The CW domain, a new histone recognition module in chromatin proteins

Verena Hoppmann^{1,3}, Tage Thorstensen^{2,3,4},
Per Eugen Kristiansen^{2,3}, Silje Veie Veiseth²,
Mohammad Aminur Rahman²,
Kenneth Finne¹, Reidunn B Aalen^{2,*}
and Rein Aasland^{1,*}

¹Department of Molecular Biology, University of Bergen, Bergen, Norway and ²Department of Molecular Biosciences, University of Oslo, Oslo, Norway

Post-translational modifications of the N-terminal histone tails, including lysine methylation, have key roles in regulation of chromatin and gene expression. A number of protein modules have been identified that recognize differentially modified histone tails and provide their proteins with the capacity to sense such modifications. Here, we identify the CW domain of plant and animal chromatin-related proteins as a novel module that recognizes different methylated states of lysine 4 on histone H3 (H3K4me). The solution structure of the CW domain of the *Arabidopsis* ASH1 HOMOLOG2 (ASHH2) histone methyltransferase provides insight into how different CW domains can distinguish different methylated histone tails. We provide evidence that ASHH2 is acting on H3K4me-marked genes, allowing for ASHH2-dependent H3K36 tri-methylation, which contributes to sustained expression of tissue-specific and developmentally regulated genes. This suggests that ASHH2 is a combined ‘reader’ and ‘writer’ of the histone code. We propose that different CW domains, dependent on their specificity for different H3K4 methylations, are important for epigenetic memory or participate in switching between permissive and repressive chromatin states.

The EMBO Journal (2011) 30, 1939–1952. doi:10.1038/emboj.2011.108; Published online 26 April 2011

Subject Categories: chromatin & transcription; plant biology

Keywords: chromatin; histone methylation; protein module

Introduction

Transcription of eukaryotic genes is not only dependent on transcription factors facilitating the recruitment of the RNA

polymerase pre-initiation complex, but also on the state of the template—the chromatin. The transcription machinery and most transcription factors cannot access promoters and enhancers when the DNA is wrapped in nucleosomes and must be aided by factors changing the chromatin structure, that is, chromatin remodelling and histone-modifying enzymes.

A number of histone tail modifications have been identified including acetylation, phosphorylation, ubiquitination, and methylation (Spotswood and Turner, 2002), and some of these modifications serve as marks in chromatin that reflect the state of gene activity. Histone acetylation and methylation on lysine 4 of histone H3 (H3K4) are generally associated with active loci, while histones methylated on H3K9, H3K27, and H4K20 correlate with silenced chromatin (Kouzarides, 2007).

Unique combinations of histone modifications mark different genic and chromatin regions, implicating cross-talk between different modifications. This in correlation with different states of gene expression is known as the histone code (Strahl and Allis, 2000). Histone-modifying enzymes ‘write’ the histone code, while the concept of this code in addition implies the existence of proteins that ‘read’ these modifications and translate the embedded information into effects on chromatin structure and/or the transcription machinery (Turner, 2007). This prediction has indeed been borne out by the identification of a growing number of nuclear proteins that are fretted with one or more small histone recognition modules (Taverna *et al.*, 2007). Prominent examples are bromodomains, specific for acetylated lysines; chromodomains, that can bind H3K9me or H3K27me₃; and PHD fingers that recognize H3K4me₃ or, in some cases acetylated or unmethylated histone tails (Chakravarty *et al.*, 2009; Zeng *et al.*, 2010). Several MBT domains have been shown to bind mono- and/or di-methylated lysines on both H3 and H4, but with less sequence selectivity than the chromodomains and PHD fingers (Bonasio *et al.*, 2010). A remarkable feature of histone recognition modules is that they often occur in a combinatorial manner, either as multiple domains within one polypeptide or on different subunits of larger protein complexes, facilitating the simultaneous recognition of different histone modifications in chromatin (Ruthenburg *et al.*, 2007).

The concerted action of ‘readers’ and ‘writers’ of a particular histone modification can explain how patterns of histone modifications can be propagated and inherited through many cell divisions, giving rise to the phenomenon of epigenetic inheritance. Similar mechanisms can explain how the transcriptional status of genes can be changed; a chromatin-modifying enzyme could ‘read’ one modification while ‘writing’ another. Furthermore, alterations of histone modification patterns may be brought about by sequence-specific transcription factors and/or non-coding RNAs that recruit different histone-modifying enzymes as cofactors

*Corresponding authors. RB Aalen, Department of Molecular Biosciences, University of Oslo, Blindernveien 31, N0371 Oslo, Norway. Tel.: +47 22854437/48218985; Fax: +47 22856041; E-mail: reidunn.aalen@imbv.uio.no or R Aasland, Department of Molecular Biology, University of Bergen, Thormøhlensgt 55, N5020 Bergen, Norway. Tel.: +47 55584531; Fax: +47 55589683; E-mail: aasland@mbi.uib.no

³These authors contributed equally to this work

⁴Present address: The Norwegian Biotechnology Advisory Board, Rosenkrantz' gt. 11, N0105 Oslo, Norway

Received: 11 August 2010; accepted: 11 March 2011; published online: 26 April 2011

(coactivators, corepressors, or silencing complexes) (Goodman and Smolik, 2000; Imhof, 2006; Muller and Kassiss, 2006; Ringrose and Paro, 2007).

Histone lysine methylation is conferred by SET-domain proteins that can be divided into several evolutionarily conserved classes (Baumbusch *et al*, 2001; Kouzarides, 2007; Wu *et al*, 2010) including: (1) the E(z) class, involved in the maintenance of a transcriptionally repressive state of genes via H3K27 tri-methylation; (2) SU(VAR)3–9 proteins, implicated in heterochromatinization via H3K9 methylation; (3) the TRXSET1 family that contribute to the active state via H3K4me3; and (4) ASH1 proteins associated with transcriptional elongation via H3K36me. In addition to the identity of the modified lysine, the number of methyl groups added is functionally significant (Fischer *et al*, 2006). For example, using genome-wide chromatin profiling in mammalian cells, it was recently shown that, while H3K4me2 and me3 marks are prominent near transcription start sites (TSS), tissue-specific enhancers are enriched for mono-methylated H3K4 (Heintzman *et al*, 2007, 2009; Kim *et al*, 2010). In the model plant *Arabidopsis thaliana*, H3K4me3 is preferentially found in the 5'-end of highly expressed genes with low tissue specificity, while H3K4me1 is highly correlated with CpG methylation in the transcribed region of genes (Zhang *et al*, 2009). Furthermore, H3K36 tri-methylation of MADS box genes involved in flowering time control and flower development, shows a positive correlation with transcription, but H3K36me1 does not (Xu *et al*, 2008; Grini *et al*, 2009).

Several SET-domain histone methyltransferases (HMTases) have histone recognition modules as co-domains, either on the same polypeptide or on another subunit in a protein complex. Examples are chromodomains in animal SU(VAR)3–9 proteins and PHD fingers in Trx/MLL proteins. Co-domains are thought to contribute to the recruitment of

the histone modifiers to relevant sites in chromatin (Ruthenburg *et al*, 2007). Alternatively, they may modulate the activity of the methyltransferase, as in the case for E(z)/EZH proteins where the H3K27me3-binding EED/Esc subunit contribute to H3K27me3 methylation on adjacent nucleosomes (Margueron *et al*, 2009).

