The EMBO Journal (2009) 28, 1965–1977 | © 2009 European Molecular Biology Organization | Some Rights Reserved 0261-4189/09
www.embojournal.org

THE EMBO JOURNAL

EMBO open

Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation or the creation of derivative works without specific permission.

Clemens Grimm[1,4], Raquel Matos[2], Nga Ly-Hartig[2], Ulrich Steuerwald[3], Doris Lindner[3], Vladimir Rybin[3], Jürg Müller[2,*] and Christoph W Müller[3,*]

[1]European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France, [2]European Molecular Biology Laboratory, Gene Expression Unit, Heidelberg, Germany and [3]European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany

Polycomb group (PcG) proteins repress transcription by modifying chromatin structure in target genes. dSfmbt is a subunit of the *Drosophila melanogaster* PcG protein complex PhoRC and contains four malignant brain tumour (MBT) repeats involved in the recognition of various mono- and dimethylated histone peptides. Here, we present the crystal structure of the four-MBT-repeat domain of dSfmbt in complex with a mono-methylated histone H4 peptide. Only a single histone peptide binds to the four-MBT-repeat domain. Mutational analyses show high-affinity binding with low peptide sequence selectivity through combinatorial interaction of the methyl-lysine with an aromatic cage and positively charged flanking residues with the surrounding negatively charged surface of the fourth MBT repeat. dSfmbt directly interacts with the PcG protein Scm, a related MBT-repeat protein with similar methyl-lysine binding activity. dSfmbt and Scm co-occupy Polycomb response elements of target genes in *Drosophila* and they strongly synergize in the repression of these target genes, suggesting that the combined action of these two MBT proteins is crucial for Polycomb silencing.

The EMBO Journal (2009) **28,** 1965–1977. doi:10.1038/emboj.2009.147; Published online 4 June 2009
Subject Categories: chromatin & transcription; structural biology
Keywords: histone modifications; MBT repeat; methyl-lysine recognition; Polycomb group proteins

*Corresponding authors. CW Müller, European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
Tel.: +49 6221 387 8320; Fax: +49 6221 387 519;
E-mail: Christoph.Mueller@embl.de or J Müller, European Molecular Biology Laboratory, Gene Expression Unit, 69117 Heidelberg, Germany.
Tel.: +49 6221 387 629; Fax: +49 6221 387 518;
E-mail: Juerg.Mueller@embl.de
[4]Present address: Institut für Biochemie, Biozentrum der Universität, Würzburg, Am Hubland, D-97074 Würzburg, Germany

Received: 21 October 2008; accepted: 11 May 2009; published online: 4 June 2009

Introduction

Polycomb group (PcG) proteins are transcriptional regulators required for the repression of developmental control genes in animals and plants. PcG proteins exist in distinct multi-protein complexes that repress transcription by modifying the chromatin of target genes and thereby generating transcriptional off states that can be stably and heritably maintained (Francis and Kingston, 2001; Schwartz and Pirrotta, 2007). To date, three principal PcG multi-protein complexes have been identified and characterized: Pho repressive complex (PhoRC), PRC2 and the two related complexes PRC1 and dRAF (Schwartz and Pirrotta, 2007; Muller and Verrijzer, 2009). Among those, the PhoRC subunit Pho is the only sequence-specific DNA-binding PcG protein. Studies in *Drosophila* showed that PcG complexes assemble at specific *cis*-regulatory sequences in target genes, called Polycomb response elements (PRE), and that PhoRC has a central function in providing a PRE-binding platform that allows for the assembly of the chromatin-binding PRC1 and PRC2 complexes (Wang *et al*, 2004; Mohd-Sarip *et al*, 2005; Klymenko *et al*, 2006).

In addition to Pho, PhoRC contains dSfmbt (Klymenko *et al*, 2006). In *Drosophila*, dSfmbt, the PRC1 subunit Sex comb on midleg (Scm) and a third protein, called L(3)mbt, form a small protein family with a very similar and unique domain architecture. The central portion of each protein contains an MBT-repeat domain that consists of two (Scm), three (L(3)mbt) or four (dSfmbt) repeats, and each protein contains Zn-finger motifs in the N-terminus and a sterile alpha motif (SAM) domain at the very C-terminus. Studies on dSfmbt, first showed that MBT-repeat domains selectively bind to mono- and dimethylated lysine residues in histones, but that they show low specificity for any particular histone lysine (Klymenko *et al*, 2006). Recent studies reported the crystal structures of the MBT domains of Scm and L3MBTL1 in complex with methylated histone-tail peptides (Grimm *et al*, 2007; Li *et al*, 2007; Min *et al*, 2007; Santiveri *et al*, 2008). In both proteins, the mono- or dimethylated histone lysine residues bind to the second MBT repeat and the interactions between the methyl-lysine side chain and an aromatic pocket in this repeat contribute the major part of the binding energy, whereas histone residues adjacent to the methyl-lysine form few interactions (Grimm *et al*, 2007; Li *et al*, 2007; Min *et al*, 2007; Santiveri *et al*, 2008). Consistent with this mode of recognition, the MBT-repeat domain of Scm binds histone-tail peptides, mono-methylated at H3-K9 or H4-K20 with a low affinity of about 500–800 μM (Grimm *et al*, 2007; Santiveri *et al*, 2008), whereas for binding of L3MBTL1 to the same mono-methylated lysines in peptides,

two studies reported different affinities ranging from 140 to 400 μM (Min *et al*, 2007) or from 5 to 10 μM (Li *et al*, 2007).

Interestingly, two distinct MBT-repeat-containing proteins, Scm and dSfmbt, are both essential components of the PcG-repression system in *Drosophila*. Functional studies on Scm showed that mutations in the MBT-repeat domain that abolish methyl-lysine binding *in vitro* impede the Polycomb-repressor function of this protein in *Drosophila* (Grimm *et al*, 2007). Intriguingly, dSfmbt binds the same methylated lysines in histones bound by Scm but with about 100-fold higher affinity than Scm (Klymenko *et al*, 2006; Grimm *et al*, 2007; Santiveri *et al*, 2008). These observations, together with the lack of knowledge of sequence-specific methyl-lysine recognition by the L3MBTL1 or Scm MBT-repeat domains prompted us to characterize the MBT-repeat domain of dSfmbt at the structural and functional level. Here, we report the crystal structure of the MBT-repeat domain of dSfmbt in complex with a histone H4 peptide, mono-methylated at lysine 20 (H4K20me1). Using isothermal calorimetry (ITC), we evaluate the binding specificity of dSfmbt for different histone-tail peptides methylated at particular lysine residues and assess the contribution of residues adjacent to the methyl-lysine residue by mutational analysis. Functional tests in *Drosophila* show that dSfmbt and Scm act in a highly synergistic manner to maintain repression at Polycomb target genes *in vivo* and suggest a role for the Scm–dSfmbt heterodimer in chromatin compaction.

Results and discussion

Overall structure of the four-MBT-repeat domain of dSfmbt

The structure of the four-MBT-repeat domain of *D. melanogaster* dSfmbt (dSfmbt-4MBT, $M_r = 51$ kDa, residues 535–977) was solved in complex with a histone H4 tail peptide centred onto H4K20me1 at 2.8 Å resolution (Table I). To favour crystallization, three point mutations (K715D, R886S and R900D) were introduced on the surface of the dSfmbt-4MBT construct; these mutations do not significantly affect H4K20me1 binding (Table II, Materials and Methods). The overall structure of the dSfmbt-4MBT–peptide complex is shown in Figure 1. As in Scm and L3MBTL1, each MBT repeat consists of a central five-stranded β-core and an elongated N-terminal arm that contacts the neighbouring repeat. Repeat 2, 3 and 4 form a propeller-like structure with three-fold pseudo-symmetry similar to L3MBTL1 (Wang *et al*, 2003). Repeat 1 is docked onto the outer rim of this propeller in the area of repeat 4 and forms most contacts with repeat 4 but also interacts with the adjacent repeat 2 through the N-terminal arm of this repeat. The arm of repeat 1 forms most of the contact surface to repeat 4 and its conformation is therefore less extended compared with the three other arms (Figure 1B).

Table I Crystallographic data collection, phasing and model refinement statistics

Data statistics	Native	Hg derivative[a]	
		Hg-λ1	Hg-λ2
Space group	$P22_12_1$	C222	
Cell axes (Å)	70.8, 97.0, 214.1	100.4, 140.0, 275.2	
Wavelength (Å)	0.9330	1.0085	1.0064
Resolution range (Å)[b]	30–2.8 (3.0–2.8)	30–3.2 (3.37–3.2)	
Reflections observed/unique	226 138/37 129	137 646/32 417	137 646/32 401
Completeness (%)	99.8 (99.9)	97.6 (91.4)	97.8 (91.7)
<I>/<sig(I)>	12.3 (3.3)	4.4 (1.7)	4.4 (1.7)
R_{sym} (%)[c]	13.7 (55.1)	8.5 (49.0)	8.2 (48.0)
Phasing statistics			
Resolution range (Å)		30–3.2	
		ano	iso/ano
R_{Cullis} (acentric) (%)		0.82	0.72/0.88
Phasing power (acentric)		0.99	0.96/0.88
Overall figure of merit		0.50	
Model refinement			
Resolution range (Å)	30–2.8 (2.87–2.8)		
R_{cryst}/R_{free} (%)[d]	22.3 (32.6)/25.4 (36.0)		
Overall B-factor [Å^2]	28.9		
Protein main/side chain [Å^2]	28.0/29.9		
Peptide ligand [Å^2]	34.7		
r.m.s.d. bond lengths [Å]	0.009		
r.m.s.d. angles (deg)	1.35		
Residues in Ramachandran plot (%):			
Most favoured region	89.4		
Allowed	9.9		
Generously allowed	1.1		
Disallowed	0.0		

[a]Hg derivative with three sites per molecule was obtained by soaking crystals for 1 h with 0.25 mM EMTS. Hg-λ1 was used as reference data set in the two-wavelengths MAD experiment.
[b]Values in parentheses refer to the outer resolution shell.
[c]$R_{sym} = \sum|I - \langle I\rangle|/\sum I$, where I is the observed intensity of a given reflection.
[d]$R_{cryst} = \sum|F_o - F_c|/\sum F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is equal to R_{cryst} for a randomly selected 5% subset of reflections not used in the refinement.

