Molecular analysis of the *armadillo* locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a *Drosophila* segment polarity gene

Bob Riggleman, Eric Wieschaus, and Paul Schedl

Department of Biology, Princeton University, Princeton, New Jersey 08544 USA

During *Drosophila* embryogenesis, the segment polarity genes are required for the formation of specific pattern domains within each segment. Mutations in the *armadillo* (*arm*) gene primarily affect the posterior part of the segment and lead to the production of anterior structures within this region. To examine the molecular basis for these effects, we have cloned the *arm* region and identified the gene by germ-line transformation. The *arm* gene produces two types of very abundant 3.2-kb transcripts that differ only in their first exons. These RNAs appear to be formed by independent transcriptional initiation but have similar patterns of expression throughout development. Both *arm* transcripts are present in virtually all of the cell types contained in embryos, third-instar larvae, and adult ovaries, suggesting that *arm* may be required in all cells. In addition, the *arm* transcripts are uniformly distributed in embryonic segments, so the regional pattern defects associated with its embryonic phenotype may result from interactions between *arm* and other localized factors. Both *arm* RNAs encode the same 91-kD polypeptide. This protein has no probable secretory or membrane-spanning regions and contains a series of novel internal repeats that are conserved in sequence, length, and spacing. Considering these results and previous genetic observations, we discuss potential roles for the *arm* gene in pattern formation processes.

During early embryogenesis, the *Drosophila* embryo is divided into a series of homologous segments along the anterior–posterior axis. Systematic mutagenesis experiments have identified a large number of zygotic genes that are required for the segmentation process (Nüsslein-Volhard and Wieschaus 1980). These genes may be divided into three general groups: those that affect large continuous regions (gap genes), those that affect double-segment intervals (pair-rule genes), and those that affect every segment (segment polarity genes). Further molecular and genetic analyses have suggested that these classes of genes act sequentially during embryogenesis to define progressively smaller units of development along the anterior–posterior axis (for review, see Akam 1987; Scott and Carroll 1987).

Genes in the segment polarity class are required for proper development of specific regions along the anterior–posterior axis of each segment. On the ventral surface of wild-type embryos, these positional cell types can be identified by the cuticle structures they produce during embryogenesis. The cells in the anterior region of each segment produce denticles, and those in the posterior region form smooth naked cuticle (Fig. 1a). Within

the segment polarity class, there are seven genes that produce anterior-type denticles in the posterior naked cuticle region when mutated, often forming a reversed polarity duplication of the anterior denticle belt. These genes are *armadillo* (*arm*) (Wieschaus et al. 1984), *cubitus interruptus*[D], *fused*, *gooseberry*, *hedgehog*, *wingless* (Nüsslein-Volhard and Wieschaus 1980), and *dishevelled* (Perrimon and Mahowald 1987). Mutations in the *arm* gene produce pattern alterations that are typical of this group (Fig. 1b), and *arm* has been used in several studies to determine how the defects associated with this phenotype arise. Genetic mosaic analyses with *arm* (Gergen and Wieschaus 1986; Wieschaus and Riggleman 1987) have shown that small patches of *arm*⁻ cells induced in the posterior part of the segment still form anterior pattern elements, even when completely surrounded by wild-type cells. This result argues that the anterior pattern duplications observed in *arm* embryos are not caused by anterior cells colonizing the posterior region following posterior cell death and are not therefore analogous to the regenerative effects produced by surgical excision of the posterior region (Wright and Lawrence 1981). Instead, the duplications appear to be
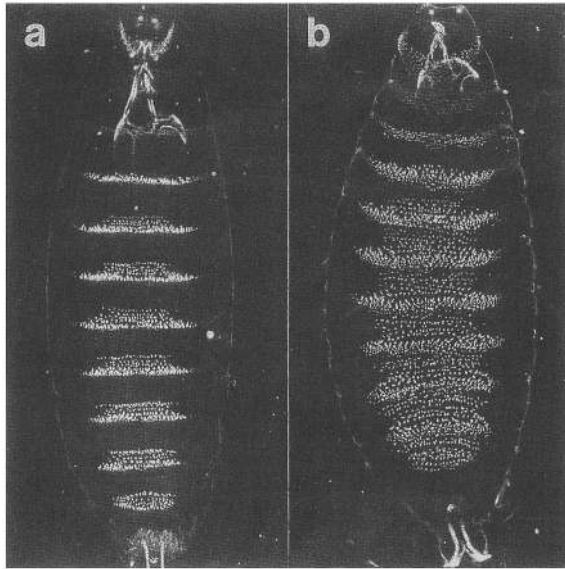
Figure 1. *arm* embryos exhibit a segment polarity phenotype. Dark-field photomicrographs of ventral cuticle preparations of a wild-type embryo (*a*) and an *arm*[25B] embryo (*b*). Loss of *arm* function produces anterior-type denticles in the posterior naked cuticle region of each segment, often forming an apparent mirror image of the anterior denticle belt.

the result of individual posterior cells becoming transformed to an anterior cell fate in the absence of *arm* function.

Given that mutations in any of the seven genes listed above produce the same basic pattern alterations, it is possible that they may function together as a system in establishing or interpeting positional information in the developing segments. In principle, such a system would require at least some gene products to be localized to specific regions within each segment. Recently, two of these genes, *wingless* (Baker 1987) and *gooseberry* (Baumgartner et al. 1987; Côté et al. 1987) have been cloned, and their spatial patterns of transcriptional expression have been analyzed. In both cases, their transcripts were detected only in the approximate region that is affected by mutations in either gene. In this respect, both of these genes are candidates for localized determinants that directly specify positional fates in the segment.

However, it does not seem necessary for all components of a positional system to be spatially localized. For example, factors that are required for producing, transporting, perceiving, or responding to positional information might be evenly distributed throughout the segment. Such products might also be required at other times during development and may not be used exclusively for pattern formation processes. Still, if these factors are required for the patterning system to function, their removal could produce specific defects that resemble the mutant phenotype associated with loss of the localized determinant itself.

In this paper, we describe the molecular cloning of the *arm* gene, its structural organization, and the spatial dis-

tribution of its transcripts throughout development. We show that in contrast to the localized transcript accumulations observed with other segment polarity genes, *arm* RNAs are uniformly distributed in all cells of the embryo. We also present the sequence of the *arm* protein, which contains a series of novel repeats. Considering these results and previous genetic observations, possible roles for the *arm* gene in pattern formation are discussed.

Results

Molecular cloning and characterization of the arm region

We have used the transposon tagging method described by Bingham et al. (1981) to clone DNA sequences from the *arm* region. A hybrid dysgenesis mutagenesis screen for embryonic lethal mutations on the X chromosome (Zusman et al. 1985) produced three dysgenic alleles of arm: 25B, 18.3, and TD5. In situ hybridizations to polytene chromosomes with a cloned P element probe showed that all three alleles contained P elements at polytene band 2B15, within the cytological region known to contain the *arm* gene (2B10–2B17; E. Wieschaus, unpubl.). A λ library was constructed using genomic DNA from one of the dysgenic alleles, *arm*[18.3], and screened with a P element probe. P-element-positive clones were isolated and hybridized to wild-type polytene chromosomes, and four of these contained sequences from the 2B15 *arm* region. Single-copy sequences from these clones were used to recover a 35-kb section of DNA from a wild-type genomic library (Fig. 2a).

To determine the location of the *arm* gene within this region, we examined four EMS *arm* alleles (XM19, XP33, XK22, and YD35; Wieschaus et al. 1984) and the three dysgenic *arm* alleles for DNA rearrangements that might have disrupted the gene. The wild-type DNA segment was subcloned as short (2–4 kb) restriction fragments into a plasmid vector and hybridized with Southern blots of genomic DNA from mutant *arm* alleles. The results of these experiments demonstrated that all three dysgenic alleles contained rearrangements, whereas none of the EMS alleles showed any detectable insertions or deletions. As shown in Figure 2a, *arm*[18.3] contains a 1.3-kb P element in the center of fragment E, and *arm*[TD5] contains a 1.2-kb insertion in the same region. The *arm*[25B] allele contains a large complex rearrangement/duplication of sequences found in the fragment E and F, and its exact structure has not been determined. However, both of the simple insertions were within 1 kb of each other in fragment E, so it is likely that this short region is critical for *arm* function.

