ELSEVIER

Gene 233 (1999) 171–179

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Cloning, mapping and tissue-specific expression of *Drosophila* clathrin-associated protein AP50 gene

Yong Q. Zhang, Kendal Broadie *

Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

Received 28 December 1998; received in revised form 25 March 1999; accepted 8 April 1999; Received by J.A. Engler

Abstract

The *Drosophila* homologue of AP50, the medium chain of clathrin-associated protein complex AP-2, was identified and characterized from the *Drosophila* Expressed Sequence Tag database. The *Drosophila* AP50 is 86% identical to that of mouse and human, and 80% identical to the *Caenorhabditis elegans* homologue. It is a single-copy gene with two mini-introns in the coding region and it maps to position 94B1–B2 on polytene chromosomes. Two P1 clones, DS01102 and DS0104, were identified that contain the AP50 gene. Alternative 5′ UTR splicing is involved in the regulation of AP50 expression. AP50 expression is highly enriched in the central nervous system and midgut caecum during embryo development, and its function is discussed. The two other *Drosophila* members of the medium-chain family of clathrin-associated protein complexes, AP47 and μ3, have also been identified and mapped to 85D20–D27 and 6E1–E4, respectively.

Keywords: Adaptor; AP47; Midgut caecum; μ2; μ3

1. Introduction

Clathrin-associated protein complexes (AP; also referred to as assembly protein or adaptors) are involved in a number of essential cellular processes, including membrane trafficking, protein sorting and endocytosis (for reviews, see Brodsky, 1988; Robinson, 1994; Südhof, 1995; Rothman and Wieland, 1996). Three clathrin-associated protein complexes, AP-1, AP-2 and AP-3, have been identified (Simpson et al., 1997; Dell'Angelica et al., 1997a), and there are indications of the existence of additional complexes or pathways (Andrews et al., 1996; Ahle and Ungewickell 1990; Murphy et al., 1991). The AP-1 complex is associated with the trans-Golgi network (TGN), the AP-2 complex with plasma membrane, and the AP-3 complex associates with the TGN as well as with endosomes (Simpson et al., 1997; Dell'Angelica et al., 1998).

All three AP complexes have similar structure and subunit organization. Each complex consists of four subunits belonging to four different families: two large-chain families, one medium-chain and one small-chain family. The γ-adaptin, α-adaptin and δ-adaptin proteins make up one group of related large subunits for AP-1, AP-2 and AP-3, respectively. The β1(β′)-adaptin, β2(β)-adaptin and β3 (β-NAP/β3B and β3A) proteins form another group of related large subunits. The μ1 (AP47), μ2 (AP50), and μ3 (p47A/μ3A and p47B/μ3B) proteins are related medium-chain subunits, and σ1 (AP19), σ2 (AP17), and σ3 (σ3A and σ3B) are related small chains found in AP-1, AP-2 and AP-3, respectively (Simpson et al., 1996, 1997; Dell'Angelica et al., 1997a,b).

We are interested in clathrin-mediated endocytosis and protein sorting in the presynaptic terminal of neurons in *Drosophila*. Genetic evidence suggests that the AP-2 complex is required for synaptic vesicle (SV) endocytosis, since mutation of the α-adaptin large subunit of AP-2 completely eliminates the presynaptic SV population (Gonzalez-Gaitan and Jackle, 1997). Yeast two-hybrid studies demonstrated that AP50, the medium chain of AP-2, plays a critical role in the complex by providing a site for AP-2 to bind to the cytoplasmic parts of membrane proteins such as TGN38, an integral

Abbreviations: aa, amino acid(s); CNS, central nervous system; EST, Expressed Sequence Tag; GCG, Genetics Computer Group; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; STS, sequence tagged site; TGN, trans-Golgi network; UTR, untranslated region(s).

* Corresponding author. Tel.: +1 801-585-9426/9425; fax: +1 801-581-4668.

E-mail address: broadie@biology.utah.edu (K. Broadie)

0378-1119/99/$ – see front matter
PII: S0378-1119(99)00142-0

membrane protein of TGN (Ohno et al., 1995) and CTLA-4, an immune response regulator (Zhang and Allison, 1997). Taken together, these studies suggest that AP50 may also be required for the sorting of presynaptic trans-membrane proteins such as synaptotagmin and synaptobrevin/VAMP during the recycling of synaptic vesicles. Here we report the molecular characterization and tissue-specific expression of the AP50 gene (*AP50*) from *Drosophila*. Our results show that *Drosophila* AP50 is highly homologous to its mammalian homologues and expressed at high levels within the central nervous system (CNS) and midgut. In addition, we also identified and mapped the other two *Drosophila* members of the medium-chain family: AP47 and μ3 of the AP-1 and the AP-3 complexes, respectively.

2. Materials and methods

2.1. Drosophila AP50 clones

The Berkeley *Drosophila* Genome Project (BDGP) EST database (http://www.fruitfly.org/) was searched for clones with homology to AP50. Two clones, GM02287 and CK02342, representing two different clots (a clot is a group of overlapping EST clones) with high homology to AP50, were obtained from the indicated sources. These two clones share identical coding sequences but have different putative 5′ UTRs. Candidate clones, LD15945 and LD08967 for AP47, LD09732 for μ3, were also obtained and subsequently sequenced and mapped.

2.2. Sequence analysis

DNA sequencing was performed on an ABI sequencer using dye terminator chemistry, and the raw sequences were processed with the program Sequencher (Gene Codes, Ann Arbor, MI, USA) to get the complete consensus sequence. Primers used for DNA sequencing and PCR reactions were: primer 1, ATA GCG CAC AAG ATG ATT GGC (Fig. 1A, bases 539–559); primer 2, CCG CAG AAC ACG GAC TCC G (bases 979–996); primer 3, GGA GAG ACC GCG TCC CTT AG (bases 1339–1358); primer 4, GAT AAT CTT GCT CCG TGA TGG (bases 2004–2024), primer 5′, CTT GTG GTT GTA GAC GAA CAG (bases 563–583); and primer 3′, CCA TCA CGG AGC AAG ATT ATC (complement of primer 4, bases 2004–2024). Primers 1, 2 and 3′ are in the sense direction; primers 3, 4 and 5′ are in the antisense direction (Fig. 1B). To sequence the plasmid insert, 600 ng DNA and 3.2 pM of each primer were used. To sequence the gel purified PCR products, 100 ng DNA and 3.2 pM primer were used. To sequence the P1 clones, 2 μg P1 DNA and 100 ng primer were used. DNA sequences were subsequently analyzed by the University of Wisconsin GCG sequence analysis software (Madison, WI, USA), and sequence alignments were done using CLUSTAL W (version 1.74) from EBI (European Bioinformatics Insititute, Cambridge, UK) server (http://www2.ebi.ac.uk/clustalw/). The protein theoretical isoelectric point and molecular weight were calculated with the program Compute pI/MW (http://expasy.hcuge.ch/ch2d/pi_tool.html). The search for protein motifs was done at http://www.motif.genome.ad.jp.