 ©2009 European Molecular Biology Organization

Table II Binding affinity of dSfmbt-4MBT for methyl-lysine-containing histone peptides

	Wild-type dSfmbt-4MBT[a] K_D (μM)	Mutant dSfmbt-4MBT protein K_D (μM)				
		D697A	D808A	D917A	E947A/Y948F	E947A/Y948F/D917A
H4K20$_{12-27}$—KGGAKRHRK$_{20}$VLRDNIQ-CONH$_2$						
H4K20	>1000					
H4K20me1	1.1 ± 0.1[b] (11 ± 0.4)	1.6 ± 0.1	2.9 ± 0.1	>1000	1.9 ± 0.1 (18 ± 0.5)	>1000
H4K20me2	2.8 ± 0.2				2.7 ± 0.3	>1000
H4K20me3	>1000					
H4K20$_{17-23}$—RHRK$_{20}$VLR-CONH$_2$						
H4K20me1 (7-mer)	1.5 ± 0.03 (12.8 ± 1.0)					
H4K20$_{18-22}$—HRK$_{20}$VL-CONH$_2$						
H4K20me1 (5-mer)	23 ± 1					
H4K20$_{19-21}$—RK$_{20}$V-CONH$_2$						
H4K20me1 (3-mer)	40 ± 2.4					
H4K20-R19A$_{12-27}$—KGGAKRHAK$_{20}$VLRDNIQ-CONH$_2$						
H4K20me1-R19A	4.6 ± 0.3 (60 ± 4)					
H4K20$_{12-27}$ (scrambled)—LNRQDIAGK$_{20}$GKHVKRR-CONH$_2$	2.8 ± 0.1 (45 ± 3)					
H3K4$_{1-15}$—ARTK$_4$QTARKSTGGKA-CONH$_2$						
H3K4me1	2.8 ± 0.1					
H3K4me2	2.2 ± 0.1					
H3K9$_{1-15}$—ARTKQTARK$_9$STGGKA-CONH$_2$						
H3K9	85 ± 3					
H3K9me1	4.7 ± 0.3 (21 ± 0.4)	1.9 ± 0.1	6.7 ± 0.4	340 ± 80		
H3K9me2	15.6 ± 0.9 (36 ± 4)					
H3K9me3	85 ± 10					
H3K9$_{8-10}$—RK$_9$S-CONH$_2$						
H3K9me1	134 ± 8 (125 ± 6)					
H3K27$_{19-35}$—QLATKAARK$_{27}$SAPATGGV-CONH2						
H3K27	40 ± 12					
H3K27me1	2.9 ± 0.4					
H3K27me2	12.0 ± 1.0					
H3K27me3	>1000					
H3K36$_{28-43}$—SAPATGGVK$_{36}$KPHRYRPG-CONH2						
H3K36me1	5.0 ± 0.1					
H3K36me2	3.4 ± 0.1					
H3K79$_{72-86}$—REIAQDFK$_{79}$TDLRFQS-CONH2						
H3K79me1	>1000					
H3K79me2	>1000					

[a]Values in parentheses were measured at 150 mM NaCl.
[b]$K_D = 2.1 \pm 0.1$ μM for dSfmbt construct K715D/R886S/R900D (residue 535–977) used for co-crystallization.

The combination of these interactions between the four repeats thus results in a compact MBT-repeat domain.

H4K20me1 peptide binds to the fourth MBT repeat of dSfmbt

In the complex structure, the H4K20me1 peptide (RHRKme1VLR) interacts with dSfmbt MBT repeat 4 (Figure 1A). Interactions between dSfmbt and the peptide are mediated by the central mono-methylated lysine, which points in the binding pocket on top of the β-barrel of the fourth MBT repeat (Figure 2) but also through a combination of polar and hydrophobic interactions of adjacent peptide residues with residues of repeat 4.

The methyl-lysine-binding pocket of the fourth repeat is formed by residues Phe941, Trp944 and Tyr948, whose aromatic planes are oriented perpendicular to each other, forming roughly the corner of a cube. The methyl-lysine side chain closely packs against the aromatic side chains of Tyr948 and Trp944. Compared with the 'aromatic cage' in Scm (Grimm *et al*, 2007), we observe a significant distortion of the ideal rectangular geometry, mainly because dSfmbt-residue Tyr948 is oriented at an angle of approximately 60° with respect to Trp944. On the other side of the binding pocket, Asp917 binds the ε-amino group of H4K20me1 through a direct hydrogen bond assisted by electrostatic interactions. Furthermore, the pocket is outlined by residue Cys925. In addition to the interactions with the mono-methylated lysine, a salt bridge connects dSfmbt Glu947 (corresponding to Scm Ala354) with Arg19 in histone H4, whereas the hydroxyl group of Tyr948 (corresponding to Scm Phe355) forms a

©2009 European Molecular Biology Organization

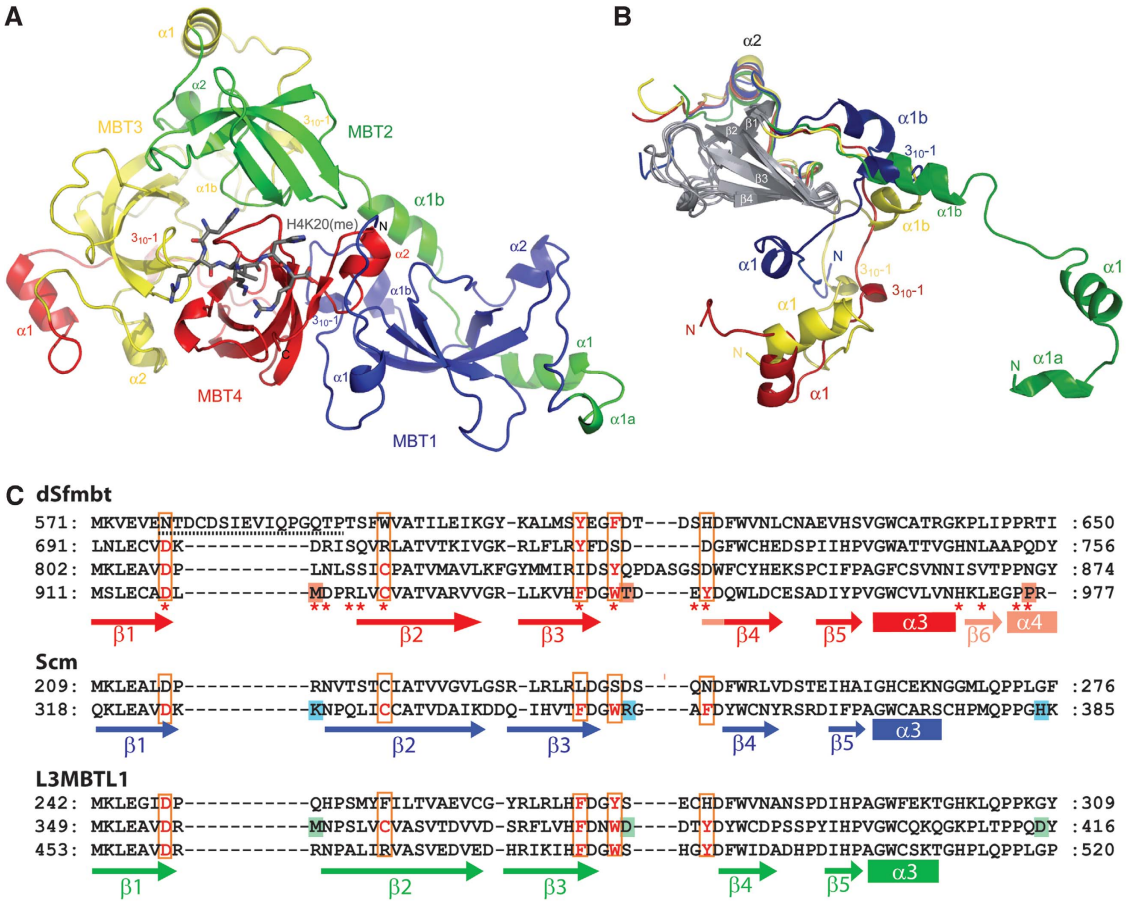

Figure 1 Structure of the four MBT-repeat domain of dSfmbt. (**A**) Ribbon diagram of the four MBT repeats of dSfmbt coloured in blue (repeat 1), green (repeat 2), yellow (repeat 3) and red (repeat 4). Histone H4K20me1 peptide is shown in grey. (**B**) Superposition of the core folds (grey) of MBT repeats 1–4. For each repeat, helix α2 and the arm regions are coloured according to (A). (**C**) dSfmbt core domains of repeat 1–4 aligned with core domains of MBT repeats in *Drosophila* Scm and human L3MBTL1. Positions corresponding to cage-forming residues in dSfmbt repeat 4, Scm repeat 2 and L3MBTL1 repeat 2 are marked with a red box and conserved cage-forming residues are depicted in red. dSfmbt residues contacting the H4K20me1 peptide are indicated with an asterisk. dSfmbt, Scm and L3MBTL1 residues important for differential peptide binding are drawn on red, blue and green background (compare text). N-terminal arm regions are less well conserved and are not included in the alignment.

hydrogen bond with the Nε atom of this arginine (Figure 2). In the dSfmbt–peptide complex, electron density can be unambiguously assigned for six of the seven peptide residues (Figure 2). A peptide surface of 480 Å^2 contacts dSfmbt, whereby 40% of the interaction surface is contributed by the mono-methylated lysine residue.

