

The *Arabidopsis* SUV4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9

Tage Thorstensen, Andreas Fischer¹, Silje V. Sandvik, Sylvia S. Johnsen, Paul E. Grini, Gunter Reuter¹ and Reidunn B. Aalen*

Department of Molecular Biosciences, University of Oslo, P.O. Box 1041 Blindern, N-0316 Oslo, Norway and ¹Institute of Genetics, Biologikum, Martin Luther University Halle, Halle, Germany

Received May 5, 2006; Revised September 4, 2006; Accepted September 7, 2006

ABSTRACT

Proteins containing the evolutionarily conserved SET domain are involved in regulation of eukaryotic gene expression and chromatin structure through their histone lysine methyltransferase (HMTase) activity. The *Drosophila* SU(VAR)3-9 protein and related proteins of other organisms have been associated with gene repression and heterochromatinization. In *Arabidopsis* there are 10 SUVH and 5 SUV4 genes encoding proteins similar to SU(VAR)3-9, and 4 SUVH proteins have been shown to control heterochromatic silencing by its HMTase activity and by directing DNA methylation. The SUV4 proteins differ from the SUVH proteins in their domain structure, and we show that the closely related SUV1, SUV2 and SUV4 proteins contain a novel domain at their N-terminus, and a SUV4 specific region preceding the SET domain. Green fluorescent protein (GFP)-fusions of these SUV4 proteins preferably localize to the nucleolus, suggesting involvement in regulation of rRNA expression, in contrast to other SET-domain proteins studied so far. A novel HMTase specificity was demonstrated for SUV4, in that monomethylated histone H3K9 is its preferred substrate *in vitro*.

INTRODUCTION

The organization of DNA into higher order chromatin structure is crucial for the correct temporal and spatial regulation of gene expression in most eukaryotic organisms. Chromatin is a dynamic DNA–protein structure that can exist as either transcriptionally permissive euchromatin or repressive heterochromatin. The difference between the two states is partly

due to different combinations of covalent post-translational modifications of the histones, including phosphorylation, acetylation, ubiquitination, ADP ribosylation and methylation (1). These modifications, constituting the so called histone code, may be interdependent, and create binding sites for chromatin-associated effector proteins (2,3) facilitating or restricting transcription.

Two of the best studied protein families responsible for histone-tail modifications are the histone acetyltransferases (HATs) and the histone deacetylases (HDACs) that function in multiprotein complexes. The HATs promote gene activation by interaction with transcriptional activators and acetylation of the conserved lysine (K) residues on the histone tails (4), while the HDACs perform transcriptional repression by deacetylation at sites targeted by transcriptional repressors (5). Lysine residues may also be modified by methylation, which can stimulate or repress transcription depending on the position of the methylated residues (6). In general, methylation of histone H3K9, H3K27 and H4K20 in combination with histone hypoacetylation and DNA methylation, is associated with heterochromatin and gene silencing. Euchromatin on the other hand, contains elevated levels of histone H3 lysine methylation at positions 4, 36 and 79 as well as hyperacetylation of histone H4 (7).

The ability to methylate lysine residues on the histone tails resides in proteins containing the evolutionarily conserved 130 amino acid SET domain named after the three *Drosophila* proteins SUPPRESSOR OF VARIATION 3-9 [SU(VAR)3-9], ENHANCER OF ZESTE [E(Z)] and TRITHORAX (TRX) (8). In the *Arabidopsis thaliana* genome there are at least 29 actively transcribed genes encoding SET-domain proteins, that can be divided into four major evolutionarily conserved classes (9,10). The diversity of these proteins suggests that they exert specific functions during *Arabidopsis* development. For example, MEDEA (MEA) is a Polycomb group protein homologous to E(Z) (9), which in *Arabidopsis* regulates seed development after fertilization (11,12). The ATX1 protein, a homologue to TRX, positively

*To whom correspondence should be addressed. Tel: +47 22857297; Fax: +47 22856041; Email: reidunn.aalen@imbv.uio.no

Present address:

Sylvia S. Johnsen, Institute of Medical Biology, University of Tromsø, 9037 Tromsø, Norway

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

regulates the expression of several flower homeotic genes in *Arabidopsis*, and *in vitro* results suggest that the protein is a histone H3K4 methyltransferase (13). The protein product of the *ASHH2* gene (At1g77300), similar to *Drosophila ASH1*, is involved in control of flowering time and may act as an H3K4 and/or H3K36 HMTase (14,15).

The two SU(VAR)3-9 homologues KRYPTONITE (KYP)/SUVH4 and SUVH2 have been studied in more detail and both have been shown to control heterochromatic H3K9 dimethylation and to function *in vivo* as heterochromatin-specific H3K9 HMTases (16–19). The most severe effects of this histone methylation mark were found in *suvh2* null mutants, and a dosage dependent effect of SUVH2 on heterochromatic gene silencing has been demonstrated (18). Two other SUVH proteins, SUVH5 and SUVH6, also show *in vitro* H3K9 HMTase activity and together with SUVH4 these HMTases control non-CpG DNA methylation and gene silencing at heterochromatic loci (16,20). Therefore, all data available strongly suggest that proteins belonging to the SU(VAR)3-9 group of SET-domain proteins are involved in multiple controls of heterochromatic H3K9 methylation and gene silencing in *Arabidopsis*.

In addition to nine active *SUVH* genes, there are five *SU(VAR)3-9* related *SUVR* genes in *Arabidopsis*. The encoded proteins have a SET domain with pre- and post-SET domains most similar to the SU(VAR)3-9 group, but are lacking the YDG domain of the SUVH proteins (9) that appears to be involved in directing DNA methylation to target sequences (18). To elucidate whether *SUVR* proteins differ in function from *SUVH* proteins we have focused the present work on *SUVR1*, *SUVR2* and *SUVR4* that constitute a subgroup amongst the *SUVR* proteins (9,10). We present their particular domain structure and splice variants. In contrast to other SET-domain proteins these *SUVR* proteins were mainly found localized in the nucleolus or nuclear bodies, and we have identified a short amino acid sequence that can direct proteins to the nucleolus. The recombinant *SUVR4* protein, but not *SUVR1* and *SUVR2*, has HMTase activity with specificity for monomethylated H3K9 *in vitro*.

MATERIALS AND METHODS

Bioinformatics

Database searches were performed using BLASTP, TBLASTN and PSI-BLAST against the nr and est databases at GenBank, or the est database at TIGR. Protein sequences or translated coding sequences (CDS) were aligned with the ClustalX program (<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html>) and manually adjusted with GeneDoc (<http://www.psc.edu/biomed/genedoc/>). The proteins or translated CDS were analysed for known motifs and domains with the InterProScan (<http://www.ebi.ac.uk/InterProScan/>) and MotifScan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) tools. Putative nuclear localization signals (NLS) were identified by using the PredictNLS tool (<http://cubic.bioc.columbia.edu/predictNLS/>).

Isolation of nucleic acids and RT-PCR

Nucleic acids were isolated from wild-type (wt) *Arabidopsis* ecotype Columbia. The AquaPure genomic DNA isolation kit

(BioRad) was used for DNA isolation. Using 1 µg total RNA extracted from 100 mg plant tissue with the RNeasy Plant Mini kit (Qiagen), first strand cDNA was synthesized with Superscript III Reverse Transcriptase (Invitrogen) and oligo dT primers and used for RT-PCR. Control reactions were run without Reverse Transcriptase.

