Protein Expression and Purification 83 (2012) 104–111

Contents lists available at SciVerse ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

# Heterologous expression and purification of *Arabidopsis thaliana* VIM1 protein: *In vitro* evidence for its inability to recognize hydroxymethylcytosine, a rare base in *Arabidopsis* DNA

Qin Yao [a], Chun-Xiao Song [b], Chuan He [b], Desigan Kumaran [a], John J. Dunn [a,*]

[a] Biology Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA
[b] Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago IL 60637, USA

ARTICLE INFO

*Article history:*
Received 27 February 2012
and in revised form 3 March 2012
Available online 20 March 2012

*Keywords:*
*Arabidopsis thaliana*
VIM1 (*V*ariant *i*n *M*ethylation 1) protein
5-Methyl-cytosine 5-hydroxymethyl-cytosine
Electrophoretic-mobility shift (EMSA) assay
Fluorescence anisotropy (FA) assays

ABSTRACT

The discovery of 5-hydroxymethyl-cytosine (5hmC) in mammalian cells prompted us to look for this base in the DNA of *Arabidopsis thaliana* (thale cress), and to ask how well the Arabidopsis Variant in Methylation 1 (VIM1) protein, an essential factor in maintaining 5-cytosine methylation (5mC) homeostasis and epigenetic silencing in this plant, recognizes this novel base. We found that the DNA of Arabidopsis' leaves and flowers contain low levels of 5hmC. We also cloned and expressed in *Escherichia coli* full-length VIM1 protein, the archetypal member of the five Arabidopsis VIM gene family. Using *in vitro* binding assays, we observed that full-length VIM1 binds preferentially to hemi-methylated DNA with a single modified 5mCpG site; this result is consistent with its known role in preserving DNA methylation *in vivo* following DNA replication. However, when 5hmC replaces one or both cytosine residues at a palindromic CpG site, VIM1 binds with approximately $\geqslant$10-fold lower affinity. These results suggest that 5hmC may contribute to VIM-mediated passive loss of cytosine methylation *in vivo* during Arabidopsis DNA replication.



## Introduction

Cytosine methylation is a post replicative process occurring primarily at the DNA dinucleotide CpG, where in both cytosines in the complementary strands of adjacent base pairs are methylated at the C′ 5 position (5mC)[1] [1,2]. Less frequently, the 5mCs in the complementary strands are separated by a single base pair (CpHpG); only rarely are they farther apart (CpHpH) (H is any base other than G) [3–6]. This pairing of 5mC on complementary strands at CpG dinucleotides affords a mechanism for maintaining the post-replication methylation marks because DNA methyltransferases, appropriately termed maintenance methyltransferases, recognize the 5mC on the parental strand, and direct the methylation of the adjacent C in the new daughter strand [7,8]. Although these maintenance methyltransferases intrinsically favor newly replicated or hemi-methylated substrates *in vitro* [9] other proteins, particularly those having so-called SET- and RING-associated (SRA) domains that bind preferentially to hemi-5mC-containing DNA, are known to be needed for high-fidelity preservation of 5mC epigenetic patterns in plant- and animal-cells.

In mammals, ubiquitin-like proteins with plant homeodomain (PHD)- and ring-finger domains 1- (UHRF1) protein [10], also known as ICBP90 or Np95 [11,12], recognize hemi-methylated CpG sites that appear during DNA replication. Structural studies revealed that the 5mC residue in hemi-methylated DNA strand bound to UHRF1 is flipped outside of the DNA helix into a specific 5mC-binding pocket within the SRA domain, and that the target cytosine in the daughter strand is then transferred to DNMT1 for maintenance methylation [13–15].

In Arabidopsis the VIM/ORTHUS gene family encodes five SRA methylcytosine-binding proteins (VIM1-5) whose functions probably are similar to those of UHRF1, viz., playing a major role in maintaining DNA-methylation patterns after DNA replication by presenting the target cytosine in the daughter strand opposite the 5mCpG dinucleotides to MET1, the plant homologue of DNMT1 [16–19]. To gain further insight into how these proteins function, we cloned, expressed, and purified full-length VIM1, the archetypical member of this gene family and carried out electrophoretic mobility-shift (EMSA) and fluorescence-anisotropy (FA) titration assays to study its interaction with model duplex DNAs containing cytosine or 5mC in one or both strands at a palindromic CpG site. In

* Corresponding author. Address: Biology Department, Bldg. 463, Brookhaven National Laboratory, Upton, NY 11973-5000, USA. Fax: +1 631 344 3407.
*E-mail addresses:* qyao@bnl.gov (Q. Yao), cxsong@uchicago.edu (C.-X. Song), chuanhe@uchicago.edu (C. He), kumaran@bnl.gov (D. Kumaran), jdunn@bnl.gov (J.J. Dunn).

[1] *Abbreviations used:* ds, double-stranded; 5mC, 5-methyl-cytosine; 5hmC, 5-hydroxymethyl-cytosine; EMSA, electrophoretic-mobility shift assay; FA, fluorescence anisotrophy assay; biotin-N3-5-gmC, biotin β-6-azide-glucosyl-5-hydroxymethyl-cytosine.

1046-5928/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.pep.2012.03.003

addition, since recent studies of genomic DNAs from human brain, neurons, and mouse embryonic stem-cells demonstrated that they contain a sixth base, oxidized 5mC or 5hydroxy-methylcytosine (5hmC) [20–22], we were interested in seeing if Arabidopsis DNA contains this base, and whether it interferes with VIM1's ability to bind methylated DNA. In our study, we detected this base using a bacteriophage T4 β-glucosyltransferase-based assay that involves transferring a modified glucose moiety on to the hydroxyl group of 5hmC [23]. We also confirmed its presence using antibodies that recognize 5hmC but not 5mC.

Our results show that 5hmC adversely impacts VIM1's DNA-binding *in vitro,* raising the possibility that, in contrast to 5mC residues at replication forks, 5hmC might not efficiently recruit MET1 to methylate the daughter strand's newly replicated CpG site. If this status is left uncorrected and these marks are not re-established by *de novo* methylation before the next round of DNA replication, it could entail VIM1-mediated passive de-methylation and loss of epigenomic homeostasis.

## Materials and methods

### *Materials*

All enzymes were obtained from New England Biolabs (NEB), Ipswich, MA, if not stated otherwise. Dual-color precision protein standards and precast 4–20% acrylamide gels for SDS–PAGE were purchased from BioRad (Hercules, CA). Antibodies to 5mC and 5hmC were purchased from Active Motif (Carlsbad, CA). We followed a published method [24] for purifying *Escherichia coli* McrA-S, which binds ds- DNA with hemi- or fully-methylated HpaII recognition sequences (CCGG), wherein the underlined C is 5mC or 5hmC. We extracted genomic DNA from *Arabidopsis thaliana* (ecotype Columbia), grown under 16 h light and 8 h dark at 30 °C. We extracted it from the leaves and flowers of 5-week-old plants using the Plant DNeasy Mini Kit (QIAGEN, Valencia, CA).