To identify transcribed sections within the *arm* region, the subcloned fragments were used to probe Northern blots of wild-type embryonic poly(A)+ RNA. At least six transcriptional units were identified in the 35-kb region, and their arrangement is shown in Figure 2b. Of these, only the 3.2-kb transcript detected by fragments D–F is directly interrupted by the simple inser-
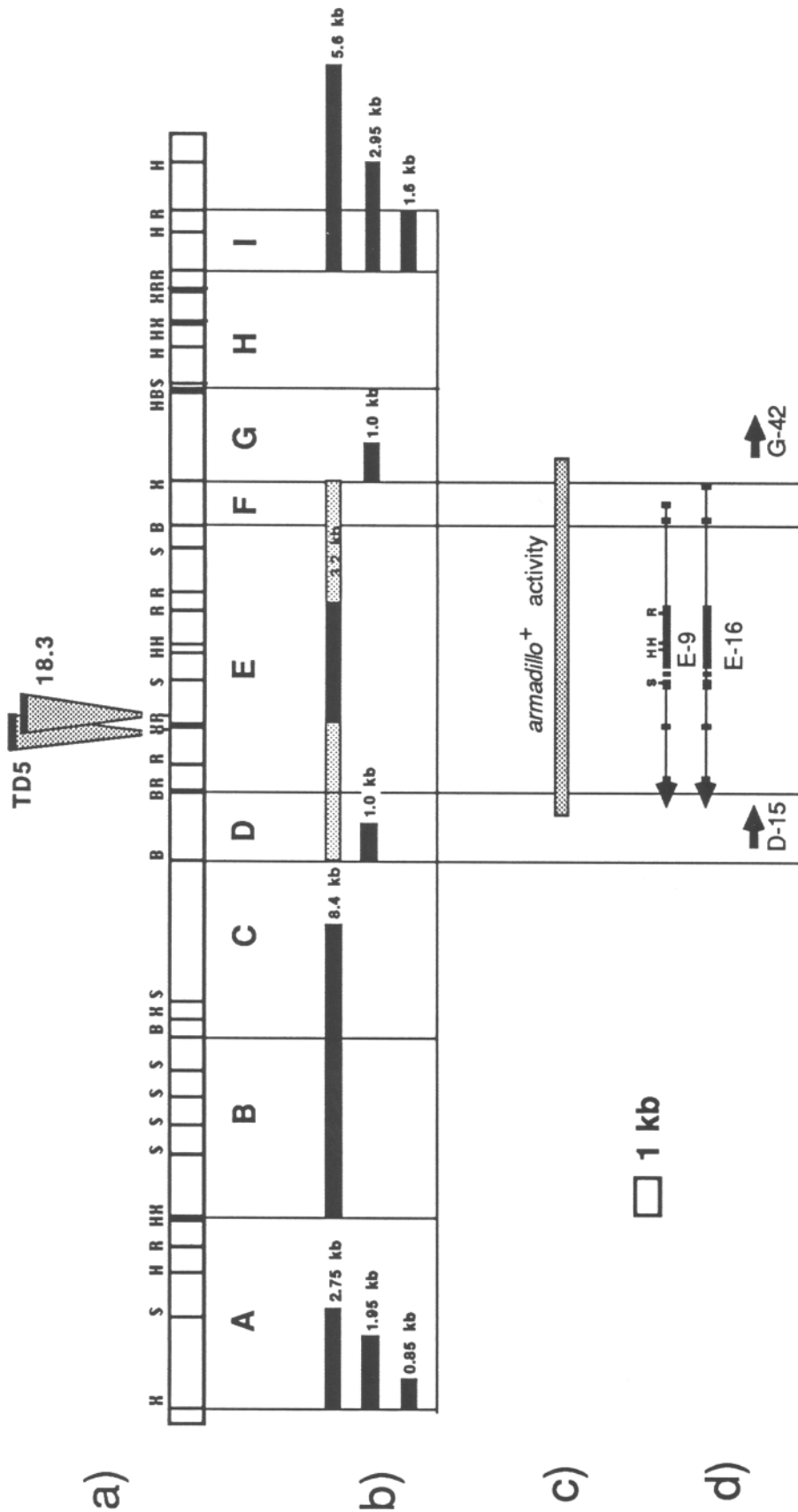
Figure 2. Transcriptional organization of the *arm* region. (*a*) Restriction map of a 35-kb segment of wild-type DNA from the *arm* region. (B) *Bam*HI; (H) *Hind*III; (R) *Eco*RI; (S) *Sac*I; (X) *Xho*I. Inverted triangles show the location of the insertions in the TD5 and 18.3 dysgenic alleles of *arm*, and capital letters A–I below the map indicate the genomic fragments subcloned for detailed analyses. (*b*) Transcriptional units detected by the subcloned genomic fragments. The sizes of transcripts are represented by dark horizontal bars aligned with the left margin of the fragment detecting them, and their size is given to the right of the bar. The 3.2-kb transcript detected by fragment E is also hybridized weakly with fragments D and F (stippled horizontal bars). (*c*) *Pst*I partial digest fragment (10.5 kb) that fully rescues *arm* mutants in germ-line transformation experiments. This fragment contains all sequences encoding the 3.2-kb transcript but only small regions of the flanking D15 and G42 genes. (*d*) Fine structure of E16 and E9 *arm* transcripts and both flanking genes. The 5′ end of the *arm* E16 cDNA is ~600 bp from the 5′ end of the G42 cDNA, and the 3′ ends of both *arm* cDNAs overlap the 3′ end of the D15 cDNA by ~120 bp. The E9 and E16 *arm* transcripts differ only in their short noncoding first exons and share six protein-coding exons contained within a 10-kb genomic region.

tions present in the dysgenic alleles, so this transcriptional unit is likely to be the *arm* gene.

Identification of the arm gene by germ-line transformation

To identify the sequences surrounding the dysgenic insertions that are necessary for *arm* gene activity, we have tested the ability of DNA from this region to rescue the *arm* mutant phenotype using the germ-line transformation system described by Spradling and Rubin (1982). Because the insertions suggest that the 3.2-kb transcript is likely to be essential for *arm* function, we tested a 10.5-kb genomic fragment (Fig. 2c), which contains all of the region encoding the 3.2-kb transcript but only very small parts of both flanking genes. Independent transformed lines containing the 10.5-kb genomic fragment at autosomal positions were established and crossed with strong (YD35) and weak (25B) alleles of *arm* to test for the ability of the 10.5-kb fragment to rescue mutations in the *arm* gene. Of nine autosomal lines that were tested, six rescued both strong and weak *arm* alleles to 100% adult viability and fertility. The remaining three lines completely rescued the embryonic phenotype but did not provide sufficient *arm* activity to allow for adult viability.

These results demonstrate that the *arm* gene is contained within the 10.5-kb genomic fragment shown in Figure 2c. Because this region contains only very small regions of the D15 and G42 neighboring transcripts (Fig. 2d), it is highly unlikely that either of these RNAs are associated with *arm* function. Thus, the 3.2-kb transcript (which is interrupted by insertions in two dysgenic alleles) is the only RNA that is entirely encoded within the rescuing region, and we conclude that this transcript represents the *arm* gene. Furthermore, because the majority of transformation lines obtained fully rescued strong *arm* alleles, all *arm*-specific regulatory sequences necessary for minimal activity appear to be contained within this 10.5-kb region.

Structural analysis of the arm gene

To understand the transcriptional organization of the *arm* region in more detail, we obtained several full-length cDNAs for the 3.2-kb *arm* transcript (E cDNAs) and the neighboring 1.0-kb transcripts located to the left and right of the *arm* gene (D15 and G42, respectively). The regions of genomic DNA encoding each transcript were determined by hybridizing cDNA probes to Southern blots of the genomic region, and this information was used to align the cDNA restriction maps with their corresponding genomic counterparts to obtain the approximate intron–exon structure of each gene. Strand-specific RNA probes from each cDNA were hybridized to Northern blots to determine the direction of transcription relative to the genomic map. Finally, the cDNAs and corresponding genomic fragments were sequenced on both strands to determine the transcriptional organization of the *arm* gene and the sequence of

any encoded protein(s). The detailed structural arrangement of the *arm* region, based on the results of these experiments, is shown in Figure 2d, and the nucleotide sequence of the 5'-nontranscribed, 5'-noncoding, protein-coding, and 3'-noncoding regions is shown in Figure 3.

