The *Drosophila fish-hook* gene encodes a HMG domain protein essential for segmentation and CNS development

Patricia A. Nambu and John R. Nambu

1Biology Department, 2Molecular and Cellular Biology Program, 3Program for Neuroscience and Behavior, University of Massachusetts, Amherst, MA 01003, USA

*Author for correspondence

SUMMARY

We describe the isolation and analysis of the *Drosophila fish-hook* (*fish*) gene, which encodes a novel member of the SOX subgroup of High Mobility Group (HMG) domain proteins that exhibit similarity to the mammalian testis determining factor, SRY. The *fish* gene is initially expressed in a pair-rule-like pattern which is rapidly replaced by strong neuroectoderm expression. *fish* null mutants exhibit severe segmentation defects, including loss and/or fusion of abdominal denticle belts and stripe-specific defects in pair-rule and segment polarity gene expression. *fish* mutant embryos also exhibit loss of specific neurons, fusion of adjacent ventral nerve cord ganglia and aberrant axon scaffold organization. These results indicate an essential role for *fish* in anterior/posterior pattern formation and nervous system development, and suggest a potential function in modulating the activities of gap and pair-rule proteins.

Key words: *Drosophila, fish-hook*, HMG domain, SOX protein, segmentation, CNS

INTRODUCTION

The process of segmentation in *Drosophila* has served as a powerful paradigm for understanding early embryonic pattern formation and specification of cell fates. This process is mediated by a complex cascade of gene regulatory and cell signaling events (reviewed in Martinez Arias, 1993; Pankratz and Jäckle 1993), which ultimately results in establishment of distinct anterior/posterior polarity both along the entire length of the embryo and within each segmental unit. Transcription regulation plays a major role in the generation of segmental periodicity, which is first indicated by stripes of pair-rule gene expression. In this regard, the maternal effect gene, *bicoid*, as well as nearly all of the gap and pair-rule genes encode transcription factors, such as homeodomain, zinc finger, basic-helix-loop-helix or basic-leucine zipper proteins (reviewed in Pankratz and Jäckle, 1993). Recent studies suggest that the stripes of most or all pair-rule genes are generated through distinct mechanisms in each segment, with direct positive and negative transcriptional regulatory input from maternal effect and gap genes (Gutjahr et al., 1993; Yu and Pick 1995). Subsequent refinement and stabilization of stripes of pair-rule gene expression involves both auto- and cross-regulatory interactions among pair-rule genes (reviewed in Pankratz and Jäckle, 1993). These complex regulatory networks are reflected in the organization of pair-rule gene regulatory regions, which typically contain multiple copies of distinct transcription factor binding sites dispersed over large DNA regions.

A key issue in understanding the segmentation process is how these many regulatory interactions are coordinated. One important mechanism is likely to be the regulation of chromatin structure to facilitate the assembly and stabilization of multiple protein/DNA and protein/protein complexes. Although this process is in general not well characterized, two clear examples are the members of the *polycomb* and *trithorax* gene families, which encode disparate types of proteins that form complexes capable of altering chromatin structure (reviewed in Simon, 1995). These ubiquitously expressed genes act as positive and negative regulators essential for proper maintenance of homeotic gene expression (e.g. Peifer and Wieschaus, 1990) and mutations in some of these genes also influence the expression of other classes of segmentation genes (e.g. McKeon et al., 1994).

Members of the High Mobility Group (HMG) protein family also regulate gene transcription through modulating chromatin structure (see Landsman and Bustin, 1993; Grosschedl et al., 1994). These proteins contain one or more HMG domains, a discrete ~80 amino acid DNA-binding structure (Jantzen et al., 1994). One major subgroup of HMG proteins includes the mammalian testis determining factor, SRY (Koopman et al., 1990) and a large number of related SOX proteins that share at least 60% amino acid identity to the HMG domain of SRY (Laudet et al., 1993). SRY and all SOX proteins have a single HMG domain and exhibit discrete ~80 amino acid DNA-binding properties, and diverse tissue specific expression patterns (e.g. Harley et al., 1992; Denny et al., 1992; Hosking et al., 1995). The developmental functions and downstream target genes of these SOX proteins are just beginning to be defined. SOX-9 has been shown to function in mammalian bone formation and sex determination, as SOX-9 mutations underlie the genetic disorder, campomelic dysplasia (Foster et al., 1994). SOX-4 mediates activation of the T cell receptor
and CD-2 genes in lymphoid cells (van de Wetering et al., 1993; Wotton et al., 1995), while SOX-2 is strongly expressed in the developing neuroepithelium of vertebrate embryos (Uwanogho et al., 1995) and regulates gene expression in developing lens cells (Kamachi et al., 1995). The strong DNA bending properties of SRY and SOX proteins (Giese et al., 1992; Harley et al., 1992; Connor et al., 1994), as well as the inability of several SOX proteins to directly transactivate transcription of target genes (Kamachi et al., 1995; Yuan et al., 1995), suggest that they may provide important architectural functions in assembly and stabilization of transcription factor/DNA complexes.