Contributions of H4K20me1 and dSfmbt residues to the peptide-binding affinity

We used ITC to evaluate binding of dSfmbt to methylated histone-tail peptides. First, we tested the binding of dSfmbt-4MBT to 16-residue peptides that were either unmodified, mono-, di- or tri-methylated at H4K20 (Table II, ITC profiles are depicted in Supplementary Figure S1). Mono- and di-methylated H4K20 peptides were bound with 1 and 3 μM affinity, respectively, whereas unmethylated and tri-methylated H4K20 peptides were bound with approximately 500-fold lower affinities ($K_D > 1000$ μM, Table II). To probe the contribution of residues flanking the methyl-lysine, we next tested binding to shorter H4K20me1 peptides. The heptameric peptide used for co-crystallization was bound with an affinity comparable to the 16-residue peptide. However, further shortening to a five-residue peptide reduced the affinity to 23 μM (Table II). This suggests that contributions provided by residues Arg17 and especially Arg23 that is well ordered in the crystal structure (Figure 2) are responsible for the approximately 15-fold higher affinity for the heptameric peptide. An even shorter three-residue peptide was bound with a K_D value of 40 μM (Table II), indicating that His18 and Leu22, both pointing away from the MBT surface, contribute little to the binding affinity. The next residue Arg19 directly adjacent to K20me1 is involved in polar interactions with dSfmbt (Figure 2) and in the context of the 16-residue H4K20me1 peptide, mutating Arg19 into alanine reduces the binding affinity by about four-fold (Table II).

In a complementary set of experiments, we mutated dSfmbt residues Glu947 and Tyr948 to generate a dSfmbt$^{E947A/Y948F}$ protein (Figure 1C). Compared with wild-type dSfmbt, the dSfmbt$^{E947A/Y948F}$ protein bound the 16-residue H4K20me1 peptide with similar affinity (Table II), presumably because the change from Tyr948 to Phe948 still permits the π−cation interaction with the guanidinium group of Arg19. However, mutating the methyl-lysine-contacting Asp917 into alanine in the single-mutant dSfmbt^{D917A} or

 ©2009 European Molecular Biology Organization

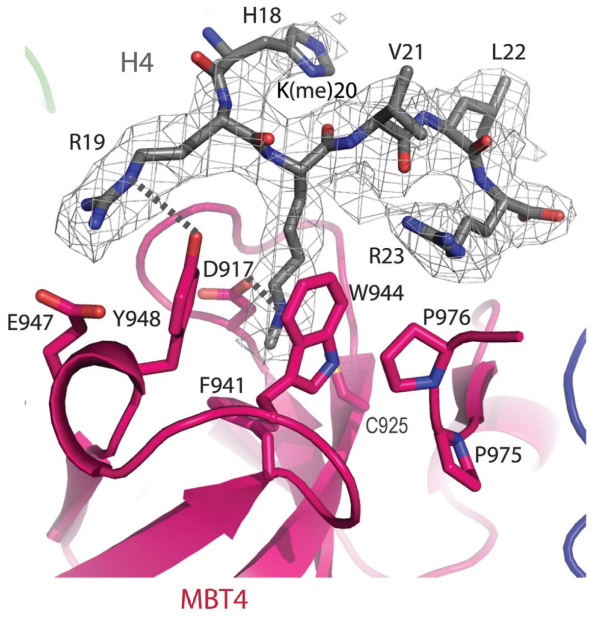

Figure 2 Methyl-lysine peptide recognition by dSfmbt. Details of the bound histone H4K20me1 peptide binding to the aromatic cage pocket within MBT repeat 4. The simulated annealing omit electron-density map for the ligand is shown in wire–frame mode.

triple-mutant dSfmbt$^{E947A/Y948F/D917A}$ proteins completely abolished their ability to bind to H4K20me1 (Table II) without affecting the overall fold and thermal stability of the domain (data not shown). As control, we also tested alanine substitutions of the conserved Asp697 or Asp808 residues at the corresponding positions in the second or third repeat, respectively, (i.e. dSfmbt^{D697A} and dSfmbt^{D808A}) but found that these mutations did not significantly affect peptide binding (Table II).

In summary, these results suggest that dSfmbt binds H4K20me1 with high affinity through the combined interaction of the MBT-binding pocket with the mono-methylated lysine and multiple contacts on the MBT surface with histone residues flanking the methyl-lysine.

Binding of dSfmbt to other methylated histone peptides

Despite the high selectivity of dSfmbt in discriminating between different lysine methylation states, it is able to recognize mono- and dimethylated lysine in a broad range of sequence contexts: dSfmbt also binds histone peptides mono- or dimethylated at H3K4, H3K9, H3K27 or H3K36 with affinities ranging between 1 and 16 μM (Table II). Furthermore, a scrambled H4K20me1 peptide is bound with similar affinity as the native H4K20me1 peptide but more negatively charged peptides such as mono- or dimethylated H3K79 peptides (pI 4.4) are bound with an affinity below 1000 μM (Table II). It thus seems that charge complementarity between the positively charged amino acids in histone-tail peptides (pI values 11–12) and the overall negatively charged dSfmbt surface (Figure 3) rather than recognition of individual residues outside the methyl-lysine-binding pocket is important for the interaction. Given the low sequence specificity, we currently cannot exclude that dSfmbt recognizes methyl-lysine residues in other proteins, although so far only interactions between MBT-repeat proteins and mono- and di-methyl-lysine-containing histone tails have been reported (Kim *et al*, 2006; Trojer *et al*, 2007; Wu *et al*, 2007).

Previous binding studies using fluorescence polarization (FP) assays suggested more pronounced sequence selectivity for dSfmbt binding to H4K20me1/2 and H3K9me1/2 as opposed to binding to H3K4me1/2 or H3K27me1/2 (Klymenko *et al*, 2006). As our ITC measurements reported here provided little evidence for such binding selectivity, we repeated the binding assays with FP assays. To this end, we used a set of peptides that had been produced during the same synthesis reaction as those used for our ITC measurements but, in addition, had been modified by coupling fluorescent carboxylic acid to the N-terminus in the final synthesis step. In FP assays with these peptides, dSfmbt bound H4K20me1/2, H3K4me1/2, H3K9me1/2, H3K27me1/2 and H3K36me1/2 with comparably low micromolar affinities and the determined K_D values were similar to those measured by ITC (Supplementary Table 1). The failure to detect high-affinity binding of dSfmbt to H3K4me1/2 or H3K27me1/2 by Klymenko *et al* (2006) might be because of differences in the method of peptide labelling (i.e. post-synthetic labelling) used in the previous study (W Fischle, personal communication). Taken together, ITC and FP assays reported here both gave comparable results and suggest that mono- and dimethylated lysines in the N-termini of H3 and H4 are all bound with similar micromolar affinities, whereas unmethylated and tri-methylated peptides are bound with much reduced affinity.

Comparison of the dSfmbt, L3MBTL1 and Scm MBT-repeat domains

The three MBT repeats of L3MBTL1 can be superimposed onto dSfmbt repeats 2, 3 and 4 (r.m.s.d.$_{300C\alpha}$ = 6.1 Å, *Z*-score = 22.6) using programme DALI (Holm and Sander, 1993), which identifies repeat 1 as the additional repeat in dSfmbt (Figure 3A). Interestingly, the N-terminal ends of the superimposed L3MBTL1 and dSfmbt structures lie in close vicinity, supporting the hypothesis that repeat 1 of dSfmbt was inserted during evolution. Scm MBT-repeats 1 and 2 can be superimposed onto repeats 1 and 3 of L3MBTL1 (r.m.s.d.$_{200C\alpha}$ = 2.0 Å, *Z*-score = 20.9) and repeats 2 and 4 of dSfmbt (r.m.s.d.$_{193C\alpha}$ = 3.8 Å, *Z*-score = 17.0). Therefore, repeat 2 of L3MBTL1 and the homologous repeat 3 of dSfmbt seem as extra features inserted between the two flanking MBT repeats of Scm (Figure 3A). The high r.m.s.d. between dSfmbt repeats 2–4 and L3MBTL1 repeats 1–3 mainly results from the more open arrangement of the three L3MBTL1 repeats, which are arranged around a central channel running along their three-fold pseudo-symmetry axis (Figure 3B). In the crystal structure, this channel is filled with solvent and bound sucrose molecules used as cryoprotectant. However, it could also serve as additional ligand-binding site as it is lined with conserved residues (Figure 3C).

