THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 10, Issue of March 7, pp. 8487–8493, 2003

Printed in U.S.A.

High Conservation of the Set1/Rad6 Axis of Histone 3 Lysine 4 Methylation in Budding and Fission Yeasts*

Received for publication, September 18, 2002, and in revised form, December 17, 2002
Published, JBC Papers in Press, December 17, 2002, DOI 10.1074/jbc.M209562200

Assen Roguev‡§, Daniel Schaft‡§, Anna Shevchenko¶, Rein Aasland‖, Andrej Shevchenko¶, and A. Francis Stewart‡ ‡‡**

From ‡BIOTEC, Technische Universitaet Dresden, c/o Max Planck Institute for Molecular Cell Biology and Genetics and the ¶Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany and the ‖Department of Molecular Biology, University of Bergen, Thromoehlenstrasse 55, N-5020 Bergen, Norway

Histone 3 lysine 4 (H3 Lys[4]) methylation in *Saccharomyces cerevisiae* is mediated by the Set1 complex (Set1C) and is dependent upon ubiquitinylation of H2B by Rad6. Mutually exclusive methylation of H3 at Lys[4] or Lys[9] is central to chromatin regulation; however, *S. cerevisiae* lacks Lys[9] methylation. Furthermore, a different H3 Lys[4] methylase, Set 7/9, has been identified in mammals, thereby questioning the relevance of the *S. cerevisiae* findings for eukaryotes in general. We report that the majority of Lys[4] methylation in *Schizosaccharomyces pombe*, like in *S. cerevisiae*, is mediated by Set1C and is Rad6-dependent. *S. pombe* Set1C mediates H3 Lys[4] methylation *in vitro* and contains the same eight subunits found in *S. cerevisiae*, including the homologue of the *Drosophila* trithorax Group protein, Ash2. Three additional features of *S. pombe* Set1C each involve PHD fingers. Notably, the Spp1 subunit is dispensable for H3 Lys[4] methylation in budding yeast but required in fission yeast, and Sp_Set1C has a novel proteomic hyperlink to a new complex that includes the homologue of another trithorax Group protein, Lid (little imaginal discs). Thus, we infer that Set1C is highly conserved in eukaryotes but observe that its links to the proteome are not.

The stable maintenance of gene expression patterns through mitotic cell divisions, termed epigenetic regulation, is essential during development of higher organisms. Searches for epigenetic mechanisms in *Drosophila* development uncovered an opposition between the trithorax (trxG)[1] and Polycomb (PcG) groups. TrxG proteins appear to maintain patterns of gene activation, whereas PcG proteins maintain patterns of gene repression. TrxG encompasses several subclasses of gene regulatory factors (1). Although there are good models for action by some trxG members (*e.g. brahma*, *moira*, *zeste*, and GAGA), the activity of one subclass of the trxG, trxG3, which includes Trx, Ash1, and Ash2 (2, 3), remains unclear. Several trxG and PcG proteins contain a SET domain (Su(var) (3–9), E(z), and Trithorax (4)), a well conserved 150-amino acid domain found in all eukaryotes, which mediates histone methyltransferase activity (5, 6). Trx and Ash1 both contain a SET domain. There is little evidence so far that Trx methylates histones; however, binding to histones has been documented (7). The third trxG3 member, Ash2, does not contain a SET domain but has a PHD finger (8) and a SPRY domain (9).

Although budding yeast does not have a Trx homologue, it has a protein, Set1, with a very similar type of SET domain (10). For this reason, we determined the composition of the Set1 complex, Set1C, and found that it has H3 Lys[4] methyltransferase activity *in vitro* (3). Concomitantly, two other groups have identified most members of Set1C (11, 12). The complete Set1C and the *in vitro* specificity for H3 Lys[4] methylation have since been confirmed (13). The requirement for Set1 in H3 Lys[4] methylation *in vivo* was identified (14) and extended to Set1C members (12, 13). Set1C includes Bre2, the protein in *Saccharomyces cerevisiae* that is most similar to the *Drosophila* trxG protein, Ash2. Because we were exploring trxG action, we speculated that the protein-protein linkage between Set1 and Ash2/Bre2 in Set1C might consolidate trxG3 action to SET domains and histone methylation. At that time, however, no Set1 homologue was apparent in the *Drosophila* genome. We mined the *Drosophila* genome with deep bioinformatic tools to identify the Set1 orthologue, buried in sequence misreads. This discovery encouraged the trxG3-histone methylation proposition and the possibility that the Set1-Ash2 association may be broadly conserved in eukaryotes (3). The work reported here was begun with the motivation to challenge our proposition by characterizing the protein complex associated with *Schizosaccharomyces pombe* Set1.

The fission yeast *S. pombe* is widely held to be a more representative model for higher eukaryotes than the budding yeast *S. cerevisiae*. Indeed many *S. pombe* proteins appear to be more similar to their mammalian homologues than to *S. cerevisiae* counterparts (15). A number of cellular aspects such as the nuclear cycle, structure of centromeres, and aspects of histone methylation are similar between *S. pombe* and higher eukaryotes but divergent in *S. cerevisiae*. Although both yeasts have H3 Lys[4] methylation, *S. cerevisiae* lacks the nearby H3 Lys[9] methylation (14, 16). Emerging evidence in *S. pombe* and higher eukaryotes indicates that methylations of H3 Lys[4] and H3 Lys[9] are mutually exclusive in chromatin domains (16, 17). Because this mechanism cannot exist in *S. cerevisiae*, extrapolations from the budding yeast to higher eukaryotes regarding H3 Lys[4] methylation, Set1C, and Ash2 remain unsafe. Hence, we examined H3 Lys[4] methylation in *S. pombe* and compared it with *S. cerevisiae*.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ These authors contributed equally to this work.

** Supported by Grant 146652/431 from the Norwegian Research Council.

‡‡ To whom correspondence should be addressed. E-mail: stewart@mpi-cbg.de.

