A functional homologue of goosecoid in Drosophila

Anne Goriely 1, Michael Stella 1, Catherine Coffinier 1, Daniel Kessler 2, Carolina Mailhos 1, Scott Dessain 3 and Claude Desplan 1, *

1 Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021, USA
2 Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA
3 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

* Author for correspondence (e-mail: desplan@rockvax.rockefeller.edu)

SUMMARY

We have cloned a Drosophila homologue (D-gsc) of the vertebrate homeobox gene goosecoid (gsc). In the Gsc proteins, the pressure for conservation has been imposed on the homeodomain, the functional domain of the protein: sequence homology is limited to the homeodomain (78% identity) and to a short stretch of 7 aminoacids also found in other homeproteins such as Engrailed. Despite this weak homology, D-gsc is able to mimic gsc function in a Xenopus assay, as shown by its ability to rescue the axis development of a UV-irradiated embryo. Moreover, our data suggest that the position of insect and vertebrate gsc homologues within a regulatory network has also been conserved: D-gsc expression is controlled by decapentaplegic, orthodenticle, sloppy-paired and tailless whose homologues control gsc expression (for BMP4 and Otx-2), or are expressed at the right time and the right place (for XFKH1/Pintallavis and Tlx) to be interacting with gsc during vertebrate development. However, the pattern of D-gsc expression in ectodermal cells of the nervous system and foregut cannot easily be reconciled with that of vertebrate gsc mesodermal expression, suggesting that its precise developmental function might have diverged. Still, this comparison of domains of expression and functions among Gsc proteins could shed light on a common origin of gut formation and/or on basic cellular processes. The identification of gsc target genes and/or other genes involved in similar developmental processes will allow the definition of the precise phylogenetic relationship among Gsc proteins.

Key words: Drosophila, goosecoid, Xenopus, gastrulation, foregut, homeodomain, phylogeny

INTRODUCTION

Formation of the anteroposterior (A-P) axis is a crucial developmental event. In Drosophila, the maternal morphogen Bicoid (Bcd) is responsible for patterning the anterior part of the embryo by differentially regulating its zygotic target genes in a concentration-dependent manner (Driever and Nusslein-Volhard 1988; Struhl et al., 1989). In vertebrates, the molecular mechanisms underlying axis formation are less well-characterized. The axis is laid down during the early stages of development and the anteroposterior and dorsoventral polarities are intimately associated with the movements of gastrulation that originate from the blastopore lip in Xenopus, primitive streak in chick and mouse and embryonic shield in zebrafish. Although it has been known since the beginning of the century that a transplanted blastopore lip can organize a secondary body axis in amphibian embryos, a region named Spemann organizer, this process is poorly understood at the molecular level. Recently, however, a number of vertebrate genes that exhibit sequence homology with Drosophila molecules (otx, emx, Dlx, BMPs, Wnts...) have been identified as putative mediators of early patterning events (reviewed in De Robertis et al., 1994; Kessler and Melton 1994; De Robertis, 1995).

The Xenopus goosecoid (gsc) gene has been proposed to play an important role in early vertebrate development (Blumberg et al., 1991; Cho et al., 1991; Niehrs et al., 1993). gsc is strongly expressed in the Spemann organizer just before the onset of gastrulation, suggesting a role in mediating the remarkable inductive properties elicited by this region of the embryo. Moreover, gsc ability to induce morphogenetic movements and axial structures when injected in ventral regions of the Xenopus embryo made Gsc a good candidate for a molecule involved in patterning anterodorsal structures of the tadpole. Highly conserved gsc homologues have been identified in chick, zebrafish and mouse. In these species, gsc is also expressed early in embryogenesis in regions homologous in function to the Spemann organizer at a time when these regions have their maximal inductive ability (Cho et al., 1991; Blum et al., 1992; IpsiIusa Belmonte et al., 1993; Stachel et al., 1993). As for the Drosophila morphogen Bcd, it has been suggested that the ability of Gsc to dorsalize tissues is achieved in a dose-dependent manner (Niehrs et al., 1994). The analogy with Bcd has also a molecular basis, which is revealed by its name: Gsc is a transcription factor containing a homeodomain (HD) which has a similar DNA-binding specificity to that of Bcd. This DNA-binding specificity is mediated by the residue at position...
50 of the HD, a lysine (K50) in the case of both Bed and Gsc (reviewed in Duboule, 1994).

