

Molecular Cloning and Complete Amino Acid Sequence of AP50, an Assembly Protein Associated with Clathrin-Coated Vesicles

CHRISTOPHE THURIEAU,^{*,**} JURGEN BROSIUS,[†] CINDY BURNE,[‡] PIERRE JOLLES,^{*}
JAMES H. KEEN,[§] ROBERT J. MATTALIANO,[‡] E. PINGCHANG CHOW,[‡]
KUZHAL L. RAMACHANDRAN,[‡] and TOMAS KIRCHHAUSEN^{**}

ABSTRACT

AP50 is the 50,000-dalton protein component found in clathrin-coated vesicles as part of the coat assembly protein (AP) complex, AP-2. AP50 cDNA clones were isolated from rat brain cDNA libraries, and their nucleotide sequence was determined. The isolated cDNA clones represent the entire coding sequence for the rat brain AP50. They encode a polypeptide containing 435 amino acids with a molecular weight of 49,612 daltons. Comparison with the partially sequenced bovine brain AP50 shows a primary structure that is highly conserved. AP50 does not have detectable sequence similarity with other known kinases or with other proteins of known sequence.

INTRODUCTION

CLATHRIN-COATED PITS and coated vesicles are organelles that mediate vesicular traffic between membrane compartments in eukaryotic cells (Goldstein *et al.*, 1979; Harrison and Kirchhausen, 1983; Pearse, 1980). The coat that surrounds the cytoplasmic face of the coated vesicles contains, in addition to clathrin heavy and light chains, a macromolecular complex known as the clathrin coat assembly protein complex (AP) and also referred to as the clathrin coat-associated proteins. Purified APs interact with clathrin and direct its assembly *in vitro* entirely into small cages under physiological conditions (Zaremba and Keen, 1983; Pearse and Robinson, 1984; Keen, 1987; Virshup and Bennett, 1988). AP-2 is the predominant form of the AP complex that can be isolated from calf brain coated vesicles. AP-2 has a Stokes' radius of about 66 Å determined by gel filtration (Virshup and Bennett, 1988) and a molecular weight of approximately 343,000 daltons determined by low-angle laser light-scattering analysis (Keen, 1987). This AP is composed of several polypep-

tide chains ranging in molecular sizes from 16,000 to 115,000 daltons as determined by NaDodSO₄-polyacrylamide gel electrophoresis. They are denoted AP17, AP50, AP100, AP102, AP104, AP112, and AP115, and they are bound to each other through noncovalent interactions with a probable molar ratio of two copies each of AP17, AP50, and the AP100-AP115 class (Keen, 1987; Pearse and Robinson, 1984), although other proportions have been proposed (Virshup and Bennett, 1988). Chemical cross-linking experiments suggest that elements of the AP100-AP115 group are in direct contact with each other (Pearse and Robinson, 1984; Virshup and Bennett, 1988). Units of AP50 in the complex cross-link to members of the AP100-AP115 group but not to other AP50s (Pauloin and Jolles, 1982; Pearse, 1984; Virshup and Bennett, 1988).

AP50 is readily phosphorylated *in vitro* on serine and threonine residues by at least two Mg-ATP-dependent, Ca²⁺-, calmodulin-, cyclic nucleotide-independent kinase activities found in purified coated vesicles or in the APs themselves (Pauloin *et al.*, 1982; Campbell *et al.*, 1984; Keen *et al.*, 1987; Manfredi and Barzari, 1987). It is be-

*Universite de Paris V, Unite associee CNRS N° 1188, France.

†Department of Genetics and Development, Columbia University, New York, NY 10032.

‡Biogen Research Corp., Cambridge, MA 02142.

§Fels Institute for Cancer Research and Molecular Biology, Temple University, Philadelphia, PA 19140.

**Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, MA 02115.

lieved that one of these modifications may result from an autophosphorylation activity of AP50 (Campbell *et al.*, 1984; Keen *et al.*, 1987). The phosphorylation of AP50 has no apparent effect on the *in vitro* interaction between AP and clathrin (Keen, 1987; T. Kirchhausen, unpublished observations).

In an ongoing effort to characterize the AP complex at a molecular level, we have obtained rat brain cDNA clones for AP50. We describe here the full-length nucleotide sequence and the complete primary structure of rat brain AP50. We have found a remarkable conservation of amino acid sequence in a comparison between rat and bovine brain AP50. In contrast, we have found no significant sequence relationship between AP50 and kinases or other proteins of known sequence.

