AD_____

AWARD NUMBER DAMD17-97-1-7206

TITLE: P/CAF Function in Transcriptional Activation by Steroid Hormone Receptors and Mammary Cell Proliferation

PRINCIPAL INVESTIGATOR: Sharon Y. Roth, Ph.D.

CONTRACTING ORGANIZATION: University of Texas M.D. Anderson Cancer Center
Houston, Texas 77030

REPORT DATE: July 1998

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1998	Annual (1 Jul 97 - 30 Jun 98)

4. TITLE AND SUBTITLE

P/CAF Function in Transcriptional Activation by Steroid Hormone Receptors and Mammary Cell Proliferation

5. FUNDING NUMBERS

DAMD17-97-1-7206

6. AUTHOR(S)

Sharon Y. Roth, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Texas M.D. Anderson Cancer Center
Houston, Texas 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research And Materiel Command
ATTN: MCMR-RMI-S
504 Scott Street
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Inappropriate gene expression and abnormal progression of the cell cycle are intimately linked to oncogenesis and tumor formation. The purpose of this proposal is to examine the role of the PCAF histone acetyltransferase as a co-factor for steroid hormone receptors in the regulation of gene expression. Five specific aims define the scope of our project: 1) to complete the isolation and cloning of mouse PCAF cDNA and genomic sequences 2) to characterize PCAF as a histone acetyltransferase 3) to examine PCAF interactions with specific steroid receptors 4) To examine whether acetyltransferase activity is required for transcriptional activation by steroid hormone receptors 5) to create mice deficient for PCAF and mice that overexpress this factor to assess the role of PCAF in mouse development, differentiation, or tumorigenesis. In the past year we have largely completed the first and second aims of our project and have made good progress towards experiments described in the fifth specific aim as well. We have also discovered that another protein, GCN5 is highly similar to PCAF in both structure and acetyltransferase functions. We conclude that both proteins are important transcriptional regulators and will expand our aims to include studies of GCN5 as well as PCAF in steroid hormone responses.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer, histone, acetylase, chromatin, steroid receptor, CBP	54
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

SR In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

SR In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

SR In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

SR In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Sharon R Roth 7/1/98

PI - Signature Date

TABLE OF CONTENTS

Front Cover	1
Standard Form 298	2
Foreword	3
Introduction	5
Body	6
Experimental Methods	6
Results and Discussion	8
Statement of Work	10
Conclusions	10
References	10
Appendix	13
Figure Legends	14
Figures	17
Xu et al. Manuscript	23

INTRODUCTION

The mitogenic effects of steroid hormones are clinically important to the proliferative stimulation of mammary and uterine tissues by estrogens and progesterone, as well as to the over-proliferation of hormone responsive tumors (Beato et al., 1995). Growth stimulation by steroid hormone receptors is directly linked to the activation of genes involved in control of the cell cycle (Sicinski et al., 1995)(Kamei et al., 1996). Hormone binding to these receptors causes their translocation to the cell nucleus, where they bind to response elements upstream of target genes and subsequently activate their expression. Several co-factors interact with the steroid hormone receptors and are required for gene activation (Kamei et al., 1996). These 'co-activators' include factors such as the CREB binding protein (CBP) and a closely related protein, called p300. CBP/p300, in turn, interact with still more proteins, including SRC-1 (Steroid receptor Cofactor 1) (Onate et al., 1995)and PCAF (p300/CBP Associated Factor) (Yang et al., 1996). All of these co-factors are important for steroid hormone responses and surprisingly, all have been shown recently to possess histone acetyltransferase activity (Yang et al., 1996)(Spencer et al., 1997).

Histones are highly basic, nuclear proteins that serve as the building blocks for nucleosomes, the fundamental repeat unit for the compaction of DNA into chromatin within the eukaryotic nucleus (van Holde, 1989). Nucleosomes consist of two molecules of each of the 4 core histones, H2A, H2B, H3 and H4, which self associate into an octamer around which DNA is spooled in roughly two turns. Nucleosome-nucleosome interactions further compact the DNA into higher order structures.

Both single nucleosomes and higher order chromatin structures inhibit access of other proteins, including transcription factors, to the DNA (Edmondson and Roth, 1996). Post-translational acetylation of the histones is associated with unfolding of the chromatin and alterations in DNA-histone interactions which increases factor access, and thus facilitates transcription (Wolffe and Pruss, 1996). Acetylation occurs on highly conserved lysines located in the amino-terminal 'tail' domains of the histones. The discovery that a transcriptional co-factor in yeast called GCN5, and its homolog in Tetrahymena (p55), is a histone acetyltransferase provided a direct molecular link between histone acetylation and transcriptional activation (Brownell et al., 1996). The subsequent discoveries that the mammalian proteins PCAF (Yang et al., 1996), which is very similar in sequence to GCN5, and CBP/p300 (Bannister and Kouzarides, 1996)(Ogryzko et al., 1996)are also histone acetyltransferases indicated that this link is conserved across evolution and that histone acetylation is a fundamental point of transcriptional regulation.

This 'IDEA' proposal is focused on understanding the role of the histone acetyltransferase PCAF in steroid hormone responses in normal mammary cell development and in abnormal events that lead to cancer. Our original Specific Aims were 1) to complete the isolation and cloning of mouse PCAF cDNA and genomic sequences 2) to characterize murine PCAF as a histone acetyltransferase 3) to examine PCAF interactions with specific steroid receptors 4) To examine whether acetyltransferase activity is required for transcriptional activation by steroid hormone receptors 5) to create mice deficient for PCAF and mice that overexpress this factor to assess the

role of PCAF in mouse development, differentiation, or tumorigenesis. Our progress towards these goals is described below.

BODY

Experimental methods, assumptions, and procedures.

cDNA Library Screening.

Nested PCR using degenerate oligomers and a mouse embryonic cDNA library (d13.5) as template was performed to generate a fragment of the mouse GCN5 cDNA. Oligomers were chosen from regions of sequence conserved between yeast and *Tetrahymena*, which correspond to amino acids 131-244 of the yeast protein sequence. A single band of 123 bp was generated and cloned into pBluescript (Stratagene). Sequencing revealed 80% nucleotide identity, and 94% identity at the amino acid level to the reported human GCN5. This PCR product and human EST clones (IMAGE clone ID#243927) with similarity to GCN5 were used together to screen a cDNA library under conditions of low stringency as previously described (Edmondson and Olson, 1989). Clones were plaque purified and rescued as per the manufacturers protocol. Sequencing revealed two types of clones, some with similarity to hsGCN5 and some with similarity to *hsP/CAF*. All the P/CAF clones contained only a short piece of P/CAF and re-screening of the library failed to isolate any longer clones. Therefore, an oligomer corresponding to the 5' most sequence of mmP/CAF was used to screen a d10.5 embryonic mouse plasmid library using GeneTrapper™ technology(Gibco BRL). Additional clones, corresponding to full length P/CAF sequences, were isolated according to the manufacturers protocol.

Genomic Library Screening.

A mouse genomic library, Lambda FIXII (Stratagene), was screened using a mixture of a 5' fragment of the *mmGCN5* cDNA and a 5' fragment of the *mmP/CAF* cDNA. Positive plaques were picked and subjected to secondary screening. Phage DNA was prepared from positive plaques according to standard procedures. Genomic inserts were released from phage DNA by NotI digestion, and subsequently subcloned into Bluescript KS(+) (Strategene).

Sequencing Analysis.

DNA sequencing was performed using the thermo-sequenase radiolabeled terminator cycle sequencing kit (Amersham LIFE Science). Sequencing amplification conditions were 94°C 30 seconds, 55°C 30 seconds and 72°C 1 minute for 40 cycles. Alternatively, automated sequencing was carried out by the sequencing core facility at the M.D. Anderson Cancer Center.

Sequence Alignment.

Published sequences were obtained by searching GenBank, PIR-Protein, and SWISS-PROT databases. Sequence alignment was carried out using the GCG (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc.) pileup program. Percentage identity between two proteins was calculated by using the GCG bestfit program.

Linkage Analysis Mapping.

Restriction fragment length polymorphisms (RFLPs) for *mmGCN5* or *mmP/CAF* in C57BL/6J and SPRET/Ei subspecies were determined using genomic DNA purchased from the Jackson Laboratory. The Jackson Laboratory interspecific backcross panel (C57BL/6JEi x SPRET/Ei)F1 x SPRET/Ei known as Jackson BSS (Rowe et al., 1994) was then used to map the chromosomal

locations of the mouse *GCN5* and *P/CAF* genes. Predigested panels (BglII for *P/CAF* or XbaI for *GCN5*) were analyzed by Southern blot using a GCN5 or P/CAF intronic probe. Typing results were processed via the Jackson Laboratory database analysis (see http://www.jax.org/resources/documents/cmdata for raw data).

RNA Analysis.

Total RNA from adult mouse tissues or whole embryos (13.5 dpc) was extracted as described (Chomcznski and Sacchi, 1987). RNAs were electrophoresed on a 1.1% agarose gel containing formaldehyde along with RNA molecular weight markers (Gibco BRL). RNA was transferred to GeneScreen Plus membrane (NEN Life Science) in 10XSSC. Hybridization was carried out using *mmGCN5* and *P/CAF* specific probes.

GCN5 Protein Analysis

12.5 dpc mouse embryos were homogenized in Ripa buffer (1XPBS, 1%NP40, 0.5% sodium deoxycholate, 0.1% SDS, 100μg/ml PMSF, 1μg/ml aprotinin), and then centrifuged at 15,000xg for 20 minutes at 4 °C. Supernatant was collected for Western blots and GCN5 was immunoprecipitated using the polyclonal human GCN5 antibody (generously provided by Shelly Berger, Wistar Inst.) following the protocol of Santa Cruz Biotech, Inc. Hela cell nuclear extract was kindly provided by Warren Liao and Yongsheng Ren (M.D. Anderson Cancer Center).

Cloning And Expression Of Full-Length Mouse GCN5 And P/CAF.

Comparison of the mouse *GCN5* genomic and cDNA clones revealed that the isolated cDNA lacks the sequences encoding the first 74 amino acids. These sequences (which lack introns) were excised from the *GCN5* genomic clone by NcoI and BssHII digestion and inserted into the appropriate position of the cDNA clone to generate a full-length mouse *GCN5* cDNA, as verified by DNA sequencing. Full-length *mmGCN5* was subcloned into the NcoI and HindIII sites of the pRSETB vectors (Invitrogen), such that an N-terminal 6Xhis-tag was fused in-frame with the coding region. Similarly, full-length *mmP/CAF* was subcloned into the BamHI and KpnI sites of the pRSETB vector. 6XHis-tagged proteins were induced in BL21-AD bacterial cells by addition of 1 mM IPTG. Recombinant protein was purified using the Nickel-NTA resin (Qiagen) according to manufacturer's protocol. Purified recombinant proteins were verified by western blot analysis using an antibody specific to the 6XHis-tag (Clonetech).

Acetyltransferase Assays.

Acetyltransferase assays were performed as previously described (Brownell and Allis, 1995)(Brownell et al., 1996). HeLa cell mononucleosomes or core histones were the kind gift of Jerry Workman, and the cysteine-linked peptides (corresponding to amino acids 1-20 of H3, or this same region with substitution of acetyl lysine at positions 9 and 14) were the gift of C. David Allis. Calf thymus histones were purchased from Worthington Biochemical Corporation (Freehold, NJ). Acetylation assays were performed in 10 to 30 μl volumes using either 10 μg of histones or the indicated amount of synthetic peptide. Following incubation at 30°C for 30 minutes, an aliquot of each reaction was processed for liquid scintillation counting (P81 filter assay) according to Brownell et al (Brownell et al., 1996) and when appropriate, another aliquot was electrophoresed on a 22% protein gel SDS-PAGE gel and histones were visualized by fluorography and Coomassie blue staining.

Results and Discussion

Our progress towards each of our specific aims is discussed below, followed by a brief description of additional findings resulting from our studies this year. Figures and figure legends are included in the appendix.

