cDNA cloning and complete sequence of porcine choline acetyltransferase: *In vitro* translation of the corresponding RNA yields an active protein

(Xenopus oocytes/rabbit reticulocyte lysate/complete amino acid sequence)

Sylvie Berrard*, Alexis Brice*, Friedrich Lottspeich[†], Axel Braun[‡], Yves-Alain Barde[‡], and Jacques Mallet^{*§}

*Laboratoire de Neurobiologie Cellulaire et Moléculaire, Centre National de la Recherche Scientifique, F-91190 Gif-sur-Yvette, France; and [†]Genzentrum and [‡]Max-Planck-Institut für Psychiatrie, D-Martinsried b. München, Federal Republic of Germany

Communicated by Jean-Pierre Changeux, August 10, 1987

ABSTRACT A cDNA clone encoding the complete sequence of porcine choline acetyltransferase (ChoAcTase; acetyl-CoA: choline O-acetyltransferase, EC 2.3.1.6.) has been identified. A cDNA library, constructed from poly(A)⁺ RNA of ventral spinal cord, was screened with a mixture of eight oligonucleotides corresponding to the N-terminal sequence of pig brain ChoAcTase. Among five positive clones, one, pChAT-1, was identified as a ChoAcTase cDNA clone based on the following criteria. (i) This clone has an open reading frame coding for a protein of the size expected for ChoAcTase (640 amino acids). (ii) The amino acid composition deduced from the nucleotide sequence of this open reading frame matches that of purified porcine ChoAcTase. (iii) When subcloned in the T7 expression system, the corresponding RNA directs the synthesis in the rabbit reticulocyte lysate of a protein that is specifically immunoprecipitated by antibodies raised against ChoAcTase. (iv) Finally and most important, this corresponding RNA, when translated in the reticulocyte lysate, as well as in the Xenopus oocyte system, directs the synthesis of a protein displaying ChoAcTase activity. This activity is inhibited by the specific ChoAcTase inhibitor 4-(1-naphthylvinyl)pyridine. Comparison of porcine ChoAcTase sequence with that of Drosophila reveals 32% identity between these proteins, when the sequences are suitably aligned. pChAT-1 probe hybridizes with a porcine mRNA species that is at least 7000 nucleotides long, whereas the equivalent rat mRNA species is 3700 nucleotides long.

The enzyme choline acetyltransferase (ChoAcTase; acetyl-CoA: choline-O-acetyltransferase, EC 2.3.1.6.) catalyzes the biosynthesis of the neurotransmitter acetylcholine and constitutes a specific marker of cholinergic systems (1). To date, there is only very limited information about the structure of the mammalian enzyme. More detailed understanding of this enzyme is particularly desirable because of the importance of the cholinergic system in neurotransmission, as well as the possible involvement of this system in certain neurological disorders, particularly Alzheimer disease (2).

In addition, ChoAcTase as well as tyrosine hydroxylase (EC 1.14.16.2), the rate limiting enzyme in catecholamine synthesis, have received much attention in examining the phenotypic expression of neurotransmitters. Studies on the ontogeny of the autonomic nervous system have revealed that neurons can change their phenotype from, for example, adrenergic to cholinergic, depending on the nature of their environment (3, 4). The analysis, in molecular terms, of the mechanisms underlying this plasticity requires the study of the genes encoding these two enzymes.

We have identified (5-7) cDNA clones corresponding to rat and human tyrosine hydroxylases. Here, we describe the isolation of a cDNA clone—pChAT-1—that encodes an active porcine ChoAcTase enzyme. The nucleotide and complete amino acid sequence is reported.[¶] Some structural characteristics of porcine ChoAcTase are discussed, and the sequence is compared with that of *Drosophila melanogaster* reported by Itoh *et al.* (8).

MATERIALS AND METHODS

Construction of a Randomly Primed cDNA Library in the λ gt10 Vector. Total RNA from porcine ventral spinal cord was extracted as described by Lomedico and Saunders (9). Poly(A)⁺ RNA was purified by oligo(dT)-cellulose chromatography. Random DNA sequences 20–50 nucleotides in length were prepared by sonication and DNase I digestion of calf thymus DNA (10) and used as primers for cDNA synthesis. First-strand cDNA was synthesized from 2.5 μ g of ventral spinal cord poly(A)⁺ RNA with 30-fold excess of random primer. The second-strand synthesis and following steps were carried out using standard procedures (11, 12). The longest cDNAs [\geq 500 base pairs (bp)] were selected on a 5–20% (wt/vol) sucrose gradient and ligated to the λ gt10 vector. The amplified library contained \approx 1.2 × 10⁶ independent recombinant phages.

Oligonucleotide Screening. The N-terminal sequence of porcine brain ChoAcTase was determined as described (13). A mixture of oligodeoxynucleotides, each containing eight different chains of 29 nucleotides, was prepared with a Biosearch DNA synthesizer model 8600 by the phosphoramidite method and purified by PAGE. The probes were end-labeled to a minimal specific activity of 8×10^8 cpm/µg. About 10⁶ recombinant phages were plated at 50,000-70,000 plaques per 13-cm plate, and duplicate filters were prepared. Filters were hybridized at 35°C with oligonucleotides in $6\times$ SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7.0), $5 \times$ Denhardt's solution (1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin), 10% (wt/vol) dextran sulfate, 0.05% sodium pyrophosphate, herring sperm DNA at 0.1 mg/ml, and Escherichia coli tRNA at 0.1 mg/ml. Filters were then washed at 35°C, 40°C, and 45°C in 6× SSC containing 0.05% NaDodSO₄. Positive clones were isolated after three successive rounds of screening. Phage DNA was prepared as

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.