Scm binds mono-methyl-lysine-containing peptides with dissociation constants of approximately 500 μM (Grimm *et al*, 2007), whereas dSfmbt binds peptides with dissociation constants in the low micromolar range and up to 500 times better than Scm. These differences probably result from their differently charged surfaces (Figure 3B). In Scm, the methyl-lysine-binding pocket is lined by several basic residues (Lys326, Arg352 and His384, Figure 1C), which point towards the positively charged histone-tail peptide. In contrast, the

©2009 European Molecular Biology Organization

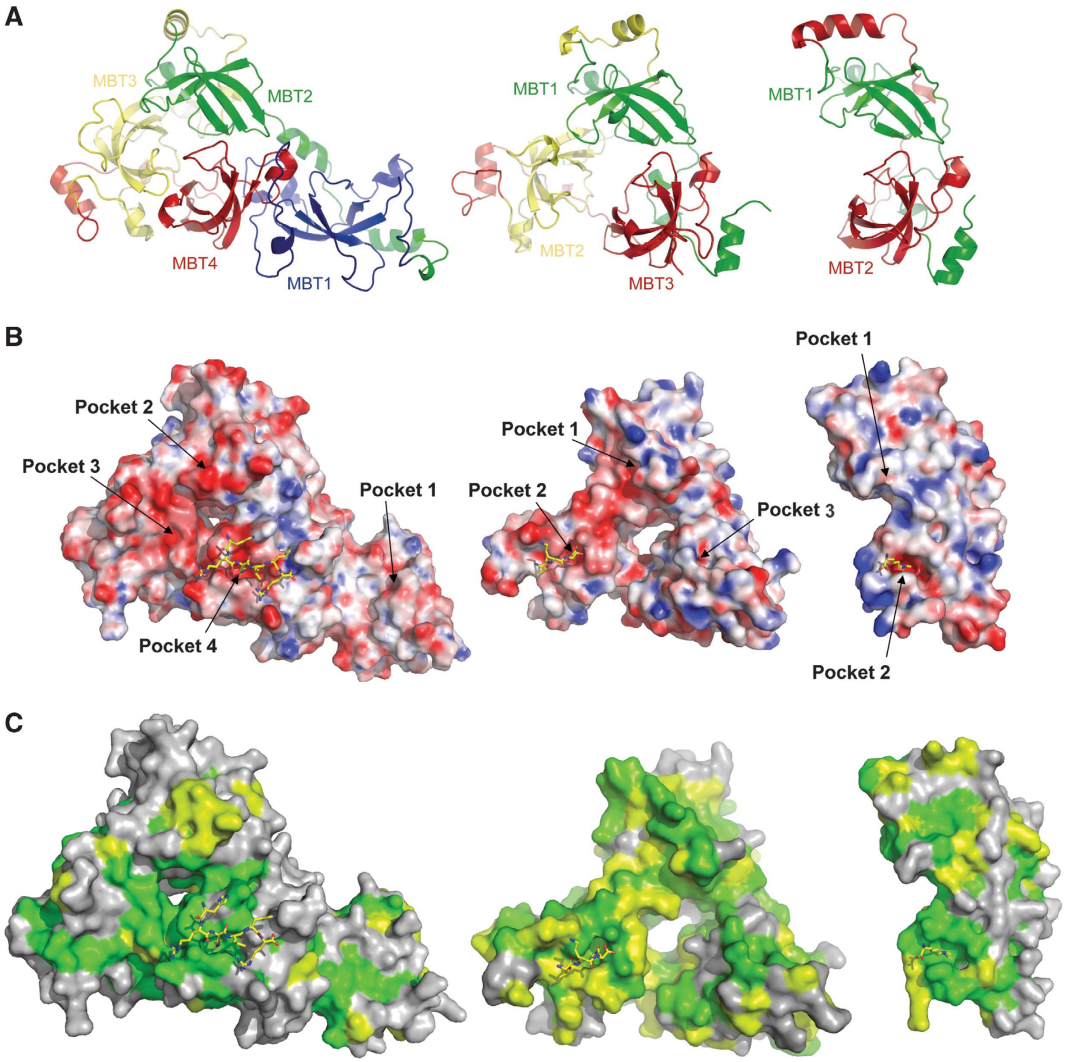

Figure 3 Comparison of the MBT-repeat-domain crystal structures of dSfmbt, L3MBTL1 and Scm. (**A**) Ribbon diagram of dSfmbt (left), L3MBTL1 (middle) and Scm (right). Equivalent MBT repeats as indicated by comparison of their tertiary structures are depicted with equivalent colours. (**B**) Electrostatic surface representation of dSfmbt, L3MBTL1 and Scm. The bound peptide ligands are depicted in yellow. In L3MBTL1, the methyl-lysine peptide is bound to MBT repeat 2. (**C**) Comparison of the surface conservation in dSfmbt, L3MBTL1 and Scm. Conserved regions with >50% sequence conservation are depicted in colour, dark green corresponds to strictly conserved residues. For surface comparison, orthologous sequences were aligned as depicted in Supplementary Figure S2. The following sequences were used for the alignments: dSfmbt: *Drosophila melanogaster*, Q9VK33, corresponds to the dSfmbt-4MBT crystal structure reported here; *Anopheles gambiae,* Q7Q0R1; *Xenopus laevis*, Q32N90; *Mus musculus*, P59178; *Homo sapiens*, Q05BQ5; *Gallus gallus*, Q5ZLC2; *Tetraodon nigroviridis*, Q4T9N5. L3MBTL1: *Homo sapiens,* Q9Y468, corresponds to the L3MBTL1 crystal structure (PDB accession code 2RHI); *Bos tauru,* Q08DF3; *Gallus gallus,* XP_417302; *Danio rerio* XP_699604. Scm: *Drosophila melanogaster,* Q9VHA0, as present in the Scm crystal structure (PDB accession code 2R57); *Xenopus tropicalis*, Q0IHT6; *Homo sapiens,* SCML2; *Ciona intestinalis* Q4H2U6.

corresponding dSfmbt residues (Met919, Thr945 and Pro976) are uncharged and assist in peptide binding. In L3MBTL1, the corresponding residues (Met357, Asp383 and Asp415) can also assist in peptide binding, however, the negatively charged area around the methyl-lysine-binding pocket is less extended compared with dSfmbt (Figure 3B), which might explain the lower binding affinity.

Multiple binding sites in MBT-repeat proteins

Superposition of the fourth MBT repeat of dSfmbt with the three other repeats (Figure 4) shows that only the fourth repeat can accommodate methyl-lysine residues. In repeat 1, the crucial aspartate is substituted by an asparagine (Figure 1C), but more importantly the conformation of the loop bearing this residue is different. In the second repeat, two of the cage-forming aromatic residues are substituted by aspartate and serine, respectively, and in the third repeat, Tyr836 blocks the access of the methyl-lysine to the binding pocket. The MBT proteins, Scm and L3MBTL1, use their second MBT repeat for methyl-lysine binding and, indeed, the cage-forming residues, including Cys925 are well conserved in the second MBT repeat of L3MBTL1 and in the second repeat of Scm. In contrast, in MBT repeats 1 and 3 of human L3MBTL1, Cys925 is substituted by bulkier residues that block the access to the binding pocket, whereas in Scm the cage-forming aromatic residues are substituted by smaller residues. In dSfmbt and Scm, conserved residues cluster around the methyl-lysine binding pocket, whereas the patch of strictly conserved residues is smaller in L3MBTL1 (Figure 3C).

