Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin.

Sebastian Eustermann[1,3], Ji-Chun Yang[1,3], Martin J. Law[2], Rachel Amos[2], Lynda M. Chapman[1], Clare Jelinska[2], David Garrick[2], David Clynes[2], Richard J Gibbons[2], Daniela Rhodes[1], Douglas R. Higgs[2,4] & David Neuhaus[1,4,5]

[1]M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K. [2]M.R.C. Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, U.K. [3]These authors contributed equally to this work. [4]Senior authorship shared equally. [5]To whom correspondence should be addressed.

Accurate read-out of chromatin modifications is essential for eukaryotic life. Mutations in the X-linked ATRX gene cause a mental retardation syndrome, while wild-type ATRX protein targets pericentric and telomeric heterochromatin for deposition of the histone variant H3.3 *via* a largely unknown mechanism. Here, we show that the ADD domain of ATRX, where most syndrome-causing mutations occur, engages the N-terminal tail of histone H3 through two rigidly oriented binding pockets, one for unmodified Lys4, the other trimethylated Lys9. *In vivo* experiments show this combinatorial readout is required for ATRX localization, with recruitment enhanced by a third interaction *via* heterochromatin protein-1 (HP1) that also recognizes trimethylated Lys9. The cooperation of ATRX ADD domain and HP1 in chromatin recruitment results in a tripartite interaction that may span neighboring nucleosomes and illustrates how the "histone-code" is interpreted by a combination of multivalent effector-chromatin interactions.

The chromatin-associated human protein ATRX was originally identified because mutations in the *ATRX* gene cause a severe form of syndromal X-linked mental retardation called ATR-X syndrome[1]. This syndrome is associated with profound developmental delay, facial dysmorphism, urogenital abnormalities and α-thalassaemia[2]. Somatic mutations are found in the pre-leukemic condition α-thalassaemia myelodysplastic syndrome[3], and more recently ATRX has also been identified as a potential tumor suppressor gene in pancreatic neuroendocrine tumors[4].

ATRX syndrome mutations or knockdown of ATRX expression cause diverse effects including altered patterns of DNA methylation[5], a telomere-dysfunction phenotype[6], aberrant chromosome segregation[7], premature sister chromatid separation[8] and changes in gene expression[1,9]. ATRX localizes predominantly to large tandemly repeated regions (*e.g.* telomeres, centromeres, rDNA) associated with heterochromatin[9-11], and recent evidence shows that it directs H3.3 deposition to pericentric and telomeric heterochromatin[10,12]. In addition, we have shown that ATRX is enriched at chromatin sites that contain tandemly repeated, GC-rich DNA sequences where it may play a role in modifying G-quadruplex DNA structures[3]. However, the mechanism of ATRX recruitment to heterochromatin remains largely unknown.

Targeting of ATRX to heterochromatin requires H3 tails bearing the H3K9me3 mark; when the Suv39H1 and Suv39H2 methyltransferases that add these methyl groups are knocked out, ATRX no longer localizes to pericentric heterochromatin (see reference [13] and our own additional studies shown in Supplementary Figure 1a). This was originally rationalized by proposing an indirect mechanism[11], whereby ATRX interacts with the C-terminal chromo shadow domains of HP1α and HP1β proteins[14,15] that in turn recognize trimethylated H3-Lys9 *via* their N-terminal chromodomains[16]. However, such a mechanism would provide little specificity since HP1 is a widely distributed, constitutive component of heterochromatin that also interacts with many other proteins[14]. We therefore considered whether ATRX might itself interact with histone tails directly. Our previous structural analysis of the ADD domain of ATRX[17] revealed that it contains a PHD zinc-finger domain packed against a GATA-like zinc-finger (Figure 1), a structure that was subsequently also found in the DNMT3 DNA methyltransferases and DNMT3L[18,19]. Given the relevance of the ADD domain for ATRX function, as evidenced by the large number of syndrome causing mutations in this domain (Figure 1), and the known role of PHD zinc-fingers as H3 histone recognition modules, we hypothesized that the ADD domain has a direct function in ATRX chromatin recruitment.

RESULTS:

Both the ADD domain and HP1 contribute to ATRX recruitment

In order to assess the relative contributions of the ADD domain and the HP1 interaction to chromatin recruitment of ATRX, we introduced point mutations in these two regions (singly or in combination) and analyzed the localization of ATRX *in vivo*. To determine the role of the ADD domain, we tested a construct containing a disease causing mutation (C240G) that eliminates a zinc binding cysteine in the

PHD zinc finger and is known to destabilize the structure of the ADD domain[17]. The role of the "Leu-X-Val-X-Leu" HP1 interaction motif[14] was tested using a construct in which the central valine of this motif (V580E) was mutated to ablate the interaction between ATRX and HP1. Transient transfections of L929 mouse mammary fibroblasts with these GFP-fused ATRX mutants were used to quantify pericentric heterochromatin localization by fixed-cell imaging (Figure 1, Supplementary Figures 1b and 1c). Whilst the combination of C240G and V580E mutations abolished ATRX localization to background levels, mutation of the HP1 interaction motif alone only reduced the localization by a factor of 2; importantly, the destabilizing C240G mutation in the ADD domain reduced localization more severely, by a factor of 5 (see Figure 1). These observations demonstrate that the delocalization of ATRX from heterochromatin in the Suv39H1 and 2 double knockout cannot be explained solely by indirect recruitment through the interaction with HP1 as previously suggested[13]. These observations point to a more elaborate network of interactions and require a direct role for the ADD domain in H3 tail recognition.

The ADD domain recognizes H3 Lys4 and H3 Lys9me3 *in vitro*

To determine the specificity of the interaction of the H3 N-terminal tail with the ATRX ADD domain, we carried out *in vitro* binding studies using a small library of modified H3 histone tail peptides in conjunction with ITC and NMR shift perturbation studies (Figure 2e–g, Table 1 and Supplementary Figures 2 and 3). Based on the solution structure of the free ATRX ADD domain, we had previously predicted that its PHD zinc finger would recognize unmodified H3-Lys4[17], as was subsequently shown for DNMT3L and DNMT3A ADD domains[18,19]. Consistent with this, ITC measurements show that the ATRX ADD domain binds unmodified H3 peptide (residues 1–15; H3$_{1-15}$-Lys4me0) with a K_D of 7.93 μM, while trimethylation of Lys4 abolishes this interaction (Figure 2f and Table 1). Various differently modified H3 histone tails bind the ATRX ADD domain with comparable affinities to that for unmodified H3 and caused similar NMR shift perturbation patterns (Table 1 and Supplementary Figure 2). Strikingly, however, we find that trimethylation of H3-Lys9 causes an increase in binding affinity by a factor of thirty; H3$_{1-15}$ trimethylated at Lys9 and unmodified at Lys4 binds the ATRX ADD domain with a K_D of 0.27 μM, and has a binding enthalpy three times larger than that of the corresponding unmodified peptide (Figure 2f; see also Supplementary Figure 3 for thermodynamic parameters). A similar affinity increase is seen for the dimethyl derivative (K_D=0.38 μM), whereas monomethylation had only a moderate effect (K_D=2.90 μM) (Table 1). Since all peptides bind the ATRX ADD domain in a 1:1 stoichiometry, our ITC data demonstrate simultaneous recognition of unmodified Lys4 and di- or trimethylated Lys9.

Structural basis for combinatorial readout by ATRX ADD

To understand the mechanism of this combinatorial readout we determined the solution structure of the ATRX ADD domain in complex with the H3$_{1-15}$ peptide containing unmodified Lys4 and trimethylated Lys9 (Figure 2a–d, Supplementary Figures 4 and 5 and Supplementary Table 1). The structure reveals how a specific combination of the PHD finger and GATA-like finger together provide an extended binding surface that gives rise to the high affinity and specificity observed. Residues 1–9 of the H3

peptide bridge the two zinc-finger domains, whereas residues 11–15 do not bind the domain and are disordered in solution. Each finger of the ADD domain provides a pocket for the recognition of different side-chains of a single H3 histone tail; a cluster of acidic sidechains in the PHD finger interact with the unmodified Lys4 sidechain (Figure 2c), while the trimethylated Lys9 sidechain is inserted into a hydrophobic pocket formed by residues in the GATA-like finger and its linker to the PHD zinc finger (Figure 2d). To our knowledge, the specific interaction of a GATA-like zinc finger with a histone mark is a novel finding; no corresponding pocket for the recognition of trimethylated Lys9 is present in the ADD domains of DNMT3s[18,19] (see Supplementary Figure 6). Overall, this arrangement gives rise to a combinatorial read-out of Lys4 and Lys9 modification status, since the two recognition sites adopt a rigid mutual orientation maintained that is by the extensive interactions of the GATA-like and PHD fingers with one another and with the C-terminal helix. This provides a scaffold that couples the two recognition events, both thermodynamically and kinetically, thus explaining the relatively large affinity increase that we observe when H3 Lys9 is di- or tri-methylated (Figure 2f). Two other cases of combinatorial readout by tandem domains have recently been reported[20,21]; our study establishes the ADD domain as a novel and distinct histone recognition module that recognizes a specific histone modification pattern in a combinatorial manner via two rigidly oriented zinc finger domains.

Dissecting the chromatin interactions of ATRX *in vivo*

Based on the structural information that defines residues important for unmodified Lys4 and trimethylated Lys9 recognition, we next tested the contribution of each binding pocket to ATRX localization *in vivo* (Figure 3a–c). Mutations in the ADD domain that would selectively compromise recognition of either Lys4me0 (E218A) or Lys9me3 (Y203K), but do not reduce the stability of the ADD domain (as shown by ITC, CD and NMR measurements; Table 1 and Supplementary Figures 3, 7 and 8), were engineered and expressed in L929 cells. These mutations each reduced the localization of ATRX to pericentric heterochromatin, demonstrating the cooperative nature of the combinatorial readout *in vivo* (Figure 3). The function of the ADD domain in histone tail recognition may also explain how syndrome-associated mutations in the ADD domain, such as C240G (Figure 1) and Q219P (Figures 1 and 2), contribute to disease, since interfering with the combinatorial histone tail recognition process, whether by altering the interaction interface or by compromising protein stability, is likely to impair appropriate ATRX recruitment.

To understand the interplay between ATRX and HP1 during interphase when both proteins localize to pericentric and telomeric heterochromatin[6,11], we used the structure-guided mutations in the binding pockets of the ADD domain described above (E218A and Y203K) together with mutation of the HP1 interaction motif (V580E) to analyze recruitment of ATRX *in vivo*. Interestingly, the results in Figure 3 show that both the ADD domain (through combinatorial readout of H3 Lys4me0 and Lys9me3) and the interaction of ATRX with HP1 (which itself recognizes H3 Lys9me3), contribute to recruitment. Thus, while individual mutations reduce recruitment substantially, double mutations (e.g. Y203K V580E) that target both recognition modes are required to abolish it to near-background levels (Figure 3a–c). In order to understand this further, we carried out NMR competition experiments. These show

that *in vitro* ATRX and HP1 bind Lys9me3 in a mutually exclusive manner; an excess of HP1 is able to displace ADD from such an H3 peptide (Figure 4). However, the situation *in vivo* is different as heterochromatin provides an array of H3 N-terminal tails enriched with the Lys9me3 histone mark. Such an arrangement of tails would permit the ADD domain and HP1 to bind to adjacent tails and hence cooperate in ATRX recruitment, rather than competing with one another for binding to the same tail. We envisage a tripartite network of ATRX interactions that spans adjacent nucleosomes: the ADD domain defines the specificity of chromatin binding by recognizing both Lys4me0 and Lys9me3 of one H3 tail, while a dimer of HP1[15] could enhance this interaction by recognition of Lys9me3 in two additional, neighboring H3 tails. Alternatively, even if the ADD domain displaces HP1 at an individual H3 tail ATRX may still form a tripartite complex, as HP1 is an abundant and integral component of pericentric heterochromatin and so is likely also to be available for ATRX interaction at adjacent nucleosomal sites. Figure 3d shows a schematic of such a tripartite interaction.