Abbreviations: ChoAcTase, choline acetyltransferase; NVP, 4-(1-naphthylvinyl)pyridine.

[§]To whom reprint requests should be addressed.

[¶]The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03021).

| ChAT N-terminal amino acids | : | H ₂ N | PRO | ILE | LEU | GLU | LYS | THR | PRO | PRO | LYS | MET | ALA | ALA | LYS |
|--------------------------------|---------|------------------|--------------------|---------------|---------------------------|------------|-----------------|--------------------|--------------------|--------------------|-----------------|------------|--------------------|--------------------|------------|
| Codons | : | | CCU C A G | AUU C A | UUA G CUU C G | GAA G | MA G | ACU C A G | CCU C A G | CCU C A G | AAA G | AUG | GCU C A G | GCU C A G | мл G |
| Synthetic oligonucleotides | : | | | | | GAA G | 444 G | ACI | CCI | CCI | 444 G | ATG | GCI | GCI | * |
| cDNA sequences | : A | (83%) | | | | GAG | AAG | AAT | ссс | ссс | AAG | ATT | ecc | CAA | CAC |
| | B | (79%) | | | | GAA | AAG | ACC | ссс | ссс | CAA | ATA | AAC | ACA | TAC |
| | C | (69%) | | | | GAG | *** | ACC | TCC | CCC | *** | TGT | CAG | CTG | CAA |
| | р((| ChAT-1 (100%) | | | | GAA GLU | AAA LYS | ACT Thr | CCC PRO | CCT PRO | AAG LYS | ATG Mét | GCA ALA | GCA ALA | AAA LYS |

FIG. 1. Sequences of the N terminus of porcine ChoAcTase (ChAT) the back-translated mRNA (codons), the mixture of 29-mer oligonucleotides, and the four identified cDNAs and pChAT-1 corresponding amino acid sequence. Deoxyinosine (I) was inserted into each third position where codon ambiguity allowed for all four nucleotides. Note the complete identity between the ChAT protein sequence obtained from the purified protein and that obtained from pChAT-1 cDNA analysis. Numbers in parentheses indicate the percentage of identity between synthetic oligonucleotides and the corresponding sequence of the isolated cDNA clones, assuming that deoxyinosine yields no mismatch.

described by Maniatis *et al.* (14), and the cDNA inserts were excised by digestion with *Eco*RI.

DNA Sequencing. Insertions were sonicated and subcloned in M13mp8 (15). Both strands were sequenced by the dideoxy method of Sanger *et al.* (16) using either the universal primer or appropriate oligonucleotides.

In Vitro Transcription. The ChoAcTase cDNA insertion was subcloned in the transcription plasmid pSPT18, which contains SP6 and T7 promoters in opposite orientations. Recombinant plasmids were linearized, and RNA was produced in the presence of 500 μ M of each NTP and [α -³²P]GTP, as tracer; capping was achieved using 2.5 mM of m⁷GpppG [7-methylguanosine(5')triphospho(5')-guanosine] (17, 18). The amount of RNA synthesized was estimated by counting the incorporated radioactivity. In these conditions, T7 RNA polymerase yielded over 6 μ g of RNA per μ g of DNA.

In Vitro Translation, Immunoprecipitation, and Oocyte Injection. In vitro translation of RNA transcripts was performed in rabbit reticulocyte lysate (19). Immunoprecipitations were carried out (20) using either monoclonal or polyclonal antibodies raised against porcine brain Cho-AcTase (21, 22). Proteins were separated by isochratic (8%) NaDodSO₄/polyacrylamide gel electrophoresis (23). Oocytes were treated and injected as described (24).

ChoAcTase Activity. ChoAcTase activity, generated from translation of RNA transcripts in oocytes and in rabbit reticulocyte lysate, was measured as described (25), in the

presence or in the absence of the specific ChoAcTase inhibitor 4-(1-naphthylvinyl)pyridine (NVP) (26).

Materials. Radioactive compounds, rabbit reticulocyte lysate, and the phage vector M13mp8 were purchased from Amersham. Restriction enzymes and oligo(dT)-cellulose were from Boehringer Mannheim. The plasmid pSPT18 was from Promega Biotec (Madison, WI). Deoxyinosine, T7 RNA polymerase, and m⁷GpppG [7-methylguanosine(5')triphospho(5')-guanosine] were purchased from Pharmacia. NVP was from Calbiochem. Nitrocellulose filters were obtained from Schleicher & Schuell.

