Structure of an atypical Tudor domain in the *Drosophila* Polycomblike protein

Anders Friberg,[1,2] Anna Oddone,[3,4] Tetyana Klymenko,[3] Jürg Müller,[3] and Michael Sattler[1,2]*

[1]Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
[2]Department Chemie, Munich Center for Integrated Protein Science at Chair of Biomolecular NMR, Technische Universität München, Lichtenbergstr. 4, 85747 Garching
[3]European Molecular Biology Laboratory (EMBL), Meyerhofstr. 1, 69117 Heidelberg, Germany
[4]Centre for Genomic Regulation (CRG), Doctor Aiguader 88, 08003 Barcelona, Spain

Received 19 May 2010; Revised 12 July 2010; Accepted 14 July 2010
DOI: 10.1002/pro.476
Published online 28 July 2010 proteinscience.org

Abstract: Post-translational modifications of histone tails are among the most prominent epigenetic marks and play a critical role in transcriptional control at the level of chromatin. The Polycomblike (Pcl) protein is part of a histone methyltransferase complex (Pcl-PRC2) responsible for high levels of histone H3 K27 trimethylation. Studies in *Drosophila* larvae suggest that Pcl is required for anchoring Pcl-PRC2 at target genes, but how this is achieved is unknown. Pcl comprises a Tudor domain and two PHD fingers. These domains are known to recognize methylated lysine or arginine residues and could contribute to targeting of Pcl-PRC2. Here, we report an NMR structure of the Tudor domain from *Drosophila* Pcl (Pcl-Tudor) and binding studies with putative ligands. Pcl-Tudor contains an atypical, incomplete aromatic cage that does not interact with known Tudor domain ligands, such as methylated lysines or arginines. Interestingly, human Pcl orthologs exhibit a complete aromatic cage, suggesting that they may recognize methylated lysines. Structural comparison with other Tudor domains suggests that Pcl-Tudor may engage in intra- or intermolecular interactions through an exposed hydrophobic surface patch.

Keywords: NMR; Polycomblike; Pcl; PRC2; Tudor; aromatic cage; methyllysine; sDMA; post-translational modification; transcriptional regulation

Introduction

Epigenetic regulation of gene expression has emerged as one of the key determinants of cell fate and of maintenance of cell identity. Linked to this, many transcriptional regulators have been found to act at the level of chromatin. Among those chromatin modifiers, the Polycomb/Trithorax system is a highly conserved machinery that is essential for controlling expression of developmental regulator genes in both animals and plants. It is implicated in transcriptional control of genes during development,[1] in stem cells,[2] in X-inactivation,[3] and tumor biology[4] of mammals, as well as in flowering time in plants.[5]

The Polycomb group (PcG) of proteins are known as developmental regulators. PcG proteins are conserved from plants to humans and are considered as general factors engaged in transcriptional

Abbreviations: aDMA, asymmetrically dimethylated arginine; HSQC, heteronuclear single quantum correlation; NMR, nuclear magnetic resonance; NOE, nuclear overhauser effect; PcG, Polycomb group; Pcl, *Drosophila* Polycomblike; Pcl-Tudor, the Tudor domain of Pcl; PRC2, Polycomb repressive complex 2; sDMA, symmetrically dimethylated arginine.

Additional Supporting Information may be found in the online version of this article.

Anders Friberg and Anna Oddone contributed equally to this work.

Grant sponsor: European Commission (3D Repertoire); Grant number: LSHG-CT-2005-512028; Grant sponsor: Deutsche Forschungsgemeinschaft.

**Correspondence to:* Michael Sattler, Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany.
E-mail: sattler@helmholtz-muenchen.de

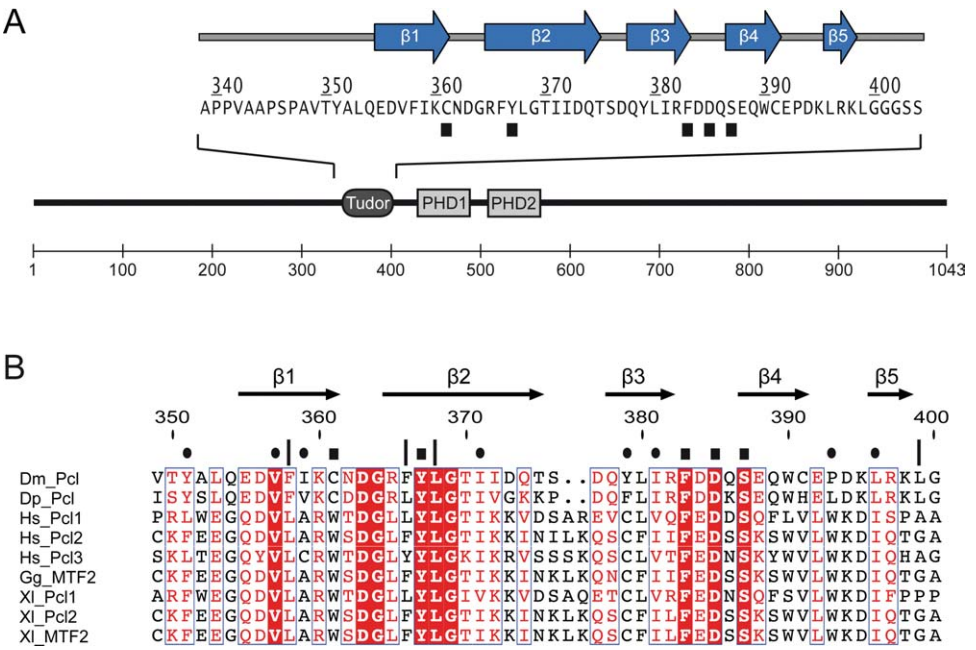

Figure 1. Domain architecture of Polycomblike (Pcl) and sequence comparison of Pcl-Tudor homologs. A: Domain architecture of full-length Pcl from *D. melanogaster* and secondary structure topology of Pcl-Tudor: β-sheets are colored blue. Filled black boxes indicate aromatic cage residues that form the ligand binding site in canonical Tudor domains. B: Multiple sequence alignment of Pcl Tudor domains from different organisms. Arrows indicate the β-strands in the Tudor domain. The numbering on top corresponds to *Drosophila* Pcl. Important residues in Pcl-Tudor are highlighted. Key: ■ putative binding site residues; ● hydrophobic core residues; | residues in an additional hydrophobic surface patch.

