THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 264, No. 21, Issue of July 25, pp. 12339-12343, 1989
Printed in U.S.A.

Molecular Cloning and Sequencing of a cDNA Encoding N^{α}-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 α-NH_2 group of eukaryotic proteins and is catalyzed by an N^{α}-acetyltransferase. Recently, a eukaryotic N^{α}-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 λgt11 cDNA library and sequencing of a full length cDNA encoding yeast N^{α}-acetyltransferase. DNA blot hybridizations of genomic and chromosomal DNA reveal that the gene (so-called *AAA1*, *a*mino-terminal, *α*-*a*mino, *a*cetyltransferase) 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^{α}-acetylation in the sorting and degradation of eukaryotic proteins to be determined.

Experimental Procedures

Protein Sequence Analyses of N^{α}-acetyltransferase

N^{α}-acetyltransferase was purified from yeast as previously described (24). N^{α}-acetyltransferase (3 nmoles) was reduced and alkylated, precipitated with cold chloroform/ methanol, redissolved in 0.1 M NH_4HCO_3, 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% CF_3COOH 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 [^{32}P]-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 polyacryamide (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' CTGGTCTAGAG 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) (Jen 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-^{32}P]-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% NaPPi), washed once at 5°C below the minimum t_d (temperature of probe dissociation based on G/C content), and exposed on x-ray film for 1 to 2 days. Maximum and minimum t_d 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 N^{α}-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 *AAA1* gene was identified by hybridization of labelled cDNA with a Saccharomyces chromo-di-hybridizer (Clonetech) (i.e., a yeast chromosomal agarose gel).

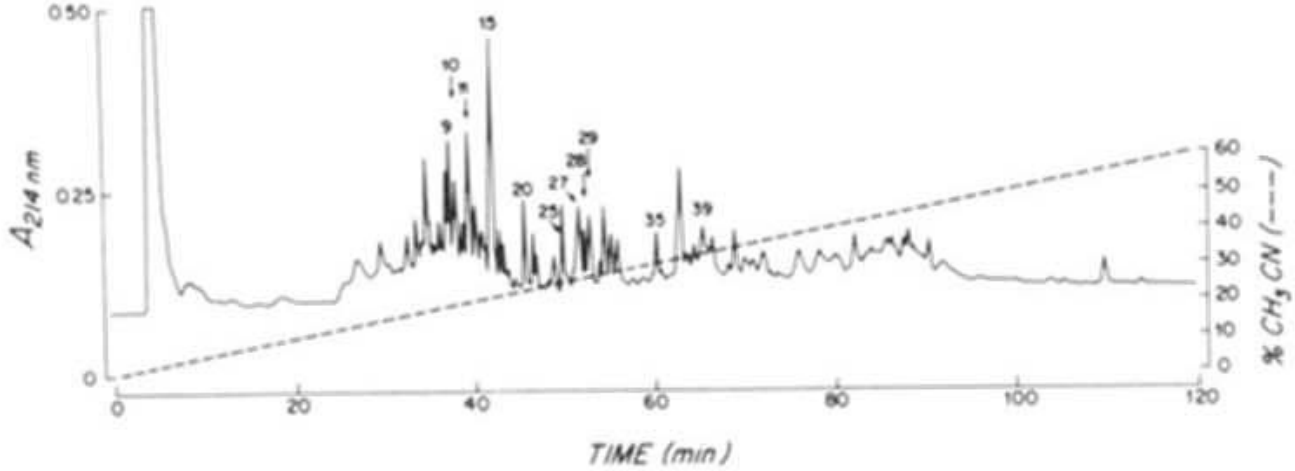


Figure 1. HPLC separation of yeast N$^{\alpha}$-acetyltransferase tryptic peptides. Three nanomoles of purified N$^{\alpha}$-acetyltransferase was reduced, alkylated, digested with trypsin, and chromatographed on a 0.46 x 25 cm Vydac phenyl HPLC column with 0.1% CF_3COOH in a linear gradient of 0-60% CH_3CN over 2 hr. Numbers refer to the tryptic peptides, the sequences of which are shown in Fig. 2C.

