

AP17 and AP19, the Mammalian Small Chains of the Clathrin-associated Protein Complexes Show Homology to Yap17p, Their Putative Homolog in Yeast*

(Received for publication, June 26, 1990, and in revised form, March 6, 1991)

Tomas Kirchhausen^{‡§}, Alvin C. Davis[¶], Steven Frucht[‡], Betsy O'Brine Greco^{||},
Gregory S. Payne^{**†‡}, and Benjamin Tubb[‡]

From the Departments of [‡]Anatomy and Cellular Biology and [¶]Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, ^{||}Biogen Incorporated, Cambridge, Massachusetts 02142, and the ^{**}Department of Biological Chemistry, University of California, Los Angeles, California 90024

AP17 and AP19 are the smallest polypeptide chain components of AP-2 and AP-1, the clathrin-associated protein complexes found in coated structures of the plasma membrane and Golgi apparatus of mammalian cells. cDNA clones representing the entire coding sequence of AP17 and AP19 were isolated from rat and mouse brain cDNA libraries, respectively. Determination of their nucleotide sequence predicts proteins of 142 and 158 amino acids with M_r 17,018 and 18,733. A sequence comparison of rat brain AP17 with mouse brain AP19 demonstrates that the small chains are highly related. A computer search for other related proteins has uncovered in yeast a previously unknown gene whose DNA sequence encodes a protein homologous to the small chain of AP complexes. The yeast sequence predicts Yap17p, a protein with 147 amino acids and a M_r of 17,373 that is slightly more related to the mammalian AP17 chain than to its AP19 counterpart.

Clathrin-coated pits and coated vesicles are involved in the sorting of membrane proteins and their vesicular intracellular traffic (reviewed in Goldstein *et al.*, 1985; Pfeffer and Rothman, 1987; Brodsky, 1988). Clathrin and its associated protein complexes (APs)¹ are the main protein components of the coat surrounding the cytoplasmic face of coated vesicles. AP-1 and AP-2, two main types of APs, are found in clathrin-coated structures located at the Golgi complex and the plasma membrane of mammalian cells, respectively (Robinson and Pearse, 1986; Keen, 1987; Ahle *et al.*, 1988).

The APs are positioned on the underside of the clathrin lattice (Heuser and Kirchhausen, 1985; Vigers *et al.*, 1986). APs are believed to interact with the cytoplasmic tails of membrane proteins, leading to their selection and concentra-

tion into coated pits and coated vesicles (Pearse, 1988; Glickman *et al.*, 1989; Kirchhausen *et al.*, 1989). APs are heterotetrameric structures containing two large chains (M_r ~100,000; γ and β' in AP-1; α and β in AP-2), a medium chain (M_r ~50,000; AP47 in AP-1 or AP50 in AP-2), and a small chain (M_r 17,000-19,000; AP19 in AP-1 or AP17 in AP-2) (Keen, 1987; Ahle *et al.*, 1988; Virshup and Bennett, 1988; Kirchhausen *et al.*, 1989; Matsui and Kirchhausen, 1990).

We describe here the molecular cloning of cDNAs for rat brain AP17 and mouse brain AP19 and present the deduced amino acid sequence of the proteins. We demonstrate here that the mammalian AP17 and AP19 are related proteins displaying significant sequence identity, and we suggest that they probably arose by gene duplication.

In addition, we have uncovered *YAP17*, a yeast DNA sequence containing a previously unknown open reading frame that could encode a protein whose sequence is highly related to the mammalian AP17 and AP19. Although coated vesicles and clathrin can also be purified from yeast cells (Mueller and Branton, 1984; Payne and Schekman, 1985; Lemmon *et al.*, 1988), there is to date no direct biochemical evidence in yeast for the existence of clathrin-associated protein complexes. We have recently uncovered *YAP54²* and *YAP80* (Kirchhausen, 1990), the putative yeast homologs of the AP47/AP50 medium chains and of the β/β' large chains of the mammalian AP-2, respectively. Taken together with these results, the identification of *YAP17* provides further support to the proposal that yeast cells also contain AP complexes.