repression.[6] Polycomb proteins exist in four distinct multiprotein complexes: Pleiohomeotic repressive complex (PHO-RC), the Polycomb repressive complexes 1 and 2 (PRC1 and PRC2)[7] and the recently described Polycomb repressive deubiquitinase (PR-DUB).[8] Among those, PRC2 is a histone methyltransferase for lysine 27 of histone 3 (H3K27). Studies in *Drosophila* suggest that mono- and dimethylation of H3K27 (H3K27me1/me2) is widespread,[9] whereas trimethylation (H3K27me3) is confined to genes regulated by the PcG machinery. Biochemical purifications isolated two different forms of the complex: PRC2 and Polycomblike (Pcl)-PRC2.[10–12] In *Drosophila*, genome-wide H3K27me1/me2 is generated by PRC2 whereas Pcl-PRC2 is responsible for the high levels of H3K27me3 at target genes to allow PcG repression.[10]

At present, the specific role of Pcl in H3K27 trimethylation by PRC2-Pcl is not clear. *In vitro*, *Drosophila* PRC2 and Pcl-PRC2 have largely similar enzymatic activities for generating H3K27me1, me2, and me3.[10] However, *in vitro* histone methylation by reconstituted human PRC2 is enhanced when supplemented by Pcl1.[12] Studies in *Drosophila* larvae suggest that Pcl is required for anchoring PRC2 at PcG target genes.[13] The human homologs of Pcl are known as PHF1 (Pcl1), MTF2 (Pcl2), and PHF19 (Pcl3).

Full-length *Drosophila* Pcl comprises 1043 amino acids [115 kDa; Fig. 1(A)] and is expressed in all cell nuclei during embryonic development, as well as in larval salivary glands where it co-localizes with other PcG proteins on polytene chromosomes.[14] Pcl contains two plant homeodomains (PHDs), which mediate binding to *E*(*z*),[15] and a Tudor domain [Pcl-Tudor; Fig. 1(A,B)]. PHD fingers and Tudor domains from several proteins are known to recognize methylated amino acids in histones or other proteins. Specifically, Tudor domains have previously been shown to bind methylated lysines in histone tails,[16,17] and methylated arginines in Sm proteins.[18–21] Recently, Tudor domains have been implicated in the selection of piRNAs by interacting with symmetrically dimethylated arginines (sDMAs) in Piwi proteins.[22] The binding of methylated ligands involves a so-called "aromatic cage." The aromatic cage of Tudor domains comprises five residues, that is, usually three to four aromatic side chains supplemented by small polar or charged residues.[21] Together these residues create a hydrophobic binding pocket with affinity for a methylated ligand.

A possible function of Pcl could be to target the Pcl-PRC2 complex via its Tudor or PHD domains by binding methylated residues in the histone tails. To address the role of the Tudor domain of *Drosophila* Pcl, we determined its three-dimensional structure by NMR spectroscopy and studied its ligand binding properties. Testing several typical Tudor domain ligands no high-affinity interaction was found. This result is rationalized based on the domain structure,

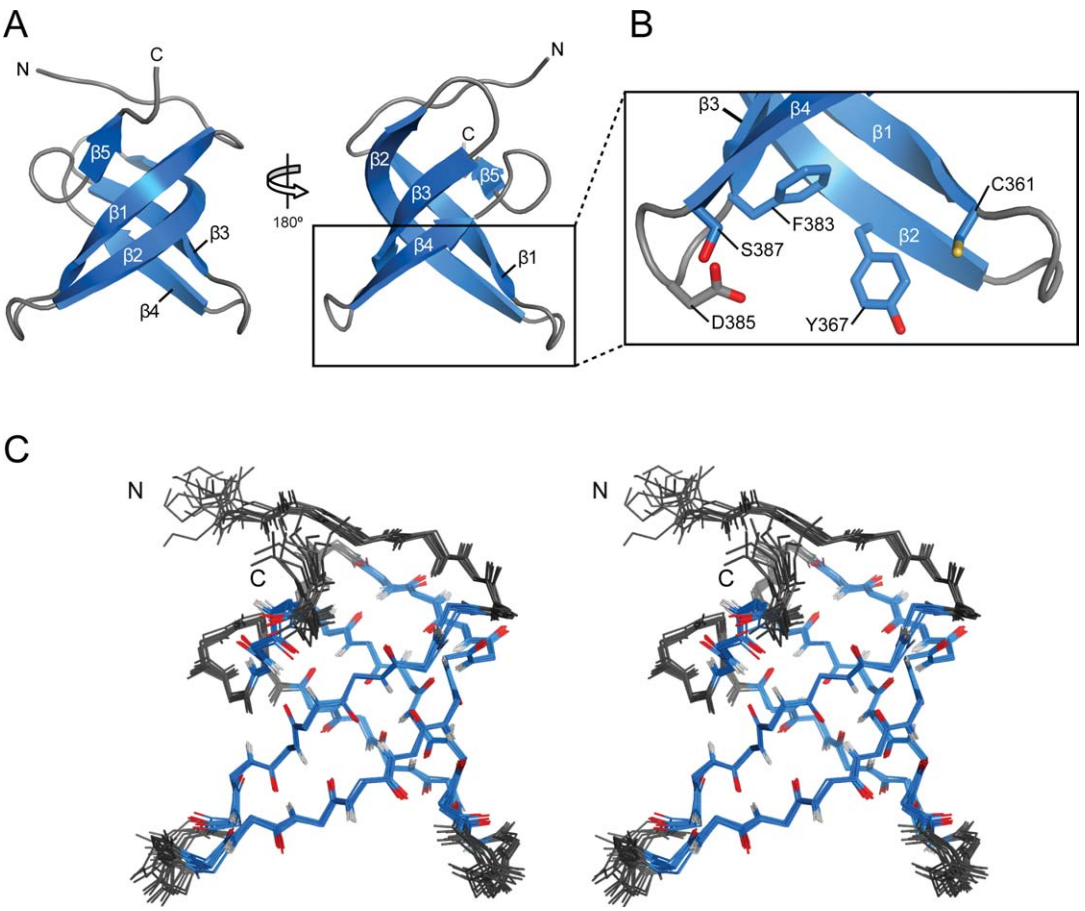

Figure 2. Solution structure of Pcl-Tudor. A: NMR structure of Pcl-Tudor. β-sheets (blue) are numbered according to Figure 1(A). B: Detailed view of the putative binding site. Residues corresponding to the "aromatic cage" are shown in stick representation. C: Stereo view of the 10 lowest energy structures from the NMR calculation, displayed as a wire model of the protein backbone. An interactive view is available in the electronic version of the article.

which reveals that *Drosophila* Pcl-Tudor contains an atypical, incomplete aromatic cage. Differences in the aromatic cages of *Drosophila* and human Pcl Tudor domains suggest divergent molecular functions. A hydrophobic surface patch on Pcl-Tudor suggests that it may engage in additional intra- or intermolecular interactions.

