

## Genome-Wide Dynamics of SAPHIRE, an Essential Complex for Gene Activation and Chromatin Boundaries<sup>∇</sup>

Matthew Gordon,<sup>1</sup> Derick G. Holt,<sup>1</sup> Anil Panigrahi,<sup>1</sup> Brian T. Wilhelm,<sup>2</sup> Hediye Erdjument-Bromage,<sup>3</sup> Paul Tempst,<sup>3</sup> Jürg Bähler,<sup>2</sup> and Bradley R. Cairns<sup>1\*</sup>

Howard Hughes Medical Institute, Department of Oncological Sciences, Huntsman Cancer Institute, 2000 Circle of Hope, University of Utah, Salt Lake City, Utah 84112<sup>1</sup>; The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom<sup>2</sup>; and Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021<sup>3</sup>

Received 1 November 2006/Returned for modification 19 December 2006/Accepted 9 March 2007

**In this study, we characterize a four-protein nucleosome-binding complex from *Schizosaccharomyces pombe*, termed SAPHIRE, that includes two orthologs of human Lsd1, a histone demethylase. The SAPHIRE complex is essential for cell viability, whereas *saphire* mutants lacking key conserved catalytic residues are viable but thermosensitive, suggesting that SAPHIRE has both an important enzymatic function and an essential nonenzymatic function. SAPHIRE is present in (or adjacent to) particular heterochromatic loci and also in the transcription start site regions of many highly active polymerase II genes. However, ribosomal protein genes are notably SAPHIRE deficient. SAPHIRE promotes activation, as target genes are selectively attenuated in *saphire* mutants. Interestingly, *saphire* mutants display increased histone H3 lysine 4 dimethylation, a modification typically associated with euchromatin. SAPHIRE localization is dynamic, as activated genes rapidly acquire SAPHIRE. Furthermore, *saphire* mutants dramatically shift a heterochromatin-euchromatin boundary in Chr1, suggesting a novel role in boundary regulation.**

Chromatin is a dynamic material that partitions chromosomes into functional domains (telomeres, centromeres, and rRNA gene arrays) and also partitions individual genes into segments, each with a distinct role in transcriptional regulation. All chromatin regions contain arrays of nucleosomes, which consist of 147 base pairs of DNA wrapped around an octameric disk of histone proteins (24, 34). However, different chromatin regions have distinctive characteristics, including (i) differences in chromatin composition, including histone variants, linker histones, and associated nonhistone chromatin architectural proteins; (ii) structural diversity (nucleosome positioning or compaction); and (iii) variation in covalent modifications of the DNA and histones (34).

Covalent histone modifications serve as docking sites for protein domains present in specific factors, which in turn endow the region with unique characteristics (46). Methylation of lysines in the unstructured termini (tails) of histones regulates a wide array of DNA-templated processes, including transcription (12, 47). Methylation of lysines in histone tails is mediated by two families of enzymes, namely, the DOT family and the SET family (30, 40). These enzymes can mono-, di-, or trimethylate lysines, and even the subtle difference between di- and trimethylation of the same lysine residue can be discriminated by recognition domains and, therefore, used to recruit distinct factors (21, 33, 52).

Dimethylation of histone H3 lysine 4 (H3K4me2) is generally thought to make chromatin permissive to transcription,

whereas trimethylation of the same residue (H3K4me3) is associated only with transcriptionally active genes (38). Set1p, the protein which is responsible for all H3K4 methylation in *Saccharomyces cerevisiae* (budding yeast) (6, 19), bears several autoregulatory domains which control whether the enzyme di- or trimethylates (6, 39). This observation argues that a balance between these two very similar methyl marks is critical for homeostasis. Indeed, chromatin immunoprecipitation (ChIP)-microarray experiments with budding yeast demonstrated that, on average, these two marks are not identically distributed across genes. H3K4me3 is concentrated at the 5' ends of transcribed genes, whereas H3K4me2 tends to peak in the middle of genes but shifts closer to the 3' end when genes are highly transcribed (23, 35). Although several factors that bind H3K4 methylated tails have been identified, it is still unclear what role H3K4me2 and H3K4me3 play in transcription. However, deletion of *SET1* leads to defects in gene activation, highlighting the importance of these marks (5).

Histone methylation is not restricted to transcriptional activation, as histone lysine methylation at alternative locations on the H3 tail promotes repression. First, methylation of H3K9 is present at sites of constitutive heterochromatin and recruits repressive proteins, such as HP1, to enforce the silent state of these domains (10). Second, H3K36 methylation in the bodies and 3' ends of transcribed genes recruits a repressive complex that inhibits spurious transcription initiation from cryptic promoters inside the gene (7, 13–15, 29).

Histone methylation is not a static mark in vivo, as shown by the discovery of histone lysine demethylases (HDMs). The largest family of HDMs is the Jumonji (JMJ) domain family. To date, five JMJ HDMs have been characterized, with substrate specificities directed toward H3K9, H3K36, or both (18, 49–51, 53). Thus, this class of HDMs appears to demethylate

\* Corresponding author. Mailing address: University of Utah, Huntsman Cancer Institute, 2000 Circle of Hope, Salt Lake City, UT 84112. Phone: (801) 585-1822. Fax: (801) 585-6410. E-mail: bcairns@hci.utah.edu.

<sup>∇</sup> Published ahead of print on 19 March 2007.

repressive methyl marks. Significantly, the active-site chemistry of JMJ domains allows for demethylation of mono-, di-, or trimethylated lysine residues.

However, only one protein, Lsd1, has been isolated that can demethylate H3K4. Lsd1 belongs to a class of proteins called nuclear amine oxidases (NAOs), which are defined by the presence of an N-terminal SWIRM domain followed by an amine oxidase domain (43). The SWIRM domain is found in many nuclear proteins, most of which are members of chromatin-associated multiprotein complexes (e.g., the SAGA histone acetyltransferase complex or the SWI/SNF remodeler complex) (1). The amine oxidase domain is a flavin adenine dinucleotide (FAD)-dependent oxidoreductase. Amine oxidases use the oxidative potential of FAD to break a carbon-nitrogen bond that forms the substrate amine. This reaction can also be thought of as the dealkylation of an amine, as is the case in a histone demethylation reaction. Importantly, this class of enzymes cannot oxidize quaternary amines; thus, Lsd1 can only demethylate mono- and dimethylated H3K4 (3, 43).

Although most of the data regarding Lsd1 point to a role in transcriptional repression, a recent study observed that Lsd1 is recruited to the prostate-specific antigen (PSA) gene during activation by the androgen receptor (AR) (20, 27, 43, 44). Small interfering RNA-mediated knockdown of the Lsd1 protein caused a drastic decrease in the activation of AR-responsive reporter genes, suggesting that Lsd1 can be an activator. Knockdown studies also demonstrated that Lsd1 is required for demethylation of repressive H3K9 methylation when PSA is activated. These studies suggested that Lsd1 demethylated H3K9 to promote activation. However, these results are also consistent with Lsd1 activity being required upstream of another H3K9 demethylase rather than Lsd1 being an H3K9 demethylase itself. Indeed, a JMJ domain protein (JHDM2A) with robust *in vitro* H3K9me2 demethylase activity using purified components is also recruited to PSA upon activation by AR. Small interfering RNA studies demonstrated that JHDM2A knockdown creates profound defects in H3K9 demethylation at the PSA gene upon activation, whereas Lsd1 knockdown effects are much more modest (53).

Thus, many important questions remain to be answered about NAOs. For example, most of the work done to date has focused on Lsd1 function at a limited number of gene targets (see Discussion), and it remains to be determined whether the principles elucidated so far are general. Because most of the data point to a repressive function for NAOs, it is particularly important to determine if the activation function seen at PSA is broadly applicable and dependent upon H3K9 demethylation. Importantly, since Lsd1 is present in large complexes, it is important to determine the catalytic and noncatalytic roles of this class of proteins and the contributions of associated proteins to Lsd1 function. Finally, the only biological function attributed to NAOs so far is transcriptional regulation, and since chromatin modifications impact many aspects of chromosome biology, it is imperative to explore other possible roles for NAOs.

Here we report the genome-wide dynamics of SAPHIRE, a novel NAO complex from *Schizosaccharomyces pombe*. Our results demonstrate that SAPHIRE promotes gene activation, most likely through enzymatic and nonenzymatic means. SAPHIRE occupancy is dynamic and shifts to newly activated

genes during heat shock. Furthermore, mutant analysis reveals that defects in transcription of SAPHIRE targets are associated with increased H3K4me2 of the affected genes. Finally, we identify a new role for NAOs in maintaining the boundary between euchromatin and heterochromatin at a telomere.

## MATERIALS AND METHODS

**Yeast strains and media.** Yeast growth, manipulations, and molecular biology were performed according to standard protocols. Strains were grown on yeast extract supplemented with histidine, uracil, leucine, and adenine at 32°C unless otherwise noted. For heat shock experiments, cells were cultured to an optical density at 600 nm of 0.5 and transferred to flasks which were preheated in a 40°C water bath. Cells were incubated, with shaking, for 15 min and harvested for cross-linking and ChIP or RNA isolation.

**Generation of anti-Saf110 antibody and Western blotting.** Full-length Saf110 with an N-terminal six-His tag was expressed in *Escherichia coli* and purified under denaturing conditions, using Ni-nitrilotriacetic acid agarose, following the manufacturer's instructions (QIAGEN). Antibodies against the protein, purified in-gel by Covance, were raised in a rabbit. The antiserum was purified using hydroxyapatite chromatography and used at a dilution of 1:1,000 for Western blotting following standard procedures.

