Comparison of the Structure and Expression of *odd-skipped* and Two Related Genes That Encode a New Family of Zinc Finger Proteins in Drosophila

Marilyn C. Hart,[1] Lei Wang and Douglas E. Coulter

E. A. Doisy Department of Biochemistry and Molecular Biology, St. Louis University Medical School, St. Louis, Missouri 63104

Manuscript received February 19, 1996
Accepted for publication May 30, 1996

ABSTRACT

The *odd-skipped* (*odd*) gene, which was identified on the basis of a pair-rule segmentation phenotype in mutant embryos, is initially expressed in the Drosophila embryo in seven pair-rule stripes, but later exhibits a segment polarity-like pattern for which no phenotypic correlate is apparent. We have molecularly characterized two embryonically expressed *odd*-cognate genes, *sob* and *bowel* (*bowl*), that encode proteins with highly conserved C_2H_2 zinc fingers. While the Sob and Bowl proteins each contain five tandem fingers, the Odd protein lacks a fifth (C-terminal) finger and is also less conserved among the four common fingers. Reminiscent of many segmentation gene paralogues, the closely linked *odd* and *sob* genes are expressed during embryogenesis in similar striped patterns; in contrast, the less-tightly linked *bowl* gene is expressed in a distinctly different pattern at the termini of the early embryo. Although our results indicate that *odd* and *sob* are more likely than *bowl* to share overlapping developmental roles, some functional divergence between the Odd and Sob proteins is suggested by the absence of homology outside the zinc fingers, and also by amino acid substitutions in the Odd zinc fingers at positions that appear to be constrained in Sob and Bowl.

GENE duplication is an important evolutionary mechanism for generating families of genes that encode structurally similar proteins with related but divergent functions. While paralogous genes (*i.e.*, homologues within a given genome) are often dispersed throughout the genome, tandem linkage is a salient feature of some gene families. Two classic examples of such linkage are provided by proteins with very disparate functions, the globin and the *Hox* gene families (for reviews, see FORGET 1980; RUDDLE *et al.* 1994). In both cases, the genes are arranged in large arrays or clusters of serially duplicated genes and/or pseudogenes. In addition, multiple unlinked clusters have evolved in the case of the alpha *vs.* beta globin genes and the *Hox* genes of higher animals.

As representative genes encoding transcription factors involved in developmental patterning, the Hox gene clusters appear to be unusually baroque but not unique. Among the zygotic segmentation genes of *Drosophila,* several examples of closely linked paralogues have been identified. These include the gene pairs *engrailed* and *invected* (COLEMAN *et al.* 1987), *knirps* and *knirps-related* (GONZALEZ-GAITAN *et al.* 1994; ROTHE *et al.* 1994), *slp1* and *slp2* at the *sloppy-paired* locus (GROSSNIKLAUS *et al.* 1992), and *gooseberry* and *gooseberry-neuro* (BAUMGARTNER *et al.* 1987; GUTJAHR *et al.* 1993). Among these are representatives of the three major segmentation gene classes (gap, pair-rule, and segment polarity) and several different DNA binding motifs (homeodomain, steroid receptor/C_4 zinc finger, forkhead domain, paired domain). In general, segmentation gene paralogues are expressed in patterns that are overtly similar, despite some differences in the timing of expression (*e.g., engrailed/invected; slp1* and *2*) and/or the specific tissue layers (*gooseberry*) or embryonic domains (*knirps*) involved. Paralogues typically exhibit the greatest conservation within putative DNA binding regions, and diverge appreciably in other coding regions.

The existence of these paralogues raises questions concerning the degree of functional divergence between them. In principle, divergence could result from alterations in coding regions that affect the biochemical properties of the protein, and/or from regulatory alterations that affect the pattern of gene expression. The possibility of redundant or overlapping roles must also be addressed, particularly where the genes appear to be coexpressed. Although in most cases the functional relatedness of segmentation gene paralogues has not been fully assessed, certain conclusions have emerged. In some cases (*e.g., engrailed*), segmentation phenotypes appear to result from point mutations in one paralogue. While this would exclude the possibility of total functional redundancy, it leaves any overlapping or exclusive role(s) for the second gene obscure in the absence of more thorough genetic and molecular analyses. Indeed, a specific role during wing development has only recently been ascribed to the *invected* gene (SIMMONDS

Corresponding author: Douglas E. Coulter, Department of Biochemistry and Molecular Biology, St. Louis University Medical School, 1402 S. Grand Blvd., St. Louis, MO 63104.
E-mail: coultede@wpogate.slu.edu

[1] *Present address:* Department of Cell Biology and Physiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110-1093.

et al. 1995). In other cases, genetic evidence indicates segmentation functions for both paralogues. The analysis of point mutations and small deficiencies in the *sloppy-paired* region indicates that *slp1* may have a unique pair-rule function, whereas a partially-redundant segment-polarity function has been ascribed to both *slp1* and *slp2* (GROSSNIKLAUS *et al.* 1992). Based upon the similar phenotypes associated with ectopic expression of *slp1* and *slp2*, it has been proposed they encode products with equivalent molecular functions, and that the unique pair-rule function of *slp1* merely reflects its differential expression (CADIGAN *et al.* 1994). Functional equivalence has also been suggested for the products of the *knirps* and *knirps-related* genes (GONZALEZ-GAITAN *et al.* 1994) and for *gooseberry* and *gooseberry-neuro* (LI and NOLL 1994; ZHANG *et al.* 1994).

We report here the isolation and initial molecular characterization of two paralogues of the pair-rule segmentation gene *odd-skipped, sob* and *bowel* (*bowl*). While all three genes share highly conserved zinc finger motifs, we find that *sob* has additional similarities to *odd* (*i.e.,* close linkage and a similar expression pattern) that suggest parallels with other segmentation gene homologues. Thus, while *odd-skipped* mutations specifically affect the *odd* transcription unit (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results), *odd* and *sob* may be regulated in common and share some overlapping functions during development. In contrast, the distinct expression of the *bowl* gene, which is less tightly linked to *odd,* together with our analysis of *bowl* mutations (WANG and COULTER 1996), indicates a more divergent developmental role.

MATERIALS AND METHODS

cDNA cloning: *odd*-cognate cDNA clones were isolated by probing a 0–3 hr Drosophila embryonic library (POOLE *et al.* 1985) using a 6.4 kb *odd-skipped* genomic clone as described (COULTER *et al.* 1990). Positive clones were plaque-purified and retested by probing a Southern blot containing the excised cDNA inserts with the same *odd* clone under high stringency. Positive clones derived from cognate genes were distinguished from *bona fide odd-skipped* cDNAs by probing a Southern blot of total Drosophila genomic DNA with each clone and determining whether the observed restriction pattern corresponded to the *odd* genomic map. Cognate clones were subsequently reprobed with a truncated *odd* cDNA clone that includes zinc finger sequences but lacks a region of repetitive "opa" sequence (WHARTON *et al.* 1985) present in the original genomic probe. Of the six initial clones (representing five distinct loci), three (clones B, C and E) hybridized strongly, two (I, K) exhibited no signal and a third (D) showed only a very faint signal (<5% of clone E) with this "opa-less" probe.

5′ end rapid amplification of cDNA ends (RACE): For reverse transcription of RNA, 170 ng of 0–8-hr embryonic mRNA was used as a template with 0.75 μM of a *sob* primer (5′ACTTGGTGAACTGTCGG3′, complementary to cDNA nucleotides 1413–1429), and 5 units of GeneAmp Rtth reverse transcriptase. After excess primer was removed with a Centricon 30 spin column, a poly A tail was added to the first strand of cDNA using terminal transferase. PCR amplification was performed with 20 pmol each of dT adapter primer (5′GACTCGAGTCGACTTTTTTTTTTTT3′) and an internal *sob* primer (5′GCTGCCAGTTCTGCGGCTGTTG3′, complementary to nucleotides 1097–1118) with the following conditions: 30 cycles of 45 sec at 94°, 60 sec 57°, 3 min 72°, followed by a 15 min extension at 72°. The reaction product was divided into two equivalent samples and electrophoresed on opposite sides of a 0.8% agarose gel, half of which was subsequently Southern blotted and probed with a *sob* cDNA to identify an appropriate fragment. DNA was extracted with Quiaex (Quiagen, Inc.) from the same region of the unused gel half, and a portion was subjected to an additional round of amplification using identical conditions, except that a *Bam*HI-tagged *sob* primer (5′GCGGATCCTGTCGCTGCCGTCAAAG3′, complementary to nucleotides 908–929) was used. The final product was digested with *Xho*I and *Bam*HI, cloned into Bluescript (Stratagene), and sequenced.

RNA isolation: Staged wild-type (Canton-S) embryos were collected on apple juice agar medium, harvested, and frozen at −70°. Samples of 1 g were ground on dry ice with a mortar and pestle followed by extraction using the acid guanidinium thiocyanate-phenol-chloroform method (CHOMCZYNSKI and SACCHI 1987). PolyA+ mRNA was isolated by oligo dT chromatography.