Results and Discussion

Solution structure of Pcl-Tudor from D. melanogaster

Recombinant Pcl-Tudor (residues 339–404) was expressed in *Escherichia coli* at high yields. This construct resulted in a well-dispersed 2D $^{1}H,^{15}N$ HSQC spectrum, indicating the protein was amenable for structural studies by NMR. A stable protein sample for further analysis required the use of a strong reducing agent to keep cysteine residues in a reduced state. The three-dimensional structure of Pcl-Tudor [Fig. 2(A–C)] was determined by NMR, using standard experiments for assignments and derivation of distance restraints.[23] Of the 69 residues in the expression construct, 52 residues (349–400) define the tertiary fold with high precision [RMSD < 1 Å; Figs. 2(C) and 3(D)] and good structural statistics (Table I).

Pcl-Tudor comprises five anti-parallel β-sheets, which together form a characteristic β-barrel [Fig. 2(A)]. The β-barrel is closed by an interaction of β5 with β1 and is stabilized by a hydrophobic core including Y351, V357, I359, I371, Y379, I381, P393, and L396 (Supporting Information Fig. 1A). Similar to other Tudor domains, the second β-strand is slightly bent around I372, thereby making a hydrogen bond possible between both of the backbone amides of I372 and D373 to the backbone oxygen of L380. Secondary chemical shift values confirm the secondary structure seen in the structure [Fig. 3(A)].

The side chains forming the putative binding pocket—the "aromatic cage"—are found in or close to the β1-β2 and β3-β4 loops [Fig. 2(B)]. The residues in Pcl-Tudor corresponding to the aromatic cage are: C361, Y367, F383, D385, and S387. Notably, the expected binding pocket of Pcl-Tudor is wider than other Tudor domains known to bind methylated ligands (Supporting Information Fig. 2A). This

Table I. *Structural Statistics*

Structure calculation restraints	
Distance restraints[a]	1579
Intraresidue	304
Inter-residue	
Short range ($\|i-j\|=1$)	333
Medium range ($1<\|i-j\|<5$)	161
Long range ($\|i-j\|>5$)	781
Dihedral restraints (PHI + PSI)	88
Quality analysis	
Coordinate precision (Å)[b,c]	
N, Cα, C′	0.29 ± 0.07
Heavy atoms	0.74 ± 0.04
Restraint RMSD[d]	
Distance restraints (Å)	0.011 ± 0.002
Dihedral restraints (°)	0.528 ± 0.043
Deviation from idealized geometry[e]	
Bond lengths (Å)	0.016
Bond angles (°)	1.3
Ramachandran values (%)[c,f]	
Preferred regions	91.9
Allowed regions	8.1
Generously allowed regions	0
Disallowed regions	0
WhatIf analysis[c,g]	
First generation packing	1.036 ± 0.145
Second generation packing	1.671 ± 0.367
Ramachandran plot appearance	0.485 ± 0.437
Chi-1/Chi-2 rotamer normality	−1.936 ± 0.442
Backbone conformation	0.051 ± 0.501

[a] A total of 3243 resonances out of 3716 were assigned by CYANA.
[b] Given as the Cartesian RMSD of the ten lowest in energy models to their mean structure.
[c] For residues 349–400.
[d] Analyzed by iCING. No restraints were violated by more than 0.2 Å or 3° in any of the models.
[e] PDB validation and deposition server (ADIT).
[f] With Procheck.
[g] Structure *Z*-scores, a positive number is better than average.

might be a result of the absence of large aromatic residues in Pcl-Tudor, which are present in other Tudor domains. Particularly, the cysteine residue in position 361 is atypical for this type of domain. In addition, Pcl-Tudor exhibits a hydrophobic patch on the surface of the β-barrel, consisting of F358, F366, L368, and L399 [Fig. 6(B)].

Protein backbone dynamics

NMR ^{15}N relaxation data (T_1, T_2 and heteronuclear {^{1}H}-^{15}N NOE) agree well with the calculated ensemble of structural models [Fig. 3(B,C)]. Our data show that the boundaries between less defined and well-defined protein backbone, at the N- and C-termini, correlate with the presence and absence, respectively, of fast motions [Fig. 3(B–D)]. No increased flexibility on fast (sub nanosecond) timescales is observed for the loops flanking the putative binding site. The reduced structural precision of these two loops probably results from paucity of distance restraints for these residues, but could also reflect slow motions in the millisecond timescale. The average ^{15}N T_1 and T_2 values were determined to be 621 ± 26 ms and 90 ± 6.0 ms, respectively. The ratio of T_1/T_2 (6.9) corresponds to a correlation time of $\tau_c^{exp} \approx 7.7$ ns. An estimation of the correlation time by HYDRONMR,[25] using our calculated structural model, gives approximately $\tau_c^{calc} \approx 5.2$ ns. Taken together, this indicates that Pcl-Tudor is a monomer in solution, consistent with size exclusion chromatography data (Supporting Information Fig. 1B).

Ligand binding studies

The role of Pcl in trimethylation of H3K27 or in recruitment of PRC2 to target genes is still unknown. To obtain insight into the molecular interactions of Pcl, we performed binding studies using our Pcl-Tudor construct with putative ligands. In a series of NMR titrations and isothermal titration calorimetry (ITC) experiments, we primarily tested known Tudor ligands and their derivatives, that is, molecules containing arginines and lysines in different methylation states. In addition, we included compounds that were suggested to bind to Tudor domains, for example, acetyl-lysine (a different epigenetic modification), Xist RNA (a proposed targeting factor for PRC2[3]), and methylated guanosines (another molecule found in different methylation states). Table II lists all the ligands tested.

Despite our efforts, no ligand showing strong affinity for Pcl-Tudor could be identified. Figure 4 shows typical results of NMR titration experiments, here using a mixture of methylated lysines and methylated arginines (a), as well as unmodified histone tails (b). No chemical shift changes could be observed, hence indicating that Pcl-Tudor does not bind any of the Tudor ligands characterized to date or any of the additional ligands tested here. Similarly, ITC measurements failed to provide evidence that Pcl-Tudor interacts with various histone tail peptides containing methylated lysine residues (Maxim Nekrasov and Jürg Müller, unpublished data).

