

Dorsal–ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the *short gastrulation* gene

Vincent François, Mark Solloway, Jason W. O'Neill, John Emery, and Ethan Bier¹

Department of Biology and Center for Molecular Genetics, University of California San Diego, La Jolla, California 92093 USA

Pattern formation in the dorsal region of the *Drosophila* embryo depends on the activity of a small group of zygotically acting genes. *dpp*, a key gene in this group, encodes a TGF- β -like product (Dpp) that has been proposed to function as a morphogen with peak levels of Dpp-specifying amnioserosa, the dorsal-most cell type, and lower Dpp levels specifying dorsal ectoderm. The *short gastrulation* gene also contributes to patterning the dorsal region, but unlike the other genes involved in this process, *sog* activity is only required in ventral cells. Genetic evidence indicates that *sog* functions to antagonize *dpp* activity. In this report we present further phenotypic characterization of *sog* mutant embryos in dorsal and lateral regions and describe the cloning of the *sog* locus. *sog* is expressed in a broad lateral stripe of cells that abuts the dorsal territory of *dpp*-expressing cells. *sog* is predicted to encode a protein with an internal signal sequence and a large extracellular domain containing four repeats of a novel motif defined by the spacing of 10 cysteine residues that is distantly related to domains present in thrombospondin and procollagen. We propose that one or more of these cysteine repeats can be liberated by proteolytic cleavage of the primary Sog protein. These putative soluble Sog peptides may then diffuse into the dorsal region to antagonize the activity of Dpp, leading to the subdivision of the dorsal territory into amnioserosa and dorsal ectoderm.

[Key Words: *short gastrulation* (*sog*); *Drosophila*; secreted protein precursor; *decapentaplegic* (*dpp*); TGF- β superfamily; antagonist; embryonic pattern formation; dorsal–ventral axis]

Received August 9, 1994; revised version accepted September 13, 1994.

Dorsal–ventral pattern formation in *Drosophila* and in vertebrates depends on the activity of a class of growth factor molecules related to transforming growth factor- β (TGF- β) (for review, see Green 1993; Smith et al. 1993; Kingsley 1994a,b). During vertebrate development, growth factors in the TGF- β superfamily function as morphogens during early axis formation and promote mesodermal differentiation (for review, see Smith et al. 1993), suppress neuronal fates during early embryogenesis (Hashimoto et al. 1990, Hemmati-Brevanlou and Melton 1994; Hemmati-Brevanlou et al. 1994), contribute to patterning the dorsal region of the neural tube (Basler et al. 1993), suppress female sexual development in males (Lee and Donahoe 1993), organize bone morphogenesis (Kingsley et al. 1994a,b), promote follicle-stimulating hormone (FSH) secretion by the pituitary (for review, see

Findlay 1993; Hillier and Miro 1993), and stimulate erythropoiesis (Murata et al. 1988).

Negative regulation of TGF- β family members is an important theme in the biology of this class of growth factors. For example, in *Xenopus*, activin functions to overcome a predisposition of undetermined ectodermal cells to differentiate as neuronal tissue (Hemmati-Brevanlou and Melton 1994). One likely candidate for a neuralizing factor produced by the Spemann organizer and by the notochord (which underlies the neural plate) is follistatin (Hemmati-Brevanlou et al. 1994), a potent antagonist of activin. Another example of a negative regulator of TGF- β family members is inhibin, a heteromer comprising one of the two identical chains of activin (β) and a related protein chain (α) that specifically interferes with the function of activin (for review, see Hillier and Miro 1993; Findlay 1993).

In *Drosophila* the polarity of the dorsal–ventral axis is determined by a cascade of maternally acting genes functioning in both the oocyte and in the surrounding follicle

¹Corresponding author.

cells (Chasan and Anderson 1993). Ultimately, these maternally acting genes establish a nuclear gradient of the *rel*-related transcription factor encoded by the *dorsal* gene (Roth et al. 1989; Rushlow et al. 1989; Steward 1989). The maternal nuclear gradient of Dorsal protein is responsible for subdividing the embryo into three primary territories of zygotic gene expression (Kosman et al. 1991; Leptin 1991): (1) a ventral region with high levels of nuclear Dorsal giving rise to mesoderm; (2) a lateral region with moderate levels of Dorsal giving rise to neuroectoderm and dorsal-lateral ectoderm; and (3) a dorsal region with little or no Dorsal giving rise to dorsal ectoderm and amnioserosa. One mechanism by which Dorsal specifies the dorsal region is by repressing the expression in ventral cells of key zygotically active genes such as *decapentaplegic* (*dpp*), *zerknüllt* (*zen*), and *tolloid* (*tld*) required for the differentiation of dorsal tissue types (Doyle et al. 1986; Rushlow et al. 1987; Rushlow and Levine 1990; Ray et al. 1991). The *dpp* gene, which encodes a protein in the TGF- β superfamily (Padgett et al. 1987), is likely to play a key role in patterning the dorsal region, as null *dpp* embryos exhibit the most severe ventralized phenotype among known zygotic mutants (Ray et al. 1991; Arora and Nüsslein-Volhard 1992). Another indication that *dpp* plays a crucial role in dorsal-ventral pattern formation is the extreme dosage sensitivity of this locus. *dpp* is one of only three known loci in *Drosophila* that is haplo-insufficient lethal (Irish and Gelbart 1987). Evidence is consistent with the proposal that the Dpp product functions as a morphogen to define a series of thresholds specifying cell fates in lateral and dorsal regions of the embryo (Ferguson and Anderson 1992a,b; Wharton et al. 1993) and then later, during imaginal disc development, to establish the anterior-posterior axis (Basler and Struhl 1994; Tabata and Kornberg 1994).

Patterning in the dorsal region also depends on the function of the *short gastrulation* (*sog*) gene (Wieschaus et al. 1984; Zusman et al. 1988; Rushlow and Levine 1990; Ray et al. 1991; Ferguson and Anderson 1992a). Unlike other genes required for patterning the dorsal region of the embryo [e.g., *dpp*, *zen*, *tld*, *twisted gastrulation* (*tsg*), *screw* (*scw*), and *shrew* (*srw*)], mosaic analysis has revealed that *sog* function is not required in dorsal cells (Zusman et al. 1988). *sog* is also unique in that it behaves genetically as an antagonist of other genes involved in patterning the dorsal region (Ferguson and Anderson 1992a; Wharton et al. 1993). These observations suggest that *sog* might encode or control the production of a diffusible factor that antagonizes the activity of Dpp.

In this report we present further characterization of the *sog* loss-of-function phenotype and describe a molecular analysis of the *sog* locus. These data indicate that *sog* is required for subdividing the dorsal region into amnioserosa and dorsal ectoderm and may also be required for normal neuronal differentiation of cells derived from the neuroectoderm. Consistent with genetic mosaic analysis, *sog* is expressed in a broad lateral stripe abutting the dorsal region. The *sog* locus encodes a predicted

protein with an internal signal sequence and a large extracellular domain containing four copies of a novel motif defined by the spacing of 10 cysteine residues. The Sog cysteine repeats are distantly related to domains present in thrombospondin and procollagen, two known TGF- β -binding proteins. These data are consistent with a model in which *sog* encodes one or more soluble peptides produced by proteolytic cleavage of the primary Sog protein. We propose that processed Sog peptides might diffuse into the dorsal region to antagonize the action of Dpp, thus setting at least two different threshold levels of Dpp activity. In this model the antagonism by Sog would contribute to establishing zones of differing *dpp* activity.

Results

Characterization of the sog mutant phenotype

In the course of analyzing promoter-*lacZ* fusion constructs carrying a fragment of the *deadpan* promoter, we obtained a P-element insertion mapping cytologically to the *sog* locus (13D) on the X chromosome (J. Emery and E. Bier, in prep.). *lacZ* expression in this line includes the typical pattern of pan-neural expression derived from the *deadpan* promoter and all features of the *sog* expression pattern shown in Figure 2, below. Embryos hemizygous for this P-element insertion die with a cuticle phenotype (data not shown) similar to existing *sog* mutants (Wieschaus et al. 1984). The lethality in this line is attributable to the P-element insertion as remobilization of the P element led to the reversion of the lethality in 19 of 20 cases. As we believe this P element is inserted into the *sog* transcription unit (see below), we designate it as the *sog*^{P1} allele. In one case, remobilization of the P element resulted in an imprecise excision that deleted flanking genomic sequences. This allele, which has a stronger phenotype than *sog*^{P1}, is likely to be a null allele as it eliminates expression of *sog* transcripts (see below). We designate this revertant allele as *sog*^{U2} (see below).

Previous analyses indicated that pattern formation is abnormal in the dorsal region of *sog* mutant embryos (Wieschaus et al. 1984; Zusman et al. 1988; Ray et al. 1991; Ferguson and Anderson 1992a). To characterize the *sog* mutant phenotype in greater detail, we examined the expression of a variety of early dorsal-ventral positional markers in embryos hemizygous for the null alleles *sog*^{U2} and *sog*⁶ (Wieschaus et al. 1984) and in embryos mutant for the slightly weaker *sog*^{P1} allele. The *rhomboid* (*rho*) gene is expressed early during the blastoderm stage in lateral stripes within the neuroectoderm (Bier et al. 1990; Ip et al. 1992) and in a dorsal stripe six to eight cells wide that gives rise to the amnioserosa (Fig. 1A). In embryos hemizygous for any of the *sog* alleles tested, the dorsal pattern of *rho* expression is broadened to a domain ~18–20 cells across (Fig. 1B). In contrast, *rho* expression in the dorsal stripe is strongly reduced or entirely absent in other zygotic mutants that disrupt patterning of the dorsal region, including *dpp* (Fig. 1C), *tld*, *scw*, *srw*, *zen*, and *tsg* (data not shown). Expression of dorsally re-