Genomic cloning: Drosophila genomic lambda (EMBL4) libraries (a gift of P. SCHEDL) were screened with chemiluminescent probes (ECL kit; Amersham) and positives were plaque purified and restriction mapped using standard procedures (MANIATIS *et al.* 1982). For an initial walking step, probes from the most 5′ (pM13H/T2.3) and 3′ (pM16H.1) regions of the cloned *odd* region (COULTER *et al.* 1990) were used to isolate two clones upstream (U.5, U.7) and three downstream (D.9, D.10, D.14) of *odd-skipped*. Two clones further downstream (B.21, B.24) were isolated in a second step by reprobing the library with a distal fragment of D14. *sob* genomic clones (E.15, E.25) were isolated using pooled subclones from the 5′ (pE-H) and 3′ (pE-B) end of cDNA clone E. Restriction mapping indicated that neither *sob* clone included sequences corresponding to the 5′ end of the cDNA or overlapped with clones from our *odd-skipped* walk.

While our lambda cloning was in progress, we obtained two P1 clones (SMOLLER *et al.* 1991) that were isolated using an *odd* cDNA probe and generously provided by D. SMOLLER. Southern analysis with cDNA probes indicated that *sob* as well as *odd* sequences were present in both P1 clones. In addition, lambda clones isolated via a chromosomal walk from *odd* toward the *Alp* locus were generously provided by S. COHEN. Homology to the *sob* cDNA (clone E) appeared to be fully contained within a 7-kb *Xho*I fragment that was present in both the P1 clones and one of the lambda clones (AW1F) from the *Alp* walk. This fragment was subcloned and further analyzed to define the orientation of *sob* and its intron/exon structure.

DNA sequencing: cDNA and genomic restriction fragments subcloned in Bluescript were isolated by the alkaline lysis method (MANIATIS *et al.* 1982) and sequenced as double-stranded templates using the dideoxynucleotide chain termination method (SANGER *et al.* 1977) with Sequenase 2.0 and other reagents provided in the Sequenase kit (United States Biochemical). Reactions were labeled with ^{35}S-dATP (New England Nuclear) and primed with the T3, T7, Reverse, and −20 primers and with custom oligonucleotides as necessary. Reactions were run on 6% polyacrylamide denaturing gels. DNA and protein sequences were analyzed using the GCG software package (Wisconsin Genetics Computer Group) on a VAX computer.

Nucleotide sequences corresponding to the four zinc fingers common to *odd, sob* and *bowl* (see Figure 1) were analyzed using MEGA (KUMAR *et al.* 1993) to identify synonymous nucleotide substitutions and estimate synonymous substitution rates (Jukes-Cantor method); frequencies of 0.70 ± 0.15, 1.24 ± 0.29, and 1.18 ± 0.27 substitutions per synonymous site were obtained for *odd vs. sob* (33.3 differences), *odd vs. bowl*

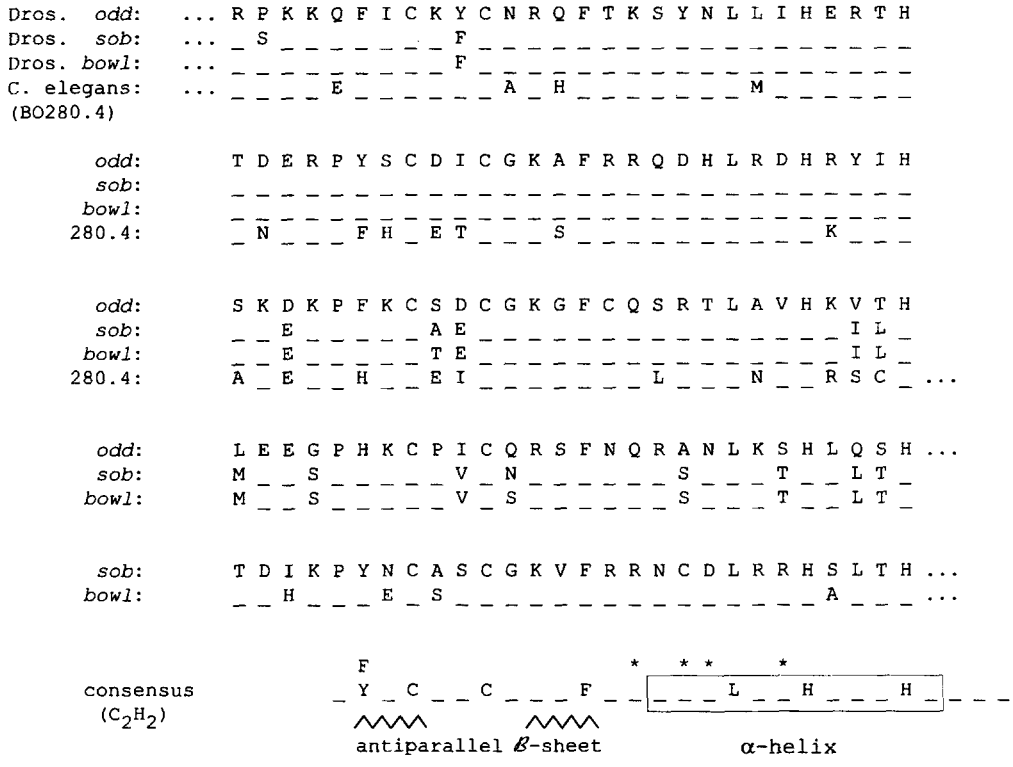

FIGURE 1.—Zinc finger sequence alignments between the Odd, Sob and Bowl proteins. Successive zinc fingers corresponding to amino acid residues 215–326 of the Odd sequence, 389–528 of the Sob sequence, and 233–372 of the Bowl sequence are shown. The *C. elegans* sequence B0280.4 corresponds to an anonymous locus identified in a 41-kb cosmid (accession no. U10438) from a sequenced region of chromosome III (WILSON *et al.* 1994); amino acids 40–123 of an inferred 159 residue protein are indicated. Dashes indicate identity with the Odd sequence, except in the case of the fifth finger where Sob is used as the reference sequence. The 28 residue consensus sequence represents the most common spacing for C_2H_2 zinc fingers (ROSENFELD and MARGALIT 1993). The indicated secondary structure elements (α-helix and antiparallel β-sheet) are inferred from the structure of model zinc fingers; asterisks (*) denote the positions of putative DNA binding residues (PAVLETICH and PABO 1993).

(45 differences), and *sob vs. bowl* (44 differences), respectively. (For sequences encoding all five fingers, *sob* and *bowl* gave an estimated rate of 1.40 ± 0.32.) While the difference between the *odd/bowl* and *sob/bowl* values is not significant, the smaller *odd/sob* value appears marginally significant compared to either *bowl* and *odd* (t = 1.65; $P \leq 0.05$) or *bowl* and *sob* (t = 1.55; P = 0.06). Qualitatively similar results were obtained by applying the Jukes-Cantor, Tamura, or Kimura two-parameter distance methods in MEGA to total nucleotide substitutions at third codon positions.

Southern and Northern analysis: Blots were prepared according to standard procedures (MANIATIS *et al.* 1982). Probes were either labeled for chemiluminescent detection with the ECL nonradioactive detection system (Amersham), or ^{32}P-labeled by nick translation or random primer labeling using commercially available reagents (United States Biochemical). For Southern analysis, hybridizations were performed overnight in ECL hybridization buffer containing 0.5 M NaCl at 42°. Membranes were washed in 0.2× SSC and 0.1% SDS at 65°. Northern blots were washed under the same conditions after hybridization in Church solution (CHURCH and GILBERT 1984) at 65°.

Embryo techniques: Wild-type embryos representing various stages throughout embryogenesis were collected and aged at 25° on apple juice/agar plates. We employed phase-partition fixation (ZALOKAR and ERK 1977) and a whole mount *in situ* hybridization protocol (TAUTZ and PFEIFLE 1989) with modifications suggested by S. DINARDO to detect *odd-skipped, sob* and *bowl* transcripts. A detailed description is available upon request. Double-stranded DNA probes were prepared by random-primer labeling of isolated cDNA fragments with digoxygenin-dUTP according to the protocol of the Genius Nonradioactive Labeling and Detection Kit (Boehringer Mannheim). For detection of *sob* transcripts, different labeled cDNA subclones that flank the zinc fingers were pooled; these included the 5′RACE clone and a 0.3-kb *Eco*RI/*Hin*dIII subclone (pEH; Figure 2) derived from the 5′ end of clone E (both of which lie upstream of the *sob* zinc fingers) and a 0.4-kb *Bam*HI/*Eco*RI subclone (pEB; Figure 2) from the 3′ end of clone E. For *bowl*, a 1.2-kb *Pst*I/*Eco*RI subclone (pB-P; WANG and COULTER 1996) from the 3′ end of clone B that lies downstream of the zinc fingers was used. The *odd* probe was generated from a near full-length cDNA clone (7.4; COULTER *et al.* 1990); although it includes the *odd* zinc finger region, the pattern detected using this probe was indistinguishable from that of truncated (3′) cDNA probes that lack the zinc fingers.