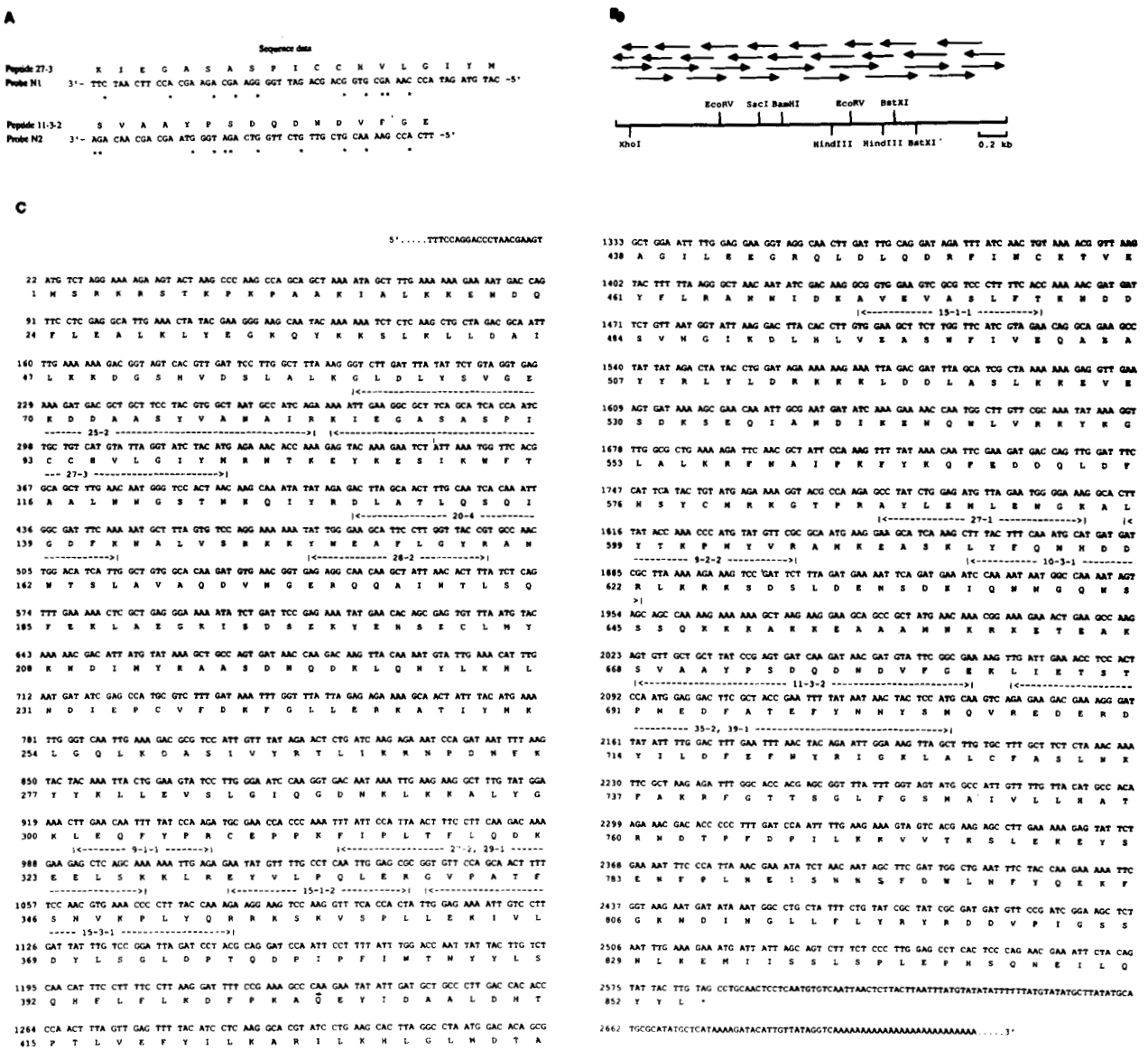


FIG. 2. **Cloning and sequencing of the cDNA encoding yeast N^{α}-acetyltransferase.** *A*, oligonucleotide probes used for initially screening the λgt11 library. The amino acid sequences of two tryptic peptides were used to construct the codon-usage frequency based oligonucleotide probes. The nucleotide 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^{α}-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.

