ELSEVIER

Gene 208 (1998) 191–199

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Isolation and characterisation of *smallminded*, a *Drosophila* gene encoding a new member of the Cdc48p/VCP subfamily of AAA proteins

Antony R. Long [a], Mingyao Yang [b], Kim Kaiser [b], David Shepherd [a,*]

[a] *School of Biological Sciences, Division of Cell Sciences, University of Southampton, Southampton, SO16 7PX, UK*
[b] *Division of Molecular Genetics, Pontecorvo Building, University of Glasgow, Glasgow, UK*

Received 2 September 1997; accepted 7 November 1997; Received by received by E. Boncinelli

Abstract

Smallminded (*smid*) encodes a new member of the cdc48p/VCP subfamily of AAA proteins in *Drosophila*. The gene was isolated by plasmid rescue from a GAL4 enhancer trap line which shows reporter gene expression in neuroblasts, imaginal disks and a subset of sensory neurons. Larvae homozygous for the insert arrest development as second instar larvae and die without pupating. The most obvious defect in these larvae is a significantly reduced CNS, hence the naming of the gene as *smallminded*. The deduced amino acid sequence of *smid* contains a tandem duplication of the AAA nucleotide binding domain characteristic of the cdc48p/VCP subfamily. Overall, *smid* shares 33% identical residues with its closest relative, yeast L0919-chrXII and 26–29% with other members of the cdc48p/VCP subfamily. The most highly conserved regions of the predicted protein structure are found in and around the nucleotide binding domains. The gene is expressed at all developmental stages. © 1998 Elsevier Science B.V.

Keywords: Cell cycle; Enhancer-trap; Gal4; Homotypic membrane fusion; Neuroblast

1. Introduction

Since the cloning of porcine valosin containing protein (VCP) by Koller and Brownstein (1987), more than 100 genes encoding related proteins have been reported. All of these proteins share either a single or duplicated highly conserved nucleotide binding domain (AAA module), which defines the superfamily, but otherwise show considerable structural heterogeneity beyond these regions (Confalonieri and Duget, 1995). This is reflected in the diversity of their known functions, hence the naming of the family members (Kunau et al., 1993) as *A*TPases *A*ssociated with diverse cellular *A*ctivities (AAA). AAA proteins play essential roles in processes as diverse as proteasome function (Dubiel et al., 1992), chaperoning (Arlt et al., 1996), gene regulation (Ohana et al., 1993), organelle biogenesis (Erdmann et al., 1991), membrane fusion (Whiteheart et al., 1994) and cell division (Fröhlich et al., 1991).

Phylogenetic analysis of the AAA family has been used to place its members into six subfamilies of structurally related proteins (AAA server, http://yeamob.pci.chemie.uni-tuebingen.de/kai.html). One such subfamily (Cdc48p/VCP) contains VCP and its homologues from mouse (mVCP; Egerton et al., 1992), *Xenopus* (p97; Peters et al., 1990) rat (transitional ER-ATPase; Zhang et al., 1994), yeast (Cdc48p; Fröhlich et al., 1991) and the higher plants *Arabidopsis* (Atcdc48p; Feiler et al., 1995) and soy bean (sVCP; Shi et al., 1995). More distantly related members include the yeast proteins AFG2 (Thorsness et al., 1993), YTA7 (Schnall et al., 1994) and L0919-chroXII (Z73139); archaebacterial proteins SAV, CdcH and mCdc48p (Confalonieri et al., 1994; Bibikov and Oesterhelt, 1994; Bult et al., 1996); *Plasmodium* CdcATP (M96757) and *C. elegans* C06A11 and C41C4. (Wilson et al., 1994).

* Corresponding author. Tel.: +44 1703 594389; Fax: +44 1703 594269; E-mail: ds@soton.ac.uk

Abbreviations: aa, amino acid(s); AAA, ATPases associated with diverse cellular activities; AFG2, AFG2 protein from *Saccharomyces cerevisiae*; Atcdc48p, AtCDC48 protein from *Arabidopsis*; bp, basepair(s); Cdc48p, cdc48 protein from *Saccharomyces cerevisiae*; CdcH, CdcH archaebacterial protein; C06A11, C06A11 protein from *C. elegans*; C41C4.8, C41C4.8 protein from *C. elegans*; CDCATP (M96757), CDCATP protein from *Plasmodium*; CNS, central nervous system; ER, endoplasmic reticulum; kb, kilobase(s); L0919-chrXII, L0919-chrXII protein from *Saccharomyces cerevisiae*; *LacZ*, bacterial β-galactosidase gene; mCDC48, mCDC48 archaebacterial protein; NB(s), neuroblast; ORF, open reading frame; p97, p97 protein from *Xenopus*; RACE, rapid amplification of cDNA ends; SAV, SAV archaebacterial protein; sVCP, VCP from soy bean; VCP, porcine VCP protein; YTA7, YTA7 protein from *Saccharomyces cerevisiae*.

PII S0378-1119(97)00652-5

The Cdc48p/VCP subfamily is typical of the AAA superfamily in that its members are characterised primarily by the presence of a duplication of the highly conserved AAA module. Relative to this, the N- and C-terminal regions are often poorly conserved and variable in length. In addition to structural aspects the subfamily also exhibits conserved cellular functions. Some subfamily members are known to be involved in homotypic membrane fusion events. For example, transitional ER-ATPase (TERA) is required for the budding of vesicles from transitional endoplasmic reticulum (Zhang et al., 1994) and p97 plays a role in the reassembly of the post-mitotic Golgi (Acharya et al., 1995; Rabouille et al., 1995). There is also evidence for their involvement in the cell cycle. Mutations in yeast *CDC48* lead to a late arrest of the cell cycle (Fröhlich et al., 1991) and this phenotype can be rescued by heterologous expression of the *Arabidopsis* homologue of cdc48p (Feiler et al., 1995).