Alterations to the pattern of histone marks *e.g.* during the cell cycle would lead to changes in the repertoire of chromatin associated proteins. For example, it has previously been shown that during mitosis, phosphorylation of H3 Ser10 by Aurora B kinase results in the ejection of HP1 from pericentric heterochromatin[22] while ATRX remains localized[11]. In order to understand how this occurs, we carried out NMR and ITC experiments using a triply modified H3 peptide (Lys9me3, Ser10P, Lys14Ac). While it had previously been shown that this peptide cannot bind HP1[23], we found that binding to the ATRX ADD domain is unaffected by these modifications ($K_D = 0.51$mM, Table 1) and causes an identical pattern of chemical shift perturbations to that seen for addition of the Lys9me3 singly modified peptide (Figure 4). Consistent with this, the competition experiment shown in Figure 4 demonstrates that only the singly modified peptide can be removed from the ADD domain by addition of excess HP1, whereas the triply modified peptide fails to bind HP1 and remains bound to the ADD domain. These results show that ATRX ADD alone can be sufficient for both Lys9me3 recognition and localization, and how ATRX recruitment and its interplay with HP1 can be selectively regulated by a pattern of histone modifications.

DISCUSSION:

Our study demonstrates how ATRX targets heterochromatin *via* a tripartite network of interactions that involves both its ADD domain and its interaction with HP1. The specificity of this network is established by the ADD domain, which, unlike other histone recognition modules, achieves localization through a combinatorial readout of unmodified Lys4 and trimethylated Lys9 on the N-terminal H3 tail. Once having thus acted as a "reader" of these chromatin marks, ATRX may then modulate the way in which epigenetic marks are "written". The presence on chromatin of both unmodified Lys4 and trimethylated Lys9 on H3 tails, as well as that of HP1, correlate with hypermethylation of CpG islands[24], and intriguingly some abnormal patterns of DNA methylation are found in ATRX patient cell lines[5]. Moreover, it was shown that ATRX interacts directly with Daxx[25], which is a histone chaperone specific for the histone variant H3.3[12]. H3.3 deposition is critical in the

de novo formation of pericentric heterochromatin during development[26] and is required for maintaining stability of subtelomeric chromatin that is enriched in tandemly repeated, GC-rich DNA sequences[6,9]. While it is Daxx that binds H3.3, ATRX is required to target this complex to heterochromatic sites for subsequent H3.3 incorporation[10,12]. Significantly, our finding that ATRX recognizes directly the H3 N-terminal tail explains how targeting of the Daxx–H3.3 complex may occur.

Our results provide insights into how mutations in the ATRX ADD domain cause mislocalization of ATRX protein to heterochromatin. While it remains unknown how closely the mechanism of such localization resembles that at other chromatin sites, it seems likely this may in turn contribute to understanding the underlying etiology of ATRX syndrome. The ATRX ADD domain is a hotspot of syndrome-associated mutations[27], and our previous structural analysis showed that most of these are located in the hydrophobic core of the domain and hence compromise its stability[17]. As exemplified by Q219P and C240G in this study, such mutations cause a breakdown of the combinatorial H3 recognition process, either by altering the ADD interaction surface directly or by destabilizing the structure. Given the central importance of the ADD domain in ATRX localization, it is most likely that such a breakdown may also impair the complex downstream functions of ATRX. Future studies aim to further dissect the regulatory function of ATRX once localized at specific chromatin sites.

In general, the collective recognition of a pattern of histone modifications gives rise to a strong and specific interaction with chromatin due to the additive enthalpies of each binding event. Such recognition of multiple histone marks not only imparts specificity but would also permit adaptation to changes in histone modification patterns during the cell cycle. For example, our results suggest that further modifications of the H3 tail (e.g. at Ser10 and Lys14) regulate the interplay between HP1 and ATRX during localization. It seems likely that related changes in such a combinatorial code might correlate with changes in cell fate during differentiation and development. Overall, our study provides a mechanistic understanding of ATRX association with heterochromatin and addresses directly the fundamental question of how the histone code may be interpreted by a combination of multivalent effector-chromatin interactions.

METHODS:

Protein Expression and Peptide Synthesis

Human ATRX ADD domain (residues 159–296) was expressed enriched with ^{13}C and ^{15}N stable isotopes in *E. coli* essentially as described previously[17]. Protein purification was carried out using Ni^{2+} NTA affinity, size exclusion and ion exchange chromatography as described in detail in Supplementary Materials and Methods. Chemically synthesized peptides listed in Table 1 were obtained from Peptide Specialty Laboratories (Heidelberg, Germany), who purified peptides by HPLC and verified chemical identity by Mass Spectrometry. Concentrations of peptide solutions were determined by quantitative amino acid analysis using a Biochrom 20 Amino Acid Analyzer (Pharmacia Biotech).

ITC and Thermal Denaturation studies

ITC experiments employed a MicroCal iTC200 calorimeter. H3-peptide solutions of 1–3mM were titrated into either 50μM or 100μM ATRX ADD protein solutions at 25°C as described in Supplementary Materials and Methods. Data analysis and curve fitting to a one-binding-site model was performed using the Origin 7 software package provided with the Microcal iTC200 calorimeter. The iTC200 was calibrated with EDTA–Ca^{2+} titrations prior to use according to the manufacturer's recommendations. The error in the reported dissociation constants and binding stoichiometries was ±20% (cumulative error of calorimeter measurements, data fitting and determination of concentrations). Equilibrium thermal denaturation experiments were carried out using CD spectroscopy using a Jasco J815 Spectropolarimeter as described in Supplementary Materials and Methods.

NMR titration experiments and structure determination

NMR signal assignments and titration experiments were carried out using ^{13}C ^{15}N isotopically enriched ATRX ADD domain (residues 159–296) and unlabelled peptides as described in Supplementary Materials and Methods. Structures were calculated in the program XPLOR-NIH using a two-stage strategy as described previously[28]. Intermolecular NOE constraints (57 distances) were derived from filtered NOESY experiments, and two intermolecular hydrogen bond restraints were defined as described in Supplementary Materials and Methods. Ensemble and structural statistics of accepted structures are shown in Supplementary Figure 4 and Supplementary Table 1.

ATRX Heterochromatin Localization Experiments

To assess the heterochromatin localization of wild type and mutant pATRX-GFP1 (encoding amino acids 85–1165 of mouse ATRX fused to the C-terminal of EGFP) L929 cells seeded onto coverslips were transfected with the constructs, fixed and then stained with the DNA marker ToPro3 (Molecular probes) and mounted with Vectashield. Mutagenesis was carried out using the QuikChange Site-directed Mutagenesis Kit (Stratagene) according to the manufacturer's protocol. Mutagenic primers are shown in Supplementary Table 2. Three transfections were completed for each construct. Confocal optical sections were collected from twenty cells for each transfection and image analysis was performed as described in Supplementary Materials and Methods. Protein enrichment ratios were calculated by dividing the mean GFP intensity at pericentromeric heterochromatin by the mean GFP nucleoplasmic intensity. Errors are given as 95% confidence limits. Localization of endogenous ATRX in wild type or Suv39h double null mouse embryonic fibroblasts was assessed using indirect immunofluorescence as described in Supplementary Materials and Methods.

Author Contributions:

SE and JCY carried out NMR experiments and structure determination. SE, JCY and DN analyzed the structures. SE and LMC prepared protein and complex samples. SE carried out *in vitro* binding experiments. MJL, RA, DC and DG carried out *in vivo* imaging experiments. SE and CJ carried out *in vitro* experiments with mutants. SE, DR, DN, RJG and DH designed the experiments and wrote the manuscript.

References:

1. Gibbons, R.J., Picketts, D.J., Villard, L. & Higgs, D.R. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* **80**, 837-45 (1995).
2. Gibbons, R.J. & Higgs, D.R. Molecular-clinical spectrum of the ATR-X syndrome. *Am J Med Genet* **97**, 204-12 (2000).
3. Gibbons, R.J. et al. Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS). *Nature Genetics* **34**, 446-449 (2003).
4. Jiao, Y. et al. DAXX/ATRX, MEN1, and mTOR Pathway Genes Are Frequently Altered in Pancreatic Neuroendocrine Tumors. *Science* (2011).
5. Gibbons, R.J. et al. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* **24**, 368-71 (2000).
6. Wong, L.H. et al. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* **20**, 351-60 (2010).
7. Ritchie, K. et al. Loss of ATRX leads to chromosome cohesion and congression defects. *J Cell Biol* **180**, 315-24 (2008).
8. De La Fuente, R., Viveiros, M.M., Wigglesworth, K. & Eppig, J.J. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev Biol* **272**, 1-14 (2004).
9. Law, M.J. et al. ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell* (2010).
10. Goldberg, A.D. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-91 (2010).
11. McDowell, T.L. et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**, 13983-8 (1999).
12. Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C. & Allis, C.D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* **107**, 14075-80 (2010).
13. Kourmouli, N., Sun, Y.M., van der Sar, S., Singh, P.B. & Brown, J.P. Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. *Biochem Biophys Res Commun* **337**, 901-7 (2005).
14. Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G. & Rauscher, F.J., 3rd. The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem Biophys Res Commun* **331**, 929-37 (2005).
15. Thiru, A. et al. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J* **23**, 489-99 (2004).
16. Nielsen, P.R. et al. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103-7 (2002).
17. Argentaro, A. et al. Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* **104**, 11939-44 (2007).
18. Ooi, S.K. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-7 (2007).
19. Otani, J. et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* **10**, 1235-41 (2009).
20. Zeng, L. et al. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-62 (2010).
21. Tsai, W.W. et al. TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**, 927-932 (2010).
22. Fischle, W. et al. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116-22 (2005).
23. Mateescu, B., England, P., Halgand, F., Yaniv, M. & Muchardt, C. Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO Rep* **5**, 490-6 (2004).
24. Ikegami, K., Ohgane, J., Tanaka, S., Yagi, S. & Shiota, K. Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. *International Journal of Developmental Biology* **53**, 203-214 (2009).

25. Xue, Y. et al. The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci U S A* **100**, 10635-40 (2003).

26. Santenard, A. et al. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* **12**, 853-62 (2010).

27. Gibbons, R.J. & Wada, T. in *Molecular Basis of Inborn Errors of Development* (eds. Epstein, C., Erickson, R. & Wynshaw-Boris, A.) 747–757 (Oxford University Press, London, 2004).