While the functional role of several recognition modules has been worked out in some detail, we are far from understanding how they contribute to maintaining or altering chromatin structure and thereby modulate gene expression. In this paper, we have identified the CW domain as a new type of histone recognition module and explored its properties. This domain, named after its conserved cysteine and tryptophan residues, was first identified as an MBD-associated domain (MAD) in a subgroup of methyl-CpG-binding proteins of *Arabidopsis* (Berg *et al*, 2003). The CW domain is found in a small number of chromatin-related proteins in animals and plants (Perry and Zhao, 2003; see Table I). Some of the genes that encode CW proteins have mutant alleles with phenotypes that underscore their functional importance: Mutation in the mouse *Morc1* causes arrested spermatogenesis (Inoue *et al*, 1999), *Morc2b* was recently shown to be involved in hybrid sterility (Mihola *et al*, 2009), and *MORC4* has been found highly expressed in large B-cell lymphomas (Liggins *et al*, 2007). The *Arabidopsis val1val2* double mutant fail to repress embryonic development during vegetative growth (Suzuki *et al*, 2007). The mammalian CW protein AOF1/LSD2 (alias KDM1B) is a H3K4me1- and me2-specific histone demethylase (Karytinos *et al*, 2009). AOF1/LSD2 has a demethylase-independent repressor function, which, on the other hand, requires the CW domain (Yang *et al*, 2010).

The best-studied CW protein is, however, the *Arabidopsis* ASH1 HOMOLOG2 (ASHH2), also known as SDG8/EFS/

Table I Proteins with CW domains in humans and *Arabidopsis*

Protein family	Protein	Chromosomal location (Hs) GeneID (At)	Function/phenotype
<i>Human</i>			
CW + PWWP	ZCWPW1	7q22.1	
	ZCWPW2	3p24.1	
MORC family	MORC1	3q13	Mouse: microrchidia (spermatogenesis)
HSP90-like ATPases			
	MORC2	22q12.2	Mouse: induced by Prdm9, a hybrid sterility gene
	MORC3	21q22.13	Mouse: CW domain required for proper localization in the nucleus
	MORC4	Xq22.3	
LSD1 histone demethylases	AOF1 (LSD2)	6p22.3	H3K4me1,2-specific histone demethylase
<i>Arabidopsis</i>			
MBD family (CW-MBD)	AtMBD1	At4g22745	
	AtMBD2	At5g35330	Euchromatic localization
	AtMBD3	At4g00416	
	AtMBD4	At3g63030	
ASHH family Histone methyltransferase (CW + SET)	ASHH2	At1g77300	Histone methyltransferase (H3K36me2,3); severe pleiotropic phenotype: small organs, early flowering, distorted development of reproductive organs
VP1/ABI3-Like (B3 + CW)	VAL1	At2g30470	Repressors of embryonic genes at germination; double mutant with embryonic seedling phenotypes
	VAL2	At4g32010	
PHD + CW	NP_179516	At2g19260	
F-box proteins SWI/SNF-ATPase putative chromatin remodelling	FB304_ARATH	At3g54460	
Solo CW	NP_191849	At3g62900	
	O23424_ARATH	At4g15730	

CCR1. ASHH2 is an ~200 K SET-domain protein considered to be a major H3K36me2/me3 HMTase in *Arabidopsis*, as chromatin of *ashh2* mutants shows a global reduction in H3K36me2/me3 levels (Zhao *et al*, 2005; Xu *et al*, 2008). In ASHH2, a CW domain precedes the AWS and SET domains. We and others have shown that mutations in ASHH2 confer pleiotropic effects like small, bushy plants with early flowering, homeotic changes of floral organs, and severely reduced fertility. The expression of the major regulator of flowering time in *Arabidopsis*, *FLOWERING LOCUS C (FLC)*, a direct target of ASHH2 (Ko *et al.*, 2010), as well as other transcription factor genes involved in these developmental processes, is repressed in the mutant correlating with a reduction in H3K36me2/me3 levels in mutant plants (Kim *et al*, 2005; Zhao *et al*, 2005; Dong *et al*, 2008; Xu *et al*, 2008; Grini *et al*, 2009). It is not clear, however, whether this mark is a prerequisite for gene expression.

In vitro, ASHH2 is active on histone H3 isolated from eukaryote nuclei but not on recombinant histones (Dong *et al*, 2008; Grini *et al*, 2009), suggesting the requirement of a pre-modified histone tail. The contrasting features of ASHH2 compared with yeast and animal H3K36 HMTases prompted us to investigate the chromatin marks and transcriptional status of putative ASHH2 target genes in more detail. We found that they are enriched for tissue-specific and developmentally regulated genes that have H3K4 marks. We therefore reasoned that the CW domain in ASHH2 might read methylated H3K4.

Here, we show that the ASHH2 CW domain as well as CW domains from other plant and animal proteins bind methylated H3K4. We have solved the solution structure of the ASHH2 CW domain and probed the putative binding site for the histone tail on the surface of CW. For ASHH2, we propose a model where the CW domain targets the enzyme to H3K4me-marked active chromatin, which subsequently introduces methylation on H3K36.

Results

H3K36me3 methylation is positively correlated with transcription of tissue-specific genes

Although ASHH2 appears to be the enzyme responsible for global di- and tri-methylation of H3K36, only a subset of genes are transcriptionally affected by *ashh2* mutation. We, therefore, investigated the effect of the *ashh2* mutation on expression and histone marks for a selected panel of genes, with the aim of identifying features of the chromatin context in which ASHH2 is acting, assuming that the function of its CW domain is to render the enzyme sensitive to this chromatin context. With antibodies against H3K4me3, H3K36me2, and H3K36me3, ChIP analyses comparing wild-type (wt) and *ashh2* mutant seedlings were used on a set of tissue-specific genes with differential expression profiles in seedlings and flowers: (1) *APETALA1 (AP1)*, *MYB99*, and the transcription factor *NAC25* predominantly expressed in the inflorescences; and (2) *DISRUPTION OF MEIOTIC CONTROL 1 (AtDMC1)*, *MADS-BOX AFFECTING FLOWERING 1 (MAF1)* and *FLC*, with low and tissue-specific expression in both vegetative and reproductive tissues (Supplementary Figure S1). We have previously reported that *AP1*, *MYB99*, *NAC25*, and *AtDMC1* are associated with mutant phenotypes, show lower transcript levels, and reduction in H3K36me3 but not

H3K4me3 and H3K36me2 levels in *ashh2* mutant inflorescences (Grini *et al*, 2009). *MAF1*, which similarly to *FLC* is involved in determination of flowering time, is transcriptionally downregulated in both *ashh2* seedlings and inflorescences (Kim *et al*, 2005; Zhao *et al*, 2005; Xu *et al*, 2008; Grini *et al*, 2009). *ACTIN2* and *GAPA (GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A SUBUNIT)*, which show high expression but little tissue specificity, were also included in the ChIP analysis. The silent *Ta3* retrotransposon was used as a control (Zhao *et al*, 2005).

H3K4me3 levels in seedlings were high in the wt for these two strongly expressed genes, and although the transcription level was not downregulated in the *ashh2* mutant (Supplementary Figure S2A), a substantial reduction in the level of H3K4me3 was evident for *GAPA* (Figure 1A; Supplementary Figure S3A). The other genes tested showed very low H3K4me3 levels and were largely unaffected by the *ashh2* mutation, except around the transcriptional start site of *MAF1* and at the beginning of the first intron of *FLC* (Supplementary Figure S3A).

H3K36me2 and H3K36me3 levels were low for the inflorescence-specific *AP1*, *MYB99*, and *NAC25* genes in wt seedlings, although significantly above background levels without antibody, and levels of the heterochromatin mark H3K9me2 (Supplementary Figures S1 and S2B). These levels were not affected in the *ashh2* mutant (Figure 1B and C).

The K36me2 and me3 levels were reduced for *MAF1* and *FLC* in *ashh2* seedlings (Figure 1B and C; Supplementary Figure S3B; see Ko *et al.* (2010) and Xu *et al* (2008)), and this correlates with reduced transcription levels in the mutants (Zhao *et al*, 2005; Xu *et al*, 2008).

