

Published in final edited form as:

Trends Biochem Sci. 2013 April ; 38(4): 177–183. doi:10.1016/j.tibs.2012.12.005.

A Common Mode of Recognition for Methylated CpG

Yiwei Liu¹, Xing Zhang¹, Robert M. Blumenthal², and Xiaodong Cheng¹

Yiwei Liu: yliu40@emory.edu; Xing Zhang: xzhan02@emory.edu; Robert M. Blumenthal: Robert.Blumenthal@utoledo.edu; Xiaodong Cheng: xcheng@emory.edu

¹Departments of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

²Department of Medical Microbiology & Immunology and Program in Bioinformatics & Proteomics/Genomics, The University of Toledo College of Medicine & Life Sciences, Toledo, Ohio 43614-2598

Abstract

Much is known about vertebrate DNA methylation, however it is not known how methylated CpG within particular sequences is recognized. Two recent structures of C2H2 zinc finger (ZnF) proteins, in complex with methylated DNA, reveal a common recognition mode for 5-methylcytosine (5mC) that involves a 5mC-Arg-G triad. In the two ZnF proteins, an arginine that precedes the first Zn-binding histidine (RH motif) can interact with 5mCpG or TpG dinucleotide. Among a family of >300 human KRAB (Krüppel-associated box) domain-containing ZnF proteins examined, two-thirds contain at least one ZnF that includes an RH motif. We propose that the RH-ZnF motifs provide specificity for 5mCpG, while the neighboring ZnF fingers recognize the surrounding DNA sequence context.

Three methylated DNA binding domains

The control of gene expression in mammals relies significantly on the modification status of DNA cytosine residues. DNA cytosine modification is a dynamic process catalyzed by specific DNA methyltransferases that convert cytosine (C) to 5-methylcytosine (abbreviated 5mC or M) [1,2]. The 5mC may then be oxidized or demethylated via either an enzyme-catalyzed active pathway [3–5] or a DNA replication-dependent passive process [6,7]. The best-known modified DNA-recognition domains are two that recognize methylated cytosine: methyl-binding domains (MBDs) recognize fully-methylated CpG dinucleotides [8,9], and “SET and RING finger-associated” (SRA) domains bind hemi-methylated CpG sites generated transiently by DNA replication [10,11]. Both MBD and SRA domains have been structurally characterized in complexes with 5mC [12–18]. A third class of mammalian proteins that recognize methylated DNA is the C2H2 zinc finger (ZnF) proteins, which preferentially bind to methylated CpG within a longer specific sequence [19]. Recently, ZnF DNA binding domains from two proteins, Kaiso and Zfp57, have been structurally analyzed in complex with their respective methylated DNA elements [20,21], allowing comparison to the other 5mCpG-binding proteins.

© 2012 Elsevier Ltd. All rights reserved.

Correspondence should be addressed to: Xiaodong Cheng (xcheng@emory.edu; phone: 404-727-8491; fax: 404-727-3746).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Kaiso is a ZnF protein of the BTB/POZ family (Broad-complex, Tramtrack, and Bric-a-brac, also known as POZ for “poxvirus and zinc finger”) [22], whereas Zfp57 belongs to the family of KRAB (Krüppel-associated box) domain-containing ZnF transcription factors [23,24]. Among the C2H2 ZnF proteins, KRAB ZnF transcription factors (KRAB-ZnFs) act mostly as chromatin-modulating transcription repressors [25]. This family has expanded greatly during vertebrate evolution: humans have about twice as many KRAB-ZnF genes as mice do [26]. There are many reviews on the variation of ZnF gene regulation between primate species [24,27–29], and on the general ZnF architecture associated with DNA binding [30]. Here we survey the two most recent structural and biochemical studies of ZnF proteins, and discuss the broader implications for recognition of DNA methylation.

Both MBD and the ZnF DNA-binding domains occupy the DNA major groove, which is generally in an ordered B-DNA conformation, and recognize specific bases through classical hydrogen bonds and specific interactions unique to each protein [12–14,20,21]. By contrast, the mammalian SRA domain flips 5mC from a hemi-methylated CpG out of the double helix and into a hydrophobic binding cage [15–18]. We do not consider the SRA domain in our analysis, due to its vastly different DNA recognition mode from MBD and ZnF. Here, we compare the structural modes of 5mC recognition by three MBD proteins (MeCP2, MBD1 and MBD2) and two ZnF proteins (Kaiso and Zfp57). While each protein has some idiosyncrasies in its interaction with DNA, we discovered that the ZnF and MBD proteins each form a conserved triad (5mC-Arg-G) for recognizing 5mC, despite their unrelated amino acid sequences and structures. Guided by the common structures of Kaiso and Zfp57, we find a conserved sequence motif, RH (an arginine prior to the first zinc-binding histidine), that may be responsible for recognition of methyl-CpG dinucleotides.