After whole mount *in situ* hybridization, some embryos were also immunochemically labeled with the anti-*engrailed* antibody according to protocols generously provided by JAN MULLEN (personal communication) and ARMEN MANOUKIAN (MANOUKIAN and KRAUSE 1993). The monoclonal mouse anti-*engrailed* hybridoma culture supernatant (PATEL *et al.* 1989) was a generous gift of IAN DUNCAN. Embryos were mounted in GMM (Canada Balsam:methyl salicylate 1:2) and staged according to the guidelines of CAMPOS-ORTEGA and HARTENSTEIN (1985).

The GeneBank accession number for the *sob* sequence is U62004.

RESULTS

Isolation of *odd-skipped* cognate gene cDNAs: cDNA clones derived from *odd*-homologous genes were identi-

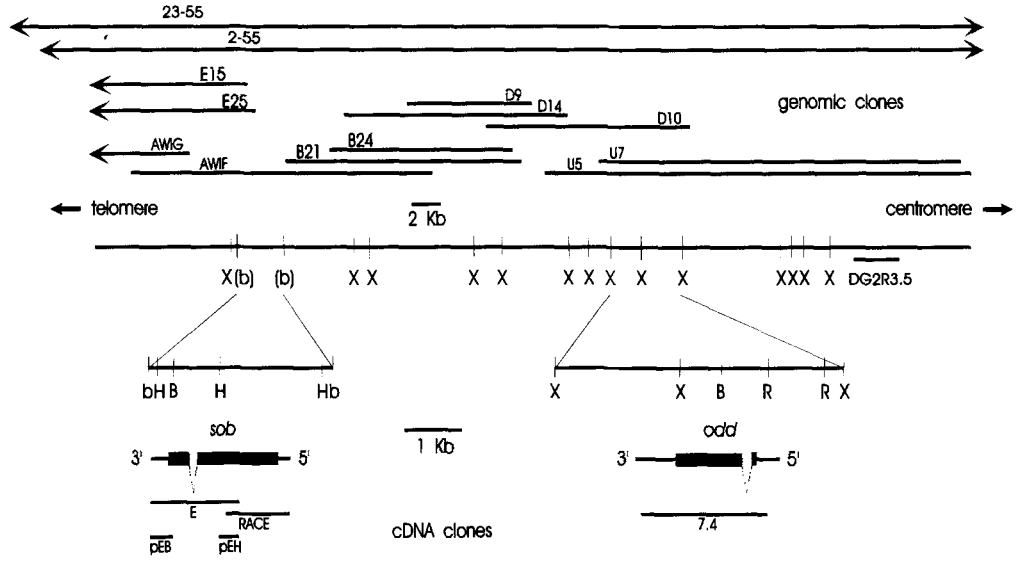

FIGURE 2.—Molecular organization of the *odd/sob* genomic region. Genomic sequences are represented by two P1 phage clones (2-55 and 23-55) that extend beyond the indicated region and by overlapping lambda clones isolated with a *sob* cDNA probe or by walking from *odd-skipped* (see Methods); additional lambda clones in the vicinity of *odd* (COULTER *et al.* 1990) are not shown. AW1F and AW1G were isolated in the laboratory of S. Cohen. DG2R3.5 corresponds to a 3.2-kb *Eco*RI genomic subclone from the distal region of the *foraging*/cGMP dependent kinase locus (KALDERON and RUBIN 1989). The *odd-skipped* and *sob* transcription units are diagramed below the genomic map. Rectangles represent protein coding regions. The *odd* intron and additional features of the *odd* sequence not previously reported (COULTER *et al.* 1990) will be presented elsewhere (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results). cDNA clones or subclones mentioned in the text are indicated below the transcription units. Restriction enzymes: B, *Bam*HI; R, *Eco*RI; H, *Hind*III; X, *Xho*I; b, *Xba*I. Parentheses are used to indicate that not all *Xba*I sites in the *odd/sob* interval are shown.

fied during a high stringency screen of a Drosophila embryonic cDNA library with a genomic fragment that included approximately $^2/_3$ of the transcribed sequences of the *odd-skipped* gene plus additional 3′ sequences (COULTER *et al.* 1990). In addition to several *odd* cDNAs, this screen yielded three clones (B, C and E) that readily hybridized to an *odd* zinc finger cDNA probe (see MATERIALS AND METHODS) but showed distinct restriction patterns when used to probe genomic Southern blots (data not shown). In some cases fainter secondary bands resulting from cross-hybridization with *odd-skipped* genomic fragments were observed. Clones B and C (3.0 and 1.4 kb, respectively) exhibited Southern patterns similar to each other, and sequencing has verified that these are derived from a single locus. We have named this locus *bowel* (*bowl*) (an acronym for brother of odd with entrails limited) based upon the phenotype of mutants (WANG and COULTER 1996); the locus represented by clone E has been named *sob* (sister of odd and bowel). [Note that *sob* and *bowl* were provisionally referred to as odd-related genes (ORG) 1 and 2, respectively, in previous unpublished communications.]

Sequence analysis indicated that the longer *bowel* cDNA clone (B) contains the entire open reading frame (ORF) and encodes a protein of 745 amino acids; a complete description is presented elsewhere (WANG and COULTER 1996). In contrast, the 1.35-kb *sob* cDNA (E) had an ORF which appeared truncated at the 5′ end, and was substantially smaller than the embryonic mRNA of approximately 2.3 kb observed on a Northern blot (data not shown). Because the 3′ end of this clone appeared intact (based on a polyadenylation signal and 18-bp poly-A tail), we used the RACE method (FROHMAN *et al.* 1988) to isolate a 5′ cDNA (see MATERIALS AND METHODS) that yielded an additional 886 bp of upstream sequence. The resulting composite cDNA sequence of 2235 bp (see Figure 3) includes a 1731 nucleotide ORF that appears to be complete (see below).

***odd, sob* and *bowel* encode closely related zinc finger proteins:** The most notable feature of the *odd-skipped* sequence is the presence of four tandem C_2H_2 zinc fingers in the carboxy-terminal third of the protein (COULTER *et al.* 1990). As expected from their hybridization to this region of the *odd* gene, both *sob* and *bowl* encode C_2H_2 zinc fingers, and these show substantial homology with the amino acid (Figure 1) and nucleotide sequences (not shown) of *odd*. Unlike *odd*, the cognates encode five (rather than four) tandem zinc fingers, such that the first four amino-terminal fingers of each align with the *odd* sequence. With the exceptions noted below, the genes show little if any significant homology beyond this region, indicating that cross-hybridization to the zinc fingers of *odd* was responsible for their isolation.

Figure 1 shows amino acid sequence alignments between the zinc fingers of the Drosophila cognates (*odd, sob* and *bowel*) plus the three fingers present in the *Caenorhabditis elegans* gene sequence B0280.4, the strongest homologue identified in a BLAST search (ALTSCHUL *et al.* 1990) of sequence databases. Each