Unexpectedly, *AtDMC1* and *GAPA* showed significant increases of K36me2 and a reduction of H3K36me3 in *ashh2* mutant background compared with wt, while *ACTIN2* showed reduction of both H3K36me2 and H3K36me3 (Figure 1B) although transcript levels of these three genes were not affected by the mutation (Supplementary Figure S2A). The significant reduction of H3K36me3 on *ACTIN2* and *GAPA* in the mutant seedling samples (Figure 1C; Supplementary Figure S3B) was also found in inflorescences (Supplementary Figure S2C).

These data suggest that the level of H3K36me3 methylation may reflect the level of expression of a gene, but also that the H3K36me3 mark may not be needed for sustained expression of genes with high and constitutive expression levels, like *ACTIN2* and *GAPA*.

Transcription profiling indicate that ASHH2 is a major regulator of transcription factor genes

To get an indication of the importance of ASHH2 for global gene transcription, a microarray experiment was conducted for wt and *ashh2* seedlings. A total of 183 genes showed >1.6-fold change, of which 135 were downregulated and 48 upregulated. Inspection of GO terms for these genes revealed a striking enrichment (18.4%; $P=1.84 \times 10^{-4}$) for genes encoding transcription factors and DNA/RNA-binding proteins among the downregulated genes (Figure 2A; Supplementary Table SI). A similar enrichment (15.5%) was also found in a previously published microarray experiment (Xu *et al*, 2008). Only 5.7 % of all *Arabidopsis* genes encode transcription factors.

Figure 1 Histone tail methylation in chromatin of selected genes from wild-type and *ashh2-1* mutant seedlings. ChIP using antibodies for (A) H3K4me3, (B) H3K36me2, and (C) H3K36me3. Data are shown as percent of input. The Ta3 retrotransposon was used as reference. Standard deviations are shown. Primers used are in the body of the genes (see Supplementary Table SIV). Results without antibody (–ab) are shown. Note that the scale on the y-axis differs in the three panels. For H3K9me2 results, see Supplementary Figure S2B.

Figure 2 Properties of genes with changed expression levels in *ashh2* seedlings. (A) Functional annotation using gene ontology (GO) of genes differentially expressed in *ashh2-1* inflorescences relative to wt. (B, C) Distribution of H3K4me marks for genes downregulated in *ashh2* mutant seedlings (c.f. Supplementary Table SIII) and genes from wt seedlings according to Zhang *et al* (2009: <http://epigenomics.mcdm.ucla.edu/H3K4m1m2m3/>). In (B) genes without any H3K4 methylation are included, and in (C) genes without H3K4 methylation were not included. 'N' is number of genes.

Global ChIP data suggest that ASHH2 has a bias for tissue-specific genes

To investigate whether the genes with ASHH2-dependent regulation had particular characteristics, data from our microarray experiment as well as previously published experiments (Xu *et al*, 2008; Cazzonelli *et al*, 2009; see Materials and methods) were surveyed for the presence of H3K4me3, H3K27me3, and H3K36me2 (Table II; Supplementary Table SII), using published global ChIP data (Oh *et al*, 2008). Interestingly, over 84% of the genes downregulated in the mutant had K4me3 marks, either alone or in combination with other marks, suggesting that ASHH2 preferentially associates with transcribed genes. Consistent with this, genes with H3K27me3 marks only, which are likely to be silent, were significantly underrepresented. Genes with all three marks, likely to be tissue specific or developmentally regulated (Oh *et al*, 2008), were most significantly

Table II Representation of *ashh2* downregulated genes with different chromatin methyl marks

Chromatin methyl marks	Global ^a		<i>ashh2-1</i> ^b			<i>sdg8-2</i> ^b			<i>ccr1</i> ^b		
	N	%	N	%	<P	N	%	<P	N	%	<P
H3K4me3	2929	12.0	33	24.4	2×10^{-5}	7	8.3	NS	12	12.9	NS
H3K4me3 + H3K36me2	12 012	49.1	60	44.4	NS	46	54.8	NS	53	57.0	NS
H3K36me2	1671	6.8	2	1.5	0.02	1	1.2	0.05	1	1.1	0.05
H3K4me3 + H3K27me3	1245	5.1	11	8.1	NS	8	9.5	NS	3	3.2	NS
H3K27me3	5724	23.4	14	10.4	0.002	11	13.1	0.05	7	7.5	0.005
H3K36me2 + H3K27me3	280	1.1	0	0.0	NS	1	1.2	NS	0	0.0	NS
H3K4me3 + H3K27me3 + H3K36me2	607	2.5	15	11.1	5×10^{-6}	10	11.9	10^{-6}	17	18.3	5×10^{-6}
Total number of genes	24 468		135			84			93		
H3K4me3 in total	16 793	68.6	119	88.1	0.001	71	84.5	NS	85	91.4	0.001
H3K27me3 in total	14 570	59.5	77	57.0	NS	58	69.0	NS	71	76.3	0.05
H3K36me2 in total	7856	32.1	40	29.6	NS	30	35.7	NS	27	29.0	NS

^aChromatin enrichment groups and number of genes in each group according to Oh *et al* (2008). In the three bottom rows, the number of genes is given for each mark, irrespective of the presence of other marks.

^bChromatin enrichment according to Oh *et al* (2008) for genes downregulated in three independent microarray experiments on *ashh2* mutants seedlings; *ashh2-1* (present study), *sdg8-2* (Xu *et al*, 2008) and *ccr1* (Cazzonelli *et al*, 2009). Values significantly higher than for the whole genome are shown in boldface, while values significantly lower than for the whole genome are shown in italicized boldface. N, number of genes; NS, not significant.

overrepresented among the downregulated genes (Table II). None of these biases were found among the upregulated genes (Supplementary Table SII). The 45 downregulated genes found common to two or three of the microarray data sets, of which nine encode transcription factors, are more likely to be direct targets of ASHH2. Of these genes, 15.7% have triple marks (Supplementary Table SIII), compared with 2.5% in the global gene set. *FLC*, known as a target of ASHH2, is among them.

The 45 genes downregulated in *ashh2* mutants and the panel of genes investigated in our ChIP experiment were also surveyed for H3K4me1, me2, me3, and H3K27me3 using a published, global data set for *Arabidopsis* seedlings (Zhang *et al*, 2009; Supplementary Table SIII; Figure 2B and C). *FLC* shows bivalent marks, both K4me2/me3 and K27me3 (Supplementary Table SIII). The inflorescence-specific genes *AP1*, *NAC25*, and *MYB99* were only marked by the repressive K27me3 in seedlings. As expected from global data analyses (Zhang *et al*, 2009), the two highly and generally expressed genes, *ACTIN2* and *GAPA*, used in our ChIP experiment (Figure 1), have K4me1/me2/me3 and K4me1/me3 marks, respectively, while the low-expressed, tissue-specific *AtDMC1* and *MAF1* are marked with K4me1/me2 and K4me1, respectively. Consistent with this, our ChIP results showed that there are much higher levels of K4me3 on *ACTIN2* and *GAPA* than on the other genes tested (Figure 1A).

From the data set of Zhang *et al*. (2009), 43% of 5839 genes investigated in detail were devoid of H3K4me marks (Figure 2B), while this category was significantly under-represented among the 45 ASHH2-dependent genes (8.9%; Figure 2B; Supplementary Table SIII). The fraction of genes with the K4me1 mark was similar to wt (31.1 versus 32%), while K4me2 and K4me3 were overrepresented (77.1 and 86.7% versus 38.8 and 55.4%, respectively). When considering combinations of K4me marks, K4me2/me3 and K4me1/me2/me3 were overrepresented among genes downregulated in the *ashh2* mutant seedlings (Figure 2B and C).

Again this suggests that ASHH2 is associated with transcribed genes, and furthermore that ASHH2 has a

particular preference for transcribed genes with K4me2 and K4me1/me2 marks.

