Molecular Basis for Gcn5/PCAF Histone Acetyltransferase Selectivity for Histone and Nonhistone Substrates[†,‡]

Arienne N. Poux and Ronen Marmorstein*

The Wistar Institute, Philadelphia, Pennsylvania 19104, and Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received September 10, 2003

ABSTRACT: Histone acetyltransferase (HAT) proteins often exhibit a high degree of specificity for lysine-bearing protein substrates. We have previously reported on the structure of the *Tetrahymena* Gcn5 HAT protein (tGcn5) bound to its preferred histone H3 substrate, revealing the mode of substrate binding by the Gcn5/PCAF family of HAT proteins. Interestingly, the Gcn5/PCAF HAT family has a remarkable ability to acetylate lysine residues within diverse cognate sites such as those found around lysines 14, 8, and 320 of histones H3, H4, and p53, respectively. To investigate the molecular basis for this, we now report on the crystal structures of tGcn5 bound to 19-residue histone H4 and p53 peptides. A comparison of these structures with tGcn5 bound to histone H3 reveals that the Gcn5/PCAF HATs can accommodate divergent substrates by utilizing analogous interactions with the lysine target and two C-terminal residues with a related chemical nature, suggesting that these interactions play a general role in Gcn5/PCAF substrate binding selectivity. In contrast, while the histone H3 complex shows extensive interactions with tGcn5 and peptide residues N-terminal to the target lysine, the corresponding residues in histone H4 and p53 are disordered, suggesting that the N-terminal substrate region plays an important role in the enhanced affinity of the Gcn5/PCAF HAT proteins for histone H3. Together, these studies provide a framework for understanding the substrate selectivity of HAT proteins.

Histone proteins have long been known to play an important role in packing DNA into the cell (*1*); however, the fact that the post-translational modification to histones plays a more direct role in the regulation of gene expression is now appreciated. Among the modifications that occur on histones are acetylation, phosphorylation, methylation, ribosylation, and ubiquitination (for reviews, see refs *2* and *3*). The first histone modification enzymes to be studied in mechanistic detail were the histone acetyltransferases (HATs), which transfer an acetyl group from coenzyme A (CoA) to the ϵ-amino group of a lysine residue within the N-terminal tails of the histone proteins. Histone acetylation has been demonstrated to be correlated with gene activation, and conversely, histone deacetylation has generally been correlated with gene repression or silencing (*4*−*7*). In addition, there is now a growing body of evidence that shows histone acetylation at some histone sites functions synergistically, and in some cases antagonistically, with other histone modifications such as phosphorylation (*8*, *9*) and methylation (*10*; reviewed in ref *2*). These and related findings have led to the "histone code" hypothesis, whereby discrete combinations of histone modifications elicit distinct transcriptional responses (*11*, *12*).

An interesting property of HAT proteins, which distinguishes them from other enzyme superfamilies that catalyze the same chemical reaction, is that they form subfamilies with little or no sequence homology. In correlation with this, different HAT subfamilies generally have distinct histone substrate specificities (reviewed in ref *13*). For example, the Gcn5/PCAF HAT family has a preference for lysine 14 on histone H3 (*14*), while they also acetylate lysines 8 and 16 on histone H4 to a lesser degree (*15*, *16*). In contrast, all but the Sas3 member of the MYST HAT family show an enhanced preference for lysine residues on histone H4 over those on histone H3 (*17*−*20*). Moreover, the substrate binding specificities of the HAT proteins are expanded by subunits within the *in vivo* HAT complexes. For example, the Gcn5-containing ADA and SAGA complexes also acetylate histone H2B (*21*). Interestingly, some HAT families, such as p300/CBP, show little specificity for histones. Finally, several HAT proteins, such as PCAF and p300/CBP, have been shown to acetylate nonhistone proteins such as the p53 tumor suppressor protein (*22*, *23*), also at distinct lysine sites.

Much of our understanding of the molecular basis for HAT specificity has been derived from structural analysis of *Tetrahymena* Gcn5 (tGcn5) in complex with various histone H3 peptides. The first such structure contained tGcn5 bound to coenzyme A (CoA) and an 11-residue histone H3 peptide centered around lysine 14 (*24*). This structure revealed a

[†] This work was supported by NIH grants to R.M. and by a grant from the Commonwealth Universal Research Enhancement Program, Pennsylvania Department of Health, awarded to The Wistar Institute. A.N.P. was supported by a National Institutes of Health predoctoral training grant, awarded to The Wistar Institute.

[‡] Coordinates for the structures described in detail here have been deposited in the Rutgers Collaborative Structural Bioinformatics (RCSB) Protein Data Bank as entries 1Q2C (tGcn5−H4p19−CoA complex) and 1Q2D (tGcn5−p53p19−CoA complex).

* To whom correspondence should be addressed: The Wistar Institute, 3601 Spruce St., Room 327, Philadelphia, PA 19104. E-mail: marmor@wistar.upenn.edu.

10.1021/bi035632n CCC: $25.00
Published on Web 11/21/2003

central core catalytic domain and N- and C-terminal domains associated with histone substrate binding. A more recent structure of tGcn5 bound to a 19-residue histone H3 substrate, preferred ~100-fold over the 11-residue histone substrate when assayed *in vitro*, revealed the molecular basis of the sequence specificity for lysine 14 of histone H3 (*25*, *26*). In brief, the tGcn5 protein crystallized with a 19-residue histone H3 peptide revealed that 15 of the 19 peptide residues are ordered and that a 12-amino acid core sequence mediates extensive main chain and side chain contacts with the protein, while residues flanking this core mediate backbone contacts that also contribute to histone binding specificity. Another recent structure of tGcn5 bound to a 19-residue histone H3 peptide phosphorylated at serine 10 (*25*), a modification that had been shown to enhance binding of histone H3 to Gcn5 (*8*), reveals that phosphorylation results in additional histone−Gcn5 contacts, both local and distal to the phosphorylation site.

