

Regulation of chromatin structure by site-specific histone H3 methyltransferases

Stephen Rea*, Frank Eisenhaber*, Dónal O'Carroll*, Brian D. Strahl†, Zu-Wen Sun†, Manfred Schmid*, Susanne Opravil*, Karl Mechtler*, Chris P. Ponting‡, C. David Allis† & Thomas Jenuwein*

* Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, Dr. Bohrgasse 7, A-1030 Vienna, Austria

† Department of Biochemistry and Molecular Genetics, University of Virginia Health Science Center, Charlottesville, Virginia 22908, USA

‡ MRC, Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, OX1 3QX, UK

The organization of chromatin into higher-order structures influences chromosome function and epigenetic gene regulation. Higher-order chromatin has been proposed to be nucleated by the covalent modification of histone tails and the subsequent establishment of chromosomal subdomains by non-histone modifier factors. Here we show that human *SUV39H1* and murine *Suv39h1*—mammalian homologues of *Drosophila Su(var)3-9* and of *Schizosaccharomyces pombe clr4*—encode histone H3-specific methyltransferases that selectively methylate lysine 9 of the amino terminus of histone H3 *in vitro*. We mapped the catalytic motif to the evolutionarily conserved SET domain, which requires adjacent cysteine-rich regions to confer histone methyltransferase activity. Methylation of lysine 9 interferes with phosphorylation of serine 10, but is also influenced by pre-existing modifications in the amino terminus of H3. *In vivo*, deregulated *SUV39H1* or disrupted *Suv39h* activity modulate H3 serine 10 phosphorylation in native chromatin and induce aberrant mitotic divisions. Our data reveal a functional interdependence of site-specific H3 tail modifications and suggest a dynamic mechanism for the regulation of higher-order chromatin.

Higher-order chromatin is essential for epigenetic gene control and for the functional organization of chromosomes. Differences in higher-order chromatin structure have been linked with distinct covalent modifications of histone tails that regulate transcriptional 'on' or 'off' states^{1–3} and influence chromosome condensation and segregation^{4,5}. Post-translational modifications of histone N termini, particularly of H4 and H3, are well documented and have functionally been characterized as changes in acetylation^{1–3}, phosphorylation⁵ and, most recently, methylation^{6,7}. In contrast to the large number of described histone acetyltransferases (HATs) and histone deacetylases (HDACs), genes encoding enzymes that regulate phosphorylation^{8,9} or methylation⁶ of histone N termini are only now being identified. Moreover, the interdependence of the different histone tail modifications for the integration of transcriptional output or higher-order chromatin organization is currently not understood.

Genetic screens for suppressors of position effect variegation (PEV) in *Drosophila*¹⁰ and *S. pombe*¹¹ have identified a subfamily of about 30–40 loci, which are referred to as *Su(var)*-group¹² genes. Several histone deacetylases^{13,14}, protein phosphatase type 1 (ref. 15) and S-adenosyl methionine synthetase¹⁶ have been classified as belonging to the *Su(var)* group. In contrast, *Su(var)2-5* (which is allelic to *HPI*)¹⁷, *Su(var)3-7* (ref. 18) and *Su(var)3-9* (refs 19, 20) encode heterochromatin-associated proteins. *Su(var)* gene function thus suggests a model, in which modifications at the nucleosomal level may initiate the formation of defined chromosomal subdomains that are then stabilized and propagated by heterochromatic SU(VAR) proteins²¹.

Su(var)3-9 is dominant over most PEV modifier mutations¹⁹, and mutants in the corresponding *S. pombe clr4* gene²², disrupt heterochromatin association of other modifying factors and result in chromosome segregation defects²³. We isolated *Su(var)3-9* homologues from humans (*SUV39H1*)²⁴ and mice (*Suv39h1*)²⁴, and showed that they encode heterochromatic proteins that associate with mammalian HP1 (ref. 24). The SU(VAR)3-9 protein family combines two of the most evolutionarily conserved domains

of 'chromatin regulators': the chromo^{25,26} and SET²⁷ domains. Whereas the 60-amino-acid chromo domain represents an ancient histone-like fold²⁸ that directs euchromatic or heterochromatic localizations²⁹, the molecular role of the 130-amino-acid SET domain has remained unknown. Here we report that mammalian SUV39H1 or Suv39h proteins are SET-domain-dependent, H3-specific histone methyltransferases (HMTases) which selectively methylate lysine 9 (Lys9) of the H3 N terminus—a modification that appears intrinsically linked to the organization of higher-order chromatin.

The SET domain and methyltransferases

The evolutionarily conserved SET domain was initially characterized as a common motif in the PEV modifier SU(VAR)3-9 (ref. 19), the *Polycomb*-group protein E(Z)³⁰ and the *trithorax*-group protein TRX³¹. In addition to these founding *Drosophila* members, we detected SET domain motifs in over 140 gene sequences³² of diverse and even antagonistic functions ranging from *Saccharomyces cerevisiae* to man, but also including some bacteria and viruses. Using the SET domains of the SU(VAR)3-9 protein family as a starting alignment, we identified significant sequence and secondary-structure similarities to six plant methyltransferases (Fig. 1). Although some of these plant sequences have been classified as potential histone lysine N-methyltransferases, only one has been functionally characterized and was found to lack HMTase activity^{33,34}. All six plant sequences contain an insertion of about 100 amino acids in the middle of the SET domain and do not comprise a SET-domain-associated cysteine-rich region or the carboxy-terminal tail with its three conserved cysteines. Whereas the secondary structure and sequence similarities extend over the entire SET domain (data not shown), highest identities are found in two short amino-acid stretches in the C-terminal half of the SET domain core: NHSC and GE(x)₅Y (see Fig. 1). Our alignment suggests a bipartite structure of the SET domain, in which these two conserved motifs may be of particular functional importance.

