Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4081–4085, June 1986
Neurobiology

Cloning of *Drosophila* choline acetyltransferase cDNA

(expression library/monoclonal antibody/amino acid sequence/DNA sequence/polytene chromosomes)

NOBUYUKI ITOH*†, J. RANDALL SLEMMON*, DAVID H. HAWKE‡, RODNEY WILLIAMSON§, ELAINE MORITA¶, KEIICHI ITAKURA¶, EUGENE ROBERTS||, JOHN E. SHIVELY‡, GARRETT D. CRAWFORD*, AND PAUL M. SALVATERRA* **

Divisions of *Neurosciences, ‡Immunology, and §Biology and Departments of ¶Molecular Genetics and ||Neurobiochemistry, Beckman Research Institute of the City of Hope, Duarte, CA 91010

Communicated by Rachmiel Levine, January 21, 1986

ABSTRACT **Choline acetyltransferase (EC 2.3.1.6) is the biosynthetic enzyme for the neurotransmitter acetylcholine. To isolate choline acetyltransferase cDNA clones, a cDNA library was constructed from poly(A)$^+$ RNA of *Drosophila melanogaster* heads, these being one of the richest known sources of the enzyme. By screening the cDNA library with a mixture of three different monoclonal antibodies to *Drosophila* choline acetyltransferase, we isolated 14 positive clones. Only 1 of these clones was identified to be a *Drosophila* choline acetyltransferase cDNA clone based on the following evidence. (*i*) The amino acid sequence deduced from the nucleotide sequence of the cDNA insert completely corresponded to that of several tryptic peptides from choline acetyltransferase. (*ii*) The cDNA insert hybridized specifically to only the region on *Drosophila* polytene chromosomes that had been identified as the site of the choline acetyltransferase (*Cha*) gene by cytogenetic analysis. The cDNA insert consisted of a coding region 2190 nucleotides long, a 3′-noncoding region 284 nucleotides long, and *Eco*RI linkers. RNA analysis of *Drosophila* head poly(A)$^+$ RNA with the cDNA insert as a probe showed the choline acetyltransferase mRNA to be ≈4700 nucleotides long.**

Choline acetyltransferase (EC 2.3.1.6) is the biosynthetic enzyme for acetylcholine, which is generally accepted as a neurotransmitter, at both peripheral and central cholinergic synapses (1). Although cholinergic phenotype expression includes a number of proteins, choline acetyltransferase is a phenotypically specific marker for cholinergic neurons. Consequently, this enzyme is an important target for studying the development of cholinergic neurons and their acquisition of this phenotype. The study of cholinergic phenotype expression in *Drosophila melanogaster* offers several significant advantages over other animals, since acetylcholine is likely to be an important sensory neurotransmitter in *Drosophila* (2), high levels of choline acetyltransferase are present in the *Drosophila* nervous system (3), and characterized primary embryonic cultures that contain developing cholinergic cells are available (4). In addition, *Drosophila* is well suited to genetic studies. Such studies have already mapped the locus for choline acetyltransferase activity on the polytene chromosomes (5), and enzyme mutants have been isolated (5).

Although cholinergic phenotype expression has been shown to be regulated by environmental factors (6, 7), the mechanisms that regulate gene transcription, control translation, and direct the posttranslational relocation of this enzyme to the synapse are not known. Evidence obtained from the earlier isolation of *Drosophila* choline acetyltransferase (8, 9) and a more recent study employing *Drosophila* mutants (10) have indicated that this enzyme probably undergoes posttranslational modification by limited proteolysis. Unfortunately, the study of this processing and the control of expression for this important enzyme has been hampered by a paucity of structural information due to the low levels of enzyme protein present in nervous tissue. The isolation of a cDNA clone for choline acetyltransferase is expected to overcome this problem and greatly enhance our ability to understand the control of neuronal phenotype expression. We describe here the isolation and characterization of a cDNA clone for *Drosophila* choline acetyltransferase that was recovered from an expression library by using our previously described monoclonal antibodies to this enzyme (11).

MATERIALS AND METHODS

Construction of cDNA Library. Total RNA was extracted from adult *Drosophila* (Canton S) heads by the urea/LiCl method (12). Poly(A)$^+$ RNA was prepared by oligo(dT)-cellulose (Pharmacia, type 7) chromatography (13). Double-stranded cDNA was synthesized from 16 μg of the poly(A)$^+$ RNA as a template by using avian myeloblastosis virus reverse transcriptase, *Escherichia coli* DNA polymerase I (Klenow fragment), and S1 nuclease (14). The double-stranded cDNA was treated with *Eco*RI methylase and *E. coli* DNA polymerase I (Klenow fragment) (15). After ligation to *Eco*RI linkers, the double-stranded cDNA was digested with *Eco*RI and fractionated by Sephacryl S-1000 (Pharmacia) chromatography. Fractions containing double-stranded cDNA that were ≈800–5000 base pairs (bp) long were pooled and precipitated by ethanol. The double-stranded cDNA was ligated to dephosphorylated, *Eco*RI-cut λgt11 DNA with DNA ligase (15). The ligated λgt11 DNA was packaged *in vitro* by using a packaging mixture. The packaged library was amplified as plate lysates on agar plates (15). The library contained 1.3×10^7 independent phage. Ninety-eight percent of the phage produced white plaques on agar plates containing 5-bromo-4-chloro-3-indolyl β-D-galactoside and isopropyl β-D-thiogalactoside, indicating 98% of the phage were recombinants.

