Vol 455|9 October 2008|doi:10.1038/nature07280

LETTERS

The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix

Hideharu Hashimoto[1], John R. Horton[1], Xing Zhang[1], Magnolia Bostick[2], Steven E. Jacobsen[2,3] & Xiaodong Cheng[1]

Maintenance methylation of hemimethylated CpG dinucleotides at DNA replication forks is the key to faithful mitotic inheritance of genomic methylation patterns. UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) is required for maintenance methylation by interacting with DNA nucleotide methyltransferase 1 (DNMT1), the maintenance methyltransferase, and with hemimethylated CpG, the substrate for DNMT1 (refs 1 and 2). Here we present the crystal structure of the SET and RING-associated (SRA) domain of mouse UHRF1 in complex with DNA containing a hemimethylated CpG site. The DNA is contacted in both the major and minor grooves by two loops that penetrate into the middle of the DNA helix. The 5-methylcytosine has flipped completely out of the DNA helix and is positioned in a binding pocket with planar stacking contacts, Watson–Crick polar hydrogen bonds and van der Waals interactions specific for 5-methylcytosine. Hence, UHRF1 contains a previously unknown DNA-binding module and is the first example of a non-enzymatic, sequence-specific DNA-binding protein domain to use the base flipping mechanism to interact with DNA.

In mammals, DNMTs belong to two structurally and functionally distinct families[3]. The DNMT3 family establishes the initial CpG methylation pattern *de novo*, whereas DNMT1 maintains this pattern during chromosome replication[4,5] and repair[6]. As a maintenance methyltransferase, DNMT1 has a preference for hemimethylated sites (discussed in ref. 7). DNMT1 is necessary but not sufficient for proper maintenance methylation. The SRA-domain-containing protein UHRF1, also known as NP95 (nuclear protein of 95 kDa) in mouse and ICBP90 (inverted CCAAT binding protein of 90 kDa) in human[8], targets DNMT1 to hemimethylated replication forks[1,2,9]. The UHRF1–DNMT1 interaction probably makes DNMT1 more specific for hemimethylated DNA than can be explained by inherent DNMT1 binding preferences alone[10].

Here we show, by means of X-ray crystallography, that the SRA domain of mouse UHRF1 (residues 419–628; Supplementary Fig. 1) binds to a hemimethylated CpG dinucleotide and increases its protein–DNA interface by flipping 5-methylcytosine (5mC) into a specific binding pocket that probably prevents the SRA domain from sliding off. A 12-base-pair oligonucleotide (plus a 5′-overhanging thymine) containing a single CpG site methylated on one strand was used for co-crystallization with the SRA domain (Fig. 1a). We crystallized the SRA–DNA complex in two different space groups ($P2_12_12_1$ and $P4_12_12$) and determined the structures to the resolutions of 2.19 Å and 1.96 Å, respectively (Supplementary Table 1). The protein components of the two structures are highly similar, with a root mean squared deviation of less than 0.3 Å when comparing 206 pairs of Cα atoms. Here we will describe the structure of $P4_12_12$ and will discuss the differences between the two. In addition, we determined the structure of a non-specific SRA–DNA complex to a lower resolution at 3.09 Å in the space group $P6_122$ (Supplementary Table 1).

The SRA domain contains two twisted β-sheets packed together to form a crescent moon-like structure (Fig. 1b). The 17-residue-long strand β8 (Asn 557–Gly 573) is part of and links together the two sheets. Four helices (α1–α4) and one 3_{10} helix are scattered through the entire domain, with helix α1 packed against the first β-sheet, helix α2 sandwiched between the two β-sheets, and helices α3 and α4

