Eur. J. Biochem. 202, 569–574 (1991)
© FEBS 1991

The medium chains of the mammalian clathrin-associated proteins have a homolog in yeast

Yasuhiro NAKAYAMA [1], Mark GOEBL [2], Betsy O'BRINE GRECO [3], Sandy LEMMON [4], Elizabeth PINGCHANG CHOW [3] and Tomas KIRCHHAUSEN [1]

[1] Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, USA
[2] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, USA
[3] Biogen Inc., Cambridge, USA
[4] Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, USA

(Received June 18/August 13, 1991) – EJB 91 0791

We have cloned and sequenced mouse brain AP47, the medium chain of the trans-Golgi network clathrin-associated protein complex AP-1. The predicted protein sequence of AP47 is closely related to rat and calf brain AP50, the corresponding medium chain of the plasma-membrane clathrin-associated protein complex AP-2. We have also identified in the yeast genome an open reading frame encoding a protein of previously unknown function. Referred to here as *YAP54*, its predicted protein sequence displays a striking homology to AP47. We therefore propose that Yap54 is the medium chain subunit of a putative AP-1 complex in yeast. From the analyses of the optimized sequence alignments of AP47, AP50 and Yap54p, we suggest a model for the domain organization of the medium chains.

AP-1 and AP-2 are the two main classes of clathrin-associated protein complexes [AP; also referred to as assembly proteins or adaptors (Ahle et al., 1988; Robinson and Pearse, 1986; Zaremba and Keen, 1983; Pearse, 1988)]. They are found in clathrin-coated pits and vesicles of the trans-Golgi network (AP-1) and at the plasma membrane (AP-2) of mammalian cells (Ahle et al., 1988; Robinson and Pearse, 1986). These complexes are heterotetrameric structures composed of two large chains of ≈ 100 kDa (γ and β' in AP-1; α and β in AP-2), one medium chain of ≈ 50 kDa (AP47 in AP-1 or AP50 in AP-2) and one small chain of ≈ 17 kDa (AP19 in AP-1 or AP17 in AP-2) (see schematic representation in Fig. 4; Ahle et al., 1988; Virshup and Bennett, 1988; Matsui and Kirchhausen, 1990). Although the function of these complexes is still not well understood, they are believed to promote the assembly of clathrin coats and to direct the sorting of membrane proteins transported by coated vesicles (Zaremba and Keen, 1983; Pearse, 1988; Glickman, 1989; Kirchhausen et al., 1989).

Recent *in vitro* experiments suggest that the medium chains may be involved in a regulatory switch of AP activity. A number of studies have shown that the mammalian AP50 self phosphorylates and that it is a good substrate for phosphorylation by a serine/threonine kinase found in coated vesicles (Pauloin et al., 1982; Campbell et al., 1984; Manfredi and Barzari, 1987; Keen et al., 1987). Furthermore, AP47 may be a serine kinase capable of phosphorylating the cytoplasmic tail of the mannose 6-phosphate receptor (Meresse et al., 1990).

Although yeast cells also contain clathrin-coated vesicles (Mueller and Branton, 1984; Payne and Schekman, 1985; Lemmon et al., 1988), the existence of AP in yeast has not been documented directly. However, their existence has been postulated based on the recent identification of putative homologs of the mammalian β/β' large chains (Kirchhausen, 1990) and AP17/AP19 small chains (Kirchhausen et al., 1991) by computer searches of the published yeast genome DNA data base.

Previously we cloned a cDNA for the rat brain AP50 subunit of AP-2 and found that the predicted sequence was essentially identical to the partially sequenced bovine brain AP50 (Thurieau et al., 1988). This remarkable sequence conservation among mammalian species has also been observed with the large chains (Kirchhausen et al., 1989; Ponnambalam et al., 1990; Robinson, 1989; Tucker et al., 1990) and small chains (Kirchhausen et al., 1991) of the AP complexes. We have assumed that this theme of sequence conservation also holds for the medium AP47 chain and have therefore used calf brain AP50 to obtain protein sequence information suitable for oligonucleotide synthesis and cDNA library screens. We now report the cloning and sequencing of a mouse brain AP47 cDNA which encodes a protein closely related to the mammalian AP50. Neither protein has significant sequence similarity with known kinases. We also report a computer search that has uncovered from the yeast genome a previously published sequence (Daignan-Fornier and Bolotin-Fukuhara, 1989), here referred to as *YAP54*, that encodes Yap54p, a protein displaying very extensive protein sequence identity with the mammalian AP47 and AP50 chains. Analysis of the

Correspondence to T. Kirchhausen, Harvard Medical School, Department of Anatomy and Cellular Biology, 220 Longwood Ave, Boston, MA 02115, USA

Abbreviations. AP, clathrin-associated protein complex; AP-1 and AP-2, AP on clathrin-coated structures from the trans-golgi network and plasma membrane, respectively; AP47 and AP50, medium chains from AP-1 and AP-2, respectively.