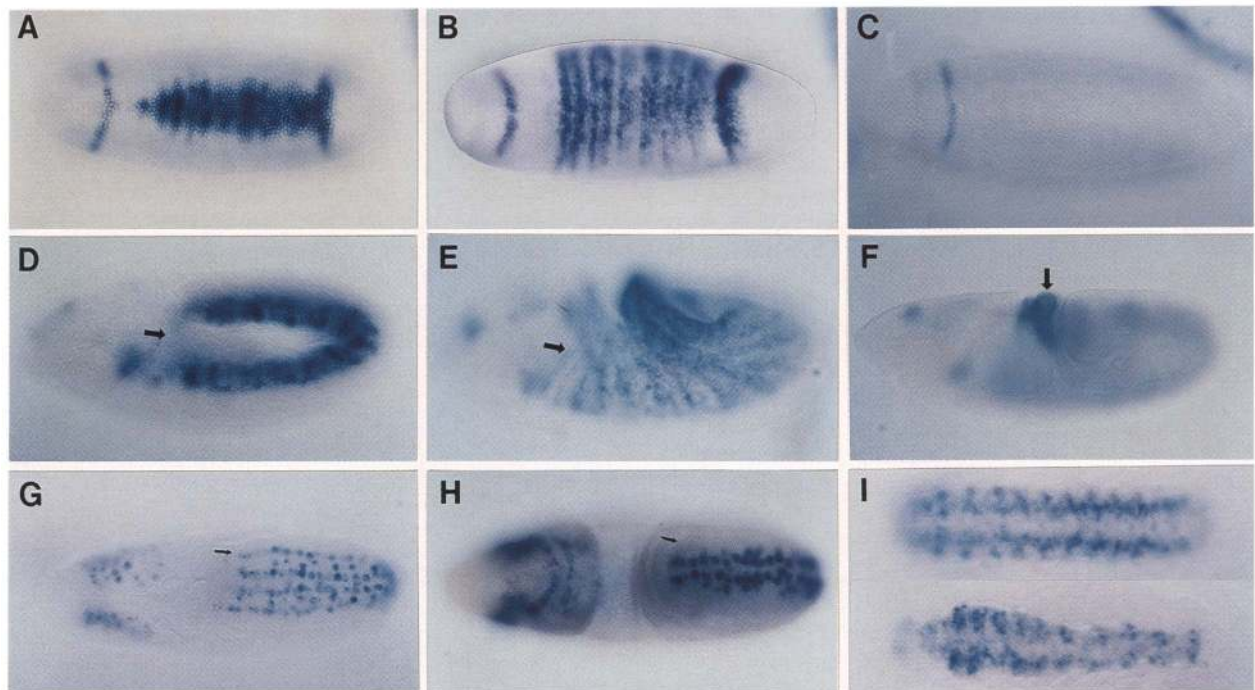


Figure 1. Analysis of the embryonic *sog* mutant phenotype. Embryos are oriented with anterior to the left. (A) Dorsal expression of *rho* in a wild-type blastoderm-stage embryo in the presumptive amnioserosa. (B) Dorsal *rho* expression expands in a null *sog*⁶ mutant embryo. (C) Expression of *rho* in a strong *dpp* mutant is missing from the thoracic and abdominal portions of the dorsal stripe, although the head stripes and the lateral stripes (out of focus) are normal in intensity and location. (D) *dpp* expression in a wild-type germ-band-extended embryo is excluded from the presumptive amnioserosa (arrow) resulting in a horseshoe appearance of staining when viewed laterally at the ectodermal surface. (E) *dpp* expression in a *sog*⁶ mutant embryo is maintained in the presumptive amnioserosa (arrow). The embryo is viewed laterally as in D. The *sog* mutant embryos in E and F are at approximately the same developmental age as the wild-type embryo in D. The *sog* mutants appear younger based on morphological criteria caused by retarded germ-band extension. (F) *dpp* expression in a *sog*⁶ mutant embryo viewed in sagittal section. The arrow indicates location of amnioserosa cells in wild-type embryos. (G) Expression of the pan-neural gene *scratch* (*sct*) (M. Roark and E. Bier, unpubl.) in the three rows of S₁ central nervous system (CNS) neuroblasts in a wild-type embryo. The arrow points to the outer row of neuroblasts. (H) Expression of *sct* in a germ-band-extended *sog*⁶ mutant embryo. The number of neuroblasts forming in the position of the outer S₁ row (arrow) is greatly reduced, especially in the posterior abdominal segments shown. (I) The final number of *sct*-expressing CNS cells in a *sog*⁶ mutant embryo (bottom) is reduced relative to wild type (top). The late mutant phenotype reflects the graded severity in anterior–posterior position observed at the germ-band-extended stage. For H and I other mutant embryos in the same collection are more severely affected.

stricted genes, such as *dpp* (St. Johnson and Gelbart 1987), *zen* (Doyle et al. 1986), and *tld* (Shimell et al. 1991), is also abnormal in *sog* mutants (Ray et al. 1991). For each of these genes the initial pattern of broad dorsal expression is normal, but subsequent pattern refinement fails to take place. For example, in wild-type embryos *dpp* fades from the dorsal-most cells giving rise to amnioserosa as it becomes confined to a broad horseshoe-shaped stripe in the dorsal ectoderm (Fig. 1D). In *sog* mutant embryos, however, *dpp* expression is maintained throughout the entire dorsal region, including cells that would normally form the amnioserosa (Fig. 1E,F). *zen* (Doyle et al. 1986) and *tld* (Shimell et al. 1991) expression domains refine in reciprocal patterns to that of *dpp* (St. Johnson and Gelbart 1987), becoming restricted to the presumptive amnioserosa. In *sog* mutants, *zen* (Ray et al. 1991) and *tld* (J.W. O'Neill and E. Bier, unpubl.) fail to refine and also fade significantly in intensity relative

to wild type, revealing that *sog* is required for both refinement and maintained expression of these two genes. In each of the cases examined, patterns of gene expression normally restricted to subdomains of the dorsal region are broadened in *sog* mutants, indicating that subdivision of the dorsal region is abnormal.

We also examined the pattern of other genes expressed in various dorsal–ventral positions at the blastoderm stage. Expression of *rho* and the *achaete-scute* transcripts T3 (*l'sc*) and T5 (*achaete*) in the neuroectoderm of *sog* mutants is essentially indistinguishable from wild type (data not shown). Similarly, expression of *single minded* (*sim*) and the *Enhancer of split* [*E(spl)*] *m7* transcript in the single row of mesectodermal cells separating the neuroectoderm from the presumptive mesoderm, and expression of genes in the presumptive mesoderm including *snail* and *twist* is unaffected during the blastoderm stage. During early gastrulation, however, the

midline does not form as a single well-ordered row (data not shown). This phenotype may arise as a secondary consequence of defects in gastrulation because of the abnormal behavior of dorsal-most cells that in *sog* mutants fail to flatten out as do wild-type amnioserosa cells.

Finally, we examined the pattern of several different tissue-specific markers in *sog* mutant embryos, including early neuronal precursor cell markers such as *deadpan*, *scratch* (M. Roark and E. Bier, unpubl.), and *snail*; markers for differentiated neurons such as anti-HRP and the 22C10 antigen; early mesodermal markers such as *tinman*, *nautilus*, *S59*, anti-FasIII, and *tld*; late muscle markers such as myosin heavy chain and mAb6D9 antigen; and epidermal markers, including *E(spl) m5* and *m7*, neurotactin, and *Egf-r*. Although the final patterns of intricate structures, such as the peripheral nervous system and somatic muscles, are somewhat disorganized, as might be expected from the profound disruption of gastrulation in *sog* mutants, expression of these markers for various stages and tissue types is largely normal with the exception of the central nervous system (CNS). In *sog* mutants many cells comprising the outer row of S_1 neuroblasts in the CNS are missing (Fig. 1G,H). Ultimately, *sog* mutant embryos exhibit defects including a reduced (and often split) CNS (Fig. 1I), a reduced lateral extent of ventral denticle belts (Wieschaus et al. 1984), irregular somatic muscles, absence of gut constrictions, various head defects, absence of amnioserosa, and failure of dorsal closure. These late phenotypes may be secondary consequences of the early failure to subdivide the dorsal ectoderm or may be attributable to other functions of *sog*.

sog antagonizes *dpp* genetically

sog is the only known mutation that has been reported to antagonize the activity of dorsally acting zygotic genes such as *dpp* (Wharton et al. 1993) and *tld* (Ferguson and Anderson 1992a). As antagonism of dorsally acting zygotic genes is an important characteristic of *sog* mutants, we investigated the ability of our newly generated weak (*sog^{P1}*) and strong (*sog^{U2}*) *sog* alleles to suppress *dpp* haplo-lethality. The result of crossing male flies carrying a *dpp* null allele (covered by a duplication of *dpp*) to females heterozygous for the *sog^{P1}*, *sog^{U2}*, and *sog⁶* mutations is summarized in Table 1. Both *sog^{U2}* and *sog⁶* rescue *dpp* haplo-insufficiency to a significant extent, consistent with evidence described below that both of these are null alleles. The *sog^{P1}* allele also rescues *dpp* haplo-insufficiency but less effectively than *sog⁶* and *sog^{U2}*, providing additional evidence that *sog^{P1}* is a hypomorphic mutation. In contrast to the rescue observed with these *sog* alleles, viable precise P-element excision revertants of *sog^{P1}* (*sog^{rev}*) and *tsg*, a mutant that has a hemizygous phenotype most similar to *sog*, do not rescue *dpp* haplo-insufficiency. Thus, rescue of *dpp* haplo-insufficiency and expansion of the dorsal *rho* stripe are unique characteristics that distinguish *sog* from other mutants disrupting the formation of dorsal structures.

Table 1. Rescue of *dpp* haplo-lethality by heterozygous *sog* alleles

Genotype of female	Genotype of test mutant				
	<i>sog^{rev}</i> (= w.t.)	<i>sog^{P1}</i>	<i>sog^{U2}</i>	<i>sog⁶</i>	<i>tsg</i>
<i>FM7/yw</i> ; <i>DP(dpp)/+</i>	232	466	196	142	439
mutant/ <i>yw</i> ; <i>Dp(dpp)/+</i>	268	593	224	182	549
<i>FM7/yw</i> ; <i>dpp^{Hin}/+</i>	11	53	24	15	4
*mutant/ <i>yw</i> ; <i>dpp^{Hin}/+</i>	1	147	78	67	0
Total female progeny	512	1259	522	406	992
Percent expected*	1%	47%	60%	66%	0%

Females carrying the different X-linked mutations balanced over *FM7c B* were crossed to males with a strong *dpp^{Hin}* allele balanced over a *CyO* chromosome (*CyO23*) carrying an extra copy of *dpp*. We tabulated the number of female progeny that were heterozygous for the *dpp⁻* chromosome and that were also heterozygous for either *FM7c B* (control) or the test mutation. The percentage of expected progeny (*) was calculated based on the quarter of zygotes having the trans-heterozygous *sog⁻/+*; *dpp⁻/+* genotype being fully viable. In all cases, heterozygosity for a *sog* allele increased the number of surviving *dpp⁻/+* individuals. The *sog^{P1}* allele in this and other experiments (not shown) is consistently less effective at rescuing *dpp⁻* haplo-lethality than either *sog^{U2}* or *sog⁶*. Rescue of *dpp⁻* haplo-lethality was not observed with viable P-element excision revertant lines [*sog^{rev}* = wild type (w.t.)] or with a *tsg* allele. *tsg* is an X-linked mutation with a phenotype superficially very similar to that of *sog*.