MATERIALS AND METHODS

Isolation and Amino-terminal Amino Acid Sequencing of Calf Brain AP17 and AP19—Partially purified AP complexes were obtained by sizing chromatography from the coats of calf brain-coated vesicles after their depolymerization by treatment with Tris (Keen *et al.*, 1979; Kirchhausen and Harrison, 1984). A sample containing AP-1 and AP-2 complexes (~0.2 mg) was concentrated by centrifugation in a Centricon-30 column (Amicon) and their polypeptide chains fractionated using preparative SDS-13% polyacrylamide gel electrophoresis (see Fig. 1; Laemmli, 1970). The band corresponding to AP17 was electrotransferred to an Immobilon membrane (Milipore) (Matsudaira, 1987), and 26 amino-terminal amino acids were determined by automated Edman degradation (Applied Biosystems). The sequence of AP17 corresponding to the first 10 amino acids was confirmed from the similar analysis of another sample of AP-2 complexes obtained after additional purification using hydroxylapatite chromatography (Pearse and Robinson, 1984). AP-1 complexes (0.1-0.2 mg) were obtained from a mixture of AP-1 and AP-2 complexes (~3 mg) after its partial purification by ionic exchange chromatography using

* This project was supported in part by grants (to T. K.) from the National Institutes of Health and the American Cancer Society, a Basil O'Connor Award from the March of Dimes, and by a contribution from the M. Peckett Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M37193, M37194, and M62418.

§ Established Investigator of the American Heart Association. To whom correspondence should be addressed.

†† Supported by a grant from the National Institutes of Health.

¹ The abbreviations used are: APs, associated protein complexes; SDS, sodium dodecyl sulfate; bp, base pair.

² Y. Nakayama, M. Goebel, B. O'Brine Greco, S. Lemmon, E. Pingchang Chow, and T. Kirchhausen, manuscript in preparation.

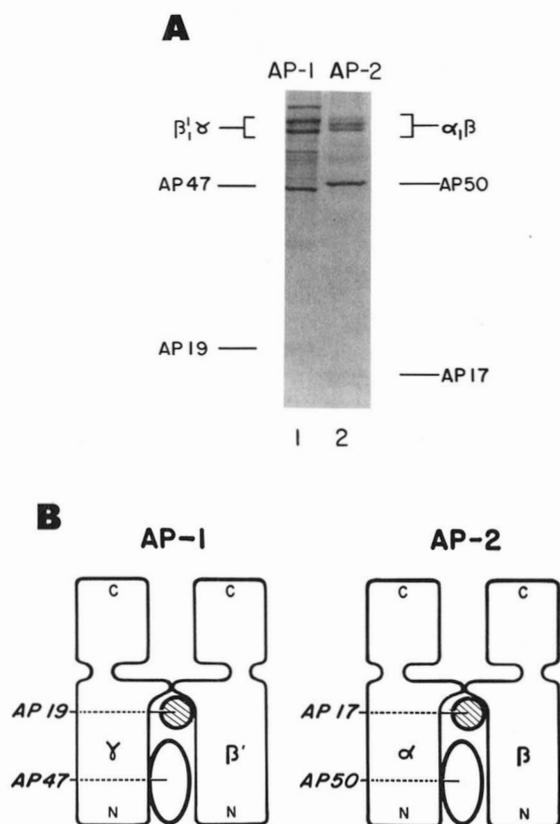


FIG. 1. Subunit composition and schematic representation of AP-1 and AP-2 complexes. *A*, representative samples of partially purified calf brain AP-1 and AP-2 complexes subjected to SDS-13% polyacrylamide gel electrophoresis. The amino-terminal sequences of the small chains AP17 and AP19 were determined by automated Edman degradation as described under "Materials and Methods." The protein sequences obtained from 26 and 30 cycles of automated Edman degradation performed on intact calf brain AP17 and AP19 samples were identical to the corresponding amino-terminal sequences of the rat brain AP17 and mouse brain AP19 deduced from their cDNAs. *B*, model for the quaternary structure for AP-1 and AP-2 complexes (Matsui and Kirchhausen, 1990). The high level of sequence identity between the primary structures of AP17 and AP19 demonstrates that they are highly related proteins.

a Sepharose Q column (Ahle *et al.*, 1988; Matsui and Kirchhausen, 1990). The polypeptide chains of AP-1 complexes were fractionated as described above (see Fig. 1) and the electrotransferred AP19 subjected to 30 cycles of automated Edman degradation. Purified AP19 was also obtained from the polyacrylamide gel by overnight diffusion followed by tryptic digestion and high performance liquid chromatography fractionation (Kirchhausen *et al.*, 1989). Several fragments were then selected for amino-terminal sequencing by automated Edman degradation.