Here we present the cloning and characterisation of a *Drosophila* gene encoding a new member of the Cdc48p/VCP subfamily of AAA proteins.

2. Materials and methods

2.1. *Isolation of* smid *mutations*

The original mutation was isolated from a screen of GAL4 enhancer-trap lines and the details published elsewhere (Smith and Shepherd, 1996). Seven excision alleles around the site of the P-element insertion *e5/e17/e41/e74/e80/e106/e122* were obtained by imprecise excision of the *pGAwB* element. All mutant stocks were maintained as balanced stocks over either TM3 or TM6b balancer chromosomes. Wild-type data are based on a single outcross of *smid* alleles to Oregon-R wild-type stock.

2.2. *Cloning and DNA sequencing*

Genomic fragments flanking the P-insertion in line C161 were plasmid rescued using the method of Bier et al. (1990). These were used to screen a genomic library constructed in the lambda vector EMBL3. Genomic clones covering the region were then used to screen a head cDNA library in the lambda vector NM1149 (Russell and Kaiser, unpublished data). Screening of lambda genomic and cDNA libraries was performed as described by Sambrook et al. (1989). Sequencing was performed using the dideoxy-chain termination method (Sanger et al., 1977). The Sequitherm cycle sequencing kit (Epicentre Technologies, Madison, WI) was used for all reactions and electrophoresis performed on a Licor 4000 automated sequencer.

2.3. *RNA isolation and Northern hybridisation*

Poly(A)$^+$ RNA was isolated using the Poly(A) Quik mRNA isolation kit (Stratagene, La Jolla, CA). Samples (10 μg) were run on 1% agarose, Mops formaldehyde denaturing gels and transferred to nitrocellulose as described by Sambrook et al. (1989). A 2.8 kb cDNA probe was prepared using the 'Ready to go' DNA labelling kit (Pharmacia Biotech, Uppsala, Sweden) and hybridised under standard conditions (Sambrook et al., 1989). Following detection, filters were stripped and re-probed with a ribosomal probe, rp49 (O'Connell and Rosbash, 1984) as a loading control.

2.4. *RACE procedures*

5′ and 3′ extensions of the 2.8 kb cDNA were obtained with RACE System kits (Gibco BRL, Gaithersburg, MD) using the following nested sets of gene-specific primers. For 5′ RACE, TACGGTGATCTTTGGAACGCG (696–717) and CAATCTTCCTCGGAACTGCTTACC (292–315) were used. For 3′ RACE, CACCATTTTGTACGTGGGTTTCC (2539–2561) and GATGAAATTGCTGCCCAAACC (2645–2665) were used.

2.5. *Southern analysis*

Genomic DNA was isolated using the DNA Extraction kit (Stratagene, La Jolla, CA). Samples (10 μg) were run on 0.7% agarose/TBE gels and transferred to nitrocellulose as described by Sambrook et al. (1989). A 2.8 kb cDNA probe was prepared as described in Section 2.3 and hybridisation was performed as described by Sambrook et al. (1989).

2.6. *Staining of tissues*

Detailed procedures have been published elsewhere (Smith and Shepherd, 1996).

2.7. *Photography*

Tissue specimens were dehydrated in glycerol and mounted in gelatin as wholemounts. Preparations were photographed on Kodak Ektachrome 160T film using a Zeiss Axioskop microscope. Images were digitised using a Nikon Coolscan and photomontages assembled using Adobe Photoshop in Macintosh computers. Images were adjusted for contrast and brightness only.

3. Results and discussion

3.1. *Enhancer-trap line C161*

As part of a screen of GAL4 enhancer-trap lines designed to identify those in which reporter gene expres-

sion is restricted to sensory neurons we uncovered one (C161) which was of particular interest. In addition to a significant expression pattern, this line also had a mutant phenotype associated with the insert. Larvae homozygous for the insertion hatch normally but exhibit poor locomotion and become developmentally arrested as second instar larvae. Homozygous larvae can survive as second instar larvae for up to 8 weeks without further development before dying. To confirm that the phenotype was due to the insertion we mobilised the P-element. Of the 123 excision lines established 116 produced viable adults which appeared wild-type in all respects. The other seven lines had lethal mutations with the same phenotype as the original P-insertion line. It is presumed that these lethal lines represent imprecise excision of the P-element and are therefore deletion mutations. The insert in this line has already been mapped to band 66A on chromosome 3 (Shepherd and Smith, 1996). The seven lethal excision lines and original insert all failed to complement one another, nor a deficiency from 65F3 to 66B9 in the *pebble* region (Df(3L)pblX1). The lethality observed is, therefore, most likely the result of the insertion. The inability to produce a more severe phenotype than the orignal allele suggests that this is a null mutation.