Expression of *sog* during embryogenesis

To examine the pattern of *sog* expression during embryogenesis we hybridized whole-mount embryos with an antisense RNA probe synthesized from a *sog* cDNA template. During cycle 14, *sog* is expressed in a broad lateral stripe 14–16 cells wide (Fig. 2A). This expression, as well as all later elements of the *sog* expression pattern, is undetectable in embryos hemizygous for the *sog^{U2}* (Fig. 2B) or *sog⁶* alleles (data not shown), consistent with phenotypic and genetic evidence described above and the molecular evidence presented below that these *sog* alleles are null mutations. *sog* is expressed in embryos mutant for the weaker *sog^{P1}* allele (data not shown), but transcripts recede prematurely from dorsal–lateral cells and are not localized apically within the cell as they are in wild-type embryos (see below).

To determine the limits of the *sog* expression domain, we used a double label RNA in situ method that we developed recently (O'Neill and Bier 1994). In this method, probes are synthesized in the presence of either digoxigenin-labeled UTP or biotin-labeled UTP. Digoxigenin and biotin probes complementary to different transcripts are hybridized simultaneously. The digoxigenin-

François et al.

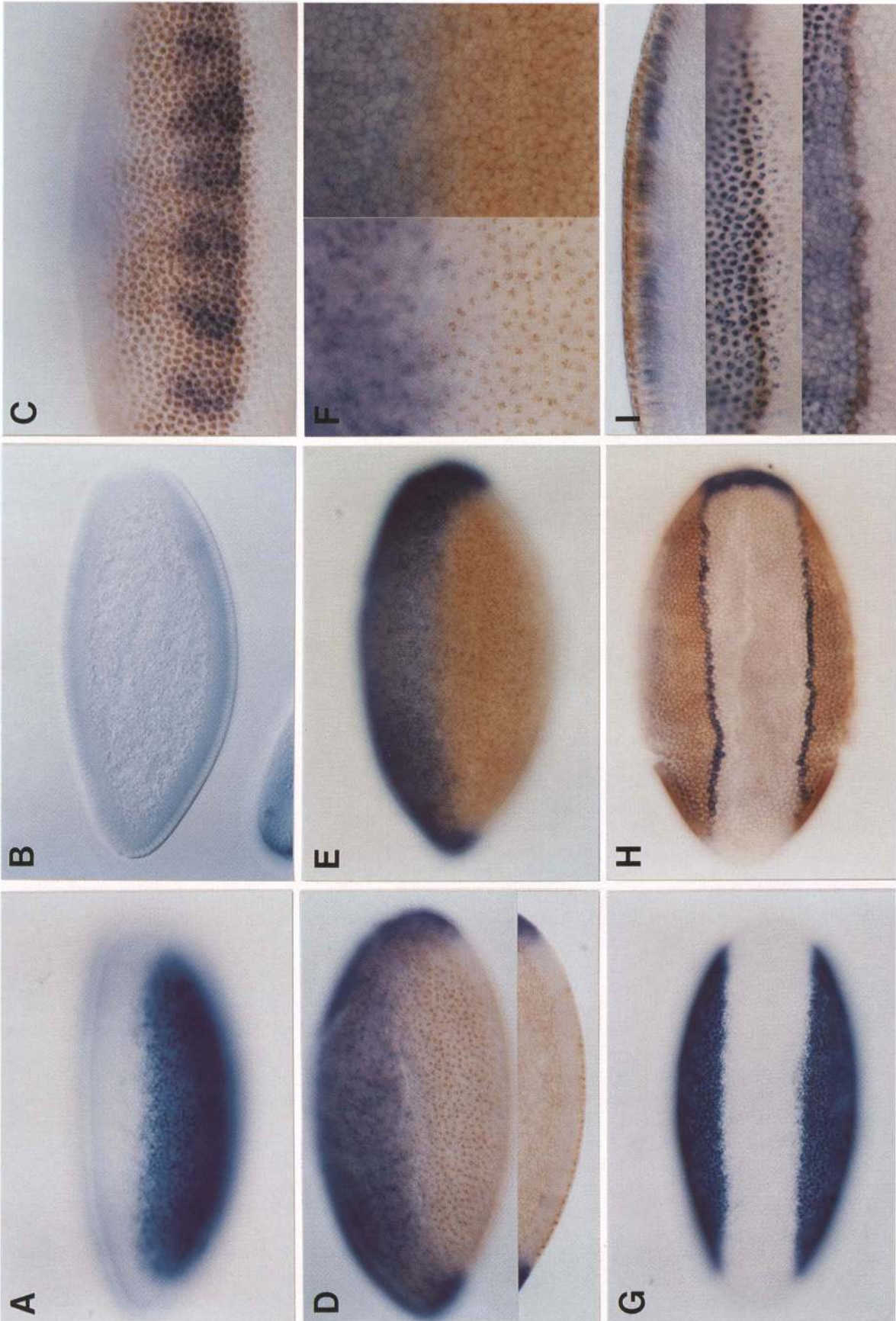


Figure 2. (See p. 2608 for legend.)

sog encodes a predicted secreted protein precursor



Figure 2. (See following page for legend.)

labeled probe was visualized as usual with a blue alkaline phosphatase reaction and the biotin probe was visualized as a brown horseradish peroxidase (HRP) reaction product. The relation of *sog* expression to that of *rho* in the neuroectoderm is revealed in embryos hybridized with both a biotin-labeled *sog* probe (brown) and a digoxigenin-labeled *rho* probe (blue) (Fig. 2C). These experiments indicate that *sog* is expressed significantly earlier than *rho* and that the *rho* lateral stripe (8–10 cells wide) is embedded in the broader *sog* expression domain (14–16 cells wide). The *sog* expression domain initially extends one to two cell diameters ventral to the lateral *rho* stripe and four to six cells beyond *rho* dorsally. A short time later the ventral limits of the *sog* and *rho* expression domains become coincident with the mesectoderm marked by *sim* expression (see below).

We examined the position of the dorsal limit of *sog* expression with respect to the ventral border of *dpp* expression with a biotin-labeled *sog* probe and a digoxigenin-labeled *dpp* probe (Fig. 2D–F). The dorsal-most cells expressing *sog* about the ventral-most cells expressing *dpp* as early as cycle 13 when these genes are first expressed (Fig. 2D). Expression of *sog* ventrally in the prospective mesoderm is rapidly lost during early cycle 14 (Fig. 2E; see below). The disappearance of *sog* expression ventrally may be the result of direct repression by *snail*,

as it begins very early and is not observed in *snail* mutant embryos (data not shown). The boundary between the *sog* and *dpp* expression domains is not absolute, as a few cells express low levels of both transcripts (Fig. 2F). *sog* expression gradually fades from its dorsal limits during mid-cycle 14 leading to the appearance of a small gap between *sog* and *dpp* expression domains by late cycle 14 (data not shown). The dorsal limit of the *sog* domain is independent of the activity of genes involved in patterning the dorsal region (e.g., *dpp*, *tld*, *scw*, *sr*, *zen*, *tsg*) or the lateral region [e.g., *rho* and *achaete-scute* complex (AS-C) genes], as *sog* expression is normal in embryos mutant for these genes (data not shown).

The position of the ventral boundary of *sog* expression (Fig. 2G) relative to the single row of *sim*-expressing mesectodermal cells that separate the neuroectoderm from the mesoderm is shown in Figure 2H (*sog*, brown; *sim*, blue). In these double-labeled preparations the ventral boundary of *sog* expression includes the mesectoderm, as cells expressing *sim* also express *sog*. Some cells ventral to the *sim*-expressing row of cells continue to express *sog* for a brief period (Fig. 2I), in contrast to *rho*, which is strictly excluded from the presumptive mesoderm before the onset of *sim* expression. An interesting feature of *sog* labeling is that the mature *sog* transcripts are localized predominantly apically within the cell in contrast to *sim*

Figure 2. Expression of *sog* relative to other genes during embryogenesis. (A) A lateral view of *sog* expression during early cycle 14 visualized with a digoxigenin-labeled antisense RNA probe synthesized from a cDNA clone. (B) *sog* expression is undetectable in an embryo hemizygous for the *sog*^{U2} allele. Approximately one-quarter of the young embryos, such as the one shown here, have this phenotype. At slightly later developmental stages (early gastrulation), when *sog* mutant embryos could be scored unambiguously based on the absence of a lacZ-producing balancer chromosome, every mutant embryo was unstained. The same absence of staining at all developmental stages is observed in *sog*⁶ mutant embryos. (C) Location of *sog* transcripts (brown staining of biotin-labeled probe) relative to *rho* transcripts (blue staining of digoxigenin-labeled probe). (D) Double label of *sog* transcripts (brown) and *dpp* (blue) during cycle 13. At this early stage *sog* is expressed in ventral (*bottom*) as well as lateral cells. (E) Double label of *sog* transcripts (brown) and *dpp* (blue) during mid-cycle 14. (F) High magnification view of embryos shown in D (*left*) and E (*right*). (G) Ventral view of *sog* expression during early to mid-cycle 14. (H) Ventral view of a mid-cycle 14 embryo double labeled with *sog* (brown) and *sim* (blue) probes. (I) High magnification views of *sog* plus *sim* double-labeled embryos. (*Top*) An optical cross section through an embryo hybridized with a biotin-labeled *sog* probe (brown) and a digoxigenin-labeled *sim* probe (blue). *sog* transcripts are localized predominantly at the apical end of the cell in contrast to the more typical accumulation of *sim* transcripts in the central portion of the cell. (*Middle*) A surface view of an early to mid-cycle 14 embryo reciprocally hybridized with a digoxigenin-labeled *sog* probe (blue) and a biotin-labeled *sim* probe (brown). The *sog* signal is strong and the *sim* signal is weak in this focal plane. (*Bottom*) An optical section of the same embryo shown in the *middle* panel taken in a slightly deeper focal plane corresponding to the approximate middle of the forming blastoderm cell layer. Note that the *sog* signal is weak and that the *sim* signal is maximal, confirming the differential localization of transcripts observed in the top panel. Similar apical localization of *tsg* transcripts has been reported recently (Mason et al. 1994). (J) *sog* expression during early gastrulation shown from a ventral perspective. (K) *sog* expression at germ-band extension is confined to mesectodermal cells along the ventral midline. (L) A sagittal section of a germ-band-extended embryo double stained with a biotin-labeled *sog* probe (brown) and a digoxigenin-labeled *rho* probe (blue). Note that the two signals are segregated to opposite ends of mesectodermal cells, which are apicobasally elongated at this stage. Expression of *lacZ* or *sog* transcripts in the mesectoderm of *sog*^{P1} embryos is not localized apically and intron-specific *sog* probes, which stain only the nuclei of expressing cells, also do not detect apically localized transcripts. (M) A lateral superficial view of a germ-band-extended embryo double stained with *sog* and *dpp* probes. (*Top*) Portion of an embryo stained with a biotin-labeled *sog* probe (brown, arrowhead) and a digoxigenin-labeled *dpp* probe (blue, arrow); (*bottom*) part of an embryo reciprocally stained with a digoxigenin-labeled *sog* probe (blue, arrowhead) and a biotin-labeled *dpp* probe (brown, arrow) to highlight *sog* labeling. (N) A double-stained late germ-band-retracting embryo hybridized with a digoxigenin-labeled *sog* probe (blue, arrowhead) and a biotin-labeled *dpp* probe (brown, arrow). (O) *sog* expression in ventral ectodermal cells after germ-band retraction. (P) *sog* expression in the gut and pharynx of a germ-band-retracted embryo viewed in horizontal section from a dorsal perspective. (Q) A double-stained germ-band-retracted embryo hybridized with a biotin-labeled *sog* probe and a digoxigenin-labeled *dpp* probe viewed in horizontal section from a dorsal perspective. At the junction of the esophagus and the gut *dpp*-expressing visceral mesoderm cells (arrow) directly overlie *sog*-expressing endodermal cells (arrowhead). (R) High magnification view of *sog* expression at the junction of the esophagus and gut. *sog* transcripts are confined to the endoderm (arrowhead) and are absent in the visceral mesoderm layer (arrow).

or *rho* transcripts, which are found in the more conventional central portion of the cell (Fig. 2I).