Specific Aim 1: to complete the isolation and cloning of mouse PCAF cDNA and genomic sequences. We successfully isolated PCAF cDNA sequences and sequences corresponding to a PCAF pseudogene. Using a fragment of the mouse PCAF cDNA as probe, we identified multiple clones from a library of mouse genomic sequences that contained PCAF sequences. These clones indicate that the mouse PCAF gene contains very large introns (16-20 kb). Because of these large introns, we have not completed cloning of the mouse PCAF genomic sequences, but were able to clone sufficient sequences to create a disruption vector for the mouse PCAF gene, which we are using to generate PCAF deficient ('knock out') mice.

In the course of these screenings, we also isolated cDNA and genomic sequences corresponding to mouse GCN5. To our surprise, we discovered that the mouse gene is significantly longer than that previously reported for human cells, and that mouse GCN5 contains an extended amino terminal domain that is highly homologous (~70% identical) to the amino terminal domain of PCAF (Fig. 1). We went to show that this 'long' form of GCN5 exists both in mouse and human cells (Fig 2). Given the similarities between GCN5 and PCAF in structure and histone acetylase activity (see below), we have now expanded our studies in order to compare the possible functions of these proteins.

Specific Aim 2: to characterize murine PCAF as a histone acetyltransferase. We expressed murine PCAF and GCN5 (both the long and short forms) in bacteria and determined the substrate specificity of these enzymes in vitro. We found that both PCAF and GCN5 acetylate free histones H3 and H4, and that PCAF and the long form of GCN5 can also acetylate these histones when incorporated into a nucleosome (Fig. 3). However, the short form of GCN5 that lacks the extended amino terminal domain can not recognize nucleosomal substrates. Thus the amino terminal domains of long GCN5 and PCAF may be important to the recognition of chromatin substrates. Using peptides corresponding to the amino terminal domains of H3 or H4, we determined that GCN5 and PCAF acetylate identical sites in these histones, specifically K14 in H3 and K8 in H4 (Fig. 4).

Specific Aim 3: to examine PCAF interactions with specific steroid receptors. We have not yet begun these experiments, although other labs have demonstrated an involvement of acetyltransferases in hormonal responses and that the acetylase activity of these proteins is an important part of gene activation (Chen et al., 1997)(Spencer et al., 1997). We anticipate beginning these experiments in the next year, expanding our studies to include GCN5 as well as PCAF.

Specific Aim 4: **to examine whether acetyltransferase activity is required for transcriptional activation by steroid hormone receptors.** We have not yet begun these experiments, but as for Sp. Aim 3, we anticipate beginning them in the coming year and expanding the scope of the experiments to include GCN5 as well.

Specific Aim 5: **to create mice deficient for PCAF and mice that overexpress this factor to assess the role of PCAF in mouse development, differentiation, or tumorigenesis.** We have successfully targeted a PCAF 'knock out' allele in embryonic stem (ES) cells and have generated chimeric mice from blastocysts injected with these ES cells. We have bred the chimeric mice to generate mice that are heterozygous for the null allele, and are currently breeding these to homozygosity. We will then examine the homozygous mice for developmental defects or increased incidence of tumors. We are also creating GCN5 deficient mice, and are currently breeding mice chimeric for the GCN5 knock out allele to generate heterozygotes, which we will then breed to obtain homozygotes. In the next year, we will concentrate on examining the consequences of these 'knock outs' and if both PCAF and GCN5 null mice are viable, we will breed them together to form the double mutants.

Additional Findings: Our discovery that mouse GCN5 and PCAF are very similar in structure raises the question of whether they provide redundant functions for the cell. To begin to address this question, we have examined the expression of these genes in various mouse tissues. We find that while they are both ubiquitously expressed, they are present in inverse ratios in most tissues (Fig. 5). That is, tissues such as muscle that express high levels of PCAF express relatively low levels of GCN5, whereas embryonic tissues that express high levels of GCN5 express very low levels of PCAF. These differences in expression suggest that GCN5 and PCAF may serve some tissue specific functions.

PCAF was originally identified via its interactions with CBP and p300 (Yang et al., 1996). We therefore asked whether the long form of GCN5 also interacted with these factors, and found that it does, at least in vitro (Fig. 6). It will be important to determine whether GCN5 and PCAF both function with CBP/p300 in gene activation in vivo. Experiments to address this question will be performed together with those of Sp. Aim 4 above, which deal with the requirement of acetyltransferase activity for gene activation by the steroid receptors.

Finally, we also mapped the chromosomal locations of mouse GCN5 and PCAF. Interestingly, we found that GCN5 cosegregated tightly with BRCA1 (1 cM interval) on chromosome 11. This region is syntenic with human chromosome 17. Interestingly, loss of heterozygosity on human chromosome 17 is a frequent genetic alteration in sporadic breast and ovarian cancers, where mutations in BRCA1 and BRCA2 are rarely found (Niederacher et al., 1997)(Tangir et al., 1996).

NOTE: A paper describing our cloning of mouse GCN5 and PCAF and all of the above findings has recently been accepted for publication in *Molecular Cellular Biology*. A copy of this manuscript (Xu et al) is included in the Appendix.

Recommendations in Relation to the Statement of Work

We have completed all the items outlined in 'Year 1' of our original statement of work, except for the initiation of receptor-P/CAF binding studies and the cloning of a MMTV-PCAF transgene. We have also made significant progress on items listed under 'Year 2', especially in regards to generating PCAF 'knock out' mice. We are on schedule to complete the remainder of experiments outlined in years 1 and 2, and data obtained in the coming year will determine the most pertinent experiments to be completed in year 3. In addition, as described above, our experimental aims have been expanded to include studies of GCN5 as well as PCAF. These studies will enhance the significance of our findings.

CONCLUSIONS

We have made significant progress towards our Specific aims and are now poised to address the importance of PCAF and GCN5 in breast cell development and tumorigenesis. Our discovery that mammalian GCN5 is larger than originally reported is very significant, and raises important questions regarding the relative functions of these two acetyltransferases. In the next two years of this funding period, we should resolve many of these questions and will determine the importance of these enzymes in regulation of cell growth and carcinogenesis. If , as we suspect, these enzymes are important to these processes, they may provide novel diagnostic tools or future therapeutic targets for treatment of breast cancer and other diseases.

REFERENCES

Bannister, A. J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641-643.

Beato, M., Herlich, P., and Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. Cell *83*, 851-857.

Brownell, J. E., and Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proceedings of the National Academy of Sciences of the United States of America *92*, 6364-8.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-51.

Chen, H., Lin, R. J., Schiltz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatain, Y., and Evans, R. M. (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. Cell *90*, 569-580.

Chomcznski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry *162*, 156-159.

Edmondson, D. G., and Olson, E. N. (1989). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes and Dev. *3*, 628-640.

Edmondson, D. G., and Roth, S. Y. (1996). Chromatin and transcription. FASEB Journal *10*, 1173-82.

Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S., Heyman, R. A., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. Cell *85*, 403-414.

Niederacher, D., Picard, F., Roeyen, C. V., An, H., Bender, H. G., and Beckmann, M. W. (1997). Patterns of allelic loss on chromosome 17 in sporadic beast carcinomas detected by fluorescent-labeled microsatellite analysis. Genes, Chromosomes, and Cancer *18*, 181-192.

Ogryzko, V. V., Schlitz, R. L., Russanova, V., Howard, B. H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953-959.

Onate, S. A., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. Science *270*, 1354-1357.

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. Mammalian genome *5*, 253-274.

Sicinski, P., Donaher, J. L., Parker, S. B., Li, T., Fazeli, A., Gardner, H., Haslam, S. Z., Bronson, R. T., Elledge, S. J., and Weinberg, R. A. (1995). Cyclin D1 provides a link between development and oncogenesis in the retina and the breast. Cell *82*, 621-630.

Spencer, T. E., Jenster, G., Burcin, M. M., Allis, C. D., Zhou, J., Mizzen, C. A., McKenna, N. J., Onate, S. A., Tsai, M. J., and O'Malley, B. W. (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. Nature *389*, 194-198.

Tangir, J., Muto, M. G., Berkowitz, R. S., Welch, W. R., Bell, D. A., and Mok, S. C. (1996). A 400 kb novel deletion unit centromeric to the BRCA1 gene in sporadic epithelial ovarian cancer. Oncogene *12*, 735-740.

van Holde, K. E. (1989). Chromatin, A. Rich, ed. (New York: Springer-Verlag).

Wolffe, A. P., and Pruss, D. (1996). Targeting chromatin disruption: Transcription regulators that acetylate histones. Cell *84*, 817-9.

Yang, X.-J., Ogryzko, V. V., Nishikawa, J.-I., Howard, B. H., and Nakatani, Y. (1996). A p300/CBP associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319-324.

Appendix

Legends to Figures 1-6

Figures 1-6

Copy of Xu et al (manuscript in press at *Molecular Cellular Biology*)

Figure Legends

FIG. 1. Alignment of mouse GCN5, mouse P/CAF, and the reported human GCN5 amino acid sequences.

Identical amino acids among mouse GCN5 (mmGCN5), human GCN5 (hsGCN5) and mouse P/CAF (mmPCAF) are shaded. Amino acid deletions are indicated with a dotted line. The locations of the HAT/acetyl CoA binding regions (38)(20)and the bromodomain motif (17) are indicated. The full bromodomain likely encompasses amino acids 363-472 of hsGCN5 (19).

FIG.2. Detection of both long and short GCN5 proteins.

A) Left panel: protein extracts were prepared from U20S cells or HeLa cell nuclei and probed with polyclonal antibodies to hsP/CAF or hsGCN5, as indicated. Right panel: Protein extracts were prepared from 12.5 dpc mouse embryos or Hela cells and probed with the hs GCN5 antibody.

B) GCN5 proteins were immunoprecipitated from 12.5 dpc mouse embryos, and then probed with hsGCN5 antibody. An unrelated, HirA polyclonal antibody was used as a negative control.

FIG. 3. Acetylation of nucleosomal histones by mouse GCN5 and P/CAF.

Acetyltransferase assays were performed using HeLa cell mononucleosomes or free histones (as indicated) and an aliquot of each assay was resolved on a 22% SDS-PAGE gel. Coomassie stained gels and corresponding autoradiographs are shown. In both assays, histones H3 and H4 were acetylated by the recombinant full-length mmGCN5.

FIG. 4. Acetylation of histone H3 synthetic peptides by mmGCN5 and mmP/CAF.

Panels A and B show the results of acetyltransferase assays using recombinant full-length mmGCN5 and synthetic peptides corresponding to the amino-terminal "tail" of histone H3. Panels C and D show results of peptide assays using recombinant mmP/CAF. Peptides were either "unacetylated" or synthesized with acetyl-lysine residues at either K9 and K18 (di 9,18) or K9 and K14 (di9,14) in H3. "Vector" indicates assay of a control extract, made from bacteria transformed by the recombinant vector without an acetyltransferase insert, subjected to the His-tag purification procedure.

FIG. 5. Ubiquitous and complementary expression of mouse GCN5 and P/CAF.

Top panel: Total RNA was isolated from various mouse tissues and embryos as indicated. Northern blot hybridization was performed using a mixture of mouse GCN5 and P/CAF cDNA probes. Two transcripts were detected and indicated by arrows on the right. The identities of the transcripts were confirmed by Northern blot hybridization using single GCN5 or P/CAF probes (data not shown). RNA molecular weight standards are shown on the left. Bottom panel: ethidium-bromide stained gel showing 18S ribosomal RNAs.