Both strands of the cDNA clone GM02287 were sequenced by subcloning the full-length *AP50* into two parts split by a *Sac*II site (base 1783, see Fig. 1A). The shorter clone, representing the 3′ part of the gene, was sequenced with vector primers T7 and T3. The longer clone, representing the 5′ part of the gene, was sequenced with vector primers and two synthetic primers (primers 2 and 3, see Fig. 1A).

2.3. In situ hybridization to polytene chromosomes

In situ hybridization to polytene chromosomes was performed essentially according to Ashburner (1989), except that the ABC kit from Vector Laboratories (Burlingame, CA, USA) was used instead of Enzo Biochemicals Detek-1-HPR kit in the post-hybridization detection step.

2.4. Intron–exon mapping of AP50 and STS mapping of P1 clones by PCR

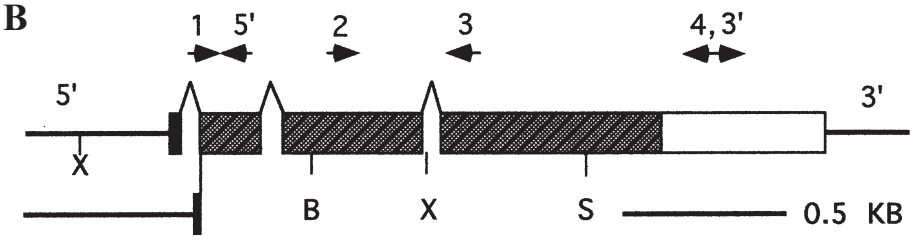
Four primers, primer 1 to primer 4, were used for the PCR-based mapping. The cDNA plasmid, genomic DNA, and P1 colonies (one 1 mm colony was resuspended in 15 μl H_2O, and 1 μl of the suspension was used in a 50 μl PCR reaction) were used as sources for

Fig. 1. (A) The nucleotide sequence of *Drosophila AP50* and its predicted protein product. The three introns are italicized. The two alternative 5′ UTRs (bases 452–478 derived from clone CK02342, bases 517–538 derived from GM02287) are in bold. One of the 5′ UTRs from the base 517–538, which is also the intron for the other transcript, is in italicized bold. Primers for PCR and sequencing, *Sac*II site for subcloning and polyadenylation signal are underlined. Primer 1 is located at base 539–559, primer 2 at 979–796, primer 3 at 1339–1358, primer 4 at 2004–2024, primer 5′ at 563–583 and primer 3′ at 2004–2024. Primers 1, 2 and 3′ are in the sense direction; primers 3, 4 and 5′ are in the antisense direction. Stop codon TAG is indicated by an asterisk. The polyadenylation site is indicated as a bold T at base 2502. The EMBL/GenBank accession number for *Drosophila AP50* is AJ005962. Amino acids in bold represent the two signatures of the medium-chain family of APs. (B) Diagram of gene structure and restriction map of *AP50*. The arrows at the top indicate the orientations of primers and the numbers at the top of arrows represent the names of primers. Pattern-filled boxes represent coding regions; blank box, 3′ UTR; black boxes, alternative 5′ UTRs. Solid bars represent functionally unknown flanking sequences of *AP50*. 5′ and 3′ at the two sides indicate the orientation of *AP50*. B refers to *Bgl*II; S, *Sac*II; X, *Xba*I.

