Structure And Function Of Histone Code Reading Proteins

Melanie A. Adams[1] and Jinrong Min[1, 2]

1. Structural Genomics Consortium, University of Toronto, 100 College St., Toronto, Ontario, M5G 1L6, Canada.

2. Department of Physiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Corresponding Author:

Jinrong Min

Structural Genomics Consortium

University of Toronto

100 College St., Toronto, On

M5G 1L6, Canada

Tel: 416-9463868

Fax: 416-946-0588

Email: jr.min@utoronto.ca

Abstract

Chromatin dynamics is regulated by chromatin remodeling factors, histone exchange, linker histone association and histone modification. Covalent modification of histones is an important factor in the regulation of the associated processes. The implementation and removal of various modifications has been implicated in replication, repair, recombination, transcription and RNA processing. In recent years, histone lysine methylation has emerged as one of the key modifications regulating chromatin function. However, the mechanisms involved are complex and not well understood. Recognition of histone methylation has been functionally ascribed to members of several protein families including the "royal family," which includes the Tudor, plant Agenet, chromo, PWWP and MBT domains, bromo domain and the WD40 repeat protein WDR5. Though the most detailed studies of histone recognition have centered on the bromo and chromo domains, a large volume of structural and biochemical information has been recently amassed for the Tudor, PHD and MBT protein families. This review summarizes current knowledge of the structures and modes of recognition employed by the PHD, Tudor and MBT domains in their interactions with target histone peptides.

Key Words: Chromatin remodeling, histone methylation, Tudor domain, MBT domain, PHD Domain

Introduction

Chromatin dynamics is regulated by chromatin remodeling factors, histone exchange, linker histone association and histone modification. Covalent modification of histones is an important factor in the regulation of the associated processes (Ehrenhofer-Murray 2004) . The implementation and removal of various modifications has been implicated in replication, repair, recombination, transcription and RNA processing (Bhaumik et al. 2007;Kouzarides 2007). Most sites of covalent modification are concentrated in the relatively unstructured N-terminal tails of histones H3, H4, H2A and H2B, with only a few distributed within the histone core (Luger et al. 1997).

One of the earliest reports linking the acetylation and methylation of histones to changes in ribonucleic acid (RNA) synthesis dates to the middle of the twentieth century (Allfrey et al. 1964), in which the authors postulated "that histone effects on nuclear RNA metabolism may involve more than a simple inhibition of RNA synthesis, and that more subtle mechanisms may exist which permit both inhibition and reactivation of RNA production at different loci along the chromosome." Since this ground-breaking work, numerous studies continue to confirm the variable and complex roles of histone modifications in normal cellular function and a vast array of human disease. The library of histone modifications has also been expanded to include phosphorylation, sumoylation, and ubiquitination along with the combinatorial expansion of the individual modifications (Bhaumik et al. 2007;Kouzarides 2007).

Histone lysine methylation has emerged as one of the key modifications regulating chromatin function. However, the mechanisms involved are complex and not well

understood. Histones may be methylated on either arginine or lysine residues through the activities of numerous enzymes. Lysine methylation occurs in three distinct forms: monomethylated (Schneider et al. 2005), dimethylated (Santos-Rosa et al. 2002), or trimethylated lysine (Wood et al. 2007). Each state may be associated with either transcriptional activation or repression with the responses likely dictated not only by the methylation state of a single lysine or arginine residue, but by the larger local or global states of histone modifications (Grewal et al. 2003;Santos-Rosa et al. 2002;Shilatifard 2006;Vakoc et al. 2005;Xiao et al. 2003). To date, seven different histone lysine residues have been identified to be functionally relevant sites of methylation (K4, K9, K27, K36 and K79 of histone H3, K20 of histone H4 and K26 of histone H1b), and each of these lysine residues can be mono-, di-, or trimethylated, often with functional consequences. Thus, histone methylation represents a major, complex system in which the interplay and global dynamics of readers and writers of the histone code may be subtly altered.

While much attention has been paid to the enzymes responsible for histone modification, including methylation, fewer examples of proteins that interpret the histone code have been extensively described. Recognition of histone methylation has been functionally ascribed to members of several protein families including the “royal family (Maurer-Stroh et al. 2003),” bromo domain (Dhalluin et al. 1999;Jacobson et al. 2000) and the WD40 repeat protein WDR5 (Fischle et al. 2003;Flanagan et al. 2005;Jacobs et al. 2002;Min et al. 2003;Nielsen et al. 2002;Wysocka et al. 2005). The “royal family” of proteins includes the Tudor, plant Agenet, chromo, PWWP and MBT domains, many of which have been found in association with chromatin (Maurer-Stroh et al. 2003). Though the most detailed studies of histone recognition have centered on the bromo and

chromo domains, a large volume of structural and biochemical information has been recently amassed for the Tudor, PHD and MBT protein families. The aim of this paper is to not to provide an exhaustive analysis of the field of histone methylation, as numerous other reviews have eloquently provided detailed overviews of both the numerous enzymes involved in the writing of the histone code and the bromo and chromo domain-containing proteins. Rather, the intent of this review is to summarize the current knowledge as it pertains to the structures and modes of recognition employed by the PHD, Tudor and MBT domains in their interactions with methylated ligands.

PHD finger as diverse histone code readers

The PHD (**P**lant **H**omeo**D**omain) finger is a conserved Cys4-His-Cys3 zinc finger domain, which is present in a large number of chromatin regulatory factors. PHD fingers exhibit diverse functions, such as histone-code reading, phosphatidylinositol phosphate binding (Gozani et al. 2003;Jones et al. 2006), and E3 ubiquitin and SUMO ligases (Dul et al. 2007;Ivanov et al. 2007;Zeng et al. 2008). This section will focus on the role of PHD fingers as histone-binding modules (Figure 1).

The PHD finger has long been proposed to function in protein-protein interactions, though the exact biochemical function of PHD was only recently established. It was demonstrated in 2004 by Aasland's group that the PHD finger of p300 cooperates with the adjacent bromodomain in binding nucleosomes (Ragvin et al. 2004). Both the bromodomain and the PHD finger were shown to bind the nucleosome while simultaneously interacting with each other (Ragvin et al. 2004). Recently, a fleet of

papers have described the ability of PHD fingers to promote both gene activation and repression through interactions with histone lysines with varying methylation states (Eberharter et al. 2004;Fiedler et al. 2008;Jia et al. 2007;Lan et al. 2007;Li et al. 2006;Liu et al. 2007;Martin et al. 2006;Matthews et al. 2007;Ooi et al. 2007;Org et al. 2008;Palacios et al. 2008;Pena et al. 2006;Pena et al. 2008;Ramon-Maiques et al. 2007;Shi et al. 2007;Shi et al. 2006;Soliman et al. 2007;Taverna et al. 2006;Vermeulen et al. 2007;Wilson et al. 2008;Wysocka et al. 2006;Xia et al. 2003;Zhou et al. 2005).

ING family of PHD fingers

The ING protein family is involved in chromatin remodeling and transcriptional regulation. Members of this family bind to and affect the activity of histone acetyltransferase, histone deacetylase, and factor acetyltransferase protein complexes (Feng et al. 2002). The ING2 protein, a core component of a repressive mSin3a-HDAC1 histone deacetylase complex, was found to bind histone H3 trimethylated at lysine 4 (H3K4me3) through the highly conserved PHD finger. In response to DNA damage, recognition of H3K4me3 by the ING2 PHD domain stabilizes the mSin3a-HDAC1 complex at the promoters of proliferation genes, resulting in histone deacetylation and the repression of the active genes (Shi et al. 2006). Interestingly, Yng1, which is a component of the NuA3 HAT complex, also binds H3K4me3 through its PHD domain. Direct in vivo assay demonstrates that NuA3 HAT complex controls transcription through a sequential trimethyl binding, acetyltransferase activity, and gene regulation (Taverna et al. 2006). The crystal structure of the ING2-PHD domain in complex with

H3K4me3 provides a molecular mechanism of histone H3K4me3 recognition by the ING2-PHD finger (Pena et al. 2006). The H3K4me3 tail is bound in an extended conformation in a deep and extensive binding site consisting of elements that are conserved among the ING family of proteins (Figure 1A). The trimethylammonium group of Lys 4 is recognized by the aromatic side chains of Y215 and W238 residues, whereas the intermolecular hydrogen-bonding and complementary surface interactions between the H3K4me3 peptide and ING2-PHD account for the PHD finger's high specificity and affinity (Pena et al. 2006). This binding mode is conserved in other ING family members as well, such as YNG1 and ING4 (Palacios et al. 2008;Taverna et al. 2006).

BPTF PHD finger

NURF (nucleosome remodeling factor) is an ISWI-containing ATP-dependent chromatin-remodeling complex. Recently, it was shown that the PHD finger of human BPTF (bromodomain and PHD domain transcription factor), the largest subunit of the NURF complex preferentially binds H3K4me3. Depletion of H3K4me3 causes partial release of BPTF from chromatin and defective recruitment of the associated ATPase, SNF2L, to the HOXC8 promoter. The co-crystal structure of BPTF-PHD-H3K4me3 complex shows that H3K4me3 peptide interacts with BPTF PHD through anti-parallel beta-sheet formation on the surface of the PHD finger, with the long side chains of R2 and K4me3 fitting nicely into adjacent pre-formed surface pockets, and bracketing an invariant tryptophan (Figure 1B) (Li et al. 2006). The observed stapling role by non-

adjacent R2 and K4me3 provides a molecular explanation for H3K4me3 site specificity, which is also conserved in ING family of PHD histone binding.