Isolation of cDNAs for Rat Brain AP17—The sequence of the degenerate oligonucleotide used to screen an amplified cytoplasmic poly(A)⁺ λ gt10 rat brain cDNA library (Mocchetti *et al.*, 1986) was 5'-(G,A)TC(G,A)TC(G,A)AA(C,T)TGCA(T,G,A)TACCA(C,T)TTNGC where N indicates G, A, T, and C. It was designed to correspond to the noncoding strand according to the protein sequence of the calf brain AP17 at amino acid positions 17–26 (see Fig. 2). The probe was labeled with [γ -³²P]ATP and T4 kinase and used for overnight hybridization at 37 °C in 6 \times SSC (sodium sodium citrate), 0.5% SDS, 25 μ g/ml boiled tRNA, 2 \times Denhardt's solution. Nitrocellulose filters were washed with 2 \times SSC, 0.5% SDS at 42 °C. Additional recombinants were isolated from an amplified plasmid rat brain cerebral cortex poly(A)⁺ RNA cDNA library (a gift from M. Brownstein, National Institute of Mental Health (NIMH)) by hybridization with the complete *Eco*RI insert of clone λ SF2117 (isolated from the λ gt10 library) that had been labeled with [α -³²P]ATP by nick translation. Overnight hybridization was performed at 65 °C in 6 \times SSC, 0.1% SDS, 0.05% Triton X-100, 2 \times Denhardt's solution, 100 μ g/ml

boiled sperm DNA. Washes were carried out at 55 °C in 1 \times SSC, 0.5% SDS.

The complete sequences of different strands from three overlapping clones (λ SF2117, pSF1112, and pSF3711) were obtained by the chain-termination method (Sanger *et al.*, 1977) using 17-mer oligonucleotide-sequencing primers.

Isolation of cDNAs for Mouse Brain AP19—Based on the amino-terminal sequences of calf brain AP19 at positions 1–23, 119–128, and 134–146 (see Fig. 2), we synthesized the degenerate oligonucleotide probes: (1) 5'-ATGCAGIGITTIATGCTICTITTTCHICGICAGGGICTIIGI(C,T)TICA(A,G)AA(A,G)TGGTA(C,T)CTIGCIAC; (2) 5'-ATGCAGIGITTIATGCTICTITTTCHICGICAGGGICTIIGI(C,T)TICA(A,G)AA(A,G)TGGTA(C,T)TTIGCIAC; (3) 5'-TTI(C,C)(A,T)IGT(A,G)TC(C,T)TGIACGTCICCCCCAT; (4) 5'-TTI(G,C)(A,T)IGT(A,G)TC(C,T)TGIACATCICCCCCAT; and (5) 5'-GCIATIGA(A,G)CA(A,G)GCIGA(C,T)(C,T)TI(C,T)TICAGGAGGAGGACGA where I indicates deoxyinosine. The probes were used to screen an amplified mouse brain poly(A)⁺ RNA cDNA library inserted in the uniZapXR vector (Stratagene). Hybond (Amersham Corp.) filters were hybridized overnight at 42 °C in 6 \times SSC, 0.5% SDS, 5 \times Denhardt's solution, 100 μ g/ml tRNA, and washed with 6 \times SSC, 0.2% SDS at 37 °C. Several hybridizing uniZapXR phage clones were converted to plasmid cDNA by rescue excision according to the manufacturer's instructions. The identity of pBT6, one of the six longer clones of same size and restriction digest pattern, and pBT8 were determined by DNA sequencing.

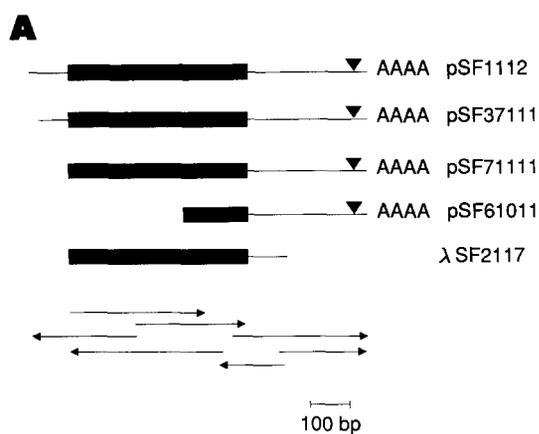
Computer Search for Related Proteins—A computer search of several protein and DNA data bases (National Biomedical Foundation, release 25; GenBank, release 65; EMBL, release 24) was performed with the FASTP (Lipman and Pearson, 1985), TFASTA (Pearson and Lipman, 1988), and WORD (Deveraux *et al.*, 1984) algorithms as modified in the Wisconsin software package for DNA analysis (Deveraux *et al.*, 1984) using the deduced protein sequence of the rat brain AP17 and mouse brain AP19 as queries.

Nucleotide Sequencing of YAP17, the Yeast Homolog of the Mammalian AP17 and AP19 Chains—Yeast DNA genomic clones spanning ~12,000 bp between the *Eco*RI sites surrounding the CDC8+ gene were kindly provided by R. Rothstein (Rothstein *et al.*, 1987). A portion of about 500 bp located at the 3' end of the CDC8+ gene encoding YAP17 was sequenced in both strands using 17-mer oligonucleotide primers and the chain-termination method (Sanger *et al.*, 1977).