**SAPHIRE purification.** SAPHIRE was purified using a tandem affinity purification (TAP) tag, which carries (in tandem) the protein A motif, a tobacco etch virus (TEV) protease recognition sequence, and a calmodulin-binding protein domain. The TAP tag was placed at the 3' end of either the Saf140p or Saf110p gene by homologous recombination (36). Whole-cell extracts were prepared in buffer A (20 mM Tris-HCl [pH 7.5], 10% glycerol, 400 mM NaCl, 1 mM EDTA, 10 mM ZnOAc, 0.1% NP-40, 0.1 mM dithiothreitol, and a protease inhibitor cocktail). Whole-cell extracts were bound to immunoglobulin G-Sepharose and washed with buffer A, and the tagged proteins were eluted by cleavage with TEV protease in buffer B (20 mM Tris-HCl [pH 8.0], 10% glycerol, 150 mM NaCl, 10 mM ZnOAc, 0.1% NP-40, 1 mM dithiothreitol, and a protease inhibitor cocktail). The TEV eluates were then bound to calmodulin-Sepharose in buffer B containing 2 mM CaCl<sub>2</sub> and eluted in buffer C (20 mM Tris-HCl [pH 7.5], 10% glycerol, 150 mM NaCl, 10 mM ZnOAc supplemented with 2 mM EGTA [no CaCl<sub>2</sub>]). Purified SAPHIRE was resolved by sodium dodecyl sulfate–7.5% polyacrylamide gel electrophoresis and silver staining. The bands were excised, and tryptic digestion was performed in-gel. The peptides were purified on reversed-phase microtip columns. Mass spectrometric analysis (matrix-assisted laser desorption ionization–time-of-flight analysis) was performed using a Bruker Reflex III instrument with delayed extraction. The experimental masses (*m/z*) were used to search a nonredundant *S. cerevisiae* or *S. pombe* protein database (NCBI, Bethesda, MD). Mass spectrometric sequencing (matrix-assisted laser desorption ionization–tandem time of flight–tandem mass spectrometry) of selected peptides was done on a Bruker Ultraflex TOF/TOF instrument.

**Nucleosome and DNA binding.** Gel mobility shift analysis was performed with a 172-bp fragment from *Xenopus laevis* 5S assembled into mononucleosomes (10 ng; 44 fmol) reconstituted with *S. cerevisiae* recombinant histones. Nucleosomes were incubated with ~16 ng of SAPHIRE (~44 fmol) in a 5- $\mu$ l reaction mix containing 20 mM Tris-HCl (pH 7.5), 10% glycerol, 150 mM NaCl, and 10 mM ZnOAc; this resulted in an equimolar concentration of DNA/nucleosomes and SAPHIRE (8.8 nM). After incubation at room temperature for 30 min, the binding reaction mix was chilled on ice and resolved in a precooled native polyacrylamide gel (4% polyacrylamide, 5% glycerol, 1 mM MgCl<sub>2</sub> in 0.5 $\times$  Tris-borate-EDTA), H3K9me2 (Abcam ab1220), and polymerase II (Pol II) (Abcam ab817) were performed using 3  $\mu$ l of antibody and 100  $\mu$ l of Dynabeads.

**ChIP.** ChIP was performed as previously described (37). Briefly, cells grown to mid-log phase were cross-linked in 1% formaldehyde for 30 min. Cells were broken, and the chromatin fraction was sheared to roughly 500-bp fragments. For TAP IP, 4  $\mu$ l of anti-protein A antibody (Sigma) was coupled to 100  $\mu$ l of Dynabeads (Invitrogen). ChIPs of histone H3 (Abcam ab1791), H3K4me2 (Upstate 03-070), H3K9me2 (Abcam ab1220), and polymerase II (Pol II) (Abcam ab817) were performed using 3  $\mu$ l of antibody and 100  $\mu$ l of Dynabeads.

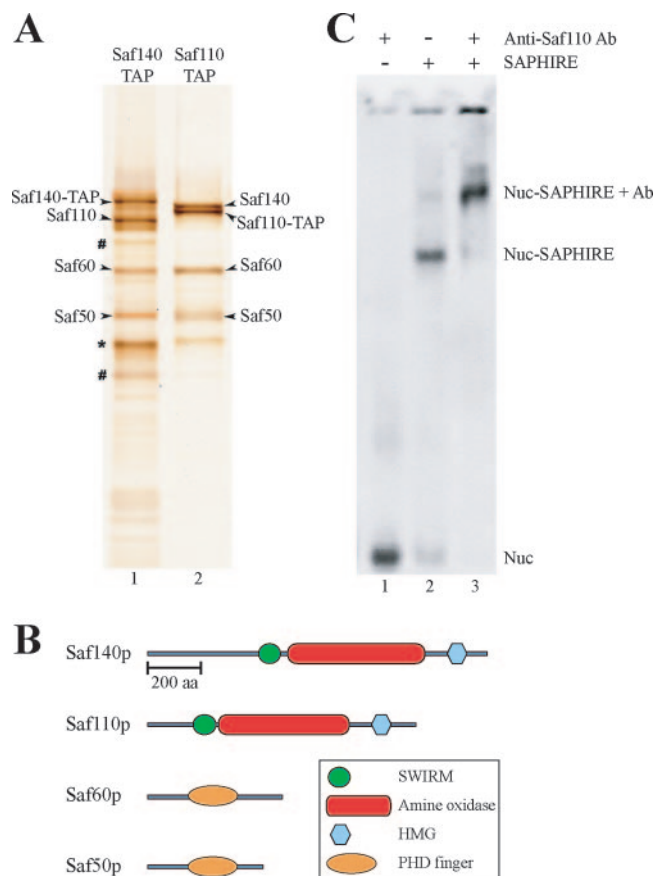
**Genomic data analysis.** S arrays were hybridized, scanned, and normalized as previously described (25). O arrays (Agilent Technologies) were hybridized, scanned, and normalized according to standard procedures. Spots flagged by the scanner as below background or saturated were removed from analysis. The data shown represent the averages for multiple biological replicates. Importantly, in order to properly interpret H3K4me2 and H3K9me2 ChIP data, signals from modification-specific H3 ChIPs were normalized to genome-wide histone H3

occupancy (determined by multiple analyses using an anti-H3 C-terminal antibody). For analysis of SAPHIRE promoter occupancy-gene expression correlations, promoters were defined as the region from  $-800$  to  $+200$  relative to the ATG of all genes. SAPHIRE ChIP ratios were averaged for all tiles falling within this promoter region, and the resulting value was used to represent SAPHIRE promoter occupancy for the downstream gene. Expression values were derived from one of two sources, either raw fluorescence intensity values from microarray data for a wild-type strain or by averaging the Pol II ChIP ratios for every tile falling within an annotated open reading frame (ORF). In order to reveal genome-wide trends, transcription (Pol II ChIP), expression (microarray), and histone methylation (H3K4me2 ChIP) data were plotted as moving averages against SAPHIRE occupancy for various data sets. In all cases, moving averages were calculated using a moving window that constituted 5% of the total data set, with a step size of 1. For BLAST analysis of subtelomeric regions, the DNA sequence spanning coordinates 5522385 to 5579828 (57 kb) of chromosome 1 was split into 60-bp segments spaced 55 bp apart (5 bp overlap on either side). Each of these segments was used to query the entire *S. pombe* genomic sequence (obtained from the Wellcome Trust Sanger Institute at [www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/)) with the BLAST algorithm available from the NIH-NCBI website ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Segments were considered to be duplicated if they generated a secondary BLAST hit that showed  $\geq 90\%$  sequence identity with the query sequence. Finally, linear regression and chi-square analyses were done using standard statistical methods.

## RESULTS

**Purification of SAPHIRE complex.** We recognized two proteins in *S. pombe* that contained three domains of interest, namely, a SWIRM domain, an FAD-dependent amine oxidase domain, and a high-mobility-group B [HMG(B)] domain. Since the SWIRM and HMG(B) domains are closely associated with chromatin function, we initially pursued these proteins as potential new chromatin enzymes that could be studied in a genetically tractable organism. The amine oxidase domains of these proteins contain the key catalytic residues present in known amine dealkylases, suggesting that these *S. pombe* proteins have the capacity for amine demethylation. The two *S. pombe* paralogs are highly similar (BLAST E value,  $2 \times 10^{-84}$ ) and share the same domain order. The human protein Lsd1 is highly similar (BLAST E value,  $4 \times 10^{-14}$ ) and shares key catalytic residues, the SWIRM domain, and the same domain order with the *S. pombe* proteins.