RAG2 PHD

Recombination activating gene (RAG) 1 and RAG2 together catalyze V(D)J gene rearrangement in lymphocytes as the first step in the assembly and maturation of antigen receptors. A large amount of evidence suggests that the regulation of chromatin structure is involved in the regulation of V(D)J recombination (Kwon et al. 1998). Recently, it was shown that RAG2 contains a PHD finger near its C terminus and the RAG2-PHD recognizes histone H3 methylated at lysine 4 and influences V(D)J recombination(Liu et al. 2007;Matthews et al. 2007;Ramon-Maiques et al. 2007). In contrast to other H3K4me3-binding PHD domains, RAG2-PHD substitutes a carboxylate that interacts with arginine 2 (R2) with a Tyr, resulting in binding to H3K4me3 that is enhanced rather than inhibited by dimethylation of R2 (Ramon-Maiques et al. 2007). This is readily explained from the H3K4me3 complex structure of the RAG2-PHD and other PHD fingers (Figure 1C). RAG2-PHD differs from other H3K4me3-binding PHDs in its interactions with R2 of the H3 peptide. In the histone complex with the PHD finger of ING2, BPTF, and Yng1, the extended side chains of R2 and K4me3 of the H3 peptide occupy two adjacent parallel grooves separated by the conserved tryptophan, and the guanidinium group of R2 forms salt bridges with a carboxylate, either Asp or Glu. Both the R2 binding groove and interacting carboxylate are absent in RAG2-PHD. Thus, the side chain of R2 extends toward solvent with no interaction with the RAG2-PHD. The structural studies of the RAG2-PHD-H3K4me3 complex structure together with binding

affinity measurements suggest that RAG2-PHD prefers symmetrically dimethylated R2 and binds the R2me2s and K4me3 doubly modified H3 peptide with higher affinity. RAG2-PHD provides an excellent example of a single protein domain recognizing a doubly modified histone. This is also the first example of a single protein domain recognizing a doubly modified histone. The structure of RAG2-PHD illustrates the versatility of a histone-binding module like the PHD fingers and explains how diverse modifications in histones might be recognized.

BHC80

Histone demethylase LSD1 represses transcription by demethylating histone H3K4 (Shi et al. 2004). BHC80 is a PHD finger-containing protein that is a core component of the LSD1 complex (Lan et al. 2007). Unlike BPTF and INGs PHD fingers (Shi et al. 2006;Wysocka et al. 2006), the PHD finger of BHC80 binds unmethylated H3K4 (H3K4me0), and this interaction is specifically abolished by methylation of H3K4 (Lan et al. 2007). Chromatin immunoprecipitation showed that BHC80 and LSD1 depend on each other to associate with chromatin, and this association is dependent on the PHD finger of BHC80. The crystal structure of the PHD finger of BHC80 in complex with an unmodified H3 peptide revealed that the H3 peptide binds to the surface of the PHD finger as an anti-parallel beta-sheet, with H3R2–H3R8 forming backbone hydrogen bonds with G498–M502 of BHC80 (Figure 1D) (Lan et al. 2007). In contrast to other domains capable of recognizing methylated lysine, including other PHD domains, unmethylated lysine recognition by BHC80 is mediated via electrostatic and hydrogen

bonding interactions.

DNMT3L

Dnmt3L is required for the de novo methylation of imprinting control regions in female germ cells and for the de novo methylation of dispersed repeated sequences in male germ cells (Bourc'his et al. 2001;Bourc'his et al. 2004;Shi et al. 2006). Dnmt3L lacks the conserved catalytic domain common to DNA methyltransferases, and acts as a general stimulatory factor for de novo methylation by Dnmt3a (Chedin et al. 2002). Binding assays showed that DNMT3L specifically interacts with the N-terminus of histone H3. This interaction was inhibited by methylation at lysine 4 of histone H3 but was insensitive to modifications at other positions. DNMT3L selectively binds unmethylated H3 lysine 4 peptides, and its PHD-like domain or ADD domain mediates this binding, which is reminiscent of the BHC80 binding H3K4me0. The basis for methylation-sensitive binding of H3 to DNMT3L is steric occlusion of the interaction between aspartic acid 90 in DNMT3L and lysine 4 of histone H3 (Figure 1E) (Jia et al. 2007).

PYGO PHD

Pygo and BCL9/Legless transduce the Wnt signal by promoting the transcriptional activity of beta-catenin/Armadillo in normal and malignant cells (Belenkaya et al. 2002;Kramps et al. 2002;Parker et al. 2002;Thompson et al. 2002). Human and Drosophila Pygo PHD fingers associate with their cognate HD1 domains from

BCL9/Legless to bind specifically to the histone H3 tail methylated at lysine 4. The crystal structures of ternary complexes between PHD, HD1, and H3K4me peptides solved by Bienz's lab show that the first five amino acids of the H3K4me peptide assume an extended conformation (Figure 1F) (Fiedler et al. 2008). The R2-K4me2 forms an antiparallel β sheet with PHD β3, as observed in all other known PHD-H3K4me structures. Residues T6 and A7 meander off the PHD surface and loop back to the surface through a stabilizing hydrogen bond between T6 and T3, which is reminiscent of the RAG2-PHD-H3K4me complex structure (Ramon-Maiques et al. 2007). K4me2 and A1 occupy two adjacent pockets separated by the PYGO-PHD W366, which is conserved in all known PHD-H3K4me structures (Fiedler et al. 2008).

The complex structures of PYGO-PHD-H3K4me2 provided many interesting insights into how PYGO-PHD finger binds H3K4me2. First, the binding of BCL9 HD1 to Pygo PHD increases its affinity to H3K4me. This is especially striking in Drosophila where the H3K4me binding of the Pygo PHD finger depends entirely on its interaction with Lgs HD1. The PHD C terminus forms the floor of the histone H3A1 binding pocket, which is thin in the PHD finger structure. These floor residues directly interact with HD1, which stabilizes the A1 binding pocket and enhances the H3K4me binding. Second, the H3K4me2 binding pocket consists of a semiaromatic cage with W366, Y341, and D352 forming the walls, and V350 and A356 the base. The two K4me2 methyl groups make hydrophobic interactions with W366 and Y341, while D352 forms an H bond with the Nζ of the K4me2 side chain. This H bond could not be formed with K4me3 because the extra methyl group would preclude the approach of the D352 carboxyl group. The K4me2

cavity is different from that in BPTF, ING2/Yng1, and RAG2 that prefer K4me3, but resembles those found in the L3MBTL1 (See below*)*, Scm (See below) and 53BP1(Botuyan et al. 2006), which selectively recognize the histones with low methylation states. PYGO-PHD shows only modest preference for H3K4me2 over H3K4me3. This is owing to the flexibility of the H3K4me2 binding pocket in PYGO-PHD, which allows for H3K4me3 binding. Third, PYGO-PHD histone binding is not sensitive to methylation of arginine 2 in H3. In the BPTF and ING2/Yng1 PHD fingers, R2 is completely anchored by its guanidinium group within a groove adjacent to the K4me-binding pocket, and H3R2 methylation abolishes its binding to PHD fingers (Li et al. 2006;Pena et al. 2006). In the PYGO-PHD ternary complex structures, only the first two carbons of the R2 side chain are fixed, while the guanidinium group at the end of this side chain is pushed out into the solvent by the side chain of L358, which again resembles the H3R2 binding mode found in the RAG2-H3K4me3 structures, and RAG2 preferentially binds the doubly methylated H3R2K4 peptides (Ramon-Maiques et al. 2007).

MBT repeat domain as low methylation histone lysine binders

The MBT repeat was first identified in *Drosophila* tumor suppressor protein L3MBT (Wismar et al. 1995). In *Drosophila*, there are 3 MBT repeat containing proteins, Scm, l(3)mbt and Sfmbt, which contain 2,3, and 4 MBT repeats respectively (Bornemann et al. 1998;Trojer et al. 2007;Usui et al. 2000). All these 3 proteins belong to polycomb group and are involved in establishment and maintenance of transcriptional repression of developmental control genes such as *Hox* genes. In human genome, there are at least 10

MBT repeat proteins, which belong to Scm, l(3)mbt and Sfmbt subclasses, respectively. The MBT (malignant brain tumor) repeat shows structural similarity with the Chromodomain, Tudor domain and PWWP domain, which are collectively referred to as the Tudor domain "Royal Family" (Figure 2). The MBT repeat proteins can also bind modified histones (Kim et al. 2006). The *Drosophila* SfMBT protein in the PhoRC complex binds mono- or di-methylated H3K9 and H4K20 peptides. PhoRC interacts with methylated histones in the chromatin flanking the Polycomb response elements (PRE) and maintains a Polycomb-repressed chromatin state (Klymenko et al. 2006). L3MBTL1, the human homolog of the *Drosophila* tumor suppressor protein l(3)mbt, forms a complex with core histones, histone H1b, HP1γ, and Rb, and the purified L3MBTL1 MBT domains compact nucleosomal arrays dependent on mono- and dimethylation of histone H4K20 and histone H1bK26 (Kalakonda et al. 2008;Trojer et al. 2007). In vitro binding assays also showed that these MBT repeat proteins selectively bind low methylated lysine histones, which are confirmed by recent structural studies.

