

Protein Interactions within the Set1 Complex and Their Roles in the Regulation of Histone 3 Lysine 4 Methylation*

Received for publication, March 31, 2006, and in revised form, August 18, 2006. Published, JBC Papers in Press, August 18, 2006, DOI 10.1074/jbc.M603099200

Pierre-Marie Dehé[‡], Bernhard Dichtl[§], Daniel Schaft[¶], Assen Roguev[¶], Mercè Pamblanco^{||}, Régine Lebrun[‡], Alfonso Rodríguez-Gil^{**}, Msau Mkandawire[¶], Katarina Landsberg[¶], Anna Shevchenko^{‡‡}, Andrej Shevchenko^{‡‡}, Lorena E. Rosaleny^{||}, Vicente Tordera^{||}, Sebastián Chávez^{**2}, A. Francis Stewart^{¶3}, and Vincent Géli^{‡4}

From the [‡]Laboratoire d'Instabilité du Génome et Cancérogénèse, Institut de Biologie Structurale et Microbiologie, CNRS, Marseille 13402, France, [§]Institute of Molecular Biology, University of Zürich, Zürich CH-8057, Switzerland, [¶]Genomics, Biotec Research Centre, Technische Universität Dresden, Dresden 01307, Germany, ^{‡‡}Max-Planck Institute of Molecular Cell Biology and Genetics, Dresden 01307, Germany, ^{||}Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot 46100, Spain, and ^{**}Departamento de Genética, Facultad de Biología, Universidad de Sevilla, Sevilla 41012, Spain

Set1 is the catalytic subunit and the central component of the evolutionarily conserved Set1 complex (Set1C) that methylates histone 3 lysine 4 (H3K4). Here we have determined protein/protein interactions within the complex and related the substructure to function. The loss of individual Set1C subunits differentially affects Set1 stability, complex integrity, global H3K4 methylation, and distribution of H3K4 methylation along active genes. The complex requires Set1, Swd1, and Swd3 for integrity, and Set1 amount is greatly reduced in the absence of the Swd1-Swd3 heterodimer. Bre2 and Sdc1 also form a heteromeric subunit, which requires the SET domain for interaction with the complex, and Sdc1 strongly interacts with itself. Inactivation of either Bre2 or Sdc1 has very similar effects. Neither is required for complex integrity, and their removal results in an increase of H3K4 mono- and dimethylation and a severe decrease of trimethylation at the 5' end of active coding regions but a decrease of H3K4 dimethylation at the 3' end of coding regions. Cells lacking Spp1 have a reduced amount of Set1 and retain a fraction of trimethylated H3K4, whereas cells lacking Shg1 show slightly elevated levels of both di- and trimethylation. Set1C associates with both serine 5- and serine 2-phosphorylated forms of polymerase II, indicating that the association persists to the 3' end of transcribed genes. Taken together, our results suggest that Set1C subunits stimulate Set1 catalytic activity all along active genes.

In *Saccharomyces cerevisiae*, methylation of histone 3 at lysine 4 (H3K4) is mediated by Set1 (1, 2). Deletion of *SET1* has

pleiotropic effects on gene expression, sporulation, DNA repair, chromosome segregation, telomere length, and cell wall integrity (1–11). Set1 is a large protein that contains a C-terminal SET domain carrying the catalytic site for H3K4 methylation and an N-terminal RNA recognition motif (2, 12). Set1 is part of a large complex termed either Set1C or COMPASS (complex associated with Set1) (2, 13, 14) that appears to be highly conserved in eukaryotes (15–17). In yeast, Set1C has eight subunits: Swd1 (Cps50 or Saf49); Swd2 (Cps35 or Saf37); Swd3 (Cps30 or Saf35); Bre2 (Cps60); Sdc1 (Cps25 or Saf19); Spp1 (Cps40 or Saf41), and Shg1. Only one subunit, Swd2, is essential in *S. cerevisiae*, because it is also a subunit of cleavage and polyadenylation factor (2, 18–20). Set1C methylates H3K4 either to mono-, di-, or trimethylated states (3). Swd1, Swd3, Bre2, and Sdc1 were initially shown to be required for bulk H3K4 dimethylation (2, 14, 21). More recently, it has been shown that Swd1 and Swd3 are required for all three methylation states, whereas Bre2, Sdc1, and Spp1 are required for H3K4 trimethylation. Furthermore, inactivation of Bre2 or Sdc1 results in a significant reduction of dimethylation (4, 22).

Set1 occupancy and trimethylation of H3K4 have been found to peak in the early coding regions of active genes transcribed by RNA polymerase II (pol II)⁵ (23). Set1 associates with the pol II C-terminal domain (CTD) through association with the Paf1 complex (23, 24). It has been proposed that Set1 is released from pol II after the loss of Ser⁵ phosphorylation and that Set1 binding to Ser⁵-phosphorylated pol II restricts H3K4 trimethylation to the 5' end of the genes (23, 24). Genome-wide studies and gene-specific chromatin immunoprecipitation experiments indicate that Set1 and dimethylation are present in the middle and 3' end of actively transcribed genes (25–27). Interestingly, under transcriptional stress, nucleosomes at 3' coding regions become trimethylated in a way depending on the chromodomain protein Chd1, suggesting a different mode of Set1 recruitment to chromatin (28).

Paf1 and Rtf1, components of the Paf1 complex, have also been shown to be required for the Rad6/Bre1-mediated monoubiquitination of H2B on lysine 123 (29, 30). Through a

* This work was supported in part by the Acciones Integradas Hispano-Francesas (HF2003-0170) and the "Picasso Program". 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.

¹ Supported by Grant BFU2005-02603 from the Ministerio de Educación y Ciencia, Spain.

² Supported by the Ministerio de Ciencia y Tecnología (Grant BMC2003-07072-C03-01) and Instituto de Salud Carlos III (Grant 02/1363).

³ To whom correspondence may be addressed: Biotec Research Centre, Technische Universität Dresden, BiInnovationsZentrum, Am Tatzberg 47, 01307 Dresden, Germany. Tel.: 49-351-46340129; Fax: 49-351-46340143; E-mail: stewart@biotec.tu-dresden.de.

⁴ Supported by "La Ligue contre le cancer" (équipe labellisée). To whom correspondence may be addressed: Instabilité du Génome et Cancérogénèse, Institut de Biologie Structurale et Microbiologie, CNRS, 31 chemin Joseph Aiguier, 13402, Marseille cedex 20, France. Tel.: 33-491-164-532; Fax: 33-491-712-124; E-mail: geli@ibsm.cnrs-mrs.fr.

⁵ The abbreviations used are: pol II, polymerase II; SET, Su(var)3-9, enhancer of zeste, and trithorax; CTD, C-terminal domain; X-gal, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside; ChIP, chromatin immunoprecipitation; TAP, tandem affinity purification.

so-called “*trans-tail*” process, di- and trimethylation of H3K4 depends on H2B ubiquitination (31, 32), and components of the Paf1 complex (Paf1, Rtf1, and Ctr9) are required for di- and trimethylation of H3K4 (24). However, although Rtf1 also appears important for monomethylation of H3K4, neither Paf1 nor ubiquitination of histone H2B are essential for H3K4 monomethylation (27, 33). Interestingly, a screen meant to identify components that selectively regulate H3K4 trimethylation identified the Bur cyclin-dependent kinase whose inactivation impairs the recruitment of the Paf elongation complex to genes and H2B ubiquitination (34, 35), suggesting that low levels of ubiquitination are sufficient to promote mono- and dimethylation of H3K4. In this study, we have investigated protein/protein interactions within Set1C and determined their role in the integrity of the complex and the regulation of H3K4 methylation.

