Molecular Cloning and Sequencing of a cDNA Encoding $N^\alpha$-Acetyltransferase from *Saccharomyces cerevisiae*

(Received for publication, March 9, 1989)

**Fang-Jen S. Lee**, Lee-Wen Lin, and John A. Smith

*From the Departments of Molecular Biology and Pathology, Massachusetts General Hospital and the Departments of Genetics and Pathology, Harvard Medical School, Boston, Massachusetts 02114*

Acetylation is the most frequently occurring chemical modification of the $\alpha$-NH$_2$ group of eukaryotic proteins and is catalyzed by an $N^\alpha$-acetyltransferase. Recently, a eukaryotic $N^\alpha$-acetyltransferase was purified to homogeneity from *Saccharomyces cerevisiae*, and its substrate specificity was partially characterized (Lee, F.-J. S., Lin L.-W., and Smith, J. A. (1988) *J. Biol. Chem.* 263, 14948–14955). This article describes the cloning from a yeast $\lambda$gt11 cDNA library and sequencing of a full length cDNA encoding yeast $N^\alpha$-acetyltransferase. DNA blot hybridizations of genomic and chromosomal DNA reveal that the gene (so-called *AAA1*, amino-terminal, $\alpha$-amino, acetyltransferase) is present as a single copy located on chromosome IV. The use of this cDNA will allow the molecular details of the role of $N^\alpha$-acetylation in the sorting and degradation of eukaryotic proteins to be determined.
Experimental Procedures

Protein Sequence Analyses of N⁰-acetyltransferase

N⁰-acetyltransferase was purified from yeast as previously described (24). N⁰-acetyltransferase (3 nmoles) was reduced and alkylated, precipitated with cold chloroform/methanol, redissolved in 0.1 M NH₄HCO₃, incubated with TPCK-treated trypsin (EC 3.4.21.4; Copper Biomedical, Malvern, PA) (120 pmol) for 24 hr at 37°C. recovered by lyophilization, and dissolved in 6 M guanidine hydrochloride in 0.1% CF₃COOH prior to HPLC.

Tryptic peptides were separated on a Vydac phenyl (0.46 x 25 cm) HPLC column, and selected fractions were rechromatographed isocratically once or twice (25). Chromatographic peaks were detected at 214 and 280 nm, collected manually, and lyophilized. The tryptic peptides were sequenced by automated Edman degradation performed with an Applied Biosystems 470A Protein Sequencer and an Applied Biosystems 120 Pth Analyzer (26).

Construction and Screening of cDNA Library.

Yeast RNA was isolated as described by Sherman et al. (27). Poly(A)⁺ RNA was selected on oligo(dT)-cellulose (28). cDNA was synthesized from 10 μg of poly(A)⁺RNA by the method of Okayama and Berg (29), as modified by Gubler and Hoffman (30), except that 10% of second strand was [³²P]-labelled. The cDNA was prepared for ligation to λgt11 arms using a method described by Aruffo and Seed (31). After the ends of the cDNA were made blunt with T4 DNA polymerase, the cDNA was ligated to adaptors consisting of two oligonucleotides: 3' CTCTAAAG 5' and 5' ACACGAGATTTC 3'. This cDNA was fractionated on a 5 to 20% linear KOAc gradient (5 ml) using a Beckman SW55 rotor centrifuged for 3 hr at 50,000 rpm at 22°C. Fractions (0.5 ml) were collected from the bottom of the tube. The cDNA was precipitated by addition of ethanol and linear polyacrylamide (20 μg/ml). The size of the cDNAs in each fraction was determined on a 1% agarose gel, and the fractions containing cDNAs between 1 and 8 kb were pooled. Ten micrograms of λgt11 DNA (32) was digested with EcoRI, ligated to adaptors (3' GTGTGACCAGATCTCTTAA 5' and 5' GTGTCTAGAG 3') and precipitated with PEG8000. 600 ng of λgt11 DNA bearing adaptors was ligated to 150 ng of size-selected cDNA bearing complementary adaptors in 2 μl and packaged in vitro (33) (Stratagene) (Jim Sheen, personal communication). Escherichia coli strain Y1088 was infected with recombinant phage, and the library was amplified once. The recombinant frequency was approximately 82%.
Among several peptides sequences, two peptides (peptides 27-3 and 11-3-2; Fig. 2A) were chosen for constructing two oligonucleotide probes (N1 and N2) based on most probable codon usage (34). The oligonucleotide probes were synthesized with an Applied Biosystems 380A DNA synthesizer by using the silica-based solid-phase method (35) and β-cyanoethyl phosphoramidite method (36). The purified oligonucleotide were isolated from the crude synthetic mixtures by PAGE and labelled to a specific activity of 2-8 x 10^8 cpm/μg by using [γ-32P]-ATP (New England Nuclear) and T4 polynucleotide kinase (New England Biolabs) (37).