Note. The novel nucleotide sequence data published here have been deposited with the EMBL/GeneBank sequence data bank and are available under accession number M 62419 for AP47 and X60288 for YAP54.

```
                                                                              <-start Region I
  -72 AGCTTGCTGAACGCCCGCAGTTCCTGCCGTTGCCGCCGCCCTCGGCTACTGCCGAGACCTTCCAAAGCCGCCATGTCCGCCAGCGCCGTCTACGTACTG   27
                                                                              M  S  A  S  A  V  Y  V  L     9
   28 GATCTGAAGGGCAAGGTGCTCATCTGCAGGAACTACCGTGGGGATGTGGACATGTCAGAGGTGGAGCACTTCATGCCCATTCTGATGGAGAAGGAGGAG  126
   10 D  L  K  G  K  V  L  I  C  R  N  Y  R  G  D  V  D  M  S  E  V  E  H  F  M  P  I  L  M  E  K  E  E    42
                                                                           ---------------------- 47.3.1
  127 GAGGGCATGCTGTCACCTATCTTGGCCCATGGTGGCGTTCGTTTCATGTGGATTAAGCACAACAACCTGTACCTGGTCGCCACTTCAAAAAAGAATGCT  225
   43 E  G  M  L  S  P  I  L  A  H  G  G  V  R  F  M  W  I  K  H  N  N  L  Y  L  V  A  T  S  K  K  N  A    75

  226 TGTGTGTCACTGGTGTTCTCCTTCCTCTACAAGGTGGTACAGGTCTTCTCCGAGTACTTTAAGGAGTTGGAGGAGGAGAGCATCCGAGACAACTTTGTC  324
   76 C  V  S  L  V  F  S  F  L  Y  K  V  V  Q  V  F  S  E  Y  F  K  E  L  E  E  E  S  I  R  D  N  F  V   108

  325 ATCATCTACGAGCTGCTAGATGAGCTCATGGACTTTGGCTACCCGCAGACCACTGACAGCAAGATCTTGCAGGAGTACATCACTCAGGAAGGCCACAAG  423
  109 I  I  Y  E  L  L  D  E  L  M  D  F  G  Y  P  Q  T  T  D  S  K  I  L  Q  E  Y  I  T  Q  E  G  H  K   141
                                                                     -------------------------X---------
                                                                     X------------------------------X---
  424 CTGGAAACGGGGGCCCCTAGGCCCCCAGCCACAGTCACCAATGCTGTGTCCTGGCGTTCAGAAGGCATCAAGTATCGGAAGAATGAAGTATTCCTGGAT  522
  142 L  E  T  G  A  P  R  P  P  A  T  V  T  N  A  V  S  W  R  S  E  G  I  K  Y  R  K  N  E  V  F  L  D   174
      --------------------------------------------- Ta84.85
      ------------------------------X-----X--------- 47.3.2
      ------------------------------------------------------ Ta72.73
  523 GTCATTGAGGCTGTTAACCTCTTGGTCAGTGCCAATGGCAACGTGCTGCGCAGTGAGATTGTGGGTTCCATCAAGATGCGGGTCTTCCTCTCAGGCATG  621
  175 V  I  E  A  V  N  L  L  V  S  A  N  G  N  V  L  R  S  E  I  V  G  S  I  K  M  R  V  F  L  S  G  M   207
                             Region I end->                                <-start Region II
  622 CCTGAGTTACGCCTGGGTCTCAATGACAAGGTCCTCTTCGACAACACAGGCCGAGGGAAGAGCAAGTCAGTGGAGCTGGAGGATGTGAAATTCCACCAG  720
  208 P  E  L  R  L  G  L  N  D  K  V  L  F  D  N  T  G  R  G  K  S  K  S  V  E  L  E  D  V  K  F  H  Q   240

  721 TGTGTGCGGCTGTCACGTTTTGAGAACGACCGCACTATCTCCTTCATCCCACCCGACGGAGAGTTTGAACTCATGTCCTACCGCCTCAACACCCATGTG  819