The ASHH2 CW domain binds methylated lysine 4 on H3

The occurrence and proximity of the CW domain to the SET domain in ASHH2 (Figure 3A) are reminiscent of many other histone-modifying enzymes and prompted us to ask whether CW is a histone recognition module that could explain the association of ASHH2 with transcribed genes and methylated H3K4 as identified above. We expressed and purified the ASHH2 CW domain as a GST fusion protein from bacteria (GST-CW_{ASHH2}) and tested it for binding to a panel of immobilized histone tail peptides (Figure 3A and B). As shown in Figure 3C, the ASHH2 CW domain can indeed bind histone tail peptides with comparable avidity as the ING2 PHD finger (Shi *et al*, 2006). Unlike the ING2 PHD finger, however, ASHH2 CW shows preference for mono- and di-methylated H3K4 peptides, while its binding to the H3K4me3 is close to background levels. The preference for mono- and di-methylated H3K4 peptides was quantified by surface plasmon resonance using BIACore (Figure 3E and F). The K_d s for H3K4me1 and H3K4me2 were 1.15 and 2.1 μ M, respectively, comparable to the affinities of ING2 PHD finger to the H3K4me3 peptide (Pena *et al*, 2006). K_d for binding to H3K4me3 in these experiments was 4 μ M.

From the data in Figure 3C, it is also evident that the ASHH2 CW domain is sequence selective, as it only binds the methylated H3K4 peptides, and not the tri-methylated H3K27 and H3K36 peptides, nor the other peptides included on the panel. The preference of ASHH2 CW for mono- and di-methylated H3K4 peptides is shared with MBT domains (Bonasio *et al*, 2010). The MBT domains show little sequence selectivity, however, as they bind several mono- and di-methylated lysines on both H3 and H4 peptides. In contrast, the ASHH2 CW domain appears to be specific for methylated H3K4 peptides, as no binding was observed for any of the three methylated forms of H3K27 peptides (Figure 3D). This suggests that the CW domain may have a more extensive interaction with the histone H3 tail residues near K4.

Figure 3 The ASHH2 CW domain binds histone H3 tails methylated on lysine 4. **(A)** Cartoon of the ASHH2 protein with CW, AWS, and KMT (SET domain) indicated. The two recombinant proteins used in this study are also illustrated. **(B)** SDS-PAGE of purified GST and GST-ASHH2 CW. **(C)** Pull-down assay using biotinylated histone peptides and recombinant GST, GST-ASHH2 CW and GST-ING2 PHD as indicated. The ratio of proteins and peptides was 0.5:0.5 μ g. **(D)** Pull-down assay using biotinylated histone peptides and recombinant GST and GST-ASHH2 CW as indicated. The ratio of proteins and peptides was 1.0:0.5 μ g. **(E)** Surface plasmon resonance traces from an experiment with immobilized, biotinylated H3K4me1 peptide and GST-ASHH2 CW. **(F)** Dissociation constants for GST-ASHH2 CW and the three H3K4me peptides indicated calculated from surface plasmon resonance data. The number of replicates are indicated.

Binding to methylated histone tails is a shared feature among CW domains

We expressed and assayed several other CW domains in the histone tail-binding assay. As is evident from Figure 4, the CW domains of human MORC4 and ZCWPW1 and *Arabidopsis* VAL1 all bind H3 peptides methylated on lysine 4. While the MORC4 CW domain shows a preference for H3K4me2, the CW domains of ZCWPW1 and VAL1 show a clear preference for H3Kme2 and me3 peptides. Based on these data, it is reasonable to predict that most, if not all, CW domains are H3K4-selective recognition modules, but their preference for the methylation states may vary. Furthermore, as the different methylated forms of H3K4 are associated with chromatin of active or poised genes and/or their enhancers, we propose that the CW proteins are acting on active or poised chromatin.

The structure of the ASHH2 CW domain

To further investigate the structure of the CW domain and its mode of interaction with the histone tail, we solved the solution structure of the ASHH2 CW domain using NMR spectrometry (Figure 5A–C; Supplementary Figure S4). The HSQC spectrum indicates that the molecule was structured (not shown). Heteronuclear 3D NMR methods were used to assign the molecule. The chemical shifts of the cysteine residues are all in agreement with metal-bonded cysteines, Cys868, 871, 893, and 904, while Cys931 is

non-oxidized. We, thus, conclude that the former four cysteines are bound to Zn^{2+} .

Comparison of this structure to the reported structure of the human ZCWPW1 CW domain (pdb:2e61; Figure 5D) revealed that both domains share a common structural core consisting of a two-stranded β -sheet and a Zn^{2+} -coordinating quartet of cysteines. Both structures also show three short helical elements (η 1–3). Inspection of the molecular surface (Figure 5C) revealed a conspicuous cleft running across the front surface with a shallow pocket lined by two of the conserved tryptophans (W865 and W874 in ASHH2). The solution structure of ZCWPW1 CW in complex with a H3K4me3 peptide was recently reported (He *et al*, 2010). This structure shows that the histone peptide traverses the cleft, and the methylated ϵ -amino group of lysine 4 is bound in the pocket above, which serves as an aromatic cage. It is reasonable to propose that the H3K4me1 peptide could bind the core of the ASHH2 CW domain in a similar manner.

In the ASHH2 CW structure, an α -helix (α 1) is formed by residues 912–919, followed by a less structured C-terminal segment (Figure 5A). This helix has a strong amphipathic character and it is situated on top of the pocket with its hydrophobic side facing the core surface and occluding the aromatic cage (Figure 5B). This helix is not part of the conserved core of CW domains (Supplementary Figure S5). In ZCWPW1 CW, the C-terminal extension provides a third tryptophan to the aromatic cage (He *et al*, 2010).

Figure 4 CW domains from four different proteins bind methylated H3K4 peptides. (A) Cartoons of the proteins studied: human MORC4, *Arabidopsis* VAL1, and human ZCWPW1. (B, C) Pull-down assay using biotinylated histone peptides and (B) recombinant GST, GST-Val1 CW and GST-MORC4 CW or (C) GST, GST-ASHH2 CW, and GST-ZCWPW1 CW, as indicated. The ratio of proteins and peptides was equimolar. 'PW' indicates the PWWP domain in ZCWPW1.

The interaction of CW with histone tail peptides

To investigate the interaction of the histone peptide with the ASHH2 CW domain, we determined the chemical shift values of the domain in the absence and presence of a mono-methylated H3K4 peptide by NMR. The HSQC spectrum showed that the domain is structured, in both the presence (Supplementary Figure S4) and absence (not shown) of histone peptide. Upon histone peptide binding, a number of discrete chemical shift changes were observed (Figure 5E). The most prominent shifts occurred around R867 and R890, which are juxtaposed on each side of the cleft. There were also moderate shifts in the central, variable region (which forms lower part of the cleft), suggesting that this part of the structure is changing conformation upon histone tail binding. The residues near the zinc-coordinating cysteines showed less changes, in agreement with their structural role in the core of the domain. Intriguingly, only one of the two conserved tryptophans in the predicted methyllysine-binding pocket showed a significant change, namely W865. There was also a significant change for the adjacent R876, which is positioned in the outer end of the cleft. The C-terminal extension, including helix $\alpha 1$, also showed significant changes, suggesting that this part of the molecule is moving when the histone peptide binds. We propose that a movement of the C-terminal helix away from the aromatic cage is taking place as the methylated H3K4 peptide binds. In summary, these NMR data confirm binding of the histone tail to the CW domain and the chemical shift changes are in good agreement with a mode of binding involving the aromatic cage.