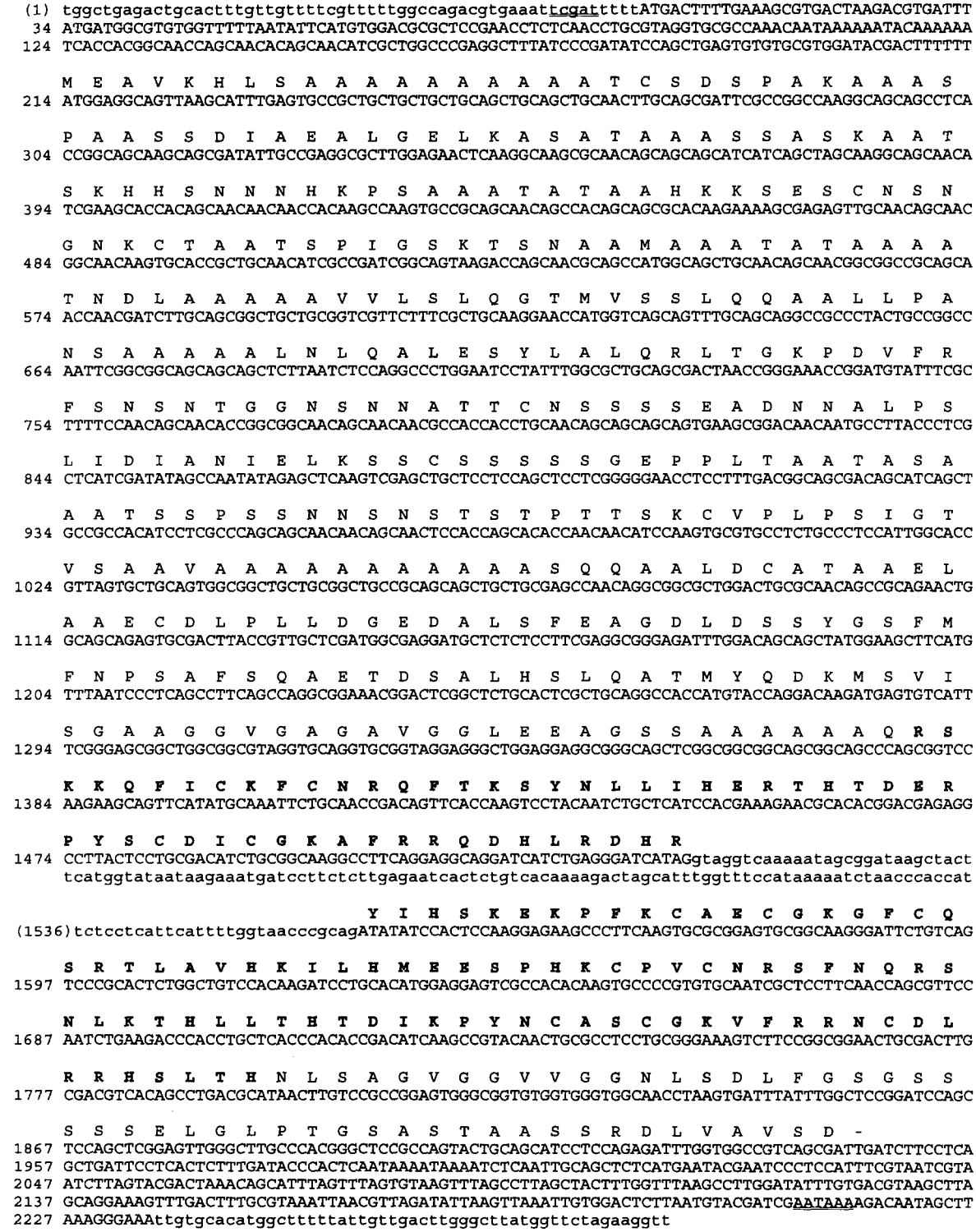

FIGURE 3.—Nucleotide and amino acid sequences of *sob*. The nucleotide sequence of the *sob* transcript is indicated in upper case letters and represents a composite of cDNA clone E and the 5′ RACE clone; at positions where the RACE and genomic sequences do not match, the latter is shown to avoid possible PCR artefacts. Lower case letters are used for the intron and additional genomic sequences upstream and downstream of the composite cDNA sequence. Underlined sequences correspond to a polyadenylation signal (AATAAA) near the 3′ end and a sequence near the putative transcription start site (5′ end of the RACE clone) with the closest match to the arthropod initiator consensus (TCAGT; CHERBAS and CHERBAS 1993). The amino acid sequence is represented using the single letter code above the cDNA sequence; zinc fingers are indicated with boldface type. The cDNA sequence is numbered from the putative transcription start site, and amino acids are numbered from the first in frame ATG. Although the sequence flanking the initiation codon at cDNA nucleotide 214 does not match the Drosophila translation start consensus sequence (TTTTATG *vs.* (C/A)AA(A/C)ATG, respectively; CAVENER 1987), the contexts of the next two in-frame methionines at nucleotide 541 and 625 are also poor. An additional poly A tail of 15 nucleotides that was present at the 3′ end of clone E is not shown.

finger is depicted as a 28-residue repeat that matches the canonical consensus for the C_2H_2 zinc finger class. However, the boundary separating adjacent fingers has been repositioned (to a more N-terminal position relative to the consensus) to maximize homology among the genes, since the sequences begin to diverge immediately outside the indicated boundaries. This suggests that evolutionary selection may operate on a repeat unit that differs somewhat from that previously inferred for the zinc finger.

A comparison of the zinc finger amino acid sequences of the three Drosophila genes indicates that the Sob and Bowl proteins are most similar. Besides having a fifth, C-terminal finger that is not present in the Odd protein, the two cognates show greater similarity among the first four fingers as well. Over this region, Sob and Bowl are 97.3% identical while sharing 86.6% and 87.5% identity with Odd, respectively. Substitutions among the three sequences occur at 15 different positions within these four fingers. At 12 of these, Sob and Bowl share the same residue, and at 2 of the 3 positions where Sob and Bowl do differ, the Odd sequence diverges from both cognates.

The conservation between Sob and Bowl is notably lower in the fifth (C-terminal) finger, where they share 85% identity, such that the zinc fingers of the two cognates are 95% identical overall. Interestingly, a similar trend toward increasing divergence in more C-terminal fingers is apparent when the first four fingers of all three proteins are compared. While substitutions are relatively rare among the first two (N-terminal) zinc fingers, they increase in frequency in the third and fourth fingers. A C-terminal bias is also apparent in the divergence between the *C. elegans* gene and the first three fingers of Odd (see Figure 1).

Beyond the zinc fingers, homology searches indicate no extensive regions of amino acid sequence conservation among the three cognates. A small region of homology occurs at the C-terminal ends of the Odd and Bowl proteins, which terminate with the sequences GFTIDEIMSR and GFSIEDIMRR, respectively. (This region of the Sob sequence is clearly distinct; see Figure 3). The relative position of the zinc fingers varies among the three proteins, falling near the C-terminal end of Odd (residues 215–326 out of 392 total) and Sob (residues 389–528 of 577), but in the amino terminal half of Bowl (residues 233–372 of 744).

***odd, sob* and *bowl* are closely linked:** Hybridization of *sob* and *bowl* cDNA clones to polytene chromosomes *in situ* (data not shown) indicated that both loci reside on the left arm of the second chromosome, with *sob* falling in the same letter division (24A) as *odd,* and *bowel* nearby at 24C. Further analysis (WANG and COULTER 1996) placed *bowl* in the cytogenetic interval between 24C2 and 24C5, with a transcriptional orientation from distal (5′) to proximal (3′) with respect to the centromere.

Although the *odd* and *sob* loci could not be resolved by our *in situ* hybridization analysis, Southern analysis indicated that *sob* lay beyond the initial 25 kb of cloned genomic DNA from the *odd* locus (COULTER *et al.* 1990). To define the relative location and orientation of *odd* and *sob,* we isolated *sob* genomic lambda clones, initiated a chromosome walk from *odd-skipped,* and obtained two P1 clones that spanned the region between *odd* and *sob* (see MATERIALS AND METHODS).

As shown in Figure 2, the *sob* transcription unit resides 24–25 kb distal to *odd,* and both genes are transcribed in the same orientation with the 5′ end positioned towards the centromere and the 3′ end towards the telomere. The assignment of a proximal/distal orientation derives from our localization of a genomic clone from the *foraging*/cGMP dependent kinase region (KALDERON and RUBIN 1989), which lies proximal to *odd,* to a 5′ position. This assignment is also consistent with the mapping of a translocation breakpoint associated with the *Alp* locus (S. COHEN, personal communication), which lies distal to *sob.*

The *sob* locus: The sequence of the *sob* locus is shown in Figure 3; this includes the complete cDNA and predicted amino acid sequences and partial genomic sequence. Comparison of genomic and cDNA sequences indicated a single small intron of 146 bp that interrupts the second zinc finger of *sob.* The compact transcription unit of *sob* is reminiscent of *odd,* which also has a single small intron (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results), and distinct from *bowl,* which has three introns of 0.8, 8.2, and 2.0 kb. However, while the *odd* intron falls well upstream of the zinc fingers, the *sob* intron and the third *bowl* intron both fall at an identical position within the second finger (WANG and COULTER 1996). Except for consensus splice signals, the zinc finger introns of *sob* and *bowl* show no obvious homology, although the complete sequence of the 2 kb *bowl* intron has not been determined.

The putative *sob* ORF encodes a 577 residue protein with a predicted molecular mass of 58 kD. Besides the five tandem zinc fingers near the carboxy-terminus (see above), additional features of the Sob protein deserve note. While the protein is basic overall (with a predicted pI of 9.3), the amino-terminal $^2/_3$ of the protein is notably more acidic, with a predicted pI of 5.5 for residues 1–386. Because the acidic residues are widely dispersed, none appear to comprise an "acid blob" characteristic of some transcriptional activation domains. The protein is also rich in alanine (20%) and serine (15%), which comprise several homopolymeric or copolymeric stretches; it is possible that some of these mediate protein-protein contacts involved in transcriptional activation or repression.