To characterize the biochemical properties of these *S. pombe* NAO paralogs and to identify associated proteins, we purified each to homogeneity. To facilitate their purification, we integrated a DNA sequence into the 3' end of each gene (separately, at the chromosomal locus) to encode an in-frame fusion of the TAP epitope. Integrations were performed separately in diploids, with sporulation and dissection generating tagged haploid progeny that grew as well as their untagged siblings, demonstrating that the tags do not impair function. Thus, the two strains used for purification contained one tagged NAO and one untagged NAO. Purification of either NAO to homogeneity yielded the same four proteins, as identified by mass spectrometry: these were the tagged NAO, the untagged NAO, and two additional proteins, of 60 and 50 kDa (Fig. 1A). Mass spectrometric analysis also showed that the additional bands in these lanes are proteolytic products of these four proteins, and thus both purifications yielded the same four-protein complex. We call the complex SAPHIRE, for a SWIRM-amine oxidase and plant homeodomain (PHD) protein complex involved in regulating gene expression (shown below), with members designated by their apparent molecular weights, i.e., Saf140p, Saf110p, Saf60p,



**FIG. 1.** Purification of SAPHIRE, a nucleosome-binding complex. (A) TAP of *S. pombe* NAO proteins. Only two NAO proteins are encoded by the *S. pombe* genome. TAP of either protein yielded the same complex, with subunits designated Saf140p, Saf110p, Saf60p, and Saf50p. Bands marked with pound signs were proteolytic products of Saf140p or Saf60p. Bands marked with asterisks appeared in a mock purification from a wild-type strain. (B) Domain architecture of SAPHIRE components. The Pfam domain structure for each SAPHIRE component was obtained from the Wellcome Trust Sanger Institute *S. pombe* gene database website ([www.genedb.org/genedb/pombe/](http://www.genedb.org/genedb/pombe/)). Saf140p and Saf110p each contain the following domains: SWIRM (*Swi3*, *Rsc8*, and *Moirai*) (green), amine oxidase (red), and HMG (blue) domains. Saf60p and Saf50p paralogs each contain a single PHD (orange). Protein sizes and domain architecture are drawn to scale. (C) SAPHIRE binds nucleosomes with a high affinity. Gel shift analysis was performed with recombinant mononucleosomes containing the 172-bp nucleosome-positioning sequence from *Xenopus* 5S DNA (lanes 1 to 3). Nucleosomes were incubated with purified SAPHIRE (lanes 2 and 3) as described in Materials and Methods. To determine whether the shifted species (lane 2) was a complex between the nucleosome (Nuc) and SAPHIRE, anti-Saf110 antibody was added to the incubation, providing a supershift (lane 3).

and Saf50p. The systematic gene names are as follows: *saf140*<sup>+</sup>, SPAC23E2.02; *saf110*<sup>+</sup>, SPBC146.09c; *saf60*<sup>+</sup>, SPAC30D11.08c, and *saf50*<sup>+</sup>, SPCC4G3.07c. Thus, the two NAO paralogs are both present in the isolated complex. Interestingly, the additional 60-kDa and 50-kDa proteins are themselves paralogs (BLAST E value,  $7 \times 10^{-44}$ ), and both contain a single PHD (Fig. 1B). Gel filtration of the entire complex provided an apparent mass of  $\sim 350$  kDa, suggesting a simple tetramer (data not shown).

**SAPHIRE binds nucleosomes.** Since SAPHIRE contains multiple domains for chromatin association, we used electrophoretic mobility shift assays to assess the interaction with mononucleosomes. We found that SAPHIRE binds to mononucleosomes with a  $K_d$  of about 8 nM (Fig. 1C and data not shown). Specificity for this interaction was confirmed with a polyclonal antibody raised against Saf110p, which supershifted the nucleosome-SAPHIRE complex (Fig. 1C, lane 3).

To date, the only Lsd1-related protein that has demonstrated histone lysine demethylation activity *in vitro* is Lsd1 itself. Extensive biochemical testing of both native and recombinant SAPHIRE complexes has not revealed appreciable activity. This is consistent with the work of others on these *S. pombe* orthologs, showing no observed activity *in vitro* (31). Therefore, we focused on the biological and genomic impact of removing the conserved catalytic residue of these enzymes, as described below.

**SAPHIRE is essential for viability.** To determine the importance of SAPHIRE, we constructed gene deletion mutations in each subunit in diploids. We then induced sporulation of the heterozygous diploids and examined the phenotypes of haploid progeny. We found that *saf140<sup>+</sup>*, *saf60<sup>+</sup>*, and *saf50<sup>+</sup>* are essential genes, as viability segregated 2:2 in four-spore tetrads (data not shown). Somewhat surprisingly, *saf110<sup>+</sup>/saf110 $\Delta$*  diploids produced four viable spores, with the *saf110 $\Delta$*  progeny always displaying extremely slow growth (Fig. 2A). Since 2:2 segregation of slow growth was observed in a large number of tetrads and was always linked to *saf110 $\Delta$ ::kanMX*, the phenotype was likely not a result of background mutation. Importantly, normal growth could be restored by transformation with a plasmid bearing only *saf110<sup>+</sup>*, confirming the null phenotype (data not shown). In addition to having slow growth, *saf110 $\Delta$*  strains are temperature sensitive (Fig. 2A).

To test whether *saf110<sup>+</sup>* and *saf140<sup>+</sup>* are partially redundant, we overexpressed *saf110<sup>+</sup>* or *saf140<sup>+</sup>* (using high-copy plasmids) in heterozygous diploids (*saf110<sup>+</sup>/saf110 $\Delta$*  or *saf140<sup>+</sup>/saf140 $\Delta$* ) and tested whether we could acquire haploid segregants where the high-copy plasmid conferred growth ability to the reciprocal null segregants. We found that a high-copy *saf110<sup>+</sup>* plasmid will not suppress the inviability of a *saf140 $\Delta$*  strain, nor will a high-copy *saf140<sup>+</sup>* plasmid suppress the extremely slow growth of a *saf110 $\Delta$*  strain. These experiments suggest that the Saf140p and Saf110p proteins make largely unique contributions to SAPHIRE.

**SAPHIRE NAO mutants lacking a conserved catalytic residue are viable.** Saf140p, Saf110p, and Lsd1 share a high degree of similarity to FAD-dependent amine oxidases, such as monoamine oxidase (MAO; dealkylates neurotransmitters) and polyamine oxidase (PAO; dealkylates spermine and spermidine). Although these enzymes have different substrates, they all dealkylate an amine, using flavin reduction to reduce the amine to an imine, followed by imine hydrolysis to oxidize two products; in the case of Lsd1, the products derived from monomethyl-lysine are formaldehyde and lysine. The most extensively studied catalytic residue is a specific lysine. In all crystal structures examined, this lysine positions the lone water molecule in the active site into a bridging position with the flavin N-5 atom of the FAD cofactor (2, 4, 9, 45). This lysine is absolutely conserved in all Lsd1 orthologues, all PAOs, and all MAOs and is the only residue in the catalytic pocket that is

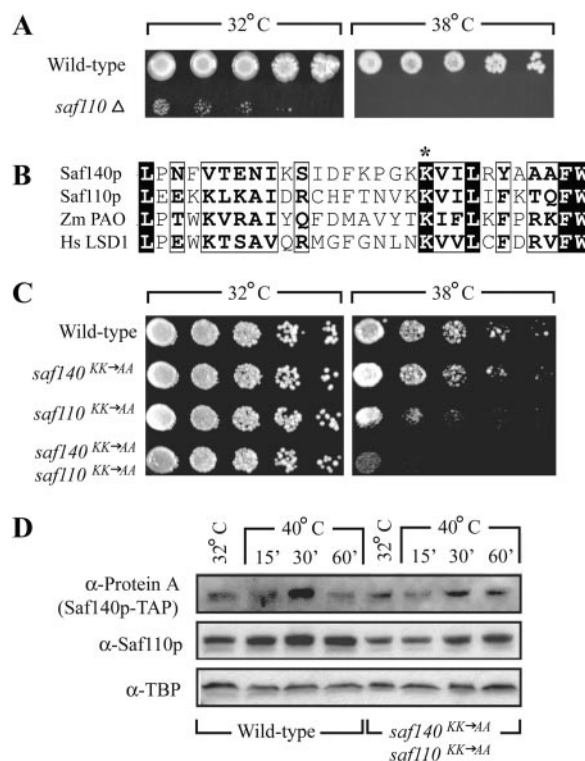


FIG. 2. Characterization of SAPHIRE mutant strains. (A) *saf110 $\Delta$*  strains have slow growth and temperature-sensitive phenotypes. Serial fivefold dilutions of wild-type and *saf110 $\Delta$*  strains were spotted onto rich medium plates containing glucose and incubated at 32°C or 38°C. (B) Conservation of key catalytic residues within SAPHIRE amine oxidase domains. Deduced amino acid sequence data for *Homo sapiens* Lsd1, *S. pombe* Saf140p and Saf110p, and *Zea mays* PAO were aligned. An asterisk highlights the lysine residue that is absolutely conserved in all Lsd1 orthologues, all PAOs, all MAOs, and all sarcosine oxidases. (C) SAPHIRE site-directed mutant strains have temperature-sensitive phenotypes. Serial fivefold dilutions of wild-type, *saf140<sup>KK→AA</sup>* single mutant, *saf110<sup>KK→AA</sup>* single mutant, and *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* double mutant strains were spotted onto rich medium plates containing glucose and incubated at 32°C or 38°C. (D) Temperature-sensitive growth of *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* strains is not due to alterations in protein stability. Western blot analysis was performed on whole-cell extracts from haploid wild-type (*saf140<sup>+</sup>::TAP saf110<sup>+</sup>*) and *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* (*saf140<sup>KK→AA</sup>::TAP saf110<sup>KK→AA</sup>*) strains grown at 32°C or subjected to heat shock at 40°C for the indicated amounts of time.

absolutely conserved in the distantly related sarcosine oxidases. Crystal structures and/or alignments identified the corresponding catalytic lysine as K661 in Lsd1, K300 in maize PAO, K862 in Saf140p, and K604 in Saf110p (Fig. 2B). In all cases tested, including Lsd1 and maize PAO, mutation of this residue alone produced no detectable activity in the resultant protein.