[1] The abbreviations used are: TrxG, trithorax Group; PcG, Polycomb Group; SET, Su(var), E(z), and Trithorax; TAP, tandem affinity purification; H, histone; RT, reverse transcriptase; MALDI, matrix-assisted laser desorption ionization; SAM, *S*-adenosyl methionine.

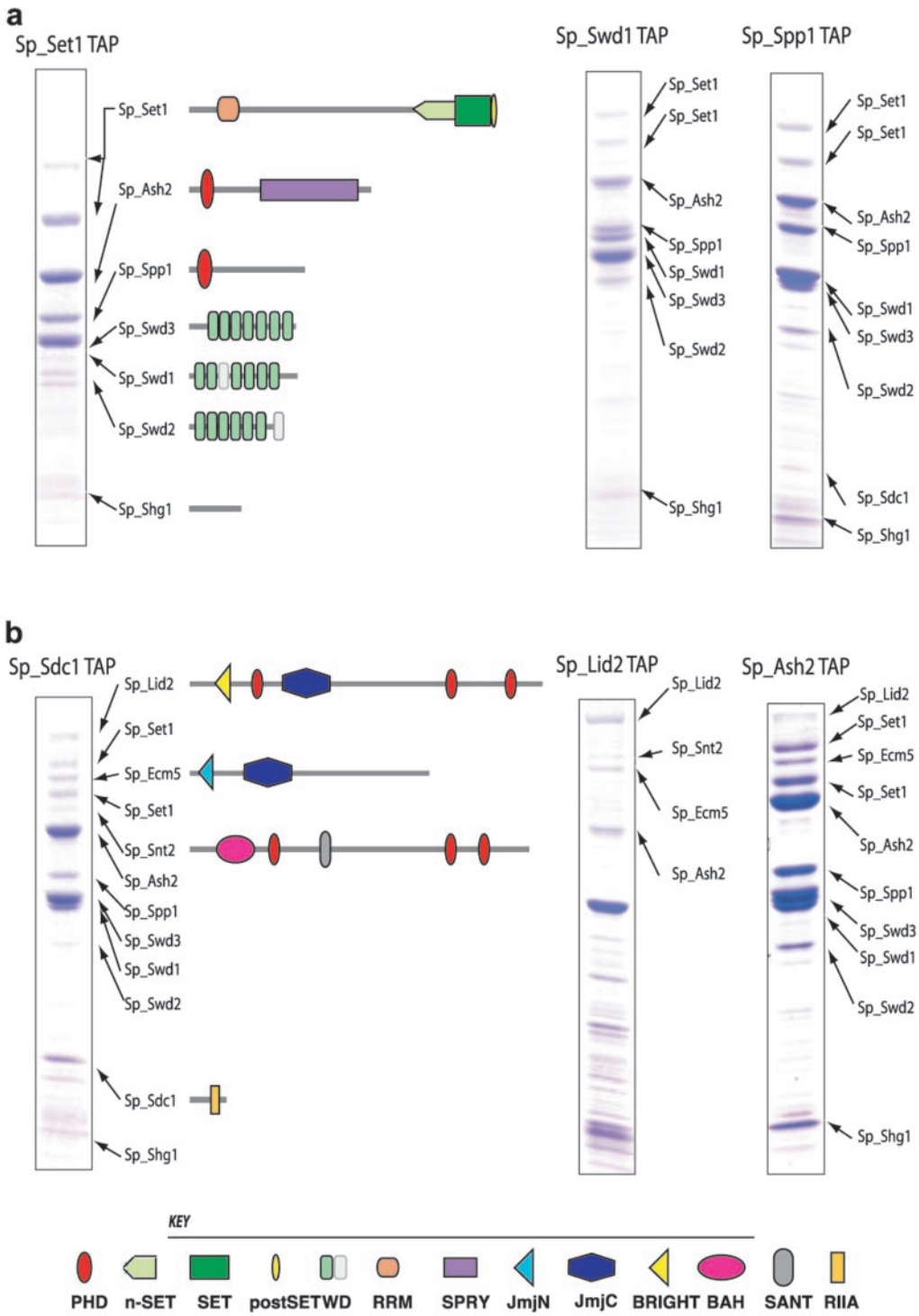

FIG. 1. **The compositions of Sp_Set1C and Sp_Lid2C.** *a*, affinity-purified Sp_Set1C using Sp_Set1-TAP (*left*), Sp_Swd1-TAP (*middle*), or Sp_Spp1-TAP (*right*) resolved on 7–25% SDS-PAGE and visualized with Coomassie Blue staining. All bands present in the gels were identified by MALDI mass spectrometry, and only those specific to Sp_Set1C are indicated. Each protein is depicted with identifiable domains and motifs, as indicated in the key at the *bottom* of the figure. *b*, same as in *a*, except proteins co-purifying with Sp_Sdc1-TAP (*left*) and Sp_Ash2 (*right*) included the entire Sp_Set1C and three new proteins (Lid2, Spbp19A11.06 (*S. pombe* genome data base reference); Ecm5, Spac3h1.12c; and Snt2, Spbc83.07) that form Sp_Lid2C. The composition of Lid2C was examined using Lid2-TAP (*middle*), which did not co-purify Set1C.

EXPERIMENTAL PROCEDURES

Strains, TAP Purification, and Mass Spectrometry—Strains used in this study are isogenic to DB241 (*h⁻; ura 4-D18; ade 6-M210; leu 1–32*). Gene disruptions and protein tagging was done as described (18). TAP purification and identification of proteins by mass spectrometry were performed essentially as described for budding yeast (19, 20). All the detectable bands in all gel lanes were analyzed, and only those specific to the complex are designated in Fig. 1. Persistent background proteins were different from those of budding yeast (20) and will be reported separately.