However, despite these very promising early results, the recent reports on the phenotype associated with gsc disruption in mouse show that gsc is dispensable for early vertebrate development. gsc developmental requirement seems to be limited to its later phase of expression, as the mutant mice only exhibit craniofacial and rib defects (Rivera-Perez et al., 1995; Yamada et al., 1995). These findings have lead to the suggestion that there could exist another redundant gsc-like activity in the early vertebrate embryo.

In this paper, we describe the identification of a Drosophila homeogene, D-goosecoid (D-gsc), sharing restricted sequence homology with gsc. We show that this gene is a functional homologue of gsc, as it is able to mimic the patterning effects of gsc in a Xenopus assay, a function that cannot be fulfilled by the morphogen Bicoid. D-gsc embryonic expression pattern occurs in two discrete domains in cells fated to become part of the nervous system (Central and Stomatogastric Nervous System) and foregut. These expression data appear to be difficult to reconcile with the vertebrate gsc expression mostly confined to mesodermal tissues. However, the later expression of D-gsc in the stomodeum might shed light on a common system (Central and Stomatogastric Nervous System) and foregut. These expression data appear to be difficult to reconcile with the vertebrate gsc expression mostly confined to mesodermal tissues. However, the later expression of D-gsc in the stomodeum might shed light on a common

MATERIALS AND METHODS

D-gsc cDNA isolation
A 4-8 hours Drosophila embryonic pNB40 library (kindly provided by Nick Brown) was screened with a 160 bp EcoRV-PstI fragment isolated from 60Mun1 (generous gift from Bill McGinnis) containing 69 bp of a homebox coding sequence (from bp 1457 to 1526 in the cDNA sequence, Fig. 1), which corresponds to the C-terminal part of the homeodomain (HD), plus 27 bp directly following it. Four clones were isolated during a low stringency screening and restriction analysis revealed that they were all identical. Sequencing on both strands of the entire length of one of the clone (D3A1) was performed by dyeoxy chain termination using Sequenase version 2.0 (USB). In order to map the position of the introns, we compared the respective lengths of several PCR fragments amplified from D-gsc cDNA and from genomic DNA of two EMBL3 lambda phages that we isolated in the region (not shown). The fragments that appeared to be longer when amplified from genomic DNA – by comparison to the cDNA template – were subcloned using TA cloning kit strategy (Invitrogen) and the position of the intron-exon boundaries were obtained by sequencing.

Frog experiment
Xenopus eggs were collected and irradiated for 75 seconds with ultraviolet light, 30 minutes postfertilization (see Steinbeisser et al., 1993 for details). At the 4- to 8-cell stage, a single blastomere was microinjected with capped synthetic RNA for bicoid (50 pg), X-gsc (50 pg), D-gsc (50 pg) or Xwnt-8 (4 pg). For preparation of the synthetic mRNAs and microinjection, see Cho et al. (1991). lacZ RNA (50 pg) was included in each sample in order to trace the lineage of the injected cells. 25 embryos were scored on the DAI (DorsoAnterior Index) at stage 35, where 0 to 4 is a measure of axis deficiency (ventralized phenotype) and 6 to 10 represents dorsoanterior enhanced embryos, (5 is wildtype) (Kao and Elision, 1988) and the average was established. The lacZ staining was revealed using X-gal as described in Smith and Harland (1991).

Antibody staining and in situ hybridization
In situ hybridization was performed with digoxigenin-labeled RNA probes (Bohringer Mannheim) as described in Simpson Brose et al. (1994). A 445 bp N-terminal BaIII/II (position 565 in cDNA sequence)-PstI (1018 fragment) (gsc BP) was used to generate the gsc probe. This construct did not include the gsc homeobox and was used to avoid cross-reactivity with other homeobox-containing genes. The double-staining gsc RNA/even-skipped antibody procedure is as described in Simpson Brose et al. (1994).

Mutant fly stocks and identification of mutant embryos
The mutant lines that we used are as follows: wildtype are w+ flies, giant (gf1ABB), buttonhead (bd5), empty-spiracles (ems225), wingless (wg1225), Kruppel (Kr5), sine oculis (mda lethal3), Distal-less (Dr4), rhomboid (rbo11), short gastrulation (sog32), huckebein (hkb3), forkhead (fkh343), orthodenticle (otio101), sloopy-paired (slpa34), decapentaplegic (dpp10), and tailless (tl1). In most cases, the mutant embryos were unambiguously identified by using marked lacZ-balancers (FM7-fztlacZ, CyO-bhlaCZ or TM3-vealaCZ) and cohybridizing the embryos with D-gsc RNA (gsc BP) and lacZ RNA probes. tl1 mutant embryos were stained with an Anti-Eve-skipped antibody (brown staining); in tl1 embryos, Eve stripes 6 and 7 are joined together. slpa34B allele is on the CyO balancer and mutant embryos were identified by the absence of staining for slpa1 and slp2 using DIG-RNA probes (Grossniklaus et al., 1992); dpp mutants were recognized based on morphological criteria.

