CMLS, Cell. Mol. Life Sci. 58 (2001) 693–703
1420-682X/01/060693-11 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Structure and function of histone acetyltransferases

R. Marmorstein

The Wistar Institute and the Department of Chemistry, University of Pennsylvania, Philadelphia (Pennsylvania 19104, USA), Fax +1 215 898 0381, e-mail: marmor@wistar.upenn.edu

Abstract. Histone acetyltranferase (HAT) enzymes are the catalytic subunit of large multisubunit HAT complexes that acetylate the ε-amino group of specific lysine residues on histone tails to promote transcriptional activation. Recent structural and functional studies on the divergent HAT enzymes Gcn5/PCAF, Esa1 and Hat1 have provided new insights into the underlying mechanism of histone binding and acetylation by HAT proteins. The three HAT enzymes contain a structurally conserved core domain that plays a functionally conserved role in binding the coenzyme A cofactor and in harboring the putative general base for catalysis. Structurally variable N- and C-terminal domains appear to contain a related scaffold that mediates histone substrate binding. These data provide a framework for understanding the structure and function of other more divergent HAT proteins such as $TAF_{II}250$ and CBP/p300, and provides a starting point for understanding how HAT proteins may cooperate with other factors within in vivo HAT complexes to promote transcriptional activation.

Key words. Histone acetyltransferases (HATs); chromatin regulation; histone modifications; GNAT (Gcn5-related acetyltranferases) superfamily; gene regulation.

Introduction

It has been over 35 years since Allfrey and co-workers proposed that the acetylation state of histones within chromatin is correlated with transcriptional regulation [1, 2], whereby hypoacetylated histones are associated with transcriptionally repressed genes and hyperacetylated histones are associated with transcriptionally active genes. Several subsequent studies reinforced this proposal by showing that hyperacetylated core histones are preferentially associated with transcriptionally active chromatin [3–7]. Despite the time that has elapsed since these initial studies, it has only been over the last 4 years that the proteins that mediate histone acetylation have been identified and characterized. A major breakthrough in this area came with the cloning of a histone acetyltransferase enzyme from *Tetrahymena* [8] that turned out to be a homologue of the previously identified Gcn5 transcriptional coactivator/adaptor from yeast [9]. Functional characterization of yeast Gcn5 mutants revealed a direct correlation between the ability of the protein to acetylate histones and to activate transcription [10,11]. Subsequently, a flurry of studies led to the discovery of a large number of HAT enzymes, many of which, such as CBP/p300 [12, 13], $TAF_{II}250$ [14] and SRC-1 [15], were previously identified transcriptional coactivators. Almost in parallel to the discovery of HAT enzymes came the identification of several histone deactylase (HDAC) enzymes [16] whose activities have been correlated with transcriptional repression. These findings underscore the key role in gene regulation that is played by the enzymes that modulate the acetylation state of histones.

The primary targets of HAT enzymes are the ε-amino groups of specific lysine residues on the N-terminal tails of the histone proteins that package DNA into chromatin [17]. Chromatin structure is generally repressive for transcription since it occludes much of the DNA surface from transcription factor binding. This packaging is mediated by nucleasome core particles that contain about 146 bp of DNA wrapped around an octomer of proteins containing two copies each of histones H2A, H2B, H3 and H4 [18]. Each of the histone proteins contain a globular C-terminal domain that is buried at the interior of the core particle and a more loosely associated, but highly conserved, N-terminal tail region that harbors nearly one-third of the protein mass. The N-terminal histone tail is the sight of covalent modifications such as methylation, ribosylation, phosphorylation and acetylation [19]. The precise mechanism that links these modifications, including acetylation, to transcriptional

regulation is still poorly understood and is an area of intense investigation.

Over the last two years significant advances have been made in the structure and function of HAT proteins. In this report, I will review what is known about the sequence and functional properties of these proteins, and what we have learned from the structure of three divergent HAT proteins, Gcn5/PCAF, Esa1 and Hat1. I will discuss how a correlation of these structural and functional studies leads to a framework for understanding how HAT enzymes catalyze acetylation and how they bind histone substrates. Finally, I will address how these studies may ultimately lead to a better understanding of how histone acetylation leads to transcriptional activation.

Sequence and functional features of HATs

A large number of transcription-related proteins are known to have HAT activity (table 1). Sequence analysis of these proteins reveal that they fall into distinct families that show high sequence similarity within families but poor to no sequence similarity between families [20]. Moreover, the HAT domain sizes between families are different (fig. 1). The first identified transcription-related HAT family includes the Gcn5/PCAF homologues in yeast [21], human (Gcn5 [22] and PCAF [23]), *Drosophila* [24], *Arabidopsis* (accession no. 2642602) and *Tetrahymena* [8]. Other HAT families include CBP/p300 from several organisms [12, 13], $TAF_{II}250$, and its homologues yeast $TAF_{II}130$ and *Drosophila* $TAF_{II}230$ proteins [14], the MYST (MOZ, YBF2/SAS3, SAS2, Tip60) family, members of the steroid receptor coactivators (SRC-1, ACTR/AIB-1/pCIP/RAC3/TRAM-1 and TIF2/GRIP1) [25, 26] and ATF-2 [27] (fig. 1 and table 1).

In addition to the sequence divergence within the HAT domains of different families, their functional contexts are also different (fig. 1 and table 1). The Gcn5/PCAF family function as coactivators/adaptors for a subset of transcriptional activators and also contain a C-terminal bromodomain that has been recently shown to be an acetyl-lysine targeting motif [28–30], thus suggesting a mechanistic connection between HAT and bromodomain function within the Gcn5/PCAF family. The steroid re-

Table 1. HAT families and their transcription-related functions.

HAT	Organism	Function
GCN5/PCAF family		
Gcn5	yeast to human	coactivator (adaptor)
PCAF	human	coactivator
MYST family		
as2	yeast	silencing
Sas3	yeast	silencing
Esa1	yeast	cell cycle progression
MOF	fruit fly	dosage compensation
Tip60	human	HIV Tat interaction
MOZ	human	leukeomogenesis
HBO1	human	origin Recognition interaction
$TAF_{II}250$ family	yeast to human	TBP-associated factor
CBP/p300 family	worm to human	global coactivator
SRC family	mice and human	steroid receptor coactivators
SRC-1		
ACTR/AIB1/pCIP/TRAM-1/RAC3		
SRC-3		
TIF-2		
GRIP1		
ATF-2	yeast to human	sequence-specific DNA-binding activator
GNAT related		
Elp3	yeast	elongation
Hpa2	yeast	unknown
HAT1	yeast to human	deposition-related (B-type)
TFIIIC	human	RNA Pol II Initiation