EXPERIMENTAL PROCEDURES

The Two-hybrid System in Escherichia coli—Protein interactions have been studied by a bacterial two-hybrid approach based on the functional reconstitution of adenylate cyclase activity of *Bordetella pertussis* in the *E. coli cya* strain (36). Pairs of proteins to be tested were fused to the two catalytic domains T18 and T25 of adenylate cyclase using plasmids pKT25link and pUT18link and their derivatives (36, 37). After co-transformation of the BTH101 strain by the two plasmids expressing the fusions, 3 ml of LB medium supplemented with ampicillin and kanamycin and 0.5 mM isopropyl 1-thio- β -D-galactopyranoside were inoculated with ten randomly chosen transformants and grown at 30 °C to an A_{600nm} of 1. Thereafter, 2- μ l drops were spotted on McConkey plates supplemented with 1% maltose and on X-gal plates. The plates were incubated 3 days at 30 °C. An interaction is considered positive if detection of β -galactosidase activity is obtained in >3 independent transformations.

Epitope Tagging of Set1, Mutant Strains, and TAP Purification—Set1 was tagged at its N terminus with a 9-Myc epitope as follows. A NotI restriction site was introduced just after the ATG codon of *SET1* into a genomic fragment extending from -497 +1179 bp in respect to the ATG of *SET1*. This DNA fragment was cloned into the integrative plasmids Yiplac128 (*LEU2*) and Yiplac204 (*TRP1*). A 9-Myc epitope was introduced in-frame into NotI. The resulting plasmids (p9Myc*SET1-LEU2* or p9Myc*SET1-TRP1*) were digested with SnaBI prior to transformation of yeast cells. The tagged 9-Myc-Set1 was fully functional for all of the *set1* phenotypes that were tested. Full deletions of *set1*, *swd1*, *swd3*, *bre2*, *sdc1*, *spp1*, and *shg1* were in the BY4741 background (Euroscarf). All experiments were done in BY4741, except TAP tag introduction and TAP purification, which were performed in strain MGD-353-13D according to Roguev *et al.* (2).

Preparation of Histones, Microsequencing Analysis, and Western Blot Analysis—Histones were isolated by a procedure described in Ref. 38. Histone preparations were subjected to 16% SDS-PAGE and either Coomassie Blue-stained or transferred onto a 0.2- μ m Protran[®] nitrocellulose membrane. For Western blot analysis, the antibodies used to detect di- and trimethylated H3K4 were anti-H3K4me2 and -H3K4me3 (Abcam). N-terminal sequence determination was performed

by stepwise Edman degradation using an automatic sequencer model Procise 494 from Applied Biosystems. Histone H3 bands from blotted protein on 0.2- μ m polyvinylidene difluoride, stained with Coomassie blue, were excised and put on a Prob-lott cartridge (Applied Biosystems). For all samples, a pulse liquid polyvinylidene difluoride protein program was performed, and phenylthiohydantoin amino acids were identified by high pressure liquid chromatography using a Brownlee C18 column (5 μ m, 220 \times 2.1 mm) with an 18-min gradient of B (acetonitrile/isopropyl alcohol, v/v) in A (3.5% H_4 folate/water) from 16–46% B at 55 °C and with the wavelength set at 269 nm. A solution of 20 phenylthiohydantoin amino acids standard (10 pmol) was used for calibrating every run.

Simultaneous Set1 and Rap1 detection in yeast whole-cell extracts was carried as follows. Whole-cell extracts were prepared by trichloroacetic acid precipitation. Identical amounts of whole-cell extracts were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and blocked in 5% milk/phosphate-buffered saline. Anti-Set1 (provided by Dr. Peter Nagy, Molecular Oncology Testing, Medical Research Center, Iowa City, IA) and anti-Rap1 (provided by Dr. David Shore, University of Genève, Geneva, Switzerland) antibodies (1:1000 dilution) were incubated for 3 h in 5% milk/phosphate-buffered saline. Infrared-labeled secondary antibodies IRDye 800 anti-rabbit goat IgG (Rockland) and Alexa Fluor[®] 680 anti-mouse goat IgG (Molecular Probes) were used at a dilution of 1:10000. Emitted fluorescence of the secondary antibodies was detected by scanning on an Odyssey Imager (LI-COR Bioscience). Antibodies BWG16, H14, and H3 used to detect the different forms of pol II CTD in Western blot were purchased from Covance.

Chromatin Immunoprecipitation (ChIP) Experiments—ChIPs were done using 50-ml cultures fixed with 1% formaldehyde for 15 min (27). Yeast cells were broken using glass beads, and fixed chromatin was broken by sonication using a Bioruptor (Diagenode). Average DNA fragment lengths were 150 bp. 360- μ l aliquots of chromatin were incubated with the indicated antibodies in 1.5 ml of siliconized eppendorfs at 4 °C for 15–20 h and immunoprecipitated with protein G-Sepharose beads (Dynabeads) for 90 min at 4 °C. After washing, the chromatin was eluted from the beads at 65 °C for 30 min. Cross-links were reversed by incubation at 65 °C for 20 h. DNA was purified using Invitrogen PCR minicolumns (Pure-Link) and eluted in 200 μ l of water. Samples were subject to real-time PCR using a light cycler (Roche Applied Science) and Sybr Green mix. ChIPs were expressed as a ratio immunoprecipitation:input (IP:input). ChIPs against the methylated states of H3K4 are normalized to the $IP_{H3} : input_{H3}$ value. Immunoprecipitation of DNA cross-linked to the Myc-Set1 was performed with 9E10 (anti-Myc)-agarose-conjugated monoclonal antibodies (Santa Cruz Biotechnology). ChIP experiments used to test the presence of methylated H3K4 were performed with anti-H3K4me (Upstate Biotechnology), anti-H3K4me2 and -H3K4me3 (Abcam).

RESULTS

Set1C Composition and Roles of the Subunits—To analyze direct interactions within Set1C and interactions with Set1, we used a bacterial two-hybrid system based on the interaction-mediated reconstruction of adenylate cyclase (36). The use of a