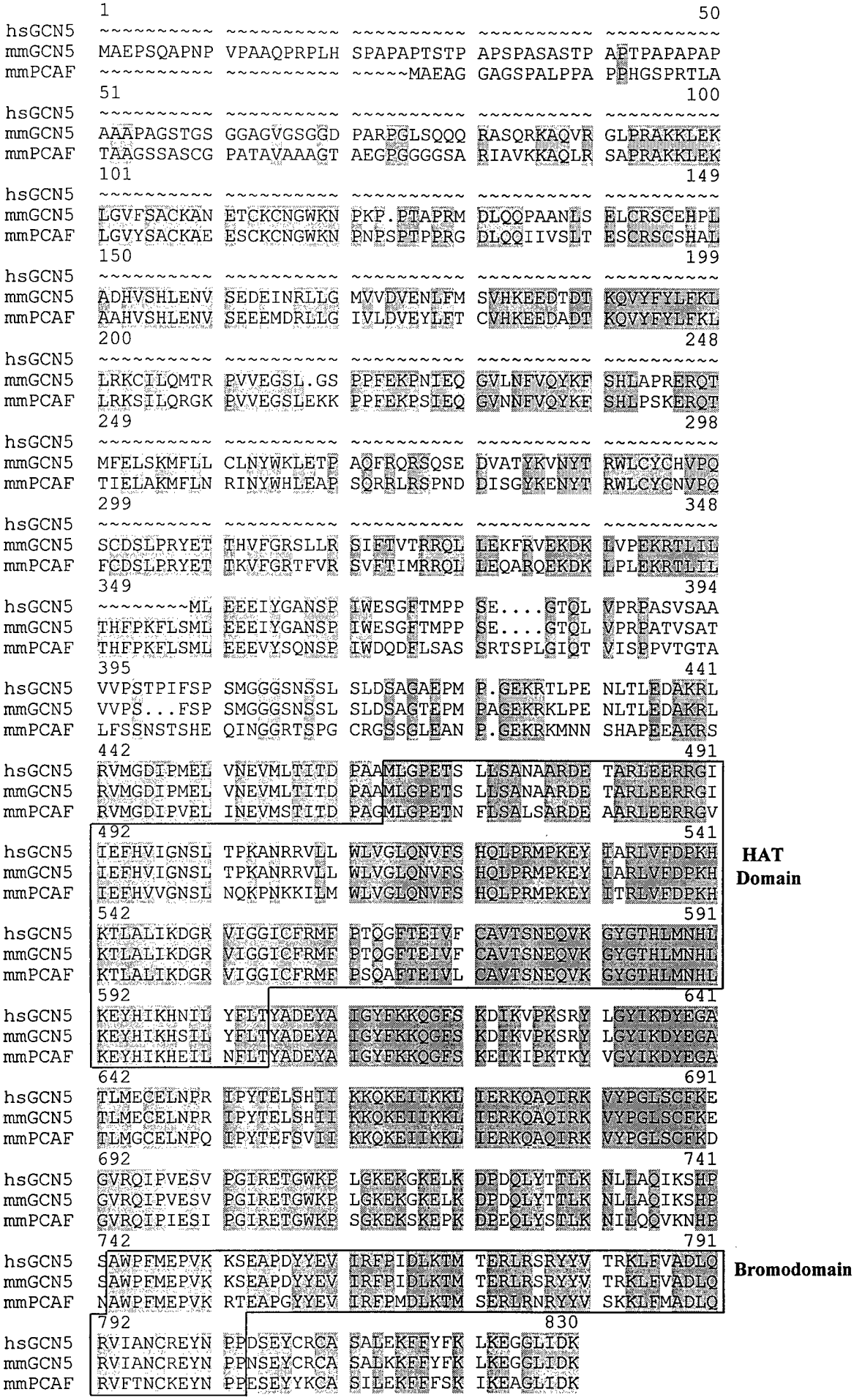
FIG 6. mmGCN5 and mmP/CAF both interact with CBP and p300 in vitro.

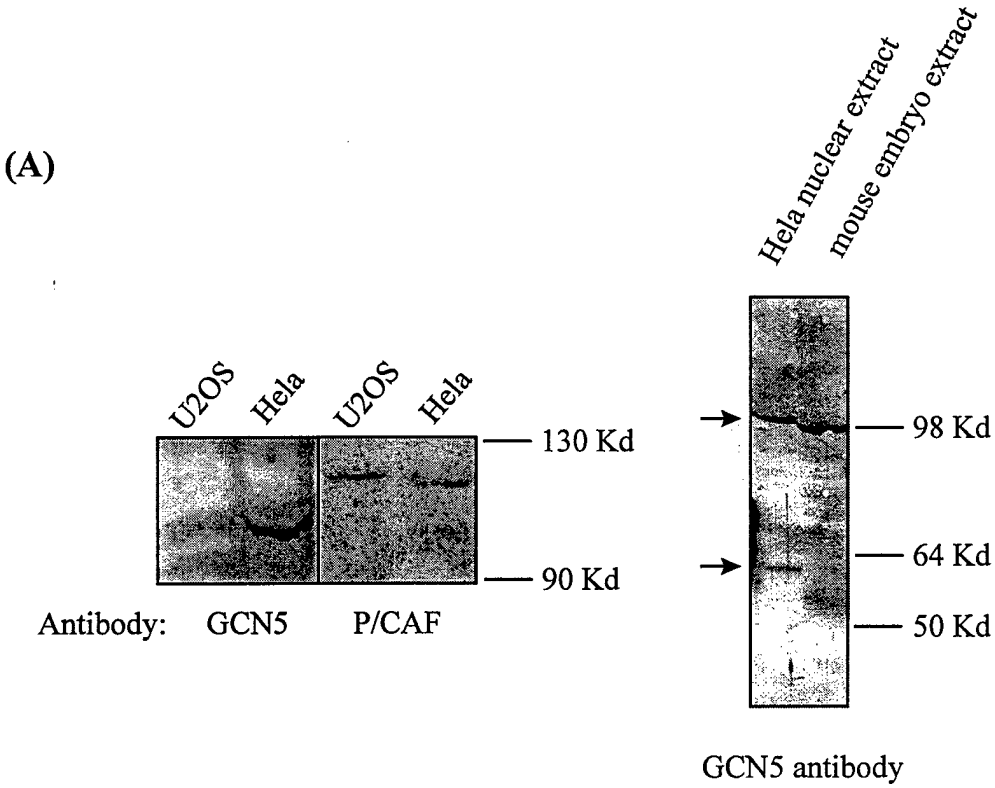
A) Fragments of CBP fused to GST that were used for the interaction assays in (C) are indicated. These fragments span the region of homology to ADA2 and extend into the transactivation domain of CBP (41).

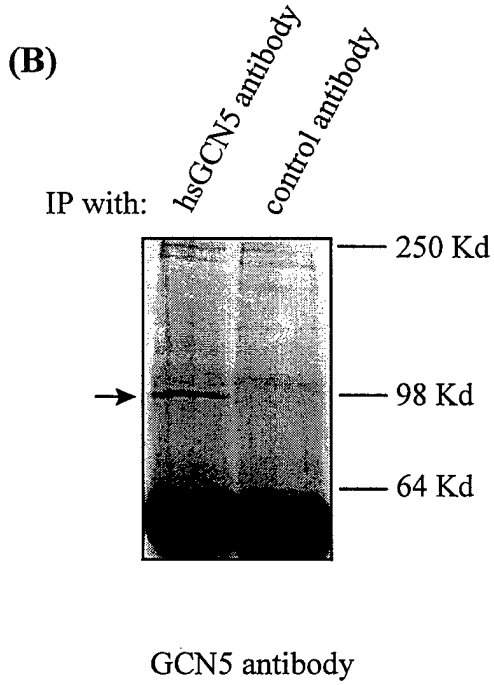
B) Recombinant, His-tagged proteins used in the interaction assays of (C) were resolved by SDS PAGE and probed with an antibody specific for the His tag. The amounts shown represent 0.25% or 1% of the protein used in the assay, as indicated.

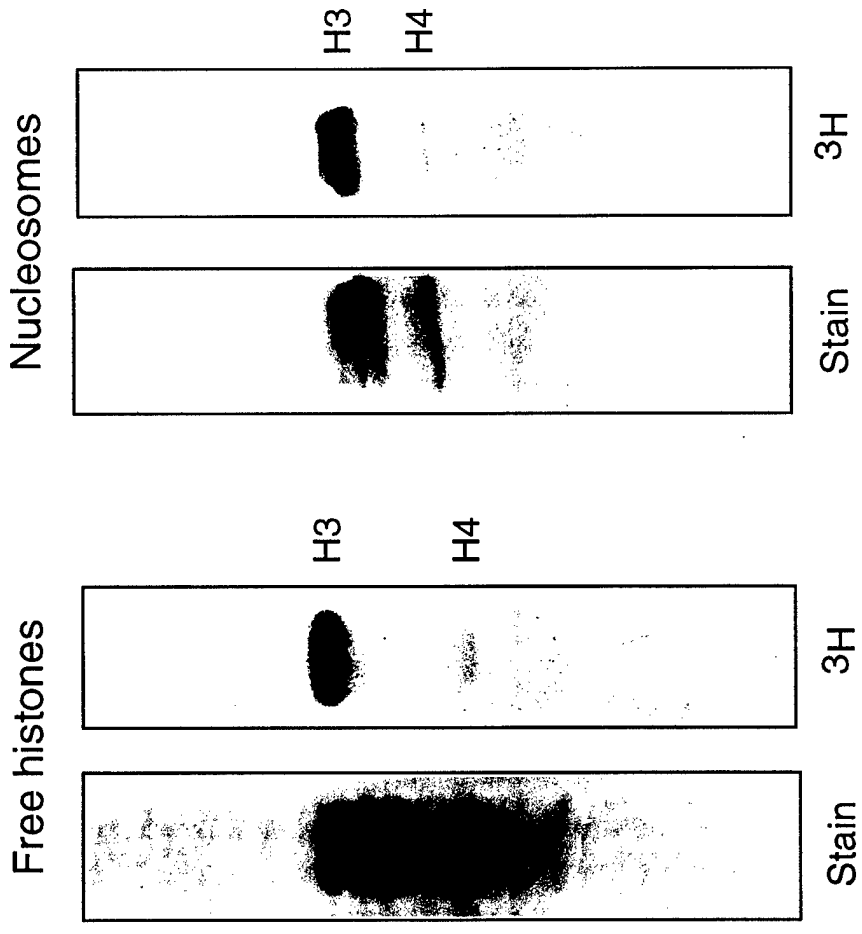
C) GST-CBP or p300 fusion proteins were mixed with crude bacterial lysates containing P/CAF, GCN5, or HIR A proteins. Interacting proteins were recovered using glutathione -sepharose and detected on

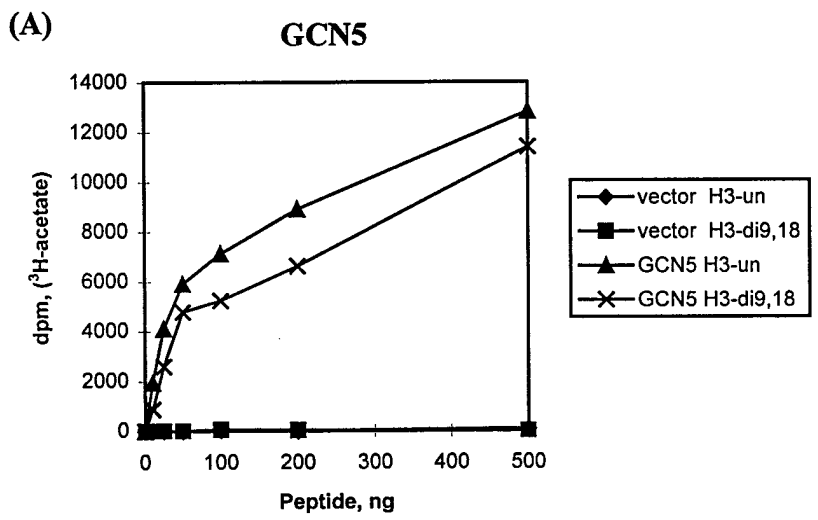
western blots with the anti-His tag antibody. The p300 B' fragment is homologous to the CBP B fragment, and the CBP ΔB fragment is missing residues 1801-1851 (41).