RESULTS

Cloning of ChoAcTase cDNA. The sequence of the first 11 amino acids from the N terminus of pig brain ChoAcTase was obtained (13). In the present study, 2 additional amino acids were used (even though their assignment was less certain), because of the high degeneration of the codons corresponding to the leucine residue at the third position (Fig. 1). A mixture of eight 29-mer oligonucleotides, reflecting all codon combinations, was synthesized using deoxyinosine where codon ambiguity involved all four nucleotides (27). These probes were used to screen 10⁶ recombinant phages of a randomly primed λ gt10 library generated from porcine ventral spinal cord $poly(A)^+$ RNA. Five positive plaques were purified after three rounds of screening. The sequences homologous to the 29-mer oligonucleotides were determined for four corresponding inserts, as shown in Fig. 1. One of them, over 2100 bp long and designated pChAT-1, contains a nucleotide pattern 128 bp from its 5' end that corresponds exactly to the N-terminal amino acid sequence of ChoAcTase purified from porcine brain. Immediately upstream from this pattern, pChAT-1 contains a putative initiation codon ATG flanked by the nucleotides cytidine and adenosine at positions -4 and -3, respectively, that fit with the consensus sequence of Kozak (28), as well as a nonsense in-frame codon 93 bp upstream from this ATG codon. These observations indicated that pChAT-1 was likely to contain the complete coding sequence of porcine ChoAcTase, since the estimated size of porcine ChoAcTase is 68 kDa (22).

Detection of ChoAcTase Activity in Oocytes and in Rabbit Reticulocyte Lysate. For these studies, pChAT-1 cDNA was subcloned in the plasmid pSPT18, and the corresponding sense RNA was synthesized. It encodes a protein of an apparent molecular weight of 68,000 that is specifically immunoprecipitated by monoclonal and polyclonal anti-ChoAcTase antibodies (results not shown).

To establish that pChAT-1 cDNA encodes an active ChoAcTase, sense RNA was first injected into frog oocytes, which have been shown to be a convenient system in which to express active rat ChoAcTase (24). Injection of pChAT-1 RNA yielded a high level of ChoAcTase activity that was inhibited by NVP, a specific ChoAcTase inhibitor (Table 1).

Table 1. Expression of ChoAcTase activity generated from pChAT-1 RNA in oocyte and in rabbit reticulocyte lysate systems

| | | AcCho, c | pm | · · · · · · · · · · · · | ······, | · · · · · · · · · · · · · · · · · · · | |
|---------|--------|----------|--------|-------------------------|----------|---------------------------------------|--------|
| | Co | ontrol | pCh | AT-1 | | % inhibition | |
| | No RNA | TPH-RNA | -NVP | +NVP | pmol/min | pmol/(min•µg of RNA) | by NVP |
| Oocytes | 150 | 160 | 13,070 | 520 | 125 | 330 | 97 |
| Lysate | 120 | 140 | 14,330 | 310 | 14 | 70 | 99 |

ChoAcTase activity was measured in 15 oocytes each injected with 25 ng of pChAT-1 RNA. Eggs were homogenized in 50 μ l of 50 mM sodium phosphate, pH 7.4/0.5% Triton X-100. [¹⁴C]Acetylcholine (AcCho) cpm represents the amount of AcCho synthesized in 5 μ l of homogenate after a 10-min reaction. ChoAcTase was also assayed after translation of 200 ng of pChAT-1 RNA in rabbit reticulocyte lysate. In this case AcCho cpm represents the total amount of [¹⁴C]Accho synthesized. In control experiments, pChAT-1 RNA was replaced by a RNA encoding rat tryptophan hydroxylase (TPH-RNA) (M. C. Darmon, personal communication). Numbers represent mean values of duplicate experiments.