### *Detection of 5hmC*

DNA samples (1 μg genomic DNA, or 5 pmol 90-mer) were denatured in 0.4 M NaOH, 10 mM EDTA at 99 °C for 10 min, then neutralized by adding an equal volume of ice-cold 2 M ammonium acetate (pH 7.0). These denatured samples were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-Rad) with 2-fold serial dilutions, according to the manufacturer's instructions. The blotted membrane was washed with 2× SSC buffer, air-dried, and fixed by Stratagene UV Stratalinker 2400 (auto-crosslink), and then blocked with 5% non-fat milk in 1×Tris buffered saline plus 0.1% Tween 20 (TBST) for 1 h at RT, and incubated with anti-5hmC (Active Motif, 1:10,000) overnight at 4 °C; thereafter, the membrane was washed four times with 1xTBST. After incubating it with HRP-conjugated anti-rabbit IgG secondary antibody (1:5000) for 1 h at room temperature, the membrane again was washed four times with 1xTBST and then retained antibody visualized by enhanced chemiluminescence. Dot-blot detection of biotinylated glucose-5hmC DNA residues after the transfer of a modified glucose moiety on to the hydroxyl group of 5hmC was undertaken as previously described using avidin–horseradish peroxidase (HRP) as detection reagent [23].

### *VIM1 cloning*

*A. thaliana* (ecotype Columbia) cDNA (the kind gift of C.J. Liu BNL) was used as template to PCR-amplify the 1938 bp VIM1 coding sequence with the following primers. 5′-ACTCGGCGCGTGACATCCAACTC-3′ and 5′- CACTGTTACCTGATGGTCGCAGAAACTGTTGCGTCAGTGTC-3′ (regions complementary to *vim1* are underlined). The PCR product was treated with T4 DNA polymerase and dCTP to produce 4-base-long single-stranded overhangs, and then directionally cloned into a $Kan^R$ T7-LacO-based expression vector, pAL13, (Studier and Kasmarcik, unpublished) digested with BsaI to fuse the entire VIM1 ORF, minus its initiating Met, to a N-terminal poly-His tag ($MAS(H)_6$ *S-start of VIM1*). Following transformation into *E. coli* XL1-Blue MR *Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*) cells (Stratagene), we purified several recombinant plasmids and sequenced them to verify that VIM1 ORF had been cloned without nucleotide errors. One correct recombinant plasmid was used as template to re-clone the VIM1 ORF without a fusion tag (rVIM1), and to generate another one with a poly-His tag fused to the C-terminus (*-end of VIM*-$SG(His)_6$), which we designated (rVIM1-H). The resulting sequence verified $Kan^R$ pET-based VIM1 plasmid DNAs were electroporated into BL21(DE3) ($F^-$ *ompT gal dcm lon* $hsdS_B(r_B^- m_B^-)$ *λ(DE3* [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*]) cells with the pACYC-RIL tRNA encoding plasmid ($Cl^R$-*argU*(AGA,AGG), *ileY*(AUA) and *leuW*(CUA)). Transformants were selected on 2xYT agar plates supplemented with 50 μg/ml kanamycin and 25 μg/ml chloramphenicol (CAM).

### *Expression and purification of VIM1*

Initially, we tested the expression of the VIM1 constructs in BL21(DE3)/pRIL cells by adding 0.5 mM IPTG to mid-log-phase cells in 2xYT medium, or by growing them to saturation in ZYM 5052 auto-induction medium at 37 and 20 °C [25]. These media contained 100 μg/ml Kan and 25 μg/ml CAM, and were supplemented with 30 μM $ZnS0_4$ because VIM1 is thought to bind several $Zn^{2+}$ ions. Induction and solubility analysis for the expressed ~71.5 kDa VIM1 protein was done by SDS–PAGE. Most VIM1 protein with or without a $(His)_6$ tag in crude extracts is soluble if expressed at 20 °C. Therefore, we induced protein expression at this temperature, and, for ease, we used auto-inducing conditions. Shake cultures (100–200 ml in 0.5 or 1 l flasks) were started initially at 37 °C by adding 1 ml of an overnight culture in 2xYT to 200 ml autoinduction medium; the temperature was lowered to 20 °C once growth became apparent ($OD_{600} \sim 0.5$). Cells were harvested after 24–36 h by centrifugation at 5000*g* for 10 min at 4 °C, washed with 1/10 vol. 1×PBS, re-centrifuged, and the cell pellets (~1.2 g/50 ml culture) stored at −20 °C.

After thawing and resuspending the cell pellets (200 ml culture) in 20 ml of buffer A (50 mM sodium phosphate, pH 7.4; 150 mM NaCl; 0.1 mM dithiothreitol (DTT); 10% glycerol), lysozyme (20 mg/ml) was added to a final concentration of 100 μg/ml, along with 1/100 vol. of a complete Protease Inhibition cocktail (Sigma). After 15 min at RT, the extract was sonicated 4–6 times in 30-s bursts alternating with chilling on ice to reduce viscosity, and then centrifuged at 10,000 rpm for 10 min at 4 °C. When purifying the untagged rVIM1, the supernatant was applied directly to a 10 ml bed of Q-Sepharose Fast Flow (Pharmacia Biotech, Uppsala, Sweden) pre-equilibrated with buffer A, and the column then was washed with another 20–30 ml of buffer A to elute the unbound proteins. This and all subsequent chromatography steps were carried out at RT. The flow-through fractions were pooled, and passed through a 10 ml bed of SP-Sepharose Fast Flow (Pharmacia Biotech, Uppsala, Sweden) pre-equilibrated with buffer A. The column (1 cm ID) was washed with ~25 ml buffer A, and bound rVIM1 eluted with a 200 ml linear NaCl gradient in buffer A (from 150 to 700 mM NaCl). Peak fractions, determined by Coomassie blue staining following SDS–PAGE, were pooled, concentrated to ⩽2 ml using an Ultracel® 10,000 MWCO centrifugal filter (Amicon), and then applied to a Sephacryl® S-400 gel filtration column (1 cm ID × 50 cm) equilibrated with buffer B (20 mM HEPES, 150 mM

NaCl, 1 mM DTT, 10% glycerol) at a flow rate of 30 ml/h. Under these conditions, rVIM1 elutes as a monomer relative to the added protein standards, alcohol dehydrogenase (150 kDa) and bovine serum albumin (66 kDa). At this stage, the peak protein-fractions (~2 ml/each) are ⩾95% pure, as assessed by Coomassie blue staining following SDS–PAGE (Fig. 1). The C-terminally tagged rVIM-H protein was purified directly from crude extracts using a PrepEase High-Yield Ni-chelate IDA maxi-prep cartridge (USB #78806) pre-equilibrated with buffer A. The resin was washed consecutively with 15 ml buffer A +20 mM imidazole, then with buffer A +250 mM imidazole to elute rVIM1-H, collecting 2-ml fractions. Peak fractions were pooled and chromatographed on of SP-Sepharose Fast Flow, eluted with salt and then dialyzed versus buffer B and concentrated as above Fig. 2.