In this study, we describe the isolation and analysis of a Drosophila Sox gene that is required for segmentation and nervous system formation. We have named this gene fish-hook (fish), based on the appearance of its expression pattern in sagittal views of early germ-band-extended embryos. The fish gene is dynamically expressed, with early expression in the entire trunk of syncitial blastoderm embryos that is rapidly refined into a series of 7 irregular ectodermal stripes in the cellular blastoderm. During germ-band extension, fish expression becomes largely restricted to the developing ventral and cephalic neuroectoderm. Generation and analysis of fish null mutant strains indicated loss and/or fusion of abdominal segments coupled with stripe-specific defects in pair-rule and segment polarity gene expression. fish mutant embryos also exhibited CNS defects which included fusion of adjacent ganglia, loss of specific neurons and aberrant axon scaffold organization. The data suggest the FISH protein may modulate the actions of other transcription factors, including gap and pair-rule proteins.

MATERIALS AND METHODS

Generation of fish-hook mutants

The fish gene was initially identified via a viable P element (the PZ, a.k.a. P[rosy+ lacZ]) vector of Y. Hiromi, see Klämbt et al., 1991) enhancer trap insertion, rJ375, on the third chromosome. rJ375 was generated in a large screen by C. Goodman and colleagues (Klämbt et al., 1991). In situ hybridization of a biotinylated P element DNA probe to rJ375 larval polytene chromosomes identified a single site of hybridization at position 70D on the left arm of the third chromosome. Imprecise excision alleles were generated by crossing rJ375 flies to a P(2;70E7), Df(3L)87 chromosome in stage 5 cellular blastoderm embryos (Fig. 1A; stages defined in Campos-Ortega and Hartenstein, 1985). The cuticles were observed and photographed under dark-field optics.

Molecular techniques

Plasmid rescue of fish DNA was accomplished by digesting genomic DNA from the rJ375 strain with XbaI and Nhel, ligating under dilute conditions, and transforming into E. coli DH5α (GIBCO/BRL). A 1.4 kb HindIII/Sacl DNA fragment was isolated from the rescued DNA, labeled with 32P and used to screen 100,000 recombinant phage from a Drosophila genomic DNA library (provided by L. Schwartz). 10 clones were identified that contain overlapping DNA from the 70D region.

fish cDNA clones were isolated by screening a 9-12 hour embryonic cgt-11 library (provided by K. Zinn) using a 32P-labeled 6.0 kb SacI fish genomic DNA probe. Approximately 300,000 phage were screened and 20 clones were identified that contain cDNA insertions ranging from 0.5 kb to 2.0 kb. These cDNA insertions were all found to derive from a 2.6 kb BamHI/EcoRI fragment located within the 6.0 kb SacI fish genomic DNA fragment. The cDNA inserts were subcloned into pBS plasmid vectors (Stratagene) and purified DNA from the longest clone, 2-5, was sent to Retrogen Inc. for double-stranded DNA sequence analysis using primer walking techniques.

Northern blots were performed using poly(A)+ mRNA isolated via oligo(dT) columns from 0-5 and 0-24 hour collections of Canton-S embryos. The RNA was electrophoresed on a formaldehyde/agarose gel, transferred to Nytran filters (Schleicher and Schuell), and hybridized to a 32P-labeled fish 2.0 kb cDNA probe and, subsequently, to a 32P-labeled ribosomal protein 49 (RP49) 600 bp DNA probe (clone provided by M. Rosbash).

