Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC

Felice Frey,1 Thomas Sheahan,1 Katja Finkl,1 Gabriele Stoehr,2 Matthias Mann,2 Christian Benda,3 and Jürg Müller1

1Laboratory of Chromatin Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; 2Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; 3Department of Structural Cell Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

Polycomb group (PcG) protein complexes repress transcription by modifying target gene chromatin. In Drosophila, this repression requires association of PcG protein complexes with cis-regulatory Polycomb response elements (PREs), but the interactions permitting formation of these assemblies are poorly understood. We show that the Smbt subunit of the DNA-binding Pho-repressive complex (PhoRC) and the Scm subunit of the canonical Polycomb-repressive complex 1 (PRC1) directly bind each other through their SAM domains. The 1.9 Å crystal structure of the Scm-SAM:Smbt-SAM complex reveals the recognition mechanism and shows that Smbt-SAM lacks the polymerization capacity of the SAM domains of Scm and its PRC1 partner subunit, Ph. Functional analyses in Drosophila demonstrate that Smbt-SAM and Scm-SAM are essential for repression and that PhoRC DNA binding is critical to initiate PRC1 association with PREs. Together, this suggests that PRE-tethered Smbt-SAM nucleates PRC1 recruitment and that Scm-SAM/Ph-SAM-mediated polymerization then results in the formation of PRC1-compacted chromatin.

Keywords: PRC1; PhoRC; SAM domain; Polycomb response element; Drosophila

Supplemental material is available for this article.

Received February 8, 2016; revised version accepted April 13, 2016.

Polycomb group [PcG] proteins are transcriptional regulators that maintain cell fate decisions in animals and plants by repressing transcription of developmental regulator genes in cells where these genes should remain inactive. In Drosophila, 18 different proteins are classified as PcG members because animals lacking any of these proteins show widespread misexpression of Hox and other developmental regulator genes. Biochemical purification of these proteins revealed that these proteins are the subunits of four distinct protein complexes: Polycomb-repressive complex 1 (PRC1), PRC2, Pho-repressive complex (PhoRC), and Polycomb-repressive deubiquitinase (PR-DUB) (Shao et al. 1999, Czermin et al. 2002, Müller et al. 2002, Klymenko et al. 2006, Scheuermann et al. 2010). Protein assemblies identical or related to PRC1, PRC2, and PR-DUB have also been purified from mammalian cells (Cao et al. 2002, Kuzmichev et al. 2002, Levine et al. 2002, Machida et al. 2009, Sowa et al. 2009, Yu et al. 2010, Gao et al. 2012).

PcG protein complexes repress transcription of target genes by modifying their chromatin (for review, see Simon and Kingston 2013). Biochemical and genetic studies have provided compelling evidence that the trimethylation of Lys27 in histone H3 by PRC2 (Cao et al. 2002, Czermin et al. 2002, Kuzmichev et al. 2002, Müller et al. 2002) and the capacity of PRC1 to compact chromatin (Francis et al. 2001, 2004, Grau et al. 2011) are critical for repression of PcG target genes in vivo (King et al. 2005, Eskeland et al. 2010, Isono et al. 2013, Pengelly et al. 2013, McKay et al. 2015). In contrast, the monoubiquitylation of histone H2A by PRC1 (Wang et al. 2004a), long considered to be a critical step for gene silencing by the PcG machinery (e.g., Stock et al. 2007), was recently shown to be dispensable for repression of canonical PRC1 target genes in both Drosophila (Pengelly et al. 2015) and mice (Illingworth et al. 2015).

The molecular interactions that permit PRC1 or PRC2 to associate with specific genomic locations are currently only poorly understood. Studies in mammalian cells have suggested that targeting of these complexes to DNA may entail binding to transcription factors, long noncoding RNAs, nonmethylated CpG dinucleotides in CpG islands, or methylated lysines in histone proteins [for review, see Klose et al. 2013; Simon and Kingston 2013]. Despite this wealth of reported interactions, they may not be sufficient forPCG activity. Genetic and biochemical analyses have not identified the molecular mechanisms underlying PRC1 recruitment, and it is still not known whether the exchange of subunits in the PRC1 complex occurs in vivo [see Klose et al. 2013; Simon and Kingston 2013]. Given recent evidence that Pol II occupancy, histone modifications, or changes in chromatin density may be needed for PRC1 recruitment (Johnson et al. 2011; Simon and Kingston 2013), these issues are likely to be important for understanding PcG function in vivo.
structural information about these is currently available only for the binding of PRC1 and PRC2 subunits to methylated lysine residues in histone proteins [Fischle et al. 2003; Min et al. 2003; Grimm et al. 2007, 2009; Santiveri et al. 2008, Guo et al. 2009, Margueron et al. 2009, Jiao and Liu 2015]. Progress toward elucidating the molecular basis of PcG protein complex targeting has come from studies in Drosophila, where PcG protein complexes assemble at Polycomb response elements [PREs]. PREs typically comprise a few hundred base pairs and contain binding sites for the PhoRC subunit Pho, the only PcG protein with sequence-specific DNA-binding activity [Brown et al. 1998; Kwong et al. 2008; Oktaba et al. 2008; Schuettengruber et al. 2009]. The molecular basis of how Pho binds to PRE DNA is known from the cocrystal structure of the human Pho homolog YY1 bound to its cognate DNA-binding site [Houbaviy et al. 1996], the YY1 residues contacting DNA are 100% conserved in Pho [Brown et al. 1998]. Early studies proposed that the Pho spacer, a region of 30 amino acids that is highly conserved in mammalian YY1, directly interacts with different subunits of PRC1 or PRC2 [Mohd-Sarip et al. 2002, 2005; Wang et al. 2004b]. However, attempts to reconstitute stable Pho:PRC2 or Pho:PRC1 assemblies with recombinant proteins have failed [Mohd-Sarip et al. 2005; Klymenko et al. 2006], and the molecular basis of the proposed interactions of Pho with the diverse PRC2 or PRC1 subunits has remained elusive. Recent structural studies revealed that the Pho spacer forms the interaction domain by which Pho binds its PhoRC partner subunit, Sfbmt [Alfieri et al. 2013]. This finding, together with the observations that Sfbmt is co-bound with Pho at PREs genome-wide [Oktaba et al. 2008] and that Sfbmt association with PREs depends on intact Pho protein-binding sites in vitro and in vivo [Klymenko et al. 2006; Alfieri et al. 2013], suggest that Sfbmt rather than Pho itself may represent the docking platform for interaction with PRC1 and/or PRC2. In support of such a scenario, the PRC1 accessory subunit Scm interacts with recombinant Sfbmt protein in vitro [Grimm et al. 2009], but the molecular basis of this interaction is not known.