A

```
ATTATACCGCATATTTTTACATTGCCAAAAATACCAGAGCGATGTCCATCAAGATAGCGACGAAATTAGA  70
ACAGTGCAATTGCCAATTGGGAATTTGTATTTTAATTTATTTTTAAATTCTGAAAGTAATTTAAATTTAA 140
AAAAAAAACTTGAGAGCTGTCTAGAAAAGAACTGATGTTTCATGATAACTTTGTCGAAGAATTAAGAAAT 210
ATTTAGTTGTAAAATAATTGTTGAATCTATTTTTTTTCCAATAACACGACTTATATATTTTTTGAAAATAT 280
TCGAGCTAAATCCCAAGAAGTAAACTCAATCTGGGATTTGAAGTGCCCAGAACTCGAATAAACACTTCTT 350
TTTAAATAATTGTAAGACCGTATCACTTATGGTATATACTGACCTCGAAGGGCCACACTAAGGGGGAGTG 420
AAAATTGATTTTCTGATAAAAATTTTCGCTTGAAGCTACAGCATCGTCCACTGTCCATGTATATATCTTA 490
TATTTGCATATAAATATATATATTACACCGACTTGGACTAACCATCAGATAGCGCACAAGATGATTGGCG 560
  primer 5'                                          primer 1   M I G G
GCCTGTTCGTCTACAACCACAAGGGCGAGGTGCTGATCTCGCGAGTTTACCGCGACGACATCGGTCGGAA 630
  L F V Y N H K G E V L I S R V Y R D D I G R N
TGCCGTGGACGCCTTTCGGGTCAACGTCATCCACGCCCGCCAGCAGGTCCGCTCGCCAGTGACCAATATT 700
 A V D A F R V N V I H A R Q Q V R S P V T N I
GCGAGGACCAGCTTCTTCCACATCAAGGTCGGTGAACAAAGTTCCTATGTCACAATCACCAACCCTCCCA 770
A R T S F F H I K
CTTATCATCAAAACTCCTTTCACAGAGAGCAAACATTTGGCTGGCGGCTGTGACCAAGCAGAATGTGAAC 840
                        R A N I W L A A V T K Q N V N
GCCGCGATGGTGTTTGAGTTCCTTTTGAAGATCATCGAGGTGATGCAATCCTACTTCGGCAAGATCTCGG 910
A A M V F E F L L K I I E V M Q S Y F G K I S E
AGGAGAACATCAAGAATAACTTCGTGCTCATCTACGAGCTGCTGGATGAGATCCTCGACTTTGGCTACCC 980
  E N I K N N F V L I Y E L L D E I L D F G Y P
GCAGAACACGGACTCCGGCACCCTGAAGACCTTCATCACACAGCAGGGCATCAAGTCGGCCACCAAGGAG 1050
 Q N T D S G T L K T F I T Q Q G I K S A T K E
primer 2
GAGCAGATGCAGATAACCTCGCAGGTTACCGGCCAGATTGGCTGGCGTCGCGAGGGCATCAAGTACCGGC 1120
E Q M Q I T S Q V T G Q I G W R R E G I K Y R R
GCAACGAGCTTTCCTGGACGTATTGGAGTACGTGAACCTGCTGATGAGCCCGCAGGGTCAGGTTTTGTC 1190
  N E L F L D V L E Y V N L L M S P Q G Q V L S
TGCCCACGTGGCCGGCAAGGTGGTAATGAAGTCGTATTTGTCGGGTAAGTTAGCATAAATAATCTAGACA 1260
 A H V A G K V V M K S Y L S
TTATTCCTTTTTAATAATCGCCATGTTGTAGGCATGCCCGAGTGCAAGTTCGGGATTAACGACAAGATCGT 1330
        primer 3            G M P E C K F G I N D K I V
GATGGAGTCTAAGGGACGCGGTCTCTCCGGAAATTCAGAGGCGGAAACCTCACGCTCCGGCAAGCCCGTC 1400
 M E S K G R G L S G N S E A E T S R S G K P V
GTGGTCATCGATGACTGCCAGTTCCATCAGTGCGTCAAGCTAAGCAAATTCGAGACGGAGCATTCGATCA 1470
V V I D D C Q F H Q C V K L S K F E T E H S I S
GCTTCATCCCGCCGGACGGGGAGTTCGAGCTGATGCGTTACCGTACCACCAAAGACATTTCGCTGCCATT 1540
  F I P P D G E F E L M R Y R T T K D I S L P F
CCGAGTCATCCCGCTGGTGCGGGAGGTGGGCCGCACCAAGATGGAGGTTAAGGTTGTGCTGAAGTCCAAC 1610
 R V I P L V R E V G R T K M E V K V V L K S N
TTTAAGCCCTCACTGCTGGGCCAAAAGATCGAGGTGAAGATACCAACCCCGCTCAATACATCGGGCGTGC 1680
F K P S L L G Q K I E V K I P T P L N T S G V Q
AGCTCATCTGCCTAAAGGGCAAAGCCAAATATAAGGCTTCGGAGAACGCGATCGTGTGGAAGATTAAGCG 1750
  L I C L K G K A K Y K A S E N A I V W K I K R
CATGGCGGGCATGAAGGAGACACAGCTGTCCGCGGAAATCGAACTTTTGGAGACGGACACCAAGAAGAAG 1820
 M A G M K E T Q L S A E I E L L E T D T K K K
TGGACTCGGCCGCCCATCTCCATGAACTTTGAGGTGCCATTCGCGCCGTCCGGCTTCAAGGTACGCTACC 1890
W T R P P I S M N F E V P F A P S G F K V R Y L
TGAAGGTGTTCGAGCCCAAGCTCAACTACTCCGACCACGATGTGGTCAAATGGGTGCGCTACATCGGACG 1960
  K V F E P K L N Y S D H D V V K W V R Y I G R
CAGTGGCTTGTATGAGACGCGCTGCTAGGGCGCCCAAGCATCCCCATCACGGAGCAAGATTATCACAATT 2030
 S G L Y E T R C *                    primer 4 and primer 3'
GAATTTAACGAGATGAATGGAGAAAGTGTTACCGCATCTATTGTTCCAGCAAGTCCGGAGGACGCCCGAG 2100
TAGCGGATCCACCCCGGACACCCTATGCCGTATACCGCACACACTGCCTCCGCGAGCACTGGTTATGCGG 2170
TGTTCCCCATAGACATACCACTACTAGGTACTATGATTTACCCTACTATGTAAATGTAAACTATGTTGTT 2240
TTTATATGCGTCGCGGGACGTCTCTCAAATCTCACGCTCACGTTTCACAACTCTAATTATTATTTGTATT 2310
ATCGTATTATAAATATATATATATATATGTATATTTAATCATTATTTTCATTGTGCAAAGAACAAAGGCG 2380
ATAAATGCAACTAAGCGTATTTAATAGACCAAAACTCTGTCAGCGCTATCAATATTGATATTCAAGTGTT 2450
AATAATATTCGAATTGTTTAATAAATAAATTAGCATGATATATACTAAAAATGATGTGTTTCTCATTGGT 2520
GAAAGGGTATCGTTATATGGCTTTTCCCTATATCGAACTGATATGCCCAACAATAGAACGGAAAGGTACG 2590
ATTACAACCTCATAAATGGGTGATATACAACTTTATGATCAGCCCACAAGTGCTGTCTTCGCTATCACGT 2660
GTGCACATTGGTTCAAGATCAGCTAAATCCTGAAATTAACGTCACAATTATTGATGATGTTCAACGAAGT 2730
GAAAACTAGTTTCTAGCCATGAATCATCGC                                          2760
```

B



PCR templates. The thermal cycle was 94°C for 45 s, 50°C for 30 s, and 72°C for 90 s for 35 cycles.

2.5. Southern blot analysis

A full-length *AP50* fragment was generated by restriction digestion of GM02287 with *Eco*RI and *Xho*I enzymes. The released 1.8 kb fragment was labeled with [α-^{32}P]dCTP using a random priming kit (cat. 300385) from Stratagene according to the product protocol. *Drosophila* genomic DNA was prepared from the wild type strain Oregon R according to Ashburner (1989). Aliquots of genomic DNA were digested with *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Xba*I or *Xho*I and then separated, blotted, hybridized and detected according to Sambrook et al. (1989). Low stringency buffer (2 × SSC, 0.5% SDS, 1 mM EDTA) was used for the post-hybridization washing.