DNA constructs

A fragment containing the C.1 Gateway cassette (Invitrogen) and the smRS-green fluorescent protein (GFP) was isolated from the pKGAW-smRSGFP vector (21), and ligated into the XhoI digested and blunted glucocorticoid-inducible pTA7002 vector (22) to create the pTA7002GWsmRSGFP vector. The *SUVR* CDS were amplified from mRNA from flowers and buds, with the following Gateway primers: splice variant *SUVR1a*—5'-attB1-ATATGAGAATAACTGGGAT-TTCATTG and 5'-attB2-CTTTGTTCAAAATCTGCATGG (SUVR1 GAW R); splice variant *SUVR1b*—5'-attB1-ATATGGATGAAGATGAATTTCCATTG and SUVR1 GAW R; splice variant *SUVR4a*—5'-attB1-ACGACGCAGTGAAA-CAGAGA and 5'-attB2-ATTTGCGCTTTTTAGACACCTC; and splice variant *SUVR2a*—5'-attB1-AATTTACCTGG-CACTGTCC and 5'-attB2-ATGCTCGCTTCTTCACATTC. The CDS were recombined by Gateway technology first into pDONR207, and then into pTA7002GWsmRSGFP to give the pTA7002-SUVR1a-GFP, pTA7002-SUVR1b-GFP, pTA7002-SUVR4a-GFP and pTA7002-SUVR2a-GFP constructs, respectively.

SUVR fragments generated to identify nucleolar localization signals were PCR amplified from sequenced full-length clones using the following Gateway primers: S1-187—5'-attB1-AATCTAGAATATGAGAATAACTGGGATTTTCATTG and 5'-attB2-TTATGCTACACTTTCTCTGGACTTC; S1-295—5'-attB1-AATCTAGAATATGAGAATAACTGGGATTTTCATTG and 5'-attB2-TTATCCACGTCTACTGCGCA-AC; S4-1535—5'-attB1-GTATGATCAGTCTCTCCGGACT and 5'-attB2-TTAGCCTTGAGATCCTTTTTATTTTTCTG; and S4-NLS—5'-attB1-TCATTTGCGCTTTTTAGACACCTC and 5'-attB2-CTCAAGGCAAGTCTATAG. The PCR products were recombined into pDONR/Zeo and subsequently in frame with an N-terminal EGFP gene in the Gateway destination vector pK7WGF2 (23).

HMTase and GST pull-down constructs were generated as follows: the *SUVR1* SACSET sequence, which include the pre-SET, SET and post-SET domains, was PCR amplified with Pfx (Invitrogen) from cDNA using the primers 5'-GGATCCGAAAGTGGTGAGTTGGCATT and 5'-CTCGAGTAGCCTCTCATGCTTTGTTCA where the underlined sequences represent the restriction endonuclease sites for BamHI and XhoI, resulting in a 1077 bp fragment encoding residues 332 to 688 in the *SUVR1* protein. The *SUVR2* SACSET fragment was made by PCR amplification of a 1054 bp cDNA sequence representing residues 348 to 697 of *SUVR2* using the primers 5'-GGATCCGTTGGTGATTCCATGGCTTT and 5'-CTCGAGCTCATGCTCGCTTCTTCA. The *SUVR1* SACSET and *SUVR2* SACSET PCR products were digested with XhoI and BamHI and ligated into pGEX-AB. The *SUVR4* CDS without the small first exon was PCR amplified with *Pfu* using the Gateway primers 5'-attB1-GTATGATCAGTCTCTCCGGACT and 5'-attB2-ATTTGCGCTTTTTAGACACCTC and recombined

into pDONR/Zeo creating the pDONR/Zeo-S4-5UTR construct. This construct was recombined into pGEX-AB GAW creating the clone pGEX-S4-5UTR.

Mutated versions of pGEX-S4-5UTR, pGEX-S4W405Y and pGEX-S4W405F, were created using the QuikChange mutagenesis kit (Stratagene) with the primers 5'-GGCCATGGATGAGTTGACATTCGATTACATGATAGACTTCAATG and 5'-CATTGAAGTCTATCATGTAATCGAATGTCAACTCATCCATGGCC; and 5'-GGCCATGGATGAGTTGACATACGATTACATGATAGACTTCAATG and 5'-CATGGAAGTCTATCATGTAATCGTATGTCAACTCATCCATGGCC, respectively.

Transgenic plants

Arabidopsis plants, ecotype Columbia, were grown under long day greenhouse conditions at 20°C. Transgenic *Arabidopsis* plants were generated by the floral dip method (24), using the *Agrobacterium tumefaciens* strain C58 pCV2260. Transgenic plants containing glucocorticoid inducible pTA7002-SUVR-smRSGFP constructs were selected on MS-2 medium (1× Murashige and Skoog salts, 0.05% 2-N-morpholino/ethanesulfonic acid, 2% sucrose, 0.8% agar) containing 15 µg/ml hygromycin. SUVR-GFP expression was induced by growing transgenic plants on agar plates containing 5 µM dexamethasone (22). Plants containing the pK7WGF2 constructs were selected on MS-2 containing 50 µg/ml kanamycin.

FISH and documentation of GFP localization

Roots of transgenic plants that expressed GFP-fusion proteins were chopped and fixed in 4% formaldehyde in phosphate-buffered saline (PBS) on a glass slide, covered with cover slip and squashed. After freezing in liquid N₂, the cover slip was removed, and the glass slide was transferred to PBS.

FISH was performed as described previously (25) on young rosette leaves expressing SUVR1-GFP. The 18S rDNA probe was PCR amplified using the primers 5'-CTGCCCGTTGCTCTGATGATTCATG and 5'-CAATAAGACCAGGAGCGTATCG and then subcloned into the pGEM vector. The probe was then DIG labelled by PCR with the DIG labelling kit (Roche Diagnostics, Mannheim). A total of 2 µl of the PCR was added to 30 µl hybridization mix [50% formamide, 2× SSC, 50 mM sodium phosphate (pH 7.0), 10% dextran sulphate, 5 µg salmon-sperm DNA] and hybridized in a wet chamber for ~14 h. The DIG labelled probe was detected with sheep anti-DIG (1:50 in 4 M buffer, Roche Diagnostics) followed by a rhodamine conjugated rabbit anti-sheep antibody (1:100, Abcam). Immunodetection of GFP was done as described (18) using a mouse-anti-GFP antibody (1:50, Molecular Probes) followed by an Alexa 488 conjugated goat-anti-mouse antibody (1:100, Molecular Probes).

All preparations were counterstained in DAPI (2 µg/ml) and inspected with a Zeiss Axiovision2 microscope equipped with epifluorescence attachment.

Protein expression, GST pull-down and histone methyltransferase assay

For histone methyltransferase assays and GST pull-down, recombinant proteins were expressed in BL21 cells, solubilized in modified RIPA buffer [20 mM Tris (pH 7.7), 150 mM NaCl, 1% NP-40 Protease Cocktail EDTA-free

(Pierce) and 0.25 mg/ml lysozyme] and immobilized on glutathione sepharose beads (Amersham).