We determined the importance of this lysine on SAPHIRE function by replacing it with alanine. Since Saf140p and Saf110p have a lysine adjacent to this position, we also replaced K861/K603 with alanine, resulting in the double mutations KK861-862AA and KK603-604AA (KK $\rightarrow$ AA), to further ensure an impact of these substitutions on catalytic activity. Strains bearing *saf140<sup>KK→AA</sup>* and *saf110<sup>KK→AA</sup>* were constructed by replacing one of the two wild-type alleles with a

KK→AA mutant allele at the endogenous genomic locus in a diploid, inducing the diploid to sporulate, and then isolating haploid KK→AA progeny. Separately, *saf140*<sup>KK→AA</sup> and *saf110*<sup>KK→AA</sup> mutants were viable, with the *saf110*<sup>KK→AA</sup> mutant displaying a moderate temperature sensitivity. Interestingly, the *saf140*<sup>KK→AA</sup> *saf110*<sup>KK→AA</sup> double mutant was also viable, though displaying pronounced temperature sensitivity (Fig. 2C). These phenotypes could not be attributed to a significant reduction in SAPHIRE stability, as SAPHIRE abundance and integrity were similar in the double mutant at the permissive and nonpermissive temperatures (Fig. 2D), though we have not rigorously quantified the impact on stability by purification of the mutant to homogeneity. No clear phenotype was detected when these strains were tested for growth in the presence of either cycloheximide, EGTA, methyl methanesulfonate, H<sub>2</sub>O<sub>2</sub>, latrunculin A, or sodium dodecyl sulfate. These results suggest that catalytic impairment does not render cells inviable but rather confers conditional growth defects. Thus, SAPHIRE exhibits an important catalytic activity and an essential noncatalytic activity.

**Genome-wide localization of SAPHIRE.** To identify the sites in the genome where SAPHIRE functions, we determined the genome-wide localization of complex members. Genome-wide SAPHIRE occupancy was initially determined with a haploid *S. pombe* strain during asynchronous growth in rich medium containing glucose. To isolate DNA segments associated with SAPHIRE, we performed ChIP using TAP-tagged Saf140p or Saf110p derivatives. ChIP-enriched fragments from TAP-tagged or untagged strains were labeled with fluorescent dyes (Cy5 and Cy3, alternated between replicates) and used to probe two different DNA array formats for the *S. pombe* genome. The first array consisted of the entire genome parsed into two types of segments (~11,000 segments; the S array), i.e., ORFs and intergenic regions (IGRs), which were used to create spots on separate glass slides. The initial results derived from this array (described below) motivated questions that required the use of a higher-density array, which consisted of 60-mer single-stranded oligonucleotide tiles (~43,000 oligonucleotides; the O array) that tiled the entire genome at a resolution of ~250 bp on a single glass slide. For each array, the normalized Cy5/Cy3 ratio (tagged/untagged) provided a measurement of SAPHIRE occupancy at each position.

SAPHIRE occupancy determinations were reproducible and specific. For example, with the S array, three biological replicates of Saf140p-TAP and (separately) Saf110p-TAP yielded Pearson correlation coefficients ( $r$ ) of  $\geq 0.97$  (data not shown). Furthermore, Saf140p-TAP and Saf110p-TAP appear to occupy the same DNA segments, as a comparison of their data sets yielded an  $r^2$  value of 0.91 (Fig. 3A). For the S array, we defined segments that displayed enrichment above the 95.5th percentile as SAPHIRE occupied, which yielded a total of 119 occupied loci. For the O array, the higher density required alternative criteria to define occupancy: a tile was considered occupied if the tile (i) displayed a signal intensity well above background, (ii) displayed an average enrichment of at least threefold (tagged/untagged) over multiple biological replicates, (iii) displayed an enrichment of at least twofold in a majority of individual replicates, and (iv) was adjacent to a tile that displayed an average enrichment of at least twofold. This stringent set of criteria was designed to ensure a very small

number of false-positive results. By these criteria, 426 separate loci are occupied by SAPHIRE genome-wide. As expected, the targets identified by the S array are a subset of those identified by the better-resolved O array.

**SAPHIRE occupancy at chromosomal loci.** In general, we found SAPHIRE distributed widely throughout the genome and dispersed at genes over all three chromosome arms. We also observed SAPHIRE at several regions associated with heterochromatin. For example, we observed SAPHIRE in the region of the silent mating (*MAT*) locus and in peaks in the subtelomeric regions of chromosomes 1 and 2 (presented below). In addition, SAPHIRE occupies the junctions between the central region of the centromeres and the dg/dh repeats (data not shown). Notably, these regions also bear clusters of tRNAs which bind TFIIC, and both tRNAs and TFIIC sites themselves have demonstrated chromatin boundary functions (32, 41). However, their relationship to SAPHIRE has not been explored. Notably deficient in SAPHIRE are the rRNA gene repeats at the distal tips of chromosome 3.

**SAPHIRE occupies promoters.** Since the majority class of SAPHIRE targets comprises genes, we examined whether SAPHIRE prefers ORFs or IGRs. In contrast to the S array, the O array included all ORF and IGR tiles on the same slide, allowing a direct comparison of these two classes. On the O array, tiles representing ORFs and IGRs comprised 65% and 35% of the total tiles, respectively. However, if we examined only SAPHIRE-occupied tiles, the distribution changed to 18% ORFs and 82% IGRs (which is highly statistically significant [ $P < 0.0001$ ]). Also, nearly all ORF tiles in the SAPHIRE-occupied class clustered at the extreme 5' and 3' ends of the ORFs, indicating that this occupancy was likely bleeding from IGR occupancy due to limitations in shearing resolution (data not shown). Thus, the vast majority of SAPHIRE occupancy is in IGRs.

We next determined the extents to which SAPHIRE occupies the three different IGR classes, namely, (i) single promoters, which are flanked by one ORF 5' end and one ORF 3' end; (ii) double promoters, which are flanked by two ORF 5' ends; and (iii) nonpromoters, which are flanked by two ORF 3' ends. In eukaryotic genomes, these three classes partition as would be predicted by random orientation relationships, i.e., 25%, 50%, and 25% (double, single, and nonpromoters, respectively). However, note that in *S. pombe*, double promoters are longer and nonpromoters are shorter, on average, than single promoters. Since the O array had a relatively fixed number of tiles per length of DNA, more tiles were present in double promoters and fewer were present in nonpromoters. Thus, the O array tile ratio was 39% to 47% to 14% for the genome as a whole (double, single, and nonpromoters, respectively). However, we found that the partitioning of SAPHIRE-occupied loci was 46%, 46%, and 8%, and this change in distribution favoring promoters was highly significant (Fig. 3B). Thus, SAPHIRE primarily occupies promoters.

**SAPHIRE occupancy is positively correlated with gene activity.** We next examined whether SAPHIRE localizes to active or inactive genes genome-wide. Here we compared SAPHIRE promoter occupancy with transcription of the downstream ORF. Promoters were defined as ~1-kb regions spanning positions -800 to +200 relative to the translation initiation codon, as our array resolution is ~250 bp. To ensure the