Expression of *odd-skipped, sob* and *bowl* in the early embryo: We used whole mount *in situ* hybridization to wild-type embryos in order to compare the transcript accumulation patterns of *odd-skipped, sob* and *bowel* during early embryogenesis. To ensure that cross-hybridization with *odd* transcripts did not interfere with our analy-

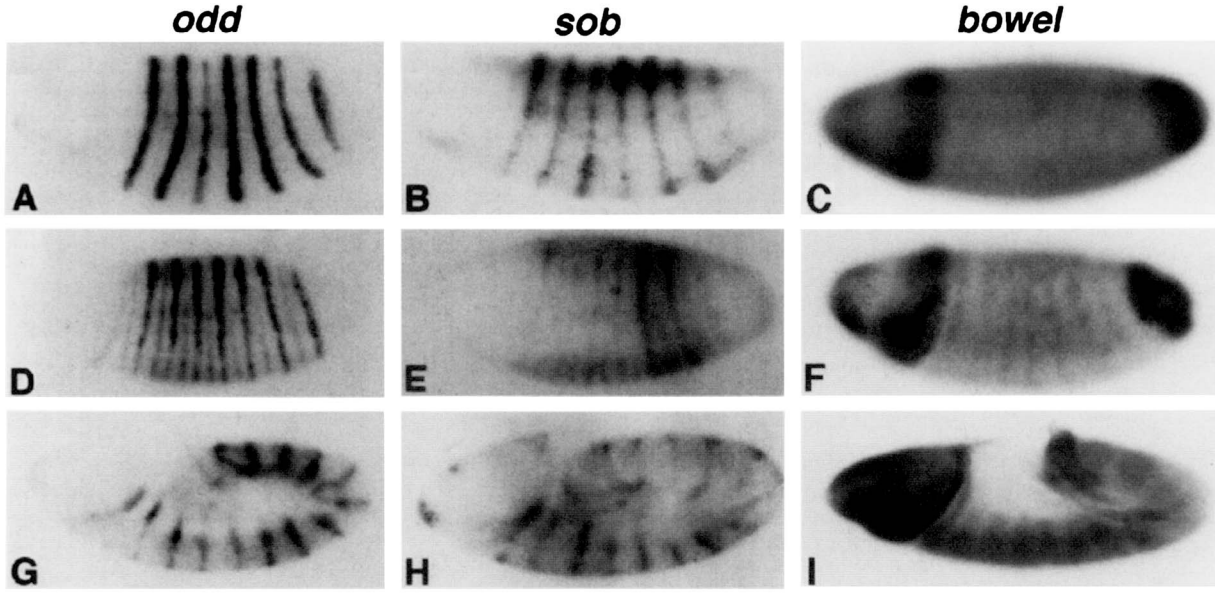

FIGURE 4.—Expression of *odd, sob* and *bowl* transcripts during early embryogenesis. Whole mounts of wild-type embryos hybridized with digoxygenein-labeled probes for *odd* (A, D, G), *sob* (B, E, H), or *bowl* (C, F, I) are shown. Embryos are oriented with the anterior to the left and dorsal side up. At the cellular blastoderm stage (A–C), both *odd* and *sob* are transcribed in seven pair rule stripes, whereas *bowl* is expressed strongly at the posterior pole and in a single broad stripe near the presumptive cephalic furrow, with a weaker domain at the anterior pole. As gastrulation commences (D–F), additional stripes of *odd* and *sob* arise between the seven pair-rule stripes, while the terminal and cephalic domains of *bowl* are supplemented with a fainter pair-rule pattern of stripes in the presumptive trunk. In extended germ band embryos (G–I), all three genes are expressed in stripes with a single-segment periodicity, although the *bowl* stripes are broader than those of *odd* and *sob* and also remain less intense than the terminal *bowl* domains.

sis, the *sob* and *bowel* probes were generated from cDNA subclones that did not include the zinc finger regions of these genes (see MATERIALS AND METHODS). As depicted in Figure 4, this analysis indicates that the expression of *sob* has striking similarities with *odd,* whereas *bowl* is expressed in a distinct pattern.

At the cellular blastoderm stage, transcripts of both *odd* and *sob* accumulate in seven transverse stripes that are approximately three or four cells wide, reflecting a double segment periodicity characteristic of the pair-rule genes (Figure 4A and B). For *odd,* these stripes correspond to the primordia of the anterior of the odd-numbered segments, where *odd* activity is required for segmentation (COULTER and WIESCHAUS 1988; COULTER *et al.* 1990). Both genes are also expressed in a nonperiodic domain at the anterior of the embryo. At the blastoderm stage, this domain extends ventrally and posteriorly from the anterior tip of the embryo (100% EL) to approximately 85% egg length (EL).

As cellularization ends and gastrulation and germ band elongation begin, the expression pattern of both *odd* and *sob* changes from the seven pair-rule stripes to a pattern of segmentally repeated stripes that is reminiscent of a segment polarity gene (Figure 4, D–E and G–H). At least two processes are responsible for this transition. First, the width of each of the seven blastoderm stripes narrows to approximately one or two cells as transcripts are eliminated from the stripe margin. Concomitant with this narrowing of the initial or "primary" stripes, secondary stripes (each one or two cells wide) arise due to *de novo* accumulation of transcripts in the interstripe regions. The secondary stripes arise in a graded fashion from anterior to posterior, and they intensify rapidly so that at about the time invagination of the ventral furrow is complete, both the primary and secondary stripes are nearly equivalent in width and intensity. Despite further dynamic changes (*e.g.,* the stripes fragment in the extended germ band due to loss of *odd* or *sob* expression at specific positions along the dorsal/ventral axis; see Figure 5), both genes continue to show equivalent expression within each segment throughout the remainder of embryogenesis.

The only observed distinction between the early expression patterns of *odd* and *sob* concerns the level of transcript accumulation. Expression of *odd* appears to be strongest at the cellular blastoderm stage when it is expressed in seven stripes. Later, when the segment polarity-like pattern develops, the intensity of the stripes diminishes. The *sob* transcripts show the reverse, appearing relatively weak at blastoderm and gaining in intensity at gastrulation when the genes are expressed with a single-segment periodicity. These differences are not due to experimental variation in staining intensity, since they were apparent with either gene when different embryos from the same experiment were compared in the same microscopic field. While both the close proximity of their transcription units and their similar expression patterns raise the intriguing possibility that *odd* and *sob* share common or redundant *cis*-regulatory elements, the observed differences in expression levels

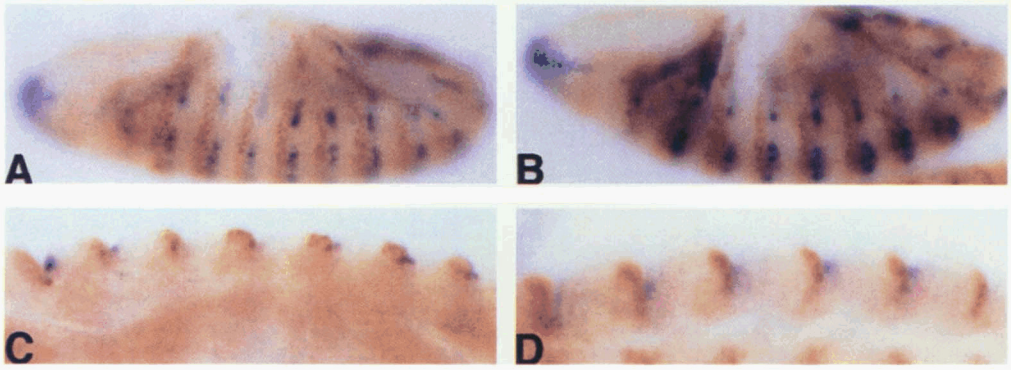

FIGURE 5.—Phasing of the *odd* and *sob* stripes. Wild type embryos were double labeled using digoxygenin labeled probes (blue) for *odd* (A, C) or *sob* (B, D) transcripts and an antibody to the Engrailed protein (brown). During germ band extension (A and B) and retraction (C and D), *odd* and *sob* are expressed immediately posterior to each *engrailed* stripe. Both the *odd* and *sob* stripes become discontinuous during germ band extension (note gaps along the dorsal/ventral axis in A and B), but later reform into coherent stripes that lie at the posterior edges of the segmental grooves (C–D).

probably reflect intrinsic quantitative differences in their response to common *trans*-acting factors. These differences are also consistent with temporally distinct functions for the two genes: The blastoderm pattern would correspond to the pair-rule function of *odd* and the later, segment polarity-like expression would indicate either a unique role for *sob* or a function that is common to both genes.