 ©2009 European Molecular Biology Organization

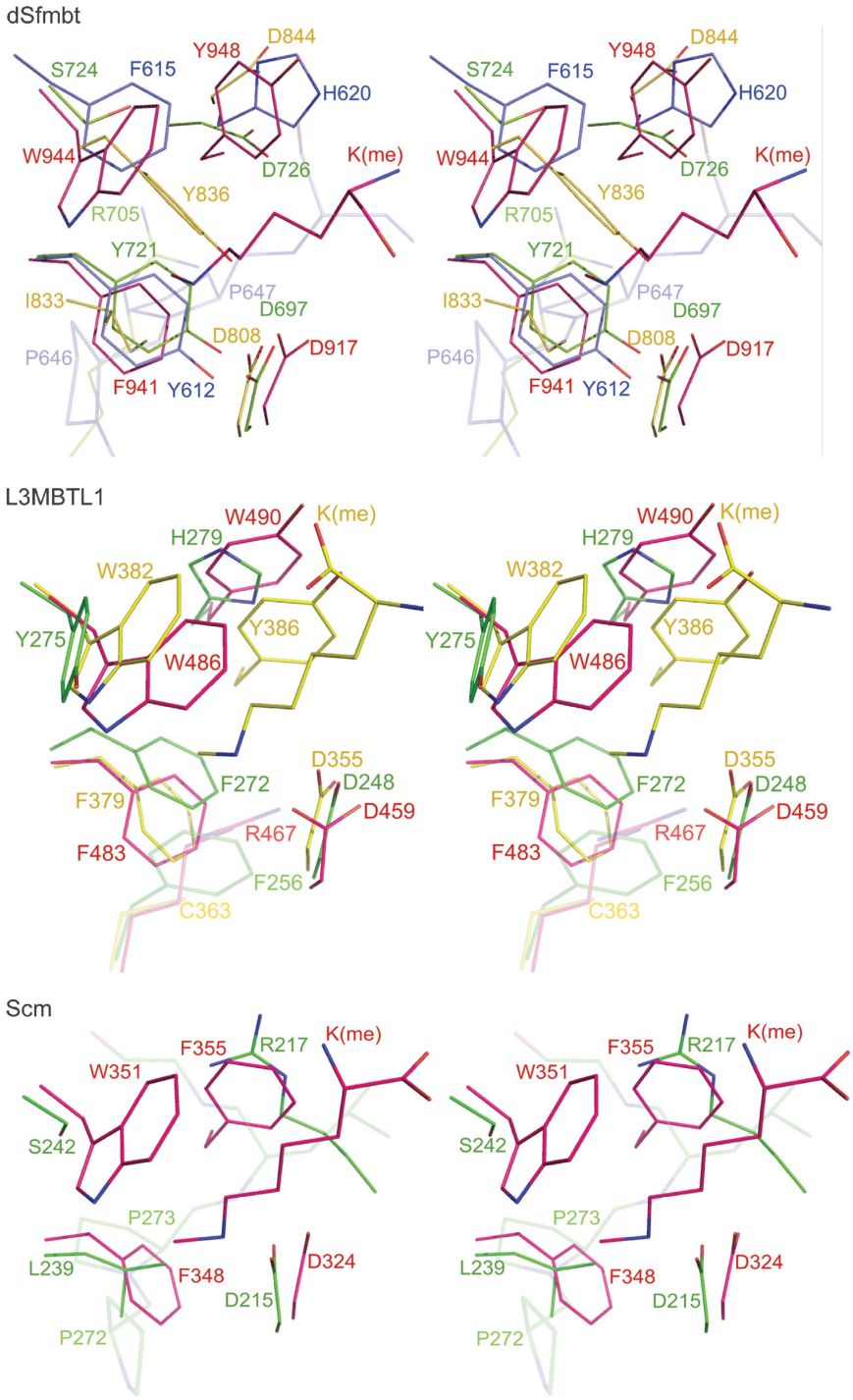

Figure 4 Stereo view of superpositons of the MBT-repeat domains of dSfmbt, L3MBTL1 and Scm. Colour code corresponds to Figure 3 with dSfmbt repeats 1, 2, 3 and 4 depicted in blue, green, yellow, and red (top), L3MBTL1 repeats 1, 2 and 3 in green, yellow and red (middle) and Scm repeats 1 and 2 depicted in green and red (bottom).

In conclusion, only a single MBT repeat in Scm, L3MBTL1 and dSfmbt can bind mono- and dimethylated lysine residues. It is possible that the other MBT repeats recognize other ligands. Indeed, in one of the crystal structures of L3MBTL1, the first MBT repeat binds a Pro–Ser-motif-containing peptide of a neighbouring molecule (Li *et al*, 2007), although the functional relevance of this interaction is not known.

dSfmbt and Scm interact functionally to maintain Polycomb repression

Previous structural/functional analyses of the MBT-repeat domain of Scm showed that a point mutation in the methyl-lysine-binding pocket that abolishes the methyl-lysine binding, or even complete deletion of the MBT-repeat domain, still permit these mutant Scm proteins to partially maintain PcG repression of target genes in a genetic-rescue

©2009 European Molecular Biology Organization

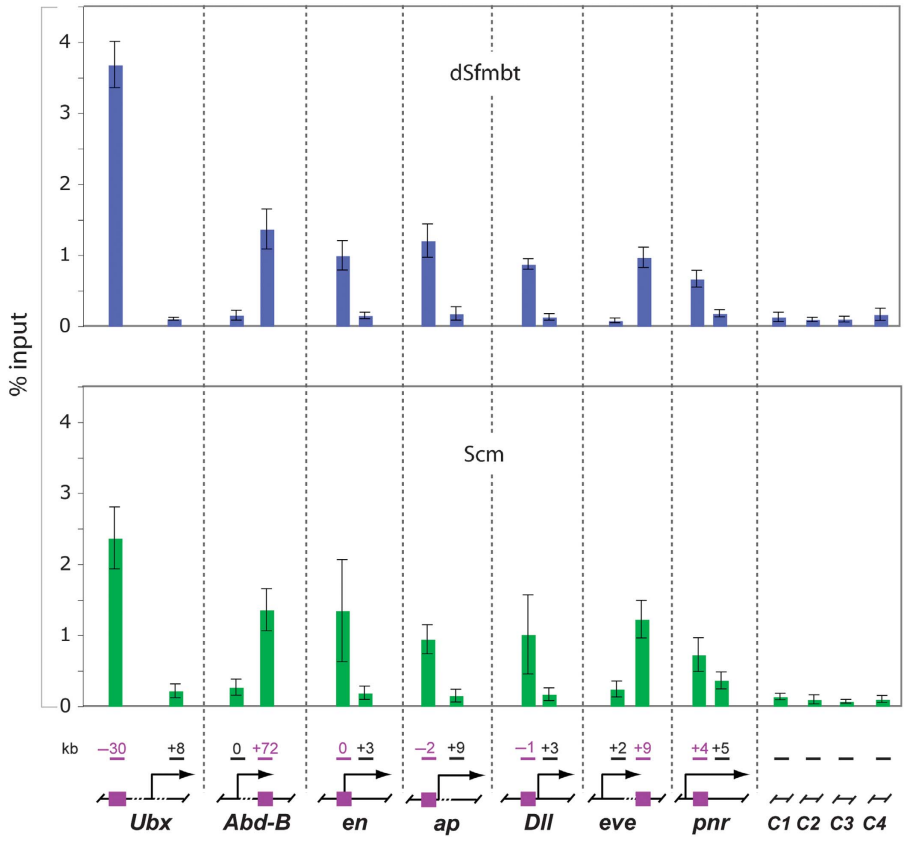

Figure 5 dSfmbt and Scm co-bind to PREs in PcG target genes. ChIP analysis monitoring dSfmbt and Scm binding in imaginal disc/CNS tissues dissected from wild-type *Drosophila* larvae. Graphs show the results from three independent immunoprecipitation reactions from different batches of chromatin preparations; ChIP signals were quantified by qPCR and are presented as percentage of input chromatin precipitated at each region, error bars correspond to s.d. The location of PREs (purple boxes) and other regions with respect to transcription start sites in the *Ubx*, *Abd-B*, *en*, *ap*, *Dll*, *eve* and *pnr* genes are indicated in kilobases; C1–C4 indicate euchromatic and heterochromatic control regions outside these genes (see Supplementary Table 2 for qPCR primer sequences). dSfmbt and Scm proteins are specifically enriched at the PRE of each gene but not at the analyzed intervals in the coding regions of the same genes or in control regions C1–C4.

assay in *D. melanogaster* (Grimm *et al*, 2007). Similar observations were made with dSfmbt; we found that not only the wild-type dSfmbt protein but also the dSfmbt$^{E947A/Y948F/D917A}$ protein (see above) is able to maintain PcG repression of target genes in a genetic-rescue assay in *dSfmbt* null mutants (data not shown). One possible explanation for these findings would be that methyl-lysine binding by the MBT domains of dSfmbt and Scm has only a minor function in PcG repression. However, as both the proteins have similar methyl-lysine-binding activities, an alternative possibility could be that the MBT-repeat domains in Scm and dSfmbt function in a partially redundant manner to maintain PcG repression.

We therefore carried out a set of experiments to test whether and how dSfmbt and Scm might interact. First, we analyzed the binding of dSfmbt and Scm at PcG target genes *in vivo*. We recently reported the genome-wide binding profile of dSfmbt in developing *Drosophila* larvae (Oktaba *et al*, 2008). However, chromatin immunoprecipitation (ChIP) assays that monitor the binding of Scm have not yet been reported. We therefore carried out ChIP assays with antibodies against Scm and dSfmbt in imaginal-disc tissues from *Drosophila* larvae. These analyses showed that both proteins are specifically bound at PREs of the PcG target genes *Ubx*, *Abd-B*, *en*, *ap*, *Dll*, *eve* and *pnr* (Figure 5). Scm and dSfmbt are thus co-bound at PREs in *Drosophila*.

We next tested for the functional redundancy between dSfmbt and Scm in the repression of these target genes. To this end, we removed *dSfmbt* function in animals that lack wild-type Scm protein and instead express the MBT-mutant protein Scm^{D215N}. Specifically, we induced clones of *dSfmbt* null-mutant cells in *Scm^{D215N}* mutant *Drosophila* larvae and analyzed the clones of *dSfmbt Scm^{D215N}* double-mutant cells for mis-expression of PcG target genes. In the wing imaginal disc, cell clones lacking dSfmbt show widespread mis-expression of the PcG target gene *Ubx* (Klymenko *et al*, 2006), but they do not show mis-expression of *Abd-B* (Figure 6). Similarly, *Abd-B* is not mis-expressed in wing imaginal discs of *Scm^{D215N}*-mutant animals (Figure 6). In striking contrast, *Abd-B* is strongly mis-expressed in clones of *dSfmbt Scm^{D215N}* double-mutant cells (Figure 6). A similar strong synergy between these two Polycomb repressor proteins is observed at the *en* gene. In imaginal discs with *dSfmbt* single-mutant clones, *en* is only mis-expressed in a subset of clones in specific regions of the disc but remains repressed in other parts of the disc, and *en* is not mis-expressed in *Scm^{D215N}* single mutants. In contrast, *en* is strongly mis-expressed in clones of *dSfmbt Scm^{D215N}* double-mutant cells (Figure 6). In addition, *dSfmbt Scm^{D215N}* double-mutant cell clones show a tumour-like phenotype that is characterized by unrestricted cell proliferation (Figure 6). This phenotype is not observed in either of the