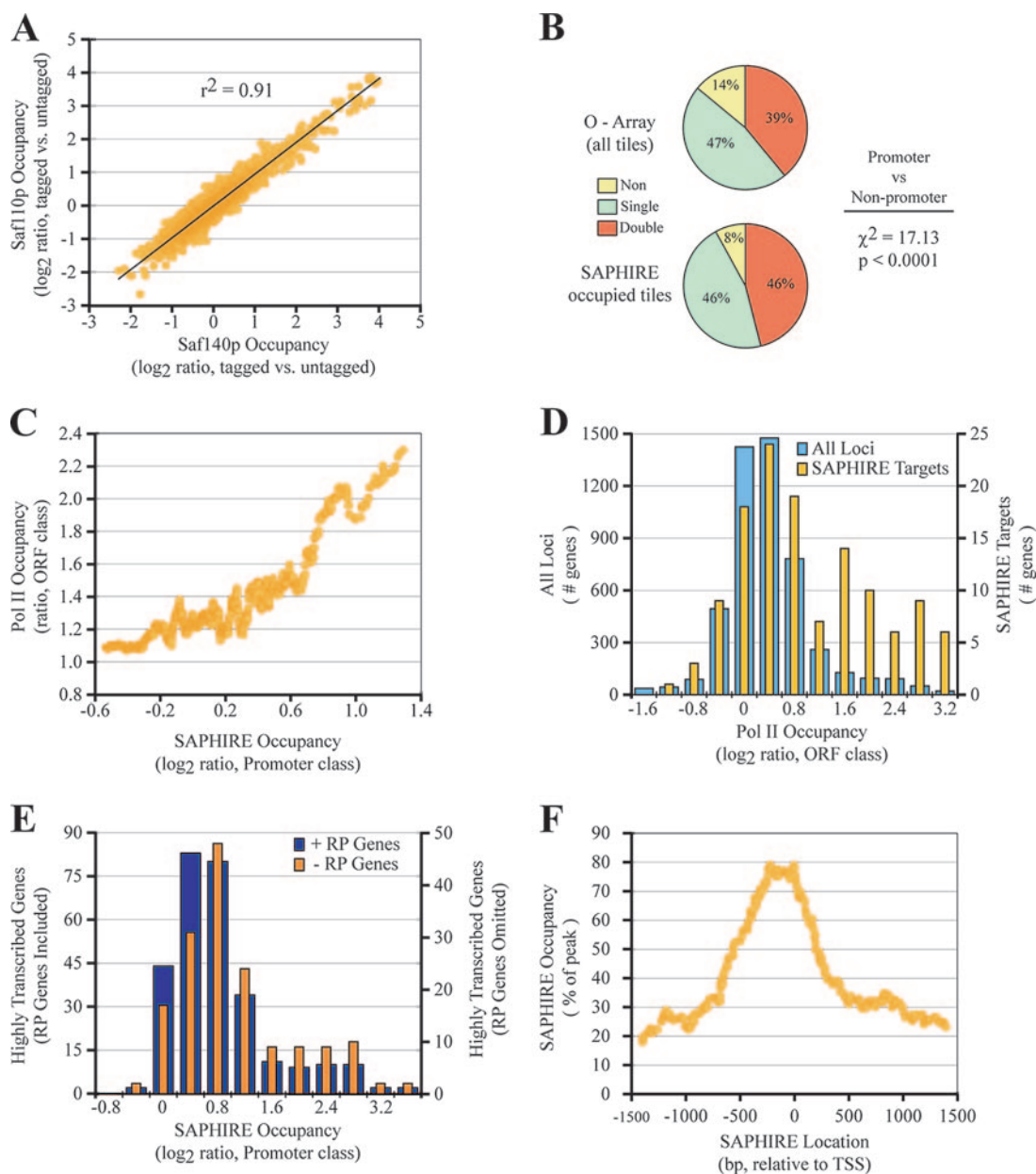


FIG. 3. Genome-wide localization of SAPHIRE. (A) Saf140p and Saf110p colocalize in the genome. The scatter plot depicts each unique tile on the S array as a point on the graph. The  $x$  and  $y$  coordinates of each point are equal to the  $\log_2$  ratio of Saf140p and Saf110p occupancy at the corresponding S array spot. The coefficient of determination ( $r^2$ ) is provided. (B) SAPHIRE-occupied tiles cluster within promoter-containing intergenic regions. Two pie graphs depict the distributions of all O array versus SAPHIRE-occupied intergenic tiles among three different classes of intergenic regions (double, single, or nonpromoters). Chi-square analysis indicates that the difference in distribution is highly significant. (C) SAPHIRE promoter occupancy scales with the transcription rate. Transcription rates of Pol II genes (as measured by the average Pol II ORF occupancy) were plotted against SAPHIRE occupancy of the upstream promoters. Data were plotted as moving averages (see Materials and Methods). (D) SAPHIRE targets are highly transcribed. Two histograms were overlaid, depicting the expression levels of all Pol II-transcribed ORFs (wide blue bars) versus the expression levels of ORFs whose promoters harbored SAPHIRE with occupancy values of  $\geq 1.5$  ( $\log_2$  ratio) (narrow yellow bars). Note the large right shift of the yellow relative to the blue histogram. (E) Not all highly transcribed genes are SAPHIRE targets. Highly transcribed genes were defined as those with an average Pol II ORF occupancy of  $\geq 1.5$  ( $\log_2$  ratio). Two histograms were overlaid, depicting the SAPHIRE promoter occupancy of highly transcribed genes, including (wide dark blue bars) or excluding (narrow orange bars) ribosomal protein genes. Note that many highly expressed genes have low SAPHIRE promoter occupancy, but removal of the coordinately regulated ribosomal protein gene class reveals that a significant fraction of the remaining highly transcribed genes are SAPHIRE targets. (F) SAPHIRE localizes to the TSS of target genes. SAPHIRE occupancy, expressed as a percentage of the peak occupancy for each promoter, was aligned relative to the TSS. The aligned SAPHIRE occupancy values for all genes were plotted versus relative genomic locations as moving averages.

proper attribution of SAPHIRE to a target gene, only single promoters were included in this analysis. SAPHIRE ChIP ratios for each tile were averaged for these single promoters and collated with their respective downstream genes for occupancy-expression correlations.

We then compared SAPHIRE occupancy to two different measures of gene activity, namely, steady-state mRNA levels and Pol II occupancy. For mRNA levels, expression profiling by microarray analysis was performed under the identical conditions utilized for SAPHIRE occupancy determinations, and the relative normalized fluorescence for the wild-type channel was used to quantify the relative amount of each transcript. Our unpublished data show that this measure is a generally accurate representation of transcript abundance. However, differences in hybridization efficiency can also affect the relative fluorescence of some transcripts with equal representation in the RNA pool. Therefore, we determined the genome-wide occupancy of RNA Pol II and used the average Pol II occupancy over the entire ORF of each gene as an alternative measure of the transcription rate. We found that SAPHIRE occupancy is strongly correlated with transcription when either Pol II occupancy or mRNA abundance is examined (Fig. 3C and data not shown).

The data presented in Fig. 3C are useful for examining activity-occupancy relationships, but as they represent moving averages for many genes, they do not reveal how uniformly individual genes conform to this correlation. To examine this, we compared the distribution of SAPHIRE-occupied loci versus that of all loci to Pol II occupancy. Note that because only ORF class Pol II was examined (intergenic regions were omitted), the distribution curve does not center around zero Pol II occupancy (Fig. 3D). We found that most SAPHIRE targets are highly transcribed, as opposed to the small percentage of the total genome with a high transcription rate. Also, although a small number of SAPHIRE-occupied genes display somewhat less than the average amount of Pol II, we did not find SAPHIRE at any highly repressed promoters. Furthermore, we found that SAPHIRE targets represent only a fraction of all highly expressed genes (defined as genes with average Pol II ORF enrichment of  $\geq 2.8$ -fold [ $\log_2$  value,  $\geq 1.5$ ]) (Fig. 3E). One particularly large class which is notably deficient in SAPHIRE is the ribosomal protein genes, the most transcriptionally active Pol II gene class in the genome. If we removed this coordinately regulated class of genes from the analysis, SAPHIRE occupied a very significant fraction of the remaining highly transcribed genes (Fig. 3E). Intriguingly, the ribosomal protein genes share a *cis* promoter sequence specific to *S. pombe*, i.e., a homology D box, that is absent from most SAPHIRE targets. Thus, promoter architecture differences may determine whether an active gene recruits SAPHIRE.

Our examination of the Pol II genome-wide occupancy allowed us to approximate the transcription start sites (TSS) of highly transcribed genes to within  $\sim 125$  bp. We defined the TSS as the location of the first tile upstream of an expressed gene with  $>2$ -fold occupancy. Then, selecting SAPHIRE-occupied genes with known TSS (45 targets), we compared the peak of SAPHIRE occupancy to the TSS. We found that SAPHIRE localizes at or just upstream of the TSS and falls sharply at locations following the TSS (Fig. 3F).

**Dynamic relocalization of SAPHIRE during heat shock.** To determine whether SAPHIRE is dynamically recruited to genes undergoing activation, we examined SAPHIRE occupancy in cells subjected to 15 min of heat shock, a treatment known to impact the transcriptional activity of hundreds of genes. Remarkably, SAPHIRE was redistributed to genes undergoing activation at a magnitude proportional to their activation (Fig. 4A). In this case, heat shock genes (such as *hsp16<sup>+</sup>*) that are bound by heat shock transcription factor (HSTF) were among the acquired targets. However, new targets also included many genes that are not believed to be bound by HSTF, suggesting that several types of promoters recruit SAPHIRE when they are activated. As a test for HSTF recruitment, we placed ectopic HSTF sites on a plasmid but did not observe heat shock-dependent recruitment of SAPHIRE, suggesting that recruitment may require both enhancer elements and promoter elements. As observed at 32°C, SAPHIRE occupied many, but not all, of the genes highly activated by heat shock, suggesting that SAPHIRE targeting is selective (data not shown). Taken together, these results demonstrate that SAPHIRE is a dynamic complex that is recruited to the region near the TSS of particular genes during the activation process.

**Target genes require SAPHIRE for full transcriptional activation.** We then took a genome-wide approach to determine whether SAPHIRE is required for the activation of target genes. If SAPHIRE assists in activation, then target genes should be selectively attenuated in *saphire* mutants. To test this, we performed mRNA expression profiling (microarray analysis) of wild-type, *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>*, and *saf110Δ* strains. Genes occupied by SAPHIRE displayed reduced transcription in the mutants, whereas unoccupied genes remained largely unaffected. The impact of the *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* double mutation was moderate, whereas that of the *saf110Δ* mutation was much more pronounced (Fig. 4B). Although the *saf110Δ* mutation confers very slow growth, the observation of a selective impact on the transcription of SAPHIRE targets strongly supports the notion that the primary effect of the *saf110Δ* mutation is a reduction in the activation of SAPHIRE targets. Furthermore, although the impact of the *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* double mutation is modest, it suggests that SAPHIRE catalytic activity also makes a contribution to target gene activation. Taken together, these results strongly suggest that SAPHIRE-occupied genes rely on SAPHIRE for their full activation and that contributions are made by both the catalytic function and the noncatalytic function.

