EMBO reports

scientific report

Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg

Clemens Grimm[1], *Andres Gaytan de Ayala Alonso*[2], *Vladimir Rybin*[3], *Ulrich Steuerwald*[1,3], *Nga Ly-Hartig*[2], *Wolfgang Fischle*[4], *Jürg Müller*[2+] & *Christoph W. Müller*[1,3++]

[1]European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France, [2]Gene Expression Unit and [3]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, and [4]Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Sex comb on midleg (Scm) is a member of the Polycomb group of proteins involved in the maintenance of repression of Hox and other developmental control genes in *Drosophila*. The two malignant brain tumour (MBT) repeats of Scm form a domain that preferentially binds to monomethylated lysine residues either as a free amino acid or in the context of peptides, while unmodified or di- or trimethylated lysine residues are bound with significantly lower affinity. The crystal structure of a monomethyl-lysine-containing histone tail peptide bound to the MBT repeat domain shows that the methyl-lysine side chain occupies a binding pocket in the second MBT repeat formed by three conserved aromatic residues and one aspartate. Insertion of the monomethylated side chain into this pocket seems to be the main contributor to the binding affinity. Functional analyses in *Drosophila* show that the MBT domain of Scm and its methyl-lysine-binding activity are required for repression of Hox genes.

Keywords: Polycomb group; MBT repeats; histone methylation

EMBO reports (2007) **8**, 1031–1037. doi:10.1038/sj.embor.7401085

INTRODUCTION

Polycomb group (PcG) proteins are required for the stable repression of different target genes in animals and plants (Müller & Kassis, 2006; Schwartz & Pirrotta, 2007). They were first identified in *Drosophila* in which they repress Hox genes outside their normal expression domains. PcG proteins form multisubunit assemblies. So far, three distinct *Drosophila* PcG complexes have been biochemically purified and characterized: Pleiohomeotic repressive complex (PhoRC), Polycomb repressive complex 1 (PRC1) and PRC2 (reviewed by Müller & Kassis, 2006; Schwartz & Pirrotta, 2007).

Although the three PcG complexes are composed of distinct proteins, the PhoRC subunit Sfmbt and the PRC1 subunit Sex comb on midleg (Scm) have strikingly similar domain architecture. Both proteins contain amino-terminal Cys_2-Cys_2 zinc fingers that are distinct from classical DNA-binding zinc fingers, and they contain an SPM domain at the carboxyl terminus. In both proteins, these domains flank a central portion of the protein that consists of malignant brain tumour (MBT) repeats. Scm contains two MBT repeats, whereas Sfmbt contains four MBT repeats (Bornemann *et al*, 1996; Klymenko *et al*, 2006). In addition to Scm and Sfmbt, the *Drosophila* genome encodes a third protein, lethal (3) malignant brain tumour (l(3)mbt), which has a similar domain architecture but contains three MBT repeats.

Both Scm and Sfmbt are essential for the repression of Hox genes. Embryos lacking Scm protein show widespread misexpression of Hox genes and die at the end of embryogenesis (Breen & Duncan, 1986; Simon *et al*, 1992). Similarly, removal of Scm protein in clones of cells during larval development results in severe misexpression of Hox genes in the mutant clones (Beuchle *et al*, 2001). Intriguingly, genetic studies also identified lethal Scm mutations that show milder phenotypes, and several of these encode proteins that are expressed as stable polypeptides *in vivo* but carry point mutations in the MBT repeats (Bornemann *et al*, 1998).

The crystal structures of the two MBT repeats of the human Scm homologue Scm-like 2 (SCML2; Sathyamurthy *et al*, 2003), and the three MBT repeats of human l(3)mbt (L3MBT; Wang *et al*, 2003), have been determined. MBT repeats consist of a five-stranded β-barrel-like structure preceded by a 30- to 45-residue N-terminal extension, which mediates the interaction between

[1]European Molecular Biology Laboratory, Grenoble Outstation, 6 Rue Jules Horowitz, 38042 Grenoble, France
[2]Gene Expression Unit, and
[3]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[4]Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[+]Corresponding author. Tel: +49 6221 387629; Fax: +49 6221 387518;
E-mail: juerg.mueller@embl.de
[++]Corresponding author. Tel: +49 6221 3878320; Fax: +49 6221 387519;
E-mail: christoph.mueller@embl.de

Received 24 May 2007; revised 6 September 2007; accepted 10 September 2007; published online 12 October 2007

adjacent repeats. In SCML2, the two MBT repeats contact each other through the core fold and the N-terminal extensions.