Figure 5 Solution structure of the CW domain and prediction of its binding site for the histone tail. (A) Backbone traces of 20 conformers of the solution structure of the ASHH2 CW domain showing residues 858–928. Secondary structure elements are indicated in blue (β -strands), red (α -helix), and pink (helix-like segments). The zinc ion is shown in green, and the two tryptophans of the aromatic cage are shown in purple sticks. (B) Spheres view of the ASHH2 CW domain highlighting hydrophobic residues (yellow), side chain carboxyl groups of glutamates and aspartates (red), nitrogens of the side chains of lysine and arginine (blue), and the zinc ion (green). The two conserved tryptophans forming the aromatic cage are shown in orange. A cartoon trace in dark blue is shown in the background. (C) Surface representations and surface potential of the ASHH2 CW domain (residues 862–913). Negative and positive potentials are shown in shades of red and blue, respectively. The predicted placement of the histone tail backbone and the aromatic cage is indicated by a dotted line and a circle, respectively. (D) Structure of the ZCWPW1 CW domain (pdb:2e61) colour coded as in (B). (E) ^1H - and ^{15}N -amide chemical shift changes for ASHH2 CW bound to the H3K4me1 peptide. The secondary structure elements are indicated in the same colour code as in (A).

To further corroborate these data, we generated several point mutations in and around the putative histone tail-binding site and tested them for binding (Figure 6A). Mutation of the three tryptophans 865, 874, and 891 to alanine abolished histone tail binding (Figure 6B). Two of these (W865 and W874) form the predicted methyllysine-binding site, while W891 is located in the presumptive histone tail-binding cleft. Mutation of the two residues Q908 and E909, positioned above the putative methyllysine-

Figure 6 Probing the predicted histone tail-binding site with mutant CW domains. (A) A space-filling model of the ASHH2 CW domain is shown with residues mutated highlighted: grey: loss of binding; light grey: reduced/altered binding. The predicted binding site for the histone tail is indicated as in Figure 5C. (B–D) Pull-down assays using biotinylated histone peptides and recombinant GST and GST-ASHH2 CW mutant proteins as indicated. The ratio of proteins and peptides was 1.0:0.5 μ g.

binding pocket also abolished binding (Figure 6C). According to our model for histone tail binding, these two residues may contribute to polar or ionic interactions with the positively charged ϵ -amino group of K4. Mutation of the non-conserved

D886 to alanine positioned in the lower part of the cleft resulted in relaxed specificity with binding also for K4me2 (Figure 6C). The significance of this result is not yet clear.

When the CW domain was C-terminally truncated from residue M910, binding was lost (Figure 6D) suggesting that the C-terminal extension, including helix α 1, has an important role in histone peptide binding.

The ASHH2 CW domain binds nucleosomal histones

To investigate whether the CW domain can bind histone H3K4me tails in a nucleosomal context, we performed pull-down experiments with chromatin prepared from *Arabidopsis* seedlings. The GST-CW_{ASHH2} fusion protein pulled down histones that were mono-, di-, or tri-methylated at H3K4 (Figure 7A). Consistent with the peptide binding data the mutant version of the CW domain (W874A) did not pull down chromatin containing H3K4 methylated histones, nor did the WIYLD domain of SUV4 (Thorstensen *et al*, 2006), included as a negative control. This suggests that the binding of CW to H3K4me is specific. CW binds H3K4me1 in chromatin somewhat stronger than H3K4me2 (4.4 versus 3.3 times of bound chromatin relative to input material (2.5%)) as compared with 2.1 times for H3K4me3. It should be noted that this chromatin is a mixture of mono-, di-, and tri-nucleosomes, and that H3K4me1, H3K4me2, and H3K4me3 marks often co-reside on ASHH2 target genes (Supplementary Table SIII; Figure 2B and C).

To investigate whether the CW domain targets genes that are regulated by ASHH2, we analysed DNA from seedling chromatin pulled down by the GST-CW_{ASHH2} fusion protein by real-time PCR. This chromatin pulldown (ChPD) experiment demonstrated that CW binds chromatin associated with these genes significantly above background levels (Figure 7B). *FLC*, proven to be targeted by ASHH2 *in vivo* (Ko *et al*, 2010), was detected in the CW-bound chromatin, thus demonstrating the ability of the CW pulldown to identify *in vivo* targets of ASHH2. The CW domain most strongly pulled down chromatin associated with genes that show substantial reduction in H3K36me3 levels in the *ashh2* mutant (*ACTIN2*, *GAPA*, *MAF1*, and *FLC*), suggesting that the CW domain may contribute to the targeting of ASHH2 to chromatin associated with these genes. For *MYB99*, *AtDMC1*, and *MAF1*, recovery was higher with primers downstream of the TSS. When comparing the CW_{ASHH2} ChPD (Figure 7B) and the ChIP experiments (Figure 7C), it is evident that the recovery profiles are very similar for ChPD and H3K4me1, and also for K36me3. Only for the very H3K4me3-rich *GAPA* gene and the transcriptional start site of *MAF1*, the H3K36me3 levels were more similar to H3K4me3 than to K4me1.

The panel of tested genes showed no dramatic difference in H3K4me1 levels between wt and the *ashh2* mutant (Supplementary Figure S3C). Western blots of ChPD experiments show that GST-CW_{ASHH2} efficiently pulled down H3K4me1 marked chromatin of wt seedlings (Figure 7D, upper panel compared with the H3 control in the lower panel). Antibodies against H3K36me3 revealed the presence of this mark on the chromatin pulled down by GST-CW_{ASHH2} (Figure 7D, middle panel), suggesting that H3K4me1 and H3K36me3 co-reside on the same or neighbouring nucleosomes. If H3K36me3 was deposited on chromatin indepen-

Figure 7 Chromatin pulldown (ChPD) using the ASHH2 CW. (A) ChPD with nucleosomes (from chromatin as prepared for ChIP from *Arabidopsis* seedlings) and immobilized GST, GST-CW_{ASHH2}, GST-CW_{ASHH2}-W874A (non-binding mutant), and GST-WYLD_{SUVR4} (negative control) as indicated. Modified histones were visualized by western blot using antibodies as indicated. Molecular size markers are shown to the right. (B) qPCR with primers for assumed ASHH2 target genes on DNA isolated from seedling chromatin pulled down with GST-CW_{ASHH2}. ChPD with GST alone was used as a negative control. Data are shown as percent of input. The Ta3 retrotransposon was used as reference. Standard deviations are shown. (C) ChIP using antibodies against H3K4me1, H3K4me3, and H3K36me3 and primers for the same genes as in (B). Data are shown relative to Ta3. The same data are shown relative to input in Supplementary Figure S3. The *FLC C* primers amplify a region at the beginning of the first intron, while *FLC E* is found more downstream in the gene. TSS indicates primers near the transcriptional start site, while other primers used are in the body of the genes (see Supplementary Table SIV). (D) ChPD from wt and *ash2* mutant seedlings as in (A) using antibodies as indicated. ChPD with GST alone was used as a negative control and histone H3 as reference.

dently of H3K4me1, one would expect that GST-CW_{ASHH2} pulled down H3K36me3 chromatin equally well in the mutant compared with the wt, relative to input. However, less H3K36me3 chromatin was pulled down from *ash2* chromatin than from wt chromatin relative to input (Figure 7D, middle panel). This suggests that the chromatin regions pulled down by CW were highly affected by the *ash2* mutation. Together with the finding that CW pulls down chromatin from genes with the most significant reduction of H3K36me3 in the mutant (Figure 7B and C), these data

suggest that the H3K36me3 mark, mediated by ASHH2 activity, is closely associated with H3K4me1-marked chromatin bound by the CW domain.

Discussion

The CW domain is a new histone recognition module with specificity for methylated H3K4

We describe here the CW domain as a new histone recognition module with specificity for histone H3 tails methylated

on lysine 4. For the ASHH2 CW, the histone tail binding was demonstrated in four different ways by: (1) histone tail-binding pulldown assays; (2) surface plasmon resonance; (3) nucleosome binding assay; and (4) NMR. The affinity for the mono- and di-methyl H3K4 peptides is in the micromolar range, comparable to PHD fingers and other histone recognition modules. We have also demonstrated histone tail binding for three additional CW domains, suggesting that this is the generic molecular function of CW domains.