L3MBTL1

L3MBTL1 represses transcription through chromatin compaction and was recently shown to bind to mono- and dimethylated H1bK26 and H4K20 (Boccuni et al. 2003). Structural, binding and mutagenesis data show that only one of the three MBT repeats is able to bind methylated histone peptides and that a unique binding pocket within the second MBT repeat accommodates mono- or dimethylated lysine histone peptides, but not unmodified or trimethylated lysine peptides (Kim et al. 2006). The crystal structures show that the K20me is accommodated in a binding pocket formed by three aromatic

residues and a negatively charged residue D355 (Figure 2A). The hydrophobic side of the dimethylated lysine interacts with the aromatic cage via cation-π interactions, and the negatively charged residue interacts with the dimethylammonium group of K20me2 via a salt bridge and a hydrogen bond. This binding mode provides a simple explanation for the selective binding of L3MBTL1 to mono- and dimethylated histone peptides. A series of structures of 3MBT alone and in complexes with methylated histone peptides reported by our laboratory reveals an unexpected mode of interaction in which a pair of methylated histone peptides bridge together two 3MBT proteins (Min et al. 2007). Our data in combination with previous studies by Trojer *et al* suggest a possible mechanism for the involvement of L3MBTL1 in histone binding and chromatin compaction (Min et al. 2007;Trojer et al. 2007). L3MBTL1 is recruited to chromatin loci with specific histone modification patterns, in this case, with H3K9 methylation and mono- and dimethylation of H4K20 and H1bK26 in a complex with Hp1γ and Rb. The protein selectively binds to mono- and dimethylation of H4K20 and H1bK26; HP1 preferentially recognizes di and trimethylated histone H3K9. These interactions are weak but specific. Therefore, the complex, endowed with two histone-binding activities, achieves avid specificity through the combinatorial recognition of multiple lysine methylation marks in histones H1b, H3 and H4 by the L3MBTL1-HP1γ complex. The full-length L3MBTL1 can also homodimerize via its SAM domain (Wismar et al. 1995). Dimerization of full length L3MBTL1 may also increase the local concentration of the histone-bound 3MBT domains, further facilitating 3MBT-histone complex dimer formation. Thus, L3MBTL1 actively bridges neighboring nucleosomes. Together, the concerted action of L3MBTL1 and HP1 foster chromatin compaction at Rb-regulated genes.

dScm

Functional analyses in Drosophila show that the MBT domain of Scm and its methyl-lysine-binding activity are required for repression of *Hox* genes (Bornemann et al. 1998). The two malignant brain tumour (MBT) repeats of Scm form a domain that preferentially binds to monomethylated lysine histones(Grimm et al. 2007). The crystal structure of a monomethyl-lysine-containing histone tail peptide bound to the MBT repeat domain shows that the methyl-lysine side chain occupies a binding pocket in the second MBT repeat formed by three conserved aromatic residues and one aspartate, which resembles the L3MBTL1 binding mode (Li et al. 2007;Min et al. 2007;Sathyamurthy et al. 2003). Insertion of the monomethylated side chain into this pocket seems to be the main contributor to the binding affinity, which also resembles L3MBTL1 (Figure 2B). But, the difference between these two MBT histone binders is that L3MBTL1 can bind mono- and di-methylated lysine histones almost equally well. The lower affinity shown by dScm for histone peptides carrying dimethylated peptides can also be explained by their dSCM-Kme2 complex structure. Superimpositon of the dScm structures bound to monomethylated and dimethylated lysine shows that monomethyl-lysine binds approximately 0.6 Å deeper into the pocket. In addition, in the dimethyl-lysine-containing complex, the water molecule that contacts the epsilon-amino group is displaced by the additional methyl group. The observed differences might explain the preferred binding for monomethylated lysine histones.

Structurally and Functionally Diverse TUDOR domains

The Tudor domain family consists of a large number of proteins with varying functions sharing a common feature – the conservation of a small 50 amino acid motif that serves to mediate intermolecular protein interactions. The Tudor domain was first identified in the Tud protein from *Drosophila*, one of the first gene products identified as part of the posterior-group genes (Boswell et al. 1985). The *tud* gene encodes a large peptide of 2515 amino acids with more than 50% of the domain-encoding segments contributing to 11 copies of the individual Tudor domain (Ponting 1997). The Tudor domain itself was first discovered through protein sequence analysis that suggested sequence motif repetition throughout the protein and an absence of any other domains of known function. The retention of this domain in proteins exhibiting RNA localization, initially suggested a role for the Tudor domain-containing proteins in pre-mRNA processing. Sequence alignments of the individual domain with other proteins also permitted identification of the Tudor domain in a variety of non-*Drosophila* proteins, including human AKAP149, p100 and Survival of Motor Neuron (SMN) protein (Ponting 1997). Coincident with this work, the p100 domain architecture was independently demonstrated via hydrophobic cluster analysis to include a C-terminal Tudor domain, in addition to four *Staphlycoccal nuclease* (Sn)-like domains (Callebaut et al. 1997). More complete analysis of proteins with predicted Tudor domains failed to support the RNA-binding function and experimental evidence from studies of the *Drosophila* Ovarian Tumor (OTU) protein clearly ascertained that the domain did not contribute to the RNA-binding ability of this protein (Glenn et al. 2001). In recent literature, both structural and biochemical evidence has demonstrated the ability of several Tudor domains to interact with methylated protein

partners. The global folds and mechanism of peptide recognition of the representative Tudor domain-containing proteins are further discussed below.

The SMN and p100 proteins: Single Tudor Domains

The mammalian SMN protein contains a single Tudor domain, mutations in which have been associated with autosomal recessive proximal spinal muscular atrophy (SMA) (Talbot et al. 1998). SMN structures exhibit a canonical Tudor domain and consist of a strongly bent antiparallel β-sheet (Figure 3A). The five strands comprise the barrel-like motif that is lined on the lower face by the lengthy β2 strand and closed by the antiparallel interaction between β1 and β5. Structural stability is promoted through the formation of a hydrophobic core via residues C98, A100, A111, I113, I116, C123, V125 and L141 with core residues exhibiting a high degree of sequence conservation between known Tudor domains (Selenko et al. 2001). The high resolution crystal and highly restrained NMR structures have permitted more detailed investigations into the ligand binding site and the unique structural properties of this domain. Biochemical studies have demonstrated that the symmetric dimethylation of arginine residues (sDMA) on the tails of SmB, SmD1 and SmD3 proteins increased the binding affinity of these proteins for the Tudor domain of SMN (Brahms et al. 2001;Friesen et al. 2001). NMR titrations based on these observations definitively showed that the SMN Tudor domain was capable of recognizing sDMA residues in arginine-glycine rich repeats and that the methyl groups of the sDMA were associated with a cluster of highly conserved aromatic residues of the Tudor domain (Sprangers et al. 2003).

Comparison of the crystal structure of the SMN Tudor domain with the highly-restrained NMR structure revealed significant structural differences in the region of the sDMA binding site. The most notable differences between the crystal and solution state structures were observed in the conformations of the aromatic residues and loops about the sDMA binding pocket. The W102 sidechain was shown to have alternative rotamers between these two structures and the coordinate root mean squared deviation for the aromatic residues in this region was approximately 2.0 Å (Sprangers et al. 2003). While the features of both structures appear to be divergent but equally valid, it is clear that the structure determination of the SMN Tudor domain in complex with the relevant peptides from the Sm protein tails, and potentially other methylated peptides, will be important in better understanding the possible binding modes and methylarginine recognition by this domain. It will also be important to assess ability of the SMN Tudor domain to interact with other methylated peptides, including those with varying degrees of lysine methylation.

In contrast to the SMN protein, which contains only the Tudor domain, the modular architecture of the p100 protein has adapted for numerous protein-protein interactions. The p100 protein has been shown to interact with STAT5TAD, STAT6TAD, CBP, RNA polymerase II, RNA helicase A, Myb and Pim1 serine/threonine kinases through SN-like domains (Leverson et al. 1998;Paukku et al. 2003;Tong et al. 1995;Valineva et al. 2005;Valineva et al. 2006;Yang et al. 2002). Biochemical investigations of the Tudor domain of this protein have shown that the expression of this domain does not affect the

transcriptional activity of the protein (Yang et al. 2002), but rather interacts with U5 sn-RNP-specific proteins and promotes pre-mRNA spicing through these intermolecular protein interactions (Shaw et al. 2007). The three-dimensional crystal structure of the C-terminal Tudor and SN (TSN) domain contained two individual protein fragments in the asymmetric unit. The larger fragment consisted of the complete interdigitated Tudor and SN domains corresponding to residues 654-870 of the full length p100 protein. The smaller peptide was the Tudor domain alone and consisted of residues 680-770 (Figure 3B). The TSN domain contained four α-helices, nine β-strands and 14 loops. Of these secondary structural elements, helix α1 and strands 3-6 were can be attributed to the Tudor domain, resulting in a structure that is highly analogous to the Tudor domain of the SMN protein. Superimposition of the SMN Tudor domain onto the Tudor domain of p100 reveals few differences between the folds and yields a root mean square deviation of 1.2 Å between the Cα atoms of the two structures (Shaw et al. 2007). Like the SMN fold, the p100 Tudor is stabilized through a hydrophobic core. Also conserved in this region are the aromatic residues predicted to be involved in the binding of methylated ligands. The p100 structure surprisingly revealed the caging of Leu808 from a neighbouring p100 molecule in the crystal by the aromatic residues, structurally confirming the protein-recognition ability of p100 and suggesting that peptide interaction with the Tudor domain likely involves methylated substrates (Shaw et al. 2007).

Esa1 – A knotted single Tudor Domain

The Sas-related acetyltransferase (Esa1) from *Sacharomyces cerevisiae* is a subunit of the nucleosomal acetyltransferase of histone H4 (NuA4) complex and is responsible for the

acetylation of histone H4 and H2A (Allard et al. 1999;Clarke et al. 1999;Lafon et al. 2007;Smith et al. 1998). In yeast, Esa1 is the only essential acetyltransferase; it is necessary for cell division (Clarke et al. 1999;Smith et al. 1998) and is also involved in DNA double-strand break repair (Utley et al. 2005), transcription initiation and elongation (Eisen et al. 2001;Morillon et al. 2005) transcriptional silencing (Clarke et al. 2006) and cell cycle regulation (Clarke et al. 1999). Early sequence analysis predicted the N-terminus of Esa1 to contain a chromodomain that was possibly involved in RNA binding and that was required full activity of the acetyltransferase domain (Letunic et al. 2006;Shimojo et al. 2008). However, subsequent NMR structural determination of two Esa1 constructs revealed the domain to be more similar to the Tudor domain family with several interesting structural variations (Figure 3C) (Shimojo et al. 2008).