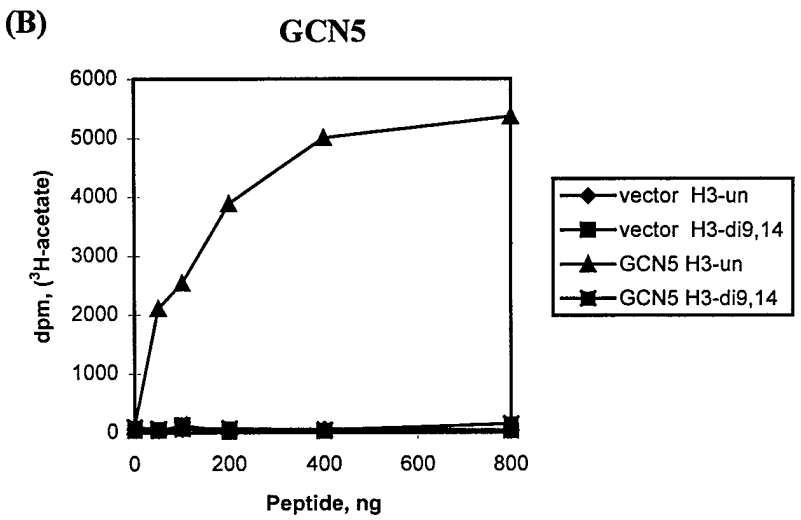

HAT Domain
Bromodomain

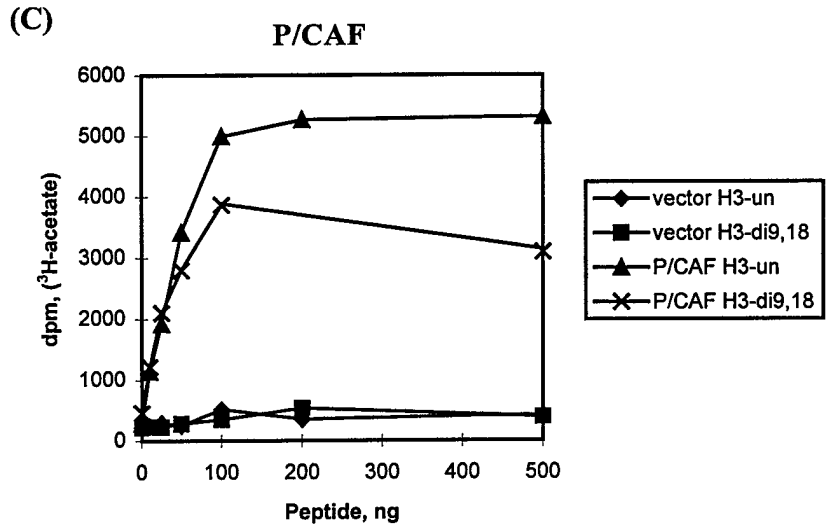

(A)
U2OS
Hela
U2OS
Hela
130 Kd
90 Kd
Antibody: GCN5 P/CAF
Hela nuclear extract
mouse embryo extract
98 Kd
64 Kd
50 Kd
GCN5 antibody

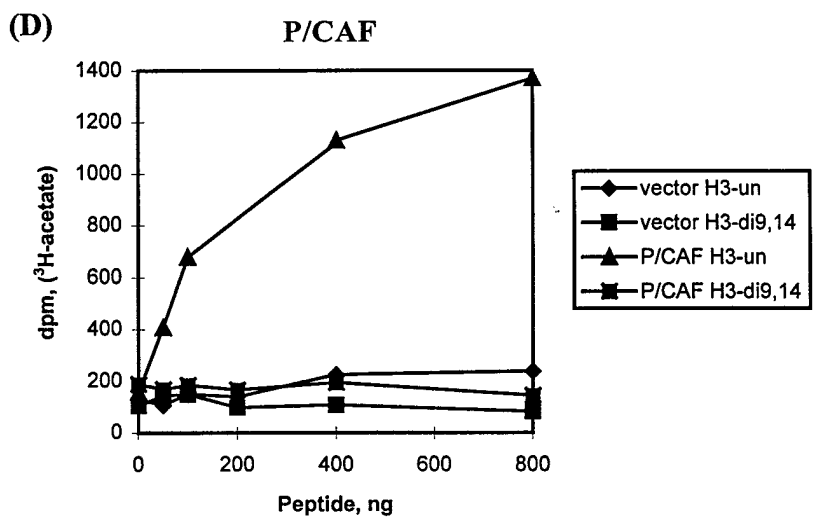

(B)
IP with: hsGCN5 antibody
control antibody
250 Kd
98 Kd
64 Kd
GCN5 antibody


Nucleosomes
H3
H4
3H
Stain
Free histones
H3
H4
3H
Stain


(A)
GCN5
dpm, (^{3}H-acetate)
Peptide, ng
vector H3-un
vector H3-di9,18
GCN5 H3-un
GCN5 H3-di9,18

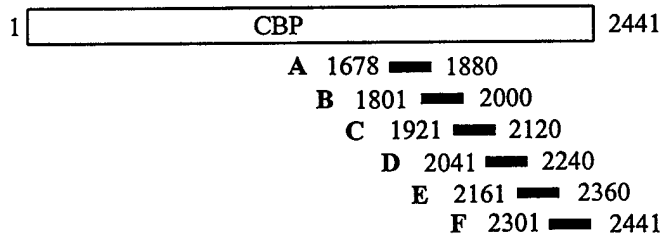

(B)
GCN5
dpm, (^{3}H-acetate)
Peptide, ng
vector H3-un
vector H3-di9,14
GCN5 H3-un
GCN5 H3-di9,14


(C)
P/CAF
dpm, (^{3}H-acetate)
Peptide, ng
vector H3-un
vector H3-di9,18
P/CAF H3-un
P/CAF H3-di9,18

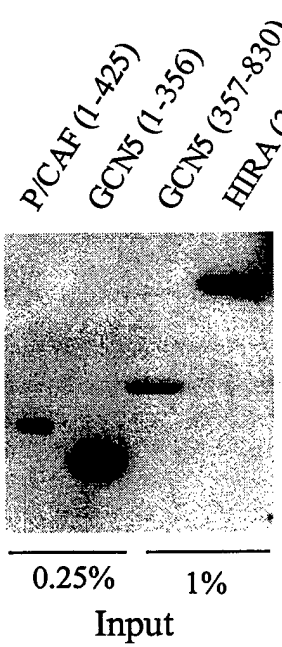

(D)
P/CAF
dpm, (^{3}H-acetate)
Peptide, ng
vector H3-un
vector H3-di9,14
P/CAF H3-un
P/CAF H3-di9,14

P/CAF
GCN5
18S rRNA
13.5 dpc embryo
brain
eye
heart
lung
liver
kidney
thymus
spleen
small intestine
skeletal muscle
ovary
testis
7.4
5.3
2.8

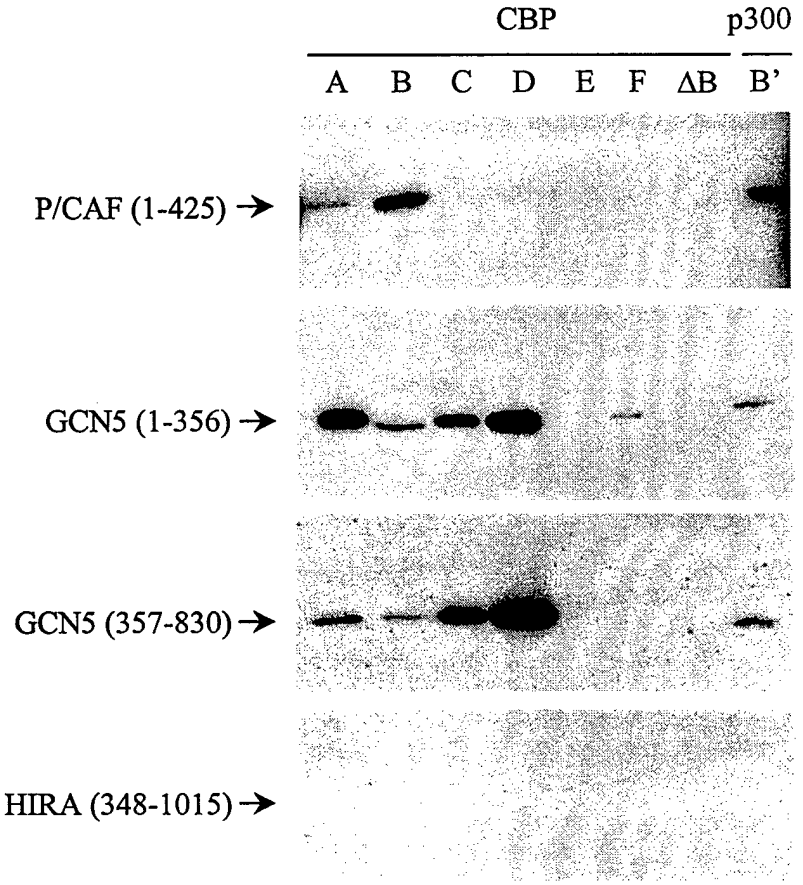
(A)


1
CBP
2441
A 1678 1880
B 1801 2000
C 1921 2120
D 2041 2240
E 2161 2360
F 2301 2441

(B)


P/CAF (1-425)
GCN5 (1-356)
GCN5 (357-830)
HIRA (348-1015)
0.25%
1%
Input

(C)


CBP
p300
A B C D E F ΔB B'
P/CAF (1-425)
GCN5 (1-356)
GCN5 (357-830)
HIRA (348-1015)

Mammalian GCN5 and P/CAF Acetyltransferases Share Homologous Amino-terminal Domains Important for the Recognition of Nucleosomal Substrates

Wanting Xu[1], Diane G. Edmondson[1], and Sharon Y. Roth*

Department of Biochemistry and Molecular Biology

University of Texas M. D. Anderson Cancer Center

Houston, Texas 77030

[1]These authors contributed equally to this work.

*Corresponding Author

Key words: chromatin, histones, acetylation, transcription, Gcn5, P/CAF

Abstract

The yeast transcriptional adaptor protein, Gcn5p, serves as a histone acetyltransferase, directly linking chromatin modification to transcriptional regulation. Two human homologs of Gcn5p have been reported previously, hsGCN5 and hsP/CAF (p300/CBP associated factor). While hsGCN5 was predicted to encode a protein close to the size of the yeast acetyltransferase, hsP/CAF encoded a protein containing an additional 356 amino-terminal residues of unknown function. Surprisingly, we have found that in mouse, both the *GCN5* and the *P/CAF* genes encode proteins containing this extended amino-terminal domain. Moreover, while a shorter version of GCN5 might be generated upon alternative or incomplete splicing of a longer transcript, mRNAs encoding the longer protein are much more prevalent in both mouse and human cells, and larger proteins are detected by GCN5 specific antisera in both mouse and human cell extracts. Mouse *GCN5* and *P/CAF* genes are ubiquitously expressed, but maximum expression levels are found in different, complementary sets of tissues. Both mmP/CAF and mmGCN5 interact with CBP/p300. Interestingly, *mmGCN5* maps to chromosome 11 and cosegregates with *BRCA1,* and *mmP/CAF* maps to a central region of chromosome 17. As expected, recombinant mmGCN5 and mmP/CAF both exhibit histone acetyltransferase activity *in vitro* with similar substrate specificities. However in contrast to yeast Gcn5p and the previously reported shorter form of hsGCN5, mmGCN5 readily acetylates nucleosomal substrates as well as free core histones. Thus, the unique amino-terminal domains of mammalian P/CAF and GCN5 may provide additional functions important to recognition of chromatin substrates and the regulation of gene expression.

Introduction

Transcription is a complex process requiring the coordinate action of multiple basal and transactivating proteins. In eukaryotic cells, this process is complicated further by the packaging of DNA into chromatin. Nucleosomes provide the fundamental repeat unit of chromatin, consisting of two molecules of each of the four core histones (H2A, H2B, H3 and H4) and ~146 bp of DNA wound in almost two turns around the exterior of the histone octamer (37). Individual nucleosomes as well as more highly folded structures are generally inhibitory to the initiation of transcription. Alterations in nucleosomal structure and in chromatin packing often accompany transcriptional activation (12).

Post-translational acetylation of the histones has long been correlated with transcriptional activation (36, 39, 40). Acetylation neutralizes the charge associated with epsilon amino groups of lysine residues, thereby loosening contacts between the histones and the negatively charged DNA. Histone acetylation also influences compaction of nucleosomal arrays, yielding less condensed chromatin structures (16). Both of these effects can increase transactivator binding to nucleosomal DNA, facilitating transcriptional activation.