RESULTS

Cloning of a homebox gene exhibiting the same DNA-binding specificity as gsc
We isolated four identical cDNAs by screening a Drosophila embryonic library with a probe containing the 3’ end of a divergent homebox sequence, originally isolated during a low stringency screen for Drosophila homeobox-containing genes (Dessain and McGinnis, 1993). Our interest in this short protein sequence (14 amino acids) arose from the presence of a lysine at the position 50 (K50) of its partial homeodomain, a landmark of proteins encoded by genes like bicoid (Driever and Nusslein-Volhard, 1988), orthodenticle (Finkelstein et al., 1990) and sine oculis (Cheyette et al., 1994) that are involved in patterning the anterior region of the Drosophila embryo (Duboule, 1994). The sequence of the isolated cDNA is shown on Fig. 1. From the first methionine (bp #465), the coding region of this clone produces a 419 amino acid protein (Fig. 1). However, a second ATG at bp #477 might be the correct initiation codon since it is in better agreement with the Drosophila consensus for translation start (CAAAATTTG) (Cavener, 1987). Comparison with the genomic sequence (not shown) reveals the presence of a TATA box and a potential transcription start consensus in the sequence just preceding (60 bp upstream) the beginning of the cDNA, supporting the interpretation that this clone is a nearly full-length cDNA.

The deduced protein sequence revealed the presence of a
homeodomain (HD) at residue position 284-345 that contains a lysine at position 50 (K50 at position 335). Fig. 2 shows a sequence comparison between the D-gsc translation product and those of the vertebrate gsc genes. The D-Gsc homeodomain sequence shows an obvious homology to that of vertebrate Goosecoid (Gsc) (78% identity). In particular, the lysine at position 50, responsible for conferring its unusual DNA-binding specificity (Hanes and Brent, 1989; Treisman et al., 1989), is conserved. This sequence homology suggests that we have cloned a *Drosophila* homologue of gsc and we therefore named this clone *D-gsc*. The vertebrate Gsc proteins are relatively small molecules of 240-250 amino acids which are remarkably conserved along their entire length (77-85% identity outside the homeodomain and 98-100% inside the homeodomain). By contrast, D-Gsc is a larger protein of 419 amino acids whose homology with its vertebrate counterparts is limited to the HD and to a small stretch (GEH for Gsc-Engrailed Homology in Fig. 2) of 7 amino acids located at the N-terminal part of the vertebrate proteins (residues 110-116 in D-Gsc). Interestingly, a similar sequence is found in several Pax genes both in vertebrate and *Xenopus* (Hemmati-Brivanlou et al., 1991; Logan et al., 1992; Aleyamma et al., 1995), where it is called the octapeptide (Noll, 1993). This short sequence is also present and conserved at a similar location in all the Engrailed homeoproteins (part of the ‘EH1 domain’) (Hemmati-Brivanlou et al., 1991; Logan et al., 1992; Aleyamma et al., 1995), where it appears to mediate part of their repression potential (J. Jaynes, personal communication). This is consistent with the fact that D-Gsc behaves as a strong repressor in a Schneider cell cotransfection assay (C. Mailhos and C. Desplan, unpublished data).

Both vertebrate (human, mouse and *Xenopus*) (Blum et al., 1994) and *Drosophila* gsc sequences are interrupted by 2 introns, one of them being located at a conserved position inside the homeodomain (between residue 47 and 48 of the HD).

Taken together these homology data strongly suggest that we have isolated a *Drosophila* homologue of goosecoid.