Expression of *sog* is progressively lost dorsally during the late blastoderm stage and during early gastrulation (Fig. 2J). By germ-band extension *sog* transcripts are confined to the ventral midline (Fig. 2K). *sog* transcripts are also localized apically at this stage as can be deduced from double-label experiments using *sim* or *rho* probes to mark the ventral midline (Fig. 2L). At full germ-band extension *sog* expression is initiated in a pattern of lateral patches dorsal to the tracheal pits (Fig. 2M) that fuse ultimately to form a lateral stripe running between the two stripes of *dpp* expression (Fig. 2N).

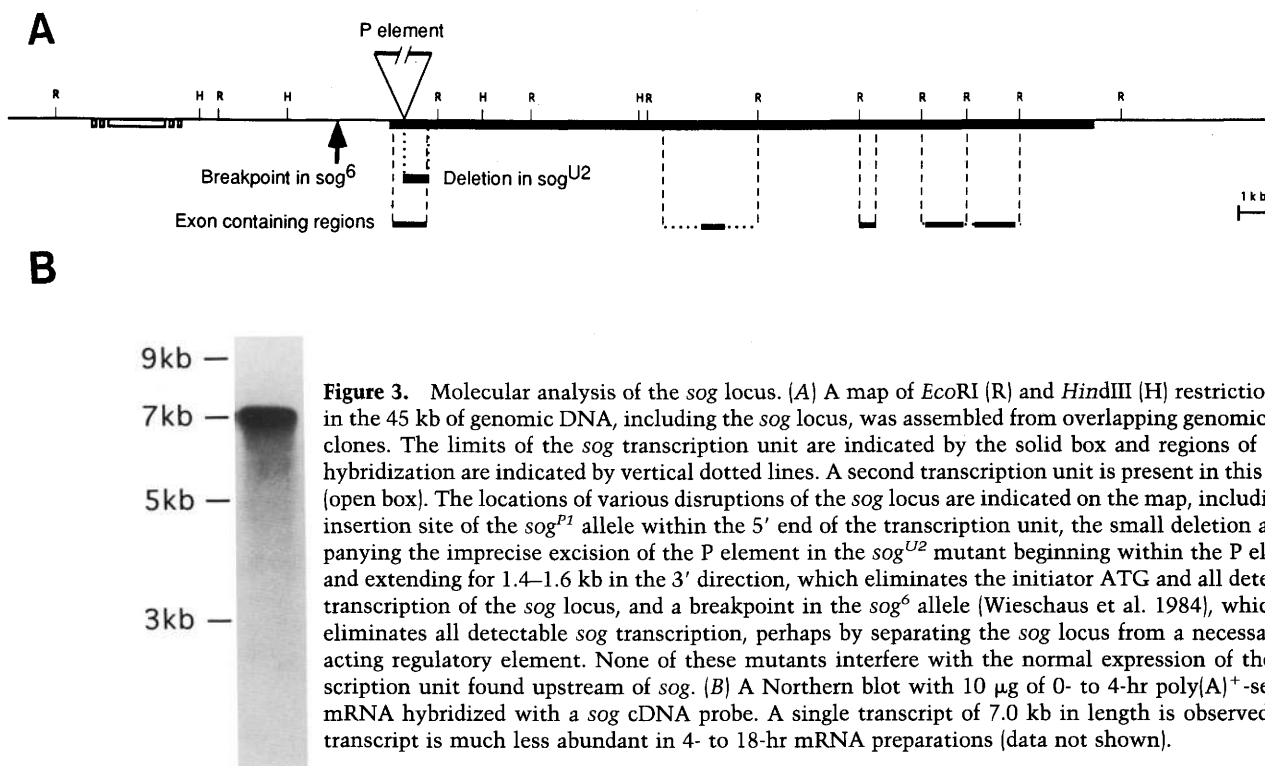
After germ-band retraction, *sog* is expressed superficially in ventral epidermal cells in a pattern resembling the outline of future denticle belts (Fig. 2O) and internally in a series of bands in the endoderm of the gut and in the esophagus (Fig. 2P–R). Endodermal expression of *sog* is generally out of phase with rings of *dpp* expression in the overlying visceral mesoderm, with the exception of the junction of the gut with the esophagus, where visceral mesoderm cells expressing *dpp* are in direct contact with internal endodermal cells expressing *sog* (Fig. 2Q).

Molecular analysis of the *sog* locus

To obtain *sog* genomic DNA flanking the site of P-element insertion we amplified a short fragment using PCR from circularized *sog*^{P1} genomic DNA template with a pair of divergent P-element primers. We then used this fragment as a probe to screen a genomic phage library. We restriction mapped several genomic *sog* clones (Fig.

3A) and used a 310-bp fragment mapping near the site of the P-element insertion as a probe to screen the 0- to 4-hr Nick Brown cDNA library to saturation (see Materials and methods). Six different cDNA clones were isolated. Northern blot analysis using a *sog* cDNA as probe identifies a single transcript of ~7 kb in length (Fig. 3B). Hybridization of the *sog* cDNA to the genomic walk and in situ hybridization of *sog* genomic fragments to embryos indicates that the *sog* locus covers >26 kb and that the mature transcript is interrupted by at least three introns.

We analyzed the structure of the *sog* locus in flies heterozygous for the *sog*⁶, *sog*^{U2}, and *sog*^{P1} alleles. The *sog*^{P1} insertion site maps ≥170 bp within the *sog* transcription unit, and the imprecise excision generated in the *sog*^{U2} allele results in a deletion of genomic sequences; one end of this deletion lies within the P-element, and the other end maps 1.4–1.6 kb 3' to the integration site. This deletion should remove sequences encoding the first 99 amino acid residues of the predicted Sog protein. Southern blot analysis of the *sog*⁶ allele, which was generated in the screen for early patterning mutants (Wieschaus et al. 1984), reveals a breakpoint mapping ~2 kb upstream from the 5' end of the longest *sog* cDNAs. The lesions in both *sog*^{U2} and *sog*⁶ are likely to represent null alleles, as no *sog* transcripts are detected by in situ hybridization to embryos mutant for either of these alleles at any developmental stage. The genomic lesions present in these three mutants support the genetic and phenotypic evidence described above that these mutants all disrupt a common transcription unit likely to correspond to the *sog* locus. Consistent with this interpretation, none of these *sog* alleles lead to



detectable alterations in the expression pattern of the nearest transcript that maps >5 kb upstream of the putative *sog* locus (data not shown).

sog encodes a predicted integral membrane protein with novel cysteine repeats in the extracellular domain

Sequence analysis of a 4.5-kb *sog* cDNA that contains the 5' end of the *sog* transcript, but lacks ~2.5 kb of 3'-untranslated sequences, predicts a single long open reading frame encoding a protein of 1038 amino acids (Fig. 4A). The predicted Sog protein does not share strong sequence similarity with any known protein in the SWISS-PROT, PIR(R), and GenPept data bases using the BLAST algorithm. Hydropathy analysis using the Kyte–Doolittle algorithm (Fig. 4B) suggests that Sog contains a single long hydrophobic domain beginning 56 amino acids from the amino terminus (overlined in Fig. 4A). The position of this single potential membrane-spanning hydrophobic domain, the excess of basic residues preceding and of acidic residues after the hydrophobic region, and the presence of two glycine residues within the hydrophobic segment are consistent with Sog being a type II membrane protein with the amino terminus facing the cytoplasm and the carboxyl terminus lying outside of the cell (Landry and Gierasch 1991). Alternatively, the internal signal sequence could serve as a cue for secretion of the carboxyl portion of the protein from the cell. In the putative extracellular domain there are five potential N-linked glycosylation sites (bold and starred in Fig. 4A), four repeats of a cysteine-rich domain (CR1–CR4; boxed in Fig. 4A; see Fig. 4C for alignment), three copies of another less conserved repeat ending in potential glycosylation sites (SR1–SR3; dotted underline in Fig. 4A; see Fig. 4D for alignment), and 20 dibasic amino acid pairs (underlined in Fig. 4A).

The four cysteine-rich repeats are defined by a fixed spacing of 10 cysteine residues (CR4 lacks the ninth cysteine) as well as several other shared amino acids represented in a consensus sequence 71 amino acids in length (Fig. 4C). Three of the cysteine repeats are closely spaced near the carboxyl end of the molecule, and one is found immediately after the putative transmembrane domain. These repeats have a similar spacing of cysteine residues and share a low degree of sequence identity with a domain referred to as the procollagen domain present in thrombospondin and pro- α -1 (I) collagen. The procollagen domain also shares more limited sequence similarity with cysteine domains found in other extracellular proteins, including von Willebrand factor, laminins, the product of the *Drosophila stranded at second (sas)* gene, and genes of the CEF-10/CTGF/ β IG-M2 family that are primary response genes induced by various growth factors (Simmons et al. 1989; Bradham et al. 1991; Brunner et al. 1991). Secondary structure analysis predicts that much of the extracellular portion of the Sog protein is likely to assume an α -helical conformation in contrast to the cysteine repeat regions that are predicted to be punctuated by many turns. The cysteine repeats are also regions unusually rich in polar and charged residues (see

hydrophilicity plot in Fig. 4B). The conserved spacing of the cysteine residues within the four repeats and predicted secondary structural features are consistent with each of the repeats folding independently, perhaps into cysteine knots (Murray-Rust et al. 1993).