Transcripts of the *bowel* gene initially accumulate at the cellular blastoderm stage in three domains that are more terminally located than the *odd* and *sob* stripes (Figure 4C); these include a strong cap at the posterior pole (from 0 to 11% EL), a relatively weak and nonuniform domain at the anterior pole (84–100% EL), and a broad transverse stripe (approximately 6 cells wide) that lies just anterior to the presumptive cephalic furrow (extending from 76% EL dorsally to 67% EL ventrally). While some cells in these regions may also express *odd* and *sob* (*e.g.,* the anterior cap of *odd* partially overlaps the anterior *bowl* domain, as do the anteriormost secondary stripes of *odd*/*sob* and the *bowl* cephalic stripe), the overall patterns have little in common. Therefore, the initial regulation of *bowl* is clearly distinct from that of *sob* and *odd.* However, at early gastrulation, the terminal expression of *bowl* is supplemented by transverse stripes in the trunk primordium (Figure 4F). As with *odd* and *sob,* these initially appear in a pair-rule pattern of seven stripes, but rapidly evolve to a segment polarity-like pattern as the number of stripes doubles. The intensity of these stripes is very low compared to the terminal domains and the single cephalic stripe, suggesting that *bowl* may retain only vestigial regulatory elements in common with *odd* and/or *sob.*

Stripes of *odd-skipped* and *sob* overlap: The apparent similarities in the striped expression of *odd* and *sob* indicate that both genes might be co-expressed in a subset of cells in the early embryo. To determine the relative phasing of the stripes, we performed *in situ* hybridization with either the *odd* or *sob* probe and double-labeled the embryos with an anti-*engrailed* monoclonal antibody (PATEL *et al.* 1989). In wild-type embryos, *engrailed* (*en*) is initially expressed in 14 stripes that mark the anterior boundaries of the gnathal and trunk parasegments (DINARDO *et al.* 1985). This pattern is initiated during the late cellular blastoderm/early gastrulation stages, with different stripes appearing in a complex temporal sequence in which even-numbered and more anterior stripes tend to precede stripes that are odd-numbered or more posterior (WEIR and KORNBERG 1985). In double-labeled embryos, we found that the even- and odd-numbered *en* stripes arise immediately anterior to the primary and secondary stripes, respectively, of both *odd* and *sob* (Figure 5, A and B). Despite changes in the pattern of both the *en* and *odd*/*sob* stripes that occur during germ band extension and retraction, the underlying anterior/posterior juxtaposition is maintained through subsequent stages of embryogenesis. These results indicate that many and possibly all of the cells in each *odd* or *sob* stripe express both genes. Also, because *engrailed* expression ultimately defines the posterior of each segment as well as the anterior of each parasegment, these results indicate that both *odd* and *sob* mark the presumptive anterior margin of each segment. This positioning is readily seen in later embryos, where the segmental grooves lie immediately anterior to either the *odd* or *sob* stripes and posterior to *en* (Figure 5, C and D). Because we were unable to detect engrailed protein prior to gastrulation (stage 6), the phasing of the broader primary stripes of *odd* and *sob* during cellularization was not directly established by this experiment. However, given the similar dynamics of the transition from a double- to single-segment periodicity and the apparent identity of the stripes in the latter mode, it is likely that the primary stripes of *odd* and *sob* are also coextensive at the blastoderm stage.

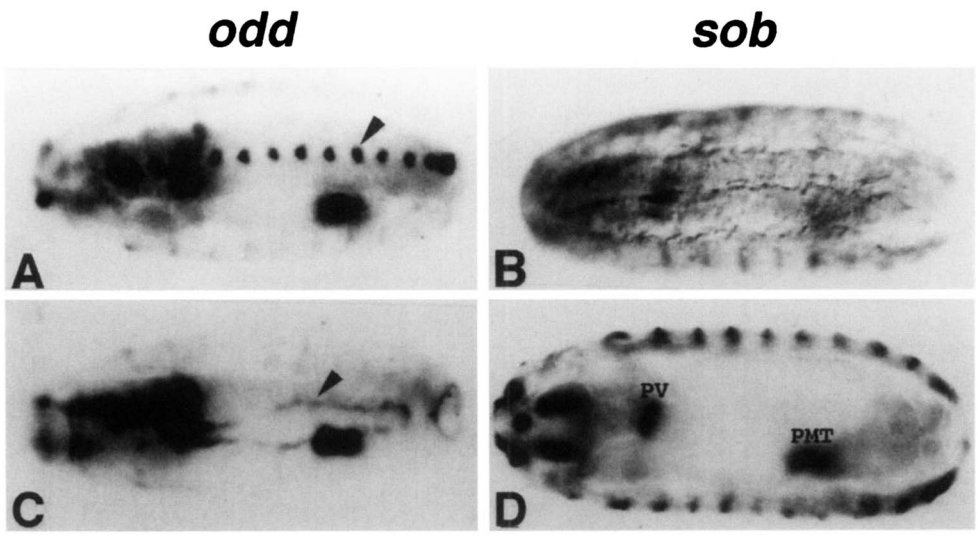

FIGURE 6.—*odd* and *sob* expression in the late embryo. Ventral (A and B) and dorsal (C and D) views of *odd* (A, C) and *sob* (B, D) transcript patterns in embryos following dorsal closure. Both probes show expression in various internal tissues including the proventriculus (PV), the proximal malpighian tubule (PMT), cells within or near the brain, the pharyngeal muscles, and the salivary duct. Expression in the ventral nerve cord (A, *arrowhead*), pericardial cells of the dorsal vessel (C, *arrowhead*), and garland cells (not shown) was detected with *odd* but not *sob*. Embryos were photographed using brightfield illumination, except for B, where Nomarski optics were used to allow visualization of the unstained ventral nerve cord.

Differences between *sob* and *odd* expression in late embryos: The similarities between the *odd* and *sob* transcript patterns during early embryogenesis suggested that the two genes might continue to be coexpressed as both patterns continue to evolve during subsequent embryonic stages. In addition to persistent expression in striped patterns in epidermal cells, we found that *odd* and *sob* are transcribed in a variety of internal organs and tissues, and in many cases, these later patterns appear to be equivalent. For example, both genes are expressed in the proventriculus, the proximal Malpighian tubules, cells within the brain, and the pharyngeal muscles. However, in contrast to *odd,* we failed to detect *sob* transcripts in the ventral nerve cord, the pericardial cells of the dorsal vessel and the garland cells (Figure 6). These differences suggest that *odd* and *sob* may maintain at least some independent *cis*-regulatory elements. Also, given the structural and functional similarities between the pericardial cells and garland cells (RIZKI 1978), it is possible that *odd* has at least one cell-type specific role that is not shared by *sob.*

DISCUSSION

The odd-skipped zinc finger family: Since the initial discovery of the C_2H_2 zinc finger as a repetitive structural motif in the Xenopus transcription factor IIIA (MILLER *et al.* 1985), hundreds of examples have been reported in genes encoding known or putative DNA binding proteins. The zinc fingers we have identified in Odd, Sob, and Bowl readily match the consensus sequence defined for this family (F/Y-X-C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_{3-5}-H-X_5; BERG 1990). They also share additional properties inferred from a statistical analysis of a subset of zinc fingers known to bind DNA (ROSENFELD and MARGALIT 1993); these include basic residues at position f2 (the second residue following the second Cys) and a high proportion of Thr and Ile at the third position in the "h-loop" (the sequence between the two histidines). Therefore, unlike some putative zinc finger proteins for which a DNA binding function appears less likely (ROSENFELD and MARGALIT 1993), Odd, Sob and Bowl appear to contain bona fide C_2H_2 zinc fingers capable of binding DNA. However, the similarity between the three proteins far exceeds their match to the consensus, such that they clearly represent a highly-conserved "subfamily" within the C_2H_2 zinc finger class.

Sequence comparisons between Odd, Sob and Bowl reveal at least three interesting trends. First, the zinc fingers of Sob and Bowl are most similar. Both proteins contain five tandem fingers, whereas Odd lacks a C-terminal finger and has only four. Also, over the four common fingers, the homology between Sob and Bowl is substantially greater than that between Odd and either cognate (3 *vs.* 15 or 14 amino acid substitutions, respectively). Because the greater conservation between Sob and Bowl does not appear to reflect a relatively recent common ancestry (see below), it is most likely due to selective constraints that are not shared by Odd. Thus, whereas the Sob and Bowl fingers are likely to have retained very similar or identical functions, some degree of functional divergence for the Odd fingers is indicated.

Second, amino acid substitutions appear to be nonrandom with respect to position within the finger. For example, none of the zinc chelating (Cys or His) amino acids or other consensus residues are affected. More than half ($^{11}/_{21}$) of the substitutions identified among all 24 fingers affect the c- and h-loops (*i.e.,* sequences

between the consensus cysteines and histidines, respectively) even though these represent less than one-fourth ($^5/_{21}$) of the nonconsensus positions in the finger sequence. Many of the differences involve conservative amino acid substitutions, and only two (in the fourth finger of the Odd protein) involve any of the four positions in the C_2H_2 zinc finger that have been most strongly implicated as probable determinants of DNA binding specificity (NARDELLI *et al.* 1991; PAVLETICH and PABO 1991, 1993). Therefore, in most cases the changes within a given finger appear unlikely to cause dramatic alterations in its structure or binding properties.