METHODS

Purification of AP50

Clathrin coat assembly protein complexes (AP-2) containing AP17, AP50, AP100, AP102, AP104, AP112, and AP115 were purified from calf brain coated vesicles by the Tris dissociation method (Zaremba and Keen, 1987). After coat depolymerization, clathrin and APs were separated from membrane vesicles by high-speed centrifugation followed by sizing chromatography (Zaremba and Keen, 1987). For most amino-terminal amino acid sequence determination of AP50, AP-2 was further purified by clathrin affinity chromatography (Keen, 1987) followed by preparative NaDodSO₄/12% polyacrylamide gel electro-

phoresis. The band containing AP50 was identified by 1 M KCl staining, excised, and then electroeluted and extensively dialyzed against 0.1% NaDodSO₄, 25 mM ammonium bicarbonate (Hunkapiller *et al.*, 1983).

Amino-terminal sequencing

To obtain the amino-terminal amino acid sequence of bovine brain AP50, about 1 nanomole of electroeluted protein was applied to a gas-phase amino acid sequencer (Applied Biosystems) and subjected to automated Edman degradation. Internal amino acid sequence was also obtained from another sample of AP50 that was succinylated, carboxymethylated, and then subjected to trypsin digestion (Huang *et al.*, 1986). The proteolytic fragments were separated by reverse-phase high-pressure liquid chromatography and selected peptides were then analyzed with the amino acid sequencer. Confirmatory sequence data corresponding to the amino terminus and peptides 5 and 7 (Fig. 2, below) were obtained from AP50 derived from a cruder sample of AP-2 (purified by sizing chromatography).

cDNA isolation

Based on the amino-terminal amino acid sequence of bovine brain AP50, a synthetic oligonucleotide probe of 71 bases was prepared by chemical methods (McBride and Caruthers, 1983 (Fig. 1)). One region of the probe was designed with a partially degenerate nucleotide sequence

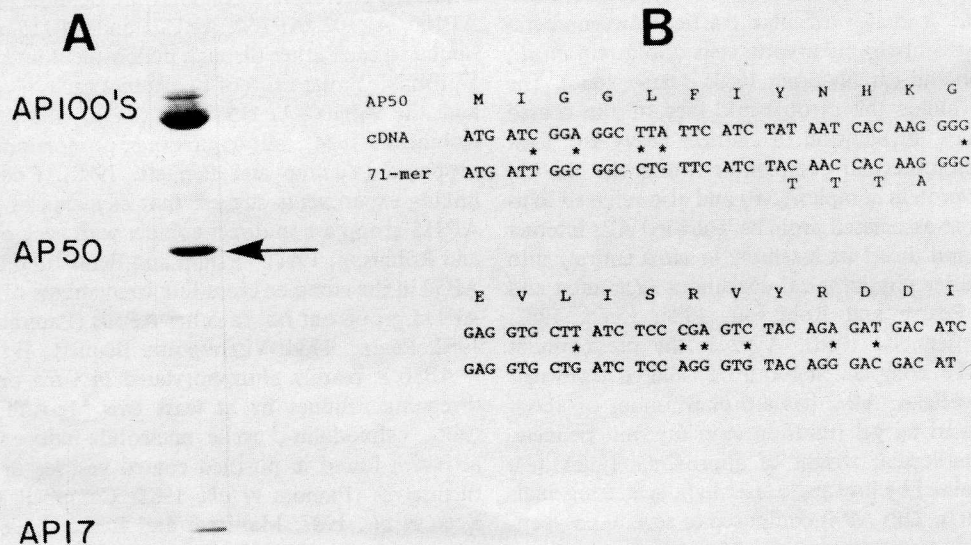


FIG. 1. Determination of the amino-terminal amino acid sequence of bovine brain AP50 and design of oligonucleotide probe. A. A representative sample of clathrin affinity-purified bovine brain AP-2 that had been subject to preparative NaDodSO₄-12% polyacrylamide gel electrophoresis and used for the isolation of AP50. B. The amino-terminal sequence of AP50 (sequence 1, Fig. 2) was obtained from the electroeluted band of *M_r* 50,000 that had been subjected to automated Edman degradation. The partially degenerate 71-mer oligonucleotide probe used in the rat cDNA library search was designed according to the pattern of codon usage in rat (Maruyama, 1986). A nucleotide identity of 85% with the rat AP50 cDNA was obtained. Mismatches are indicated by a star.