Both Esa1 constructs contain similar β-barrel structures, each containing five β-strands, as opposed to the predicted three strands plus one helix that would be expected for a chromodomain (Shimojo et al. 2008). The overall structures of the Esa1 variants align well with the Tudor domain of SMN (Shimojo et al. 2008), though significant differences are observed at both the N- and C-terminal tails of the longer construct when compared with SMN and the shorter Esa1 construct. The long construct contains an additional β-sheet formed by an N-terminal β-strand (residues 11-13) and a C-terminal β-strand (residues 78-80), which is described by the authors as acting as a “knot” for the Tudor domain (Shimojo et al. 2008). Heteronuclear {1H}-15N NOE experiments revealed the NOEs of the long construct to be significantly higher than those of the short Esa1 construct, suggesting the additional structural features contribute added stability to the

core domain (Shimojo et al. 2008). Accordingly, several hydrophobic residues found in the longer construct pack into the hydrophobic corresponding to the complete core of the shorter construct. The added stability in the core in turn induces the formation of an additional helical turn in the loop region connecting strands β3 and β4 (Shimojo et al. 2008). To date, there has been no structural evidence supporting the interaction of Esa1 with any particular ligand. From analysis of the NMR structure, a slight depression is found in a region corresponding to the location of the aromatic cage seen in other Tudor domains. Two tyrosine residues are also located in this region, along with a glutamate residue, which may suggest that the Esa1 Tudor is more similar to other Tudor domains than initially hypothesized. Recognition of the extended conformations of either methylated lysine or arginine residues, however, would require significant induced fit that may be precluded by the overall stability of the hydrophobic Tudor core.

53BP1 and Crb2: Orthologous tandem Tudor domains

The mammalian protein p53 binding protein (53BP1) was originally identified in a yeast two-hybrid screen as a protein that interacts with the DNA binding domain of p53 through its two C-terminal BRCT domains (Derbyshire et al. 2002;Iwabuchi et al. 1994;Joo et al. 2002). The protein rapidly forms foci at the sites of DNA double strand breaks (DSBs) upon exposure to ionizing radiation and is phosphorylated by ATM kinase, a protein central in the signaling responses to DSBs (Anderson et al. 2001;Rappold et al. 2001;Schultz et al. 2000). Taken together with a multitude of other studies, 53BP1 has been shown to be a central mediator of the DNA damage checkpoint (Charier et al. 2004). Based on sequence analysis, the region residing between the N-

terminal phosphorylated region and the C-terminal BRCT domains of 53BP1 was predicted to contain a single Tudor domain (Charier et al. 2004). Initial NMR characterization of this region of the mouse 53BP1 homologue, which exhibits 99% sequence identity to the human homologue, revealed the presence of two Tudor domains (Figure 3D). Similar to the single Tudor domains of SMN and p100, the individual 53BP1 Tudor domains each contain a strongly bent β-sheet that folds into a compact β-barrel. Each domain also contains a long β-strand that is bent into a 90° angle through the presence of classical β-bulges at positions 30, 31 and 39 in domain 1 and positions 82, 83 and 94 in the second Tudor domain. The individual domains are connected by a six-residue linker and are largely superimposable, with the exception of loops joining β1 and β2 and β2 with β3. Despite the relatively low sequence identity of 18%, the root mean square deviation between the backbone atoms of the NMR structure was calculated to be 1.9 Å (Charier et al. 2004).

Crystallographic analysis of the 53BP1 yeast orthologue Crb2 also revealed the presence of tandem Tudor domains, despite the limited sequence similarity in the homologous region to that of 53BP1 (Figure 3D) (Botuyan et al. 2006). As with sequence analysis of 53BP1, the amino acid sequence of Crb2 alone was insufficient to predict the existence of the second Tudor domain. Structural alignments of the tandem Tudor domains of 53BP1 and Crb2 shows a similar orientation of the individual domains relative to one another, though both Tudor domains of Crb2 are significantly larger than those of 53BP1 (Botuyan et al. 2006). Additional residues are located in the loops connecting the third and fourth strands from each domain when compared to the corresponding loops of

53BP1. The detailed structural comparison of 53BP1 and Crb2 has highlighted the high degree of structural conservation within the Tudor fold and between tandem Tudor domains of orthologous proteins, even when domain annotation of proteins is precluded by a lack of sequence similarity. This work has also indirectly suggested the existence of a much larger family of Tudor domain-containing proteins, with additional members remaining unidentified owing to poor sequence conservation with known family members.

The hybrid tandem Tudor of JMJD2A

JMJD2A is a member of a conserved family of Jmjc domain-containing proteins belonging to the Jmjc histone deacetylase superfamily (Tsukada et al. 2006). With regard to its specific function, JMJD2A has been reported to interact with histone deacetylase complexes and Rb and to function as a transcription repressor (Gray et al. 2005;Zhang et al. 2005). In contrast to the tandem Tudor domains of 53BP1 and Crb2, the crystal structure of JMJD2A revealed an alternate mode of tandem Tudor assembly, termed a hybrid tandem Tudor (Figure 3E). The two domains of JMJD2A are interdigitated forming a saddle-shaped structure with each lobe resembling the canonical Tudor domain of SMN and p100 (Huang et al. 2006). The hybrid Tudor domains are formed by the exchange of β-strands 3 and 4 with respect to the canonical Tudor domain. The swapped β3 strand forms a long contiguous strand with the nonswapped β2. The two long strands bridge the two individual domains (Huang et al. 2006).

Triple Tudor Repeat: Spindlin 1

Spindlin is a major maternal transcript in the mouse that has been found to associate and co-migrate with the meiotic spindle in the first meiotic cell cycle (Oh et al. 1997). Spindlin1, the human homologue, exhibits up-regulated expression in ovarian cancer cells, where it localizes in the cell nuclei. The over-expression of Spindlin1 is believed to correlate with variations in cell cycle distribution during mitosis, which is consistent with studies of other spindlin1 homologues (Fletcher et al. 2002). The crystal structure of spindlin1 consists of three tandem repeats of Tudor-like domains (Figure 3F) (Zhao et al. 2007). The individual domains adopt similar folds with a total sequence identity of 32% among three domains. The spindlin1 and Tudor domains all compose two sheets that are made of five anti-parallel β-strands with portions of the second strains participating in both β-sheets. As seen in the canonical Tudor domains, the long β-strands are bent by approximately 90° about classical β-bulges, though the orientations of the fifth β-strand are somewhat divergent between the spindlin1 and Tudor domains. Globally, the β-sheets of each domain are arranged such that they lie perpendicular to one another with the first β-sheet of each domain exposed to the solvent and the second sheet buried in the protein interior (Zhao et al. 2007). To date, spindlin1 is the largest of the Tudor domain-containing proteins to be studied structural, thus further comparisons for the global topology of Tudor domain association cannot be made. As with Esa1, there virtually no experimental evidence for the role of Spindlin1 in protein-protein interactions, the identity of its binding partners, or the ability of the protein to interact with a methylated ligand through one of more of its canonical Tudor domains.

Methyl lysine recognition through the aromatic cage

Histones undergo numerous types on post-translational modification, including methylation. The combination of modifications is believed to extend heritable information beyond the DNA sequence alone (Strahl et al. 2000). Specific methylation states of the histones may result in either up-regulation or down-regulation of transcriptional activities of various genes (Kouzarides 2002;Sims, III et al. 2003;Sims, III et al. 2006). In general, prior studies of the Tudor, chromodomains and PHD fingers have suggested that the mode of methyl-histone recognition involves a cage of two to four aromatic amino acids and an acidic residue to balance the charge of the methylammonium ion (Corsini et al. 2007). The best studied examples of such an interaction are the crystal structures of the 53BP1 and JMJC2A proteins in complex with various histone peptides.

The JMJD2A hybrid tandem Tudor domain has the novel property of binding two distinct peptides, H3K4me3 and H4K20me3, with equal affinity. Both peptides interact with the second of the Tudor domains and the methyllysine of each peptide is also caged by F932, W967 and Y973 (Huang et al. 2006;Lee et al. 2008). In each case, the resolution of the crystallographic data permitted unambiguous assignment of the majority of residues in the peptide and showed the peptides to bind in opposite orientations along the protein surface (Figure 3E, panel). For the H3K4me3 peptide, R2 recognition by D945 of JMJC2A was one of the primary stabilizing interactions. The H4K20me3 peptide was also stabilized through an interaction of a peptide arginine with the Tudor domain. In this case, however, R19 of the peptide bound D939 of JMJD2A. No interaction was observed

between D945 and the H4K20me3. Likewise, N940 of JMJD2A hydrogen bonds with the backbone amide and hydroxyl groups of residue T3 from H3K4me3 but did not participate in any interactions with H4K20me3. Residues Y942 and T968 are also involved in weaker intermolecular hydrogen bonds and selectively contact only one of the peptides. Through the hydroxyl group, Y942 forms a hydrogen bond with the terminal amino group of the H3K4me3 peptides but is distant from the binding site of the H4K20me3 peptide. The hydroxyl group of T968, however, forms a hydrogen bond with the guanidino group of R23 from H4K29me3 but does not contact the H3K4me3 peptide. It is important to note that despite major differences in the binding modes of these peptides, no structural changes were noted on the JMJC2A protein on peptide binding (Lee et al. 2008). It has been suggested that the binding modes ensure the dual specificity of the protein while preventing simultaneous recognition of both histone peptides. While the *in vivo* relevance of the dual specificity of JMJD2A has yet to be established, these studies have highlighted the complexity of methyllysine recognition and inherent difficulty in definitively determining the *in vivo* specificity of the methyllysine recognition domains.