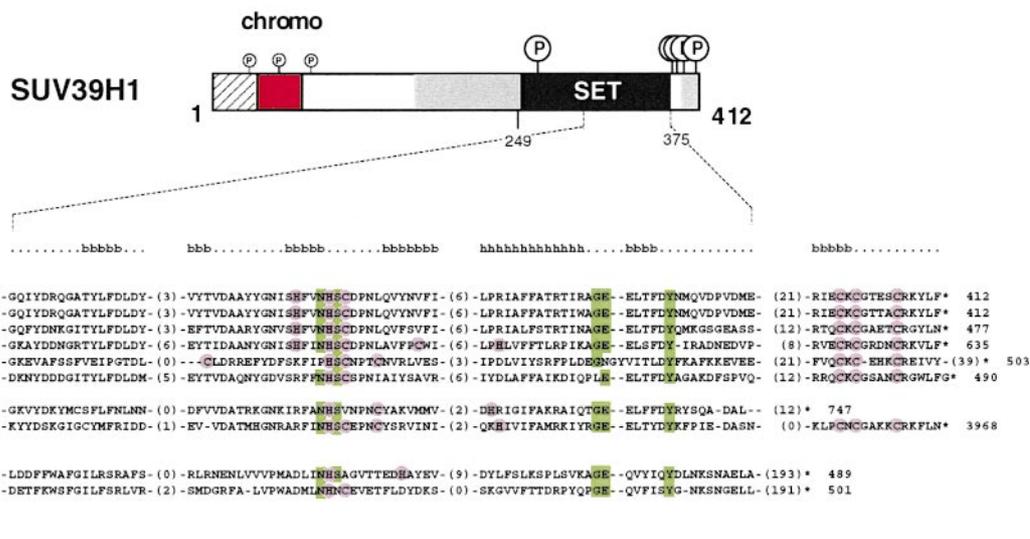


Figure 1 Sequence similarity of SET domains with plant methyltransferases. Top, the 412 amino acids of human SUV39H1 protein, highlighting the chromo domain (red box), the C-terminal SET domain (black box; residues 249–375) and SET domain-associated cysteine-rich regions (grey shading)^{24,35}. Bottom, amino-acid and secondary structure, β -sheet (b) or α -helix (h), similarities of the C-terminal halves of SET domain sequences from human SUV39H1 (ref. 24), murine Suv39h1 (ref. 24), murine Suv39h2 (D. O’C. *et al.*, manuscript in preparation), *Drosophila* SU(VAR)3-9 (ref. 19), a *C. elegans* SU(VAR)3-9-like

open reading frame C15H1.5 (CAB02737), *S. pombe* CLR4 (ref. 22), human EZH2 (ref. 36), the human trithorax homologue HRX³⁷ and MTases from *P. sativum* (rubisco ls-MT)^{33,34} and *A. thaliana* (O65218). The plant MTase sequences contain an insertion of about 100 amino acids in the middle of the SET domain (arrow). Cysteine and histidine residues are highlighted in pink, and identical amino-acid positions in at least 9 of the 10 SET domains are highlighted in green.

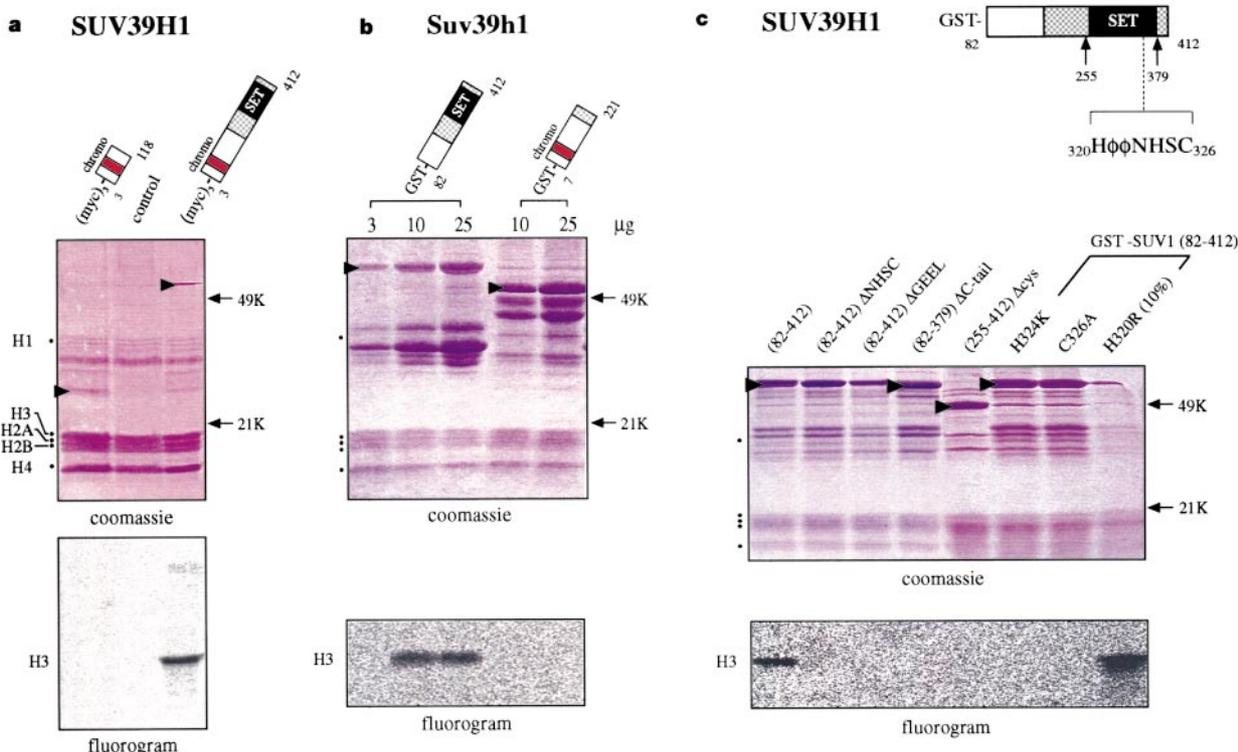


Figure 2 Histone methyltransferase activity of transfected and recombinant SUV39H1 and Suv39h1 proteins. **a**, Triple myc-tagged full-length human SUV39H1 (residues 3–412) or a C-terminally truncated SUV39H1 protein (residues 3–118) were immunoprecipitated from ‘stably’ transfected HeLa cell lines³⁵ and used in *in vitro* HMTase reactions with free histones as substrates and S-adenosyl-[methyl-¹⁴C]-L-methionine as methyl donor. Coomassie stain (top panel) shows purified proteins (arrowheads) and free histones (dots). Fluorography (bottom panel) indicates H3 HMTase activity of (myc)₃-SUV39H1.

b, Recombinant GST fusion proteins encoding different domains of murine Suv39h1 were used in increasing protein concentrations for *in vitro* HMTase reactions with free histones. **c**, About 10 μ g of the indicated fusion proteins encoding GST-SUV1 (82–412) (human SUV39H1) and seven GST-SUV39H1 mutants were analysed in *in vitro* HMTase reactions with free histones. For the hyperactive H320R mutant, only 1 μ g (10%) of the corresponding fusion product was used.