2.6. In situ hybridization to whole-mount embryos

RNA probes were labeled with digoxigenin-11-dUTP using a DIG RNA labeling kit (cat. 1 175 025, Boehringer Mannheim) following the product protocol. Briefly, the clone GM02287 with full-length insert of *AP50* was linearized with *Pst*I for synthesizing the antisense RNA probe with T7 polymerase, and with *Xho*I for synthesizing the sense RNA probe with T3 polymerase. The newly generated RNA probes were hydrolyzed in alkaline condition at 60°C for 20 min to reduce their lengths, allowing for better penetration of the probes. Under our labeling and detection conditions, both sense and anti-sense probes were equally sensitive assayed by dot blot test. In situ hybridization was performed according to Tautz and Pfeifle (1989), except that proteinase K treatment was omitted in the pretreatment step and levamisole treatment was omitted in the staining step. The nervous system specific expression gene *embryonic lethal abnormal visual system* (*elav*, Campos et al., 1987) was used as a positive control. The *elav* clone (HL03451) was ordered from the same source as the AP50 clones and its identity was confirmed by sequencing. The clone was linearized with *Kpn*I and *Sac*I to make antisense and sense probes with T7 polymerase and T3 polymerase, respectively.

3. Results

3.1. Identification of cDNA clones encoding the Drosophila homologue of AP50

A search of the BDGP EST database as of 21 December 1998 revealed the existence of five clones in two clots (clots 4561 and 1129), which encoded polypeptides with significant homology to AP50, the medium chain of AP-2. Two EST clones, GM02287 and CK02342, representing the two different clots, were obtained and analyzed. When the clones were ordered, GM02287 from the ovary cDNA library had the longest putative 5′ UTR and CK02342 from the rough endoplasmic reticulum cDNA library was the only clone in the clot. According to the *Drosophila* EST sequences, the putative 5′ UTRs from the two overlapping clones, CK02342 and GM02287, are different (EMBL/GenBank accession number AA141728 for CK02342, AA567297 for GM02287). Our sequencing of the two clones showed that they were identical except for the putative 5′ UTRs. To answer the question as to whether the 5′ UTR difference was the result of a cloning artifact or represented alternative splicing, we sequenced the genomic DNA upstream of the start codon ATG using P1 clones as sequencing templates. The results show that the 5′ end difference is indeed from alternative splicing of the 5′ UTRs. A 60 bp intron was found in the 5′ UTR of CK02342, whereas GM02287's 5′ UTR links to the coding region without interruption (see Fig. 1A). This finding demonstrates that 5′ UTR alternative splicing is involved in the regulation of *AP50* expression.

We have sequenced the full-length cDNA of AP50 from the EST clones and the resultant sequence is presented in Fig. 1. The predicted AP50 protein consists of 437 aa residues with a deduced molecular weight of 49 856 Da. The protein is basic with a calculated isoelectric point of 9.52. A search for protein motifs showed that there are two medium-chain signatures, one at 157–177 and another at 265–279 (see Fig. 1A). The two fly AP50 signatures are identical to the mammalian homologues (data not shown). Sequence comparison shows that the *Drosophila* AP50 is 86% identical to human AP50 (accession No. U36188), 87% to mouse AP50 (Ohno et al., 1995) and 82% to *C. elegans* AP50 (Lee et al., 1994). Interestingly, the human AP50 is 99.8% identical to mouse AP50. This remarkable homology is consistent with the previous observation that proteins involved in specific and complex interactions with many macromolecules are highly conserved (Ponnambalam et al., 1990). Comparison to different members of medium-chain family in *Drosophila* showed that AP50 is 41% identical to AP47 (accession No. AJ006219) and 25% to μ3 (accession No. AJ009657) of AP-1 and AP-3, respectively. This homology among the three members of the medium-chain family in *Drosophila* is similar to the homology observed in their counterparts in rat and mouse (Pevsner et al., 1994).

3.2. AP50 maps at 94B1–B2

Polytene chromosome in situ hybridization with an *AP50* cDNA probe lit up a unique band at 94B1–B5

(see Fig. 2). To confirm the cytology, three different P1 clones, DS01102 (94B1–B2), DS04104 and DS05321 (94B2–B4), listed in Berkeley Fly Database (a subset database of BDGP) were used as PCR templates for STS mapping. These three P1 clones overlap and constitute a physical contig, Dm 1595 (Kimmerly et al., 1996). Two P1 clones, DS01102 and DS04104, gave expected bands using combinations of primer 1 and primer 4 (see lanes 6 and 7 of Fig. 3), and primer 1 and 3 (data not shown). Whereas, P1 clone DS05321 gave no positive bands (lane 8 of Fig. 3) under the same conditions. The PCR-based mapping of these three P1 clones confirms the location of *AP50*. Furthermore, it narrows down the gene location to 94B1–B2, inferred from the cytology and the STS contig of the three P1 clones.

The other two *Drosophila* members of the AP medium-chain family, AP47 and μ3, were identified using an identical approach to that described above. The full-length cDNA sequences for the two genes were deposited in the EMBL/GenBank with accession numbers AJ006219 for AP47 and AJ009657 for μ3. The two genes, *AP47* and *μ3*, were mapped to 85D20-D27 and 6E1-E4, respectively, and the genomic positions were confirmed by PCR mapping of P1 clones in the regions (data not shown). These two genes were not investigated further.

3.3. AP50 is a single-copy gene and contains two mini-introns in the coding region

The unique site displayed by polytene chromosome in situ hybridization implies that there is only one copy of *AP50* in the genome. Southern hybridization using a full-length *AP50* probe shows a very simple pattern (see Fig. 4). Single positive bands were observed in single *Eco*RI, *Hin*dIII, *Pst*I and *Xho*I digestions. Two bands were observed in the *Bgl*II digest (a strong 8 kb band and a weak band larger than 12 kb), and two bands in the *Xba*I digest (a strong 10 kb band and a weak 1.1 kb band). These are expected from the sequence shown in Fig. 1A, as there is a *Bgl*II site (base 902 of Fig. 1A) and two *Xba*I sites (bases 160 and 1253, the 5′ most 160 bp *Xba*I fragment did not show up in the Southern, as the cDNA probe did not extend that far). The results from both polytene chromosome in situ hybridization and Southern hybridization are consistent with the presence of one gene copy and the absence of pseudogenes coding for AP50. None of the bands correspond to a cross hybridization with the related and unique *Drosophila* AP47 gene (accession No. AJ006219), since a different pattern was obtained in a parallel hybridization with a full length *Drosophila* AP47 cDNA probe (data not shown).