which introduced a 16-fold redundancy and guaranteed a stretch of 14 uninterrupted bases without mismatch. The remainder of the probe sequence was based on a guess of the third position according to the pattern of preferred codon usage in rat (Maruyama *et al.*, 1986). The probe was labeled with [γ - 32 P]ATP and T4 kinase and used to search for hybridizing recombinants in a λ gt10 rat brain cytoplasmic poly(A)⁺cDNA library (Mocchetti *et al.*, 1986). Hybridization conditions were 20% formamide, 5 \times SSPE (SSPE is 0.15 M NaCl, 0.01 M Na phosphate, 0.001 M EDTA), 0.1% NaDodSO₄, and 100 μ g/ml of boiled salmon sperm DNA at 42°C overnight. Nitrocellulose filters were washed with 2 \times SSC (SSC is 0.15 M NaCl, 0.015 M Na citrate), 0.5% NaDodSO₄ at 42°C.

Since the initial rat brain cDNA clones that were identified did not code for the entire AP50 amino acid sequence, additional recombinants were isolated from a search of a plasmid cDNA library made from rat brain cerebral cortex poly(A)⁺RNA (a gift from M. Brownstein, NIMH). This library was probed with the complete *Eco* RI insert of clone λ 5511 (isolated from the λ gt10 library) that had been labeled with [α - 32 P]dATP by nick-translation. Hybridization was performed as before except for the presence of 50% formamide and the washes were carried out at 55°C.

DNA sequencing, DNA and RNA blots

The complete sequence of four overlapping and independent clones was determined by a combination of the chemical (Maxam and Gilbert, 1980) and the chain-termination (Sanger *et al.*, 1977) methods. The 17-mer oligonucleotide primers were synthesized and used to sequence al-DNA and RNA blots were carried out using the same conditions described previously (Devlin *et al.*, 1986; Wallner *et al.*, 1986).

RESULTS

To obtain cDNA clones that code for AP50, we determined the amino-terminal amino acid sequence, from which an appropriate oligonucleotide probe was designed, synthesized, and used to screen the library.

Since the amino terminus of AP50 purified from calf brain coated vesicles is not blocked, we were able to determine the sequence of the first 56 amino acids by automated microsequencing (sequence 1, Fig. 1 and Fig. 2). We selected a region of 24 amino acids to design an oligonucleotide probe of 71 bases of partially degenerate nucleotide sequence (Fig. 1). Seven independent clones were identified after the initial screen of a λ gt10 cDNA library from rat brain. The deduced amino acid sequence of the *Eco* RI inserts in clones λ 5511 and λ 5311 (about 300 and 600 bp in length, Fig. 3) confirmed their identity as partial length AP50 cDNA clones. The agreement included the complete amino acid sequence of the first 56 residues located at the amino terminus of bovine brain AP50.

To obtain full-length clones, additional recombinants

were isolated from a search of the rat brain plasmid cDNA library using the *Eco* RI insert of clone λ 5511 as a probe. Two independent cDNA clones (pCT3a and pCT6b, Fig. 3) were obtained and fully characterized. The longest AP50 clone (pCT3a) has 1814 bases and is fully described in Fig. 2. It contains the ATG translational start codon, defined as position 1, followed by an open reading frame that starts with the amino terminus of the bovine brain AP50. The ATG is immediately preceded by a stop codon at position -12 and a 5' untranslated region of at least 62 bases. The open reading frame has 1305 bases and defines a protein of 435 amino acids, molecular weight of 49,612 daltons, and pI of 10. The predicted size and pI of rat AP50 is in excellent agreement with both the apparent molecular weight of 50,000 daltons obtained for bovine AP50 by NaDodSO₄-polyacrylamide gel electrophoresis (Keen, 1987) and its behavior as a highly basic protein on two-dimensional nonequilibrium pH gradient gel electrophoresis (Keen and Black, 1986). The stop codon at position 1306 defines a 3' untranslated region of 452 bp, including the consensus polyadenylation signal AATAAA at position 1745 and a poly(A)⁺ tail. Comparison of the two plasmid clones pCT3a and pCT6b indicates that clone pCT6b is slightly shorter at its 5' end (Fig. 3) and that it has a silent mutation (C-T, nucleotide position 467), probably due to allelic variation. Clone pCT6b also shows 7 additional bases (at position 1754) preceding the poly(A)⁺ tail (Fig. 2).