SUV39H1 and Suv39h1 possess HMTase activity

To investigate whether the SET domain of human SUV39H1 has enzymatic activity, we tested histones as possible substrates for *in vitro* methylation. Using HeLa cell lines 'stably' expressing³⁵ triple myc-tagged full-length SUV39H1 (residues 3–412), we enriched the ectopic protein from nuclear extracts by immunoprecipitation with anti-myc beads (Fig. 2a, top panel, arrowhead) and probed for activity to transfer a labelled methyl group from *S*-adenosyl-[methyl-¹⁴C]-L-methionine to free histones according to described conditions⁷. Reaction products were separated by polyacrylamide gel electrophoresis with SDS (SDS-PAGE) and visualized by fluorography, indicating selective transfer of the methyl label to H3 (see below). In contrast, no signals were detected with extracts from a HeLa cell line that expresses only the N-terminal third of SUV39H1 (residues 3–118) or with extracts from HeLa control cells.

To confirm that the HMTase activity is an intrinsic property of SUV39H1 and not mediated by a SUV39H1-associated factor, we repeated the *in vitro* HMTase reactions with recombinant products that were purified as glutathione-*S*-transferase (GST) fusion proteins from *Escherichia coli* (Fig. 2b, top panel, arrowheads). For this analysis, we used murine Suv39h1, which is 95% identical to human SUV39H1 (ref. 24). A purified GST product comprising residues 82–412 maintained HMTase activity (although at a reduced level as compared with transfected SUV39H1), whereas a purified GST product comprising residues 7–221 proved negative, even at higher protein concentrations (Fig. 2b, lower panel).

A catalytic motif in the SET domain

Similar to the recombinant murine GST–Suv1(82–412) product, the corresponding human SUV39H1 fusion protein, GST–SUV1(82–412), is catalytically active (Fig. 2c). We introduced short internal deletions (Δ NHSCDPN_{323–329}; Δ NHSC and Δ GEELTFDY_{358–365} Δ GEEL) into the two conserved regions of the SET domain core in GST–SUV1(82–412), and also generated mutants that lack the C-terminal tail (Δ C-tail) or the SET-associated cysteine-rich region (Δ cys). None of the mutant proteins showed HMTase activity.

These results suggest that the cysteine-rich regions have a significant contribution, although their apparent absence in the plant methyltransferases does not prevent catalytic activity. To investigate enzyme function of the SET domain in more detail, we therefore introduced point mutations into the most highly conserved ³²⁰H ϕ ϕ NHSC₃₂₆ motif (where ϕ are hydrophobic residues) of GST–SUV1(82–412). In particular, we mutated the invariant histidine 324 to leucine (H324L; data not shown) or lysine (H324K), and also changed cysteine 326 to alanine (C326A) or replaced histidine 320 by arginine (H320R). *In vitro* HMTase assays indicated that all point mutations, with the exception of H320R, abolished enzymatic activity. Unexpectedly, the H320R mutation resulted in an hyperactive enzyme with activity increased 20-fold or more (Fig. 2c). These data define the ³²⁰H ϕ ϕ NHSC₃₂₆ motif in the SET domain as an important catalytic site.

HMTase activity within the SU(VAR)3-9 family

Because the SET domain is one of the most conserved protein motifs in chromatin regulators^{27,31}, we analysed whether SU(VAR)3-9 family members or other SET domain proteins contain HMTase activity. We generated GST fusion products of the extended SET domains of *S. pombe* CLR4 (ref. 22), human EZH2 (ref. 36) and human HRX³⁷ that would correspond to GST–SUV1(82–412) (Fig. 3a). GST–CLR4(127–490) displayed pronounced HMTase activity at 3–5-fold increased levels (Fig. 3b) compared with the recombinant SUV39H1 product, consistent with CLR4 carrying an arginine at the hyperactive position (see Fig. 1). In contrast, both GST–EZH2(382–747) and GST–HRX(3643–3966) had undetectable HMTase activity towards free histones (Fig. 3b), whereas a comparable GST product generated from the murine *Suv39h2* gene (D. O’C.

et al., manuscript in preparation), GST–Suv2(157–477), was as active as GST–SUV1(82–412). EZH2 lacks the C-terminal cysteines, and HRX does not contain the SET domain-associated cysteine-rich region. Both of these cysteine domains are present in CLR4, Suv39h2 and SUV39H1. In agreement with the mutational analysis of SUV39H1, it thus appears that HMTase activity towards free histones requires the combination of the SET domain with adjacent cysteine-rich regions, and is a quality found in only a restricted number of proteins containing SET domains.

H3 Lys 9 is the main site for the Suv39h1 HMTase

The above experiments indicated that the HMTase activity of mammalian SU(VAR)3-9 related proteins is selective for H3 under our assay conditions. To examine this finding in more detail, we carried out *in vitro* methylation reactions with individual histones, using GST–Suv1(82–412) as enzyme. H3 is specifically methylated, whereas no signals are detected with H2A, H2B or H4 (Fig. 4a). A weak signal is present if H1 was used as the sole substrate; the significance of H1 methylation remains to be determined.

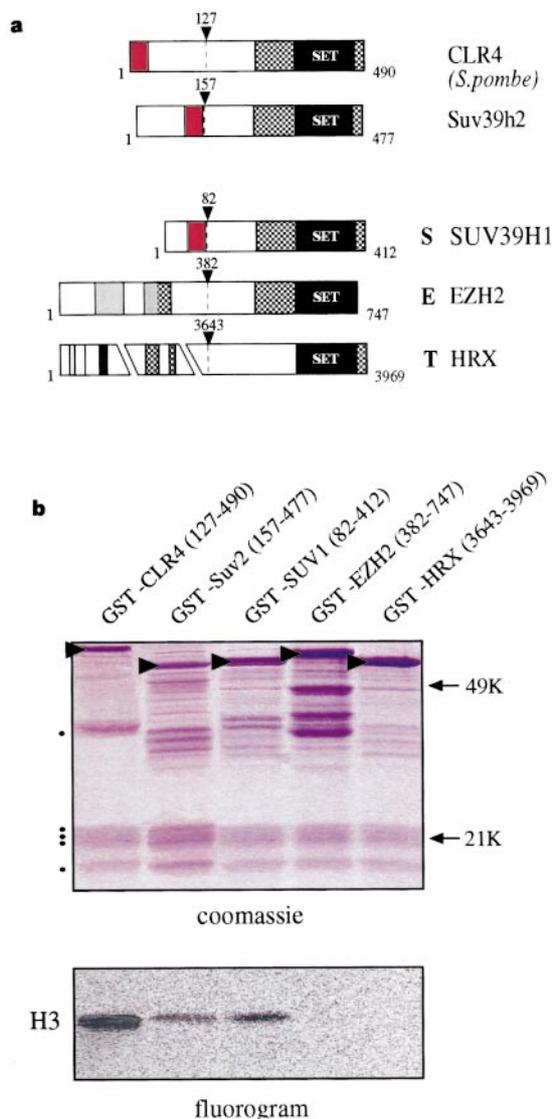


Figure 3 Specific HMTase activity of SU(VAR)3-9 related proteins. **a**, Diagram representing the domain structures of *S. pombe* CLR4, murine Suv39h2, human SUV39H1, human EZH2 and human HRX proteins. Arrowheads indicate the N-terminal fusion to GST. Cysteine-rich regions are indicated by grey stippling. **b**, About 10 μ g of the indicated GST fusion proteins were analysed in *in vitro* HMTase reactions with free histones. Coomassie stain (top panel) shows purified proteins (arrowheads) and free histones (dots). The murine Suv39h2 product is abbreviated as GST–Suv2(157–477).