A conserved methyl-Arg-G interaction for 5mCpG and TpG dinucleotides

Zfp57 recognizes the sequence TGCMGCR (M = 5mC, R = A or G) [31]. Structural analysis of the complex between fully-methylated DNA and the DNA-binding domain of mouse Zfp57, which has two ZnF motifs in tandem, revealed that the 5mCs of the two DNA strands engage in very different interactions. In the top strand a layer of ordered water molecules enclose the 5mC methyl group (Figure 1a). By contrast, the methyl group of the bottom strand 5mC makes van der Waals contacts with the guanidino moiety of Arg178, which in turn makes bifurcated hydrogen-bonding interactions (via the O6 and N7 atoms) with the 3' guanine (Figure 1b). We refer to this interaction as the 5mC-Arg-G triad.

Kaiso recognizes either two consecutive methylated CpG dinucleotides [22] or a specific intrinsically-methylated DNA element TCCTGCNA (T contains a methyl group at C5) [32]. Structures of Kaiso's three-ZnF DNA binding domain, in complex with its methylated CGCG or TpG-containing cognate sequences, have recently been analyzed by X-ray crystallography and NMR spectrometry [20]. Strikingly, Arg511 of Kaiso interacts with the methyl-CpG and TpG dinucleotides in a similar fashion to Arg178 of Zfp57, forming a methyl-Arg-G triad [20] (Figures 1c–d). The triad is maintained for TpG because the thymine methyl group is in the equivalent position (5-carbon) to that in 5mC.

Moving from ZnFs to the unrelated MBD domains, these bind the symmetrical, fully-methylated CpG site using two arginines, each of which hydrogen-bonds one guanine in a bifurcated pattern (Figures 1e–f). Like Arg178 of Zfp57 and Arg511 of Kaiso, the two MBD arginines are also engaged in van der Waals contacts with the methyl group of the neighboring 5' 5mC of the same DNA strand. Interestingly, Zfp57 and Kaiso each form one 5mC-Arg-G triad to bind the fully-methylated DNA in an asymmetric way, while the MBD domain forms two 5mC-Arg-G triads to symmetrically bind the palindromic fully-methylated CpG duplex (Figures 1e–f). For each triad in MBD1, the calculated binding

energy difference between methylated and non-methylated CpG is approximately -0.8 ± 0.4 kcal/mol [33]. As an aside, the report by Zou *et al.* [33] on the molecular dynamics simulations of MBD-DNA complexes also suggested that cation- π interactions play an important role in the binding, with Arg providing the cationic moiety and 5mC the π electrons; however the high-resolution structure of Zfp57-DNA complex indicates that cation- π interactions are, at the least, not universal in 5mC-Arg-G triads.

In addition, both families can use water-mediated bonding: Zfp57 (though not Kaiso) uses a layer of hydration for one of the two methyl groups, whereas MBD uses a tyrosine to mediate water-associated interaction with one of the two methyl groups that further enhances the methyl DNA binding [12,13] (compare Figure 1, panels a and e). In the latter case, the hydroxyl oxygen atom of the tyrosine mimics a water molecule in the network surrounding the methyl group. In summary, Zfp57, Kaiso and MBD proteins appear to use a conserved mechanism (5mC-Arg-G triad) to recognize 5mC, despite their unrelated protein sequences and structures.

That Kaiso uses the same mechanism to recognize both methyl-CpG and TpG dinucleotides implies that Kaiso does not distinguish T:A from 5mC:G at the corresponding base pair (Figure 1c, d). The sequence element TGCM \underline{M} GCR, which is recognized by Zfp57, further illustrates this 5mC/T equivalence. Specifically, this sequence contains a TpG dinucleotide at base pair positions 1 and 2 in addition to the MpG and, interestingly, Arg157 forms a thymine methyl-Arg-G triad at that location (Figure 1g). Furthermore, a third methyl-Arg-G triad can be involved in Zfp57-DNA binding: the seventh position of the recognized sequence can be either an A or G (or less often C) [31], and an A at that position would generate a TpG dinucleotide on the opposite strand. It happens that an A was used at position 7 in a structural study [21] and, strikingly, an identical thymine-Arg-G triad was found for the TpG on the bottom strand (Figure 1h). We next sought to assess the relatedness of Kaiso and Zfp57 use of the 5mC/T-Arg-G triad by determining how well their triad structures could be superimposed.