Screening the cDNA Library with Antibody. The library was plated on a lawn of *E. coli* 1090, induced with isopropyl β-D-thiogalactoside-soaked nitrocellulose filters (15), and screened for antigen-producing clones according to de Wet *et al.* (16) by using a mixture of three different monoclonal antibodies to *Drosophila* choline acetyltransferase (1C8, 1G4, 14) (11).

Recloning of cDNA Insert in pBR328 or pUC13 and DNA Sequencing. Phage DNA was prepared from positive clones by a plate-lysate method (17). The cDNA insert was excised

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bp, base pairs.

†Present address: Department of Biological Chemistry, Faculty of Pharmaceutical Science, Kyoto University, Kyoto, 606, Japan.
**To whom reprint requests should be addressed.

by digestion with *Eco*RI and recloned into pBR328 or pUC13. Appropriate restriction fragments from the insert were subcloned into M13 mp18 and M13 mp19 for sequencing by the dideoxy method (18). Ambiguous regions of the sequence were confirmed by the Maxam–Gilbert method (19). The cDNA sequence was aligned with known peptide sequence as indicated in Fig. 3.

Choline Acetyltransferase Purification and Amino Acid Sequencing. Choline acetyltransferase (156 μg) was purified from the heads (300 g) of *Drosophila* (Canton S) in the same manner as described (8). Seventy-six micrograms of the enzyme protein was oxidized and then subjected to digestion with trypsin according to the method of Yuan *et al.* (20). Separation of the fragments was accomplished with reverse-phase HPLC (18), which yielded ≈61 peaks. Several of these were subjected to sequence analysis according to the method of Hawke *et al.* (21, 22). In general, each peak represented ≈1 nmol of material, of which about one-third was used for sequence analysis.

***In Situ* Hybridization to Polytene Chromosomes.** *Drosophila* salivary gland polytene chromosome squashes were prepared for *in situ* hybridization as described (23). The DNA probe was nick-translated by using a nick-translation reagent kit (Bethesda Research Laboratories) and dATP [α-^{35}S] (1200 Ci/mmol; 1 Ci = 37 GBq; Amersham). Hybridization with ^{35}S-labeled DNA (3×10^8 cpm/μg) was carried out and hybrids were detected according to Pardue and Gall (23).

RNA Analysis. *Drosophila* head poly(A)$^+$ RNA was electrophoresed in a 1% agarose gel containing 5 mM methylmercuric hydroxide (24). The gel was stained with ethidium bromide (0.5 μg/ml) in 0.5 M ammonium acetate and washed twice with distilled H_2O for 15 min, followed by drying in a gel-slab dryer at 60°C. The DNA probe was nick-translated using a nick-translation reagent kit (Bethesda Research Laboratories) and [α-^{32}P]dCTP (5000 Ci/mmol; New England Nuclear). The dry gel was hybridized with the ^{32}P-labeled DNA probe (10 ng/ml, 5×10^8 cpm/μg) in a hybridization buffer (50% formamide, 0.9 M NaCl/50 mM sodium phosphate, pH 7.4/5 mM EDTA, 5× concentrated Denhart's solution, yeast RNA at 100 μg/ml, 0.1% $NaDodSO_4$) at 42°C for 20 hr. The gel was washed by carrying it through a sequence consisting of four times with 0.36 M NaCl/20 mM sodium phosphate/2 mM EDTA and 0.1% $NaDodSO_4$ at room temperature for 1 hr, twice with 18 mM NaCl/1 mM sodium phosphate/0.5 mM EDTA and 0.1% $NaDodSO_4$ at room temperature for 15 min, twice with 18 mM NaCl/1 mM sodium phosphate/0.5 mM EDTA and 0.1% $NaDodSO_4$ at 50°C for 15 min and once with 0.36 M NaCl/20 mM sodium phosphate/2 mM EDTA and 0.1% $NaDodSO_4$ at room temperature for 15 min.

RESULTS

Screening of cDNA Library. A *Drosophila* head cDNA library was screened for cDNA clones coding for *Drosophila* choline acetyltransferase by using a mixture of three different monoclonal antibodies (1C8, 1G4, 14) to the choline acetyltransferase. Fourteen positive clones were isolated by screening 6×10^6 phage. The positive clones were examined by using individual antibodies. The 1C8 and 1G4 antibodies could bind all of the 14 positive clones, but the 14 antibody could not (data not shown). Phage DNA was prepared from the positive clones, and the cDNA inserts were excised by digestion with *Eco*RI. The cDNA inserts ranged in size from approximately 1000 to 2500 bp (data not shown). The longest cDNA insert (≈2500 bp) was recloned into pBR328. The recombinant DNA was designated pCha-2.