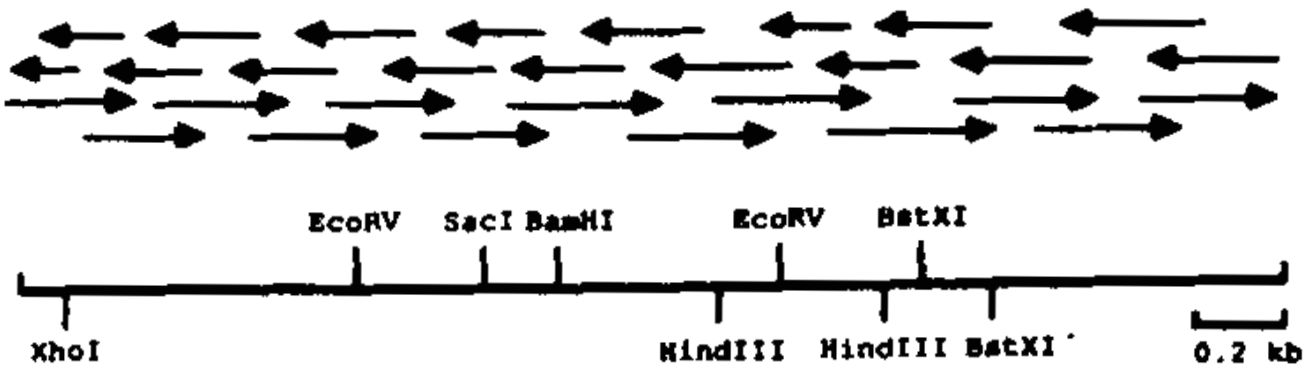
FIG. 2. **Cloning and sequencing of the cDNA encoding yeast N^{α}-acetyltransferase.** *A*, oligonucleotide probes used for initially screening the λgt11 library. The amino acid sequences of two tryptic peptides were used to construct the codon-usage frequency based oligonucleotide probes. The nucleotide 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^{α}-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.