28. Muto, Y. et al. The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol* **341**, 185-98 (2004).

29. Canzio, D. et al. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell* **41**, 67-81 (2011).

Figure Captions

Figure 1 The ADD domain of ATRX is involved in localization *in vivo*. **(a)** Missense mutations (small circles) associated with ATRX syndrome cluster in the highly conserved ADD and SWI-SNF ATPase domains of ATRX protein. Residue numbering is shown for human and mouse ATRX; mutants are named using mouse numbering. **(b)** Localization of fusion protein pATRX-GFP1 (here called wildtype, see Supplementary Methods) to pericentric heterochromatin foci as marked by the DNA stain ToPro3. Panels show representative examples of cells transfected with the indicated mutant proteins. **(c)** Quantification of ATRX localization. Each bar represents the ratio of fluorescence intensity at enriched foci relative to background (averaged across 60 cells per construct in three independent transfections; see Supplementary Materials and Methods). Absence of localization results in a homogenous distribution within the nucleus, corresponding to a ratio of 1 (dashed grey line). Wild type protein is highly enriched at foci (ratio 9.6±2.0), while mutations in the ADD domain or HP1 interaction motif show reduced localization (2.3±0.2 for C240G and 5.1±0.9 for V580E respectively). Double mutation delocalizes ATRX to background levels, as for the EGFP control.

Figure 2 Structural basis for recognition of an N-terminal H3 tail by the ADD domain of ATRX. **(a)** Two views of human ATRX ADD domain (surface representation) complexed to Lys9me3 H3 peptide (sticks, orange). The ADD domain structure holds the GATA-like and PHD zinc fingers in a rigid mutual orientation, facilitating combinatorial readout of the H3 histone tail modification status; the PHD finger recognizes unmodified Lys4, while trimethylated Lys9 is inserted in a hydrophobic pocket formed by parts of the GATA-like finger and inter-finger linker. The interface is formed by H3 residues 1–9; residues 11–15 of the peptide (not shown) are disordered in solution. **(c)** Cartoon representation of the same complex. Residues 3–5 of the H3 peptide extend the β-sheet of the PHD finger. Q219P is a syndrome-associated mutation located midway between the Lys4 and Lys9me3 pockets; it abolishes H3 peptide binding *in vitro* and interferes with localization *in vivo*. **(c)** and **(d)** Detailed views of the H3 Lys4 and H3 Lys9me3 recognition pockets. **(e)** Sequence of the N-terminal tail of histone H3, showing posttranslational modifications assessed in this study. **(f)** Isothermal calorimetry titration data. In each case, the indicated H3 peptides (residues 1–15) were added to the wild-type or Q219P mutant ADD domain (100 μM for the upper two curves, 50 μM for the lower two). Raw data are shown on the left, fitted curves on the right (see also Table 1). **(g)** Superimposed ^{15}N-^{1}H HSQC spectra of the free ATRX ADD (grey), ATRX ADD mixed 1:1 with unmodified H3$_{1-15}$ (red) and ATRX ADD mixed 1:1 with H3$_{1-15}$ Lys9me3 (blue). Many perturbations are common for either peptide; however, some are additional for H3$_{1-15}$ Lys9me3, showing that this modification is specifically recognized. The insets show three particularly clear cases. Amide group chemical shift differences [calculated as $((\Delta\delta^{15}N/5)^2 + (\Delta\delta^1H)^2)^{1/2}$] for ATRX ADD for the complex with H3$_{1-15}$ Lys9me3 relative to those for the complex with unmodified H$_{1-15}$, are plotted below as a function of sequence and mapped to the surface of the structure. The strongest perturbations are all in the region of the pocket into which the Lys9me3 sidechain is inserted in the structure of the complex.

Figure 3 Structure-guided mutations reveal tripartite chromatin interaction of ATRX *in vivo*. **(a)** Localization of pATRX-GFP1 fusion protein, comparing wildtype with ADD mutants that target selectively the Lys4 (E218A) and Lys9me3 (Y203K) interactions, or both (Q219P). **(b)** As for (a), but in combination with the V580E mutation. **(c)** Quantification of ATRX localization averaged over 60 cells as described for Figure 1. Each bar represents the ratio of intensity at enriched foci relative to background and absence of localization corresponds to a ratio of 1 (dashed grey line). The impact of these mutations on pericentric enrichment is summarized by the crosses below the histogram, showing which protein interactions are affected. **(d)** The schematic summarizes interactions of ATRX that have been identified and tested *in vitro* and *in vivo:* The ADD domain reads both H3 Lys4me0 and H3 Lys9me3, and in addition an HP1 dimer recognizes H3 Lys9me3. Pericentric heterochromatin provides an array of H3 tails highly enriched in Lys9me3 as well as HP1, permitting an ADD domain and HP1 to bind adjacent tails and hence cooperate in ATRX recruitment. HP1 is drawn bridging nucleosomes as in a recent report[29], although of course different arrangements of the tails may occur. ATRX has recently been shown to be required for localization of Daxx–H3.3 at heterochromatin and to catalyze incorporation of H3.3[10,12].

Figure 4 ADD Lys9me3 recognition provides mechanistic evidence for HP1 independent recruitment of ATRX in mitosis. **(a)** Three superimposed ^{15}N-^{1}H HSQC spectra. Grey: free ATRX ADD. Red: ATRX ADD bound to H3$_{1-15}$ peptide with trimethylated Lys9. Blue: ATRX ADD bound to H3 peptide with trimethylated Lys9, phosphorylated Ser10 and acetylated Lys14. ATRX ADD binds both peptides with similar affinities (see Table 1) and in the same conformation, as shown by the closely corresponding peak positions in the two bound-state ^{15}N-^{1}H HSQC spectra. In combination, these data suggest strongly that ATRX ADD binding is independent of the Ser10 and Lys14 modification status of H3 histone tails. **(b)** Competition experiments with HP1β. Unlabelled HP1β was titrated to the peptide-protein complexes described in panel (a). The panel shows a series of ^{15}N-^{1}H HSQC spectra recorded at the given ADD:peptide:HP1β stoichiometries. The region selected (amide resonance of Gly277) is highlighted in panel a by a black box and reports the peptide-binding status of ADD ATRX; the color code is as for panel (a). The H3 Lys9me3 peptide was competed off its binding site in ATRX ADD upon addition of excess HP1. Phosphorylation of Ser10 and acetylation of Lys14 interfere with HP1 binding as reported previously[23]. Consequently, ATRX ADD remains bound to the triply modified H3 peptide as excess HP1 is added.
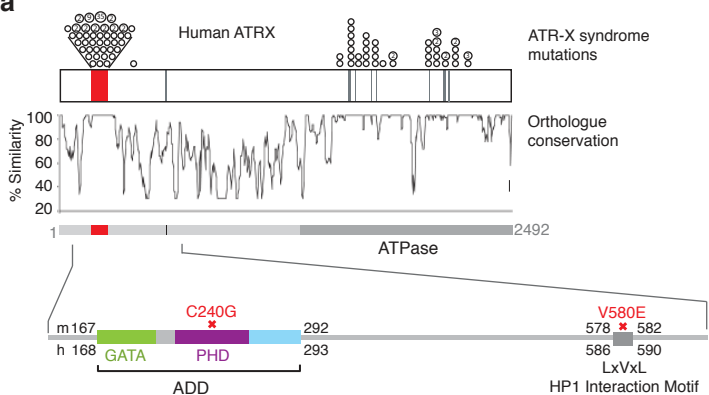
Table 1 Summary of ITC and NMR titration data

Binding Experiment	H3 peptide modification	ITC K_D [μM]*	ITC Binding Stoichiometry	NMR [1]H-[15]N Shift Perturbation**
ADD (wt) + H3 (1–15)				
(1)	none	7.94±0.38	1.01	Lys4
(2)	Lys4(me3)	large	n.d.	minor
(3)	Lys9(me1)	2.90±0.17	0.99	n.d.
(4)	Lys9(me2)	0.38±0.04	1.02	n.d.
(5)	Lys9(me3)	0.27±0.02	0.96	Lys4+Lys9(me3)
(6)	Arg2(me2)	12.5±0.07	0.99	Lys4
(7)	Lys9(me3)Ser10(P)Lys14(Ac)	0.59±0.09	0.91	Lys4+Lys9(me3)
ADD (wt) + H3.3 (1–33)				
(9)	Ser31P	n.d.	n.d.	Lys4
ADD Q219P + H3 (1–15)				
(10)	Lys9me3	large	n.d.	n.d.
ADD Y203K + H3 (1–15)				
(11)	Lys9me3	11.1±0.12	0.89	n.d.
ADD E218A + H3 (1–15)				
(12)	Lys9me3	1.1±0.08	0.99	n.d.

* All K_D values are reported with their associated fitting errors. The experimental error was estimated to be below 20% (cumulative error of calorimeter measurements and determination of protein and peptide concentrations).

** [1]H-[15]N Chemical Shift Perturbation (CSP) categories (see also Figure 2g and Supplementary Figure 2):

Lys4:   Characteristic CSP for ADD signals upon binding of unmodified H3 1–15;

Lys4+Lys9me3:   Characteristic CSP for ADD signals upon binding of H3 1–15 peptide unmodified at Lys4 and trimethylated at Lys9.

a

Human ATRX

ATR-X syndrome mutations

% Similarity

Orthologue conservation

1 — ATPase — 2492

m 167 — C240G — 292
h 168 — GATA PHD — 293
ADD

578 — V580E — 582
586 — 590
LxVxL
HP1 Interaction Motif

b

	WT	V580E	C240G	C240G V580E	EGFP
GFP					
DNA					

c

Pericentromeric GFP Enrichment

WT   V580E   C240G   C240G V580E   GFP

a

Linker

Ser10

Lys9me3

Lys4

PHD
Zinc Finger

GATA-like
Zinc Finger

Helix 4

Ala1

Arg2

Lys4

Thr3

Thr6

Gln5

Lys9me3

Ala7

Arg8

Ser10

b

Gln219

Lys4

Ser10

Ala1

Lys9me3

Zn²⁺

Zn²⁺

Helix 3

Helix 2

Zn²⁺

Helix 1

N

Helix 4

C

c

Asp212

Asp214

Glu218

Asp217

Lys4

d

Ile209

Ala224

Lys9me3

Tyr203

Asp207

e

H3 N-terminal tail

me2 me3 | me1 me2 me3 | P | Ac | P

ARTKQTARKS TGGKAPRKQL ATKAARKSAP STGGVKK | H3

10   20   30

f

ΔP (µcal s⁻¹) vs Time (s)   ΔH (kcal mol⁻¹) vs [Protein]/[Ligand]

ADD (wt) +H3

ADD (wt) +H3 Lys4me3

ADD (wt) +H3 Lys9me3

ADD (Q219P) +H3 Lys9me3

g

ADD

ADD + H3₁₋₁₅

ADD + H3₁₋₁₅ Lys9me3

¹⁵N ppm / ¹H ppm

A Ser205

B Asp206

C Ile209

Δδ vs residues

Lys9me3
recognition

a

b

c

	WT	E218A	Y203K	Q219P	V580E	E218A V580E	Y203K V580E	Q219P V580E
ADD PHD (Lys4me0)		✗		✗		✗		✗
ADD GATA (Lys9me3)		✗	✗				✗	✗
HP1 (Lys9me3)			✗	✗		✗	✗	✗

d

a

b

Supplementary Material

Combinatorial Readout of Histone H3 Modifications Specifies Localization of ATRX to Heterochromatin.