RESULTS

Molecular Cloning of the cDNAs for Rat AP17 and Mouse AP19—In order to identify cDNA clones encoding AP17, oligonucleotide probes were designed based on the amino-terminal amino acid sequence of AP17 purified from calf brain-coated vesicles (see "Materials and Methods"). Several hybridizing cDNA clones were detected in two rat brain cDNA libraries. The nucleotide sequence of the longest rat brain AP17 cDNA clone detected (pSF1112) overlaps the sequence of the other isolated clones for AP17 and is described in Fig. 2. It has an open reading frame of 426 bases, coding for a protein of 142 amino acids and predicted M_r 17,017 (pI = 6.12). The ATG translational start codon at position 1 conforms to the consensus sequence for eukaryotic translational start sites A/GXXATGG/A (Kozak, 1989) and is preceded by a 5' untranslated region of 69 nucleotides. The stop codon at position 427 defines a 3'-untranslated region of 295 nucleotides, including a polyadenylation signal AATAAA at position 687 and a poly(A)⁺ tail.

A similar strategy was used to isolate AP19 cDNA clones from a mouse brain cDNA library (see "Material and Methods"). The nucleotide sequence compiled from examples of the two sizes of mouse brain AP19 cDNA clones (pBT6 and pBT8) is shown in Fig. 3. The sequences of these cDNA clones are identical except for 34 nucleotides missing at the 5'-terminal end of pBT6 and for 627 nucleotides missing in pBT8 between positions 533 and 1159 at the 3'-untranslated region of pBT6. The clones contain 1266 and 674 nucleotides and predict a protein of 158 amino acids and M_r 18,733 (pI =



B

```

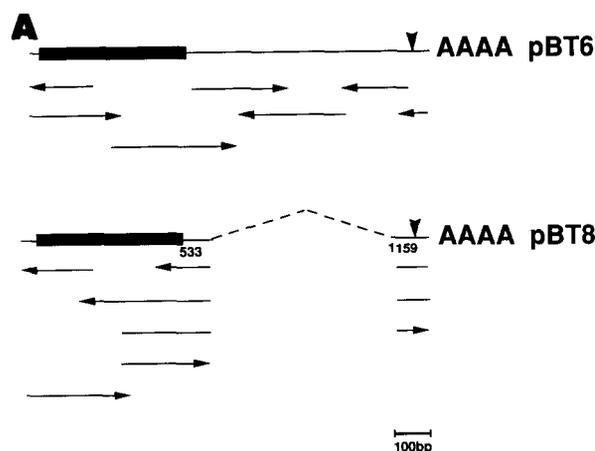
-69  TCGAAGCAGAGCCGAGCCGCCAGATTACAGGGGACCGGGCCCTCGGGTCCAGACA  -10
-9   GGGCTCCGCATGATCGGATTCATTCTCATCCAGAACCGGCGAGGCAAGACCCCTGGCC
      M I R F I L I Q N R A G K T R L A
.....
52  AAGTGGTATATCCAGTTGGATGATCAGCAGAAGCAGAGTTGATGAGGAAGTCCAGCCG  111
18  K M Y M Q F D D D E K Q K L I E E V H A  37
.....
112 GTGCTCAGCGTCAGGGATGCCAAGCAGCAACTTCTGCGATTGGGAACTTCAAGATC  171
38  V V T V R D A K H T H F V E F R N F K I  57
172 ATCTACGGCCCTAGCTGGCTTACTTTCGATCTGGTTGAGTCAAGCAGCAACAAT  231
58  I Y R R Y A G L Y F C I C V D V N D N N  77
232 CTGGCTATCTTGAGCCATCCAGCACTTCTGAGAAGTTTAAATGAATCTCCACAAT  291
78  L A Y L E A I H N F V E V L N E Y F H N  97
292 GTCTGTGAAGTGCAGCTGGTTCAGCTTCTACAGGTTTACAGCGTGTGATAGATG  351
98  V C E L D L V F R F Y K V Y T V V D E N  117
352 TTCCTGGCAGCAGATCCGAGACAGCAGCAGCAAGCTGCTGAGGAGCTGCTGATG  411
118 F L A G E I R E T S Q T K V L K Q L L M  137
412 CTGCACTGCGTGGATGAGCCGGCCCTCCGCCAGCCGGCCCTCTGGGCCAACCTGCT  471
138 L Q S L E  142
472 CCTGCTGCGTGGAGGCTCGAGGGCAGCCGACAGCCGCTGCTGCTCACTGCGCC  531
532 CAGAGCACTGACAGCAGCAGCAGCAGCAGCAGCTTACGGCCACAGGGAGGAGGAG  591
592 AACCCACTGCTGCGCCCTAGCTGATGGAGGAGGAGCCAGCTGATGCTCCGAGTAAT  651
652 CTGGATTATCTGTGATGCTGCACTTGGCTGCTGCGAGCCCAATAAACCTGTGCT  711
712 CCTGCTCGCAAAAAAATAAAAAA  736
    
```