  241 C  V  R  L  S  R  F  E  N  D  R  T  I  S  F  I  P  P  D  G  E  F  E  L  M  S  Y  R  L  N  T  H  V   273
                                                                                          ---------X---
  820 AAGCCTTTGATCTGGATTGAGTCCGTGATTGAGAAGCATTCCCACAGCCGCATTGAGTACATGGTCAAGGCCAAGAGCCAGTTCAAGAGGCGGTCAACA  918
  274 K  P  L  I  W  I  E  S  V  I  E  K  H  S  H  S  R  I  E  Y  M  V  K  A  K  S  Q  F  K  R  R  S  T   306
      ------------X-----X--X--------- 47.2.5                                                    ----
  919 GCCAACAATGTAGAGATCCATATACCAGTCCCCAACGATGCTGATTCACCCAAGTTCAAGACTACAGTGGGGAGTGTCAAGTGGGTCCCTGAAAACAGT 1017
  307 A  N  N  V  E  I  H  I  P  V  P  N  D  A  D  S  P  K  F  K  T  T  V  G  S  V  K  W  V  P  E  N  S   339
      ------X--------------------------------X-----X------ Tb72.73                    ----------------
 1018 GAGATCGTGTGGTCCGTCAAGTCCTTTCCGGGTGGCAAGGAGTACCTGATGCGGGCCCACTTTGGCCTTCCCAGTGTGGAAGCTGAAGACAAGGAGGGA 1116
  340 E  I  V  W  S  V  K  S  F  P  G  G  K  E  Y  L  M  R  A  H  F  G  L  P  S  V  E  A  E  D  K  E  G   372
      ---------------I--- Tb84.85                               ---X------------------------------------
 1117 AAGCCCCCCATCAGCGTCAAGTTTGAGATCCCCTACTTCACTACCTCTGGCATCCAGGTGCGCTACCTGAAAATCATTGAGAAGAGTGGGTACCAGGCC 1215
  373 K  P  P  I  S  V  K  F  E  I  P  Y  F  T  T  S  G  I  Q  V  R  Y  L  K  I  I  E  K  S  G  Y  Q  A   405
      ------------------- 47.3.1                      --------------------------------------------------
                           ---------------------X------------------ 47.3.2                -------------
                                           Region II end->
 1216 CTGCCCTGGGTACGATATATCACACAGAACGGAGATTATCAGCTCCGGACCCAGTGAGAGGCCTCTGCCACCAGCCCAGCCCACCTAGCCTCAGGGACA 1314
  406 L  P  W  V  R  Y  I  T  Q  N  G  D  Y  Q  L  R  T  Q                                                 423
      --------------------------------------------- T121.122
      ------------- T77
 1315 CCTGCACACTCACCAAAACTGAAGCTAGAGGGTGGCCCTGGACATGCAGCCACCCTCTCCTCAAGCCTGAGTGGACTCACACAGAACCCCTTCCCCGGT 1413
 1414 CCCATTCTGATCCGAGGGTGGGAAGGAAGGGCCTGCCAGCCTCCCCCAGGGACAAGCCAGCTGGAGCTGTGCTCCTGGTGCCTCATGACTGGCCACCCA 1512
 1513 CCACCCTGTAGACGCCATCCCTGCCCACCTCCGAAGCCTCCTTCTGTTGCCATTTTGCTGAGCATAGTTGGCTCATTGTCATTGTCATCTGTGGCTGTC 1611
 1612 CTGTCACCTTTCTTCAAGTGTCCTGTGCAGCTGCCATGCTGCACTTAATGAGGGCAGCTGTCCCCTATCCATGCTGGTCTGTATGCCGGATAGTTGCTC 1710
 1711 TGCCTGGGCCGGCTGTGCCTCCCTTGCCAGCCCTTGACCACATCAGTGTTTTCTCAGAAGGGCACATGGCCTCAGCCTTTGCCCTAAAATTCCTGGGCA 1809
 1810 GGCACAGGCTACCCTCACTGTACAAGGCCTCGTTGTCCTGGACCCATGTGTGTGTGTGACAGTTATATTAAAATCCATTTGTTACAAAAAAAAAAAAAA 1908
 1909 AAAAA  1913
```