Here, we set out to purify proteins that associate with canonical PRC1 in Drosophila embryos and identified PhoRC as a major interactor of this complex. We found that the SAM domains of Sfbmt and Scm mediate the interaction between the two complexes and determined the crystal structure of the Sfbmt-SAM:Scm-SAM complex to reveal the recognition mechanism. Functional tests in Drosophila show that the Sfbmt-SAM:Scm-SAM interaction is critical for long-term PcG repression of target genes and that PRC1 association with PREs requires DNA-bound PhoRC. Together, these studies thus reveal the molecular basis of how PhoRC targets PRC1 to PREs.

Results

Biochemical purification identifies PhoRC as an interaction partner of PRC1

To identify proteins that associate with the PRC1 subunit polyhomeotic-proximal [Ph-p] or its paralog, polyhomeotic-distal [Ph-d], we performed tandem affinity purification [TAP] [Rigaut et al. 1999] on nuclear extracts from transgenic Drosophila embryos expressing TAP-tagged Ph-p or Ph-d, respectively [Supplemental Fig. S1A]. The TAP-Ph-p and TAP-Ph-d fusion proteins were expressed at levels comparable with endogenous Ph-p and Ph-d proteins [Supplemental Fig. S1B], and either protein was able to rescue the severe phenotype of ph0 mutant embryos [Supplemental Fig. S1C], demonstrating that the fusion proteins could functionally substitute Ph-p and Ph-d. Mass spectrometric analyses of the purified material identified the PRC1 core subunits Psc, Su[z]2, Sce, and Pc and, intriguingly, also the PhoRC subunit Sfbmt as possible interaction partners of either Ph-p or Ph-d [Supplemental Fig. S1D, Supplemental Table S1]. Western blot analysis confirmed that PRC1 and Sfbmt are enriched in TAP-Ph-p and TAP-Ph-d preparations and revealed that the purified assembly also contains the PRC1 accessory subunit Scm (Fig. 1A). The finding that PhoRC subunits are associated with PRC1 in Drosophila nuclear extracts is consistent with earlier studies that identified PhoRC subunits in
Pc protein assemblies purified from *Drosophila* embryonic nuclear extracts (Strübbe et al. 2011) and in Pc and Scm chromatin assemblies purified from embryos after cross-linking [Kang et al. 2015]. Moreover, this association appears to be conserved in vertebrates where canonical PRC1 subunits were identified in purifications of Sfmbt1 and Sfmbt2 from human cells [Zhang et al. 2013].

**Sfmbt and Scm interact through the C-terminal SAM domains**

We next wanted to identify the molecular basis of the PRC1–PhoRC interaction. We previously found that, upon coexpression in insect cells, Scm and Sfmbt can be isolated as a stable dimeric complex [Grimm et al. 2009]. Scm itself associates with PRC1 by binding to Ph, an interaction that is mediated by the SAM domains that are present in the C termini of both proteins (Fig. 1B; Peterson et al. 1997, 2004; Kim et al. 2005). To test whether Sfmbt may also interact with Ph, we coexpressed Sfmbt with Flag-tagged Ph-Δ1298–1589, a C-terminal fragment of Ph-p that has high sequence identity with Ph-d and contains the three structural domains in Ph proteins (Fig. 1B; Supplemental Fig. S1A). However, Flag affinity purification resulted in the isolation of Flag-Ph–Δ1298–1589 alone (Fig. 1C, lane 1), suggesting that Sfmbt and Ph-p do not interact directly. In contrast, coexpression of Scm together with Sfmbt and Flag-Ph-Δ1298–1589 protein permitted isolation of a stable complex containing all three proteins (Fig. 1C, lane 2). This suggested that Scm, binding to both Ph and Sfmbt, may act as the physical link that mediates the interaction between PRC1 and PhoRC.

These results prompted us to examine how Sfmbt and Scm bind to each other. Initial studies using C-terminally truncated Sfmbt and Scm proteins had suggested that the two proteins interact through poorly defined regions in their N termini [Grimm et al. 2009]. However, we found that a Flag-SfmbtΔ530–1220 protein that lacks the previously described N-terminal Scm-interacting region is also able to form a stable complex with Scm (Fig. 1C, lane 4), suggesting an additional Scm-interacting region in the C terminus of Sfmbt. Deletion of the C-terminal SAM domain in Sfmbt substantially reduced the ability of this truncated Flag-SfmbtΔ530–1136 protein to bind to Scm (Fig. 1C, lane 3). This suggested that the Sfmbt SAM domain is important for binding to Scm.