Protein Assays and Antibodies—Assays were performed as described previously (3). The antibodies used were peroxidase anti-peroxidase (PAP; Sigma), rabbit polyclonal anti-dimethylated Lys[4]-H3 (Abcam), and rabbit polyclonal anti-acetyl histone H4 (Upstate Biotechnology Inc.). Superose 6 size exclusion column (Amersham Biosciences) was loaded with 500 μl of cleared crude cell extract from a TAP-tagged strain and run in Buffer E (20 mM Na-HEPES, pH 8.0, 350 mM NaCl, 10% glycerol, 0.1% Tween 20). Fractions were resolved on 10% SDS-PAGE and analyzed by immunoblotting against the TAP tag. Size

TABLE I
Relative molar subunit composition of Set1C

Estimations were made by densitometry of Coomassie-stained gel images taking Set1 as 1.0. Because retrieval of the tagged protein itself can be relatively overestimated if the complex dissociates during purification, these values are presented in italics. Relative overestimation by approximately 2-fold was observed when Set1 or Swd1 was tagged. In other cases no overestimation was observed, thus indicating that the complex did not usually disassemble. —, denotes an inability to determine a relative estimation due to the co-migration of another protein.

Set1C	Set1-TAP	Ash2-TAP	Spp1-TAP	Swd1-TAP	Swd2-TAP	Sdc1-TAP
Set1	*1.0*	1.0	1.0	1.0	1.0	1.0
Ash2	1.4	2.8	2.8	2.8	2.8	7.0
Spp1	1.1	2.1	*2.1*	2.1	1.3	2.1
Swd3	—	1.4	1.4	1.2	?	1.4
Swd1	—	1.5	1.5	2.5	1.5	1.5
Swd2	0.4	0.7	1.3	1.3	*1.1*	0.7
Sdc1	0	0	—	0	0	*18.9*
Shg1	—	—	—	—	—	—

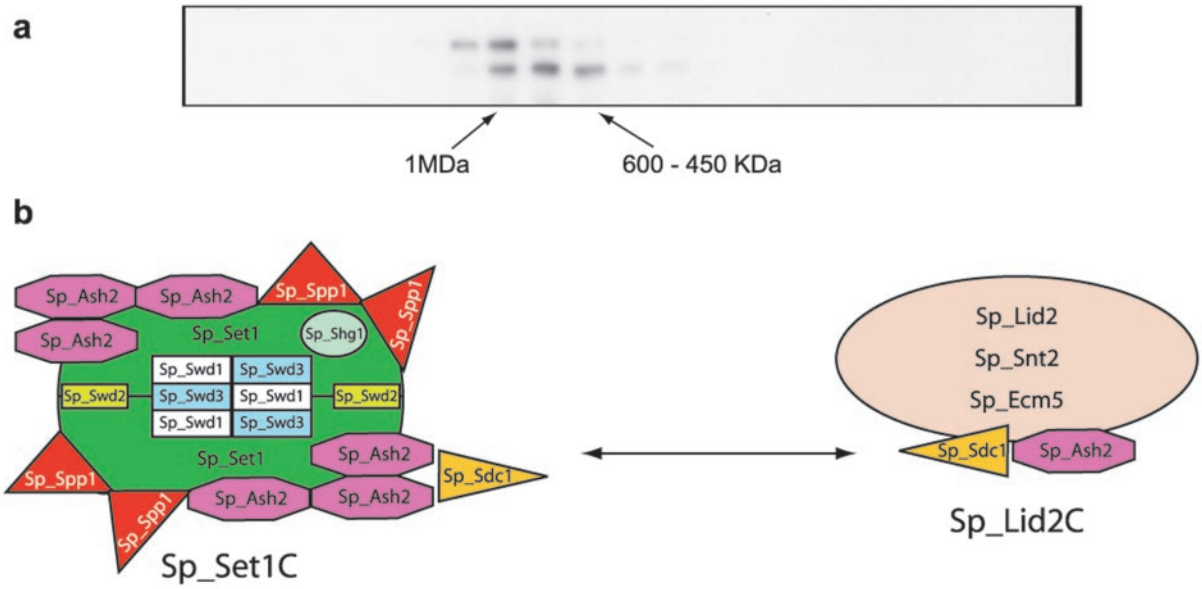

FIG. 2. **Size estimation and proteomic environment of Sp_Set1C.** *a*, size estimation of Sp_Set1C using a crude cell extract from the Sp_Set1-TAP strain and Superose 6 sizing column. Fractions were analyzed by Western blotting against protein A of the TAP tag. *b*, the proteomic environment of Set1C is depicted, which includes a working model of Set1C based on densitometric quantification of Coomassie staining intensities as presented in Table I. The proteomic hyperlink between Set1C and Lid2C is shown by a *double-headed arrow*. The protein-protein interactions included in the diagram are based on information from Sc_Set1C (3).

standards were run in parallel under the same conditions.

RNA and DNA Assays—Total RNA from exponentially growing yeast cultures was isolated (Qiagen) and subjected to RT-PCR. As a control for DNA contamination, a reaction without reverse transcriptase (RT) was run in parallel. Before reaching reaction plateau, (30 cycles) aliquots were analyzed by agarose gel electrophoresis. Genomic DNA was isolated from exponentially growing yeast cultures, digested with *Eco*RI, and analyzed by standard methods using a 1.3% agarose/TAE (Tris acetate-EDTA) gel. pAMP1 (21) was digested with *Apa*I and *Eco*RI (New England Biolabs), gel-purified (Qiagen), and labeled by random priming (Amersham Biosciences).

RESULTS

Composition of S. pombe Set1C—We purified the complex associated with *S. pombe* Set1 (Sp_Set1C) using tandem affinity purification (TAP) and mass spectrometry (19, 20, 22). Sp_Set1C is composed of the following eight proteins (Fig. 1) named in descending molecular weight order as follows (with *S. pombe* genome database reference in parentheses): Sp_Set1 (spcc306.04c), Sp_Ash2 (spbc13g1.08c), Sp_Spp1 (spcc594.05c), Sp_Swd3 (spbc354.03), Sp_Swd2 (spbc18h10.06c), Sp_Swd1 (spac23h3.05), Sp_Shg1 (spac17g8.09), and Sp_Sdc1 (spcc-18.11c). The overall composition differs only slightly from *S. cerevisiae* Set1C (3, 13). The composition of *S. pombe* Set1C (Sp_Set1C) was confirmed by tagging each member except Sp_Shg1 (Fig. 1 and data not shown).