3.2. *Reporter gene expression in line C161 reveals neuroblasts*

The GAL4-induced *lacZ* expression pattern of line C161 in wandering third instar larvae and pupae has already been described showing expression restricted exclusively to a subset of sensory neurons (Smith and Shepherd, 1996; Shepherd and Smith, 1996). The expression pattern in early to mid stage larvae (< 100 h after hatching), however, has not been previously described. At these early larval stages *lacZ* expression is still seen in sensory neurons (not shown) but is also seen in neuroblasts (NBs) in the cephalic and thoracic neuromeres of the central nervous system (Fig. 1A). Expression is also seen in the imaginal disks (Fig. 1C). Reporter gene expression was not detected in embryonic stages (not shown).

In larvae homozygous for the insert, the pattern of reporter gene expression shows significant differences from the heterozygote. Reporter gene expression still reveals NBs but the number of NBs detectable is reduced. During early stages of larval development (24–60 h), a normal complement of *lacZ* expressing NBs can be seen (not shown). By 70 h, however, the number of visible NBs is decreased massively, for example, the preparation in Fig. 1B shows only four detectable NBs in the cephalic neuromeres. Examination of numerous preparations ($n > 20$) suggests that the distribution of detectable NBs in the mutation is random. Furthermore, the NBs in the homozygotes also have

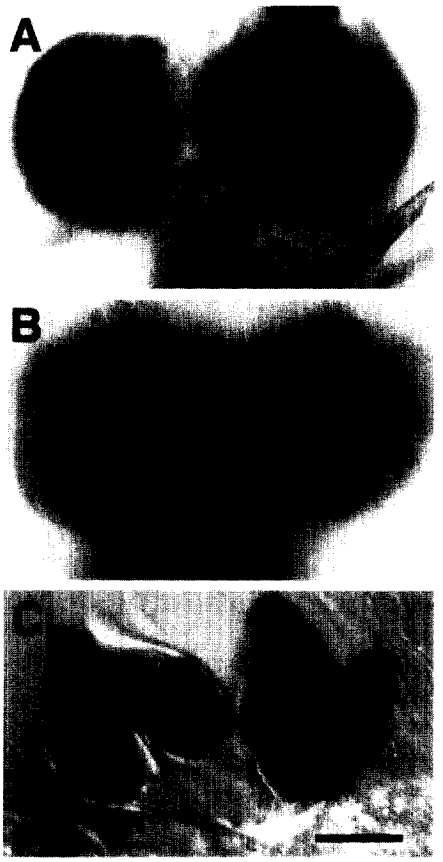

Fig. 1. *lacZ* Expression in line C161 revealed immunochemically (**A and B**) and with X-gal (**C**). (**A**) Wholemount larval brain showing *lacZ* expression in a wild-type larval background, 70 h after hatching which reveals arrays of neuroblasts (NBs) and progeny in cephalic neuromeres. (**B**) Wholemount larval brain showing *lacZ* expression in *smid* mutant background, 70 h after hatching reveals NBs but, in contrast to wild-type (**A**) there is a significant decrease in the number of NBs. In this specimen only four clusters of expressing cells are seen, each representing a single NB and its progeny. The morphology of these clusters is different from wild-type. (**C**) *lacZ* Expression in wild-type third instar leg imaginal disks reveals expression throughout. Note that (**A**) and (**B**) are not to the same scale. Scale: (**A**) 30 μm; (**B**) and (**C**) 15 μm.

abnormal morphology. This contrasts with heterozygous larvae which at the same developmental stage show a large number of expressing NBs in the cephalic neuromeres (Fig. 1A). There are no changes evident in the sensory neurons. These results clearly suggest that the phenotype is likely to be due to some failure in the functioning of the NBs.

Examination of the larval CNS shows that the CNS of C161 homozygous larvae is markedly smaller than that of an identically aged wild-type larva (Fig. 2A and Fig. 2B); identical observations were made of the lethal excision alleles (not shown). To confirm this a detailed analysis of the CNS size at different stages of larval development in both *smid* mutants and wild-type animals was undertaken (Fig. 2C and Fig. 2D). During the earliest stages of larval development (0–24 h after hatch-