**D-gsc is a functional homologue of gsc**

We investigated the ability of *D-gsc* to mimic the well-documented effect of the vertebrate gsc molecules in *Xenopus* develop-

---

**Fig. 1.** Nucleotide and predicted amino acid sequences of *D-gsc* cDNA cloned from pNB40 embryonic 4-8 hours library. The sequence of the cDNA clone encompasses 2372 bp including 16 adenyl residues at the 3’ end. The ATG at the position 465-468 represents a putative start codon of the D-Gsc homeodomain-containing protein. However, the ATG at position 477-480 could be an alternative starting methionine (see results). A potential signal sequence for polyadenylation is underlined (2327-2334). An asterisk indicates the termination codon. The homeodomain (284-345 in the protein sequence) appears in bold and the lysine at position 50 (K50), landmark of homeodomain molecules like Gsc, Otd and Bcd, is conserved. The position of the 2 introns is indicated by a A sign and is located at position 1095 of the cDNA sequence (intron 1 of 1 kb) and 1458 (intron 2 of 8 kb in the middle of the homeodomain). Sequence accession number: X95420.
opment: gsc is able to rescue the axis polarity of a UV-irradiated Xenopus embryo (Cho et al., 1991). Embryos irradiated with ultraviolet light shortly after fertilization fail to develop an axis as shown by their DAI (Dorsal Anterior Index, see Materials & Methods) of 0.6 (Fig. 3A). At 4- to 8-cell stage, embryos were microinjected with RNA for either bicoid (Fig. 3B), Xwnt-8 (Fig. 3C), D-gsc (Fig. 3E,F) or X-gsc (Fig. 3D). The average DAI for these experiments is as follows: UV-irradiated embryos (not shown): 0.8; lacZ alone: 0.6; bicoid: 0.6; Xwnt-8: 4.2; X-gsc: 1.7; D-gsc: 1.5. As shown by these results, the injection of D-gsc mRNA (Fig. 3E,F) resulted in axial rescue similar to that obtained with Xenopus gsc mRNA (Fig. 3D). As for vertebrate gsc, cells injected with D-gsc were able to recruit neighboring cells into the axis, as indicated by the restriction of the lacZ staining to the most anterior region of the partial axis (see arrowhead in Fig. 3F). These data demonstrate that D-gsc is able to function alone in a Xenopus assay and is therefore a functional homologue of gsc. It should be noted that, despite similarity of binding specificity with Gsc, Bcd is unable to mimic Gsc function in the Xenopus assay, as the embryos injected with bicoid mRNA do not exhibit any sign of axis rescue (Fig. 3B).

D-gsc is expressed in two distinct domains during Drosophila embryogenesis

In the Drosophila embryos, D-gsc expression is first detected at the cellularization stage (2 hours of development at 25°C, stage 4) as a dorsoanterior stripe, 2-3 cells wide (Fig. 4A). D-gsc expression is reinforced and appears as a strong horseshoe-like pattern across the dorsal side of the embryo, 3-4 cell diameters anterior to the appearing cephalic furrow, now visible (Fig. 4B). At germband extension stage (4 hours, stage 8), D-gsc is quickly down-regulated in the dorsal-most region of the embryo (Fig. 4C), green arrow) and the horseshoe pattern resolves into a large cluster of cells on each side of the presumptive head (Fig. 4D, E, red arrow). These positive cells migrate laterally and, by stage 10 (6 hours), D-gsc is restricted to three smaller clusters of cells on each side of the embryo (Fig. 4F, red arrow). During the germband retraction stage (9-13 hours, stage 12-14), these D-gsc-expressing cells migrate inside the embryo and are associated with the brain hemispheres (Fig. 4G, red arrows).

D-gsc second domain of expression is seen as a unique anterior cluster of positive cells at the time the stomodeum is first visible (5 hours, stage 9, Fig. 4D, black arrowhead). This cell cluster follows the movements of foregut invagination and is incorporated into the roof of the stomodeum (black arrowhead in Fig. 4D to G). By the end of germband retraction (10 hours), the anterior cluster of D-gsc expressing cells abut the endodermic anterior midgut territory (esophagus). Based on their origin and their location at the end of embryogenesis, these cells appear to be part of the anterior foregut, the ring gland and the stomatogastric nervous system (SNS) -
a tissue that is affected in a mutant lacking D-gsc function (Hahn and Jäckle, 1996).