Dibasic pairs of amino acids are scattered throughout the extracellular domain, two of which define the ends of CR2 and CR4. Cleavage by many trypsin-like serine proteases occurs after dibasic residues, as is the case in processing TGF- β family members to yield mature active factors (Kingsley 1994b). Processing of membrane-bound precursors also occurs at other sites (Massagué and Pandiella 1993). The proposed structure of the predicted Sog protein, with the potential for liberating diffusible peptides by proteolytic processing, is consistent with the nonautonomous genetic behavior of *sog* mutations.

Discussion

sog functions nonautonomously to antagonize Dpp activity leading to the early subdivision of the dorsal ectoderm

Phenotypic analysis suggests that the primary defect in *sog* mutants is a failure to subdivide the dorsal region of the blastoderm (e.g., into amnioserosa and dorsal ectoderm). For example, *rho* expression, which is normally confined to the amnioserosa, is expanded ventrally in *sog* mutants, and *dpp* expression, which normally withdraws from the amnioserosa, fails to do so. As *sog* transcripts occupy a broad lateral domain but are absent in abutting dorsal cells that require the activity of this gene, *sog* must act over a significant distance (>12 cells) to influence dorsal cell fates. This nonautonomous behavior of *sog* is consistent with previous mosaic analysis that defined a ventral or lateral focus for the dorsal defects observed in *sog* mutants (Zusman et al. 1988).

sog behaves genetically as an antagonist of other genes required for patterning the dorsal region of the embryo. *sog* is unique in this regard, as in comprehensive genetic screens no autosomal mutations were recovered (other than duplications of Dpp itself) that could dominantly suppress *dpp* haplo-lethality (W.M. Gelbart, pers. comm.). Similarly, *sog* was the only mutation isolated in a screen for dominant suppressors of lethality due to a combination of hypomorphic *tld* alleles (Ferguson and Anderson 1992a). The abutting domains of *sog* and *dpp* expression and the potent antagonism of *dpp* by *sog* suggests that at least two different threshold levels of Dpp activity may be set by putative diffusible Sog peptides (i.e., Dpp alone in the presumptive amnioserosa versus Dpp plus Sog in the dorsal ectoderm). In simple versions of this model Dpp protein would not need to be present in a concentration gradient to participate in subdividing the dorsal region.

sog may also function later during embryogenesis

In addition to patterning the dorsal ectoderm there is evidence that *sog* plays a role in the development of the

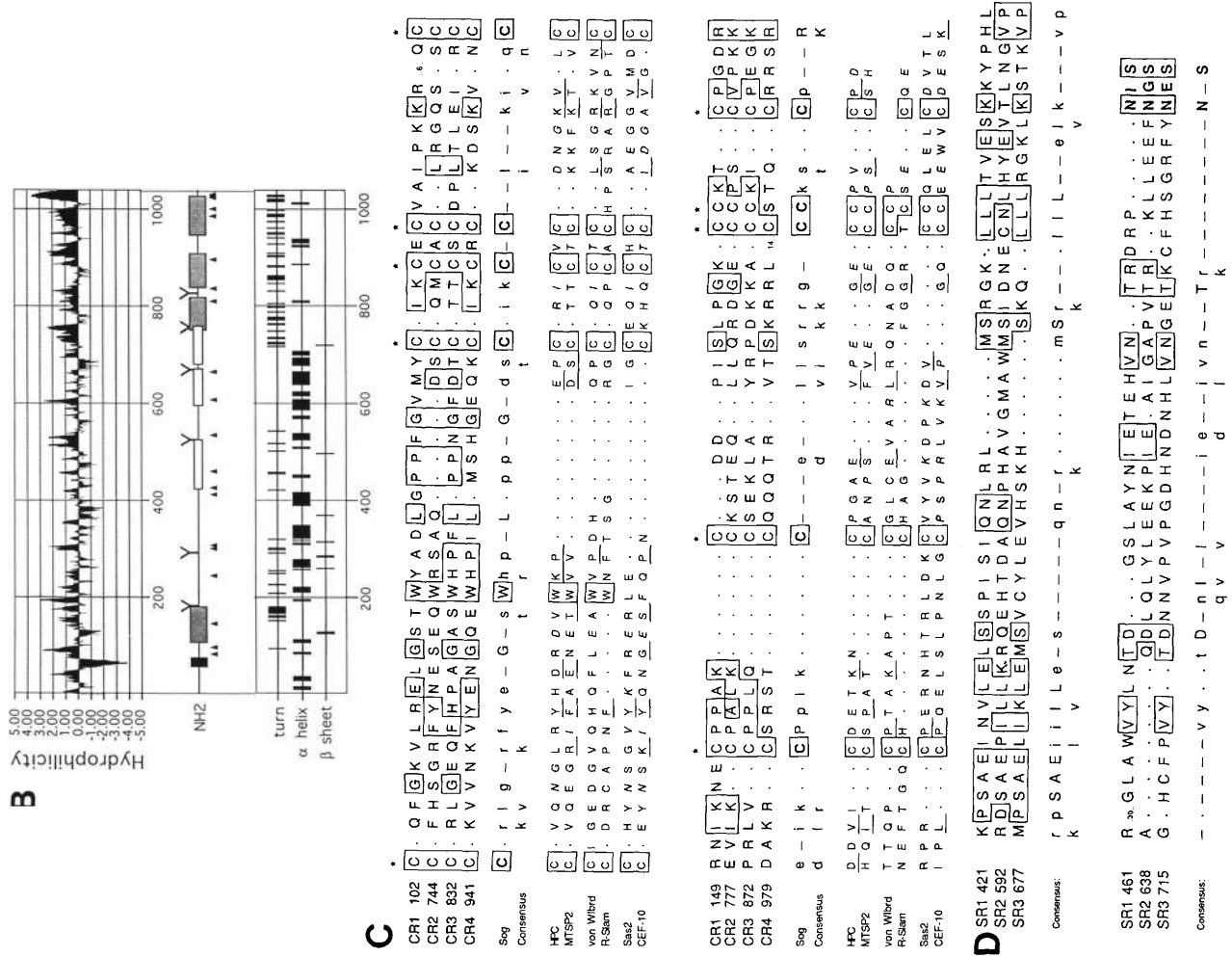
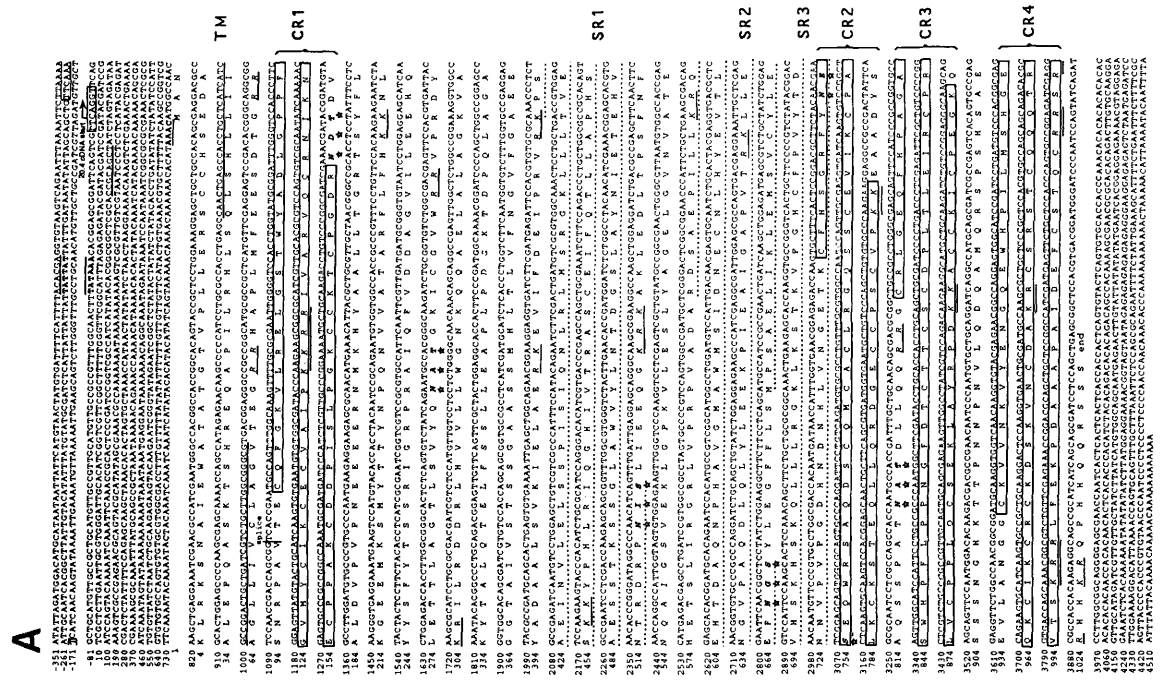


Figure 4. (See following page for legend.)



neuroectoderm in which it is expressed. For example, the number of outer row S_1 neuroblasts is reduced in *sog* mutants. Epidermal structures derived from the lateral ectoderm also exhibit defects visible in the cuticles produced by *sog* mutant embryos. The ventral cuticle defects are exacerbated by increasing the dosage of *dpp* (Ferguson and Anderson 1992a), suggesting that *sog* may antagonize *dpp* in the neuroectoderm as well as in the dorsal region.

sog may also play additional roles later in embryogenesis, as it is expressed in sharp longitudinal stripes alternating with *dpp* stripes during germ-band retraction and then in a series of rings in the endoderm spaced along the anterior–posterior axis of the developing gut. An intriguing feature of the late *sog* expression patterns is that *dpp* is expressed in similar late patterns usually alternating with *sog*. Whether these two genes interact during these late stages to establish fine positional values remains to

be determined. It is worth noting that *dpp* expressed in the visceral mesoderm has been shown to play a key inductive role in formation of the second midgut constriction, which involves activation of the *labial* gene in the underlying endoderm (Panganiban et al. 1990; Reuter et al. 1990; Hirsch et al. 1993; Capovilla et al. 1994).

sog encodes a potential precursor protein for a novel class of cysteine-type secreted factors

The potential of releasing short peptides from the primary Sog protein is consistent with the nonautonomous requirement for *sog* function. There are four copies of a repeat unit defined by a fixed spacing of 10 cysteine residues. One or more of these cysteine repeats might be liberated from the extracellular domain of the primary Sog protein as free peptides after cleavage at dibasic residues or other sites. If diffusible Sog peptides are gener-