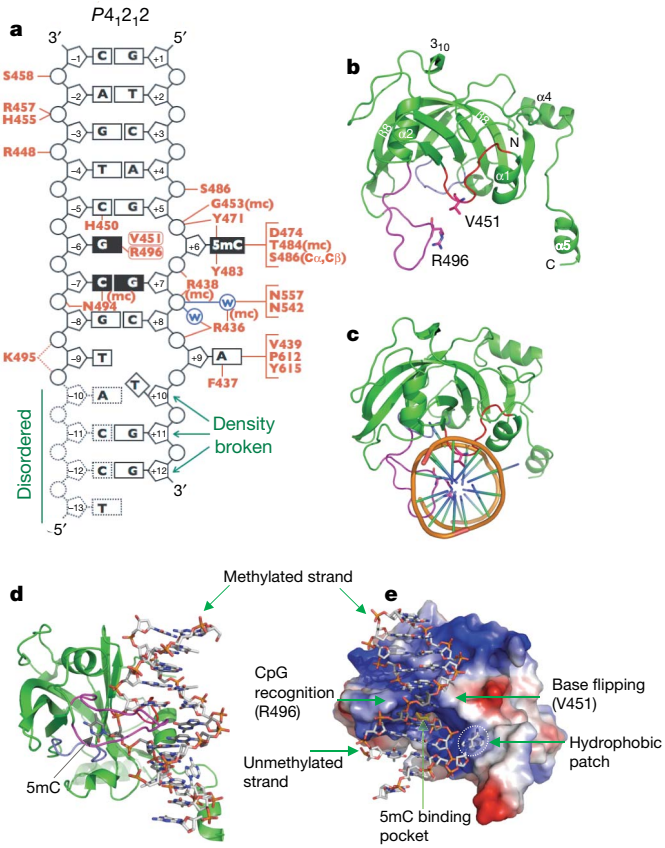

Figure 1 | Structure of SRA–DNA complex. a, Summary of the SRA–DNA interactions; mc, main-chain-atom-mediated contacts; w, water-mediated hydrogen bonds. Black boxes represent CpG recognition sequence and K495-associated dotted lines represent weak hydrogen bonds. **b**, The side chains of V451 of the base flipping loop and R496 of the CpG recognition loop are in direct van der Waals contact. **c**, The two loops—CpG recognition and base flipping—penetrate into the DNA helix from opposite directions. **d**, The 5mC flips out and binds in a cage-like pocket. **e**, The surface charge at neutral pH is displayed as blue for positive ($20\ k_BT$), red for negative ($-20\ k_BT$), and white for neutral, where k_B is the Boltzmann's constant and T is the temperature.

[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA. [2]Department of Molecular Cell and Developmental Biology, [3]Hughes Medical Institute, University of California, Los Angeles, 621 Charles E. Young Dr. South, Los Angeles, California 90095, USA.

positioned on the inner and outer surfaces of the crescent, respectively. The carboxyl helix α5 protrudes away from the rest of the domain. The inner surface of the crescent contains three loops, one between helix α1 and strand β2 (red), one between strands β3 and β4 (light blue), and one between strand β5 and helix α3 (magenta). These loops are responsible for CpG recognition (Asn 494 and Arg 496 in magenta), promotion of base flipping (His 450 and Val 451 in red), and formation of a 5mC-binding pocket (Tyr 483–Ser 486 and Tyr 471–Asp 474; see later).

The DNA is bound to the basic inner surface of the crescent (Fig. 1c–e), with phosphate contacts spanning 8 base pairs but mostly concentrated on the 5 phosphates surrounding the 5mC (two 5′ and three 3′; Fig. 1a). The SRA residues involved in phosphate interactions are scattered throughout an amino-terminal region of ~50 residues (436–486) by forming secondary structures of both β-sheets and their associated loops (Supplementary Fig. 2). The tips of two hairpin loops, Asn 494–Arg 496 and His 450–Val 451, approach the DNA from opposite directions. They contact major and minor grooves, respectively, and penetrate into the DNA helix by forming van der Waals contacts between side chains of Arg 496 and Val 451 in the centre of the DNA helix (Fig. 1c–e).