Fig. 1. *Complete nucleotide and deduced protein sequence of the mouse brain cDNA for AP47.* The nucleotide sequence corresponds to the longest cDNA clone (*pYN101*) and is otherwise identical to three other shorter and overlapping cDNA clones. The underlined sequences indicate the identity between the N-terminal sequences determined from selected tryptic fragments of calf brain AP47 and the corresponding regions of the mouse cDNA; X denotes amino acids whose identity by Edman degradation was ambiguous. The putative polyadenylation signal ATTAAA is underlined. Regions I and II define the two most conserved portions among the medium chains (see sequence alignment in Fig. 3)

alignment of AP47, AP50 and *YAP54* sequences lead us to identify portions of the proteins that may be specific for the function of AP47 and AP50 and to propose a hypothesis for a possible domain organization of the medium chains in the AP complexes.

MATERIALS AND METHODS

Purification of calf brain AP47

Samples of AP complexes were obtained from depolymerized coats of calf-brain coated vesicles by sizing chromatography as described previously (Matsui and Kirchhausen, 1990). AP-1 complexes were separated from AP-2 complexes using Sepharose-Q (Ahle et al., 1988; Matsui and Kirchhausen, 1990) followed by hydroxylapatite ionic-exchange chromatography (Pearse and Robinson, 1984). The samples containing AP-1 complexes were pooled and its polypeptide chains fractionated by preparative 6% or 12% SDS/PAGE (Laemmli, 1970). The band corresponding to calf brain AP47 was obtained from each type of gel by either overnight diffusion at 37 °C (into 10 vols 20 mM Tris, 1 mM $CaCl_2$, 0.01% SDS, pH 7.9) or electroelution at 25 °C (into 50 mM Tris, 384 mM glycine, 0.01% SDS, pH 8.3) and concentrated by centrifugation in a Centricon-30 filter (Amicon). Internal amino acid sequence information for the purified calf brain AP47 was obtained from the automated Edman degradation of fragments generated by tryptic digestion and HPLC frac-

tionation (Huang et al., 1986). N-terminal sequencing of the intact AP47 was performed after preparative 6% SDS/PAGE (Laemmli, 1970) and electrotransfer onto a poly(vinyldifluoride) membrane (Immobilon, Milipore) (Matsudaira, 1987; Kirchhausen et al., 1989; Kirchhausen et al., 1991).

cDNA isolation, DNA sequencing and DNA blots

Based on the sequences of the tryptic fragments T121.122, 47.3.1 and Tb84.85, three oligonucleotide probes of degenerate sequence 3′-TTI AGI CCI ATG/A GTC/T CGI GAI GGI ACC CAI G/TCI ATG/A TAI TGI GTC/T TTG/A CCI CTG/A ATG/A GTC/T GAI, 5′-AAG/A TGG GGI AAG/A CCI CCI ATI T/AC/GI GTI AA and 5′-TGG GTI CCI GAG/A AAC/T T/AC/GI GAG/A ATI GTI TGG T/AC/GI ATI AA were prepared and used at low stringency conditions to screen a mouse brain cDNA library inserted in the uniZapXR vector (Stratagene). The probes were labeled with [γ-^{32}P]ATP and T4 kinase and used for overnight hybridization at 37°C in 6 × NaCl/Cit (3 M NaCl, 0.3 M NaCit), 0.5% SDS, 25 µg/ml boiled tRNA, 2 × Denhardt's solution (0.24 × NaCl/Cit, 0.02% SDS, 0.4 g/l Ficoll, 0.4 g/l polyvinylpyrrolidone, 0.4 g/l serumalbumin). Nitrocellulose filters were washed with 2 × NaCl/Cit, 0.5% SDS at 42°C. Limited sequencing of two hybridizing uniZapXR phage clones (pYN611 and pYN614), converted to plasmid cDNA by rescue excision according to the instructions of the manufacturer, indicated that they were partial length cDNA. The full length clone pYN101 was obtained from an additional screen of the same mouse cDNA library using as a probe the anti-sense oligonucleotide 5′-AGT GAT GTA CTC CTG CAA GAT CTT GCT GTC AGT GG starting at nucleotide position 408. The complete sequence of pYN101 was determined in both strands by the chain termination method using synthetic 17-mer primers (Sanger et al., 1977). Genomic mouse DNA blots were carried out as previously described (Reed and Mann, 1985) using a cDNA AP47 probe spanning nucleotide positions 535–1913.

RESULTS AND DISCUSSION

Molecular cloning and the sequence of the mouse AP47 medium chain

Since the N terminus of AP47 purified from AP-1 complexes of calf-brain clathrin-coated vesicles appears to be blocked, we obtained internal amino acid sequence information from chemical microsequencing of several tryptic fragments and used T/21.122, 47.3.1 and Tb84.85 (Fig. 1) to design the oligonucleotide probes described in Materials and Methods. Using these probes, we screened a mouse brain cDNA library, and identified three hybridizing independent and overlapping clones corresponding to AP47. The sequence of the longest cDNA clone pYN101, shown in Fig. 1, contains 1985 nucleotides and predicts a protein of 423 amino acids (M_r 48543; pI 7.32). The first ATG translational start codon at position 1 of the nucleotide sequence partially conforms to the consensus sequence for eukaryotic translational start sites GCCATGA/G (Kozak, 1989). Use of a more 5′-proximal ATG start codon is unlikely given that upstream of the suggested start codon there is no significant protein sequence relationship with the related mammalian AP50 chain nor with the putative yeast Yap54p homolog (see below). The 3′-untranslated region (625 nucleotides) contains a possible polyadenylation signal (ATTAAA) at nucleotide position 1876 followed by a poly(A) tail.