Regulation of Histone H3 Lysine 4 Methylation

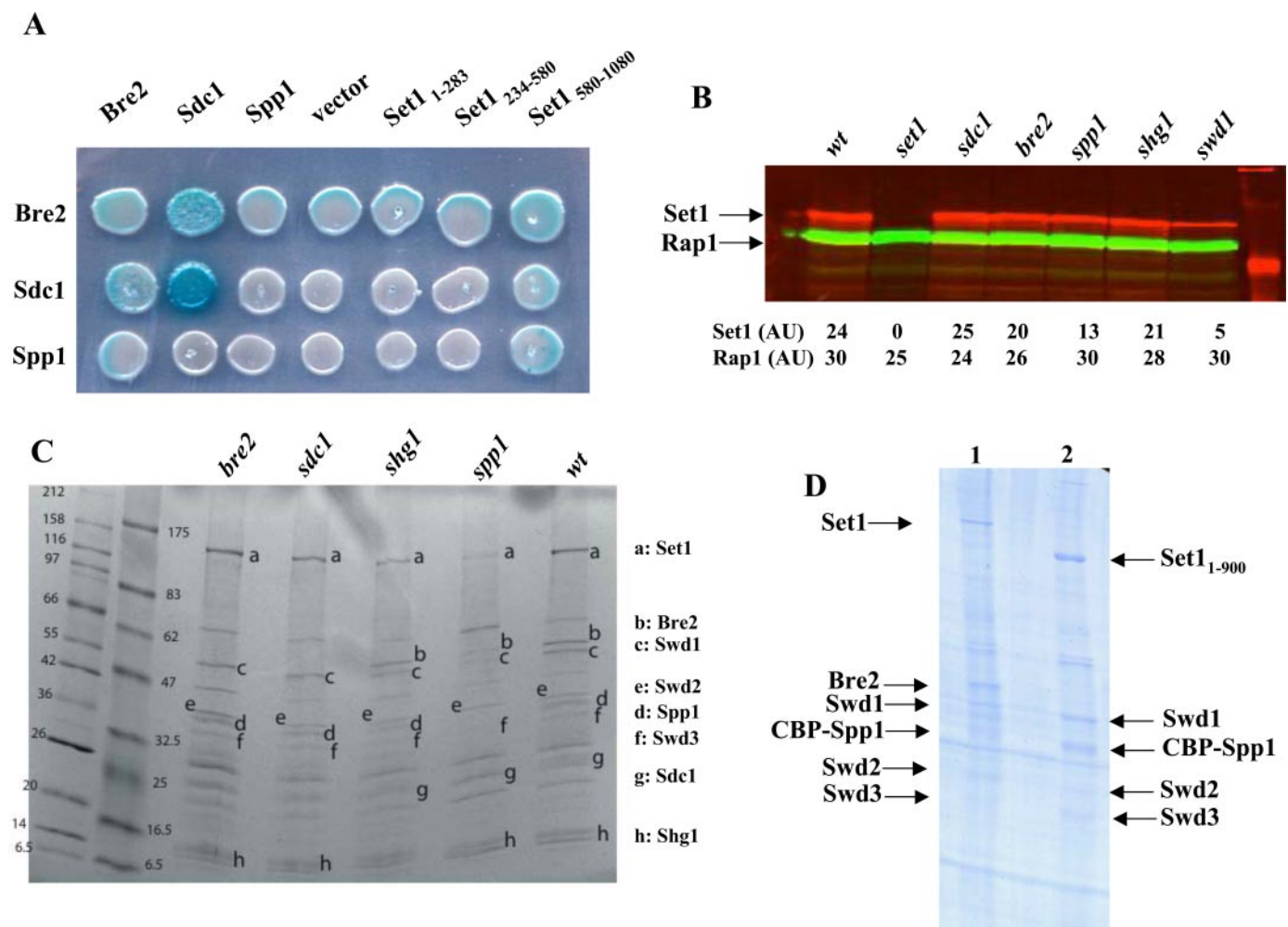


FIGURE 1. Role of Set1C subunits in Set1 stability and complex assembly. *A*, protein/protein interactions within Set1C detected by the *E. coli* two-hybrid system. The indicated proteins or protein fragments were genetically fused to the two complementary fragments T25 and T18 from the catalytic domain of *B. pertussis* and co-expressed in *E. coli cya* cells. Interaction between two-hybrid proteins resulted in functional complementation between the T25 and the T18 fragments leading to cAMP synthesis and activation of *E. coli* catabolic operons. Interactions were revealed by the presence of β -galactosidase activity visualized in X-gal plates. The efficiencies of functional complementation between the indicated proteins were quantified by measuring β -galactosidase activities in suspensions of toluene-treated *E. coli* cells expressing the indicated T18 and T25 fusion proteins. *B*, Set1 stability is affected in the absence of Spp1 and Swd1. The presence of Set1 and Rap1 proteins were checked by Western blotting with monoclonal anti-Set1 antibodies and anti-Rap1 polyclonal antibodies. Infrared fluorescence-labeled secondary antibodies IRDye 800 anti-rabbit IgG and Alexa Fluor® 680 anti-mouse IgG were used to allow simultaneous direct infrared fluorescence detection of Set1 and Rap1 with an Odyssey imaging system. Quantified signals expressed in arbitrary units (AU) are shown below the image of each blot. *C*, TAP purification of Set1 in Set1C mutants. The TAP tag was fused to the N terminus of Set1 and expressed in strains lacking the indicated Set1C subunits. Affinity-purified Set1 and its associated proteins were separated on 7–25% SDS-PAGE and visualized by staining with Coomassie Blue. The bands present in the gel are identified in the figure. *D*, TAP purification of Spp1 in strains expressing full-length Set1 (lane 1) and Set1-(1–900) (lane 2). Mass spectrometry-identified subunits are indicated by arrows. Although not observed as visible bands, the presence of Sdc1 and Shg1 in lane 1 and Shg1 only in lane 2 was established by mass spectrometry analysis of slices of the low-molecular-weight regions of the gel. CBP represents the calmodulin-binding protein tag.

bacterial two-hybrid system avoids the detection of indirect interactions between yeast proteins mediated by endogenous proteins acting as bridging molecules. In this assay, the proteins of interest were fused to two fragments (T25 and T18) of the catalytic domain of *B. pertussis* and co-expressed in an *E. coli* strain deficient for adenylate cyclase. Interaction of the two proteins resulted in a functional complementation between T25 and T18, leading to cAMP synthesis and in turn to transcriptional activation of catabolic operons, such as the lactose operon. The N terminus of Set1 (Set1-(1–283)), the double RNA recognition motif (RRM1-RRM2, Set1-(234–580)) (39), and a large C-terminal domain (Set1-(580–1080)), as well as Bre2, Sdc1, Spp1, and Swd3 were cloned either into the T25 or T18 expression plasmids (or in both) to generate recombinant plasmids expressing hybrid proteins in which the polypeptides

were fused at the C terminus of either the T25 or T18 fragment of the *B. pertussis* adenylate cyclase. The efficiencies of functional complementation were detected on McConkey and X-gal plates. In this assay, Bre2 and Sdc1 interact with each other, and Sdc1 also has the ability to interact with itself (Fig. 1A). Furthermore, the large C-terminal Set1-(580–1080) fragment interacts with Bre2 and Spp1.