Structural comparison of ZnF DNA binding domains of Kaiso and Zfp57

The DNA sequences recognized by Zfp57 and Kaiso have some similar elements, facilitating structural comparison of their complexes. The Kaiso recognition sequence TGCC resembles the first half of the Zfp57 recognition sequence (TGCM of the top strand), and MGMG, the alternative Kaiso recognition sequence, resembles the second half of the Zfp57 cognate sequence (TGMG of the bottom strand) (Figure 2a). Thus, the DNA components of the Kaiso-DNA complex structures were superimposed onto the first- and second-half sequence elements recognized by Zfp57. Only the superimposition of the second-half 5'-TGMG-3' resulted in alignment of ZnFs, with the second ZnF of Zfp57 superimposed on the first ZnF of Kaiso (Figure 2a), but this was sufficient to allow comparison of the DNA-methyl recognition.

As in previously-characterized ZnF structures [30], the base contacts are made by the side chains in the N-terminal portion of the helix, particularly those prior to the first zinc-binding histidine (Figure 2a). The amino acids at positions -1 to -8 (relative to the first zinc-binding histidine) are well positioned to make base contacts. Interestingly, pairwise comparison of the aligned zinc fingers (the second ZnF of Zfp57 bound to TpG and the first ZnF of Kaiso bound to MpG; Figure 2b) revealed only a single conserved arginine at the -1 position (RH). The RH dipeptide is also conserved in the first ZnF of Zfp57, which recognizes a TpG dinucleotide on the opposite strand (Figure 2a). Bidentate arginine recognition of the guanine in a TpG dinucleotide has been noted previously in the yeast transcription factor Ndt80 [34,35]. Thus the functionally and structurally conserved methyl-Arg-G triads are

associated with a sequence motif (RH) that, while small, allows a bioinformatic search of ZnF motifs in specific protein families for candidate methyl-binding regions.

Predicting RH-ZnF proteins that recognize methyl-CpG dinucleotides

The limited sequence conservation between Kaiso and Zfp57 makes the prediction of methyl-CpG binding ZnF proteins more challenging. However, all structurally characterized methyl-CpG binding proteins (except the base-flipping SRA domain proteins) involve a 5mC-Arg-G triad (Figure 1), and the zinc-binding C2H2 residues can be located accurately, therefore ZnF proteins could be surveyed for those that contain the RH dipeptide at the appropriate position.

KRAB-ZnF proteins were the target of this analysis for the following reasons: first, humans have about twice as many KRAB-ZnF genes as mice [26], so these transcription factors seem likely to play significant roles in human biology. (This twofold human/mouse difference begs for an explanation, but none is currently in the literature and the topic is beyond the scope of this analysis.) Second, a substantial number of these genes are preferentially expressed in brain and embryonic cells, where epigenetic reprogramming of DNA methylation takes place [36,37]. Third, KRAB-ZnF proteins are in part responsible for the establishment during early embryogenesis of site-specific DNA methylation patterns that are maintained through development [38].

The C2H2 Zinc Finger Gene database (SysZNF) was searched for ZnFs containing RH motifs at the proper position [39]. Among the 330 human KRAB-ZnF proteins examined, 224 (about 2/3) of them have at least one RH-ZnF (Figure 3a, left panel). These KRAB-ZnF proteins contain tandem ZnFs ranging from 3 to 35 fingers (ZNF91 contains 34–35 fingers [40]), with a mode of around 11–13 fingers (Figure 3a, right panel). The same trend is observed for mouse KRAB-ZnF proteins (Figure 3b).

A few examples of RH-containing mammalian KRAB-ZnF proteins with known biological roles are shown (Figure 3c). ZFP57 mutations - either deletions or missense mutations - have been found in patients with transient neonatal diabetes [23]. Two missense mutations, which change base-interacting R157 to H and zinc ligand H186 to N, each abolishes DNA-binding activity [21]. The absence of Zfp57 at its binding site might provide an opportunity for other proteins to compete and further modify 5mC (such as Tet-catalyzed oxidation, see below), resulting in the hypomethylation phenomenon associated with these human ZFP57 mutations [23]. Like Zfp57 [31,41], ZNF274 recruits the histone H3 lysine 9 methyltransferase SETDB1 (SET domain, bifurcated 1) [42] via the co-repressor TRIM28 (tripartite motif-containing 28; also known as KAP1 for Krüppel-associated protein 1), an essential regulator of genomic imprinting [43]. PRDM9 (PR domain zinc finger protein 9), a major determinant of meiotic recombination hotspots, contains a SET domain responsible for histone H3 lysine 4 methyltransferase activity [44]. These examples further illustrate the coordinated chromatin controls involving DNA methylation and the lysine methylation status of histone H3 (at residues 4 and 9) [45]. Zfp568 (also known as Chato) regulates extraembryonic tissue morphogenesis [46]. Zfp809 restricts retroviral transposition in embryonic stem cells [47], and retroviral silencing has been suggested to be the ancestral role of KRAB-ZnFs [48]. The 18-base pair DNA sequence recognized by Zfp809 contains two CpG sites, and the single base pair change of C-to-G at the corresponding cytosine residue disrupted a reconstituted protein-DNA complex [47]. Regulator of sex limitation (Rsl1) regulates sex- and tissue-specific promoter methylation [49]. In summary, where the functions of the KRAB-ZnFs containing appropriate RH motifs are known, they are consistent with a role that is modulated by DNA methylation. This makes it more likely that many of the numerous RH-containing KRAB proteins of unknown function are also