Purification of Sp_Set1C from the Sp_Set1-TAP strain indicated that all cellular Set1 is incorporated in the complex because no uncomplexed, free, cellular protein was apparent (Fig. 1*a*). As for Sc_Set1, Sp_Set1 contains an RNA recognition motif and a SET domain flanked by an n-SET domain (3) and postSET peptide. In all Set1C preparations, Sp_Set1 appeared in two forms, the shorter of which lacked the N terminus (as determined by MALDI peptide mass maps; data not shown). Whether this is a result of specific intracellular processing or nonspecific degradation during the isolation of the complex remains to be determined.

Sp_Ash2 is orthologous to *Drosophila melanogaster* Ash2 protein and is more closely related to it than to its counterpart in Sc-Set1C, Sc_Bre2, which has no PHD finger. Homologies flanking the SPRY domains of Sp_Ash2, Bre2, Dm_Ash2, Mm_Ash2l, Hs_Ash2L, and Hs_Ash2l2 extend beyond the defined limits of the SPRY domain (3, 9). The Sp_Ash2 PHD finger is closely related to the PHD finger of Dm_Ash2 and to the unconventional PHD fingers of Mm_Ash2l, Hs_Ash2L, and Hs_Ash2l2 but not to the PHD finger of Spp1 (data not shown).

Purification of Sp_Set1C from the Sp_Spp1-TAP strain indicated that all cellular Spp1 is incorporated in the complex, because no uncomplexed free cellular protein was apparent. Both yeast Spp1s contain nearly identical PHD fingers, which

are closely related to the PHD finger of CGBP, a protein that binds preferentially to unmethylated CpGs (23). Thus, in contrast to Sc_Set1C, the complex from *S. pombe* contains two PHD fingers (in Sp_Ash2 and Sp_Spp1).

Sp_Swd1, Sp_Swd2, and Sp_Swd3 belong to the WD40 β propeller protein superfamily (24). Based on the Hidden Markov model and profile-based self-dot-plot analyses, each of these three proteins contains seven statistically significant WD40 repeats (Fig. 1*a* and data not shown). They also share significant sequence similarity with their homologues from *S. cerevisiae* and several species including flies, *Arabidopsis*, and man, indicating that they are individually very conserved (data not shown).

Sp_Shg1, like its homologue in the Set1C of *S. cerevisiae*, appeared in Sp_Set1C as a minor component. Only two other proteins with similarity to Sp_ and Sc_Shg1s are evident in the data bases, one in *Candida albicans* and the other in *Caenorhabditis elegans*, with highest similarity at the N termini (data not shown).

In Sc_Set1C, Sdc1 appeared as a minor component. We did not identify its *S. pombe* homologue in the Sp_Set1 TAP preparations, probably because of its small size and poor mass spectrometry signature. However, when the Sp_Sdc1 homologue was tagged, Sp_Set1C was retrieved, thus proving its presence in the complex. Notably, a significant excess of Sp_Sdc1 was retrieved over Set1C members and other proteins, indicating that Sp_Sdc1 may also exist as free protein in the cell (Fig. 1*b*; Table I). Sp_Sdc1 and Sc_Sdc1 show similarity to the *C. elegans* dosage compensation protein, Dpy-30. The similarity includes a short motif related to the dimerization motif in the regulatory subunit of protein kinase A (3).

Identification of a New Protein Complex Containing Lid—Interestingly, when either Sp_Ash2 or Sp_Sdc1 was TAP-tagged and purified, Sp_Set1C was retrieved along with three new high molecular weight proteins, now called Sp_Lid2 (Spbp19A11.06), Sp_Ecm5 (Spbc83.07), and Sp_Snt2 (Spac3h1.12c). Sp_Lid2 is the *S. pombe* homologue of the *Drosophila* trxG protein, Lid (little imaginal discs (25)). The Lid family of proteins, which includes Sc_Lid2 and mammalian Xe169 and Rbp2, contains three PHD fingers, a BRIGHT domain, and a JmjC domain. The BRIGHT domain is a helix-turn-helix DNA binding domain with preference for AT-rich regions (26). The JmjC domain is found in a wide variety of organisms from bacteria to humans in at least seven families of proteins. The domain has no known function but may be involved in regulation of chromatin remodeling (27, 28). Sp_Snt2 contains three PHD fingers, a SANT domain (29), and a BAH domain (30). In budding yeast, the homologues of Ecm5 and Snt2 interact physically; however, they do not appear to interact with Sc_Sdc1 or Sc_Lid2 (data not shown). Sp_Ecm5 contains a JmjC domain together with the N-terminal domain JmjN (27), often associated with JmjC.

To confirm the composition of the Sp_Lid2 complex (Sp_Lid2C), Sp_Lid2 was TAP-tagged and purified. Although Sp_Sdc1 was not identified, again presumably because of its small size and consequent difficulties with a clear mass spectrometry signature, Sp_Ash2, Sp_Ecm5, and Sp_Snt2 were retrieved (Fig. 1*b*). By dissection of protein-protein interactions within Sc_Set1C, we showed previously that Bre2 and Sdc1 directly interact with each other (3). Here we show that Sp_Ash2 and Sp_Sdc1 both associate with Sp_Set1C and Sp_Lid2C. Thus, we conclude that Sp_Ash2 and Sp_Sdc1 serve as proteomic hyperlinks between two complexes and probably form a module through interaction with each other.