We examined the effect of each Set1C member on Set1 stability. Set1 protein levels were compared with Rap1 (as an internal control) by Western blots in strains lacking each of the nonessential Set1C subunits. We used anti-Set1 mouse monoclonal antibodies (14) and anti-Rap1 rabbit polyclonal antibodies and two different secondary antibodies allowing the simultaneous direct infrared fluorescence detection of the Rap1 and Set1 antibodies (Fig. 1B). The amount of Set1

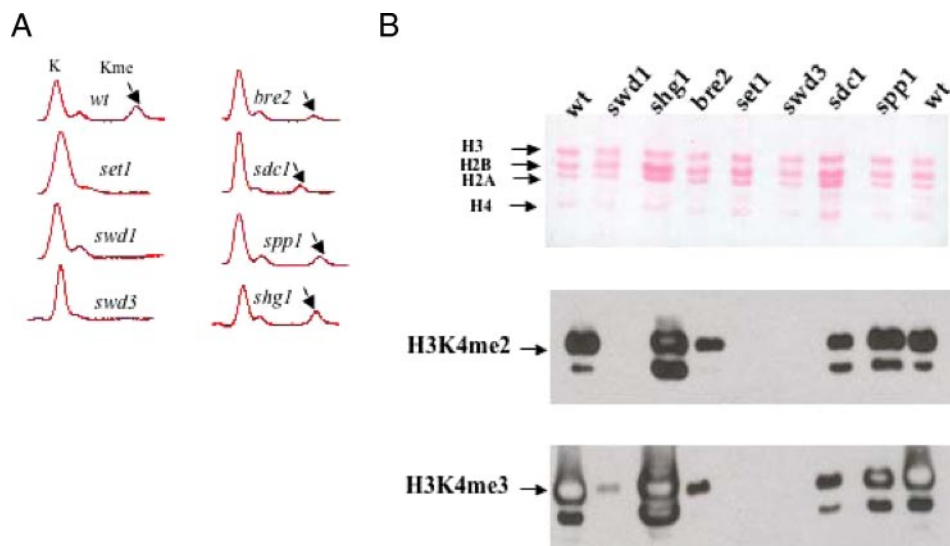


FIGURE 2. **Histone H3 methylation at lysine 4 in Set1C mutants.** A, histones were purified and separated by SDS-PAGE. Upper panel, the N-terminal sequence determination of histone H3 isolated from the indicated strain was performed by stepwise Edman degradation. The chromatograms corresponding to cycle 4 are shown. The elution positions of K and Kme (monomethylated lysine) are indicated on the first chromatogram. Kme is shown by an arrow. Lower panel, the ratio K:Kme and the relative ratios of the different lysine forms are indicated in Table 1. wt, wild type. B, purified histones from the indicated strains were transferred onto nitrocellulose and stained with Ponceau red (top). The histones were analyzed on Western blots with the indicated antibodies for the presence of the dimethylated H3K4 (*H3K4me2*) and trimethylated H3K4 (bottom, *H3K4me3*).

TABLE 1
Quantification of histone H3 methylation

Strain	Kme/K	K	Kme	Kme(2 + 3)
		%	%	%
wt	0.34	49	17	34
<i>set1</i>	0	100	0	0
<i>swd1</i>	0	100	0	0
<i>swd3</i>	0	100	0	0
<i>bre2</i>	0.08	87	7	6
<i>sdcl</i>	0.10	85	9	6
<i>spp1</i>	0.20	72	14	14
<i>shg1</i>	0.34	44	15	41

was similar to wild type in *sdcl*, *bre2*, and *shg1* mutants. In *swd1* and *spp1* mutants, the amount of Set1 was reduced by 6- and 2-fold, respectively. Similar results were obtained in an experiment where TAP-tagged Set1 levels were evaluated by the detection of the TAP tag. In this experiment, deletion of either Swd1 or Swd3 reduced TAP-Set1 levels by >10-fold (data not shown). The expression of Set1 was also tested in several *swd2 ts* mutants (19). In all of the tested *swd2* mutants, Set1 levels were reduced by >5-fold (not shown). We conclude that the Swd1-Swd3 heterodimer and Swd2 are required for Set1 stability and consequently for Set1C integrity.

We next analyzed the protein complexes associated with TAP-tagged Set1 by mass spectrometry. Because the absence of Swd1 or Swd3 reduces Set1 stability, these experiments were only performed with strains lacking Bre2, Sdc1, Spp1 and Shg1 (Fig. 1C). In accordance with the existence of a Bre2-Sdc1 heteromer, the deletion of either abolishes the association of the other. However, the rest of the complex remained. Therefore, no other subunit requires Bre2-Sdc1 for association with Set1C. Furthermore, the complex does not require Bre2-Sdc1 for integrity. The absence of either of the two other deletable sub-

units, Spp1 or Shg1, did not destroy the complex, and all of the other subunits were still associated with Set1 (Fig. 1C). These results are in agreement with previous results showing that, in the absence of Set1, two heteromeric remnants of the complex, Swd1-Swd3 and Bre2-Sdc1, were found (2).

Interactions within the complex were further analyzed in a *set1* strain in which endogenous Spp1 was TAP-tagged and a centromeric plasmid either expressed the full-length Set1 or the Set1-(1-900) fragment (1). In both experiments, Swd1, Swd2, Swd3, and Shg1 were retrieved with Spp1-TAP (Fig. 1D), whereas Bre2 and Sdc1 were only retrieved when the full-length Set1 was expressed. Thus, the association of Spp1, Swd1, Swd2, Swd3, and Shg1 to Set1 do not depend on the C-terminal fragment Set1-(900-1080), whereas the association of

Bre2 and Sdc1 requires the presence of Set1-(900-1080).

Distinct Roles of Set1C Subunits in H3K4 Methylation—Lysines can be mono-, di-, or trimethylated. Previously, we developed a method to quantify the unmethylated and monomethylated forms of H3K4 on histones purified from *S. cerevisiae* by N-terminal sequencing of H3 (27). This method determines the ratio between the unmethylated and monomethylated H3K4 lysine by measuring the area under the K and Kme peaks at cycle 4 (Fig. 2A). Consequently, the amounts of di- and trimethylated H3K4 can be deduced (Table 1). Histones were purified from mutant strains disrupted for each one of the seven nonessential members of the Set1 complex. No peaks corresponding to monomethylation of H3K4 in the *swd1* and *swd3* mutants were detected. The *bre2* and *sdcl* mutants exhibited ~50% of the wild-type monomethylation level, and the levels of di- and trimethylated H3K4 were reduced by ~5-fold. The *spp1* mutant did not affect monomethylation, and the di- and trimethylated H3K4 amount was reduced by ~3-fold. Unexpectedly, the *shg1* mutation provoked a small but reproducible increase of the di- and trimethylation H3K4 level and a slight reduction of the amount of unmethylated H3K4. Because N-terminal sequencing could not distinguish between di- and trimethylation, we analyzed these methylations by Western blot with H3K4me2 and H3K4me3 antibodies (Fig. 2B). For precision, these analyses were performed with purified histones (Fig. 2B). They agreed with and extended previous studies with whole-cell extracts (4, 14, 15, 16). Trimethylation was severely affected (but not abolished) in the *bre2* and *sdcl* mutants, whereas dimethylation was reduced by ~4-fold. In the *spp1* mutant, dimethylation was not affected; however, trimethylation was reduced by ~3-fold.