Previous structural studies showed that the SAM domains of Scm and Ph both contain a mid-loop (ML) and an end helix (EH) surface [Fig. 2A] and have the propensity to form homopolymers or heteropolymers in a head-to-tail fashion where the EH surface of one SAM domain interacts with the ML surface of the SAM domain-binding partner [Kim et al. 2002, 2005]. Work on Scm-SAM:Ph-SAM heteropolymers revealed that this assembly occurs in a defined orientation with preferential binding of the Scm-SAM EH surface to the Ph-SAM ML surface [Kim et al. 2005]. Intriguingly, we found that recombinant Sfmbt-SAMΔ1137–1220 (referred to as Sfmbt-SAM), if expressed alone, eluted in a single peak corresponding to a monomer when purified on a gel filtration column [Supplemental Fig. S2A]. The Sfmbt-SAM domain thus lacks the capacity to form homopolymers. We next tested whether Sfmbt-SAM might bind to the Scm-SAM domain. Considering that Sfmbt, Scm, and Ph can form a trimeric complex and that the EH surface of Scm-SAM engages in binding to Ph-SAM, we hypothesized that Sfmbt might bind to the ML surface of Scm-SAM. In a first set of pull-down experiments, we found that Scm-SAMΔ803–877 in which the EH surface had been mutated [Scm-SAMΔ855E/L859E] to prevent the formation of Scm-SAM homopolymers bound to GST-Sfmbt-SAM [Supplemental Fig. S2B]. In contrast, Ph-SAMΔ1561E/L1565E containing the corresponding EH surface mutations to prevent Ph-SAM homopolymer formation interacted only very poorly with GST-Scm-SAM in the same assay [Supplemental Fig. S2B]. We then coexpressed Sfmbt-SAM and Scm-SAMΔ855E/L859E and found that the Scm-SAM:Scm-SAMΔ855E/L859E complex [referred to here as Scm-SAM:Scm-SAM] eluted as a single peak corresponding to a dimer on a gel filtration column (Fig. 2B; Supplemental Fig. S2A), suggesting that these two SAM domains form a stable complex.

**Structure of the Sfmbt-SAM:Scm-SAM complex**

We obtained diffracting crystals of the Sfmbt-SAM:Scm-SAM dimer. We solved the structure by molecular replacement with Ph-SAM:Scm-SAM as a search model and refined it to 1.9 Å resolution [R Fischer of 27.7% and Rwork of 25.5%]. The asymmetric unit contained two Scm-SAM dimers that interacted laterally. This lateral interaction between the two dimers is likely imposed by crystal packing because Sfmbt-SAM:Scm-SAM behaves as a heterodimer in gel filtration chromatography even at high concentrations (Fig. 2B), and, moreover, one of the two interactions between the dimers is a salt bridge formed by the mutated ScmΔE599 residue [data not shown]. In the following, we thus consider the Sfmbt-SAM:Scm-SAM complex to exist as a dimer.

The structure revealed that the Sfmbt-SAM domain is a helical bundle with the canonical SAM domain fold [Fig. 2C]. The major interaction in the dimer is between the EH surface of Sfmbt-SAM and the ML surface of Scm-SAM and covers an area of ~550 Å2 [calculated with PDBePISA [proteins, interfaces, structures, and assemblies]] [Fig. 2C]. The interface comprises a hydrophobic region and an adjacent polar region. The hydrophobic interactions involve the side chains of A838, L841, L842, M846, and Y850 on the Scm ML surface and M1180, V1187, G1188, and L1191 on the Sfmbt EH surface (Fig. 2C, top panel). The adjacent polar interactions are formed by three salt bridges (E833Scm–K1186Smbt, D835Scm–K1192Smbt, and K849Scm–D1177Smbt) and hydrogen bonds between backbone [H832Scm–O1–G1188Smbt] and side chains [Y850Scm–D1177Smbt] [Fig. 2C, right, bottom panel]. Two water molecules bridging Y850Scm with D1177Smbt and E833Scm with K1192Smbt further stabilize the interface [data not shown].
The ML surface of Sfmbt-SAM lacks the capacity for forming SAM–SAM interactions

Inspection of the ML surface of the Sfmbt-SAM domain explains why this domain lacks polymerization capacity. In Sfmbt, a number of residues at positions corresponding to the apolar residues in the Scm ML surface are substituted with polar residues. In particular, R1170 in the ML surface of Sfmbt-SAM, corresponding to a conserved Ala in Scm-SAM and Ph-SAM (Fig. 2A), would block interaction with the EH surface of another SAM domain. This provides a likely explanation of why Sfmbt-SAM can engage in interaction with only a single SAM domain partner via the EH surface.

Molecular basis for the formation of extended Sfmbt–Scm–Ph assemblies

Because Scm-SAM adopts the same conformation in Sfmbt-SAM:Scm-SAM and Scm-SAM:Ph-SAM [Protein Data Bank [PDB] ID 1PK1], the two complexes can easily be superimposed via their common Scm-SAM domains [root mean square deviation [RMSD] of 0.42 over 67 Ca]. This superposition permitted us to generate a model of the trimeric Sfmbt-SAM:Scm-SAM:Ph-SAM complex [Fig. 2D]. In support of this, we reconstituted a complex containing all three SAM domains by coexpressing Sfmbt-SAM with an Scm-SAM–Ph-SAM fusion protein [Supplemental Fig. S2C]. On a gel filtration column, the major peak contained stoichiometric amounts of the two polypeptides [Supplemental Fig. S2C]. We presume that this defined complex containing all three SAM domains is formed because the fusion protein promotes the intramolecular interaction of Scm-SAM with Ph-SAM and thus effectively limits the oligomerization capacity of Scm-SAM, leaving its ML surface available for interaction with Sfmbt-SAM.

In each of the three proteins, the SAM domain is located at the very C terminus and is separated from adjacent domains by linker regions that are predicted to be mostly...
disordered [Fig. 1B]. Therefore, even though the SAM polymer assumes a rigid conformation, these adjacent linker sequences likely allow flexibility for orientation of the rest of each protein. It is important to recall that both Scm-SAM and Ph-SAM have the capacity to form homopolymers or heteropolymers. PRE-tethered Sfmbt, lacking this polymerization capacity, may therefore act as an assembly platform to initiate formation of longer polymers containing Scm and/or Ph proteins that emanate from PREs.