A

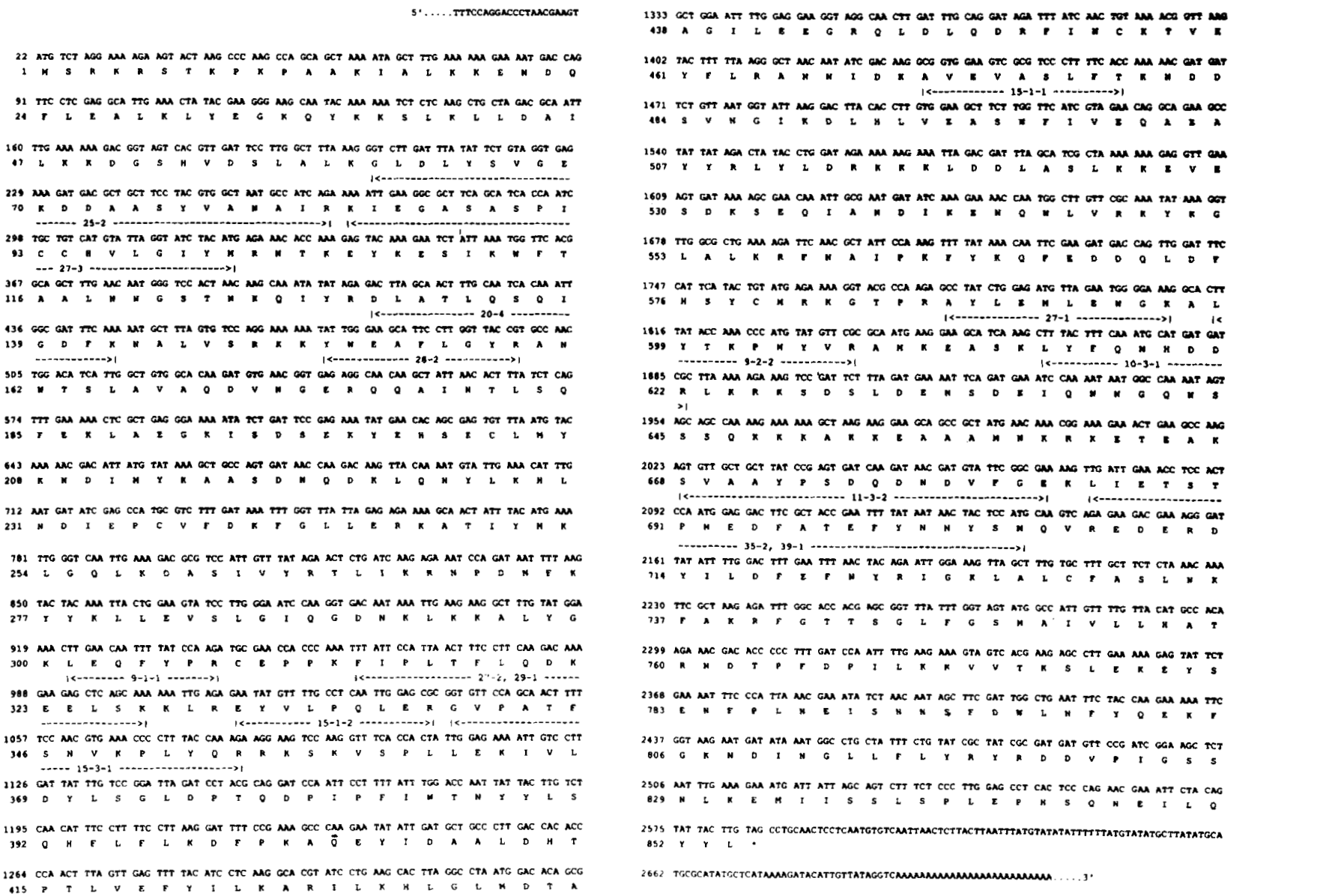
Sequence data

Peptide 27-3		K	I	E	G	A	S	A	S	P	I	C	C	N	V	L	G	I	Y	M
Probe N1	3'-	TTC	TAA	CTT	CCA	CGA	AGA	CGA	AGG	GGT	TAG	ACG	ACG	GTG	CGA	AAC	CCA	TAG	ATG	TAC	-5'
			*			*		*	*	*				*	*	**	*				
Peptide 11-3-2		S	V	A	A	Y	P	S	D	Q	D	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'			
		**				*	*	**	*		*		*	*		*					