In the initial screen, 500,000 recombinant clones in λgt11 yeast cDNA library were plated on E. coli Y1088. Duplicate transfers of the clones were made onto nitrocellulose, and the filters were prepared for hybridization (37). Afterward, the filters were washed twice at room temperature in 6xSSC (0.15 M NaCl/15 mM sodium citrate (NaCl/Cit) containing 0.1% SDS and 0.05% NaP Pi ), washed once at 50°C below the minimum tº (temperature of probe dissociation based on G/C content), and exposed on x-ray film for 1 to 2 days. Maximum and minimum tº were determined for two pools of redundant oligonucleotide probes (N3 and N4) (38).

DNA Sequencing and Blot Analysis.

cDNA fragments were cleaved out from recombinant λgt11 phage DNA by EcoRI digestion. The cDNA fragments were separated by gel electrophoresis in low melting point agarose. The correct DNA band was sliced out, the gel was melted at 65°C, and the DNA was extracted with phenol. The purified cDNA fragments were cloned into the Bluescript plasmid (Stratagene). The complete sequence of the yeast Nα-acetyltransferase cDNA was determined by exonuclease III deletion (39), the dideoxy chain termination method of Sanger (40) modified for double-stranded sequencing by Guo et al. (41), and specific priming with synthetic oligonucleotides. All restriction enzymes were purchased from New England Biolabs. RNA and DNA markers were obtained from Bethesda Research Laboratories. Biotrans nylon membrane was from ICN. Poly(A)+RNA was analyzed by RNA hybridization (42,43). Genomic DNA was isolated from yeast (27), digested with restriction enzymes, and analyzed by DNA blot hybridization (44). The chromosome bearing the αA/αA′ gene was identified by hybridization of labelled cDNA with a Saccharomyces chromo-di-hybridizor (Clonetech) (i.e., a yeast chromosomal agarose gel).
Figure 1. HPLC separation of yeast Nα-acetyltransferase tryptic peptides. Three nanomoles of purified Nα-acetyltransferase was reduced, alkylated, digested with trypsin, and chromatographed on a 0.46 x 25 cm Vydac phenyl HPLC column with 0.1% CF₃COOH in a linear gradient of 0-60% CH₃CN over 2 hr. Numbers refer to the tryptic peptides, the sequences of which are shown in Fig. 2C.
Molecular Cloning and Sequencing of a cDNA Encoding N*-Acetyltransferase and a 16-residue (peptide 27-3) peptide and the known codon-usage frequencies (34) (Fig. 2A). In addition, two degenerate oligonucleotide probes of 23 (5'CCXTTGATYTTYTTRCAAGAYAAA3') and 20 (3'TCRTCRTGCA-TYTGRAARTA5') bases with 64- and 32-fold redundancy, designated N3 and N4, were synthesized based on sequence data for peptides 29-1 and 10-3-1, respectively.

Cloning of the Yeast N*-Acetyltransferase cDNA—After initial screening of 500,000 recombinant cDNA clones in the yeast Xgtll cDNA library, 11 clones were detected which hybridized to both oligonucleotides N1 and N2. These clones, designated AN1 to XN11, also hybridized with oligonucleotide N3 and N4, and their cDNA inserts were analyzed by restriction enzyme digestions and DNA blot analyses. The use of four oligonucleotide probes derived from four discrete amino acid sequences allowed the unequivocal identification of the cDNA clones encoding N*-acetyltransferase. EcoRI digestion revealed inserts that lacked internal EcoRI sites and ranged from 2.0 to 2.7 kilobase pairs. The six longest cDNA inserts were subcloned as EcoRI fragments into the Bluescript plasmid, and additional restriction enzyme mapping, DNA blot analyses, and nucleotide sequence analyses were carried out.

+*-
-4-4
-4+4-
44

FIG. 2. Cloning and sequencing of the cDNA encoding yeast N*-acetyltransferase. A. oligonucleotide probes used for initially screening the Xgtll library. The amino acid sequences of two tryptic peptides were used to construct the codon-usage frequency based oligonucleotide probes. The nucleotide positions indicated by the asterisks differ from the actual DNA sequence shown in C. The numbering of the tryptic peptides is as follows: the first number refers to the corresponding peak in Fig. 1, the second number refers to the peak in the first isocratic HPLC separation (data not shown), and the third number refers to the peak in the second isocratic HPLC separation (data not shown). B. restriction map and DNA sequencing strategy for the cDNA clones. The arrows indicate the direction and extent of sequence determination for each fragment after exonuclease III digestion. C. nucleotide and deduced amino acid sequence of N*-acetyltransferase cDNA clones. The amino acid sequences of HPLC-purified tryptic peptides determined by automated protein sequence analysis are also shown. The protein sequence analyses were completed with repetitive yields between 87 and 93% for 100-200 pmol of each peptide.
Fig. 2. Cloning and sequencing of the cDNA encoding yeast N°-acetyltransferase. A, oligonucleotide probes used for initially screening the λgt11 library. The amino acid sequences of two tryptic peptides were used to construct the codon-usage frequency based oligonucleotide probes. The nucleotide positions indicated by the asterisks differ from the actual DNA sequence shown in C. The numbering of the tryptic peptides is as follows: the first number refers to the corresponding peak in Fig. 1, the second number refers to the peak in the first isocratic HPLC separation (data not shown), and the third number refers to the peak in the second isocratic HPLC separation (data not shown). B, restriction map and DNA sequencing strategy for the cDNA clones. The arrows indicate the direction and extent of sequence determination for each fragment after exonuclease III deletion. C, nucleotide and deduced amino acid sequence of N°-acetyltransferase cDNA clones. The amino acid sequences of HPLC-purified tryptic peptides determined by automated protein sequence analysis are also shown. The protein sequence analyses were completed with repetitive yields between 87 and 93% for 100–200 pmol of each peptide.
Sequence data

