2948–2954 *Nucleic Acids Research, 1998, Vol. 26, No. 12*

Cloning of *Drosophila* GCN5: conserved features among metazoan GCN5 family members

Edwin R. Smith, John M. Belote[1], R. Louis Schiltz[2], Xiang-Jiao Yang[2], Paul A. Moore[3], Shelley L. Berger[4], Yoshihiro Nakatani[2] and C. David Allis*

Department of Biology, University of Rochester, Rochester, NY 14627, USA, [1]Department of Biology, Syracuse University, Syracuse, NY 13244, USA, [2]Laboratory of Molecular Growth Regulation, National Institutes of Child Health and Development, National Institutes of Health, Bethesda, MD 20892, USA, [3]Department of Molecular Biology, Human Genome Sciences, Rockville, MD 20850, USA and [4]Wistar Institute, Philadelphia, PA 19104, USA

Received March 3, 1998; Accepted April 22, 1998

DDBJ/EMBL/GenBank accession nos AF029776, AF029777

ABSTRACT

***PCAF* and *hGCN5* are distinct human genes that encode proteins related to the yeast histone acetyltransferase and transcriptional adapter GCN5. The PCAF protein shares extensive similarity with the 439 amino acids of yGCN5, but it has an ~350 amino acid N-terminal extension that interacts with the transcriptional co-activator p300/CBP. Adenoviral protein E1a can disrupt PCAF–CBP interactions and prevent PCAF-dependent cellular differentiation. In this report, we describe the cloning and initial characterization of a *Drosophila* homolog of yGCN5. In addition to the homology to yGCN5, the *Drosophila* protein shares sequence similarity with the N-terminal portion of human PCAF that is involved in binding to CBP. In the course of characterizing dGCN5, we have discovered that hGCN5 also contains an N-terminal extension with significant similarity to PCAF. Interestingly, in the case of the h*GCN5* gene, alternative splicing may regulate the production of full-length hGCN5. The presence of the N-terminal domain in a *Drosophila* GCN5 homolog and both human homologs suggests that it was part of the ancestral form of metazoan GCN5.**

INTRODUCTION

It is becoming increasingly clear that a response to many cellular and developmental signals is the recruitment of histone acetyltransferases (HATs) or deacetylases for the regulation of gene expression (1,2). A major advance in understanding the connection between histone acetylation and gene transcription came with cloning of the first transcription-associated HAT, HATA1 from *Tetrahymena*, finding that it shared significant homology to the *Saccharomyces cerevisiae* adapter GCN5 (3). *In vitro*, GCN5 acetylates specific lysines in a pattern associated with transcriptionally active chromatin (4). Furthermore, the HAT activity of GCN5 is required for activation of GCN5-responsive genes *in vivo* (5–7). Together, these studies strengthen the link between histone acetyltransferase activity and activation of transcription.

To date, two human homologs of GCN5 have been cloned and shown to have HAT activity (8,9). One homolog (hGCN5) was reported to have the same structural organization as the *Tetrahymena thermophila* and *S.cerevisiae* proteins. The second homolog (hPCAF) is highly similar to hGCN5 (75% identity), but it was found to have a unique N-terminal extension required for binding to CBP and p300 (9). CBP and p300 are transcriptional co-activators that interact with a large number of developmentally important transcription factors (10). Since inactivating one copy of CBP causes severe developmental defects in Rubenstein–Taybi syndrome (11), it has been proposed that CBP is a limiting integrator of multiple signal transduction pathways at the level of gene activation (10). Recently, CBP/p300 was reported to have intrinsic HAT activity (12,13). PCAF and CBP are also found in a complex with a third family of HATs, the hormone receptor co-activators SRC and ACTR (15,14). Importantly, the *in vitro* substrate specificity of each of these HATs is distinct, suggesting that multiple HATs could act in concert at or near a single promoter.

To gain further understanding into the role of GCN5 in activating transcription, we set out to clone its *Drosophila* homolog. The combination of cytology and genetics in *Drosophila* has already proven to give worthwhile insight into relationships between states of acetylation and levels of gene activity (see for example 16). In this paper, we report the cloning and initial characterization of a *Drosophila* homolog of the histone acetyltransferase and transcriptional co-activator GCN5. We find that this protein shares significant homology to the N-terminal half of human PCAF. In addition, we find that the h*GCN5* gene is subject to alternative splicing, a process that may be responsible for generating at least two distinct protein products.

MATERIALS AND METHODS

GenBank accession numbers

d*GCN5* complete coding sequence, accession no. AF029776; h*GCN5* complete coding sequence, AF029777; h*GCN5* genomic DNA sequence, AC003104; *Arabidopsis*, AF031958; y*GCN5*, Q03330; Tet p55/*HATA1*, U47321; *PCAF*, U57317; h*GCN5*-S, U57316.

*To whom correspondence should be addressed. Tel: +1 716 275 7586; Fax: +1 716 275 2070; Email: cdaa@uhura.cc.rochester.edu

Degenerate oligo PCR

An aliquot of 5 µg total RNA from KC_0 cells was reverse transcribed with 200 ng degenerate oligo complementary to VPDYYD (TCR RAR TAR TCI GGI AC) in a final volume of 20 µl. The reaction was carried out for 2 h with 1 µl Superscript II (Gibco BRL) at 42°C. The cDNA was used at a final concentration of 1:500 for degenerate oligo PCR. Samples of 300 ng each primer were used in a 50 µl reaction. For the first round of PCR, we used a sense primer for MPKEY (ATG CCN AAR GAR TAY AT) and an antisense primer for YIK(D/E)Y (CCN TCR TAR TCY TTD AT). A second round was a hemi-nest using a nested sense primer for QVRGYG (CAR GTN AGR GGN TAY GG) and the same antisense primer for YIK(D/E)Y. First round product was used as template for the second round of PCR at a final concentration of 1:1000.