Methylation of H3 occurs predominantly at Lys 4 in a wide range of organisms, as well as at Lys 9 in HeLa cells, although the responsible HMTase(s) have not been defined⁷. To investigate the methylation-site profile of Suv39h1, we tested unmodified peptides as substrates encoding the wild-type H3 N terminus (residues 1–20) and a mutant K9L peptide, changing Lys 9 to leucine. In addition, we included insulin and peptides encoding the N termini of CENP-A³⁸ and macroH2A³⁹. Peptides were methylated *in vitro* by GST–Suv1(82–412), and reaction products were separated by high percentage SDS–PAGE and visualized by fluorography. These *in vitro* assays revealed selective methylation of the wild-type H3 peptide, whereas no signals were detected with the CENP-A or macroH2A peptides, or with insulin (Fig. 4b). Notably, the mutated H3 (K9L) peptide was not a substrate, suggesting that Lys 9 of the H3 N terminus is a preferred residue for Suv39h1-dependent HMTase activity.

To determine more definitively this site preference, the wild-type H3 N-terminal peptide was methylated *in vitro* by GST–Suv1(82–

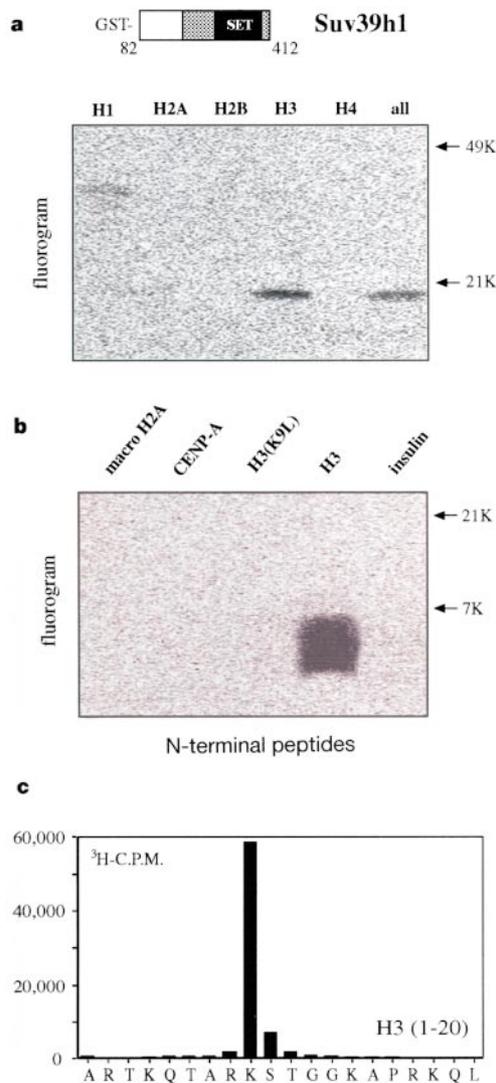


Figure 4 Lysine 9 of the H3 N terminus is the principal site for *in vitro* methylation by recombinant Suv39h1. **a**, About 10 μg of GST–Suv1(82–412) (murine Suv39h1) were used in *in vitro* HMTase reactions with individual histones. Only the fluorogram is shown. **b**, *In vitro* methylation assays using GST–Suv1(82–412) as enzyme and the indicated N-terminal peptides of wild-type H3, mutated H3(K9L), CENP-A, macroH2A or insulin as substrates. **c**, Automated sequencing of the wild-type H3 N-terminal peptide (residues 1–20) that had been methylated *in vitro* by recombinant GST–Suv1(82–412). The tritium incorporation of individual amino acids identified at each successive round of microsequencing is shown.

412), using S-adenosyl-[methyl-³H]-L-methionine. The labelled peptide, purified by reverse-phase high performance liquid chromatography, was then directly microsequenced, and ³H-incorporation associated with each individual amino acid was analysed by scintillation counting. The results confirmed selective transfer of methyl-label to Lys 9 (Fig. 4c), showing that Suv39h1 is a highly site-specific HMTase for the H3 N terminus *in vitro*.

Lys 9 methylation and Ser 10 phosphorylation

Position-specific modifications of the H3 N terminus have been correlated with distinct chromatin-based outputs, such as transcriptional regulation (Lys 14 acetylation), histone deposition (Lys 9 acetylation) and chromosome condensation/segregation (Ser 10 phosphorylation)¹. We analysed whether methylation of Lys 9 would possibly interfere with this ‘histone code’ and whether it might be influenced by pre-existing modifications at these sites. We performed *in vitro* HMTase assays with GST–SUV1(82–412) and H3 N-terminal peptides that were either unmodified, phosphorylated at Ser 10 (S10-phos), di-methylated at Lys 9 (K9-dimeth), or acetylated at Lys 9 (K9-Ac) or Lys 14 (K14-Ac). We visualized the processed reaction products by fluorography and quantified them with a Phosphorimager, setting the amount of SUV39H1-dependent methylation of the unmodified peptide as 100% (Fig. 5a). As expected, a Lys 9 dimethylated position severely compromised enzymatic activity, and a Lys 9 trimethylated peptide was no longer a substrate (data not shown). Similarly, pre-existing acetylation at Lys 9 prevented SUV39H1-dependent methylation, whereas acetylation at Lys 14 did not significantly interfere. Unexpectedly, the S10-phos peptide could also not be methylated, suggesting that Ser 10 phosphorylation inhibits SUV39H1-dependent methylation of the adjacent Lys 9.

A mitotic kinase (Ipl1/aurora) that selectively phosphorylates Ser 10 in H3 has recently been identified⁹. We used recombinant Ipl1/aurora kinase to examine a potential reciprocal interference of Ser 10 phosphorylation by pre-existing methylation at Lys 9 with our set of modified H3 N-terminal peptides. These *in vitro* kinase assays indicated that the K14-Ac peptide represents an even better substrate than the unmodified N terminus, whereas a Ser 10 phosphorylated position allowed only residual activity. Notably, di-methylation at Lys 9, but not acetylation at Lys 9, significantly reduced the substrate quality of the H3 N terminus, and 35% or less of Ipl1/kinase activity was detected with the K9-dimeth peptide (Fig. 5b). Together with the above enzymatic analyses of SUV39H1 HMTase preference, these data provide strong experimental evidence for an interdependence of site-specific histone tail modifications.