The SAM domains of Sfmbt and Scm are essential for target gene repression

We next investigated the requirement of the Sfmbt-SAM and Scm-SAM domains for transcriptional repression of PcG target genes in Drosophila. In a first set of experiments, we tested whether SfmbtΔSAM lacking the SAM domain can replace the endogenous Sfmbt protein in a genetic rescue assay in larvae, as follows: In the wing imaginal disc, clones of cells that are homozygous for the SfmbtΔ-null mutation fail to maintain PcG repression of the Hox gene Ultrabithorax (Ubx), and the Ubx protein becomes strongly misexpressed in the mutant clones [Fig. 3A, left]. When such SfmbtΔ mutant clones were induced in animals carrying a transgene with a genomic fragment expressing wild-type Sfmbt protein, the transgene-encoded Sfmbt protein fully rescued repression, and no Ubx protein was detected in the SfmbtΔ homozygous cells [Fig. 3A, middle]. In contrast, the SfmbtΔSAM protein expressed from the same genomic fragment was inefficient in rescuing repression, and the Ubx protein was strongly misexpressed in a large fraction of SfmbtΔ homozygous cells [Fig. 3A, right]. Importantly, the truncated SfmbtΔSAM protein was stable and present at levels comparable with that of the wild-type Sfmbt protein [Fig. 3B]. Together, these results show that the SAM domain is critical for Sfmbt to function in PcG repression. The comparison of the repression defects in SfmbtΔSAM and Sfmbt-null mutant clones nonetheless suggests that the SfmbtΔSAM protein retains partial repressor function [Fig. 3A], suggesting that deletion of the SAM domain does not completely incapacitate the protein.

Figure 3. SAM domains of Sfmbt and Scm are required for PcG repression. (A) Imaginal wing discs with clones of SfmbtΔ homozygous cells from animals that carried no transgene [no TG] or the indicated genomic transgenes expressing wild-type Sfmbt (Sfmbtwt) or SfmbtΔSAM stained with antibody against Ubx protein [red]. Clones of SfmbtΔ homozygous cells are marked by the lack of GFP [green] and were induced 72 h before analysis. Only SfmbtΔ homozygous cells [no TG] in the wing pouch [arrowheads] but not in the notum or hinge show strong misexpression of Ubx [Klymenko et al. 2006], and rescue of Ubx repression by Sfmbtwt or SfmbtΔSAM was therefore only analyzed in clones in the wing pouch area. In no TG animals, 98.4% of SfmbtΔ homozygous clones [n = 63 clones] showed misexpression of Ubx [arrowheads]. In animals carrying Sfmbtwt, repression of Ubx was rescued in most SfmbtΔ homozygous clones [empty arrowheads], and only 1% of clones [n = 104 clones] showed misexpression of Ubx. In animals carrying SfmbtΔSAM, 55% of SfmbtΔ homozygous clones [n = 169 clones] showed misexpression of Ubx [arrowheads]. Thus, even though the SfmbtΔSAM protein largely fails to rescue, the repression of Ubx in 45% of clones [empty arrowheads] suggests that the SfmbtΔSAM protein retains some repressor activity. An asterisk marks normal Ubx expression in a trachea attached to the disc. (B) Western blots on serial dilutions [9:3:1] of extracts from whole nuclei of 0- to 12-h-old embryos of the indicated genotypes probed with antibodies against Sfmbt and, as a loading control on the same membrane, Lamin. The asterisk marks normal Ubx expression in a trachea attached to the disc. (C) Analysis of Ubx repression as in A but in clones of ScmΔF homozygous cells in animals carrying no transgene [no TG] or transgenes expressing Flag-tagged wild-type Scm [F-Scmwt] or F-ScmΔSAM. ScmΔF homozygous clones marked by the lack of GFP [green] were induced 72 h before analysis. Only clones in the pouch were analyzed for statistics. In no TG animals, 100% of ScmΔF homozygous clones [n = 94 clones] showed misexpression of Ubx [arrowheads]. F-Scmwt rescued repression of Ubx in all clones [empty arrowheads], and none of the clones [n = 42 clones] showed misexpression of Ubx, consistent with the lack of a phenotype in Scm-null mutant adults that are rescued by this transgene [Peterson et al. 2004]. In animals carrying F-ScmΔSAM, 100% of ScmΔF homozygous clones [n = 130 clones] showed misexpression of Ubx [arrowheads]. See also Supplemental Figure S3. (D) Western blots on serial dilutions [9:3:1] of total extracts from 14- to 18-h-old embryos of the indicated genotypes probed with antibodies against Scm and, as a loading control on the same membrane, Caf1. An asterisk marks normal Scm protein that persists in ScmΔF homozygous embryos, in lanes 7–12, this band is partially obscured [l*] by the bands from the transgene-encoded F-Scmwt and F-ScmΔSAM proteins [top and bottom arrows, respectively].
Next, we assessed the requirement of the Scm-SAM domain. Previous studies showed that a transgene containing a genomic Scm fragment rescues Scm-null mutant animals into viable and fertile adults but that an Scm\textsuperscript{SSAM} protein expressed from the same genomic fragment fails to rescue viability of these animals [Peterson et al. 2004]. To extend this finding, we tested the capacity of the Scm\textsuperscript{SSAM} protein to repress PcG target genes in clones of Scm-null mutant cells. Only the transgene expressing wild-type Scm protein was able to rescue repression of Ubx in clones of cells that were homozygous for the Scm\textsuperscript{H1}-null mutation [Fig. 3C]. The Scm\textsuperscript{SSAM} protein completely lacked Scm repressor activity, and the Ubx protein was as widely misexpressed as in Scm\textsuperscript{H1} mutant clones in animals carrying no transgene [Fig. 3C]. In parallel, we also assayed the rescue capacity of the Scm\textsuperscript{SSAM} protein in embryos that were homozygous for the Scm\textsuperscript{H1} protein in embryos that were homozygous for Scm\textsuperscript{H1} (Supplemental Fig. S3). Only the transgene-encoded wild-type Scm protein, but not the Scm\textsuperscript{SSAM} protein, was able to rescue repression of the Hox gene Abdominal-B (Abd-B) in Scm\textsuperscript{H1} homozygous embryos [Supplemental Fig. S3]. As previously reported [Peterson et al. 2004], we found that the Scm\textsuperscript{SSAM} protein is stable and expressed at levels comparable with that of the wild-type Scm protein [Fig. 3D]. Together, these results show that the Scm-SAM domain is essential for Scm to function in PcG repression.