Hence Sp_Set1C and Sc_Set1C show almost the same polypeptide and domain compositions, differing only by the absence of a PHD finger in Bre2 and unexpected proteomic hyperlinks

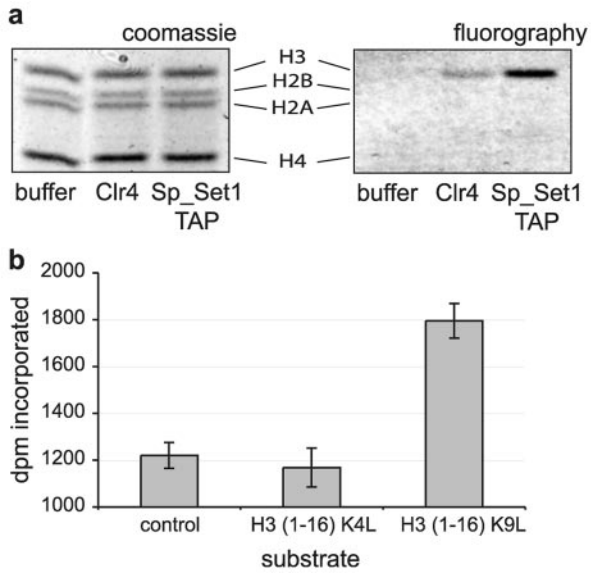

FIG. 3. **Sp_Set1C methylates histone 3 lysine 4 *in vitro*.** *a*, Sp_Set1C methyltransferase activity was assayed with Sp_Set1C-TAP-purified extracts using core histones as substrates in the presence of [^{3}H]SAM. Following incubation, the reactions were resolved by SDS-PAGE followed by Coomassie Blue staining and fluorography. *Clr4*, the *S. pombe* H3 Lys^9 methyltransferase was used as a positive control. *b*, Sp_Set1C has a preference for H3 Lys^4. The complex isolated from the Sp_Set1-TAP strain was incubated with H3 N-terminal (1–16) peptides mutated at lysine 4 (K4L) or lysine 9 (K9L) in the presence of [^{3}H]SAM, and incorporated ^{3}H was determined by scintillation counter.

through Sp_Ash2 and Sp_Sdc1 to Sp_Lid2C. Except for Shg1, homologues and orthologues are present in diverse eukaryotic databases, indicating that Set1C is highly conserved.

The Apparent Sizes of the Two Yeast Set1Cs and Their in Vitro Methylation Specificities Are Identical—We estimated the size of the Sp_Set1C by means of Superose 6 size-exclusion column chromatography (Fig. 2*a*). Two bands were observed, corresponding to the full-length and shorter versions of Set1 with the longer Set1 migrating in a complex(es) at ~1 MDa and the shorter migrating around 800 kDa (assuming an overall globular configuration). We do not know whether the smaller complex(es) represents an authentic second complex(es) or a degradation product. The estimated size of 1 MDa is the same as Sc_Set1C (11)[2] but differs from the smaller apparent estimation of ~500 kDa made after cell extraction in higher salt (12).

The relative stoichiometry of Sp_Set1C subunits was estimated by densitometric measurement of images from the Coomassie-stained gels and then adjusted for molecular weight (Table I). Given a 1-MDa complex, we propose the complex composition presented in Fig. 2*b*. This model is concordant with our data; however, it should be regarded as an initial proposition. Mass estimations from Coomassie staining intensities can be only approximations for several reasons, including understaining because of the highly acidic content. This may not be a problem for Sp_Set1C because all subunits have isoelectric points between 5 and 6.5 except for Set1, which is 9.0. We also point out that the question as to whether Set1C is present in cells as a single complex or in several slightly different forms cannot be answered yet.

The histone methyltransferase activity of Sp_Set1C was tested using the complex isolated from the Sp_Set1-TAP strain and incubation with core histones in the presence of [^{3}H]SAM. It showed methyltransferase activity directed toward histone H3 (Fig. 3*a*), which was specific for Lys^4 (Fig. 3*b*). In this case,

[2] A. Roguev, D. Schaft, A. Shevchenko, R. Aasland, A. Shevchenko, and A. Francis Stewart, unpublished observations.

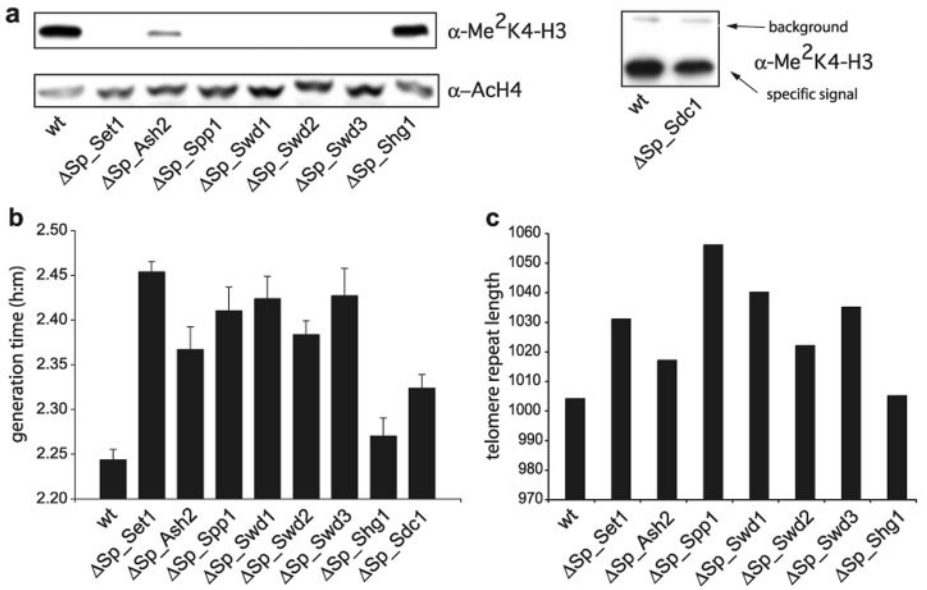

FIG. 4. **H3 Lys[4] methylation *in S. pombe* requires Sp_Set1C members and correlates with phenotypic markers.** *a*, Western blot using an antibody specific to methylated H3 Lys[4] showing the effect of disruption of all the members of Sp_Set1C on the levels of H3 Lys[4] methylation *in vivo*. An antibody specific to acetylated H4 was used as a loading control. *b*, doubling times of strains lacking Sp_Set1C members. Results are from three experiments with *error bars* showing standard deviations. *c*, the length of the telomere repeats in strains lacking Sp_Set1C members determined from Southern blots and scans from phosphorimaging data.

unlike Sc_Set1C (3), fusion of the TAP tag to the C terminus next to the SET domain did not inhibit the enzymatic activity *in vitro*.