sensitive to CpG methylation (or specific for TpG dinucleotides). Furthermore, some of these RH-containing KRAB proteins may respond to the oxidative derivatives of 5mC.

Can KRAB-ZnF proteins recognize the oxidative products of 5mC?

The ten–eleven translocation (Tet) proteins can oxidize 5mC to 5-hydroxymethylcytosine (5hmC) [3], and then further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [4,5]. Both 5mC and 5hmC occur in almost all human tissues and cell types examined [50], but 5hmC is relatively enriched in embryonic stem cells [3] and brain [50–52]. The genomic levels of 5fC and 5caC may also be similarly enriched, as both are derived from 5hmC oxidation. Interestingly KRAB-ZnF genes are also expressed at higher levels in embryonic cells and brain [26], so it seems entirely justifiable to suspect that some of the KRAB-ZnF proteins might have gained the ability to bind the oxidative products of 5mC. Some support for this suspicion comes from *in vitro* binding study of Zfp57 itself.

On the one hand, Zfp57 binds most strongly to methylated (5mC) DNA, with decreasing affinity in the order 5mC > 5hmC > 5fC > 5caC [21]. The effects of 5hmC and 5fC on binding by Zfp57 are similar to that of demethylation (loss of methyl groups altogether), whereas 5caC is similar to non-specific binding (*i.e.*, changing one or more base pairs from the recognition sequence). On the other hand, a single spatially-conserved Glu residue in Zfp57 and Kaiso may explain the discrimination against oxidative products of 5mC, and even provide a possible means to predict which RH-containing KRAB proteins might recognize these oxidative products. Zfp57 has a Glu at position 182, which spatially corresponds to Glu535 of Kaiso (Figures 2c–d). Like Kaiso Glu535, Zfp57 Glu182 forms a van der Waals contact with the methyl group of 5mC and one of its carboxylate oxygen atoms interacts with the N4 atom of the same 5mC base. Substitution of Glu182 with its corresponding amide (E182Q) had no effect on methylated DNA binding [21] because the mutant could maintain the same interactions with the 5mC. Surprisingly, this E-to-Q mutant loses the ability to discriminate against unmethylated DNA as well as 5mC oxidative products and, most dramatically, has ~100-fold higher binding affinity to 5caC DNA (in comparison to wild-type protein; Figure 2e). It is possible that the uncharged amide group of E182Q can interact more favorably with the carboxylate group of 5caC. This observation raises the possibility that some KRAB-ZnF proteins might preferentially recognize the oxidative products of 5mC.

Concluding remarks

Despite the large number of KRAB-ZnF proteins, few of them have known biological roles (Figure 3c), and even fewer have known target DNA sequences. The discovery that Zfp57 recognizes a TGCMGCR element provided one of the first examples, among KRAB-ZnF proteins, of DNA recognition sensitive to sequence as well as methylation status [31]. It should be a very high priority to experimentally determine the target DNA sequence for each KRAB-ZnF protein.

Here, we propose that the two-thirds of human KRAB-ZnF proteins containing RH-ZnF motifs might provide specificity for 5mCpG or its oxidized derivatives, within a sequence context determined by the other associated tandem ZnF fingers. Asymmetric binding by one arginine (Kaiso and Zfp57) might be related to the asymmetric nature of the TpG sequence (CpA on the opposite strand). By contrast, MBDs use two arginines for the symmetrical 5mCpG dinucleotide. This prediction provides a testable model of how methylated DNA is recognized by various DNA binding proteins, once the specific target sequence is known, for example, by ChIP-Seq analyses. Structural analyses will clearly continue to play a central role, synergistically with biochemical and genetic studies, in addressing the

recognition code of 5mCpG or its oxidized derivatives. Importantly, with ~200 human KRAB proteins having appropriately-located RH motifs, if even a fraction of them turn out to be truly responsive to DNA methylation status then this would represent a large increase in the number of such known proteins.