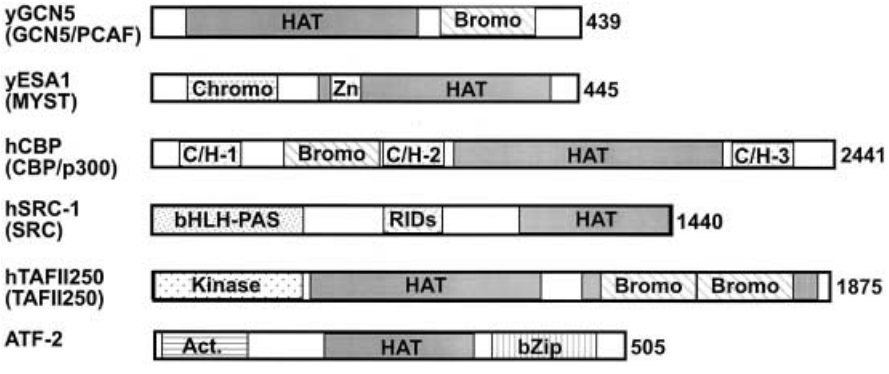

Figure 1. Bar diagram of the different HAT families and their associated domains. A representative member from each HAT family is shown with the family designation indicated in parenthesis. HAT, histone acetyltransferase domain; Bromo, bromodomain; Chromo, chromodomain; Zn, zinc binding domain; C/H, Cys/His rich regions (1, 2 and 3); bHLH-PAS, basic helix-loop-helix and PAS A and PAS B domains; RIDs, LXXLL receptor interacting motifs; Act, transcriptional activation domain; bZip, basic zipper DNA binding domain.

ceptor coactivators enhance transcription of genes responsive to liganded nuclear receptors [31]. In addition to a C-terminal HAT domain these proteins contain an N-terminal basic-helix-loop-helix-PASA-PASB (bHLH-PAS) domain [32] found in proteins that regulates various developmental and physiological events in vertebrates and invertebrates [32], and central receptor interacting domains (RIDs). The CBP/p300 proteins are global regulators of transcription [12] and contain several other protein domains, including a bromodomain and three cystine-histidine rich domains (C/H-1, C/H-2 and C/H-3) that are believed to mediate protein-protein interaction. The $TAF_{II}250$ family of HAT proteins are subunits of the TFIID general transcription factor complex that contains TBP (TATA-binding protein) [14]. Interestingly, these HAT proteins contain two kinase domains, one at each end of the protein, in addition to a double bromodomain that colocalizes to the region of the C-terminal kinase domain and that has been shown to have high affinity for a doubly acetylated histone H4 substrate [29]. The ATF-2 protein is the only sequence-specific DNA-binding transcriptional activator that has been shown to have HAT activity [27]. At the N-terminus of ATF-2 is a potent transcriptional activation domain and near the C-terminus is a basic-zipper (bZip) DNA binding domain [33]. Finally, the MYST family of HAT proteins have divergent biological functions. For example, members such as Sas2 and Sas3 are positive regulators of transcriptional silencing [34], MOZ and TIF2 were identified as translocations in acute leukemias [35, 36] and MOF function is associated with dosage compensation in *Drosophila* [37]. Many of the MYST members have a chromodomain at their N-termini and a small zinc binding domain within the N-terminal region of the HAT domain. Interestingly, it has recently been reported that the chromodomain of MOF binds RNA [38].

All of the HAT proteins that have been characterized in vivo are associated with large multiprotein complexes. For example, yeast Gcn5 (yGcn5) is part of at least two multiprotein complexes, Ada and SAGA [39], and PCAF is part of a SAGA-like complex in vivo [40]. Within the MYST family of proteins, Esa1 is the catalytic subunit of the histone H4-specific NuA4 (Nucleosome acetyltransferase of histone H4) histone acetyltransferase complex [41]; MOF is a component of the MSL complex that regulates dosage compensation in *Drosophila* [37]; and Sas3 is the catalytic subunit of the histone H3-specific HAT complex, NuA3 (Nucleosome acetyltransferase of histone H3) [42].

A general feature of HAT proteins is that although the recombinant proteins will acetylate free histones, nucleosomal acetylation only occurs in the context of the in vivo HAT complexes [26]. Detailed biochemical characterization of the acetylation activity of HAT proteins reveals that each HAT family generally has a unique substrate specificity [20] (table 2). In addition, these specificities are further modulated within the context of the multisubunit in vivo HAT complexes (table 2). For example, recombinant yGcn5 will preferentially acetylate lysine 14 on histone H3, but will also acetylate to a lesser extend lysines 8 and 16 of histone H4 [10, 11, 43]. However, in the context of the Ada and SAGA complexes, the sites that are acetylated on nucleosomal H3 are expanded [44], and histone H2B is also a substrate [10, 11]. Recombinant yeast Esa1 (yEsa1) will efficiently acetylate several lysines on histone H4 as well as lysine 14 of histone H3, whereas the Esa1-containing NuA4 complex has a strong preference for histone H4 only [41].

Table 2. Substrate Specificity For HAT Proteins And HAT Complexes.

HAT	Substrate
GCN5/PCAF family	
Gcn5	H3 (14), H4 (8, 16)
PCAF	H3 (14), p53 (320), MyoD (99, 102, 105) E2F1 (117, 120, 125), HMGI(Y) (71)
GCN5/Ada	H3 (14, 18), H2B (nucleosomes)
GCN5/SAGA	H3 (9, 14, 18), H2B (nucleosomes)
PCAF/complex	H3, H4 (nucleosomes)
MYST family	
Esa1	H4 (5, 8, 16), H3 (14), H2A (5)
MOF	H4 (16), H3, H2A
Tip60	H4, H3, H2A
Sas3	H3, H4, H2A
Esa1/NuA4	H4, H2A (nucleosomes)
Sas3/NuA3	H3 (nucleosomes)
MOF/MSL	H4 (nucleosomes)
CBP/p300	H2A (5), H2B (12, 15), H3 (14, 18), H4 (5, 8), p53 (373, 382), HMG I(Y) (65), TFIIF
SRC-1	H3 (9, 14)

In addition to catalyzing histone acetylation, a number of HAT proteins, including CBP/p300 and PCAF, possess intrinsic transcription factor acetyltransferase (FAT) activity (table 2). PCAF acetylates specific lysines on the transcriptional activators, MyoD [45], E2F1 [46] and p53 [47,48], and on the transcriptional architectural protein HMGI(Y) [49]. CBP/p300 also acetylates specific lysines on p53 [47,48] and HMGI(Y) [49], and on the general transcription factors TFIIEβ and TFIIF [50]. In many, but not all of these cases, acetylation has been shown to enhance the DNA binding affinity of the affected protein.

In addition to the HAT proteins that function in RNA Pol II transcription, several that effect other biological processes have also been identified. Cytoplasmic, or B-type HAT proteins, acetylate newly synthesized histones H3 and H4 prior to their deposition in replicating chromatin [51]. The most well studied B-type HAT, Hat1 from yeast (yHat1), is the catalytic subunit of the Hat1/Hat2 complex that acetylates histone H4 for replication-dependant chromatin assembly [52, 53]. Hat1 homologues from hu-

man [54] and maize [55, 56] have also been identified. Intriguingly, Hat1 from yeast and human has also been localized to the nucleus [54, 57], suggesting a role in transcription- or replication-related histone acetylation. There is also evidence that HAT proteins are used for Pol III transcription since subunits of the TFIIIC complex (TAFIIIC90, TAFIIIC 110 and TAFIIIC 220), that are required for initiation of Pol III transcription also have HAT activity [58, 59].