Finally, we note that similar structure/function analyses of the Ph protein in Drosophila recently demonstrated that the Ph SAM domain is essential for all functions of Ph [Gambetta and Müller 2014]. Taken together, these data thus show that deletion of the SAM domain in Sfmbt, Scm, or Ph in each case ablates protein function.

**PhoRC DNA-binding is critical for PRC1 recruitment to a PRE**

We next investigated the requirement of PhoRC for recruitment of PRC1 to PREs in developing Drosophila. A number of limitations precluded us from monitoring PRC1 binding in Sfmbt or pho phol mutant animals by chromatin immunoprecipitation (ChiP) assays. First, it is not possible to generate embryos lacking maternally deposited PhoRC because the complex is essential for germ cell development [Breen and Duncan 1986; Girton and Jeon 1994; Brown et al. 2003; Klymenko et al. 2006]. Consequently, in Sfmbt or pho phol zygotic mutant embryos, maternally deposited wild-type PhoRC permits almost normal establishment of PcG repression [Simon et al. 1992; Brown et al. 2003; Klymenko et al. 2006]. Second, as the Sfmbt and pho phol zygotic mutant animals that develop into larvae are becoming depleted of maternally supplied PhoRC, the diploid larval cells begin to show defects in proliferation and form only poorly developed central nervous system (CNS) and rudimentary imaginal disc tissues [Brown et al. 2003; Klymenko et al. 2006], making these tissues unreliable material for ChiP analysis.

As an alternative approach to analyze PRC1 recruitment to PREs in the absence of PhoRC, we analyzed binding at a PRE transgene with mutated Pho/Phol-binding sites. PRE\textsubscript{PRED}, a 570-base-pair (bp)-long fragment containing the core of the bxd PRE in Ubx, contains six Ph protein-binding sites [Chan et al. 1994; Fritsch et al. 1999]. Previous studies in embryos showed that Pho, Sfmbt, and Ph bind to the wild-type PRE\textsubscript{PRED} in a transgene but that this binding was strongly reduced at a PRE\textsubscript{PRED pho mut} transgene in which the Pho-binding sites had been mutated [Fritsch et al. 1999; Klymenko et al. 2006]. Here, we performed ChiP assays in imaginal wing discs from PRE\textsubscript{PRED} or PRE\textsubscript{PRED pho mut} transgenic larvae to monitor binding of the Pho, Sfmbt, Scm, and Ph proteins in the same tissues where we investigated the function of the Sfmbt-SAM and Scm-SAM domains [Fig. 3]. In addition, we also analyzed binding of the Trithorax (Trx) protein, the PRC2 core subunit E[z], and the levels of the H3K27me3 mark. Binding of each of these proteins at the native bxd PRE in Ubx, at the iab-7 PRE in Abd-B, and at other genomic regions not bound by PcG proteins served as internal controls in the two transgenic lines [Fig. 4A]. The Pho, Sfmbt, Scm, and Ph proteins were all bound at the wild-type PRE\textsubscript{PRED} transgene, but, for each protein, only very low-level binding was detected at the PRE\textsubscript{PRED pho mut} transgene [Fig. 4A]. In contrast, binding of the Trx protein was comparable at the PRE\textsubscript{PRED} and PRE\textsubscript{PRED pho mut} transgenes [Fig. 4A]. E[z] and H3K27me3 were strongly enriched at the PRE\textsubscript{PRED} transgene and drastically reduced at the PRE\textsubscript{PRED pho mut} transgene, consistent with earlier findings that suggested that Pho is required for association of PRC2 with PREs [Wang et al. 2004b]. In summary, these analyses show that recruitment of PRC1 and also PRC2 to the bxd PRE critically depends on the ability of PhoRC to bind to Pho DNA-binding sites.

**Discussion**

Atomic-level information on the molecular interactions by which PcG protein complexes bind to the genes that they regulate is essential for understanding how the PcG system works. In Drosophila, the different PcG protein complexes assemble at PRE sequences in target genes. Previous structural studies showed how YY1, the mammalian ortholog of the PhoRC subunit Pho, recognizes its cognate DNA-binding site, providing an atomic model of how PhoRC binds to PRE DNA [Fig. 4B]. More recent studies then revealed how Sfmbt interacts with the Pho spacer to form PhoRC [Fig. 4B]. Here, we present the structural basis of how PhoRC binds to canonical PRC1 and provide functional evidence that this PhoRC–PRC1 interaction targets PRC1 to PRE DNA [Fig. 4].