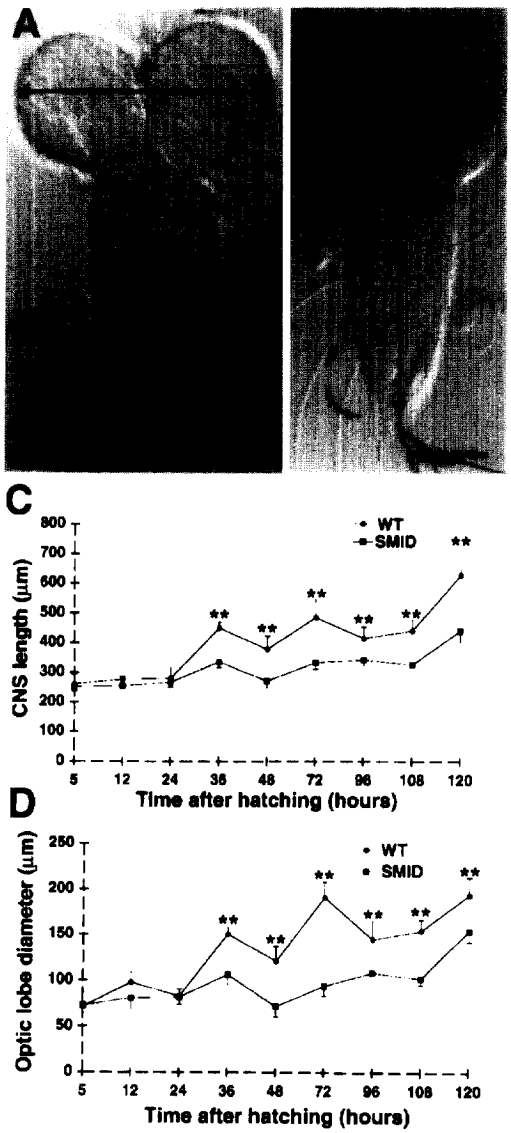

Fig. 2. Analysis of brain size in wild-type and *smid* larvae. **(A)** Wholemount of a wild-type larva 108 h after hatching. **(B)** Wholemount of an identically treated and aged *smid* mutant larva. Arrows indicate the dimensions measured. **(C)** and **(D)** Graphical presentation of the quantitative measurements of brain size through development. **(C)** represents the CNS length and **(D)** the diameter of the optic lobes. Each point is the average of at least five specimens ±1 standard deviation. ** indicates statistical significance (*t*-test. $P<0.001$). Scale 50 μm.

ing) the CNS of wild-type and homozygous C161 larvae are comparable in size and indistinguishable. Later in development (post 24 h), significant differences in the size of the CNS become apparent. Over the next 96 h both wild-type and C161 nervous systems increase in size but the rate and extent of the growth of the wild-type CNS is far greater than its C161 equivalent. As early as 36 h after hatching the difference is such that the wild-type CNS is significantly larger than the equivalently aged CNS from C161 larvae ($P<0.001$). The wild-type CNS can be seen to be significantly larger than the C161 CNS at all subsequent stages of larval development ($P<0.001$, *t*-test). It is for this reason that we have named the mutated gene *smallminded* (*smid*). This observation supports the hypothesis that mutations of *smid* disrupt the normal patterns of neurogenesis.

3.3. *Isolation of the* smid *gene*

Having identified a potentially interesting phenotype, we isolated the affected gene. Plasmid rescue (Bier et al., 1990) was employed to isolate 3.5 and 1.3 kb of genomic DNA flanking the P-element insert (Fig. 3). These fragments were used to obtain larger genomic clones covering the region. Northern blotting indicated that one of these hybridised with a 3.2 kb transcript present in RNA isolated from adult heads (data not shown) and was used to screen a *Drosophila* head cDNA library, yielding a 2.8 kb cDNA. The C161 P-insertion has previously been mapped to 66A on the third chromosome (Smith and Shepherd, 1996) and we have mapped this cDNA to the same position. In a high stringency Southern analysis (Fig. 4) this cDNA hybridised to a single 10 kb band in *Kpn*I digested wild-type (Oregon-R) genomic DNA. Probing of *Kpn*I-digested DNA from the C161 heterozygotes detected the same 10 kb band and a smaller band representing the P-insertion allele which contains a *Kpn*I site present within that pGawB element. Evidently these data indicated the *smid* is present as a single copy in the *Drosophila* genome.

3.4. *Sequencing of the* smid *cDNA*

The cDNA was sequenced on both strands and RACE procedures used to extend the 5′ terminus and complete the 3′ terminus, giving a cDNA sequence of 3184 bp containing an open reading frame of 2841 bp (Fig. 3). This was flanked by 285 bp of 3′ untranslated sequence containing a polyadenylation signal consensus sequence, beginning 207 bp downstream of the stop codon

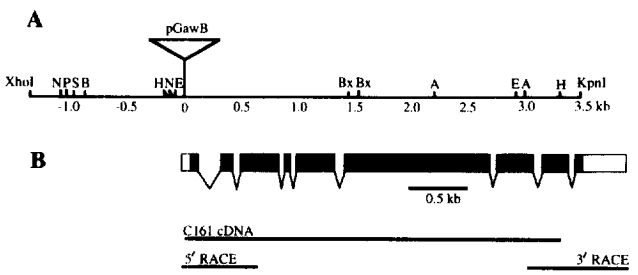

Fig. 3. Structure of the *smallminded* locus. **(A)** Restriction map of the rescued 1.3 kb *Xho*I and 3.5 kb *Kpn*I genomic DNA fragments flanking the P-element insertion (co-ordinate 0). The insertion is between bases 2 and 3 of the composite cDNA sequence. **(B)** Structure of the composite *smid* cDNA. Filled boxes indicate predicted protein coding regions. The direction of transcription is from left to right. Restriction sites: A, *Ava*I; B, *Bam*HI; Bx, *Bst*XI; E, *Eco*RV; H, *Hin*dIII; N, *Nco*I; P, *Pst*I; S, *Sal*I.

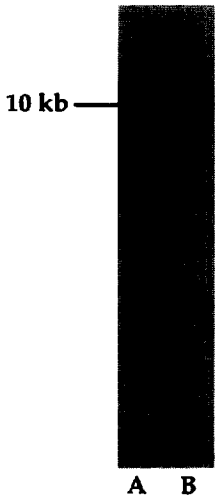

Fig. 4. Southern blot of adult wild-type **(A)** and C161 heterozygote **(B)** genomic DNA. In each case 10 μg of *Kpn*I-digested DNA was probed with the *smid* cDNA.