Although histone H4 and p53 have been shown to be biologically relevant substrates for Gcn5/PCAF HAT proteins *in vivo* (*22*, *23*, *27*, *28*), biochemical studies employing the PCAF HAT domain reveal that they are poor substrates relative to histone H3 (*26*). The amino acid sequence around lysines 8 and 320 of histone H4 and p53, respectively, exhibit little homology in sequence to the amino acids proximal to lysine 14 of histone H3, possibly explaining the reduced preference of the Gcn5/PCAF HAT proteins for the histone H4 and p53 substrates. To investigate the molecular basis for how Gcn5/PCAF HAT proteins can accommodate binding to histone H4 and p53, we now report on the crystal structures of tGcn5 bound to 19-residue histone H4 and p53 peptides. Our data indicate that the Gcn5/PCAF HATs can accommodate such varied substrates by utilizing analogous interactions with the lysine target and two C-terminal residues with a related chemical nature, suggesting that these conserved interactions play a general role in tGcn5/PCAF substrate binding selectivity. Comparison of the tGcn5 peptide structures also reveals that substrate regions N-terminal to the target lysine play an important role in the enhanced affinity of the Gcn5/PCAF HAT proteins for histone H3 over the histone H4 and p53 substrates. Together, these studies provide a framework for understanding the substrate binding specificity of HAT proteins.

MATERIALS AND METHODS

Protein Overexpression and Purification. The HAT domain of *Tetrahymena* Gcn5 (tGcn5) (residues 48−210) was overexpressed in bacteria and purified essentially as described previously (*24*). The purified protein was concentrated in a buffer containing 20 mM sodium citrate (pH 6.0), 150 mM NaCl, and 10 mM β-mercaptoethanol to ~20 mg/mL, aliquoted, flash-frozen, and stored at −20 °C.

Cocrystallization and Data Collection. The tGcn5 HAT domain was cocrystallized with a 19-residue histone H4 peptide (residues 1−19) and coenzyme A (CoA), using hanging drop vapor diffusion at 4 °C. A 1.0 μL complex solution containing 0.4 mM tGcn5, 1.5 mM histone H4 19-mer peptide (H4p19), and 1 mM Na-CoA was combined with 1.0 μL of a reservoir solution containing 1.6 M ammonium sulfate, 50 mM Tris-HCl (pH 8.5), and 25 mM magnesium sulfate and equilibrated over 1.0 mL of the reservoir solution. Crystals grew to typical dimensions of 25 μm × 50 μm × 50 μm over 3−5 days and were harvested at 4 °C into a buffer of 1.7 M ammonium sulfate, 50 mM Tris-HCl (pH 8.5), and 25 mM magnesium sulfate. Crystals were cryoprotected by transfer to harvest buffer supplemented with glycerol at concentrations increasing steadily from 5 to 25%. Crystals were subsequently flash-frozen in liquid nitrogen-condensed propane for storage prior to data collection.

Crystals of the tGcn5 complex with a 19-residue p53 peptide (residues 311−329) and CoA were cocrystallized using hanging drop diffusion at room temperature. A 1.0 μL complex solution containing 0.4 mM tGcn5, 1.5 mM p53 19-mer peptide (p53p19), and 1.0 mM Li-CoA was combined with 1.0 μL of the reservoir solution containing 1.7 M ammonium sulfate, 100 mM Hepes (pH 7.5), and 100 mM sodium chloride and equilibrated over 1.0 mL of the reservoir solution. Crystals grew to typical dimensions of 50 μm × 100 μm × 100 μm over 2−3 days. Crystals were harvested at room temperature into a buffer of 1.8 M ammonium sulfate, 100 mM Hepes (pH 7.5), and 100 mM sodium chloride, in the manner described above for the tGcn5−H4p19−CoA complex, and cryoprotected using successive exposure to the harvest buffer supplemented with glycerol at concentrations increasing from 5 to 25%, followed by flash-freezing in liquid propane.

Diffraction data from both crystal forms was collected at the A1 beamline at the Cornell High Energy Synchrotron Source. Data were collected at −180 °C using a single wavelength (λ = 0.9213 Å) and were processed and scaled using the MOSFLM (*29*) and SCALA (*30*) program suites. Both complex crystals were isomorphous in the $P3_221$ space group with similar unit cell dimensions: $a = b = 65.132$ Å and $c = 96.568$ Å and $a = b = 64.740$ Å and $c = 96.413$ Å for the tGcn5−h4p19−CoA and tGcn5−p53p19−CoA complexes, respectively.

Structure Determination and Refinement. The structures of both tGcn5 complexes were determined by molecular replacement with AMORE (*31*), using the tGcn5 protein extracted from the ternary tGcn5−H3p11−CoA complex (PDB entry 1QSN) as a search model (*24*). The structures were refined with CNS (*32*) and the models manually adjusted with O (*33*), using Sigma A-weighted $F_o - F_c$ maps. Advanced stages of refinement incorporated individual *B*-factor refinement, introduction of the coenzyme A and peptide models at 2.5 and 2.25 Å resolution, respectively, and careful addition of solvent molecules. The final ligand models were verified against $F_o - F_c$ maps generated prior to addition of these molecules for refinement. The final protein models for the two tGcn5 complexes have excellent fits to their respective electron density maps, with the exception of N-terminal residue 48, which was not modeled in either complex. In addition, in the case of the tGcn5−p53p19−CoA model, carboxyl-terminal residue Arg210 could not be modeled and Arg61 was modeled as an alanine because of poor electron density in these solvent-exposed regions. The CoA molecules of both complexes had excellent fits to the electron density, particularly in the pantothenic arm region, consistent with previous tGcn5 complex models (*24*, *34*). Only eight of the 19 H4p19 residues and six of the 19 p53p19 residues could be modeled because of poor electron density in other regions of the substrates. Both the tGcn5−H4p19−CoA and tGcn5−p53p19−CoA models were

Table 1: Data and Refinement Statistics

	tGcn5−CoA−H4p19	tGcn5−CoA−p53p19
	Crystal Parameters	
unit cell dimensions	$a = 65.13$ Å, $b = 65.13$ Å, $c = 96.57$ Å, $\alpha = 90.0°$, $\beta = 90.0°$, $\gamma = 120.0°$	$a = 64.74$ Å, $b = 64.74$ Å, $c = 96.41$ Å, $\alpha = 90.0°$, $\beta = 90.0°$, $\gamma = 120.0°$
space group	$P3_221$	$P3_221$
	Data Collection	
resolution range (Å)	27.1−2.0	27.8−1.9
total no. of reflections	81056	91754
no. of unique reflections	23766	23773
R_{symm} (%)	6.9	3.9
I/σ (%)	4.6	11.3
completeness (%)	93.5	87.0
	Refinement Statistics	
resolution range (Å)	27.1−2.25	27.1−2.25
$I/\sigma(I)$ cutoff	0.0	0.0
final model		
protein atoms	1361	1344
water atoms	47	62
CoA atoms	48	48
peptide atoms	48	46
$R_{working}$ (%)	24.1	24.7
R_{free} (%)	27.2	27.7
rms deviation		
bond lengths (Å)	0.008	0.009
bond angles (deg)	1.41	1.51
average *B*-factor (Å^2)		
overall	29.60	26.09
protein	27.42	24.09
water	29.86	32.80
CoA	49.49	29.25
peptide	71.22	64.62

refined to 2.25 Å resolution with excellent crystallographic statistics (Table 1).