GST pull-down was done according to (26) and the *in vitro* HMTase assay essentially performed as described in (19,27) using 10 µg of matrix-bound GST-SUVR proteins and 5–10 µg of histones from calf thymus (Roche), recombinant histone H3 (Upstate) and methylated histone H3 peptides mono- or dimethylated at K4, K9 or K27 (Upstate or Abcam). The presence of bound core histones in the pull-down assay was confirmed by Coomassie staining or western blotting using antibodies against dimethylated histone H3K9 (1:1000, Upstate #07-212). Peptides from the *in vitro* HMTase assay were transferred to nitrocellulose membranes and immunodetected with α-dimethyl-H3K9 (1:1000, from Thomas Jenuwein's lab) to confirm specificity of the SUVR4 activity. Detection of primary antibody was performed with peroxidase-conjugated secondary antibodies (1:2000, Abcam) using the ECL kit (Amersham).

Sequences and accession numbers

The *SUVR1* (At1g04050) gene has two different splice variants: *SUVR1a* (AF394239, 2506 bp) and *SUVR1b* (2629 bp). In addition, there is a predicted splice variant that when translated contains the WIYLD domain in the N-terminus (AAD10665). *SUVR2* (At5g43990) has three different splice variants *SUVR2a* (AY045576, 2568 bp), *SUVR2b* (2508 bp) and *SUVR2c* (NM_203151, 2595 bp). *SUVR4* (At3g04380) has two splice variants *SUVR4a* (2004 bp) and *SUVR4b* (AF408062, 2085 bp).

The sequences used in the alignments (Figure 3 and Supplementary Figure 1) have the following accession numbers referring to the GenBank protein database at NCBI (<http://www.ncbi.nlm.nih.gov/>): *SUVR1* (AAD10665), *SUVR2* (AAK92218), *SUVR4* (NP_974217), *SUVH4* (Q8GZB6), *SUVH6* (AAK28971), *SUVR2_Os* (XP_466798), *G9a_Hs* (Q96KQ7). The TC170256_Ta sequence in Figure 3B, and the TC166832_Le and TC206578_Ta sequences in Supplementary Figure 1, refer to the assembled est sequences at TIGR (<http://www.tigr.org/>). The accession numbers in Figure 3B refer to the GenBank nucleotide database.

RESULTS

The *SUVR* transcripts are subject to alternative splicing

The *SUVR1*, *SUVR2* and *SUVR4* genes were originally identified using BLAST searches with SET-domain sequences from *Drosophila* proteins against the *Arabidopsis* genomic database (9). To investigate their expression patterns more in detail, RT-PCR was performed using RNA from young roots, seedlings, rosette leaves, inflorescences and green full-grown siliques. This analysis demonstrated highly similar expression patterns for the three genes, with expression in all tissues examined; strongest in inflorescences, weakest in leaves and relatively weaker expression in roots than in seedling (Figure 1A). The ubiquitous expression pattern suggests that these *SUVR* genes are of importance during the whole life cycle of the plant.

RT-PCR, cloning of long cDNA sequences using primers designed for the 5'- and 3'-untranslated regions (5'- and 3'-UTRs), and additional available GenBank sequences revealed

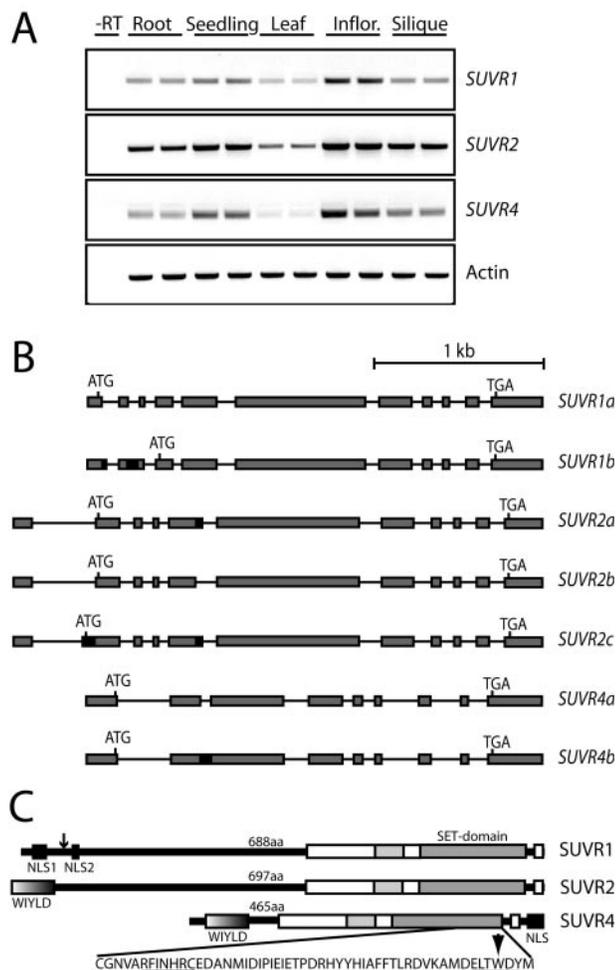


Figure 1. *SUVR1*, 2 and 3 transcripts and encoded proteins. (A) RT-PCR analysis of *SUVR* transcript levels in wild-type tissues as indicated, using gene specific primers. Two parallels are shown for each tissue. Actin was used as a positive control and -RT is a negative control without Reverse Transcriptase. (B) Schematic presentation of *SUVR* splice variants. Grey boxes, exons; black boxes, alternatively spliced exons; lines, introns. Positions of start and stop codons are indicated. (C) Domain architecture of *SUVR1*, *SUVR2* and *SUVR4* proteins. The amino acid sequence of the C-terminal part of *SUVR4*'s SET domain is given. Motifs shown in other SET-domain proteins to be important for HMTase activity (27) are underlined, and a residue of importance for product specificity (28) is indicated by an arrowhead. The arrow indicates the start of the *SUVR1b* splice variant. Boxes in black, NLS; white, pre-SET and post-SET domains; light grey, *SUVR*-specific pre-SET region; dark grey, SET domain; gray gradient, WYLD domain.

that there for all three genes were transcripts with alternative splicing at the N-terminus (Figure 1B). The *SUVR1* gene expressed two splice variants in flowers and buds with alternative start codons. The *SUVR1a* transcript consisted of 11 exons, with the putative start codon in exon 1, giving rise to an open reading frame (ORF) of 2064 bp (688 amino acids, Figure 1C). The *SUVR1b* transcript contained an alternative GT donor site in the first intron, resulting in a 39 bp longer first exon. In addition, the second intron, which has a stop codon in the last triplet in frame with the rest of the ORF, was retained in this transcript. If the *SUVR1b* transcript is to be translated into a SET-domain containing protein, an alternative start codon in exon 4 must be used, giving an

ORF size of 1884 bp (628 amino acids, Figure 1C). This splice variant was found in five independent clones amplified from first strand cDNA generated by Reverse transcriptase primed with oligo dT. Thus, it was unlikely that these clones represented an incompletely processed transcript.

We identified three *SUVR2* transcripts (Figure 1B): *SUVR2a* contained 11 exons with an ORF of 2151 bp (717 amino acids). A splice variant with an alternative GT donor site reduced the length of exon 5 by 60 bp, resulting in the *SUVR2b* transcript with an ORF of 2091 bp (697 amino acids, Figure 1C). A third transcript (*SUVR1c*) revealed an alternative AG acceptor site in the first intron which extended the second exon by 74 bp, and would add 23 residues to the N-terminus of the translation product compared to the other splice variants.

For the *SUVR4* gene at least two different transcripts were expressed (Figure 1B). *SUVR4a* contained 9 exons with an ORF of 1395 bp (465 amino acids, Figure 1C). An alternative splice variant, *SUVR4b*, had retained the 81 bp second intron, which can be translated in frame without disrupting the ORF, thereby resulting in an ORF of 1476 bp (492 amino acids).