FIG. 2. Complete nucleotide sequence and deduced primary structure of rat brain AP17. A, schematic representation of the cDNA clones isolated from λ gt10 (λ) and plasmid (*p*) rat cDNA brain libraries. The horizontal arrows indicate the sequenced DNA strands. The coding region (*heavy line*), the polyadenylation signal (*arrow*), and poly(A) stretches are indicated. B, the nucleotide sequence corresponds to the overlap of the longest cDNA clone pSF1112 with other shorter cDNA clones isolated from the search of the rat brain plasmid cDNA library. The sequence of the amino terminus of calf brain AP17 determined by automated Edman degradation and the polyadenylation signal are *underlined*.

5.51). The ATG initiation codon also conforms to the consensus sequence for translational starts. The 3'-untranslated region contains a polyadenylation signal AATAAA located 23 nucleotides upstream of the corresponding poly(A)⁺ tail.

The identity of the cDNA clones was established by the perfect agreement between the deduced sequence of the rat brain AP17 and mouse AP19 with the amino-terminal sequences determined by automated Edman degradation of calf brain AP17 (26 amino acids) and AP19 (65 amino acids), respectively. The predicted sizes of the rat AP17 and the mouse AP19 are in close agreement with the apparent molecular weights of 16,000–17,000 and 19,000 obtained for calf brain AP17 and AP19 by SDS-polyacrylamide gel electrophoresis (Keen, 1987; see also Fig. 1). The deduced protein sequences of AP17 and AP19 have no internal repeats nor do they display an unusual distribution of amino acids.

The Protein Sequences of the Mammalian AP17 and AP19 Chains Are Closely Related—Sequence comparison of AP17 and AP19 indicates that they are closely related proteins displaying significant sequence similarities throughout the entire length of their polypeptide chains. The optimal sequence alignment of AP17 and AP19, shown in Fig. 4, demonstrates absence of gaps, identity of 45%, and similarity of 71% allowing for conservative changes. With the exception of the additional 16 amino acids at the carboxyl-terminal end of AP19, most of the sequence differences are uniformly distrib-



B

```

-50  CTGGCTTCTACGGTCCGCGAGGGACCCACCGGGCTGGAGGCTGGAGATGATCGGAT  10
      M M R F  4
.....
11  TCATGCTACTGTTGAGCCGCGAGGAAAACCTCGGTCACAAAAATGCTACTCGGCCACT  70
5   M L L F S R Q Q K L R L L Q K W Y L A T S  24
.....
71  CAGACAAGGAGAGCAGCAAGATGCTCCGAGACCTCATGCAACTGCTGCTGGTACGAAGC  130
25  D E R K K H V R E L H Q V V L A R K F  44
.....
131 CGAAAAATGCGAGGCTCGGAGTGGAGGACCTCAAACCTGCTATAAGAGATACGGCA  190
45  K M C S F L E R R D L K V Y Y K R Y A S  64
191 GTCTCTATTGCTGCTGGCCATCGAAGCGCAGCAAGCAGCTGATGACACTGGAGCTGA  250
65  L Y F C C A I E G Q D N E L I T L E L I  84
251 TCACCGTACGCTAGAGCTTGGCAAGCACTTCCGGAGCTATCTGATGATGACATCA  310
85  H R Y V E L L D K Y F G S V C E L D I I  104
311 TCTTCAACTTGGAAAGCTACTTTATCTGAGCAGTTCCTGAGCTGGTGGGATGCTCC  370
105 F N F E K A Y F I L D E F L H G C D V Q  124
.....
371 AGCACACTCCAAAGAGTGTCTGCAAGGCTTAGCAGCTGACCTGCTGCGAGGAGC  430
125 D T S X K S V L K A I E A D L L Q E E  144
.....
431 AGGATGAGTGGCCGCGACTCTCTGAGGAGATGGCCCTGGATAGCCGCTCAGCAG  490
145 D E S P R S V L E E M G L A  158
.....
491 CCGCGGCGCTCTCGACCGGGGGAAGCAAGCAAGCAGCACTTCGGCGCTCTGCTCT  550
551 CTCAGCTTCCCGACCGCCTTCTTGGAGGACTGGCTGCTTTCCTGGCCAGCTT  610
611 ACCTACTTTAGAAAGCTGTGGCTCCGGCTTCTGTAAGAAATTCCTGCTCCAAATC  670
571 TCCCTTCCCTCAGCTGACTCTCCGAGCGAGCTTTGAGATCTAGAGCGAGGAAA  730
731 AAGTGAAGTGGAGTGGGCTGCTCACTGCTGATTTGCTGCTTCAAGAGGATTC  790
791 TTCAGCCAGGACTGCTCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  850
851 TCAGTTAAAGGAAAAGGTTTCAGGCGCTTCTTCTGCTGCTGCTGCTGCTGCTGCT  910
911 CCAGCTCCCGGATAAGCAGCCAGGCGCTTCTGCTGGAGGAGGCTTTCCTGCTCC  970
971 GCGCCACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  1030
1031 TAAGCTTTTAAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  1090
1091 GCTTCTGGAGCACTGACCTTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  1150
1151 GCAAGGAGCTGGTGGCCCACTGCTGCTTATCTGCTTCTTCTGCTGCTGCTGCTGCT  1210
1211 AAAAAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  1250
    
```