The following main conclusions can be drawn from the work presented in this study. First, biochemical purification of canonical PRC1 from nuclear extracts of Drosophila identifies PhoRC but no other DNA-binding proteins among the most highly enriched PRC1 interactors. This observation is consistent with previous studies [Strübbe et al. 2011] and suggests that, in Drosophila, PhoRC is the main PRC1 interaction partner with sequence-specific DNA-binding activity. Second, biochemical reconstitution shows that the PRC1 subunit Scm and
The PRC1 bind to each other in the nucleoplasm independently of the Sfmbt interaction. PhoRC constitutes the binding platform for PRC1 association with intact DNA-binding sites for Pho, suggesting that Drosophila prepared by high-salt extraction from embryonic nuclei of PRC1 to a Hox gene PRE in this PRE because deletion of the Sfmbt-SAM and Scm-SAM complex reveals the recognition mechanism for repression of PcG target genes, providing functional evidence for the importance of the Sfmbt-SAM:Scm-SAM interaction. Fifth, binding of both PhoRC and PRC1 to a Hox gene PRE in Drosophila critically depends on intact DNA-binding sites for Pho, suggesting that PhoRC constitutes the binding platform for PRC1 association with this PRE.

The Sfmbt–Scm interaction links PhoRC and PRC1

The PRC1–PhoRC assemblies isolated in this study and by Strübbe et al. (2011) were identified in affinity purifications of PRC1 from soluble nuclear extracts that had been prepared by high-salt extraction from embryonic nuclei of Drosophila. Although it is possible that PRC1 and PhoRC bind to each other in the nucleoplasm independently of DNA, we consider it more likely that the PRC1-associated PhoRC in the purified material represents PRC1–PhoRC assemblies that had formed at target gene DNA and were solubilized from chromatin during extract preparation. This scenario is also supported by ChIP profiling and proteomic analyses on cross-linked chromatin that showed that PhoRC and PRC1 colocalize at a large fraction of genomic target sites (Kwong et al. 2008; Oktaba et al. 2008; Schuettengruber et al. 2009; Kang et al. 2015).

The requirement of the Sfmbt-SAM and Scm-SAM domains for target gene repression supports the importance of their interaction. Nevertheless, we found that the SfmbtSAM protein is not completely incapacitated for repression [Fig. 3A]. It is possible that interaction of SfmbtSAM and Scm through a second interaction site in their N termini (Grimm et al. 2009) accounts for the repression in a fraction of SfmbtSAM mutant cells. The observation that the ScmSAM protein is nonetheless completely incapacitated for repression [Fig. 3C; Supplemental Fig. S3] is not at odds with this interpretation because deletion of the SAM domain in Scm not only compromises Scm binding to Sfmbt but also prevents Scm binding to the Ph SAM domain and thus association with PRC1 (Fig. 4B; Peterson et al. 2004; Kim et al. 2005).
In this scenario, the Scm-SAM domain provides the critical physical link that mediates the interaction between PhoRC and PRC1 [Fig. 4B]. It is possible that other, unidentified DNA-binding proteins are also able to recruit Scm to PREs (Wang et al. 2010). However, at least at the bxd PRE, the PhoRC–PRC1 link reported here appears to be critical for recruitment of PRC1 because binding of both Scm and Ph is strongly reduced if PhoRC binding to PRE DNA is compromised [Fig. 4A].

**PRE-tethered Sfmbt-SAM:Scm-SAM as a nucleation site for short- and long-range chromatin organization**

Current models suggest that PRC1 represses target gene transcription by altering chromatin organization at two different levels. First, Psc and CBX2 in Drosophila and vertebrate PRC1, respectively, act at the level of nucleosomes by compacting arrays of nucleosomes and inhibiting their remodeling in vitro (Francis et al. 2001, 2004; Grau et al. 2011). This activity relies on low-complexity regions in the Psc and CBX2 proteins that are rich in basic amino acids, and, in the case of Psc, these regions were shown to be essential for target gene repression in Drosophila (King et al. 2005). The PRC1 subunit Ph, on the other hand, appears to alter chromatin at a higher level of organization by changing chromatin topology in part by forming long-range contacts between distant Ph-bound chromosomal sites (Cheutin and Cavalli 2012; Isono et al. 2013; Boettiger et al. 2016; Wani et al. 2016). This activity requires the Ph-SAM domain and its ability to form homo-oligomers or hetero-oligomers (Isono et al. 2013; Wani et al. 2016). The capacity of Ph-SAM to form oligomers is also strictly required for repression of PcG target genes in Drosophila (Gambetta and Müller 2014). The Sfmbt-SAM:Scm-SAM interaction reported here may thus provide the molecular basis toward understanding not only how PRE-tethered PRC1 locally compacts arrays of nucleosomes in the chromatin flanking PREs but also how it interacts with other PRC1 assemblies bound at distant chromosomal sites to form topologic domains of PcG-repressed chromatin.

**Materials and methods**

**TAP of Ph complexes**

The α-tubulin1-TAP-Ph-p and a-tubulin1-TAP-Ph-d transgenes in the previously described Drosophila transformation vector CaSpeR-NTAP (Nekrasov et al. 2007) contained a 2.6-kb fragment of the a-tubulin1 gene, including promoter and 5' untranslated region sequences, followed by the TAP tag and either the Ph-pΔ1-1599 or Ph-dΔ1-1531 coding regions (plasmid maps are available on request). TAP from 0- to 14-h-old embryonic nuclear extracts was performed as described (Klymenko et al. 2006).

**Mass spectrometric analysis of protein isolated by TAP**

We performed two independent purifications from extracts of TAP-Ph-d transgenic embryos and one from TAP-Ph-p transgenic embryos, in each case, a mock purification from wild-type embryos was performed in parallel. Calmodulin eluates were analyzed by SDS-PAGE, and proteins were visualized by Coomassie staining. For each sample, lanes were cut into five equal gel slices, and peptides were generated by tryptic in-gel digestion (Shevchenko et al. 2006).