Among the families of H3K4-specific recognition modules, CW has a novel profile of ligand selectivity, with members showing preference for me1 and me2 (ASHH2), me2 and me3 (VAL1 and ZCWPW1) or for me2 (MORC4). This is distinct from, for example, PHD fingers, which bind either tri-methylated or unmethylated H3K4 peptide; the trimethyllysine-specific double tudor domain and double chromo domains; and the MBT domains, which also bind mono- and di-methylated lysines, but in several different sequence contexts (Bonasio *et al*, 2010).

Remarkably, none of the mammalian and plant CW proteins are orthologous, yet, they appear to be involved in phenomena related to chromatin and gene regulation (Table I). Surprisingly, CW domains are found in plants and chordates as well as certain protist lineages and the cnidarian *Nematostella vectensis*, but not in insects and nematodes (see Pfam:zn-CW, PF07496, for details). We suggest that CW proteins allow plants and chordates to employ methylated H3K4 marks in different ways. It is tempting to draw a parallel to the phylogenetic distribution and usage of cytosine methylation in DNA, also absent from several lineages including insects and nematodes (Jeltsch, 2010; Zemach *et al*, 2010). Intriguingly, four of the CW proteins in *Arabidopsis* have methyl-CpG binding domains (see Table I).

Structure of the CW domain and its mode of interaction with histones

Comparing the structures of the CW domain of ASHH2 and that recently reported for ZCWPW1 (He *et al*, 2010) reveals that both domains share a common structural core built around two β -strands and a Zn^{2+} -binding site, reminiscent of PHD fingers. A cleft traverses one side of the domain, just underneath a pocket containing two conserved tryptophans forming an aromatic cage reminiscent of the aromatic cages of other methyllysine-binding domains, as on the PHD fingers, the chromodomains and in the bottom of the cavities in the MBT domains (Taverna *et al*, 2007). In ZCWPW1 CW, this is the binding site for the H3K4me3 peptide (He *et al*, 2010). It is reasonable to suggest that these conserved features form the binding site on ASHH2 CW. This is supported by both NMR spectrometry and site-directed mutagenesis (Figures 5 and 6).

For ZCWPW1 CW, it was also shown that the N-terminal amino group of the histone tail is interacting with an aspartate carbonyl oxygen. In ASHH2 CW, D869 is placed in an equivalent position. This may be a critical determinant for the CW domain's preference for N-terminal H3 tails. It is intriguing, though, that the lower part of the cleft, which interacts with the histone peptide, shows so much sequence variation (see Supplementary Figure S5; residues 870–880 in ASHH2). One explanation could be that some CW domains also recognize methylated lysines on non-histone proteins. Another possibility is that the CW domains are differentially

sensitive to other modifications on the histone tail (i.e., R2, T3, and T6).

The most remarkable feature of the two CW domain structures is, however, the structure of their non-conserved C-terminal extensions. Each subfamily of CW domains has a unique C-terminal extension, e.g. presenting a third tryptophan for the aromatic cage in ZCWPW1, while an amphipathic helix in ASHH2. Given our observations that different CW domains show different preferences for the three states of H3K4 methylation, we are led to propose that the family-specific C-terminal embellishments serve as determinants for recognition of the differentially methylated H3K4 tails, a novel feature among histone recognition modules.

ChIP analyses indicate that the major activity of ASHH2 is H3K36 tri-methylation

For the so far best-studied CW protein (and gene), *Arabidopsis* ASHH2, we can consider in more detail the functional implications of a H3K4me1/me2-reading module in conjunction with its role as a HMTase. Recently, experiments have indicated that ASHH2 can confer methylation of both H3K36 and H3K4 *in vitro* and that the FLC activator protein FRIGIDA (FRI) can stimulate the H3K4me3 activity (Ko *et al*, 2010). This may suggest that the reduction in H3K4me3 seen in the *ashh2* mutant near the TSS of the *MAF1* and *FLC* genes, could be conferred by ASHH2. However, this could also be an indirect effect of reduced transcriptional activity, especially since our experiments have been conducted in the Col ecotype, which is mutant for *FRI*.

It should be noted, though, that reduction of H3K4me3 and H3K36me3 for *ACTIN2* and *GAPA* was not accompanied by reduced expression levels. All seedling-expressed genes had lower levels of H3K36me3 in the mutant, but unexpectedly, *GAPA* and *AtDMC1* showed an increase in H3K36me2 methylation. This may suggest that *Arabidopsis* has another SET-domain protein that is responsible for di-methylation of H3K36, and/or that ASHH2 uses H3K36me2 as a substrate. In mutant inflorescences we did, however, not detect any changes in H3K36me2 in the tissue-specific genes expressed there (Grini *et al*, 2009). All in all our data support that the major activity of ASHH2 is H3K36 tri-methylation in *fri* background. Thus, ASHH2 seems to be a protein with an H3K4me1,2 reading module and a H3K36me3 writing module.

ASHH2 activity correlates with transcriptional output of tissue-specific and developmentally regulated genes

The importance of the K4me reading function of the CW domain is indicated by the underrepresentation of genes without K4me marks among putative ASHH2 target genes. Furthermore, the inflorescence-specific genes *API1*, *MYB99*, and *NAC25* are not affected by *ashh2* in seedlings where they are silent and marked with H3K27me3. For the expressed tissue-specific genes, mutation in *ashh2* leads to a reduction both in transcript levels and in H3K36me3 levels. Using global ChIP data, we have been able to survey the chromatin marks of a larger number of genes affected by mutation in *ashh2*. This analysis showed a significant overrepresentation of H3K4me3/H3K36me2/H3K27me3 triple-marked genes (Table II), which indicates tissue specificity or developmental regulation, with H3K27me3 associated with silent genes in cells where the gene is not expressed (Oh *et al*, 2008). Alternatively, the three marks might reside on the same

chromatin as a specific means of controlling expression of genes involved in differentiation. *FLC* is such an ASHH2-regulated gene with triple marks (Pien *et al*, 2008; Xu *et al*, 2008; Schmitz *et al*, 2009). Interestingly, genes encoding transcription factors, many of which are tissue specific, are overrepresented among genes depending on ASHH2 for maintenance of transcription levels in seedlings. A substantial number of transcription factors were also found down-regulated in *ashh2* inflorescences (Grini *et al*, 2009).

The CW domain may contribute to ASHH2's preference for genes with H3K4 methylation

In *Arabidopsis*, H3K4 methylation generally localizes to the promoters and transcribed regions of genes (Zhang *et al*, 2009). H3K4me3 is in particular associated with transcribed genes, and H3K4me2 often co-occurs with H3K4me3 in the 5'-end of genic regions, while H3K36me2 increases towards the 3'-end (Oh *et al*, 2008). H3K4me1 on the other hand, is found in internal regions especially in long genes (>4 kb) and correlates with CpG DNA methylation in transcribed regions (Zhang *et al*, 2009).

Our analyses suggest that ASHH2 has a strong preference for H3K4 methylated genes, especially those with combinations of K4 methylation marks (Supplementary Table SIII) associated with moderate expression levels and moderate tissue specificity (Zhang *et al*, 2009). Interestingly, ChPD indicated a reduction in H3K36me3 chromatin pulled down with the CW domain from the *ashh2* mutant, compared with the total *ashh2* chromatin. This suggests that H3K36me3 is generally associated with H3K4me1 which is the preferred target for ASHH2 CW. qPCR with DNA from ChPD by the CW domain showed that CW interacts with the supposed target genes that we have used in our study, including *FLC*, which has also been shown to bind ASHH2 in ChIP (Ko *et al*, 2010). The lowest qPCR levels were found for the genes that are not affected by mutation in *ASHH2* with respect to transcription

levels and chromatin marks. Furthermore, the profile of abundance of the putative target genes including *FLC* is similar in the ChPD, H3K4me1, and H3K36me3 ChIP experiments.