In Vivo Methylation Requirements for Set1C Subunits—The contribution of individual complex members to the enzymatic activity *in vivo* was examined. Disruption strains were readily obtained for all complex members except for Sp_Sdc1, which was elusive until recently. Hence, no complex member is essential for *S. pombe*. Individual disruption of all members totally ablated H3 Lys[4] methylation except for Sp_Ash2, Sp_Sdc1, and Sp_Shg1, which showed about one-tenth, one-half, or wild type levels of H3 Lys[4] methylation, respectively (Fig. 4*a*). These results are identical to those obtained from the same experiments in *S. cerevisiae*[2] (Refs. 12 and 13) except for Spp1. Deletion of Spp1 in *S. cerevisiae* has no effect on H3 Lys[4] methylation, whereas in fission yeast it leads to complete loss of H3 Lys[4] methylation. All of the WD40 repeat proteins are required for H3 Lys[4] methylation, demonstrating that they are not redundant to each other. WD40 repeat proteins have been shown to serve as platforms for protein complex assembly; however, some of these proteins bind to histones (*e.g.* TBL1, Groucho, TUP1, and RCC1; *e.g.* Ref. 31). As seen in Fig. 4*a*, no dimethylated H3 Lys[4] was present upon disruption of the complex, indicating that Sp_Set1C is the major, possibly only, H3 Lys[4] methyltransferase activity in *S. pombe* and extending the recently published result of Noma and Grewal (32) who show that deletion of set1 from *S. pombe* abolishes H3 Lys[4] dimethylation.

Disruption of Set1 and other Sc_Set1C components causes complex phenotypes in *S. cerevisiae* including defects in cell wall formation, growth, and DNA repair, as well as alterations in silencing and telomere length (10, 14, 33–36). We examined two of these phenotypic markers, growth and telomere length, for Sp_Set1C members (Fig. 4, *b* and *c*). Concordance between loss of H3 Lys[4] methylation and phenotypic impact was observed in both assays. Loss of Shg1 and Sdc1 had the least impact on H3 Lys[4] methylation and phenotype. The next mildest impact in all three assays was loss of Ash2. The only notable deviation from qualitative equivalence in all three assays is the effect of loss of Spp1 on telomere length. This suggests selectivity for Spp1 in the regulation of telomere length. Notably, in *S. cerevisiae*, loss of Set1C members results in shorter telomeres (3, 10, 34), whereas we observed lengthening in *S. pombe*. Either the telomere length changes due to Set1C disruption are not directly due to disturbance of chromatin at the telomeres, or the difference between *S. cerevisiae* and *S. pombe* may reflect the possibility that loss of H3 Lys[4] methylation in *S. pombe* alters the boundaries of H3 Lys[9] methylation domains.

H3 Lys[4] Methylation in S. pombe Requires the Rad6 Homologue—In *S. cerevisiae*, H3 Lys[4] methylation is dependent upon both Set1C and ubiquitinylation of H2B by Rad6 (37, 38). When the *S. pombe* Rad6 orthologue Rhp6 (Spac18b11.07c) was disrupted, there was a complete loss of H3 Lys[4] methylation (Fig. 5*a*). The possibility that this effect was indirect via abolished expression of a Set1C member was excluded by semiquantitative RT-PCR (Fig. 5*b*). Hence, the regulation and enzymology of H3 Lys[4] methylation is highly conserved between budding and fission yeasts.

DISCUSSION

Set1C in *S. cerevisiae* includes Bre2, the homologue of the *Drosophila* trxG member, Ash2. To enquire whether the association between Set1 and Ash2 may also be found in other eukaryotes, we characterized Set1C from *S. pombe*. We found that both yeast Set1Cs are highly conserved, hence suggesting that Set1C will be highly conserved throughout eukaryotes, including the retention of the Set1-Ash2 linkage in flies. Our observations also strengthen the proposition that H3 Lys[4] methylation is a common mechanistic feature of the trxG3 subgroup. Very recently, further evidence for this proposition has emerged with the biochemical identification of H3 Lys[4] methyltransferase activity in both of the other trxG3 members, Ash1 (42) and Mll (43, 44). We also found that Ash2 associates with Lid2 in *S. pombe*. Because Lid is a trxG protein in *Drosophila* (25), the identification of Ash2 as a trxG protein may relate to its interaction with Lid rather than with Set1 or possibly to its involvement in both complexes.

In both yeasts, mutations of both Ash2 (Bre2) and Sdc1 decreased but did not abolish H3 Lys[4] methylation. Whether these reductions represent a global reduction of methylation or loss of methylation at specific sites and not others remains to be determined. In any event, the protein-protein linkage of Ash2 and Sdc1 suggests that these reductions are based on a similar mechanistic loss. If Set1C selectively targets specific nucleosomes, then it is likely that the nonessential Ash2, Sdc1 or Shg1 subunits play roles in selectivity. Conversely, if Set1C methylates nucleosomes by a general mechanism, for example