Modification of H3 in native chromatin

Murine *Suv39h* genes are encoded by two loci (*Suv39h1* and *Suv39h2*; D.O’C. *et al.*, manuscript in preparation), both of which have recently been disrupted in the mouse germ line. Although single *Suv39h1* and *Suv39h2* null mice are viable, double *Suv39h* deficient mice are born at only about 20–25% of the expected mendelian ratios and are growth retarded, showing that *Suv39h* has an important function during mouse development (D.O’C. *et al.*, manuscript in preparation). To begin to uncover the role(s) of Suv39h-dependent H3 methylation *in vivo*, we generated primary mouse embryonic fibroblasts (PMEFs) from *Suv39h* double null fetuses. Cell-cycle profiles of wild-type and *Suv39h* double null PMEFs indicated that similar percentages of cells are in S and G2/M phases (Fig. 6a), but that *Suv39h* double null PMEFs have a reduced G1 index and an increased proportion of cells with aberrant nuclear morphologies, reminiscent of division defects during mitosis. For example, *Suv39h* double null PMEFs contain twofold or more elevated numbers of cells with micro- and polynuclei, and are further characterized by cell subpopulations with oversized nuclei or a weak definition of heterochromatin that appears in only a few unusually condensed foci (Fig. 6b).

Phosphorylation at Ser 10 in the N-terminal tail of H3 (phosH3) is required for condensation and subsequent segregation of chromosomes⁵. During the cell cycle, phosH3 initiates within pericentric heterochromatin in late G2 and then progresses along

the entire chromosomes during mitosis⁴⁰. In wild-type PMEFs, about 7% of cells stain positive for the characteristic, heterochromatin-associated phosH3 foci, as detected by indirect immunofluorescence with anti-phosH3 antibodies (Fig. 6c, right panel). In

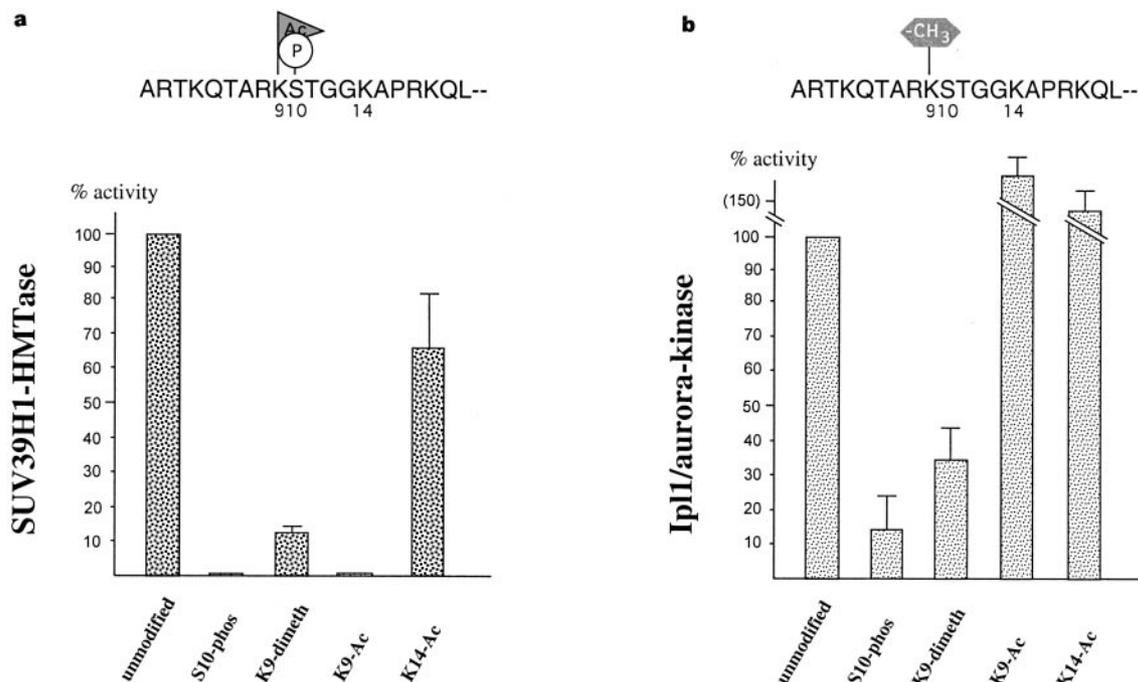


Figure 5 Interdependence of Lys 9 methylation and Ser 10 phosphorylation in the H3 N terminus. **a**, *In vitro* HMTase assays with GST-SUV1(82–412) and H3 N-terminal peptides that are either unmodified or phosphorylated (S10-phos), di-methylated (K9-dimeth) or acetylated (K9-Ac or K14-Ac) at the indicated positions. **b**, *In vitro* kinase

assays with recombinant His₁₀-Ipl1/aurora and the same set of H3 N-terminal peptides. For **a** and **b**, the percentage of enzyme activity detected with each individual peptide is plotted, setting the reaction with unmodified H3 peptide to 100%. For both enzymes, inhibiting modifications in the H3 N terminus are shown on top of the respective panels.