To further probe the gene structure of *AP50*, PCR was performed using both genomic DNA and cDNA as templates with various combinations of primers (see Fig. 3). There was approx. 70 bp difference between PCR products from cDNA and genomic DNA templates using a combination of primers 2 and 3 (see lanes 3, 4 of Fig. 3), and a combination of primer 2 and primer 4 (data not shown). This demonstrated that a mini-intron of about 70 bp exists in the gene between primer 2 and primer 3, but there is no intron between primer 3 and primer 4. Also, there was approx. 150 bp difference between PCR products from the cDNA and genomic DNA templates using a combination of primer 1 and primer 3, and a combination of primer 1 and primer 4 (lane 1, 2, 5, 9 of Fig. 3). This suggested that there is a 70 bp intron between primer1 and primer 2, in addition to the 70 bp intron between primer 2 and primer 3.

Full-length sequencing of the PCR products from genomic DNA templates with a combination of primer 1 and primer 4 confirmed the presence of a 68 bp intron between primer 1 and primer 2, and a 56 bp intron between primer 2 and primer 3 (see Fig. 1A). The genomic sequences containing the 5′ UTR or the 3′ UTR were obtained by sequencing the P1 clone DS4104 with primer 5′ or primer 3′, respectively. In addition to the two introns in the coding region, one intron is found in one of the two alternative 5′ UTR regions (see the first paragraph in Section 3.1 and Fig. 1A). Therefore, the gene structure of *Drosophila* AP50 is much simpler than that of the mouse AP50 gene, which has 12 exons and 11 introns (Ohno et al., 1998), and that of the *C. elegans* AP47 gene, which has seven exons and six introns (Lee et al., 1994).

3.4. Spatial and temporal expression of AP50 during embryonic development

AP50 message RNA is highly enriched in neurons of the CNS and in the midgut during embryogenesis. Both tissues are specialized for high levels of exocytosis/endocytosis cycling: neurotransmitter release in CNS and secretion of digestive enzymes in gut. No appreciable expression was observed in either tissue before stage 8 (staging according to Campos-Ortega and Hartenstein, 1985). By stage 9, weak but consistent expression of *AP50* was first detected in the invaginated midgut and neuroblasts of the CNS germ band (Fig. 5A). The strength of expression increased in both locations with embryonic development from this stage. By stage 13 (Fig. 5B), expression in the midgut and CNS was striking, especially in the midgut caecum (see Poulson, 1950 and Skaer 1993), which had by far the strongest expression (Fig. 5C–E). The expression pattern is consistent with that described for the EST clone CK02342 in the BDGP database (Kopczynski et al., 1998).

As both AP50 and α-adaptin are subunits of AP-2, we expected to see that the two genes had identical expression patterns. However, this is not the case. The

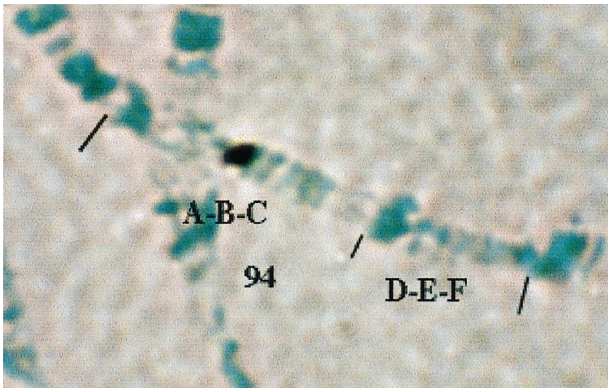


Fig. 2. In situ hybridization of *AP50* cDNA to polytene chromosomes. Polytene map divisions (Lindsley and Zimm, 1992) around the target region are shown.

most striking difference is that *AP50*, but not α-adaptin, is abundantly expressed in midgut caecum (Fig. 5; Dornan et al., 1997; Gonzalez-Gaitan and Jackle, 1997). The reason for this remarkable difference in expression of two subunits of the same protein complex is unknown at present. Presumably, AP50 is a member of another complex in the midgut caecum, or an unidentified AP2 large subunit substitutes for the α-adaptin in the midgut caecum. However, both genes are highly expressed in the CNS from a similar stage of development. These

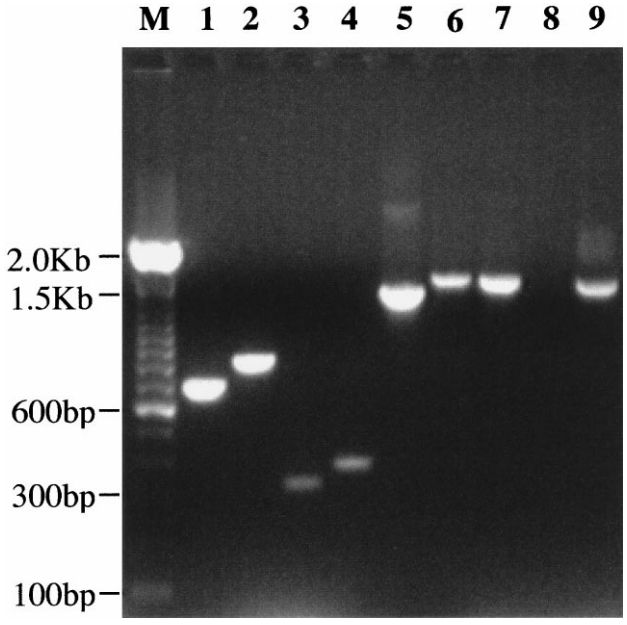


Fig. 3. Intron–exon mapping of *AP50* and PCR-based mapping of P1 clones. Lane M contained 100 bp marker from Gibco BRL. The sizes are marked on the left. Lanes 1, 3 and 5 used cDNA as templates; lanes 2, 4 and 9 used genomic DNA as templates; lanes 6, 7 and 8 used P1 clones DS04104, DS01102 and DS05321 as templates, respectively. Lanes 1 and 2 used a combination of primer 1 and primer 3; lanes 3 and 4 used a combination of primer 2 and primer 3; lanes 5–9 used a combination of primer 1 and primer 4.