The DNA sequence analysis demonstrated the existence of two classes of *arm* RNAs, designated E9 and E16 for their cDNA counterparts. The cDNAs for E9 and E16 begin with different 142-bp (E9) or 104-bp (E16) 5' sequences that correspond to regions in the genomic sequence that are 500 bp apart (Fig. 3). To obtain further information on the origin and structure of the E9 and E16 RNAs, we performed S1 and primer extension analyses using probes or primers that were specific to either E9 or E16 sequences. These experiments demonstrated that the 5' ends of the E9 and E16 cDNAs represent actual 5' ends of transcripts found in vivo and that these transcripts begin with different first exons (Figs. 2d and 3). Furthermore, no class of RNA was detected that contains both first exons. Thus, it is likely that the E9 and E16 *arm* transcripts arise through independent transcriptional initiation, although neither 5' end is preceded by the typical RNA polymerase II promoter (TATA) consensus sequence (Benoist and Chambon 1981).

The *arm* gene covers a region of ~10 kb between two closely flanking genes, G42 and D15 (Figs. 2d and 3). The 5' end of the E16 first exon is separated from the approximate 5' end of the neighboring gene G42 (represented by the G42 cDNA) by only 622 bp, and the G42 and *arm* gene are transcribed in opposite directions. The E9 first exon begins within the first intron of E16 and is spliced to a short second exon shared with E16. Altogether, the unique E9 and E16 first exons are spliced to a series of six common exons that contain a 2529-bp open reading frame and 3'-nontranslated regions. The 3' end of the *arm* gene is also very close to the flanking transcriptional unit and overlaps the 3' end of the adjacent D15 gene (represented by the D15 cDNA (Fig. 2d)) by 121 bp. It is possible that this close juxtaposition of opposed transcriptional units may influence their respective transcriptional termination sites. Sequence analysis of various polyadenylated *arm* cDNAs suggests that the 3' ends of *arm* RNAs are processed at two alternative positions (Fig. 3), both of which are preceded by variations of the consensus AATAAA sequence (for review, see Birnstiel et al. 1985). It is interesting to note that the overall structural organization of the *arm* gene closely resembles that of the *Drosophila* actin 5C gene, which also produces transcripts that differ in the use of alternative noncoding first exons and 3'-processing sites (Bond and Davidson 1986; Vigoreaux and Tobin 1987).

The arm gene encodes a protein containing novel internal repeats

The E9 and E16 transcripts both encode a potential polypeptide of 843 amino acids within their only long open reading frame (Fig. 3). This protein is large and acidic,