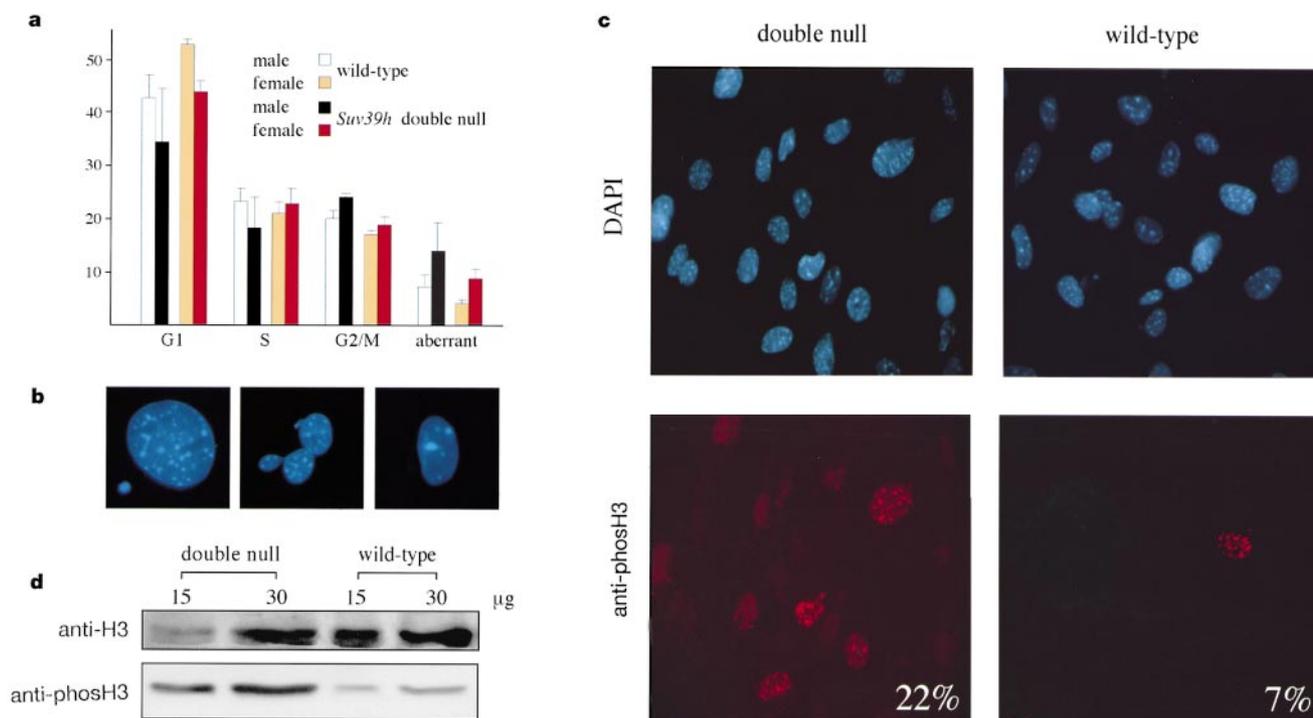


Figure 6 *Suv39h*-dependent modification of the H3 N terminus in native chromatin. **a**, Percentages of cell-cycle stages of passage 3 wild-type and *Suv39h* double null primary mouse embryonic fibroblasts (PMEFs). **b**, Representative images of aberrant mitoses in *Suv39h* double null PMEFs detected by β -tubulin (not shown) and DAPI staining (blue). Also shown (right image) is a nucleus exemplifying the unusual definition of heterochromatin in a subpopulation of *Suv39h* double null PMEFs. Original magnification

of all images, $\times 630$. **c**, Interphase chromatin staining with anti-phosH3 antibodies and CY3-conjugated secondary antibodies (red). DNA was counterstained with DAPI (blue). Roughly 1,000 cells were counted to evaluate the percentage (shown bottom right) of anti-phosH3 positive cells. **d**, Quantitative western analysis with 15 μ g and 30 μ g of total nuclear proteins immuno-blotted with anti-H3 and anti-phosH3 antibodies.

contrast, in *Suv39h* double null PMEfs about 22% of cells contain phosH3-positive foci (Fig. 6c, left panel), although their definition appears in many small speckles that do not always overlap with DAPI-dense material (data not shown). This result suggested that the overall levels of phosH3 may be enhanced in *Suv39h* double null PMEfs. We therefore determined the relative abundance of phosH3 in precalibrated nuclear extracts with anti-phosH3-specific antibodies. This quantification indicated a significantly higher level of phosH3 in *Suv39h* double null cells than in wild-type controls (Fig. 6d). These *in vivo* data complement our enzymatic *in vitro* analyses, and confirm a model in which methylation of Lys 9 in H3 impairs phosphorylation of adjacent Ser 10 at specific regions in native chromatin.

HMTase versus MTase

We have shown that SU(VAR)3-9-related proteins possess a selective H3 HMTase activity towards free histones *in vitro*. This H3 HMTase activity, however, was not detected with other constructs containing SET domains, for example EZH2 or HRX. Although both EZH2 and HRX contain clustered cysteines in their N-terminal halves^{30,31,36,37}, they lack either the C-terminal tail (EZH2) or the SET domain-associated cysteine-rich region (HRX). As observed for HATs, it is likely that the targets for SET-domain-dependent methyltransferase (MTase) activity are probably not restricted to histones. For example, the functionally characterized *Pisum sativum* lysine MTase methylates the large subunit of a metabolic enzyme^{33,34}, and an *S. cerevisiae* gene encoding a cytochrome *c* MTase activity⁴¹ also contains a SET domain (see Methods). We therefore propose a generic MTase quality for the SET domain, whose substrate specificity would largely depend upon combinations with additional protein modules. Alternatively, some SET domains may intrinsically lack MTase activity, similar to described anti-phosphatases that can counteract the roles of the catalytically active enzyme(s)⁴². This interpretation is particularly intriguing, given that an anti-phosphatase has been shown to be a SET domain binding factor^{42,43}, and that the SET domain is present in chromatin regulators of apparently antagonistic functions^{19,27,30,31}.

A 'histone code' for higher-order chromatin

SUV39H1 and *Suv39h1* are modular proteins that are enriched at heterochromatin^{24,35}, thereby directing their highly selective H3 HMTase activity to localized chromatin regions. Indeed, Lys 9 methylation appears nearly unchanged in bulk H3 preparations from nuclear extracts of wild-type and *Suv39h* double null PMEfs (data not shown), indicating that Lys 9 is more widely methylated by additional HMTases. Unexpectedly, methylation of Lys 9 interferes with Ipl1/aurora-dependent phosphorylation of Ser 10, but is also inhibited by pre-existing modifications in the H3 N terminus. These data argue that the *in vivo* substrate specificity for *Suv39h*-dependent methylation would be maximal for an unmodified H3 N terminus, which is primarily presented after DNA replication and nucleosome assembly. Notably, chromatin assembly factor 1 interacts with murine HP1 proteins⁴⁴, which have previously been shown to complex with SUV39H1 or *Suv39h1* (ref. 24).