Regulation of Histone H3 Lysine 4 Methylation

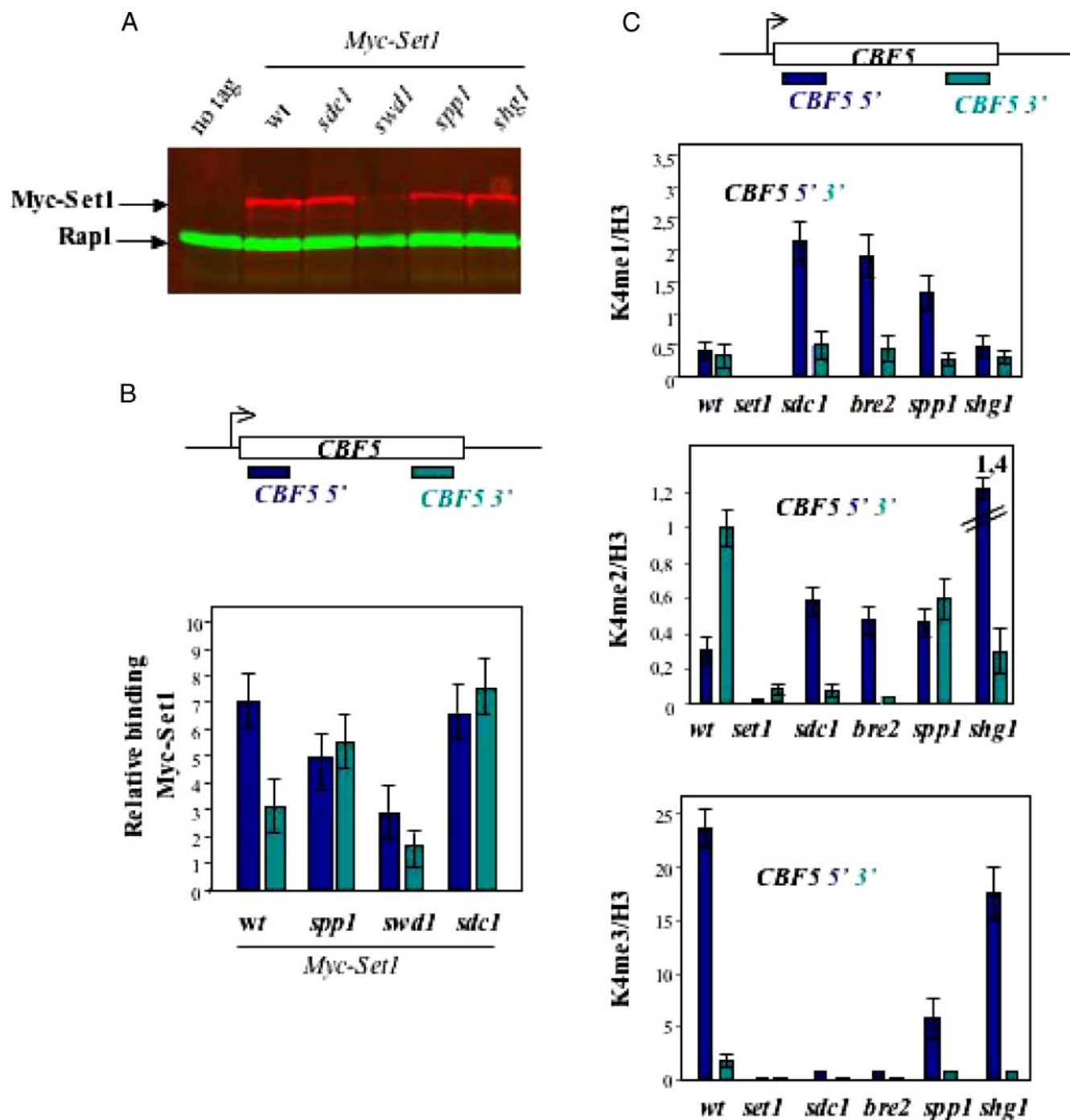


FIGURE 3. Loss of the individual subunits of the Set1 complex differentially affects the distribution of H3K4 methylation. *A*, a 9-Myc tag was placed just after the ATG of *SET1* (see "Experimental Procedures") and introduced in wild-type (*Myc-Set1*) and in the indicated mutant strains (*sdcl*, *swd1*, *spp1*, and *shg1*). The expression of *Myc-Set1* was checked as in Fig. 2. *B*, ChIP of *Myc-Set1* (9E10, Santa Cruz Biotechnology) at positions on *CBF5* corresponding to either early 5'-coding regions (*blue*) and 3' regions of the coding sequence (*green*). ChIPs were performed from cells grown in YPD medium (yeast extract, peptone, dextrose). The specificity of the *Myc-Set1* ChIP was controlled by performing the ChIP experiment in the untagged *Set1* strain (no tag). Quantitative PCR analyses were performed by real-time PCR using a Light Cycler (Roche Applied Science). The values (IP:input) obtained for the binding of the tagged *Myc-Set1* were normalized to the values obtained when the ChIP was carried out with the same antibodies in the untagged *Set1* strain. This value thus reflects the *Myc-Set1*-specific binding. *C*, ChIP experiments were performed with anti-H3K4me1 (Upstate Biotechnology), anti-H3K4me2, -H3K4me3, and -H3Ct (Abcam) antibodies at the same positions in *CBF5* as in Fig. 4*B*. The amounts of methylated H3K4 were normalized to the level of H3. In these experiments, the ChIPs against H3 and the methylated states of H3K4 were performed on the same input; therefore, the values represented correspond to IP:IP_{H3}. Errors bars were calculated from three independent ChIP experiments. *wt*, wild type.

Therefore, *Set1C* subunits can be grouped into four classes. 1) *Swd1* and *Swd3* form a heterodimer that is essential, together with *Swd2*, for *Set1* stability and integrity of the complex and therefore for any H3K4 methylation. 2) *Sdc1* and *Bre2* form a

heteromer that is required for trimethylation of H3K4 and efficient dimethylation but not for *Set1C* integrity. 3) *Spp1* stabilizes *Set1* and promotes efficient H3K4 trimethylation. 4) *Shg1* seems to slightly inhibit H3K4 di- and trimethylation.

Loss of Individual Set1C Subunits Differentially Affects the Distribution of H3K4 Methylation at the 5' and 3' Coding Regions of Active Genes—We looked for the effects caused by the loss of individual Set1C subunits on the distribution of Set1 protein as well as H3K4 methylation along the coding region of two active genes, *RPL17A* and *CBF5*. Results for both were similar; therefore, only *CBF5* is presented (Fig. 3). To locate Set1, it was Myc-tagged at its N terminus. The tag did not interfere with the ability of Set1 to methylate H3K4 (27). We first tested the expression of the Myc-tagged Set1 in *swd1*, *sdc1*, and *spp1* cells. As expected, the amount of the Myc-tagged Set1 in *swd1* was compromised (Fig. 3A), and occupancy on *CBF5* was diminished (Fig. 3B). The presence of Myc-Set1 at the 5' end of the coding region was not affected in *sdc1* or *spp1* mutant cells but was slightly increased at the 3' end of *CBF5*. These results indicate that Spp1 and Sdc1 were not required for the association of Set1 with the coding region of *CBF5* (and *RPL17A*).

Next, we looked at mono-, di-, and trimethylation of H3K4 at the 5' and 3' ends of *CBF5* in *sdc1*, *bre2*, *spp1*, and *shg1* cells (Fig. 3C). A *set1* mutant was used to control the specificity of the ChIP. The amounts of methylated H3K4 are normalized to the level of H3. As can be seen, loss of Bre2 or Sdc1 had the same effect, namely an increase of monomethylation and a loss of trimethylation at the 5' end and a loss of dimethylation at the 3' end. Loss of trimethylation at the 5' end was the largest effect by far. Loss of Spp1 produced a similar profile; however, both the losses of 5' or 3' end dimethylation were less pronounced. Loss of Shg1 produced a different profile. Only moderate changes in mono- and trimethylation at either the 5' or 3' ends were observed. However, dimethylation was increased at the 5' end and decreased at the 3' end.