A molecular basis for the linkage between histone acetylation and gene activation was provided by the discovery that the yeast transcriptional adaptor protein, Gcn5p, serves as the catalytic subunit of a histone acetyltransferase type A (HAT A) activity (5). Gcn5p is associated with two multisubunit complexes in yeast, which include Ada proteins (Ada2p, Ada3p, and Ada5p) and/or certain Spt proteins (6, 14, 18, 24, 25, 31). These complexes are required for transcriptional activation by particular transactivators, including heterologous VP16 derivatives and endogenous Gcn4p (3, 13, 24, 34). Components of the Gcn5p/Adap complex contact both

transactivator proteins and basal transcription proteins, thus providing an adaptor or coactivator function, in addition to histone acetyltransferase activity (2, 34). Association with both Ada2p and acetyltransferase activity is required for Gcn5p function *in vivo* (8, 38).

Human homologs of *GCN5* have been cloned based on sequence and functional similarities to the yeast protein. A cDNA predicted to encode a protein of similar size and overall homology to yeast Gcn5p has been described (7, 41). A human *ADA2* gene has also been cloned, indicating a conservation of adaptor and histone acetyltransferase functions across species (7). In addition, a cDNA encoding a second, larger Gcn5-related protein has been identified that possesses unique sequences in its amino terminal half. This protein, P/CAF, associates with two highly related proteins, p300 and CBP, that share a region of homology with *ADA2* (41). Interestingly, p300 and CBP are also histone acetyltransferases (1, 29). Interactions between P/CAF and p300 or CBP are disrupted by the viral E1A oncogene product, and this disruption is required for cellular transformation by E1A (41). Proper association of these histone acetyltransferase activities, then, is extremely important for normal cell growth (32).

In order to further study the functions of histone acetyltransferases in the growth and development of mammalian cells, we endeavored to isolate sequences encoding mouse Gcn5p and P/CAF. To our surprise, although our mouse GCN5 clone exhibited 98% identity with the reported human GCN5 sequence, the mouse cDNA encoded an extended amino terminal domain with high similarity to a corresponding domain in P/CAF. Upon further examination, we found that the reported human GCN5 cDNA (41) may result from an incompletely spliced transcript, and that a more prevalent transcript exists that potentially encodes a longer human GCN5 protein similar to that encoded by the mouse cDNA that we isolated. Moreover, in contrast to previous reports that yeast and human GCN5 proteins only acetylate free core histones, the full-length recombinant mouse GCN5 protein containing this extended amino-terminal region acetylates both "free" and nucleosomal histones H3 and H4. These results suggest that this additional

domain in the mammalian GCN5 acetyltransferase facilitates chromatin recognition. Interestingly, P/CAF and GCN5 are expressed in inverse ratios in many mouse tissues, indicating these proteins may serve tissue specific functions.

Materials And Methods

cDNA Library Screening.

Nested PCR using degenerate oligomers and a mouse embryonic cDNA library (d13.5) as template was performed to generate a fragment of the mouse GCN5 cDNA. Oligomers were chosen from regions of sequence conserved between yeast and *Tetrahymena*, which correspond to amino acids 131-244 of the yeast protein sequence. A single band of 123 bp was generated and cloned into pBluescript (Stratagene). Sequencing revealed 80% nucleotide identity, and 94% identity at the amino acid level to the reported human GCN5. This PCR product and human EST clones (IMAGE clone ID#243927) with similarity to GCN5 were used together to screen a cDNA library under conditions of low stringency as previously described (11). Clones were plaque purified and rescued as per the manufacturers protocol. Sequencing revealed two types of clones, some with similarity to hsGCN5 and some with similarity to *hsP/CAF*. All the P/CAF clones contained only a short piece of P/CAF and re-screening of the library failed to isolate any longer clones. Therefore, an oligomer corresponding to the 5' most sequence of mmP/CAF was used to screen a d10.5 embryonic mouse plasmid library using GeneTrapper™ technology(Gibco BRL). Additional clones, corresponding to full length P/CAF sequences, were isolated according to the manufacturers protocol.

Genomic Library Screening.

A mouse genomic library, Lambda FIXII (Stratagene), was screened using a mixture of a 5' fragment of the *mmGCN5* cDNA and a 5' fragment of the *mmP/CAF* cDNA. Positive plaques were picked and subjected to secondary screening. Phage DNA was prepared from positive plaques according to standard procedures. Genomic inserts were released from phage DNA by NotI digestion, and subsequently subcloned into Bluescript KS(+) (Strategene).

Sequencing Analysis.

DNA sequencing was performed using the thermo-sequenase radiolabeled terminator cycle sequencing kit (Amersham LIFE Science). Sequencing amplification conditions were 94°C 30 seconds, 55°C 30 seconds and 72°C 1 minute for 40 cycles. Alternatively, automated sequencing was carried out by the sequencing core facility at the M.D. Anderson Cancer Center.

Sequence Alignment.

Published sequences were obtained by searching GenBank, PIR-Protein, and SWISS-PROT databases. Sequence alignment was carried out using the GCG (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc.) pileup program. Percentage identity between two proteins was calculated by using the GCG bestfit program.

Linkage Analysis Mapping.

Restriction fragment length polymorphisms (RFLPs) for *mmGCN5* or *mmP/CAF* in C57BL/6J and SPRET/Ei subspecies were determined using genomic DNA purchased from the Jackson Laboratory. The Jackson Laboratory interspecific backcross panel (C57BL/6JEi x SPRET/Ei)F1 x SPRET/Ei known as Jackson BSS (33) was then used to map the chromosomal locations of the mouse *GCN5* and *P/CAF* genes. Predigested panels (BglII for *P/CAF* or XbaI for *GCN5*) were analyzed by Southern blot using a GCN5 or P/CAF intronic probe. Typing results were

processed via the Jackson Laboratory database analysis (see http://www.jax.org/resources/documents/cmdata for raw data).

Reverse Transcriptase PCR (RT-PCR).

Isolation of total RNA from various mouse tissues was performed as described (10). RNA was digested with RNaseA-free DNase I (Ambion) for 30 minutes at 37°C. RT-PCR was performed using the RT-PCR kit (Perkin Elmer) according to manufacturer's protocols. Reverse transcription was carried out at 42°C for 15 min followed by heating at 95°C for 5 minutes. PCR reactions were carried out at 95°C 60 seconds and 60°C 60 seconds for 35 cycles as suggested by the manufacturer. Primer A (see figure 3B for sequence location) for RT-PCR is CTGGTGCCTGAGAAGAGGAC; Primer B (see figure 3B) is CTCCGAAGGTGGCATGGTGAAG.

RNA Analysis.

Total RNA from adult mouse tissues or whole embryos (13.5 dpc) was extracted as described (10). RNAs were electrophoresed on a 1.1% agarose gel containing formaldehyde along with RNA molecular weight markers (Gibco BRL). RNA was transferred to GeneScreen Plus membrane (NEN Life Science) in 10XSSC. Hybridization was carried out using *mmGCN5* and *P/CAF* specific probes.

GCN5 Protein Analysis

12.5 dpc mouse embryos were homogenized in Ripa buffer (1XPBS, 1%NP40, 0.5% sodium deoxycholate, 0.1% SDS, 100μg/ml PMSF, 1μg/ml aprotinin), and then centrifuged at 15,000xg for 20 minutes at 4 °C. Supernatant was collected for Western blots and GCN5 was immunoprecipitated using the polyclonal human GCN5 antibody (generously provided by Shelly Berger, Wistar Inst.) following the protocol of Santa Cruz Biotech, Inc. Hela cell nuclear extract was kindly provided by Warren Liao and Yongsheng Ren (M.D. Anderson Cancer Center).

Cloning And Expression Of Full-Length Mouse GCN5 And P/CAF.

Comparison of the mouse *GCN5* genomic and cDNA clones revealed that the isolated cDNA lacks the sequences encoding the first 74 amino acids. These sequences (which lack introns) were excised from the *GCN5* genomic clone by NcoI and BssHII digestion and inserted into the appropriate position of the cDNA clone to generate a full-length mouse *GCN5* cDNA, as verified by DNA sequencing. Full-length *mmGCN5* was subcloned into the NcoI and HindIII sites of the pRSETB vectors (Invitrogen), such that an N-terminal 6Xhis-tag was fused in-frame with the coding region. Similarly, full-length *mmP/CAF* was subcloned into the BamHI and KpnI sites of the pRSETB vector. 6XHis-tagged proteins were induced in BL21-AD bacterial cells by addition of 1 mM IPTG. Recombinant protein was purified using the Nickel-NTA resin (Qiagen) according to manufacturer's protocol. Purified recombinant proteins were verified by western blot analysis using an antibody specific to the 6XHis-tag (Clonetech).

Acetyltransferase Assays.

Acetyltransferase assays were performed as previously described (4, 5). HeLa cell mononucleosomes or core histones were the kind gift of Jerry Workman, and the cysteine-linked peptides (corresponding to amino acids 1-20 of H3, or this same region with substitution of acetyl lysine at positions 9 and 14) were the gift of C. David Allis. Calf thymus histones were purchased from Worthington Biochemical Corporation (Freehold, NJ). Acetylation assays were performed in 10 to 30 µl volumes using either 10 µg of histones or the indicated amount of synthetic peptide. Following incubation at 30°C for 30 minutes, an aliquot of each reaction was processed for liquid scintillation counting (P81 filter assay) according to Brownell et al (5) and when appropriate, another aliquot was electrophoresed on a 22% protein gel SDS-PAGE gel and histones were visualized by fluorography and Coomassie blue staining.

GST-fusion protein interaction assays.

GST-CBP/p300 interaction assays were performed as described by Yang et al (41) except that crude bacterial lysates containing his-tagged recombinant P/CAF, GCN5, or HIRA were used and the interactions were detected by western blot using the His-tag antibody.

Results

Cloning of *mmGCN5*

In order to study the function of acetyltransferases in a mammalian system, we endeavored to clone mouse *GCN5* homologues. First we generated a fragment of the mouse *GCN5* cDNA using a nested PCR strategy employing degenerate primers homologous to conserved regions of the yeast *GCN5* and the *Tetrahymena* p55. To further enhance the probability of identifying GCN5 related sequences, this fragment was used together with a human GCN5 EST to screen a 13.5 dpc mouse embryonic cDNA library under conditions of low stringency (11). Multiple positive clones were identified and upon sequencing, these were found to contain open reading frames predicted to encode proteins with significant homology to either hsGCN5 or hsP/CAF (Fig. 1).

One cDNA clone contained an open reading frame of 756 amino acids and the C-terminal portion of this predicted amino acid sequence exhibited 98% identity with the reported hsGCN5 sequence, but only 71% homology to the hsP/CAF sequence, over the length of the predicted proteins (Fig. 2A). We tentatively concluded that this cDNA clone likely encodes the mouse homolog of *GCN5* (*mmGCN5*), as confirmed below.

We next used a fragment from the 5' end of this clone to screen a library of mouse genomic sequences. Three different clones were isolated, and restriction analysis and sequencing indicated that all three clones harbored the entire *mmGCN5* gene. Comparison of the

genomic and cDNA clones of mmGCN5 revealed that the cDNA clone isolated above actually lacked the first 74 amino terminal codons and that the *mmGCN5* gene is divided into 19 exons and contains relatively small (85 bp to 1Kb) introns (Fig. 3A). We inserted sequences from the genomic clone containing the missing amino terminal codons into the cDNA clone to generate a full length (830 amino acids), recombinant *mmGCN5* cDNA.

The two previously reported hsGCN5 sequences differ in the position of the initiating methionine, such that one reported sequence contains 49 additional amino-terminal amino acids relative to the other (7, 41). The mmGCN5 open reading frame also encodes these additional amino acids, but the open reading frame is further extended for some distance upstream of these sequences, potentially encoding 356 additional amino acids. The context of the predicted translation initiation site in this extended region of mmGCN5 matches well the Kozak consensus sequence (Fig. 2B) (21, 22). Moreover, the amino acids encoded in this amino terminal extension exhibit more than 66% identity to sequences in the corresponding region of both mouse (see below) and human P/CAF, and the length of this extended region is similar to that of the P/CAF proteins. These data indicate that *mmGCN5* encodes a protein that is very homologous to yeast Gcn5p and is almost identical to the previously reported hsGCN5, but that contains an extended N-terminal domain homologous to P/CAF in both size and sequence.

Incomplete splicing might yield a shorter GCN5 protein in mouse and human cells.