Sebastian Eustermann,[1*] Ji-Chun Yang,[1*] Martin J. Law,[2] Rachel Amos,[2] Lynda M. Chapman,[1] Clare Jelinska,[2] David Garrick,[2] David Clynes,[2] Richard J Gibbons,[2] Daniela Rhodes,[1] Douglas R. Higgs[2†] and David Neuhaus.[1†‡]

[1] M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K.

[2] M.R.C. Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, U.K.

* These authors contributed equally to this work.

† Senior authorship shared equally.

‡ To whom correspondence should be addressed.

Contents:
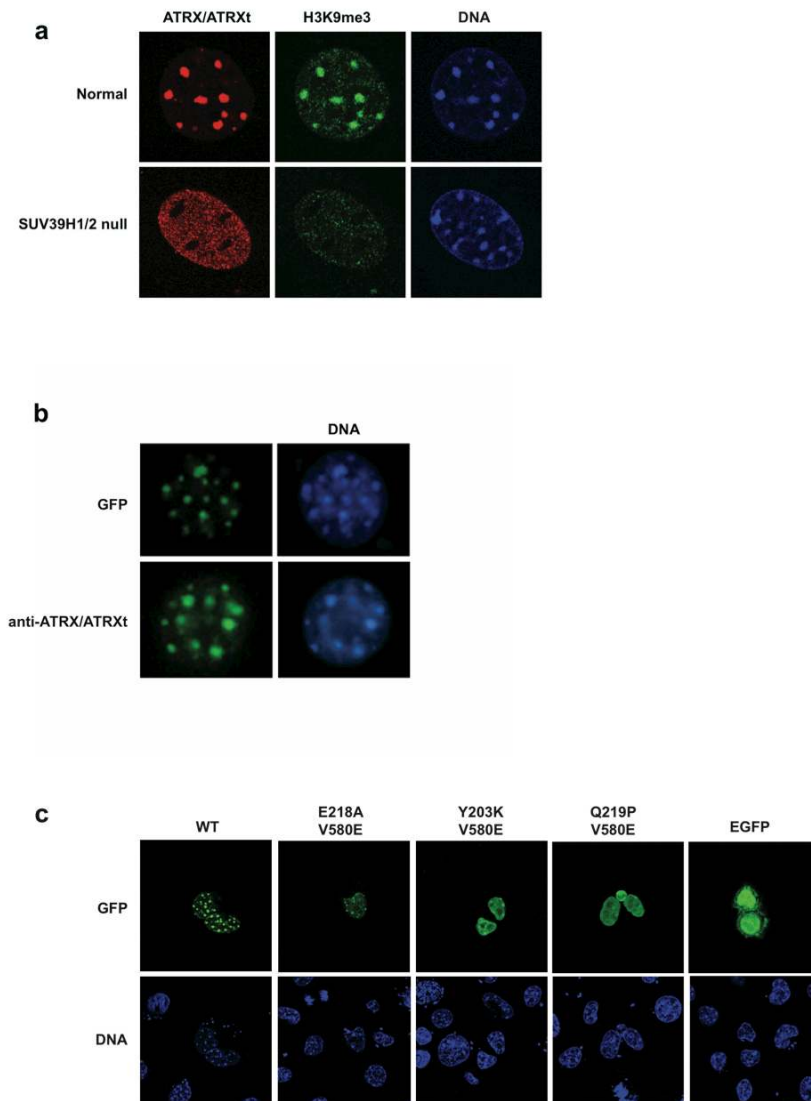
Supplementary Figure 1

Localization of ATRX to pericentric heterochromatin

(a) Demonstrates delocalization of endogenous ATRX (as detected by the anti-ATRX/ATRXt antibody 23c) from pericentric heterochromatin in Suvar39H1/2 double knockout mouse embryonic fibroblasts.

(b) The localization of the pATRX-GFP1 fusion protein (GFP) to pericentric heterochromatin (DNA) is identical to the distribution of endogenous ATRX as detected by the anti-ATRX/ATRXt antibody 23c.

(c) The localization of wildtype (wt) and double mutant pATRX-GFP1, comparing their pericentric localization with the diffuse localization of EGFP protein and absence of GFP signal (autoflourescence) in untransfected cells.

a

ADD
ADD + H3(1-15) unmodified

b

ADD
ADD + H3(1-15) K4me3
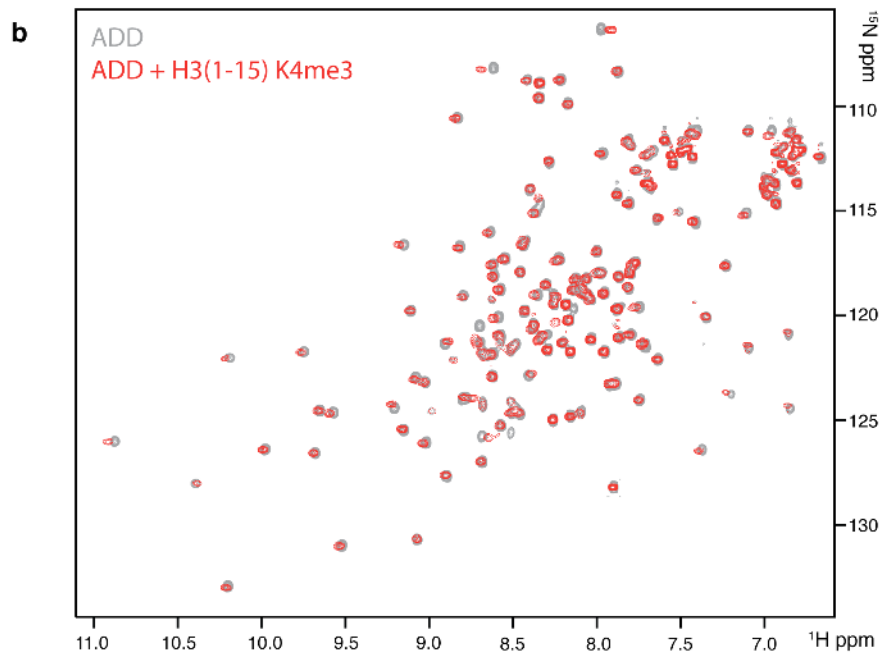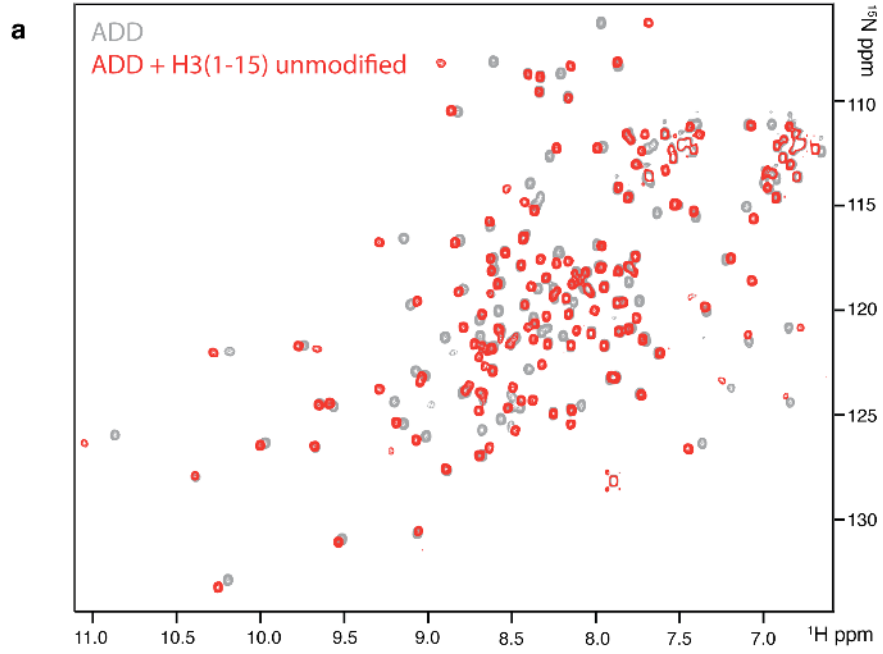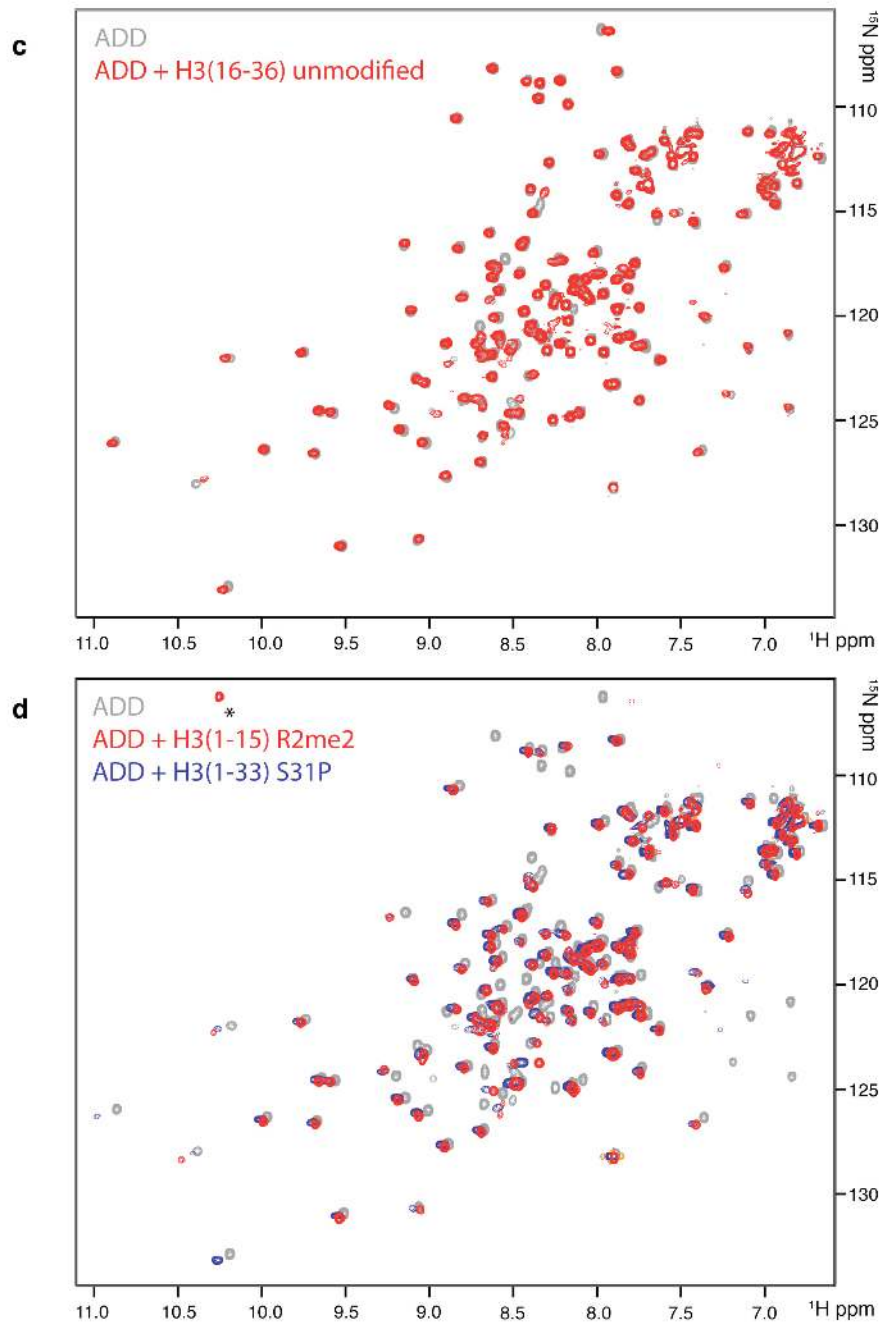
Supplementary Figure 2

^{15}N-^{1}H HSQC titrations showing interactions of other H3 tail peptides with ATRX ADD.