These results are in reasonable agreement with the global analyses (Fig. 2). The Bre2-Sdc1 heteromer appears to be required for 5' and 3' end dimethylation. In its absence, 5' end monomethylation increases, reflecting the possibility that the transition of monomethylation to trimethylation at 5' ends depends on Bre2-Sdc1. Potentially, the loss of 5' end trimethylation precedes the loss of 3' end dimethylation. Some support for this proposition can be found in the similar effects caused by loss of Spp1, where losses of 5' end trimethylation and 3' end dimethylation are both milder.

Set1C Associates with Both Phosphorylated Forms of pol II—The Paf1 complex recruits Set1C to chromatin (probably because of a physical interaction between the two complexes and pol II phosphorylated at Ser⁵ of its CTD) (23, 24). The Paf1 complex associates with elongating pol II over the entire mRNA-coding region (40), and several reports show that Set1 and dimethylation can be found on the middle and end of actively transcribed genes (25–27). However CTD Ser⁵ phosphorylation is limited to the initiating pol II, and the elongating pol II found toward the 3' end of genes is characterized by CTD Ser² phosphorylation (41). We therefore examined which pol II CTD-phosphorylated forms co-purify with Set1C (Fig. 4). Although Set1C was poorly associated with nonphosphorylated pol II (Fig. 4, upper panel; compare the signals in cell extracts with the ones of TAP-purified proteins), we found that Set1C associates with pol II phosphorylated at both Ser⁵ and Ser² (Fig. 4). Although a significant fraction of Set1C may be

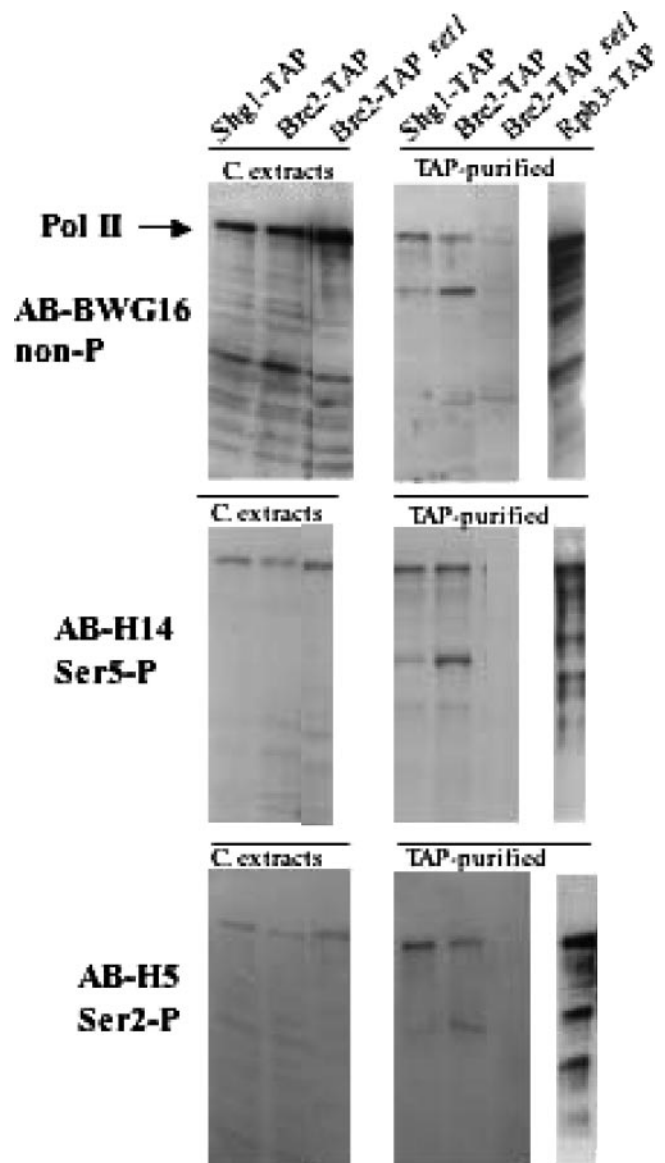


FIGURE 4. Set1 interacts with both forms of phosphorylated pol II. Shg1-TAP, Bre2-TAP, and Rbp3-TAP were affinity-purified with their associated proteins. Whole-cell extracts (*left*) and affinity-purified proteins (*right*) of the indicated strains were analyzed by Western blot, with the indicated antibodies (BWG16, H14, and H5) recognizing nonphosphorylated pol II, pol II CTD phosphorylated at Ser⁵, and pol II CTD phosphorylated at Ser², respectively. The Rbp3-TAP lanes are from the same blot; however, the exposure was for 10× less time.

released from pol II after the loss of Ser⁵ phosphorylation, this result suggests that Set1C can also associate with pol II, which is CTD-phosphorylated on Ser².

DISCUSSION

The subunit composition of the Set1 complex has been comprehensively characterized in both budding and fission yeasts (this study and in Refs. 2, 3, 13–15, 21, and 22), and recently the human counterpart has been shown to be nearly identical (16, 17). These data indicate that Set1C is strongly conserved in eukaryotes, as suggested in Ref. 2. Although it has been purified several times from different laboratories and sources, the size of the complex has not been established yet. By sizing column or sucrose gradient centrifugation, the complex has been twice

Regulation of Histone H3 Lysine 4 Methylation

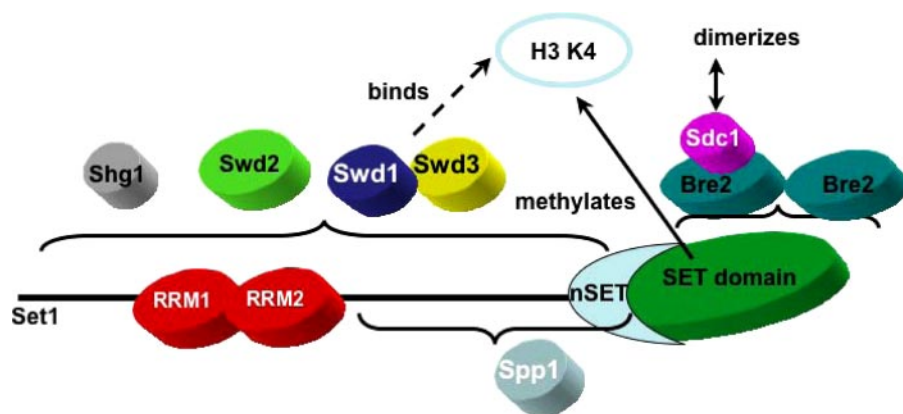


FIGURE 5. A model for Set1C organization. The core complex is made by Set1, the Swd1-Swd3 heterodimer, and probably Swd2. According to results obtained with higher eukaryotes (46), this core complex may mediate binding to dimethylated H3K4 (dashed lines), although it has not been shown that any of the Swd proteins is able to bind dimethylated H3K4. Shg1, Spp1, and the heteromer Bre2-Sdc1 are not required for complex assembly. The interaction of the heteromer Bre2-Sdc1 with Set1C requires the presence of the SET domain (Set1-900–1040), whereas the region of interaction with Spp1 lies in the Set1-(560–900) region. The catalytic SET domain likely integrates different inputs from Bre2-Sdc1 and Spp1 (12). Sdc1 is shown to contact only one of the two molecules of Bre2, because previous observations indicate that the stoichiometry of TAP-tagged Bre2 and Sdc1 is 2:1, whereas the ratio of TAP-tagged Sdc1 and Bre2 is 1:1 (2). In this model, the Set1 complex is proposed to dimerize through the ability of Sdc1 to make homodimers.

estimated to be ~1 MDa and twice ~500 kDa (2, 13, 14, 16). Because the smallest possible size for the 8-subunit yeast complex is 360 kDa (that is, only 1 molecule of each subunit/complex), the size estimates permit three different propositions. Is the complex (i) monomeric with one active site, (ii) dimeric, as suggested (15), with two active sites, or (iii) agglomerative?