RESULTS AND DISCUSSION

Overall Structure of the Complexes. The tGcn5 structures with CoA and the 19-residue histone H4 and p53 substrates described here were both determined at high resolution (2.25 Å) and are crystallographically isomorphous to the corresponding complex with the 19-residue histone H3 complex determined at 2.3 Å resolution (*25*). We therefore feel confident that a structural comparison of these complexes reflects inherent structural differences as opposed to crystal packing artifacts. Both the tGcn5−H4p19−CoA and tGcn5−p53p19−CoA complexes show overall architectural features that are typical of previously described Gcn5/PCAF structures (Figure 1) (*24*, *25*, *35*, *36*). Briefly, the HAT domain consists of a mixed α/β topology, comprised of five α-helices and six β-strands. The N-terminal ($\beta1-\alpha1-\alpha2$) and C-terminal ($\alpha4-\alpha5-\beta6$) regions of the protein lie on opposite sides of a catalytic core ($\beta2-\beta3-\beta4-\alpha3-\beta5$) that is structurally conserved among the Gcn5/PCAF proteins as well as each of the other HAT proteins structurally characterized to date, including Esa1 (*37*) and Hat1 (*38*). Together, the central core and the flanking N- and C-terminal domains create an "L-shaped" cleft into which the CoA and peptide substrates bind nearly orthogonally to the short and long part of the L-shaped cleft, respectively. When compared to the tGcn5−H3p19−CoA complex, the protein components of the tGcn5−H4p19−CoA and tGcn5−p53p19−CoA complexes have rms deviations of only 0.8960 and 0.8011 Å for all atoms, respectively (*25*), emphasizing their high degree of structural homology. The highest degree of variability among the protein component of the three complexes occurs for solvent-exposed residues 183−188 of tGcn5, which form the α5 helix in the complexes with histone H3 and p53, but form a loop in the histone H4 complex. We conclude that this region of the protein is inherently flexible.

Within the catalytic core of the tGcn5−H4p19−CoA and tGcn5−p53p19−CoA complexes, the CoA binds in a very analogous position, showing very little divergence in the position of the pantothenic arm, consistent with the structure of other HAT complexes, including Hat1 (*38*), Esa1 (*37*), and other tGcn5 complex structures (*24*, *25*, *35*, *36*). In contrast, the adenosine base of CoA varies greatly between the structures, also consistent with other HAT complex structures mentioned above.

The greatest structural divergence among the tGcn5 complexes with histones H3 and H4 and p53 maps to the peptide substrates (Figure 1). While 15 of the 19 histone H3 residues crystallized with tGcn5 are ordered in the tGcn5−CoA−H3p19 complex, only the reactive lysine and either seven or five of the residues C-terminal to the catalytic lysine of the histone H4 and p53 peptide substrates could be confidently modeled in their respective complexes (Figure 2). The corresponding ordered peptide residues of the three complexes overlay well with rms deviations of 0.921 and 0.578 Å, respectively, for the Cα atoms of the histone H4 and p53 peptides relative to the histone H3 peptide within the complexes (Figure 1C).

Mode of Histone H4 and p53 Substrate Binding. Figure 3 shows a sequence comparison of the Gcn5/PCAF HAT substrates, H3p19 (Lys14), H4p19 (Lys8), and p53p19 (Lys320). Remarkably, except for the lysine acetylation target, there are no residues that are strictly conserved between the substrates. Lys14 of histone H3 and Lys8 of histone H4 have been shown to be substrates for *Tetrahymena* Gcn5 (tGcn5), yeast Gcn5 (yGcn5), and human PCAF (hPCAF) both *in vitro* and *in vivo* (*14*−*16*, *27*, *39*).