Finally, substitutions appear to be biased toward more C-terminal fingers (see RESULTS), suggesting that C-terminal regions may be less constrained evolutionarily than the N-terminal fingers. Although this trend could reflect true functional diversification of C-terminal regions, the absence of a fifth finger from the Odd sequence suggests that some C-terminal fingers become dispensable as paralogues diverge, and it is possible that the sequence alterations in a subset of the extant fingers reflect the early stages of a similar degradation. In any event, the cross-species comparison with the *C. elegans* homologue, B0280.4, indicates that the processes responsible for preferential diversification and loss of C-terminal fingers may be a general feature of genes in this family.

Our identification of three paralogues in Drosophila raises the question of the extent of the *odd-skipped* family. Although our cDNA screens were not exhaustive, a preliminary analysis of genomic Southern blots using zinc finger specific probes for each of the three cognates (data not shown) indicated that few if any additional immediate members with an equivalent degree of conservation exist in Drosophila: whereas each probe detected restriction fragments corresponding to the other two genes under high stringency conditions, no consistent pattern that would represent an additional locus was observed. However, reduced stringency revealed a few additional bands that might represent more distantly-related genes. Indeed, COHEN *et al.* (1991) have identified a gene with two zinc fingers in the vicinity of the *Alp* locus, approximately 35 kb distal to *sob,* that appears to represent an additional *odd*-paralogue. While this gene is transcribed in a pattern that partially overlaps that of *odd* and *sob,* its zinc fingers show appreciable divergence from those of Odd, Sob and Bowl (S. COHEN, personal communication).

Evolution of *odd, sob* and *bowl*: The structural homologies and linkage between *odd, sob* and *bowl* indicate that all three genes derive from a common ancestor. In the simplest scenario, unequal crossovers would generate tandemly duplicated transcription units that subsequently diversified through changes in regulatory and/or structural sequences. While the duplicated copies would initially be oriented in parallel, a small chromosomal inversion could account for the divergent orientation and less tight linkage of *bowl* relative to *odd* and *sob.* By separating and/or introducing regulatory sequences from the transcription unit, such an inversion could also contribute to the diversification of *bowl,* which appears to have only a remnant of the stripe-specific regulation in the early embryo that is shared by *odd* and *sob.*

At least two independent duplication events must have occurred to generate the three paralogues described here, which raises the question of which gene diverged first. Based on their highly conserved zinc finger amino acid sequences and the small intron at a common position in the two sequences, the extant *sob* and *bowl* genes appear more similar than *odd.* However, the inference that *sob* and *bowl* share a much more recent common ancestor than *odd* is not supported by the pattern of silent (synonymous) nucleotide changes within the zinc finger coding sequences; because these are more likely to be selectively neutral than amino acid changes, they provide a less biased measure of evolutionary distance between paralogues whose biochemical functions may have diverged. Nucleotide comparisons (see MATERIALS AND METHODS) indicate that all three genes have diverged substantially at synonymous nucleotide positions within the zinc fingers. Interestingly, while substitutions between *bowl* and either *sob* or *odd* have approached saturation, somewhat fewer silent changes appear to have accumulated between the *odd* and *sob* zinc finger sequences. Although this might represent residual conservation between *odd* and *sob* at selectively neutral sites (and an earlier divergence for *bowl*), the apparently complete divergence of the two proteins beyond the zinc fingers raises the possibility that differential codon bias or some other selective process is responsible for this similarity. In any event, while a definitive resolution of the evolutionary lineage of these genes will require analysis of other species, we conclude that the extreme conservation between the Sob and Bowl zinc fingers most likely reflects a higher degree of purifying selection than that experienced by the Odd protein, rather than a recent common ancestry.

Functions of *odd, sob* and *bowl*: Although we have yet to determine their target specificities, the amino acid sequence conservation between the zinc fingers of Odd, Sob and Bowl suggests that they are likely to recognize similar nucleotide sequences. Current structural and functional studies indicate a modular function for zinc fingers, with successive tandem fingers contacting adjacent regions ("subsites") of the DNA target site. Thus, whereas the absence of a fifth finger suggests that Odd might have a less extensive binding site than Sob or Bowl, the subsites targeted by its conserved N-terminal fingers might correspond to a portion of the Sob or Bowl recognition sequences. Indeed, it is possible that the substitutions in the third and fourth Odd fingers partially compensate for the loss of the fifth finger by increasing the binding affinity at subsites common to all three proteins. In this case, high affinity binding sites for Sob or Bowl (*i.e.,* those contacted by all five

fingers) might represent a subset of the Odd sites. Alternatively, amino acid changes may have altered the subsite specificities of particular fingers, such that the preferred binding sites of the three proteins would differ despite overall similarities in DNA sequence.

Of course, functional divergence between Odd, Sob and Bowl could have occurred independently of conservation or divergence in their DNA binding domains. The absence of homology beyond the zinc fingers might indicate that the three proteins interact with different components of the transcriptional machinery, thereby affecting the target specificity and/or transcriptional effects of each protein. In the case of *odd-skipped*, a role in transcriptional repression is implied by the ectopic expression of several different segmentation genes in *odd* mutants (*e.g.* DINARDO and O'FARRELL 1987; BAUMGARTNER and NOLL 1990; MULLEN and DINARDO 1995), although it is not yet certain which of these represent direct targets of the Odd protein. It is possible that Sob or Bowl could function as either activators or repressors of the same or different genes. A preliminary genetic analysis of *bowl* did not identify potential regulatory targets (WANG and COULTER 1996), and no *sob* mutants are yet available.

Differences in the expression of the three genes represent another source of functional divergence, particularly in the case of *bowl*. Whereas the nearly equivalent patterns of *odd* and *sob* in the early embryo imply roles in segmentation, the expression of *bowl* at or near the poles of the embryo is consistent with a function in the development of terminally derived structures. Indeed, our analysis of the *bowl* mutant phenotype (WANG and COULTER 1996) indicates such a role, and comparison with the *odd* phenotype indicates that the functions of *odd* and *bowl* during development are distinct. Thus, while some genes might be targeted by both proteins, the embryonic regions involved and the ultimate effects of any common regulation must not be the same for *odd* and *bowl*.

The possibility of overlapping or redundant functions is substantially greater in the case of *odd* and *sob*, given the obvious similarities in their expression patterns. The pair-rule phenotype associated with *odd-skipped* mutants, which results from lesions that are limited to the *odd* transcription unit (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results), indicates that at least some functions of the *odd/sob* region must be unique to *odd*. While a similar conclusion for *sob* awaits the isolation of *sob*-specific mutations, a synthetic deficiency of the *odd/sob* region (NÜSSLEIN-VOLHARD *et al.* 1985; D. COULTER, unpublished observation) exhibits a pair-rule phenotype that is overtly similar to *odd*. However, given the poor cuticle differentiation of these embryos, any subtle defects affecting every segment that might reflect redundant or *sob* specific functions might have gone undetected.

Several possible regulatory mechanisms are suggested by the similar patterns of *odd* and *sob*. Equivalent patterns could be explained by assuming a hierarchical scheme in which one gene is activated solely by the other. However, *sob* is transcribed at apparently normal levels in *odd* mutant embryos (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results). Also, activation of *odd* by *sob* is not consistent with either the relatively weak blastoderm stripes of *sob* or the expression of *odd* transcripts in some regions of the late embryo that appear not to express *sob* (see RESULTS). Therefore, parallel regulation by a common set of *trans*-acting factors appears more likely. These in turn could work through *cis*-elements in the *odd/sob* vicinity that are either redundant, shared between the two genes, or gene-specific. While we have not surveyed *sob* expression in all of the segmentation mutants that affect the *odd* pattern, the spatial patterns of both genes do respond similarly to point mutations in *odd* (M. C. HART, S. CALLACI, E. J. WARD and D. E. COULTER, unpublished results).

Ultimately, any regulatory model must account for the differences as well as the similarities between *odd* and *sob* expression. These include both differences in late embryos, where certain tissues appear to express *odd* exclusively, and quantitative differences in the early (pair-rule) *vs.* later (segment polarity-like) stripes of the two genes (see RESULTS). Based upon its distinct tissue specificity, we assume that at least some gene-specific *cis*- and *trans*-regulatory elements exist for *odd*. On the other hand, while the quantitative differences in the early embryo do not necessarily imply gene-specific *cis*-regulatory sequences, they would indicate a differential response to any shared or redundant elements. Because *odd* and *sob* transcripts are similar in size and do not include large introns, it is unlikely that differences in the time required to synthesize and process complete transcripts are responsible for the greater relative intensity of the *odd* stripes at the cellular blastoderm stage. Such size differences do appear to be responsible for the differential blastoderm expression of the segmentation gene paralogues *knirps* and *knirps-related*, whose primary transcripts are 3 and 21 kb, respectively (ROTHE *et al.* 1992).