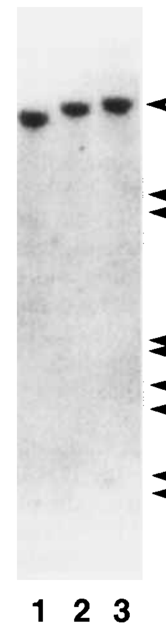

Fig. 2. *Southern blot analysis of DNA.* Mouse genomic DNA was digested with *Eco*RI (lane 1), *Eco*RV (lane 2) and *Hin*dIII (lane 3) and probed with a mouse-brain AP47 cDNA fragment from nucleotide positions 535–1913 spanning part of the coding and the complete 3′ untranslated regions. Size markers (arrowheads) were M_r: 21226, 5148, 4268, 2027, 1904, 1584, 1375, 947 and 831

The identity of the three hybridizing clones as mouse brain AP47 cDNA clones was established by the perfect match between their predicted protein sequence and all the non-ambiguous amino acid sequence determinations obtained from the tryptic fragments of calf brain AP47 (141 determinations out of 142 amino acids compared). Since these perfect matches correspond to a random sampling of about 30% along the open reading frame of the mouse AP47 chain, we suggest that the entire primary structures of the mouse and calf AP47 chains are highly conserved. In fact, in mammalian species, the same pattern of strong sequence conservation has also been observed among the other major protein components of clathrin-coated vesicles including all the other chains of the AP-1 and AP-2 complexes (Fig. 4; Kirchhausen et al., 1989; Kirchhausen et al., 1991; Ponnambalam et al., 1990; Robinson, 1989; Robinson, 1990; Thurieau et al., 1988) as well as clathrin heavy and light chains (Kirchhausen et al., 1987a; Kirchhausen et al., 1987b).

The predicted size deduced from the mouse AP47 cDNA (M_r 48543) is slightly larger than the apparent molecular mass (47000) obtained for calf brain AP47 by SDS/PAGE (Ahle et al., 1988). Since the N terminus of the calf AP47 chain is blocked, we cannot eliminate an anomalous electrophoretic behavior nor a post-translational cleavage at the N-terminal end as a possible explanation for this apparent size reduction. However, we can rule out an extensive cleavage at the C-terminal end of the calf brain AP47, since the N-terminal sequence for tryptic fragment T121.122 extends to Arg421 located within two amino acids of the C terminus predicted for the mouse brain AP47 chain (Fig. 1).

The DNA blot analysis of the mouse genomic DNA, performed with an AP47 cDNA probe including part of the coding region and the complete 3′-untranslated region, is consistent with the presence of one gene and the absence of pseudo genes coding for AP47 (Fig. 2). None of the bands correspond to a cross hybridization with the related and unique rat AP50

gene (Thurieau et al., 1988) since a different pattern was obtained in the mouse genomic DNA blot probed with the coding region of rat brain AP50 (data not shown).

Recently it was reported that calf AP47 binds GTP and phosphorylates the cytoplasmic tail of the mannose 6-phosphate receptor (Meresse et al., 1990). To determine if AP47 is a member of the family GTP-binding proteins with kinase activity, we performed several computer searches using FASTA and TFASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). We could not detect significant alignments with known GTP-binding proteins, kinases or indeed with any other proteins, (except mammalian AP50 and yeast Yap54p; see below) whose sequence has been deposited in the protein data bank (National Biomedical Research Foundation, version 25). We also could not find alignments with any open reading frame contained in the DNA data bank (GenBank, version 66). Moreover, inspection of the predicted protein sequence of mouse AP47 did not reveal the characteristic set of three consensus sequence elements GXXXXGK, DXXG and NKXD normally found in the GTP-binding domain of proteins having a high affinity for GTP (reviewed in Dever et al., 1987), nor did it show the consensus sequence GXGXXG followed by a downstream lysine 15–20 residues away that can be found in the nucleotide binding domain of other proteins (reviewed by Hunter, 1987). These results are consistent with our analysis of the mammalian AP50 (Thurieau et al., 1988), which also lacks detectable sequence similarity with known kinases and nucleotide binding proteins, although it is considered to have autokinase activity (Campbell et al., 1984; Keen et al., 1987).

Identification in the yeast genome of Yap54p, a putative homolog of the mammalian AP47 and AP50 medium chains of the AP-1 and AP-2 complexes

Recently, Daignan-Fornier and Bolotin-Fukuhara have reported the sequence of a yeast genomic DNA clone which has not yet appeared in the DNA data bank with three open reading frames (A, B, C) but only frame C encoded a protein of known function (Daignan-Fornier and Bolotin-Fukuhara, 1989). Using the program T FASTA (Deveraux et al., 1984; Pearson and Lipman, 1988) to compare the predicted protein sequences of the open reading frames A and B against a limited protein data base compiled by one of us (M. G.), it was noticed that the open reading frame A is a yeast homolog of the mammalian AP47 and AP50 chains. The deduced protein, referred to here as Yap54p contains 475 amino acids and has a predicted M_r 53865 (pI 9.43). Its expression in yeast cells is likely, since a mRNA of the expected size is transcribed (Daignan-Fornier and Bolotin-Fukuhara, 1989). Using the programs GAP or BESTFIT (Deveraux et al., 1984), we found that the optimal alignment of Yap54p and mouse AP47 introduces eight relatively small gaps and results in 56% sequence identity and 75% sequence similarity allowing for conservative changes (Fig. 3). The alignment of Yap54p with rat AP50 (M_r 49612; pI 10.35) (Thurieau et al., 1988) introduces eight gaps and shows 39% sequence identity and 59% sequence similarity. This level of relationship is similar to the relatedness between the mammalian AP47 and AP50 chains, whose optimal alignment requires five gaps for 40% sequence identity and 64% sequence similarity.