 ©2009 European Molecular Biology Organization

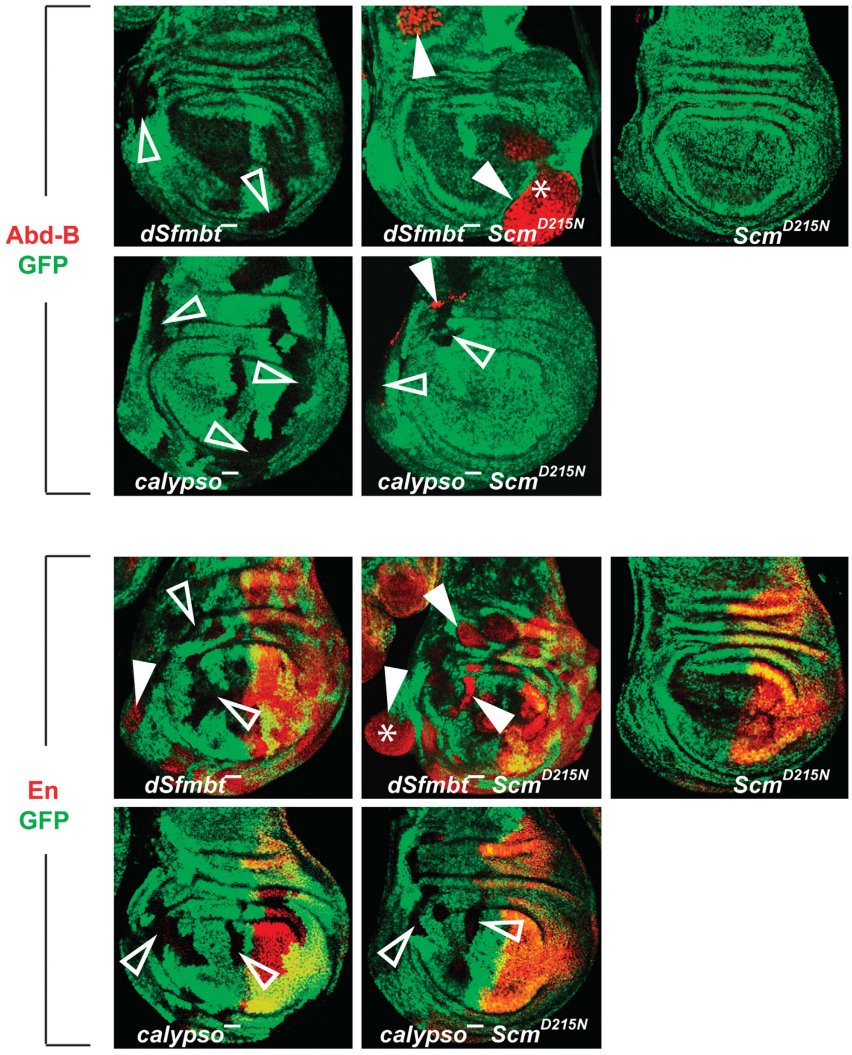

Figure 6 dSfmbt and Scm interact functionally to maintain Polycomb repression. dSfmbt and Scm act redundantly to maintain repression of Polycomb target genes *Abd-B* and *en* in *Drosophila*. Wing imaginal discs stained with antibodies against Abd-B (red, top) or En protein (red, bottom) as indicated. Left: discs with clones of *dSfmbt* or *calypso* single-mutant cells that are marked by the absence of nuclear GFP. Right: disc from Scm^{D215N}-mutant larvae; these animals were trans-heterozygous for Scm^{D215N} and the protein null mutation Scm^{D1} (Bornemann *et al*, 1998) and all cells thus express Scm^{D215N} instead of wild-type Scm protein, nuclear GFP was used here to show all nuclei. Middle: Scm^{D215N}/Scm^{D1} mutant discs with clones of *dSfmbt* or *calypso*-mutant cells; the *dSfmbt* Scm^{D215N} double-mutant and *calypso* Scm^{D215N} double-mutant cells, respectively, are GFP-negative. Abd-B is not expressed in wild-type wing discs, remains repressed in *dSfmbt* or *calypso* single-mutant cells (left, empty arrowheads) or in Scm^{D215N}-mutant discs (right) but is strongly mis-expressed in *dSfmbt* Scm^{D215N} double-mutant cells (middle, arrowheads). In clones of *calypso* Scm^{D215N} double-mutant cells (middle), Abd-B is mis-expressed in a small fraction of clone cells (arrowhead) but remains repressed in the majority of clone cells (empty arrowheads). En expression is confined to the posterior-compartment cells of wild-type imaginal discs and this pattern is unchanged in Scm^{D215N}-mutant discs (right); En remains repressed in *dSfmbt* or in *calypso* single-mutant clones in the anterior compartment (left, empty arrowheads) with the exception of some *dSfmbt*-mutant clones in the hinge that show mis-expression of En (filled arrowhead). Note that En is strongly mis-expressed in almost all *dSfmbt* Scm^{D215N} double-mutant clones in the anterior compartment (middle, arrowheads) but remains repressed in *calypso* Scm^{D215N} double-mutant clones. Note that only *dSfmbt* Scm^{D215N} but not *calypso* Scm^{D215N} double-mutant clones show the tumour-like phenotype (asterisks, see text for details).

single mutants (Figure 6) but is characteristic of cell clones lacking the PRC1 components Psc–Su(z)2 or Ph (Oktaba *et al*, 2008).

To test whether this strong genetic interaction between *dSfmbt* and *Scm* was specific, we used the same strategy to remove the function of the PcG gene *calypso* (Gaytán de Ayala Alonso *et al*, 2007) in Scm^{D215N}-mutant *Drosophila* larvae. Like in the case of *dSfmbt*, clones of *calypso* single-mutant cells in the wing imaginal disc show mis-expression of *Ubx* (Gaytán de Ayala Alonso *et al*, 2007) but maintain repression of *Abd-B* and *en* (Figure 6). In clones of *calypso* Scm^{D215N} double-mutant cells, *en* remains fully repressed, and the clones do not show the tumour-like phenotype observed in *dSfmbt* Scm^{D215N} double-mutant clones (Figure 6). *Abd-B* becomes mis-expressed in a fraction of *calypso*-Scm^{D215N} clone cells but mis-expression is much less extensive than in *dSfmbt* Scm^{D215N} double-mutant clones (Figure 6). Removal of *dSfmbt* function in Scm^{D215N}-mutant animals therefore results in a much more severe Polycomb phenotypes compared with when *calypso* is removed in this genetic background. Taken together, these results suggest a particularly strong synergy between the PhoRC-component dSfmbt and the PRC1-component Scm in the repression of target genes and the control of cell proliferation.