Accordingly, ligand binding studies for the 53BP1 protein have provided several suggestions of relevant peptides. Initial NMR titrations of 53BP1 suggested that the tandem Tudor domain was capable of interaction with both unmethylated and symmetrically dimethylated arginine-glycine-rich peptides (Charier et al. 2004) though the fast exchange rate yielded dissociation constants in the millimolar range. Despite the low affinity of the protein for the peptides, the interaction was mapped to a cavity

containing three aromatic and two charged residues (Charier et al. 2004). More detailed analysis of the interaction of 53BP1 with a series of H3 and H4 peptides demonstrated that the protein was capable of interaction with dimethylated H4K20 peptides with micromolar affinity (Botuyan et al. 2006) and bound dimethylated H3K79 with much lower affinity (Kd, 2.0 mM) (Botuyan et al. 2006). The high resolution crystal structures of 53BP1 in its apo and H4K20me2-bound forms permitted extensive analysis of the induced fit mechanism employed by 53BP1 in the interaction (Figure 3D, panel). The dimethyllysine binding pocket of 53BP1 is distinct from any of the other known methyllysine recognition domains, which all preferentially bind trimethylated peptides in addition to binding other methylation sites with high affinity. Relative to such domains, the 53BP1 peptide binding site is much narrower and involves a greater number of interactions that ensure a optimal complementarity between the dimethyllysine and the binding pocket (Botuyan et al. 2006). The dimethyllysine is caged by four aromatic residues (W1495, Y1502, F1519, and Y1523) and D1521 in the first Tudor domain. The aromatic amino acid participate in can der Waals and cation-π interactions with the methyllysine ammonium group while a salt bridge is formed between the carboxylate group of D1521 and the dimethylammonium ion of the H4K20me2 peptide. A direct hydrogen bond is also apparent between the carboxylate group of D1521 and the amino protein of the methylated lysine (Botuyan et al. 2006). The lack of amino proton in the trimethyllysine variant of the peptide likely contributes to the specificity of the interaction for the dimethylated peptide, as the loss of this proton precludes formation of this hydrogen bond. With regard to the unmethylated and monomethylated peptides, it is likely that the less optimal van der Waals and cation-π interactions with the four aromatic

rings in the binding cage of 53BP1 not entirely compensated for by the presence of the D1521-mediated hydrogen bonding. Additionally, the increased affinity of 53BP1 for the h4K20me2 peptide is possible due to the presence of a histidine residue that immediately before the methyllysine in this peptide while the H3K9 and H3K27 peptides all retain an arginine in the this position, though the lack of density observed in the crystal structure suggests that the peptide exhibits significant conformational flexibility outside of the methyllysine recognition site. Additional binding and structural studies are required to fully delineate the interactions involved in peptide recognition that lie on the surface of 53BP1.

In addition to JMJD2A and 53BP1, other Tudor domains have received some attention in the search for relevant cellular ligands. Classical biochemical and cell biology studies have demonstrated the interaction of SMN with fibrillarin, a highly conserved nucleolar protein that is associated with all Box C/D small nucleolar RNAs and functions in processing and modification of pre-rRNA (Jones et al. 2001). SMN has also been shown to interact with Sm proteins, core components of the snRNPs (Sprangers et al. 2003). NMR titrations and deletion studies have identified the symmetrically dimethylated arginine residues in the arginine-glycine-rich tails of the Sm proteins and fibrillin as the binding sites for SMN, respectively (Jones et al. 2001;Sprangers et al. 2003). The p100 protein has also been identified as a member of the U snRNP complexes and the interaction is directly mediated through the TSN domain. In a recurring theme, SMN, and likely p100, employs an aromatic cage in the recognition of the methylarginine ligands.

Concluding Remarks

The structural studies of PHD, MBT and Tudor domains have revealed similarities in the mechanisms by which these proteins recognize methylated lysines. In all cases described, the conserved utilization of aromatic caging has been identified as a necessary feature in methyl lysine binding. The exploitation of limited variation in the number and identity of the aromatic and charged residues, in combination with variations in amino acid sequence in the histone tails, appears confer specificity of binding to these recognition domains. The conservation of the aromatic cage for recognition of diverse histone peptides and degrees of methylation raises several questions with regard to the *in vivo* functions of these protein domains. Foremost of the outstanding problems is probably the determination of contribution of the individual to the binding specificity. How do simple amino acid substitutions in the aromatic cage lend to the differentiation between the lysine methylation states? To what extent do the surface complementary and inter-residue interactions determine the global specificity the protein-protein interaction? The contribution of steric restraint on the specificity of the protein-histone interaction when the conformation of the entire histone tail is considered rather than the peptide alone is also of interest. The intramolecular interactions between the a single methyl-lysine-binding domain and other domains, which may also exhibit methyl-lysine affinity, add another significant complication to any attempts at elucidating the relevant *in vivo* activity of the larger proteins of which these domains are often only a small part. While the above studies have contributed a large amount of knowledge to an area of research that remains in its infancy, it is clear that there is much, even with respect to methyl-

lysine-binding proteins alone, that remains to be investigated. The multiple layers of complexity beginning with the multiple potential ligands for each domain up to the dynamic interchange of histone modifications and the association of the histone code reading proteins within numerous multi-subunit protein complexes add pose numerous challenged in any attempt to dissect the fundamental mechanisms of Epigenetic inheritance.

Acknowledgements

This research was supported by the Structural Genomics Consortium, a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust.

References

Allard,S., Utley,R.T., Savard,J., Clarke,A., Grant,P., Brandl,C.J., Pillus,L., Workman,J.L., and Cote,J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. **18,** 5108-5119.

Allgrey,V.G., Faulkner,R., and Mirsky,A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U. S. A **51,** 786-794.

Anderson,L., Henderson,C., and Adachi,Y. (2001). Phosphorylation and rapid relocalization of 53BP1 to nuclear foci upon DNA damage. Mol. Cell Biol. **21,** 1719-1729.

Belenkaya,T.Y., Han,C., Standley,H.J., Lin,X., Houston,D.W., Heasman,J., and Lin,X. (2002). pygopus Encodes a nuclear protein essential for wingless/Wnt signaling. Development **129,** 4089-4101.

Bhaumik,S.R., Smith,E., and Shilatifard,A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat. Struct. Mol. Biol. **14,** 1008-1016.

Boccuni,P., MacGrogan,D., Scandura,J.M., and Nimer,S.D. (2003). The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). J. Biol. Chem. **278,** 15412-15420.

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

Boswell,R.E., and Mahowald,A.P. (1985). tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell **43,** 97-104.

Botuyan,M.V., Lee,J., Ward,I.M., Kim,J.E., Thompson,J.R., Chen,J., and 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.

Bourc'his,D., and Bestor,T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature **431,** 96-99.

Bourc'his,D., Xu,G.L., Lin,C.S., Bollman,B., and Bestor,T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science **294,** 2536-2539.

Brahms,H., Meheus,L., de,B., V, Fischer,U., and Luhrmann,R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA. **7,** 1531-1542.

Callebaut,I., and Mornon,J.P. (1997). The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in Drosophila melanogaster development. Biochem. J. **321** (Pt 1)**,** 125-132.

Charier,G., Couprie,J., pha-Bazin,B., Meyer,V., Quemeneur,E., Guerois,R., Callebaut,I., Gilquin,B., and Zinn-Justin,S. (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure. **12,** 1551-1562.

Chedin,F., Lieber,M.R., and Hsieh,C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc. Natl. Acad. Sci. U. S. A **99,** 16916-16921.

Clarke,A.S., Lowell,J.E., Jacobson,S.J., and Pillus,L. (1999). Esa1p is an essential histone acetyltransferase required for cell cycle progression. Mol. Cell Biol. **19,** 2515-2526.

Clarke,A.S., Samal,E., and Pillus,L. (2006). Distinct roles for the essential MYST family HAT Esa1p in transcriptional silencing. Mol. Biol. Cell **17,** 1744-1757.

Corsini,L., and Sattler,M. (2007). Tudor hooks up with DNA repair. Nat. Struct. Mol. Biol. **14,** 98-99.

Derbyshire,D.J., Basu,B.P., Serpell,L.C., Joo,W.S., Date,T., Iwabuchi,K., and Doherty,A.J. (2002). Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. EMBO J. **21,** 3863-3872.

Dhalluin,C., Carlson,J.E., Zeng,L., He,C., Aggarwal,A.K., and Zhou,M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature **399,** 491-496.

Dul,B.E., and Walworth,N.C. (2007). The plant homeodomain fingers of fission yeast Msc1 exhibit E3 ubiquitin ligase activity. J. Biol. Chem. **282,** 18397-18406.

Eberharter,A., Vetter,I., Ferreira,R., and Becker,P.B. (2004). ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. EMBO J. **23,** 4029-4039.

Ehrenhofer-Murray,A.E. (2004). Chromatin dynamics at DNA replication, transcription and repair. Eur. J. Biochem. **271,** 2335-2349.

Eisen,A., Utley,R.T., Nourani,A., Allard,S., Schmidt,P., Lane,W.S., Lucchesi,J.C., and Cote,J. (2001). The yeast NuA4 and Drosophila MSL complexes contain homologous subunits important for transcription regulation. J. Biol. Chem. **276,** 3484-3491.

Feng,X., Hara,Y., and Riabowol,K. (2002). Different HATS of the ING1 gene family. Trends Cell Biol. **12,** 532-538.

Fiedler,M., Sanchez-Barrena,M.J., Nekrasov,M., Mieszczanek,J., Rybin,V., Muller,J., Evans,P., and Bienz,M. (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. Mol. Cell **30,** 507-518.

Fischle,W., Wang,Y., Jacobs,S.A., Kim,Y., Allis,C.D., and Khorasanizadeh,S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. **17,** 1870-1881.

Flanagan,J.F., Mi,L.Z., Chruszcz,M., Cymborowski,M., Clines,K.L., Kim,Y., Minor,W., Rastinejad,F., and Khorasanizadeh,S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature **438,** 1181-1185.

Fletcher,B.S., Dragstedt,C., Notterpek,L., and Nolan,G.P. (2002). Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. Leukemia **16,** 1507-1518.

Friesen,W.J., Massenet,S., Paushkin,S., Wyce,A., and Dreyfuss,G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol. Cell **7**, 1111-1117.

Glenn,L.E., and Searles,L.L. (2001). Distinct domains mediate the early and late functions of the Drosophila ovarian tumor proteins. Mech. Dev. **102,** 181-191.

Gozani,O., Karuman,P., Jones,D.R., Ivanov,D., Cha,J., Lugovskoy,A.A., Baird,C.L., Zhu,H., Field,S.J., Lessnick,S.L., Villasenor,J., Mehrotra,B., Chen,J., Rao,V.R., Brugge,J.S., Ferguson,C.G., Payrastre,B., Myszka,D.G., Cantley,L.C., Wagner,G., Divecha,N., Prestwich,G.D., and Yuan,J. (2003). The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell **114,** 99-111.