The SRA side chain intercalation is associated with 5mC flipping out of the DNA helix (Fig. 1d). Val 451 from the minor groove side occupies the space left by the everted base, and we refer to the Val 451-containing loop as the base-flipping-promotion loop (Fig. 1b–e). On the major groove side, Arg 496 makes bifurcated hydrogen bonding interactions with the intrahelical orphaned guanine (Fig. 2a). Asn 494 makes bridging hydrogen bonds between Arg 496 and the 5′ phosphate of unmethylated cytosine on the opposite strand (Fig. 2a). Furthermore, the backbone carbonyl oxygen of Asn 494 is in close proximity to the ring carbon atom C5 of the unmethylated cytosine, forming a C=O···H–C type of hydrogen bond. We name the Asn 494/Arg 496-containing loop the CpG recognition loop, as it seems to determine specificity for hemimethylated CpG. The addition of a methyl group to the ring carbon C5 of the unmethylated cytosine would cause a steric clash between the methyl group and Asn 494, and could account for the approximately sevenfold decrease in binding to fully methylated DNA[1] (Supplementary Fig. 3).

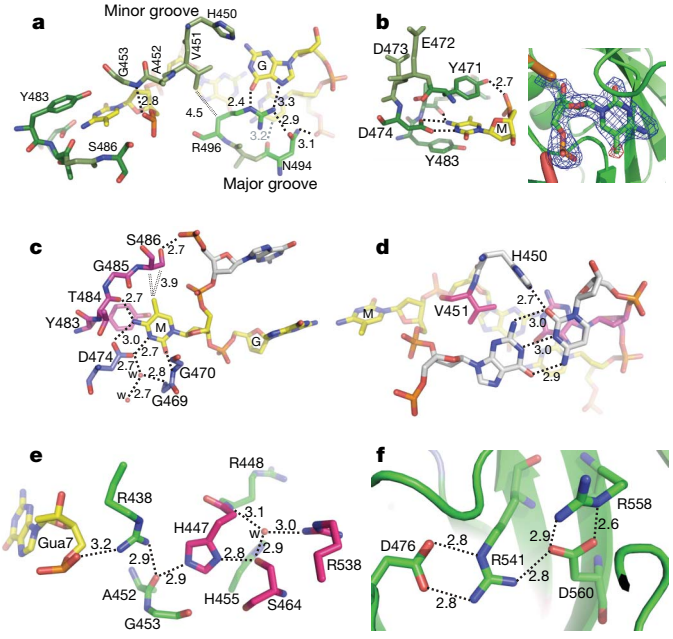

Figure 2 | Details of SRA–DNA interactions. **a**, The 5mC·G base pair is shown in the front, and the adjoining G·C base pair is in the back. **b**, Planar stacking contacts of the extrahelical 5mC with Y471 and Y483 (left image). Omit electron densities, contoured at 4σ and 5σ above the mean, respectively, are shown for omitting 5mC (blue) or the methyl group (red) (right image). **c**, The hydrogen bond interactions with the polar atoms of 5mC. The double-dotted lines indicate van der Waals contacts with the methyl group of ring carbon C5. **d**, H450 forms a hydrogen bond from the minor groove side with cytosine of G·C pair at position 5 (see Fig. 1a). **e**, Network of internal polar interactions centred on residues H447 and S464. Gua, guanine. **f**, Network of internal charged interactions centred on residues R541 and D560. Distances are shown in angstroms.

The extrahelical 5mC is bound in a hydrophobic cage formed by two tyrosines (Tyr 471 and Tyr 483) stacking the flipped base in between (Fig. 2b). The polar groups of the 5mC ring that normally form the Watson–Crick pair with guanine are now involved in hydrogen bonds with the main-chain amide nitrogen atoms of Ala 468 and Gly 470 (interacting with O2 oxygen atom), the side-chain carboxylate oxygen atoms of Asp 474 (interacting with N3 and N4 nitrogen atoms) and the main-chain carbonyl oxygen of Thr 484 (interacting with N4 nitrogen atom; Fig. 2c). Interactions with the exocyclic amino group N4 (NH_2) define the binding pocket base specificity for a cytosine. The distance between Asp 474 and the ring nitrogen N3 (2.7 Å) strongly suggests the presence of a hydrogen bond and therefore the presence of a proton either at the carboxylate oxygen or at N3. The most likely source of that proton is an ordered water molecule, which is in fact found in direct contact with the side chain of Asp 474 and a network of other water molecules. The methyl group at the C5 position in 5mC is in van der Waals contact with the Cα and Cβ atoms of Ser 486, the hydroxyl oxygen of which interacts with the phosphate of the 5′ nucleotide (Fig. 2c).