The concentration of rVIM1 and rVIM1-H were determined by spectrophotometry. rVIM1 has a calculated molar extinction coefficient of 73,910 from its amino acid composition (<http://encorbio.com/protocols/Prot-MW-Abs.htm>) ignoring the contribution of the protein's 16 cysteine residues. The absorbance of a 1 mg/ml solution at 280 nm is calculated to be 1.034. The molar extinction coefficient for rVIM1-H is calculated to be 73,910; the absorbance of a 1 mg/ml solution at 280 nm is 1.020. Approximately 4 mg of purified rVIM1 is obtained from 200 ml of autoinduced cells. Based on SDS–PAGE of various fractions most loses take place because of poor retention on SP-Sepharose Fast Flow. Passing the flow-through back over the column does not significantly increase retention.

*Test DNAs*

HPLC-purified 5′ FAM (6-carboxylfluorescine)-labeled oligonucleotides with cytosine, 5mC, or 5hmC were purchased from integrated DNA technologies. By design, all oligonucleotides with one exception had a central HpaII recognition-sequence (CCGG), with the modifications occurring at the second C (underlined) in the recognition sequence. The exception was when the CpG palindrome was replaced by CAG. Purified DNAs were dissolved in 10 mM Tris–HCl, 0.01 mM EDTA buffer, pH 7.5, and, as needed, annealed at 36 μM with an equal amount of their complements, in 1× One-Phor-All buffer (10 mM Tris–acetate (pH 7.5), 10 mM Mg-acetate, 50 mM K-acetate) (GE Healthcare) by heating for 2 min at 98 °C, and thereafter, cooling slowly to room temperature. With 10% PAGE analysis, we verified their conversion into duplexes, leaving only minimal amounts (<5%) of residual single-stranded oligonucleotides. Table 1 lists the oligonucleotides we used, and gives the sequences of their resulting fully base-paired ds-cassettes. We confirmed the presence of 5hmC by incubating the DNA on appropriate ds-cassettes with purified phage-T4 β-glycosyl transferase and UDPG to convert 5hmC to 5-glycosylated-mC; this then was shown, by EMSA, to prevent the binding of *E. coli* McrA that can shift DNA that has a 5hmC containing a hemi- or fully-modified HpaII site (Dunn, unpublished). We also verified 5hmC content by showing that MspI could cleave duplexes in which the cytosine was modified, either symmetrically and asymmetrically, at the CpG site with 5mC, but not when they contained 5hmC (data not shown). HpaII could only cleave the unmodified cytosine-containing duplexes [26,27].

Following the methodology of Jin et al. [28], we generated modified base-containing synthetic 90-mer DNA fragments using two rounds of polymerase chain reaction (PCR) amplification, using the modified deoxycytidine triphosphates, 5-methyl-2′-deoxycytidine 5′-triphosphate (5mdCTP) (Fermentas; Glen Burnie, MD), and 5-hydroxymethyl-2′-deoxycytidine 5′-triphosphate (5hmdCTP) (Bioline; Taunton, MA). A single-stranded 90-mer oligonucleotide

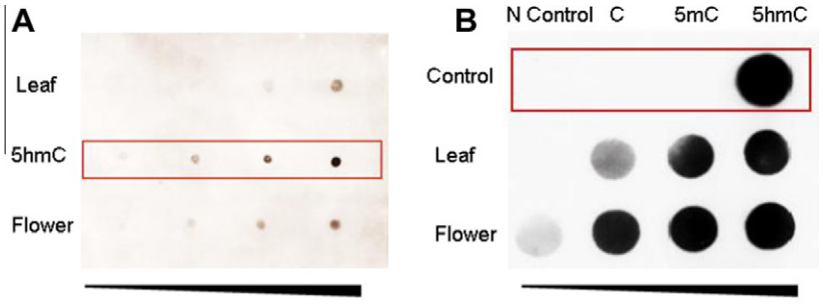


**Fig. 1.** Dot-blot detection of 5hmC in Arabidopsis DNA. (A) Dot-blot assay of synthetic DNA and Arabidopsis genomic DNA containing biotin-N3-5-gmC. Detection used avidin–horseradish peroxidase (HRP) [23]. The amount of loaded DNA increased from left to right. Detection top row: leaf DNA (25, 50, 100, 200 ng); middle row: 5hmC containing DNA standard (boxed) (1–8 ng); bottom row: flower DNA (25, 50, 100, 200 ng). (B) Dot-blot assay using antibody to 5hmC to detect 5hmC in synthetic DNAs or Arabidopsis genomic DNA. Spotted samples in top row (boxed): N Control: $H_2O$; C: Synthetic DNA with C; 5mC: Synthetic DNA with 5mC; 5hmC: Synthetic DNA with 5hmC; 0.5 ng of each synthetic DNA were loaded; middle row and bottom rows (leaf and flower DNA, respectively): The amount of Arabidopsis DNA (0.5ng, 1ng, 2ng, 4ng) increased from left to right.

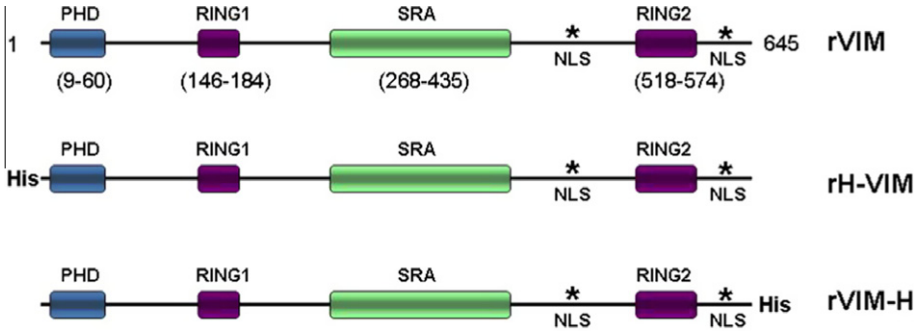


**Fig. 2.** Schematic representation of recombinant Arabidopsis VIM1-methylcytosine binding protein. The Arabidopsis genome contains five VIM genes, each of which encodes a protein containing a PHD (plant homeodomain), two RING (Really Interesting New Gene) domains, an SRA (SET and RING associated) domain) and two NLS nuclear localization signals. The amino acid coordinates for the 645 residue long VIM1 protein, excluding the $(His)_6$ affinity tag, are shown in parenthesis.