(Fig. 5). When the 3.5 kb genomic fragment was sequenced it was found to containg all but the last 16 bases of this ORF, interspersed with eight small introns (Fig. 3). Alignment of the genomic sequences flanking the P-element with that of the cDNA indicated that the P-insertion was close to the transcriptional start site, within the 5′ untranslated region of the transcription unit and thus would be expected to produce a null allele. The first methionine codon with the ORF was assigned as the initiation codon, since none of the methionine codons present formed part of the consensus sequence said to be most favoured by the eukaryotic ribosome (Kozak, 1989). With this assignment we predict a protein product of 104 kDa from *smid*.

3.5. Sequence analysis of the predicted gene product

Analysis of the predicted primary structure of the protein encoded by *smid* (Fig. 5) indicates that it is a member of the AAA superfamily of proteins and contains a tandem duplication of the characteristic AAA module, each of which is conserved within a region of 170 amino acid residues (290–459 and 702–873). Alignment of these regions shows that they share 38% identical and 52% similar residues. The N-terminal region of each domain contains Motif-A (GPPGCGKT in both cases) and Motif B (VLFIDE and VIFFDE) nucleotide binding consensus sequences (Walker et al., 1982; Gorbalenya and Koonin, 1989). This is followed by the AAA protein family signature (Fig. 5). The first domain contains two deviations from the consensus, whereas the second domain meets the consensus in full.

A search for related sequences using the BLAST algorithm (Altschul et al., 1990) shows Smid to be a new member of the Cdc48p/VCP subfamily of AAA

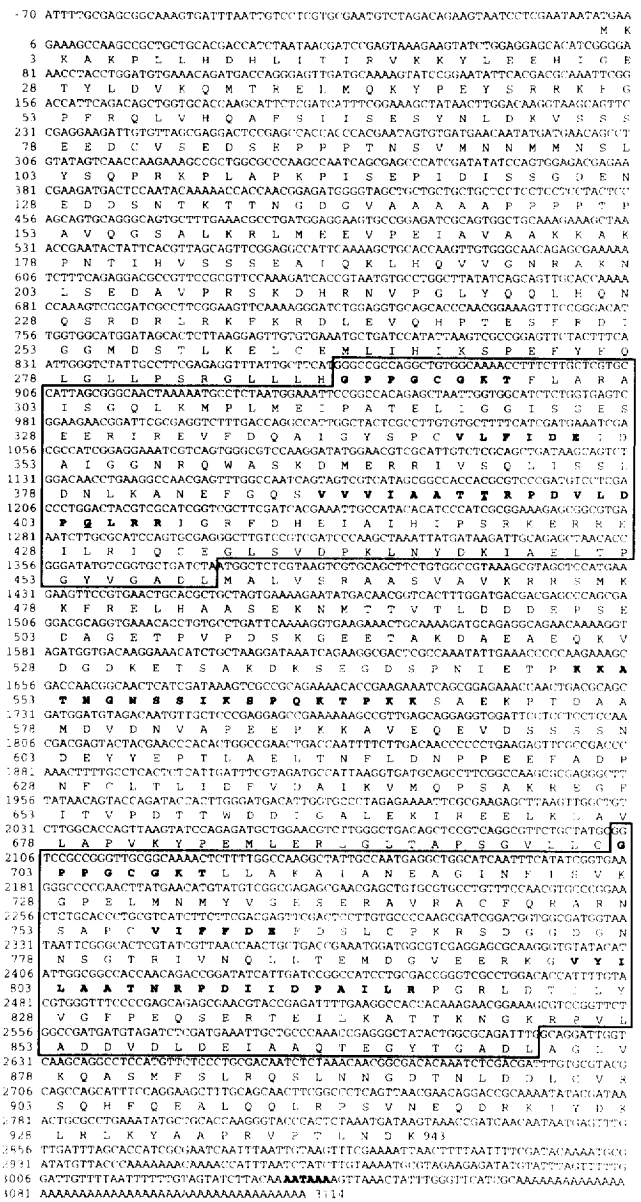

Fig. 5. Nucleotide sequence of *smallminded* and predicted primary structure of its gene product. Boxed regions contain the duplicated AAA modules with nucleotide binding consensus motifs and AAA signature in bold. Deviations from consensus are underlined. Putative nuclear localisation and polyadenylation signals are also shown in bold.

proteins. It is most closely related to the final AAA family member to be identified in *S. cerevisiae*, L0919-chroXII; SAV, CdcH and mCdc48p proteins from the archaebacteria *S. acidocaldarus*, *H. salinarum* and *Methanococcus jannaschii*, respectively; porcine VCP and its homologues from mouse, rat, *Xenopus*, *Arabidopsis*, *Glycine max*, and *S. cerevisiae*. Alignment of Smid with these sequences shows that they all contain a duplication of the characteristic AAA module (Fig. 6). which is the definitive feature of this subfamily. Overall, Smid shares 33% identical residues with its closest