Riggleman et al.

```
                                    G42 gene 5' end ◄
GGATCCGCCGCCAGCTGCTGTGACCATAACCGATGAACCCACTTCTACGACTGCATAAACCAATATTGCTTGCCCAACTTAAATATGTAT    90
TAAGAAATTAAAGACCAGTGTTTGAATACTTTGGAAATTTAAAAAGTGCGCTTTCTATTTTGAAGCAAGTTGGCAGCCACCTGCGATAGA   180
GGAGTTATCGGTTATCTGTTCTGCGCTCATATACACAGTTTTCATAGTCGGGGATTTTTGTCGAATTTAAATATCTAAGTGCGTCAAAAG   270
TGCGTTAACCTGCCTTTGAGCAGATAAATTTACTTTCGAAGCGGTAAGTAATAGTAAAGTCATAGATATTCTTTAATTCGTCGCCCAAA   360
AACAATAGTTTTTGGGCGTACCTGTGTTGTCATACACCCTCATGTACACTCAGGTGCTTGTGTTGGTAAATGTTTATACAAGATAAACGG   450
TGAGTTCGAGATAACATGTAACATGAACTAGAAATAAATCGCCTTAGCTAAATTATAAATATTGTGAATTGAGTTTCTTCCGGATGAAAT   540
GCTATACAAATTTCGTGAGATTTTTGTGTTCTCAGAACAAAAAAAAATTTGAAAAATTCCTTTGGCTATCGCACAGGGCCCATCGAGTTA   630
                               ►  5' end, arm E16 first exon
TCGGGACAGCAGCAGGCCATCGCTAGTTGGGTTCGTCGCCATATACACAGTTTTTCTCGAGCAGACAATTCGCTGTTATTTTGTTTCTT   720
ATAGCGTTTTAAATAAATAAATCCTGCTCTAAAGCAGACTTAAATTCCTTGGACGCGGTGAGTACTGTTAGTCGGTAGTGAATCCGTTTG   810
GAAAGCCAAAGTTTCTGTGTGCCATGAAAGTTGGAAAGCCATGGCAAACAAGTCGGCGCGGGGGTGGGCAGGGGTTCGCTGGATGTGCGC   900
CTGTGTATGTGTGTGTGTATGATCTGGATACACACAAACGGCATTTTTATGTATTCCTCCCCGTAAATCGGGCGCTTGTTTTGTTGCTCG   990
CTGGAATGTAAACAATGCCACAGACAGGAAATCGTTTTCTTCGAATCTAGAGGCAGGCTTATCGTTAGGTGTTTTATCTAATCGACAACT  1080
GAAATGGAGTTGGGATTCTTTCTGAGCTGCCTACGCGAGTATGTGTGTGTAGGCACAATCGTTTGGATGTGTGGCAACTCTGCGCAGCAC  1170
TGCATTTGTATGCTGCACTGCCGCGGCCCTTTACTTTCTCTCCCAGTCTCTTTCCCTGTCCCTTTCCCGCACTCGCACGCGCGGTGTTGC  1260
                               ►  5'end, arm E9 first exon
AAGTGCCTTGCTTTCGTTGCGCTCTCTCGCATCGTGATTCTGTTCCGTTTAAACGAGCGTGCTGGTCGGTCGCATTTTCCTGGCGCCTT  1350
TTTCGAGAGAAGTGAAAAATCCATTTTGTTTTTTGTTGGAGTGCATGAGATTTGAAAGAGTGCTAAACAAAGTAATTAAATAACTCAAGG  1440
TTCGTGGCCAAAGTCTTCGGCAGGTAGACACACACATACCCACAAATACCCGTTCTACCTGCGTAAGAATGTATTAGTGCGTTTGTTTGT  1530
ATACGCAGCATATTTCATATTATGCTTCGGCATCTCAAAGCGGAAATGAGTTAAAAAAAAAGACCAACGACTAGAATAAGTGACTAACGG  1620
ACGATATCTGCGTCTGTTGGCAGTCTGCAAACACAACAAGCCCTTCTTTCCGGGGCCCCCATTTAGTGGAGTTCCGTGGTCTCCATCCGA  1710
                                                            E9 and E16 exon 2
AGCACACACGAACTCCATGTTATTGACTGTCCAGCACATTACTCATATACTCTCCGTCTCTTTCTTGCAGGTGTGGTGTAAAGCGCCTCA  1800
                10                              20
        M  S  Y  M  P  A  Q  N  R  T  M  S  H  N  N  Q  Y  N  P  P  D  L  P  P  M  V  S
ATCACCAAGATGAGTTACATGCCAGCCCAGAATCGAACCATGTCGCATAATAATCAATACAATCCACCTGATCTGCCGCCGATGGTATCC  1890
                          exon 2  exon 3
        30                      40                              50
   A  K  E  Q  T  L  M  W  Q  Q  N  S  Y  L  G  D  S  G  I  H  S  G  A  V  T  Q  V  P  S  L
GCCAAGGAGCAGACCCTCATGTGGCAGCAGAATTCGTACTTGGGCGACTCCGGCATCCACTCGGGTGCCGTGACCCAGGTTCCATCGCTG  1980
           60                      70                              80
   S  G  K  E  D  E  E  M  E  G  D  P  L  M  F  D  L  D  T  G  F  P  Q  N  F  T  Q  D  Q  V
TCTGGCAAGGAGGACGAGGAGATGGAGGGAGATCCACTTATGTTCGACCTGGACACCGGTTTCCCGCAGAATTTCACACAAGACCAAGTG  2070
           90                      100                             110
   D  D  M  N  Q  Q  L  S  Q  T  R  S  Q  R  V  R  A  A  M  F  P  E  T  L  E  E  G  I  E  I
GACGATATGAACCAGCAACTGAGCCAGACACGCTCCCAACGTGTACGTGCTGCCATGTTTCCGGAAACCCTGGAGGAGGGCATTGAGATT  2160
           120                     130                             140
   P  S  T  Q  F  D  P  Q  Q  P  T  A  V  Q  R  L  S  E  P  S  Q  M  L  K  H  A  V  V  N  L
CCCTCCACCCAGTTTGATCCCCAACAGCCGACGGCAGTGCAACGTCTTTCGGAACCGTCACAAATGCTAAAGCACGCGGTGGTCAATCTG  2250
           150                     160                             170
   I  N  Y  Q  D  D  A  E  L  A  T  R  A  I  P  E  L  I  K  L  L  N  D  E  D  Q  V  V  V  S
ATCAACTACCAGGACGACGCTGAGCTGGCAACCAGGGCCATACCCGAGTTGATCAAGCTGCTGAACGATGAGGATCAGGTGGTAGTTTCC  2340
           180                     190                             200
   Q  A  A  M  M  V  H  Q  L  S  K  K  E  A  S  R  H  A  I  M  N  S  P  Q  M  V  A  A  L  V
CAGGCCGCCATGATGGTCCACCAGCTGTCTAAGAAGGAGGCCTCGCGACATGCCATTATGAACAGCCCTCAGATGGTAGCCGCTTTGGTG  2430
           210                     220                             230
   R  A  I  S  N  S  N  D  L  E  S  T  K  A  A  V  G  T  L  H  N  L  S  H  H  R  Q  G  L  L
CGTGCCATCTCTAACAGCAACGATCTGGAGAGCACCAAGGCAGCGGTAGGAACACTGCACAACTTATCACACCATCGCCAGGGTCTGCTG  2520
           240                     250                             260
   A  I  F  K  S  G  G  I  P  A  L  V  K  L  L  S  S  P  V  E  S  V  L  F  Y  A  I  T  T  L
GCCATCTTCAAGAGTGGCGGCATCCCGGCACTCGTCAAGTTGCTCTCCTCGCCAGTGGAGAGTGTGCTGTTCTATGCAATTACCACTCTG  2610
           270                     280                             290
   H  N  L  L  L  H  Q  D  G  S  K  M  A  V  R  L  A  G  G  L  Q  K  M  V  T  L  L  Q  R  N
CACAATTTGCTGCTCCACCAGGATGGCTCTAAGATGGCTGTGCGCCTGGCCGGCGGGCTTCAGAAGATGGTTACTCTGCTGCAACGAAAC  2700
           300                     310                             320
   N  V  K  F  L  A  I  V  T  D  C  L  Q  I  L  A  Y  G  N  Q  E  S  K  L  I  I  L  A  S  G
AACGTTAAGTTTCTGGCTATCGTCACAGATTGCTTGCAAATTCTGGCCTATGGTAACCAGGAGAGCAAGTTAATAATTCTTGCCTCCGGA  2790
           330                     340                             350
   G  P  N  E  L  V  R  I  M  R  S  Y  D  Y  E  K  L  L  W  T  T  S  R  V  L  K  V  L  S  V
GGGCCCAACGAACTCGTGCGCATTATGCGCTCCTACGACTACGAGAAGCTTCTGTGGACCACTTCGCGGGTACTGAAAGTGCTCTCCGTT  2880
           360                     370                             380
   C  S  S  N  K  P  A  I  V  D  A  G  G  M  Q  A  L  A  M  H  L  G  N  M  S  P  R  L  V  Q
TGCTCCAGCAACAAGCCGGCCATCGTGGATGCCGGTGGAATGCAGGCGCTGGCTATGCACTTGGGTAACATGTCACCGCGCCTTGTGCAA  2970
           390                     400                             410
   N  C  L  W  T  L  R  N  L  S  D  A  A  T  K  V  E  G  L  E  A  L  L  Q  S  L  V  Q  V  L
AACTGTTTGTGGACGCTCCGCAATCTTTCGGATGCAGCCACTAAGGTGGAGGGCCTTGAAGCTTTGCTCCAATCTCTCGTCCAGGTTCTG  3060
```
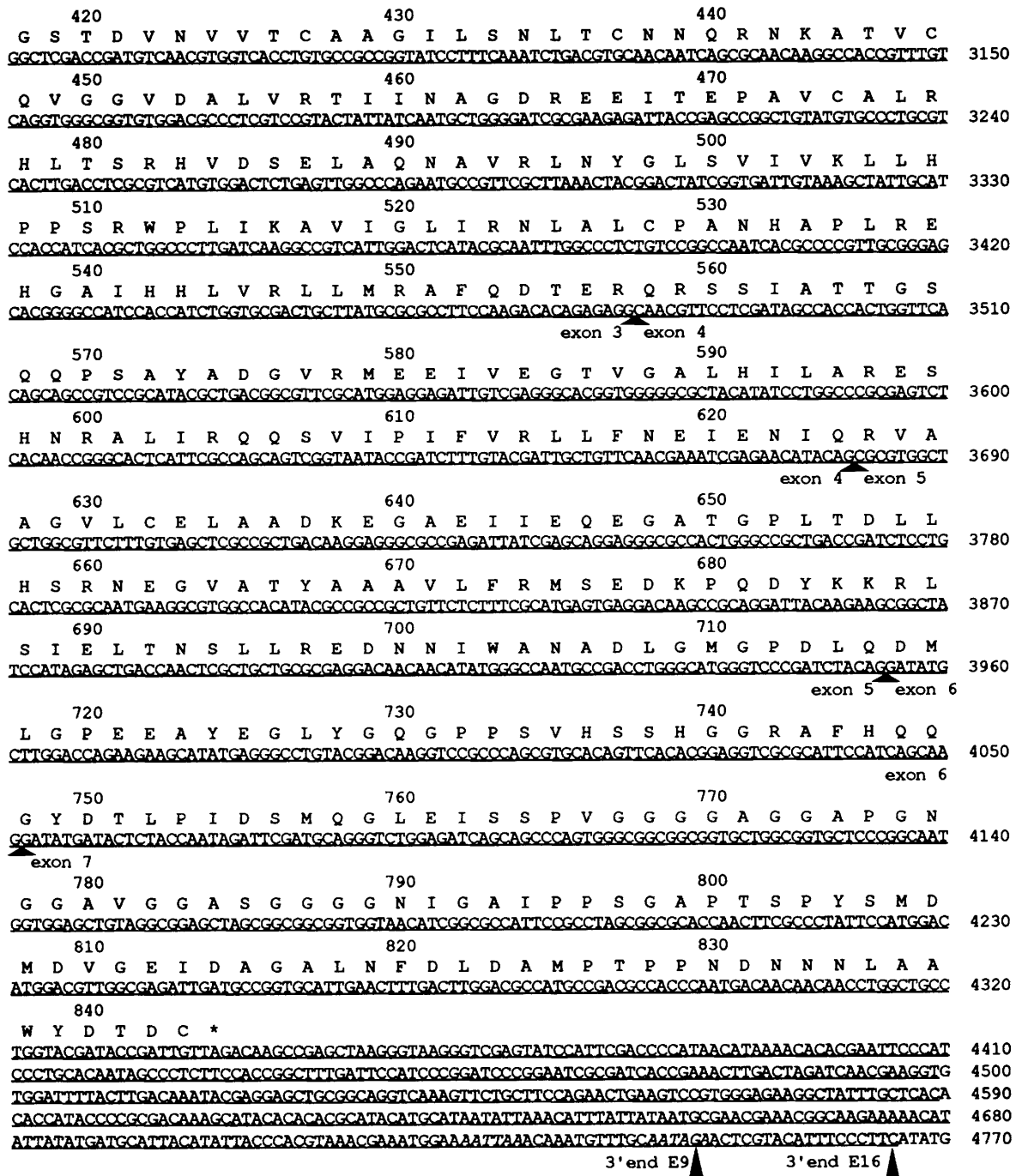
```
            420                        430                        440
     G  S  T  D  V  N  V  V  T  C  A  A  G  I  L  S  N  L  T  C  N  N  Q  R  N  K  A  T  V  C
     GGCTCGACCGATGTCAACGTGGTCACCTGTGCCGCCGGTATCCTTTCAAATCTGACGTGCAACAATCAGCGCAACAAGGCCACCGTTTGT  3150
            450                        460                        470
     Q  V  G  G  V  D  A  L  V  R  T  I  I  N  A  G  D  R  E  E  I  T  E  P  A  V  C  A  L  R
     CAGGTGGGCGGTGTGGACGCCCTCGTCCGTACTATTATCAATGCTGGGGATCGCGAAGAGATTACCGAGCCGGCTGTATGTGCCCTGCGT  3240
            480                        490                        500
     H  L  T  S  R  H  V  D  S  E  L  A  Q  N  A  V  R  L  N  Y  G  L  S  V  I  V  K  L  L  H
     CACTTGACCTCGCGTCATGTGGACTCTGAGTTGGCCCAGAATGCCGTTCGCTTAAACTACGGACTATCGGTGATTGTAAAGCTATTGCAT  3330
            510                        520                        530
     P  P  S  R  W  P  L  I  K  A  V  I  G  L  I  R  N  L  A  L  C  P  A  N  H  A  P  L  R  E
     CCACCATCACGCTGGCCCTTGATCAAGGCCGTCATTGGACTCATACGCAATTTGGCCCTCTGTCCGGCCAATCACGCCCCGTTGCGGGAG  3420
            540                        550                        560
     H  G  A  I  H  H  L  V  R  L  L  M  R  A  F  Q  D  T  E  R  Q  R  S  S  I  A  T  T  G  S
     CACGGGGCCATCCACCATCTGGTGCGACTGCTTATGCGCGCCTTCCAAGACACAGAGAGGCAACGTTCCTCGATAGCCACCACTGGTTCA  3510
                                                            exon 3 ▲ exon 4
            570                        580                        590
     Q  Q  P  S  A  Y  A  D  G  V  R  M  E  E  I  V  E  G  T  V  G  A  L  H  I  L  A  R  E  S
     CAGCAGCCGTCCGCATACGCTGACGGCGTTCGCATGGAGGAGATTGTCGAGGGCACGGTGGGGGCGCTACATATCCTGGCCCGCGAGTCT  3600
            600                        610                        620
     H  N  R  A  L  I  R  Q  Q  S  V  I  P  I  F  V  R  L  L  F  N  E  I  E  N  I  Q  R  V  A
     CACAACCGGGCACTCATTCGCCAGCAGTCGGTAATACCGATCTTTGTACGATTGCTGTTCAACGAAATCGAGAACATACAGCGCGTGGCT  3690
                                                               exon 4 ▲ exon 5
            630                        640                        650
     A  G  V  L  C  E  L  A  A  D  K  E  G  A  E  I  I  E  Q  E  G  A  T  G  P  L  T  D  L  L
     GCTGGCGTTCTTTGTGAGCTCGCCGCTGACAAGGAGGGCGCCGAGATTATCGAGCAGGAGGGCGCCACTGGGCCGCTGACCGATCTCCTG  3780
            660                        670                        680
     H  S  R  N  E  G  V  A  T  Y  A  A  A  V  L  F  R  M  S  E  D  K  P  Q  D  Y  K  K  R  L
     CACTCGCGCAATGAAGGCGTGGCCACATACGCCGCCGCTGTTCTCTTTCGCATGAGTGAGGACAAGCCGCAGGATTACAAGAAGCGGCTA  3870
            690                        700                        710
     S  I  E  L  T  N  S  L  L  R  E  D  N  N  I  W  A  N  A  D  L  G  M  G  P  D  L  Q  D  M
     TCCATAGAGCTGACCAACTCGCTGCTGCGCGAGGACAACAACATATGGGCCAATGCCGACCTGGGCATGGGTCCCGATCTACAGGATATG  3960
                                                                 exon 5 ▲ exon 6
            720                        730                        740
     L  G  P  E  E  A  Y  E  G  L  Y  G  Q  G  P  P  S  V  H  S  S  H  G  G  R  A  F  H  Q  Q
     CTTGGACCAGAAGAAGCATATGAGGGCCTGTACGGACAAGGTCCGCCCAGCGTGCACAGTTCACACGGAGGTCGCGCATTCCATCAGCAA  4050
                                                                            exon 6
            750                        760                        770
     G  Y  D  T  L  P  I  D  S  M  Q  G  L  E  I  S  S  P  V  G  G  G  A  G  G  A  P  G  N
     GGATATGATACTCTACCAATAGATTCGATGCAGGGTCTGGAGATCAGCAGCCCAGTGGGCGGCGGCGGTGCTGGCGGTGCTCCCGGCAAT  4140
            ▲ exon 7
            780                        790                        800
     G  G  A  V  G  G  A  S  G  G  G  G  N  I  G  A  I  P  P  S  G  A  P  T  S  P  Y  S  M  D
     GGTGGAGCTGTAGGCGGAGCTAGCGGCGGCGGTGGTAACATCGGCGCCATTCCGCCTAGCGGCGCACCAACTTCGCCCTATTCCATGGAC  4230
            810                        820                        830
     M  D  V  G  E  I  D  A  G  A  L  N  F  D  L  D  A  M  P  T  P  P  N  D  N  N  N  L  A  A
     ATGGACGTTGGCGAGATTGATGCCGGTGCATTGAACTTTGACTTGGACGCCATGCCGACGCCACCCAATGACAACAACAACCTGGCTGCC  4320
            840
     W  Y  D  T  D  C  *
     TGGTACGATACCGATTGTTAGACAAGCCGAGCTAAGGGTAAGGGTCGAGTATCCATTCGACCCCATAACATAAAACACACGAATTCCCAT  4410
     CCCTGCACAATAGCCCTCTTCCACCGGCTTTGATTCCATCCCGGATCCCGGAATCGCGATCACCGAAACTTGACTAGATCAACGAAGGTG  4500
     TGGATTTTACTTGACAAATACGAGGAGCTGCGGCAGGTCAAAGTTCTGCTTCCAGAACTGAAGTCCGTGGGAGAAAGGCTATTTGCTCACA  4590
     CACCATACCCCGCGACAAAGCATACACACACGCATACATGCATAATATTAAACATTTATTATAATGCGAACGAAACGGCAAGAAAAACAT  4680
     ATTATATGATGCATTACATATTACCCACGTAAACGAAATGGAAAATTAAACAAATGTTTGCAATAGAACTCGTACATTTCCCTTCATATG  4770
                                        3'end E9 ▲            3'end E16 ▲
```