Figure 4. Sequence analysis of *sog*. (A) Composite *sog* sequence derived from genomic DNA (nucleotides –351 to +1105) and cDNA clones (nucleotides 0 to +4541). The longest *sog* cDNA clone contains the full predicted open reading frame. The 5' end of the longest cDNA (boxed nucleotides from –2 to +5) matches well (5 of 7) with the consensus mRNA cap site ATCA(G/T)T(C/T) for non-heat shock genes (Hultmark et al. 1986). Another match to the consensus cap site is also boxed (nucleotides –177 to –171). Two potential TATA boxes, spaced 25 and 30 nucleotides, respectively, from each of these two consensus cap sites, are overlined. The P-element insertion site in *sog*^{PI} is located 140–200 bp downstream from the 5' end of the longest cDNA and is therefore within the 5'-untranslated region of the *sog* transcription unit. A good match to the consensus for preferred P-element insertion sites is indicated in this interval (underlined nucleotides +169 to +176). The codon for the predicted initiator methionine (amino acid +1) of the Sog protein is preceded by AAAC (bold and underlined; nucleotides +807 to +810), which conforms to the consensus for a *Drosophila* translation start site (C/A)AA(C/A) (Cavener 1987). A single long open reading frame contained two independent cDNA isolates predicts a protein 1038 amino acids in length. The encoded Sog protein is predicted to have an internal signal sequence (overlined, amino acids 56–74). The large predicted extracellular domain contains four repeats of a domain (boxed) defined by a fixed spacing of cysteine residues (CR1–CR4, see Fig. 4C), three copies of a second type of repeat (SR1–SR3, dotted underline, see Fig. 4D), five potential glycosylation sites (bold, italic, with stars below), and 20 pairs of dibasic residues (solid underline, italic) that could potentially serve as cleavage recognition sites for serine-type proteases (Massagué and Pandiella 1993). (B) Secondary structure predictions. (Top) Hydrophilicity plot (Kyte and Doolittle 1982) of the predicted Sog protein as calculated using the McVector hydrophilicity program with the window size set to 10 amino acids. The majority of the predicted protein is hydrophilic, with only a single region of sufficient length and hydrophobicity to span the membrane. Note the relatively polar values of the cysteine repeats in the carboxy-terminal portion of the protein. (Middle) Diagram of the predicted Sog protein with the amino terminus at left. The solid box indicates the relative location of the single potential hydrophobic membrane-spanning region, shaded boxes indicate the locations of the cysteine repeats, open boxes denote SR repeats, Y indicates potential N-linked sites, and arrowheads indicate the occurrences of dibasic residues. (Bottom) Secondary structure prediction for the Sog protein using the combined Chou–Fasman and Robson–Garnier algorithms calculated by the McVector program. The top line represents the tendency to introduce turns; the middle line, the likelihood of assuming an α -helical conformation; and the bottom line, the propensity to form a β -sheet. Note that the majority of the extracellular domain is predicted to form an α -helix but that this probability drops off abruptly in the regions containing the cysteine repeats. The cysteine repeats are predicted to be riddled with turns, consistent with the view that they may form highly folded structures. (C) Sequence alignment of the four cysteine repeats, CR1 (amino acids 102–178), CR2 (amino acids 744–807), CR3 (amino acids 832–902), and CR4 (amino acids 941–1023), of Sog with domains from other proteins. Domains with related sequences were identified by use of the algorithm of Rohde and Borke (1993). Boxed residues indicate two or more identities between repeats, stars indicate the location of the cysteine residues in Sog repeats that can all be aligned according to the fixed spacing $Cx_{24}Cx_2Cx_1Cx_{6-8}Cx_4Cx_4Cx_{12}CCx_2Cx_3R/K$. Single dots denote gaps in the sequence introduced to preserve alignment. A consensus for the cysteine repeats is shown below the aligned sequences. Cysteine residues and one tryptophan residue shared by all repeats are boxed; uppercase letters in the consensus indicate three or more identities at that position; lowercase letters indicate that two or more repeats contain an amino acid or a similar amino acid using the stringent substitution criteria: R or K, E or D; S or T; L, V, or I; and Q or N; italic letters indicate a frequently occurring amino acid among the compared sequences that is not favored in the Sog repeats; dashes indicate positions in the consensus where no clear residue is favored. In the alignment of CR1 there is an insertion of 6 amino acids relative to the consensus after amino acid 140 (R I V A R V), and in CR4 an insertion of 14 amino acids after residue 1000 (F E K P D A A A P A A I D E F). The insert of 14 amino acids in CR4 after the Cx_{12} spacer can itself be aligned with the four Cx_{12} spacer sequences, suggesting that this region of the CR4 repeat may have arisen by a short tandem duplication. Abbreviations for aligned proteins are (HPC) Human pro- α -1 (I) collagen (Chu et al. 1984); (MTSP2) mouse thrombospondin 2 (Bornstein et al. 1991); (von Wlbrd) von Willebrand factor (Verweij et al. 1986); (R-Slam) rat S-laminin (Hunter et al. 1989); (Sas2) second repeat (of four) of the predicted *Drosophila stranded at second* (*sas*) gene product (Schonbaum et al. 1992); (CEF-10) CEF-10 (Simmons et al. 1989). (D) Sequence alignment of the three SR repeats (SR1–SR3). Insert of 30 amino acids (Q G H I V T R A S C E I F Q T L L A P H S A E S S T K S S).

ated by processing of the primary predicted product they could function by one of several obvious mechanisms. For example, Sog peptides could activate a yet unknown Sog receptor or might bind to and inactivate some component required for dorsal patterning.

Sog repeats share the spacing of cysteine residues and limited sequence identity with various known or predicted extracellular proteins. The procollagen domains of thrombospondin and of procollagen are the closest relatives of the Sog repeats and share all 10 conserved cysteines, albeit with slightly altered spacing. The sequence similarity between Sog, thrombospondin, and procollagen is particularly interesting, as type IV collagen (Paralkar et al. 1991) and a region of thrombospondin containing the procollagen domain (Murphy-Ullrich et al. 1992) have been shown to bind TGF- β . The procollagen domain is biologically active as a soluble factor, as peptides corresponding to the isolated domains from thrombospondin or pro- α -1 collagen have anti-angiogenic activity (Tolsma et al. 1993). It is also noteworthy that procollagen expression is tightly linked to that of TGF- β (Nakatsukasa et al. 1990; Kulozik et al. 1990) and that TGF- β can induce procollagen gene expression (McAnulty et al. 1991; Ritzenthaler et al. 1993). Whether the distant evolutionary relationship between the Sog cysteine repeats and the procollagen domain reflects a conserved function for binding TGF- β family members will be an important question to address in the near future.

In addition to Dpp, there are other known extracellular proteins with which Sog could potentially interact, including Tld (Shimell et al. 1991), Scw (another TGF- β family member; Arora et al., this issue), Tsg (Mason et al. 1994), or Dpp receptors such as *saxophone* or *thickveins* (Brummel et al. 1994; Nellen et al. 1994; Penton et al. 1994; Xie et al. 1994). With respect to Tsg, it is noteworthy that the amino-terminal portion of one of the two cysteine rich domains in the CTGF/CEF-10/ β IG-M2 family of proteins shares sequence similarity with the predicted Tsg protein, whereas the carboxyl portion of this domain is related to the Sog CR repeats. It is also interesting in this regard that β IG-M2 (or *cyr61*) mRNA (β IG-M2 shares 94% amino acid identity with CEF-10 in this region) is induced by TGF- β 1 in mouse AKR-2B cells.

Finally, Sog might also function as a cell-bound transmembrane protein rather than as a secreted factor, possibly by binding one of the diffusible proteins required for patterning the dorsal region. A role for Sog as a membrane-bound receptor is similar to one proposed function of TGF- β type III receptors (Wang et al. 1991), although type III receptors are believed to facilitate rather than antagonize TGF- β signaling. In these latter scenarios the nonautonomous action of *sog* mutants would necessarily be indirect.

Negative regulation of TGF- β class growth factors

Negative regulation of TGF- β class growth factor molecules is important for modulating the activity of this family of peptides in a wide variety of biological settings.

One of the most thoroughly studied examples is the specific blocking of activin by inhibin. As inhibin and activin share a common subunit and form dimers of related structure, it is not clear whether inhibin functions to block activin binding to its receptor or by binding to an independent receptor that antagonizes the activin receptor at the level of intracellular signaling cascades. The other well-studied class of inhibitors of the TGF- β class of growth factors is follistatin (Robertson et al. 1987; Ueno et al. 1987), which inactivates activin by binding to it stoichiometrically (Nakamura et al. 1990). It is interesting to note that recent evidence indicates that follistatin plays a key role in promoting neurogenesis in *Xenopus* by blocking the activin pathway in ectodermal cells (Hemmati-Brevalou et al. 1994). Activin may be a general antagonist to neurogenesis in vertebrates as retinoic acid-induced neural differentiation of mouse P19 and neuroblastoma cell lines can be blocked by activin (Hashimoto et al. 1990). *sog* may play an analogous role during early dorsal ventral patterning in the *Drosophila* embryo, as it is expressed specifically in the neuroectoderm. Although the most obvious phenotype in *sog* mutant embryos is the failure to subdivide the dorsal region, various evidence (see above) indicates that *sog* is also required to antagonize *dpp* in the neuroectoderm. Neutralizing signaling by TGF- β class factors may therefore be a conserved feature of early neural development.

Components of the TGF- β signaling pathway have been highly conserved during evolution. For example, the homologous *Drosophila* Dpp and vertebrate BMP-4 proteins can substitute functionally for each other in early *Drosophila* development (Padgett et al. 1993) and in vertebrate bone morphogenesis (Sampath et al. 1993). In addition, the *Drosophila* *tld* gene, which functions in concert with *dpp* to pattern the dorsal region of the embryo, encodes a product homologous to vertebrate BMP-1, which binds to and may process TGF- β family members of the BMP subtype. It will be interesting to determine the extent to which these pathways are conserved as coherent units and, more specifically, whether there is a homolog of *sog* that functions in vertebrate development.