Gray,S.G., Iglesias,A.H., Lizcano,F., Villanueva,R., Camelo,S., Jingu,H., Teh,B.T., Koibuchi,N., Chin,W.W., Kokkotou,E., and Dangond,F. (2005). Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. J. Biol. Chem. **280,** 28507-28518.

Grewal,S.I., and Moazed,D. (2003). Heterochromatin and epigenetic control of gene expression. Science **301,** 798-802.

Grimm,C., de Ayala Alonso,A.G., Rybin,V., Steuerwald,U., Ly-Hartig,N., Fischle,W., Muller,J., and Muller,C.W. (2007). Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. EMBO Rep. **8,** 1031-1037.

Huang,Y., Fang,J., Bedford,M.T., Zhang,Y., and Xu,R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science **312,** 748-751.

Ivanov,A.V., Peng,H., Yurchenko,V., Yap,K.L., Negorev,D.G., Schultz,D.C., Psulkowski,E., Fredericks,W.J., White,D.E., Maul,G.G., Sadofsky,M.J., Zhou,M.M., and Rauscher,F.J., III (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Mol. Cell **28,** 823-837.

Iwabuchi,K., Bartel,P.L., Li,B., Marraccino,R., and Fields,S. (1994). Two cellular proteins that bind to wild-type but not mutant p53. Proc. Natl. Acad. Sci. U. S. A **91,** 6098-6102.

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

Jacobson,R.H., Ladurner,A.G., King,D.S., and Tjian,R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science **288,** 1422-1425.

Jia,D., Jurkowska,R.Z., Zhang,X., Jeltsch,A., and Cheng,X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature **449,** 248-251.

Jones,D.R., Bultsma,Y., Keune,W.J., Halstead,J.R., Elouarrat,D., Mohammed,S., Heck,A.J., D'Santos,C.S., and Divecha,N. (2006). Nuclear PtdIns5P as a transducer of stress signaling: an in vivo role for PIP4Kbeta. Mol. Cell **23,** 685-695.

Jones,K.W., Gorzynski,K., Hales,C.M., Fischer,U., Badbanchi,F., Terns,R.M., and Terns,M.P. (2001). Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. J. Biol. Chem. **276,** 38645-38651.

Joo,W.S., Jeffrey,P.D., Cantor,S.B., Finnin,M.S., Livingston,D.M., and Pavletich,N.P. (2002). Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. Genes Dev. **16,** 583-593.

Kalakonda,N., Fischle,W., Boccuni,P., Gurvich,N., Hoya-Arias,R., Zhao,X., Miyata,Y., MacGrogan,D., Zhang,J., Sims,J.K., Rice,J.C., and Nimer,S.D. (2008). Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. Oncogene. (E-published ahead of print)

Kim,J., Daniel,J., Espejo,A., Lake,A., Krishna,M., Xia,L., Zhang,Y., and Bedford,M.T. (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., and Muller,J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. **20,** 1110-1122.

Kouzarides,T. (2007). Chromatin modifications and their function. Cell 128**,** 693-705.

Kouzarides,T. (2002). Histone methylation in transcriptional control. Curr. Opin. Genet. Dev. **12,** 198-209.

Kramps,T., Peter,O., Brunner,E., Nellen,D., Froesch,B., Chatterjee,S., Murone,M., Zullig,S., and Basler,K. (2002). Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. Cell **109,** 47-60.

Kwon,J., Imbalzano,A.N., Matthews,A., and Oettinger,M.A. (1998). Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. Mol. Cell **2,** 829-839.

Lafon,A., Chang,C.S., Scott,E.M., Jacobson,S.J., and Pillus,L. (2007). MYST opportunities for growth control: yeast genes illuminate human cancer gene functions. Oncogene **26,** 5373-5384.

Lan,F., Collins,R.E., De,C.R., Alpatov,R., Horton,J.R., Shi,X., Gozani,O., Cheng,X., and Shi,Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature **448,** 718-722.

Lee,J., Thompson,J.R., Botuyan,M.V., and 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.

Letunic,I., Copley,R.R., Pils,B., Pinkert,S., Schultz,J., and Bork,P. (2006). SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. **34,** D257-D260.

Leverson,J.D., Koskinen,P.J., Orrico,F.C., Rainio,E.M., Jalkanen,K.J., Dash,A.B., Eisenman,R.N., and Ness,S.A. (1998). Pim-1 kinase and p100 cooperate to enhance c-Myb activity. Mol. Cell **2,** 417-425.

Li,H., Fischle,W., Wang,W., Duncan,E.M., Liang,L., Murakami-Ishibe,S., Allis,C.D., and Patel,D.J. (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. Mol. Cell **28,** 677-691.

Li,H., Ilin,S., Wang,W., Duncan,E.M., Wysocka,J., Allis,C.D., and Patel,D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature **442,** 91-95.

Liu,Y., Subrahmanyam,R., Chakraborty,T., Sen,R., and Desiderio,S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. Immunity. **27,** 561-571.

Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389,** 251-260.

Martin,D.G., Baetz,K., Shi,X., Walter,K.L., MacDonald,V.E., Wlodarski,M.J., Gozani,O., Hieter,P., and Howe,L. (2006). The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. Mol. Cell Biol. **26,** 7871-7879.

Matthews,A.G., Kuo,A.J., Ramon-Maiques,S., Han,S., Champagne,K.S., Ivanov,D., Gallardo,M., Carney,D., Cheung,P., Ciccone,D.N., Walter,K.L., Utz,P.J., Shi,Y., Kutateladze,T.G., Yang,W., Gozani,O., and Oettinger,M.A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature **450,** 1106-1110.

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

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

Min,J., Zhang,Y., and Xu,R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. **17,** 1823-1828.

Morillon,A., Karabetsou,N., Nair,A., and Mellor,J. (2005). Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. Mol. Cell **18,** 723-734.

Nielsen,P.R., Nietlispach,D., Mott,H.R., Callaghan,J., Bannister,A., Kouzarides,T., Murzin,A.G., Murzina,N.V., and Laue,E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature **416,** 103-107.

Oh,B., Hwang,S.Y., Solter,D., and Knowles,B.B. (1997). Spindlin, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo. Development **124,** 493-503.

Ooi,S.K., Qiu,C., Bernstein,E., Li,K., Jia,D., Yang,Z., Erdjument-Bromage,H., Tempst,P., Lin,S.P., Allis,C.D., Cheng,X., and Bestor,T.H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature **448,** 714-717.

Org,T., Chignola,F., Hetenyi,C., Gaetani,M., Rebane,A., Liiv,I., Maran,U., Mollica,L., Bottomley,M.J., Musco,G., and Peterson,P. (2008). The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep. **9,** 370-376.

Palacios,A., Munoz,I.G., Pantoja-Uceda,D., Marcaida,M.J., Torres,D., Martin-Garcia,J.M., Luque,I., Montoya,G., and Blanco,F.J. (2008). Molecular Basis of Histone H3K4me3 Recognition by ING4. J. Biol. Chem. **283,** 15956-15964.

Parker,D.S., Jemison,J., and Cadigan,K.M. (2002). Pygopus, a nuclear PHD-finger protein required for Wingless signaling in Drosophila. Development **129,** 2565-2576.

Paukku,K., Yang,J., and Silvennoinen,O. (2003). Tudor and nuclease-like domains containing protein p100 function as coactivators for signal transducer and activator of transcription 5. Mol. Endocrinol. **17,** 1805-1814.

Pena,P.V., Davrazou,F., Shi,X., Walter,K.L., Verkhusha,V.V., Gozani,O., Zhao,R., and Kutateladze,T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature **442,** 100-103.

Pena,P.V., Hom,R.A., Hung,T., Lin,H., Kuo,A.J., Wong,R.P., Subach,O.M., Champagne,K.S., Zhao,R., Verkhusha,V.V., Li,G., Gozani,O., and Kutateladze,T.G. (2008). Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J. Mol. Biol. **380,** 303-312.

Ponting,C.P. (1997). Tudor domains in proteins that interact with RNA. Trends Biochem. Sci. **22,** 51-52.

Ragvin,A., Valvatne,H., Erdal,S., Arskog,V., Tufteland,K.R., Breen,K., OYan,A.M., Eberharter,A., Gibson,T.J., Becker,P.B., and Aasland,R. (2004). Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. J. Mol. Biol. **337,** 773-788.

Ramon-Maiques,S., Kuo,A.J., Carney,D., Matthews,A.G., Oettinger,M.A., Gozani,O., and Yang,W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc. Natl. Acad. Sci. U. S. A **104,** 18993-18998.

Rappold,I., Iwabuchi,K., Date,T., and Chen,J. (2001). Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. J. Cell Biol. **153,** 613-620.

Santos-Rosa,H., Schneider,R., Bannister,A.J., Sherriff,J., Bernstein,B.E., Emre,N.C., Schreiber,S.L., Mellor,J., and Kouzarides,T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature **419,** 407-411.

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

Schneider,J., Wood,A., Lee,J.S., Schuster,R., Dueker,J., Maguire,C., Swanson,S.K., Florens,L., Washburn,M.P., and Shilatifard,A. (2005). Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. Mol. Cell **19,** 849-856.

Schultz,L.B., Chehab,N.H., Malikzay,A., and Halazonetis,T.D. (2000). p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. J. Cell Biol. **151,** 1381-1390.

Selenko,P., Sprangers,R., Stier,G., Buhler,D., Fischer,U., and Sattler,M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat. Struct. Biol. **8,** 27-31.

Shaw,N., Zhao,M., Cheng,C., Xu,H., Saarikettu,J., Li,Y., Da,Y., Yao,Z., Silvennoinen,O., Yang,J., Liu,Z.J., Wang,B.C., and Rao,Z. (2007). The multifunctional human p100 protein 'hooks' methylated ligands. Nat. Struct. Mol. Biol. **14,** 779-784.