RESULTS

Drosophila fish-hook mutants exhibit segmentation defects

In an attempt to identify genes important for embryonic CNS development, in situ hybridization was used to detect the earliest transcription of β-galactosidase (β-gal) in several P[lacZ] enhancer trap strains. One strain, rJ375 (obtained from C. Goodman), exhibited prominent early β-gal transcription in a series of seven irregular ectodermal stripes and the procephalic region in stage 5 cellular blastoderm embryos (Fig. 1A; stages defined in Campos-Ortega and Hartenstein, 1985). At this point, insufficient β-gal protein was present to be detected via anti-β-gal immunocytochemistry (data not shown). During germ-band extension, β-gal transcripts became largely restricted to the developing neuroectoderm (Fig. 1B) and the expression of β-gal protein could first be detected. This

In situ hybridization, immunocytochemistry and larval cuticle preparations

In situ hybridizations were performed by generating digoxigenin-labeled antisense RNA probes from fish, even-skipped, wingless and β-galactosidase DNA clones using reagents from the Genius Labeling Kit (Boehringer Mannheim) and T3 and T7 RNA polymerases. Whole-mount in situ hybridizations to embryos collected from wild-type and fish mutant strains were performed as described by Tautz (1992). The stained embryos were cleared in 80% glycerol and viewed and photographed using Nomarski optics.

Immunocytochemistry was performed as described in Zhou et al. (1995) using: a mouse anti-β-gal monoclonal antibody (Promega) at 1:800 dilution, mAb BP102 (provided by Karen Jensen, Developmental Studies Hybridoma Bank) at 1:10 dilution and a rabbit anti-FITC antiserum (provided by Y. Hiromi) at 1:1000 dilution. Visualisation of antibody binding was performed using biotinylated secondary antibodies, streptavidin-horseradish peroxidase and diaminobenzidine reactions using the Vectastain ABC kit (Vector Labs). Stained embryos were dehydrated in an ethanol series, cleared in methyl salicylate and mounted in Permout (Fisher). Stained embryos were viewed and photographed using Nomarski optics.

1st instar larval cuticles were prepared from non-hatching fish mutant embryos essentially as described in Wieschaus and Nüsslein-Volhard (1985). The cuticles were observed and photographed under dark-field optics.
expression pattern hinted that the corresponding gene, which we have named \textit{fish-hook} (\textit{fish}), might play a role in segmentation and nervous system development, and led us to pursue further analyses. The rj375 strain was found to contain a single P element insertion that mapped to 70D on the third chromosome. Genetic analyses were initiated via generation of P element excision alleles and six embryonic lethal excision strains were isolated that all fail to complement each other (see Materials and Methods). Additional complementation tests to four deficiency strains, Df(3L)\textit{fz}-D21, Df(3L)\textit{fz}-GF3b, Df(3L)\textit{fz}-GS1a and Df(3L)\textit{fz}-M21, localized the site of lethality to 70D1-3 (Fig. 2).

The potential role of \textit{fish} in segmentation was assayed by larval cuticle preparations from two \textit{fish} null mutant alleles, \textit{fish}87 and \textit{fish}96 (see Materials and Methods and below). Strikingly, both alleles exhibited severe segmentation defects, including loss or fusion of abdominal denticle belts (Fig. 3A-C). Although the defects were variable, in strongly affected mutant embryos only 4 abdominal denticle belts were present. Unlike pair-rule mutants, however, this loss of denticle belts does not correspond to a segmentally repeated pattern. Typically, the A2 and A8 denticle belts were missing, the A3 denticle belt was narrowed and there were fusions between the A4-A5 and A6-A7 denticle belts. Many of the \textit{fish} mutants also exhibited defects in the organization of head structures. Alternatively, the thoracic segments appear largely unaffected.

**Fig. 3.** \textit{fish} is essential for segmentation. Larval cuticle preparations from wild-type and \textit{fish} mutant strains. (A) Wild-type 1st instar larval cuticle. Note 8 abdominal denticle belts. (B) Unhatched \textit{fish}87 mutant cuticle exhibiting severe segmentation defects in abdominal denticle belts. The A2 and A8 denticle belts are deleted (arrows), while A3 is narrowed, and A4,5 and A6,7 are fused (arrowheads). (C) Unhatched \textit{fish}96 mutant cuticle exhibiting similar segmentation defects. (D) Unhatched \textit{fish}87/Df(3L)\textit{fz}-GF3b mutant embryo showing similar segmentation defects. All views ventral with anterior to right.