We thank STEVE COHEN, DAVE SMOLLER, and MARLA SOKOLOWSKI and ELIZABETH BURGESS for sharing unpublished results and clones, IAN and DIANE DUNCAN for the *engrailed* antibody, PAUL SCHEDL for genomic libraries, MI ZHOU for suggestions concerning the RACE procedure, JAN MULLEN and ARMEN MANOUKIAN for the double-label protocols, and ELLEN WARD for assistance in analyzing expression in late embryos. We also thank CLAUDETTE KLEIN and PAM HOPPE for their careful reading and suggestions on the manuscript. Initial cDNA analyses were carried out in the laboratory of PAUL SCHEDL with the assistance of DAVID GOLDBERG, DOUG KERR and DAN VITTUM. This work was supported by grants GM-42621 from the National Institutes of Health, NP-763 from the American Cancer Society and by a Biomedical Research Support Grant from St. Louis University.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

BAUMGARTNER, S., D. BOPP, M. BURRI and M. NOLL, 1987 Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. Genes Dev. **1:** 1247–1267.

BAUMGARTNER, S., and M. NOLL, 1990 Network of interactions among pair-rule genes regulating *paired* expression during primordial segmentation of *Drosophila*. Mech. Dev. **33:** 1–18.

BERG, J. M., 1990 Zinc finger domains: hypotheses and current knowledge. Annu. Rev. Biophys. Biophys. Chem. **19:** 405–421.

CADIGAN, K. M., U. GROSSNIKLAUS and W. J. GEHRING, 1994 Functional redundancy: the respective roles of the two *sloppy paired* genes in *Drosophila* segmentation. Proc. Natl. Acad. Sci. USA **91:** 6324–6328.

CAMPOS-ORTEGA, J. A., and V. HARTENSTEIN, 1985 *The Embryonic Development of Drosophila melanogaster*. Springer-Verlag, Berlin.

CAVENER, D. R., 1987 Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Res. **15:** 1353–1361.

CHERBAS, L., and P. CHERBAS, 1993 The arthropod initiator: the capsite consensus plays an important role in transcription. Insect Biochem. Mol. Biol. **23:** 81–90.

CHOMCZYNSKI, P., and N. SACCHI, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

CHURCH, G. M., and W. GILBERT, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

COHEN, B., E. A. WIMMER and S. M. COHEN, 1991 Early development of leg and wing primordia in the *Drosophila* embryo. Mech. Dev. **33:** 229–240.

COLEMAN, K. G., S. J. POOLE, M. P. WEIR, W. C. SOELLER and T. KORNBERG, 1987 The *invected* gene of *Drosophila:* sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. Genes Dev. **1:** 19–28.

COULTER, D. E., and E. WIESCHAUS, 1988 Gene activities and segmental patterning in *Drosophila:* analysis of *odd-skipped* and pair-rule double mutants. Genes Dev. **2:** 1812–1823.

COULTER, D. E., E. A. SWAYKUS, M. A. BERAN-KOEHN, D. GOLDBERG, E. WIESCHAUS *et al.*, 1990 Molecular analysis of *odd-skipped*, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. EMBO J. **9:** 3795–3804.

DINARDO, S., and P. H. O'FARRELL, 1987 Establishment and refinement of segmental pattern in the *Drosophila* embryo: spatial control of *engrailed* expression by pair-rule genes. Genes Dev. **1:** 1212–1225.

DINARDO, S., J. M. KUNER, J. THEIS and P. H. O'FARRELL, 1985 Development of embryonic pattern in D. melanogaster as revealed by accumulation of the nuclear *engrailed* protein. Cell **43:** 59–69.

FORGET, B. G., 1980 Structure and organization of the human globin genes. Texas Rept. Biol. Med. **40:** 77–86.

FROHMAN, M. A., M. K. DUSH and G. R. MARTIN, 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

GONZALEZ-GAITAN, M., M. ROTHE, E. A. WIMMER, H. TAUBERT and H. JÄCKLE, 1994 Redundant functions of the genes *knirps* and *knirps-related* for the establishment of anterior *Drosophila* head structures. Proc. Natl. Acad. Sci. USA **91:** 8567–8571.

GROSSNIKLAUS, U., R. KURTH PEARSON and W. J. GEHRING, 1992 The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev. **6:** 1030–1051.

GUTJAHR, T., N. H. PATEL, X. LI, C. S. GOODMAN and M. NOLL, 1993 Analysis of the *gooseberry* locus in *Drosophila* embryos: *gooseberry* determines the cuticular pattern and activates *gooseberry neuro*. Development **118:** 21–31.

KALDERON, D., and G. M. RUBIN, 1989 cGMP-dependent protein kinase genes in *Drosophila*. J. Biol. Chem. **264:** 10738–10748.

KUMAR, S., K. TAMURA and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01*. The Pennsylvania State University, University Park, PA.

LI, X., and M. NOLL, 1994 Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. Nature **367:** 83–87.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MANOUKIAN, A. S., and H. M. KRAUSE, 1993 Control of segmental asymmetry in *Drosophila* embryos. Development **118:** 785–796.

MILLER, J., A. D. MCLACHLAN and A. KLUG, 1985 Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4:** 1609–1614.

MULLEN, J. R., and S. DINARDO, 1995 Establishing parasegments in *Drosophila* embryos: roles of the *odd-skipped* and *naked* genes. Dev. Biol. **169:** 295–308.

NARDELLI, J., T. J. GIBSON, C. VESQUE and P. CHARNAY, 1991 Base sequence discrimination by zinc-finger DNA-binding domains. Nature **349:** 175–178.

NÜSSLEIN-VOLHARD, C., H. KLUDING and G. JURGENS, 1985 Genes affecting the segmental subdivision of the *Drosophila* embryo. Cold Spring Harbor Symp. Quant. Biol. **50:** 145–154.

PATEL, N. H., E. MARTIN-BLANCO, K. G. COLEMAN, S. J. POOLE, M. C. ELLIS *et al.*, 1989 Expression of *engrailed* proteins in arthropods, annelids, and chordates. Cell **58:** 955–968.

PAVLETICH, N. P., and C. O. PABO, 1991 Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science **252:** 809–817.

PAVLETICH, N. P., and C. O. PABO, 1993 Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science **261:** 1701–1707.

POOLE, S. J., L. M. KAUVAR, B. DREES and T. KORNBERG, 1985 The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell **40:** 37–43.

RIZKI, T. M., 1978 The circulatory system and associated cells and tissues, pp. 397–452 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER and T. R. F. WRIGHT. Academic Press, New York.

ROSENFELD, R., and H. MARGALIT, 1993 Zinc fingers: conserved properties that can distinguish between spurious and actual DNA-binding motifs. J. Biomol. Struct. Dynam. **11:** 557–570.

ROTHE, M., M. PEHL, H. TAUBERT and H. JÄCKLE, 1992 Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo. Nature **359:** 156–159.

ROTHE, M., E. A. WIMMER, M. J. PANKRATZ, M. GONZALEZ-GAITAN and H. JÄCKLE, 1994 Identical transacting factor requirement for *knirps* and *knirps-related* gene expression in the anterior but not in the posterior region of the *Drosophila* embryo. Mech. Dev. **46:** 169–181.

RUDDLE, F. H., J. L. BARTELS, K. L. BENTLEY, C. KAPPEN, M. T. MURTHA *et al.*, 1994 Evolution of Hox genes. Annu. Rev. Genet. **28:** 423–442.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SIMMONDS, A. J., W. J. BROOK, S. M. COHEN and J. B. BELL, 1995 Distinguishable functions for *engrailed* and *invected* in anterior-posterior patterning in the *Drosophila* wing. Nature **376:** 424–427.

SMOLLER, D. A., D. PETROV and D. L. HARTL, 1991 Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75–100 kilobase pairs. Chromosoma **100:** 487–494.

TAUTZ, D., and C. PFEIFLE, 1989 A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma **98:** 81–85.

WANG, L., and D. E. COULTER, 1996 *bowel*, an *odd-skipped* homologue, functions in the terminal pathway during *Drosophila* embryogenesis. EMBO J. **15:** 3182–3196.

WEIR, M. P., and T. KORNBERG, 1985 Patterns of *engrailed* and *fushi tarazu* transcripts reveal novel intermediate stages in *Drosophila* segmentation. Nature **318:** 433–439.

WHARTON, K. A., B. YEDVOBNICK, V. G. FINNERTY and S. ARTAVANIS-TSAKONAS, 1985 *opa:* a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in D. melanogaster. Cell **40:** 55–62.

WILSON, R., R. AINSCOUGH, K. ANDERSON, BAYNES, C., BERKS, M. *et al.*, 1994 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature **368:** 32–38.

ZHANG, Y., A. UNGAR, C. FRESQUEZ and R. HOLMGREN, 1994 Ectopic expression of either the *Drosophila gooseberry-distal* or *proximal* gene causes alterations of cell fate in the epidermis and central nervous system. Development **120:** 1151–1161.

Communicating editor: T. SCHÜPBACH