Outside of the hemimethylated CpG dinucleotides, two additional bases are in direct contact with the SRA domain. The cytosine of the G•C pair at position 5 has a hydrogen bond, by means of its hydrogen-acceptor O2 in the minor groove, with His 450 (Fig. 2d). The His 450-mediated interaction is probably not base specific at this position, as in theory the hydrogen bond could also form with thymine O2 of an A•T pair, guanine N3 of a C•G pair, or adenine N3 of a T•A pair.

Surprisingly, the adenine at the position 9, two bases 3′ to the 5mC, also adopts an extrahelical conformation (Fig. 1a) and stacks against a hydrophobic surface patch (Fig. 1e and Supplementary Fig. 4). The flipping of adenine seems to destabilize the rest of the DNA, leaving the last four nucleotides of the unmethylated strand (positions −10 to −13) disordered. In addition, the last two nucleotides of the methylated strand (positions 11–12)—having only broken, residual densities—appear as if they are single-stranded DNA. The flexibility of the last four base pairs (positions 9–12) is probably due to the fact that the SRA-domain–DNA interactions are shifted towards one end of the DNA, leaving the disordered end unattended (Fig. 1a). In the space group $P2_12_12_1$ crystal, these flexible base pairs (positions 9–12) were more ordered, stabilized by contacts with the crystallographic symmetry-related molecule (Fig. 3a and Supplementary Fig. 5). Interestingly, in this complex the adenine 9 adopts two conformations—an intrahelical one with normal pairing and the extrahelical one described for the $P4_12_12$ complex, with adenine 9 stacking against the hydrophobic patch. It is also possible that the hydrophobic patch normally interacts with residues outside of the SRA domain used here for crystallization[11]. The absence of such residues and the weak crystallographic restraining forces on the end of the DNA may have caused adenine 9 to be extrahelical and interact with the hydrophobic patch in the crystal.

In the complex with unmethylated DNA, the CpG recognition loop (residues 490–500) appears disordered (Fig. 3b). The SRA domain does not bind the equivalent non-methylated CpG site, but instead it binds the junction between the two DNA duplexes, arranged head-to-head by the crystallographic twofold symmetry, suggesting that the SRA binds unmethylated DNA non-specifically and only the non-heterogeneous end complex is trapped in crystal form. The side chains of Val 451 and His 450 are in close proximity to the first two base pairs, which are unpaired and non-stacked. These observations are reminiscent of some DNA methyltransferases that

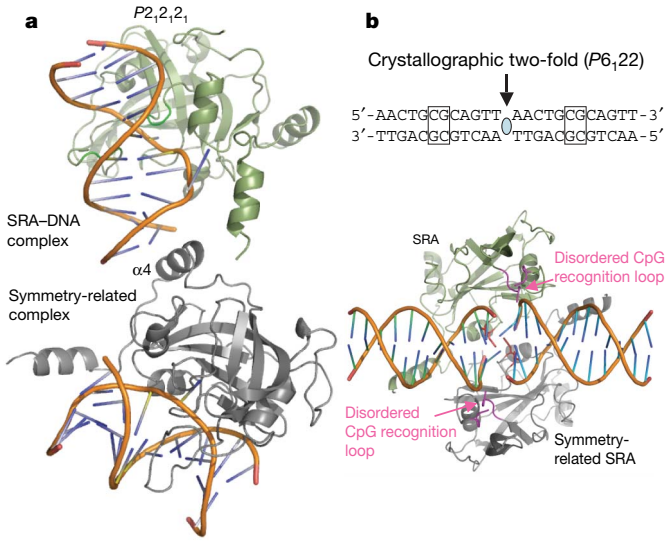

Figure 3 | Structure of the SRA–DNA specific ($P2_12_12_1$) and non-specific ($P6_122$) complexes. a, The crystal packing interactions in the $P2_12_12_1$ space group may stabilize the DNA duplex. The extra contacts are made by the helix α4 of a symmetry-related molecule. These interactions are not present in the $P4_12_12$ complex. **b**, A 12-base-pair duplex containing a single, centrally located unmethylated CG site was used to generate a non-specific complex with the SRA domain. A V451/H450-associated loop approaches the DNA from the minor groove at the junction of two head-to-head DNA molecules.

also bind at the junction of two duplexes[12–14], which mimics damaged DNA or altered recognition sites[15].