CCGÃA 125 140 185 200 ATG CCC ATC CTG GAA AAA ACT CCC CCT AAG ATG GCA GCA AAA AGT CCC AGC AGT GAG GAG GAG CCT GGG CTG CCC AAA CTC CCT GTG CCC MET <u>PRO ILE LEU GLU LYS THR PRO PRO LYS MET ALA ALA LYS</u> SER PRO SER SER GLU GLU GLU PRO GLY LEU PRO LYS LEU PRO VAL PRO 215 230 245 260 275 290 CCA TTG CAG CAG ACC CTG GCC ACC TAC CTG CGG TGC ATG CAG CAC CTG GTA CCT GAG GAG CAA TTT AGG AGG AGC CAG GCC ATT GTG CAG PRO LEU GLN GLN THR LEU ALA THR TYR LEU ARG CYS MET GLN HIS LEU VAL PRO GLU GLU GLN PHE ARG ARG SER GLN ALA ILE VAL GLN 305 320 - 335 350 365 380 CAG TTT GGG GCC CCT GGT GGC CTT GGC GAG GAC CTG CAG CAG GAA CGC CAG GAG CAG GAC GCT AAC TGG GTG TCT GAG TAC GLN PME GLY ALA PRO GLY GLY LEU GLY GLU THR LEU GLN GLN LYS LEU LEU GLU ARG GLM GLU GLN THR ALA ASM TRP VAL SER GLU TYR 95 410 425 440 455 470 TGG CTG AAC GAC ATG TAT CTC AAT AAC CGT CTG GCC CTG CCT GTC AAC TCC AGC CCA GCT GTG ATT TTT GCC CGG CAG CAC TTC CAA GAC TRP LEU ASN ASP MET TYR LEU ASN ASN ARG LEU ALA LEU PRO VAL ASM SER SER PRO ALA VAL ILE PHE ALA ARG GLN HIS PHE GLM ASP 85 545 560 515 530 545 560 ACC MAT GAC CAG CTA AGG TTT GCA GCC MAC CTT ATC TCT GGT GTG GTC AGC TAC AGG GCC CTG GAC AGC CAC TCC ATC CCC ATT GAC THR ASH ASP GLN LEU ARG PHE ALA ALA ASH LEU ILE SER GLY VAL LEU SER TYR LYS ALA LEU LEU ASP SER HIS SER ILE PRO ILE ASP 485 575 590 605 620 635 650 TGT GCC AAG GGC CAG CTG TCA GGA CAG CCT CTC TGT ATG AAG CAA TAC TAT GGG CTT TTC TCC TCT TAC CGG CTC CCT GGC CAC ACC CAG CYS ALA LYS GLY GLN LEU SER GLY GLN PRO LEU CYS MET LYS GLN TYR TYR GLY LEU PHE SER SER TYR ARG LEU PRO GLY HIS THR GLN 665 55 770 785 800 815 830 GTC ATT AAT TTC CGC CGT CTC AGT GAG GGG GAT CTG TTC ACT CAG TTG AGA AAG ATA GTC AGA ATG GCT TCC AAC GAG GAT GAA CGC TTG VAL ILE ASN PHE ARG ARG LEU SER GLU GLY ASP LEU PHE THR GLN LEU ARG LYS ILE VAL ARG MET ALA SER ASN GLU ASP GLU ARG LEU 755 AS 860 920 CCT CCA ATC GGC CTG CTG ACG TCA GAC GGG AGG AGC GAG TGG GCT GAG GCC AGG ACG GTC CTC GTG AAA GAC TCC ACC AAT CGG GAC TCT PRO PRO 1LE GLY LEU LEU THR SER ASP GLY ARG SER GLU TRP ALA GLU ALA ARG THR VAL LEU VAL LYS ASP SER THR ASN ARG ASP SER 35 950 965 900 995 1010 CTG GAT ATG ATC GAG CGG TGC ATC TGC CTG GTG TGC CTG GAT GCC CCT GGA GGC ATG GAG CTC AGC GAC ACC AAC AGG GCG CTC CAG CTT LEU ASP MET ILE GLU ARG CYS ILE CYS LEU VAL CYS LEU ASP ALA PRO GLY GLY MET GLU LEU SER ASP THR ASN ARG ALA LEU GLN LEU 935 25 1040 1055 1070 1085 1100 CTT CAC GGC GGA GGC TGC AGC AAG AAT GGA GCC AAC CGC TGG TAC GAG AAG TCC CTA CAG TTT GTG GTG GGC CGA GAT GGC ACC TGC GGC LEU HIS GLY GLY GLY CYS SER LYS ASN GLY ALA ASN ARG TRP TYR ASP LYS SER LEU GLN PHE VAL VAL GLY ARG ASP GLY THR CYS GLY 1025 15 1130 1145 1160 1175 1190 GTG GTG TGC GAA CAC TCC CCT TTT GAT GGC ATT GTC CTG GTG CAG TGC ACG GAG CAT CTG CTC AAA CAC ATG GTG AAG AGC AGC AAG AAG VAL VAL CYS GLU HIS SER PRO PHE ASP GLY ILE VAL LEU VAL GLN CYS THR GLU HIS LEU LEU LYS HIS MET VAL LYS SER SER LYS LYS 1205 1220 1235 1250 1265 1280 ATG GTC CGA GCT GAC TCG GTC AGC GAG CTC CCA GCA CCC AGA AGG CTG AGG TGG AAG TGT TCC CCG GAA ATC CAA GGC CTC TTA GCT TCC MET VAL ARG ALA ASP SER VAL SER GLU LEU PRO ALA PRO ARG ARG LEU ARG TRP LYS CYS SER PRO GLU ILE GLN GLY LEU LEU ALA SER 95 1310 1325 1340 1355 1370 TGG GCA GAA AAA CTC CAA CAA ATA GTC AAG AAT CTT GAC TTC ACT GTT TAT AAA TTT GAC GAC TAT GGG AAG ACT TTC ATT AAG CAG CAG SER ALA GLU LYS LEU GLN GLN ILE VAL LYS ASN LEU ASP PHE THR VAL TYR LYS PHE ASP ASP TYR GLY LYS THR PHE ILE LYS GLN GLN 1295 1385 1400 1415 1430 1445 1460 AMA TGC AGC CCC GAT GCC TTT ATT CAG GTG GCC CTC CAG CTA GCC TTC TAC AGG CTC CAT GGG AGA CTC GTG CCT ACC TAT GAG AGC GCG LYS CYS SER PRO ASP ALA PHE ILE GLN VAL ALA LEU GLN LEU ALA PHE TYR ARG LEU HIS GLY ARG LEU VAL PRO THR TYR GLU SER ALA 75 1490 1505 1520 1535 1550 TCC ATC CGC CGA TTC CAC GAG GGA CGG GTG GAC AAC ATT CGA TCG GCC ACT CCG GAG GCA CTG CAT TTT GTG AAA GCC ATT ACT GAC CAT SER ILE ARG ARG PHE HIS GLU GLY ARG VAL ASP ASN ILE ARG SER ALA THR PRO GLU ALA LEU HIS PHE VAL LYS ALA ILE THR ASP HIS 1565 1580 1595 1610 1625 1640 GCA TCT GCC ATG CCG GAT TCG GAG AAG CTG CTG CTG CTG CTG AAG GAT GCC ATC CGA GCC CAG ACC CAG TAC ACA GTC ATG GCC ATC ACG GGG ALA SER ALA MET PRO ASP SER GLU LYS LEU LEU LEU LEU LYS ASP ALA ILE ARG ALA GLN THR GLN TYR THR VAL MET ALA ILE THR GLY 55 1670 1685 1700 1715 1730 ATG GCC ATC GAC AAC CAC CTG CTG GGG CTG CGG GAA CTG GCC CGA GAA GTG TGC AAG GAA CTG CCT GAG ATG TTC ACG GAT GAA ACA TAC MET ALA ILE ASP ASN HIS LEU LEU GLY LEU ALG AGG GLU LEU ALA ARG GLU VAL CYS LYS GLU LEU PRO GLU MET PHE THR ASP GLU THR TYR 1655 15 1760 1275 1790 1805 1820 CTG ATG AGC AAC CGC TTT GTC CTC TCC ACC CAG GTG CCC ACC ACC ATG GAG ATG TTT TGC TGC TAT GGT CCT GTG GTA CCC AAT LEU MET SER ASN ARG PHE VAL LEU SER THR SER GLN VAL PRO THR THR MET GLU MET PHE CYS CYS TYR GLY PRO VAL VAL PRO ASN GLY IS 1850 1865 1880 1895 1910 TAC GGT GCC TGC TAC AAC CCC CAG CCA GAG AGC ATC CTT TTC TGC ATC TCC AGC TTT CAC GGC TGC AAA GAA ACC TCT TCA ACC AAG TTT TYR GLY ALA CYS TYR ASN PRO GLN PRO GLU SER ILE LEU PHE CYS ILE SER SER PHE HIS GLY CYS LYS GLU THR SER SER THR LYS PHE 25 1940 1955 1970 1985 2000 GCA AMA GCT GTG GAA GAA AGC TTT ATT GAA ATG AAA GGT CTC TGC AGT CTG TCC CAG TCT GGC ATG GGC AAG CCC CTG GCA ACA AMG GAA ALA LYS ALA VAL GLU GLU SER PHE ILE GLU MET LYS GLY LEU CYS SER LEU SER GLN SER GLY MET GLY LYS PRO LEU ALA THR LYS GLU 2015 2030 2045 2060 2075 2090 2105 2120 MAA GTA ACA AGG CCT AGC CAG GTA CAC CAA CCT TGACTGCTGCCGCCTCAGTTTCGCCTCCCCAAACCCAGGCCTCTGCAGGCCGCCGGAGACCCCTGCTGAACCCCTGCTG LYS VAL THR ARG PRO SER GLN VAL HIS GLN PRO 640