©2009 European Molecular Biology Organization

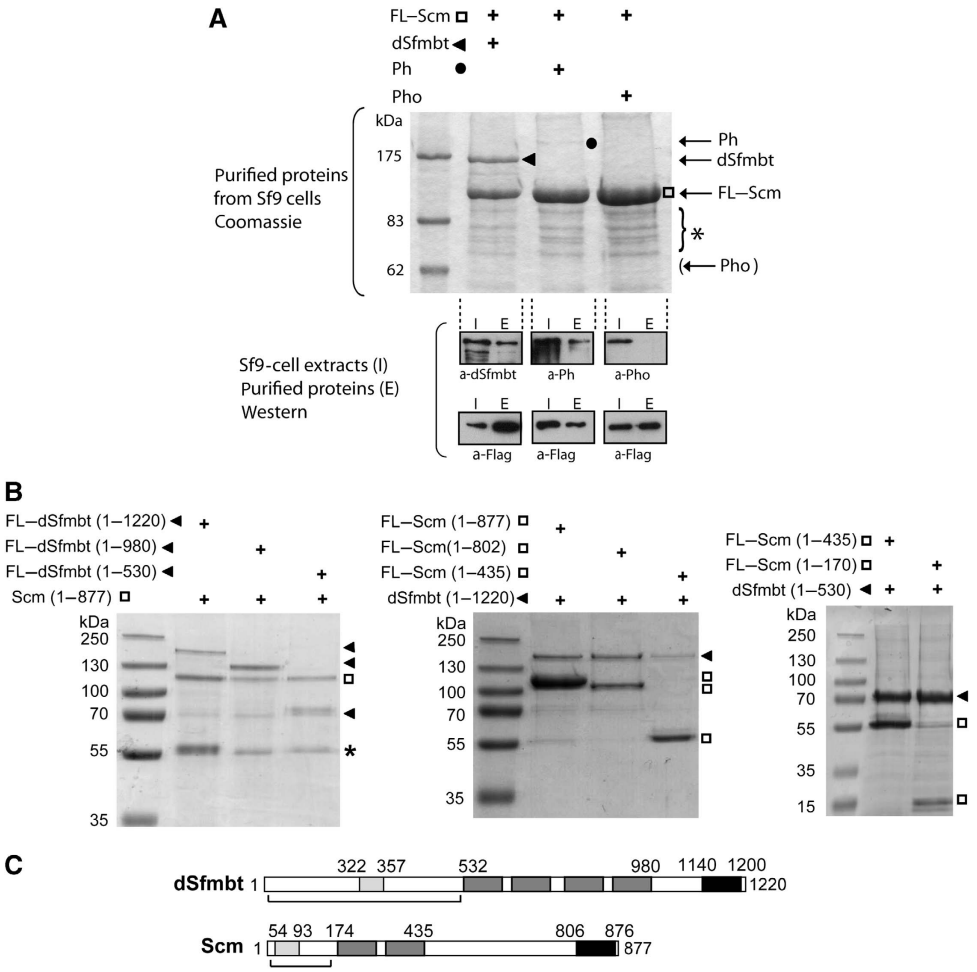

Figure 7 Reconstitution of Scm–dSfmbt complexes. (**A**) FLAG-tagged Scm and untagged dSfmbt, Ph or Pho proteins were affinity purified by FLAG-tag, separated by SDS–PAGE and visualized by Coomassie staining (top). Western blot of corresponding Sf9 total cell-extract input before purification (I) and eluted purified proteins (E) to show relative enrichment of proteins after purification (below). Note that dSfmbt (black arrowhead) forms a stable complex with Flag–Scm (square) whereas Ph (dot) co-purifies less efficiently with Flag–Scm than dSfmbt. Also note that co-expression of Pho with Flag–Scm results in the purification of only Flag–Scm and Pho is not detected by Coomassie staining or western-blot analysis of the eluted purified material. Asterisk marks Flag–Scm degradation products. (**B**) Immunopurification of different Flag-tagged dSfmbt constructs with full-length Scm (left panel), Flag-tagged Scm constructs with full-length dSfmbt (middle panel) and of Flag-tagged Scm constructs with C-terminally truncated dSfmbt (right panel). Arrowheads and squares indicate dSfmbt and Scm constructs, respectively. Degradation products of Scm protein are indicated by an asterisk. (**C**) Domain organization of dSfmbt and Scm. Zn-finger domain, MBT repeats and SAM domain are depicted in light grey, grey and black, respectively. Domain borders used for the Scm and dSfmbt constructs are indicated. Brackets indicate the regions of dSfmbt and Scm minimally required for interaction.

Direct interaction between dSfmbt and Scm proteins

The strong genetic interaction between dSfmbt and Scm prompted us to test whether these two proteins might also physically interact with each other. To this end, we co-expressed Scm and dSfmbt in Sf9 cells using baculovirus and tested whether they form a stable complex, which can be purified from Sf9 cell extracts. As controls, we co-expressed Scm along with the PhoRC-component Pho or with Ph, the PRC1 component that had been reported to interact with Scm (Peterson *et al*, 1997, 2004). Flag-affinity purification from extracts of Sf9 cells that co-express Flag–Scm and untagged dSfmbt resulted in the isolation of a stable Scm–dSfmbt complex (Figure 7A). Ph also interacted weakly with Scm under the same assay condition but Pho did not form any complex with Scm (Figure 7A).

In the next step, we used C-terminal truncations of Scm and dSfmbt to define the interacting regions between the two proteins with a greater precision. N-terminal Flag-tagged dSfmbt constructs lacking the C-terminal SAM domain and the MBT repeats were still able to interact with Scm (Figure 7B, left panel), and the N-terminal Flag-tagged Scm constructs still interacted with the full-length and C-terminally truncated dSfmbt, also lacking the SAM domain and the MBT repeats (Figure 7B, middle and right panel). Our results identify the N-terminal moieties of dSfmbt and Scm containing Zn-finger motifs as the interacting regions (Figure 7C). Interestingly, interaction between Scm and dSfmbt does not seem to depend on the SAM domains. SAM domains of Scm and Ph form homo-polymeric structures, but are also thought to form Scm–Ph hetero-polymers (Kim *et al*, 2005). C-terminally truncated Scm lacking the SAM domain does no longer interact with Ph (Supplementary Figure S3), although it still binds to dSfmbt.

Our finding that Scm and dSfmbt can be isolated as a stable complex from Sf9 cells was somewhat unexpected because the biochemically purified PhoRC from *Drosophila* embryos

 ©2009 European Molecular Biology Organization

does not include Scm and similarly biochemically purified PRC1 contains substoichiometric quantities of Scm but no dSfmbt (Saurin *et al*, 2001; Klymenko *et al*, 2006). The failure to isolate dSfmbt–Scm complexes from *Drosophila* embryonic nuclear extracts may have different reasons. It could be that the dSfmbt–Scm interaction is weaker and becomes disrupted during complex purification. Alternatively, Scm and dSfmbt might interact only under certain conditions (i.e. once both are tethered to chromatin). Taken together, our genetic data, ChIP experiments and physical-interaction data show that dSfmbt and Scm interact directly and cooperate in a highly synergistic manner to maintain Polycomb repression.

Concluding remarks

Our results show how the MBT-repeat domain of dSfmbt binds mono- or dimethyl-lysine–containing histone-tail peptides. The binding affinity of dSfmbt for methylated lysines in the histone H3 and H4 N-termini is in the low micromolar range and is thus comparable to that of heterochromatin protein-1 or the double bromodomain of TAF250 that recognize modified histone lysines in specific sequence contexts (Ruthenburg *et al*, 2007). However, despite its high selectivity for different states of lysine methylation, dSfmbt-MBT recognizes mono- and dimethylated lysines in various sequence contexts.

This broad binding specificity may be important for dSfmbt function within the PhoRC complex. Genome-wide-binding profiling showed that dSfmbt occupies 50% of its targets sites together with Pho, suggesting that dSfmbt is bound to those regions as a part of the PhoRC complex (Oktaba *et al*, 2008). Previous studies showed that dSfmbt binding at HOX genes crucially depends on Pho-protein-binding sites in PREs, and it is thus the DNA-binding activity of Pho that targets PhoRC to the genes it regulates (Klymenko *et al*, 2006). Similarly, L3MBTL1 is associated with an E2F–RBF complex (Lewis *et al*, 2004) and it thus seems likely that the association of L3MBTL1 with E2F target genes (Trojer *et al*, 2007) is mediated by the DNA-binding factor E2F. Histone methyl-lysine binding by these MBT-repeat proteins thus does not seem to be involved in targeting. Instead, the chromatin environment flanking Pho target sites may dictate which particular mono- and dimethylated lysines are recognized by dSfmbt *in vivo*.

What is the role of methyl-lysine binding of MBT-repeat proteins? It has been proposed that DNA-tethered MBT proteins use this binding activity for interactions with modified nucleosomes in the flanking chromatin to maintain a repressed-chromatin state (Klymenko *et al*, 2006; Trojer *et al*, 2007). The repeat structure of MBT-domain proteins also led to the suggestion that a single MBT-repeat domain could simultaneously recognize several methylation marks (Li *et al*, 2007; Trojer *et al*, 2007), which would provide a molecular mechanism for the observed chromatin compaction by L3MBTL1 *in vitro* (Trojer *et al*, 2007). However, the structure of the dSfmbt MBT-repeat domain bound to the H4K20me1 peptide and also the structures of L3MBTL1 and Scm bound to methyl-lysine-containing peptides (Grimm *et al*, 2007; Li *et al*, 2007; Min *et al*, 2007; Santiveri *et al*, 2008) argue against such a model. Only a single methyl-lysine-binding pocket is present in all MBT-repeat proteins, whereas the corresponding 'pockets' in the other repeats are shallower and less well conserved. Moreover, there is no biophysical evidence for simultaneous interaction with multiple methylated histone-tail peptides.

The physical and genetic interaction between dSfmbt and Scm suggests a close cooperation of these two proteins in Polycomb repression. Both proteins possess a similar methyl-lysine-binding capacity because of their MBT domains. It is therefore tempting to speculate that dSfmbt–Scm complexes may recognize methylated lysines in two different nucleosomes. Heterodimerization of dSfmbt and Scm with the MBT-repeat domain of each protein bound to a methylated-histone tail could provide a plausible mechanism for chromatin compaction.

Materials and methods

Protein expression and purification

Wild-type and mutant constructs of the four-MBT-repeat domain from *D. melanogaster* dSfmbt were generated using standard PCR and restriction-cloning techniques and the bacterial expression vector pETM11. The dSfmbt-4MBT protein and all variants were overexpressed in *Escherichia coli* strain BL21(DE3) as TEV-protease-cleavable N-terminal His6-fusion proteins at 18°C for 15 h. The cleared bacterial lysate in 50 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10% (v/v) glycerol, 10 mM imidazole and 2 mM β-mercaptoethanol was incubated with Ni^{2+}–NTA Sepharose (Qiagen) and the recombinant protein recovered by elution with imidazole followed by incubation with His-tagged TEV protease (0.01% w/w, overnight, 4°C). After dialysis to remove the imidazol, the protease was removed by incubation with Ni^{2+}–NTA Sepharose. The final purification step comprised a gel-filtration step using a Superdex-200 column (GE Healthcare) in a buffer containing 10 mM Tris–HCl (pH 8.0), 150 mM NaCl and 5 mM DTT and protein at concentration of 30 mg/ml.

Crystallization and data collection

Wild-type dSfmbt-4MBT protein (residues 532–980) was crystallized using the hanging-drop method by mixing 5 μl of protein solution at 50 mg/ml with 5 μl of reservoir solution (0.8 M sodium acetate, 100 mM imidazole, pH 6.5). Crystals were cooled for data collection to 100K in the mother liquor containing 25% (v/v) glycerol as cryoprotectant. Crystals diffracted only to 3.2 Å resolution at the ESRF synchrotron and belonged to space-group C222 with three molecules in the asymmetric unit. Three point mutations (K715D, R886S and R900D) were introduced on the surface of a slightly shorter dSfmbt-4MBT construct (residues 535–977). Co-crystals of this mutant construct with peptide RHRKme1VLR were obtained by mixing protein solution at 15 mg/ml in presence of 3 mg/ml peptide with 3.7 M NaCl as the precipitant. These crystals diffracted to 2.8 Å at 100K in the mother liquor containing 35% (w/v) sucrose as cryoprotectant and belonged to space group $P22_12_1$ with two molecules in the asymmetric unit.