Finally, use of the methyl-Arg-G triad to recognize both 5mCpG and TpG raises an intriguing possibility. Perhaps TpG dinucleotides are selected for when it is advantageous for a particular DNA sequence to be treated as if it is *permanently* methylated; the dual binding (5mC/T) triad mechanism provides that option.

Acknowledgments

Y.L. performed all experiments and sequence analyses and all were involved in analyzing data and preparing the manuscript. The U.S. National Institutes of Health (grant GM049245-19 to X.C. and X.Z.) and National Science Foundation (grant MCB-0964728 to R.M.B.) supported this work. X.C. is a Georgia Research Alliance Eminent Scholar and R.M.B. is a U.T. Distinguished University Professor. The authors thank the editor and anonymous reviewers for their helpful suggestions.

References

- Bestor T, et al. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol.* 1988; 203:971–983. [PubMed: 3210246]
- Okano M, et al. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet.* 1998; 19:219–220. [PubMed: 9662389]
- Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 2009; 324:930–935. [PubMed: 19372391]
- Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science.* 2011; 333:1300–1303. [PubMed: 21778364]
- He YF, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science.* 2011; 333:1303–1307. [PubMed: 21817016]
- Inoue A, et al. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* 2011; 21:1670–1676. [PubMed: 22124233]
- Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science.* 2011; 334:194. [PubMed: 21940858]
- Dhasarathy A, Wade PA. The MBD protein family—reading an epigenetic mark? *Mutat Res.* 2008; 647:39–43. [PubMed: 18692077]
- Guy J, et al. The role of MeCP2 in the brain. *Annual review of cell and developmental biology.* 2011; 27:631–652.
- Hashimoto H, et al. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics.* 2009; 4:8–14. [PubMed: 19077538]
- Sharif J, Koseki H. Recruitment of Dnmt1: roles of the SRA protein Np95 (Uhrf1) and other factors. *Progress in molecular biology and translational science.* 2011; 101:289–310. [PubMed: 21507355]
- Scarsdale JN, et al. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* 2011; 39:6741–6752. [PubMed: 21531701]
- Ho KL, et al. MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol Cell.* 2008; 29:525–531. [PubMed: 18313390]
- Ohki I, et al. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell.* 2001; 105:487–497. [PubMed: 11371345]
- Hashimoto H, et al. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature.* 2008; 455:826–829. [PubMed: 18772888]

16. Arita K, et al. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*. 2008; 455:818–821. [PubMed: 18772891]
17. Avvakumov GV, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*. 2008; 455:822–825. [PubMed: 18772889]
18. Rajakumara E, et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev*. 2011; 25:137–152. [PubMed: 21245167]
19. Sasai N, et al. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res*. 2010; 38:5015–5022. [PubMed: 20403812]
20. Buck-Koehntop BA, et al. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc Natl Acad Sci U S A*. 2012; 109:15229–15234. [PubMed: 22949637]
21. Liu Y, et al. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev*. 2012; 26:2374–2379. [PubMed: 23059534]
22. Prokhortchouk A, et al. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev*. 2001; 15:1613–1618. [PubMed: 11445535]
23. Mackay DJ, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet*. 2008; 40:949–951. [PubMed: 18622393]
24. Collins T, et al. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol Cell Biol*. 2001; 21:3609–3615. [PubMed: 11340155]
25. Meylan S, et al. A gene-rich, transcriptionally active environment and the pre-deposition of repressive marks are predictive of susceptibility to KRAB/KAP1-mediated silencing. *BMC Genomics*. 2011; 12:378. [PubMed: 21791101]
26. Vinogradov AE. Human more complex than mouse at cellular level. *PLoS ONE*. 2012; 7:e41753. [PubMed: 22911852]
27. Huntley S, et al. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res*. 2006; 16:669–677. [PubMed: 16606702]
28. Thomas JH, Emerson RO. Evolution of C2H2-zinc finger genes revisited. *BMC Evol Biol*. 2009; 9:51. [PubMed: 19261184]
29. Nowick K, et al. Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS ONE*. 2011; 6:e21553. [PubMed: 21738707]
30. Wolfe SA, et al. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct*. 2000; 29:183–212. [PubMed: 10940247]
31. Quenneville S, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell*. 2011; 44:361–372. [PubMed: 22055183]
32. Daniel JM, et al. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res*. 2002; 30:2911–2919. [PubMed: 12087177]
33. Zou X, et al. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res*. 2012; 40:2747–2758. [PubMed: 22110028]
34. Lamoureux JS, et al. Recognition of 5'-YpG-3' sequences by coupled stacking/hydrogen bonding interactions with amino acid residues. *J Mol Biol*. 2004; 335:399–408. [PubMed: 14672650]
35. Lamoureux JS, Glover JN. Principles of protein-DNA recognition revealed in the structural analysis of Ndt80-MSE DNA complexes. *Structure*. 2006; 14:555–565. [PubMed: 16531239]
36. Gregg C, et al. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science*. 2010; 329:682–685. [PubMed: 20616234]
37. Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*. 2010; 329:643–648. [PubMed: 20616232]