Structural similarity between MBT repeats, tudor domains and chromodomains led to the suggestion that MBT repeats also recognize methylated lysine or arginine residues in histone tails (Maurer-Stroh *et al*, 2003; Sathyamurthy *et al*, 2003). Data from protein-array screens show binding of MBT repeat proteins to monomethylated lysine residues in histone tails (Kim *et al*, 2006), whereas fluorescence polarization experiments showed that the MBT repeats of Sfmbt bind to mono- and dimethylated histone H3 lysine 9 (H3K9) and H4K20 peptides (Klymenko *et al*, 2006). Thus, in contrast to the well-characterized binding of the chromodomains of heterochromatin protein 1 (HP1), Polycomb protein, the plant homeodomain (PHD) finger of BPTF (bromodomain and PHD finger transcription factor) or ING2 (inhibitor of growth family 2) to specific trimethylated lysine residues in histone tails (reviewed by Sims & Reinberg, 2006), MBT repeats seem to prefer binding to lower methylated states of lysine residues. They also seem to recognize these methyl-lysine residues in the context of different flanking amino-acid sequences.

In this study, we show how the MBT repeats of Scm selectively recognize monomethylated lysine residues and also show that this methyl-lysine-binding activity is required for Polycomb repression in *Drosophila*.

RESULTS AND DISCUSSION

Methyl-lysine-binding activity of the MBT repeats of Scm

We used isothermal titration calorimetry (ITC) to test whether a fragment of the Scm protein that contains the two MBT repeats binds to unmodified or mono-, di- or trimethylated lysine residues or to methylated arginine. We found that the K_D values for interaction with unmodified or di- or trimethylated lysine or with mono- and dimethyl-arginine were too high to allow a reliable determination of the binding constant. By contrast, the MBT repeats clearly interacted with monomethylated lysine. The affinity could be determined with sufficient accuracy and was found to be 2.5 mM (Table 1).

We next asked whether the MBT repeats would also recognize monomethylated lysine in the context of a peptide. As Scm is a component of the chromatin-binding PRC1 complex, we measured binding to a set of histone H3 or H4 peptides that were unmodified or mono-, di- or trimethylated at specific lysine residues. K_D values for binding to unmodified or di- and trimethylated peptides were too high (>2 mM) to allow a reliable determination of the binding constant, whereas peptides monomethylated at H3K4, H3K9, H3K27, H3K36 and H4K20 were bound with K_D values of 500–1,200 μM (Table 1). A representative titration curve is depicted in Fig 1A. Similar results were obtained using fluorescence polarization assays (data not shown). Taken together, the MBT domain of Scm thus selectively binds to several different monomethylated lysine residues in histones H3 and H4. The affinity of the MBT domain for these peptides was, however, only about two- to fivefold higher than that for the free amino acid. This relatively modest increase in binding affinity and the apparent lack of specificity for residues flanking the methylated lysine suggest that adjacent residues in the analysed peptides contribute very little to the overall binding affinity. At this point, we cannot exclude that Scm MBT repeats also bind to other methylated lysine-containing proteins, perhaps even with higher affinities, although so far such proteins have not been identified. Our results contrast with observations made with *Drosophila* Sfmbt. The MBT repeats of Sfmbt selectively recognize histone peptides containing mono- or dimethylated H3K9 or H4K20 with low micromolar affinity, whereas other mono- and dimethylated lysine residues are recognized with significantly lower affinity (Klymenko *et al*, 2006).

Table 1 | Binding constants for various Scm ligands

	K_D (μM)*			
Amino acids	**Scm protein‡**			
Lys	>10,000			
Lys-me1	2,500 ± 60			
Lys-me2	>10,000			
Lys-me3	>10,000			
Arg-me1	>10,000			
symArg-me2	>10,000			
asymArg-me2	>10,000			

	K_D (μM)*			
Peptides	**Scm protein‡**	**Asp215Ala**	**Asp215Asn**	**Asp324Ala**
H3K4me1	1,220 ± 130			
H3K9me1	710 ± 110	1,200 ± 80	1,600 ± 100	>3,000
H3K9me2	>2,000			
H3K9me3	>2,000			
H3K27me1	1,200 ± 110			
H3K36me1	510 ± 40			
H4K20me1	550 ± 40	950 ± 90	1,100 ± 40	>3,000
H4K20me2	>2,000			

H, histone; K, Lys, lysine; me, methylation; Scm, Sex comb on midleg.
*Reported errors are fitting errors and were calculated using the program Origin version 5.0 provided by the manufacturer. K_D values for peptides H3K4me1, H3K9me1, H3K27me1, H3K36me1 and H4K20me1 were measured up to 72%, 77%, 71%, 83% and 80% saturation, respectively.
‡Scm protein corresponds to the construct used for co-crystallization with surface mutation Arg277Cys. Comparison of H4K20me1 peptide binding to wild-type and mutant proteins yielded comparable K_D values of 640 ± 30 and 550 ± 40 μM, respectively.