The SUVR proteins localize to the nucleolus

To identify the subcellular localization of the *SUVR* proteins and the effect of ectopic *SUVR* expression, *Arabidopsis* plants were transformed with glucocorticoid-inducible *SUVR*-smRSGFP (*SUVR*-GFP) fusion constructs (see Materials and Methods). After induction with dexamethasone, plants expressing the *SUVR*-GFP fusion proteins were inspected by fluorescence microscopy. No GFP was detected in any non-induced plants, and induced plants with high *SUVR*-GFP expression did not display any visible aberrant morphological phenotype. For each fusion construct the same subnuclear localization was seen in aboveground tissues and in roots. The localization was documented in roots (Figure 2) as visualization of the GFP signal was better in cells without chlorophyll.

Plants expressing the *SUVR1a*-GFP splice variant showed a strong GFP signal in the nucleolus and a very weak signal in the nucleoplasm (Figure 2A). In DAPI-stained nuclei, the nucleolus appears as a black hole (Figure 2A, left, arrow). FISH, performed with a probe against the 18S rDNA repeats, was used as a control to confirm the specific localization to the nucleolus. This clearly demonstrated that the *SUVR1a*-GFP protein (green) associated with the nucleolus and did not overlap with the heterochromatic nucleolus organizing regions (NOR) detected by the 18S rDNA probe (red) (Figure 2A, insert). In plants transformed with the *SUVR1b*-GFP construct, strong GFP expression was in contrast detected in the nucleoplasm, and the protein was excluded from the nucleolus and the densely DAPI-stained heterochromatin (Figure 2B). The differences in subnuclear localization between the two splice variants indicate that *SUVR1* distribution is regulated by alternative splicing.

SUVR2a-GFP expression was mainly seen in subdomains associated with or within the nucleolus, and a weaker signal was present in the nucleoplasm (Figure 2C). Additionally, a variable number of nuclear bodies of unequal size were observed in *SUVR2a*-GFP expressing cells, suggesting specific association with subnuclear regions. Notably, these

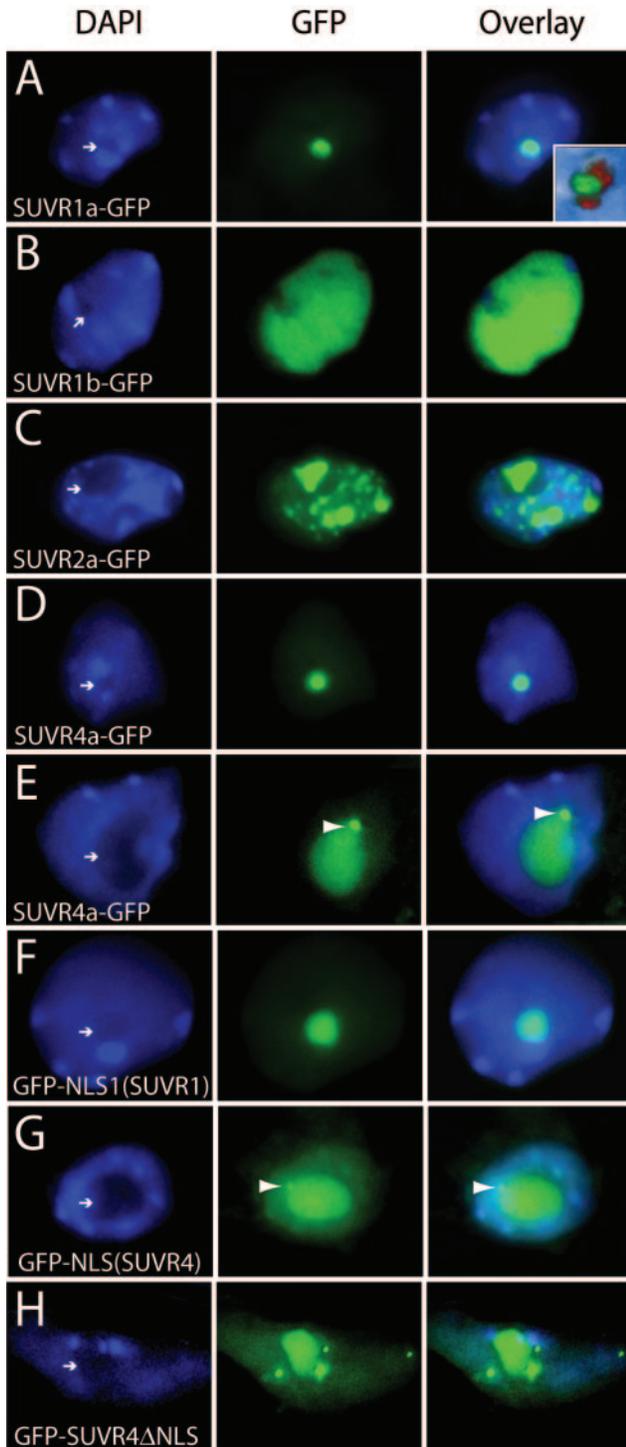


Figure 2. Subnuclear localization of the SUVR proteins. Root cells from *Arabidopsis* seedlings expressing the indicated SUVR-GFP fusion proteins (green) were investigated using fluorescence microscopy. Nuclei were stained with DAPI (blue) to visualize heterochromatin (left panels). The nucleolus is indicated by an arrow, and spots at the edge of the nucleolus with an arrowhead. (A) SUVR1a-GFP. The insert in the right panel shows, as a control, the nucleolus of a cell expressing SUVR1a-GFP (green) exposed to FISH with an 18S rDNA probe (red). (B) SUVR1b-GFP. (C) SUVR2a-GFP. (D) SUVR4a-GFP localization pattern found in the majority of cells inspected. (E) Alternative subnuclear localization of SUVR4a-GFP. (F) GFP fused to the N-terminus (40 amino acid) of SUVR1a. (G) GFP fused to the C-terminus (27 amino acid) of SUVR4. (H) GFP fused to SUVR4 without the C-terminus.

nuclear bodies showed weaker DAPI staining than the rest of the nucleoplasm, and did not overlap with the DAPI-stained chromocenters.

Like SUVR1a-GFP, SUVR4a-GFP displayed a very specific localization, with uniform expression in the whole nucleolus in most of the nuclei investigated and much weaker expression in the nucleoplasm (Figure 2D). However, in some nucleoli the GFP signal was not covering the whole nucleolus, but seen as a smaller spot inside the nucleolus (data not shown), while for other nuclei the signal was an intermediate of the two preceding situations. In these nuclei, GFP was observed in the whole nucleolus, with a stronger signal in a spot at the edge of the nucleolus (Figure 2E).

A short amino acid sequence from SUVR1 can function as a nucleolus localization signal

Putative NLS were predicted to be present in both the SUVR1 and SUVR4 proteins by the PredictNLS server. The SUVR4a protein has a potential NLS with 21 residues at the C-terminus, while the SUVR1a sequence contains two putative NLS, a 17 amino acid long motif in position 16 (NLS1), and a short signal (NLS2) with 7 residues in position 66 (Figure 1C). The SUVR1b splice variant contains only NLS2. The difference in subnuclear localization between SUVR1a and SUVR1b suggested that the longer N-terminal fragment (58 residues) of SUVR1a containing NLS1 was responsible for the nucleolar targeting. Interestingly, NLS1 with flanking residues has 37% sequence identity to the predicted NLS motif of the SUVR4 protein (Figure 3A), which also has nucleolar localization.