38. Quenneville S, et al. The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. *Cell reports*. 2012
39. Ding G, et al. SysZNF: the C2H2 zinc finger gene database. *Nucleic Acids Res*. 2009; 37:D267–273. [PubMed: 18974185]
40. Hamilton AT, et al. Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. *Genome Res*. 2006; 16:584–594. [PubMed: 16606703]
41. Zuo X, et al. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J Biol Chem*. 2012; 287:2107–2118. [PubMed: 22144682]
42. Frietze S, et al. ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. *PLoS ONE*. 2010; 5:e15082. [PubMed: 21170338]
43. Messerschmidt DM, et al. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science*. 2012; 335:1499–1502. [PubMed: 22442485]
44. Mihola O, et al. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*. 2009; 323:373–375. [PubMed: 19074312]
45. Cheng X, Blumenthal RM. Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. *Biochemistry*. 2010; 49:2999–3008. [PubMed: 20210320]
46. Garcia-Garcia MJ, et al. Chato, a KRAB zinc-finger protein, regulates convergent extension in the mouse embryo. *Development*. 2008; 135:3053–3062. [PubMed: 18701545]
47. Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*. 2009; 458:1201–1204. [PubMed: 19270682]
48. Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res*. 2011; 21:1800–1812. [PubMed: 21784874]
49. Krebs CJ, et al. The KRAB zinc finger protein RSL1 regulates sex- and tissue-specific promoter methylation and dynamic hormone-responsive chromatin configuration. *Mol Cell Biol*. 2012; 32:3732–3742. [PubMed: 22801370]
50. Globisch D, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*. 2010; 5:e15367. [PubMed: 21203455]
51. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009; 324:929–930. [PubMed: 19372393]
52. Khare T, et al. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol*. 2012; 19:1037–1043. [PubMed: 22961382]
53. Williams AJ, et al. Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. *J Biol Chem*. 1995; 270:22143–22152. [PubMed: 7673192]
54. Jenuwein T, et al. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci*. 1998; 54:80–93. [PubMed: 9487389]

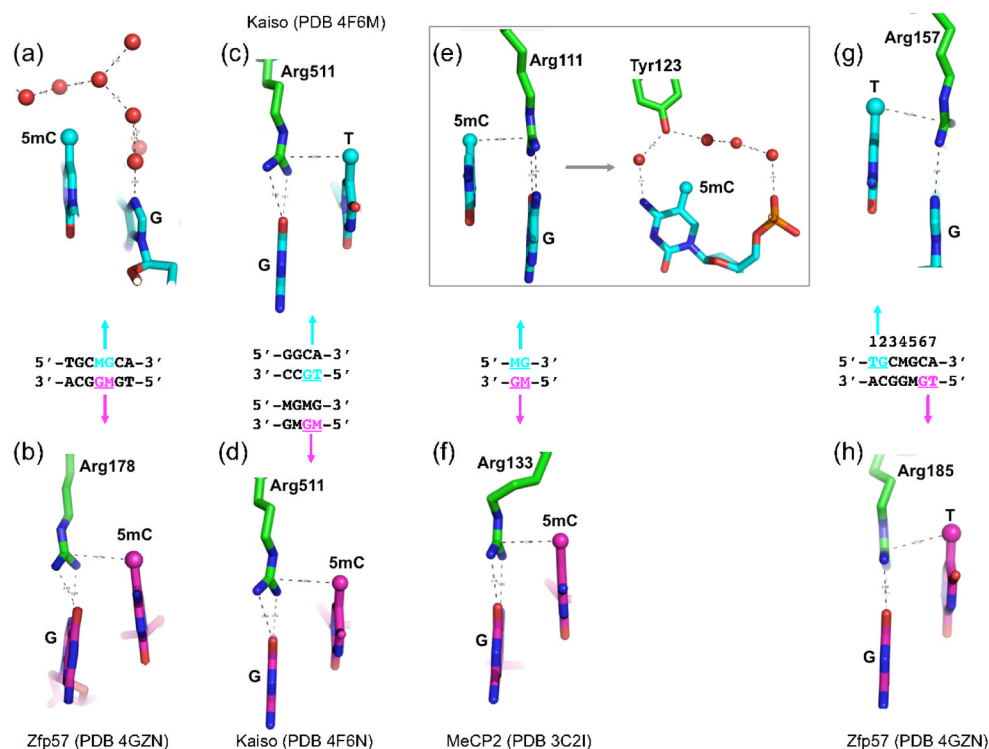


Figure 1. A methyl-Arg-G triad forms during recognition of methyl-CpG and TpG dinucleotides in double stranded DNA