Amplification of cDNA ends

3′ RACE was performed according to Frohman (17) and the anchor primers were kindly provided by Rafaat El Maghrabi (Stony Brook, NY). A complete 3′-end of d*GCN5* was cloned as evidenced by a polyadenylation signal 29 bases upstream of the priming site of oligo(dT) (18). cRACE (19) was used to obtain additional 5′ sequence of d*GCN5*.

h*GCN5* cloning and cDNA analysis

Additional 5′ sequence of h*GCN5* cDNA was obtained by the cRACE method (19) using human liver total RNA (a gift from Peter Gibbs, Rochester, NY) for reverse transcription. A full-length cDNA was identified by searching a large database of ESTs generated by Human Genome Sciences (Rockville, MD). Pairwise similarity calculations and multiple sequence analysis were performed with the GCG software package. Shading was done with the computer program Boxshade (http://ulrec3.unil.ch:80/software/BOX_form.html). An amino acid is shaded if three of the six proteins are similar (gray) or identical (black).

RT-PCR of hGCN5 message was performed on human liver total RNA, followed by TA Cloning (Invitrogen) and sequencing. The reverse transcription primer was 5′-TGCTGGGGCTGAA-GATGGG-3′. Primers for amplification were the sense oligo 5′-CATGTCTTTGGGCGAAGCC-3′ from exon 6 and the anti-sense oligo 5′-GTGCTGGGAACAACCGCTG-3′ from exon 8.

Genomic library screen

A *Drosophila melanogaster* genomic library, constructed in λ EMBL3 using DNA from a *th st tra cp in ri pP*/TM3 strain (J.M.Belote, unpublished results), was screened using the d*GCN5* cDNA clone as probe and two overlapping recombinant phage were isolated. A 4.9 kb *Bam*HI fragment containing the hybridizing sequences was subcloned into pGEM3 (Promega) for subsequent analysis. From the genomic sequence we realized that we still lacked the full 5′-end of the cDNA. We predicted the start of the coding sequence and this prediction was confirmed by sequencing a cDNA clone LD17356 from the Berkeley *Drosophila* Genome Project (clone obtained from Genome Systems Inc., St Louis, MO).

In situ hybridization

Hybridization of the chromosomes was carried out according to standard procedures (20). Probes were prepared for *in situ* hybridization by random primed labeling of a gel-purified restriction fragment containing the entire d*GCN5* cDNA. The modified deoxynucleotide used was biotinylated dUTP from Gibco BRL and signal detection was done using a Detek-I-hrp Kit (ENZO).

Northern blot analysis

Total RNA extracted from several different stages of *Drosophila* development was generously provided by David Sullivan (Syracuse University, Syracuse, NY). Poly(A)$^+$ RNA was isolated and electrophoresed through a formaldehyde–agarose gel as described (21). The RNA was blotted onto GeneScreen (Dupont) nylon membrane and probed with a random primed ^{32}P-labeled *Cla*I fragment from the d*GCN5* cDNA. Following autoradiography, counts were removed from the filter by treating with boiling TE buffer. The filter was autoradiographed to confirm that no residual radioactivity remained. The filter was then hybridized with a probe specific for the ribosomal protein rp49 (22) to assess the levels of intact RNA in each lane.

Histone acetyltransferase assays

Full-length (amino acids 1–813) and a fragment corresponding to the catalytic domain (amino acids 469–634) were expressed as 6×His N-terminal tagged fusion proteins using pRSET vector (Invitrogen) and BL21 (DE3) *Escherichia coli*. HAT assays were performed as previously described (23).

Western blotting

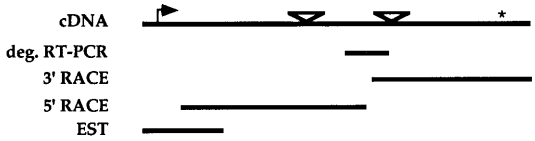
Rabbit polyclonal antibodies raised against recombinant hGCN5 and the N-terminal half of hPCAF were described previously (9). Anti-dGCN5 antibodies were raised in rabbits, using 6×His-tagged recombinant dGCN5 (amino acids 349–813) as immunogen. Affinity purification was done after coupling recombinant protein to CNBr-activated Sepharose under denaturing conditions as described (24). HeLa nuclear extracts were made according to Dignam *et al.* (25). *Drosophila* embryo nuclear extracts were prepared from 0–12 h old embryos by incubating nuclei on ice in extraction buffer for 30 min then removing debris by centrifugation at 15 000 *g* for 20 min. Extraction buffer was 0.5 M KCl, 100 mM HEPES, pH 7.4, 5 mM $MgCl_2$, 0.5 mM EDTA, 10% glycerol, 1 mM DTT and 1 mM PMSF. Proteins were resolved by SDS–PAGE and transferred to PVDF membrane and detected with the ECL-Plus detection system (Amersham).

RESULTS

Cloning, mapping and expression of *Drosophila GCN5*

To clone the *Drosophila* homolog of *GCN5* we took advantage of conserved sequences between y*GCN5*, Tetp55/*HATA1* and partial cDNA sequences of two human genes represented in the Merck/WashU. expressed sequence tag (EST) database. As shown in Figure 1A, we designed degenerate primers to amplify a *Drosophila* cDNA by RT-PCR, then used 5′ and 3′ RACE to obtain the ends of the cDNA. Cloning and sequencing of these RT-PCR products confirmed that they encode a protein closely related to GCN5. Additional 5′ sequence was obtained from a