Based on this striking pattern of sequence similarities, we suggest that Yap54p is the yeast homolog of mammalian AP47. Direct biochemical confirmation that Yap54p is a constituent of yeast-coated vesicles is not yet available. However, additional identification in the yeast genome of the open reading frames encoding for Yap80p and Yap17p as homologs of the mammalian β/β' large chains (Kirchhausen, 1990) and of the AP17/AP19 small chains (Kirchhausen et al., 1991), respectively, provides further support for the presence of AP complexes in yeast. The alignment of the Yap80p and Yap17p with their mammalian counterparts did not allow the establishment of whether these yeast homologs are clearly more related to the AP-1 or AP-2 components. In contrast the homology of Yap54p is much closer to the AP47 subunit, which suggests yeast cells may contain a clathrin-associated AP-1 complex involved in vesicular traffic from the trans-Golgi network. It is interesting to note that although the complete role for clathrin in yeast has not been elucidated, recent studies indicate that clathrin plays a key role in the selective retention of membrane proteins in an intracellular trans-Golgi compartment (Payne and Schekman, 1989). What are the chances that yeast cells contain more than one type of AP complex? Assuming that the proposed yeast AP complex also contains four polypeptide chains, that yeast cells contain 5000 distinct proteins and that only about 10% of the yeast genome is currently available for sequence searches in different data bases, then the probability of randomly identifying any three out four chains from an AP complex is about $(1/10)(1/10)(1/10)(9/10)(4) = \approx 1/300$. Against these odds, it is remarkable that three homologies were found. This calculation is therefore consistent with the possibility that yeast cells, like mammalian cells, also contain several types of AP complexes so that one might expect to find in the future other gene versions of *YAP17, YAP54, YAP80* as well as the missing homologs of the mammalian α and γ large chains.

Sequence analysis of the medium chains; possible domain organization

The high level of sequence relationship among Yap54p, AP47 and AP50 allows us to detect those regions that may contain sequence information unique to each class of chains. Based on the optimal alignment of their sequences (Fig. 3), we can divide the primary structures into two regions of similar length but with distinct characteristics. Region I, located at the N terminus, spans about 230 amino acids that are mostly conserved among the three chains. Region I of each chain is separated from the C-terminal region II by different linkers of 10–42 amino acids in length. Region II spans about 190 amino acids and in contrast to region I, contains several sequence stretches specific to AP50 interspersed with other strongly conserved sections common to all three chains. One prediction from this analysis is that the linker between regions I and II may declineate the boundary between two domains in the medium chains. Indeed, limited tryptic proteolysis of AP-2 complexes bound to clathrin coats results in the cleavage and degradation of the C terminus of AP50 while the N terminus, of about M_r 24000, remains intact (Matsui and Kirchhausen, 1990). A second prediction is that the interactions determining the specificity of association of the medium chains with the other subunits of AP-1 or AP-2 complexes may principally be located in the more constant region I. The more variable region II could participate in interactions with other proteins to be found in specific association with coated structures at either the plasma membrane or the trans-Golgi network.

Use of genomic data bases

The yeast DNA sequences deposited in the available data bases correspond mostly to specific genes and their flanking