B



c


5'.....TTTCCAGGACCCTAACGAAGT
22 ATG TCT AGG AAA AGA AGT ACT AAG CCC AAG CCA GCA GCT AAA ATA GCT TTG AAA AAA GAA AAT GAC CAG
1 M S R K R S T K P K P A A K I A L K K E N D Q
91 TTC CTC GAG GCA TTG AAA CTA TAC GAA GGG AAG CAA TAC AAA AAA TCT CTC AAG CTG CTA GAC GCA ATT
24 F L E A L K L Y E G K Q Y K K S L K L L D A I
160 TTG AAA AAA GAC GGT AGT CAC GTT GAT TCC TTG GCT TTA AAG GGT CTT GAT TTA TAT TCT GTA GGT GAG
47 L K K D G S H V D S L A L K G L D L Y S V G E
229 AAA GAT GAC GCT GCT TCC TAC GTG GCT AAT GCC ATC AGA AAA ATT GAA GGC GCT TCA GCA TCA CCA ATC
70 K D D A A S Y V A N A I R K I E G A S A S P I
298 TGC TGT CAT GTA TTA GGT ATC TAC ATG AGA AAC ACC AAA GAG TAC AAA GAA TCT ATT AAA TGG TTC ACG
93 C C H V L G I Y M R N T K E Y K E S I K W F T
367 GCA GCT TTG AAC AAT GGG TCC ACT AAC AAG CAA ATA TAT AGA GAC TTA GCA ACT TTG CAA TCA CAA ATT
116 A A L N N G S T N K Q I Y R D L A T L Q S Q I
436 GGC GAT TTC AAA AAT GCT TTA GTG TCC AGG AAA AAA TAT TGG GAA GCA TTC CTT GGT TAC CGT GCC AAC
139 G D F K N A L V S R K K Y W E A F L G Y R A N
505 TGG ACA TCA TTG GCT GTG GCA CAA GAT GTG AAC GGT GAG AGG CAA CAA GCT ATT AAC ACT TTA TCT CAG
162 W T S L A V A Q D V N G E R Q Q A I N T L S Q
574 TTT GAA AAA CTC GCT GAG GGA AAA ATA TCT GAT TCC GAG AAA TAT GAA CAC AGC GAG TGT TTA ATG TAC
185 F E K L A E G K I S D S E K Y E H S E C L M Y
643 AAA AAC GAC ATT ATG TAT AAA GCT GCC AGT GAT AAC CAA GAC AAG TTA CAA AAT GTA TTG AAA CAT TTG
208 K N D I M Y K A A S D N Q D K L Q N Y L K H L
712 AAT GAT ATC GAG CCA TGC GTC TTT GAT AAA TTT GGT TTA TTA GAG AGA AAA GCA ACT ATT TAC ATG AAA
231 N D I E P C V F D K F G L L E R K A T I Y M K
781 TTG GGT CAA TTG AAA GAC GCG TCC ATT GTT TAT AGA ACT CTG ATC AAG AGA AAT CCA GAT AAT TTT AAG
254 L G Q L K D A S I V Y R T L I K R N P D N F K
850 TAC TAC AAA TTA CTG GAA GTA TCC TTG GGA ATC CAA GGT GAC AAT AAA TTG AAG AAG GCT TTG TAT GGA
277 Y Y K L L E V S L G I Q G D N K L K K A L Y G
919 AAA CTT GAA CAA TTT TAT CCA AGA TGC GAA CCA CCC AAA TTT ATT CCA TTA ACT TTC CTT CAA GAC AAA
300 K L E Q F Y P R C E P P K F I P L T F L Q D K
988 GAA GAG CTC AGC AAA AAA TTG AGA GAA TAT GTT TTG CCT CAA TTG GAG CGC GGT GTT CCA GCA ACT TTT
323 E E L S K K L R E Y V L P Q L E R G V P A T F
1057 TCC AAC GTG AAA CCC CTT TAC CAA AGA AGG AAG TCC AAG GTT TCA CCA CTA TTG GAG AAA ATT GTC CTT
346 S N V K P L Y Q R R K S K V S P L L E K I V L
1126 GAT TAT TTG TCC GGA TTA GAT CCT ACG CAG GAT CCA ATT CCT TTT ATT TGG ACC AAT TAT TAC TTG TCT
369 D Y L S G L D P T Q D P I P F I W T N Y Y L S
1195 CAA CAT TTC CTT TTC CTT AAG GAT TTT CCG AAA GCC CAA GAA TAT ATT GAT GCT GCC CTT GAC CAC ACC
392 Q H F L F L K D F P K A Q E Y I D A A L D H T
1264 CCA ACT TTA GTT GAG TTT TAC ATC CTC AAG GCA CGT ATC CTG AAG CAC TTA GGC CTA ATG GAC ACA GCG
415 P T L V E F Y I L K A R I L K H L G L M D T A
1333 GCT GGA ATT TTG GAG GAA GGT AGG CAA CTT GAT TTG CAG GAT AGA TTT ATC AAC TGT AAA ACG GTT AAG
438 A G I L E E G R Q L D L Q D R F I N C K T V K
1402 TAC TTT TTA AGG GCT AAC AAT ATC GAC AAG GCG GTG GAA GTC GCG TCC CTT TTC ACC AAA AAC GAT GAT