In addition to modulating phosH3 distribution, overexpression of SUV39H1 also induces ectopic heterochromatin³⁵, whereas disruption of *Suv39h* function prevents heterochromatin association of HP1 proteins (data not shown), suggesting that Lys 9 methylation in H3 might impart a preferred binding surface for heterochromatin-specific proteins. This 'heterochromatic' affinity would be highest in the absence of additional modifications and may even be self-propagating within transcriptionally less-permissive chromatin regions, consistent with a maximal H3 HMTase activity of SUV39H1 towards an unmodified H3 N terminus. Our data predict that a regional increase in the methylation profile of Lys 9 in H3, probably in conjunction with a reduced acetylation status of the H4 N terminus¹⁻³, represents a key determinant in defining a 'histone

code' for heterochromatin. On the basis of our results, we propose that SU(VAR)3-9 related proteins provide important enzymatic activities towards the induction and assembly of higher-order chromatin. □

METHODS

Sequence alignments and secondary-structure predictions

The SET domains of human SUV39H1 (ref. 24), *Drosophila* SU(VAR)3-9 (ref. 19) and *S. pombe* CLR4 (ref. 22) were used as a multiple-starting alignment for database similarity searches using Profile⁴⁵, hidden Markov⁴⁶ and position-specific iterative BLAST⁴⁷ methods (representative listings are available from the SET domain page of the SMART WWW server²³). These searches revealed significant similarities to six plant proteins (GenBank accession numbers Q43088, O65218, P94026, O80013, AAC29137 and AC007576_12) described as putative lysine N-methyltransferases. For example, a PSI-BLAST search using the *S. pombe* hypothetical protein SPAC3c7.09 as query identified these plant sequences and well known SET domain sequences within ten rounds using an E-value inclusion threshold of 0.001. The same search also revealed the presence of a SET domain in YHR109w (which is known to encode a cytochrome *c* MTase⁴¹) within three rounds. Consensus secondary structures were predicted by described algorithms⁴⁸.

Epitope-tagged SUV39H1 proteins in HeLa cells

The HeLa cell lines overexpressing full-length (myc)₃-SUV39H1 (3-412) or (myc)₃-Nchromo (3-118) have been described³⁵. Nuclear extracts were immunoprecipitated with anti-myc antibody beads²⁴, and about 1-3 µg of matrix-bound (myc)₃-tagged SUV39H1 proteins were used for *in vitro* HMTase assays.

Generation and purification of GST fusion proteins

The GST-Suv1(82-412) product expressed from the pGEX-2T vector (Pharmacia) has been described²⁴. Additional GST constructs were generated by transferring *Bam*HI-EcoRI PCR amplicons into pGEX-2T, encoding in-frame fusions for Suv1(7-221), SUV1(82-412), SUV1(82-378)ΔC-tail, SUV1(255-412)Δcys, Suv2(157-477) (D. O'C. *et al.*, manuscript in preparation), CLR4(127-490)²², EZH2(382-747)³⁶ and HRX(3643-3969)³⁷. Short internal deletions (ΔNHSCDPN³²³⁻³²⁹ and ΔGEELTFDY³⁵⁸⁻³⁶⁵) or point mutations (see Fig. 2c) within the ³²⁰HφφNHSC₃₂₆ motif were directly engineered in the GST-SUV1(82-412) plasmid by double PCR mutagenesis. All constructs were confirmed by sequencing.

Recombinant proteins were expressed in 11 cultures of *E. coli* strain BL21 and solubilized in 10 ml RIPA buffer (20 mM Tris pH 7.5, 500 mM NaCl, 5 mM EDTA, 1% NP-40 and 0.5% sodium deoxycholate containing a full set of protease inhibitors (Boehringer Mannheim) and lysozyme (5 mg ml⁻¹; Sigma)) by freeze-thawing in liquid N₂, followed by sonication. Soluble proteins were cleared by centrifugation, purified with 800 µl glutathione Sepharose beads (Pharmacia) and washed twice in RIPA buffer. Protein concentration was determined by Coomassie staining of SDS-PAGE gels. Matrix-bound fusion proteins were used immediately for *in vitro* HMTase assays or stored at 4 °C.

In vitro histone methyltransferase assay

In vitro HMTase reactions were modified on the basis of described protocols⁷ and carried out in a volume of 50 µl of methylase activity buffer (MAB: 50 mM Tris pH 8.5, 20 mM KCl, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 250 mM sucrose), containing 10 µg of free histones (mixture of H1, H3, H2B, H2A and H4; Boehringer Mannheim) as substrates and 300 nCi S-adenosyl-[methyl-¹⁴C]-L-methionine (25 µCi ml⁻¹) (Amersham) as methyl donor. We routinely used 10 µg of matrix-bound GST fusion proteins to assay for HMTase activity. After incubation for 60 min at 37 °C, reactions were stopped by boiling in SDS loading buffer, and proteins were separated by 15% or 18% SDS-PAGE and visualized by Coomassie staining and fluorography.

HMTase assays with individual histones (Boehringer Mannheim), insulin (Sigma) or N-terminal peptides were performed with 5 µg of substrate. We used the following peptides: wild-type N terminus of human histone H3 (ARTKQTARKSTGGKAPRKQL) and mutant peptide which changes Lys 9 (underlined) to leucine; N terminus of human CENP-A (MGPRRRSRKPEAPRRRSPSP)³⁸; N terminus of rat macro-H2A (MSSRGKKKTKTSRSKAKG)³⁹. Modified H3(1-20) peptides (see Fig. 5) were synthesized using phosphorylated serine, di- or tri-methylated lysine and acetylated lysine (all Bachem or NovaBiochem).

Peptide microsequencing of the *in vitro* methylated wild-type H3 N-terminal peptide and determination of ³H-incorporation of individual amino acids by scintillation counting was done as described⁷.

In vitro Ipl1/aurora kinase assay

Purification of recombinant His₁₀-Ipl1 and analysis of kinase assays with H3 N-terminal peptides by scintillation counting was done as described⁹.

Generation and analysis of *Suv39h* double null PMEfs

Primary mouse embryonic fibroblasts were derived from day E12.5 *Suv39h* double null fetuses obtained after intercrossing *Suv39h1*^{-/-}/*Suv39h2*^{-/-} compound mutant mice (D.O'C. *et al.*, manuscript in preparation). As controls, PMEfs were prepared from wild-type fetuses of the same genetic background. For cell-cycle profile analysis by FACS, passage three PMEfs were analysed as described⁴⁹. Staining of unextracted PMEf

interphase chromatin with anti-phosH3 (ref. 40) antibodies and counterstaining of DNA with 4',6'-diamidino-2-phenylindole (DAPI; Sigma) was done as described³⁵. For the biochemical analysis, total nuclear extracts were precalibrated by Ponceau staining, immuno-blotted with anti-H3 (Upstate Biotechnology) and anti-phosH3 (ref. 40) antibodies and visualized by peroxidase staining using enhanced chemiluminescence (Amersham).

Received 24 February; accepted 9 June 2000.