Structures of Scm bound to methyl-lysine peptides

A *Drosophila* Scm construct spanning the two MBT repeats and 40 additional C-terminal residues (174–435) was crystallized not only in its free form but also together with seven different peptides varying in length and centred on monomethylated K9 of histone H3 or monomethylated K20 of histone H4 (Fig 1B). Only a construct carrying the PCR-induced mutation Arg277Cys on the protein surface yielded crystals diffracting up to 2.2 Å not only with all peptides present but also in their absence (supplementary Table 1 online). The mutation is located in a loop preceding helix α1 of the N-terminal extension of the second MBT repeat and is more than 30 Å distant from the methyl-lysine-binding pocket of the second MBT repeat. A comparison of the binding of the wild-type and the mutant protein to one of the monomethylated

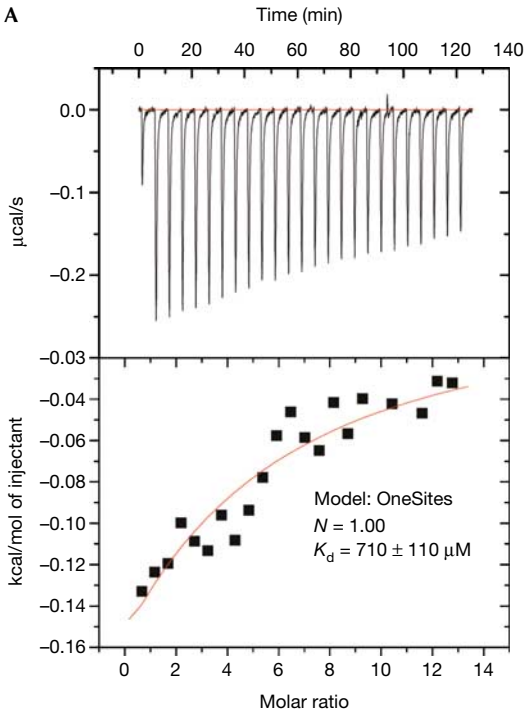

Fig 1 | Binding of methyl-lysine peptides to the MBT repeats of Scm. (**A**) ITC profile for the binding of a 16-mer H4 tail peptide with monomethylated lysine (K) 9 residue to the Scm MBT repeat domain. Data were fitted to a one-site model with stoichiometry of 1:1. The K_D value is 710 µm (see Table 1). (**B**) Methyl-lysine peptides used for the ITC experiments as summarized in Table 1. Sequences of mono-methylated H3K9 and H4K20 peptides used for co-crystallization experiments are underlined. ITC, isothermal titration calorimetry; MBT, malignant brain tumour; Scm, Sex comb on midleg.

histone peptides (H4K20me1) showed no marked difference in the K_D values (Table 1).

The structure of the Scm MBT repeat domain was solved by molecular replacement using the structure of the human Scm-like MBT repeats (Sathyamurthy *et al*, 2003; Protein Data Bank ID 1OI1), which shares 64% identical residues with *Drosophila* Scm. As expected, the two structures are very similar with an r.m.s. difference of 1.0 Å for 205 aligned residues.

Electron density maps computed from various data sets of crystals co-crystallized with monomethyl-lysine-containing peptides showed well-defined additional density within an 'aromatic cage' situated in the second MBT repeat, which was readily attributed to monomethyl-lysine (Fig 2A). By contrast, continuous electron density was not observed for the flanking residues in these peptides (Fig 2B); this is consistent with a minor contribution of these residues to the overall binding affinity as shown by ITC. The corresponding region of the first MBT repeat also showed no additional density features.

The methyl-lysine-binding site in the C-terminal MBT repeat of Scm comprises an aromatic cage formed by side chains of residues Phe 348, Trp 351 and Phe 355, where the planes of the three aromatic side chains are oriented perpendicular to each other, forming roughly the corner of a cube. Asp 324 forms the limit at the opposite side of the binding pocket. The ε-amino group of the bound monomethylated lysine points into the pocket and forms a hydrogen bond to Asp 324. In addition, hydrophobic interactions and also π-cation interactions between the three aromatic cage residues and the ammonium moiety largely contribute to the binding. The shaft of the amino acid formed by methylene groups is nearly perfectly aligned with the planar surface of the aromatic ring systems of Trp 351 and Phe 355 at the sides of the cage, whereas the ε-methyl group interacts with Phe 348 at its base (Fig 2B,C). The binding pocket is also outlined by three water molecules, with the closest forming a hydrogen bond to the ε-amino group of the lysine side chain (Fig 2B). The binding pocket identified in our experiments coincides with the binding pocket identified in human L3MBT, where in one crystal form a 2-(*N*-morpholino)ethanesulphonic acid molecule was bound to each of the three MBT repeats (Wang *et al*, 2003). It is likely that L3MBT also recognizes methyl-lysine residues, as the crucial methyl-lysine-contacting aspartate and most residues forming the 'aromatic cage' are conserved in all repeats. By contrast, in *Drosophila* Sfmbt, only the fourth repeat contains the methyl-lysine-contacting aspartate and all aromatic residues, whereas these residues are less conserved in the other repeats (supplementary Fig 1 online). Despite the overall conservation of the fourth repeat in Sfmbt, Phe 355 of Scm corresponds to a tyrosine, which could interact through its hydroxyl group with residues adjacent to the methyl-lysine.