FIG. 2. Nucleotide and predicted amino acid sequences of porcine ChoAcTase as deduced from pChAT-1. Nucleotides are numbered in the $5'\rightarrow 3'$ direction, starting with the first residue following the *Eco*RI cloning site. The N-terminal proline of the mature protein and the proline at the C terminus are numbered 1 and 640, respectively. Peptide sequence derived from purified porcine ChoAcTase is underlined.

The ChoAcTase levels were high enough to detect the activity in a single oocyte (data not shown). Because of the high sensitivity of the ChoAcTase assay, the activity was also measured after translation of the sense RNA in rabbit reticulocyte lysate. Again, a clear positive response was detected (Table 1). These results clearly demonstrate that pChAT-1 can direct the translation of an active, NVP-sensitive ChoAcTase enzyme. No activity was detected in either system following translation of a rat RNA coding for tryptophan hydroxylase.

pChAT-1 Nucleotide and Amino Acid Sequences. The

pChAT-1 complete nucleotide sequence of 2120 bp is displayed in Fig. 2. The ATG codon at position 125 specifies an open reading frame of 1923 nucleotides followed by 73 bp of 3'-untranslated sequence. The 3'-untranslated region is incomplete, since it contains neither a poly(A) sequence nor a polyadenylylation signal. The open reading frame encodes a protein of 640 amino acids with a calculated molecular weight of 71,517 and an isoelectric value of 7.72. Six putative serine or threonine phosphorylation sites (29, 30) are located at positions 217, 233, 327, 365, 466, and 500.