Shi,X., Hong,T., Walter,K.L., Ewalt,M., Michishita,E., Hung,T., Carney,D., Pena,P., Lan,F., Kaadige,M.R., Lacoste,N., Cayrou,C., Davrazou,F., Saha,A., Cairns,B.R., Ayer,D.E., Kutateladze,T.G., Shi,Y., Cote,J., Chua,K.F., and Gozani,O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature **442,** 96-99.

Shi,X., Kachirskaia,I., Walter,K.L., Kuo,J.H., Lake,A., Davrazou,F., Chan,S.M., Martin,D.G., Fingerman,I.M., Briggs,S.D., Howe,L., Utz,P.J., Kutateladze,T.G., Lugovskoy,A.A., Bedford,M.T., and Gozani,O. (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J. Biol. Chem. **282,** 2450-2455.

Shi,Y., Lan,F., Matson,C., Mulligan,P., Whetstine,J.R., Cole,P.A., Casero,R.A., and Shi,Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell **119,** 941-953.

Shilatifard,A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu. Rev. Biochem. **75,** 243-269.

Shimojo,H., Sano,N., Moriwaki,Y., Okuda,M., Horikoshi,M., and Nishimura,Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J. Mol. Biol. **378,** 987-1001.

Sims,R.J., III, Nishioka,K., and Reinberg,D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet. **19,** 629-639.

Sims,R.J., III, and Reinberg,D. (2006). Histone H3 Lys 4 methylation: caught in a bind? Genes Dev. **20,** 2779-2786.

Smith,E.R., Eisen,A., Gu,W., Sattah,M., Pannuti,A., Zhou,J., Cook,R.G., Lucchesi,J.C., and Allis,C.D. (1998). ESA1 is a histone acetyltransferase that is essential for growth in yeast. Proc. Natl. Acad. Sci. U. S. A **95,** 3561-3565.

Soliman,M.A., and Riabowol,K. (2007). After a decade of study-ING, a PHD for a versatile family of proteins. Trends Biochem. Sci. **32,** 509-519.

Sprangers,R., Groves,M.R., Sinning,I., and 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.

Strahl,B.D., and Allis,C.D. (2000). The language of covalent histone modifications. Nature **403,** 41-45.

Talbot,K., Miguel-Aliaga,I., Mohaghegh,P., Ponting,C.P., and Davies,K.E. (1998). Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. Hum. Mol. Genet. **7,** 2149-2156.

Taverna,S.D., Ilin,S., Rogers,R.S., Tanny,J.C., Lavender,H., Li,H., Baker,L., Boyle,J., Blair,L.P., Chait,B.T., Patel,D.J., Aitchison,J.D., Tackett,A.J., and Allis,C.D. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Mol. Cell **24,** 785-796.

Thompson,B., Townsley,F., Rosin-Arbesfeld,R., Musisi,H., and Bienz,M. (2002). A new nuclear component of the Wnt signalling pathway. Nat. Cell Biol. **4,** 367-373.

Tong,X., Drapkin,R., Yalamanchili,R., Mosialos,G., and Kieff,E. (1995). The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. Mol. Cell Biol. **15,** 4735-4744.

Trojer,P., Li,G., Sims,R.J., III, Vaquero,A., Kalakonda,N., Boccuni,P., Lee,D., Erdjument-Bromage,H., Tempst,P., Nimer,S.D., Wang,Y.H., and Reinberg,D. (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell **129,** 915-928.

Tsukada,Y., Fang,J., Erdjument-Bromage,H., Warren,M.E., Borchers,C.H., Tempst,P., and Zhang,Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature **439,** 811-816.

Usui,H., Ichikawa,T., Kobayashi,K., and Kumanishi,T. (2000). Cloning of a novel murine gene Sfmbt, Scm-related gene containing four mbt domains, structurally belonging to the Polycomb group of genes. Gene **248,** 127-135.

Utley,R.T., Lacoste,N., Jobin-Robitaille,O., Allard,S., and Cote,J. (2005). Regulation of NuA4 histone acetyltransferase activity in transcription and DNA repair by phosphorylation of histone H4. Mol. Cell Biol. **25,** 8179-8190.

Vakoc,C.R., Mandat,S.A., Olenchock,B.A., and Blobel,G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol. Cell **19,** 381-391.

Valineva,T., Yang,J., Palovuori,R., and Silvennoinen,O. (2005). The transcriptional co-activator protein p100 recruits histone acetyltransferase activity to STAT6 and mediates interaction between the CREB-binding protein and STAT6. J. Biol. Chem. **280,** 14989-14996.

Valineva,T., Yang,J., and Silvennoinen,O. (2006). Characterization of RNA helicase A as component of STAT6-dependent enhanceosome. Nucleic Acids Res. **34,** 3938-3946.

Vermeulen,M., Mulder,K.W., Denissov,S., Pijnappel,W.W., van Schaik,F.M., Varier,R.A., Baltissen,M.P., Stunnenberg,H.G., Mann,M., and Timmers,H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell **131,** 58-69.

Wilson,D.R., Norton,D.D., and Fugmann,S.D. (2008). The PHD domain of the sea urchin RAG2 homolog, SpRAG2L, recognizes dimethylated lysine 4 in histone H3 tails. Dev. Comp Immunol. **32,** 1221-1230.

Wismar,J., Loffler,T., Habtemichael,N., Vef,O., Geissen,M., Zirwes,R., Altmeyer,W., Sass,H., and Gateff,E. (1995). The Drosophila melanogaster tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. Mech. Dev. **53,** 141-154.

Wood,A., Shukla,A., Schneider,J., Lee,J.S., Stanton,J.D., Dzuiba,T., Swanson,S.K., Florens,L., Washburn,M.P., Wyrick,J., Bhaumik,S.R., and Shilatifard,A. (2007). Ctk complex-mediated regulation of histone methylation by COMPASS. Mol. Cell Biol. **27,** 709-720.

Wysocka,J., Swigut,T., Milne,T.A., Dou,Y., Zhang,X., Burlingame,A.L., Roeder,R.G., Brivanlou,A.H., and Allis,C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell **121,** 859-872.

Wysocka,J., Swigut,T., Xiao,H., Milne,T.A., Kwon,S.Y., Landry,J., Kauer,M., Tackett,A.J., Chait,B.T., Badenhorst,P., Wu,C., and Allis,C.D. (2006). A PHD

finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature **442,** 86-90.

Xia,Z.B., Anderson,M., Diaz,M.O., and Zeleznik-Le,N.J. (2003). MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc. Natl. Acad. Sci. U. S. A **100,** 8342-8347.

Xiao,T., Hall,H., Kizer,K.O., Shibata,Y., Hall,M.C., Borchers,C.H., and Strahl,B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. **17,** 654-663.

Yang,J., Aittomaki,S., Pesu,M., Carter,K., Saarinen,J., Kalkkinen,N., Kieff,E., and Silvennoinen,O. (2002). Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. EMBO J. **21,** 4950-4958.

Zeng,L., Yap,K.L., Ivanov,A.V., Wang,X., Mujtaba,S., Plotnikova,O., Rauscher,F.J., III, and Zhou,M.M. (2008). Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. Nat. Struct. Mol. Biol. **15,** 626-633.

Zhang,D., Yoon,H.G., and Wong,J. (2005). JMJD2A is a novel N-CoR-interacting protein and is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2). Mol. Cell Biol. **25,** 6404-6414.

Zhao,Q., Qin,L., Jiang,F., Wu,B., Yue,W., Xu,F., Rong,Z., Yuan,H., Xie,X., Gao,Y., Bai,C., Bartlam,M., Pei,X., and Rao,Z. (2007). Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation. J. Biol. Chem. **282,** 647-656.

Zhou,Y., and Grummt,I. (2005). The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. Curr. Biol. **15,** 1434-1438.

Figure Legends

Figure 1. Histone lysine binding pockets in PHD finger proteins. **A.** ING2-H3K4me3 complex. The K4me3 is accommodated in a two-aromatic-residue cage. **B.** BPTF-H3K4me3 complex. The K4me3 is accommodated in a four-aromatic-residue cage. **C.** RAG2-H3K4me3 complex. The K4me3 is accommodated in a two-aromatic-residue channel. **D.** BHC80-H3K4me0 complex. The unmethylated lysine is recognized by hydrogen bonds and electrostatic interactions. **E.** DNMT3L-H3K4me0 complex. The unmethylated lysine is recognized by hydrogen bonds and electrostatic interactions. **F.** PYGO1-H3K4me3 complex. The methyl-lysine binding pocket consists of aromatic residues and a negatively charged residue, which preferentially bind H3K4me2 over H3K4me3

Figure 2. MBT domain functions as a lowly methylated lysine recognition motif. A. L3MBTL1-H4K20me2 complex. The second MBT repeat in L3MBTL1 selectively binds mono or di-methylated histone lysine peptides. **B.** The crystal structure of Drosophila Scm in complex with a mono-methylated peptide RKme1S. The mono-methylated lysine is inserted in a binding pocket formed by three aromatic residues and a negatively charged residue from the C-terminal MBT repeat.

Figure 3. Structural diversity and variation in the modes of peptide recognition amongst members of the Tudor domain family. For each structure, the residues forming the aromatic cage involved in ligand recognition are shown with corresponding

numbering of the full-length protein. The canonical Tudor domain is colored in cyan and accessory helices in chartreuse. Additional features or Tudor domains are colored in light cyan. **A.** Survival of Motor Neuron (SMN) protein exhibits what has been described as the canaonical Tudor domain (PDB 1MHN). **B.** The p100 Tudor domain (PDB 2HQX). **C.** The knotted Tudor domain of the Esa1 protein (PDB 2RO0). The additional β-strands of the knot are coloring light cyan at the N- and C-termini. **D.** 53BP1 contains a competent ligand binding site in the first of two tandem Tudor domains. The interaction of 53BP1 with the H4K20me2 peptide is depicted in greater detail in the left panel (PDB 2IGO). **E.** The hybrid tandem Tudor domains of JMJD2A associate in a saddle structure. The second Tudor domain is involved in the recognition of the H3K4me3 (PDB 2GFA) and H4K20me3 (PDB 2QQS) peptides and does so through alternative utilization of surface residues, shown in the lower left panel. For the peptide binding figures of 53BP1 and JMJD2A, the Blue numbers denote amino acids comprising the bound peptide and the Black numbering corresponds to the proximal Tudor domain side chains. **F.** The association of three canonical Tudor domains is shown for the Spindlin1 crystal structure (PDB 2NS2). The first Tudor is colored in cyan and the second and third domains in darker and lighter shades, respectively.