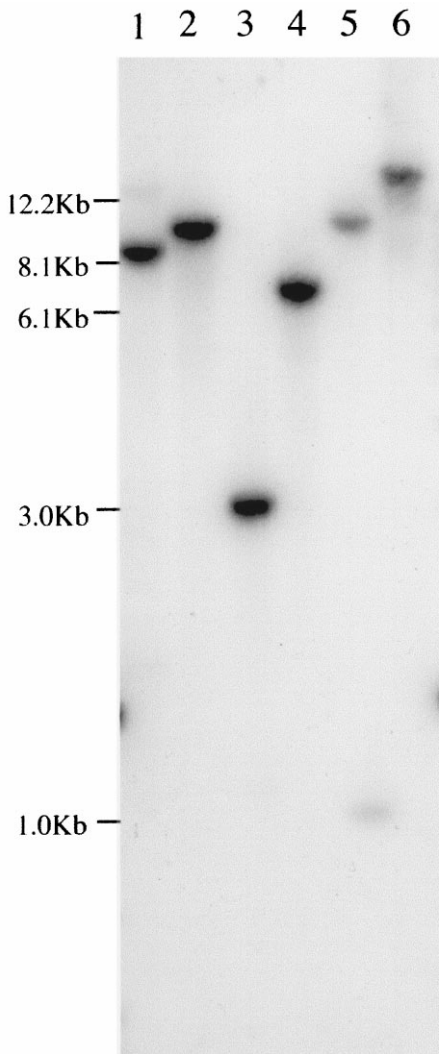


Fig. 4. Southern hybridization of genomic DNA from wild-type strain Oregon R with full-length *AP50* insert as probe. Lanes 1–6 contained DNA digested with *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Xba*I and *Xho*I, respectively. The sizes of DNA marker are indicated on the left.

data suggest that AP50 and α-adaptin form components of the AP2 complex in the CNS, as expected.

AP50 appeared to display a pan-neural expression pattern from the onset of neurogenesis (stage 9–10, Fig. 5A). However, in late embryo development, only the CNS, but not the peripheral nervous system, showed detectable AP50 expression (Fig. 5C–E). We noticed that the CNS expression of *AP50* is much lower than that of *elav* (Fig. 5F; note the color development time difference between the *AP50* probe and the *elav* probe), and slightly lower than that of *unc-13*, a gene involved in neuronal transmission (Aravamudan and Broadie, unpublished) based on the staining intensities obtained from parallel experiments. These results show that the AP50 transcript is at relatively low abundance in the CNS, although expressed in all identifiable central neurons.

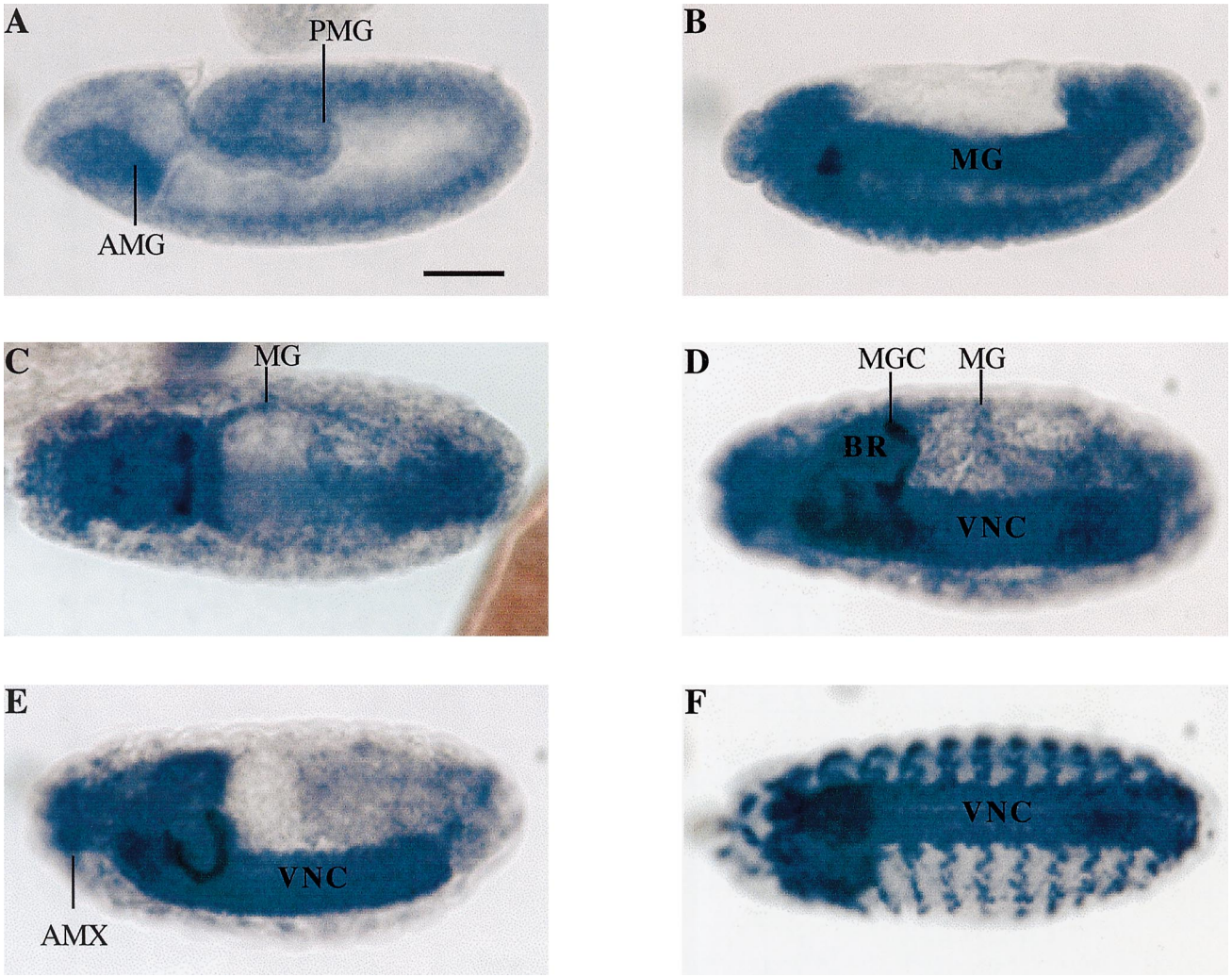


Fig. 5. Expression of *Drosophila AP50* during embryonic development. (A–E) Hybridized with *AP50* antisense probes and stained for 60 min; (F) hybridized with *elav* antisense probe and stained for 10 min. Hybridization with sense probes in the same condition did not give any appreciable signals (data not shown). Embryos are oriented with anterior to the left (A–F), dorsal to the top (A, B, D and E), dorsal facing up (C) or dorsal facing down (F). (A) Embryo at stage 9; (B) stage 13; (C) stage 14; (D) stage 15; (E) stage 16; (F) stage 15. AMG, anterior midgut; AMX, antenno-maxillary complex; BR, brain; MG, midgut; MGC, midgut caecum; PMG, posterior midgut; VNC, ventral nerve cord. Scale bar represents 90 μm.