Desalted peptide samples were analyzed using liquid chromatography-tandem mass spectrometry ([LC-MS/MS] on a LTQ Orbitrap Velos instrument online coupled to an Easy-nLC system [both Thermo Fisher Scientific]). Samples were analyzed using 65- and 90-min reversed-phase chromatography, respectively. Settings were set to 200 nL/min flow and gradient from 5% to 30% solvent B [80% ACN, 0.5% AcOH] for 43 and 65 min, respectively. Mass spectrometric settings were as follows: full scan: 10^6 ions, 100-msec maximum filling times; MS2 scan: 4 × 10^6 ions, 150-msec maximum filling times, isolation width: 4; minimal signal required: 10^5; and NCE: 40. The 10 most intense ions were fragmented using HCD mode and fixed at m/z 100. Unassigned charges and singly charged ions were rejected from MS2 fragmentation. Dynamic exclusion was enabled.

All .RAW files were analyzed together using the MaxQuant software suite, including the Andromeda search engine (version 1.5.0.0), (Cox and Mann 2008; Tyanova et al. 2014). The analysis was performed on tryptic peptides, including fixed modifications of carboxymethylation (Cys) and variable modifications of N-terminal acetylation and oxidation of methionine. Files were searched against the UniProt database of Drosophila (18,796 entries) as well as the self-defined FASTA sequences of Ph-d and Ph-p. Peptide, protein, and site false discovery rates were set to 0.01. Label-free algorithm (Cox et al. 2011) was included, and the “match between runs” function was enabled.

Data were filtered and further analyzed using the in-house software tool Perseus. Intensity values were filtered for contaminant and reverse entries. The three experiments were split into sub-files, and each experiment was filtered for one valid value. Missing values of the LFQ intensities were imputed using a normal distribution at the lower-intensity range. Values are represented using the free software environment R [http://www.r-project.org].

A detailed list of peptide sequences obtained from MS analysis of the proteins is presented in Supplemental Table S1.

**Protein expression and purification**

Baculoviruses expressing Sfmbt, Scm, and Flag-PhΔ1299–1589 have been described previously [Klymenko et al. 2006; Grimm et al. 2009; Gambetta and Müller 2014]. For this study, new viruses for Flag-SfmbtΔ130–1220 and Flag-Scm-SAMΔ137–1220 expression were generated after cloning the appropriate coding fragments into pFastBac. Flag affinity purification of protein [complexes] was performed as described [Grimm et al. 2009] with the following modifications: Instead of whole-cell extracts, cytosolic, nuclear, and chromatin extract were prepared in extraction buffer (25 mM HEPES at pH 7.9, 150 mM NaCl, 20 µM ZnCl2, 10% glycerol, 0.5 mM DTT, 0.1% pefabloc-SC, 1× Complete protease inhibitor [Roche] and subsequently combined to get the maximal amount of recombinant protein for purification. Insect cell extracts were precleared with mouse IgG-agarose beads (Sigma) for 45 min at 4°C on a rotating wheel prior to incubation with anti-Flag M2 agarose beads (Sigma) overnight at 4°C. For 12 mL of extracts, 0.3 µL of anti-Flag beads was used.

For GST pull-downs, GST-His-Sfmbt-SAMΔ137–1220, His-Scm-SAMΔ122–877, and His-Ph-SAMΔ877–1220 were cloned into pET-derived vectors, individually expressed in *Escherichia coli* [Rosetta [DE], Novagen], and purified by nickel affinity purification. The purified GST-His-Sfmbt-SAMΔ137–1220 domain was mixed with His-Scm-SAMΔ122–877 or His-Ph-SAMΔ877–1220, coupled to glutathione beads for 2 h in binding buffer [50 mM phosphate buffer at pH 7.5, 250 mM
NaCl, 10 mM MgCl₂, 10% glycerol, 0.01% IGEPA-LCA-630, 1 mM pefablok-CS, 4 mM dithiothreitol [DTT], 1× Complete protease inhibitor), and then washed three times with the same buffer but containing 200 μM instead of 250 mM NaCl. Beads were then resuspended in 150 μL of 1× NuPAGE LDS sample buffer and incubated for 5 min at 95°C prior to SDS-PAGE. Binding assays with purified GST-His-GFP served as control.

The Smbt-SAM:Scm-SAM complex was obtained by cloning Smbt-SAM₁₁₃₇₋₁₂₂₀ and Scm-SAM₈₀₃₋₈₇₇ into pET-derived vectors and coexpression in E. coli (Rosetta [DE], Novagen) as GST-His fusion and His fusion proteins, respectively. One liter of expression culture was resuspended in 25 mL of lysis buffer [50 mM phosphate buffer at pH 7.5, 250 mM NaCl, 10 mM MgCl₂, 25 mM imidazole, 10% glycerol, 0.1% Triton-X100, 1 mM pefablok-CS, 2 mM β-mercaptoethanol, 1× Complete protease inhibitor], and cells were disrupted by sonication. SAM domains were purified by nickel affinity, GST affinity, ion exchange, and size exclusion chromatography. After the last purification step, samples were concentrated with Amicon Ultra centrifugal filters [3-kDa cutoff, Millipore] to 7–15 mg/mL.

The same strategy was used to purify the Smbt-SAM₁₁₃₇₋₁₂₂₀-Scm-SAM₈₀₃₋₈₇₇-Ph-SAM₁₄₉₉₋₁₅₇₇ complex [Supplemental Fig. S2C].

Crystallization and X-ray structure determination

All diffraction data were collected at the Swiss Light Source (SLS) synchrotron facility at beamline PXII. The data were processed with XDS (Kabsch 2010). All crystal structures were solved using Phaser from the PHENIX suite [Adams et al. 2010]. The atomic models were built with Coot (Emsley et al. 2010) and refined using either the PHENIX suite [Adams et al. 2010] or the CCP4 suite (Winn et al. 2011). Validation was performed with MolProbity [Davis et al. 2007]. Figures were made with PyMOL (version 1.2).