Phase determination and refinement

The structure of wild-type dSfmbt-4MBT (residues 532–980) was solved by a two-wavelength MAD experiment in crystal form C222 using a mercury derivate. Four heavy-atom sites were identified using program SOLVE (Terwilliger and Berendzen, 1999). Coordinates for these sites were refined and five more sites were identified using program SHARP (de la Fortelle and Bricogne, 1997). The resulting experimental phases were further improved by solvent flattening and averaging using program DM (Cowtan and Zhang, 1999). In the resulting electron density, the MBT core fold could be located and a partial model could be built. However, the remaining parts of the molecule were disordered and the poor quality of the electron density in these regions prevented us from building a complete model. Molecular replacement was carried out using this partial model and a dataset from the $P22_12_1$ crystals at 2.8 Å resolution using program PHASER (McCoy *et al*, 2005), which yielded a solution for two molecules. The resulting electron density maps allowed us to complete the missing parts of the model and to locate and to build the bound peptide. Several rounds of manual building using program O (Jones and Kjeldgaard, 1997) and

©2009 European Molecular Biology Organization

automated refinement using program REFMAC, including TLS refinement (Murshudov *et al*, 1997) led to a final model with excellent geometry (Table I).

ITC and FP measurements

ITC was carried out using a VP-ITC Microcal calorimeter (Microcal, Northhampton, MA, USA). Peptides were purified by reverse-phase HPLC in the presence of trifluoroacetic acid. To remove traces of trifluoroacetic acid, dry-peptide samples were treated with 25 mM ammonium bicarbonate followed by lyophilization and resuspended in ITC buffer. Before all titrations, proteins were dialysed extensively against ITC buffer (20 mM Tris–HCl (pH 8.0), 20 mM or 150 mM NaCl, 2 mM β-mercaptoethanol). The experiments were carried out at 25°C. A typical titration consisted of injecting 5–10 μl aliquots of 1–5 mM peptide into a solution of 50–200 mM dSfmbt-4MBT protein at time intervals of 5 min to ensure that the titration peak returned to the baseline. The ITC data were analyzed and corrected for the heat of dilution of peptides in the absence of protein using program Origin version 5.0 provided by the manufacturer.

Fluorescein-labelled peptides were synthesized at Protein Specialty Laboratories, Heidelberg. FP assays were carried out at 20 mM Tris–HCl (pH 8.0), 20 mM NaCl, 2 mM β-mercaptoethanol using fluorescein-labelled peptides at an 80 nM concentration on a Synergy 4 instrument (BioTek Instruments). To calculate the K_D values the experimental data were imported and analyzed by program Origin 7.5 as previously described (Jacobs *et al*, 2004).

Flag-affinity purification of Scm–dSfmbt complexes

Baculoviruses expressing full-length Ph, Pho and dSfmbt have been described earlier (Francis *et al*, 2001; Klymenko *et al*, 2006). Flag–Scm_{1-877} was a gift from Jeff Simon. The detailed plasmid maps of Scm and dSfmbt constructs used in this study are available on request.

Sf9 cells were co-infected for 48 h with untagged dSfmbt and with different Flag–Scm constructs or with untagged Scm and Flag–dSfmbt construct. The whole-cell extracts were prepared according to Klymenko *et al* (2006). 0.2 ml anti-Flag beads (Sigma) were used for 10 ml of extracts. Binding was carried out overnight at 4°C in extraction buffer A (20 mM Tris–HCl (pH 8.0), 300 mM NaCl, 20% (v/v) glycerol, 4 mM $MgCl_2$, 0.4 mM EDTA and 2 mM DTT) with 0.05% NP40, 10 μM $ZnCl_2$ and 1 tablet complete protease inhibitor cocktail (Boehringer) for 50 ml lysis buffer. Beads were extensively washed with increasing concentrations of KCl up to 1.2 M in buffer B (20 mM Hepes (pH 7.9), 0.4 mM EDTA and 20% (v/v) glycerol with 0.05% NP40, 0.2 mM protease inhibitors and 0.5 mM DTT). Beads were eluted at 4°C with 0.4 mg/ml Flag peptide in buffer B, containing 300 mM KCl. The supernatant was analyzed by SDS–PAGE followed by Coomassie staining.

Functional analysis of dSfmbt and Scm in imaginal discs

Imaginal discs were dissected from third instar larvae that were produced by crossing the appropriate mutant fly strains listed below:

yw hs–flp; hs–nGFP FRT40
yw hs–flp; [hs–nGFP FRT40; $Scm^{Su(z)302}$]/SM5-TM6B
w; $dSfmbt^1$ FRT40/SM6B
w; FRT82 Scm^{D1}/TM6C
w; [$dSfmbt^1$ FRT40; FRT82 Scm^{D1}]/SM5-TM6B
yw; FRT40 FRT42D P[y^+] $calypso^2$/SM6B
yw hs–flp; [FRT40 FRT42D P[y^+] $calypso^2$; $Scm^{Su(z)302}$]/SM5-TM6B
yw hs–flp; [FRT42D hs–nGFP; Scm^{D1}]/SM5-TM6B

Note, the $Scm^{Su(z)302}$ allele encodes Scm^{D215N}.

Clone induction and staining of imaginal discs with antibodies against Abd-B (Celniker *et al*, 1989) or En (mouse monoclonal 4D9) was done as described earlier (Beuchle *et al*, 2001).

Accession code

Protein Data Bank: Atomic coordinates and structure factors for the dSfmbt-4MBT–histone H4K20me1 peptide complex have been deposited under accession code 3H6Z.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank J Simon for the gift of the Flag–Scm baculovirus expression vector. RM is supported by a grant from the Deutsche Forschungsgemeinschaft. We thank the EMBL–ESRF Joint Structural Biology Group for access and support at the ESRF beamlines. We also acknowledge the support of the crystallization facility of the Partnership for Structural Biology, Grenoble and the proteomic core facility at EMBL, Heidelberg.

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

Celniker SE, Keelan DJ, Lewis EB (1989) The molecular genetics of the bithorax complex of Drosophila: characterization of the products of the Abdominal-B domain. *Genes Dev* **3:** 1424–1436

Cowtan KD, Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* **72:** 245–270

de la Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276:** 472–494

Francis NJ, Kingston RE (2001) Mechanisms of transcriptional memory. *Nat Rev Mol Cell Biol* **2:** 409–421

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

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

Grimm C, de Ayala Alonso AG, Rybin V, Steuerwald U, Ly-Hartig N, Fischle W, Muller J, Muller 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

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233:** 123–138

Jacobs SA, Fischle W, Khorasanizadeh S (2004) Assays for the determination of structure and dynamics of the interaction of the chromodomain with histone peptides. *Methods Enzymol* **376:** 131–148

Jones TA, Kjeldgaard M (1997) Electron-density map interpretation. *Methods Enzymol* **277:** 173–208

Kim CA, Sawaya MR, Cascio D, Kim W, Bowie JU (2005) Structural organization of a Sex-comb-on-midleg/Polyhomeotic copolymer. *J Biol Chem* **280:** 27769–27775

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

Lewis PW, Beall EL, Fleischer TC, Georlette D, Link AJ, Botchan MR (2004) Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev* **18:** 2929–2940

Li H, Fischle W, Wang W, Duncan EM, Liang L, Murakami-Ishibe S, Allis CD, Patel DJ (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

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61:** 458–464

 ©2009 European Molecular Biology Organization

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

Mohd-Sarip A, Cleard F, Mishra RK, Karch F, Verrijzer CP (2005) Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes Dev* **19:** 1755–1760

Muller J, Verrijzer CP (2009) Biochemical mechanisms of gene regulation by Polycomb group protein complexes. *Curr Opin Genet Dev* **19:** 150–158

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255

Oktaba K, Gutierrez L, Gagneur J, Girardot C, Sengupta AK, Furlong EEM, Muller J (2008) Dynamic regulation by Polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev Cell* **25:** 877–889

Peterson AJ, Kyba M, Bornemann D, Morgan K, Brock HW, 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 AJ, Mallin DR, Francis NJ, Ketel CS, Stamm J, Voeller RK, Kingston RE, Simon JA (2004) Requirement for sex comb on midleg protein interactions in Drosophila Polycomb group repression. *Genetics* **167:** 1225–1239

Ruthenburg AJ, Li H, Patel DJ, Allis CD (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* **8:** 983–994

Santiveri CM, Lechtenberg BC, Allen MD, Sathyamurthy A, Jaulent AM, Freund SM, Bycroft M (2008) The Malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J Mol Biol* **382:** 1107–1112

Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE (2001) A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature* **412:** 655–660

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

Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55:** 849–861

Trojer P, Li G, Sims III RJ, Vaquero A, Kalakonda N, Boccuni P, Lee D, Erdjument-Bromage H, Tempst P, Nimer SD, Wang YH, Reinberg D (2007) L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **129:** 915–928

Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS (2004) Hierarchical recruitment of Polycomb group silencing complexes. *Mol Cell* **14:** 637–646

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

Wu S, Trievel RC, Rice JC (2007) Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. *FEBS Lett* **581:** 3289–3296

SOME RIGHTS RESERVED

The EMBO Journal **is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Licence. [http://creativecommons.org/licenses/by-nc-nd/3.0]**