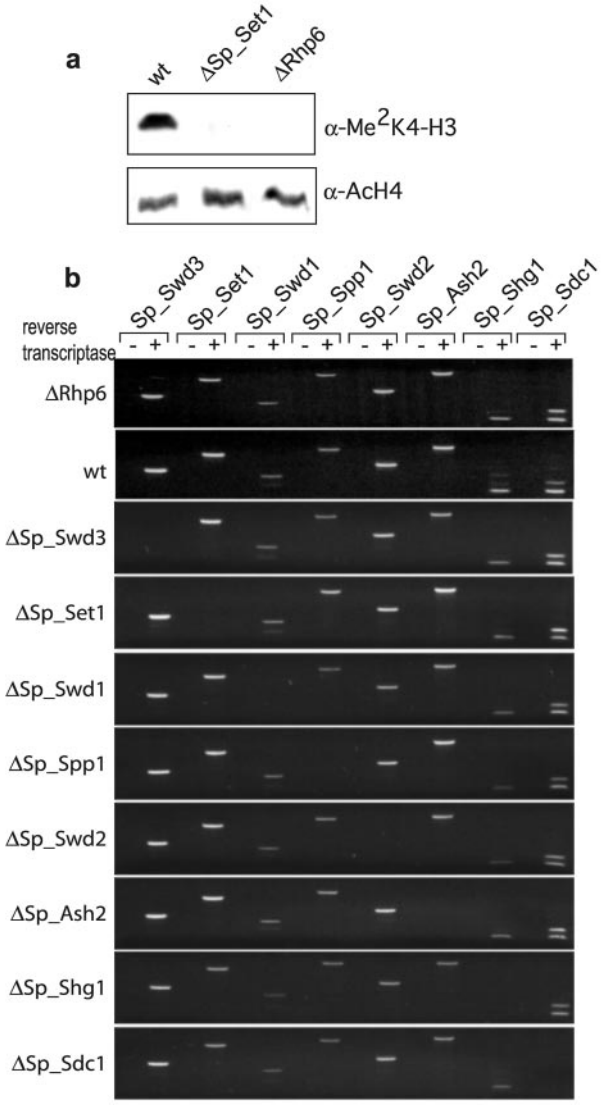

FIG. 5. **H3 Lys[4] methylation in *S. pombe* requires the Rad6 homologue, Rhp6.** *a*, Western blot using an antibody specific to methylated H3 Lys[4] showing the effect disruption of Rhp6 or Set1 on H3 Lys[4] methylation *in vivo*. *b*, RT-PCR analysis for transcripts of different members of Sp_Set1C in ΔRhp6 (*top*), wild type, and ΔSet1C member strains, as indicated, showing that all members of Set1C are expressed unless specifically deleted. Each RT-PCR reaction was run in duplicate (with and without reverse transcriptase) as a control for DNA contamination.

to methylate every nucleosome in a chromatin domain, then it is likely that the three WD40 proteins, Swd1, Swd2, and Swd3, play roles in substrate binding.

Our results reveal the striking conservation surrounding H3 Lys[4] methylation. However, prior evidence from the two yeasts highlight differences rather than similarities. Budding yeast lacks H3 Lys[9] methylation, but it appears that H3 Lys[4] and Lys[9] methylations are mutually exclusive in other eukaryotes (16, 17). Furthermore, the first evidence from budding yeast indicates that Set1 and H3 Lys[4] methylation is associated with chromatin repression at telomeres and rDNA (13, 14, 33, 36), whereas it correlates with active chromatin in fission yeast (16, 32). Resolution of this discrepancy awaits further work; however, recent evidence from budding yeast also associates H3 Lys[4] methylation with active chromatin, in particular with actively transcribed coding regions (39, 40).

Although strikingly conserved, the two yeast Set1Cs differ in three ways, each of which relates to PHD fingers. First, Ash2 in Sp_Set1C includes a PHD finger, whereas Bre2 in Sc_Set1C does not. Second, the PHD finger protein Spp1 is required for H3 Lys[4] methylation in *S. pombe* but not in *S. cerevisiae*. Third, Sp_Ash2 and Sp_Sdc1 form a proteomic hyperlink between Sp_Set1C and Lid2C, which includes several PHD fingers. Although PHD fingers have been found exclusively in chromatin proteins so far, their mode of action remains elusive. Possibly the missing PHD finger relationships of Sc_Set1C relate to the absence of H3 Lys[9] methylation in budding yeast. This correlation suggests that the PHD fingers of Ash2 and Spp1 may help in the recognition, perhaps directly, of the post-translational states of the nucleosomes upon which Set1C acts. If so, then some PHD fingers would interpret the histone code like certain chromo and bromo domains (8, 40). Furthermore the intriguing proteomic association of Set1C to another trxG protein, Lid, is either a highly unusual coincidence involving two trxG proteins or a hint that Lid proteins and complexes may also involve reading of the histone code. The close linkage of Sdc1 to Set1C and Lid2C also has implications for dosage compensation in *C. elegans* through the Sdc1 homologue Dpy30 (45).

As for Set1C, ubiquitinylation of H2B (46) and Rad6 homologues are apparent in all eukaryotes examined. Because the protein machinery for H3 Lys[4] methylation and its relationship to Rad6 is highly conserved between budding and fission yeasts, it is likely that the Set1C/Rad6 axis in H3 Lys[4] methylation is highly conserved in eukaryotes. If so, it will be interesting to understand why mammals appear to have two orthologues of Set1 (3). Because Set1C appears to be built on a dimeric platform, both mammalian Set1s may be incorporated into a single Set1C, or there could be two or more Set1Cs with possible differences in function. It will also be interesting to understand how the putative mammalian Set1Cs relate to the other known H3 Lys[4] activities mediated by Set7/9 (47, 48) and Ash1 (42).

Our work also presents new proteomic insights. The remarkable conservation of the nonessential Set1C in the two yeasts bodes well for future extrapolations from lower to higher eukaryotes regarding other protein complexes. However the proteomic environment surrounding the two Set1Cs appears to differ. The only identified proteomic hyperlink in Sc_Set1C was Swd2, which is also a subunit of cleavage polyadenylation factor (3). This proteomic hyperlink is not conserved in *S. pombe*.[2] Furthermore, we have identified a new hyperlink between Sp_Set1C and Sp_Lid2. The *S. cerevisiae* orthologues of Ecm5 and Snt2 also interact with each other but do not appear to interact with Sc_Bre2, Sc_Sdc1, or Sc_Lid2.[2]

This work, in conjunction with our previous paper (3), presents, to our knowledge, the first case in which fine mapping of one complex and the surrounding proteomic environment from two organisms has been compared. Although we find that the Set1Cs from *S. cerevisiae* and *S. pombe* are highly conserved, their proteomic environments appear to differ, thus pointing to an inherent limitation of the "orthologous proteome" concept (49) and the need for comparative proteomic analyses (50). Whether our observations define a new proteomic principle regarding strong conservation within complexes and weaker conservation of linkages between complexes remains to be determined, but this explanation certainly makes intuitive sense.