Comparison of Porcine and Drosophila ChoAcTase Se-

| 1 | | |
|---|--|------------|
| 2 | IPDPKGANVASNEAST SAAGSGPE SAALFSKLRSFSIGSGPNSPQRVVSNLRGFLTHRLSNITPSDTGWKDSI | 73 |
| 1 | PILEKTPPKMAAKSPSSEEEPGLPKLPVPPLQQTLATYLRCMQHLVPEEQFRRSQAIVQQFGAPGGLGETLQQKLLERQEQTANWVSEYWLNDMYLNNRLALPVNSSPAVIFAROHFQDT | 120 |
| 2 | LS IPKKWLSTAESVDEFGFPDTLPKVPVPALDETMADY IRALEPITTPAQLERTKELIRQFSAPQGIGARLHQYLLDKREARITGPITTGSTRCTWIFAFPLPINSNPGIGVPAASLQDR | 193 |
| 1 | NDQLRFAANLISGVLSYKALLDSHSIPIDCAKGQLSGQPLCMKQYYGLFSSYRLPGHTQDTLVAQKSSVMP-EPEHVIVACCNQFFVLDVVINFRRLSEGDLFTQLRKIVRMASNED | 236 |
| 2 | PRRAHFAARLLDGILSHREMLDSGELPLFRAL-AEKNQPLCMAQYYRLLGSCRRPGVKQDSQFLPSRERLNDEDRHVVVICANQM—YKVVLQASDRGKLSESEIASQILYVLSDAPCLP | 310 |
| 1 | ERLPPIGLLTSDGRSEWAEARTVLVKDSTNRDSLDMIERCICLVCLDAPGGMELSDTNRALQLLHGGGCSKNGANRWYDKSLQFVVGRDGTCGVVCEHSPFD | 338 |
| 2 | AKPVPVGLLTAEPRSTWARDREMLQEDERNQRNLELIETAQVVLCLDEPLAGNFNARGFTGATPTVHRAGDRDETNMAHEMIHGGGSEYNSGNRWFDKTMQLIICTDGTWGLCYEHSCSE | 430 |
| 1 | GIVLVQCTEHLLKHMVKSSKKMVRADSVSELPAPRRLRWKCSPEIOGLLASSAEKLQQIVKNLDFTVYAFDDYGKTFIKQQKCSPDAFIQVALQLAFYRLHGRLVPTYES | 448 |
| 2 | GIAVVQLLEKITKKIEEHPDEDNGLPQHHLPPPEKLEMHVGPQLQLKFAQASKSVDKCIDDLDFTVYKYQSYGKIFIKSCQVSPDVYIQLAIAIGSLQVVKISGGHLKKCVHSIIS | 540 |
| 1 | ASIRRFHEGRVDNIRSATPEALHFVKAITDHASAMPDSEKLLLLKDAIRAQTQYTVMAITGMAIDNHLLGLRELAREVCKELPEMFTDETYLMSNRFVL | 547 659 |
| • | | 0.00 |
| 1 | STSQVPTTMEMFCCYGPVVPNGYGACYNPQPESILFCISSFHGCKETSSTKFAKAVEESFIEMKGLCSLSQSGMGKPLATKEKVTRPSQVHQP | 640 728 |
| - | איזאייראיזיגעינעערערא איזאייראיזיגעערערא איזאיירא איזאיירא איזאיירא איזאיירא איזאיירא איזאיירא איזאיירא איזאיי | , 20 |

FIG. 3. Comparison of amino acid sequences for porcine ChoAcTase and *Drosophila* ChoAcTase 1 and 2, respectively. Amino acid position is at the right. Vertical bar, identical residues; discontinuous bar, homologous amino acid replacement (31).

quences. Comparison of amino acid sequences deduced from porcine and *Drosophila* (8) cDNAs reveals 32% identity; this value reaches 51% when homologous amino acid replacement (31) is considered. Note that, to achieve maximum homology, 11 gaps were inserted into the amino acid sequences (Fig. 3). Six regions, located on porcine ChoAcTase at positions 22–80, 125–181, 241–285, 292–344, 369–430, and 545–605, are highly conserved and display homologies ranging from 64% to 79%. No other significant homology was found with known proteins using the GenBank^{II} and National Biomedical Research Foundation** protein sequence data bases.

RNA Analysis. RNA gel blot experiments (32) were performed with ventral spinal cord $poly(A)^+$ RNAs. Filters were hybridized with pChAT-1 cDNA labeled by nick-translation. Surprisingly, the probe reveals RNA species of quite different sizes: in pig, ChoAcTase mRNA is at least 7000 nucleotides long as compared to 3700 in rat (Fig. 4). In preliminary experiments, cross-reactivity was also detected with human spinal cord poly(A)⁺ RNAs (result not shown).

DISCUSSION

The present study has led to the isolation of a cDNA clone that encodes the entire amino acid sequence of an active mammalian ChoAcTase enzyme. The isolation of this cDNA coding for a protein of very low abundance (21) was based on the following strategy. (*i*) The sequence of 13 amino acids at the N terminus was determined from the pig brain enzyme. (*ii*) Synthetic oligonucleotide probes were synthesized, in which deoxyinosine was inserted in every position where all four nucleotides might be found. This was an important point since the codons corresponding to the N terminus of porcine ChoAcTase turned out to be mostly of low frequency (33). (*iii*) The λ gt10 cDNA library was derived from the ventral spinal cord, a region shown (24) to be a suitable source of ChoAcTase mRNA in the central nervous system. Oocvtes injected with mRNA from ventral spinal cord generated about 10 times more ChoAcTase activity than those injected with striatal mRNA, although the latter structure was found to contain the highest ChoAcTase activity in the central nervous system (24). (*iv*) The synthesis of the first-strand cDNA was performed with random primers to ensure that the 5' portions of the mRNAs would be included in the library. This particular point was crucial in considering that the porcine ChoAcTase mRNA is >7000 nucleotides long.