\textit{fish} regulates specific stripes of pair-rule and segment polarity gene expression

In order to understand the basis of segmentation defects in \textit{fish}
mutants, we examined the expression of pair-rule and segment polarity genes. In situ hybridizations using a fushi-tarazu (ftz) antisense RNA probe and immunocytochemistry using an anti-FTZ serum indicated that ftz expression is initiated normally in fish96 mutant syncitial blastoderm embryos (data not shown), however, by the onset of cellularization specific stripes of ftz expression became altered (Fig. 4A,B). Typically, fish mutant embryos exhibited complete or partial fusions between ftz stripes 3 and 4, and a weakening of stripes 5 and 6. These results indicate that fish function is essential for either repressing or maintaining ftz expression in different segmental domains. These distinct roles are consistent with proposed stripe-specific mechanisms of ftz gene regulation (Yu and Pick, 1995). The effects of fish mutations on even-skipped (eve) expression were also examined. eve transcription is also initiated normally in fish96 mutant syncitial blastoderm embryos but, by the onset of cellularization, eve stripe 4 is severely weakened or lost (Fig. 4C,D), and eve stripes 5, 6 and 7 exhibit variable weakening and/or partial fusion.

Expression of the segment polarity gene wingless (wg) was initiated normally in the fish96 mutant blastoderm embryos; however, during gastrulation, alterations in the intensity of several stripes became evident and, in fully germ-band-extended fish mutant embryos, there were clear segment-specific defects in wg expression (Fig. 4E-H). These defects typically included a loss of wg stripes in ventral regions of the maxillary and labial segments, fusions between the A1 and A2 stripes, and fusion or reduction of other abdominal stripes. Interestingly, the loss of wg expression in head segments was independent of corresponding defects in ftz or eve expression. Several segments also exhibited aberrant spacing between wg stripes at the ventral midline (see Xiao et al., 1996), and there is an enlargement of the expression domain at posterior tip of the germ band at the site of hindgut invagination. Similar segment-specific defects were also detected in engrailed expression, as immunocytochemistry using the mAb 4D9 revealed loss and/or fusion of specific stripes in germ-band-extended embryos (data not shown).

fish-hook mutants exhibit defects in CNS development

The strong neuroectodermal expression of fish suggested a potential role in nervous system development. This was assayed via immunostaining with mAb BP102, which labels all CNS axons (Elkins et al., 1990). fish87 mutant embryos exhibited severe and variable defects in CNS organization (Fig. 5A,B). Typically, there were fusions between several adjacent

Fig. 4. fish mutants exhibit stripe-specific defects in pair-rule and segment polarity gene expression patterns. (A) Expression of FTZ protein in a stage 5 wild-type embryo as detected via anti-FTZ immunostaining. Note regular array of 7 pair-rule stripes in even parasegments. (B) Similarly stained stage 5 fish96 mutant embryo. Note fusion of FTZ stripes 3 and 4 (arrowhead) and weakened stripes 5 and 6 (arrows). (C) Expression of eve mRNA in stage 5 wild-type embryo as detected via in situ hybridization using an eve antisense RNA probe. Note regular array of pair-rule stripes in odd parasegments. (D) Similarly stained stage 5 fish96 mutant embryo. Note weakened eve stripe 4 (arrowhead) and partial fusions between stripes 5-7 (arrow). (E) Expression of wingless mRNA in stage 11 wild-type embryo as detected via in situ hybridization using a wingless antisense RNA probe. Note regularly spaced segment polarity stripes at the anterior border of each segment. (F) Similarly stained stage 11 fish96 mutant embryo. Note loss of expression in wingless stripes 2 and 3 in head segments (small arrowhead) as well as fusions of stripes 7,8 (arrow) and 13,14. There is also ectopic wingless expression in cells at the end of the germ band near the site of hindgut invagination (large arrowhead). (G) Dorsal view of a stage 11 wild-type embryo showing regular array of wingless stripes in the ventral and lateral neuroectoderm, but not in CNS midline cells. Anterior is to left. (H) Dorsal view of a similarly stained stage 11 fish96 mutant embryo. Note loss of stripes 10 and 12 in this specimen (arrows) and defects in spacing of some wingless stripes at the midline (arrowheads). (A-F) Sagittal views with anterior to left.
neuromeres resulting in 3-4 fewer ganglia as compared to wild type. In some segments, often including the thoracic segments, there was a moderate to severe narrowing of the longitudinal axon connectives and fusion of the anterior and posterior axon commissures. This phenotype is similar to that of mutations in slit, as well as single-minded and other spitz class genes, where there are defects in differentiation or migration of CNS midline cells (e.g. Klämbt et al., 1991). The CNS defects were further analyzed via immunostaining with mAb 4D9 to identify subsets of CNS cells that express the engrailed gene (Patel et al., 1989). There did not appear to be defects in the formation of engrailed-expressing cells in the CNS, which were normal in number in germ-band-extended embryos (data not shown). However, during germ-band retraction nearly all segments began to exhibit loss and/or fusion of midline and lateral engrailed-expressing CNS cells (Fig. 5C,D). Loss of CNS midline cells, including the median neuroblast and VUM neurons, is likely responsible for at least some of the axon scaffold defects observed. These results suggest that fish is essential for proper differentiation and/or survival of specific CNS cells.