A role for UHRF1 in DNA methylation control is widely conserved because mutation of the *Arabidopsis* homologue VIM1 (also known as ORTH2) reduces CG methylation and its SRA domain binds methylated CG sites[16,17]. In addition, three missense mutations arising in *Arabidopsis* SRA domains are required for binding methylated DNA[18]. The corresponding residues in mouse UHRF1 are Ser 464, Glu 472 and Arg 541 (Supplementary Fig. 2). Ser 464 and Arg 541 are involved in intramolecular interactions that probably confer stability to the molecule (Fig. 2e, f). Glu 472 is part of the 5mC-binding pocket, but is not directly involved in interaction with the flipped 5mC (see Fig. 2b).

Base flipping is a conserved mechanism that is widely used by nucleotide-modifying enzymes, including DNA methyltransferases[19,20], DNA repair enzymes[21–23] and RNA modification enzymes[24]. This mechanism, first discovered in the bacterial 5mC methyltransferase M.HhaI (ref. 19), involves enzyme binding to the DNA and eversion of the target nucleotide so that it projects out of the double helix and into the active-site pocket. The SRA domain is the first non-enzyme sequence-specific DNA-binding protein domain that uses the base flipping mechanism in its interaction with DNA.

There is no apparent sequence or structural similarity between the SRA and the DNA methyltransferase domain (or of DNA repair enzymes). However, the phosphodiester backbone pinching[25] caused by extensive protein–phosphate contacts surrounding the flipped nucleotide, the use of two loops to approach DNA from the major and minor grooves simultaneously, and the binding of the flipped base in a concave pocket are analogous (Supplementary Fig. 6)[26]. Furthermore, enzymes use base flipping to gain access to a DNA base to perform chemistry on it, but the SRA domain probably uses base flipping to increase its protein–DNA interface and to prevent the SRA domain from sliding off the DNA. This may be particularly important for the SRA domain, as its recognition sequence is only two base pairs. The surface area buried at the SRA–DNA interface is approximately 2,500 Å^2, an ~70% increase from what is buried at the MBD1–DNA interface (1,480 Å^2), which does not involve base flipping[27,28] (Supplementary Fig. 7). The 5mC base flipping by the SRA domain might also provide a more general mechanism to distinguish the methylated parental strand from the unmethylated daughter strand, an ability particularly important for mismatch repair if an error occurs during DNA replication. Supporting this hypothesis, the expression of ICBP90 (the human orthologue of UHRF1) is deregulated in cancer cells[29], and mouse UHRF1-null cells are more sensitive to DNA-damaging agents and DNA replication arrest[30]. We therefore suggest that the SRA–DNA interaction (through recognition and flipping of the 5mC) serves as a placeholder to keep UHRF1 at hemimethylated CpG site where it recruits DNMT1 for maintenance methylation, and perhaps other proteins such as DNA repair enzymes for mismatch repair.

METHODS SUMMARY

We generated a hexahistidine–SUMO (small ubiquitin-like modifier)-tagged construct containing mouse UHRF1 residues 419–628 (pXC666). The fusion protein was cleaved and the SRA domain was crystallized with DNA containing either a hemimethylated CpG site or a non-specific DNA. The structures were solved by molecular replacement.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 March; accepted 23 July 2008.
Published online 3 September 2008.

1. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317,** 1760–1764 (2007).
2. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908–912 (2007).
3. Cheng, X. & Blumenthal, R. M. Mammalian DNA methyltransferases: a structural perspective. *Structure* **16,** 341–350 (2008).
4. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481–514 (2005).
5. Chen, T. & Li, E. Establishment and maintenance of DNA methylation patterns in mammals. *Curr. Top. Microbiol. Immunol.* **301,** 179–201 (2006).
6. Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M. C. & Leonhardt, H. Recruitment of DNA methyltransferase I to DNA repair sites. *Proc. Natl Acad. Sci. USA* **102,** 8905–8909 (2005).
7. Jeltsch, A. Molecular enzymology of mammalian DNA methyltransferases. *Curr. Top. Microbiol. Immunol.* **301,** 203–225 (2006).
8. Unoki, M., Bronner, C. & Mousli, M. A concern regarding the current confusion with the human homolog of mouse Np95, ICBP90/UHRF1. *Radiat. Res.* **169,** 240–244 (2008).
9. Achour, M. *et al.* The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* **27,** 2187–2197 (2008).
10. Ooi, S. K. & Bestor, T. H. Cytosine methylation: remaining faithful. *Curr. Biol.* **18,** R174–R176 (2008).
11. Arita, K. Ariyoshi, M. Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* doi:10.1038/nature07249 (this issue).
12. Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A. & Cheng, X. Structure and substrate recognition of the *Escherichia coli* DNA adenine methyltransferase. *J. Mol. Biol.* **358,** 559–570 (2006).
13. Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A. & Cheng, X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell* **121,** 349–361 (2005).
14. Yang, Z. *et al.* Structure of the bacteriophage T4 DNA adenine methyltransferase. *Nature Struct. Biol.* **10,** 849–855 (2003).
15. Klimasauskas, S. & Roberts, R. J. M. HhaI binds tightly to substrates containing mismatches at the target base. *Nucleic Acids Res.* **23,** 1388–1395 (1995).
16. Woo, H. R., Pontes, O., Pikaard, C. S. & Richards, E. J. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* **21,** 267–277 (2007).
17. Johnson, L. M. *et al.* The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17,** 379–384 (2007).
18. Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21,** 6842–6852 (2002).
19. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76,** 357–369 (1994).
20. Cheng, X. & Roberts, R. J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **29,** 3784–3795 (2001).
21. Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452,** 961–965 (2008).
22. Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449,** 570–575 (2007).
23. Parker, J. B. *et al.* Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449,** 433–437 (2007).

24. Lee, T. T., Agarwalla, S. & Stroud, R. M. A unique RNA fold in the RumA–RNA–cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120,** 599–611 (2005).
25. Werner, R. M. *et al.* Stressing-out DNA? The contribution of serine–phosphodiester interactions in catalysis by uracil DNA glycosylase. *Biochemistry* **39,** 12585–12594 (2000).
26. Cheng, X. & Blumenthal, R. M. Finding a basis for flipping bases. *Structure* **4,** 639–645 (1996).
27. Ohki, I. *et al.* Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105,** 487–497 (2001).
28. Ho, K. L. *et al.* MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).
29. Mousli, M. *et al.* ICBP90 belongs to a new family of proteins with an expression that is deregulated in cancer cells. *Br. J. Cancer* **89,** 120–127 (2003).
30. Muto, M. *et al.* Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *J. Biol. Chem.* **277,** 34549–34555 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. M. Blumenthal for critical comments. The Emory University School of Medicine supported the use of SER-CAT beamlines. This work was supported by grant GM049245 to X.C. from the National Institutes of Health (NIH). Work in the Jacobsen laboratory is funded by the NIH grant GM060398. M.B. is funded by NIH-NSRA Fellowship number CA1263022. S.E.J. is an Investigator of the Howard Hughes Medical Institute and X.C. is a Georgia Research Alliance Eminent Scholar.

Author Information The X-ray structures (coordinates and structure factor files) of mouse UHRF1 SRA with bound DNA have been submitted to PDB under accession numbers 2ZO0 ($P2_12_12_1$), 2ZO1 ($P4_12_12$) and 2ZO2 ($P6_122$), respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.C. (xcheng@emory.edu).