**Control of D-gsc expression during embryogenesis**

Several molecules that are likely to interact with gsc have been identified in vertebrates, among which are BMP-4 (Jones et al., 1992; Fainsod et al., 1994; Winnier et al., 1995), Otx2 (Blitz and Cho, 1995; Pannese et al., 1995), chordin (Sasai et al., 1994), XFKHI/Pintallavis (Dirksen and Jamrich, 1992; Ruiz i Altaba and Jessell, 1992), Xnot/th (von Dassow et al., 1993; Talbot et al., 1995) and XLIM1 (Taira et al., 1993; Shawlot and Behringer, 1995). Most of these molecules are homologues to Drosophila genes whose mutations affect embryonic development. These mutants allowed us to investigate the control of D-gsc expression in Drosophila embryos. We selected the putative regulators of D-gsc expression based on their pattern of expression relative to D-gsc. For instance, the early D-gsc stripe overlaps the expression domains of orthodenticle (otd), sloppy-paired (slp), tailless (tll), empty spiracles (ems), rho (rhomboid), Distal-less (Dll) and abuts the buttonhead (btd) domain (data not shown), while wingless (wg), forkhead (fkh) and Krüppel (Kr) are known to be expressed in the invaginating stomodeum, the second site of D-gsc expression (reviewed in Skaer, 1993). Other mutants, such as decapentaplegic (dpp) or short gastrulation (sog) were analyzed because they are targets of the dorsoventral system and we have shown that D-gsc expression is restricted both ventrally and dorsally. Furthermore, interactions between dpp and sog homologues (BMP4 and chordin, respectively) and gsc have been described in the vertebrate system (Fainsod et al., 1994; Sasai et al., 1994).

Not surprisingly, the three maternal systems (anterior-posterior (bicoid); terminal (torso); dorsoventral (dorsal)) control the early expression of D-gsc (not shown): in embryos issued from bicoid− females, the early D-gsc stripe never appears, showing that D-gsc is a bicoid target. The horseshoe stripe is shifted anteriorly in embryos derived from torso− females, in such a way that it is now located closer to the anterior pole and appears thicker. The ventral repression of the stripe is abolished in embryos derived from dorsal− mothers, resulting in D-gsc being expressed around the circumference of the embryo. However, the maternal control of D-gsc expression is likely to be indirect, due to the time at which D-gsc first appears.

Zygotic mutations in giant, buttonhead, empty-spiracles, wingless, Krüppel, sine oculis, Distal-less, rhomboid, short gastrulation, Huckebein and forkhead did not affect the expression of D-gsc.

An effect on the early stripe of expression of D-gsc was observed in the sloppy-paired (slp), orthodenticle (otd), tailless (tll) and decapentaplegic (dpp) mutants. Both slp and otd affect D-gsc in a similar way: the early stripe of D-gsc appeared normally (Fig. 5A,C) but at the end of the cellularization stage, there was no reinforcement of its expression. This leads to the premature loss of D-gsc in otd− embryos (Fig. 5B) and to a very weak expression (with only a few D-gsc-expressing cells in the lateral region) in slp− embryos (Fig. 5D).

As seen in Fig. 5E, decapentaplegic (dpp) is necessary to bring about D-gsc repression in the dorsal-most region of the horseshoe, which does not split in the mutant embryos (Fig. 5EF). In tailless (tll−) mutant embryos, the horseshoe pattern of D-gsc is limited to a dorsal region which suggests that tll is required to promote D-gsc expression in the lateral region, or to prevent its repression by the dorsoventral patterning system (data not shown). All four of these mutants only affect the early horseshoe-like D-gsc expression. We have not been able to identify any zygotic gene that controls the second domain of D-gsc expression in the anterior cluster.

**DISCUSSION**

**D-gsc, a functional homologue of gsc, shares restricted homology with gsc**

We have cloned a homeobox gene whose 419 amino acids predicted protein product contains a homeodomain exhibiting 78% identity with the homeodomain of vertebrate Gsc (Fig. 2). Interestingly, these homeodomains share a lysine at position 50, a rare characteristic which confers a particular DNA-binding specificity to these types of molecules (Treisman et al., 1989). So far, only 5 such homeodomains have been described and are found in important developmental regulators (Bicoid (Drieve et al., 1988), Otd/Otx (Finkelstein et al., 1990; Simeone et al., 1993), Sineoculis/Six (Cheyette et al., 1994; Oliver et al., 1995); UNC-30 (Jin et al., 1994) and Gsc (Blumberg et al., 1991)). Outside this conserved region, only a small stretch of homology of 7 amino acids is found among Gsc proteins (GEH domain in Fig. 2), a region also present and conserved across species in other proteins such as Engrailed and which may be important in mediating the repression potential of these types of molecules (Logan et al., 1992; Han and Manley, 1993; Aleyamma et al., 1995; C. Mailhos and C. Desplan, unpublished data).

Despite this weak sequence conservation between vertebrate and Drosophila, we showed that D-gsc is able to mimic the effect of its vertebrate counterparts in a Xenopus assay. The injection of D-gsc mRNA is able to rescue the axis polarity of a UV-irradiated Xenopus embryo to a level similar to that of the X-gsc mRNA injection (Fig. 3). Taken together, these data strongly suggest that we have isolated a functional homologue of gsc in Drosophila.