4. Discussion

4.1. All three AP complexes are present in Drosophila

Drosophila contains all three identified AP complexes. We have identified the homologue of AP47 (accession No. AJ006219) and putative clones for AP19 (LD14109, GM05816 and LD10283 from the BDGP EST database) of AP-1. Likewise, α-adaptin (Gonzalez-Gaitan and Jackle, 1997), β-adaptin (Camidge and Pearse, 1994) and AP50 (this report) of the AP-2 complex have now been identified. The β3 (accession No. AJ011778), δ-adaptin (Ooi et al., 1997; Simpson et al., 1997) and μ3 (accession No. AJ009657) of the AP-3 complex have also been identified in *Drosophila*. Thus, there is no doubt that all three AP complexes identified so far in other organisms are present in *Drosophila*, although the complete set of subunits for all three AP complexes have yet to be described.

In addition to the three medium-chain members we describe here, we have also characterized an additional, special member of the medium-chain family in *Drosophila* called Stoned B (STNB) (Fergestad et al., in press). The N-terminal one-third of STNB is 23% identical to AP50/AP47 family (Andrews et al., 1996). This protein is strongly expressed in the presynaptic terminal of *Drosophila* motor neurons as well as interneurons, and is involved in synaptic vesicle biogenesis and protein sorting (Fergestad et al., in press). However, STNB's low sequence homology to the AP50/AP47 family and its much larger size indicate that it is not an orthologue of AP50 or AP47. STNB (1260 aa) is about three times

bigger than AP50/AP47. Thus, STNB may be a subunit of a novel AP complex or participate in a novel pathway involved in vesicle trafficking.

4.2. AP50 is a critical subunit of AP-2 and provides a binding site for plasma membrane proteins

Different models for the AP complex have been postulated based on the EM analysis, proteolysis and structure analysis of adaptins. AP50 has been suggested to be either buried in the middle of the complex (Ponnambalam et al., 1990; Robinson, 1994) or at the N-terminal side of the adaptins (Kirchhausen et al., 1989). However, these models cannot explain recent findings regarding AP50 function. The new data demonstrate that AP50 interacts with the cytoplasmic domain of many trans-membrane proteins, such as CTLA-4, a T cell surface receptor expressed transiently after T cell activation (Zhang and Allison, 1997) and tyrosine-signal containing membrane proteins (Ohno et al., 1995). In addition, Zhang et al. (1994) showed that AP2 binds synaptotagmin I, an intrinsic membrane protein of synaptic vesicles, although the actual binding subunit(s) within AP2 has not been elucidated.

These data suggest that AP50 must be exposed to interact with plasma membrane proteins, instead of buried in the middle of the AP2 complex. However, the membrane protein binding domain(s) of AP50 has yet to be defined. Moreover, the molecular interactions between AP50 and other subunits of AP2 complex remain unresolved. The molecular characterization of *AP50* from *Drosophila* will be instrumental to the future structural and functional study of AP50 to answer these questions.

4.3. Comments on mutants of AP50

Searching through FlyBase, the comprehensive database on *Drosophila* (The Flybase Consortium, 1999), we found that 94B, where *AP50* maps, is a relatively poor region in terms of mutants, P element lines and chromosome aberrations, compared to other regions in the genome. The only lethal mutation induced by P element insertion in the region, *l(3)L4910*, was plasmid rescued and the rescued sequence was deposited in the EMBL/GenBank database with accession number AJ132670. The rescued sequence showed no homology to *AP50* (data not shown). The cytological localization and expression pattern of *AP50* suggest that it may correspond to the previously described gene Nc94B and/or anon-J7 (Kokoza et al., 1991a,b), which are also expressed in the CNS and map at 94B. We are currently investigating these possibilities and preparing for a mutagenesis of the *AP50* locus to facilitate the functional study of the protein.

Acknowledgements

This work was supported by a NIH grant GM54544 and a Searle Scholarship to K.B.. We thank Emma Rushton and Chris Rodesch for aid in interpreting the patterns of whole-mount embryo in situ. We are grateful to Carl Thummel's laboratory for the P1 clones, Bharathi Aravamudan for the *unc-13* antisense probe, and the Berkeley *Drosophila* EST Project led by Gerald Rubin for generating and distributing the EST clones that made this study possible.

References

Ahle, S., Ungewickell, E., 1990. Auxilin, a newly identified clathrin-associated protein in coated vesicles from bovine brain. J Cell Biol. 111, 19–29.

Andrews, J., Smith, M., Merakovsky, J., Coulson, M., Hannan, F., Kelly, L.E., 1996. The stoned locus of *Drosophila melanogaster* produces a dicistronic transcript and encodes two distinct polypeptides. Genetics 143, 1699–1711.

Ashburner, M., 1989. *Drosophila*: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brodsky, F.M., 1988. Living with clathrin: its role in intracellular membrane traffic. Science 242, 1396–1402.

Camidge, D.R., Pearse, B.M., 1994. Cloning of *Drosophila* beta-adaptin and its localization on expression in mammalian cells. J Cell Sci. 107, 709–718.

Campos-Ortega, J.A., Hartenstein, V., 1985. The Embryonic Development of *Drosophila mangaster*. Springer-Verlag, Berlin.

Campos, A.R., Rosen, D.R., Robinow, S.N., White, K., 1987. Molecular analysis of the locus *elav* in *Drosophila melanogaster*: a gene whose embryonic expression is neural specific. EMBO J. 6, 425–431.

Dell'Angelica, E.C., Klumperman, J., Stoorvogel, W., Bonifacino, J.S., 1998. Association of the AP-3 adaptor complex with clathrin. Science 280, 431–434.