METHODS

Expression and purification. We generated a hexahistidine–SUMO-tagged construct[31] containing mouse UHRF1 residues 419–628 (pXC666). Protein was expressed in *Escherichia coli* BL21(DE3)-Gold cells (Stratagene) with the RIL-Codon plus plasmid. Expression cultures were grown at 25 °C in autoinduction medium[32]. Cells were lysed as a 20% (v/v) suspension in 50 mM sodium phosphate, pH 7.4, 300 mM NaCl, 5% (v/v) glycerol, 0.1% 2-mercaptoethanol and 40 mg ml^{-1} phenylmethylsulphonyl fluoride by two passes through an ice-cold French pressure cell press. The lysate was clarified by centrifugation twice at 50,000*g* for 30 min. Hexahistidine–SUMO fusion proteins were isolated on a nickel-charged chelating column (GE Healthcare). After imidazole elution, the fusion protein was cleaved with Ulp1 protease[33] at 25 U ml^{-1} with 16 h dialysis at room temperature (approximately 20 °C). Only two extraneous N-terminal amino acids (HisMet) were left as a result of the NdeI restriction site. The protein was further purified by ion exchange (HiTrap-S) and gel filtration chromatography (Superdex-75, GE Healthcare). A maximum yield of 2.8 mg l^{-1} for the UHRF1 SRA domain was obtained. The domain is monomeric and showed little aggregation during gel filtration in the presence of 5% (v/v) glycerol and 0.1% 2-mercaptoethanol.

Crystallography. For co-crystallization with DNA, the SRA was concentrated to approximately 42 mg ml^{-1} in 20 mM HEPES, pH 7.0, 150 mM NaCl, 0.1% 2-mercaptoethanol and 5% glycerol, and incubated with an annealed oligonucleotide duplex (5′-TCCATGCGCTGAC-3′ and 5′-GTCAGMGCATGG-3′, where M = 5mC) at a 1:1.2 ratio for 30 min before setting up crystallization drops. The crystals appeared in various morphologies, sometimes in the same drops. Diamond-shaped ($P4_12_12$) and needle-shaped ($P2_12_12_1$) crystals were grown under a 20% (v/v) polyethylene glycol 3350, 0.4 M NaCl condition. X-ray diffraction data were collected, at the SER-CAT beamline, and processed with HKL2000[34], from crystals cryoprotected by the mother liquor supplemented with 40% ethylene glycerol and allowing crystals to soak for several minutes.

For crystallization with unmethylated oligonucleotide, SRA protein (~32.5 mg ml^{-1}) and annealed DNA (5′-AACTGCGCAGTT-3′) were mixed at a 1:1.5 ratio and were incubated for 30 min before setting up crystallization drops. Diamond-shaped crystals grown under 20% (v/v) polyethylene glycol 3350 or 10000, 0.2–0.4 M NaCl and 0.1 M MES (pH 5.8–6.2 or without buffer). However, only one crystal (out of 40–50 crystals screened) diffracted X-ray strong enough and allowed us to collect a complete data set to 3.09 Å resolution.

During the course of study, the Structural Genomics Consortium deposited coordinates of the apo structure of the human SRA domain (PDB accession number 3BI7), which encouraged us to focus on the SRA–DNA complex. We used the molecular replacement program PHASER[35] to obtain crystallographic phases. The resulting electron density map for DNA and the structured portion of the CpG recognition loop (residues 483–496) was easily interpretable, using the model-building program O[36]. CNS[37] scripts were used for refinement, and the statistics shown in Supplementary Table 1 were calculated for the entire resolution range. The R_{free} and R_{work} values were calculated for 5% (randomly selected) and 95%, respectively, of observed reflections. The structures of $P2_12_12_1$ and $P4_12_12$ were solved, built and refined independently. For the non-specific complex structure, discontinued densities do exist but we were not able to unambiguously distinguish between solvent molecules and disordered protein densities. The assignment of solvent molecules to these residual densities would reduce values of R_{factor} and R_{free}; however, we took a conservative approach without including such solvent molecules in the final model.

31. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448,** 718–722 (2007).
32. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
33. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5,** 75–86 (2004).
34. Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A* **59,** 228–234 (2003).
35. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60,** 432–438 (2004).
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
37. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).