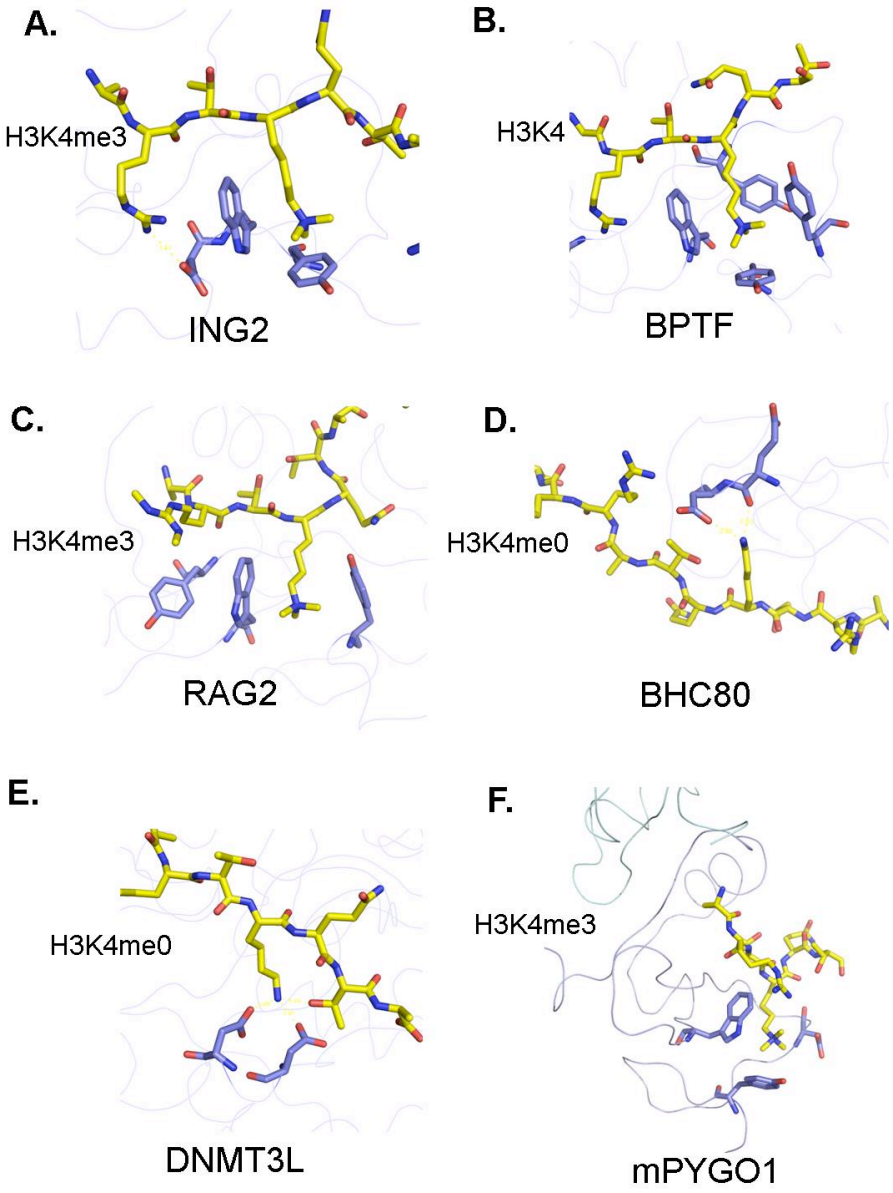
A.
H3K4me3
ING2
B.
H3K4
BPTF
C.
H3K4me3
RAG2
D.
H3K4me0
BHC80
E.
H3K4me0
DNMT3L
F.
H3K4me3
mPYGO1

Figure 1.

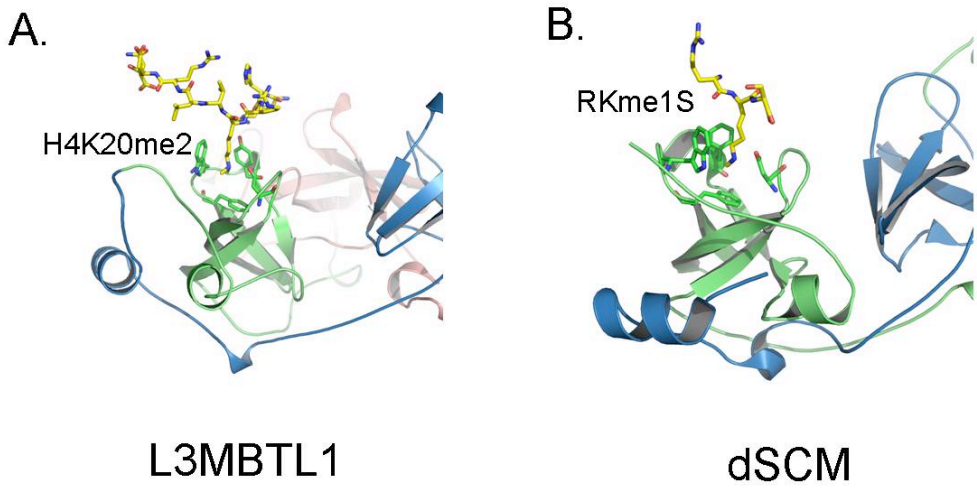

Figure 2.

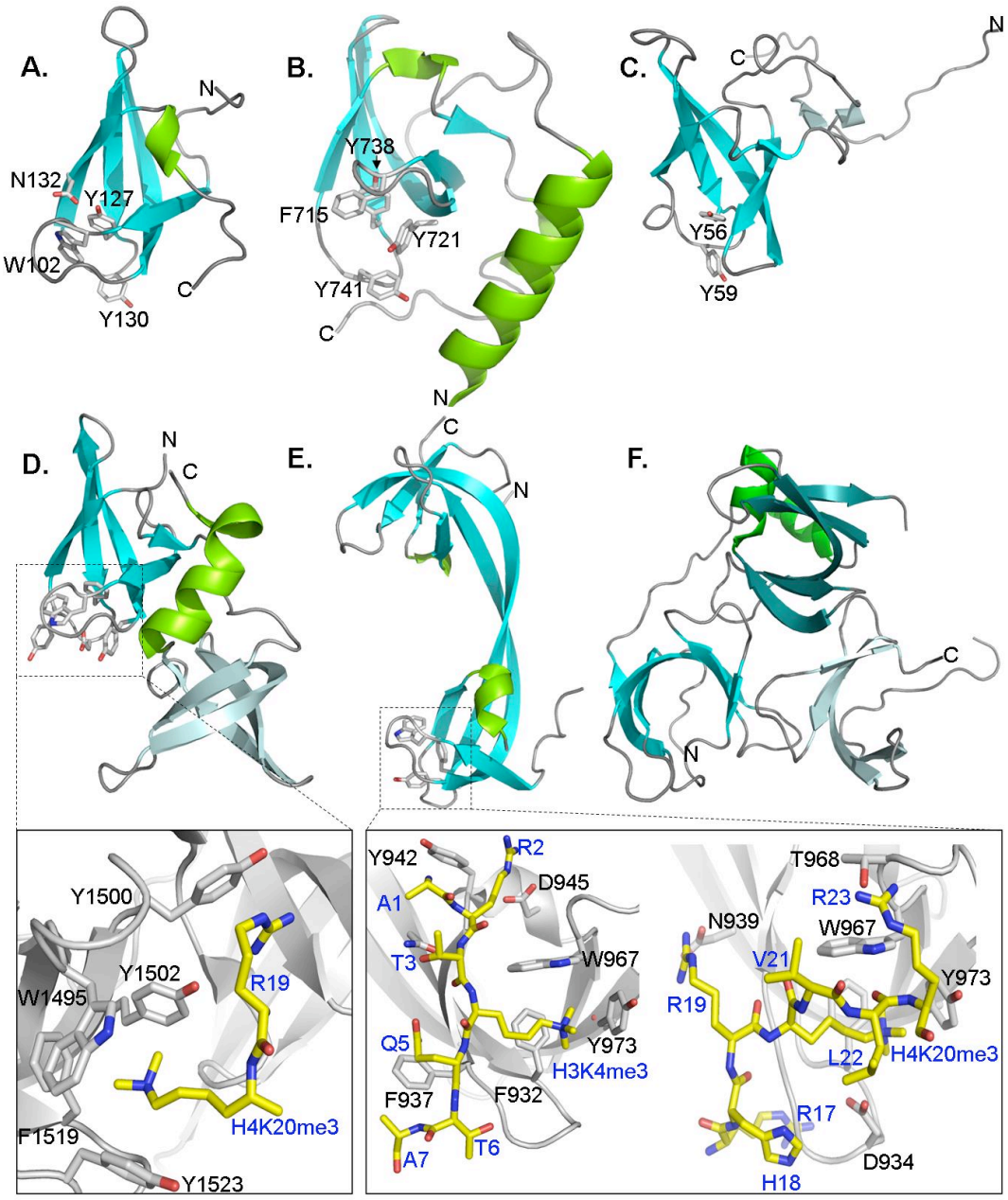
A.
N
N132
Y127
W102
C
Y130
B.
Y738
F715
Y721
Y741
C
N
C.
C
N
Y56
Y59
D.
N
C
E.
C
N
F.
C
N
Y1500
Y1502
W1495
R19
H4K20me3
F1519
Y1523
Y942
R2
D945
A1
T3
W967
Q5
Y973
H3K4me3
F937
F932
A7
T6
T968
R23
N939
W967
V21
Y973
R19
L22
H4K20me3
R17
D934
H18

Figure 3.