FIG. 3. Complete nucleotide sequence and deduced primary structure of mouse brain AP19. A, schematic representation of the independent cDNA clones isolated from the uniZapXR mouse cDNA brain library and their sequencing strategy. Captions are as in Fig. 1. B, combined nucleotide sequences of the cDNA clones pBT6 and pBT8. pBT6 is the longest cDNA and starts at nucleotide position -15. pBT8 starts at nucleotide position -50 and is identical to pBT6 except for 627 untranslated nucleotides missing between positions 533 and 1159 of pBT6. The polyadenylation signal as well as the amino-terminal sequences of intact calf brain AP19 and of several tryptic fragments are *underlined*.

uted along the polypeptide chains of AP17 and AP19. The striking level of sequence similarity between AP17 and AP19 suggests that their genes probably arose from a gene duplication event.

Identification of Yap17p, the Putative Yeast Homolog of the Small Chains of APs—In an attempt to detect other proteins related to AP17 and AP19, we performed several computer searches of different data banks. A search conducted using FASTP (Lipman and Pearson, 1985) and Word (Deveraux *et al.*, 1984) failed to indicate any significant sequence similarity with other proteins whose sequences have been deposited in the National Biomedical Foundation protein data base. However, a search conducted using TFASTA (Pearson and Lipman, 1988) with the rat AP17 protein sequence against the six reading frames deduced from the DNA data bases GenBank and EMBL identified one previously unknown matching sequence. This DNA sequence is located immediately

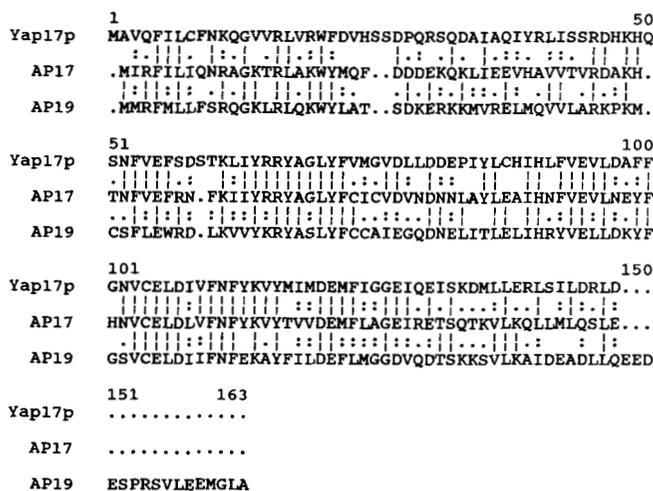


FIG. 4. Sequence alignment of the mammalian small chains AP17 and AP19 with the yeast homolog Yap17p. Several cycles of sequence comparisons were alternatively performed between the sequences of AP17, AP19, and Yap17p using the program GAP (Deveraux *et al.*, 1984). The optimal alignment introduced gaps at the positions indicated by dots. Identities (vertical bars) and conservative replacements (colons and dots) are indicated.

downstream of the yeast *Saccharomyces cerevisiae* thymidylate kinase *CDC8+* gene and is transcribed in the opposite direction (accession number M15468; Rothstein *et al.*, 1987; see also the map in Fig. 5). This published DNA sequence contains an open reading frame of 42 amino acids that aligns with the carboxyl-terminal end of the rat brain AP17. Resequencing of the yeast genomic DNA clone corresponding to this region revealed two missing bases and one additional base in the published sequence (insertion of T residues following G₂₀₆₄, G₂₀₉₀ and removal of C₂₂₅₈ in Fig. 2 of Rothstein *et al.*, 1987). This correction indicated that the open reading frame for the yeast candidate for the small chain continued further downstream of the published sequence. After additional sequencing, we determined the complete yeast nucleotide sequence and the predicted primary structure for the yeast homolog of the mammalian AP17 chains, here referred to as Yap17p (shown in Figs. 4 and 5). The open reading frame of *YAP17* contains 441 bases and encodes a putative protein of 147 amino acids and *M_r* of 17,373 (*pI* = 4.85).