We found that the three-dimensional structure of the Tudor domain of *Drosophila* Pcl (Pcl-Tudor) comprises the typical β-barrel fold of Tudor domains, but lacks an intact aromatic cage. This particular structural feature of Pcl-Tudor provides an explanation for the lack of binding to known ligands of Tudor domains: without a complete aromatic cage, the methylated residue cannot be properly coordinated. Methylated amino acids can exist in different methylation states, which demands a specific and well-tuned recognition. Slight differences in the aromatic cage of Tudor domains provide selectivity for certain ligands.[21] To date, only structures of methylated lysines in complex with Tudor domains have been reported. A structural comparison of Pcl-Tudor with available structures of Tudor domain-ligand complexes as well as to its human homologs provides

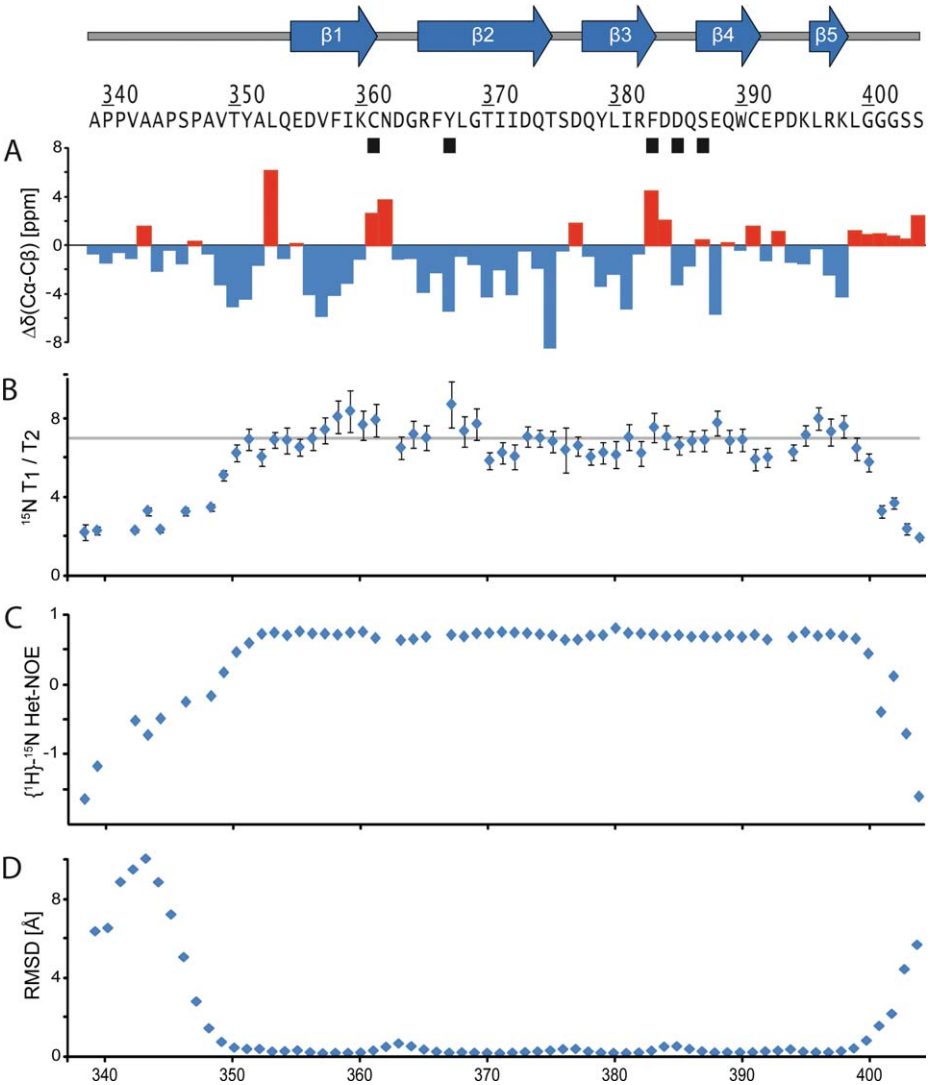

Figure 3. Secondary chemical shifts and ^{15}N relaxation data. A: Secondary chemical shifts, $\Delta\delta(^{13}C\alpha\text{-}^{13}C\beta)$. Positive (red) and negative (blue) values indicate α-helical and β-strand conformation, respectively. B: ^{15}N NMR relaxation data. The average ratio of T_1/T_2 (6.9) is indicated by a gray line. The error bars are derived by Monte Carlo simulations in NMRviewJ[24] (v. 8.0). C: {^{1}H-}-^{15}N heteronuclear NOE indicates flexible N- and C-termini of Pcl-Tudor. D: Backbone RMSD. The RMSD of the backbone atoms (N, Cα and C′) in the calculated ensemble of 10 lowest energy structures.

a rationale for the distinct binding properties and highlights the diversity of Tudor domains, as described in the following.

Structural comparison with methyllysine recognition by 53BP1 Tudor

The first Tudor domain of 53BP1 has been proposed to be required for targeting the protein to DNA double-strand breaks by recognition of H3K20me2.[16] The interaction with dimethyllysine (K_D = 20 μ*M*) is mediated by an intact aromatic cage, comprising W1495, Y1502, F1519, D1521, and Y1523. The aspartate is thought to be critical for high affinity and selectivity, due to its capability of forming a hydrogen bond to the side-chain amino group of the ligand. Pcl-Tudor contains an aspartate in the equivalent position. However, Pcl-Tudor only has two aromatic side chains in the binding pocket, Y367 and F383. W1495 of 53BP1, shown by mutational analysis to be essential for ligand binding, is replaced by a cysteine in Pcl-Tudor [Fig. 5(A)]. Also Y1523 of 53BP1 is substituted in Pcl-Tudor by a small nonaromatic residue, namely S387. The overall sequence identity between Pcl-Tudor and 53BP1 is 23% [Fig. 5(F)].

Structural comparison of Drosophila *and human Pcl Tudor domains*

Structures of the Tudor domains of Pcl1, Pcl2, and Pcl3, the human homologs of *Drosophila* Pcl, have been deposited in the PDB (accession codes: 2E5P, 2EQJ, and 2E5Q). The sequence identity between the *Drosophila* Pcl-Tudor and its human homologs is 28%, 32%, and 24% for Pcl1, Pcl2, and Pcl3, respectively [Fig. 1(B)]. The structures of human and *Drosophila* Pcl Tudor domains are similar and all contain a rather wide putative binding pocket

Table II. *NMR Binding Studies of Pcl-Tudor with Putative Ligands*

Ligands tested by NMR	Final protein to ligand ratio
sDMA, aDMA, Kme1, Kme2, Kme3 (mixture)	1:5
R (unmodified)	1:20
AGRGRG	1:5
AGR*GR*G (R* = sDMA)	1:5
H3 (2–29)	1:5
H3K27me3	1:5
H3R2 sDMA	1:5
H3R17 sDMA	1:5
H3R26 sDMA	1:5
H4 (2–21)	1:5
H4R3 sDMA	1:5
Acetyl-lysine	1:5
Xist RNA 14mer	1:2
7-methylguanosine	1:5
N^2,N^2,7-trimethylguanosine	1:5

sDMA, symmetrically dimethylated arginine; aDMA, asymmetrically dimethylated arginine; Kme1, monomethylated lysine; Kme2, dimethylated lysine; Kme3, trimethylated lysine.