Figure 3. Sequence of the *arm* gene. The region shown begins at the right end of the 10.5-kb fragment used for germ-line transformation (Fig. 2c) and terminates 5 bp past the 3' end of the E16 transcript. Horizontal arrowheads mark the 5' end of the G42, E9, and E16 cDNAs, and vertical arrowheads designate the 3' ends of the E9 and E16 transcripts. Exon sequences are underlined, potential polyadenylation sites for E9 and E16 are italicized, and intron/exon junctions are marked with triangles; the sequence of introns 2–6 is not shown. The DNA sequence shown includes (in 5' → 3' order) the 5' end of the neighboring G42 gene (1–52); the G42-*arm* E16 intergenic region (53–671); the E16 first exon (672–777); the E16 first intron (778–1780), which contains the E9 first exon (1299–1439) and the E9 first intron (1440–1780); the six protein-coding exons shared by E9 and E16 (1781–4765); and the E9 and E16 3' ends (4746 and 4765, respectively). The 843-amino-acid sequence of the potential *arm* protein is shown from the first ATG codon (1810) in the only long open reading frame to the first termination codon (4339). This protein has a molecular mass of 91.1 kD, acidic regions of ~150 amino acids at both ends, and a hydrophobic, repetitive central region.

with a predicted molecular mass of 91.1 kD and an estimated isoelectric point of 5.86. Sequence homology searches of current versions of the NBRF, SWISSPROT, and translated GenBank protein data bases failed to reveal any significant similarities between the *arm* protein and other sequences that might suggest a role for the protein in the segmentation process. Structural analysis of the amino acid sequence, however, identified several features of the protein that may relate to its biological function. First, the protein does not contain any long continuous stretches of hydrophobic amino acids that resemble signal or membrane-spanning sequences. Thus, the *arm* protein is expected to be found inside the cell and is unlikely to be an integral membrane protein. Second, a 23-amino-acid glycine-rich region that contains no charged amino acids is found near the carboxyl terminus (Fig. 3). Such sequences, often referred to as *pen* repeats (Haynes et al. 1987), have been found in many other proteins, including intermediate filaments (Steinert et al. 1985). This sequence is probably not hydrophobic enough to serve as a membrane-spanning domain but may act as a highly flexible section that connects different regions in the protein. Third, the charged and hydrophobic amino acids are distributed asymmetrically in the protein so that regions of ~150 amino acids flanking both the amino and carboxyl termini are highly acidic and hydrophilic, whereas the central region is basic and hydrophobic. Finally, dot matrix analysis demonstrated that the central hydrophobic/basic region of the protein contains a series of internal repeats.