A



B

```
  1  MSGGPSITIK  SQPIDGNNTG  NAAAQQQQQA  ANGAATAGAS  GAAGSAQNPG
 51  HGGAASGAGS  VPAEGTRQNS  LQRIQQRKQK  VFNLPVPQKL  AKLSMYSACQ
101  SEGCRCTGWK  TPQENRHRDV  ESSYCPEFNE  ECRNTSCRHS  LRSHIAHLDN
151  ISSSSMNELL  GAIIDMENLF  MSMQRVEDED  TKKVYQYLFR  LLRQCVLTRQ
201  QAVIRGPLGD  PPFETPCITK  AVLSLVFYKY  NHLSTPELQT  MTEVAKTFLN
251  FLNHYNFESP  STRRGDLTHE  DASNYKINYT  RWLVFCHVPA  FCNSLRQCET
301  SLVFGRTLLR  TVFQCVSQQL  KKKCISERDR  FPEDKRSIIT  LMPKFLETLR
351  AELLKDDSPI  WDTSYRPSNS  FVIQQRKRNQ  EVASVPIGPS  AASIGGNKRT
401  SVGEPLHKRI  KKEPTDRPSS  ENLDDLPADV  VMRAMKSVSE  SKTTNKAEIL
451  FPVNVSRDEN  VKAEEQKRAI  EFHVVGNSLT  KPVDKQTVLW  LLGLQLVFAY
501  QLPEMPREYI  SQLVFDTKHK  TLALIKENQP  IGGICFRPFP  SQGFTEIVFC
551  AVTMSEQVKG  YGTHLMNHLK  DYSIQRGIKH  LLTFADCDAI  GYFKKQGFSK
601  DVKLARPVYA  GYIKEYDSAT  LMHCELHPSI  VNTQFIAVIR  SQSEILKELI
651  AQRHNEVQKV  RPGLTCFKEG  LPVIPVESIP  GLREIGWKPQ  NRPARSSRPL
701  EESTDPEKLA  TSFASVLQSV  RQHTTAWPFL  RPVTAAEVPD  YYDHIKYPMD
751  LKTMGERLKK  GYYQTRRLFM  ADMARIFSNC  RFYNSPDTEY  YRCANSLERY
801  FQTKMRELGL  WDK
```

Figure 1. Cloning of a *Drosophila* homolog of GCN5, dGCN5. (**A**) Diagram of the various clones used to obtain the 2.8 kb cDNA. The arrow represents the proposed start methionine, the open triangles 55 and 62 base introns respectively and the asterisk denotes the end of the open reading frame. (**B**) The predicted protein translated from d*GCN5* is shown with the catalytic domain underlined and the bromodomain double underlined.

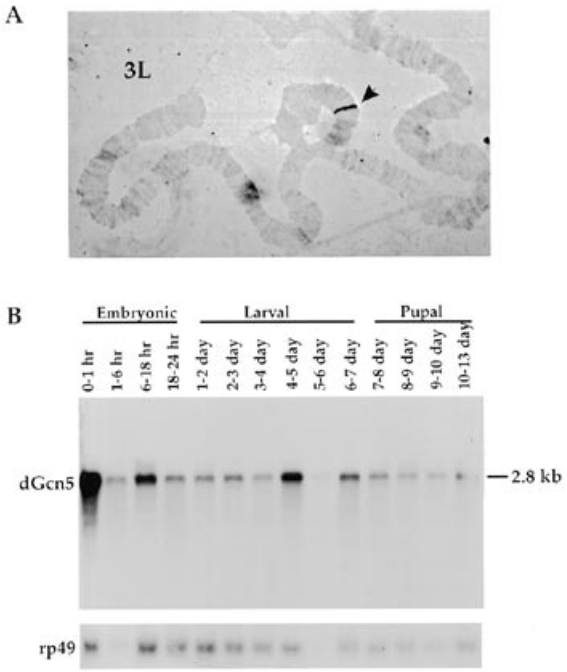


Figure 2. Cytological location of the d*GCN5* gene and northern analysis. (**A**) A cDNA probe was random primer labeled with biotin nucleotides. The probe was hybridized to polytene chromosomes from late third instar larvae. The left arm of the third chromosome (3L) is shown. A single site of hybridization was detected at 69C8-12 (marked by arrowhead). (**B**) 5 µg poly(A)$^+$ RNA from various developmental stages were separated on a 1.2% agarose–formaldehyde gel, transferred to a nylon membrane and probed with a random primed ^{32}P-labeled d*GCN5* cDNA. A 2.8 kb band of hybridization is detected throughout development (top). The blot was reprobed for the rp49 message as a loading control (bottom). After taking into account loading differences, it is apparent that there is a large amount of message in early embryos before zygotic transcription has started and there is also a second peak of message in late larval development (4–7 days).

cDNA clone found by searching the Berkeley *Drosophila* Genome EST database. A cDNA probe was used to isolate the corresponding gene from a *Drosophila* genomic library and two overlapping recombinant phage were obtained. A 4.9 kb *Bam*HI fragment was subcloned into a plasmid vector and sequenced for determination of exon–intron boundaries.

The deduced translated protein is shown in Figure 1B. The putative start methionine is in-frame with an upstream stop codon. The open reading frame of 2.4 kb encodes a protein of 93 kDa. Protein database searches retrieved hGCN5 and hPCAF as the strongest matches, while yGCN5, Tet p55 (HATA1) and an *Arabidopsis* GCN5-related protein are the next highest matches. Other significant matches are restricted to the bromodomain (26), which is a motif found in GCN5 family members and in a wide variety of co-activators and other transcription factors, including the HATs TAFII250 (27) and CBP (12,13). Because the derived protein sequence shown in Figure 1B predicts the presence of an N-terminal extension with significant homology to hPCAF, a domain thought to be missing in hGCN5 (8,9), we originally considered the *Drosophila* protein to be a homolog of hPCAF. However, in the course of this study, we have discovered that hGCN5 also contains an N-terminal extension with significant similarity to PCAF (see below and Fig. 5). For clarity, we have elected to refer to the *Drosophila* gene as d*GCN5*.