The optimal protein sequence alignment of the rat AP17, mouse AP19, and their putative yeast homolog Yap17p was obtained after several cycles of sequence alignment with the program GAP (Deveraux *et al.*, 1984). The comparison, shown in Fig. 4, indicates that the sequence of Yap17p is slightly more related to AP17 (50% identity or 68% similarity) than to AP19 (39% identity and 65% similarity). In addition, we note that the carboxyl-terminal ends of Yap17p and AP17 are coincident and that both lack the 16 amino acid extension characteristic of the AP19 chain.

DISCUSSION

What is the extent of relationship between the AP-1 and AP-2 complexes of the Golgi apparatus and the plasma membrane? It is known that these complexes, isolated from calf brain-coated vesicles, will bind *in vitro* to clathrin and coassemble to form coats (Zaremba and Keen, 1983; Pearse and Robinson, 1984). The AP complexes are heterotetrameric structures of similar size and of related polypeptide composition (Keen, 1987; Ahle *et al.*, 1988; Virshup and Bennett, 1988; Matsui and Kirchhausen, 1990). Molecular cloning of cDNAs for the two sets of large chains and alignment of the predicted

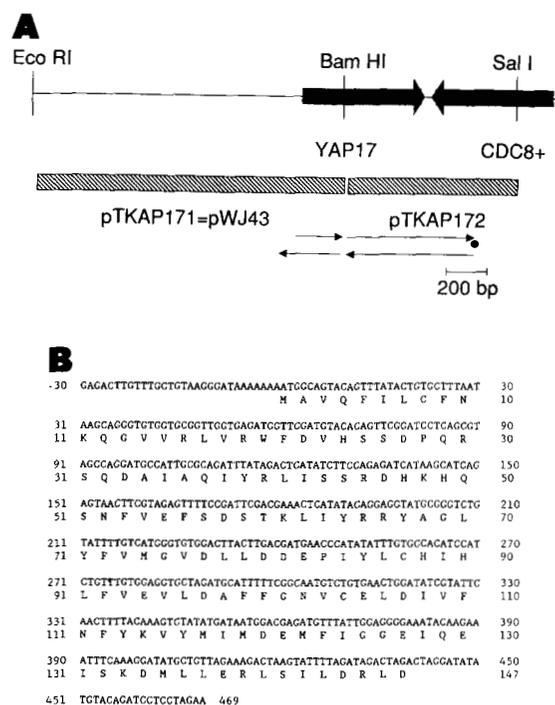


FIG. 5. Nucleotide sequence and deduced primary structure of *YAP17*, the putative yeast homolog of the small chains. **A**, schematic representation of the yeast genome containing the gene for the putative homolog of the small chain, *YAP17* and for *CDC8+*. The orientation of the genes indicate their direction of transcription. The hatched horizontal bars indicate the two yeast genomic clones pTKAP171 and pTKAP172 used for sequencing. pTKAP171 contains the ~1800-bp *BamHI*-*EcoRI* insert described as pWJ43 in Rothstein *et al.*, (1987) and includes the 40 amino-terminal amino acids of the yeast homolog of the small chains. pTKAP172 is the clone containing the ~700-bp *BamHI*-*SalI* insert at the 3' end of the already sequenced SUP4 region (Fig. 1; Rothstein *et al.*, 1987). It includes part of the gene for *CDC8+* and the 107 carboxyl-terminal amino acids of the putative yeast small chain. The horizontal arrows indicate the sequenced DNA strands. **B**, nucleotide sequence and primary structure of the putative yeast homolog of the small chains, Yap17p. The new DNA sequence data includes nucleotides -30 through 119.

protein sequences show that the large chains β' (in AP-1) and β (in AP-2) are closely related to each other (formerly denoted AP105a and AP105b; Kirchhausen *et al.*, 1989), while the large chains γ (in AP-1) and α (in AP-2) are only weakly related to each other (Robinson, 1989; Robinson 1990). These results, taken together with the extensive sequence identity of the small chains described here and of the medium AP47 and AP50 chains,² lead us to conclude that AP-1 and AP-2 complexes must have equivalent structures and similar functions.

The striking similarity between the predicted amino acid sequences of the mammalian AP17 and AP19 chains and the putative Yap17p yeast protein uncovered here is consistent with the possibility that yeast cells also contain AP complexes. Although direct evidence for the existence of AP complexes in yeast is still lacking (Mueller and Branton, 1984; Lemmon *et al.*, 1988), additional support for this hypothesis is provided by the recent identification in the yeast genome of *YAP54*² and *YAP80* (Kirchhausen, 1990) as the putative homologs of the mammalian AP47/AP50 medium chains and of the β/β' large chains of AP complexes, respectively.