The central region (positions 159–689, Fig. 3) can be aligned in a series of twelve and one-half 42 amino acid repeats, which are linked continously from end to end (Fig. 4a). A consensus sequence can be formed at 25 of the 42 positions, and the individual repeats have between 28% and 80% identical matches with this consensus (Fig 4a). The structure and significance of the repeated region can be further evaluated by considering the general type of amino acid (hydrophobic, hydrophilic, or polar) used at each position in the aligned repeats (Koenig et al. 1988). Using these guidelines, the same type of amino acid is found in a majority of repeats at 38 of 42 positions, and the same type is used by virtually all of the repeats at 18 of 42 positions. Hence, the general chemical and structural features of the repeats have been conserved in the evolution of the *arm* protein (especially at some positions), so it is likely that they are important for its function.

The general-type consensus can also be used to consider the potential structural characteristics of conserved regions within the repeats (Fig. 4b). The consensus suggests that the repeats can be divided into two similar subdomains (A and B). These each contain two hydrophobic regions (P1 and P2) and two polar/hydrophilic regions (T1 and T2). The P1 and P2 regions are highly conserved and contain many hydrophobic residues that are likely to participate in helical or extended secondary structures (Garnier et al. 1978). Conversely, the T1 and T2 regions contain a high percentage of polar amino acids (and certain hydrophilic amino acids) that

often break or deform secondary structures (Garnier et al. 1978). Computer analysis of the individual repeats confirmed that the actual T1 and T2 sequences are likely to form turns, whereas the P1 and P2 regions have a high probability of forming organized secondary structures. Because the T1 and T2 regions are arranged between the short hydrophobic P1 and P2 sequences, any continuous secondary structures in the repeat region are likely to be distorted or disrupted.

Because of the similarity between the A and B subdomains, it seems possible that the 42 amino acid repeat itself may have been formed by duplication. In fact, direct alignment of all A and B region sequences shows that they often use identical amino acids at the same position. However, the A region sequences are much more related to other A regions than they are to any B regions, and vice versa. This implies that the 42-amino-acid unit itself was repeated to provide the 12.5 repeats found in the central region of the *arm* protein. Because the A and B regions are still conserved in their general structure, twofold symmetry of the 42 amino acid repeat may still be important for its biological function.

In addition to the amino acid sequence conservation noted above, another significant feature of the repeats is that they are highly conserved in length, the majority being exactly 42 residues long. Furthermore, they are linked together in a continuous fashion without any separating regions, which are often found in other proteins containing internal repeats. The conservation of these features may allow the *arm* protein to interact with itself or other molecules in a highly ordered fashion.

Transcriptional analysis shows that arm RNAs are abundant and present throughout development

Genetic analysis has shown that the *arm* gene is required for oogenesis (Wieschaus and Noell 1986), embryogenesis, and the development of imaginal discs (M. Williams et al., in prep.). To determine when *arm* transcripts are actually present, we hybridized a 2.5-kb *arm* cDNA fragment to Northern blots of poly(A)$^+$ RNA from various stages of *Drosophila* development. The results (Fig. 5A) show that the *arm* gene produces very abundant transcripts, which are found at all stages but accumulate to the highest levels during early to mid-embryogenesis and early pupal development. The general abundance and developmental distribution of *arm* transcripts are similar to those of the actin 5C transcripts (shown as a hybridization control in Fig. 5A). Direct comparisons show that *arm* transcripts are ~50% more abundant than the actin 5C 2.0-kb transcripts in 4- to 12-hr embryos. Thus, *arm* transcripts are among the most prevalent species in the cell, equaling ~35% of total actin transcripts at this stage or ~0.5% of total poly(A)$^+$ RNA (Fyrberg et al. 1983; Anderson and Lengyel 1984).

To determine the relative abundance of the E9 and E16 transcripts at different stages, short (105 and 107 bp, respectively) transcript-specific probes were constructed from the unique first exons of the E9 and E16 cDNAs.

a.

NUMBER / POSITION		SEQUENCE	SCORE
1	159	RAIPELIKLLNDE-DQVVVSQAAM-MVHQLSKKEASRHAIMNSP	64
2	201	QMVAALVRAISNS-NDLESTKAAVGTLHNLSHHRQGLLAIFKS-	68
3	243	GGIPALVKLLSSP-VESVLFYA-ITTLHNLLLHQDGSKMAVRLA	80
4	285	GGLQKMVTLLQRN-NV-KFLAIVTDCLQILAYGNQESKLIILAS	64
5	327	GGPNELVRIMRSY-DYEKLLWTTSRVLKVLSVC-SSNKPAIVDA	68
6	369	GGMQALAMHLGNM-SP-RLVQNCLWTLRNLSDAATKVEGLEALL	60
7	411	---QSLVQVLGST-DVN-VVTCAAGILSNLTCNNQRNKATVCQV	48
8	450	GGVDALVRTIINAGDREEITEPAVCALRHLTSRHVDSELAQNAVRLN	72
9	497	YGLSVIVKLLHPP-SRWPLIKAVIGLIRNLALC-PANHAPLREH	48
10	539	GAIHHLVRLLMRA-FQDTERQRSSIATTGSQQPSAYADGVRMEEIVE	32
11	585	GTV-GALHILARE-SHNRALIRQQSV-----------------	28
12	609	--IPIFVRLLFNE-IENIQRVAA-GVLCELAADKEGAEII-EQE	60
13	648	GATGPLTDLLHSR-NEGVATYAA-AVLFRMSEDKPQDYKKRLSI	56

```
CONSENSUS    GGIPALVRLL-N--D---LL-AA--VLH-LS-----NKAI----
             VQE  K   R   N    AT    T R  A       SELA
                      S   S    V
```

b.


```
PHOBIC    *     ** **       **           ***   *  **   *              ***
POLAR    **  **                               **  *              *
PHILIC         *          ****              **     **   ***    *
         T1A  P1A      T2A       P2A  T1B  P1B      T2B       P2B
         |     REGION A      |        REGION B        |
```

Figure 4. Analysis of the *arm* protein repeats. (*a*) The central region of the *arm* protein (amino acids 159–689) can be aligned in a continuous series of twelve and one-half 42 amino acid repeats. A single space was added between positions 13 and 14 of each repeat to allow for an extra amino acid in repeat 8, and a second space was added to most other repeats to optimize alignment. Most repeats are exactly 42 amino acids long, and a consensus sequence can be formed at 25 of the 42 positions. Each position in the consensus sequence lists one or two amino acids that occur at that position in a majority of repeats; in several cases, three amino acids are listed if a third amino acid is used as often as the first or second residue. Amino acids that match the consensus in each repeat are underlined; the individual repeats have between 28% and 80% identity with this consensus (SCORE). (*b*) The general chemical and structural features of the repeats can be analyzed by considering the type of amino acid used at each position (Koenig et al. 1988). When the amino acids are classified as hydrophilic (D, E, R, K, H, Q, N), hydrophobic (L, I, V, F, Y, W, M, T, A), or polar (G, A, P, C, P, S, T), a single type is conserved in the majority of repeats at 38 of 42 positions. The repeats can be divided into two similar subdomains (A and B), which each contain four short characteristic regions (two hydrophobic regions, P1 and P2, and two polar-rich regions, T1 and T2). The arrangement and predicted structural properties of these regions suggest that the repeated region contains a series of short structured hydrophobic regions that are distorted or disrupted by alternating turns.