Materials and methods

Fly stocks

The *sog*^{P1} allele was isolated as an insertion of a P element carrying a *deadpan* promoter-*lacZ* fusion cloned into the pCaSpeR-AUG- β -gal vector [marked with *white*⁺ (*w*⁺)]. To test whether the *sog*^{P1} P-element insertion is responsible for the X-chromosomal linked lethality associated with the *sog*^{P1} chromosome, females of the genotype *sog*^{P1}/*FM7a w*^a were crossed to males of the genotype *FM7a w*^a; Δ 2-3 *Sb*/+. Females of the genotype *sog*^{P1}/*FM7a w*^a; Δ 2-3 *Sb*/+ were recovered from this cross and mated individually to *FM7c* males in 130 pair matings. The progeny of these crosses were screened for either *w*⁻ non-*FM7c* males (viable P-element excisions) or for pale-eyed female progeny (*w*⁻/*w*^a). Twenty-two viable *w*⁻ males and 20 *w*⁻/*FM7c w*^a females were recovered. Each *w*⁻/*FM7c w*^a female was then crossed to *FM7c* males to establish stocks. Of these lines, 19 of 20 are homozygous viable and 1 line is ho-

mozygous lethal. Southern blot analysis of seven viable revertant lines demonstrates that these viable revertants have a wild-type pattern of hybridizing *sog* genomic restriction fragments, indicating that clean P-element excisions result in viable revertants of *sog^{P1}* lethality. The *sog^{U2}* allele, on the other hand, is associated with a deletion of P-element and flanking genomic sequences (see Fig. 3A). The *sog⁶* allele, isolated in the screen for embryonic patterning mutants (Wieschaus et al. 1984), was obtained from the Bloomington Indiana Stock Center. Other balancers and chromosomal markers (Lindsley and Grell 1968; Lindsley and Zimm 1992) were obtained from either the Bloomington Indiana Stock Center or the Bowling Green Stock Center.

PCR amplification of genomic DNA flanking the *sog^{P1}* insertion and isolation of *sog* genomic phage clones

DNA was isolated from *sog^{P1}* flies, digested with *AseI*, and circularized with ligase. Divergent P-element primers were used on this circularized template to amplify with PCR a 700-bp fragment including 500 bp of flanking *sog* genomic DNA. This PCR fragment was labeled and used to screen a λ -Dash genomic library. The filters were counterscreened with a P-element probe, as this library is a rich source of P-element sequences. Three overlapping genomic phage were recovered, and *EcoRI* subfragments from each of these phage were subcloned into Bluescript. The extent and orientation of the *sog* transcription unit contained within these clones was determined by in situ hybridization to whole-mount embryos using digoxigenin-labeled single-stranded RNA probes to each strand of the subclones. Sequence analysis of this genomic DNA identified several potential coding regions based on the presence of open reading frames and on codon usage bias. cDNAs were isolated using a short genomic probe to one of these likely coding regions. A fragment from the 3' end of one of these cDNAs was used to isolate additional overlapping genomic phage from the *sog* locus. Genomic phage clones were mapped, and probes synthesized from these phage were used for Southern blot analysis to determine the nature of lesions in the *sog* locus in various *sog* mutant flies.

Saturation screening the Nick Brown plasmid library

sog cDNAs containing the 5' end of the transcript are not abundant in any library that we have screened. Therefore, we screened the complex 0- to 4-hr Nick Brown plasmid cDNA library to saturation by fractionating it into 50 pools containing ~50,000 different plasmids each. DNA was prepared from each pool, digested with *NotI* (*NotI* cuts once within the PNB40 vector), electrophoresed on a 1% agarose gel, and analyzed by Southern blot using probes synthesized from various *sog* genomic subclones. One probe, synthesized from a 310-bp fragment lying 630–940 bp 3' to the site of P-element insertion, contains a likely open reading frame based on codon usage bias. This probe hybridized to six different bands on Southern blots of the library subfractions, ranging in insert size from 3.0–4.5 kb. Subfractions corresponding to lanes with specific hybridizing bands were then plated at low density (5,000–10,000 colonies per plate) and screened on duplicate filters. Representative clones for each different insert size were repurified and fine structure restriction mapped. All clones but one have a common restriction map. The 5' end of the longest cDNA (4.5 kb) maps near the beginning of the *sog* transcription unit (<500 bp), and sequence analysis reveals that the 3' ends of cDNAs correspond to initiation of reverse transcription at one of two short poly(A) stretches found within the transcript. The remaining 2.5 kb

comprising the 3' end of the 7-kb mRNA maps immediately downstream of the 3' end of the longest cDNA on the genomic walk. The longest cDNA was sequenced on both strands using the dideoxy chain termination method on double-stranded and single-stranded DNA templates.

Double label in situ hybridization

The method that we have developed for double in situ hybridization with digoxigenin- and biotin-labeled RNA probes is described in detail in O'Neill and Bier (1994).

Acknowledgments

We thank Elizabeth Little and Russell Doolittle for help with computer analysis to identify the *Sog* CR repeats and sequence similarity with the procollagen domain and other proteins displayed in Figure 4C. We thank Michael Levine, Kathryn S. Burton, Mark A. Sturtevant, and Margo Roark for critical comments on the manuscript, Tony Ip for helpful discussions, Nickolina Cataulina for tirelessly optimizing color printing, and Kathryn S. Burton for assembling the figures. This work was supported by National Institutes of Health grant RO1-NS29870-01, and research grant 5-FY92-1175 from the March of Dimes Birth Defects Foundation. V.F. was supported in part by grants from the Centre National de la Recherche Scientifique and NATO. E.B. was supported by funds from the McKnight Neuroscience Foundation, Sloan Foundation, and an American Cancer Society Junior Faculty Award.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Arora, K. and C. Nüsslein-Volhard. 1992. Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes. *Development* **114**: 1003–1024.
- Arora, K., M. Levine, and M.B. O'Connor. 1994. The *screw* gene encodes a ubiquitously expressed member of the TGF- β family required for specification of dorsal cell fates in the *Drosophila* embryo. *Genes & Dev.* (this issue).
- Basler, K. and G. Struhl. 1994. Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein. *Nature* **368**: 208–214.
- Basler, K., T. Edlund, T.M. Jessell, and T. Yamada. 1993. Control of cell pattern in the neural tube: Regulation of cell differentiation by dorsalin-1, a novel TGF beta family member. *Cell* **73**: 687–702.
- Bier, E., L.Y. Jan, and Y.N. Jan. 1990. *rhomboid*, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*. *Genes & Dev.* **4**: 190–203.
- Bornstein, P., K. O'Rourke, K. Wikstrom, F.W. Wolf, R. Katz, P. Li, and V.M. Dixit. 1991. A second, expressed thrombospondin gene (*Thbs2*) exists in the mouse genome. *J. Biol. Chem.* **266**: 12821–12824.
- Bradham, D.M., A. Igarashi, R.L. Potter, and G. Grotendorst. 1991. Connective tissue growth factor: A cysteine-rich secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10. *J. Cell Biol.* **114**: 1285–1294.
- Brummel, T.J., V. Twombly, G. Marqués, J.L. Wrana, S.J.

- Newfield, L. Attisano, J. Massagué, M.B. O'Connor, and W.M. Gelbart. 1994. Characterization and relationship of Dpp receptors encoded by the *saxophone* and *thick veins* genes in *Drosophila*. *Cell* **78**: 251–261.
- Brunner, A., J. Chinn, M. Neubauer, and A.F. Purchio. 1991. Identification of a gene family regulated by transforming growth factor. *DNA Cell Biol.* **10**: 293–300.
- Capovilla, M., M. Brandt, and J. Botas. 1994. Direct regulation of *decapentaplegic* by *Ultrabithorax* and its role in *Drosophila* midgut morphogenesis. *Cell* **76**: 461–475.
- Cavener, D.R. 1987. Comparison of the consensus sequence flanking translation start sites in *Drosophila* and vertebrates. *Nucleic Acids Res.* **15**: 1353–1361.
- Chasan, R. and K. Anderson. 1993. Maternal control of dorsal-ventral polarity and pattern in the embryo. In *The Development of Drosophila melanogaster* (ed. M. Bate and A. Martinez-Arias), vol. 1, pp. 387–424. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Chu, M.-L., W. de Wet, M. Bernard, J.-F. Ding, M. Morabito, J. Meyers, C. Williams, and F. Ramirez. 1984. Human pro-alpha-1(I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. *Nature* **310**: 337–340.
- Doyle, H.J., K. Harding, T. Hoey, and M. Levine. 1986. Transcripts encoded by a homeobox gene are restricted to dorsal tissues of *Drosophila* embryos. *Nature* **323**: 76–79.
- Ferguson, E.L. and K.V. Anderson. 1992a. Localized enhancement and repression of the activity of the TGF- β family member, *decapentaplegic*, is necessary for dorsal-ventral pattern formation in the *Drosophila* embryo. *Development* **114**: 583–597.
- . 1992b. *Decapentaplegic* acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell* **71**: 451–461.
- Findlay, J.K. 1993. An update on the roles of inhibin, activin and follistatin as local regulators of folliculogenesis. *Biol. Reprod.* **48**: 15–23.
- Green, J.B. 1993. Mesodermal growth factor candidates elected. *BioEssays* **15**: 129–130.
- Hashimoto, M., S. Kondo, T. Sakuri, Y. Etoh, H. Shibai, and M. Muramatsu. 1990. Activin/EDF as an inhibitor of neural differentiation. *Biochem. Biophys. Res. Comm.* **173**: 193–200.
- Hemmati-Brivanlou, A. and D.A. Melton. 1994. Inhibition of activin receptor signaling promotes neuralization in *Xenopus*. *Cell* **77**: 273–281.
- Hemmati-Brivanlou, A., O.G. Kelly, and D.A. Melton. 1994. Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell* **77**: 283–295.
- Hillier, S.G. and F. Miro. 1993. Inhibin, activin, and follistatin. Potential roles in ovarian physiology. *Ann. New York Acad. Sci.* **687**: 29–38.
- Hultmark, D., R. Klemenez, and W. Gehring. 1986. Translational and transcriptional control elements in the untranslated leader of the heat shock gene *hsp22*. *Cell* **44**: 429–438.
- Hunter, D.D., V. Shah, J.P. Merlie, and J.R. Sanes. 1989. A laminin-like adhesive protein concentrated in the synaptic cleft of the neuromuscular junction. *Nature* **338**: 229–234.
- Hursh, D.A., R.W. Padgett, and W.M. Gelbart. 1993. Cross regulation of *decapentaplegic* and *Ultrabithorax* transcription in the visceral mesoderm of *Drosophila*. *Development* **117**: 1211–1222.
- Ip, Y.T., R.E. Park, D. Kosman, E. Bier, and M. Levine. 1992. The dorsal gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes & Dev.* **6**: 1728–1739.
- Irish, V.F. and W.M. Gelbart. 1987. The *decapentaplegic* gene is required for dorsal-ventral patterning of the *Drosophila* embryo. *Genes & Dev.* **1**: 868–879.
- Kingsley, D.M. 1994a. The TGF-beta superfamily: New members, new receptors, and new genetic tests of function in different organisms. *Genes & Dev.* **8**: 133–146.
- . 1994b. What do BMPs do in mammals? Clues from the mouse short ear mutation. *Trends Genet.* **10**: 16–21.
- Kosman, D., Y.T. Ip, M. Levine, and K. Arora. 1991. Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254**: 118–122.
- Kulozik, M., A. Hogg, B. Lankat-Buttgereit, and T. Krieg. 1990. Co-localization of transforming growth factor β 2 with procollagen mRNA in tissue sections of patients with systemic sclerosis. *J. Clin. Invest.* **86**: 917–922.
- Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**: 105–132.
- Landry, S.J. and L.M. Gierasch. 1991. Recognition of nascent polypeptides for targeting and folding. *Trends Biochem. Sci.* **16**: 159–163.
- Lee, M.M. and P.K. Donahoe. 1993. Mullerian inhibiting substance: A gonadal hormone with multiple functions. *Endocr. Rev.* **14**: 152–164.
- Leptin, M. 1991. *twist* and *snail* as positive and negative regulators during *Drosophila* mesoderm development. *Genes & Dev.* **5**: 1568–1576.
- Lindsley, D.L. and E.H. Grell. 1968. *Genetic variations in Drosophila melanogaster*. *Carnegie Inst. Washington Publ.* **27**.
- Lindsley, D.L. and G.G. Zimm. 1992. *The genome of Drosophila melanogaster*. Academic Press, San Diego, CA.
- Massagué, J. and A. Pandiella. 1993. Membrane anchored growth factors. *Annu. Rev. Biochem.* **62**: 515–541.
- Mason, E.D., K.D. Konrad, C.D. Webb, and L.J. Marsh. 1994. Dorsal midline fate in *Drosophila* embryos requires *twisted gastrulation*, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes & Dev.* **8**: 1489–1501.
- McAnulty, R.J., J.S. Campa, A.D. Cambrey, and G.J. Laurent. 1991. The effect of transforming growth factor β on rates of procollagen synthesis and degradation in vitro. *Biochem. Biophys. Acta* **1091**: 231–235.
- Murata, M., Y. Eto, H. Shibai, M. Sakai, and M. Muramatsu. 1988. Erythroid differentiation factor is encoded by the same mRNA as that of the inhibin beta chain. *Proc. Natl. Acad. Sci.* **85**: 2434–2438.
- Murphy-Ullrich, J.E., S. Schultz-Cherry, and M. Hook. 1992. Transforming growth factor- β complexes with thrombospondin. *Mol. Biol. Cell.* **3**: 181–188.
- Murray-Rust, J., N.Q. McDonald, T.L. Blundell, M. Hosang, C. Oefner, F. Winkler, and R.A. Bradshaw. 1993. Topological similarities in TGF-beta-2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure* **1**: 153–159.
- Nakamura, T., K. Takio, Y. Eto, H. Shibai, K. Titani, and H. Sugino. 1990. Activin-binding protein from rat ovary is follistatin. *Science* **247**: 836–838.
- Nakatsukasa, H., R.P. Everts, C.C. Hsia, and S.S. Thorgeirsson. 1990. Transforming growth factor- β 1 and type I procollagen transcripts during regeneration and early fibrosis of rat liver. *Lab. Invest.* **63**: 171–180.
- Nellen, D., M. Affolter, and K. Basler. 1994. Receptor serine/threonine kinases implicated in the control of *Drosophila* body pattern by *decapentaplegic*. *Cell* **78**: 225–237.
- O'Neill, J.W. and E. Bier. 1994. Double in situ hybridization