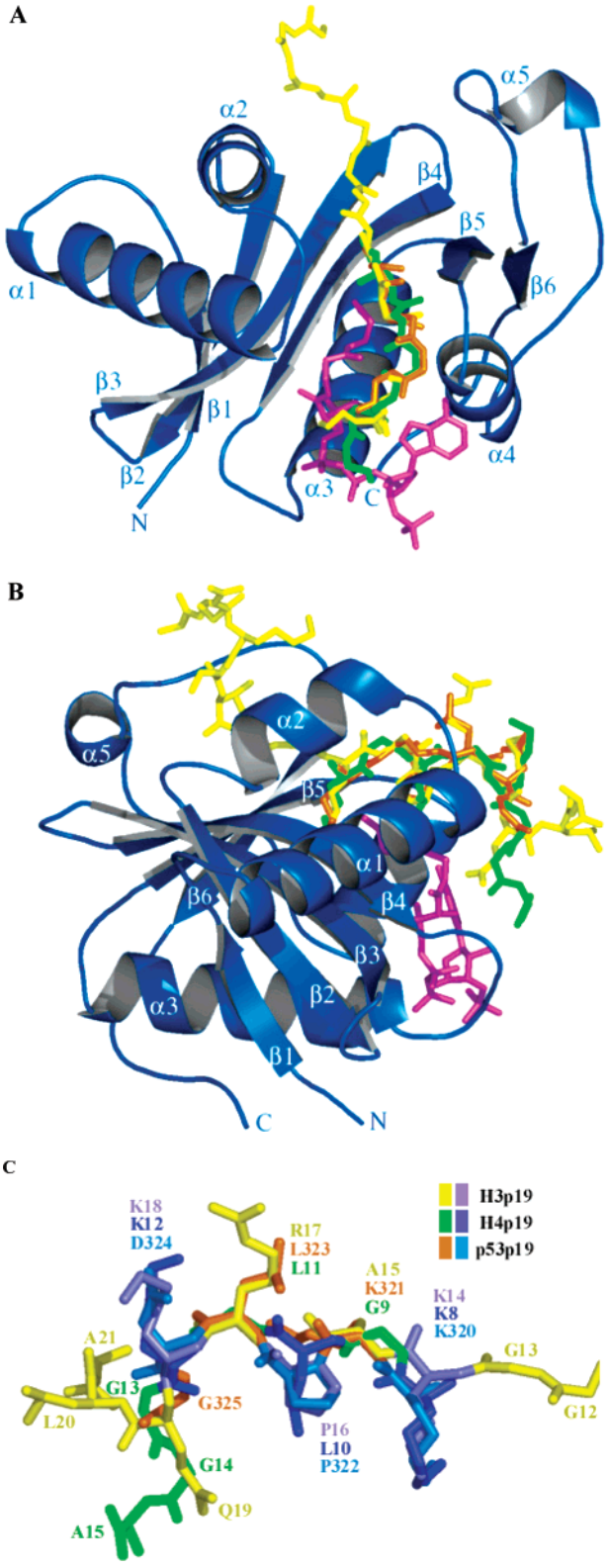


FIGURE 1: Overlay of peptide substrates bound in a ternary complex with tGcn5 and CoA. (A) Superposition of histone H3, histone H4, and p53 19-residue peptides, with the corresponding main chains shown in yellow, green, and orange, respectively. The protein, from the H3 complex model, is shown in blue with secondary structural elements labeled, and coenzyme A is colored purple. The respective proteins were used to align the superposition. (B) Like panel A, but rotated approximately 90° about the vertical axis. (C) Superposition of histone H3, histone H4, and p53 19-residue peptides. Histone H3 is shown in yellow, histone H4 in green, and p53 in orange. Residues in positions 0, +2, and +4 of each are highlighted in various shades of blue (as indicated).

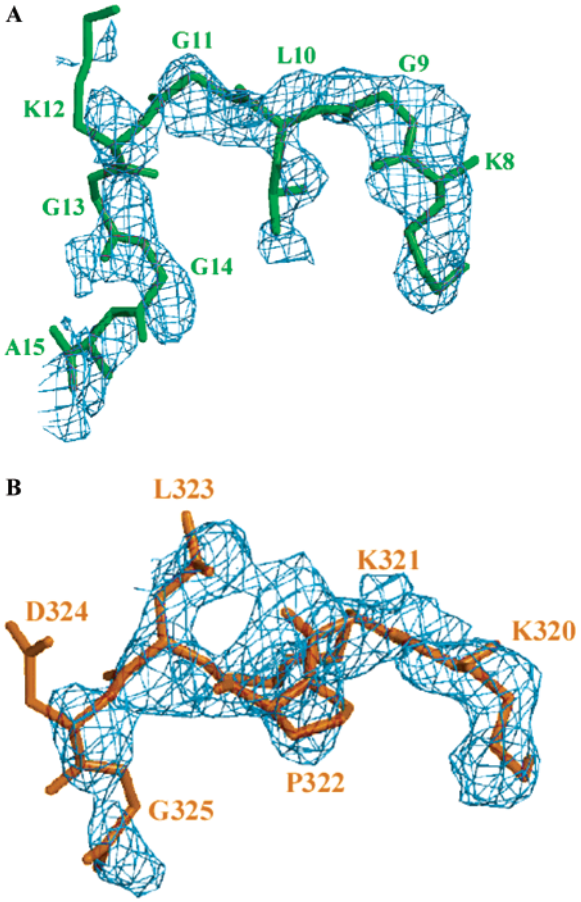


FIGURE 2: Electron density map of histone H4 and p53 peptides bound to tGcn5. (A) A Sigma A-weighted $F_o - F_c$ map omitting the entire modeled histone H4 peptide (green) is shown in blue at a level of 1.75σ. (B) A Sigma A-weighted $F_o - F_c$ map omitting the entire modeled p53 peptide (orange) is shown in blue at a level of 2.0σ.

However, PCAF is the only member of the Gcn5/PCAF family of HAT enzymes that has been demonstrated to acetylate p53 at Lys320. In this study, we used tGcn5 for our structural studies because of its amenability for cocrystallization with peptide substrates. Since tGcn5, yGcn5, and hPCAF show a high degree of sequence and structural homology, in addition to similarities in substrate preferences for histones H3 and H4 (*16*), we feel that tGcn5 provides a suitable model for understanding substrate binding specificity by the Gcn5/PCAF family of HAT proteins.