Northern blot analyses were performed with these probes under conditions that would allow for accurate quantitation of the two messages. These blots (Fig. 5B,C) show that both transcripts have very similar patterns of developmental expression and are found at the highest levels during early to mid-embryogenesis. Direct measurements and longer exposure times show that the E16 RNA is about two (at 4–12 hr) to five times (at 0–4 hr and in larvae, pupae, and adults) more abundant than E9, but both transcripts are present at all stages. Thus, the ratio of E9 to E16 transcripts varies only ~2.5-fold during development, and neither transcript is found exclusively at one stage. These analyses show that the difference in developmental expression between the two *arm* RNAs is subtle and resembles expression patterns of the different actin 5C transcripts (Bond and Davidson 1986; Vigoreaux and Tobin 1987). In addition, any preferential requirement for the different RNA isoforms that may exist does not appear to change greatly throughout

development. The Northern blots also demonstrate that the small transcript-specific probes and the full-length cDNA probe detect only the expected 3.2-kb *arm* transcripts and thus should be *arm* specific for in situ hybridization experiments.

arm transcripts are uniformly distributed and expressed in a wide range of cell types

To determine whether *arm* RNAs are expressed in tissue-specific or localized patterns, in situ hybridizations were performed on sectioned ovaries, embryos, and larval tissues. Labeled antisense RNAs from either a 2.5-kb cDNA subclone (for greatest sensitivity in detecting general accumulation) or the E9 or E16 0.1-kb transcript-specific subclones were used as probes. Alternate sections of the same tissues were probed with *engrailed* antisense RNA (Poole et al. 1985) as a control for localized expression (embryos and larval tissue only) or with
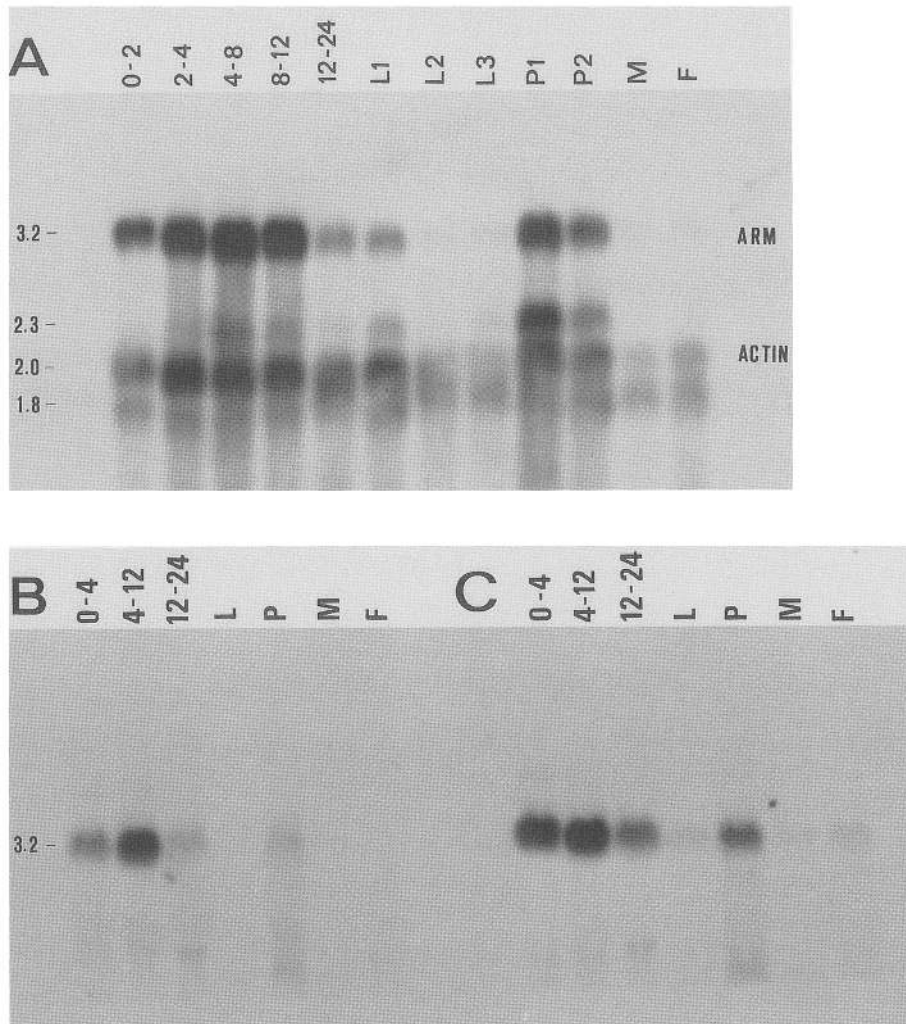
Figure 5. Expression of *arm* transcripts during development. (*A*) Northern blot of poly(A)+ RNAs from different stages of development were hybridized simultaneously with probes that detect the E9 and E16 *arm* 3.2-kb transcripts and the 1.7-, 2.0-, and 2.3-kb actin 5C transcripts. Stages shown are 0- to 2-hr, 2- to 4-hr, 4- to 8-hr, 8- to 12-hr, 12- to 24-hr, 1.5-day (L1), 3.5-day (L2), and 5.5-day (L3) larvae, 6.5- (P1) and 8.5-day (P2) pupae, and male (M) and female (F) adults. Like the actin 5C hybridization control, *arm* transcripts are detected at all stages of development but are most abundant during embryogenesis and pupation. (*B*) 0- to 4-hr; 4- to 12-hr; 12- to 24-hr; 1.5-day, 3.5-day, and 5.5-day combined larval RNAs (L); 6.5- and 8.5-day combined pupal RNAs (P); and male (M) and female (F) adult RNAs probed with the E9 first exon probe. (*C*) Identical blot to *b*, probed with the E16 first exon probe. The patterns of E9 and E16 RNA accumulation are similar during development, but the E16 RNA is between twofold (at 4–12 hr) and fivefold (at pupation) more abundant than the E9 RNA at different stages.

a sense-strand RNA probe from the 2.5-kb cDNA construct as a control for nonspecific background.

In embryos, *arm* transcripts were present at all stages but reached the highest levels of accumulation during early to mid-embryogenesis, in agreement with the Northern blot data. Using the 2.5-kb probe, *arm* RNAs could be detected at low levels during early cleavage stages (cycle 1–9, Foe and Alberts 1983), and it is likely that these transcripts are maternal in origin (Fig. 6a–c). By the blastoderm stage (cycles 10–14), a much stronger