**Table 1**
Base-paired oligonucleotide cassettes used for binding studies.

| | |
|---|---|
| 5′-FAM-AGGGATATC**C**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**C**CGCCTTTCC-5′ | C/C |
| 5′-FAM-AGGGATATC**M**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**C**CGCCTTTCC-5′ | M/C |
| 5′-FAM-AGGGATATC**M**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**M**CGCCTTTCC-5′ | M/M |
| 5′-FAM-AGGGATATC**H**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**C**CGCCTTTCC-5′ | H/C |
| 5′-FAM-AGGGATATC**C**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**H**CGCCTTTCC-5′ | C/H |
| 5′-FAM-AGGGATATC**H**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**H**CGCCTTTCC-5′ | H/H |
| 5′-FAM-AGGGATATC**H**GGCGGAAAGG-3′<br>3′-TCCCTATAGG**M**CGCCTTTCC-5′ | H/M |
| 5′-FAM-AGGGATATC**C**AGGCGGAAAGG-3′<br>3′-TCCCTATAGGT**C**CGCCTTTCC-5′ | CAG/C |
| 5′-FAM-AGGGATATC**M**AGGCGGAAAGG-3′<br>3′-TCCCTATAGGT**C**CGCCTTTCC-5′ | MAG/C |

**C**=cytosine; **M**=5mC (5-methylcytosine; **H**=5hmC (5-hydroxymethylcytosine)

(5′- CCTCACCATCTCAACCAATATAATAATAAGGTTGAGGACTTTTTCC GGATGCCCGGAATGGGTTCAAAGGTATTGAGGGAGAAGTGGTGA-3′) was used as the template for 25 cycles of first-round PCR using the forward-primer 5′-CCTCACCATCTCAACCAATA-3′, and the reverse-primer 5′-TCACCACTTCTCCCTCAAT-3′. To effectively remove unmodified DNA templates from the final products, we undertook another 30 cycles of PCR amplifications using 1 μl of the first-round PCR products per 50 μl of reaction volume under the same reaction conditions [28]. The upper strand of the resulting ds-DNA has seven 5mC- or 5hmC-residues, while the complementary bottom strand has fifteen. We employed a PCR reaction with normal dCTP to prepare a control ds-90-mer without 5mC or 5hmC.

*EMSA binding-assays*

Fluorescently labeled ds-DNA probes (7.2 pmol) were incubated with increasing amounts of full-length untagged rVIM1 protein in 10 mM HEPES, pH 7.2, 75 mM NaCl, 0.5 mM DTT, 5% glycerol, 100 μg/ml bovine serum albumin, and an excess (100 ng) of poly d(I-C) competitor DNA for 30 min at room temperature before analyzing them on a 12% acrylamide gel in Tris–acetate–EDTA buffer. The non-stained gels then were photographed under a UV light.

*Fluorescence anisotrophy assays*

The binding affinities of rVIM1 and rVIM1-H for 5′-FAM-labeled ds-oligonucleotides were determined at 25 °C using an ISS fluorometer with excitation at 495 nm and emission at 530 nm. Anisotrophy values were obtained in 20 mM Tris, pH 8, 75 mM NaCl and 0.025% *n*-Dodecyl-β-D-maltoside. The anisotropy data were fitted by nonlinear least-squares regression to the following equation: $A = A_{min} + ((E + D + K_d) - ((E + D + K_d)2 - 4DE)1/2)(A_{max} - A_{min})/(2D)$, where $A$ is the anisotropy, $E$ is the total protein concentration, $D$ is the total DNA concentration, $A_{min}$ is the anisotropy of free DNA, $A_{max}$ is the anisotropy of the DNA–protein complex, and $K_d$ (nM) is the dissociation constant. We assumed a 1:1 stoichiometry for the DNA–protein complex.

*Homology modelling*

SRA domain of Arabidopsis VIM1 (amino acids 268–435) was modeled with the program MODELLER [29] using default settings as implemented in MODELLER package. The SRA domain of human UHRF1 (DNA bound complex: pdb id 3CLZ) was used as the template for the homology modelling since its sequence identity is ∼50% with the SRA domain of Arabidopsis.

## Results

*Arabidopsis DNA contains 5hmC*

To investigate whether 5hmC is present in Arabidopsis DNA, we performed two complementary dot-blot assays on sonicated (∼100–500 base-pair long) genomic DNA fragments. One involves a two-step procedure, using bacteriophage T4 β-glucosyltransferase to transfer a glucose moiety containing an azide group from UDP-6-$N_3$-glucose onto the hydroxyl group of 5hmC [23,30]. The azide group then is modified chemically with biotin so it can be detected by its interaction with avidin conjugated to horseradish peroxidase (HRP) (Bio-Rad), which then is visualized by enhanced chemiluminescence [23]. The other assay uses antibody against 5hmC or 5mC for direct dot-blot detection of these modified cytosines in the original fragments. As shown in Fig. 1a and b, both assays revealed low but measurable amounts of 5hmC, estimated to be ∼0.068–0.075% of the total cytosine nucleotides of the genome in the DNA of Arabidopsis leaves and flowers from plants grown under the conditions specified in Materials and methods. Quantification was calculated using a working curve generated by 1–8 ng of 32 bp synthetic biotin-$N_3$-5-gmC-containing DNA as described previously [23].

*Purification of recombinant VIM1*

Initially, we attempted to purify a N-terminally tagged VIM1 fusion (rH-VIM1) (Fig. 1) from induced *E. coli* lysates using commercial $Co^{2+}$ and $Ni^{2+}$ affinity supports. Surprisingly, the N-terminally tagged version, although soluble, is not efficiently retained on these columns unless chromatography is done under denaturing conditions. Under non-denaturing conditions, only a very small amount (⩽1%) of the applied recombinant protein was retained and eluted with an imidazole-containing buffer as judged by SDS–PAGE analysis of various fractions. In contrast, most of it was retained and eluted from these matrices with 250 mM imidazole when the extracts and chromatographic buffers contained 6 M urea. We interpret this finding as indicating that under non-denaturing conditions, steric- or electrostatic-interactions impair the N-terminal fusion tag's ability to interact with the supports' immobilized metal-ions. All our attempts failed to recover soluble rH-VIM1 from samples purified under denaturing conditions by removing urea by slow dialysis or rapid dilution. In contrast, we purified to near homogeneity both untagged rVIM1 and a C-terminal tagged form (rVIM1-H) from induced extracts under standard non-denaturing conditions (Fig. 3). In both cases the yield was approximately 4 mg final product from 200 ml of autoinduced cells. To our knowledge this is the first reported purification of full-length recombinant VIM1 with or without an affinity tag expressed in *E. coli*. For most of the studies reported here we used the untagged protein since we considered that it best preserved biological activity.