The fish gene encodes a novel SOX protein

Molecular analysis of the fish gene was initiated via plasmid rescue techniques to isolate genomic DNA flanking the P element insertion. A 1.4 kb HindIII/SacI fragment was recovered and used to screen a Drosophila genomic DNA library, resulting in the isolation of 34 kb of genomic DNA from the fish locus at 70D (Fig. 6A). In situ hybridization experiments identified a 6.0 kb SacI genomic DNA fragment adjacent to the site of P element insertion that yielded a hybridization pattern closely mimicking the β-gal expression pattern from the rj375 strain. Significantly, no other DNA fragment in this region detected any embryonic expression. In Northern blots, the 6.0 SacI fragment detected a single 2 kb mRNA transcript in poly(A)+ SacI fragment detected a single 2 kb mRNA transcript in poly(A)+

sacI genomic DNA fragment. This genomic fragment was then used to isolate several fish cDNA clones from an embryonic cDNA library. These clones all exhibited embryonic expression patterns identical to each other (see below) and to the 6.0 kb genomic DNA fragment.

DNA sequence analysis was performed on the longest fish cDNA clone, 2-5, to define the structure of the fish gene product. The sequenced region of the cDNA insert is 1914 bp in length, not including an approximately 40 nucleotide poly(A) tail, the presence of which permitted unambiguous assignment of the direction of translation. The sequence encodes a predicted polypeptide of 382 amino acids (Fig. 7A). This amino acid sequence was used for BLAST and FASTA searches of protein sequence databases, and a single ~80 amino acid region was identified that shares over 60% identity to the HMG domain of the mammalian SRY protein. Higher levels of identity were found with several SRY-related SOX proteins. In particular, the FISH protein is closely related to the vertebrate SOX-2 and SOX-3 proteins. BESTFIT analysis revealed that FISH and the human SOX-2 protein share 42% overall sequence identity with several introduced gaps (data not shown). Significantly, the HMG domains of FISH and SOX-2 are 88% identical with 94% similarity. Strong similarity is also seen with the human SOX-3 protein, which shares 35% overall sequence identity with FISH. The HMG domains of these two proteins are 83% identical with 91% similarity.

Based on homology within the HMG domains of SRY and SOX proteins, 6 distinct subgroups, A-F, have been proposed (Wright et al., 1993). The sequence of its HMG domain places FISH in the B subgroup, which includes SOX-1, SOX-2, SOX-3, SOX-11, SOX-14 and SOX-19. Comparison of the HMG domains from representatives of each of the 6 SOX subgroups is presented in Fig. 7B There are 25 positions at which an invariant residue is present in the HMG domains of SRY, FISH and the 13 SOX proteins analyzed. In addition, there are several other positions where there are strong consensus residues. Interestingly, the only position where FISH differs from an otherwise invariant residue is at position 18, which, except for the most variant F subgroup, is a lysine (it is an arginine in the F subgroup), but is a glutamine in the FISH HMG domain. Outside the HMG domain, the FISH protein possesses several short alanine-, glutamine- and serine-rich stretches that may serve as transcriptional activation domains, as well as 11 copies of a consensus G/S O/S G/S M pentapeptide sequence (O= hydrophobic residue) (Fig. 7C).
The fish-hook gene is transcribed in a dynamic pattern during embryogenesis

