Gene expression pattern

Identification of maverick, a novel member of the TGF-β superfamily in Drosophila

Minh Nguyen, Louise Parker, Kavita Arora *

Department of Developmental and Cell Biology, University of California, Irvine, CA 92697, USA

Received 21 December 1999; received in revised form 17 March 2000; accepted 21 March 2000

Abstract

The transforming growth factor-β (TGF-β) superfamily of structurally related ligands regulates essential signaling pathways that control many aspects of cell behavior in organisms across the phylogenetic spectrum. Here we report the identification of maverick (mav), a gene that encodes a new member of the TGF-β superfamily in Drosophila. Phylogenetic analysis and sequence comparison suggest that Mav cannot be easily assigned to any one sub-family, since it is equally related to BMP, activin and TGF-β ligands. mav maps to the fourth chromosome and is expressed throughout development. In situ hybridization experiments reveal the presence of maternally derived mav transcript in precellular blastoderm embryos. Later in development, mav is expressed in a dynamic pattern in the developing gut, both in endodermal and visceral mesodermal cells. In adult females, high levels of mav mRNA are present in late stage egg chambers.

Keywords: Drosophila; Transforming growth factor-β; Bone morphogenetic proteins; Activin

1. Results and discussion

Based on sequence similarity the TGF-β superfamily can be subdivided into three main groups: the prototypical TGF-βs, activins and bone morphogenetic proteins (BMPs). Until recently the only representatives of the TGF-β superfamily in Drosophila were the BMP-related ligands, Decapentaplegic (Dpp), Screw (Scw) and Glass bottom boat (Gbb; Padgett et al., 1987; Wharton et al., 1991; Doctor et al., 1992; Arora et al., 1994). Genetic and functional studies have established that these ligands are involved in critical developmental events such as patterning of the body axes and determination of cell fates, as well as regulation of cell proliferation and apoptosis (Raftery and Sutherland, 1999). More recently, an activin B ortholog and Myoglianin, a ligand related to GDF8-Myostatin, were identified in Drosophila, but the biological roles of these ligands are not known (Kutty et al., 1998; Lo and Frasch, 1999). In this study we report the identification and expression pattern of Maverick (Mav), a new BMP/TGF-β related ligand in Drosophila.

A BLAST search of the Berkeley Drosophila Genome Project (BDGP) EST database for sequences sharing similarity to the carboxyl-terminal ligand domain of Scw identified two uncharacterized clones, CK00014 and CK00025, that contained the same 1.3 kb insert (Fig. 1A). Sequence analysis of CK00014 revealed that it encodes a 378 amino acid peptide with significant similarity to members of the TGF-β superfamily. We have named this protein Maverick (Mav). An approximately 1.1 kb PCR amplified genomic fragment corresponding to the open reading frame of CK00014 was used to screen a λgt10 cDNA library derived from imaginal discs. This resulted in the isolation of a 1.7 kb mav cDNA, designated 6A1. Sequence analysis revealed that 6A1 extends 440 nucleotides further 5′ than CK00014. BLAST searches with the entire 1.7 kb sequence identified six additional cDNA clones that originate 5′ to 6A1, as well as a 17.6 kb genomic clone (AC014858) containing the mav locus (Fig. 1A). The longest EST clones LD13618 and LD22618 contain a common 2.7 kb insert. Comparison of cDNA sequences with the genomic sequence indicated that the mav locus encodes at least two alternatively spliced transcripts. The transcript corresponding to cDNAs LD22618, LD13618 and LD46352 is obtained by splicing out a 108 nucleotide intron (Fig. 1A). In the alternatively spliced variant represented by 6A1, an additional 49 nucleotide intron is removed. Both introns are flanked by consensus splice donor and acceptor sites, suggesting that they represent authentic introns. Northern blot analysis of adult male and female mRNAs detected an approximately 2.8 kb band (data not shown), suggesting that the cDNA...
clones LD13618, LD22618 are close to full length, assuming the presence of a 50–100 nucleotide poly(A) tail. Primer extension analysis using RNA from adult females supports this assertion, since we detected a prominent transcription start site 18 nucleotides upstream of the first nucleotide of LD13618 (Fig. 1B). The presence of a single band on the Northern blot suggests that the alternatively spliced mav transcripts may be of similar length, or that the transcript corresponding to 6A1 is of low abundance.

In situ hybridization indicates that mav is located on the fourth chromosome. Cross-hybridization to a cosmid contig generated by the Canadian Drosophila Genome Project allowed a more precise localization of mav to the cytogenetic region 102C. Given that chromosome four contains
only 2% of the Drosophila euchromatin, it is striking that mav is the third member of the TGF-β superfamily to map to the same region, the other two being myoglianin (102C) and dActivin (102F; Kutty et al., 1998; Lo and Frasch, 1999).

The longest mav isoform (Mav1) encodes a putative protein 701 amino acids in length (Fig. 1C). Hydroxyapatite analysis showed that the first in-frame Methionine is followed by a stretch of hydrophobic amino acids indicative of a signal sequence. In isoform 2 represented by 6A1, initiation of translation would occur at Met 269 and retain the same reading frame, to yield a predicted protein of 433 amino acids (Fig. 1C). Mav2 lacks a signal sequence suggesting that this form of the protein may not be secreted. The consensus multibasic proteolytic cleavage site (RKDK in residues 586-89 in Mav1), would generate a mature ligand of 112 amino acids (Massagué, 1998).