*Differential binding of rVIM1 to DNA substrates containing 5mC or 5hmC*

To determine the relative binding affinity of rVIM1, we used electrophoretic-mobility shift assays (EMSA) to investigate binding

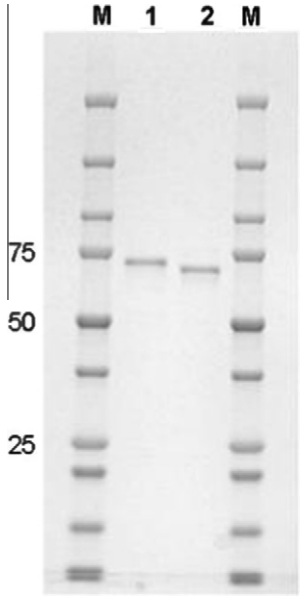


**Fig. 3.** SDS 4–20% PAGE Coomassie stained gel showing rVIM1-H and non-tagged VIM1 (rVIM1) purified from *E. coli* BL21(DE3)/pRIL cells after autoinduction of T7-based expression plasmids for 24 h. at 20 °C. Aliquots of the samples (~1.5 μg protein) were separated on a 4–20% SDS–gel and stained with Coomassie blue. Lane 1, rVIM1-H (72.45 kDa), lane 2, rVIM1 (71.48 kDa). Precision plus protein standards in kDa are indicated (M).

to DNA duplexes (Table 1) that have centrally located cytosine modifications. Although the substrate-binding specificity of several VIM1 constructs carrying various domain combinations was reported previously [16,18] to our knowledge no one has used purified, recombinant-untagged full-length protein expressed in *E. coli*. For these studies, we employed fluorescently labeled 20-base-pair-long, ds-DNAs with a singly modified CpG palindromic site in one or both of the DNA strands, or a similar 21-mer with a normal- or hemi-modified-CAG sequence (Table 1). As expected, rVIM1 bound well to our test DNA with a single hemi-methylated 5mCpG site to form protein–DNA complexes with reduced mobility in a non-denaturing polyacrylamide gel (Fig. 4, lanes 1–4). Under the EMSA conditions used here rVIM1 does not produce a well defined DNA–protein complex, however, binding to the hemi-methylated test DNA is readily evident as judged by the nearly complete disappearance of the free probe band in lane 4. This type of analysis indicates that the rVIM1 protein binds less well to unmethylated DNA (lanes 1–2), or to DNA with a fully-methylated CpG site (lanes 7–8). It

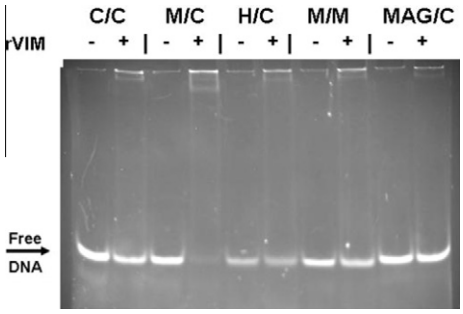


**Fig. 4.** EMSA of rVIM1 DNA binding specificity. DNA and protein–DNA complexes were fractionated by native PAGE as described in Materials and methods. The molar ratio of DNA to VIM for each reaction is ~1:6. Note that rVIM1 preferentially binds to hemi-methylated 5mCpG containing DNA. C/C = 5′-CG-3′/3′-GC-5′; M/C = 5′-5mCG-3′/3′-GC-5′; H/C = 5′-5hmCG-3′/3′-GC-5′; M/M = 5′-5mCG-3′/3′-G5mC-5′; MAG/C = 5′-5mCAG-3′/3′- GTC-5′.

also binds less well to hemi-methylated DNA when the 5mC is replaced by 5hmC (lanes 5–6). Under our conditions, it also binds poorly to DNA with a hemi-methylated 5mCpApG sequence (lanes 9–10). In other experiments (data not shown) rVIM1 bound poorly to DNA with 5hmC in both strands, or when the CpG palindromic site contains 5mC in one strand and 5hmC in the complementary strand. Overall, these binding affinities seem to be independent of strand-orientation or if the binding buffer contains mM concentrations of $Mg^{2+}$ ions. Placing a poly(His)$_6$ at the protein's C-terminus minimally impacted VIM1 binding to these test DNAs (data not shown), although its binding to the test DNAs with 5hmC seemed to increase slightly when the binding buffer contained 10 mM of $Mg^{2+}$ ions.

In subsequent experiments, we measured the binding affinities of rVIM1 to DNAs with a modified or unmodified CpG sites using fluorescence-anisotropy titrations. This technique, which reflects the tumbling rate of molecules in solution, is ideal for studying protein–DNA interactions, provided that the protein–DNA complex formed is larger than the unbound oligonucleotide, and tumbles more slowly than does the unbound oligonucleotide. In addition, since these experiments are done in solution without a subsequent separation step, they may yield true equilibrium-binding constants [31]; Fig. 5a–c illustrates some examples of the results we obtained with rVIM1. Under these conditions, i.e. [DNA] $\leqslant K_d$, the concentration of the protein that results in 50% maximal formation of a protein–DNA complex is approximately equal to $K_d$. The $K_d$ = [protein][DNA]/[protein–DNA] was measured as the concentration of VIM at which half of the maximal FA change is obtained.

Consistent with our EMSA DNA-binding data, we observed that rVIM1 binds preferentially to DNA when the target CpG site is hemi-methylated. We calculated that rVIM1 has a high affinity ($K_d \sim 15$ nM) for a test DNA with a single hemi-methylated 5mCpG site, but a much (~10-fold) lower affinity ($K_d \sim 140$ nM) for DNA with unmodified cytosines (Fig. 4B). It binds with still lower affinity ($K_d \sim 216$ nM) to test DNA with a hemi-methylated 5hmCpG site (Fig. 5C). Table 2 summarizes these data, and also shows that placing a (His)$_6$ tag at the VIM1 C-terminus had little effect on the binding of the protein to fluorescent-labeled DNA containing a hemi-methylated 5hmCpG site ($K_d \sim 28$ nM).