Therefore, a plausible model for ASHH2 function is that the CW domain first positions the protein near the TSS by binding to H3K4me2 (and/or weakly to H3K4me3), and that binding to K4me2, and in longer genes K4me1 along the body of the gene, is accompanied with H3K36me3 methylation (Figure 8). H3K4me2 often co-resides with the repressive mark H3K27me3 (Zhang *et al*, 2009), and we suggest that K36me3 marks are needed to maintain expression when such repressive marks are present. For genes with high expression levels and devoid of repressive marks, this maintenance function may not be needed. Therefore, reduction in H3K36me3 marks does not affect the transcription level of highly expressed genes like *ACTIN2* or *GAPA*.

The different specificities of the CW domains may contribute either to maintenance or to changes in gene expression and chromatin status

Intriguingly, different CW domains show preference for different methylation states of H3K4. We have argued above that, in the case of ASHH2, reading of K4me1 and me2 may direct ASHH2 HMTase activity to transcribed genes, which could contribute to sustained gene expression (Figure 8). VAL1 on the other hand, a transcriptional repressor of the related ABI3/FUS3/LEC2 (AFL) transcription factors controlling the maturation program of *Arabidopsis* seed development (Aichinger *et al*, 2009), has strongest preference for K4me3 found at TSSs. Thus, in the case of VAL1, the CW domain may, in contrast to CW in ASHH2, contribute to a switch from an expressed to a repressed chromatin status.

Future investigation will hopefully elucidate the relationship between CW specificity and the function of all CW proteins in chromatin maintenance and remodelling.

Figure 8 A model for ASHH2 function and the role of its CW domain. The CW domain of ASHH2 binds H3K4me2 found close to transcription start site (TSS) in both highly expressed housekeeping genes and weakly expressed tissue-specific genes, but not in silent genes. In addition CW binds H3K4me1 found along the transcribed gene regions. This facilitates H3K36me3 methylation both near the TSS and along the gene body. Loss of H3K36me3 has little consequences for the transcription level of highly expressed housekeeping genes. However, loss of H3K36me3 due to mutation in *ASHH2* results in non-sustainable transcription of genes that may carry both active and repressive chromatin marks. Due to the lack of H3K4me marks ASHH2 is not active on silent genes and thus there is no effect on transcription when ASHH2 is mutated. Icons for each of the histone modifications are shown to the right. The ASHH2 protein and its SET domain are shown in green, while the CW domain engaged in H3K4me binding is shown in yellow. RNA polymerase II occupancy is illustrated with blue pentagons, while the extent of transcription is indicated by arrows.

Among the most intriguing questions to address is the unusual phylogenetic distribution of this domain and whether this is related to DNA methylation.

Materials and methods

Microarray experiment and data analysis

Experiments were performed essentially as in Grini *et al* (2009) using five biological replicas with 8–10-days-old seedlings of both *Arabidopsis thaliana* plants Col ecotype and the *ashh2-1* mutant (identical to *sdg8-1*) (Grini *et al*, 2009).

All data are MIAME compliant and the raw data have been deposited in the MIAME compliant database GEO (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22990>).

We compared our microarray data with those of Cazzonelli *et al* (2009) and Xu *et al* (2008). The mutant alleles and conditions for these two sets were green rosette leaves from 10-day-old *ccr1-1* seedlings using ATH1 Arrays (Affymetrix) (Cazzonelli *et al*, 2009), and 6-day-old *sdg2-1* and *sdg2-2* seedlings analysed on the CATMA 24k chip (Xu *et al*, 2008). The gene lists were generated with log2 absolute value 0.7 for *ashh2-1* ($P < 0.01$), 0.8 for *sdg8-2* ($P < 0.05$), and 0.3 for *ccr1-1* ($P < 0.05$). The TAIR GO annotation tool (<http://www.Arabidopsis.org/tools/bulk/go/index.jsp>) was used to assign functional classes to the genes.

Real-time quantitative PCR (qPCR)

qPCR was performed essentially as in Grini *et al* (2009). Expression levels of target genes in the *ashh2* mutant were calculated relative to wt levels with normalization to *TUB8*. Primers are given in Supplementary Table SIV.

ChIP

Wt and *ashh2-1* mutant plants were cultivated in growth chambers at 20°C for 8 h of dark and 16 h of light ($100 \mu\text{E m}^{-2} \text{s}^{-1}$). For each experiment, 2–3 g of 15-days-old seedlings was crosslinked in 1% formaldehyde under vacuum until the tissue was translucent.

ChIP was done as described in Gendrel *et al* (2005). The antibodies used for immunoprecipitation were anti-H3K9me2 (#07-212, Millipore), anti-H3K4me3 (#07-473, Upstate), anti-H3K36me2 (#07-369, Upstate), or anti-H3K36me3 (ab9050, Abcam). Further details are given in the legends to Supplementary Figures S2 and S3.

Expression and purification of GST fusion proteins

For all GST fusion protein expression constructs the CW domains were cloned via *EcoRI* and *BamHI* restriction sites into pSXG vector (Ragvin *et al*, 2004). The ASHH2 (nts 2547–2811) and VAL1 (nts 1575–1797) CW domains were cloned by PCR from *Arabidopsis* cDNA, MORC4 (nts 1236–1422) and ZCWPW1 (nts 714–942) CW domains were cloned from HEK293 cDNA.

Protein expression was performed in YT-G medium supplemented by 2 μM Zn-acetate by incubation with 0.4 mM IPTG (Isopropyl β -D-1-thiogalactopyranoside) for 4 h at 26°C and purified by affinity chromatography using glutathione Sepharose as previously described (Ragvin *et al*, 2004).

Mutant versions of ASHH2 CW were generated by PCR using mutation-specific primers (Supplementary Table SIV) and subsequent annealing and primer extension to generate full-length, double-stranded mutant DNA. Mutant GST–CW proteins were cloned and expressed in pSXG as described above. All constructs were verified by DNA sequencing.

Histone peptide binding assays

Histone peptide binding assays were performed as described by Shi *et al* (2006) with biotinylated histone peptides from Upstate Biotechnology. The protein–peptide ratio used is indicated in the legends to the figures. Bound proteins were visualized by immunoblotting using rabbit anti-GST antibodies Z-5 (SC-456) from Santa Cruz at a 1:20 000 dilution, and a donkey anti-rabbit HRP conjugate (Amersham NA934) at a 1:10 000 dilution.

Surface plasmon resonance

Surface plasmon resonance binding assays were performed on a Biacore T100 biosensor according to the manufacturer's protocols

using immobilized, biotinylated H3 peptides mono-methylated (0.54 ng), di-methylated (0.24 ng), or tri-methylated (0.48 ng) on lysine 4. GST-tagged ASHH2 CW protein in five different concentrations in a range from 0.1 to 10 μM was injected for 2 min at a flow rate of 10 $\mu\text{l min}^{-1}$. Each sample injection was followed by injection of HBS-P buffer (10 mM HEPES pH 7; 150 mM NaCl) alone for 5 min at a flow rate of 5 $\mu\text{l min}^{-1}$. K_d values were obtained using the Biacore T100 Evaluation software 2.0.1. Measurements were repeated 2–5 times.

NMR spectroscopy

^{13}C - and ^{15}N -labelled ASHH2 CW was expressed, purified and subjected to NMR spectroscopy in the absence or presence of a histone H3K4me1 peptide as described in the legend to Supplementary Figure S4. The final structure ensemble is deposited in the protein data bank accession code 2L7P and the chemical shifts have been deposited in the Biological Magnetic Resonance Data Bank with the accession code 17365.

Structures were rendered in MacPyMOL (<http://www.pymol.org/>) and surface potentials were calculated and displayed with PMV-1.5.4 (<http://mgltools.scripps.edu/packages/pmv>).