Dell'Angelica, E.C., Ohno, H., Ooi, C.E., Rabinovich, E., Roche, K.W., Bonifacino, J.S., 1997a. AP-3: an adaptor-like protein complex with ubiquitous expression. EMBO J. 16, 917–928.

Dell'Angelica, E.C., Ooi, C.E., Bonifacino, J.S., 1997b. Beta3A-adaptin, a subunit of the adaptor-like complex AP-3. J. Biol. Chem. 272, 15078–15084.

Dornan, S., Jackson, A.P., Gay, N.J., 1997. Alpha-adaptin, a marker for endocytosis, is expressed in complex patterns during *Drosophila* development. Mol. Biol. Cell 8, 1391–1403.

Fergestad, T.J., Davis, W.S., Broadie, K.S., The stoned proteins regulate synaptic recycling in the presynaptic terminal. J. Neurosci., in press.

Gonzalez-Gaitan, M., Jackle, H., 1997. Role of *Drosophila* alpha-adaptin in presynaptic vesicle recycling. Cell 88, 767–776.

Kimmerly, W., Stultz, K., Lewis, S., Lewis, K., Lustre, V., Romero, P., Benke, J., Sun, D., Shirley, G., Martin, C., Palazzolo, M., 1996. A P1-based physical map of the *Drosophila* euchromatic genome. Genome Res. 6, 414–430.

Kirchhausen, T., Nathanson, K.L., Matsui, W., Vaisberg, A., Chow, E.P., Burne, C., Keen, J.H., Davis, A.E., 1989. Structural and functional division into two domains of the large (100- to 115-kDa) chains of the clathrin-associated protein complex AP-2. Proc. Natl. Acad. Sci. USA 86, 2612–2616.

Kokoza, V.A., Baritcheva, E.M., Katokhin, A.V., 1991a. Cloning and chromosomal localization of conservative brain-specific sequences in *Drosophila melanogaster*. *Drosophila* Inf. Serv. 70, 121–123.

Kokoza, V.A., Baricheva, E.M., Katokhin, A.V., 1991b. Cloning and cytogenetic localization of evolutionarily conserved genomic sequences which are expressed in the head of the *Drosophila melanogaster* imago. Genetika 27, 51–60.

Kopczynski, C.C., Noordermeer, J.N., Serano, T.L., Chen, W.Y., Pendleton, J.D., Lewis, S., Goodman, C.S., Rubin, G.M., 1998. A high throughput screen to identify secreted and trans-membrane proteins involved in *Drosophila* embryogenesis. Proc. Natl. Acad. Sci. USA 95, 9973–9978.

Lee, J., Jongeward, G.D., Sternberg, P.W., 1994. unc-101, a gene required for many aspects of *Caenorhabditis elegans* development and behavior, encodes a clathrin-associated protein. Genes Dev. 8, 60–73.

Lindsley, D.L., Zimm, G.G., 1992. The Genome of *Drosophila melanogaster*. Academic Press, London.

Murphy, J.E., Pleasure, I.T., Puszkin, S., Prasad, K., Keen, J.H., 1991. Clathrin assembly protein AP-3. The identity of the 155K protein, AP 180, and NP185 and demonstration of a clathrin binding domain. J. Biol. Chem. 266, 4401–4408.

Ohno, H., Poy, G., Bonifacino, J.S., 1998. Cloning of the gene encoding the murine clathrin-associated adaptor medium chain mu 2: gene organization, alternative splicing and chromosomal assignment. Gene 210, 187–193.

Ohno, H., Stewart, J., Fournier, M.C., Bosshart, H., Rhee, I., Miyatake, S., Saito, T., Gallusser, A., Kirchhausen, T., Bonifacino, J.S., 1995. Interaction of tyrosine-based sorting signals with clathrin-associated proteins. Science 269, 1872–1875.

Ooi, C.E., Moreira, J.E., Dell'Angelica, E.C., Poy, G., Wassarman, D.A., Bonifacino, J.S., 1997. Altered expression of a novel adaptin leads to defective pigment granule biogenesis in the *Drosophila* eye color mutant garnet. EMBO J. 16, 4508–4518.

Pevsner, J., Volknandt, W., Wong, B.R., Scheller, R.H., 1994. Two rat homologs of clathrin-associated adaptor proteins. Gene 146, 279–283.

Ponnambalam, S., Robinson, M.S., Jackson, A.P., Peiperl, L., Parham, P., 1990. Conservation and diversity in families of coated vesicle adaptins. J. Biol. Chem. 265, 4814–4820.

Poulson, D.F., 1950. Histogenesis, organogenesis and differentiation in the embryo of *Drosophila melanogaster* Meigen. In: Demerec, M. (Ed.), Biology of *Drosophila*. John Wiley, New York, pp. 168–274.

Robinson, M.S., 1994. The role of clathrin, adaptors and dynamin in endocytosis. Curr. Opin. Cell Biol. 6, 538–544.

Rothman, J.E., Wieland, F.T., 1996. Protein sorting by transport vesicles. Science 272, 227–234.

Sambrook, J., Maniatis, T., Fritsch, E.F., 1989. Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Simpson, F., Bright, N.A., West, M.A., Newman, L.S., Darnell, R.B., Robinson, M.S., 1996. A novel adaptor-related protein complex. J. Cell. Biol. 133, 749–760.

Simpson, F., Peden, A.A., Christopoulou, L., Robinson, M.S., 1997. Characterization of the adaptor-related protein complex, AP-3. J. Cell. Biol. 137, 835–845.

Skaer, H., 1993. The alimentary canal. In: Bate, M., Martines Arias, A. (Eds.), The Development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 941–1012.

Südhof, T.C., 1995. The synaptic vesicle cycle: a cascade of protein–protein interactions. Nature 375, 645–653.

Tautz, D., Pfeifle, C., 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. Chromosoma 98, 81–85.

The Flybase Consortium, 1999. The flybase database of the *Drosophila* genome projects and community literature. Nucleic Acids Res. 27, 85–88.

Zhang, J.Z., Davletov, B.A., Südhof, T.C., Anderson, R.G., 1994. Synaptotagmin I is a high affinity receptor for clathrin AP-2: implications for membrane recycling. Cell 78, 751–760.

Zhang, Y., Allison, J.P., 1997. Interaction of CTLA-4 with AP50, a clathrin-coated pit adaptor protein. Proc. Natl. Acad. Sci. USA 94, 9273–9278.