In each case ^{15}N-^{1}H HSQC spectra are shown superimposed for the free ADD domain (grey peaks) and a mixture containing the ADD domain and an added peptide at a 1:1 ratio (red or blue peaks). **(a)** ATRX ADD with unmodified H3$_{1-15}$. Substantial perturbations of the chemical shifts upon peptide addition show clearly that there is an interaction; see main paper Figure 2 for comparison with the case of H3$_{1-15}$ Lys9me3. **(b)** ATRX ADD with H3$_{1-15}$ Lys4me3. Only very minor perturbations indicate a very weak interaction. **(c)** ATRX ADD with H3$_{16-36}$. Essentially no interaction occurs. **(d)** ATRX ADD with H3$_{1-15}$ Arg2me2 (red peaks) or H3$_{1-33}$ Ser31P (blue peaks). Both of these peptides cause essentially identical interactions to those caused by unmodified H3$_{1-15}$ (panel a), showing that neither the Arg2me2 modification nor the Ser31P modification is recognized by ATRX ADD.

a

ADD [100uM]
H3(1-15)

K_D $7.94\pm0.38\times10^{-6}$ M
N 1.01 ± 0.007
ΔH -4930.0 ± 47.5 cal·mol^{-1}
ΔS 6.82 cal·mol^{-1}·K^{-1}
$-T\Delta S$ -1998.3 cal·mol^{-1}
ΔG -6928.3 cal·mol^{-1}

b

ADD [100uM]
H3(1-15) K9me1

K_D $2.90\pm0.17\times10^{-6}$ M
N 0.99 ± 0.005
ΔH -5444.0 ± 75.0 cal·mol^{-1}
ΔS 3.75 cal·mol^{-1}·K^{-1}
$-T\Delta S$ -1098.7 cal·mol^{-1}
ΔG -6542.7 cal·mol^{-1}

c

ADD [50uM]
H3(1-15) K9me2

K_D $0.38\pm0.04\times10^{-6}$ M
N 1.02 ± 0.005
ΔH -11380.0 ± 96.0 cal·mol^{-1}
ΔS -8.80 cal·mol^{-1}·K^{-1}
$-T\Delta S$ 2578.4 cal·mol^{-1}
ΔG -8801.6 cal·mol^{-1}

d

ADD [50uM]
H3(1-15) K9me3

K_D $0.27\pm0.02\times10^{-6}$ M
N 0.95 ± 0.003
ΔH -13160.0 ± 56.1 cal·mol^{-1}
ΔS -14.0 cal·mol^{-1}·K^{-1}
$-T\Delta S$ 4102.0 cal·mol^{-1}
ΔG -9058.0 cal·mol^{-1}

e

ADD [100uM]
H3(1-15) K9me2

K_D $0.33\pm0.06\times10^{-6}$ M
N 1.00 ± 0.007
ΔH -10750.0 ± 200.1 cal·mol^{-1}
ΔS -6.28 cal·mol^{-1}·K^{-1}
$-T\Delta S$ 1840.0 cal·mol^{-1}
ΔG -8910.0 cal·mol^{-1}

f

ADD [100uM]
H3(1-15) K9me3

K_D $0.28\pm0.04\times10^{-6}$ M
N 0.91 ± 0.004
ΔH -13140.0 ± 191.0 cal·mol^{-1}
ΔS -14.0 cal·mol^{-1}·K^{-1}
$-T\Delta S$ 4102 cal·mol^{-1}
ΔG -9038 cal·mol^{-1}

g

ADD [100uM]
H3(1-15) R2me2

K_D $12.39\pm0.07\times10^{-6}$ M
N 0.96 ± 0.012
ΔH -4408.0 ± 73.7 cal·mol^{-1}
ΔS 7.71 cal·mol^{-1}·K^{-1}
$-T\Delta S$ -2259.0 cal·mol^{-1}
ΔG -6667.0 cal·mol^{-1}

h

ADD [50uM]
H3(1-15) K9me3 S10P K14Ac

K_D $0.59\pm0.09\times10^{-6}$ M
N 1.06 ± 0.01
ΔH -14530.0 ± 197.1 cal·mol^{-1}
ΔS -20.2 cal·mol^{-1}·K^{-1}
$-T\Delta S$ 5918.6 cal·mol^{-1}
ΔG -8611.4 cal·mol^{-1}

i

ADD Y203K [50uM]
H3(1-15) K9me3

K_D $11.10\pm0.12\times10^{-6}$ M
N 0.89 ± 0.025
ΔH 4164.0 ± 147.6 cal·mol^{-1}
ΔS 8.71 cal·mol^{-1}·K^{-1}
$-T\Delta S$ -2552.0 cal·mol^{-1}
ΔG -6716.0 cal·mol^{-1}

j

ADD E218A [50uM]
H3(1-15) K9me3

K_D $1.10\pm0.08\times10^{-6}$ M
N 0.99 ± 0.005
ΔH -134000 ± 106.9 cal·mol^{-1}
ΔS -17.7 cal·mol^{-1}·K^{-1}
$-T\Delta S$ -5186.1 cal·mol^{-1}
ΔG -8213.9 cal·mol^{-1}
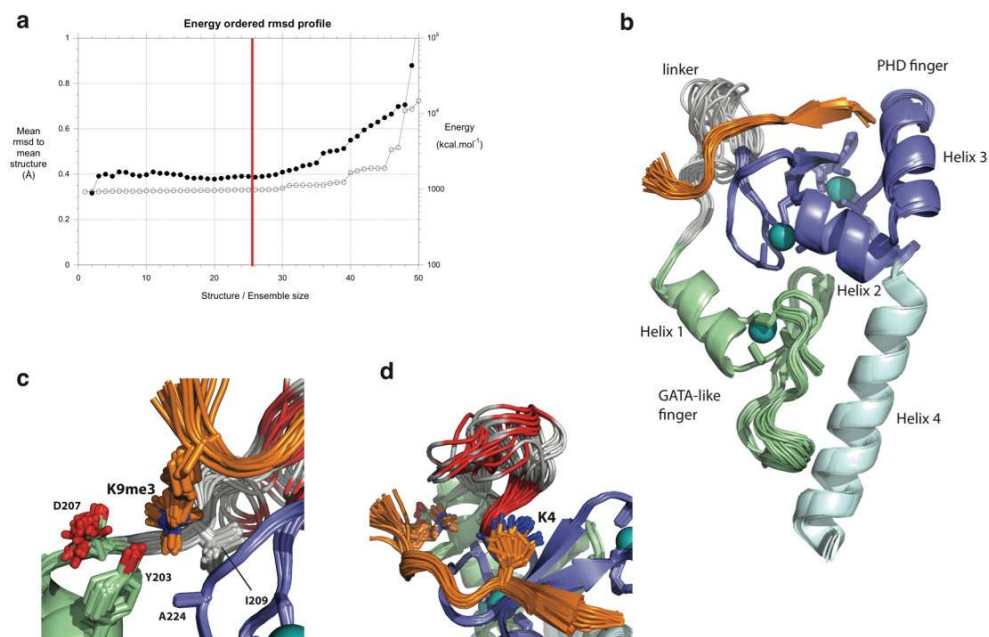
Supplementary Figure 3

Isothermal titration calorimetry data for addition of H3 tail peptides into ATRX ADD.

In each case, raw ITC data (upper panel) was integrated and molar heat of binding per injection (lower panel) was calculated using the respective peptide concentrations as determined by amino acid analysis (see Supplementary Materials and Methods section). Solid grey lines represent best fit of data to a one-binding-site model. The thermodynamic parameters derived for each fit appear below the corresponding panel, and the fitted K_D values and stoichiometries also appear in main paper Table 1.

(a-d) Comparison showing increasing strength of binding as the number of methyl groups on Lys9 of $H3_{1-15}$ is increased: **(a)** unmodified, **(b)** Lys9me1, **(c)** Lys9me2 and **(d)** Lys9me3. Data for Lys9me2 and Lys9me3 were carried out using $50\mu M$ concentration of ATRX ADD **(c and d)** because experiments using $100\mu M$ ATRX ADD **(e and f)** did not characterize the early parts of the curves sufficiently well. The thermodynamic parameters show that the increase of binding affinity upon methylation of Lys9 results from a strong trend in the enthalpy values that overcomes a weaker opposing trend in the entropy terms.

Panel **(g)** shows that peptide $H3_{1-15}$ Arg2me2 binds to ATRX ADD with essentially identical affinity and thermodynamic parameters as the unmodified peptide (panel a), while panel **(h)** shows that the triply modified peptide $H3_{1-15}$ Lys9me3 Ser10P Lys14Ac binds with essentially identical affinity and thermodynamic parameters as the $H3_{1-15}$ Lys9me3 peptide (panel f); these data show that the Arg2me2, Ser10P and Lys14Ac modifications do not affect binding to ATRX ADD.

As expected from the structure, the Tyr203Lys mutation **(i)** affects recognition of the Lys9me3 sidechain, and significantly reduces binding of the $H3_{1-15}$ Lys9me3 peptide relative to binding of the same peptide by wild-type (panel f). The Glu218Ala mutation **(j)** affects recognition of the unmodified Lys4 sidechain and also reduces binding, though to a lesser degree because the Lys4 sidechain probably interacts with more than one nearby acidic residue of ADD ATRX.
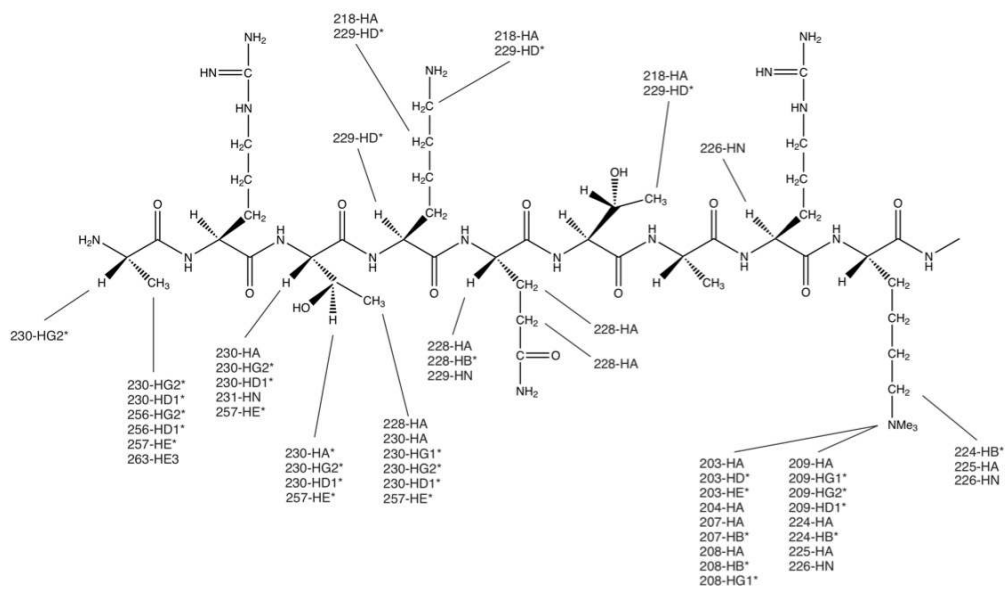
Supplementary Figure 4

NMR Ensemble of structures for the complex of ATRX ADD with H3$_{1-15}$ Lys9me3.