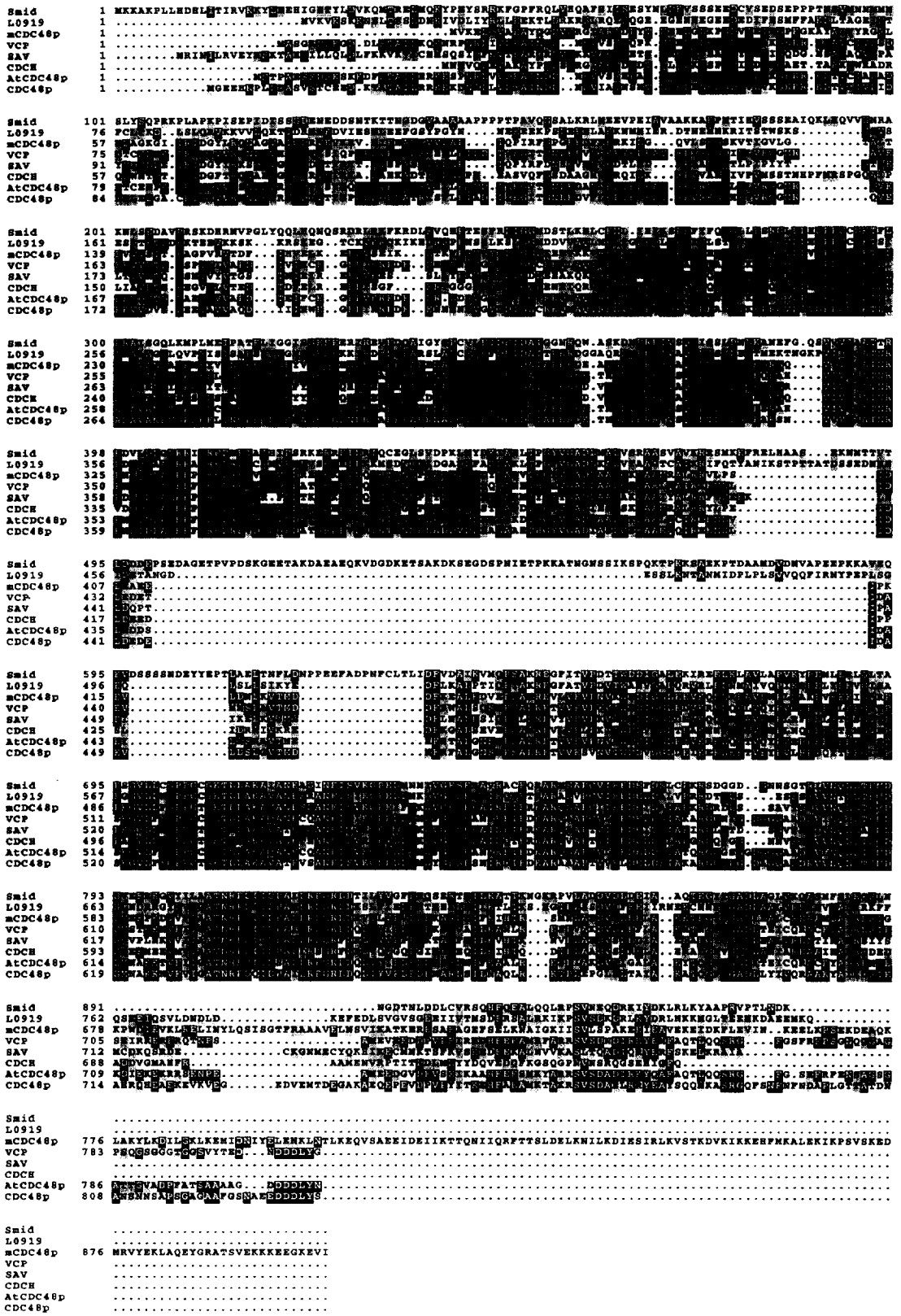

Fig. 6. Protein sequence alignment of Smid, L0919, mCdc48p, VCP. Sav, CdcH, Atcdc48p and Cdc48p. Identical residues present in two or more sequences are boxed in black. Grey shading denotes similar residues.

relative, L0919-chrXII, and 25–29% for the other proteins listed. The most highly conserved regions, however, are found in and around the nucleotide binding domains. Here Smid displays similar high levels of identity with the proteins listed and in each case the second AAA module shows greater conservation than the first. The following identities are for the first and second modules, respectively (regions 290–459 and 702–873 of Smid); L0919-chrXII 48.3 and 56.3%, mCdc48p 44.7 and 59.3%; Sav 43.5 and 58.1%; CdcH 44.1 and 57%; pVCP 42.4 and 59.5%; sVCP 43.5 and 57.8%; Cdc48p 42.9 and 53.2%; Atcdc48p 42.4 and 57.5%. Conservation of this module also exists between members of different subfamilies. For example, the C-terminal module of Smid shares 49.7% identical residues with that of the human biogenesis disorder protein Pxaaa1p and 47.4% with the single module of the *E. coli* metalloprotease Ftsh.

Such a high degree of structural conservation indicates a common function for the module in all AAA proteins, suggesting that other regions are the primary determinants of the broad range of functional specificities observed for the family members. In this respect it is interesting to note that the N- and C-termini are poorly conserved. Here Smid and L0919-chrXII share only 14.8% and 24.7% identical residues, respectively, compared with 10.4% and 8.7% for Smid and Cdc48p. To illustrate the evolutionary divergence of Smid from Cdc48p and its vertebrate homologues, it is notable that yeast Cdc48p and porcine VCP share 61% and 35% identical residues in these same regions. This difference may simply indicate a similar divergence of unidentified interacting proteins in *Drosophila* or perhaps that Cdc48p and Smid perform the same roles via differing molecular mechanisms.

The most striking difference between Smid and the other members of the AAA protein family is found in the region separating its AAA modules. Most family members have around 100 aa separating the conserved domains (as defined in this paper). The only significant deviations from this are found in L0919-chrXII, which has 156 aa and Smid, with 239 aa. Here Smid has a region of 150 aa (478–634) with little relation to any of the other family members except L0919-chrXII, which shows some sequence similarity but few identical residues (Fig. 6). It shows no significant homology with any other sequences in the database.