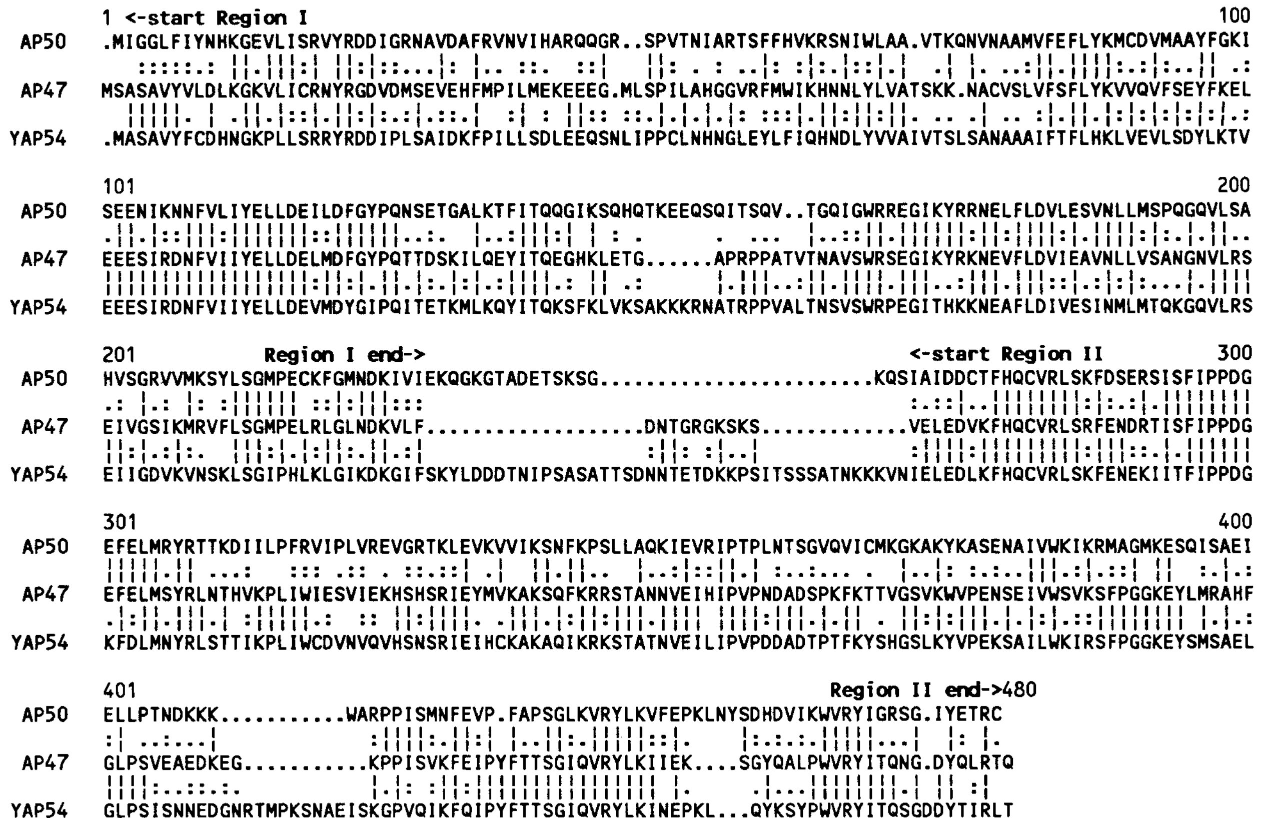

Fig. 3. *Alignment of the predicted sequences of the mouse brain AP47 (this paper) and rat brain AP50 (Thurieau et al., 1988) chains and the yeast Yap54p (this paper) homolog.* The alignment was optimized using the program GAP (Deveraux et al., 1984). Regions I and II denote two conserved portions separated by linkers of variable length. The sequence homology of region I does not show preference for any type of chain. In contrast, the primary structure of region II contains some portions that are unique to AP47 and Yap54p or to AP50. Region II may be involved in interactions specific to AP-1 and AP-2 complexes. Vertical bars represent sequence identities. Semicolons and full stops represent conservative replacements with high (0.50) and low (0.10) thresholds, respectively (Deveraux et al., 1984)

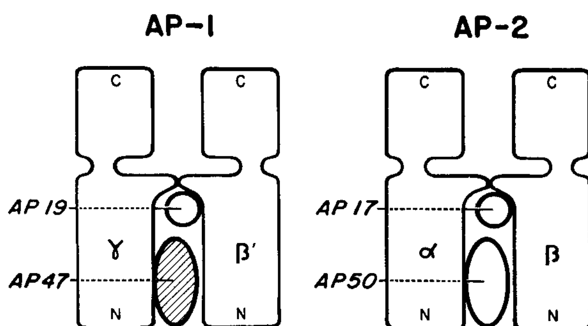

Fig. 4. *Schematic representation of the AP complexes.* AP-1 and AP-2 complexes are found in clathrin-coated structures of the trans-Golgi network and the plasma membrane, respectively. Each type of complex contains two large chains, one medium chain and one small chain encoded by different genes. The model is based on the proposed quaternary structure of AP-1 an dAP-2 complexes (Ahle et al., 1988; Matsui and Kirchhausen, 1990; Virshum and Bennett, 1988) and on the high level of sequence identity among the primary structures of AP17 and AP19 (Kirchhausen et al., 1991), AP47 and AP50 (this paper), β and β' (Kirchhausen et al., 1989) and α and γ (Robinson, 1989; Robinson, 1990)

sequences, and they therefore do not represent a random sampling of the yeast genome. Nevertheless, an interesting consequence of the existence of a yeast DNA data base, albeit limited in length, has been the rapid identification by computer searches of *YAP17, YAP54* and *YAP80*, the genes encoding for Yap17p, Yap54p and Yap80p. From these data we have postulated the existence in yeast of AP complexes and we have also used sequence comparisons with the mammalian counterparts to present models for their structural and functional organization. In addition, we are in the process of studying their functions based on gene disruption techniques. In the future, access to a more complete DNA data base will be an effective shortcut to the identification of genes homologous to ones already known in other organisms.