The embryonic transcription pattern of the fish gene was examined via in situ hybridization using a probe derived from fish cDNA clone 2-5. fish transcription was first detected in cycle 13 syncitial blastoderm embryos as a wide circumferential band, corresponding to the entire trunk region (15-65% egg length) (Fig. 8A). There does not appear to be any maternally deposited fish mRNA. This trunk expression rapidly split into two subdomains and, by early cycle 14, high levels of fish transcripts were present in a narrow stripe at approximately 50% egg length and a wider stripe from about 15-30% egg length (Fig. 8B). Lower levels of fish expression persist in the intervening regions. At this time, fish transcripts also became detectable in the procephalic region. During cellularization (stage 5), fish expression was quickly refined into a series of seven irregular stripes and a strong dorsal 'saddle' (Fig. 8C). The intensity of these stripes varies, with stripes 1, 5, 6 and 7 more intense than stripes 2, 3 and 4. These stripes do not exhibit even spacing, as they both overlap and flank specific ftz stripes. Thus, fish stripe 6 corresponds to parasegment 11, between ftz stripes 5 and 6, while fish stripe 7 is coincident with ftz stripe 7 in parasegment 14 (data not shown).

During gastrulation and early germ-band extension there is a rapid change in the fish transcription pattern, as the seven ectodermal stripes diminish and are replaced by two longitudinal columns of fish expression that are approximately 4 cells wide and flank the invaginating mesoderm (Fig. 8D). Expression is maintained in the developing cephalic neuroectoderm. In stage 10 germ-band-extended embryos, medial neuroectodermal cells exhibit high levels of fish transcripts, and this expression is maintained in stage 11 embryos (Fig. 8E). During germ-band retraction, there is an overall decrease in the level of fish transcripts. Expression is detected in subsets of cells in the brain and CNS midline, the hingdut and segmentally repeated stripes of cells along the ventral epidermis (Fig. 8F). Overall, the pattern of fish transcription corresponds well to the pattern of β-gal expression in the rJ375 strain. Analysis of fish transcription in the fish excision mutants indicated that the fish alleles do not express any detectable fish mRNA transcripts (data not shown).

DISCUSSION

A novel segmentation gene

The elaboration of the Drosophila body pattern into discrete segments involves a series of transcriptional regulatory pathways that require complex yet precise interactions of regulatory proteins and target DNA sites. The fish gene appears to function parallel to or downstream of the gap genes, as defects in anterior/posterior patterning are first detected in the elaboration of pair-rule stripes. fish does not appear to be required for the initial activation of pair-rule or segment polarity genes, but is instead essential for their proper refinement by maintaining or repressing specific stripes. fish could function to modulate the activities of abdominal gap proteins, such as KRUPPEL, KNIRPS or GIANT, in their regulation of pair-rule gene expression. For example, because proper elaboration of eve stripes 4 through 6 requires the gap gene Krüppel (Frasch and Levine, 1987), the eve expression defects in fish mutants could be due to an alteration in Krüppel function. This does not appear to occur at the level of regulating early Krüppel transcription, which is normal in fish mutant blastoderm embryos (P. A. N. and J. R. N., unpublished data). fish might also modulate the function of pair-rule proteins in the refinement and stabilization of pair-rule stripes. For example, because the HAIRY protein represses ftz expression in odd parasegments (Ish-Horowicz and Pinchin 1987), the fusion of ftz stripes 3 and 4 in fish mutants could be due to a localized defect in hairy function. In this regard, it is notable that there are several hairy alleles that exhibit segment-specific defects in abdominal regions (Howard et al., 1988). Finally, because the defects in wg (and en) expression in fish mutants do not strictly correspond to the defects in pair-rule gene expression, fish may also directly modulate the function of regulatory complexes required for segment polarity gene expression. fish appears to be a member of an emerging class of novel segmentation genes that includes hopscotch and marelle, two components of the JAK/STAT signaling pathway, which act to regulate the generation of specific individual segments (Binari and Perrimon, 1994; Hou et al., 1996; Yan et al., 1996).