The position of the d*GCN5* gene on *Drosophila* polytene chromosomes was localized by *in situ* hybridization. Consistent with our detection of only a single gene by genomic Southern analysis (data not shown), we detect a single strong site of hybridization on the left arm of the chromosome 3 at cytological position 69C8-12 (Fig. 2A). Northern analysis shows that there is a single message detected of ~2.8 kb (Fig. 2B, top), consistent with the open reading frame of 2.439 kb and the 239 bp 3′-UTR and 98 bp 5′-UTR. Message is detected at all developmental stages. After controlling for loading differences (Fig. 2B, bottom), two peaks of particularly high message levels are apparent. One peak is detected in early embryos, which can be explained by maternal deposition; the other peak is detected in late larval development.

To demonstrate that dGCN5 possesses HAT activity, both full-length dGCN5 and a fragment corresponding to the catalytic domain of dGCN5 (amino acids 469–634) were expressed as 6×His N-terminal tagged fusion proteins in *E.coli* and purified on Ni$^+$–agarose. As expected and in excellent agreement with other GCN5 family members (4), each protein acetylates histone H3 when presented with a mixture of core histones (Fig. 3). The similar substrate specificity between the full-length and catalytic fragment is consistent with evidence that a corresponding catalytic fragment from yGCN5 contains all of the sequences required for full HAT activity (5). An additional substrate utilized by hPCAF is nucleosomal histones; the ability to weakly acetylate nucleosomal substrates was attributed to the N-terminal domain (9). However, despite sequence similarities between the N-termini of hPCAF and dGCN5, we have not been able to detect significant acetylation of nucleosomal histones by dGCN5 (data not shown).

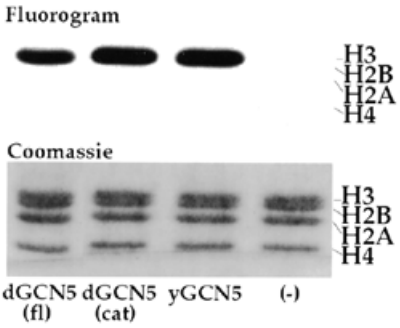


Figure 3. HAT activity of dGCN5. Full-length dGCN5 (fl), the dGCN5 catalytic domain (cat, amino acids 469–634), yGcn5p or no enzyme (–) was incubated with chicken erythrocyte histones and [^{3}H]acetyl-CoA and the histones were then analyzed by SDS–PAGE and fluorography. As expected from homology to GCN5, full-length dGCN5 and the catalytic domain of dGCN5 have HAT activity with a preference for histone H3.

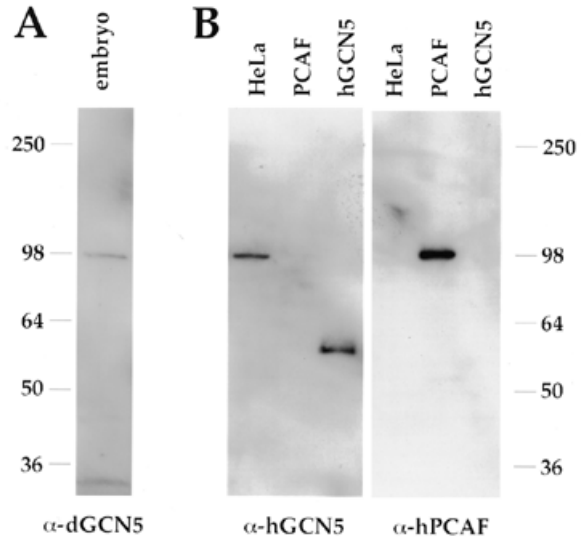


Figure 4. Western analysis of dGCN5 and hGCN5. (**A**) Antibodies raised against recombinant dGCN5 were used to probe nuclear extracts from 0–12 h old *Drosophila* embryos. A single band of 98 kDa is detected, consistent with the predicted molecular weight of 93 kDa. (**B**) Antibodies specific for hGCN5 (left) or hPCAF (right) were used to probe a western blot of HeLa nuclear extracts (10 µg), recombinant hPCAF (5 ng) and recombinant hGCN5 (1 ng). While the anti-hGCN5 antibody did not cross-react with recombinant PCAF protein, a 98 kDa band was readily detected in HeLa extracts. Although the anti-PCAF antibody reacts strongly with recombinant PCAF, this antibody fails to react (or in some experiments weakly intereacts) with this band in HeLa nuclear extracts.

Conservation of an N-terminal domain in metazoan GCN5

To confirm the size predicted for dGCN5 (93 kDa), antibodies to the C-terminus of dGCN5 (amino acids 348–813) detect a single protein with an apparent molecular mass of 98 kDa in nuclear extracts from *Drosophila* embryos (Fig. 4A), consistent with the migration of recombinant dGCN5 (data not shown). While we have only been able to find a single GCN5-related protein in *Drosophila*, two GCN5-related polypeptides have been described in humans (8,9). The human homologs differ in the presence (PCAF) or absence (hGCN5) of the 350 amino acid N-terminal domain. To confirm the predicted sizes of human GCN5-related proteins, we probed HeLa nuclear extracts with antibodies specific for either hGCN5 (Fig. 4B, left) or PCAF (Fig. 4B, right).

As shown in Figure 4B, the hGCN5 antibodies used here react well with a short form of recombinant hGCN5 (see below) and do not react with recombinant PCAF, demonstrating selectivity of the antibody for the appropriate polypeptide. Unexpectedly, when HeLa nuclear extracts are probed with these antibodies, a single polypeptide is detected with an apparent molecular weight of 98 kDa, the same size as recombinant PCAF. In contrast, PCAF antibodies react well with recombinant PCAF (Fig. 4B, right), but only react weakly, if at all, with the 98 kDa protein in HeLa nuclei. These results suggest that the 98 kDa HeLa nuclear protein detected by hGCN5 antibodies is distinct from PCAF, is more closely related to hGCN5 and yet is larger than that previously predicted for hGCN5 (8,9).