The high sensitivity of the ChoAcTase enzymatic assay facilitated the functional identification of pChAT-1. The corresponding RNA directed the synthesis of an active enzyme both in frog oocytes and in rabbit reticulocyte lysate. Interestingly, the detection of a functional ChoAcTase in the latter translation system suggests that specific post-translational modifications are not required to generate enzyme activity.

The molecular weight of 71,517 and the isoelectric value of

123

6.56

4.36

FIG. 4. RNA gel blot hybridization analysis using pChAT-1 cDNA probe. Ventral spinal cord poly(A)⁺ RNA was fractionated by electrophoresis on 1% agarose gels in the presence of 1 M formaldehyde, transferred to nitrocellulose filter, and hybridized with nicktranslated pChAT-1 cDNA probe (specific activity, 2×10^8 cpm/µg). Lanes: 1, rat RNA (5 μ g); 2 and 3, porcine RNA (5 and 10 μ g, respectively). The autoradiogram was obtained after exposure for 48 hr at -70° C with an intensifying screen. λ wild-type phage DNA digested with HindIII was used as a size marker (shown in kilobases).

EMBL/GenBank Genetic Sequence Database (1987) GenBank (Bolt, Beranek, and Newman Laboratories, Cambridge, MA), Tape Release 50.

^{**}Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 12.

7.72, deduced from the ChoAcTase cDNA sequence, are in good agreement with that of 68,000 for the apparent molecular weight of the purified porcine enzyme (22) and that of 8.1 for the isoelectric value determined for the human enzyme (34). Also, the amino acid composition reported for this enzyme (13) corroborates our data. Indeed, most of the amino acids that can be reliably determined using the ophthaldialdehyde method are within $\pm 10\%$ of the sequence data, when the total of 640 amino acids reported here is used as a base of calculation.

Eckenstein et al. (21, 22) purified porcine brain ChoAcTase using a procedure that selected for soluble proteins. However, several biochemical experiments suggest that the enzyme also exists in a membrane-bound form, with a distinct isoelectric point and molecular weight (35, 36), although this apparent heterogeneity of ChoAcTase may be due to interactions with proteins or cleavage by proteolytic enzymes (37, 38). Analysis of the hydropathy profile and secondary structure predicted from the pChAT-1 sequence displayed in Fig. 2 does not allow confirmation that this protein is either the soluble or the membrane-bound form of ChoAcTase. Both forms could arise from the same ChoAcTase pre-mRNA through alternative splicing, as has been described in several instances (39). The fact that only one band is observed in RNA gel blot experiments does not necessarily mean that there is only one messenger since small, but perhaps functionally significant, differences cannot be detected by this method. Generation of molecular diversity through differential splicing could also provide a means to generate multiple ChoAcTase molecules endowed with distinct properties, whether they are soluble or membrane-bound. Such a mechanism was suggested for tyrosine hydroxylase (7).

Even when optimally aligned, with the creation of several gaps, the amino acid sequence of porcine ChoAcTase exhibits only 32% identity with that of Drosophila (8). However, six domains are more highly conserved: they display up to 79% homology (see Results) and are likely to contain the structural features necessary for catalytic activity. In this regard, Malthe-Sorenssen reported (40) that histidine residue plays a crucial role in the enzymatic reaction. Among the six domains mentioned above, only one (amino acids 292-344) contains conserved histidine residues, which could participate in the catalytic reaction. The importance of these various structural features can now readily be tested by mutagenesis experiments, taking advantage of the rabbit reticulocyte expression system. Although Drosophila and porcine ChoAcTase have similar apparent molecular weights, the Drosophila cDNA sequenced by Itoh et al. (8), in which no initiation codon was found, is at least 73 amino acids longer than that of porcine ChoAcTase. In view of the apparently abnormal length of the inferred Drosophila amino acid sequence, Itoh et al. (8) suggested that Drosophila ChoAcTase may be derived from a larger precursor that is enzymatically inactive. It is not clear whether this size discrepancy between porcine and Drosophila ChoAcTase results from species differences or from molecular diversity of the protein within a given species. It is of interest that two enzymes serving the same function differ substantially in their primary structure. In this context, it might be of relevance to note that the specific activity of Drosophila ChoAcTase is considerably higher than ChoAcTase from mammals (41).

The porcine cDNA hybridizes with rat and human mRNAs. This cross-reactivity should facilitate the identification of the corresponding genes. In human, the ChoAcTase clone could be of great interest in the study of degenerative diseases in which cholinergic systems are implicated.

We are grateful to J. F. Mayaux and T. Ciora (Rhône-Poulenc Santé) for providing the oligonucleotides. We thank A. Menez, J. Smith, and colleagues in the laboratory for helpful discussion; J. Randle for critical reading of the manuscript; and C. Bréant, J. Clot, J. Le Flohic, G. Peudevin, and D. Samolyk for technical assistance. This work was supported by grants from the Bundesministerium für Forschung und Technik, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Fondation pour la Recherche Médicale, the Ministère de la Recherche et de l'Enseignement Supérieur, and Rhône-Poulenc Santé. A.B. received fellowships from the Fondation pour la Recherche Médicale and the Fonds d'Etude du Corps Médical des Hopîtaux de Paris.