**Catalytic impairment increases H3K4 methylation levels at SAPHIRE targets.** Since Lsd1 is a known H3 demethylase, we correlated H3 methylation levels with SAPHIRE occupancy genome-wide. To this end, we determined the genome-wide patterns of H3K4me2 and H3K9me2. To properly interpret H3K4me2 and H3K9me2 ChIP experiments, we normalized all data to H3 occupancy, determined by multiple independent genome-wide ChIP analyses (using a C-terminal anti-H3 antibody). Genome-wide examination revealed an extremely weak correlation with H3K9 (data not shown), which was due to SAPHIRE occupancy at or adjacent to particular heterochromatic loci. However, we observed that the levels of H3K9me2 at promoters and ORFs were extremely low compared to the levels at heterochromatic loci ( $<1\%$ ) or in comparison to

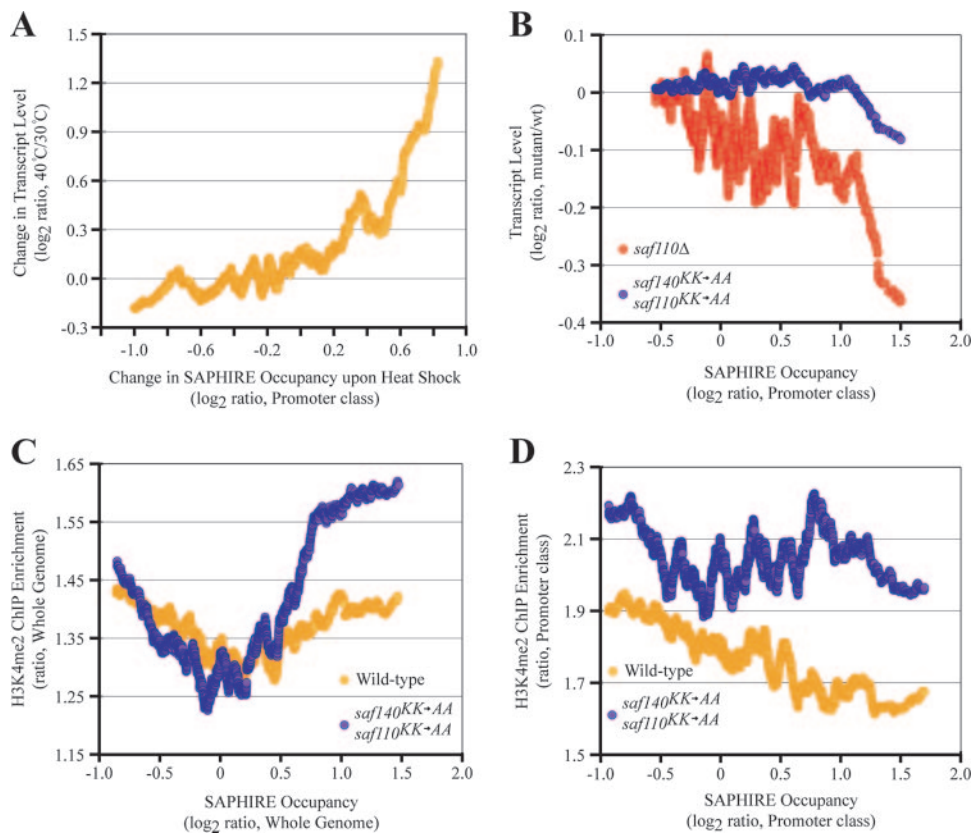


FIG. 4. SAPHIRE regulation of target gene transcript levels is associated with changes in histone methylation. (A) During heat shock, SAPHIRE dynamically relocates to new sites of active transcription. The  $\log_2$  ratio of the change in transcript level was plotted against the  $\log_2$  ratio of the change in SAPHIRE promoter occupancy for all ORFs after 15 min of heat shock (40°C). The data are plotted as moving averages. (B) SAPHIRE targets depend on SAPHIRE for their transcriptional activation. Genome-wide microarray ratios ( $\log_2$ ) comparing gene expression in *saf110* $\Delta$  versus wild-type strains (red circles) and *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> versus wild-type strains (purple circles) were plotted as moving averages against SAPHIRE promoter occupancy for the corresponding genes. Note that SAPHIRE target genes (right edge of the graph) tend to be expressed at lower levels (downward trend) in SAPHIRE mutant strains. (C and D) SAPHIRE targets gain H3K4me2 in *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> mutant strains. H3-normalized H3K4me2 levels (as determined by genome-wide ChIP-microarray analysis) for wild-type (gold circles) and *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> (purple circles) strains were plotted against SAPHIRE occupancy ( $\log_2$  ratio) as moving averages. Data sets including all tiles (C) or promoter tiles only (D) are shown. In panel C, SAPHIRE-occupied tiles (right edge of the graph) have increased methylation levels (upward trend) in *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> versus wild-type strains. In panel D, SAPHIRE-occupied promoter tiles tend to have lower H3K4me2 levels in wild-type strains (downward trend). However, in *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> mutants, this downward trend is abolished.

H3K4me2 levels. Therefore, we focused our analysis on the impact on H3K4 methylation. For H3K4 methylation, we focused on the trends observed at loci that contain SAPHIRE (i.e., display a  $\log_2$  occupancy ratio of  $\geq 0$ ). In this analysis, we observed a modest positive correlation of SAPHIRE occupancy with H3K4me2 genome-wide (Fig. 4C) but a modest negative correlation at promoters (Fig. 4D). This raises the possibility that SAPHIRE interacts with H3K4me2 regions but promotes H3K4 demethylation selectively at promoters. To address this possibility, we examined H3K4me2 genome-wide in wild-type cells and in our mutants bearing substitutions in the conserved catalytic residues. Interestingly, H3K4me2 increased at SAPHIRE targets in the *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> double mutant, consistent with impairment of an H3K4 demethylation function (Fig. 4C and D). Taken together, these results suggest that misregulation of H3K4 demethylation at promoters is correlated with reduced transcriptional activation (see Discussion).

***saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> mutations cause spreading of telomeric heterochromatin.** At present, little is known about the factors which restrict the spreading of telomeric heterochromatin in *S. pombe*. The telomere-proximal regions of all three *S. pombe* chromosomes bear repetitive segments with high levels of H3K9 methylation, which decrease in telomere-distal regions as unique segments are encountered. Although H3K9me2 was not significantly affected at SAPHIRE targets genome-wide (data not shown), perturbations were observed at particular loci. The clearest example, by far, is a major change in the heterochromatin-euchromatin boundary of the right telomere of chromosome 1 in the *saf140*<sup>KK-AA</sup> *saf110*<sup>KK-AA</sup> double mutant (Fig. 5). In this double mutant, H3K9me2 spread  $\sim 30$  kb past its normal boundary (Fig. 5B). We recognized that certain telomeric repeat elements (and certain genes and pseudo-genes) in *S. pombe* are also located on other telomeres, which can prevent the attribution of ChIP enrichment to a particular locus. However, the affected region mapped almost exclusively

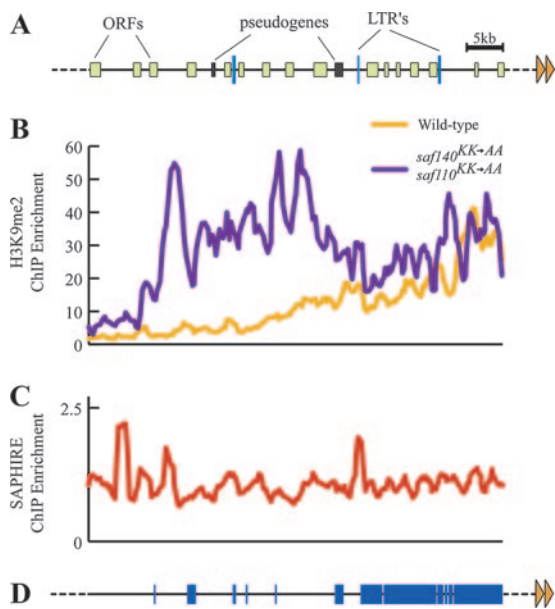


FIG. 5. SAPHIRE regulates telomeric heterochromatin. (A) Physical map of the right subtelomeric region of chromosome 1. The map corresponds to coordinates 5522385 to 5579828 in the Sanger Institute chromosome 1 virtual contig. A scale drawing of genetic elements of interest is depicted, including ORFs (light green boxes), pseudogenes (gray boxes), and Tf2-type long terminal repeats (thick blue lines). Telomeric repeats (not included within the virtual contig) are indicated by orange arrowheads. (B) Spreading of telomeric heterochromatin in *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* mutant strains. H3-normalized H3K9me2 levels (as determined by genome-wide ChIP-microarray analysis) for wild-type (gold line) and *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* (purple line) strains were aligned to the physical map (A) and subjected to a five-tile moving average to reduce noise. Note that telomeric heterochromatin extends ~30 kb into the chromosome arm in *saf140<sup>KK→AA</sup> saf110<sup>KK→AA</sup>* strains. (C) SAPHIRE localizes to telomeric heterochromatin boundaries. SAPHIRE occupancy (as determined by genome-wide ChIP-microarray analysis) was aligned and plotted as described for panel B. Note that a distinct peak of SAPHIRE occupancy coincides with the wild-type heterochromatin-euchromatin boundary. (D) Map of unique versus repeated subtelomeric regions. The entire right subtelomeric region of chromosome 1 was analyzed using the BLAST algorithm to distinguish unique regions (thin black lines) from duplicated regions (thick blue lines) with significant homology ( $\geq 90\%$  sequence identity) to other parts of the genome (see Materials and Methods). The results were aligned onto the physical map (A).

with unique O array tiles, showing that the spreading observed can be attributed unambiguously to the right subtelomeric region of chromosome 1 (Fig. 5D). Interestingly, examination of this subtelomeric region reveals a clear peak of SAPHIRE that coincides with the location where H3K9me2 begins to spread in the mutant (Fig. 5C). This peak of SAPHIRE occupancy includes three consecutive tiles with  $\geq 2$ -fold occupancy that are unique in the genome. This raises the interesting possibility that the catalytic activity of SAPHIRE helps to promote boundary activity at this telomere.