(a) Energy-ordered rmsd profile for the calculated ensemble of structures of the complex of ATRX ADD with H3$_{1-15}$ Lys9me3. Rmsd values (filled circles) are independently calculated using each ensemble size, adding successive structures in order of increasing XPLOR total energy term. Open circles represent the XPLOR total energy terms. Only the 25 structures to the left of the vertical red line were deposited and included when calculating the structural statistics. Rmsd calculations employed the program CLUSTERPOSE[1].

(b) Ensemble view of the structure of the complex, in the same orientation and coloring as Figure 3 of the main paper.

(c-d) Close up ensemble views of the Lys9me3 and Lys4 binding pockets, respectively. In panel d acidic residues are colored in red.

Supplementary Figure 5

Intermolecular NOE contacts in the complex of ATRX ADD with H3$_{1-15}$ Lys9me3.

The intermolecular NOE interactions summarized here were those used when calculating the deposited ensemble of NMR structures of the complex.
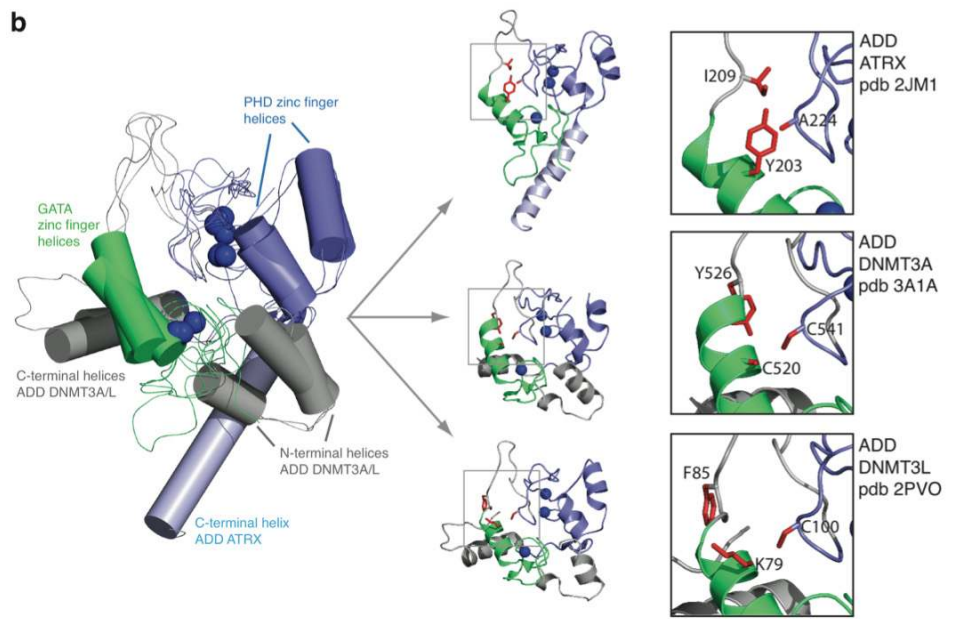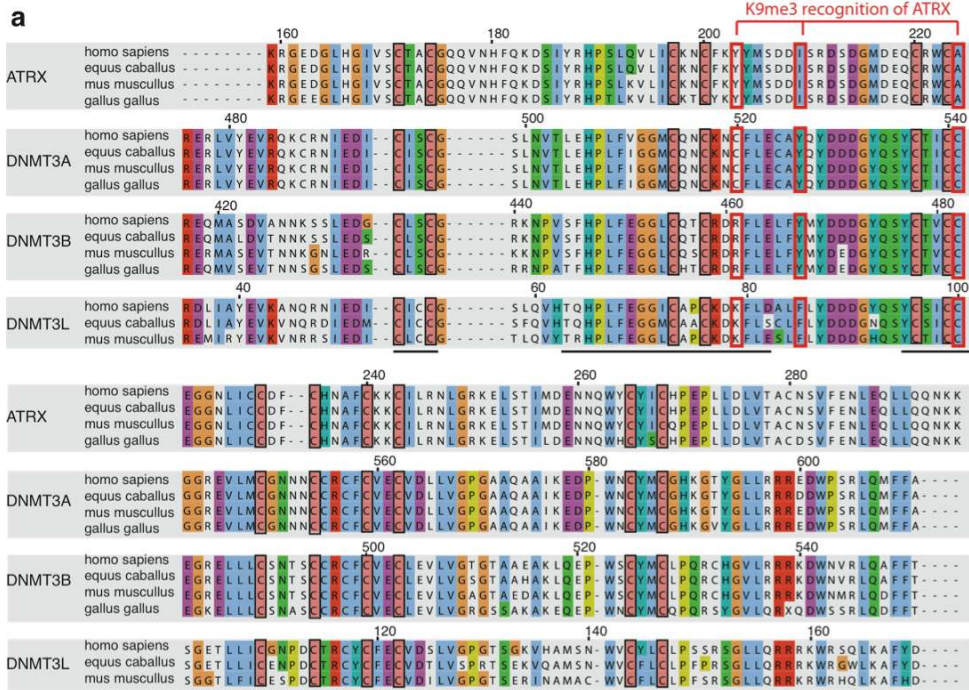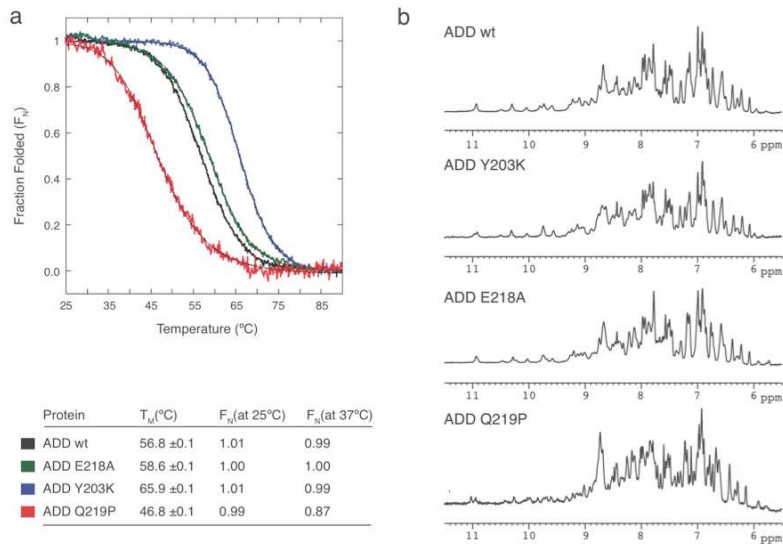
Supplementary Figure 6

Structure-guided sequence alignment for the ADD domains of ATRX, DNMT3A, DNMT3B and DNMT3L.

(a) Structure guided sequence alignment. Note that DNMT3L is not described for *gallus gallus*. (b) Structural elements used for the guided alignment: only the helices of the GATA (green) and PHD (purple) fingers were used; the additional helices shown in grey differ between the ADD domain of ATRX and those of the DNMT3 cases, and consequently were not used for guiding the alignment.

The three expansions show the regions corresponding to the Lys9me3 binding pocket in ATRX. Residues Tyr203, Ile209 and Ala224 form the binding pocket in ATRX (shown using red sticks). DNMT3A and DNMT3L do not recognize H3 K9me3 and residues in corresponding positions are differently conserved. Unlike Tyr203 of ATRX, residues Tyr526 of DNMT3A and Phe85 of DNMT3L both make hydrophobic contacts to the C-terminal helices of their respective proteins. Note that our choice of the Tyr203Lys mutant to interfere with the K9me3 bonding pocket of ATRX was made based on the occurrence of Lys79 at the corresponding position in DNMT3L; this mutant reduces the affinity of the ATRX ADD domain for H3 K9me3 to approximately the level of binding found for the unmodified H3 peptide.

Supplementary Figure 7 (a) Melting curves for ATRX ADD wild-type and E218A, Y203K and Q219P mutants, with table of T_M values and fraction folded at 25°C and 37°C. (b) One-dimensional NMR spectra for ADD wt, ADD Y203K, ADD E218A and ADD Q219P.

Protein	T_M(°C)	F_N(at 25°C)	F_N(at 37°C)
ADD wt	56.8 ±0.1	1.01	0.99
ADD E218A	58.6 ±0.1	1.00	1.00
ADD Y203K	65.9 ±0.1	1.01	0.99
ADD Q219P	46.8 ±0.1	0.99	0.87

Supplementary Figure 7

Melting curves and one-dimensional NMR spectra for ATRX ADD wild-type and E218A, Y203K and Q219P mutants.