To gain further insight into the binding specificity of Scm MBT repeats, native crystals were soaked with lysine or mono-, di- or trimethylated lysine, which did not change the diffraction quality of the crystals. Similar electron density in the C-terminal MBT repeat was observed for the mono- and dimethyl-lysine compounds as for crystals co-crystallized with monomethyl-lysine-containing peptides, whereas no additional density was observed with unmodified lysine or trimethylated lysine. Mono- or dimethylation of the ε-amino group establishes the necessary hydrophobic interactions and simultaneously maintains the hydrogen bond with Asp 324, whereas in trimethyl-lysine this hydrogen bond can no longer form. Comparison of the structures bound to monomethylated (Fig 2C) and dimethylated lysine (Fig 2D) shows that monomethyl-lysine binds around 0.6 Å deeper into the pocket. In addition, in the dimethyl-lysine-containing complex, the water molecule that contacts the ε-amino group is displaced by the additional methyl group (Fig 2B,C). The observed differences might explain the preferred binding of monomethylated lysine in ITC experiments (Table 1).

Different functional roles of the two MBT repeats

Comparison of the methyl-lysine-bound C-terminal MBT repeat structure with the unbound form shows the rearrangements of side

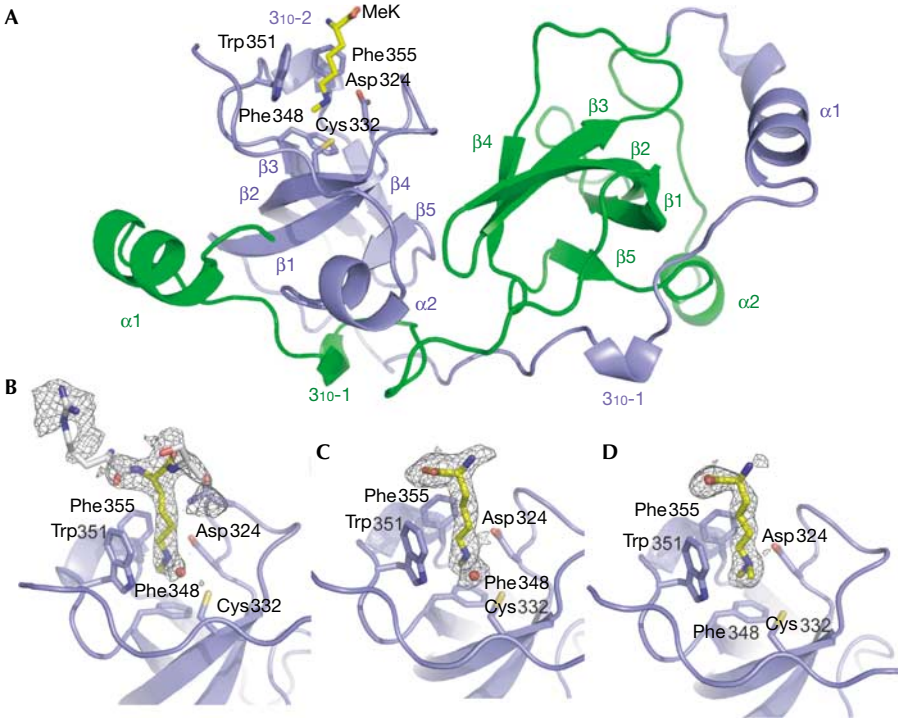

Fig 2 | Structure of the Scm MBT repeat domain. (**A**) Amino- and carboxy-terminal MBT repeats are depicted in green and blue. Methyl-lysine bound to the second MBT repeat is depicted in yellow. (**B**) Arg(Kme1)Ser peptide bound to the second MBT repeat. The electron density for a $2F_{obs}- F_{calc}$-simulated annealing omit map is contoured at 0.7σ. The well-ordered mono-methyl-lysine moiety is depicted in yellow, and disordered flanking residues are shown in grey. (**C**) Mono-methyl-lysine-bound structure contoured at 0.7σ. (**D**) Dimethyl-lysine-bound structure. MBT, malignant brain tumour; Scm, Sex comb on midleg.

chains in the binding pocket, while the rest of the molecule remains unchanged (Fig 3A). Specifically, Phe 355 occupies the binding pocket in the unbound structure and only on binding moves towards its perpendicular 'aromatic cage' conformation. In addition, the loop harbouring Phe 355 and also Trp 351 and Asp 324 on the other side of the binding pocket move slightly towards the monomethyl-lysine.