Role of fish-hook in CNS formation

Similar to many other segmentation genes (see Goodman and Doe, 1993), the fish gene exhibits prominent expression in cells of the developing embryonic nervous system and fish mutants exhibit severe alterations in CNS organization. fish likely plays a direct role in these processes as the CNS defects observed in fish mutants do not precisely correspond to the epidermal...
defects. Clearly this issue will need to be further examined using conditional mutants. Because fish expression is activated in medial neuroectodermal cells shortly after gastrulation, it could play an early role in nerve cell development. However, because CNS defects in fish mutants were first detected during germ-band retraction, fish may instead function in nerve cell differentiation. One issue that will be important to address is whether fish has common functions in mediating gene expression during segmentation and nervous system development, for example, by interacting with the gap gene Krüppel, which is also widely expressed in many cells of the developing CNS (Hoch et al., 1990). In addition, because the vertebrate SOX-2 protein can interact with the POU domain protein, OCT-3 to regulate FGF-4 expression, perhaps fish interacts with POU domain proteins expressed within overlapping regions of the Drosophila nervous system, such as PDM-1, PDM-2/MITI-MERE and CFl/a DRIFTER (Lloyd and Sakonju, 1991; Billin et al., 1991; Anderson et al., 1995), pdm-1, pdm-2/miti-mere and CFla/driver all appear to play important roles in specificity or differentiation of specific CNS cell types (Yang et al., 1993; Bhat and Schell, 1994; Anderson et al., 1995; Bhat et al., 1995; Yeo et al., 1995). Interestingly, the expression of the pdm-1 and pdm-2/miti-mere genes also overlap with fish in abdominal segments of blastoderm embryos (e.g. Lloyd and Sakonju, 1991; Billin et al., 1991) and expression of a dominant negative pdm-2/miti-mere transgene results in segmentation defects that include loss of A2 and A6 (Bhat and Schell 1994).

A Drosophila Sox gene

This study provides the first functional analysis of an invertebrate member of the Sox gene family and may provide a useful paradigm for analyzing the functional roles of Sox genes in specific developmental contexts. SRY and related SOX proteins regulate gene expression by acting as architectural proteins and/or classical transcription factors. Thus a crucial aspect to understanding their functions will be to characterize the mechanisms by which these proteins act and the identities of downstream target genes. SRY and SOX proteins bind to A/T AT CAAAG or AACAAT consensus sequences (Harley et al., 1992; van de Wetering et al., 1993; Connor et al., 1994; Wotton et al., 1995; Yuan et al., 1995), and the binding of SRY and SOX-5 to target DNA sites induces strong DNA bending, at angles of 85° for SRY and 73-90° for SOX-5 (Giese et al., 1992; Harley et al., 1992; Connor et al., 1994). The strong similarity between the HMGI domain of FISH and other SOX proteins suggests similar DNA-binding and transcriptional regulatory properties

Fig. 7. The fish gene encodes a novel HMG domain protein. (A) Amino acid sequence of the predicted FISH protein. The HMGI domain (bold and underlined), pentapeptide repeats (1-11), and glycin-serine-[G/S], alanine-[A], glutamine-[Q], and serine-[S] residues are denoted. (B) Pentapeptide repeats in FISH protein. (C) Comparison of HMGI domains from FISH, SRY, and representative SOX proteins. Subgroupings of SRY and SOX proteins (Wright et al., 1993) are denoted by letters (A-F). INV = positions at which there are invariant residues. Note that FISH has a glutamine (Q) at position 18 which is a lysine (K) in all the other HMGI domains except those of the divergent F subgroup.

| A | 1 | MATLSTHFNYGFHLGQAQLGLDFY implications of downstream target genes. SRY and SOX proteins bind to A/T AT CAAAG or AACAAT consensus sequences (Harley et al., 1992; van de Wetering et al., 1993; Connor et al., 1994; Wotton et al., 1995; Yuan et al., 1995), and the binding of SRY and SOX-5 to target DNA sites induces strong DNA bending, at angles of 85° for SRY and 73-90° for SOX-5 (Giese et al., 1992; Harley et al., 1992; Connor et al., 1994). The strong similarity between the HMGI domain of FISH and other SOX proteins suggests similar DNA-binding and transcriptional regulatory properties