**Specific effect of Gsc in the Xenopus rescue assay**

As the sequence conservation between D-Gsc and its vertebrate counterparts is mostly confined to the HD (78% identity), the functional rescue in the Xenopus assay emphasizes the crucial role of this DNA-binding motif in mediating the functional specificity of a HD-containing protein (McGinnis and Krumlauf, 1992). It is formally possible, however, that injected D-Gsc mimics the function of another homeoprotein present in the early frog embryo and exhibiting a similar DNA-binding specificity. Otx-2 could be such a protein: it not only contains a K50, but also belongs to the same class of HDs as Gsc (Prd/Pax class), i.e. it is able to bind DNA cooperatively as homo- or heterodimer with any members of the Paired-class HDs (Wilson et al., 1993, 1995). Bicoid is another K50 HD-containing protein, which exhibits a similar DNA-binding specificity to Gsc and Otx/Otd, even though it is only able to bind DNA as a monomer. However, these three K50 homeoproteins are likely to have different molecular functions: Gsc
behaves as a strong repressor while Bcd and Otd act as activators of transcription in a Schneider cell culture assay (C. Mailhos and C. Desplan, unpublished data). Moreover, Gsc, Bcd and Otx-2 have different effects when injected into UV-irradiated Xenopus embryos. We have shown that Bcd cannot mediate the axis rescue in a Xenopus assay (Fig. 3B), while otx-2 injection, in addition to the partial rescue of the ventralized phenotype, causes the production of a cement gland, the anterior-most structure of the embryo. The axis rescue that is observed upon injection of both otx-2 and gsc is in agreement with the fact that gsc expression is turned on following otx-2 injection (Pannese et al., 1995). Thus, it is likely that otx-2 axis rescue is mediated through gsc activation.

Taken together, these data strongly suggest that the effect of D-gsc in the Xenopus assay is intrinsically due to the Gsc molecule and not to the mere presence of a HD with a K50.

**Commonality of expression pattern among Gsc proteins**

D-gsc is expressed in two discrete domains during embryogenesis. The early cellularization horseshoe-like pattern gives rise to cells located in the brain hemispheres, while the second unique domain invaginates inside the stomodeum along with ectodermal cells fated to be a part of the stomatogastric nervous system (SNS), ring gland and foregut. gsc expression in vertebrates also appears in two distinct phases: the first, taking place in the organizer, is intimately associated with the gastrulation process (De Robertis et al., 1994), while the later one is restricted to the branchial arches, the limbs and the body walls (Gaunt et al., 1993). Despite their diphasic expression, at first sight, the comparison of gsc expression domains in Drosophila and vertebrates shows no obvious phylogenetic conservation. However, there could be an element of evolutionary conservation in the fact that gsc expression in both classes of organisms is associated with gastrulation. The comparison of the gastrulation processes is complicated by the fact that the vertebrates are Deuterostomes (the mouth forms as a secondary consequence far away from the blastopore, which gives rise to the anus), while insects are Protostomes (the mouth arises at or near the blastopore). In all Protostomes, the stomodeum is the anterior end of the blastopore. However, as the head of the Drosophila embryo invaginates during development, it is difficult to precisely locate its anterior portion. Comparative studies of arthropod embryology indicate that the cells invaginating in the stomodeum – and which express D-gsc – constitute the anterior-most structure of the embryo. The SNS is also believed to arise from the labrum, the most anterior segment of the embryo head (Schmidt-Ott et al., 1994). In vertebrates, gsc is precisely expressed in the dorsal lip of the blastopore, the region leading the invagination of the mesoderm during gastrulation and which will give rise to the head process and the anterior-most tissues of the vertebrate embryo (prechordal plate mesoderm) (Niehrs et al., 1993). Thus, gsc expression in Drosophila and vertebrates could be evolutionary related.