To analyze H3 Lys4me0 and H3 Lys9me3 recognition ADD mutants E218A and Y203K were engineered. Note that we have chosen to mutate tyrosine at position 203 of ATRX ADD into a lysine, as this is the conserved residue type at the corresponding position in the ADD of DNMT3L (see Supplementary Figure 6). Both mutations compromise H3 peptide binding (main paper Table 1 and Supplementary Figure 3). The E218A mutant is essentially equally stable as the wt, while the Y203K mutant is stabilized by 9°C. ATRX syndrome-associated Q219P mutant is destabilized by 10°C. In all cases, the 1D NMR spectrum shows that the protein is well-folded at 25°C as expected from the CD analysis and that the structure is essentially preserved.
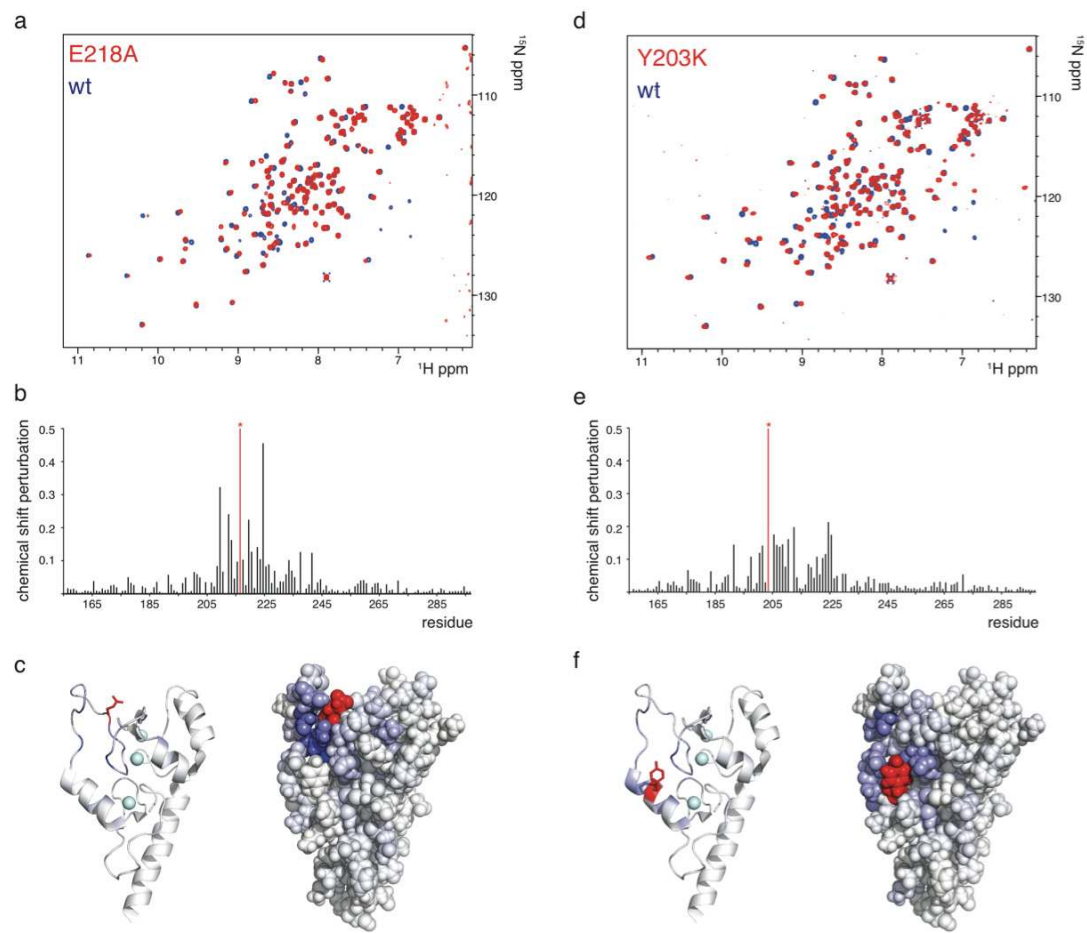
Supplementary Figure 8

(^{15}N, ^{1}H) HSQC spectra of E218A and Y203K mutants.

In each case we show an overlay of the HSQC spectrum of the engineered ADD mutants (see also Supplementary Figure 7) with that of wild-type ADD domain **(a and d)**, a histogram of the chemical shift perturbations by sequence **(b and e)** and a mapping of the perturbations onto the surface of the wild-type ADD domain **(c and f)**.

These data show that for both of these mutants the structure of the wild-type ADD domain is largely preserved, with chemical shift perturbations limited to regions close in space to the mutated residue (Lys4me0 and Lys9me3 recognition pockets; see main text). Assignment of the mutants was made using the principle of minimum shift difference[2]; for the mutated residue in each case the shift perturbation is expected to be large and no assignment was made (red line on histogram).

Supplementary Table 1 Structural statistics of ADD – histone H3$_{1\text{-}15}$ K9me3 complex

Structural restraints			
NOE-derived distance restraints		**ADD**	**H3 tail**
Intraresidue		497	
Sequential		340	6
Medium	(2≤|i-j|≤4)	314	
Long	(|i-j|>4)	371	
Total		1522	
Dihedral restraints			
Chi1		30	
H-bond restraints		36 distances	
		18 H-bonds	
Intermolecular NOE-derived distances		57	
Intermolecular H-bonds (unambiguous)		2 distances	
		1 H-bond	
Intermolecular H-bonds (ambiguous)		2 distances	
		1 H-bond	

Statistics for accepted structures	
Number of accepted structures	25
Mean XPLOR energy terms (kcal mol^{-1} ± S.D.)	
E(total)	957.2 ± 15.8
E(van der Waals)	403.6 ± 6.3
E(distance restraints)	60.9 ± 4.1
E(dihedral restraints)	0.0 ± 0.0
Distance restraint viols. > 0.2Å (average number per structure)	1.4 ± 0.6
Angle restraint viols. > 10° (average number per structure)	9.9 ± 1.2
	(0.24 ± 0.42 excl. Zn)

RMS deviations from the ideal geometry used within XPLOR	
Bond lengths	0.0064 Å
Bond angles	0.84°
Improper angles	0.43°

Ramachandran Statistics	
Most favored	69.5%
Additionally allowed	29.2%
Generously allowed	1.0%
Disallowed	0.3%

Average atomic RMS deviations from the average structure (± S.D.)	Res. 168-209, 218-293
(N, C$^{\alpha}$, C atoms)	0.39 ± 0.09 Å
(All heavy atoms)	0.87 ± 0.09 Å

Supplementary Table 2 Primer sequences used for mutations

Primer name	Sequence 5' – 3'
Y203K	GTAAGAACTGCTTTAAGAAATATATGAG
E218A	CTCAGATGGAATGGATGCACAATGTAGATGGTGTG
Q219P	GATGGAATGGATGAACCATGTAGATGGTGTG
C240G	CATAACGCCTTCGGTAAGAAATGCAT
V580E	AAAGTAAGAAAAGAATTATTTGAGAAACTTACCCCAGTTTCCC

Highlighted in red are the mutated codons, which are numbered according to their position in mouse ATRX.

SUPPLEMENTARY MATERIALS AND METHODS:

Protein Expression and Peptide Synthesis

Human ATRX ADD domain (residues 159-296) was expressed in *E. coli* in a modified pet30A vector as described in reference [3], and contained an N-terminal 6xHis tag, an S-tag and a Tobacco etch virus (TEV) protease cleavage-site. Plasmid DNA was transformed into *E. coli* Rosetta (DE3)pLysS cells (Novagen). Freshly transformed cells were cultured in either LB or M9 medium, each containing 100μM ZnSO$_4$, 30μg/ml Kanamycin and 15μg/ml Chloramphenicol. For stable isotope labeling M9 medium was supplemented with ^{15}NH$_4$Cl (0.5g/L) and [^{13}C$_6$]-glucose (2g/L) (Sigma Aldrich Isotec) as sole nitrogen and carbon sources, respectively. Cells were grown at 25°C to OD 0.6 and protein expressed overnight at 16°C after induction using 200μM IPTG. Protein purification was carried out at 4°C throughout. Harvested cells were resuspended in protein purification buffer A (25mM Tris pH 7.4, 100μM ZnSO$_4$, 1mM DTT, 10% Glycerol, 500mM NaCl and Complete protease inhibitor mix [Roche Complete Protease Inhibitor Cocktail EDTA free; 1 tablet per 50ml]). Following sonication, the lysate was cleared by centrifugation and filtered using a 0.22μM PVDF Stericup filter (Millipore). Initial purification was *via* a 5ml HisTrap HP column (GE Healthcare) and a linear gradient of imidazole in protein purification buffer A. Imidazole and major impurities were removed by gel filtration on a Superdex 200 column 26/60 equilibrated in protein purification buffer A. Digestion with TEV was carried out overnight at room temperature (TEV:protein ratio of 1:50), leaving vector-derived residues Gly-Pro-Gly-Ser N-terminal to the ATRX ADD construct. Uncleaved protein, cleaved N-terminal fragment (His and S-tag) as well as His-tagged TEV protease were removed by repeating the nickel His-tag affinity purification (but using 300mM NaCl) and collecting cleaved ADD in the flow-through. Finally, the protein was passed through a 5ml DEAE anion-exchanger column (GE Healthcare) and purified to homogeneity by gel filtration using a Superdex-S75 column (GE Healthcare) equilibrated with protein purification buffer A. Protein was concentrated and buffer exchanged for subsequent analysis using Vivaspin 20 columns (MWCO 5000Da, Sartorius).

Chemically synthesized peptides listed in main paper Table 1 were obtained from Peptide Specialty Laboratories (Heidelberg, Germany), who purified peptides by HPLC and verified chemical identity by Mass Spectrometry. Concentrations of peptide solutions were determined by quantitative amino acid analysis using a Biochrom 20 Amino Acid Analyzer (Pharmacia Biotech).

Isothermal Titration Calorimetry

ITC experiments employed a MicroCal iTC200 calorimeter. H3-peptide solutions of 1-3mM were titrated into either 50μM or 100μM ATRX ADD protein solutions at 25°C. Typically, one initial injection of 0.5μl was followed by 19 injections of 2μl with 120 seconds between each step. Before each experiment, ATRX ADD protein (residues 159-296) solutions were dialyzed overnight at 4°C into ITC buffer (50mM Tris pH 7.0, 100μM ZnSO$_4$, 2mM DTT and 100mM NaCl) using Slide-A-Lyzer cassettes (Pierce) with a MWCO of 3500Da. Protein concentrations were determined photometrically and adjusted by dilution. Excess heat upon each injection was integrated using a manually adjusted base line and derived values corrected by heat of dilution. The latter was derived either from the

endpoints of peptide-protein titrations under saturating conditions or by conducting separate peptide-buffer titrations.

Data analysis and curve fitting to a one-binding-site model was performed using the Origin 7 software package provided with the Microcal iTC200 calorimeter. The iTC200 was calibrated with EDTA-Ca^{2+} titrations prior to use according to the manufacturer's recommendations. The error in the reported dissociation constants and binding stoichiometries was ±20% (cumulative error of calorimeter measurements, data fitting and determination of concentrations).

ATRX ADD thermal denaturation experiments

Equilibrium thermal denaturation experiments employed circular dichroism spectroscopy using a Jasco J815 Spectropolarimeter. Thermal denaturation was monitored in the far-UV absorption band at 220nm (2nm bandwidth). Wild-type and mutant ATRX ADD (residues 159-296) were adjusted to a buffer containing 40mM pyrophosphate pH 7.0, 0.5mM DTT, 50μM ZnSO$_4$. Experiments were performed using a protein concentration of 10μM in a 1mm pathlength cell. Temperature was increased using a Jasco PFD 425S controller from 2°C to 91°C with slope of 0.2°C/min, measuring the CD signal after every temperature increase of 0.2°C. CD spectra of ATRX ADD before and after thermal denaturation were nearly identical and experiments were repeated at different protein concentrations to exclude protein aggregation artifacts.