Although we cannot completely rule out the possibility that other proteins with partial sequence homology may exist, we believe that AP50 is a unique protein of the clathrin coat-associated proteins system. The RNA blot in Fig. 4A shows that an AP50 transcript of the expected size (about 2 kb) is present in several rat tissues including brain, heart, lung, liver, testis, and spleen. Within the resolution of the blot, this result shows the absence of related mRNA molecules of different size that could have been generated either by alternative splicing or by transcription from other related gene(s). The DNA blot analysis of rat genomic DNA shown in Fig. 4B, performed with either a DNA probe that spans coding nucleotide positions 1273-1305 or most of the AP50 3' untranslated region (not shown) confirms that only one gene codes for AP50. On the basis of these data, we suggest that AP50 is unrelated (or at least not highly related) to another less abundant protein species of about 47,000 daltons (Keen, 1987; Manfredi and Barzari, 1987; Ahle *et al.*, 1988) also found in bovine brain coated vesicles. It is believed that this polypeptide is part of an additional AP complex that can also bind *in vitro* to clathrin and induce the formation of coats of different morphology (Pearse and Robinson, 1984; Keen *et al.*, 1987). Unlike AP50, this 47,000-dalton protein is not susceptible to *in vitro* phosphorylation (Manfredi and Barzari, 1987).

DISCUSSION

Analysis of the primary structure of AP50 shows that there are no unusual clusters of amino acids or repeating

elements. No significant amino acid sequence homology was found in a computer search between AP50 and the other major components of coated vesicles [AP17, (S. Frucht and T. Kirchhausen, unpublished data), AP100-115 (T. Kirchhausen *et al.*, unpublished data), and clathrin (Kirchhausen *et al.*, 1987)] and light chains (Kirchhausen *et al.*, 1987; Jackson *et al.*, 1987) nor with other proteins whose sequences have been deposited in the National Bio-

medical Research Foundation protein data bank (release XII, 1987). The search conducted using FASTP (Lipman and Pearson, 1985) also failed to indicate any significant relationship with the microtubule binding protein Tau from mouse brain (Lee *et al.*, 1988) that has been described as immunologically cross-reactive with AP50 (Pfeffer *et al.*, 1983). In addition, no significant similarity of primary structure was found with the sequence of known kinases or of the various nucleotide binding proteins (Hunter, 1987). This is a surprising result since the AP50 may undergo autophosphorylation (Campbell *et al.*, 1984; Keen *et al.*, 1987), and several groups (Pauloin *et al.*, 1982; Campbell *et al.*, 1984; Keen and Black, 1986; Manfredi and Barzari, 1987) have shown that it can be phosphorylated *in vitro* by an Mg-ATP-dependent and Ca²⁺-calmodulin- and cAMP-independent kinase activity that copurifies with coated vesicles and APs.

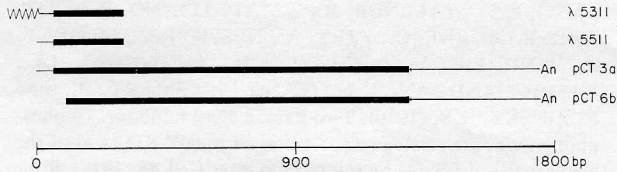


FIG. 3. Schematic representation of sequenced rat brain cDNA clones for AP50. cDNA clones were derived from a λgt10 and plasmid rat brain libraries. Poly(A)⁺ tails are marked by An. Zigzag lines indicate unrelated cDNA portion probably derived by the fusion prior to linker addition. DNA sequence determination was performed on both DNA strands by chemical (Maxam and Gilbert, 1980) and chain termination (Sanger *et al.*, 1977) methods.