Previously, it was shown that the complex is based on a scaffold of Set1 (2). Here, we extended the scaffold to include the Swd1-Swd3 heterodimer and probably Swd2. Unexpectedly, we report that the Swd1-Swd3 heterodimer does not need the SET domain for stable interaction with Set1C. This was unexpected, because the mammalian homologues of these proteins (Rbbp5 and WDR5, respectively) have been found in partially characterized complexes with the mammalian H3K4 methyltransferases Mll and Mll2 (42–44) as well as with the mammalian homologues of Set1 (16–17). The only obvious region of homology shared by Mll and Set1 is the SET domain, and even their preSET domains are divergent (2). Therefore, logical simplicity would infer that the highly conserved Swd1-Swd3 (Rbbp5-WDR5) heterodimer and its highly conserved interaction with H3K4 methyltransferases should interact with the common element, the highly conserved SET domain. However, stable interaction between *S. cerevisiae* Set1C and the Swd1-Swd3 heterodimer is not reliant on the Set1 SET domain. Set1 must therefore have, outside of the SET domain, a conserved structural motif required for the interaction with the heterodimer Swd1-Swd3.

Recently, it was discovered that WDR5 binds to dimethylated H3K4 (45). Hence, all characterized complexes conveying H3K4 methylation appear to include a subunit that binds the methylated site. Furthermore, we show that this subunit in yeast is essential for complex integrity. Because the Swd1-Swd3 heterodimer is essential to complex integrity and Set1 protein expression, the total loss of H3K4 methylation in their absence is explained.

By several criteria, including a bacterial two-hybrid system, we conclude that Sdc1 and Bre2 interact with each other and Sdc1 interacts with itself. In most published studies, both mutants *bre2* and *sdc1* share similar phenotypes (2, 4, 14, 22). Furthermore, in the two-hybrid system, Bre2 interacts with Set1-(580–1080), and TAP purification indicates that Bre2 and Sdc1 do not associate with a Set1C that lacks the SET domain. These results suggest that Bre2 interacts directly with Set1 and that this interaction requires the SET domain (Set1-(900–1048)).

We conclude that the Bre2-Sdc1 heteromer associates onto the Set1/Swd1/Swd3 scaffold. This association requires the SET domain. Again, the evidence from partially

characterized mammalian H3K4 methyltransferase complexes indicates that the association of the Bre2-Sdc1 heteromer (human names ASH2-DPY30) with H3K4 methyltransferases (17, 42) is highly conserved. Furthermore, the heteromer also exists independently of H3K4 methyltransferase complexes. In *Schizosaccharomyces pombe*, it is also found associated with the Jumonji domain protein Lid2 (15), which is the homologue of the trithorax group (trxG) protein Lid (little imaginal discs), (46). Ash2 (Bre2) is itself in the trxG and is highly conserved. Sdc1 is also highly conserved. In *Caenorhabditis elegans*, its homologue Dpy30 is involved in dosage compensation (47). We report here that Sdc1 has a very strong ability to interact with itself. This is in concordance with fact that the most conserved part of Sdc1 is very similar to the dimerization interface of the regulatory subunit of cAMP-dependent type II protein kinases (2). This interface consists of two α -helices that dimerize into a four-helix bundle (48). Hence, we suggest that Sdc1 mediates a dimerization of the Bre2-Sdc1 heteromer, which may contribute to dimerization of the Set1, Swd1-Swd3 scaffold. If this proposition is correct, then the absence of the Bre2-Sdc1 heteromer may cause Set1C to fall into two halves. If so, the residual H3K4 di- and trimethylation seen in the absence of Bre2 or Sdc1 may reflect methylation by a half-complex with only one active site. Conversely, the presence of two active sites in the complex may confer a degree of coordination, which could play a role in a process such as hemimethylation and epigenetic maintenance.

The absence of Bre2 or Sdc1 reduces the dimethylation of H3K4 and severely affects trimethylation, consistent with previous results (4, 22). ChIP experiments showed that mutation of *sdc1* did not reduce the presence of Set1 in the 5' and 3' coding regions of active genes. Consistent with this observation, we found that inactivation of Bre2 or Sdc1 resulted in an increase of H3K4 mono- and dimethylation at the 5' coding regions of active genes at the expense of trimethylation. These results suggest that Bre2 and Sdc1 regu-

late Set1 histone methyltransferase activity rather than affecting the presence of Set1 in specific chromosomal regions. We further observed that deleting *BRE2* or *SDC1* abolished H3K4 dimethylation at the 3' end of coding regions. Therefore Bre2 and Sdc1 promote the catalytic activity of Set1 in 3' gene regions transcribed by the elongating pol II.⁶

Consistent with this proposition, we showed that Set1C associates with both serine 5 and serine 2 phosphorylated forms of pol II. Recent studies indicate that trimethylated H3K4 peaks at the beginning of the transcribed portions of genes, dimethylated H3K4 is most enriched in the middle of genes, and monomethylated H3K4 is found predominantly at the end of genes (26). Our results indicate that Set1C can associate with pol II along the complete length of transcribed regions, with different subunits differentially stimulating catalytic activity according to the phase of the transcriptional cycle.

In addition to the highly conserved heteromers that interact with both Set1- and Mll-type H3K4 methyltransferases, Set1C includes three individual proteins that do not appear to interact with any subunit other than Set1. They include the plant homeodomain finger protein Spp1, the WD40 repeat protein Swd2, and the pioneer protein Shg1. The two-hybrid analysis also revealed an interaction between Spp1 and Set1-(580–1080). However, in contrast to Bre2, Spp1 co-purifies with Set1-(1–900), suggesting that it interacts with a region of Set1 lying between residues 580 and 900, which includes the N-SET domain (12). The counterpart for Spp1 in the human Set1C is CFP (CXXC finger protein), also known as CGBP for CpG-binding protein. Both proteins include a very similar plant homeodomain finger. Additionally CFP possesses a domain that binds unmethylated CpG dinucleotides.

Swd2 is probably a homodimer (2) whose function remains unclear; however, it is also a subunit of CPF (cleavage and polyadenylation factor) and therefore could play a similar role for Set1C and CPF at the 3' end of genes. The smallest subunit, Shg1, does not appear to be well conserved in evolution. The *S. pombe* Set1C has a counterpart; however, our data base searches have only found homologues among other fungi (not shown). It appears to play a moderately repressive role on Set1C methylation activity.