461 Y F L R A N N I D K A V E V A S L F T K N D D
1471 TCT GTT AAT GGT ATT AAG GAC TTA CAC CTT GTG GAA GCT TCT TGG TTC ATC GTA GAA CAG GCA GAA GCC
484 S V N G I K D L H L V E A S W F I V E Q A E A
1540 TAT TAT AGA CTA TAC CTG GAT AGA AAA AAG AAA TTA GAC GAT TTA GCA TCG CTA AAA AAA GAG GTT GAA
507 Y Y R L Y L D R K K K L D D L A S L K K E V E
1609 AGT GAT AAA AGC GAA CAA ATT GCG AAT GAT ATC AAA GAA AAC CAA TGG CTT GTT CGC AAA TAT AAA GGT
530 S D K S E Q I A N D I K E N Q W L V R K Y K G
1678 TTG GCG CTG AAA AGA TTC AAC GCT ATT CCA AAG TTT TAT AAA CAA TTC GAA GAT GAC CAG TTG GAT TTC
553 L A L K R F N A I P K F Y K Q F E D D Q L D F
1747 CAT TCA TAC TGT ATG AGA AAA GGT ACG CCA AGA GCC TAT CTG GAG ATG TTA GAA TGG GGA AAG GCA CTT
576 H S Y C M R K G T P R A Y L E M L E W G K A L
1816 TAT ACC AAA CCC ATG TAT GTT CGC GCA ATG AAG GAA GCA TCA AAG CTT TAC TTT CAA ATG CAT GAT GAT
599 Y T K P M Y V R A M K E A S K L Y F Q M H D D
1885 CGC TTA AAA AGA AAG TCC GAT TCT TTA GAT GAA AAT TCA GAT GAA ATC CAA AAT AAT GGC CAA AAT AGT
622 R L K R K S D S L D E N S D E I Q M N G Q N S
1954 AGC AGC CAA AAG AAA AAA GCT AAG AAG GAA GCA GCC GCT ATG AAC AAA CGG AAA GAA ACT GAA GCC AAG
645 S S Q K K K A K K E A A A M N K R K E T E A K
2023 AGT GTT GCT GCT TAT CCG AGT GAT CAA GAT AAC GAT GTA TTC GGC GAA AAG TTG ATT GAA ACC TCC ACT
668 S V A A Y P S D Q D N D V F G E K L I E T S T
2092 CCA ATG GAG GAC TTC GCT ACC GAA TTT TAT AAT AAC TAC TCC ATG CAA GTC AGA GAA GAC GAA AGG GAT
691 P M E D F A T E F Y N N Y S M Q V R E D E R D
2161 TAT ATT TTG GAC TTT GAA TTT AAC TAC AGA ATT GGA AAG TTA GCT TTG TGC TTT GCT TCT CTA AAC AAA
714 Y I L D F E F N Y R I G K L A L C F A S L N K
2230 TTC GCT AAG AGA TTT GGC ACC ACG AGC GGT TTA TTT GGT AGT ATG GCC ATT GTT TTG TTA CAT GCC ACA
737 F A K R F G T T S G L F G S M A I V L L H A T
2299 AGA AAC GAC ACC CCC TTT GAT CCA ATT TTG AAG AAA GTA GTC ACG AAG AGC CTT GAA AAA GAG TAT TCT
760 R N D T P F D P I L K K V V T K S L E K E Y S
2368 GAA AAT TTC CCA TTA AAC GAA ATA TCT AAC AAT AGC TTC GAT TGG CTG AAT TTC TAC CAA GAA AAA TTC
783 E N F P L N E I S N N S F D W L N F Y Q E K F
2437 GGT AAG AAT GAT ATA AAT GGC CTG CTA TTT CTG TAT CGC TAT CGC GAT GAT GTT CCG ATC GGA AGC TCT
806 G K N D I N G L L F L Y R Y R D D V P I G S S
2506 AAT TTG AAA GAA ATG ATT ATT AGC AGT CTT TCT CCC TTG GAG CCT CAC TCC CAG AAC GAA ATT CTA CAG
829 N L K E M I I S S L S P L E P H S Q N E I L Q
2575 TAT TAC TTG TAG CCTGCAACTCCTCAATGTGTCAATTAACTCTTACTTAATTTATGTATATATTTTTTATGTATATGCTTATATGCA
852 Y Y L *
2662 TGCGCATATGCTCATAAAAGATACATTGTTATAGGTCAAAAAAAAAAAAAAAAAAAAAAAAAAA.....3'
|<---------- 25-2 ----------
|<---------- 20-4 ----------
|<---------- 26-2 ----------
|<---------- 15-1-1 ----------
|<---------- 27-1 ----------
|<---------- 9-2-2 ----------
|<---------- 10-3-1 ----------
|<---------- 11-3-2 ----------
|<---------- 35-2, 39-1 ----------
|<---------- 9-1-1 ----------
|<---------- 27-2, 29-1 ----------
|<---------- 15-1-2 ----------
|<---------- 15-3-1 ----------