The GNAT superfamily of acetyltansferases

Neuwald and Landsman have noted that Gcn5/PCAF and Hat1 belong to a functionally diverse superfamily of *N*-acetyltransferases (called GNATs for GCN5 related N-acetyltransferases), containing limited sequence homology within four 15–33-amino acid segments called motifs A, B, C and D [60]. However, Gcn5, P/CAF and Hat1 do not harbor statistically significant sequence homology within motif C. The crystal structure of two members of the GNAT family, yHat1 bound to acetyl-coenzyme A [61], and the *Serratia marcescens* aminoglycoside 3-*N*-acetyltransferase (SmAAT) bound to coenzyme A (CoA) [62] reveal that motifs A, B and D form a structurally conserved domain (fig. 2). Within this conserved domain, contiguous motifs A–D form three antiparallel β strands followed by an α helix on the underside of the β sheet and roughly parallel to its direction. Motif B adopts a different structure in the two proteins, although in both cases the CoA cofactor is bound between the A–D and B submotifs [63], and CoA interactions with the A–D motif are very similar in the two proteins.

The recent structure determination of the HAT GNAT members Gcn5/PCAF [64, 65] and yEsa1 (which only shows sequence homology in motif A) [66] also reveals a homologous structure corresponding to the A–D motif with analogous CoA interactions mediated by these motifs. Interestingly, the structure of *N*-myristoyl transferase (NMT) [67, 68], which is not a member of the GNAT proteins, also contains structural homology in the region corresponding to motifs A–D of the GNAT proteins [63]. These observations imply that other HAT proteins that show sequence divergence from the GNAT proteins, will, nonetheless, contain a structurally homologous A–D motif for CoA interaction.

Overall structure of HATs

A detailed comparison of the HAT domains of various members of the Gcn5/PCAF family in several different liganded forms [64, 65, 69], with the MYST family member yEsa1 bound to CoA [66] and yHat1 bound to Ac-CoA reveals structural homology that extends beyond the A–D motif of the GNAT proteins (fig. 2). The HAT proteins also show a structurally conserved loop-β-strand region immediately C-terminal to the helix of the A–D motif of the GNAT proteins (figs 2 and 3a). In each of the HAT proteins, this structurally homologous region occurs roughly at the center of the molecule and forms a scaffold for the folding of the corresponding N- and C-terminal domains, and I therefor refer to this region of the HAT proteins as the core domain.

In contrast to the structurally similar core domains of the HAT proteins, the protein segments N- and C-terminal to the corresponding core domains show structural divergence (fig. 2). Interestingly, the histone H4-specific yEsa1 and yHat1 structures are more similar to each other than to the histone H3-specific Gcn5/PCAF structure. Both the yEsa1 and yHat1 structures show a β-rich N-terminal protein segment and an α-helix-rich C-terminal segment with similar overall topologies, whereas the Gcn5/PCAF structure shows more dramatic differences. Despite the overall structural differences within the N- and C-terminal segments of the three HAT proteins, subregions within the N- and C-terminal segments show good alignment upon superposition of the core domains.

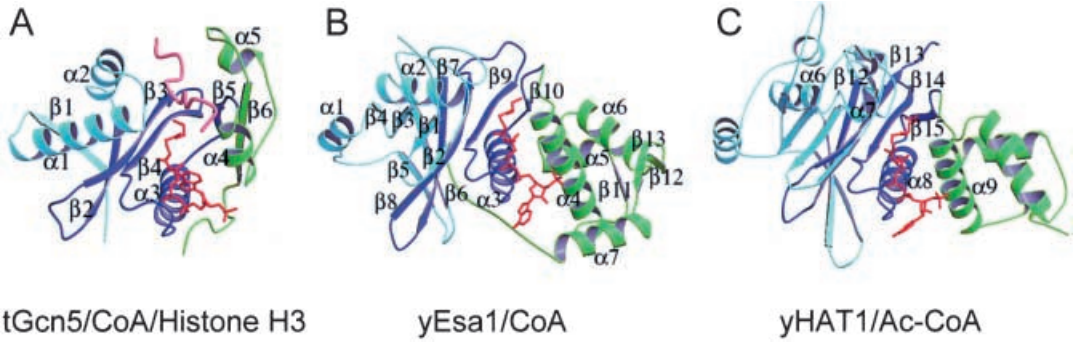

Figure 2. Overall structure of HAT proteins. (*A*) Structure of the *Tetrahymena* Gcn5/CoA/histone H3 peptide complex. The central core domain is colored in blue, the N-terminal subdomain is colored in aqua and the C-terminal subdomain is colored in green. CoA is shown as a stick figure in red, and the 11-amino acid histone H3 peptide is indicated in pink. (*B*) Structure of the MYST HAT family member, yeast Esa1, bound to CoA. (*C*) Structure of the deposition-related (B-type) HAT, yeast Hat1, bound to Ac-CoA.

Specifically, an α-helix-loop region just N-terminal to the core domain and a loop-α-helix region just C-terminal to the core domain superimpose well (fig. 5d).

Role of the central core domain in CoA binding and catalysis

The core domain in each of the HAT proteins plays a conserved role in binding the CoA cofactor (fig. 3b). The CoA is bound in a cavity formed on the surface of the core domain and buries over one-half of the CoA-accessible surface area and approximately 500 Å^2 of protein surface area. Most of the CoA contacts are mediated by a β-strand-loop-α-helix segment of the central core domain, corresponding to motif A of the GNAT proteins. In each case, the CoA is bound in a bent conformation which helps facilitate an extensive set of protein interactions that are mediated predominantly by the pantetheine arm and the pyrophosphate group of the CoA. Strikingly, nearly every functional group of the 16-member pantheteine arm-pyrophosphate chain is contacted by the protein. In addition, nearly all of these contacts are mediated by either protein backbone hydrogen bonds or protein side chain van der Waals interactions. Surprisingly, the adenine base of CoA does not mediate protein contacts in either of the structures, and as a result it adopts a somewhat different orientation within each structure. Interestingly, the structure of a putative HAT from yeast, Hpa2, bound to CoA [70], shows a tetrameric arrangement in which the adenine bases of one dimer mediate base-pair interactions with the other dimer to form the tetramer [70]. The physiological relevance of this is not known; however, it suggests that HAT proteins may employ the adenine moieties of Ac-CoA for other in vivo functions such as to stabilize higher-order structures.

A correlation of the structural and functional data demonstrates that the core domain of yGcn5 also plays an important role in catalysis. The structure of yGcn5 reveals that Glu173, located in the β4 strand of the core domain and strictly conserved within the Gcn5/PCAF family of HAT proteins, is in an ideal position to function as a general base for catalysis [64, 69] (fig. 4a). Glu173 is centered within an electronegative patch, which would be an ideal docking site of a positively charged lysine substrate, and the Glu73 side chain is partially buried within a hydrophobic patch which would serve to raise its pK_a value to facilitate proton extraction from the ε-amino group of the target lysine. Correlating with the structural studies, mutagenesis studies find that a FAE to AAA triple mutation in positions 171–173 of yGCN5 is one of the most debilitating mutations made within the HAT domain [10]. In vivo and in vitro analysis of a E173Q mutant of yGcn5 further supports the role of Glu173 in catalysis [69]. In vivo, the E173Q mutation (yGcn5-E173Q) results in a debilitated growth phenotype and low levels of transcription, similar to that seen for the yGcn5 deletion strain. Detailed kinetic analysis of yGcn5-E173Q [71] reveals that the protein has K_m values for CoA and histone H3 that are indistinguishable from the wild-type protein, showing that the mutation does not effect substrate recognition. In contrast, the K_{cat} value is down 360-fold for the mutant relative to the wild-type protein at pH 7.5, demonstrating that the catalytic step is inhibited by the mutation. At pH 9.5, where deprotonation of the target lysine is not required, K_{cat} is only down by six-fold for the mutant relative to the wild-type protein, further supporting the catalytic role of Glu173 in yGcn5.