To test whether these NLS motifs with associated residues were involved in targeting SUVR1a and SUVR4a to the nucleolus, 35S::GFP-fusion constructs were made containing, (i) the first 176 amino acids of SUVR1 including both NLS1 and NLS2 (pK7WGF2-S1-295), (ii) the first 40 amino acids of SUVR1 including NLS1 (pK7WGF2-S1-187), (iii) the last 27 amino acids of SUVR4 including the NLS (pK7WGF2-S4-NLS) and (iv) SUVR4 without the terminal NLS (pK7WGF2-S4-1535) (cfr. Figure 1C). Roots from *Arabidopsis* plants transformed with these constructs were inspected for subcellular and subnuclear localization of the GFP signal. GFP was not found in the cytoplasm for any of the constructs. The 176 amino acid fragment (data not shown) and also the 40 amino acid fragment of SUVR1a could direct GFP to the nucleolus (Figure 2F), confirming that the difference in subnuclear localization between the SUVR1a and SUVR1b proteins was due to sequences in the unique N-terminus of the SUVR1a splice variant. NLS2, on the other hand, is not involved in nucleolar targeting. The SUVR4 27 amino acid peptide did also direct GFP to the nucleolus, but the GFP signal was in addition seen in the nucleoplasm (Figure 2G). In the nucleolus a stronger signal was usually present in a distinct spot, as also seen in some nuclei expressing the SUVR4a-GFP fusion protein (Figure 2E). The signal from GFP-SUVR4 devoid of the terminal NLS was also found localized in such spots, however also in distinct spots in the nucleoplasm (Figure 2H). Thus, this NLS plays a role in the subnuclear localization of SUVR4, although other parts of the SUVR protein seem to be needed to accomplish a wild-type subnuclear localization pattern.

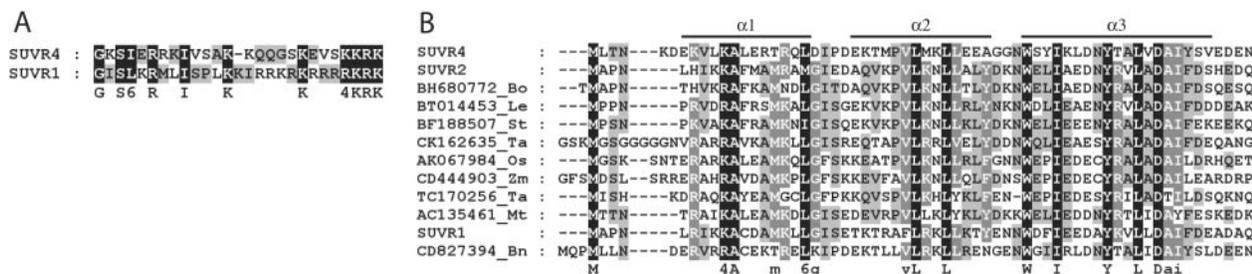


Figure 3. Plant specific regions in SUVR proteins. (A) Alignment of SUVR1 NLS1 and the SUVR4 NLS. (B) Alignment of the WIYLD domain. Three alpha helices predicted by PHDsec and JPRED software are indicated above the alignment. The SUVR1 sequence represented here is AAD10665. Bo, *Brassica oleracea*; Le, *Lycopersicon esculentum*; St, *Solanum tuberosum*; Ta, *Triticum aestivum*; Os, *Oryza sativa*; Zm, *Zea mays*; Mt, *Medicago truncatula*; Bn, *Brassica napus*.

SUVR1, SUVR2 and SUVR4 contain a new conserved plant specific protein domain

We have previously reported that the SUVR subgroup contains no other domains than the SET domain and the associated cysteine-rich regions (9), but the increased number of sequenced genomes and est sequences warranted a renewed search. We used an N-terminal sequence of SUVR4a extending to the first amino acid in the pre-SET domain (119 amino acids) in BLASTP, TBLASTN and PSI-BLAST searches against nr and est databases of GenBank. Protein sequences or translated est sequences that showed significant sequence similarity to the SUVR4 sequence were aligned (Figure 3B). A conserved region (residues 21–77 in SUVR4) was identified, and named the WIYLD domain based on conserved residues. This domain, which according to PHDsec and JPRED secondary structure prediction consists of three alpha helices, was only found in plant sequences, and so far only in proteins that possess a SET domain or are without other known domains. The latter situation could, however, be due to the fact that most of the sequences found to contain this domain were translations of partial CDS.

Putative homologues of the *Arabidopsis* SUVR1, 2 and 4 proteins were identified in rice, tomato and wheat (Supplementary Figure 1). The SUVR proteins belong to the SU(VAR)3-9 subgroup of SET-domain proteins (9), but alignment of these proteins to other proteins in this subgroup revealed a number of SUVR specific characteristics. The SET domains with their pre- and post-SET domains, and SET-I region (amino acid 319–349 in the SUVR4a sequence) were highly conserved within the SUVR group (Supplementary Figure 1), but differ from other SET domain proteins at sites known to be of functional importance (28,29). The pre-SET domains are extended with a small insertion that contains three additional conserved cysteine residues (SUVR pre-SET, Figure 1C and Supplementary Figure 1). The SUVR group has the aa DAN instead of the common aa PNL or PNV just C-terminal to the catalytic core (NHRC), and a tryptophan (W) in the motif ELx[FYW]DY (Figure 1C, arrowhead and Supplementary Figure 1, arrow). The domain composition, the SUVR specific pre-SET domain and particularities of the SUVR SET domain, substantiate that SUVR1, SUVR2 and SUVR4 are members of a particular subgroup of SET-domain proteins, which seems to be plant specific.

SUVR4 shows enzymatic specificity for H3K9 *in vitro*

Since the SET domains of SUVR1, SUVR2 and SUVR4 proteins show high sequence similarity with the SET domains of known HMTases (Supplementary Figure 1) we tested whether these proteins had HMTase activity *in vitro*. The SUVR1 and SUVR2 C-terminal fragments encompassing the pre-SET, SET and post-SET region (SACSET constructs, see Materials and Methods) did not methylate calf thymus histones (Figure 4A). The full-length GST-SUVR4a protein was, however, able to methylate calf thymus histone H3, but no methylation was seen when using recombinant full-length H3 as substrate (Figure 4B). The same results were seen for a 374 amino acid C-terminal fragment (data not shown).

These results indicated that posttranslational modification of H3 (30,31) was necessary for SUVR4 activity. In an initial approach, variously methylated histone H3 peptides were tested as substrates. As expected, the SUVR4 protein, showed very low HMTase activity against the unmethylated H3 1–20 peptide, again suggesting that unmethylated histone H3 is a poor substrate (Figure 4C). SUVR4 was also unable to methylate H3 1–20 peptides monomethylated at K4 (data not shown) or dimethylated at K9 (Figure 4C). In contrast, the H3 1–20 peptide monomethylated at K9 was significantly methylated by SUVR4 (Figure 4C). Using antibodies against dimethylated H3K9 on methylated peptide products from the HMTase assay, it was evident that monomethylated histone H3K9 became dimethylated at this position when incubated with active SUVR4 protein (Figure 4D). SUVR4 was also able to methylate histone H3 peptides of variable size when monomethylated at K9, but not when monomethylated at K4 or dimethylated on K9 (data not shown). The smallest peptide methylated, residues 5–11 mono-methylated at K9, contains only the lysine in position 9. Together these data demonstrate that SUVR4 specifically methylates histone H3 position K9, and has a substrate preference for monomethylated H3K9. This was supported by the lack of SUVR4 HMTase activity when using a histone H3 peptide (residues 23–34) monomethylated at K27 as a substrate (Supplementary Figure 2A).