Smid also contains four putative bipartite nuclear localisation signals (Robbins et al., 1991) in regions 2–19, 160–177, 406–426 and 550–568. The most likely candidate for a functional signal is that found within the region 550–568, KKATNGNSSIKSPQKTPKK, since this also contains four consensus sequences (TPKK, SPQK, TPKK and SAEK) for cyclin-dependent kinase (Nigg, 1993), all of which are present within, or overlap with the localisation signal itself. The phosphorylation by Cdc28p of such a site adjacent to the nuclear localisation signal of the yeast transcription factor SW15 (Moll et al., 1991) is known to be a means of regulating its entry into the nucleus. Two such potential signal sequences have also been noted in the N-terminal region of Atcdc48p, and a nuclear localisation has been reported for Atcdc48p, Cdc48p and p97 (Feiler et al., 1995; Latterich et al., 1995; Peters et al., 1990). If, like Cdc48p and Atcdc48p, Smid performs a mitotic role, one would expect its presence in the nucleus to be tightly regulated. The presence of four potential sites of phosphorylation by cyclin-dependent kinase within this region of Smid makes it a strong candidate as a functional signal for the cell-cycle regulation of the protein. A definitive assignment, however, will require both a confirmation of its phosphorylation by cyclin-dependent kinase and its mutagenesis.

As reported for other AAA proteins, numerous potential phosphorylation sites of the calmodulin kinase II, casein kinase II and protein kinase C types are also present in the Smid sequence. The amino acid composition indicates that it is a hydrophiic protein with a predicted pI of 5.13, lacking potential membrane spanning regions or signal sequences for ER import.

3.6. *Expression of* Smid

A developmental Northern analysis was performed to assess transcription of *Smid* at specific development stages (Fig. 7). Probing with the *smid* cDNA indicates that a 3.2 kb transcript is expressed at all key developmental stages: its expression being highest during embryogenesis and at lower but similar levels in larvae, pupae and adults. Northern analysis of Oregon-R wild-type and heterozygous C161 adults showed an approx. 50% reduction in the intensity of the 3.2 kb transcript detected in the heterozygote relative to the wild-type signal, demonstrating the disruption of transcription from the P-insertion allele (Fig. 7).

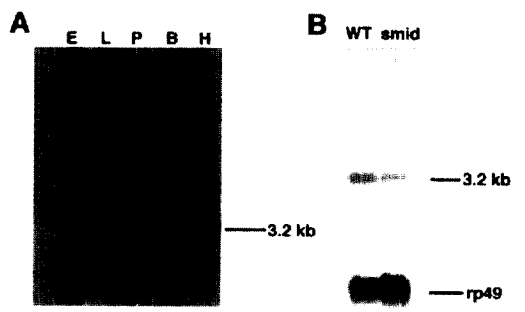

Fig. 7. **(A)** Northern blot of poly(A)$^+$ RNA extracted at various stages of development: (E) embryos, (L) larvae, (P) pupae, adult head (H) and body (B). **(B)** Northern blot of poly-A$^+$ RNA extracted from wild-type flies and flies heterozygous for the C161 P-element insertion. In each case 10 μg of RNA was used and probed with the 2.8 kb cDNA.

4. Conclusions

We have described the cloning and sequencing of a new *Drosophila* gene, *smallminded*, encoding a member of the AAA superfamily of ATPases. Flies homozygous for mutations of this locus arrest as second instar larvae and have a severely reduced CNS. The predicted amino acid sequence of Smid contains a duplication of the characteristic nucleotide binding domain and is most closely related to members of the Cdc48p/VCP subfamily, suggesting possible roles in the cell cycle and/or homotypic membrane fusion. Overall, Smid shares 33% identical residues with its closest relative, yeast L0919-chrXII and 26–29% with other members. Its most highly conserved regions are found in and around the nucleotide binding domains. *Smid* is expressed at all developmental stages and localises to 66A on chromosome 3.

Acknowledgement

The authors would like to thank Douglas Armstrong and Simon Smith for their efforts in the original screen of the Gal4 line. We are also grateful to Mike Bate and Kai Fröhlich for useful discussions. This work was supported by grants from BBSRC, The Wellcome Trust and The Hasselblad Foundation. The nucleotide sequence of the *smid* cDNA is present in the EMBL database, accession number X99207.

References

Acharya, U., Jacobs, R., Peters, J.-M., Watson, N., Farquhar, M.G., Malhotra, V., 1995. The formation of golgi stacks from vesiculated golgi membranes requires two distinct fusion events. Cell 82, 895–904.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arlt, H., Tauer, R., Feldmann, H., Neupert, W., Langer, T., 1996. The YTA10-12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. Cell 85, 875–885.

Bibikov, S., Oesterhelt, D., 1994. Light-dependent tyrosine phosphorylation of the VCP analog in archaebacterium *Halobacterium salinarum*. 6th International Conference on Retinal Proteins, Leiden, The Netherlands, p. 67.

Bier, E., Jan, L.Y., Jan, Y.N., 1990. *Rhomboid*, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila*. Genes Dev. 4, 190–203.

Bult, C.J., White, O., Olsen, G.J., Zhou, L.X., Fleischmann, R.D., Sutton, G.G. et al., 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273, 1058–1073.

Confalonieri, F., Duguet, M., 1995. A 200-amino acid ATPase module in search of a basic function. BioEssays 17, 639–650.