***In Situ* Hybridization.** *In situ* hybridization to *Drosophila* salivary gland polytene chromosomes with ^{35}S-labeled pCha-2 showed it to hybridize specifically to only the region 91B-D

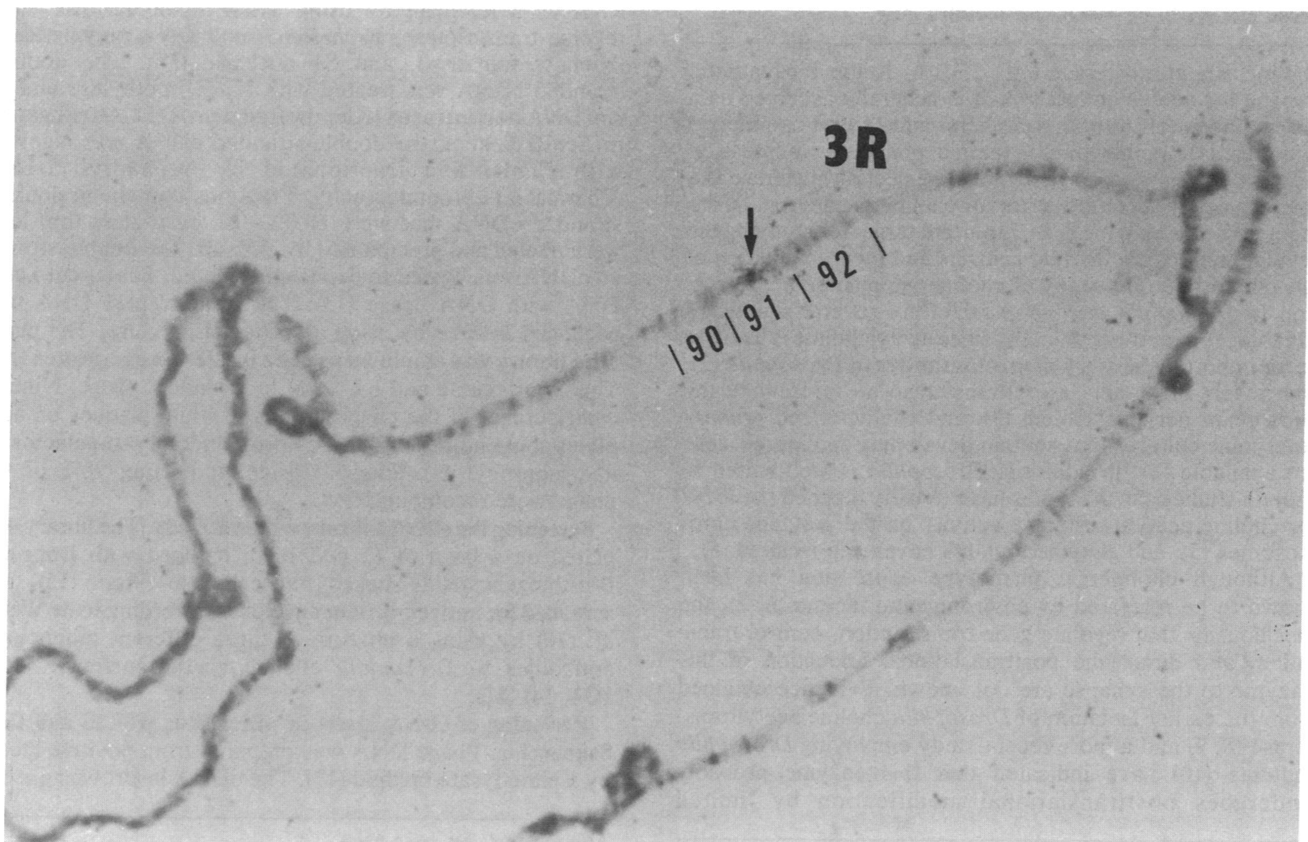

FIG. 1. *In situ* hybridization of ^{35}S-labeled pCha-2 to salivary gland polytene chromosomes. The chromosomes of *Drosophila* (Canton S) are shown. The arrow indicates the site of hybridization. No other site of hybridization was observed.

on polytene chromosome 3R (Fig. 1). This region had been identified as the site of the choline acetyltransferase (*Cha*) gene by cytogenetic analysis (5).

cDNA Nucleotide Sequence and Amino Acid Sequence of Tryptic Peptides from Choline Acetyltransferase. The cDNA insert was prepared from pCha-2, and the restriction map and nucleotide sequence were determined (Figs. 2 and 3). The cDNA insert was found to be composed of 2474 bp. We found a coding region 2190 nucleotides long and deduced the amino acid sequence (728 amino acid residues) from the nucleotide sequence. The insert also contained a 3′-noncoding region 284 nucleotides long and *Eco*RI linkers.

Purified *Drosophila* choline acetyltransferase was digested with L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin. The tryptic peptides from the choline acetyltransferase were resolved on reverse-phase HPLC (Fig. 4), and the amino acid sequences of 11 tryptic peptides were determined. The amino acid sequence of 11 of the tryptic peptides were found in the cDNA sequence and are indicated in Fig. 3.

RNA Analysis. *Drosophila* head poly(A)$^+$ RNA along with human rRNA for molecular weight markers were resolved in a 1% agarose gel containing methylmercuric hydroxide (Fig. 5*A*). The gel was dried and hybridized with ^{32}P-labeled pCha-2. Only one band was detected in the poly(A)$^+$ RNA (Fig. 5*B*). The size of the band was deduced to be ≈4700 nucleotides long from the RNA markers.