As the fourth residue of the ELx[FYW]DY motif has been shown to determine product specificity, i.e. the number methyl groups added to the acceptor lysine (20,32), the tryptophan (W405) of SUVR4 was mutated into phenylalanine and tyrosine. GST-fusion proteins with these mutant versions of SUVR4 (S4W405F and S4W405Y) were used in the

HMTase assay against recombinant H3 and methylated H3 peptides. However, none of these modified proteins showed any HMTase activity (Supplementary Figure 2B), demonstrating that the tryptophan residue is absolutely necessary for the HMTase activity of SUVR4 *in vitro*.

Some proteins require certain histone tail modifications to bind histones (33,34). To investigate whether posttranslational modifications were required for targeting and binding of SUVR4, a pull-down experiment was performed. In this *in vitro* binding assay the GST-SUVR4 full-length protein was able to bind recombinant histone H3 which is devoid of posttranslational modifications (Figure 5A). In addition, we tested binding to calf thymus histones (Figure 5B), and by using antibodies against dimethylated histone H3K9 (Figure 5C) we could show that SUVR4 also bound histone H3 with this modification (Figure 5C), although dimethylated H3K9 is not a substrate for SUVR4 (Figure 4C). A 374 amino acid C-terminal SUVR4 fragment lacking the WIYLD domain pulled down calf thymus histone H3 equally well as the full-length GST-SUVR4 protein (data not shown).

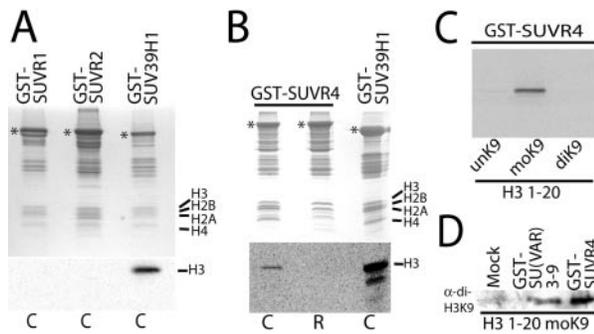


Figure 4. Histone methyltransferase activity of SUVR proteins. (A) Assay for GST-SUVR1 (amino acid 332–688), GST-SUVR2 (amino acid 348–697) and the positive control GST-SUV39H1 (amino acid 82–412) on calf thymus histones (C). (B) Assay for the full-length SUVR4a GST-fusion protein (1–465 amino acid) and GST-SUV39H1 (amino acid 82–412) on calf thymus histones (C) or recombinant histones (R). Upper panels in (A) and (B) show Coomassie stained SDS–PAGE gel with GST-fusion proteins of expected size indicated with an asterisk. Lower panels show fluorograms with the specific localization of the transferred ^{14}C labelled methyl groups. (C) Fluorogram of assay for GST-SUVR4 on histone H3 peptides residues 1–20 unmodified, monomethylated or dimethylated at K9. (D) Western analysis using a diMeH3K9 antibody after incubation of monomethylated H3K9 peptides with GST-SUVR4 and as a control SU(VAR)3-9.

In conclusion, these pull-down experiments demonstrate that SUVR4 binds histone H3 irrespective of the methylation status of lysine 9.

DISCUSSION

The SUVR proteins are plant-specific SET-domain proteins

Based on domain composition and sequence alignments of the SET domains, we have earlier classified the SUVR1, SUVR2 and SUVR4 proteins as SU(VAR)3-9 related proteins most similar to the human G9a (9). Later this small sub-group has been referred to as class V-6 SET-domain proteins (10). SUVR1, SUVR2 and SUVR4 diverge from the *Arabidopsis* SUVH members of the SU(VAR)3-9 class both in the SET domain itself and in the pre-SET region (Supplementary Figure 1). The importance of the SET flanking regions for enzyme activity has been demonstrated for other SET-domain proteins (29). The SUVR pre-SET adds three invariant cysteines to the triangular zinc-binding cluster of nine invariant cysteines found in the pre-SET of the rest of the SU(VAR)3-9 group (35). Thus, the SUVR pre-SET may confer a new type of binding, or be involved in substrate specificity (36). Notably, the variable SET-I region of the SET domain also contains SUVR-specific conserved motifs (Supplementary Figure 1).

These SUVR-specific pre-SET and SET-I regions were only found in plant proteins (Supplementary Figure 1), as was also the case for the novel N-terminal WIYLD domain (Figure 3B). Secondary structure prediction indicates that this domain consists of three alpha helices with four conserved Leu/Ile residues (Figure 3B), suggesting that it may be involved in dimerization, as is the case for the N-terminal region of SU(VAR)3-9 (37). SUVR4 histone H3 binding is independent of the WIYLD domain, and the nucleolar localization of the SUVR1a splice variant, lacking this domain, excludes a major role for this domain in nucleolar localization. However, structural similarity to the RuvA C-terminal domain (38) (Rein Aasland, personal communication), may point to a role in binding of DNA. The YDG domain of the *Arabidopsis* SUVH2 protein appears to be involved in directing DNA methylation to target sequences (18). Similarly, the WIYLD domain may be involved in directing proteins to

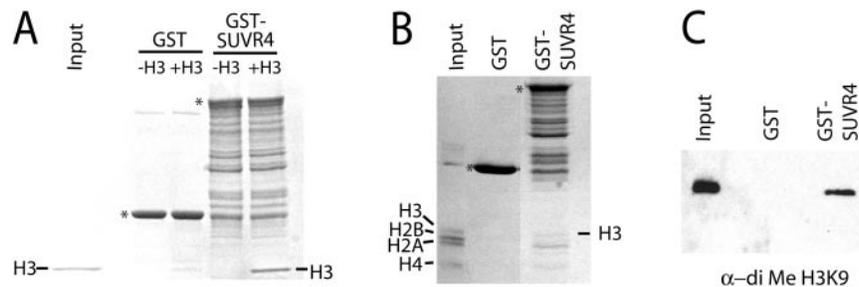


Figure 5. SUVR4 interaction with histones. (A) Coomassie stained SDS–PAGE gel after GST pull-down of recombinant histone H3 using full-length GST-SUVR4. GST alone, and mock pull-down reactions without H3 input (-H3), were used as a negative controls. The undegraded GST-fusion proteins are indicated by asterisks. (B) Coomassie stained SDS–PAGE gel after GST pull-down with full-length GST-SUVR4 using calf thymus core histones as input. GST was used as a negative control. The undegraded GST-fusion proteins are indicated by asterisks. (C) Western analysis of the reactions in (B) using a diMeH3K9 antibody.

their targets, or conversely be directed to its targets through interactions with the WIYLD domain. The plant dialect of chromatin modulation differ from other phyla in several aspects (39,40), and the SUVR proteins may play a role in some of these processes.