The primary structure of AP50 is highly conserved in the two mammalian species we have studied. We have compared the predicted amino acid sequence deduced from the cDNA clones of rat brain AP50 with the amino-terminal sequence from intact AP50 and from all the tryptic fragments of bovine brain AP50 whose sequences were determined by automated Edman degradation (a total of 128

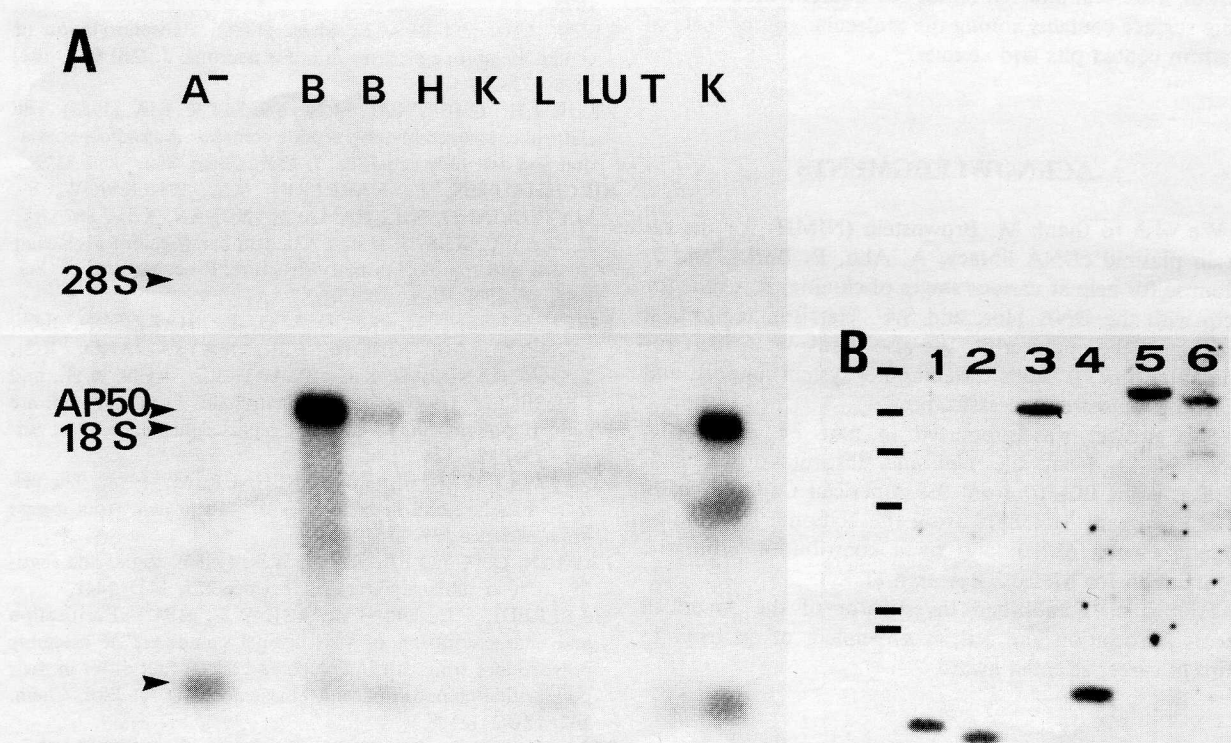


FIG. 4. RNA and DNA blot analysis of rat samples. **A.** For the RNA blot analysis, poly(A)⁻RNA (~1 μg) from rat brain and poly(A)⁺RNA (~1 μg) from rat brain (B), heart (H), kidney (K), liver (L), lung (LU), and testis (T) were probed with the complete length of the rat AP50 cDNA. A mRNA species of similar size and of the same mobility as β-actin mRNA is present in all tissues studied. A small RNA species, also present in poly(A)⁻RNA is suggestive of the existence of a repetitive sequence element. 28S and 18S rRNA size markers are indicated. **B.** The genomic DNA blot analysis was performed with rat liver genomic DNA cut with *Pst* I, *Sac* I, *Pvu* II, *Bam* HI, *Eco* RV, and *Eco* RI, and probed with an AP50 oligonucleotide probe spanning nucleotide positions 1273-1305 of the coding region. Size markers: 23,000, 9400, 6700, 4300, 2300, 2000, and 500 bp.