## DISCUSSION

In this study, a search for new chromatin-modifying enzymes in *S. pombe* yielded a four-protein complex called SAPHIRE, which binds nucleosomes and resides at the promoters of par-

ticular active genes. SAPHIRE bears two subunits, Saf140p and Saf110p, which are highly similar to the human histone demethylase Lsd1. Our genetic data identify SAPHIRE as an essential complex, providing evidence that Lsd1 orthologues are essential for viability. Our *in vivo* work also provides several lines of evidence that SAPHIRE promotes gene activation. While we were preparing this paper for submission, others purified a complex of largely the same composition, localized the complex to promoters, and demonstrated that transcription is reduced at three selected targets in *saf110 $\Delta$*  mutants (31). However, our work is unique in its (i) evidence for nucleosome binding, (ii) examination of knockouts of all four complex members, (iii) preparation and analysis of catalytically impaired mutants, (iv) observation of occupancy of the complex at heterochromatin regions, (v) genome-wide analysis of the impact of both *saf110 $\Delta$*  and catalytic mutants on transcription, (vi) observation of changes in histone methylation that occur in catalytic mutant strains at occupied genes, (vii) observation of dynamic relocalization of SAPHIRE during gene activation, and (viii) demonstration of a large shift in the euchromatin-heterochromatin boundary at chromosome 1 in catalytic mutants, suggesting a novel role in boundary formation. We discuss aspects of these features and the insights gained below.

**SAPHIRE is an *S. pombe* complex that resembles the BHC complex in metazoans.** Gel filtration of purified SAPHIRE as well as recombinant versions of the complex (data not shown) suggested that SAPHIRE consists of a heterodimer of Saf140p-Saf110p bound to a heterodimer of Saf60p-Saf50p. One curiosity is that while *saf140<sup>+</sup>*, *saf60<sup>+</sup>*, and *saf50<sup>+</sup>* are essential genes, the *saf110 $\Delta$*  strain is viable, though extremely slow growing. One possibility is that the complex can partially form and function without Saf110p. Another possibility is that Saf140p can homodimerize at a low frequency, with the homodimer taking the place of the Saf140p-Saf110p heterodimer in the complex.

Lsd1 is purified in a multiprotein complex named BHC (11). Comparison of BHC and SAPHIRE reveals that SAPHIRE bears proteins with domain architectures similar to those of a subset of BHC subunits. For example, BHC80 is purified with Lsd1 and regulates demethylation by Lsd1 *in vitro* (44). BHC80 bears a single PHD, analogous to the proteins Saf50p and Saf60p. Also, although Lsd1 does not have an HMG domain (present on Saf140p and Saf110p), the BHC complex contains the HMG domain protein BRAF35 (11). Thus, SAPHIRE may constitute a conserved core demethylase-related complex.

**SAPHIRE interaction with nucleosomes.** SAPHIRE binds nucleosomes with a fairly high affinity ( $K_d$ , ~8 nM). However, our interaction experiments were conducted with recombinant nucleosomes which lacked modifications of their N-terminal tails. Notably, Saf60p and Saf50p both contain a PHD, a motif that was recently shown to bind directly to methylated lysine residues, including H3K4me2 (21, 33, 42, 52). Furthermore, we detected a moderate correlation of SAPHIRE with this mark genome-wide. Thus, a clear line of investigation is to test whether H3K4 methylation enhances SAPHIRE interaction with nucleosomes due to its recognition by the PHD.

**SAPHIRE promotes transcription.** Our work provides several lines of evidence that SAPHIRE promotes transcription.

First, SAPHIRE occupancy is well correlated with transcriptional activity. However, SAPHIRE is present at only a few hundred genes, and many genes with high levels of transcription are not occupied by SAPHIRE. This suggests that SAPHIRE recruitment and action are selective. At present, factors or DNA sites that account for this selectivity have not been obtained but represent an important area of future work. Interestingly, SAPHIRE localization is dynamic, as SAPHIRE is recruited to promoters during activation. This suggests that SAPHIRE is not a factor that assembles at repressed/basal genes to promote subsequent activation (as observed with the histone variant Htz1) but, rather, is dynamically recruited at the time of activation.

We impaired Saf140p and Saf110p catalytic activities in combination and observed conditional growth phenotypes but not the inviability observed with null alleles. Furthermore, *saf110*<sup>+</sup> deletion had a strong and selective impact on SAPHIRE target transcription, whereas the effect with the impaired alleles was modest. This suggests that Saf140p and Saf110p have very important nonenzymatic roles in transcriptional activation. One possibility is that SAPHIRE components recruit proteins that are important for full transcriptional activation. Another possibility is that SAPHIRE binding to chromatin helps to establish a chromatin state that promotes transcription.

Interestingly, the combined catalytic mutants conferred two effects on histone methylation at SAPHIRE targets, namely, a clear increase in H3K4me2 levels (Fig. 4C and D) and a very slight increase in H3K9me2 levels (data not shown). First, we address the impact on H3K9me. H3K9 methylation is generally correlated with gene repression, and thus removing this mark might promote activation. This is in keeping with one proposed transcriptional activation function for human Lsd1, where Lsd1-mediated H3K9 demethylation contributes to AR target gene activation (27). However, our genome-wide experiments indicated that H3K9me2 is extremely low or absent from virtually all promoters, making H3K9 methylation an unlikely player in general gene activation in *S. pombe*, and therefore H3K9 demethylation by SAPHIRE is an unlikely regulatory mechanism. Although we did observe a slight increase in H3K9 methylation in our catalytic mutants, this acquired H3K9 methylation signal remained <4% of the levels observed at pericentric heterochromatin or the silent mating loci. In contrast, H3K4 dimethylation is generally correlated with gene activity and/or a chromatin environment permissive for activation. However, in our catalytic mutants, the acquisition of H3K4 methylation was considerable and was correlated with downregulation of SAPHIRE target transcription genome-wide. This result raises the intriguing possibility that H3K4 demethylation can help to activate certain genes. In support of this notion, H3K4 methylation was recently shown to participate in a novel histone deacetylase-dependent gene repression pathway (42), and loss of H3K4 methylation has been correlated with transcriptional activation at certain loci (8, 16, 22, 26). Indeed, since H3K4 methylation is highly regulated both spatially and temporally at transcribed genes, changes in H3K4me2 might have opposite effects at distinct times and places along the gene. For example, at a typical gene in *S. cerevisiae*, H3K4me2 peaks in the middle of the ORF, and the peak moves closer to the 3' end as transcription in-

creases. Furthermore, as H3K4me3 peaks at the very 5' end of the ORF, the promoter becomes devoid of H3K4me2. This orchestration of H3K4 methylation states may need to be properly balanced by the opposing activities of histone methyltransferases and demethylases. Our observation of SAPHIRE at activated genes, promoting their activation through (at least in part) H3K4 demethylation, is strongly supported by genome-wide studies of human Lsd1 showing very similar results, which were published while this paper was in revision (8).

One observation that may shed light on the H3K4 methylation issue is that *S. pombe* lacks chromatin remodelers of the ISWI family, a family present in both budding yeasts and vertebrates. Remodelers of the ISWI class are primarily linked to the ordering of nucleosomes following replication and also to the ordering of nucleosomes at the promoters of genes. For example, the ISW1a complex in *S. cerevisiae* is localized to the 5' end of the *MET16* gene, where it helps to promote repression (28). Interestingly, *S. pombe* has a larger number of chromodomain (CHD)- and PHD-containing remodelers. Since CHDs and PHDs interact with methylated lysines, we speculate that *S. pombe* may utilize CHD/PHD remodelers to help organize gene 5' ends (and/or the proximal promoter). We further speculate that demethylation of lysines by SAPHIRE might prevent this repressive remodeling, thus promoting activation.

**SAPHIRE and telomere boundaries.** Our finding that *saf140*<sup>KK→AA</sup> *saf110*<sup>KK→AA</sup> mutations lead to spreading of telomeric heterochromatin on the right arm of chromosome 1 suggests that SAPHIRE may act as a boundary element. One clear question is why only one telomere is affected. Studies of *S. cerevisiae* provide the precedent that mutations in telomere boundary elements do not always affect chromatin at all telomeres (17, 48, 54). SAPHIRE does not strongly occupy all telomeric boundaries, although subtelomeric regions are enriched for SAPHIRE targets in comparison to the genome overall (data not shown). In principle, SAPHIRE might directly demethylate histones at the boundary region on Chr1, helping to prevent the propagation of heterochromatin. Alternatively, a locus of active transcription mediated by SAPHIRE in the subtelomeric region might provide a barrier to heterochromatin spreading. Another possibility, consistent with our speculation above regarding H3K4 methylation, is that impaired SAPHIRE activity leads to increased remodeling by CHD class remodelers. These CHD class remodelers might organize nucleosomes in a manner that both promotes repression and favors heterochromatin propagation. These issues can be addressed in future studies.

In conclusion, we describe SAPHIRE as an essential nucleosome-binding complex that dynamically relocates to active promoters and demonstrates both catalytic and noncatalytic functions that promote gene expression and chromatin boundary function. Future studies on SAPHIRE will address how it is targeted to particular loci, whether SAPHIRE affects the recruitment of CHD/PHD remodelers, how SAPHIRE contributes to the chromatin boundary at Chr1, and whether SAPHIRE functions at other loci, such as centromeres, ORFs, and the *MAT* locus.

## ACKNOWLEDGMENTS

We thank Brian Dalley for ChIP array development and processing and Brett Milash for database development. We thank Robin Allshire for strains and plasmids.