The previously reported cDNAs for hGCN5 encode proteins with a predicted molecular weight of 55 kDa, more similar to yGCN5 than to dGCN5 or hPCAF (8,9). We investigated whether the h*GCN5* gene could also encode a protein similar in size to PCAF and dGCN5, as was suggested by our western analysis. 5′ RACE and searches of an extensive database of ESTs maintained by Human Genome Sciences (Rockville, MD) confirm that the h*GCN5* gene can encode a larger polypeptide with additional homology to the N-terminal portion of dGCN5 and hPCAF (Fig. 5A). The new predicted molecular weight of hGCN5 is 93 kDa, which is similar in size to the 92 kDa PCAF. For clarity, this longer form of hGCN5 will be referred to as hGCN5-L (Fig. 5B). Others have found evidence for the existence of a hGCN5-related protein of ~60 kDa (28–30), as was expected from the previously isolated cDNAs (8,9). This shorter product will be referred to as hGCN5-S (Fig. 5B). We suggest that a more precise and informative nomenclature be chosen for the metazoan GCN5 family members when more rigorous data clarifies which polypeptides are produced from the h*GCN5* gene in various biological settings, as well as what function(s) they might serve (see below).

The h*GCN5* gene was recently sequenced by the human genome project (accession no. AC0033104). Comparison of the cDNA and genomic sequences allows us to draw exon–intron boundaries, as shown in Figure 6A. Of particular note is the sixth intron of 840 bases. Since the first 191 bases of this intron (see shaded box in Fig. 6) are found in the cDNA clone used to predict the hGCN5-S protein (9), we performed RT-PCR to confirm that the hGCN5 message is alternatively spliced.

Human liver RNA was used for RT-PCR, with primers in exons 6 and 8. Both a fully spliced product and a partially spliced product that contains the first 191 bases of intron 6 were amplified and sequenced, while amplification of the same region from genomic DNA yields a fragment that contains all of introns 6 and 7 (Fig. 6B and data not shown). One explanation for the two cDNA products involves alternative 5′ splice site usage in intron 6 of the h*GCN5* transcript. If the 5′ splice site at nt 1 of the 840 base intron is used, the entire intron can be removed in a single step (Fig. 6B). Alternatively, if a 5′ splice site at nt 192 of the