DISCUSSION

We initially screened *Drosophila* embryonic genomic and adult head cDNA libraries, constructed in pUC13, with several oligonucleotide probes that were chemically synthesized based on the amino acid sequence of the tryptic peptide 44 from choline acetyltransferase. We isolated positive clones, but, based on the nucleotide sequences of these clones, none of them was found to be a clone coding for choline acetyltransferase (unpublished observation). This is most likely due to the difficulties in screening a large number of colonies in a cDNA library, constructed by using a plasmid vector. Subsequently, we screened an adult *Drosophila* head cDNA library constructed in λgt11, using our monoclonal antibodies to *Drosophila* choline acetyltransferase, and were successful in recovering the clone for the enzyme described in this study. From the following evidence we conclude that the pCha-2 clone described here is a cDNA clone coding for *Drosophila* choline acetyltransferase. (*i*) Monoclonal antibodies to the choline acetyltransferase could bind the product of the recombinant λgt11 clone containing the cDNA insert. (*ii*) Extensive amino acid sequence (Fig. 3) deduced from the nucleotide sequence of the cDNA insert corresponded to that of 11 tryptic peptides (a total of 107 amino acids) from choline acetyltransferase. (*iii*) pCha-2 hybridized specifically only to the region of the *Drosophila* polytene chromosome that had been identified as the site of the choline acetyltransferase (*Cha*) gene by cytogenetic and enzymatic analysis (5).