We were interested in determining the basis of the incongruity in size between *mmGCN5* and the reported human cDNA. Inspection of the *mmGCN5* genomic sequence revealed the presence of an intron (intron 6 in Fig. 3A) 10 bp upstream of the previously reported upstream-most hsGCN5 translation initiation site (41). Sequences highly similar (91% identical) to these intronic sequences are also present in the predicted 5' untranslated region of the reported hsGCN5 cDNA but are absent in the mouse cDNA we isolated above. These comparisons suggest either that the mouse and human GCN5 genes are subject to differential splicing events,

in which this intron is either removed (mouse) or retained (human), or that the previously identified human cDNA sequence is incomplete. Interestingly, a conserved, in frame stop codon is found near the beginning of intron 6, and retention of this intron would prevent translation of the larger protein in both mouse and human cells, perhaps yielding a smaller protein corresponding to that previously predicted for hsGCN5.

To investigate the possibility of alternative (or incomplete) splicing of mouse and human *GCN5* transcripts, we performed RT-PCR on total RNA isolated from human HeLa cells, human hepatoma cells, mouse kidney, mouse ovary, and a 13.5 dpc mouse embryo. All RNAs were treated with a RNAse free DNAseI before RT-PCR to remove any genomic DNA sequences from the samples. A *mmGCN5* genomic DNA clone was used in a separate reaction, as a positive control for the presence of the intron sequences. Two primers corresponding to conserved sequences in the exons 6 and 8, which flank the introns 6 and 7 (Fig. 3 A and B) were used for the amplification. The RT-PCR products were separated on an agarose gel, transferred to a membrane and then probed sequentially with *mmGCN5* cDNA sequences or intron 6 sequences.

A predominant RT-PCR product of a size corresponding to the spliced cDNA (lacking the intron) was amplified from mouse embryonic, kidney and ovarian RNAs (lower band, Fig. 3B). As expected, this product was significantly smaller (126 bp) than the amplification product from the genomic DNA (about 1 Kb), which contains introns 6 and 7. This small product hybridized to the *mmGCN5* cDNA sequences but not to the intron 6 probe, consistent with the removal of these intronic sequences by splicing. In contrast, two less abundant, closely spaced bands were detected by both the cDNA and the intron 6 probes. An intron 7 probe only hybridized to the genomic DNA, but failed to detect any of the RT-PCR products (data not shown), suggesting that intron 7 had been removed in all the transcripts. Sequencing of the larger, closely spaced RT-PCR products revealed that they represent two alternatively spliced

variants of *mmGCN5* (Fig. 3C). Both of these variants retained intron 6, but one also contained a novel 25bp exon (exon 7) located between introns 6 and 7. Intron 7 was removed from both of these alternatively spliced products, bringing the stop codons in intron 6 to a position just upstream of the ATG sequence corresponding to the previously predicted translation start site of the hsGCN5. Together these data indicate that the predominant form of the mouse cDNA is completely spliced, lacks these stop codons, and therefore is predicted to encode the longer version of GCN5. However, the two minor RT-PCR products we observe might encode shorter GCN5 proteins, consisting of the amino-terminal, P/CAF-like domain in isolation or of the C-terminal domain, which is most similar to yeast GCN5.

RT-PCR of total RNA from human cells revealed a similar mixture of completely and incompletely spliced RNAs. For example, two RT-PCR products were generated from the human HeLa and hepatoma RNAs. The size of the more abundant, smaller product again is consistent with a spliced cDNA lacking sequences homologous to the mouse intron 6 and exon 7, and this product only hybridizes to cDNA sequences. The less abundant, larger product hybridizes to both intron and exon sequences (Fig. 3B middle panel). We suggest that the longer product likely corresponds to the *hsGCN5* cDNA sequences previously reported, whereas the more prevalent, shorter form represents a spliced product predicted to encode a longer protein analogous to that encoded by the mouse cDNA isolated above.

Long GCN5 proteins are present in both human and mouse cells

To identify the size of the native mammalian GCN5 protein(s), total cell extracts prepared from a 12.5 dpc mouse embryo or human Hela cells were probed with a polyclonal sera raised against the previously described hsGCN5 (generously provided by Dr. Shelley Berger, Wistar Inst.). The hsGCN5- specific antisera detected a 98 Kd protein in the HeLa cell nuclear extracts, consistent with the predicted size of the full length GCN5 protein containing the extended amino-terminal region (Fig 4A, left panel). To ensure that this band corresponded to

mouse GCN5 and that the hGCN5 antibody did not cross react with P/CAF, we compared the relative signals obtained with the hGCN5 antibody and a P/CAF antibody (generously provided by Yoshihiro Nakatani, NIH) with extracts from U2OS cells or HeLa cells. The P/CAF antibody recognized a single band in the U2OS extract, consistent with previous reports that P/CAF is well expressed in these cells (41), and in the HeLa cell nuclear extract. The hGCN5 antibody, however, did not recognize any proteins of a similar size in either extract, but did recognize a prominent band of ~98 kDa in the HeLa nuclear extract. Therefore, the hGCN5 antibody does not appear to cross react significantly with PCAF, and we conclude that the 98 kDa protein recognized by this antibody in HeLa cell extracts is GCN5.

The hGCN5 antibody also recognized a faint 60 kDa band (lower arrow in right panel, Fig. 4A) in the HeLa cell extracts, close to the predicted size of the shorter GCN5 protein described previously (38) and above. Thus, both the long and short forms of GCN5 appear to be expressed in these cells, but the longer form appears predominant. Interestingly, the long form of GCN5 was the only form detected in mouse embryo extracts. The expression of GCN5 protein in the embryonic extracts is consistent with high levels of GCN5-specific RNA detected in these tissues (see Fig 5). Moreover, since only very low levels of P/CAF RNA were detected at this (or any) stage of mouse embryogenesis (data not shown and Fig 5), these data further support our conclusion that the hGCN5 antibody recognizes mouse GCN5 rather than P/CAF. Neither the long nor the short forms of GCN5 were detected by control, pre-immune serum in either the mouse or human extracts (data not shown).

We also used the anti-hsGCN5 sera to immunoprecipitate GCN5 proteins from the mouse embryo extract. Precipitated proteins were then detected by western blot using the same sera. Again, a 98 kDa protein was detected by the hsGCN5 antibody but not by a control rabbit sera (Fig. 4B). Unfortunately, the shorter form of GCN5, if it were present, would comigrate

with the IgG band and thus could not be detected by this approach. Nevertheless, these experiments confirm the presence of the longer GCN5 protein in mouse embryos.

Cloning of *mmP/CAF*

A second GCN5-related cDNA clone was isolated in our screen of the mouse cDNA library that contained a high degree of similarity to hsP/CAF. Since all initial clones appeared to be incomplete, containing a 867 bp fragment of the cDNA (relative to the human sequence), a second library was screened using GeneTrapper™ technology. Multiple full-length cDNAs were obtained containing a predicted open reading frame of 813 amino acids. This ORF exhibited 93% identity to the hsP/CAF cDNA sequence, but only 75% identity to the reported hsGCN5 cDNA sequence (41). We therefore designated this clone as *mmP/CAF*. Both the mmGCN5 and the mm P/CAF sequences possess predicted catalytic domains and bromodomains identified in a number of recently identified histone acetyltransferases , including several highly conserved amino acids near the putative the catalytic center (Fig 1).

Using a fragment from the 5' region of the *mmP/CAF* cDNA as probe, we identified multiple clones from a library of mouse genomic sequences that contained P/CAF sequences. Four of these contained different portions of the cDNA sequence. These clones indicate that in contrast to the *mmGCN5* gene, which contains small introns (a few hundred base pairs each), the *mmP/CAF* gene contains very large introns (16-20 kb). Because of these large introns, we have not completed cloning of *mmP/CAF* genomic sequences.

Interestingly, several clones identified in our genomic screens apparently contain a *P/CAF* pseudogene. No intronic sequences are present in these clones and several base substitutions, relative to the cDNA sequence, are scattered throughout the predicted coding region of the pseudogene. RT-PCR analysis indicates that the pseudogene is not expressed in several mouse tissues examined, including brain, eye, heart, lung, liver, kidney, thymus, spleen, fat, diaphragm, small intestine, ovary, testis, or a 13.5 dpc embryo (data not shown).

Ubiquitous but complementary expression of *mmGCN5* and *mmP/CAF*

To examine and compare the expression of mmGCN5 and mmP/CAF, total RNA was extracted from various mouse tissues, subjected to denaturing electrophoresis, transferred to a membrane and then probed with mmGCN5 or mmP/CAF specific sequences.

A single transcript of 3.3 kb was detected in all tissues with the GCN5 probe, consistent with size of the cDNA clone we isolated. Similarly, a single, ubiquitous transcript was detected with the P/CAF probe, and the size of this RNA, 4.4 kb, is similar to the P/CAF cDNA we isolated. Interestingly, the P/CAF RNA always exhibited a broader banding pattern than did the GCN5 RNA. These two RNAs were clearly distinguished from one another when probed on the same blot, and a differential pattern of expression was detected (Fig. 5). For example, the ratio of mmGCN5 to P/CAF expression is higher in brain, thymus, spleen, testis, and 13.5 dpc embryonic tissue, while this ratio is much lower in heart, liver, kidney, and skeletal muscle. Western blot analysis of GCN5 protein levels (using the polyclonal antiserum to hGCN5 described above) in various mouse tissues confirmed the general pattern of expression indicated by this RNA analysis (data not shown).

Chromosomal location of the *mmGCN5* and *mmP/CAF* genes

The chromosomal location of the mmGCN5 gene was mapped by standard linkage analysis using the Jackson Laboratory interspecific backcross panel (C57BL/6Jei x SPRET/Ei)F1 x SPRET/Ei, also known as Jackson BSS (33). mmGCN5 mapped cleanly to a distal region on chromosome 11 and cosegregated tightly with *BRCA1* , as well as a number of other genes previously mapped to that locus (data not shown but raw data from the Jackson Lab is available at http://www.jax.org/resources/documents/cmdata). Interestingly, the human GCN5 gene was recently mapped by FISH analysis to a syntenic region of human chromosome 17 (9) and also found to cosegregate with human *BRCA1*.

The location of mmP/CAF was mapped in a similar fashion, using the same backcross panel. In this case we used a probe specific for intronic sequences to ensure we mapped the authentic mmP/CAF gene and not the P/CAF pseudogene. This analysis indicated that mmP/CAF is located 32 cM from the centromere of mouse chromosome 17 and that it cosegregates with the DNA marker D17Bir8 (see www address above).

mmGCN5 encodes a histone acetyltransferase with similar substrate specificity to P/CAF

The high degree of homology between the mouse, human, and yeast GCN5 proteins strongly predicts that mmGCN5 and mmP/CAF will exhibit histone acetyltransferase activity. We confirmed this initially by examining the activity of the isolated, conserved acetylase domain of mmGCN5 and mmP/CAF, expressed as recombinant protein in *E. coli.* As expected, this domain of mmGCN5 was quite active as a histone acetylase, and it preferentially acetylated free (non-nucleosomal) histone H3, and to a lesser degree H4, as does yGcn5p (23) and the previously reported form of the hsGCN5 protein (41). Full-length mmGCN5 and mmP/CAF recombinant proteins (also expressed in bacteria) exhibited this same substrate specificity towards free histones (Fig 6 and data not shown).

To determine which residues of histone H3 were acetylated by mmGCN5, we performed assays with synthetic peptides corresponding to the amino terminal 'tail' of this histone. As expected, we found the full-length GCN5 protein efficiently acetylated peptides corresponding to the first twenty amino acids of histone H3 (Fig. 6A and B). This domain alone, then is sufficient for binding to the enzyme and subsequent catalysis. However, mmGCN5 could not acetylate a peptide that contained acetyl-lysine moieties at positions 9 and 14 (Fig. 6B), suggesting that one or both of these lysines may be a target site for mmGCN5. In contrast, mmGCN5 readily acetylated a peptide containing acetyl-lysine at positions 9 and 18 (Fig 6A). Taken together, these data suggest that K14 is the preferred acetylation site in H3 for mmGCN5. Similar assays performed with H4 peptides indicate that K8 is the preferred site of acetylation in H4 (data not

shown). These results are consistent with the site specificity determined for recombinant yeast Gcn5p, confirmed by protein sequencing of acetylated histones (23). Importantly, these results indicate that the extended amino terminal domain of mmGcn5 does not change the histone or lysine residue specificity of the enzyme.