- Rossier, J. (1977) Int. Rev. Neurobiol. 20, 284-337.
- 2. Coyle, J. T., Price, D. L. & DeLong, M. R. (1983) Science 219, 1184-1190.
- Black, I. B. & Patterson, P. H. (1980) Curr. Top. Dev. Biol. 15, 27-40. 3.
- Le Douarin, N. M. (1980) Nature (London) 286, 663-669.
- Lamouroux, A., Faucon Biguet, N., Samolyk, D., Privat, A., Salomon, 5. J. C., Pujol, J. F. & Mallet, J. (1982) Proc. Natl. Acad. Sci. USA 79, 3881-3885.
- 6. Grima, B., Lamouroux, A., Blanot, F., Faucon Biguet, N. & Mallet, J. (1985) Proc. Natl. Acad. Sci. USA 82, 617-621
- Grima, B., Lamouroux, A., Boni, C., Julien, J. F., Javoy-Agid, F. & Mallet, J. (1987) Nature (London) 326, 707-711. 7.
- 8. Itoh, N., Slemmon, J. R., Hawke, D. H., Williamson, R., Morita, E., Itakura, K., Roberts, E., Shively, J. E., Crawford, G. D. & Salvaterra, P. M. (1986) Proc. Natl. Acad. Sci. USA 83, 4081-4085.
- Lomedico, P. T. & Saunders, G. F. (1976) Nucleic Acids Res. 3, 9. 381-391.
- 10. Dudley, J. P., Butel, J. S., Socher, S. H. & Rosen, J. M. (1978) J. Virol. 28, 743-752.
- Gubler, U. & Hoffman, B. J. (1983) Gene 25, 263-269. 11.
- Huynh, T. V., Young, R. A. & Davies, R. W. (1985) in DNA Cloning: A 12. Practical Approach, ed. Glover, D. M. (IRL, Oxford), Vol. 1, 49-78.
- Braun, A., Barde, Y.-A., Lottspeich, F., Mewes, W. & Thoenen, H. (1987) J. Neurochem. 48, 16-21. 13.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A 14. Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Deininger, P. L. (1983) Anal. Biochem. 129, 216-223. 15
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. 16. USA 74, 5463-5467.
- Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & 17. Green, M. R. (1984) Nucleic Acids Res. 12, 7035-7056.
- Krieg, P. A. & Melton, D. A. (1984) Nucleic Acids Res. 12, 7057-7070. 18.
- 19.
- Pelham, H. R. B. & Jackson, R. J. (1976) Eur. J. Biochem. 67, 247-256. Dobberstein, B., Garoff, M., Warren, G. & Robinson, P. J. (1979) Cell 20. 17, 759-769.
- 21. Eckenstein, F., Barde, Y.-A. & Thoenen, H. (1981) Neuroscience 6, 993-1000.
- 22. Eckenstein, F. & Thoenen, H. (1982) EMBO J. 1, 363-368.
- 23. Laemmli, U. K. (1970) Nature (London) 227, 680-685.
- 24. Berrard, S., Faucon Biguet, N., Gregoire, D., Blanot, F., Smith, J. & Mallet, J. (1986) Neurosci. Lett. 72, 93-98.
- 25. Smith, J., Fauquet, M., Ziller, C. & Le Douarin, N. M. (1979) Nature (London) 282, 853-855.
- White, H. L. & Cavallito, C. J. (1970) Biochim. Biophys. Acta 206, 26. 343-358.
- Martin, H. M. & Castro, M. M. (1985) Nucleic Acids Res. 13, 27. 8927-8938.
- 28. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
- Cohen, P. (1985) Eur. J. Biochem. 151, 439-448.
- Pearson, R. B., Woodget, J. R., Cohen, P. & Kemp, B. E. (1985) J. 30. Biol. Chem. 260, 14471-14476.
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in Atlas of 31. Protein Sequence and Structure, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345-352
- 32. Faucon Biguet, N., Buda, M., Lamouroux, A., Samolyk, D. & Mallet, J. (1986) EMBO J. 5, 287-291.
- Lathe, R. (1985) J. Mol. Biol. 183, 1-12. 33.
- Bruce, G., Wainer, B. H. & Hersch, L. B. (1985) J. Neurochem. 45, 34. 611-620.
- Benishin, C. G. & Carrol, P. T. (1983) J. Neurochem. 41, 1030-1039. 35.
- Eder-Colli, L. & Amato, S. (1985) Neuroscience 15, 577-589. 36.
- Malthe-Sorenssen, D. (1976) J. Neurochem. 26, 861-865. 37.
- Hersh, L. B., Wainer, B. H. & Andrews, L. P. (1984) J. Biol. Chem. 259, 1253-1258. 38.
- Leff, S. E. & Rosenfeld, M. G. (1986) Annu. Rev. Biochem. 55, 1091-1117. 39.
- 40. Malthe-Sorenssen, D. (1976) J. Neurochem. 41, 1030-1039.
- Slemmon, J. R., Salvaterra, P. M., Crawford, G. D. & Roberts, E. 41. (1982) J. Biol. Chem. 257, 3847-3852.