Use of *Drosophila* to purify and clone choline acetyltrans-

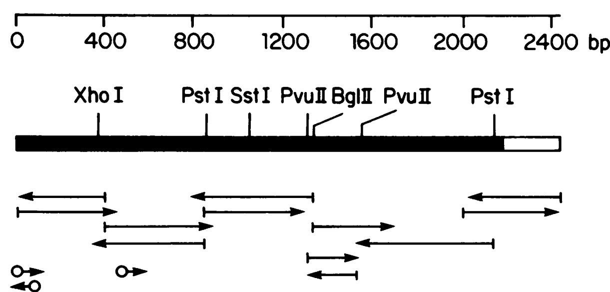

FIG. 2. Restriction map and sequencing strategy of the cDNA insert of pCha-2. Arrows (↦) and (o→) indicate the direction and extent of sequence determined by the dideoxy method and the Maxam–Gilbert method, respectively. The closed bar and open bar indicate the coding region and 3′-noncoding region, respectively.

```
                1                                  10
   AA TTC CGG ATT CCG GAT CCG AAA GGA GCG AAC GTG GCG TCC AAC GAG GCC AGC ACC AGC    59
              Ile Pro Asp Pro Lys Gly Ala Asn Val Ala Ser Asn Glu Ala Ser Thr Ser
        20                                   30
GCA GCG GGC AGT GGC CCG GAG TCC GCC GCC CTG TTC TCC AAG TTG CGT AGC TTC TCC ATT GGC    122
Ala Ala Gly Ser Gly Pro Glu Ser Ala Ala Leu Phe Ser Lys Leu Arg Ser Phe Ser Ile Gly
    40                                   50
AGC GGG CCC AAC TCG CCG CAG CGC GTC GTC TCC AAT CTC CGA GGA TTC CTC ACC CAT CGC CTC    185
Ser Gly Pro Asn Ser Pro Gln Arg Val Val Ser Asn Leu Arg Gly Phe Leu Thr His Arg Leu
60                                   70                                      80
AGC AAC ATC ACA CCG AGC GAT ACA GGA TGG AAA GAC TCG ATT CTG TCG ATA CCA AAG AAA TGG    248
Ser Asn Ile Thr Pro Ser Asp Thr Gly Trp Lys Asp Ser Ile Leu Ser Ile Pro Lys Lys Trp
                                90                                  100
CTC TCA ACG GCC GAG TCT GTG GAC GAG TTT GGA TTC CCT GAC ACT CTA CCC AAG GTG CCC GTT    311
Leu Ser Thr Ala Glu Ser Val Asp Glu Phe Gly Phe Pro Asp Thr Leu Pro Lys Val Pro Val    T44
                            110                                 120
CCA GCA CTG GAT GAA ACG ATG GCC GAC TAC ATC CGC GCC CTG GAA CCG ATT ACC ACT CCG GCG    374
Pro Ala Leu Asp Glu Thr Met Ala Asp Tyr Ile Arg Ala Leu Glu Pro Ile Thr Thr Pro Ala    T41
                        130                                 140
CAG CTC GAG CGG ACC AAG GAG CTG ATC AGG CAG TTC TCG GCT CCC CAG GGA ATC GGA GCG CGG    437
Gln Leu Glu Arg Thr Lys Glu Leu Ile Arg Gln Phe Ser Ala Pro Gln Gly Ile Gly Ala Arg
                    150                                 160
CTG CAT CAG TAT CTG CTG GAC AAG CGT GAG GCG AGG ATA ACT GGG CCT ATT ACT ACT GGC TCA    500
Leu His Gln Tyr Leu Leu Asp Lys Arg Glu Ala Arg Ile Thr Gly Pro Ile Thr Thr Gly Ser
                170                                 180
ACG AGA TGT ACA TGG ATA TTC GCA TTC CCT TTG CCG ATC AAC TCG AAT CCG GGC ATT GGT GTT    563
Thr Arg Cys Thr Trp Ile Phe Ala Phe Pro Leu Pro Ile Asn Ser Asn Pro Gly Ile Gly Val
                190                                 200
CCC GCC GCG TCG CTT CAA GAC CGT CCA CGA CGT GCC CAC TTC GCC GCT CGC CTG CTG GAC GGC    626
Pro Ala Ala Ser Leu Gln Asp Arg Pro Arg Arg Ala His Phe Ala Ala Arg Leu Leu Asp Gly    T45
            210                                 220
ATT CTG AGC CAC CGC GAG ATG CTG GAC AGT GGG GAG CTG CCG CTG GAG CGG GCG CTC GCG GAG    689
Ile Leu Ser His Arg Glu Met Leu Asp Ser Gly Glu Leu Pro Leu Phe Arg Ala Leu Ala Glu    T40
    230                                 240
AAG AAT CAG CCG CTG TGC ATG GCG CAG TAC TAC CGC CTG CTG GGC TCC TGT CGT CGG CCT GGT    752
Lys Asn Gln Pro Leu Cys Met Ala Gln Tyr Tyr Arg Leu Leu Gly Ser Cys Arg Arg Pro Gly    T17
    250                                 260
GTC AAG CAG GAC TCG CAG TTC CTG CCG TCG CGG GAG CGA CTG AAC GAC GAG GAT CGC CAT GTG    815
Val Lys Gln Asp Ser Gln Phe Leu Pro Ser Arg Glu Arg Leu Asn Asp Glu Asp Arg His Val
270                                 280                                     290
GTG GTT ATT TGC CGC AAC CAA ATG TAT TGC GTC GTG CTG CAG GCT AGC GAT CGT GGA AAG TTG    878
Val Val Ile Cys Arg Asn Gln Met Tyr Cys Val Val Leu Gln Ala Ser Asp Arg Gly Lys Leu    T35
                                300                                 310
TCG GAG AGT GAG ATC GCC TCA CAG ATC CTC TAT GTG CTC AGT GAT GCT CCC TGT CTG CCA GCT    941
Ser Glu Ser Glu Ile Ala Ser Gln Ile Leu Tyr Val Leu Ser Asp Ala Pro Cys Leu Pro Ala
                            320                                 330
AAA CCA GTG CCG GTG GGT CTG CTG ACC GCT GAA CCG AGG AGC ACG TGG GCA CGG GAC CGG GAA    1004
Lys Pro Val Pro Val Gly Leu Leu Thr Ala Glu Pro Arg Ser Thr Trp Ala Arg Asp Arg Glu
                        340                                 350
ATG CTT CAG GAG GAC GAA CGC AAT CAA CGC AAT CTG GAG CTC ATC GAG ACG GCA CAG GTG GTC    1067
Met Leu Gln Glu Asp Glu Arg Asn Gln Arg Asn Leu Glu Leu Ile Glu Thr Ala Gln Val Val
                    360                                 370