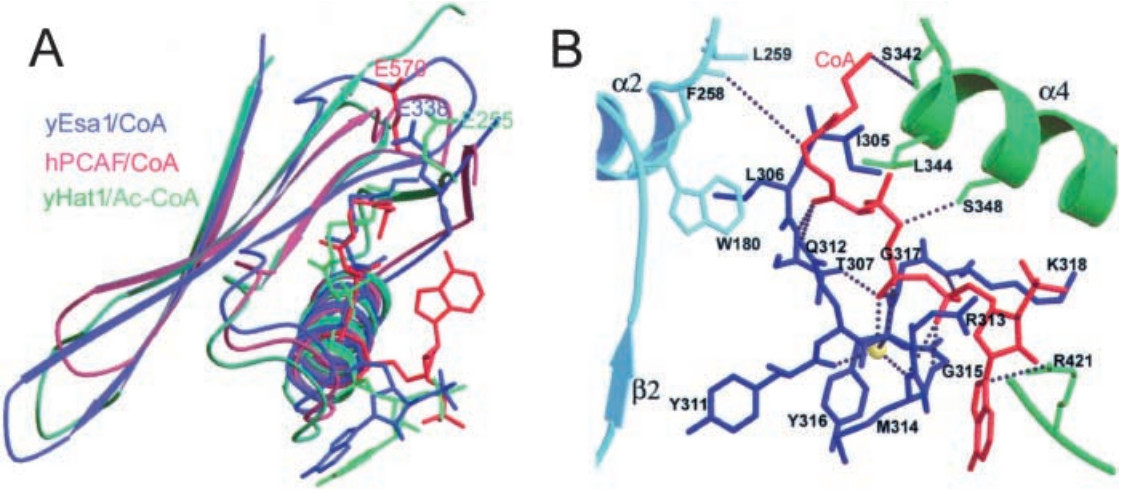

Figure 3. The core domain of HAT proteins. (*A*) Schematic superposition of the core domains of P/CAF (red), Esa1 (blue) and Hat1 (green) highlighting the close alignment of the core domain, the associated CoA cofactor (stick figure) and the corresponding putative catalytic bases. (*B*) A representative view of HAT core domain-CoA interactions are illustrated for the yEsa1/CoA complex. Only protein regions and side-chain residues that mediate direct hydrogen bonds or van der Waals interactions with CoA (red) are indicated. Esa1 color coding is as indicated in the legend to figure 2, and a bound water molecule is shown in yellow.

The ternary *Tetrahymena* Gcn5 (tGcn5)/CoA/histone H3 complex allows one to derive a detailed mechanism of catalysis that is mediated by the glutamate within the core domain of Gcn5 [64] (fig. 4). Glu122 of tGcn5 (Glu173 in yGcn5) is bound to a water molecule located between the Glu122 side chain and the reactive Lys14 residue of the histone H3 peptide, and is held in place by hydrogen bonds from the backbone carbonyl of Val123 and the backbone NH of Tyr160. This water molecule is ideally located to shuttle a proton from the reactive Lys14 of the histone to Glu122 of the protein. Once the lysine proton is extracted, the acetyl group of the Ac-CoA, which is hydrogen-bonded to the backbone NH of Leu126 (in the tGCN5/acetyl-CoA structure), is transferred to the reactive Lys14 side chain of the histone. The backbone NH of Lue126 probably functions to polarize the carbonyl group of the thioester prior to nucleophilic attack of the amino group and stabilize the negative charge that develops on the oxygen in the tetrahedral transition state. Once Lys14 of the histone substrate is acetylated, the acetylated histone substrate is released, probably due to a steric clash between the acetylysine and nearby backbone carbonyls of the protein.

Recent studies on the HAT domain of yEsa1 indicate that its core domain also plays an important role in catalysis [66]. A superposition of the yEsa1 core domain with the core domains of Gcn5/PCAF reveals that the residue corresponding to Glu173 in yGcn5 (570 in PCAF) superimposes most closely in three-dimensional space to Glu338 in yEsa1 (fig. 3a). This superposition occurs despite the fact that the two glutamate residues arise from different secondary structure elements of their corresponding proteins. Glu338 of yEsa1 is conserved within the entire MYST family and is also in an ideal chemical environment to function as a catalytic residue. Functional characterization of an E338Q mutant of yEsa1 (yEsa1-

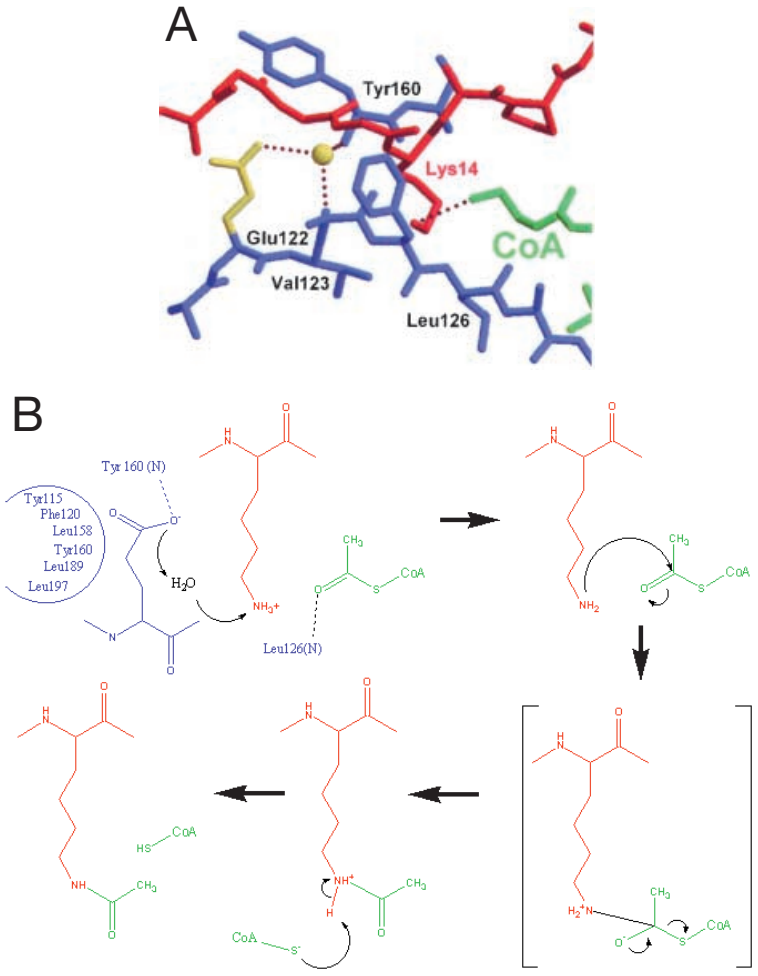

Figure 4. Catalytic mechanism for *Tetrahymena* Gcn5. (*A*) A detailed view of the active site of the ternary tGCN5/CoA/histone H3 complex. The region around histone H3 Lys14 targeted for acetylation is shown in red. Relevant regions of CoA (green) and protein (blue) involved in catalysis or that make important hydrogen bonds (dotted line) or van der Waals interactions are also shown. The putative catalytic base, Glu122, and the water molecule proposed to play a catalytic role are shown in yellow. (*B*) Proposed reaction mechanism for tGCN5-mediated catalysis of histone H3. Protein residues and CoA functionalities that play a direct role in the catalytic mechanism, and the Lys14 substrate of histone H3, are indicated.