Although the core fold of the N- and C-terminal MBT repeats of Scm is similar (Fig 3B), we observe differences between them. Phe 348, Trp 351 and Phe 355 in the C-terminal repeat correspond to Leu 239, Ser 242 and Asn 246 in the N-terminal repeat, respectively, which do not form an aromatic cage. In addition, in the N-terminal repeat, a loop region from Glu 271 to Phe 276 occupies most of the volume corresponding to the binding pocket in the C-terminal repeat. Thus, it is unlikely that the first repeat would bind to amino-acid residues in a similar way as the second repeat.

Residue Asp 215 in the first MBT repeat corresponds to the methyl-lysine that contacts Asp 324 in the second repeat (supplementary Fig 1 online). As expected, mutating these two aspartate residues yields different results: mutant proteins $Scm^{Asp215Ala}$ and $Scm^{Asp215Asn}$ bind to monomethyl-lysine peptides H3K9 and H3K20 about twofold weaker, whereas $Scm^{Asp324Ala}$ protein no longer shows detectable binding (Table 1). Circular dichroism (CD) spectra analyses indicate that mutations in either MBT repeat do not perturb the overall fold, but melting curves show that all three mutant proteins $Scm^{Asp215Ala}$, $Scm^{Asp215Asn}$ and $Scm^{Asp324Ala}$ are less stable than the wild-type protein (see supplementary Fig 3 online for CD spectra and melting curves). The strongest effect on thermostability is observed in the case of $Scm^{Asp215Ala}$, probably because the two hydrogen bonds of Asp 215 with the backbone amides of residues 217 and 274, observed in the wild-type structure, can no longer form. $Scm^{Asp215Asn}$ corresponds to the protein encoded by the mutant allele $Scm^{Su(z)302}$ (Bornemann *et al*, 1998), and it also shows reduced thermostability compared with the wild-type protein (supplementary Fig 3 online), which could explain the mild phenotype of $Scm^{Su(z)302}$ homozygotes (Bornemann *et al*, 1998).

Comparison with chromo-, tudor- and PHD finger domains

The structure of the methyl-lysine-bound Scm MBT repeat domain was compared with structures of chromo- (Jacobs & Khorasanizadeh, 2002), tudor- (Huang *et al*, 2006) and PHD finger domains (Pena *et al*, 2006) bound to lysine methylated histone tail peptides (supplementary Fig 2 online). This shows that in these complexes—unlike in Scm MBT complexes—several histone residues flanking the methyl-lysine interact with the protein. The 'radial' orientation of the methyl-lysine residue bound by the Scm MBT repeat with the side chain pointing towards the molecular centre contrasts with a more 'tangential' binding mode observed in these other structures. In the latter mode, residues directly neighbouring the methyl-lysine are also gauged, whereas reading out residues directly adjacent to the

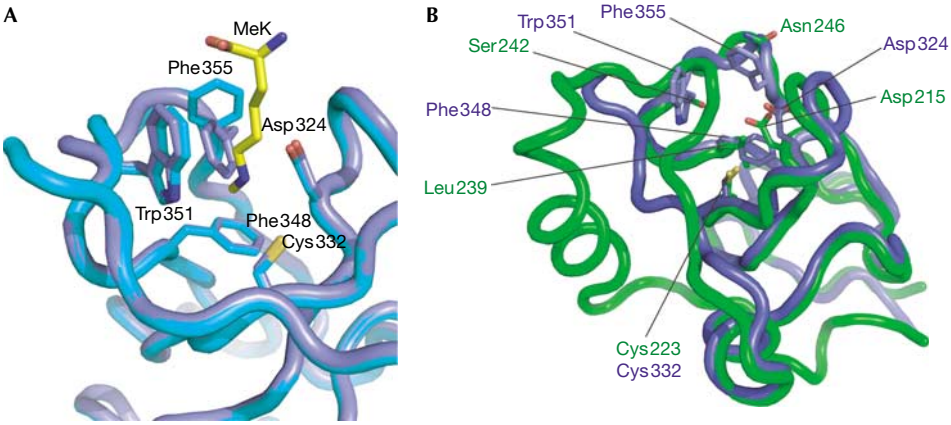

Fig 3 | Ribbon representation of the Scm-binding pocket. (**A**) View of the binding pocket with superimposed apo (turquoise) and monomethylated lysine-bound structures (light blue). The monomethylated lysine is shown in yellow. (**B**) Superposition of the amino-terminal (green) with the carboxy-terminal MBT (blue) repeat of Scm. MBT, malignant brain tumour; Scm, Sex comb on midleg.

methyl-lysine seems more difficult in the 'radial' binding mode, which forces adjacent side chains to point away from the protein surface. Similarly, the binding affinities measured in this work are at least one order of magnitude below those of PHD finger, tudor and chromodomains for their cognate peptides.