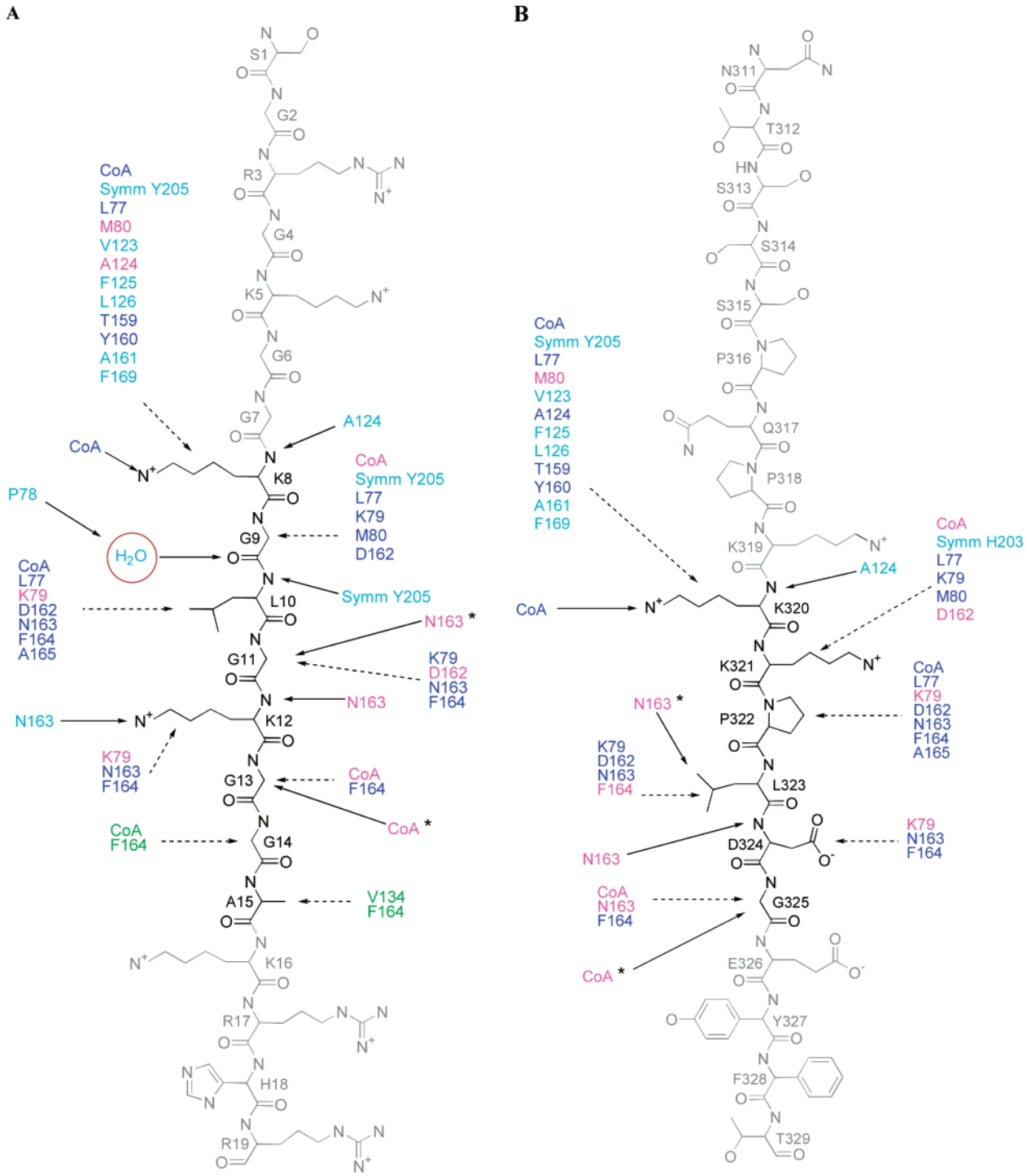
Like the tGcn5−CoA−H3p19 complex, the ternary tGcn5 complexes with the histone H4 and p53 peptide substrates show that the N-terminal (α1−α2) and C-terminal (α5−β6) loops, which comprise the sides of the binding cleft, as well as the β-sheet of the central core (β-strands 2−4), which makes up the floor of the binding cleft, together anchor the substrate (Figure 1). These interactions serve as the scaffold for substrate binding.

The reactive lysine (K8) and seven C-terminal substrate residues of the histone H4 substrate (G9−A15) make interactions with the tGcn5 protein, while seven residues N-terminal to the reactive lysine (S1−G7) and four C-terminal residues (K16−R19) could not be modeled in the structure and are presumably disordered (Figure 2A). Most of the protein interactions are mediated by van der Waals contacts with Lys8 and the three C-terminal residues, Gly9,

	-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9
H3p19	T	A	R	K	S	T	G	G	K_{14}	A	P	R	K	Q	L	A	S	K
H4p19 (8)		S	G	R	G	K	G	G	K_8	G	L	G	K	G	G	A	K	R
p53p19	T	S	S	S	P	Q	P	K	K_{320}	K	P	L	D	G	E	Y	F	T

FIGURE 3: Sequence alignment of Gcn5/PCAF substrates. The substrates, histone H3 (K14), histone H4 (K8), and p53 (K320), are aligned with the reactive lysine residues shown in red and the residues that make analogous protein interactions highlighted in blue. Other residues that are conserved in at least two of the substrates are shown in aqua. Regions of structural overlap are boxed.

Leu10, and Gly11 (Figure 4A,C). Lysine 8 contacts 10 different protein residues (Leu77, Met80, Val123, Ala124, Phe125, Leu126, Thr159, Tyr160, Ala161, and Phe169) and CoA. Glycines 9 and 11 are each contacted by residues Lys79 and Asp162, while Gly9 is also contacted by CoA and protein residues Leu77 and Met80; Gly11 is also contacted by Asn163 and Phe164 and forms an H-bond from Asn163. Leu10 is contacted by CoA and protein residues Leu77, Lys79, Asp162, Asn163, Phe164, and Ala165. Strikingly, the vast majority of protein residues that contact the histone