The subnuclear localization of SUVR proteins may be regulated by alternative splicing

All the SUVR-GFP fusion proteins were exclusively localized to the nucleus, with no expression in the cytoplasm. SUVR1a and SUVR4a were almost exclusively found within the nucleolus, and only a weak GFP signal was detectable in the nucleoplasm, while SUVR2a was associated with the nucleolus and other subnuclear regions (Figure 2A–E). When expressed in the nucleoplasm, however, the SUVR-GFP signals were excluded from densely DAPI stained heterochromatin and localized to euchromatin. Altogether, the subnuclear localization to the nucleolus and weakly DAPI-stained subdomains suggest that the SUVR proteins are associated with rDNA and/or euchromatin. This is in contrast to many SU(VAR)3-9 class proteins that function as heterochromatic stabilizers at the chromocenters (6,16,18).

The primary function of the nucleolus is rDNA transcription, pre-rRNA processing and modification, and ribosome assembly (41,42). Moreover, the activity and availability of proteins involved in cell-cycle progression may be regulated by sequestration in the nucleolus (43). The SUVH genes of *Arabidopsis*, with the exception of SUVH4, are intronless, suggesting that they have evolved via retrotransposition (9). In contrast, the SUVR genes all have introns and alternative splicing results in formation of protein isoforms (Figure 1B and C) that may regulate subnuclear spatial distribution. We have demonstrated that the extended N-terminus in SUVR1a compared to SUVR1b is responsible for localization of GFP-fusion proteins to the nucleolus (Figure 2A and B), and that 40 amino acid of this terminus encompassing the 17 amino acid NLS1 is sufficient for the nucleolar targeting (Figure 2F). The SUVR4 C-terminus containing a predicted NLS similar to NLS1 of SUVR1a (Figure 3A), could also direct GFP to the nucleolus, although the GFP signal was also seen in the rest of the nucleus (Figure 2G). Interestingly, the GFP signal of the SUVR4 NLS and the full-length SUVR4 was sometimes seen in a spot within or close to the nucleolus (Figure 2E and G). The nature of these spots as well as the subnuclear foci in the nucleoplasm observed for SUVR2 remains to be determined.

The SUVR4 protein methylates histone H3 with preference for monomethylated K9

Alignment of the SET-domain sequences [Supplementary Figure 1 and (9)] show that the SUVR proteins are most closely related to G9a, SUVH2, SUVH6 and SUVH4 of the known HMTases. These proteins methylate H3K9 and H3K27, H3K9 and H4K20 or H3K9, respectively (18,19,44). We were not able to identify any *in vitro* HMTase activity for SUVR1 and SUVR2. This may reflect the need for a cofactor or for a particular molecular context of a presumptive target histone peptide, as demonstrated for the mammalian EED-EZH2 complex. H3K27 methylation by this complex requires a minimum of three components including EZH2,

EED and SUZ12 (45). Therefore, we cannot exclude that the SUVR1 and SUVR2 proteins may function as HMTases *in vivo* at the nucleolar/euchromatic sites specified by our GFP data.

We have, however, demonstrated that at least SUVR4 acts on H3K9. *In vitro* SUVR4 has a novel specificity, in that it acts as an efficient dimethyltransferase specifically adding the second methyl group to monomethylated H3K9 (Figure 4C and D), but not to monomethylated H3K27 (Supplementary Figure 2A). Only very weak HMTase activity was seen when unmethylated H3 protein or tail peptides were used as substrates. Other known HMTases, e.g. SUVH4, SUVH5 and SUVH6, are in contrast very efficient monomethyltransferases but moderately efficient dimethyltransferases *in vitro* (16,20). These proteins have a tyrosine (Y) in position 4 of SET motif IV (ELx[FYW]DY) (35), while the SUVR proteins have a tryptophan (W) (arrowhead in Figure 1C) like the SETDB1/ESET protein, which is a di/tri HMTase (46). The DIM-5 of *Neurospora crassa* and human G9a that have phenylalanine (F) in this position can add three methyl groups to unmethylated H3K9 (28). However, conversion of the SUVR4 tryptophan to either tyrosine or phenylalanine both resulted in total loss of HMTase activity. In contrast, conversions from phenylalanine to tyrosine or vice versa in other SET-domain proteins only change the product specificity (20,28). Possibly, other particularities in the SUVR4 SET domain, for instance the amino acid DAN instead of PLN close to the important NH[RSH]C motif (Figure 1C) may impose a 3D conformation that is incompatible with substrate binding or methyl transfer when the W is exchanged with Y or F.

SUVR4 was able to bind both core and recombinant histone H3 *in vitro* (Figure 5) indicating that the binding to the histone substrate itself is independent of posttranslational modifications. The monomethylation at H3K9 thus seems to be important for the methyltransferase activity of SUVR4, but not for substrate binding. Thus, binding to histone H3 and HMTase activity may be separate functions of the same protein. If SUVR4 HMTase activity is dependent on a monomethylated histone H3 at position K9 *in vivo*, this suggests that SUVR4 is reliant on the cooperative action of another monomethyltransferase. Several examples of interdependency of posttranslational modifications have been reported [reviewed in (3)]. Furthermore, recent data shows that the DNA-methylating activity of CMT3 is regulated by the combined activity of three HMTases (20).

All available data for the SUVH proteins show that these HMTases associate with heterochromatin and are involved in heterochromatic gene silencing (18–20). In contrast, the localization of the SUVR proteins to the nucleolus or non-condensed nuclear bodies suggests that these proteins are not involved in heterochromatic gene silencing. This is also suggested by the close sequence similarity of the SUVR proteins to the G9a HMTase, which function as a repressor in euchromatin (47). Similarly, the SUVR proteins may function as regulators in euchromatin or in the nucleolus. Although the rDNA is mainly decondensed in the nucleolus, untranscribed foci of condensed rDNA may also be found inside the nucleolus (48). Based on the specific dimethylation of H3K9 *in vitro*, which is a marker of repressive chromatin domains, we suggest that SUVR4 may function as a repressor of rDNA

gene clusters in the decondensed part of the nucleolus, and possibly act to regulate rDNA expression together with HDACs also present in this compartment (49).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Rein Aasland for useful comments on the bioinformatics work, Thomas Jenuwein for the α -dimethyl-H3K9 antibody and the methylated H3 peptides. Thanks also to T. J. Meza and M. A. Butenko for the initial RACE experiments on SUVRI and SUVRI2, G. E. Stenvik for constructing the pGEX-AB-GAW vector and to B. Sæther, R. Falleth, A. Dietzel and Solveig H. Engebretsen for technical assistance. This work has been supported by grants from the Research Council of Norway (projects no. 146652/431, 166057/V40 and 151991/150), as well as grants from SFB648 of the Deutsche Forschungsgemeinschaft (DFG). The work was facilitated by the Norwegian *Arabidopsis* Research Centre (NARC), a national technology platform supported by the Research Council of Norway's Programme for Research in Functional Genomics (FUGE). Funding to pay the Open Access publication charges for this article was provided by the Research Council of Norway.

Conflict of interest statement: None declared.