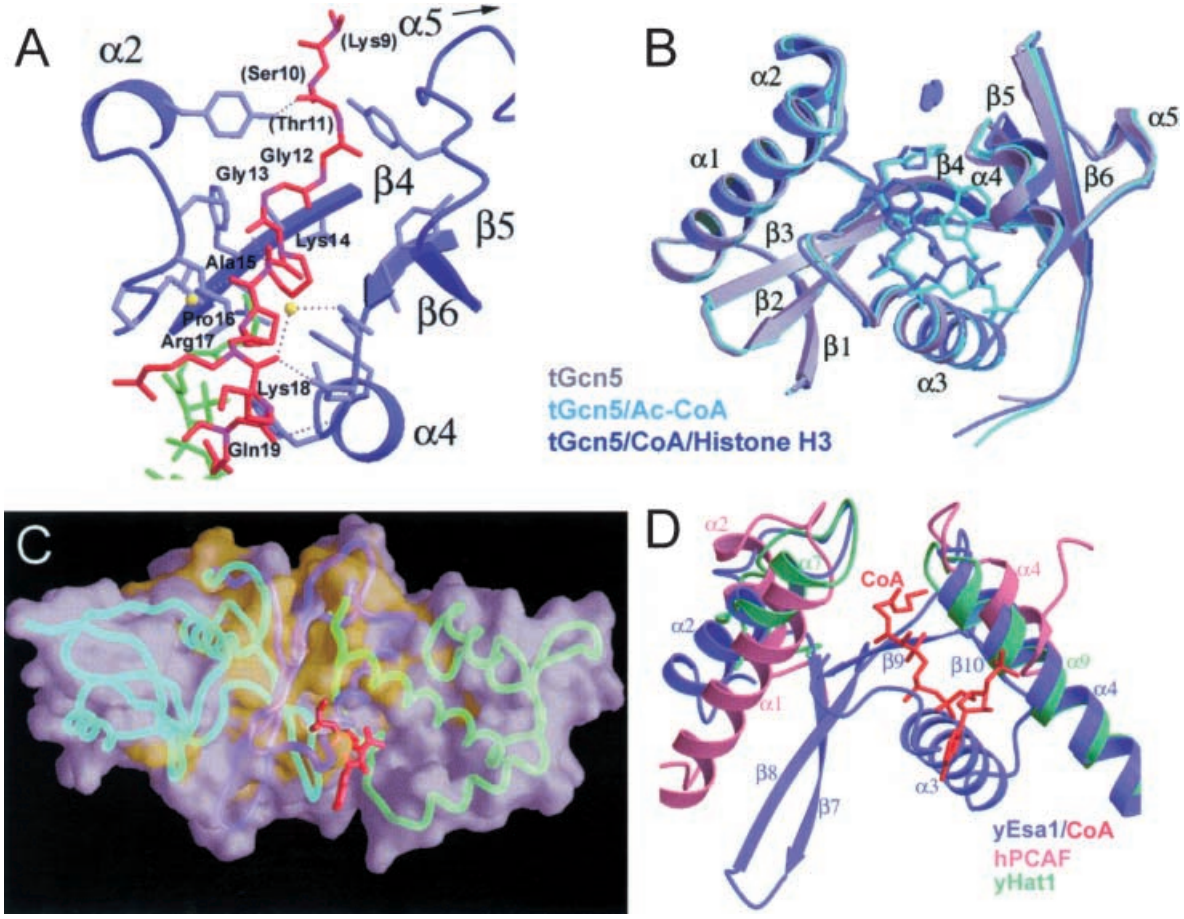

Figure 5. The histone binding site of HAT proteins. (*A*) Detailed view of tGcn5 interactions with histone H3 in the tGcn5/CoA/histone H3 complex. Relevant protein regions are shown in blue, the histone H3 peptide is shown in red, the CoA molecule is shown in green and water molecules that mediate protein-peptide interactions are shown in yellow. (*B*) Superposition of tGcn5 in various liganded states: tGCN5/CoA/histone H3 complex (blue), tGCN5/ acetyl-CoA (aqua) and tGCN5 (gray). (*C*) Conservation of residues within the MYST family are mapped on the surface of yEsa1/CoA structure to illustrate the putative histone substrate binding site. Only residues that are conserved in at least five of six MYST members are shown in yellow coloring. The color coding of Esa1 domains (internal ribbon) are as indicated in the legend to figure 2, and CoA is indicated as a red stick model. (*D*) Schematic superposition of the histone substrate binding site for P/CAF (pink) (human Gcn5 homologue) with Esa1 (blue, with CoA in red) and HAT1 (green). Only the core domain and CoA for Esa1 is shown for clarity.

E338Q) further supports its catalytic role. An in vitro HAT assay reveals that wild-type yEsa1 has robust activity, whereas yEsa1-E338Q has an activity that is only marginally higher than background levels. Moreover, in vivo studies demonstrate that whereas wild-type yEsa1 effectively complements an Esa1 deletion strain, while yEsa1-E338Q does not. Transformation of a wild-type gene into a strain bearing yEsa1-E338Q also does not complement the deletion, further suggesting that the mutation acts as a dominant negative.

Interestingly, a superposition of the core domain of yHat1 with the core domains of Gcn5/PCAF and yEsa1 also reveals that Glu254 of Hat1 superimposes well with the putative general base residues of the other HAT proteins, suggesting that Glu255 may play an important catalytic role in yHat1. In support of this, within the Hat1 protein family, Glu254 is strictly conserved within a stretch of sequence that shows poor conservation. Taken together, these observations suggest that the structurally conserved core domains of Gcn5/PCAF, yEsa1 and yHat1 mediate functionally conserved roles in CoA binding and catalysis.

Role of the N- and C-terminal domains in histone substrate binding

The mode of histone binding by Gcn5/PCAF is visualized in the structure of the ternary tGcn5/CoA/histone H3 complex [64]. The structure reveals that the 11-amino acid histone H3 peptide (centered around the Lys14 target of histone H3) is bound in a pronounced protein cleft of the tGCN5 protein above the core domain and flanked on opposite sites by the N- and C-terminal protein seg-

ments (fig. 5a). The 11-residue histone H3 peptide adopts a random coil structure and buries a total of 1562 Å^2 of solvent-exposed surface area upon complex formation. Most of the protein-peptide interactions are mediated by the loop-α2 N-terminal segment and the loop-α4 C-terminal segment. Remarkably, most of these interactions involve the backbone of the histone H3 peptide, and about 75% of these interactions involve Lys14 and the five residues C-terminal to it.