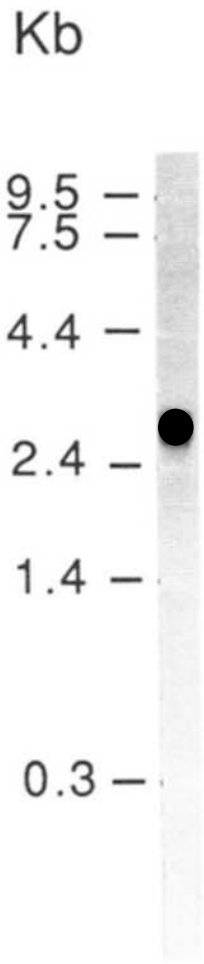


Figure 3. Autoradiogram of Northern blot analysis of yeast poly(A)$^+$ RNA. Yeast poly(A)$^+$ RNA (10 μg) was electrophoresed on a 1.2% agarose/formaldehyde gel (33). The mRNA was transferred onto a nylon membrane and hybridized with random primed, [^{32}P]-cDNA (derived from pBN1) for 24 hr and washed (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.

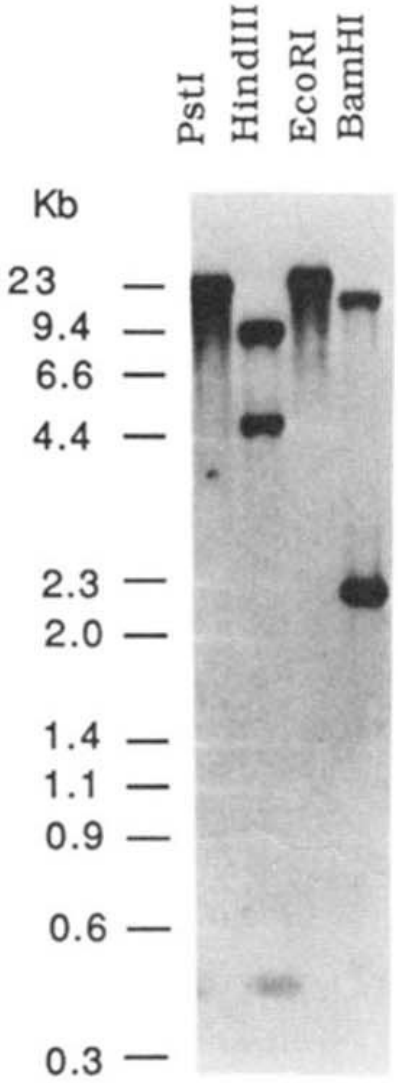


Figure 4. Autoradiogram of Southern blot analysis of restriction fragments of yeast genomic DNA. Yeast DNA (10 μg) was digested with indicated restriction enzymes. The restriction fragments were electrophoresed on a 0.8% agarose gel in Tris-borate buffer. The DNA was transferred onto a nylon membrane and hybridized with random primed, [^{32}P]-cDNA (derived from pBN1) for 24 hr and washed (44).