## Discussion

Our research shows that Arabidopsis DNA contains low but measurable amounts of 5hmC, making this newly rediscovered base one that has been identified in plant and mammals, as well as in certain bacteriophages, such as T4. In phage DNA, 5hmC, especially when it is glucosylated, acts as barrier to the degradation of the infecting viral genome by the host's nucleases [32]. Here, the substitution of 5hmC for normal cytosine occurs during viral DNA-replication and requires the phage-directed synthesis of several new enzymes in the infected bacterium [33]. In mammalian cells, 5hmC is formed after DNA replication from 5mC by a reaction with reactive oxygen species (ROS), such as superoxide anions ($\cdot O_2^-$), hydroxyl radicals (OH.), and hydrogen peroxide ($H_2O_2$); it also can be formed enzymatically by a family of 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenases (the TET proteins) [20–22,34]. These same protein(s) may act in removing 5mC to generate hypomethylated CpG sites through further TET-catalyzed oxidation of 5hmC to form 5-carboxylcytosine, followed by decarboxylation to regenerate normal cytosine [35]. Exactly how 5hmC is formed in Arabidopsis DNA is unknown, although presumably, it is generated only via some type of post-replicative mechanism. To date, no putative homologue(s) of the mammalian TET enzymes has been unambiguously identified in the Arabidopsis genome, nor has this type of cytosine modifying enzymatic

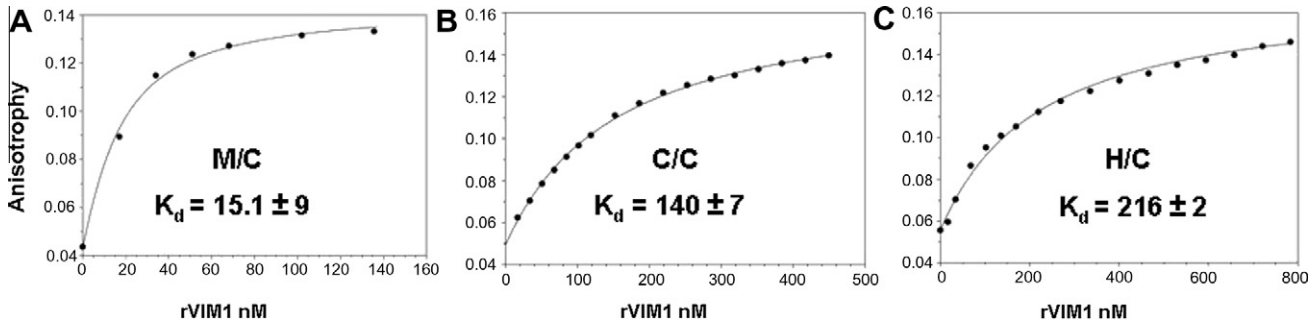


**Fig. 5.** Titration of fluorescein (FAM)-labeled DNAs. Different 5′-FAM-labeled ds-oligonucleotides were titrated with increasing amounts of rVIM1. (A) M/C = 5′-5mCG-3′/3′-GC-5′; (B) C/C = 5′-CG-3′/3′-GC-5′; (C) H/C = 5′-5hmCG-3′/3′-GC-5′. *Note:* larger amounts of rVIM1 are needed to titrate the DNAs with unmodified or hydroxymethylated CpG dinucleotides.

**Table 2**
$K_d$ (nM) values determined by fluorescence anisotrophy assays.

| Protein | DNA | $K_d$[a] |
|---|---|---|
| rVIM1 | M/C | 15.1 ± 9 |
| | C/C | 140 ± 7 |
| | M/M | 161 ± 7 |
| | H/C | 216 ± 2 |
| | H/M | 140 ± 2 |
| | H/H | 132 ± 20 |
| | MAG/C | 187 ± 2 |
| rVIM1-H | M/C | 28.5 ± 3 |

[a] Average of ⩾3 determinations.

activity been demonstrated in any plant. Clearly, further studies are needed to determine how 5hmC in Arabidopsis DNA is generated under standard growth conditions. Also unknown is the exact sequence-context in which 5hmC is present, its position relative to genes and regulatory regions, and if its level changes under different growth conditions. Efforts are underway in our lab to answer these questions.

Current thinking is that 5hmC in mammalian DNA does not merely mimic 5mC residues but likely plays a role in a yet-unknown epigenetic regulation of various biological processes or in active DNA demethylation [36–40]. In this regard, recent studies demonstrated that 5hmC significantly inhibits binding of the several mammalian methyl-DNA binding proteins that normally bind to methylated CpG sites to promote chromatin condensation, transcriptional silencing, and gene inactivation [37,38]. A similar role for this unusual epigenetic modification seems likely in Arabidopsis since, as we show herein, it adversely affects the interaction between rVIM1 and hemi-methylated 5mCpG sites in DNA. Structural data show that the homologous mammalian UHRF1 SRA-domain increases its protein–DNA interface and affinity for these sites by flipping the 5mC out of the DNA duplex into a specific binding-pocket in the protein. This change causes DNA looping, allowing the SRA domain's N-terminal region to interact with the DNA's major and minor grooves [13–15]; the cytosine residue in the daughter strand requiring modification then is flipped out of the helix and presented to the DNA methyl-transferase, DNMT1, which then adds the methyl group to the nearby cytosine in the daughter strand. During this step it is envisioned that the hemi-methylated cytosine in the template-strand returns to its more usual position [13–15].

Such a dual base-flipping mechanism forms the basis for a successive DNA-transfer model for maintaining DNA methylation in mammals by promoting direct successive interactions between UHRF1 and DNMT1 proteins at replication forks. We believe that a similar mechanism is highly likely during DNA replication in plants wherein VIM1, and perhaps other members of this protein family, aid in recruiting MET1, the maintenance methyltransferase,

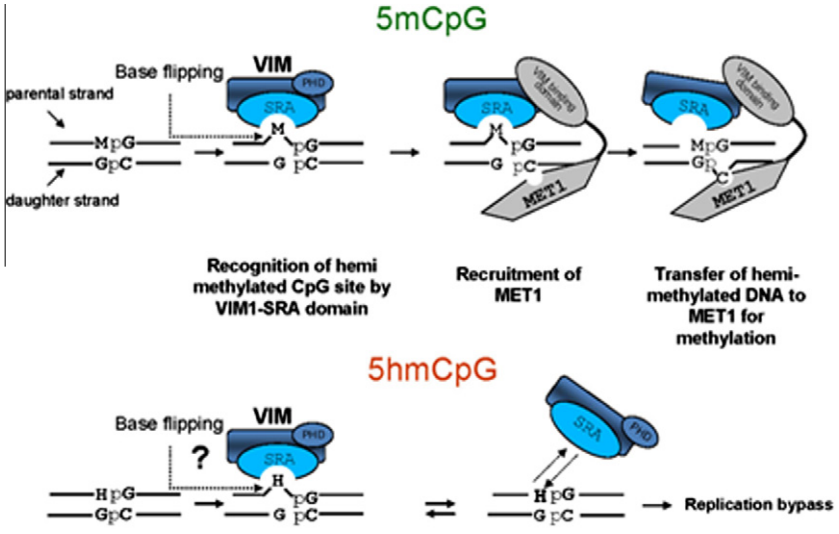


**Fig. 6.** Successive DNA transfer model for maintenance of DNA methylation by VIM1 and MET1. Structural data show that the homologous mammalian UHRF1-SRA domain increases its protein–DNA interface and affinity for hemi-methylated sites by flipping the parental strand 5mC out of the DNA duplex into a binding pocket on the protein adapted for recognizing the methyl group. As depicted in the lower portion, the presence of 5hmC is expected to allow VIM1 to disassociate more readily resulting in methylation bypass and passive loss of 5mC in the daughter stand during DNA replication.