In addition to Lys14 of the histone H3 substrate, Gly13 and Pro16 play important roles in histone H3 binding specificity. The requirement for Lys14 and Pro16 stem from the extensive protein interactions, and the requirement for Gly13 appears to derive from its unusually constrained packing environment within the protein. Significantly, of the mapped acetylation sites of recombinant yGcn5 for K14 in histone H3, and for K8 and K16 in histone H4, only positions 13 and 16 (and the reactive lysine at position 14) show homology among the three sequences. Position 13 is a glycine in all three cases and position 16 is either a proline, leucine or histidine, each of which can provide hydrophobic surfaces to mimic the interactions mediated by Pro16 of the histone H3 peptide in the complex. Thus, the histone H3 binding determinants of tGCN5 appear to be restricted to a small G-**K**-X-P recognition sequence.

Comparison of the ternary tGcn5/CoA/histone H3 complex with the binary tGcn5/CoA and apo-tGcn5 structures also demonstrates that CoA plays an unanticipated structural role in configuring the HAT domain for histone H3 binding (fig. 5b). A superposition of the three structures reveals that CoA binding widens the histone H3 peptide binding groove by moving the C-terminal protein segment outward. This groove is further widened upon binding of the histone H3 peptide. A comparison of CoA-protein contacts between the binary and ternary complexes also shows that CoA interactions to the core domain are left largely unchanged, whereas CoA interactions with the N- and C-terminal protein segments, which also interact with the histone H3 peptide, make different CoA interactions. Taken together, it appears that CoA-mediated protein contacts that in turn facilitate histone H3 contacts.

Neither the yHat1/Ac-CoA [61] or yEsa1/CoA [66] structures have bound peptide, so one cannot directly visualize HAT protein-histone interactions in these cases. However, a mapping of conserved and surface-exposed residues in the respective HAT families highlights the importance of the α-helix-loop region in the N-terminal segment (α2-loop in yEsa1 and α7-loop in yHat1) and a loop-α-helix region in the C-terminal segment (loop-α4 in yEsa1 and loop-α9 in yHat1). This is shown for Esa1 in figure 5c. These segments flank opposite sides of the core domain, and as in Gcn5, form a pronounced cleft that is proximal to the CoA and the putative general base for catalysis. Correlatively, a superposition of the core domains of Gcn5/PCAF, yEsa1 and yHAT1 shows that these same regions that show sequence conservation within the corresponding HAT families superimpose remarkably well. This is surprising given the overall structural divergence of the corresponding N- and C-terminal segments between these proteins. Taken together, the three HAT proteins analyzed here suggest a conserved structural framework for histone substrate binding whereby sequence divergence within this framework may modulate the binding to specific histone targets. This proposal is consistent with the observation that the analogous structural regions of GNAT proteins that bind non-histone targets have significantly more divergent N- and C-terminal structures.

Conclusions and implications for other HATs

Despite the fact that yEsa1 shows weak sequence homology to PCAF/Gcn5 and yHat1, the entire core domain of the three HAT proteins superimpose remarkably well and also mediate very similar CoA interactions. The studies described above also suggest that the core domains of yEsa1 and PCAF/Gcn5, and possibly also yHAT1, contain a structurally superimposable general base for catalysis and thereby share a related catalytic mechanism. Taken together, it is reasonable to propose that protein members from other HAT families, such as CBP/p300 and $TAF_{II}250$, also contain a structurally conserved core domain that mediates functionally conserved roles in CoA binding and catalysis.

Regions N- and C-terminal to the core domain of the HAT proteins Gcn5/PCAF, yEsa1 and yHAT1 show overall structural divergence. Nonetheless, a superposition of the core domains of these proteins reveal striking superposition between a N-terminal α-helix-loop and a C-terminal loop-α-helix. These superimposed regions colocalize to the precise regions of tGcn5 within the ternary tGcn5/CoA/histone H3 peptide complex that play particularly critical roles in histone substrate binding. Moreover, in each case, these regions contain surface-exposed residues that are highly conserved within their respective HAT families, and directly flank the core domain which harbors the putative catalytic base. Taken together, the three HAT proteins analyzed here appear to have a conserved structural framework for histone substrate binding that likely extends to other more divergent HAT proteins. Moreover, it is likely that sequence divergence within this framework modulates the binding to specific histone targets.

A comparison of tGcn5 in various liganded forms reveals that the CoA cofactor plays an unanticipated structural role in histone substrate binding. CoA makes direct interactions with the same helix-loop and loop-helix segments

from the N- and C-terminal protein domains, respectively, that mediate histone substrate interactions. Moreover, these interactions appear to influence structural movement of the C-terminal segment of tGCN5. Correlatively, CoA also makes interactions with these homologous N- and C-terminal segments in both the yEsa1/CoA and yHat1/Ac-CoA structures, and it is reasonable to assume that CoA also mediates structural movements in these respective HAT proteins to facilitate histone-specific binding. Since the histone H4-specific HAT proteins, Esa1 and Hat1, are more similar to each other than to the histone H3-specific Gcn5/PCAF, it is likely that the N- and C-terminal segments of yEsa1 and yHat1 represent a more histone H4-specific structure, whereas the corresponding segments of Gcn5/PCAF and PCAF represent a more histone H3-specific structure. Whether this is indeed true will require the structural analysis of other more divergent HAT proteins such as CBP/p300 or $TAF_{II}250$. Nonetheless, the structural and functional studies presented here provide a framework for understanding how HAT proteins acetylate and bind histone substrates.

Perspectives

Although we have learned much over the last few years about catalysis and substrate binding by HAT enzymes, several questions about HAT function still remain. Most notably, it is not clear how histone target specificity is achieved. The ternary tGcn5/CoA/histone H3 complex reveal that protein-peptide interactions are mediated predominantly by not-discriminative peptide backbone contacts and sequence specificity determinants are restricted to a small G-K-X-P recognition sequence. Functional data demonstrate that HAT proteins and HAT complexes have similar but distinct substrate specificities, supporting the notion that substrate specificity is modulated, at least in part, by other proteins and/or other protein domains, within the in vivo HAT complexes. In this vain, both the Gcn5/PCAF [64] and yEsa1 [66] structures reveal conserved patches of residues that do not appear to be directly involved in substrate binding but may highlight contact sites for other proteins or protein domains within the in vivo HAT complexes. A resolution of the substrate specificity issue will likely require the structural analysis of the larger in vivo HAT complexes. A related question is, what directs different HAT complexes to different histones?

The observation that aberrant HAT activity is associated with diseased states, such as the association of the HAT proteins MOZ and CBP in acute leukemias [35, 36, 72], leads to the prospect of designing HAT specific inhibitors to effect disease progression. In theory, such inhibitors should be possible to design since bisubstrate peptide inhibitors that exhibit HAT specificity have already been reported [73]. Aside from the disease connection, the availability of HAT-specific inhibitors would be a useful research tool since it would help one to sort out which HAT is acetylating which substrate and when. Finally, the million dollar question is, how does histone acetylation lead to transcriptional activation? In order to answer this question, it may be necessary to understand why large multimeric HAT complexes are required to carry out the simple task of transferring an acetyl group to the ε-amino groups of histone lysines.