(Supporting Information Fig. 2D). However, an interesting difference is that the aromatic cage of all three human homologs comprises a conserved and characteristic tryptophan, which in *Drosophila* Pcl is replaced by a cysteine (C361) [Fig. 5(B,F) and Supporting Information. Fig. 2B,C]. By structural comparisons, we find that the aromatic cage residues of the homologous Pcl1-3 proteins are identical to the tandem hybrid Tudor domain of JMJD2A [Fig. 5(C)].[17] The JMJD2A Tudor domain binds trimethylated lysines in histone tails, both H3K4me3 and H4K20me3,[26] and it was shown that a mutation of the tryptophan to histidine abolishes binding of JMJD2A to H3K4me3.[17] These results are in line with our data since we do not observe any binding of H3K4me3 by ITC or of Kme3 by NMR to Pcl-Tudor, which lacks the conserved tryptophan. On the other hand, it is tempting to hypothesize that Pcl1-3 would interact with trimethylated lysines, considering the striking similarity of the putative binding sites in Pcl1-3 and JMJD2A [Fig. 5(C)].

Structural comparison to methylarginine-binding Tudor domains

The multifunctional Tudor-SN protein contains a methylarginine-binding Tudor domain also comprising an intact aromatic cage [Fig. 5(D)]. The binding site is composed of F760, Y767, Y783, Y786, and N788, providing the protein with selectivity for sDMA over aDMA (K_D = 720 μ*M* and ~5 m*M*, respectively).[21] The major difference compared to the methyllysine-binding 53BP1, is a spatial rearrangement of a small polar residue, D1521 and N788 in 53BP1 and Tudor-SN, respectively. Again, as in the case of 53BP1, compared to Tudor-SN Pcl-Tudor displays substitutions making an interaction with methylated arginines unlikely. F760, Y783, Y786, and N788 of Tudor-SN are replaced by C361, F383, D385, and S387 in Pcl-Tudor, respectively, and the overall sequence identity between Pcl-Tudor and the Tudor domain of Tudor-SN is only 13% [Fig. 5(F)].

Similarly to Tudor-SN, also the survival of motor neuron (SMN) protein interacts with methylated arginines.[18,19,27,28] Apart from one substitution, their aromatic cages are identical in composition.[21] Compared to SMN, Pcl-Tudor exhibits major differences in the aromatic cage [Fig. 5(E)]. C361, F383, D385, and S387 in Pcl-Tudor are replaced by W102, Y127, Y130, and N132 in SMN, respectively, and the overall sequence identity between these two Tudor domains is 13% [Fig. 5(F)].

A putative interaction surface in Pcl-Tudor

Taken together, Pcl-Tudor adopts the characteristic overall fold of other Tudor domains, but exhibits major differences in the putative binding site that most likely renders it incapable of binding any of

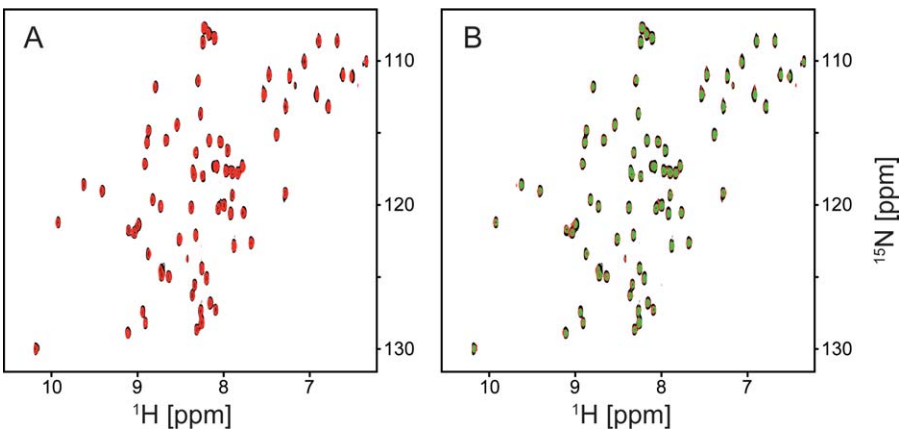

Figure 4. Ligand titrations followed by NMR. A reference ^{1}H,^{15}N HSQC spectrum (black) was measured on a ^{15}N-labeled 100 μ*M* Pcl-Tudor sample. A: Red spectrum, 1:5 protein:ligand ratio using a mixture of various modified amino acids (sDMA, aDMA, mono-, di-, and trimethylated lysine). B: Red spectrum, 1:5 protein:ligand ratio with unmodified H3 peptide. Green spectrum, 1:5 protein:ligand ratio of unmodified H4 peptide.