<table>
<thead>
<tr>
<th>B</th>
<th>1) LSFGM</th>
<th>2) GSSGM</th>
<th>3) SISOG</th>
<th>4) GSAAG</th>
<th>5) GSLGS</th>
<th>6) SSILO</th>
<th>7) SSOOS</th>
<th>8) GSOOG</th>
<th>9) TSGQQ</th>
<th>10) SSPSS</th>
<th>11) SSPGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCONSIS:</td>
<td>G/S S G/S G/S G/S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/12</td>
<td>12/12</td>
<td>12/12</td>
<td>12/12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>@ = hydrophobic residue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>(A) HSRY:VKKPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B) FISH:IKRPMNAFIIWISDQRHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(C) CSX2:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(D) CSX3:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(E) MSK1:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(F) MSK2:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(G) MSK3:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(H) MSK4:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(I) MSK5:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(J) MSK6:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
<tr>
<td>(K) MSK7:IKRPMNAFIVWSDRQHAKMELNFMHMSIESQQLGQWNLTEAREKRPFTQEAQQQLKHREKHFVYKYPRRRKAK</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 7.** The fish gene encodes a novel HMG domain protein. (A) Amino acid sequence of the predicted FISH protein. The HMGI domain (bold and underlined), pentapeptide repeats (1-11), and glycin-serine-[G/S], alanine-[A], glutamine-[Q], and serine-[S] rich stretches are denoted. (B) Pentapeptide repeats in FISH protein. (C) Comparison of HMGI domains from FISH, SRY, and representative SOX proteins. Subgroupings of SRY and SOX proteins (Wright et al., 1993) are denoted by letters (A-F). INV = positions at which there are invariant residues. Note that FISH has a glutamine (Q) at position 18 which is a lysine (K) in all the other HMGI domains except those of the divergent F subgroup.**
Fig. 8, fish mRNA transcripts are dynamically expressed during embryogenesis. Embryonic in situ hybridization to wild-type embryos with an antisense RNA probe derived from fish cDNA clone 2-5. (A) Stage 5 cycle 13 blastoderm embryo. Note uniform expression of fish mRNA throughout entire trunk region of embryo. (B) Stage 5 early cycle 14 blastoderm embryo. Note splitting of trunk expression domain into 2 subdomains at approximately 50% (small arrowhead) and 15-30% egg length (arrow). Lower levels of fish mRNA are still detected in intervening regions. fish expression is also detected in the procephalic region (large arrowhead). (C) Stage 5 cellular blastoderm embryo. fish mRNA is now present in a dorsal saddle (arrow) and 7 irregular stripes that display variable spacing and intensity. (D) Stage 7 gastrulating embryo. The fish stripes have been replaced by two columns of fish expression in neuroectodermal cells flanking the invaginating mesoderm (arrowhead). (E) Stage 11 fully germ-band-extended embryo. High levels of fish mRNA are present in the medial neuroectoderm. (F) Stage 15 germ-band-retracted embryo. fish mRNA is present in subsets of cells within the ventral nerve cord (short arrow), brain (small arrowhead), pharynx (large arrowhead) and hindgut (long arrow). (A-C,F) Sagittal views with anterior to left. (D,E) Ventral views with anterior to left.

-bending capabilities. However, it is notable that the FISH HMG domain has a glutamine substitution in an otherwise invariant basic residue (generally a lysine). A similar glutamine versus lysine difference in the recognition helix of several homeodomain proteins, including BICOID, FTZ and ORTHODENTICLE, appears to be responsible for determining DNA-binding specificity (Driever et al., 1989; Schier and Gehring, 1992; Pankratz and Jäckle, 1993). Thus FISH could recognize target sequences distinct from those bound by other SOX proteins. Identification of these binding sites should prove useful in determining whether conformational changes in DNA that may be brought about by FISH binding are important for regulating pair-rule and segment polarity gene expression, perhaps by facilitating interactions between distantly located complexes of gap and pair-rule proteins bound to DNA.

The fish gene has been independently isolated by Michael Ashburner’s group and they report similar findings on protein sequence, transcription patterns and mutant segmentation phenotypes (see Russell et al., 1996, this issue of Development). In addition, they have generated data indicating that fish mutants are allelic to Dichaete. We have found that all of the fish mutant alleles that we have generated are lethal over Dichaete. As Dichaete appears to be a dominant regulatory mutant defined by inversion breakpoints, we propose referring to Dichaete mutants as fish alleles. The fish cDNA clone sequence has been deposited in GenBank, accession number U68056.

We are very grateful to Dr Corey Goodman for initially providing us with rJ375 and other enhancer trap strains. In addition, we thank Drs Yash Hiromi, Tony Ip, Karen Jensen, Judy Lengyel, Roel Nusse, Michael Rosbash, Larry Schwartz and Kai Zinn for providing us with DNA clones, antibodies, genomic DNA and cDNA libraries, Kathy Matthews for providing fly stocks, and Jun Xu for assistance with mapping fish genomic DNA clones. The critical reading of the manuscript by Vivian Budnik, Tony Ip and Hong Xiao is greatly appreciated. This work was supported by March of Dimes Basil O’Connor and Phase-In Awards and NIH grant R29NS31225 to J. R. N.

REFERENCES


Driever, W., Thoma, G. and Nüsslein-Volhard, C. (1989). Determination of