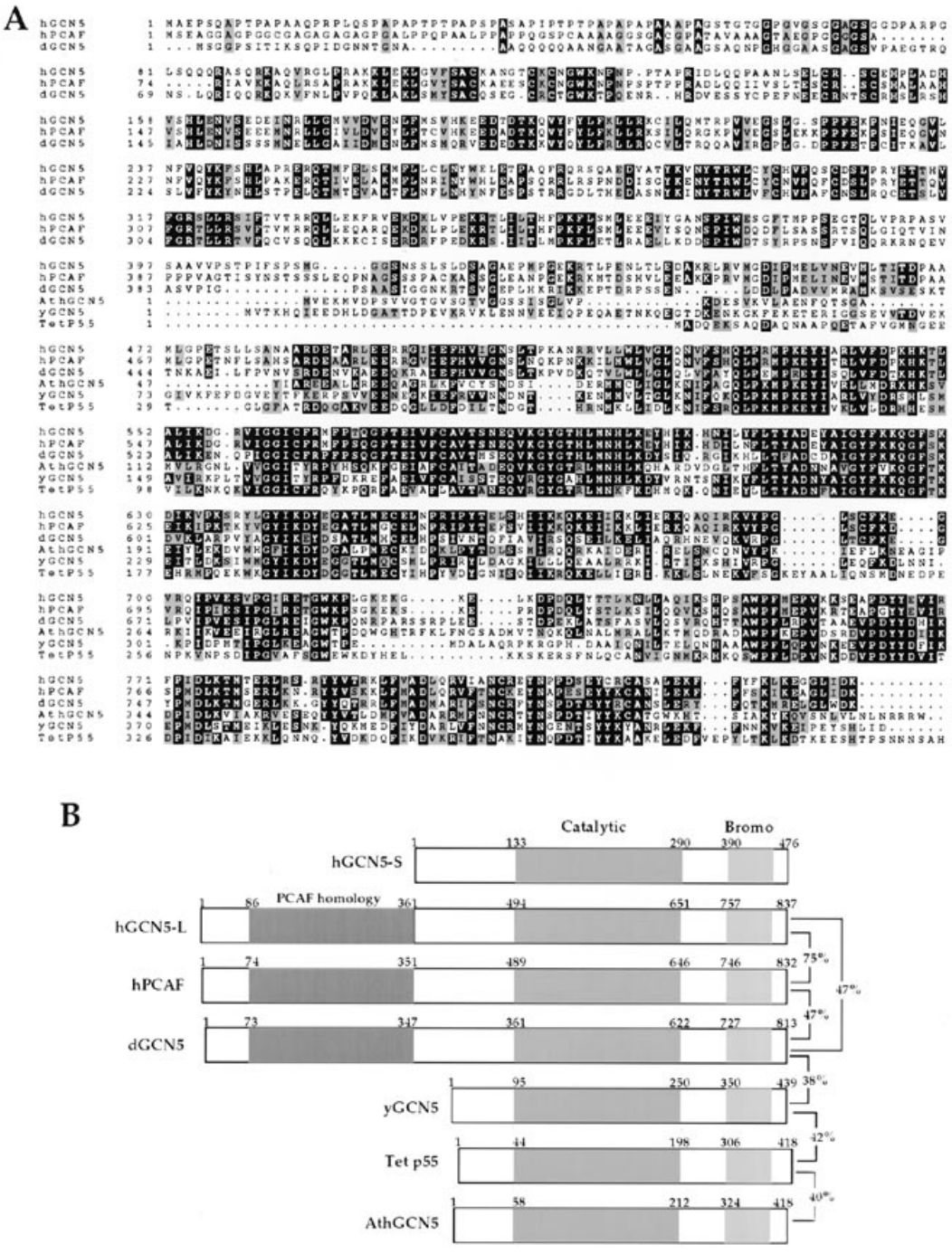


Figure 5. A conserved N-terminal extension exists in metazoan GCN5/PCAF family members. (**A**) 5′ RACE and searches of expressed sequence tag (EST) databases were used to obtain additional N-terminal sequence of hGCN5, which shares significant similarity to the corresponding region of hPCAF and dGCN5. Shown is an alignment of the known GCN5-related sequences. Amino acids are shaded if three proteins have similar (gray) or identical (black) residues at that position. (**B**) A schematic representation of the GCN5 family is shown, along with the percentage of identical amino acid residues shared between members. All family members share regions of significant similarity, including the region required for full catalytic activity of yGcn5p (9) and the bromodomain. The metazoan members of the family share an N-terminal domain not found in the yeast or *Tetrahymena* proteins (PCAF homology region). The h*GCN5* gene can encode long (hGCN5-L) and short (hGCN5-S) proteins. Amino acid one of hGCN5-S corresponds to amino acid 362 of hGCN5-L.

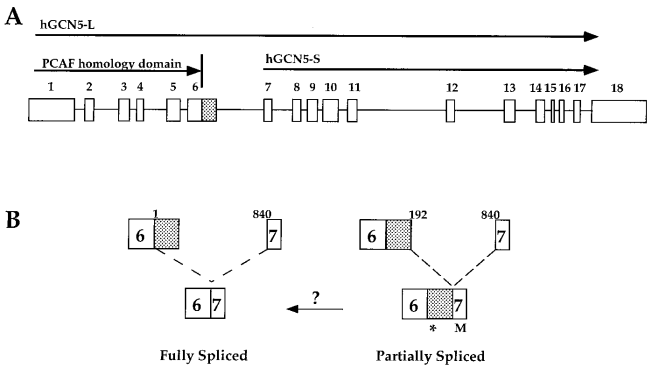


Figure 6. The hGCN5 mRNA is alternatively spliced. (**A**) Schematic representation of the intron–exon boundaries of the h*GCN5* gene. Numbered boxes represent exons. Arrows show which exons encode hGCN5-L, HGCN5-S and the PCAF homology domain (Fig. 5). The shaded box adjacent to exon 6 represents a 191 base portion of intron 6 that is not removed in some cDNA clones. (**B**) A schematic representation of intron 6 alternative splicing. Two 5′ splice sites are located at nt 1 and nt 192 of the 840 base intron. If the 5′ splice site at nt 1 is used, the whole 840 base intron is removed to generate a fully spliced product. If the 5′ splice site at nt 192 is used, only the last 649 bases of the intron are removed, to make a partially spliced product. The 191 base segment from intron 6 that remains in the partially spliced message may serve as an alternative exon for hGCN5 translational control (see Discussion). An asterisk marks a stop codon within this 191 base element that would cause premature termination of translation of the hGCN5-L protein, while the M below exon 7 represents a proposed start site of translation for hGCN5-S (9). An interesting feature of the 191 base segment is that it begins with a 5′ splice site and ends with a potential 3′ splice site, allowing for the possibility that intron 6 is removed in two steps (arrow with question mark).

intron is chosen, the last three fourths of the intron would be removed, leaving behind the 191 base segment found in some of our cDNA clones. Notably, this 191 base element ends with a CAG, itself a canonical 3′ splice site, raising the possibility that this intron can be spliced in a two-step mechanism (Fig. 6B). Inclusion of the 191 base intron-like element in the hGCN5 mRNA interrupts the reading frame of the h*GCN5* cDNA; a stop codon at nt 2 of the intron would terminate translation of hGCN5 at amino acid 361 (as numbered in hGCN5-L, Fig. 5B). The ATG predicted to be the start site of translation of hGCN5-S (9) is found at the beginning of exon 7. Therefore, inclusion of the 191 base element splits h*GCN5*-L into two ORFs, one encoding the N-terminal domain and one encoding hGCN5-S.

DISCUSSION

GCN5 and associated ADA proteins are needed for full activation by some transcriptional activators in *S.cerevisiae* (31,32). The cloning of human relatives of y*GCN5* and y*ADA2* raised the intriguing possibility that the encoded proteins play a similar role in metazoans (8,9). The single homolog of GCN5 in *Drosophila* described here (dGCN5) shows equal relatedness (47% amino acid identity) to the two human GCN5 homologs, while the two human proteins share 75% identity with each other (Fig. 5B). The close relationship between the two human homologs is consistent with a gene duplication of a single *GCN5*-related gene in the vertebrate lineage that resulted in the two genes found in humans. Murine homologs of h*GCN5* and h*PCAF* have been cloned recently, and their relationship to each other (~75%) and to each of their human counterparts (~90%) supports the idea of a gene duplication in the vertebrate lineage (S.Roth, personal communication). Other developmental regulators such as CBP, E2F, Rb and myb have only one known family member in *Drosophila*, while two or more family members exist in mammals (33–37). However, our failure to find a second *GCN5*-related gene in *Drosophila* does not exclude the possibility that such a protein exists.