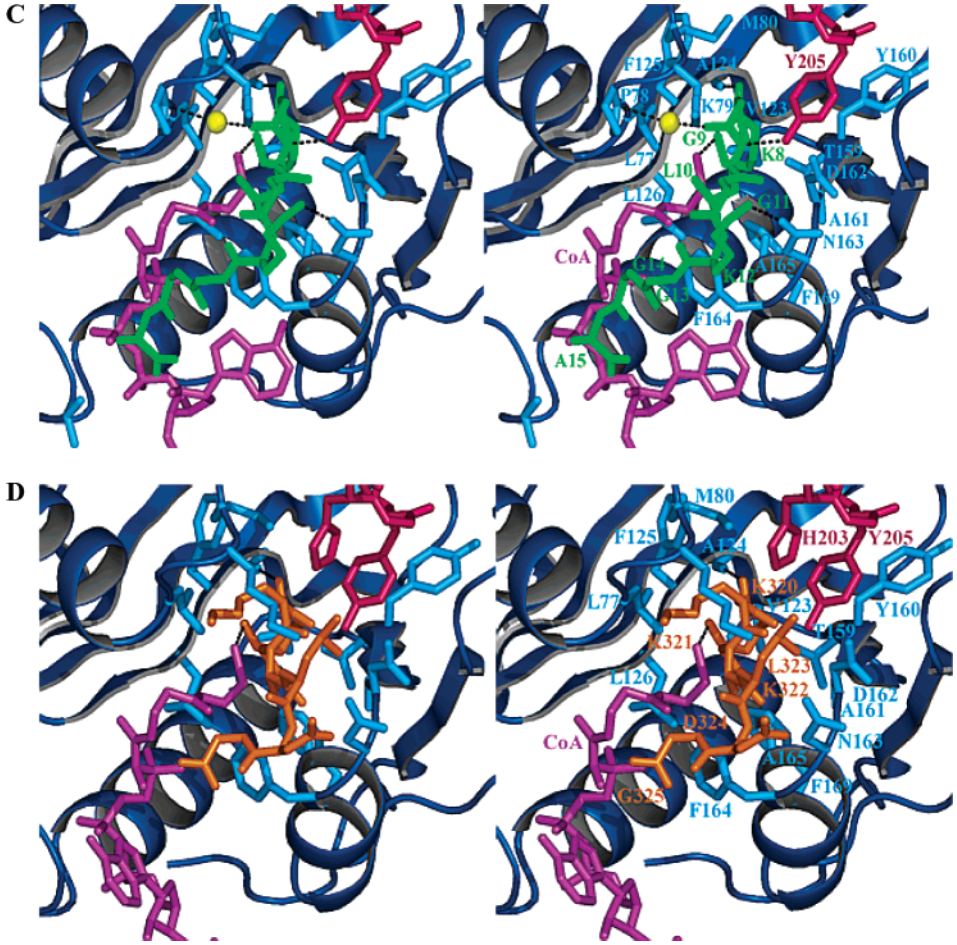


FIGURE 4: Detailed interactions of tGcn5 with histone H4 and p53 peptide substrates. (A) Schematic diagram of the comparative tGcn5−H4p19−CoA vs tGcn5−H3p19−CoA protein−peptide interactions. Protein residues making hydrogen bonds are shown with a solid line, and van der Waals contacts are shown with a dashed line. Peptide residues modeled in the structure are shown in black, and peptide residues that could not be modeled are shown in gray. Residues colored blue make the same interactions in both the tGcn5−H4p19−CoA and tGcn5−H3p19−CoA structures; cyan denotes residues that make histone H4-specific interaction, and magenta denotes residues that make H3-specific interactions. Residues shown in green depict interactions where H4 residues are not in orientations proximal for comparison. (B) Schematic diagram of the comparative tGcn5−p53p19−CoA and tGcn5−H3p19−CoA interactions. All elements are colored as described for panel A. (C) Detailed stereodiagram of the tGcn5−H4p19−CoA complex interface. The protein secondary structural elements are shown in dark blue. Protein residues making hydrogen bond or van der Waals interactions with H4p19 are shown with side chains (main chain for glycines) in light blue. H4p19 is shown in green. CoA is shown in purple. Water is represented by a yellow sphere. Hydrogen bond interactions are depicted with a dashed line. (D) Detailed stereodiagram of the tGcn5−p53p19−CoA complex interface. All elements are depicted as described for panel C, except p53p19 is colored orange.

H4 substrate are conserved with the Gcn5/PCAF HAT family or are mutationally sensitive in yGcn5 (*40*), emphasizing the importance of these interactions for substrate binding by the Gcn5/PCAF HAT proteins.

The reactive lysine (Lys320) and seven C-terminal substrate residues of the histone p53 substrate (G9−A15) make interactions with the tGcn5 protein, while nine residues N-terminal to the reactive lysine (N311−K319) and four C-terminal residues (K16−R19) could not be modeled in the structure and are presumably disordered (Figure 2B). As with the tGcn5 complex with histone H4, the reactive Lys320 and the three residues C-terminal to it (Lys321, Pro322, and Leu323) are the most extensively contacted residues of the tGcn5 protein (Figure 4B,D). Interestingly, many of the same protein residue interactions are conserved, despite the divergence in sequence of the three residues C-terminal to the reactive lysine.

Comparison of tGcn5 Bound to Different Peptide Substrates. A comparison of the structures of tGcn5 cocrystallized with 19-residue peptides from histone H3, H4, and p53 substrates reveals that while the histone H3 substrate makes extensive contacts with 15 of 19 residues, the histone H4 and p53 complexes show interactions with only eight and six residues, respectively. The finding that histone H4 and p53 substrates make less extensive contacts with the tGcn5 protein than with the histone H3 substrate is consistent with biochemical studies demonstrating that histone H3 is preferred over the other substrates. Specifically, enzymatic studies show that H3p19 has a specificity constant (K_{cat}/K_m) that is increased by ~1000- and ~10000-fold over those of the H4p19 and p53p19 peptide substrates, respectively (*26*). The same enzymatic studies reveal that that the 19-residue histone H3 peptide is preferred over an 11-residue histone H3 substrate by ~100-fold. Consistent with these biochemical studies, the structure of tGcn5 crystallized with the 19-residue histone H4 and p53 peptides is more structurally similar to tGcn5 crystallized with the 11-residue histone H3 peptide than the crystallographically isomorphous tGcn5 complex structure crystallized with the histone H3 19-residue peptide.

In the tGcn5−H3p11−CoA complex (cocrystallized with an 11-residue histone H3 peptide), the C-terminus of the peptide is well-ordered with each of the five side chains modeled, three of which are contacted by the tGcn5 protein (*24*). In contrast, residues N-terminal to the reactive lysine are less well-ordered. While the backbone of the N-terminus could be modeled, three of the extreme N-terminal residues could only be modeled as shorter side chains; furthermore, there were no side chain contacts observed in the N-terminus. Together, the tGcn5 structures with H3p11, H4p19, and p53p19 each reveal five to seven side chain residues C-terminal to the reactive lysine that are ordered and participate in protein contacts. In addition, each of these structures exhibits poorly ordered side chains in residues N-terminal to the reactive lysine, where H3p11 exhibits only an ordered backbone region. Consistent with the structural results, H3p11, H4p19, and p53p19 each exhibit significantly reduced substrate specificity constants relative to that of H3p19, where the values for H4p19 and p53p19 peptides are 10−100-fold lower than that for H3p11. On the basis of these correlations, we hypothesize that residues C-terminal to the reactive lysine of the tGcn5 substrates play a particularly important role in substrate selectivity, while residues N-terminal to the reactive lysine play a more important role in substrate affinity.