- using biotin and digoxigenin tagged RNA probes. *BioTechniques* (In press).
- Padgett, R.W., R.D. St. Johnson, and W.M. Gelbart. 1987. A transcript from a *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor- β family. *Nature* **325**: 81–84.
- Padgett, R.W., J.M. Wozney, and W.M. Gelbart. 1993. Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo. *Proc. Natl. Acad. Sci.* **90**: 2905–2909.
- Panganiban, G.E., R. Reuter, M.P. Scott, and F.M. Hoffmann. 1990. A *Drosophila* growth factor homolog, *decapentaplegic*, regulates homeotic gene expression within and across germ layers during midgut morphogenesis. *Development* **110**: 1041–1050.
- Paralkar, V.M., R.G. Hammonds, and A.H. Reddi. 1991. Transforming growth factor beta type 1 binds to collagen IV of basement membrane matrix: Implications for development. *Dev. Biol.* **143**: 303–308.
- Penton, A., Y. Chen, K. Stachling-Hampton, J.L. Wrana, L. Attisano, J. Szidonya, J.A. Cassill, J. Massagué, and F.M. Hoffmann. 1994. Identification of two bone morphogenetic protein type I receptors in *Drosophila* and evidence that *brk25D* is a *decapentaplegic* receptor. *Cell* **78**: 239–250.
- Ray, R.P., K. Arora, C. Nüsslein-Volhard, and W.M. Gelbart. 1991. The control of cell fate along the dorsal-ventral axis of the *Drosophila* embryo. *Development* **113**: 35–54.
- Reuter, R., G.E. Panganiban, F.M. Hoffmann, and M.P. Scott. 1990. Homeotic genes regulate the spacial expression of putative growth factors in the visceral mesoderm of *Drosophila* embryos. *Development* **110**: 1031–1040.
- Ritzenthaler, J.D., R.H. Goldstein, A. Fine, and B.D. Smith. 1993. Regulation of the alpha-1(I) collagen promoter via a transforming growth factor-beta activation element. *J. Biol. Chem.* **268**: 13625–13631.
- Robertson, D.M., R. Klein, F.L. De Vos, R.I. McLachlan, E.H. Wettenthal, M.T.W. Hearn, H. Burger, and D.M. De Krester. 1987. The isolation of polypeptides with FSH suppressing activity from bovine follicular fluid which are structurally different to inhibin. *Biochem. Biophys. Res. Commun.* **149**: 744–749.
- Rohde, K. and P. Bork. 1993. A fast sensitive pattern-matching approach for protein sequences. *Comp. Appl. Biosci.* **9**: 183–189.
- Roth, S., D. Stein, and C. Nüsslein-Volhard. 1989. A gradient of nuclear localization of the Dorsal protein determines dorso-ventral pattern in the *Drosophila* embryo. *Cell* **59**: 1189–1202.
- Rushlow, C. and M. Levine. 1990. Role of the *zerknüllt* gene in dorsal-ventral pattern formation in *Drosophila*. *Adv. Genet.* **27**: 277–307.
- Rushlow, C.A., M. Frasch, H. Doyle, and M. Levine. 1987. Maternal regulation of *zerknüllt*: A homeobox gene controlling differentiation of dorsal tissues in *Drosophila*. *Nature* **330**: 583–586.
- Rushlow, C.A., K. Han, J.L. Manley, and M. Levine. 1989. The graded distribution of the *dorsal* morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* **59**: 1165–1177.
- Sampath, T.K., K.E. Rashka, J.S. Doctor, R.F. Tucker, and F.M. Hoffmann. 1993. *Drosophila* transforming growth factor beta superfamily proteins induce endochondrial bone formation in mammals. *Proc. Natl. Acad. Sci.* **90**: 6004–6008.
- Schonbaum, C.P., E.L. Organ, S. Qu, and D. Cavener. 1992. The *Drosophila melanogaster stranded at second (sas)* gene encodes a putative epidermal cell surface receptor required for larval development. *Dev. Biol.* **151**: 431–445.
- Shimell, M.J., E.L. Ferguson, S.R. Childs, and M.B. O'Connor. 1991. The *Drosophila* dorsal-ventral patterning gene *tolloid* is related to human bone morphogenetic protein 1. *Cell* **67**: 469–481.
- Simmons, D.L., D.B. Levi, Y. Yannoni, and R.L. Erikson. 1989. Identification of a phorbol ester-repressible *v-src*-inducible gene. *Proc. Natl. Acad. Sci.* **86**: 1178–1182.
- Smith, J.C., V. Cunliffe, J.B. Green, and H.V. New. 1993. Intercellular signaling in mesoderm formation during amphibian development. *Philos. Trans. R. Soc. Lond.* **340**: 287–296.
- St. Johnson, R.D. and W.M. Gelbart. 1987. Decapentaplegic transcripts are localized along the dorsal-ventral axis of the *Drosophila* embryo. *EMBO J.* **6**: 2785–2791.
- Steward, R. 1989. Relocalization of the *dorsal* protein from the cytoplasm to the nucleus correlates with its function. *Cell* **59**: 1179–1188.
- Tabata, T. and T.B. Kornberg. 1994. Hedgehog is a signaling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**: 89–102.
- Tolsma, S.S., O.V. Volpert, D.J. Good, W.A. Frazier, P.J. Polverini, and N. Bouck. 1993. Peptides from two separate domains of the matrix protein thrombospondin-1 have anti-angiogenic activity. *J. Cell Biol.* **122**: 497–511.
- Ueno, N., N. Ling, S.Y. Ying, F. Esch, S. Shimasaki, and R. Guillemin. 1987. Isolation and partial characterization of follistatin: A single chain Mr 35,000 monomeric protein that inhibits the release of follicle-stimulating hormone. *Proc. Natl. Acad. Sci.* **84**: 8282–8286.
- Verweij, C.L., P.J. Diergaarde, M. Hart, and H. Pannekoek. 1986. Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably larger than the mature vWF. *EMBO J.* **5**: 1839–1847.
- Wang, X.F., H.Y. Lin, E. Ng-Eaton, J. Downward, H.F. Lodish, and R.A. Weinberg. 1991. Expression cloning and characterization of the TGF-beta type III receptor. *Cell* **67**: 797–805.
- Wieschaus, E., C. Nüsslein-Volhard, and G. Jürgens. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's Arch. Dev. Biol.* **193**: 296–307.
- Wharton, K.A., R.P. Ray, and W.M. Gelbart. 1993. An activity gradient of *decapentaplegic* is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**: 807–822.
- Xie, T., A.L. Finelli, and R.W. Padgett. 1994. The *Drosophila saxophone* gene: A serine threonine kinase receptor of the TGF-beta superfamily. *Science* **263**: 1756–1759.
- Zusman, S.B., D. Sweeton, and E.F. Wieschaus. 1988. *short gastrulation*, a mutation causing delays in stage specific cell shape changes during gastrulation in *Drosophila melanogaster*. *Dev. Biol.* **129**: 417–427.



Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the short gastrulation gene.

V Francois, M Solloway, J W O'Neill, et al.

Genes Dev. 1994, **8**:

Access the most recent version at doi:[10.1101/gad.8.21.2602](https://doi.org/10.1101/gad.8.21.2602)

References

This article cites 72 articles, 25 of which can be accessed free at:
<http://genesdev.cshlp.org/content/8/21/2602.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Dharmacon[™] Reagents
Custom synthesis, RNAi, and CRISPR solutions

Infinite Reliability

More

horizon
a PerkinElmer company

The advertisement features a dark background with a colorful, abstract image of what appears to be a DNA double helix or a similar molecular structure. The text is arranged in a clean, modern layout.