We thank Dr. M. Warman for help with DNA blots, S. Hamilton for help in sequencing the yeast DNA and L. Scott for secretarial assistance. T. K. acknowledges the financial support from the National Institutes of Health, the American Cancer Society and a Basil O'Connor Award from the March of Dimes. T. K. is an Established Investigator for the American Heart Society.

REFERENCES

Ahle, S., Mann, A., Eichelsbacher, U. & Ungewickell, E. (1988) *EMBO J. 7*, 919–929.

Campbell, C., Squicciarini, J., Shia, M., Pilch, P. F. & Fine, R. E. (1984) *Biochemistry 23*, 4420–4426.

Daignan-Fornier, B. & Bolotin-Fukuhara, M. (1989) *Yeast 5*, 259–269.

Dever, T. E., Glynias, M. J. & Merrick, W. C. (1987) *Proc. Natl Acad. Sci USA 84*, 1814–1818.

Deveraux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acid Res. 12*, 387–395.
Glickman, J. N., Conibear, E. & Pearse, B. M. F. (1989) *EMBO J. 8*, 1041–1047.
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. & Pepinsky, R. B. (1986) *Cell 46*, 191–199.
Hunter, T. (1987) *Cell 50*, 823–829.
Keen, J. H., Chesnut, M. H. & Beck, K. A. (1987) *J. Biol. Chem. 262*, 3864–3871.
Kirchhausen, T. (1990) *Mol. Cell. Biol. 10*, 6089–6090.
Kirchhausen, T., Davis, A. C, Frucht, S., Greco, B., Payne, G. S. & Tubb, B. (1991) *J. Biol. Chem. 266*, 11153–11157.
Kirchhausen, T., Harrison, S. C., Chow, E. P., Mattaliano, R. S., Ramachandran, K. L., Smart, J. & Brosius, J. (1987a) *Proc. Natl Acad. Sci USA 84*, 8805–8809.
Kirchhausen, T., Nathanson, K. L., Matsui, W., Vaisberg, A., Chow, E. P., Burne, C., Keen, J. H. & Davis, A. E. (1989) *Proc. Natl Acad. Sci USA 86*, 2612–2616.
Kirchhausen, T., Scarmato, P., Harrison, S. C., Monroe, J. J., Chow, E. P., Mattaliano, R. J., Ramachandran, K. L., Smart, J. E., Ahn, A. & Brosius, J. (1987b) *Science 236*, 320–324.
Kozak, M. (1989) *J. Cell Biol. 108*, 229–241.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
Lemmon, S. K., Lemmon, V. P. & Jones, E. W. (1988) *J. Cell. Biochem. 36*, 329–340.
Lipman, D. J. & Pearson, W. R. (1985) *Science 227*, 1435–1441.
Manfredi, J. J. & Barzari, W. L. (1987) *J. Biol. Chem. 262*, 12182–12188.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 11035–11038.
Matsui, W. & Kirchhausen, T. (1990) *Biochemistry 29*, 10791–10798.
Meresse, S., Ludwig, T., Frank, R. & Hoflack, B. (1990) *J. Biol. Chem. 31*, 18833–18842.
Mueller, S. & Branton, D. (1984) *J. Biol. Chem. 98*, 341–346.
Pauloin, A., Bernier, I. & Jolles, P. (1982) *Nature 298*, 574–576.
Payne, G. & Schekman, R. (1985) *Science 230*, 1009–1014.
Payne, G. & Schekman, R. (1989) *Science 245*, 1358–1365.
Pearse, B. M. F. (1988) *EMBO J. 7*, 3331–3336.
Pearse, B. M. F. & Robinson, M. S. (1984) *EMBO J. 3*, 1951–1957.
Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl Acad. Sci USA 85*, 2445–2448.
Ponnambalam, S., Robinson, M. A., Jackson, A. P., Peiperl, L. & Parham, P. (1990) *J. Biol. Chem. 265*, 4814–4820.
Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res. 13*, 7207–7221.
Robinson, M. (1989) *J. Cell Biol. 108*, 833–842.
Robinson, M. (1990) *J. Cell Biol. 111*, 2319–2326.
Robinson, M. S. & Pearse, B. M. F. (1986) *J. Cell Biol. 102*, 48–54.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci USA 74*, 5463–5467.
Thurieau, C., Brosius, J., Burne, C., Jolles, P., Keen, J. H., Mattaliano, R. J., Chow, E. P., Ramachandran, K. L. & Kirchhausen, T. (1988) *DNA 7*, 663–669.
Tucker, K. L., Nathanson, K. & Kirchhausen, T. (1990) *Nucleic Acid Res. 18*, 5306.
Virshup, D. & Bennett, V. (1988) *J. Cell Biol. 106*, 39–50.
Zaremba, S. & Keen, J. H. (1983) *J. Cell Biol. 97*, 1339–1347.