Chromatin pulldown

ChPD was performed using crude chromatin in ChIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris–HCl pH 8, 167 mM NaCl) prepared as for ChIP (Gendrel *et al.*, 2005) incubated with 8 μg GST–CW_{ASHH2} fusion protein and control proteins (only GST alone, or together with GST–CW_{ASHH2}–W874A and GST–WYLD) overnight. For western blotting, the pulled down chromatin was washed three times in ChIP dilution buffer, run on a 15% SDS–PAGE and blotted onto a PVDF membrane. Blots were probed with the following antibodies: anti-H3 (ab1791; Abcam, 1:1000), anti-H3K4me1 (ab8895; Abcam, 1:1000), anti-H3K4me2 (07-030; Millipore, 1:1000), anti-H3K4me3 (ab8580; Abcam, 1:10 000) and anti-H3K36me3 (ab9050; Abcam, 1:2000). For ChPD followed by qPCR on ASHH2 target genes, the complete ChIP protocol (Gendrel *et al*, 2005) was used, exchanging antibodies with GST fusion proteins. Pulled down chromatin–CW complexes were eluted in a total of 250 μl elution buffer. The subsequent procedures were performed as for ChIP (see section above).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank Lill Knudsen, Roy Falleth, Charles Albin-Amiot, and Solveig H Engebretsen for technical assistance; Ole Horvli for help with Biacore; and Steven van Nocker for providing lists of genes enriched in H3K4me3, H3K27me3, and H3K36me2 according to Oh *et al* (2008). The work was facilitated by the services provided by the Norwegian *Arabidopsis* Research Centre (NARC, <http://www.narc.no/>), a part of the Research Council of Norway's National Program for Research in Functional Genomics (FUGE). Regarding the microarray experiment, we are grateful for the efforts of Vibeke Alm at the University of Oslo and Per Winge, Tommy S Jørstad, Torfinn Sparstad, Herman Midelfart, and Atle Bones at NARC. We also thank Drs Valeria De Marco, Corina Guder, and Gro EK Bjerga for helpful discussions. The Research Council of Norway has supported this work (grant 146652/431), and VH, TT, and SVV (grant 183609/S10). RA has had grant support from The Norwegian Cancer Society (grant PR-2006-0455) and from the University of Bergen.

Author contributions: The overall study was conceived and designed by RBA, RA, VH, and TT with important contributions from PEK. Experiments were performed by VH, TT, PEK, SVV, MAR, and KF, while VH, TT, RBA, and RA analysed the data, with contributions from SVV, MAR, and KF. RA and RBA wrote the paper with substantial contributions from VH, TT, and PEK.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Aichinger E, Villar CBR, Farrona S, Reyes JC, Hennig L, Köhler C (2009) CHD3 proteins and polycomb group proteins antagonistically determine cell identity in *Arabidopsis*. *PLoS Genet* **5**: e1000605
- Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (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
- Berg A, Meza TJ, Mahic M, Thorstensen T, Kristiansen K, Aalen RB (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
- Bonasio R, Lecona E, Reinberg D (2010) MBT domain proteins in development and disease. *Semin Cell Dev Biol* **21**: 221–230
- Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogson BJ (2009) Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell* **21**: 39–53
- Chakravarty S, Zeng L, Zhou MM (2009) Structure and site-specific recognition of histone H3 by the PHD finger of human auto-immune regulator. *Structure* **17**: 670–679
- Dong G, Ma D-P, Li J (2008) The histone methyltransferase SDG8 regulates shoot branching in *Arabidopsis*. *Biochem Biophys Res Commun* **373**: 659–664
- Fischer A, Hofmann I, Naumann K, Reuter G (2006) Heterochromatin proteins and the control of heterochromatic gene silencing in *Arabidopsis*. *J Plant Physiol* **163**: 358–368
- Gendrel A-V, Lippman Z, Martienssen R, Colot V (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat Methods* **2**: 213–218
- Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**: 1553–1577
- Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jørstad TS, Wilson ZA, Aalen RB (2009) The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis*. *PloS One* **4**: e7817
- He F, Umehara T, Saito K, Harada T, Watanabe S, Yabuki T, Kigawa T, Takahashi M, Kuwasako K, Tsuda K, Matsuda T, Aoki M, Seki E, Kobayashi N, Güntert P, Yokoyama S, Muto Y (2010) Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* **18**: 1127–1139
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LE, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318
- Imhof A (2006) Epigenetic regulators and histone modification. *Brief Funct Genomic Proteomic* **5**: 222–227
- Inoue N, Hess KD, Moreadith RW, Richardson LL, Handel MA, Watson ML, Zinn AR (1999) New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* **8**: 1201–1207
- Jeltsch A (2010) Molecular biology. Phylogeny of methylomes. *Science* **328**: 837–838
- Karytinos A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, Mattevi A (2009) A novel mammalian flavin-dependent histone demethylase. *J Biol Chem* **284**: 17775–17782
- Kim SY, He Y, Jacob Y, Noh YS, Michaels S, Amasino R (2005) Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyltransferase. *Plant Cell* **17**: 3301–3310
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187
- Ko JH, Mitina I, Tamada Y, Hyun Y, Choi Y, Amasino RM, Noh B, Noh YS (2010) Growth habit determination by the balance of histone methylation activities in *Arabidopsis*. *EMBO J* **29**: 3208–3215
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**: 693–705
- Liggins AP, Cooper CD, Lawrie CH, Brown PJ, Collins GP, Hatton CS, Pulford K, Banham AH (2007) MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. *Br J Haematol* **138**: 479–486
- Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury III WJ, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**: 762–767
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**: 373–375
- Muller J, Kassis JA (2006) Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr Opin Genet Dev* **16**: 476–484
- Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in *Arabidopsis*. *PLoS Genet* **4**: e1000077
- Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**: 100–103
- Perry J, Zhao Y (2003) The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. *Trends Biochem Sci* **28**: 576–580
- Pien S, Fleury D, Mylne JS, Crevillen P, Inze D, Avramova Z, Dean C, Grossniklaus U (2008) ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. *Plant Cell* **20**: 580–588
- Ragvin A, Valvatne H, Erdal S, rskog V, Tufteland KR, Breen K, yan AM, Eberharter A, Gibson TJ, Becker PB, Aasland R (2004) Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol* **337**: 773–788
- Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* **134**: 223–232
- Ruthenburg AJ, Li H, Patel DJ, Allis CD (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* **8**: 983–994
- Schmitz RJ, Tamada Y, Doyle MR, Zhang X, Amasino RM (2009) Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in *Arabidopsis*. *Plant Physiol* **149**: 1196–1204
- Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Cote J, Chua KF et al (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**: 96–99
- Spotswood HT, Turner BM (2002) An increasingly complex code. *J Clin Invest* **110**: 577–582
- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**: 41–45
- Suzuki M, Wang HHY, McCarty DR (2007) Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol* **143**: 902–911
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**: 1025–1040
- Thorstensen T, Fischer A, Grini PE, Sandvik SV, Johnsen SS, Reuter G, Aalen RB (2006) The *Arabidopsis* SUV4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. *Nucleic Acids Res* **34**: 5461–5470
- Turner BM (2007) Defining an epigenetic code. *Nat Cell Biol* **9**: 2–6

- Wu H, Min J, Lunin VV, Antoshenko T, Dombrowski L, Zeng H, Allali-Hassani A, Campagna-Slater V, Vedadi M, Arrowsmith CH, Plotnikov AN, Schapira M (2010) Structural biology of human H3K9 methyltransferases. *PLoS One* **5**: e8570
- Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou J-P, Steinmetz A, Shen W-H (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol Cell Biol* **28**: 1348–1360
- Yang Z, Jiang J, Stewart DM, Qi S, Yamane K, Li J, Zhang Y, Wong J (2010) AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. *Cell Res* **20**: 276–287
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919
- Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**: 258–262
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: R62
- Zhao Z, Yu Y, Meyer D, Wu C, Shen WH (2005) Prevention of early flowering by expression of *FLOWERING LOCUS C* requires methylation of histone H3 K36. *Nat Cell Biol* **7**: 1156–1160