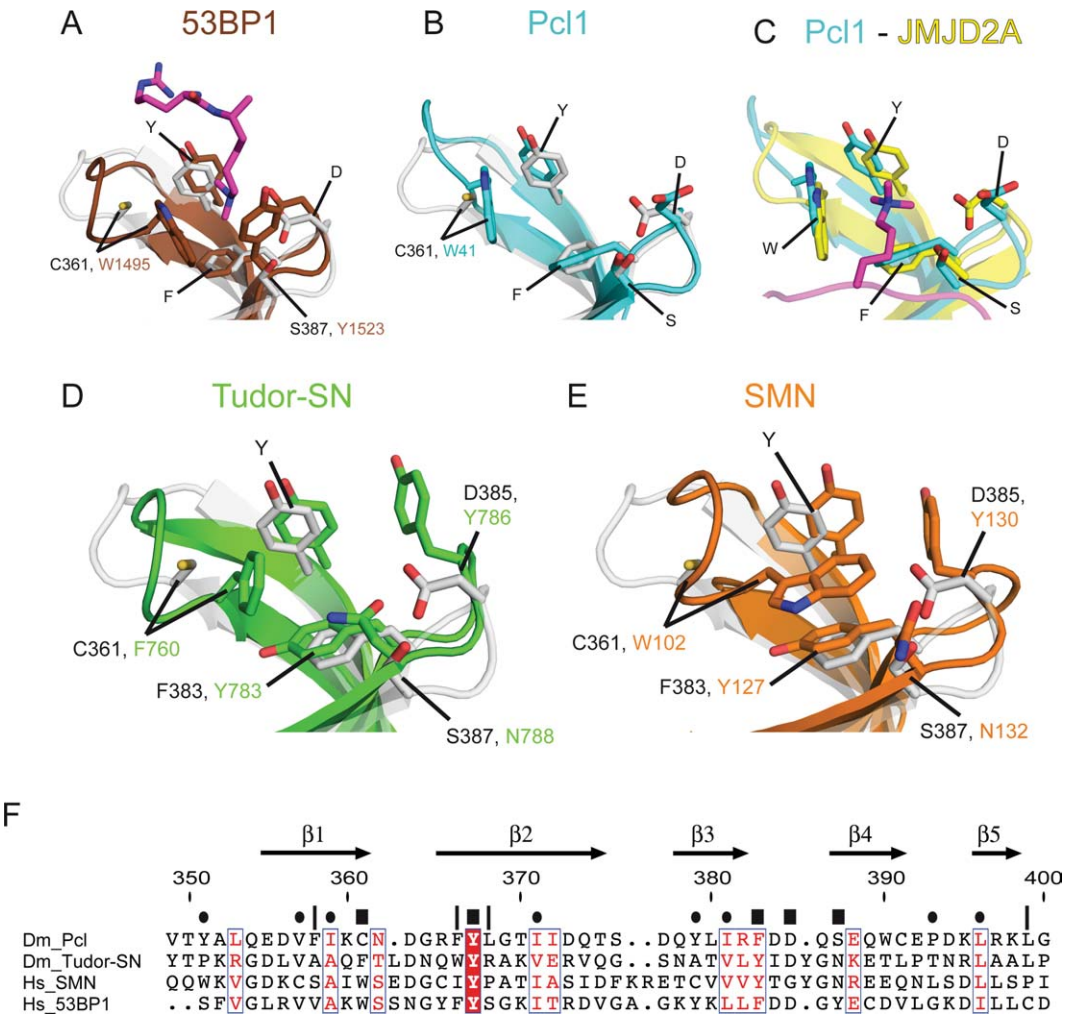

Figure 5. Comparison of Pcl-Tudor with other Tudor domains. (A, B) Side chains corresponding to the aromatic cage are highlighted as sticks in Pcl-Tudor and other Tudor domains, only substitutions are labeled with residue numbers. Superposition of Pcl-Tudor (lightgray) with (A) the first 53BP1 tandem Tudor domain in complex with H4K20me2 (brown; 2IGO.pdb; ligand in magenta) and (B) Pcl1 (cyan; 2E5P.pdb). (C) Superposition of the human homolog Pcl1 (cyan; 2E5P.pdb) and the hybrid Tudor domain of JMJD2A in complex with H3K4me3 (yellow; 2GFA.pdb). Note the identical composition of residues in the binding site. The peptide ligand is shown in magenta. (D, E) Similar to (A, B), superposition of Pcl-Tudor (gray) and, (D) Tudor-SN (green; 2WAC.pdb), (E) SMN (orange; 1MHN.pdb). (F) Multiple sequence alignment of *Drosophila* Pcl-Tudor and Tudor domains known to bind methylated lysines or methylated arginines. Symbols and numbering as in Figure 1(B). An interactive view is available in the electronic version of the article.

the established Tudor ligands. We cannot exclude the possibility that the atypical aromatic cage of Pcl-Tudor could recognize a ligand distinct from the ones included tested in this study. Also, additional domains in Pcl or in other binding partners might be needed for ligand recognition. These could, for example, associate with Pcl-Tudor to complete the aromatic cage, thus recreating the canonical binding pocket of Tudor domains, or they could recognize additional parts of the histone tails, leading to efficient substrate recognition. Methyllysine-binding modules such as PHD fingers are often found associated in proteins that play an important role in epigenetic regulation and are likely to function cooperatively. Recently, PHF8, a human histone demethylase was reported to function in such modular fashion.[29] The demethylase activity of PHF8 resides in a Jumonji domain, with H3K9me2 and H3K27me2 as substrates. It was shown that the demethylase activity is enhanced and more specific if an H3K4me3 mark is present that interacts with a neighboring PHD domain.

We note that a distinct hydrophobic patch at the surface of Pcl-Tudor, remote from the aromatic cage, could be used as an interaction site for other domains or proteins (Fig. 6; Supporting Information Fig. 3). Hydrophobic residues in this region are found in other Tudor domains, for example, in SMN, 53BP1 and in the human homologs of Pcl [Figs. 1(B) and 5(F)]. In some structures of Tudor-containing proteins, for example in Tudor-SN (2WAC.pdb)[21] and TDRD2 (3FDR.pdb),[30] the hydrophobic patch is covered by secondary structure elements extending

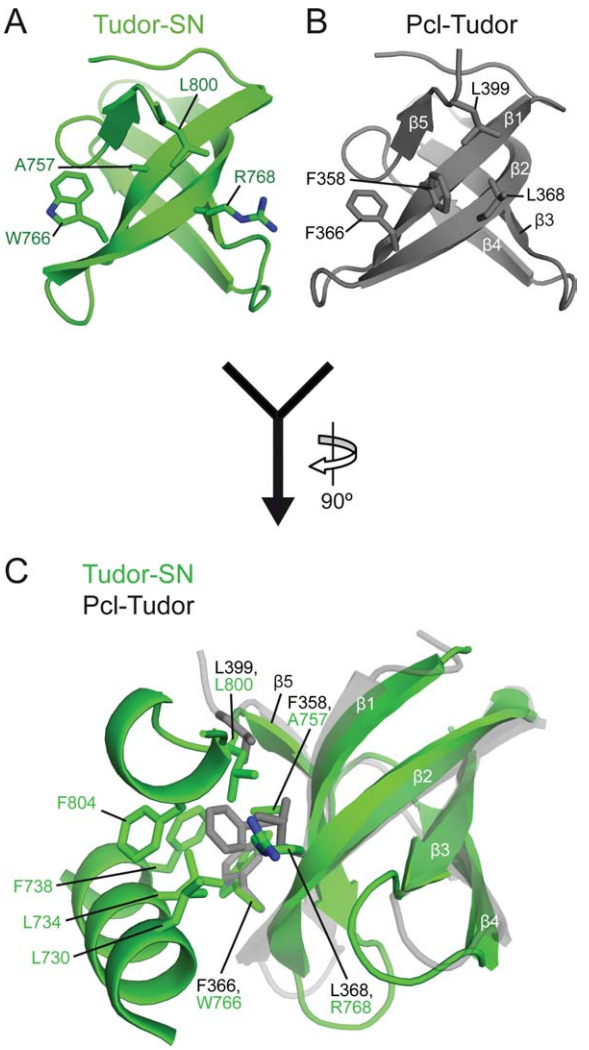

Figure 6. Potential interaction site on *Drosophila* Pcl-Tudor. A: Hydrophobic patch residues in Tudor-SN (green). B: A corresponding hydrophobic patch can be found in Pcl-Tudor (gray). C: The Tudor domain of Tudor-SN forms a hydrophobic interaction with residues from neighboring secondary structures. An interactive view is available in the electronic version of the article.