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
2. Grunstein, M. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* **93**, 325–328 (1998).
3. Turner, B. M. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cell. Mol. Life Sci.* **54**, 21–31 (1998).
4. Karpen, G. & Allshire, R. C. The case for epigenetic effects on centromere identity and function. *Trends Genet.* **13**, 489–496 (1997).
5. Wei, Y., Lanlan, Y., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**, 99–109 (1999).
6. Chen, D. *et al.* Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174–2177 (1999).
7. Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl Acad. Sci. USA* **96**, 14967–14972 (1999).
8. Sassone-Corsi, P. *et al.* Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* **285**, 886–891 (1999).
9. Hsu, J.-Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 1–20 (2000).
10. Reuter, G. & Spierer, P. Position-effect variegation and chromatin proteins. *BioEssays* **14**, 605–612 (1992).
11. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–233 (1995).
12. Wallrath, L. L. Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**, 147–153 (1998).
13. De Rubertis, F. *et al.* The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**, 589–591 (1996).
14. Grewal, S. I., Bonaduce, M. J. & Klar, A. J. Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. *Genetics* **150**, 563–576 (1999).
15. Baksa, K. *et al.* Mutations in the protein phosphatase 1 gene at 87B can differentially affect suppression of position effect variegation and mitosis in *Drosophila melanogaster*. *Genetics* **135**, 117–125 (1993).
16. Larsson, J., Zhang, J. & Rasmuson-Lestander, Å. Mutations in the *Drosophila melanogaster* gene encoding S-adenosyl methionine suppress position-effect variegation. *Genetics* **143**, 887–896 (1996).
17. Eissenberg, J. C., Morris, G. D., Reuter, G. & Hartnett, T. The heterochromatin-associated protein HP-1 is an essential protein in *Drosophila* with dosage-dependent effects on position-effect variegation. *Genetics* **131**, 345–352 (1992).
18. Cléard, F., Delattre, M. & Spierer, P. SU(VAR)3-7, a *Drosophila* heterochromatin-associated protein and companion of HP1 in the genomic silencing of position-effect variegation. *EMBO J.* **16**, 5280–5288 (1997).
19. Tschiersch, B. *et al.* The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13**, 3822–3831 (1994).
20. Schotta, G. & Reuter, G. Controlled expression of tagged proteins in *Drosophila* using a new modular P-element vector system. *Mol. Gen. Genet.* **262**, 916–920 (2000).
21. Henikoff, S. in *Epigenetic Mechanisms of Gene Regulation* (eds Russo, V. A. E., Martienssen, R. A. & Riggs, A. D) 319–334 (CSHL, New York, 1996).
22. Ivanova, A. V., Bonaduce, M. J., Ivanov, S. V. & Klar, A. J. S. The chromo and SET domains of the *Clr4* protein are essential for silencing in fission yeast. *Nature Genet.* **19**, 192–195 (1998).
23. Ekwall, K. *et al.* Mutations in the fission yeast silencing factors *clr4^Δ* and *riki1^Δ* disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *J. Cell Sci.* **109**, 2637–2648 (1996).
24. Aagaard, L. *et al.* Functional mammalian homologues of the *Drosophila* PEV modifier *Su(var)3-9* encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
25. Aasland, R. & Stewart, A. F. The chromo shadow domain, a second chromodomain in heterochromatin-binding protein1, HP1. *Nucleic Acids Res.* **23**, 3168–3173 (1995).
26. Koonin, E. V., Zhou, S. & Lucchesi, J. C. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic Acids Res.* **23**, 4229–4232 (1995).

27. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET-domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* **54**, 80–93 (1998).
28. Ball, L. J. *et al.* Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO J.* **16**, 2473–2481 (1997).
29. Platero, J. S., Harnett, T. & Eissenberg, J. C. Functional analysis of the chromo domain of *HP-1*. *EMBO J.* **14**, 3977–3986 (1995).
30. Jones, R. S. & Gelbart, W. M. The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol. Cell. Biol.* **13**, 6357–6366 (1993).
31. Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V. & Harte, P. J. The *Drosophila trithorax* protein contains a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52**, 209–223 (1995).
32. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234 (2000).
33. Klein, R. R. & Houtz, R. L. Cloning and developmental expression of pea ribulose-1,5-biphosphate carboxylase/oxygenase large subunit N-methyltransferase. *Plant Mol. Biol.* **27**, 249–261 (1995).
34. Zheng, Q., Simel, E., Klein, P., Royer, M. & Houtz, R. L. Expression, purification, and characterization of recombinant ribulose-1,5-biphosphate carboxylase/oxygenase large subunit N^ε-methyltransferase. *Protein Exp. Purif.* **14**, 104–112 (1998).
35. Melcher, M. *et al.* Structure–function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation and mitotic progression. *Mol. Cell. Biol.* **20**, 3728–3741 (2000).
36. Laible, G., *et al.* Mammalian homologues of the Polycomb-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. *EMBO J.* **16**, 3219–3232 (1997).
37. Tkachuk, D. C., Kohler, S. & Cleary, M. L. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691–700 (1992).
38. Sullivan, K. F., Hechenberger, M. & Masri, K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* **127**, 581–592 (1994).
39. Pehrson, J. R. & Fried, V. A. MacroH2A, a core histone containing a large nonhistone region. *Science* **257**, 1398–1400 (1992).
40. Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).
41. Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
42. Cui, X. *et al.* Association of SET domain and myotubularin-related proteins modulates growth control. *Nature Genet.* **18**, 331–337 (1998).
43. Firestein, R., Cui, X., Huie, P. & Cleary, M. L. SET domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of *Drosophila* Su(var)3-9. *Mol. Cell. Biol.* **20**, 4900–4909 (2000).
44. Murzina, N., Verrault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
45. Birney, E., Thompson, J. D. & Gibson, T. J. PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res.* **24**, 2730–2739 (1996).
46. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
47. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **17**, 3389–3402 (1997).
48. Frishman, D. & Argos, P. Seventy-five percent accuracy in protein secondary structure prediction. *Protein Struct. Funct. Genet.* **27**, 329–335 (1997).
49. Xu, X. *et al.* Centrosome amplification and a defective G₂-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).

Acknowledgements

We would like to thank M. Cleary for providing a partial *HRX* cDNA, R. Allshire for the *Clr4* cDNA, I. Gorny for peptide synthesis and R. G. Cook for automated sequencing of the H3 N-terminal peptide. During the course of this work, E. V. Koonin and L. Aravind have independently discovered homology of the SET domain with plant methyltransferases. We acknowledge M. Doyle for the contribution to the *Suv39HZ* knockout and thank K. Nasmyth and M. Busslinger for helpful comments and critical reading of the manuscript. Research in the laboratory of C.D.A. is funded by grants from the NIH to C.D.A. and B.D.S. Z.-W.S. is supported by a postdoctoral cancer training grant from the University of Virginia Cancer center. Research in the laboratory of T.J. is supported by the IMP, the Austrian Research Promotion Fund and the Vienna Business Agency.

Correspondence and requests for materials should be addressed to T.J. (e-mail: jenuwein@nt.imp.univie.ac.at).