Peptide 27-3  K I E G A S A S P I C C H V L G I Y M
Probe N1  3' - TTC TAA CTT CCA CGA AGA CGA AGG GGT TAG ACG AGC GTG CGA AAC CCA TAG ATG TAC -5'  

Peptide 11-3-2  S V A A Y P S D Q D W N D V F G E
Probe N2  3' - AGA CAA CGA CGA ATG GGT AGA CTG GTT CTG TTG CTG CAA AAG CCA CTT -5'  

Diagram:

- EcoRV
- SacI
- BamHI
- EcoRV
- BatXI
- XhoI
- HindIII
- HindIII
- BatXI

0.2 kb
Molecular Cloning and Sequencing of a cDNA Encoding N*-Acetyltransferase and a 16-residue peptide and the known codon-usage frequencies (34) (Fig. 2A). In addition, two degenerate oligonucleotide probes of 23 (5' CCXTTGA-CYTTYTTRCAAGAYAA 3') and 20 (3' TCRTCRTGCA-TYTGRAARTA 5') bases with 64- and 32-fold redundancy, designated N3 and N4, were synthesized based on sequence data for peptides 29-1 and 10-3-1, respectively.

Cloning of the Yeast N*-Acetyltransferase cDNA—After initial screening of 500,000 recombinant cDNA clones in the yeast Xgtll cDNA library, 11 clones were detected which hybridized to both oligonucleotides N1 and N2. These clones, designated AN1 to XN11, also hybridized with oligonucleotide N3 and N4, and their cDNA inserts were analyzed by restriction enzyme digestions and DNA blot analyses. The use of four oligonucleotide probes derived from four discrete amino acid sequences allowed the unequivocal identification of the cDNA clones encoding N*-acetyltransferase. EcoRI digestion revealed inserts that lacked internal EcoRI sites and ranged from 2.0 to 2.7 kilobase pairs. The six longest cDNA inserts were subcloned as EcoRI fragments into the Bluescript plasmid, and additional restriction enzyme mapping, DNA blot analyses, and nucleotide sequence analyses were carried out.
Figure 3. Autoradiogram of Northern blot analysis of yeast poly(A)$^+$ RNA. Yeast poly(A)$^+$ RNA (10 μg) was electrophoresed on a 1.2% agarose/formaldehyde gel (33). The mRNA was transferred onto a nylon membrane and hybridized with random primed, $[^{32}P]$-cDNA (derived from pBN1) for 24 hr and washed (42,43). The gel lane containing the RNA markers was sliced out, visualized by staining with ethidium bromide, and used for determining the molecular size of the yeast poly(A)$^+$ RNA.
Figure 4. Autoradiogram of Southern blot analysis of restriction fragments of yeast genomic DNA. Yeast DNA (10 μg) was digested with indicated restriction enzymes. The restriction fragments were electrophoresed on a 0.8% agarose gel in Tris-borate buffer. The DNA was transferred onto a nylon membrane and hybridized with random primed, [32P]-cDNA (derived from pBN1) for 24 hr and washed (44).
<table>
<thead>
<tr>
<th>Origin</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome #</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td></td>
</tr>
<tr>
<td>XV</td>
<td></td>
</tr>
<tr>
<td>XVI</td>
<td></td>
</tr>
<tr>
<td>XIII</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
</tr>
<tr>
<td>XIV</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>XI</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Chromosome identification of yeast N<sup>α</sup>-acetyltransferase gene. An agarose gel of yeast chromosomal DNA was hybridized with random primed, [<sup>32</sup>P]-cDNA (derived from pBN1) for 24 hr and washed according to the manufacturer's recommendations. The position of the individual yeast chromosomes on the gel is indicated by the numbers.
Figure 6. Hydrophobicity plot of yeast Nα-acetyltransferase. The plot was calculated by the method of Kyte and Doolittle (51) with a window of 9.