This commonality of expression can possibly be seen later in development as well, when gsc expression is found in the foregut of both flies (Fig. 4F,G, black arrowheads) and vertebrates. Indeed, gsc is expressed in the foregut of zebrafish (Stachel et al., 1993) and chick (Izpisua Belmonte et al., 1993). Moreover, injections of gsc mRNA in Xenopus induce cells to migrate into the anterior endomesoderm, occupying the entire foregut of the resulting tadpole (see red arrow in Fig. 3E,F where lacZ colocalizes with the gsc injected cells). Thus, gsc expression may shed light on a common origin of the foregut of these organisms, which would arise from the anterior-most region of the blastopore. This point will need to be confirmed.
**Fig. 4.** *D-gsc* expression during *Drosophila* embryonic development, as revealed by in situ hybridization with a *D-gsc* antisense RNA probe labelled with DIG-UTP. (A) Stage 4 embryo in dorsal view, *D-gsc* is expressed as an anterior stripe; (B) stage 6 in lateral view, horseshoe expression pattern of *D-gsc*; (C) stage 8, dorsal view, the horseshoe splits dorsally as indicated by the green arrow; (D) stage 9, the second domain of *D-gsc* expression (black arrowhead) appears as the stomodeum opens; (E) stage 10; (F) stage 11, the cells of the anterior domain invaginate into the stomodeum (black arrowhead); (G) stage 14, the early *D-gsc*-expressing cells are located on the brain hemispheres (red arrows), while the cells of the *D-gsc* anterior domain (black arrowhead) abut the anterior midgut territory; (H) stage 15. All the embryos are in lateral view with anterior to the left and dorsal to the top, except for A, C, G which are in dorsal view.

**Fig. 5.** *D-gsc* expression in mutant embryos for *orthodenticle* (*otd*Δ101) (A,B), *sloppy-paired* (*slp*Δ34B) (C,D), *decapentaplegic* (*dpp*HIN) (E,F) and *tailless* (*tll*1) (G,H). (A-H) In situ hybridization with a *D-gsc* antisense RNA probe labelled with DIG-UTP (blue staining). (G,H) Embryos have been double labelled with an anti-Evenskipped antibody (brown staining) in order to identify the mutant embryos (see Material and Methods). The early stripe of *D-gsc* expression is affected in all mutant combinations: (A) Ventral view, (C) dorsal view; stage 4; (B,D,E,G) stage 7-8; (F,H) dorsal view at stage 10. Anterior is to the left.
by comparing the expression of other genes involved in similar processes that have been conserved between insects and vertebrates (for example, fkh genes) and/or by comparing the functional requirement of Gsc proteins during development (see later).

Commonality of genetic control among Gsc proteins
Among all the zygotic mutants that we tested and which could be involved in regulating D-gsc, only four had an effect on D-gsc early expression, while none of them affected the later expression of D-gsc in the SNS/stomodeum anterior domain. The regulation of the early D-gsc pattern is, to a large extent, analogous to the situation observed in the vertebrate system where several genes have been shown, or are likely, to interact with gsc in the organizer region (fkh1, otx-2, BMP-4) or later in the head region (Tlx).

Both otd and slp expression overlap the early D-gsc stripe which is affected in mutants for these two genes. otx-2, one of two vertebrate homologues of otd (Blitz et al., 1995; Pannese et al., 1995), is expressed in the Spemann organizer at the same time and in a pattern very similar to that of gsc. Consistent with the observation that otd is required to maintain D-gsc early expression (Fig. 5B), the injection of otx-2 mRNA (mouse or Xenopus) activates gsc expression in a Xenopus animal cap assay (Pannese et al., 1995). Injection of gsc mRNA is also able to induce X-otx-2 expression (Blitz et al., 1995), suggesting that these two genes regulate each other during early development. The recent mouse knock-out of otx-2 should clarify the details of the epistasis between these two genes (Matsuo et al., 1995).

The slp locus encodes two highly related genes containing a HNF3/FKH domain (Grossniklaus et al., 1992), which is also present in XFKH1/Pintallavis, a gene known to be expressed in the Spemann organizer at the same time as gsc (Dirksen et al., 1992; Ruiz i Altaba et al., 1992). However, it is not yet known whether or not slp and XFKH1/Pintallavis are functional homologues and the effect of XFKH1/Pintallavis on gsc expression has not been documented so far.

dpp, a homologue of vertebrate BMP-4, is a gene that affects multiple patterning processes in Drosophila, including the specification of dorsal fate. In Xenopus, injection of BMP-4 results in the induction of ventral mesoderm and the inhibition of dorsoanterior structures which is associated with the repression of gsc expression (Fainsod et al., 1994). It has been suggested that the dorsoventral axes of vertebrates and invertebrates have been inverted during evolution even though the molecular mechanisms of dorsoventral patterning are conserved (Arendt and Nübler-Jung, 1994; Holley et al., 1995). In agreement with this proposition, we showed that dpp represses D-gsc in the dorsal-most region of the Drosophila embryo.

Finally, tailless (tl), a zygotic effector of the torso terminal system, encodes a transcription factor of the nuclear orphan receptor class. Tlx, a functional vertebrate homologue of tailless is expressed in the head region of mouse and chick embryos, rostral to the anterior gsc domain of expression (Liaw and Lengyel, 1993; Yu et al., 1994). An interaction between these two genes in the vertebrate system is therefore probable, but remains to be determined.