The occurrence of the N-terminal domain in all of the known metazoan GCN5-related proteins suggests that this domain was present in the ancestral metazoan GCN5, while the absence of the domain in the known fungal, plant and protozoan GCN5s suggests that this domain is unique to animals. Although PCAF requires this N-terminal domain for *in vitro* acetylation of nucleosomal histones, the failure of dGCN5 to acetylate similar substrates suggests that other functions are responsible for conservation within this domain. A likely function of this conserved region is to interact with other transcription factors and co-activators for the targeted acetylation of chromatin for the control of gene activation (38,39).

One intriguing outcome of this study is that multiple protein products may be expressed from the h*GCN5* gene due to alternative splicing. Inclusion of the 191 base element in the mRNA creates a di-cistronic message consisting of an ORF that corresponds to the N-terminal portion of hGCN5-L (amino acids 1–161) and a C-terminal portion that corresponds to hGCN5-S protein (Fig. 5B). An N-terminal domain protein, lacking the HAT catalytic domain, might bind CBP, as does the corresponding region from PCAF. The hGCN5 N-terminal product might compete with PCAF for binding to CBP, as was proposed for the adenoviral protein E1a (9). Expression of the C-terminal ORF (h*GCN5*-S) could produce a histone acetyltransferase free of the control provided by p300/CBP (9). Recent studies have used western analysis to demonstrate the existence of the h*GCN5*-S product (28–30), but protein sequencing or mass spectrometry of purified hGCN5-S would help discriminate between proteolysis of hGCN5-L and alternative start sites of translation.

An additional reason to further investigate the translational control of h*GCN5* is the unusually high conservation of the 191 base element in intron 6 between the mouse and human *GCN5* genes (91% identity; S.Roth, personal communication). This nucleotide conservation is higher than the amino acid conservation in any of the reading frames within this sequence. We have not observed any significant matches to other sequences in the public databases and do not recognize any motifs within this sequence. Of potential interest is the occurrence of a 19 base palindromic sequence, since palindromes of this length could form RNA hairpins and affect translation (40).

The 191 base element could be required for translation of the C-terminal ORF (hGCN5-S protein) in at least three ways. First, the element may function as a nuclear export signal for the mRNA. Typically, the presence of an intron prevents transport of an mRNA to the cytoplasm; the HIV protein Rev binds to a particular structure in the viral RNA and facilitates transport of incompletely spliced RNAs to the cytoplasm for translation (reviewed in 41). Thus, nuclear export of the partially spliced hGCN5 mRNA would be required for translation of either the N- or C-terminal ORFs. A second translational function of the intron could be to serve as an internal ribosomal entry site (IRES); an IRES is needed for translation of a downstream ORF in a

di-cistronic message (reviewed in 42). A third potential function of the intron element may be to terminate translation after the upstream ORF in a manner that ensures ribosomal re-initiation at the correct start site of hGCN5-S (reviewed in 43). Since the conserved N-terminal domain of hGCN5 is likely to have important regulatory functions, it will be important to investigate, in further detail, the regulation of protein expression from the h*GCN5* gene.

The growing list of GCN5/PCAF family members has already produced unexpected connections between these HAT activities and cellular transformation. For example, disruption of interactions between PCAF and CBP by the viral *E1a* oncogene product is required for E1a-mediated cellular transformation. Furthermore, overexpression of PCAF has been shown to have a modest inhibitory effect on cell cycle progression (9), while a catalytically active PCAF is required for muscle differentiation (38). Mutations in the histone acetyltransferases p300/CBP are associated with developmental disorders (11,44) and a specific subtype of leukemia (45). Understanding the nature, regulation and specificity of these highly conserved chromatin-modifying activities will undoubtedly contribute to our understanding of both normal cellular and developmental processes as well as abnormal processes which lead to oncogenesis or other developmental disorders. Isolation of the *Drosophila GCN5* gene will enable us to explore some of these relationships in a powerful genetic model system.

ACKNOWLEDGEMENTS

We are grateful to Sharon Roth for communicating data on murine GCN5/PCAF before publication. We would like to thank Kerrie Ann Smyth for help with the *in situ* hybridizations and Jing Ma for help with the northern blotting. The contributions of the Merck/Washington University EST Project, Human Genome Sciences Sequencing Facility and the Whitehead/MIT Center for Genomic Research are appreciated. This work was supported by NIH grant RO1GM53512 to C.D.A. and NSF grant MCB-9506885 to J.M.B.