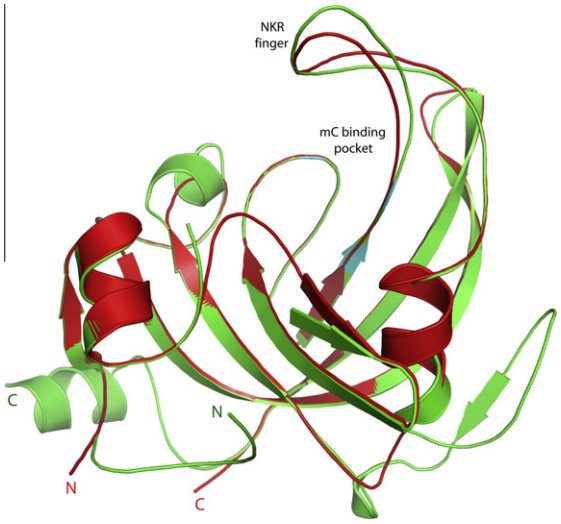


**Fig. 7.** Super-positioning of SRA domains of human UHRF1 and Arabidopsis VIM1 proteins. The root mean square deviation (RMSD) for Cα carbon atoms between the Arabidopsis (red) and human (green) SRA regions is 0.85 Å. Structurally important elements, such as the NKR finger and 5-methylcytosine (mC) binding pocket, are shown. N and C termini of the SRA domains are labeled as N and C in their respective colors.

to newly replicated DNA via an interaction between the VIM homeodomain (PHD) with MET1 (Fig. 6). Our data imply that oxidation of the hydrophobic methyl group of 5mC to a more hydrophilic and slightly larger hydroxymethyl group in 5hmC may interfere with the binding of the VIM1-SRA domain to DNA. X-ray structural data suggest UHFR1's SRA domain, and by extension that of VIM1 (Fig 7), have insufficient space in the binding pocket of SRA to assure that 5hmC is tightly bound, thereby preventing it from efficiently targeting the methylation of newly replicated DNA. As depicted in Fig. 6, HmC may be flipped, but not bound as tightly in the pocket as is mC; therefore, the HmC containing DNA would be released faster (the Kon will be the same but Koff will be shorter). If this failure is uncorrected, methylation might well be lost at these bypassed sites after one more round of DNA replication, since, as shown by Valinluck and Sowers [41], oxidative damage to 5mC causes the formation of 5hmC that prevents methylation of the cytosine in the daughter strand by purified DNMT1 (and, presumably, by MET1). Using modified ds-DNAs similar to the ones used here, Hashimoto et al. [42] showed that DMNT1 prefers hemi-methylated substrates by a factor of >60 over substrates with unmodified C or hemi-hydroxymethylated C. Related *in vitro* studies by Zahn et al. [43] have recently shown enhanced efficiency (~5-fold) for mis-incorporation incorporation of dAMP opposite 5hmC but not 5mC by RB69 gp43, a high-fidelity DNA polymerase sharing homology with human replicative DNA polymerases. This finding may explain in part the predominance of C → T transitions observed in aerobic organisms [44–46].

A limitation of our studies and similar ones is that only a subset of variations in the oligonucleotide sequence-context that might impact binding values was examined; however, a clear preference for rVIM1 binding to hemi-methylated DNA was apparent. Our data directly contrast the results of Frauer et al. [47] who recently reported that recombinant constructs of the mammalian UHRF1 protein expressed in human embryonic kidney cells, bind 5mC- and 5hmC-containing DNAs with similar affinities regardless whether one (hemi-modified) or both (fully-modified) cytosine residues at a palindromic CpG site are methylated or hydroxymethylated. However, they used intact chimeric GFP/YFP N-terminal fusion proteins. The presence of the fusion partners may have adversely influenced the fusions' biological activity since, as we discovered, an N-terminal His tag was ineffective in affinity purification of otherwise native rVIM1 protein. Overall our results are more closely in agreement with those of Hashimoto et al. [42] who demonstrate that both DNMT1 and full-length UHRF1 lose their intrinsic preference for hemi-methylated 5mCpG sites when 5mC is replaced by 5hmC.

## Conclusions

We have developed a procedure for purification of full-length recombinant Arabidopsis VIM1 protein (rVIM1) without affinity tags from induced *E. coli* cells and show that Arabidopsis genomic DNA contains low levels of 5hmC. Both EMSA and FA assays demonstrate that rVIM1 has a preference for DNA containing hemi-methylated CpGs. It dose not bind well to unmethylated or fully methylated DNA or to hemi-methylated CpApGs sites. Significantly, rVIM1 does not bind well to DNA containing hydroxymethylated CpG sites suggesting that *in vivo* oxidation of 5mC to 5hmC will prevent or drastically reduce methylation of daughter strand DNA complements which, if uncorrected, would lead to passive loss of cytosine methylation following Arabidopsis DNA replication.

## Authors' contributions

Q.Y. together with C.X.S. and C.H. designed and carried out all of the genomic 5hmC studies and analyzed the data. D.K. constructed the computer overlays of the human UHRF1 and VIM1 SRA domains. J.J.D. conceived the VIM experiments, coordinated this aspect of the project and drafted the manuscript. All authors read and approved the final manuscript.

## Conflict of interest statement

None declared.

## Acknowledgments

This work was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. We gratefully acknowledge the technical assistance of Eileen Kasmarcik, Barbara Lade, Laura-Li Loffredo, Judi Romeo, and Vito Graziano.