amino acids, see Fig. 2). This analysis shows that in these regions both proteins are practically identical, including the basic residues that are the sites of tryptic cleavage. Since the peptides reflect a random sampling along the amino acid sequence of AP50, we suggest that a high level of identity also occurs throughout the complete AP50 molecule. The theme of a highly conserved primary structure confirms sequence studies on the other major protein components of mammalian coated vesicles where an interspecies sequence identity of at least 93% was detected. These proteins include the AP17 protein components of the AP complex and the AP100-115 group, as well as the clathrin heavy chain (Kirchhausen *et al.*, 1987) and the various classes and types of clathrin light chains (Kirchhausen *et al.*, 1987; Jackson *et al.*, 1987). Thus, we believe that this new result obtained with AP50 further substantiates our previous assertion that there must exist a strong pressure to maintain the structure of these proteins. This, we think is a consequence of the extensive intermolecular surface interactions that must exist between these major protein components of coated pits and coated vesicles. It seems likely that these interactions participate in the coordinated cycles of assembly and disassembly of the coat as well as in the process of receptor sorting. The information presented in this work will therefore be useful in the pursuit of structural and cell biological studies aimed at mapping surface contacts among the molecular components of clathrin coated pits and vesicles.

ACKNOWLEDGMENTS

We wish to thank M. Brownstein (NIMH) for the rat brain plasmid cDNA library; A. Ahn, P. Burke, and J. Monroe for help at various stages of cloning; R. Cathe for help with the DNA blot; and S.C. Harrison for critical reading of the manuscript. We also thank W. Gilbert for computer time, L. Pops for computer system support, and L. Scott for secretarial assistance.

This project was supported in part by grant R01-GM36548-01 from the National Institutes of Health (T.K.); grant BC-567 from the American Cancer Society (J.H.K.); grant MH38819 from the National Institute of Mental Health (J.B.); and by a contribution from the Foundation for Medical Research (T.K.).

T.K. is an Established Investigator of the American Heart Association and J.B. is a recipient of an Irma T. Hirsch career scientist award.

REFERENCES

- AHLE, S., MANN, A., EICHELNBACHER, U., and UNGEWICKELL, E. (1988). Structural relationships between clathrin assembly proteins from the Golgi and the plasma membrane. *EMBO J.* **7**, 919-929.
- CAMPBELL, C., SQUICCIARINI, J., SHIA, M., PILCH, P.F., and FINE, R.E. (1984). Identification of a protein kinase as an intrinsic component of rat liver coated vesicles. *Biochemistry* **23**, 4420-4426.
- DEVLIN, P.E., RAMACHANDRAN, K.L., and CATE, R.L. (1988). Southern analysis of genomic DNA with unique and degenerate oligonucleotide probes: A method for reducing probe degeneracy. *DNA* **7**, 499-507.
- GOLDSTEIN, J.L., ANDERSON, R.G.W., and BROWN, M.S. (1979). Coated pits, coated vesicles, and receptor mediated endocytosis. *Nature* **279**, 679-685.
- HARRISON, S.C., and KIRCHHAUSEN, T. (1983). Clathrin, cages and coated vesicles. *Cell* **33**, 650-652.
- HUANG, K.S., WALLNER, B.P., MATTALIANO, R.J., TIZARD, R., BURNE, C., FREY, A., HESSION, C., McGRAY, P., SINCLAIR, L.K., CHOW, E.P., BROWNING, J.L., RAMACHANDRAN, K.L., TANG, J., SMART, J., and PEPINSKY, R.B. (1986). Two human 35kd inhibitors of phospholipase a_2 are related to substrates of pp60V-*src* and of the epidermal growth factor receptor/kinase. *Cell* **46**, 191-199.
- HUNKAPILLER, M.W., LUJAN, E., OSTRANDER, F., and HOOD, L.E. (1983). *Methods Enzymol.* **91**, 227-236.
- HUNTER, T. (1987). A thousand and one protein kinases. *Cell* **50**, 823-829.
- JACKSON, A.P., SEOW, J.F., HOLMES, N., KRICKAMER, K., and PARHAM, P. (1987). Clathrin light chains contain brain-specific insertion sequences and a region of homology with intermediate filaments. *Nature* **326**, 154-159.
- KEEN, J.H. (1987). Clathrin assembly proteins: Affinity purification and a model for coat assembly. *J. Cell Biol.* **105**, 1989-1998.
- KEEN, J.H., and BLACK, M.M. (1986). Phosphorylation of coated membrane proteins in intact neurons. *J. Cell Biol.* **102**, 1325-1333.
- KEEN, J.H., CHESTNUT, M.H., and BECK, K.A. (1987). The clathrin coat assembly polypeptide complex: Autophosphorylation and assembly activities. *J. Biol. Chem.* **262**, 3864-3871.
- KIRCHHAUSEN, T., HARRISON, S.C., PINGCHOW, E., MATTALIANO, R.J., RAMACHANDRAN, K.L., SMART, J., and BROSIUS, J. (1987). Clathrin heavy chain: Molecular cloning and complete primary structure. *Proc. Natl. Acad. Sci. USA* **84**, 8805-8809.
- KIRCHHAUSEN, T., SCARMATO, P., HARRISON, S.C., MONROE, J.S., CHOW, E.P., MATTALIANO, R.S., RAMACHANDRAN, K.L., SMART, J.E., AHN, A.H., and BROSIUS, J. (1987). Clathrin light chains LCA and LCB are similar, polymorphic, and show repeated heptad motifs. *Science* **236**, 320-324.
- LEE, G., COWAN, J., and KIRSCHNER, M. (1988). The primary structure and heterogeneity of Tau protein from mouse brain. *Science* **239**, 285-288.
- LIPMAN, D.J., and PEARSON, W.R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435-1441.
- MANFREDI, J.J., and BARZARI, W.L. (1987). Purification and characterization of two distinct complexes of assembly polypeptides from calf brain coated vesicles that differ in their polypeptide composition and kinase activities. *J. Biol. Chem.* **262**, 12182-12188.
- MARUYAMA, T., GOJOBORI, T., AOFA, S., and IKEMURA, T. (1986). Codon usage tabulated from the genebank genetic sequence data. *Nucleic Acids Res.* **14**, r151-r197.
- MAXAM, A.M., and GILBERT, W. (1980). Sequencing N-labelled DNA with base-specific chemical cleavage. *Methods Enzymol.* **65**, 499-560.
- McBRIDE, L.J., and CARUTHERS, M.H. (1983). An investigation of several deoxynucleotide phosphoramidites useful for synthesizing deoxyoligonucleotides. *Tetrahedron Lett.* **24**, 245-248.