REFERENCES

1 Wu,C. (1997) *J. Biol. Chem.*, **272**, 28171–28174.
2 Grunstein,M. (1997) *Nature*, **389**, 349–352.
3 Brownell,J.E., Zhou,J., Ranalli,T., Kobayashi,R., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) *Cell*, **84**, 843–851.
4 Kuo,M.-H., Sobel,R.E., Brownell,J.E., Ranalli,T.A., Cook,R.G., Roth,S.Y. and Allis,C.D. (1996) *Nature*, **383**, 269–272.
5 Candau,R., Zhou,J., Allis,C.D. and Berger,S.L. (1997) *EMBO J.*, **16**, 555–565.
6 Kuo,M.-H., Zhou,J., Jambeck,P., Churchill,M.E. and Allis,C.D. (1998) *Genes Dev.*, **12**, 627–639.
7 Wang,L., Liu,L. and Berger,S.L. (1998) *Genes Dev.*, **12**, 640–653.
8 Candau,R., Moore,P., Wang,L., Barlev,N., Ying,C., Rosen,C. and Berger,S. (1996) *Mol. Cell. Biol.*, **16**, 593–602.
9 Yang,X.-J., Ogryzko,V.V., Nishikawa,J.-I., Howard,B.H. and Nakatani,Y. (1996) *Nature*, **382**, 319–324.
10 Kamei,Y., Xu,L., Heinzel,T., Torchia,J., Kurokawa,R., Gloss,B., Lin,S.C., Heyman,R.A., Rose,D.W., Glass,C.K. and Rosenfeld,R.G. (1996) *Cell*, **85**, 403–414.
11 Petrij,F., Giles,R.H., Dauwerse,H.G., Saris,J.J., Hennekam,R.C.M., Masuno,M., Tommerup,N., van Ommen,G.B., Goodman,R.H., Peters,D.J.M. and Breuning,M.H. (1995) *Nature*, **376**, 348–351.
12 Ogryzko,V.V., Schiltz,R.L., Russanova,V., Howard,B.H. and Nakatani,Y. (1996) *Cell*, **87**, 953–959.
13 Bannister,A.J. and Kouzarides,T. (1996) *Nature*, **384**, 641–643.
14 Spencer,T.E., Jenster,G., Burcin,M.M., Allis,C.D., Zhou,J., Mizzen,C.A., McKenna,N.J., Onate,S.A., Tsai,S.Y., Tsai,M.-J. and O'Malley,B.W. (1997) *Nature*, **389**, 194–208.
15 Chen,H., Lin,R.J., Schiltz,R.L., Chakravarti,D., Nash,A., Nagy,L., Privalsky,M.L., Nakatani,Y. and Evans,R.M. (1997) *Cell*, **90**, 569–580.
16 Bone,J.R., Lavender,J.S., Richman,R., Palmer,M.J., Turner,B.M. and Kuroda,M.I. (1994) *Genes Dev.*, **8**, 96–104.
17 Frohman,M.A. (1994) In Mullis,K.B., Ferre,F. and Gibbs,R.A. (eds), *The Polymerase Chain Reaction.* Birkhauser, Boston, MA, pp. 14–37.
18 Wickens,M. and Stephenson,P. (1984) *Science*, **226**, 1045–1051.
19 Maruyama,I.N., Rakow,T.L. and Maruyama,H.I. (1995) *Nucleic Acids Res.*, **23**, 3796–3797.
20 Ashburner,M. (1989) *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22 O'Connell,P.O. and Roshbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495–5513.
23 Brownell,J.E. and Allis,C.D. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6234–6238.
24 Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
26 Haynes,S.R., Dollard,C., Winston,F., Beck,S., Trowsdale,J. and Dawid,I.B. (1992) *Nucleic Acids Res.*, **80**, 2603.
27 Mizzen,C.A., Yang,X.-J., Kokubo,T., Brownell,J.B., Bannister,A.J., Owen-Hughes,T., Workman,J., Wang,L., Berger,S.L., Kouzarides,T., Nakatani,Y. and Allis,C.D. (1997) *Cell*, **87**, 1261–1270.
28 Forsberg,E.C., Lam,L.T., Yang,X.-J., Nakatani,Y. and Bresnick,E.H. (1997) *Biochemistry*, **36**, 15918–15924.
29 Currie,R.A. (1998) *J. Biol. Chem.*, **273**, 1430–1434.
30 Barlev,N.A., Poltoratsky,V., Owen-Hughes,T., Ying,C., Liu,L., Workman,J., Berger,S.L. (1998) *Mol. Cell. Biol.*, **18**, 1349–1358.
31 Georgakopoulos,T. and Therios,G. (1992) *EMBO J.*, **11**, 4145–4152.
32 Berger,S.L., Pina,P., Silverman,N., Marcus,G.A., Agapite,J., Reiger,J.L., Triezenberg,S.J. and Guarente,L. (1992) *Cell*, **70**, 251–265.
33 Akimaru,H., Chen,Y., Dai,P., Hou,D.-X., Nonaka,M., Smolik,S.M., Armstrong,S., and Goodman,R.H. (1997) *Nature*, **340**, 735–738.
34 Katzen,A.L., Kornberg,T.B. and Bishop,J.M. (1985) *Cell*, **41**, 449–456.
35 Peters,C.W., Sippel,A.E., Vingron,M. and Klempnauer,K.H. (1987) *EMBO J.*, **6**, 3085–3090.
36 Dynlacht,B.D., Brook,A., Dembski,M., Yenush,L. and Dyson,N. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 6359–6363.
37 Du,W., Vidal,M., Xie,J.E. and Dyson,N. (1996) *Genes Dev.*, **10**, 1206–1218.
38 Puri,P.L., Sartorelli,V., Yang,X.-J., Hamamori,Y., Ogryzko,V.V., Howard,B.H., Kedes,L., Wang,J.Y.J., Graessmann,A., Nakatani,Y. and Leverno,M. (1997) *Mol. Cell*, **1**, 35–45.
39 Kozus,E. Torchia,J., Rose,D.W., Xu,L., Kurokawa,R., McInerney,E.M., Mullen,T.-M., Glass,C.K. and Rosenfeld,M.G. (1998) *Science*, **279**, 703–707.
40 Zhang,J. and Maquat,L.E. (1997) *EMBO J.*, **16**, 826–833.
41 Ernst,R.K., Bray,M., Rekosh,D. and Hammarskjold,M.L. (1997) *Mol. Cell. Biol.*, **17**, 135–144.
42 Nambru,C., Lafon,I., Audigier,S., Gensac,M.-C., Vagner,S., Huez,G. and Prats,A.-C. (1997) *J. Biol. Chem.*, **272**, 32061–32066.
43 Walker,W.H., Girardet,C. and Habener,J.F. (1996) *J. Biol. Chem.*, **271**, 20145–21050.
44 Tanaka,Y., Naruse,I., Maekawa,T., Masuya,H., Shiroishi,T. and Ishii,S. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 10215–10220.
45 Borrow,J., Stanton,V.P., Andresen,J.M., Becher,R., Behm,F.G., Chaganti,R.S., Civin,C.I., Disteche,C., Dube,I., Frischauf,A.M., Horsman,D., Mitelman,F., Volinia,S., Watmore,A.E. and Housman,D.E. (1996) *Nature Genet.*, **14**, 33–41.