## References

[1] Y. Gruenbaum, R. Stein, H. Cedar, A. Razin, Methylation of CpG sequences in eukaryotic DNA, FEBS Lett. 124 (1981) 67–71.
[2] R.L. Sinsheimer, The action of pancreatic deoxyribonuclease II. Isomeric dinucleotides, J. Biol. Chem. 215 (1955) 579–583.
[3] S.W. Chan, I.R. Henderson, S.E. Jacobsen, Gardening the genome: DNA methylation in Arabidopsis thaliana, Nat. Rev. Genet. 6 (2005) 351–360.
[4] I.R. Henderson, S.E. Jacobsen, Epigenetic inheritance in plants, Nature 447 (2007) 418–424.
[5] B.H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A.P. Bird, R. Jaenisch, Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, Proc. Natl Acad. Sci. USA 97 (2000) 5237–5242.
[6] D. Zilberman, M. Gehring, R.K. Tran, T. Ballinger, S. Henikoff, Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription, Nat. Genet. 39 (2007) 61–69.
[7] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.
[8] M.G. Goll, T.H. Bestor, Eukaryotic cytosine methyltransferases, Annu. Rev. Biochem. 74 (2005) 481–514.
[9] S. Pradhan, A. Bacolla, R.D. Wells, R.J. Roberts, Recombinant human DNA (cytosine-5) methyltransferase I. Expression, purification, and comparison of de novo and maintenance methylation, J. Biol. Chem. 274 (1999) 33002–33010.
[10] M. Bostick, J.K. Kim, P.O. Esteve, A. Clark, S. Pradhan, S.E. Jacobsen, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, Science 317 (2007) 1760–1764.
[11] R. Hopfner, M. Mousli, J.M. Jeltsch, A. Voulgaris, Y. Lutz, C. Marin, J.P. Bellocq, P. Oudet, C. Bronner, ICBP90, a novel human CCAAT binding protein, involved in

the regulation of topoisomerase II alpha expression, Cancer Res. 60 (2000) 121–128.
[12] J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T.A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, Nature 450 (2007) 908–912.
[13] K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura, M. Shirakawa, Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism, Nature 455 (2008) 818–821.
[14] G.V. Avvakumov, J.R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C.H. Arrowsmith, S. Dhe-Paganon, Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, Nature 455 (2008) 822–825.
[15] H. Hashimoto, J.R. Horton, X. Zhang, M. Bostick, S.E. Jacobsen, X. Cheng, The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix, Nature 455 (2008) 826–829.
[16] L.M. Johnson, M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis, S.E. Jacobsen, The SRA methyl-cytosine-binding domain links DNA and histone methylation, Curr. Biol. 17 (2007) 379–384.
[17] E. Kraft, M. Bostick, S.E. Jacobsen, J. Callis, ORTH/VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases, Plant J. 56 (2008) 704–715.
[18] H.R. Woo, O. Pontes, C.S. Pikaard, E.J. Richards, VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization, Genes Dev. 21 (2007) 267–277.
[19] H.R. Woo, T.A. Dittmer, E.J. Richards, Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis, PLoS Genet. 4 (2008) e1000156.
[20] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, Science 324 (2009) 929–930.
[21] M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, A. Rao, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, Science 324 (2009) 930–935.
[22] S. Ito, A.C. D'Alessio, O.V. Taranova, K. Hong, L.C. Sowers, Y. Zhang, Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification, Nature 466 (2010) 1129–1133.
[23] C.X. Song, K.E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T.J. Looney, B. Zhang, L.A. Godley, L.M. Hicks, B.T. Lahn, P. Jin, C. He, Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine, Nat. Biotechnol. 29 (2011) 68–72.
[24] E.A. Mulligan, J.J. Dunn, Cloning, purification and initial characterization of *E. coli* McrA, a putative 5-methylcytosine-specific nuclease, Protein Expr. Purif. 62 (2008) 98–103.
[25] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234.
[26] S.M. Kinney, H.G. Chin, R. Vaisvila, J. Bitinaite, Y. Zheng, P.O. Esteve, S. Feng, H. Stroud, S.E. Jacobsen, S. Pradhan, Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes, J. Biol. Chem. 286 (2011) 24685–24693.
[27] K. Ichiyanagi, Inhibition of MspI cleavage activity by hydroxymethylation of the CpG site: a concern for DNA modification studies using restriction endonucleases, Epigenetics 7 (2012) 131–136.
[28] S.G. Jin, S. Kadam, G.P. Pfeifer, Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine, Nucleic Acids Res. 38 (2010) e125.
[29] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.
[30] A. Szwagierczak, S. Bultmann, C.S. Schmidt, F. Spada, H. Leonhardt, Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA, Nucleic Acids Res. 38 (2010) e181.
[31] B.J. Anderson, C. Larkin, K. Guja, J.F. Schildbach, Using fluorophore-labeled oligonucleotides to measure affinities of protein–DNA interactions, Methods Enzymol. 450 (2008) 253–272.
[32] C.L. Bair, L.W. Black, A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs, J. Mol. Biol. 366 (2007) 768–778.
[33] J.G. Flaks, S.S. Cohen, Virus-induced acquisition of metabolic function. III. Formation and some properties of thymidylate synthetase of bacteriophage-infected *Escherichia coli*, J. Biol. Chem. 234 (1959) 2981–2986.
[34] C. Loenarz, C.J. Schofield, Oxygenase catalyzed 5-methylcytosine hydroxylation, Chem. Biol. 16 (2009) 580–583.
[35] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–1303.
[36] K.E. Szulwach, X. Li, Y. Li, C.X. Song, J.W. Han, S. Kim, S. Namburi, K. Hermetz, J.J. Kim, M.K. Rudd, Y.S. Yoon, B. Ren, C. He, P. Jin, Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells, PLoS Genet. 7 (2011) e1002154.
[37] S.G. Jin, X. Wu, A.X. Li, G.P. Pfeifer, Genomic mapping of 5-hydroxymethylcytosine in the human brain, Nucleic Acids Res. 39 (2011) 5015–5024.
[38] M. Munzel, D. Globisch, T. Carell, 5-Hydroxymethylcytosine, the sixth base of the genome, Angew. Chem. Int. Ed. Engl. 50 (2011) 6460–6468.
[39] P. Hajkova, S.J. Jeffries, C. Lee, N. Miller, S.P. Jackson, M.A. Surani, Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway, Science 329 (2010) 78–82.
[40] J.U. Guo, Y. Su, C. Zhong, G.L. Ming, H. Song, Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain, Cell 145 (2011) 423–434.
[41] V. Valinluck, L.C. Sowers, Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1, Cancer Res. 67 (2007) 946–950.
[42] H. Hashimoto, Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, X. Zhang, X. Cheng, Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic Acids Res. (2012), http://dx.doi.org/10.1093/nar/gks155.
[43] K.E. Zahn, A. Averill, S.S. Wallace, S. Doublie, The miscoding potential of 5-hydroxycytosine arises due to template instability in the replicative polymerase active site, Biochemistry 50 (2011) 10350–10358.
[44] R.M. Schaaper, R.L. Dunn, Spontaneous mutation in the *Escherichia coli* lacI gene, Genetics 129 (1991) 317–326.
[45] M.E. Huang, A.G. Rio, A. Nicolas, R.D. Kolodner, A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations, Proc. Natl Acad. Sci. USA 100 (2003) 11529–11534.
[46] S. Nakken, E.A. Rodland, E. Hovig, Impact of DNA physical properties on local sequence bias of human mutation, Hum. Mutat. 31 (2010) 1316–1325.
[47] C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M.C. Cardoso, I. Antes, H. Leonhardt, Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain, PLoS ONE 6 (2011) e21306.