Acknowledgments. I would like to thank members of my laboratory, Jeannie Rojas, Yuan Yan, Raymond Trievel and Adrienne Clements, who have contributed to the work described here and to the preparation of several of the illustrations. Figures 2, 3, 5B, 5C, and 5D have been reprinted from reference 66, with permission from Elsevier Science – Figure 4A has been reprinted from reference 64.

1 Pogo B. G. T., Allfrey V. G. and Mirsky A. E. (1966) RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. Proc. Natl. Acad. Sci. USA **55:** 6212–6222
2 Allfrey V. G., Faulkner R. and Mirsky A. E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. USA **51:** 786–794
3 Vidali G., Boffa L. C., Bradbury E. M. and Allfrey V. G. (1978) Butyrate suppression of histone deacetylation leads to accumulation of nultiacetylated forms of histones H3 and H4 and increased DNAase I sensitivity of the associated DNA sequences. Proc. Natl. Acad. Sci. USA **75:** 2239–2243
4 Hebbes T. R., Thorne A. W. and Crane-Robinson C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. **7:** 1395-1403
5 Hebbes T. R., Clayton A. L., Throne A. W. and Crane-Robinson C. (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO **13:** 1823–1830
6 Vettese-Dadey M., Grant P. A., Hebbes T. R., Crane-Robinson C., Allis C. D. and Workman J. L. (1996) Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J. **15:** 2508–2518
7 Sealy L. and Chalkley R. (1978) DNA associated with hyperacetylated histone is preferentially digested by DNasa I. Nucleic Acids Res. **5:** 1863–1876
8 Brownell J. E., Zhou J., Ranalli T., Kobayashi R., Edmondson D. G., Roth S. Y. et al. (1996) *Tetrahymena* histone acetyltransferase A: a homologue of yeast Gcn5p linking histone acetylation to gene activation. Cell **84:** 843–851
9 Berger S. L., Pina B., Silverman N., Marcus G. A., Agapite J., Regier J. L. et al. (1992) Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. Cell **70:** 251–265
10 Wang L., Liu L. and Berger S. L. (1998) Critical residues for histone acetylation by Gcn5, functioning in Ada and SAGA complexes, are also required for transcriptional function in vivo. Genes Dev. **12:** 640–653
11 Kuo M. H., Zhou J. X., Jambeck P., Churchill M. E. A. and Allis C. D. (1998) Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev. **12:** 627–639
12 Shikama N., Lyon J. and LaThangue N. B. (1997) The p300/CBP family: integrating signals with transcription factors and chromatin. Trends Cell Biol. **7:** 230–236

13 Bannister A. J. and Kouzarides T. (1996) The CBP co-activator is a histone acetyltransferase. Nature **384:** 641–643
14 Mizzen C. A., Yang X.-J., Kokubo T., Brownell J. E., Bannister A. J. and Owen-Hughes T. (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell **87:** 1261–1270
15 Spencer T. E., Jenster G., Burcin M. M., Allis C. D., Zhon J. X. and Mizzen C. A. (1997) Steroid receptor coactivator-1 is a histone acetyltransferase. Nature **389:** 194–198
16 Cress W. D. and Seto E. (2000) Histone deacetylases, transcriptional control and cancer. J. Cell. Physiol. **184:** 1–16
17 Wolffe A. P. (1992), Chromaton: Structure Function, Academic Press, London
18 Luger K., Mader A. W., Richmond R. K., Sargent D. F. and Richmond T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature **389:** 251–260
19 Wolffe A. P. and Hayes J. J. (1999) Chromatin disruption and modification. Nucleic Acids Res. **27:** 711–720
20 Kuo M. H. and Allis C. D. (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. Bioessays **20:** 615–626
21 Marcus G., Silverman N., Berger S., Horiuchi J. and Guarente L. (1994) Functional similarity and physical association between GCN5 and ADA2-putative transcriptional adaptors. EMBO J. **13:** 4807–4815
22 Candau R., Moore P. A., Wan, L., Barlev N., Ying C. Y., Rosen C. A. and Berger S. L. (1996) Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5. Mol. Cell. Biol. **16:** 593–602
23 Yang X.-J., Ogryzko V. V., Nishikawa J., Howard B. H. and Nakatani Y. (1996) A p300/CBP-associated factor that competes with the adenoviral E1A oncoprotein. Nature **382:** 319–324
24 Smith E. R., Belote J. M., Schiltz R. L., Yang X. J., Moore P. A., Berger S. L. et al. (1998) Cloning of *Drosophila* GCN5: conserved features among metazoan GCN5 family members. Nucleic Acids Res. **26:** 2948–2954
25 Neal K. C., Pannuti A., Smith E. R. and Lucchesi J. C. (2000) A new human member of the MYST family of histone acetyl transferases with high sequence similarity to *Drosophila* MOF. Biochim. Biophys. Acta-Gene Struct. Expression **1490:** 170–174
26 Sterner D. E. and Berger S. L. (2000) Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. **64:** 435–459
27 Kawasaki H., Schiltz L., Chiu R., Itakura K., Taira K., Nakatani Y. et al. (2000) ATF-2 has intrinsic histone acetyltransferase activity which is modulated by phosphorylation. Nature **405:** 195–200
28 Dhalluin C., Carlson J. E., Zeng L., He C., Aggarwal A. K. and Zhou M. M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. Nature **399:** 491–496
29 Jacobson R. H., Ladurner A. G., King D. S. and Tjian R. (2000) Structure and function of a human TAF(II)250 double bromodomain module. Science **288:** 1422–1425
30 Ornaghi P., Ballario P., Lena A. M., Gonzalez A. and Filetici P. (1999) The bromodomain of Gcn5p interacts in vitro with specific residues in the N terminus of histone H4. J. Mol. Biol. **287:** 1–7
31 Leo C. and Chen J. D. (2000) The SRC family of nuclear receptor coactivators. Gene **245:** 1–11
32 Crews S. T. and Fan C. M. (1999) Remembrance of things PAS: regulation of development by bHLH- PAS proteins. Curr. Opin. Genet. Dev. **9:** 580–587
33 Burley S. K. (1994) DNA-binding motifs from eukaryotic transcription factors. Curr. Opin. Struct. Biol. **4:** 3–11
34 Reifsnyder C., Lowell J., Clarke A. and Pillus L. (1996) Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases. Nat. Genet. **14:** 44–49
35 Borrow J., Stanton V. P., Andresen J. M., Becker R., Behn F. G. and Chaganti R. S. K. (1996) The translocation t(8;l6)(p11, p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB binding protein. Nat. Genet. **14:** 33–41
36 Carapeti M., Aguiar R. C. T., Watmore A. E., Goldman J. M. and Cross N. C. P. (1999) Consistent fusion of MOZ and TIF2 in AML with inv(8)(p11q13). Cancer Genet. Cytogenet. **113:** 70–72
37 Hilfiker A., HilfikerKleiner D., Pannuti A. and Lucchesi J. C. (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila.* EMBO J. **16:** 2054–2060
38 Akhtar A., Zink D. and Becker P. B. (2000) Chromodomains are protein-RNA interaction modules. Nature **407:** 405–409
39 Grant P. A., Duggan L., Côté J., Roberts S. M., Brownell J. E., Candau R. et al. (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. **11:** 1640X–1650
40 Ogryzko V.V., Katani T., Zhang X., Schiltz R. L., Howard T. and Yang X. T. (1998) Histone-like TAFs within the PCAF histone acetylase complex. Cell **94:** 35–44
41 Allard, S., Utley R. T., Savard J., Clarke A., Grant P. and Brandl C. J. (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. **18:** 5108-5119
42 John S., Howe L., Tafrov S. T., Grant P. A., Sternglanz R. and Workman J. L. (2000) The Something About Silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)- FACT complex. Genes Dev. **14:** 1196–1208
43 Kuo M. H., Brownell J. E., Sobel R. E., Ranalli T. A., Cook R. G., Edmondson D. G. et al. (1996) Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature **383:** 269-272
44 Grant P. A., Eberharter A., John S., Cook R. G., Turner B. M. and Workman J. L. (1999) Expanded lysine acetylation specificity of Gcn5 in native complexes. J. Biol. Chem. **274:** 5895–5900
45 Puri P. L., Sartorelli V., Yang K. J., Hamamori Y. and Ogryzko V. V. (1997) Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. Mol. Cell **1:** 35–45
46 Martinez-Balbas M. A., Bauer U.-M., Nielson S. J., Brehm A. and Kouzarides T. (2000) Regulation of E2F1 activity by acetylation. EMBO J. **19:** 662–671
47 Gu W. and Roeder R. G. (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell **90:** 595–606
48 Liu L., Scolnick D. M., Trievel R. C., Zhang H.-B., Marmorstein R., Halazonetis T. D. et al. (1999) p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. Mol. Cell. Biol. **19:** 1202–1209
49 Munshi N., Merika M., Yie J. M., Senger K., Chen G. Y. and Thanos D. (1998) Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome. Mol. Cell. **2:** 457–467
50 Imhof A., Yang X. J., Ogryzko V. V., Nakatani Y., Wolffe A. P. and Ge H. (1997) Acetylation of general transcription factors by histone acetyltransferases. Curr. Biol. **7:** 689–692
51 Brownell J. and Allis C. (1996) Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. **6:** 176–184
52 Kleff S., Andrulis E. D., Anderson C. W. and Sternglanz R. (1995) Identification of a gene encoding a yeast histone H4 acetyltransferase. J. Biol. Chem. **270:** 24674–24677
53 Parthun M. R., Widom J. and Gottschling D. E. (1996) The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell **87:** 85–94