This work was supported by the Howard Hughes Medical Institute (supplies and support of B.R.C. and D.H.), an NIH Developmental Biology training grant (support of M.G.), and the Huntsman Cancer Institute (support of A.P. and genomics resources). B.T.W. was supported by Sanger postdoctoral and Canadian NSERC fellowships. Research in the Bähler laboratory is funded by Cancer Research UK (CUK) under grant no. C9546/A6517.

## REFERENCES

- Aravind, L., and L. M. Iyer. 2002. The SWIRM domain: a conserved module found in chromosomal proteins points to novel chromatin-modifying activities. *Genome Biol.* 3:RESEARCH0039.
- Binda, C., A. Coda, R. Angelini, R. Federico, P. Ascenzi, and A. Mattevi. 1999. A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Struct. Fold Des.* 7:265–276.
- Binda, C., A. Mattevi, and D. E. Edmondson. 2002. Structure-function relationships in flavoenzyme-dependent amine oxidations: a comparison of polyamine oxidase and monoamine oxidase. *J. Biol. Chem.* 277:23973–23976.
- Binda, C., P. Newton-Vinson, F. Hubalek, D. E. Edmondson, and A. Mattevi. 2002. Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nat. Struct. Biol.* 9:22–26.
- Boa, S., C. Coert, and H. G. Patterson. 2003. *Saccharomyces cerevisiae* Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. *Yeast* 20:827–835.
- Briggs, S. D., M. Bryk, B. D. Strahl, W. L. Cheung, J. K. Davie, S. Y. Dent, F. Winston, and C. D. Allis. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* 15:3286–3295.
- Carrozza, M. J., B. Li, L. Florens, T. Sugauma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123:581–592.
- Garcia-Bassets, I., Y. S. Kwon, F. Telese, G. G. Prefontaine, K. R. Hutt, C. S. Cheng, B. G. Ju, K. A. Ohgi, J. Wang, L. Escoubert-Lozach, D. W. Rose, C. K. Glass, X. D. Fu, and M. G. Rosenfeld. 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* 128:505–518.
- Geha, R. M., K. Chen, J. Wouters, F. Ooms, and J. C. Shih. 2002. Analysis of conserved active site residues in monoamine oxidase A and B and their three-dimensional molecular modeling. *J. Biol. Chem.* 277:17209–17216.
- Grewal, S. I. 2000. Transcriptional silencing in fission yeast. *J. Cell Physiol.* 184:311–318.
- Hakimi, M. A., D. A. Bochar, J. Chenoweth, W. S. Lane, G. Mandel, and R. Shiekhattar. 2002. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl. Acad. Sci. USA* 99:7420–7425.
- Jenuwein, T., and C. D. Allis. 2001. Translating the histone code. *Science* 293:1074–1080.
- Joshi, A. A., and K. Struhl. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* 20:971–978.
- Kaplan, C. D., L. Laprade, and F. Winston. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301:1096–1099.
- Keogh, M. C., S. K. Kurdastani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski, and N. J. Krogan. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123:593–605.
- Kim, J., L. Jia, W. D. Tilley, and G. A. Coetzee. 2003. Dynamic methylation of histone H3 at lysine 4 in transcriptional regulation by the androgen receptor. *Nucleic Acids Res.* 31:6741–6747.
- Kimura, A., T. Umehara, and M. Horikoshi. 2002. Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nat. Genet.* 32:370–377.
- Klose, R. J., K. Yamane, Y. Bae, D. Zhang, H. Erdjument-Bromage, P. Tempst, J. Wong, and Y. Zhang. 2006. The transcriptional repressor JHD3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 442:312–316.
- Krogan, N. J., J. Dover, S. Khorrani, J. F. Greenblatt, J. Schneider, M. Johnston, and A. Shilatifard. 2002. COMPASS, a histone H3 (lysine 4) methyltransferase required for telomeric silencing of gene expression. *J. Biol. Chem.* 277:10753–10755.
- Lee, M. G., C. Wynder, N. Cooch, and R. Shiekhattar. 2005. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437:432–435.
- Li, H., S. Ilin, W. Wang, E. M. Duncan, J. Wysocka, C. D. Allis, and D. J. Patel. 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442:91–95.
- Li, J., Q. Lin, H. G. Yoon, Z. Q. Huang, B. D. Strahl, C. D. Allis, and J. Wong. 2002. Involvement of histone methylation and phosphorylation in regulation of transcription by thyroid hormone receptor. *Mol. Cell. Biol.* 22:5688–5697.
- Liu, C. L., T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman, and O. J. Rando. 2005. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.* 3:e328.
- Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260.
- Lyne, R., G. Burns, J. Mata, C. J. Penkett, G. Rustici, D. Chen, C. Langford, D. Vetrie, and J. Bahler. 2003. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics* 4:27.
- Ma, H., C. T. Baumann, H. Li, B. D. Strahl, R. Rice, M. A. Jelinek, D. W. Aswad, C. D. Allis, G. L. Hager, and M. R. Stallcup. 2001. Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr. Biol.* 11:1981–1985.
- Metzger, E., M. Wissmann, N. Yin, J. M. Muller, R. Schneider, A. H. Peters, T. Gunther, R. Buettner, and R. Schule. 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* 437:436–439.
- Morillon, A., N. Karabetsov, J. O'Sullivan, N. Kent, N. Proudfoot, and J. Mellor. 2003. Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell* 115:425–435.
- Morris, S. A., Y. Shibata, K. Noma, Y. Tsukamoto, E. Warren, B. Temple, S. I. Grewal, and B. D. Strahl. 2005. Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryot. Cell* 4:1446–1454.
- Ng, H. H., Q. Feng, H. Wang, H. Erdjument-Bromage, P. Tempst, Y. Zhang, and K. Struhl. 2002. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* 16:1518–1527.
- Nicolas, E., M. G. Lee, M. A. Hakimi, H. Cam, S. S. Grewal, and R. Shiekhattar. 2006. Fission yeast homologs of human histone H3 lysine 4 demethylase regulate a common set of genes with diverse functions. *J. Biol. Chem.* 281:35983–35988.
- Noma, K., H. P. Cam, R. J. Maraia, and S. I. Grewal. 2006. A role for TFIIIC transcription factor complex in genome organization. *Cell* 125:859–872.
- Pena, P. V., F. Davrazou, X. Shi, K. L. Walter, V. V. Verkhusha, O. Gozani, R. Zhao, and T. G. Kutateladze. 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442:100–103.
- Peterson, C. L., and M. A. Laniel. 2004. Histones and histone modifications. *Curr. Biol.* 14:R546–R551.
- Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527.
- Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17:1030–1032.
- Roberts, D. N., A. J. Stewart, J. T. Huff, and B. R. Cairns. 2003. The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc. Natl. Acad. Sci. USA* 100:14695–14700.
- Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* 419:407–411.
- Schlichter, A., and B. Cairns. 2005. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *EMBO J.* 24:1222–1231.
- Schotta, G., A. Ebert, V. Krauss, A. Fischer, J. Hoffmann, S. Rea, T. Jenuwein, R. Dorn, and G. Reuter. 2002. Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* 21:1121–1131.
- Scott, K. C., S. L. Merrett, and H. F. Willard. 2006. A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. *Curr. Biol.* 16:119–129.
- Shi, X., T. Hong, K. L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Pena, F. Lan, M. R. Kaadige, N. Lacoste, C. Cayrou, F. Davrazou, A. Saha, B. R. Cairns, D. E. Ayer, T. G. Kutateladze, Y. Shi, J. Cote, K. F. Chua, and O. Gozani. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442:96–99.
- Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstone, P. A. Cole, and R. A. Casero. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119:941–953.
- Shi, Y. J., C. Matson, F. Lan, S. Iwase, T. Baba, and Y. Shi. 2005. Regulation

- of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**:857–864.
45. **Stavropoulos, P., G. Blobel, and A. Hoelz.** 2006. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol.* **13**: 626–632.
46. **Strahl, B. D., and C. D. Allis.** 2000. The language of covalent histone modifications. *Nature* **403**:41–45.
47. **Strahl, B. D., R. Ohba, R. G. Cook, and C. D. Allis.** 1999. 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.
48. **Suka, N., K. Luo, and M. Grunstein.** 2002. Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat. Genet.* **32**:378–383.
49. **Takeuchi, T., Y. Watanabe, T. Takano-Shimizu, and S. Kondo.** 2006. Roles of jumonji and jumonji family genes in chromatin regulation and development. *Dev. Dyn.* **235**:2449–2459.
50. **Tsukada, Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst, and Y. Zhang.** 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**:811–816.
51. **Whetstone, J. R., A. Nottke, F. Lan, M. Huarte, S. Smolikov, Z. Chen, E. Spooner, E. Li, G. Zhang, M. Colaiacovo, and Y. Shi.** 2006. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**:467–481.
52. **Wysocka, J., T. Swigut, H. Xiao, T. A. Milne, S. Y. Kwon, J. Landry, M. Kauer, A. J. Tackett, B. T. Chait, P. Badenhorst, C. Wu, and C. D. Allis.** 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**:86–90.
53. **Yamane, K., C. Toumazou, Y. Tsukada, H. Erdjument-Bromage, P. Tempst, J. Wong, and Y. Zhang.** 2006. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**:483–495.
54. **Zhang, H., D. O. Richardson, D. N. Roberts, R. Utley, H. Erdjument-Bromage, P. Tempst, J. Cote, and B. R. Cairns.** 2004. The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres. *Mol. Cell Biol.* **24**:9424–9436.