An overlay of the 19-residue peptides from H3, H4, and p53 reveals that despite the divergence in sequence between the substrates (Figure 3), the C-terminal regions overlay surprisingly well (Figure 1C). Examination of the protein−peptide interactions in this region reveals that the majority of these interactions involve H-bond and van der Waals contacts with the peptide main chain and predominantly van der Waals interactions with the peptide side chains (Figure 4A,B). This comparison also reveals that three nonadjacent side chains in the peptide at positions 0, +2, and +4 (where 0 denotes the reactive lysine) show the best superposition (Figure 1C). The substrate interactions with the protein are maximized at these positions by utilizing residues analogous in chemical nature to those present in the preferred histone H3 substrate. In particular, each of these residues contains an analogously positioned aliphatic region for protein contact. In the case of histone H4, residues at the 0, +2, and +4 positions in the peptide are Lys8, Leu10, and Lys12, respectively, compared to Lys14, Pro16, and Lys18 residues found in histone H3, respectively. Figure 4A shows in detail the interactions of H4p19 with tGcn5 and compares them with H3p19−tGcn5 interactions. Many of the interactions of the peptide with the protein at these positions are conserved (Figure 4A). The catalytic lysines in both the histone H3 and H4 complexes utilize hydrophobic interactions within the binding cleft with CoA, Leu77, Thr159, and Tyr160. Additional hydrophobic interactions are also present between histone H4 and Val123, Phe125, Leu126, Ala161, and Phe169. Furthermore, the main chain of this residue has a hydrogen bond with Ala124 to secure binding of the catalytic residue. At position +2, the histone H4 Leu10 residue is buried into the floor of the binding site. It is in position to make buried, hydrophobic contacts in the same manner as Pro16 of histone H3. Leu10 of histone H4 makes virtually all of the same van der Waals contacts as Pro16 of histone H3, which includes interactions with Leu77, Asp162, Asn163, Phe164, Ala165, and CoA. As a result of a shift in the side chain density of Lys79, Leu10 of histone H4 does not make contacts with this residue. Lys12, at position +4 of histone H4, maintains two of the three van der Waals contacts observed with histone H3, specifically with Asn163 and Phe164. However, as mentioned before, the shift of the Lys79 side chain precludes interaction with Lys12 of histone H4. An additional H-bonding interaction with Asn163 and histone H4 helps secure the Lys12 side chain within the binding site. The histone H4 sequence is very glycine-rich, and the glycines at positions −1 and +1 of the histone H4 peptide make additional stabilizing interactions with the protein that are analogous to those of the histone H3 complex. The Gly9 residue has conserved hydrophobic interactions with Leu77, Lys79, Met80, and Asp162 and a water-mediated hydrogen bond with Pro79. Gly11 makes conserved hydrophobic interactions with Lys79, Asn163, and Phe164. The structural overlap with histone H3 of the histone H4 substrate is relatively poor C-terminal to Lys12 (position +4) of histone H4, and the two substrates make relatively divergent interactions with the tGcn5 protein.

The residues of the p53 peptide are very bulky relative to the glycine-rich residues of the H4p19 and H3p19 peptides, yet regions of the p53p19 peptide show structural overlap with these peptides. Of the six residues in the p53p19 peptide that were modeled, as found with the histone H4 peptide, the residues in positions 0, +2, and +4 show the best superposition when overlaid with the histone H3 peptide (Figures 1C and 4B). In p53p19, these residues correspond to Lys320 (position 0), Pro322 (position +2), and Asp324 (position +4). Similar to Lys14 of histone H3, Lys320 of p53 makes van der Waals interactions with Leu77, Ala124, Thr159, Tyr160, and CoA. Lys320 also makes additional van der Waals contacts with Val123, Phe125, Leu126, Ala161, and Phe169. Lys320 of p53p19 has an additional hydrogen bonding interaction with Ala124, like that found in H4p19. Pro322, at position +2 of p53, has interactions nearly identical to those found with Pro16 of H3p19, excluding one. Both prolines are found buried within the protein binding cleft making extensive hydrophobic contacts with Leu77, Asp162, Asn163, Phe164, Ala165, and CoA. Asp324 of p53, at position +4, which is in the same orientation as Lys18 of H3p19, also makes conserved hydrophobic contacts with Asn163 and Phe164 and hydrogen bonds to Asn163. Analogous to histone H4, the p53p19 residues in positions +1, +3, and +5 make some hydrophobic interactions that are conserved with histone H3, despite the bulkier nature of these residues in the p53 substrate. These include interactions of Lys321 of the p53 peptide with Leu77, Lys79, and Met80 of tGcn5; Leu323 of the p53 peptide with Lys79, Asp162, and Asn163 of tGcn5; and G325 of the p53 peptide with Phe164 of tGcn5.

General Features of Substrate Selectivity by the Gcn5/PCAF HAT Family. Taken together, a comparison of tGcn5 bound to three different cognate substrates suggests how the Gcn5/PCAF family can accommodate a variety of substrates with divergent sequence flanking the reactive lysine. Specifically, the following correlations can be made. (1) The reactive lysine itself at position 0 and regions C-terminal to the reactive lysine, and in particular residues at positions +2 and +4, play a particularly important role in substrate selectivity. Interestingly, these residues make strongly conserved contacts, which take advantage of their similar

chemical nature and permit similar aliphatic interactions, despite the divergence in sequence. The backbone of the substrate in this region in positions +1 and +3 also contributes to substrate binding. (2) The substrate region N-terminal to the reactive lysine plays a particularly important role in modulating substrate affinity.

Members of the Gcn5/PCAF HAT family have also been demonstrated to acetylate lysine residues within other substrates, including lysines 9 and 18 of histone H3 (*14*), lysine 16 of histone H4 (*15*, *16*), and a variety of other nonhistone substrates such as HMGI (*41*), MyoD (*42*), and HIV-1 Tat (*43*, *44*). These residues also exhibit significant sequence divergence within their respective cognate sites; however, residues present at positions +2 and +4 in these substrates could make homologous aliphatic interactions with the Gcn5/PCAF HAT proteins. Another notable feature of the known substrates for the Gcn5/PCAF HAT proteins is that they usually contain a small side chain at position −2, usually glycine or alanine (Figure 3). Although position −2 is disordered in histone H4 and p53 peptides bound to tGcn5, it may be that the absence of significant side chain contact in this region may nonetheless be important for substrate binding. Consistent with this conclusion, the structure of the tGcn5−CoA−H3p19 complex, which has a glycine at this position, has relatively poor electron density (*25*), and modeling of a large side chain in this position would require significant protein rearrangement.

In vivo, members of the tGcn5/PCAF HAT family function as multisubunit protein complexes to acetylate their protein substrates (*21*), and substrate binding specificity is modulated in the context of such complexes (*14*, *16*). It may be that the role of one of more of the subunits within these HAT complexes is to facilitate more optimal contacts with substrates that have inherently poor affinity for the isolated HAT protein, such as the case with the histone H4 and p53 Gcn5/PCAF substrates. Addressing this possibility will require a structure determination of an intact HAT complex bound to a physiologically relevant substrate. Nonetheless, the studies presented here delineate the fundamental requirements for substrate binding by the Gcn5/PCAF HAT proteins, that may then be further elaborated by other protein subunits to extend the substrate selectivity, and possibly also the biological activities, of HAT proteins.