We extend pre-existing data on the structure of Set1C (2, 4, 13–16, 22), which permits the advancement of a model (Fig. 5). The core complex is made by Set1, the Swd1-Swd3 heterodimer, and probably Swd2. Hence, the association between the H3K4 enzyme and the H3K4 binding domain is fundamental to complex integrity. The C-terminal region of Set1 contacts Bre2. In turn, Bre2 forms a heteromer with Sdc1, which has the ability to interact with itself. Spp1 interacts with a region of Set1 lying between residues 580 and 900. Both the Bre2-Sdc1 heteromer and Spp1 differentially regulate H3K4 di- and trimethylation during the transcription cycle of pol II. Further work is required to resolve the size

and composition of the complex as well as the role of Swd2. This information will advance the understanding of the most evolutionarily conserved, and to date best understood, SET domain complex.

Acknowledgments—We thank the laboratory of Dr. Emmanuelle Bouveret for help with the *E. coli* two-hybrid system. We thank Drs. Peter Nagy and David Shore for anti-Set1 monoclonal and anti-Rap1 antibodies, respectively. We thank Drs. David Allis, Scott Briggs, and Pierre Luciano for reagents, helpful comments, and advice.

REFERENCES

- 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
- Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) *EMBO J.* **20**, 7137–7148
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) *Nature* **419**, 407–411
- Morillon, A., Karabetsou, N., Nair, A., and Mellor, J. (2005) *Mol. Cell* **18**, 723–734
- Bryk, M., Briggs, S. D., Strahl, B. D., Curcio, M. J., Allis, C. D., and Winston, F. (2002) *Curr. Biol.* **12**, 165–170
- Santos-Rosa, H., Bannister, A. J., Dehé, P. M., Géli, V., and Kouzarides, T. (2004) *J. Biol. Chem.* **279**, 47506–47512
- Carvin, C. D., and Klädde, M. P. (2004) *J. Biol. Chem.* **279**, 33057–33062
- Nislow, C., Ray, E., and Pillus, L. (1997) *Mol. Biol. Cell* **8**, 2421–2436
- Sollier, J., Lin, W., Soustelle, C., Suhre, K., Nicolas, A., Géli, V., and de La Roche Saint-André, C. (2004) *EMBO J.* **23**, 1957–1967
- Corde, Y., Schramke, V., Longhese, M. P., Smokvina, T., Paciotti, V., Brevet, V., Gilson, E., and Géli, V. (1999) *Nat. Genet.* **21**, 204–208
- Zhang, K., Lin, W., Latham, J. A., Riefler, G. M., Schumacher, J. M., Chan, C., Tatchell, K., Hawke, D. H., Kobayashi, R., and Dent, S. Y. (2005) *Cell* **122**, 723–734
- Schlichter, A., and Cairns, B. R. (2005) *EMBO J.* **24**, 1222–1231
- 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
- Nagy, P. L., Griesenbeck, J., Kornberg, R. D., and Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 90–94
- Roguev, A., Schaft, D., Shevchenko, A., Aasland, R., and Stewart, A. F. (2003) *J. Biol. Chem.* **278**, 8487–8493
- Lee, J. H., and Skalnik, D. G. (2005) *J. Biol. Chem.* **280**, 41725–41731
- Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) *Genes Dev.* **17**, 896–911
- Roguev, A., Shevchenko, A., Schaft, D., Thomas, H., Stewart, A. F., and Shevchenko, A. (2004) *Mol. Cell. Proteomics* **3**, 125–132
- Dichtl, B., Aasland, R., and Keller, W. (2004) *RNA (Cold Spring Harbor)* **10**, 965–977
- Cheng, H., He, X., and Moore, C. (2004) *Mol. Cell. Biol.* **24**, 2932–2943
- Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shilatifard, A. (2002) *J. Biol. Chem.* **277**, 10753–10755
- Schneider, J., Wood, A., Lee, J. S., Schuster, R., Dueker, J., Maguire, C., Swanson, S. K., Florens, L., Washburn, M. P., and Shilatifard, A. (2005) *Mol. Cell* **19**, 849–856
- Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) *Mol. Cell* **11**, 709–719
- Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2003) *Mol. Cell* **11**, 721–729
- 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
- Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J.,

⁶ J. Mellor, A. Morillon, and B. Cairns, unpublished results.

Regulation of Histone H3 Lysine 4 Methylation

- Lewitter, F., Gifford, D. K., and Young, R. A. (2005) *Cell* **122**, 517–527
27. Dehé, P. M., Pamblanco, M., Luciano, P., Lebrun, R., Moinier, D., Sendra, R., Verreault, A., Tordera, V., and Géli, V. (2005) *J. Mol. Biol.* **353**, 477–484
28. Zhang, L., Schroeder, S., Fong, N., and Bentley, D. L. (2005) *EMBO J.* **24**, 2379–2390
29. Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003) *J. Biol. Chem.* **278**, 34739–34742
30. Ng, H. H., Dole, S., and Struhl, K. (2003) *J. Biol. Chem.* **278**, 33625–33628
31. Sun, Z. W., and Allis, C. D. (2002) *Nature* **418**, 104–108
32. Dover, J., Schneider, J., Tawiah-Boateng, M. A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002) *J. Biol. Chem.* **277**, 28368–28371
33. Shahbazian, M. D., Zhang, K., and Grunstein, M. (2005) *Mol. Cell* **19**, 271–277
34. Larabee, R. N., Krogan, N. J., Xiao, T., Shibata, Y., Hughes, T. R., Greenblatt, J. F., and Strahl, B. D. (2005) *Curr. Biol.* **15**, 1487–1493
35. Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2005) *Mol. Cell* **20**, 589–599
36. Karimova, G., Dautin, N., and Ladant, D. (2005) *J. Bacteriol.* **187**, 2233–2243
37. Gully, D., and Bouveret, E. (2005) *Proteomics* **6**, 282–293
38. Poveda, A., Pamblanco, M., Tafrov, S., Tordera, V., Sternglanz, R., and Sendra, R. (2004) *J. Biol. Chem.* **279**, 16033–16043
39. Tresaugues, L., Dehe, P. M., Guerois, R., Rodriguez-Gil, A., Varlet, I., Salah, P., Pamblanco, M., Luciano, P., Quevillon-Cheruel, S., Sollier, J., Leulliot, N., Couprie, J., Tordera, V., Zinn-Justin, S., Chavez, S., van Tilbeurgh, H., and Geli, V. (2006) *J. Mol. Biol.* **359**, 1170–1181
40. Mueller, C. L., Porter, S. E., Hoffman, M. G., and Jaehning, J. A. (2004) *Mol. Cell* **14**, 447–456
41. Buratowski, S. (2003) *Nat. Struct. Biol.* **10**, 679–680
42. Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., Hayes, D. N., Shanmugam, K. S., Bhattacharjee, A., Biondi, C. A., Kay, G. F., Hayward, N. K., Hess, J. L., and Meyerson, M. (2004) *Mol. Cell* **13**, 587–597
43. Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004) *Mol. Cell Biol.* **24**, 5639–5649
44. Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005) *Cell* **121**, 873–885
45. Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005) *Cell* **121**, 859–872
46. Gildea, J. J., Lopez, R., and Shearn, A. (2000) *Genetics* **156**, 645–663
47. Hsu, D. R., Chuang, P. T., and Meyer, B. J. (1995) *Development (Camb.)* **121**, 3323–3334
48. Newlon, M. G., Roy, M., Morikis, D., Hausken, Z. E., Coghlan, V., Scott, J. D., and Jennings, P. A. (1999) *Nat. Struct. Biol.* **6**, 222–227