CTC TGT CTG GAC GAA CCG TTG GCT GGG AAC TTT AAT GCG CGC GGT TTT ACG GGT GCC ACG CCC    1130
Leu Cys Leu Asp Glu Pro Leu Ala Gly Asn Phe Asn Ala Arg Gly Phe Thr Gly Ala Thr Pro    T26
                380                                 390
ACA GTT CAT CGG GCG GGA GAT AGG GAC GAG ACG AAC ATG GCC CAC GAG ATG ATC CAC GGC GGA    1193
Thr Val His Arg Ala Gly Asp Arg Asp Glu Thr Asn Met Ala His Glu Met Ile His Gly Gly
            400                                 410
GGC AGC GAG TAC AAC TCC GGA AAT CGC TGG TTT GAC AAG ACC ATG CAG CTC ATT ATT TGC ACC    1256
Gly Ser Glu Tyr Asn Ser Gly Asn Arg Trp Phe Asp Lys Thr Met Gln Leu Ile Ile Cys Thr
        420                                 430
GAT GGA ACC TGG GGC CTT TGC TAT GAG CAC TCC TGT TCC GAA GGC ATT GCT GTT GTC CAG CTG    1319
Asp Gly Thr Trp Gly Leu Cys Tyr Glu His Ser Cys Ser Glu Gly Ile Ala Val Val Gln Leu
    440                                 450
CTG GAG AAG ATC TAC AAA AAA ATC GAG GAG CAC CCG GAC GAG GAT AAC GGT CTA CCG CAA CAC    1382
Leu Glu Lys Ile Tyr Lys Lys Ile Glu Glu His Pro Asp Glu Asp Asn Gly Leu Pro Gln His
    460                                 470
CAC TTG CCA CCA CCG GAG CGT CTG GAG TGG CAT GTG GGT CCG CAA TTG CAA TTG CGC TTT GCC    1445
His Leu Pro Pro Pro Glu Arg Leu Glu Trp His Val Gly Pro Gln Leu Gln Leu Arg Phe Ala
480                                 490                                     500
CAA GCC TCC AAG AGT GTG GAC AAA TGC ATC GAT GAC CTG GAC TTC TAT GTG TAC CGC TAC CAG    1508
Gln Ala Ser Lys Ser Val Asp Lys Cys Ile Asp Asp Leu Asp Phe Tyr Val Tyr Arg Tyr Gln    T13
                                510                                 520
AGT TAC GGA AAG ACC TTT ATC AAA TCG TGC CAG GTC AGT CCG GAT GTG TAC ATT CAG CTG GCA    1571
Ser Tyr Gly Lys Thr Phe Ile Lys Ser Cys Gln Val Ser Pro Asp Val Tyr Ile Gln Leu Ala
                                530                                 540
ACT GCA ACT GGC TCA CTA CAA GTT GTA CGG ACG TCT GGT GGC CAC CTA CGA AAG TGC GTC CAC    1634
Thr Ala Thr Gly Ser Leu Gln Val Val Arg Thr Ser Gly Gly His Leu Arg Lys Cys Val His
                        550                                 560
TCG ACG ATT TCT GCA CGG CCG CGT AGA CTG CAT CAG AGC GGC CAG CAC GGA GGC ATT GGA GTG    1697
Ser Thr Ile Ser Ala Arg Pro Arg Arg Leu His Gln Ser Gly Gln His Gly Gly Ile Gly Val
                    570                                 580
GGC CAG GCC ATG TGC CAG GGT GAG GGT GCA AAC GTG CCC CTG GAG AGC GAT CGC GAG GAT GAG    1760
Gly Gln Ala Met Cys Gln Gly Glu Gly Ala Asn Val Pro Leu Glu Ser Asp Arg Glu Asp Glu
                    590                                 600
GAG GAG TCG CGA AAG GTC AAG TTT AGC ATT TAC AGT AAG GAT CAT CTC CGT GAG CTT TTC CGG    1823
Glu Glu Ser Arg Lys Val Lys Phe Ser Ile Tyr Ser Lys Asp His Leu Arg Glu Leu Phe Arg
                610                                 620
TGC GCC GTC GCC CGC CAG ACT GAG GTG ATG GTG AGA ATA TCC TGG GCA ATG GCA TCG ACA TCC    1886
Cys Ala Val Ala Arg Gln Thr Glu Val Met Val Arg Ile Ser Trp Ala Met Ala Ser Thr Ser
        630                                 640
CGC TGC TGG CCT GCG AGA GGC CAG TAT AGA GGT CAC CGG CGA GAT GCA CGA GCT GTT CAA AGA    1949
Arg Cys Trp Pro Ala Arg Gly Gln Tyr Arg Arg His Arg Arg Asp Ala Arg Ala Val Gln Arg
        650                                 660
CGA GTC TTA CAA CAG TGC TCG CAG TGC AAC CTG CTC TCC ACC AGT CAG GTC GCC TGC TCT ACG    2012
Arg Val Leu Gln Gln Cys Ser Gln Cys Asn Leu Leu Ser Thr Ser Gly Val Ala Cys Ser Thr
    670                                 680
GAC AGC TTT ATG GGA TAC GGA CCG GTA ACG CCA CGT GGT TAT GGC TGC TCC TAC AAT CCG CAT    2075
Asp Ser Phe Met Gly Tyr Gly Pro Val Thr Pro Arg Gly Tyr Gly Cys Ser Tyr Asn Pro His    T45a
690                                 700                                     710
CCG GAA CAG ATC GTC TTC TGC GTC TCG GCC TTC TAC TCA TGT GAG GAT ACG AGT GCC TCG CGA    2138
Pro Glu Gln Ile Val Phe Cys Val Ser Ala Phe Tyr Ser Cys Glu Asp Thr Ser Ala Ser Arg
                                    720
TAC GCC AAA TCG CTG CAG GAC TCG CTG GAC ATA ATG CGT GAT CTA CTG CAA AAC TAG ACG AAC    2201
Tyr Ala Lys Ser Leu Gln Asp Ser Leu Asp Ile Met Arg Asp Leu Leu Gln Asn                T38

TAG ACT AGA ATG TCG CTA GGA TTG GGG TCC ACC AGA AAA AAA AAA ACA TAT CAG TTA ATG TAC    2264
CTA AGC CGG TTA GCG AAC GAA AGT AAG TAA GTG TAA CTA GCG ACC ACA CAC GGG CGT TTC ATT    2327
TGT GAC CAA GCA CCA CCA AGC GAG CCA CCC AAC GAG GGG TGG GGA TCT GAG GAA TAA GAT GGA    2390
TAA CGA TAA CGA AAT CGG AAT CGA AAA CTA ATG TCA ATC ACT ATC AAG TTG AGA CAA AAA AAA    2453
AAC ACA AAA AAA AAG GGA ATT
```