(a–b) Zfp57 recognizes the 5mCs (M) of the two DNA strands via very different interactions. The upper strand 5mC methyl is surrounded by an ordered network of water molecules (red spheres), while the lower strand 5mC methyl is part of a 5mC-Arg-G triad. (c–d) Kaiso recognizes either a specific nonmethylated DNA element containing a TpG dinucleotide (panel c) or two consecutive methylated CpG dinucleotides (panel d). In both cases, a methyl-Arg-G triad is involved.

(e–f) MeCP2 forms two 5mC-Arg-G triads to symmetrically bind the palindromic fully-methylated CpG duplex. In addition, the upper strand 5mC methyl group is surrounded by an ordered network of water molecules that also includes a tyrosine hydroxyl (panel e, right).

(g–h) Zfp57 uses an identical pair of thymine-Arg-G triads for the TpG dinucleotides on the two strands.

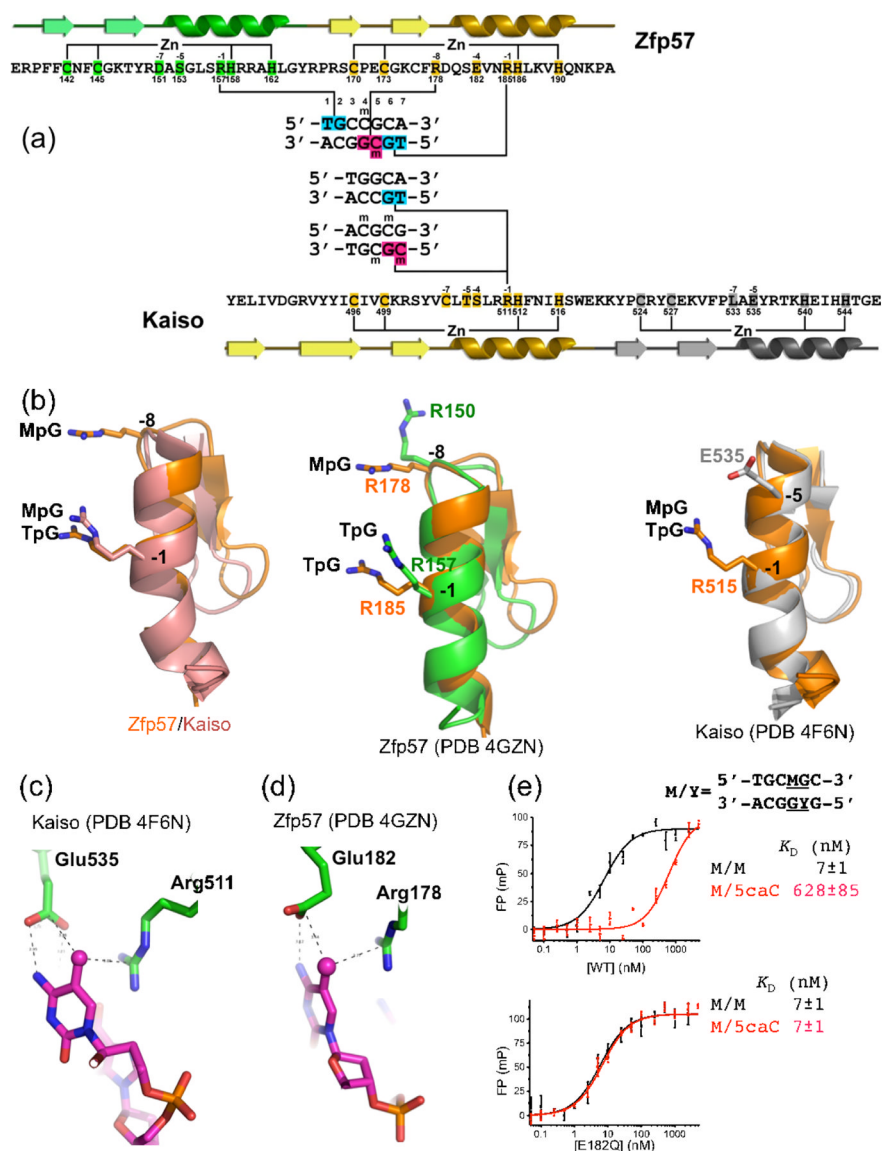


Figure 2. Structural comparison of Zfp57- and Kaiso-DNA interactions

(a) Structure-based sequence alignment of Zfp57 and Kaiso. The secondary structure elements (arrows for β strands and ribbons for α helices) are shown above (Zfp57) or below (Kaiso) the protein sequence. The third ZnF of Kaiso is not shown. The amino acid positions highlighted are responsible for Zn ligand binding (C2H2) and DNA base-specific interactions (-1 to -8). The arginines responsible for TpG or MpG recognition are located in the -1 or -8 positions (relative to the first Zn-binding His).