REFERENCES

- Turner,B.M. (2002) Cellular memory and the histone code. *Cell*, **111**, 285–291.
- Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
- Fischle,W., Wang,Y. and Allis,C.D. (2003) Histone and chromatin cross-talk. *Curr. Opin. Cell. Biol.*, **15**, 172–183.
- Marmorstein,R. (2001) Structure and function of histone acetyltransferases. *Cell. Mol. Life Sci.*, **58**, 693–703.
- Ng,H.H. and Bird,A. (2000) Histone deacetylases: silencers for hire. *Trends Biochem. Sci.*, **25**, 121–126.
- Lachner,M. and Jenuwein,T. (2002) The many faces of histone lysine methylation. *Curr. Opin. Cell. Biol.*, **14**, 286–298.
- Kouzarides,T. (2002) Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.*, **12**, 198–209.
- Jenuwein,T., Laible,G., Dorn,R. and Reuter,G. (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.*, **54**, 80–93.
- Baumbusch,L.O., Thorstensen,T., Krauss,V., Fischer,A., Naumann,K., Assalkhou,R., Schulz,I., Reuter,G. and Aalen,R.B. (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.*, **29**, 4319–4333.
- Springer,N.M., Napoli,C.A., Selinger,D.A., Pandey,R., Cone,K.C., Chandler,V.L., Kaeppler,H.F. and Kaeppler,S.M. (2003) Comparative analysis of SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol.*, **132**, 907–925.
- Köhler,C., Hennig,L., Spillane,C., Pien,S., Grussem,W. and Grossniklaus,U. (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev.*, **17**, 1540–1553.
- Köhler,C., Hennig,L., Bouveret,R., Gheyselinck,J., Grossniklaus,U. and Grussem,W. (2003) *Arabidopsis* MSII is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J.*, **22**, 4804–4814.
- Alvarez-Venegas,R., Pien,S., Sadler,M., Witmer,X., Grossniklaus,U. and Avramova,Z. (2003) ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. *Curr. Biol.*, **13**, 627–637.
- Kim,S.Y., He,Y., Jacob,Y., Noh,Y.S., Michaels,S. and Amasino,R. (2005) Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyl transferase. *Plant Cell*, **17**, 3301–3310.
- Zhao,Z., Yu,Y., Meyer,D., Wu,C. and Shen,W.H. (2005) Prevention of early flowering by expression of *FLOWERING LOCUS C* requires methylation of histone H3 K36. *Nature Cell. Biol.*, **7**, 1156–1160.
- Jackson,J.P., Johnson,L., Jasencakova,Z., Zhang,X., PerezBurgos,L., Singh,P.B., Cheng,X., Schubert,I., Jenuwein,T. and Jacobsen,S.E. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma*, **112**, 308–315.
- Jasencakova,Z., Soppe,W.J., Meister,A., Gernand,D., Turner,B.M. and Schubert,I. (2003) Histone modifications in *Arabidopsis*—high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. *Plant J.*, **33**, 471–480.
- Naumann,K., Fischer,A., Hofmann,I., Krauss,V., Phalke,S., Irmeler,K., Hause,G., Aurich,A.C., Dorn,R., Jenuwein,T. *et al.* (2005) Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. *EMBO J.*, **24**, 1418–1429.
- Jackson,J.P., Lindroth,A.M., Cao,X. and Jacobsen,S.E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, **416**, 556–560.
- Ebbs,M.L. and Bender,J. (2006) Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell*, **18**, 1166–1176.
- Berg,A., Meza,T.J., Mahic,M., Thorstensen,T., Kristiansen,K. and Aalen,R.B. (2003) Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, *AtMBD11*, is crucial for normal development. *Nucleic Acids Res.*, **31**, 5291–5304.
- Aoyama,T. and Chua,N.H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.*, **11**, 605–612.
- Karimi,M., De Meyer,B. and Hilson,P. (2005) Modular cloning in plant cells. *Trends Plant Sci.*, **10**, 103–105.
- Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.*, **16**, 735–743.
- Fransz,P., Armstrong,S., Alonso-Blanco,C., Fischer,T.C., Torres-Ruiz,R.A. and Jones,G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.*, **13**, 867–876.
- Katsani,K.R., Arredondo,J.J., Kal,A.J. and Verrijzer,C.P. (2001) A homeotic mutation in the trithorax SET domain impedes histone binding. *Genes Dev.*, **15**, 2197–2202.
- Rea,S., Eisenhaber,F., O'Carroll,D., Strahl,B.D., Sun,Z.W., Schmid,M., Opravil,S., Mechtler,K., Ponting,C.P., Allis,C.D. *et al.* (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, **406**, 593–599.
- Collins,R.E., Tachibana,M., Tamaru,H., Smith,K.M., Jia,D., Zhang,X., Selker,E.U., Shinkai,Y. and Cheng,X. (2004) *In vitro* and *in vivo* analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J. Biol. Chem.*, **280**, 5563–5570.
- Xiao,B., Wilson,J.R. and Gamblin,S.J. (2003) SET domains and histone methylation. *Curr. Opin. Struct. Biol.*, **13**, 699–705.
- DeLange,R.J., Hooper,J.A. and Smith,E.L. (1973) Histone 3. 3. Sequence studies on the cyanogen bromide peptides; complete amino acid sequence of calf thymus histone 3. *J. Biol. Chem.*, **248**, 3261–3274.
- Schurter,B.T., Koh,S.S., Chen,D., Bunick,G.J., Harp,J.M., Hanson,B.L., Henschen-Edman,A., Mackay,D.R., Stallcup,M.R. and Aswad,D.W. (2001) Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry*, **40**, 5747–5756.
- Zhang,X., Yang,Z., Khan,S.I., Horton,J.R., Tamaru,H., Selker,E.U. and Cheng,X. (2003) Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell*, **12**, 177–185.

33. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, **410**, 116–120.
34. Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I. *et al.* (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.*, **23**, 4286–4296.
35. Cheng, X., Collins, R.E. and Zhang, X. (2005) Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 267–294.
36. Marmorstein, R. (2003) Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem. Sci.*, **28**, 59–62.
37. Eskeland, R., Czermin, B., Boeke, J., Bonaldi, T., Regula, J.T. and Imhof, A. (2004) The N-terminus of *Drosophila* SU(VAR)3-9 mediates dimerization and regulates its methyltransferase activity. *Biochemistry*, **43**, 3740–3749.
38. Parsons, C.A., Stasiak, A., Bennett, R.J. and West, S.C. (1995) Structure of a multisubunit complex that promotes DNA branch migration. *Nature*, **374**, 375–378.
39. Tariq, M. and Paszkowski, J. (2004) DNA and histone methylation in plants. *Trends Genet.*, **20**, 244–251.
40. Loidl, P. (2004) A plant dialect of the histone language. *Trends Plant Sci.*, **9**, 84–90.
41. Huang, S. (2002) Building an efficient factory: where is pre-rRNA synthesized in the nucleolus? *J. Cell. Biol.*, **157**, 739–741.
42. Olson, M.O., Dunder, M. and Szebeni, A. (2000) The nucleolus: an old factory with unexpected capabilities. *Trends Cell. Biol.*, **10**, 189–196.
43. Visintin, R. and Amon, A. (2000) The nucleolus: the magician's hat for cell cycle tricks. *Curr. Opin. Cell. Biol.*, **12**, 752.
44. Tachibana, M., Sugimoto, K., Fukushima, T. and Shinkai, Y. (2001) SET domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.*, **276**, 25309–25317.
45. Cao, R. and Zhang, Y. (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell*, **15**, 57–67.
46. Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R.G. and Zhang, Y. (2003) mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol. Cell*, **12**, 475–487.
47. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H. *et al.* (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.*, **16**, 1779–1791.
48. Neves, N., Delgado, M., Silva, M., Caperta, A., Morais-Cecilio, L. and Viegas, W. (2005) Ribosomal DNA heterochromatin in plants. *Cytogenet. Genome Res.*, **109**, 104–111.
49. Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W. and Pikaard, C.S. (2004) A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell*, **13**, 599–609.