from the Tudor domain, thereby forming an intramolecular hydrophobic interface (Fig. 6). In the case of Tudor-SN, this arrangement stabilizes the interdomain interaction to a neighboring nuclease domain, making the two domains tumble together as one unit in solution.[21] Moreover, in the structure of the human histone methyltransferase SETD1B (3DLM.pdb) such additional hydrophobic interaction surfaces are found between three consecutive Tudor domains (Supporting Information Fig. 3). Hence, it could be suggested that other domains or proteins interact with Pcl-Tudor using this structural feature. An additional factor interacting with Pcl-Tudor might increase ligand affinity and affect its binding specificity, possibly by complementing the incomplete aromatic cage of the Tudor domain.

Conclusions

In this study, we reported the structure of the Tudor domain of Pcl from *Drosophila melanogaster* and, based on its structural features, rationalized our ligand binding results. Although the overall structure represents a canonical Tudor fold, it harbors an atypical incomplete aromatic cage. Pcl-Tudor shows no affinity for any of the typical known Tudor domain ligands, that is, methylated lysines, methylated arginines, or other putative ligands tested. The Tudor domain of *Drosophila* Pcl may thus not directly participate in the recognition of post-translational modifications on histone proteins. However, it cannot be excluded that full-length Pcl contains such a function. Future studies, including the analysis of the Tudor domain in the context of larger portions of the Pcl protein should help to establish the role of Pcl in transcriptional repression and perhaps identify other, currently uncharacterized Tudor ligands.

Materials and Methods

Cloning, protein expression and purification

Residues 339–404 of Pcl from *Drosophila melanogaster* (Uniprot: *Q24459*) were cloned into a modified pET-24d vector using standard protocols. The fusion protein comprises a green fluorescence protein (GFP) tag to facilitate purification. This protein construct was expressed in *Escherichia coli* BL21 (DE3) pLysS (Novagen) using kanamycin for selection. A 10 mL lysogeny broth (LB) preculture was inoculated with a single colony from a transformation plate. The preculture was used to start larger 1 L cultures, containing LB or M9 minimal medium for labeling with ^{15}N or $^{15}N/^{13}C$. Upon reaching optical density (OD) of 0.6 cultures were put at 20°C and, after 30 min of cooling, induced over night with 0.2 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG).

Recombinant protein was purified by sonicating the harvested cell pellet in 25 mL lysis buffer (20 m*M* TRIS pH 7.5, 300 m*M* NaCl, 10 m*M* imidazole, 1 m*M* DTT, and 0.02% NaN_3), also including protease inhibitors, RNase, lysozyme, and 0.2% IGEPAL. After high-speed centrifugation (20,000 rpm, 30 min) and filtering, the supernatant was applied three times to Ni-NTA Agarose resin (Qiagen). Several rounds of washing were performed with: lysis buffer including 0.2% IGEPAL, lysis buffer, lysis buffer with high salt concentration (1*M* NaCl), and lysis buffer with high imidazole concentration (30 m*M* imidazole). Finally, the protein was eluted by applying 10 mL of a buffer containing 20 m*M* TRIS pH 7.5, 300 m*M* NaCl, 330 m*M* imidazole, 1 m*M* DTT, and 0.02% NaN_3. Tobacco etch virus (TEV) protease was added to the sample and incubated overnight at 4°C. To remove the cleaved GFP tag, the sample was passed three times over a second Ni-NTA column. A last purification step included size exclusion chromatography

(HiLoad, Superdex 75 16/60, GE Healthcare). In this step, the protein was buffer-exchanged into the NMR buffer (20 m*M* sodium phosphate pH 6.3, 25 m*M* NaCl, and 2 m*M* fresh DTT or TCEP, 0.02% NaN_3).

NMR spectroscopy and structure determination

Protein backbone and amino acid side-chain assignments were done by using $^{15}N/^{13}C$ labeled samples at a concentration of 0.25 m*M*. For this purpose, several multidimensional heteronuclear experiments were acquired: $^{1}H,^{15}N$ HSQC, $^{1}H,^{13}C$ HSQC, HNCA, HNCACB, CBCA(CO)NH, (H)CC(CO)NH-TOCSY, H(CC)(CO)NH-TOCSY, and H(C)CH-TOCSY.[23] To obtain distance restraints for structure calculation, a series of NOE-based experiments were recorded: 2D ^{1}H homonuclear NOESY, $^{1}H,^{15}N$ HSQC-NOESY, $^{1}H,^{13}C$ HMQC-NOESY (aliphatic and aromatic versions). Assignment of the aromatic ring systems was enabled by another two experiments, (HB)CB(CG,CD)HD and (HB)CB(CG,CD,CE)HE.[31] All experiments were performed at 298.5 K on Bruker 900, 750, 600, and 500 MHz spectrometers equipped with pulsed field gradients. The data were processed with NMRPipe[32] and analyzed in NMRView.[24]

Automatic assignment of the NOESY spectra and derivation of distance restraints were accomplished using CYANA v2.1.[33] TALOS+ was used to predict dihedral angle restraints.[34] Hundred structures were calculated and further water-refined[35] by use of RECOORD scripts[36] with CNS.[37] The 10 lowest energy structures were selected as a representative ensemble. Structure validation was performed using Molprobity[38] and iCing[39] (including PROCHECK[40] and WHATCHECK[41]). Secondary structure definitions were based on DSSP algorithms[42] as implemented in Procheck and Pymol, together with manual inspection.

To study the dynamical properties of the protein backbone ^{15}N T_1, T_2 and $\{^{1}H\}$-^{15}N heteronuclear NOE were measured on a 1 m*M* ^{15}N-labeled sample (600 MHz proton Larmor frequency) as described previously.[43] T_1 was measured in an interleaved fashion with 14 different relaxation delays, varying from 21 ms to 2160 ms. Similarly, T_2 was determined by using eight time points ranging from 11 ms to 190 ms. Duplicate time points were used for error estimation. The data were analyzed using the relaxation module integrated in NMRViewJ (v. 8.0). Average values were based on residues selected on specific criteria[44] and errors estimated as one standard deviation of included values. The correlation time (τ_c) of the protein molecule is estimated using the ratio of averaged T_1 and T_2 values.[45] HYDRONMR was used with standard parameter settings at 298.5 K.[25]

Images for structure comparisons were generated with Pymol.[46] The structure of Pcl2 (2EQJ.pdb), one of the human homologs, was deposited as originating from mouse. However, the amino acid sequence is identical to human and used accordingly. Sequence alignments were performed with ProbCons[47] (http://probcons.stanford.edu) and SSM[48] (www.ebi.ac.uk/msd-srv/ssm) using standard settings. Only the sequence of Pcl-Tudor that was well-defined in the NMR ensemble (residues 349–400) was included. The alignments were compared with structural data if available.