Confalonieri, F., Marsault, J., Duguet, M., 1994. SAV, an archaebacterial gene with extensive homology to a family of highly conserved eukaryotic ATPases. J. Mol. Biol. 235, 396–401.

Dubiel, W., Ferrell, K., Pratt, G., Rechsteiner, M., 1992. Subunit 4 of the 26S protease is a member of a novel eukaryotic ATPase family. J. Biol. Chem. 267, 22699–22702.

Egerton, M., Ashe, O.R., Chen, D., Druker, B.J., Burgess, W.H., Samelson, L.E., 1992. VCP, the mammalian homolog of cdc48, is tyrosine phosphorylated in response to T cell antigen receptor activation. EMBO J. 11, 3533–3540.

Erdmann, R., Wibel, F.F., Flessau, A., Rytka, J., Beyer, A., Fröhlich, K.-U., Kunau, W.-H., 1991. *PAS1*, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. Cell 64, 499–510.

Feiler, H.S., Desprez, T., Santoni, V., Kronenberger, J., Caboche, M., Traas, J., 1995. The higher plant *Arabidopsis thaliana* encodes a functional *CDC48* homologue which is expressed in dividing and expanding cells. EMBO J. 14, 5626–5637.

Fröhlich, K.-U., Freis, H.-W., Rüdiger, M., Erdmann, R., Botstein, D., Mecke, D., 1991. Yeast cell cycle protein Cdc48p shows a full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. J. Cell Biol. 114, 443–453.

Gorbalenya, A., Koonin, E., 1989. Viral proteins containing the purine NTP-binding sequence pattern. Nucl. Acids Res. 17, 8413–8440.

Koller, K.J., Brownstein, M.J., 1987. Use of a cDNA clone to identify a supposed precursor protein containing valosin. Nature 325, 452–545.

Kozak, M., 1989. The scaning model for translation. An update. J. Cell Biol. 108, 229–241.

Kunau, W.-H., Beyer, A., Franken, T., Gotte, K., Marzioch, M., Saidowsky, J., Skaletzrorowski, A., Wiebel, F.F., 1993. Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. Biochimie 75, 209–224.

Latterich, M., Fröhlich, K.-U., Schekman, R., 1995. Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. Cell 82, 885–893.

Moll, T., Tebb, G., Surana, U., Robitsch, H., Nasmyth, K., 1991. The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SW15. Cell 66, 743–758.

Nigg, E., 1993. Cellular substrates of $p34^{cdc2}$ and its companion cyclin-dependent kinases. Trends Cell Biol. 3, 296–301.

O'Connell, P., Rosbash, M., 1984. Sequence, structure, and codon preference of the *Drosophila* ribosomal protein-49 gene. Nucl. Acids Res. 12, 5495–5513.

Ohana, B., Moore, P.A., Ruben, S.M., Southgate, C.D., Green, M.R., Rosen, C.A., 1993. The type 1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes. Proc. Natl. Acad. Sci. USA 90, 138–142.

Peters, J.-M., Walsh, M.J., Franke, W.W., 1990. An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. EMBO J. 9, 1757–1767.

Rabouille, C., Levine, T.P., Peters, J.-M., Warren, G., 1995. An NSF-like ATPase, p97, and NSF mediate cisternal regrowth from mitotic golgi fragments. Cell 82, 905–914.

Robbins, J., Dilworth, S., Laskey, R., Dingwall, C., 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64, 615–623.

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

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Schnall, R., Mannhaupt, G., Stucka, R., Tauer, R., Ehnle, S., Schwarzlose, C., Vetter, I., Feldmann, H., 1994. Identification of a set of yeast genes coding for a novel family of putative ATPases with high similarity to constituents of 26S protease complex. Yeast 10, 1141–1155.

Shepherd, D., Smith, S.A., 1996. Central projections of persistent larval sensory neurons prefigure adult sensory pathways in the CNS of *Drosophila*. Development 122, 2375–2384.

Shi, J., Dixon, R., Gonzales, R., Kjellbom, P., Bhattacharyya, M., 1995. Identification of cDNA clones encoding valosin-containing protein and other plant plasma membrane-associated proteins by a general immunoscreening strategy. Proc. Natl. Acad. Sci. USA 92, 4457–4461.

Smith, S.A., Shepherd, D., 1996. Central afferent projections of proprioceptive sensory neurons in *Drosophila* revealed with the enhancer-trap technique. J. Comp. Neurol. 364, 311–323.

Thorsness, P.E., White, K.H., Ong, W.-C., 1993. AFG2, an essential gene in yeast, encodes a new member of the Sec18p, Pas1p, TBP-1 family of putative ATPases. Yeast 9, 1267–1271.

Walker, J.E., Saraste, M.J., Runswick, J.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATPase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

Whiteheart, S.W., Rossnagel, K., Buhrow, S.A., Brunner, M., Jaenicke, R., Rothman, J.E., 1994. *N*-ethylmaleimide-sensitive fusion protein: a trimeric ATPase whose hydrolysis of ATP is required for membrane fusion. J. Cell Biol. 126, 945–954.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Burton, J. et al., 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 368, 32–38.

Zhang, L., Ashendel, C.L., Becker, G.W., Morre, D.J., 1994. Isolation and charcterization of the principle ATPase associated with transitional endoplasmic reticulum of rat liver. J. Cell Biol. 127, 1871–1883.