Role of the MBT repeats in PcG repression

To study the function of the MBT repeat domain of Scm in PcG repression, we analysed Hox gene expression in animals in which the wild-type Scm protein had been replaced by mutant forms containing lesions in the MBT domain. Specifically, we used a genetic rescue assay in *Drosophila* larvae as follows: clones of imaginal disc cells that are homozygous for an *Scm*-null mutation and therefore lack Scm protein fail to maintain PcG repression, and Hox genes such as *Ultrabithorax* (*Ubx*) are strongly mis-expressed in the mutant cells (Fig 4, left-most panels; cf. Beuchle *et al*, 2001). Repression of Ubx in such Scm mutant clones is fully rescued by a transgene expressing wild-type Scm protein under the control of a heat-inducible heat-shock protein 70 (*hsp70*) promoter (*hsp70-Scm*) (Fig 4, second from left; cf. Beuchle *et al*, 2001). We therefore generated *hsp70-Scm*$^{\Delta MBT}$ and *hsp70-Scm*^{Asp324Ala} transgenes and tested their rescuing capacity in the same assay. The Scm$^{\Delta MBT}$ protein contains a deletion of the entire MBT repeat domain, whereas in the Scm^{Asp324Ala} protein, replacement of Asp 324 by an alanine residue in the binding pocket of the second MBT repeat abolishes peptide binding without perturbing the overall fold of the MBT domain (see above). On heat-shock-induced expression in larvae, the Scm$^{\Delta MBT}$ and Scm^{Asp324Ala} proteins were both expressed at levels comparable with the wild-type Scm protein expressed from the *hsp70-Scm* control transgene (Fig 4B, bottom row). However, both the Scm$^{\Delta MBT}$ and the Scm^{Asp324Ala} protein showed reduced functionality in the rescue assay, and neither protein was able to repress *Ubx* as efficiently as the wild-type Scm protein. In particular, every animal expressing Scm$^{\Delta MBT}$ or Scm^{Asp324Ala} showed mis-expression of *Ubx* in at least a fraction of clone cells (Fig 4A, white arrowheads). Unexpectedly, the Scm$^{\Delta MBT}$ and Scm^{Asp324Ala} proteins were still able to rescue repression of *Ubx* in a fraction of clone cells (Fig 4A, empty arrowheads), suggesting that even the deletion of both MBT repeats (that is, in the case of Scm$^{\Delta MBT}$) does not completely disable the PcG repressor function of Scm. This result might seem surprising. However, it should be kept in mind that the PcG protein *Drosophila* Sfmbt has monomethyl-lysine-binding activity similar to Scm (Klymenko *et al*, 2006). Although removal of either the Scm or the Sfmbt protein causes a severe PcG phenotype (Breen & Duncan, 1986; Simon *et al*, 1992; Klymenko *et al*, 2006), both proteins are bound to regulatory sequences of Hox genes *in vivo* (Bornemann *et al*, 1998; Klymenko *et al*, 2006) and physically interact with each other *in vitro* (R. Matos & J. Müller, unpublished data). It is thus possible that the MBT repeats in the two proteins function in a partly redundant manner and that, in the absence of a functional MBT domain in Scm, methyl-lysine binding by Sfmbt can to some extent compensate for the lack of Scm-mediated methyl-lysine binding.

Interestingly, the expression of Scm$^{\Delta MBT}$ or Scm^{Asp324Ala} in larvae resulted in developmental arrest shortly after puparium formation in 100% of the animals, whereas larvae expressing wild-type Scm protein from the control transgene developed into healthy adult flies that showed no obvious phenotypes (see Methods). This suggests that the two mutant proteins not only have reduced functionality but also interfere with normal development. In this context, it is intriguing to note that while genetic screens for lethal *Scm* mutations identified five *Scm* alleles with amino-acid substitutions or deletions in the first MBT repeat, no mutations with lesions in the second MBT repeat have been recovered (Bornemann *et al*, 1998). It is possible that lesions disrupting the function of the second MBT repeat cannot be isolated, because they would cause lethality already in heterozygotes, akin to what is observed on expression of Scm^{Asp324Ala}.