FIG. 3. Nucleotide and deduced amino acid sequences of the cDNA insert of pCha-2. Nucleotide residues are numbered in the 5′ to 3′ direction. The cDNA insert has *Eco*RI linkers at both ends. The deduced amino acid sequence is shown below the nucleotide sequence, and the residues are underlined when it was confirmed by amino-terminal sequencing of tryptic peptides from the enzyme. Numbers preceded by "T" in the right margin identify which tryptic peptide begins its sequence on that line.

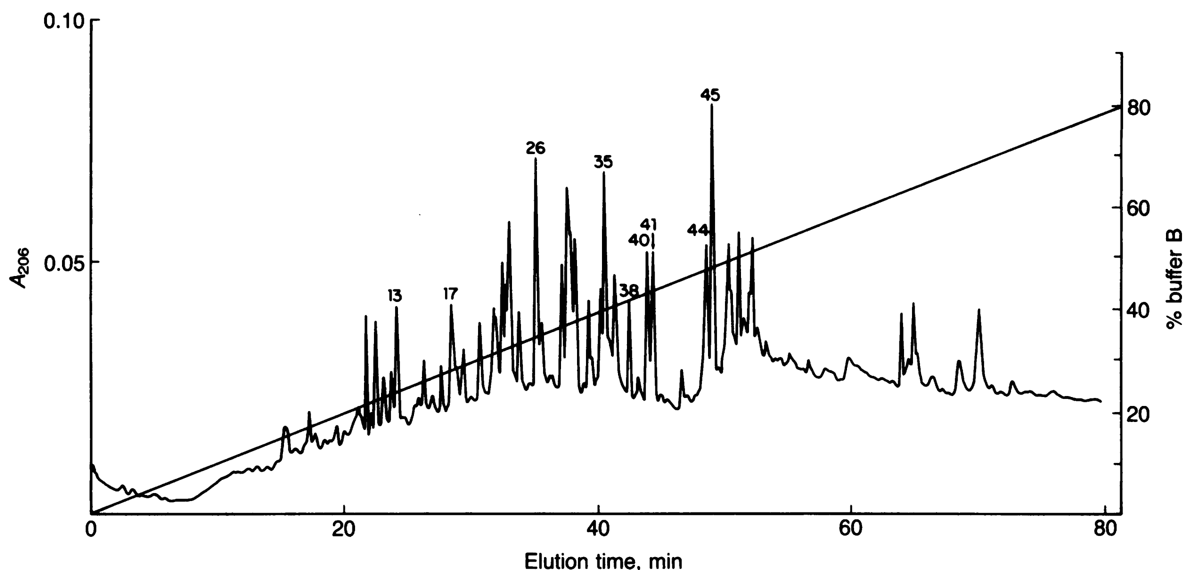

FIG. 4. Tryptic peptides of *Drosophila* choline acetyltransferase resolved by reverse-phase HPLC. Seventy-six micrograms of tryptic digest was injected onto an Altex ODS column (0.45 × 25 cm) (solvent A, 0.09% trifluoroacetic acid; solvent B, trifluoroacetic acid/H_2O/CH_3CN, 0.09/9.91/90). A linear gradient program from 0 to 80% solvent B was run immediately after sample injection. The relevant peak numbers are indicated.

ferase offered several advantages over other animals, since *Drosophila* is relatively rich in this enzyme and it is possible to identify the positions of genes in the absence of structural information. This study originally yielded 14 antibody-positive clones, which would have required randomly sequencing many clones in order to find one that contained corresponding amino acid sequence. Fortunately, the enzyme's gene position had been identified by cytological methods (5). We were able to perform *in situ* hybridization on *Drosophila* polytene chromosomes and demonstrate that only the pCha-2 clone hybridized to the correct position (Fig. 1). Southern analysis confirmed this result by demonstrating that the pCha-2 clone hybridized only to itself and not to any of the remaining 13 clones. These other clones appeared to represent a different and apparently more abundant message, since the longest insert from this group hybridized to all of the 13 remaining clones, but not to pCha-2 (unpublished observation). We were able to eliminate these as authentic choline acetyltransferase clones by additional *in situ* hybridization, using the longest cDNA insert (1500 nucleotides) from this group of related clones. This cDNA hybridized to region 29C-F of chromosome 2L (unpublished observation), whereas the pCha-2 clone hybridized to position 91B-D, where the choline acetyltransferase gene is located (5). Using this approach, we were able to identify the single choline acetyltransferase clone we had recovered from our antibody screening and eliminate all of the remaining clones early in the study.

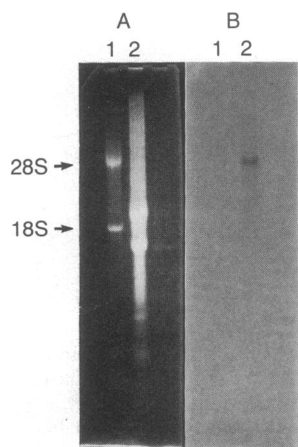

FIG. 5. RNA transfer analysis of *Drosophila* head poly(A)$^+$ RNA by using ^{32}P-labeled pCha-2 as a probe. (*A*) Human rRNA (lane 1; 3.5 μg) and *Drosophila* head poly(A)$^+$ RNA (lane 2; 25 μg) were resolved on a 1% agarose gel containing 5 mM methylmercuric hydroxide. The gel was stained with ethidium bromide. (*B*) The gel was dried and hybridized with ^{32}P-labeled pCha-2. The positions of 28S rRNA and 18S rRNA are indicated.