Sequence comparisons and phylogenetic analysis of the ligand domain of Mav suggests that this protein cannot be easily assigned to either the TGF-β, the BMP or the activin subfamilies (Fig. 2). Mav shows the highest sequence conservation with the BMP related ligands, human BMP3 (32% identity), human GDF10 (31% identity), and mouse Nodal (29% identity). Interestingly, Mav is only slightly less related to human TGF-β2 and Activin A (28 and 23% identity respectively). Among the Drosophila ligands, Mav shares the highest sequence identity with Myoglianin (28%), and the BMPs, Dpp and Gbb (27%), followed by Sccw (21%). Mav is only 19% identical to dActivin. Given that phylogenetic analysis places Mav in a cluster of BMP ligands, it is interesting that the putative Mav ligand domain contains nine invariant cysteine residues that are typical of TGF-β and activin ligands (see Fig. 1C; Daopin et al., 1992; Schlunegger and Grütter, 1992). One possible interpretation is that Mav is part of an emerging group of BMP-like ligands that contain nine cysteines, such as mammalian GDF8, GDF11 and GDF15, as well as C. elegans Daf-7 (Ren et al., 1993; Letsou et al., 1995; Brummel et al., 1999), raising the possibility that they are involved in Mav signaling.

In conclusion, we have identified a novel member of the TGF-β superfamily in Drosophila. Phylogenetic analysis and sequence comparison precludes the assignment of Mav to a specific subfamily. Thus, Mav may be an orphan molecule or the prototypic member of a new subfamily of ligands.

2. Materials and methods

2.1. Cloning and sequencing of mav cDNA

Primers QK3 (5’CCTACTACGGCCGAA3’) and QK4 (5’AGAATAATCGGTAGAAG3’), flanking the ORF of the EST CK00014 were used to amplify a genomic DNA

Fig. 1. (A) Molecular organization of the mav genomic region and cDNAs. A restriction map of the region containing the mav locus is shown. (C, Clal; E, EcoRI; H, HindIII; P, PstI). The locations of primers used are marked. The cDNAs analyzed in this study are shown to scale. The closed boxes denote an ORF, the 49 nucleotides comprising the alternatively spliced intron are underlined. A potential poly adenylation signal and a poly(A) addition site are in bold letters.

The mav genomic region contains a transcription start site (marked with an arrow), 18 nucleotides upstream of the start of the LD13618 cDNA. The first four lanes on the left indicate a DNA sequencing ladder using the same primer. (C) Sequence of the near full-length mav cDNA and deduced amino acid sequence (Genbank accession no. AF252386). The first methionine of Mav1 (Met 1) and the hydrophobic stretch of amino acid residues following it are underlined. The proteolytic site (KDDK, residues 586–589) is boxed and the nine invariant cysteines are encircled. The first in-frame ATG of the isoform Mav 2 (Met 269) is highlighted in gray. The 49 nucleotides comprising the alternatively spliced intron are underlined. A potential polyadenylation signal and a poly(A) + addition site are in bold letters.
fragment for cDNA library screens. Clone 6A1 as well as two shorter clones were obtained from 200,000 plaques screened. cDNA clones LD22618, LD13618 and LD46352, were obtained from Genome Systems. The entire mav cDNA sequence was derived from sequencing cDNAs 6A1 and CK00014 in their entirety, and the 5' sequence of LD13618 past the region of overlap with 6A1. All 5 cDNAs have identical 3' ends although they differed in the length of the poly(A) + tail. Restriction mapping and sequencing of PCR amplified DNA was used to confirm that LD22618, LD13618 and LD46352, contain sequences corresponding to the first intron.

2.2. Primer extension analysis

Primer extension analysis was performed as described in Stathakis et al. (1999) using the primer mavPE (5'GATCACAGTTTCCGTGAATGTG3') located 68 nucleotides from the 5' end of LD13618.
2.3. RT-PCR analysis

Poly(A)^+ RNA was isolated from yw flies using RNAeasy columns and the oligotex mRNA Midi Kit (Qiagen), and cDNA prepared with the First Strand cDNA Synthesis Kit (Amersham Pharmacia Biotech). The Mav specific primers QK3 and QK4 amplified a 1068 bp fragment present in both alternatively spliced forms, and were used to detect overall levels of mav expression. Primers PACAI1 (5’CCAGGGATTTGCGTGCAACTGCTGGTGCTATTC3’) and PACAT2 (5’ATTTGAAAGGCTCATGCTGCTACGC3’) flanking an intron in...
\(\alpha\)-catenin (Genbank accession no. D13964) were used both as a positive control and to rule out genomic DNA contamination.

### 2.4. In situ hybridization

Embryos and ovaries were collected from \(yw\) flies. DIG-labeled SP6 sense and T7 anti-sense riboprobes were generated from Mav cDNA clone 6A1. In situ hybridization was performed as described in Tautz and Pfeifle (1989), except that hybridization was carried out at 55\(^\circ\)C for 60 h, followed by 6\(\times\)12 h washes in hybridization buffer, also at 55\(^\circ\)C. Embryos and ovarioles were mounted in canada balsam/methyl salicylate (GMM) for observation and photography. In situ hybridization to salivary gland chromosomes was carried out using standard protocols.

### Acknowledgements

We would like to thank the Canadian Drosophila Genome Project for help with mapping the \(mav\) gene to the fourth chromosome contig. We are grateful to Dean Stathakis and Rahul Warrior for their thoughtful contributions to the work presented in this manuscript. This research was supported by an NIH-RO1 grant to K.A. (GM55442) and a Searle Scholar Award from the Kinship Foundation.

### References


Ren, P., Lim, C., Johnsen, R., Albert, P.S., Pilgrim, D., Riddle, D.L., 1996. Control of \(C.\) \(e\)\(g\)\(e\)\(a\)\(n\)\(s\) larval development by neuronal expression of a TGF\(\beta\) homolog. Science 274, 1389–1391.


