
AGTAIDNIEEDFR		+	
ALDFIASK		+	
DQTLANELR		+	
FAIQDISVEEMTAK		+	
GKEEMLQSQDFR		+	
HRPDLIDYAK		+	
IDILHLEFAK		+	
ILAADKPYILPDELRL		+	
LETNFNTLQTK	+	+	
LMEEYER		+	
LSNRPAYLPTEGK		+	
NVNVQNFHLSFK	+		
QADNSLAGVQK		+	
YTNYTMETLR	+	+	
RpS20 (40S ribosomal protein S20)			
DIEKPHVGDASVHR		+	+
IIDLHSPSEIVK	+	+	+
Bap55 (Brahma associated protein 55kD)			
NAVLAAFSSGR		+	
SPLGGDFLSR	+	+	
YNVPAFFLVK	+		
RpS18 (Ribosomal protein S18)			
IMNTNIDGK		+	+
SLVIPEKFQHILR			
VGIAMTAIK	+	+	
VPNWFLNR	+	+	+
VVTIISNPLQYK		+	
YWQLTSSNLDSK		+	
Ef1beta (Elongation factor 1 beta)			
APSADNVNVAR	+	+	+
HIASFEEAER		+	+
LVPVGYGINK		+	+
Asp (Protein abnormal spindle)			
DLSLLSPQTK	+		
DTSIQPSVK	+		

EIFLSTIR	+		
ETLQLIDNR	+		
FHAAATVLQK	+		
FIASDALAVLSR	+		
FNLHEVGR	+		
IDGQPHTPLNK	+		
LGLEVVFGEK	+		
LGYVLQHR	+		
LLAAAETAR	+		
LSSLSIQR	+		
SAFEITVTPSR	+	+	
SVVSAAVQQK	+		
TIAVHLQKQ	+		
VLNPSDDDIEVK	+		
VMEVILLR	+		
Tcp-1eta			
LLEIIHPAAK	+		
tacc (transforming acidic coiled-coil protein)			
AIAELISEK	+		
ELPLPDGVSTAGK	+		
LSPLAVKPSPK	+	+	
LTLSALDSTK	+		
TAELEFLALR	+		
VELDQLAK	+		
Scra/ Anil (Actin-binding protein anillin)			
ALGISNASK		+	
ATLEASPAKPLR		+	
DITIPLR		+	
DQLPHELK		+	+
GTTASNSFSFR	+	+	
HLLSADTK		+	
LVMPPVQSPAGPHVVR		+	
MGALYSNTDDLSSPIHR		+	
NFASSAPAPK	+		
PAPAPAVSVK		+	
QACLDEVQR		+	
SPGDAPTTDEDSKR		+	+
TVPTMPGLLSVK	+	+	
VAALVANAQSSAETR		+	
VENSIRPVGAPK		+	
VLSSLEAQGFQR		+	
YHINLNK		+	
YNEHVLATK	+	+	

RpLP2 (Ribosomal protein LP2) ILSSVGVEVDAER	+	+	+
ApoLI/ RfaBp (Retinoid-and fatty acid binding glycoprotein) AISEILEK IAGIVDDINK ISLDHQTFLIK IVNAAEGVTHEFK YSFNNEVSYK	+		
ARP14D/ Arp2 (Actin-related protein 2) DLMVGDEASQLR DMVFIGGAVLAEVTK GYAFNHSADFETVR HIVLGGSTMYPLPSR ILLTEPPMNPTK KMDIDPTNTK LALETTVLVESYTLPDGR MIEVMFEK NVIVCDNGTGFK QEYQEQGLK SLLEVSYPMENGVVR	+	+	
Tcp1-like YISEHLTAPVDELGR	+		

Acknowledgements

I express my sincere gratitude to the following people:

To Jürg, for giving me the opportunity to be part of his research group and do my PhD work in his laboratory, as well as for his supervision, critical discussions and support during my PhD.

To my thesis defense committee Prof. Heinrich Leonhardt, Prof. Nicolas Gompel, Prof. Peter Becker, Dr. Bettina Bölter as well as Dr. Steffen Dietzel and Prof. Barbara Conradt.

To my Thesis Advisory Committee Prof. Heinrich Leonhardt, Prof. Peter Becker and Dr. Wolfgang Zarachia.

To my collaborators Christian Benda and Gabriele Stöhr. Christian Benda for his many explanations about structure biology, patience and time to answer all my questions. Gabriele Stöhr for performing the mass spectrometry for my purifications and spending time on the analysis.

To the crystallization facility at the MPI for their help with crystallization trials and the core facility for their services.

To all members of the Müller lab, for helpful discussions, sharing expertise and creating a friendly and enthusiastic atmosphere in the lab. A special thanks to Katharina for all her help with the Baculo work.

To Ana for enjoying science together and passing through good and tougher times.

To Ira for always being helpful and ready to discuss my work and its obstacles.

To Pierre, for always listening, caring and giving advices.

To my family for being there.

I dedicate my PhD thesis to my family.