Finally, we note that in the rescue assay shown in Fig 4, the Scm^{Asp324Ala} and Scm$^{\Delta MBT}$ proteins show a comparably compromised rescuing activity. This strongly suggests that the Asp324Ala mutation in the methyl-lysine-binding pocket renders the MBT repeats of Scm nonfunctional. Methyl-lysine binding thus seems to be the main activity that the MBT domain contributes to the PcG repressor activity of the Scm protein.

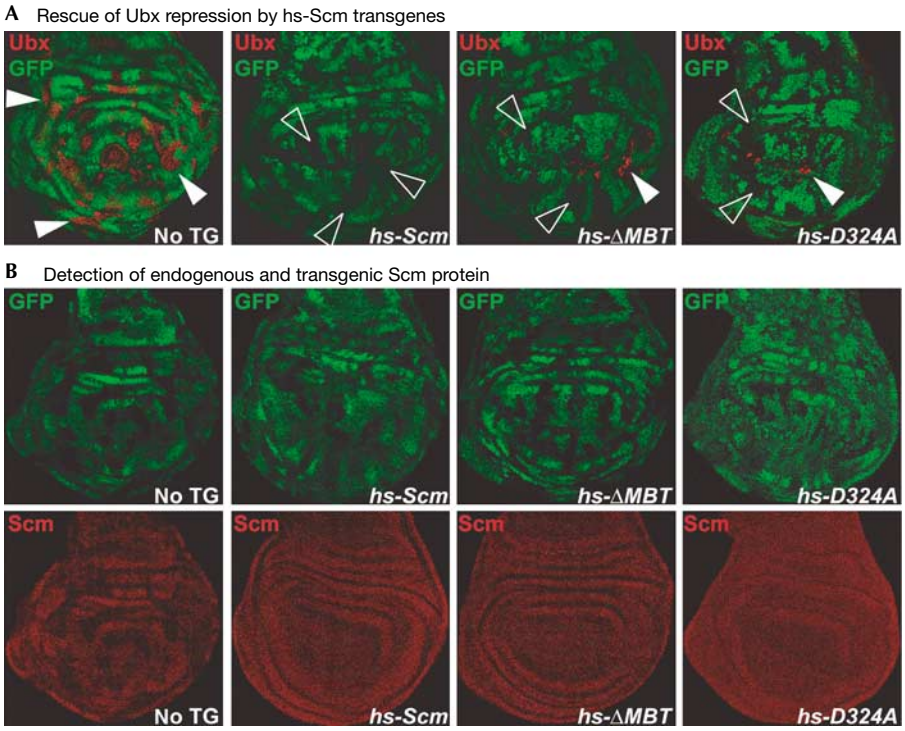

Fig 4 | Requirement for the MBT repeat domain of Scm in Hox gene repression. Wing imaginal discs with Scm^{D1} mutant clones from animals carrying no transgene (no TG) or the indicated *hsp70-Scm* transgene were used. Discs were stained with antibodies against Ubx (red, top row) or against Scm protein (red, bottom row); Scm^{D1} mutant clones are marked by the absence of GFP (green, top and middle rows). In the middle and bottom rows, the GFP and Scm protein signals of the same disc are shown separately. Animals were repeatedly heat-shocked for 1 h every 12 h beginning at the time of clone induction, and discs were analysed 96 h after clone induction. (**A**) In the absence of an *hsp70-Scm* transgene (no TG), Ubx is strongly misexpressed in most *Scm* mutant clones in the wing pouch. In animals carrying *hsp70-Scm*, Ubx stays repressed in all mutant cells. The *hsp70-*$Scm^{\Delta MBT}$ and *hsp70-*$Scm^{Asp324Ala}$ transgenes are unable to fully rescue repression of Ubx, and small clusters of Ubx-positive cells are present in the region of the presumptive wing margin (white arrowheads), but note that Ubx repression is nevertheless rescued in a considerable fraction of clone cells (empty arrowheads) compared with animals lacking a transgene. (**B**) The transgene-encoded hs-Scm, hs-$Scm^{\Delta MBT}$ and hs-$Scm^{Asp324Ala}$ proteins are all expressed at comparable levels (bottom row). GFP, green fluorescent protein; hs, heat shock; hsp70, heat-shock protein 70; MBT, malignant brain tumour; Scm, Sex comb on midleg; Ubx, Ultrabithorax.