- MOCCHETTI, I., EINSTEIN, R., and BROSIUS, J. (1986). Putative diazepam binding inhibitor peptide: cDNA clones from rat. *Proc. Natl. Acad. Sci. USA* **83**, 7221-7225.
- PAULOIN, A., BERNIER, I., and JOLLES, P. (1982). Presence of cyclic nucleotide-Ca independent protein kinase in bovine brain coated vesicles. *Nature* **298**, 574-576.
- PEARSE, B.M.F. (1980). Coated vesicles. *Trends Biochem. Sci.* **5**, 131-134.
- PEARSE, B.M.F., and ROBINSON, M. (1984). Purification and properties of 100-kd proteins from coated vesicles and their reconstitution with clathrin. *EMBO J.* **3**, 1951-1957.
- PFEFFER, S.R., DRUBIN, D.G., and KELLY, R.B. (1983). Identification of three coated vesicle components as α - and β -tubulin linked to a phosphorylated 50,000-dalton polypeptide. *J. Cell Biol.* **97**, 40-47.
- SANGER, F., NICKLEN, S., and COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- VIRSHUP, D., and BENNETT, V. (1988). Clathrin coated vesicle assembly polypeptides: Physical properties and reconstitution studies with brain membranes. *J. Cell Biol.* **106**, 39-50.
- WALLNER, B.P., MATTALIANO, R.J., HESSION, C., CATE, R., TIZARD, R., SINCLAIR, L., FOELLER, C., CHOW, E.P., BROWNING, J., RAMACHANDRAN, K.L., and PEPINSKY, R.B. (1986). Cloning and expression of human lipocortin, a phospholipase A₂ inhibitor with potential antiinflammatory activity. *Nature* **320**, 77-81.
- ZAREMBA, S., and KEEN, J.H. (1983). Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J. Cell Biol.* **97**, 1339-1347.

Address reprint requests to:

Dr. T. Kirchhausen
Department of Anatomy and Cellular Biology
Harvard Medical School
220 Longwood Avenue
Boston, MA 02115

Received for publication June 2, 1988, and in revised form August 15, 1988.