Based on sequence conservation and on the overlapping expression patterns of insects and vertebrates, we propose that not only otx2 and BMP-4, but also XFKH1/Pintallavis and Tlx, are involved in the control of gsc expression in vertebrates in a manner similar to their Drosophila homologues. These similarities allow us to suggest that the position of Drosophila and vertebrate gsc genes within a regulatory network has been conserved during evolution.

Commonality of biological function among Gsc proteins?
When gsc was first isolated from Xenopus dorsal blastopore lips, much attention was given to its early phase of expression in the organizer: its RNA distribution in this tissue, the nature of its protein product and the remarkable conservation of expression domains and sequence among vertebrates strongly suggested an important role for this homeoprotein in patterning the early embryo. Overexpression and molecular data confirmed that this molecule had indeed the potential for being a key regulator of vertebrate development (reviewed in De Robertis et al., 1994). However, the recent reports of gsc gene disruption in mouse show that the early phase of gsc expression in the organizer region is dispensable for mouse development. The phenotype observed in mutant mice is limited to craniofacial and rib defects which are associated with the second phase of gsc expression in the branchial arches and ventrolateral body walls (Rivera-Perez et al., 1995; Yamada et al., 1995). It is possible that a second gsc-like gene might mask the early developmental requirement for this locus. Consistent with this possibility, when a sequence containing the gsc homeobox is probed on a chick genomic DNA southern, two extra bands are observed, suggesting the existence of at least one related gene (Izpisua Belmonte et al., 1993). As in vertebrates, the absence of D-gsc in Drosophila appears to specifically affect the second phase of D-gsc expression in the stomodeum. The mutation in D-gsc results in the loss of one of the four stomatogastric ganglia (Hahn and Jäckle, 1996).

It is difficult to reconcile these data on mutant phenotypes. Examples of genes for which sequence conservation is not paralleled by an apparent functional conservation are multiple, i.e. the vertebrate T (Brachyury) gene is required for the formation of posterior mesoderm and for axial development, while its insect homologues are involved in development of the procotodeum (anal pads and hindgut) (Kispert et al., 1994). However, a commonality of biological function between vertebrate and insect Gsc may exist at the cellular level and could be reflected by the fact that the cells expressing gsc undergo morphogenetic movements. In vertebrates, early gsc-expressing cells take part in morphogenetic movements associated with gastrulation, in such a way that these cells migrate to the anterior end of the head process (Niehrs et al., 1993). gsc might therefore not be specifying a particular cell fate but rather may be involved in inducing cell fate changes and/or cell adhesion processes (De Robertis et al., 1990; Niehrs et al., 1993).

Finally, it might be worth noting that the deficiency phenotype of D-gsc (specifically, deficiency D-Gsc2) exhibits not only the SNS mutant phenotype seen in the P(D-gsc) allele, a P-element insertion that is likely to represent only a partial loss of function, but is also associated with an ‘abnormal midgut’ phenotype (Hahn and Jäckle, 1996). In Drosophila, the anterior gut is made up of ectodermal cells (foregut) and endodermal cells (anterior midgut) that arise from two separate primordia. The endodermal midgut primordium at the anterior tip of the ventral furrow and the stomodeum only fuse later in
development after they have moved into the embryo (reviewed in Skaer, 1993). This midgut phenotype could therefore reflect a requirement for D-gsc in cells involved in anterior gut morphogenesis. However, it still remains to be determined whether this developmental anomaly is specifically due to the absence of D-Gsc locus or to an adjacent gene (Hahn and Jäckle, 1996).

We are indebted to Bill McGinnis for the gift of the 60Mun1 homeobox sequence from which all this work originated. We wish to thank Jim Jaynes for drawing our attention to the sequence conservation between Engrailed and Gsc proteins, the members of the Desplan laboratory for their encouragement, Mounou Hahn and Herbert Jäckle for communicating their results prior to publication and also for the many enthusiastic interactions that we had with them. Finally, we wish to thank an anonymous reviewer for very constructive suggestions on an earlier version of this manuscript. A. G. was supported by an EMBO fellowship, C. M. is a Wellcome Trust travelling fellow.

REFERENCES


Drosophila goosecoid gene 1649
the organizer and midline cells of frog embryos: involvement in the development of the neural axis. *Development* 116, 81-93.


(Accepted 5 February 1996)