A prominent defect associated with the deletion of the clathrin heavy chain in yeast is mislocalization of the endopeptidase Kex 2p from internal vesicles to the cell surface and consequent failure to process the α -mating factor precu-

sor (Payne and Schekman, 1989). This phenotype led to the suggestion that at least one principal cellular site for clathrin/AP coat formation in yeast is in the Golgi apparatus (Payne and Schekman, 1989). It appears, however, that Yap17p is marginally more similar to the mammalian AP17 of the AP-2 complex (plasma membrane) than it is to the related mammalian AP19 of the AP-1 complex (Golgi apparatus) (50 versus 39% identity, respectively; see Fig. 4). Therefore it is possible that Yap17p may be the small chain component of an AP-2-like complex associated with endocytic coated structures of the yeast plasma membrane. Alternatively, clathrin may participate only in the yeast exocytic pathway, thus requiring only AP-1 type complexes. In this view, the gene duplication event that led to the appearance of the mammalian AP17 and AP19 chains may have occurred only after yeast and mammals diverged from a common ancestor.

Acknowledgments—We thank C. Hsu, G. Jung, W. Matsui, and Y. Nakayama for the purification of AP19 chains, M. Brownstein (NIMH) for the rat brain plasmid cDNA library, R. Rothstein (Columbia University) for providing the yeast genomic clones, A. C. Davis (Harvard Medical School) and E. Pingchang Chow (Biogen Inc.) for making available their microsequencing facilities, K. Tucker for help sequencing the yeast DNA, J. Brosius for useful discussions, and L. Scott for secretarial assistance.

REFERENCES

- Ahle, S., Mann, A., Eichelsbacher, U., and Ungewickell, E. (1988) *EMBO J.* **7**, 919–929
- Brodsky, F. M. (1988) *Science* **242**, 1396–1402
- Deveraux, J., Haerberli, P., and Smithies, O. (1984) *Nucleic Acid Res.* **12**, 387–395
- Glickman, J. N., Conibear, E., and Pearse, B. M. F. (1989) *EMBO J.* **8**, 1041–1047
- Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., and Schneider, W. J. (1985) *Annu. Rev. Cell Biol.* **1**, 1–39
- Heuser, J., and Kirchhausen, T. (1985) *J. Ultrastruct. Res.* **92**, 2457–2460
- Keen, J. H. (1987) *J. Cell Biol.* **105**, 1989–1998
- Keen, J. H., Willingham, M. C., and Pastan, I. (1979) *Cell* **16**, 303–312
- Kirchhausen T. (1990) *Mol. Cell. Biol.* **10**, 6089–6090
- Kirchhausen, T., and Harrison, S. C. (1984) *J. Cell Biol.* **99**, 1725–1734
- Kirchhausen, T., Nathanson, K. L., Matsui, W., Vaisberg, A., Chow, E. P., Burne, C., Keen, J. H., and Davis, A. E. (1989) *Proc. Natl. Acad. Sci. (U. S. A.)* **86**, 2612–2616
- Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241
- Laemmli, U. K. (1970) *Nature* **227**, 680–685
- Lemmon, S. K., Lemmon, V. P., and Jones, E. W. (1988) *J. Cell Biochem.* **36**, 329–340
- Lipman, D. J., and Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 11035–11038
- Matsui, W., and Kirchhausen, T. (1990) *Biochemistry* **29**, 10791–10798
- Mocchetti, I., Einstein, R., and Brosius, J. (1986) *Proc. Natl. Acad. Sci. (U. S. A.)* **83**, 7221–7225
- Mueller, S. C., and Branton, D. (1984) *J. Cell Biol.* **98**, 341–346
- Payne, G., and Schekman, R. (1985) *Science* **230**, 1009–1014
- Payne, G., and Schekman, R. (1989) *Science* **245**, 1358–1365
- Pearse, B. M. F. (1988) *EMBO J.* **7**, 3331–3336
- Pearse, B. M. F., and Robinson, M. S. (1984) *EMBO J.* **3**, 1951–1957
- Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2445–2448
- Pfeffer, S. R., and Rothman, J. E. (1987) *Annu. Rev. Biochem.* **56**, 829–852
- Robinson, M. S., and Pearse, B. M. F. (1986) *J. Cell Biol.* **102**, 48–54
- Rothstein, R., Helms, C., and Rosenberg, N. (1987) *Mol. Cell Biol.* **7**, 1198–1207
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
- Virshup, D., and Bennett, V. (1988) *J. Cell Biol.* **106**, 39–50
- Vigers, G. P. A., Crowther, R. A., and Pearse, B. M. F. (1986) *EMBO J.* **5**, 2079–2085
- Zaremba, S., and Keen, J. H. (1983) *J. Cell Biol.* **97**, 1339–1347