ACKNOWLEDGMENT

We thank J. Rojas, A. Clements, J. Rux, M. Fitzgerald, and K. Zhao for useful discussions, R. Burnett and F. Xue for access to and assistance with computer equipment, and I. Kriksunov, M. Szebenyi, and the staff scientists of MacCHESS at Cornell University (Ithaca, NY).

REFERENCES

1. Wolffe, A. P. (1992) *Chromatin: Structure and Function*, Academic Press, London.
2. Berger, S. L. (2002) *Curr. Opin. Genet. Dev. 12*, 142−148.
3. Wolffe, A. P., and Hayes, J. J. (1999) *Nucleic Acids Res. 27*, 711−720.
4. Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 786−794.
5. Kuo, M.-H., Zhou, J., Jambeck, P., Churchill, M., and Allis, C. D. (1998) *Genes Dev. 12*, 627−639.
6. Vidal, M., and Gaber, R. F. (1991) *Mol. Cell. Biol. 11*, 6317−6327.
7. Taunton, J., Hassig, C., and Schreiber, S. (1996) *Science 272*, 408−411.
8. Lo, W.-S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J.-Y., Allis, C. D., Marmorstein, R., and Berger, S. L. (2000) *Mol. Cell 5*, 917−926.
9. Lo, W.-S., Duggan, L., Emre, N. C. T., Belotserkovskya, R., Lane, W. S., Sheikhattar, R., and Berger, S. L. (2001) *Science 293*, 1142−1146.
10. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) *Science 292*, 110−113.
11. Strahl, B. D., and Allis, D. C. (2000) *Nature 403*, 41−45.
12. Jenuwein, T., and Allis, C. D. (2001) *Science 293*, 1074−1080.
13. Roth, S. Y., Denu, J. M., and Allis, C. D. (2001) *Annu. Rev. Biochem. 70*, 81−120.
14. Grant, P. A., Eberharter, A., John, S., Cook, R. G., Turner, B. M., and Workman, J. L. (1999) *J. Biol. Chem. 274*, 5895−5900.
15. Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996) *Nature 383*, 269−272.
16. Schiltz, L., Mizzen, C. A., Vassilev, A., Cook, R. G., Allis, C. D., and Nakatani, Y. (1999) *J. Biol. Chem. 274*, 1189−1192.
17. Clarke, A. S., Lowell, J. E., Jacobson, S. J., and Pillus, L. (1999) *Mol. Cell. Biol. 19*, 2515−2526.
18. Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J. C. (1997) *EMBO J. 16*, 2054−2060.
19. Kimura, A., and Horikoshi, M. (1998) *Genes Cells 3*, 789−800.
20. Takechi, S., and Nakayama, T. (1999) *Biochem. Biophys. Res. Commun. 266*, 405−410.
21. Grant, P. A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., and Workman, J. L. (1997) *Genes Dev. 11*, 1640−1650.
22. Sakaguchi, K., Herrera, J. E., Saito, S. i., Miki, T., Bustin, M., Vassilev, A., Anderson, C. W., and Appella, E. (1998) *Genes Dev. 12*, 2831−2841.
23. Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D., and Berger, S. L. (2001) *Mol. Cell 6*, 1243−1254.
24. Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, D. C., and Marmorstein, R. (1999) *Nature 401*, 93−98.
25. Clements, A., Poux, A. N., Lo, W.-S., Pillus, L., Berger, S. L., and Marmorstein, R. (2003) *Mol. Cell 12*, 461−473.
26. Trievel, R. C., Li, F.-Y., and Marmorstein, R. (2000) *Anal. Biochem. 287*, 319−328.
27. Ogryzko, V. V., Kotani, T., Zhang, X., Schiltz, R. L., Howard, T., Yang, X. J., Howard, B. H., Qin, J., and Nakatani, Y. (1998) *Cell 94*, 35−44.
28. Liu, L., Scolnick, D. M., Trievel, R. C., Zhang, H. B., Marmorstein, R., Halazonetis, T. D., and Berger, S. L. (1999) *Mol. Cell. Biol. 19*, 1202−1209.
29. Leslie, A. G. W. (1990) *Crystallographic Computing*, Oxford University Press, New York.
30. Evans, P. R. (1993) *Proceedings of the CCP4 Study Weekend: On Data Collection and Processing*, pp 114−122, Daresbury Laboratory, Warrington, U.K.
31. Navaza, J. (2001) *Acta Crystallogr. D57*, 1367−1372.
32. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
33. Jones, T. A., Zou, J. Y., and Cowen, S. W. (1991) *Acta Crystallogr. A47*, 110−119.
34. Poux, A. N., Cebrat, M., Kim, C. M., Cole, P. A., and Marmorstein, R. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 14065−14070.
35. Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, D. C., Berger, S. L., and Marmorstein, R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 8931−8936.
36. Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., and Marmorstein, R. (1999) *EMBO J. 18*, 3521−3532.
37. Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L., and Marmorstein, R. (2000) *Mol. Cell 6*, 1195−1205.
38. Dutnall, R. N., Tafrov, S. T., Sternglanz, R., and Ramakrishnan, V. (1998) *Cell 94*, 427−438.
39. Brownell, J. E., and Allis, C. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6364−6368.
40. Langer, M. R., Tanner, K. G., and Denu, J. M. (2001) *J. Biol. Chem. 276*, 31321−31231.
41. Munshi, N., Merika, M., Yie, J., Senger, K., Chen, G., and Thanos, D. (1998) *Mol. Cell 2*, 457−467.

42. Puri, P. L., Sartorelli, V., Yang, X. J., Hamamori, Y., Ogryzko, V. V., Howard, B. H., Kedes, L., Wang, J. Y., Graessmann, A., Nakatani, Y., and Levrero, M. (1997) *Mol. Cell 1*, 35−45.
43. Benkirane, M., Chun, R. F., Xiao, H., Ogryzko, V. V., Howard, B. H., Nakatani, Y., and Jeang, K. T. (1998) *J. Biol. Chem*. *273*, 24898−24905.
44. Kiernan, R. E., Vanhulle, C., Schiltz, L., Adam, E., Xiao, H., Maudoux, F., Calomme, C., Burny, A., Nakatani, Y., Jeang, K. T., Benkirane, M., and Van Lint, C. (1999) *EMBO J. 18*, 6106−6118.

BI035632N