NMR experiments

All NMR data was acquired on Bruker Avance 800, DMX600 and DRX500 spectrometers, each equipped with a triple resonance (^{1}H/^{15}N/^{13}C) cryoprobe. All experiments were conducted at 27°C and ^{1}H, ^{15}N and ^{13}C chemical shifts were calibrated using sodium 3,3,3-trimethylsilylpropionate (TSP) as an external ^{1}H reference [4]. For NMR titration experiments, lyophilized H3 peptides were dissolved in NMR buffer containing 50mM [^{2}H$_{11}$] Tris pH 7.0 (Sigma Isotec), 200mM NaCl, 150μM ZnSO$_4$, 1mM [^{2}H$_6$] DTT (Sigma Isotec), 5% ^{2}H$_2$O (v/v) and EDTA free Complete protease inhibitor mix (Roche). 100μM ^{15}N labeled ATRX ADD domain (residues 159-296) was adjusted to the same buffer and peptide added stepwise up to a molar ratio of 1:1. Chemical shift perturbations were followed at each titration point by recording [^{15}N-^{1}H] HSQC spectra at 27°C. In the same way, complexes of ATRX ADD with either Lys9me3 modified H3 peptide (residues 1-15) or Lys9me3 Ser10P and Lys14Ac triple-modified peptide were reconstituted in a 1:1 stoichiometry for HP-1 competition experiments. Subsequently, full length HP-1β was added stepwise as shown in main paper Figure 6 and chemical shift perturbations were followed in [^{15}N-^{1}H] HSQC spectra. Recombinant expressed full length HP-1β was generous gift from Maria Garcia Alai (MRC Laboratory of Molecular Biology, UK) and was expressed and purified as described previously[5].

For NMR structure determination 200μM ^{13}C ^{15}N labeled ATRX ADD domain (residues 159-296) was reconstituted with Lys9me3 H3 peptide (residues 1-15) in a 1:1 stoichiometry as described above for the titration experiments. The buffer contained 50mM [^{2}H$_{11}$] Tris pH 7.0 (Sigma Isotec), 200mM

NaCl, 150μM ZnSO$_4$, 1mM [^{2}H$_6$] DTT (Sigma Isotec), Complete protease inhibitor mix (Roche) and 95% H$_2$O/5% ^{2}H$_2$O (or 99% ^{2}H$_2$O in those cases specified below).

The following spectra were acquired: 2D: [^{15}N-^{1}H] HSQC, [^{13}C-^{1}H] HSQC covering the full ^{13}C spectral width, constant-time [^{13}C-^{1}H] HSQC covering only the aliphatic ^{13}C region, constant-time [^{13}C-^{1}H] HSQC covering only the aromatic ^{13}C region, [^{1}H-^{1}H] NOESY experiments (without heteronuclear filtering; τ_m = 150 ms), [^{1}H-^{1}H] NOESY recorded in 99% ^{2}H$_2$O with filters in F$_1$ and F$_2$ both set to reject signals coupled to ^{13}C, [^{1}H-^{1}H] TOCSY recorded in 99% ^{2}H$_2$O with filter in F$_2$ set to reject signals coupled to ^{13}C, [^{1}H-^{1}H] NOESY with filter in F$_2$ set to reject signals coupled to ^{15}N; 3D data sets: HNCA, HNCOCA, HBHACONH, [^{1}H-^{13}C-^{1}H] HCCH-TOCSY, [^{13}C-^{13}C-^{1}H] HCCH-TOCSY, ^{15}N NOESY-HSQC (τ_m = 150ms), ^{13}C NOESY-HSQC (τ_m = 150ms), separate datasets acquired for ^{13}C aliphatic and aromatic spectral regions. Acquisition of ^{13}C NOESY-HSQC was repeated using a sample that contained 99% ^{2}H$_2$O. Intermolecular NOE constraints were derived from [^{1}H, ^{1}H] NOESY (τ_m = 150ms) and ^{13}C NOESY-HSQC (τ_m = 150ms) with half-filters set to reject NOE crosspeaks between ^{13}C coupled protons and between ^{12}C coupled protons (reference). Intrapeptide NOEs were derived from [^{1}H, ^{1}H] NOESY (τ_m = 150ms) with half-filters set to reject crosspeaks to ^{13}C coupled protons of the ^{13}C labeled protein in t$_1$ and t$_2$. Resonance assignments of ATRX ADD were made by comparing data from the experiments listed above with those from corresponding experiments for the free protein[3]. NMR signals of the H3 peptide in the complex state were assigned using the filtered 2D [^{1}H, ^{1}H] TOCSY and [^{1}H, ^{1}H] NOESY spectra described above.

Structure Calculation

Structures were calculated in the program XPLOR-NIH using a two-stage strategy, introducing explicit zinc atoms and geometrical binding constraints at the start of the second stage once the zinc-binding site was approximately formed. NOE distance restraints were derived from analysis of all of the data from NOE-based experiments. Cross peak intensities were measured using the program SPARKY[6] and grouped into four categories. The strongest d$_{\alpha N}$ (i, i+1) and inter-strand d$_{\alpha \alpha}$ (i, j) connectivities in β-sheet regions were used to set the upper limit for the category "very strong" (0-2.3 Å; 37 intraprotein, 5 intermolecular, 3 intrapeptide), strong d$_{NN}$ (i, i+1) connectivities in α-helices defined the category "strong" (0-2.8 Å; 244 intraprotein, 6 intermolecular, 3 intrapeptide), d$_{\alpha N}$ (i, i+3) cross-peaks in helices defined the category "medium" (0-3.5 Å; 560 intraprotein, 26 intermolecular), and all remaining peaks were classified as weak (0-5 Å; 684 intraprotein, 20 intermolecular). Lower bounds for all NOE restraints were set to zero[7], and no multiplicity corrections were required since r^{-6} summation was used for restraints involving groups of equivalent or non-stereoassigned spins[8,9]. Stereospecific assignments and hydrogen bond restraints (all within regular secondary structural elements) were taken from the assignment for the free protein[3]. In addition, two intermolecular hydrogen bond restraints were defined, one from the amide proton of Leu229 to the CO of Lys4 of the peptide, and one ambiguous constraint for the amide proton of residue Gly226 to the CO of either Ala7, Arg8 or Lys9 of the peptide; in both cases, the relevant protein amide signal moves strongly downfield upon complexation ($\Delta\delta$ ^{1}H Leu229 = +0.69ppm, $\Delta\delta$ ^{15}N Leu229 = +3.1ppm; $\Delta\delta$ ^{1}H Gly226 = +1.26ppm, $\Delta\delta$ ^{15}N Gly226 = +3.6ppm).

Structures were calculated starting from polypeptide chains with randomized ϕ and ψ torsion angles, and with a 30Å separation introduced between the centers of masses of the protein and peptide chains. Parameter and topology files for the trimethyllysine residue were obtained by adding the necessary entries for the methyl groups to a copy of the standard entries for lysine taken from the files topallhdg.pro and parallhdg.pro in the XPLOR-NIH distribution. A two-stage simulated annealing protocol was used within the program XPLOR, essentially as described elsewhere[10,11] but employing larger numbers of cycles as follows: First stage calculations comprised Powell energy minimization (500 steps), dynamics at 1000K (25000 steps), increase of the van der Waals force constant and tilting of the NOE potential function asymptote (4000 steps), switching to a square-well NOE function then cooling to 300K in 2000 step cycles, and final Powell minimization (1000 steps). Second stage calculations used Powell minimization (500 steps), increasing dihedral force constant during 4000 step cycles of dynamics at 1000K (with a strong van der Waals force constant and square-well NOE potential function), cooling to 300K in 1000 step cycles, and 2000 steps of final Powell minimization.

Suv39H1/2 double knockout cells

Wild type or Suv39h double null mouse embryonic fibroblasts immortalized with a 3T3 protocol have been previously described[12]. Cells grown on glass coverslips were fixed in 3% formaldehyde and permeabilized in 0.1% Triton X-100. Indirect immunofluorescence was carried out using anti-ATRX mouse monoclonal antibodies (23C), anti-HP1β rat monoclonal antibodies (MAC353, Biocarta), with appropriate secondary antibodies. Coverslips were mounted in Vectashield (Vector Laboratories) containing 0.25 mg/ml 4',6-diamidino-2-phenylindole (DAPI; Sigma).

Site directed mutagenesis

Mutagenesis was carried out using the QuikChange Site-directed Mutagenesis Kit (Stratagene) according to the manufacturer's protocol. Mutagenic primers are shown in Supplementary Table 2. DNA template pATRX-GFP1 (which encodes amino acids 85-1165 of mouse ATRX fused to the C-terminal of EGFP) was used to introduce mutations. The amplification product was incubated with *Dpn*I (New England Biolabs) to digest the template DNA prior to subcloning.

ATRX heterochromatin localization experiments

To assess the localization of the ATRX-GFP fusion constructs to pericentromeric heterochromatin, L929 cells seeded onto coverslips were transfected with the constructs, fixed and then stained with the DNA marker ToPro3 (Molecular probes) and mounted with Vectashield. Three transfections were completed for each construct. Confocal optical sections were collected, from twenty cells for each transfection using a Radiance 2000 (Biorad) confocal device, an Olympus BX51 microscope with a 100x /1.4 PlanApo oil immersion objective lens (Olympus) and Laser Sharp software (Biorad). Image analysis was performed with Photoshop (Adobe). To measure protein enrichment at pericentromeric heterochromatin, a single random pericentromeric focus of blue fluorescent ToPro3 staining was circled using the freehand lasso tool. The histogram function was then used to measure the mean intensity in the green channel, giving a value for GFP intensity, and therefore protein enrichment, at the

pericentromeric focus. This procedure was repeated in the nucleoplasm; the freehand lasso tool was used to select an area of nucleoplasm without any heterochromatic spots and the histogram function used to measure the mean intensity of the green channel in this area. Protein enrichment was calculated by dividing the mean GFP intensity at pericentromeric heterochromatin by the mean GFP nucleoplasmic intensity, and error bars in the figures correspond to ±95% confidence limits.

Supplementary References:

1.   Diamond, R. Coordinate-based cluster analysis. *Acta Cryst. D* **51**, 127-135 (1995).
2.   Farmer, B.T., 2nd et al. Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat Struct Biol* **3**, 995-7 (1996).
3.   Argentaro, A. et al. Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* **104**, 11939-44 (2007).
4.   Wishart, D.S., Bigam, C. G., Yao, J., Abildaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., Sykes, B. D. ^{1}H, ^{15}N and ^{13}C chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135-140 (1995).
5.   Nguyen, D.P., Garcia Alai, M.M., Kapadnis, P.B., Neumann, H. & Chin, J.W. Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *J Am Chem Soc* **131**, 14194-5 (2009).
6.   Goddard, T.D., Kneller, D. G. SPARKY 3. (University of California, San Francisco).
7.   Hommel, U., Harvey, T.S., Driscoll, P.C. & Campbell, I.D. Human epidermal growth factor: high resolution solution structure and comparison with human transforming growth factor α. *Journal of Molecular Biology* **227**, 271-282 (1992).
8.   Nilges, M. A calculation strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins: Struct. Funct. Genet.* **17**, 297-309 (1993).
9.   Fletcher, C.M., Jones, D. N. M., Diamond, R., Neuhaus, D. Treatment of NOE constrains involving equivalent or nonstereoassigned protons in calculations of biomolecular structures. *J. Biomol. NMR* **8**, 292-310 (1996).
10.   Dutnall, R.N., Neuhaus, D. & Rhodes, D. The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold. *Structure* **4**, 599-611 (1996).
11.   Muto, Y. et al. The structure and biochemical properties of the human spliceosomal protein U1C. *Journal of Molecular Biology* **341**, 185-198 (2004).
12.   Peters, A.H. et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-37 (2001).