METHODS

Data collection and processing and structure solution. Data were collected at the European Synchrotron Radiation Facility (ESRF) beamline ID29, recorded with an Area Detector Systems Corporation Q315r detector and processed with program XDS (Kabsch, 1993). Crystals diffracted up to 2.0 Å and belong to space group $P4_32_12$ with cell axes of $a=83.2$ Å and $c=75.6$ Å. Molecular replacement with PHASER using the human SCML2 model (Sathyamurthy *et al*, 2003) yielded a clear solution with one molecule in the asymmetric unit. After two rounds of automated refinement with REFMAC (CCP4, 1994) including TLS refinement and manual rebuilding, the R/R_{free} factor converged at 18.8%/22.5% for the apo structure. Crystallographic statistics are summarized in supplementary Table 1 online. Residue Asp 341 falls into the disallowed region of the Ramachandran plot, although close to a generously allowed region. Asp 341 is located in a tight turn connecting strands β2 and β3 of the second MBT repeat and is not directly involved in ligand binding. The electron density corresponding to this residue is well defined. In SCML2, the C_α trace for this turn region is similar; however, the position corresponding to residue Asp 341 is occupied by a glycine residue with higher conformational flexibility. The complete model contains 210 residues and 171 water molecules. The last 40 C-terminal residues after the second MBT repeat are disordered.

***Drosophila* transgenes and functional analyses.** *hs-*$Scm^{Asp324Ala}$ and *hs-*$Scm^{\Delta MBT}$ contain the Scm complementary DNA fragment described by Beuchle *et al* (2001). In the case of the Asp324Ala mutant, the GAC codon for Asp 324 was replaced by a GCC codon for the alanine residue, and in the case of the ΔMBT mutant the codons from Val 174 to Met 379 were substituted by the sequence GCCGCGGGC to generate an Ala-Ala-Gly linker.

For the clonal analysis shown in Fig 4, y w *hs-flp*; *FRT82B hs-GFP* females were crossed to males of the genotype *w*; *hsp70-Scm**; *FRT82B* Scm^{D1}/TM6C (*hsp70-Scm** indicates the transgene encoding wild-type Scm, $Scm^{Asp324Ala}$ or $Scm^{\Delta MBT}$) or to *w*; *FRT82B* Scm^{D1}/TM6C males (no-transgene control). The progeny of these crosses was reared at 25 °C, and 1 h heat shocks (37 °C) were applied every 12 h over a period of 96 h, starting

30 h after egg laying. Before dissection, larvae were subjected to another 1 h heat shock followed by a 1 h recovery period to induce expression of the green fluorescent marker protein. Imaginal discs were stained with a mouse monoclonal anti-Ubx FP3.38 antibody or with an antibody against the Scm protein.

We also subjected *hsp70-Scm*, *hsp70-Scm*$^{\Delta MBT}$ and *hsp70-Scm*^{Asp324Ala} transgenic animals to extended heat-shock regimes throughout larval and pupal development (that is, 1 h heat shocks at 37 °C every 12 h, starting 30 h after egg laying). Among 241 *hsp70-Scm* transgenic animals that developed into third instar larvae, 121 developed into adult flies and these animals showed no obvious phenotype. Of the 314 *hsp70-Scm*$^{\Delta MBT}$ and 201 *hsp70-Scm*^{Asp324Ala} transgenic animals, none developed beyond puparium formation.

Coordinates. The atomic coordinates and structure factors for the apo Scm MBT repeat domain, bound to monomethyl-lysine, dimethyl-lysine and to a monomethylated peptide, have been deposited with the Protein Data Bank under the accession codes 2R57, 2R5A, 2R58 and 2R5M, respectively.

Supplementary information is available at *EMBO reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

We thank J. Simon for antibodies and B. Gärtner for assistance in plasmid construction, C. Petosa for discussion, the European Molecular Biology Laboratory (EMBL)/European Synchrotron Radiation Facility (ESRF) Joint Structural Biology Group for access and support at the ESRF beamlines, and the EMBL Grenoble crystallization facility for support. A.G.A.A. is supported by a grant from the Deutsche Forschungsgemeinschaft.

REFERENCES

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

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

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

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

CCP4 (1994) Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50:** 760–776

Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312:** 748–751

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

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry cell constants. *J Appl Crystallogr* **26:** 795–800

Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7:** 397–403

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

Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP (2003) The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28:** 69–74

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

Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442:** 100–103

Sathyamurthy A, Allen MD, Murzin AG, Bycroft M (2003) Crystal structure of the malignant brain tumor (MBT) repeats in Sex Comb on Midleg-like 2 (SCML2). *J Biol Chem* **278:** 46968–46973

Schwartz YB, Pirrotta V (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* **8:** 9–22

Simon J, Chiang A, Bender W (1992) Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development* **114:** 493–505

Sims III RJ, Reinberg D (2006) Histone H3 Lys 4 methylation: caught in a bind? *Genes Dev* **20:** 2779–2786

Wang WK, Tereshko V, Boccuni P, MacGrogan D, Nimer SD, Patel DJ (2003) Malignant brain tumor repeats: a three-leaved propeller architecture with ligand/peptide binding pockets. *Structure* **11:** 775–789