(b) (Left panel) Superimposition of the recognition helices of Zfp57 (second ZnF) and Kaiso (first ZnF). (Middle panel) Superimposition of the two helices of Zfp57. Arg150 of mouse Zfp57, at the -8 position of the first ZnF, makes a DNA phosphate interaction but is not conserved in human ZFP57. (Right panel) Superimposition of the first two helices of Kaiso.

(c) In Kaiso, in addition to Arg511, the side chain of Glu535 forms a van der Waals contact with the methyl group of 5mC and one of its carboxylate oxygen atoms interacts with the N4 atom of the same 5mC base. The hydrogen bond could also be formed with the O4 atom of the thymine of TpG (presumably through protonation of the carboxylate side chain) [20].

(d) In Zfp57, in addition to Arg178, the side chain of Glu182 forms a van der Waals contact with the methyl group of 5mC and one of its carboxylate oxygen atoms interacts with the N4 atom of the same 5mC base. The hydrogen bond could also be formed with the O4 atom of the thymine of TpG (presumably through protonation of the carboxylate side chain) [20].

(d) Although not aligned at the primary sequence level, Zfp57 has a spatially-conserved glutamate (Glu182) with respect to Glu535 of Kaiso.

(e) DNA binding activity of Zfp57 Glu182 mutants. The 5-carboxylation (5caC) has a large effect on DNA binding by WT Zfp57, but no such effect of 5caC is seen with the E182Q (glutamate to glutamine) mutant.

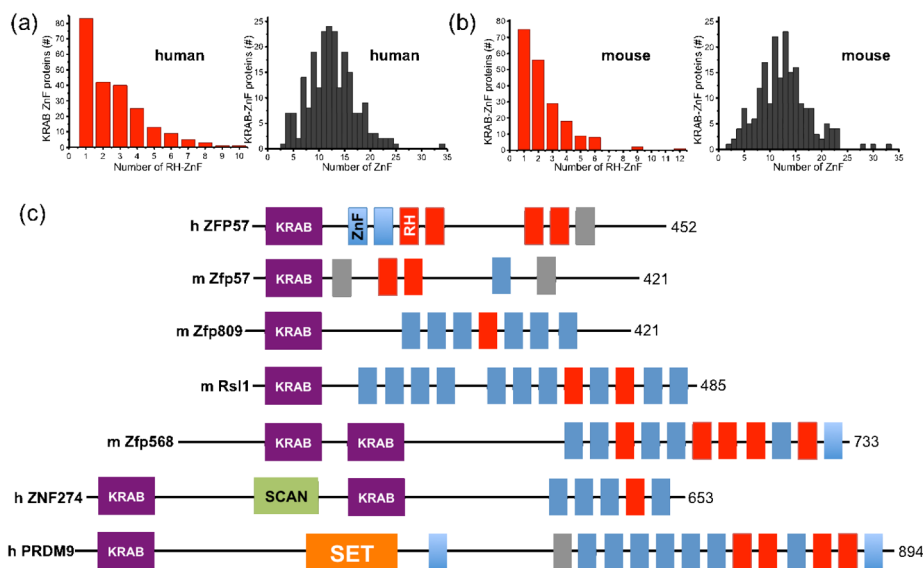


Figure 3. KRAB-ZnF proteins containing RH-ZnF motifs

(a–b) The SysZNF database (<http://lifecenter.sgst.cn/SysZNF/>) was examined for KRAB (Krüppel-associated box)-ZnF proteins in human (a) or mouse (b). In each case, the distribution of KRAB-ZnF proteins containing a given number of RH-ZnF motifs is shown in the left panel and the distribution of KRAB-ZnF proteins containing a given number of ZnF repeats is shown in right panel.

(c) Examples of mammalian KRAB-ZnF proteins with known biological roles. The size of each protein (in amino acids) is shown at right. The classic C2H2 ZnF motifs are shown in blue boxes and the RH-ZnF motifs are shown in red boxes. The grey boxes indicate degenerate ZnFs that include mutations affecting zinc coordination. The SCAN box, a leucine-rich region, was named after SRE-ZBP, CTfin51, AW-1 (ZNF174), and Number 18 cDNA (ZNF20) [53]. The SET domain was named after Su(var)3–9, En(zeste), and TriThorax [54].