Although not a focus of the present study, the mRNA represented by the 13 clones recovered independently of the choline acetyltransferase clone is by itself interesting. It appears to code for an immunologically related protein that is more abundant than choline acetyltransferase, due to the large number of clones recovered. The importance of this observation has two facets. (*i*) It demonstrates the power of cDNA cloning for studying the expression of proteins in the nervous system by revealing related proteins that would otherwise go undetected or unappreciated. (*ii*) It demonstrates the problem of protein purification based on immunological affinities. Had *Drosophila* choline acetyltransferase originally been purified by using immobilized-antibody columns, it is likely that the abundant protein species would not have been choline acetyltransferase, but rather the translation product of this other gene. The enzyme would have existed only as a minor contaminant. This observation emphasizes the importance of having structural information when identifying clones for a specific polypeptide. The sequencing of these other related clones and the production of a fusion polypeptide directed from the longest of the 13 clones await further investigation.

The choline acetyltransferase cDNA isolated in this study consists of a coding region that is 2190 nucleotides long, followed by a 3′-noncoding region 284 nucleotides in length. The coding region spans 728 amino acids, which is ≈50–100 amino acids more than is required for an average M_r 67,000 protein. The 5′ region of the cDNA sequence contains no methionine residues for the choline acetyltransferase protein sequence. This may indicate that *Drosophila* choline acetyltransferase has a larger precursor that is enzymatically inactive, since the only activity ever observed had a M_r of

67,000 (7, 8). RNA transfer analysis on *Drosophila* head poly(A)$^+$ RNA, using the enzyme cDNA as a probe, estimated the enzyme's mRNA to be ≈4700 nucleotides long (Fig. 5), leaving ≈2300 bases in the mRNA to be elucidated. Since our clone does not contain a poly(A) tail, we cannot be sure how much of the remaining sequence is in the 3′ region. Considering the length of the insert and that cDNA production was directed against poly(A) tails, there could be as many as 2000 bases more in the 5′ direction. Anything near this length would be an unusually long 5′-noncoding sequence and would further suggest that *Drosophila* choline acetyltransferase is derived from a much larger precursor. The content of the 5′ region upstream of our present sequence remains to be elucidated, but, with the information gained through this study, the remaining sequence of the enzyme's message can be determined directly. An analysis of this sequence will no doubt greatly facilitate our understanding of the structure and processing of this important enzyme and of the control of cholinergic phenotypic expression.

We thank Dr. E. B. Lewis for confirming the cytogenetic position of pCha-2. We also thank Noel Neighbor and Ting Huang for their technical assistance and Sharon Shafer for preparing the manuscript. This work was supported by grants (NS19482 to P.M.S., GM31259 to K.I., and NS18858 and NS18859 to E.R.) from the National Institutes of Health.

1. Tuček, S. (1983) in *Handbook of Neurochemistry*, ed. Lajitha, A. (Plenum, New York), Vol. 4, pp. 219–249.
2. Klemm, N. (1976) *Prog. Neurobiol.* **7**, 99–169.
3. Dewhurst, S. A., McCaman, R. E. & Kaplan, W. D. (1970) *Biochem. Genet.* **4**, 499–508.
4. Dewhurst, S. A. & Seecof, R. L. (1973) *Comp. Biochem. Physiol. C.* **50**, 53–58.
5. Greenspan, R. J. (1980) *J. Comp. Physiol.* **137**, 83–92.
6. Patterson, P. H. (1978) *Annu. Rev. Neurosci.* **1**, 1–17.
7. Landis, S. C. & Keefe, D. (1983) *Dev. Biol.* **98**, 349–371.
8. Slemmon, J. R., Salvaterra, P. M., Crawford, G. D. & Roberts, E. (1982) *J. Biol. Chem.* **257**, 3847–3852.
9. Slemmon, J. R., Salvaterra, P. M. & Roberts, E. (1984) *Neurochem. Int.* **6**, 519–525.
10. Salvaterra, P. M. & McCaman, R. E. (1985) *J. Neurosci.* **5**, 903–910.
11. Crawford, G. D., Slemmon, J. R. & Salvaterra, P. M. (1982) *J. Biol. Chem.* **257**, 3853–3856.
12. Auttray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
13. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
14. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483–2495.
15. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. (Information Retrieval Limited, Oxford), pp. 49–78.
16. de Wet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. & Helinski, D. R. (1984) *DNA* **3**, 437–447.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 371–372.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
20. Yuan, P.-M., Pande, H., Clark, B. R. & Shively, J. E. (1982) *Anal. Biochem.* **120**, 289–301.
21. Hawke, D. H., Yuan, P.-M. & Shively, J. E. (1982) *Anal. Biochem.* **120**, 302–311.
22. Hawke, D. H., Harris, D. C. & Shively, J. E. (1985) *Anal. Biochem.* **147**, 315–330.
23. Pardue, M. L. & Gall, J. G. (1975) *Methods Cell Biol.* **10**, 1–16.
24. Bailey, J. M. & Davidson, N. (1976) *Anal. Biochem.* **70**, 75–85.