The specificity of mmP/CAF was also tested using the peptide substrates. In all respects, mmP/CAF exhibited a substrate specificity identical to that of mmGCN5 (Fig. 6C and D).

One striking difference between the previously reported, shorter form of recombinant hsGCN5 (or yGcn5p) and recombinant hsP/CAF was the ability of P/CAF to acetylate nucleosomal substrates (23)(41). Given the homology between the amino-terminal portions of P/CAF and mmGCN5, we asked whether the full-length recombinant mmGCN5 could also acetylate histones within a nucleosome. We found that mmGCN5, like hsP/CAF, can acetylate nucleosomal H3, and to a lesser degree, H4 (Fig. 7). In agreement with previously reported results (23)(41), we also found that the short form of mmGCN5 or yGcn5p was unable to acetylate nucleosomes (data not shown). These results suggest that one function of the amino-terminal domains of mammalian GCN5 and P/CAF may be to facilitate the recognition of chromatin templates.

mmGCN5 and mmP/CAF both interact with CBP and p300

hsP/CAF interacts with CBP and p300 (41). Given the similarity between mmGCN5 and mmP/CAF, we examined the ability of both of these proteins to bind to CBP or p300 in vitro.

Whole cell lysates from bacteria expressing fragments of CBP fused to glutathione-S-transferase (GST; fusion constructs were kindly provided by Y. Nakatani, NIH) (41) were mixed with lysates from cells expressing the amino terminal domain of mmP/CAF, the amino-terminal domain of mmGCN5, or the C-terminal domain of mmGCN5. The CBP fragments (A-F) spanned the ADA2 homology domain and extended into the transcriptional activation domain (41). A fragment of p300 (B') homologous to the B fragment of CBP was also tested. GST

fusion proteins were purified together with any interacting proteins using glutathione-sepharose, and the interacting proteins were identified by western blot using an antiserum specific for the six histidine tag present in the recombinant mmP/CAF or mmGCN5 proteins.

The amino terminal domain of mmP/CAF selectively bound to fragments A and B of CBP and the corresponding B' fragment of p300. In some experiments, we also observed binding to the D fragment, but we never observed binding to fragments C, E, or F. A deletion within the B fragment (ΔB) of CBP that removed residues 1801-1851 eliminated binding. This pattern of binding to the CBP/p300 fragments is extremely similar to that previously reported for hsP/CAF (41), as expected.

A recombinant form of hsGCN5, which lacked the amino terminal domain reported here for mmGCN5, failed to bind CBP or p300 in previous experiments by Yang et al (41). This form of hsGCN5 corresponds to the C-terminal region of mmGCN5. We therefore compared binding of the amino-terminal and the C-terminal halves of mmGCN5 to the GST-CBP and -p300 fragments. Surprisingly, we found that both of these mmGCN5 domains bound to CBP fragments A-D, with little or no binding to fragments E, F, or the ΔB fragment. Both the amino-terminal and C-terminal regions of mmGCN5 also bound to the p300 B' fragment. The amount of the GST-fusion proteins recovered from the GST columns that did not exhibit binding to the GCN5 fragments was greater than or equal to that of the GST fusions that did exhibit binding (data not shown), so the absence of binding was not due to lowered amounts of the E, F, or ΔB fragments. In addition, the selective binding of the mmGCN5 peptides to CBP fragments A-D indicates that these interactions are not nonspecifically mediated by the GST moiety, since this moiety is also present in fragments E, F and ΔB. The specificity of the interactions was further tested using an unrelated protein, HIR A, which failed to bind to any of the GST-CBP or p300 fragments. Thus, CBP fragments A-D do not exhibit general, nonspecific binding to random proteins. We conclude that mmGCN5 contains two distinct CBP/p300 interaction domains, and

these domains interact with a broader region in CBP than does P/CAF. Importantly, our finding that mmGCN5 and mmP/CAF can both interact with CBP/p300 indicates that these proteins are very similar in function as well as in structure.

Discussion

The recent identification of nuclear histone acetyltransferases has directly linked chromatin modification with transcriptional regulation (1, 5, 26, 29). We report here the cloning of mouse GCN5 and P/CAF sequences. We find that mmGCN5 differs from yeast *GCN5* and the previously reported human GCN5 sequences in that it encodes a large N-terminal domain similar to that found in P/CAF. Our data indicate that human GCN5 contains this extra domain as well. While this domain does not appear to affect the histone specificity of the acetyltransferase, it does afford the enzyme the ability to modify nucleosomal substrates *in vitro*.

In vivo, both the yeast and the mammalian enzymes must interact with and modify nucleosomal histones. In yeast, this is accomplished by association of Gcn5p into high molecular weight protein complexes that can modify nucleosomes and that recognize additional histones (14). At least some of the Gcn5p-associated proteins are conserved in higher eukaryotes, and Gcn5-Ada complexes have been identified in human cells (7), further indicating that these enzymes serve similar functions across species. We scanned the yeast genome database to determine whether a protein homologous to the amino-terminal domains of mmGCN5 or mmP/CAF might exist that could be a component of the Gcn5p-containing complexes. However, we found no such homologues.

Interestingly, a single GCN5-related gene has been identified in *Drosophila*. This gene exhibits high similarity to mammalian P/CAF (E. Smith and D. Allis, personal communication)

and encodes the extended N-terminal domain. Although this domain is apparently not needed in yeast, its functions are not restricted to mammals.

Our work indicates that multiple differentially spliced forms of GCN5 transcripts co-exist in both mouse and human cells, which may generate different isoforms of GCN5 proteins. Of course, we can not rule out the possibility that the less abundant products represent incompletely spliced RNAs, but it is interesting that intron 6 and the stop codons therein are conserved between human and mouse. Since we detected transcripts containing intron 6 in both mouse and human tissues, we are intrigued by the possibility that shortened GCN5 proteins, containing either the N-terminal domain alone or the C-terminal domain alone, may provide an additional level of regulation of GCN5 functions. We detected, for example, the full length GCN5 protein in mouse embryo extracts and some human cells, but we detected a shorter, less abundant protein in Hela cells in addition to the full length protein. It will be especially interesting to determine whether various forms of GCN5 proteins are differentially regulated in different cell types or at different developmental stages.

The long form of mmGCN5 is very similar to mmP/CAF in structure, in acetyltransferase activity and substrate specificity, and in interactions with CBP/p300. Additional experiments are needed to determine whether these two proteins are functionally redundant in vivo. Even if GCN5 and P/CAF perform the same functions, they might be utilized at different developmental stages or in different cell types or tissues. The similarity between these proteins is somewhat reminiscent of that between CBP and p300. These two proteins also appear to be functionally equivalent in vitro, but mutations in p300 and CBP exhibit different phenotypes (27, 30), indicating they are not functionally redundant in vivo. It will be interesting to determine whether the same is true for mmGCN5 and mmP/CAF.

Several histone acetyltransferases, including p300/CBP and P/CAF, have been implicated in growth control and tumorigenesis (30, 32, 41). p300/CBP physically interacts with

the tumor suppressor p53, and potentiates sequence specific DNA binding and transactivation by p53 through acetylation of its C-terminal domain (15). Moreover, mutations in p300 have been found in certain colorectal and gastric cancers (27). CBP mutations are also involved in the etiology of certain acute myeloid leukemias and Rubinstein-Taybi syndrome (30), a developmental disorder with high incidence of neoplasms. In addition, P/CAF counteracts the transforming activity elicited by oncoprotein E1A, and overexpression of P/CAF has been shown to inhibit cell cycle progression (41). Therefore, histone acetyltransferases have been postulated to be negative regulators of cell growth, and possibly, tumor suppressors. We (this study) and others (9) have found that *GCN5* cosegregates with the tumor suppressor *BRCA1* gene (1 cM interval) in a highly syntenic region in mouse (Chr 11) and human (Chr 17). Interestingly, loss of heterozygosity on human chromosome 17 is a frequent genetic alteration in sporadic breast and ovarian cancers, where mutations in *BRCA1* and *BRCA2* are rarely found (28, 35). Indeed, a novel tumor suppressor gene involved in these cancers has been postulated to be located adjacent to the *BRCA1* locus (28, 35). *GCN5* may provide an attractive candidate for this novel tumor suppressor. The isolation and characterization of murine GCN5 and P/CAF reported here should facilitate further study of the role of these genes and of histone acetylation in normal mammalian development, as well as abnormal events leading to tumorigenesis.

Acknowledgments

We thank Jerry Workman and Patrick Grant for the kind gift of HeLa cell mononucleosomes and histones. We thank David Allis for the gift of H3 amino-terminal peptides, and E. Smith and D. Allis for the gift of oligomers for PCR and for sharing results prior to publication. We also thank Shelley Berger for antisera specific for human GCN5, Yongshen Ren for the gift of Hela cell nuclear extracts, and Yoshihiro Nakatani for the GST-CBP and p300 fusion constructs and hsP/CAF antibodies. Some DNA sequencing was performed by the UTMDACC Sequencing

Core Facility. We are grateful to Lucy Rowe and Mary Barter at the Jackson Laboratory for their assistance in the mouse chromosome mapping analysis. We thank Karen Hensley for help in preparation of some graphics, and Aurora Diaz for help in preparing the manuscript. W.X. is supported by a Rosalie B. Hite Fellowship and D.G.E. is supported by a Theodore Law UCF Scientific Fund Fellowship. This work was supported by grants to S.Y.R. from the Robert A. Welch Foundation, the USARMC, and the Breast Cancer Research Center at UTMDACC.

References

1. **Bannister, A. J., and T. Kouzarides.** 1996. The CBP co-activator is a histone acetyltransferase. Nature. **384:**641-643.

2. **Barlev, N., R. Candau, L. Wang, P. Darpino, N. Silverman, and S. Berger.** 1995. Characterization of physical interactions of the putative transcriptional adaptor, ADA2, with acidic activation domains and TATA binding protein. J. Biological Chemistry. **270:**19337-19344.

3. **Berger, S. L., B. Pina, N. Silverman, G. A. Marcus, J. Agapite, J. L. Regier, S. J. Triezenberg, and L. Guarente.** 1992. Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. Cell. **70:**251-265.

4. **Brownell, J. E., and C. D. Allis.** 1995. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proceedings of the National Academy of Sciences of the United States of America. **92**(14)**:**6364-8.

5. **Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth, and C. D. Allis.** 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell. **84**(6)**:**843-51.

6. **Candau, R., and S. L. Berger.** 1996. Structural and functional analysis of yeast putative adaptors: evidence for an adaptor complex in vivo. Journal Biological Chemistry. **271:**5237-5243.

7. **Candau, R., P. A. Moore, L. Wang, N. Barlev, C. Y. Ying, C. A. Rosen, and S. L. Berger.** 1996. Identification of functionally conserved human homologues of the yeast adaptors ADA2 and GCN5. Molecular and Cell Biology. **16:**593-602.

8. **Candau, R., J. Zhou, C. D. Allis, and S. L. Berger.** 1997. Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo. EMBO J. **16:**555-565.

9. **Carter, K. C., L. Wang, B. K. Shell, I. Zamir, S. L. Berger, and P. A. Moore.** 1997. The human transcriptional adaptor genes TADA2L and GCN5L2 colocalize to chromosome 17q12-q21 and display a similar tissue expression pattern. Genomics. **40:**497-500.

10. **Chomcznski, P., and N. Sacchi.** 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry. **162:**156-159.

11. **Edmondson, D. G., and E. N. Olson.** 1989. A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes and Dev. **3:**628-640.

12. **Edmondson, D. G., and S. Y. Roth.** 1996. Chromatin and transcription. FASEB Journal. **10**(10):1173-82.

13. **Georgakopoulos, T., and G. Thireos.** 1992. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO J. **11:**4145-4152.

14. **Grant, P. A., L. Duggan, J. Cote, S. Roberts, J. E. Brownell, R. Candau, R. Ohba, T. Owen-Hughes, C. D. Allis, F. Winston, S. L. Berger, and J. L. Workman.** 1997. Yeast GCN5 functions in two multisubunit complexes to actylate nucleosomal histones: characterization of an ADA complex and the SAGA (SPT/ADA) complex. Genes and Dev. **11:**1640-1650.