Binding studies

Binding of putative ligands was investigated by NMR titrations. All NMR experiments were performed at 298.5 K with the protein in: 20 m*M* sodium phosphate pH 6.3, 25 m*M* NaCl and 2 m*M* fresh DTT or TCEP and 0.02% NaN_3. For NMR titrations samples of 100–200 μ*M* of ^{15}N labeled protein were prepared. Ligands were then added in increasing amounts and binding followed by consecutive acquisition of 2D $^{1}H,^{15}N$ HSQC spectra. All ligands investigated are listed in Table II. All compounds were purchased except the 14mer Xist RNA, which was produced by *in vitro* transcription.[49]

Accession codes

Atom coordinates and restraint files have been deposited at the Protein Data Bank (accession code: *2XK0*). NMR chemical shifts have been deposited at Biological Magnetic Resonance Data Bank (accession code: *17050*).

Acknowledgments

The authors thank Gunter Stier (EMBL Heidelberg) for help with designing and preparing the Tudor-Pcl expression plasmids; Maxim Nekrasov and Vladimir Rybin (EMBL Heidelberg) for ITC measurements. They also thank Kostas Tripsianes, Alex Beribisky, and Iren Wang for discussions and reading of the manuscript. A.F. is supported by a PhD fellowship from Helmholtz Zentrum München (HMGU), and by the International PhD program in Protein Dynamics from Elitenetzwerk Bayern. They acknowledge NMR measurement time at the Bavarian NMR Centre, Garching, Germany.

References

1. Breiling A, Sessa L, Orlando V (2007) Biology of polycomb and trithorax group proteins. Int Rev Cytol 258:83–136.
2. Pietersen AM, van Lohuizen M (2008) Stem cell regulation by polycomb repressors: postponing commitment. Curr Opin Cell Biol 20:201–207.
3. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322:750–756.
4. Bracken AP, Helin K (2009) Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat Rev Cancer 9:773–784.
5. Henderson IR, Dean C (2004) Control of Arabidopsis flowering: the chill before the bloom. Development 131: 3829–3838.

6. Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. Nat Genet 38:700–705.
7. Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10:697–708.
8. Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Muller J (2010) Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature 465:243–247.
9. Ebert A, Schotta G, Lein S, Kubicek S, Krauss V, Jenuwein T, Reuter G (2004) Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. Genes Dev 18:2973–2983.
10. Nekrasov M, Klymenko T, Fraterman S, Papp B, Oktaba K, Kocher T, Cohen A, Stunnenberg HG, Wilm M, Muller J (2007) Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J 26:4078–4088.
11. Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D (2008) Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol Cell Biol 28: 2718–2731.
12. Cao R, Wang H, He J, Erdjument-Bromage H, Tempst P, Zhang Y (2008) Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol Cell Biol 28: 1862–1872.
13. Savla U, Benes J, Zhang J, Jones RS (2008) Recruitment of *Drosophila* Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. Development 135:813–817.
14. Lonie A, D'Andrea R, Paro R, Saint R (1994) Molecular characterisation of the Polycomblike gene of *Drosophila melanogaster*, a trans-acting negative regulator of homeotic gene expression. Development 120:2629–2636.
15. O'Connell S, Wang L, Robert S, Jones CA, Saint R, Jones RS (2001) Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. J Biol Chem 276:43065–43073.
16. Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell 127:1361–1373.
17. 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.
18. Sprangers R, Groves MR, Sinning I, Sattler M (2003) High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. J Mol Biol 327:507–520.
19. Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M (2001) SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol 8: 27–31.
20. Shaw N, Zhao M, Cheng C, Xu H, Saarikettu J, Li Y, Da Y, Yao Z, Silvennoinen O, Yang J, Liu ZJ, Wang BC, Rao Z (2007) The multifunctional human p100 protein 'hooks' methylated ligands. Nat Struct Mol Biol 14:779–784.
21. Friberg A, Corsini L, Mourao A, Sattler M (2009) Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN. J Mol Biol 387: 921–934.
22. Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS (2009) Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol 16:639–646.
23. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed. Prog Nucl Magn Reson Spectrosc 34:93–158.
24. Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614.
25. Garcia de la Torre J, Huertas ML, Carrasco B (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J Magn Reson 147:138–146.
26. Lee J, Thompson JR, Botuyan MV, Mer G (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. Nat Struct Mol Biol 15:109–111.
27. Brahms H, Raymackers J, Union A, de Keyser F, Meheus L, Luhrmann R (2000) The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. J Biol Chem 275:17122–17129.
28. Friesen WJ, Massenet S, Paushkin S, Wyce A, Dreyfuss G (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol Cell 7:1111–1117.
29. Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol 17:38–43.
30. Chen C, Jin J, James DA, Adams-Cioaba MA, Park JG, Guo Y, Tenaglia E, Xu C, Gish G, Min J, Pawson T (2009) Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. Proc Natl Acad Sci USA 106:20336–20341.
31. Yamazaki T, Forman-Kay JD, Kay LE (1993) Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. J Am Chem Soc 115:11054–11055.
32. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293.
33. Guntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378.
34. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223.
35. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506.
36. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59:662–672.
37. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905–921.
38. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity:

all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–W383.
39. Geerten W, Vuister JFD, Alan WS da Silva. iCing v2.0, Available at: http://nmr.cmbi.ru.nl/icing/iCing.html.
40. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486.
41. Vriend G, Sander C (1993) Quality control of protein models: directional atomic contact analysis. J Appl Crystallogr 26:47–60.
42. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.
43. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 33:5984–6003.
44. Tjandra N, Kuboniwa H, Ren H, Bax A (1995) Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements. Eur J Biochem 230:1014–1024.
45. Daragan VA, Mayo KH (1997) Motional model analyses of protein and peptide dynamics using and NMR relaxation. Prog Nucl Magn Reson Spectrosc 31:63–105.
46. DeLano WL. (2002). The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific.
47. Do CB, Mahabhashyam MS, Brudno M, Batzoglou S (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res 15:330–340.
48. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.
49. Duszczyk MM, Zanier K, Sattler M (2008) A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. Nucleic Acids Res 36:7068–7077.