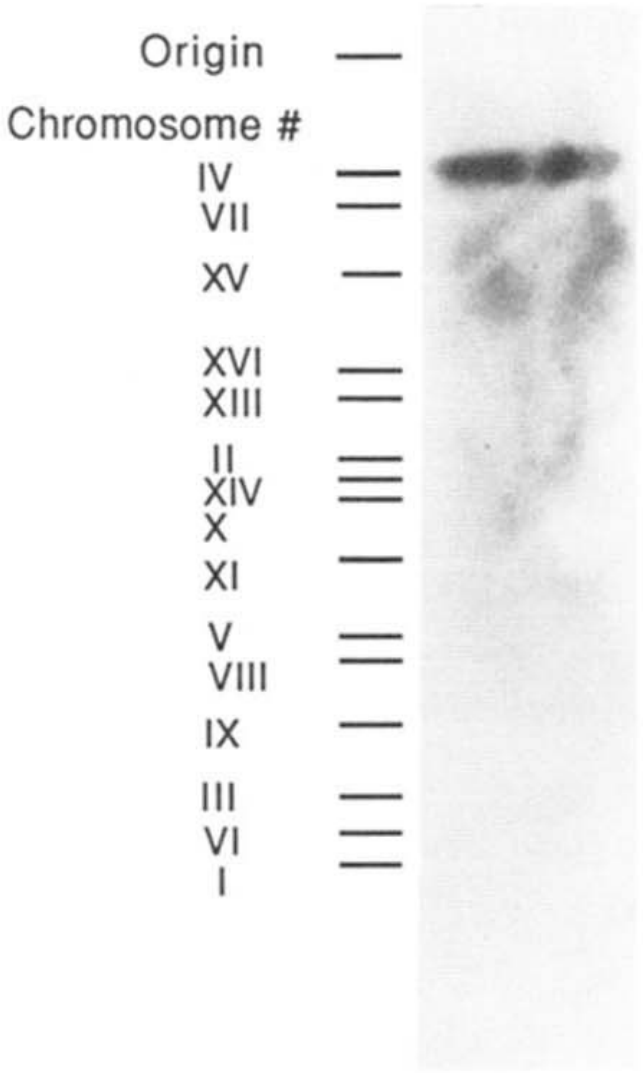


Figure 5. Chromosome identification of yeast N$^{\alpha}$-acetyltransferase gene. An agarose gel of yeast chromosomal DNA was hybridized with random primed, [^{32}P]-cDNA (derived from pBN1) for 24 hr and washed according to the manufacturer's recommendations. The position of the individual yeast chromosomes on the gel is indicated by the numbers.

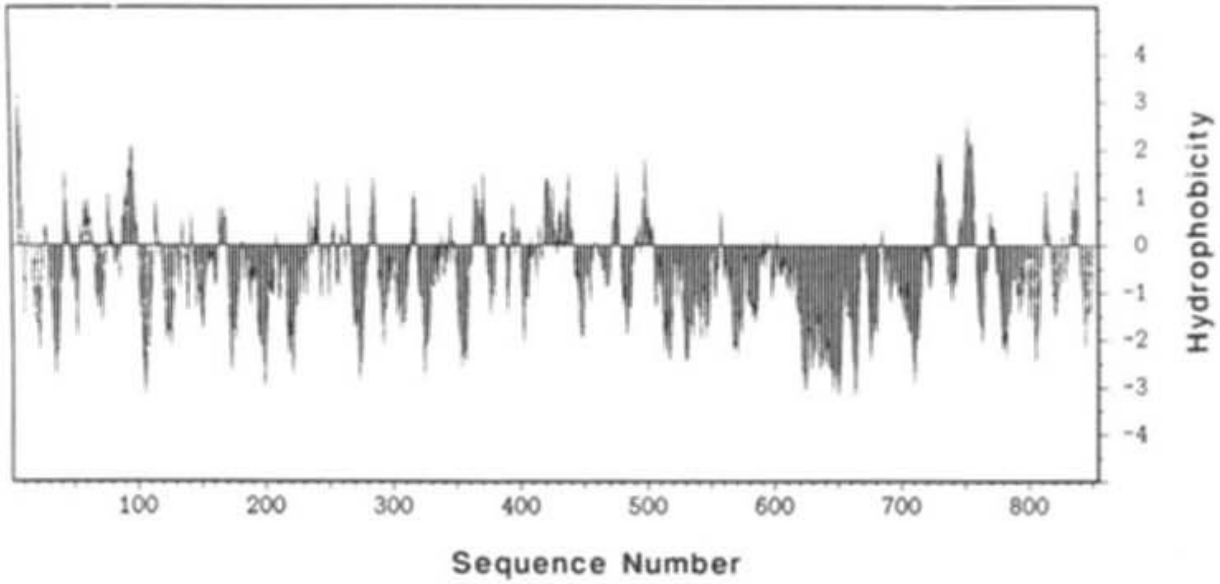


Figure 6. Hydrophobicity plot of yeast N^{α}-acetyltransferase. The plot was calculated by the method of Kyte and Doolittle (51) with a window of 9.