Crystals of the Smbt-SAM:Scm-SAM heterodimeric complex were grown at 4°C in 0.05 M Tris/HCl (pH 7.5), 4% MPD, 0.2 M ammonium acetate, and 32.5% PEG3350. They contained two copies of the complex per asymmetric unit (space group P1 2₁ 2₁ 1). A complete data set of an Sfmbt-SAM:Scm-SAM crystal, flash-frozen in liquid nitrogen, was collected to a resolution of 1.975 Å. The structure was solved by molecular replacement with the Ph-SAM:Scm-SAM complex (PDB ID 5J8Y) as a search model. Interaction surfaces were calculated with the PDBePISA (proteins, interfaces, structures, assemblies) service at the European Bioinformatics Institute [http://www.ebi.ac.uk/pdbe/prot_int/pstart.html; Krissinel and Henrick 2007].

Accession Numbers

The accession number for the coordinates and structure factors reported in this study is PDB ID 5J8Y.

Acknowledgments

We thank the Max Planck Institute Crystallography Facility for crystal screening and optimization, and the beamline scientists at Swiss Light Source for excellent assistance with data collection. We are most grateful to Jeffrey Simon and Renato Paro for the generous gift of Drosophila strains and antibodies. This work was supported by the European Commission Seventh Framework Program 4DCellFate [grant no. 277899], the Deutsche Forschungsgemeinschaft (SFB1064) and the Max-Planck Society. F.F., C.B., and J.M. conceived the project, designed the experiments, discussed and interpreted the data, and prepared the manuscript. T.S. performed the ChIP analysis. K.F. provided technical support. G.S. and M.M. performed the mass spectrometric analysis.

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, ...
et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221.

Alfieri C, Gambetta MC, Matos R, Glatt S, Sehr P, Fraterman S, Wilm M, Müller J, Müller CW. 2013. Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. Genes Dev 27: 2367–2379.

Beuchle D, Struhl G, Müller J. 2001. Polycomb group proteins and heritable silencing of Drosophila Hox genes. Development 128: 993–1004.

Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu C-T, Zhuang X. 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529: 418–422.

Breen TR, Duncan IM. 1986. Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Dev Biol 118: 442–456.

Brown JL, Mucci D, Whiteley M, Dirksen ML, Kassis JA. 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 JL, Fritsch C, Mueller J, Kassis JA. 2003. The Drosophila ph-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development 130: 285–294.

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

Chan CS, Rastelli L, Pirrotta V. 1994. A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J 13: 2553–2564.

Cheutin T, Cavalli G. 2012. Progressive polycomb assembly on pre-compact chromatin repressor Sfmbt to Polycomb response elements. Development 139: W375–W385.

Eskeland R, Lee CH, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoulitchi AI, Wutz A, et al. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell 38: 452–464.

Francis NJ, Saurin AJ, Shao Z, Kingston RE. 2001. Reconstitution of a functional core polycomb repressive complex. Mol Cell 8: 545–556.

Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a polycomb group protein complex. Science 306: 1574–1577.

Fritsch C, Brown JL, Kassis JA, Müller J. 1999. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development 126: 3905–3913.

Gambetta MC, Müller J. 2014. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. Dev Cell 31: 629–639.

Gao Z, Zhang J, Bonasio R, Stringo F, Sawal A, Parisi F, Kluger Y, Reinberg D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45: 344–356.

Girton JR, Jeon SH. 1994. Novel embryonic and adult homeotic phenotypes are produced by pleiohomeotic mutations in Drosophila. Dev Biol 161: 393–407.

Grau DJ, Chapman BA, Garlick JD, Borowsky M, Francis NJ, Kingston RE. 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev 25: 2210–2221.

Grimm C, de Ayala Alonzo AG, Rybin V, Steuerwald U, Ly-Hartig N, Fischle W, Müller J, Müller CW. 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, Müller CW. 2009. Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. EMBO J 28: 1965–1977.

Guo Y, Nady N, Qi C, Allali-Hassani A, Zhu H, Pan P, Adams-Cioaba MA, Amaya MF, Dong A, Vedadi M, et al. 2009. Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. Nucleic Acids Res 37: 2204–2210.

Houbavib H, Ushave A, Shenk T, Burley SK. 1996. Corecrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc Natl Acad Sci USA 93: 13577–13582.

Il lingworth RS, Mofat M, Mann AR, Read D, Hunter CJ, Pradepa MM, Adams IR, Bickmore WA. 2015. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev 29: 1897–1902.

Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE, Koseki H. 2013. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev Cell 26: 565–577.

Jiao L, Liu X. 2015. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science 350: aac4383.

Kabsch W. 2010. XDS. Acta Crystallogr D Biol Crystallogr 66: 125–132.

Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, Kuroda MI. 2015. Sex comb on midleg (Scm) is a functional core polycomb repressive complex. Nature 529: 13577–13582.

Kim CA, Gingery M, Pilpa RM, Bowie JU. 2002. The SAM domain of Polycomb group gene pleiohomeotic encodes a helical polymer. Nat Struct Biol 9: 453–457.

Kim CA, Sawaya MR, Casadio D, Kim W, Bowie JU. 2005. Structural organization of a Sex-comb-on-midleg/polyhomeotic co-polymer. J Biol Chem 280: 27769–27775.
Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235–242.

Yu H, Mashtalir N, Daou S, Hammond-Martel I, Ross J, Sui G, Hart GW, Rauscher FJ, Drobetsky E, Milot E, et al. 2010. The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* 30: 5071–5085.

Zhang J, Bonasio R, Strino F, Kluger Y, Holloway JK, Modzelewski AJ, Cohen PE, Reinberg D. 2013. SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes Dev* 27: 749–766.