54 Verreault A., Kaufman P. D., Kobayashi R. and Stillman B. (1997) Nucleosomal DNA regulates the core-histone-binding subunit of the human HAT1 acetyltransferase. Curr. Biol. **8:** 96–108
55 Eberharter A., Lechner T., GoralikSchramel M. and Loidl P. (1996) Purification and characterization of the cytoplasmic histone acetyltransferase B of maize embryos. FEBS Lett. **386:** 75–81
56 Kolle D., Sarg B., Lindner H. and Loidl P. (1998) Substrate and sequential site specificity of cytoplasmic histone acetyltransferases of maize and rat liver. FEBS Lett. **421:** 109–114
57 Ruiz-Garcia A. B., Sendra R., Galiana M., Pamblanco M., Perez-Ortin J. E. and Tordera V. (1998) HAT1 and HAT2 proteins are components of a yeast nuclear histone acetyltransferase enzyme specific for free histone H4. J. Biol. Chem. **273:** 12599–12605
58 Kundu T. K., Wang Z. and Roeder R. G. (1999) Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity. Mol. Cell. Biol. **19:** 1605–1615
59 Hsieh Y. J., Kundu T. K., Wang Z., Kovelman R. and Roeder R. G. (1999) The TFIIIC90 subunit of TAFIIIC interacts with multiple components of the RNA polymerase III machinery and contains histone-specific acetyltransferase activity. Mol. Cell. Biol. **19:** 7697–7704
60 Neuwald A. F. and Landsman D. (1997) GCN5-related histone N-acetyltransferases belong to a diverse superfamily that include the yeast SPT10 protein. Trends Biochem. Sci. **22:** 154–155
61 Dutnall R. N., Tafrov S. T., Sternglanz R. and Ramakrishnan V. (1998) Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. Cell **94:** 427–438
62 Wolf E., Vassilev A., Makino Y., Sali A., Nakatani Y. and Burley S. K. (1998) Crystal structure of a GCN5-related N-acetyltransferase: *Serratia marcescens* aminoglycoside 3-N-acetyltransferase. Cell **94:** 439–449
63 Modis Y. and Wierenga R. (1998) Two crystal structures of N-acetyltransferases reveal a new fold for CoA-dependent enzymes. Structure **6:** 1345–1350
64 Rojas J. R., Trievel R. C., Zhou J., Mo Y., Li X., Berger S. L. et al. (1999) Structure of the *Tetrahymna* GCN5 bound to coenzyme-A and a histone H3 peptide. Nature **401:** 93-98
65 Clements A., Rojas J. R., Trievel R. C., Wang L., Berger S. L. and Marmorstein R. (1999) Crystal structure of the histone acetyltransferase domain of the human P/CAF transcriptional regulator bound to coenzyme-A. EMBO J. **18:** 3521–3532
66 Yan Y., Barlev N. A., Haley R. H., Berger S. L. and Marmorstein R. (2000) Crystal structure of yeast Esa1 suggests a unified mechanism of catalysis and substrate binding by histone acetyltransferases. Mol. Cell **6:** 1195–1205
67 Bhatnagar R. S. et al. (1998) Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. Nat. Struct. Biol. **5:** 1091–1097
68 Weston S. A., Camble R., Colls J., Rosenbrock G. and Taylor I. (1998) Crystal structure of the anti-fungal target N-myristoyltransferase. Nat. Struct. Biol. **5:** 213–221
69 Trievel R. C., Rojay J. R., Sterner D. E., Venkataramani R., Wang L. and Zhon J. (1999) Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. Proc. Natl. Acad. Sci. USA **96:** 8931–8936
70 Angus-Hill M. L., Dutnall R. N., Tafrov S. T., Sternglanz R. and Ramakrishnan V. (1999) Crystal structure of the histone acetyltransferase Hpa2: A tetrameric member of the Gcn5-related N-acetyltransferase superfamily. J. Mol. Biol. **294:** 1311-1325
71 Tanner K. G., Trievel R. C., Kuo M.-H., Howard R., Berger S. L., Allis C. D. et al. (1999) Catalytic mechanism and function of invariant glutamic acid-173 from the histone acetyltransferase GCN5 transcriptional coactivator. J. Biol. Chem. **274:** 18157–18160
72 Carapeti M., Aguiar R. C. T., Goldman J. M. and Cross N. C. P. (1998) A novel fusion between MOZ and the nuclear receptor coactivator TIF2 in acute myeloid leukemia. Blood **91:** 3127–3133
73 Lau O. D., Kundu T. K., Soccio R. E., Ait-Si-Ali S., Khalil E. M. and Vassilev A. (2000) HATs off: Selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF. Mol. Cell **5:** 539–595