Acknowledgments—We are grateful to Damian Brunner for the yeast strains and the crash course on fission yeast, Thomas Jenuwein for reagents and discussion, David Drechsel for help with the sizing columns, Vincent Geli for communication of unpublished data, Junko Kanno for pAMP1, Henrik Thomas for assistance with mass spectrometric sample preparation, and to all members of the Stewart laboratory and Sabine Goldmann for useful discussions.

REFERENCES

1. Kennison, J. A. (1995) *Annu. Rev. Genet.* **20,** 289–303
2. Shearn, A. (1989) *Genetics* **121,** 517–525
3. Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) *EMBO J.* **20,** 7137–7148
4. Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994) *EMBO J.* **13,** 3822–3831
5. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) *Nature* **406,** 593–599
6. Zhang, Y., and Reinberg, D. (2001) *Genes Dev.* **15,** 2343–2360
7. Katsani, K. R., Arredondo, J. J., Kal, A. J., and Verrijzer, C. P. (2001) *Genes Dev.* **15,** 2197–2202
8. Aasland, R., Gibson, T. J., and Stewart, A. F. (1995) *Trends Biochem. Sci.* **20,** 56–59
9. Ponting, C., Schultz, J., and Bork, P. (1997) *Trends Biochem. Sci.* **22,** 193–194
10. Nislow, C., Ray, E., and Pillus, L. (1997) *Mol. Biol. Cell* **8,** 2421–2436
11. Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 12902–12907
12. Nagy, P. L., Griesenbeck, J., Kornberg, R. D., and Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 90–94
13. Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shilatifard, A. (2002) *J. Biol. Chem.* **277,** 10753–10755
14. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) *Genes Dev.* **15,** 3286–3295
15. Sipiczki, M. (2000) *Genome Biol*. **1,** REVIEWS1011
16. Noma, K., Allis, C. D., and Grewal, S. I. (2001) *Science* **293,** 1150–1155
17. Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D., and Felsenfeld, G. (2001) *Science* **293,** 2453–2455
18. Tasto, J. J., Carnahan, R. H., McDonald, W. H., and Gould, K. L. (2001) *Yeast* **18,** 657–662
19. Pijnappel, W. W., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R., and Stewart, A. F. (2001) *Genes Dev.* **15,** 2991–3004
20. Shevchenko, A., Schaft, D., Roguev, A., Pijnappel, W. W. M. P., Stewart, A. F., and Shevchenko, A. (2002) *Mol. Cell. Proteomics* **1,** 204–212
21. Matsuura, A., Naito, T., and Ishikawa, F. (1999) *Genetics* **152,** 1501–1512
22. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) *Nat. Biotechnol.* **17,** 1030–1032
23. Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A., and Skalnik, D. G. (2000) *Mol. Cell. Biol.* **20,** 2108–2121
24. Paoli, M. (2001) *Prog. Biophys. Mol. Biol.* **76,** 103–130
25. Gildea, J. J., Lopez, R., and Shearn, A. (2000) *Genetics* **156,** 645–663
26. Kortschak, R. D., Tucker, P. W., and Saint, R. (2000) *Trends Biochem. Sci.* **25,** 294–299
27. Balciunas, D., and Ronne, H. (2000) *Trends Biochem. Sci.* **25,** 274–276
28. Clissold, P. M., and Ponting, C. P. (2002) *Trends Biochem. Sci.* **26,** 7–9
29. Aasland, R., Stewart, A. F., and Gibson, T. (1996) *Trends Biochem. Sci*. **21,** 87–88
30. Callebaut, I., Courvalin, J. C., and Mornon, J. P. (1999) *FEBS Lett.* **446,** 189–193
31. Nemergut, M. E., Mizzen, C. A., Stukenberg, T., Allis, C. D., and Macara, I. G. (2001) *Science* **292,** 1540–1543
32. Noma, K. I., and Grewal, S. I. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** Suppl. 4, 16438–16445
33. Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L., and Jenuwein, T. (1997) *EMBO J.* **16,** 3219–3232
34. Corda, Y., Schramke, V., Longhese, M. P., Smokvina, T., Paciotti, V., Brevet, V., Gilson, E., and Geli, V. (1999) *Nat. Genet.* **21,** 204–208
35. Schramke, V., Neecke, H., Brevet, V., Corda, Y., Lucchini, G., Longhese, M. P., Gilson, E., and Geli, V. (2001) *Genes Dev.* **15,** 1845–1858
36. Bryk, M., Briggs, S. D., Strahl, B. D., Curcio, M. J., Allis, C. D., and Winston, F. (2002) *Curr. Biol.* **12,** 165–170
37. Sun, Z-W., and Allis, C. D. (2002) *Nature* **418,** 104–108
38. Dover, J., Schneider, J., Boateng, M. A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002) *J. Biol. Chem.* **277,** 28368–28371
39. Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8695–8700
40. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) *Nature* **419,** 407–411
41. Deleted in proof
42. Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002) *Nature* **419,** 857–862
43. Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., and Canaani, E. (2002) *Mol. Cell* **10,** 1119–1128
44. Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., and Hess, J. L. (2002) *Mol. Cell* **10,** 1107–1117
45. Hsu, D. R., Chuang, P. T., and Meyer, B. J. (1995) *Development* **121,** 3323–3334
46. Jason, L. J. M., Moore, S. C., Lewis, J. D., Lindsey, G., and Ausio, J. (2002) *Bioessays* **24,** 166–174
47. Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001) *Mol. Cell* **8,** 1207–1217
48. Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Reinberg, D. (2002) *Genes Dev.* **16,** 479–489
49. Rubin, G. M., *et al.* (2000) *Science* **287,** 2204–2215
50. Liska, A., and Shevchenko, A. (2003) *Proteomics*, in press