15. **Gu, W., and R. G. Roeder.** 1997. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell. **90:**595-606.

16. **Hansen, J. C.** 1997. The core histone amino-termini: combinatorial interaction domains that link chromatin structure with function. Chem Tracts: Biochemistry and Molecular Biology. **10:**56-69.

17. **Haynes, S. R., C. Dollard, F. Winston, S. Beck, J. Trowsdale, and I. B. Dawid.** 1992. The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. Nucleic Acids Research. **20**(10):2603.

18. **Horiuchi, J., N. Silverman, G. A. Marcus, and L. Guarente.** 1995. ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA2 and GCN5 in a trimeric complex. Mol. Cell Biol. **15:**1203-1209.

19. **Jeanmougin, F., J.-M. Wurtz, B. Le Douarin, P. Chambon, and R. Losson.** 1997. The bromodomain revisited. TIBS. **22:**151-153.

20. **Korzus, E., J. Torchia, D. W. Rose, L. Xu, R. Kurokawa, E. M. McInerney, T.-M. Mullen, C. K. Glass, and M. G. Rosenfeld.** 1998. Transcription factor-specific requirements for co-activators and their acetyltransferase functions. Science. **279:**703-707.

21. **Kozak, M.** 1991. An analysis of vertebrate mRNA sequences: intimations of translational contro. J. Cell Biol. **115:**887-904.

22. **Kozak, M.** 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266:**19867-19870.

23. **Kuo, M. H., J. E. Brownell, R. E. Sobel, T. A. Ranalli, R. G. Cook, D. G. Edmondson, S. Y. Roth, and C. D. Allis.** 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature. **383**(6597):269-72.

24. **Marcus, G., N. Silverman, S. Berger, J. Horiuchi, and L. Guarente.** 1994. Functional similarity and physical association between GCN5 and ADA2, putative transcriptional adaptors. EMBO J. **13:**4807-4815.

25. **Marcus, G. A., J. Horiuchi, N. Silverman, and L. Guarente.** 1996. ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. Molecular and Cellular Biology. **16:**3197-3205.

26. **Mizzen, C. A., X. J. Yang, T. Kokubo, J. E. Brownell, A. J. Bannister, T. Owen-Hughes, J. Workman, L. Wang, S. L. Berger, T. Kouzarides, Y. Nakatani, and C. D. Allis.** 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell. **87**(7)**:**1261-70.

27. **Muraoka, M., M. Konishi, R. Kikuchi-Yanoshita, K. Tanaka, N. Shitara, J. Chong, T. Iwama, and M. Miyaki.** 1996. p300 gene alterations in colorectal and gastric carcinomas. Oncogene. **12:**1565-1569.

28. **Niederacher, D., F. Picard, C. V. Roeyen, H. An, H. G. Bender, and M. W. Beckmann.** 1997. Patterns of allelic loss on chromosome 17 in sporadic beast carcinomas detected by fluorescent-labeled microsatellite analysis. Genes, Chromosomes, and Cancer. **18:**181-192.

29. **Ogryzko, V. V., R. L. Schlitz, V. Russanova, B. H. Howard, and Y. Nakatani.** 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. **87:**953-959.

30. **Petrij, F., R. H. Giles, H. G. Dauwerse, J. J. Saris, R. C. M. Hennekam, M. Masuno, N. Tommerup, G.-J. B. van Ommen, R. H. Goodman, D. J. M. Peters, and M. H. Breuning.** 1995. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature. **376:**348-351.

31. **Roberts, S. M., and F. Winston.** 1996. SPT20/ADA5 encodes a novel protein functionally related to the TATA-binding protein and important for transcription in *Saccharomyces cerevisiae*. Molecular and Cellular Biology.

32. **Roth, S. Y.** 1996. Something about silencing [news; comment]. Nature Genetics. **14**(1):3-4.

33. **Rowe, L. B., J. H. Nadeau, R. Turner, W. N. Frankel, V. A. Letts, J. T. Eppig, M. S. H. Ko, S. J. Thurston, and E. H. Birkenmeier.** 1994. Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. Mammalian genome. **5:**253-274.

34. **Silverman, N., J. Agapite, and L. Guarente.** 1994. Yeast ADA2 binds to the VP16 protein activation domain and activates transcription. Proceedings of the National Academy of Sciences. **91:**11665-11668.

35. **Tangir, J., M. G. Muto, R. S. Berkowitz, W. R. Welch, D. A. Bell, and S. C. Mok.** 1996. A 400 kb novel deletion unit centromeric to the BRCA1 gene in sporadic epithelial ovarian cancer. Oncogene. **12:**735-740.

36. **Turner, B. M.** 1991. Histone acetylation and control of gene expression. Journal of Cell Science. **99**(Pt 1):13-20.

37. **van Holde, K. E.** 1989. Chromatin. Springer-Verlag, New York.

38. **Wang, L., C. Mizzen, C. Ying, R. Candau, N. Barlev, J. Brownell, C. D. Allis, and S. L. Berger.** 1997. Histone acetyltransferase activity is conserved between yeast and human GCN5 and is required for complementation of growth and transcriptional activation. Molecular & Cellular Biology. **17**(1):519-27.

39. **Wolffe, A.** 1992. Chromatin Structure and Function. Academic Press, London.

40. **Wolffe, A. P., and D. Pruss.** 1996. Targeting chromatin disruption: Transcription regulators that acetylate histones. Cell. **84**(6):817-9.

41. **Yang, X.-J., V. V. Ogryzko, J.-I. Nishikawa, B. H. Howard, and Y. Nakatani.** 1996. A p300/CBP associated factor that competes with the adenoviral oncoprotein E1A. Nature. **382:**319-324.

Figure Legends

FIG. 1. Alignment of mouse GCN5, mouse P/CAF, and the reported human GCN5 amino acid sequences.

Identical amino acids among mouse GCN5 (mmGCN5), human GCN5 (hsGCN5) and mouse P/CAF (mmPCAF) are shaded. Amino acid deletions are indicated with a dotted line. The locations of the HAT/acetyl CoA binding regions (38)(20)and the bromodomain motif (17) are indicated. The full bromodomain likely encompasses amino acids 363-472 of hsGCN5 (19).

FIG. 2. Comparison of GCN5 and P/CAF sequences across species.

(A) Schematic comparisons of mouse GCN5 and mouse P/CAF sequences with GCN5 and P/CAF proteins from other species. Published sequences were obtained by searching the PIR-Protein and SWISS-PROT databases. Positions of the putative catalytic domains and the bromodomains are indicated above the diagram. Percentage identities between proteins are indicated on the right of the diagram. mmGCN5: mouse GCN5; mmP/CAF: mouse P/CAF; hsGCN5: human GCN5; hsP/CAF: human P/CAF; yGcn5p: yeast Gcn5p;. The dotted box indicates the existence of a predicted extended amino-terminal region in the human GCN5 protein. (B) Comparison of the Kozak consensus sequence for translation start sites (top), the predicted mouse GCN5 translation start sites (middle), and the previously reported human GCN5 translation start site (bottom) (41). The underlined AUG is the codon for the initiator methionine. The A of the AUG is designated as +1. In the consensus sequence, the nucleotides at the position –3 and +4 have the greatest impact on translation efficiency. 97% of vertebrate mRNAs have a purine (A or G, preferably A) in position –3, and 46% have a G in position +4 (21, 22).

FIG. 3. (A) Genomic structure of mouse *GCN5* gene. Exons in the mouse *GCN5* gene are shown in boxes, while introns are represented by the intervening lines. The thickened line indicates the intron that is retained in some alternatively spliced variants and which is homologous with sequences found in the previously reported 5'UTR of the human GCN5 cDNA. The *'s indicate in frame stop codons that would prevent translation of the full length protein. The shadowed box represents an exon that is skipped in the mouse cDNA. The positions of the translation start codon (ATG) and the termination codon (TAG) are indicated. (B) Co-existence of multiple forms of mammalian GCN5 transcripts. Left panel: DNaseI-treated RNA from mouse kidney, ovary and embryo, human Hela and hepatoma cell lines were RT-PCR amplified using primer A and B shown in the diagram on right. RT-PCR products were resolved on an ethidium-bromide stained agarose gel. Mouse *GCN5* genomic DNA was also amplified under the same conditions. A prevalent product corresponding to the size of the mouse cDNA (without introns 6 and 7) was amplified from all the RNA samples, while other larger products corresponding to the size of the reported human cDNA (containing intron 6, but not intron 7) were barely detected. Middle panel: The RT-PCR products were transferred to nylon membrane, and hybridized with mouse GCN5 cDNA sequences. Right panel: The same blot from the middle panel was stripped and re-hybridized with a probe (probe A) specific to the conserved intron 6. (C) The above RT-PCR products were gel purified and then amplified by PCR using a nested pair of primers. The PCR products were then subjected to DNA sequencing. Nucleotide sequence of the 317 bp fragment is shown. The smaller (292 bp) fragment has the same sequence as the larger fragment except that it lacks exon 7 sequences. Exons are boxed, and introns are numbered underneath. In-frame stop codons are in bold and marked by asterisks. The mouse GCN5 reading frame separated by intron 6 and exon 7 is shown. The initiator methionine codon in the reported human GCN5 (41) protein is in bold and underlined.

FIG.4. Detection of both long and short GCN5 proteins.

A) Left panel: protein extracts were prepared from U20S cells or HeLa cell nuclei and probed with polyclonal antibodies to hsP/CAF or hsGCN5, as indicated. Right panel: Protein extracts were prepared from 12.5 dpc mouse embryos or Hela cells and probed with the hs GCN5 antibody.

B) GCN5 proteins were immunoprecipitated from 12.5 dpc mouse embryos, and then probed with hsGCN5 antibody. An unrelated, HirA polyclonal antibody was used as a negative control.

FIG. 5. Ubiquitous and complementary expression of mouse GCN5 and P/CAF.

Top panel: Total RNA was isolated from various mouse tissues and embryos as indicated. Northern blot hybridization was performed using a mixture of mouse GCN5 and P/CAF cDNA probes. Two transcripts were detected and indicated by arrows on the right. The identities of the transcripts were confirmed by Northern blot hybridization using single GCN5 or P/CAF probes (data not shown). RNA molecular weight standards are shown on the left. Bottom panel: ethidium-bromide stained gel showing 18S ribosomal RNAs.

FIG. 6. Acetylation of histone H3 synthetic peptides by mmGCN5 and mmP/CAF.

Panels A and B show the results of acetyltransferase assays using recombinant full-length mmGCN5 and synthetic peptides corresponding to the amino-terminal "tail" of histone H3. Panels C and D show results of peptide assays using recombinant mmP/CAF. Peptides were either "unacetylated" or synthesized with acetyl-lysine residues at either K9 and K18 (di 9,18) or K9 and K14 (di9,14) in H3. "Vector" indicates assay of a control extract, made from bacteria

transformed by the recombinant vector without an acetyltransferase insert, subjected to the His-tag purification procedure.

FIG. 7. Acetylation of nucleosomal histones by mouse GCN5 and P/CAF.

Acetyltransferase assays were performed using HeLa cell mononucleosomes or free histones (as indicated) and an aliquot of each assay was resolved on a 22% SDS-PAGE gel. Coomassie stained gels and corresponding autoradiographs are shown. In both assays, histones H3 and H4 were acetylated by the recombinant full-length mmGCN5.

FIG 8. mmGCN5 and mmP/CAF both interact with CBP and p300 in vitro.

A) Fragments of CBP fused to GST that were used for the interaction assays in (C) are indicated. These fragments span the region of homology to ADA2 and extend into the transactivation domain of CBP (41).

B) Recombinant, His-tagged proteins used in the interaction assays of (C) were resolved by SDS PAGE and probed with an antibody specific for the His tag. The amounts shown represent 0.25% or 1% of the protein used in the assay, as indicated.

C) GST-CBP or p300 fusion proteins were mixed with crude bacterial lysates containing P/CAF, GCN5, or HIR A proteins. Interacting proteins were recovered using glutathione - sepharose and detected on western blots with the anti-His tag antibody. The p300 B' fragment is homologous to the CBP B fragment, and the CBP ΔB fragment is missing residues 1801-1851 (41).