Figure 6. Spatial localization of *arm* RNAs in sectioned embryos. (*a–c*) Preblastoderm embryos in bright field (*a*), or dark field and hybridized with sense strand *arm* probe (*b*) or antisense strand *arm* probe (*c*). Low levels of *arm* RNA are detected at this stage, which are probably maternal in origin. (*d–f*) Blastoderm embryos in bright field (*d*) or dark field and hybridized with sense strand *arm* probe (*e*) or antisense strand *arm* probe (*f*). Moderate levels of *arm* RNA are found in the cortical cytoplasmic (cc) region and around the pole cells (pc). (*g–l*) Germ-band-extended embryos in bright field (*g*) or dark field and hybridized with a sense strand *arm* probe (*h*), an antisense strand engrailed probe (*i*), the E9 first exon antisense strand probe (*j*), the E16 first exon antisense strand probe (*k*), or the 2.5 kb *arm* antisense strand probe (*l*). These results show that although other RNA species (*engrailed, i*) are found only in specific tissues and are localized to defined regions within each segment, both forms of *arm* RNA (*j–l*) are found in all cell types contained in these stages of embryonic development and are uniformly distributed within these tissues. Slightly higher levels of transcripts appear to accumulate in the posterior midgut (pmg) than in the germ band (gb).
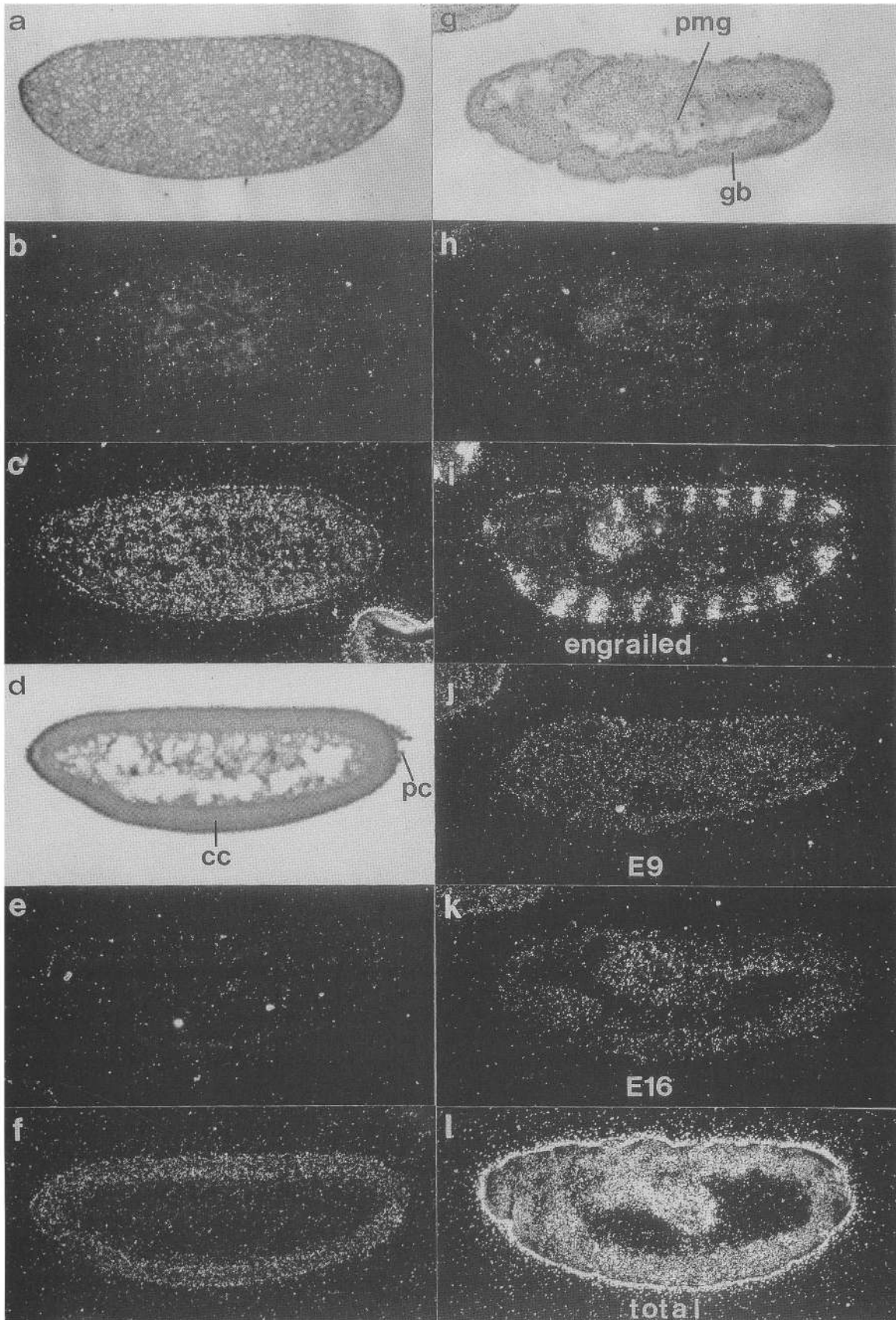
Figure 6. {*See facing page for legend.*}

signal is seen over the cytoplasmic cortex and around the nuclei (Fig. 6d–f). It is difficult to say, however, whether the increased grain density is caused by the transport of cytoplasm to the surface or by early zygotic transcription (Fig. 6d–f). The abundance of *arm* message increases to the highest levels during germ band extension and declines as the germ band retracts.

During germ-band extension stages, both *arm* transcripts are found in all germ layers at similar levels (Fig. 6g–l). Slightly higher levels of E9 and E16 RNAs appear to be present in the posterior midgut than in the other parts of the developing embryo. However, neither the E9, E16, or 2.5-kb probes detected any segmental localization of *arm* transcripts at any stage (Fig. 6c,f,j–l). In contrast, embryos in control sections reproducibly exhibited a segmental pattern of *engrailed* expression (Fig. 6i). Hence, if *arm* transcripts accumulated in this fashion, it is likely such localization would have been detected in these sections. Because *arm* transcripts do not show discrete segmental patterns of expression, the segmental defects associated with loss of *arm* activity may be due to interactions between *arm* and other localized factors, perhaps some of the other segment polarity genes.

Loss of *arm* function also results in pattern defects in tissues derived from the imaginal discs (M. Williams et al., in prep.), so we examined sections of third-instar larvae to analyze the distribution of *arm* RNAs in discs. As was the case with embryos, the E9 and E16 transcripts are also uniformly distributed in imaginal discs (Fig. 7a–d), as compared with the *engrailed* control slides. Thus, the adult positional defects are also likely to be the result of interactions between *arm* and localized factors. Unexpectedly, both *arm* transcripts were also found in most other cell types present in third-instar larval sections. The *arm* RNAs are present at the highest levels in diploid brain and imaginal disc cells (Fig. 6a–d) and at moderate levels in polyploid gut (Fig. 8), salivary gland, and fat body cells. In addition, it is likely that the weak signal seen over muscle cell bodies and hypodermal cells represents specific hybridization also. The presence of *arm* transcripts in these tissues suggests that *arm* may actually be required by all cells, including some that are terminally differentiated and some that are not associated with segmentation processes.

Genetic analyses have shown that the *arm* gene is required for oogenesis and that maternal *arm* activity is contributed to the developing embryo (Wieschaus and Noell 1986), so we also examined sectioned ovaries to analyze the expression of *arm* transcripts in these tissues. Similar to the situation in embryos and larvae, the E9 and E16 transcripts were found in all major cell types in ovaries (nurse cells, follicle cells, and oocyte

chambers) throughout their development (Fig. 7e–h). Likewise, both RNAs have no distinct patterns of spatial distribution within any given cell type, although they are found at much higher levels in the nurse cells at early and middle stages of egg chamber development. The general pattern of *arm* maternal expression differs from that of maternal genes which are required primarily for zygotic development. For example, the *rudimentary* (Ambrosio and Schedl 1984) and *dorsal* (Steward et al. 1985) transcripts are both produced in the nurse cells, primarily during middle to late oogenesis (stages 8–11; King 1970) and are transported continuously to the growing oocyte where they accumulate at the highest levels. In contrast, *arm* transcripts are present at the highest levels in early to mid-oogenesis nurse cells (stage 5–10a), and only relatively small amounts are deposited into the developing oocyte chamber. Thus, *arm* transcripts are associated primarily with the growing nurse cells during oogenesis, and only a fraction of these transcripts are contributed to zygotic development.

Discussion

Identification and structural analysis of the *arm* gene

We have used a dysgenic allele to clone a 35-kb segment of DNA from the *arm* locus and have identified DNA insertions within this region in two *arm* alleles. The insertions interrupt a transcriptional unit encoding two related 3.2-kb mRNAs, and germ-line transformation experiments demonstrated that this unit is the *arm* gene. Because of the high frequency of fully rescuing lines obtained in these studies, it is probable that all *arm*-specific regulatory and coding sequences are located within the 10.5-kb genomic segment contained in the transposon. This region includes the entire RNA coding regions for *arm* transcripts and a suprisingly small amount (~670 bp) of 5′-flanking sequence.

The *arm* gene is present in only one copy per haploid genome and produces two forms of transcripts (referred to as E9 and E16) that might have different functions. Both forms appear to encode the same protein, but it is possible that the use of different transcripts may allow for optimal expression of this protein in various developmental situations. Because they appear to arise through independent transcriptional initiation, the two forms could be regulated by different promoters. Other genes that have a similar transcriptional organization produce RNAs that are tissue specific (mouse amylase; Schibler et al. 1983) or stage specific (*Drosophila* alcohol dehydrogenase; Benyajati et al. 1983). Likewise, sequences in the first exons could influence the relative stability of the two forms. However, Northern blot analyses showed that the relative expression of the E9 and E16 RNAs

Figure 7. Spatial distribution of *arm* transcripts in larval and ovarian tissues. Sectioned larval heads (*a–d*) and stage 10A egg chambers (*e–h*) are shown in dark field after hybridization with a 2.5-kb *arm* sense strand control probe (*a,e*), an E9 first exon antisense strand probe (*b,f*), an E16 first exon antisense strand probe (*c,g*), and a 2.5-kb antisense strand *arm* probe (*d,h*). These results show that both forms of *arm* transcripts are uniformly distributed and present at high levels in larval brains (br), eye-antennal imaginal discs (ed), nurse cells (nc), oocyte chambers (oc), and follicle cells (fc).
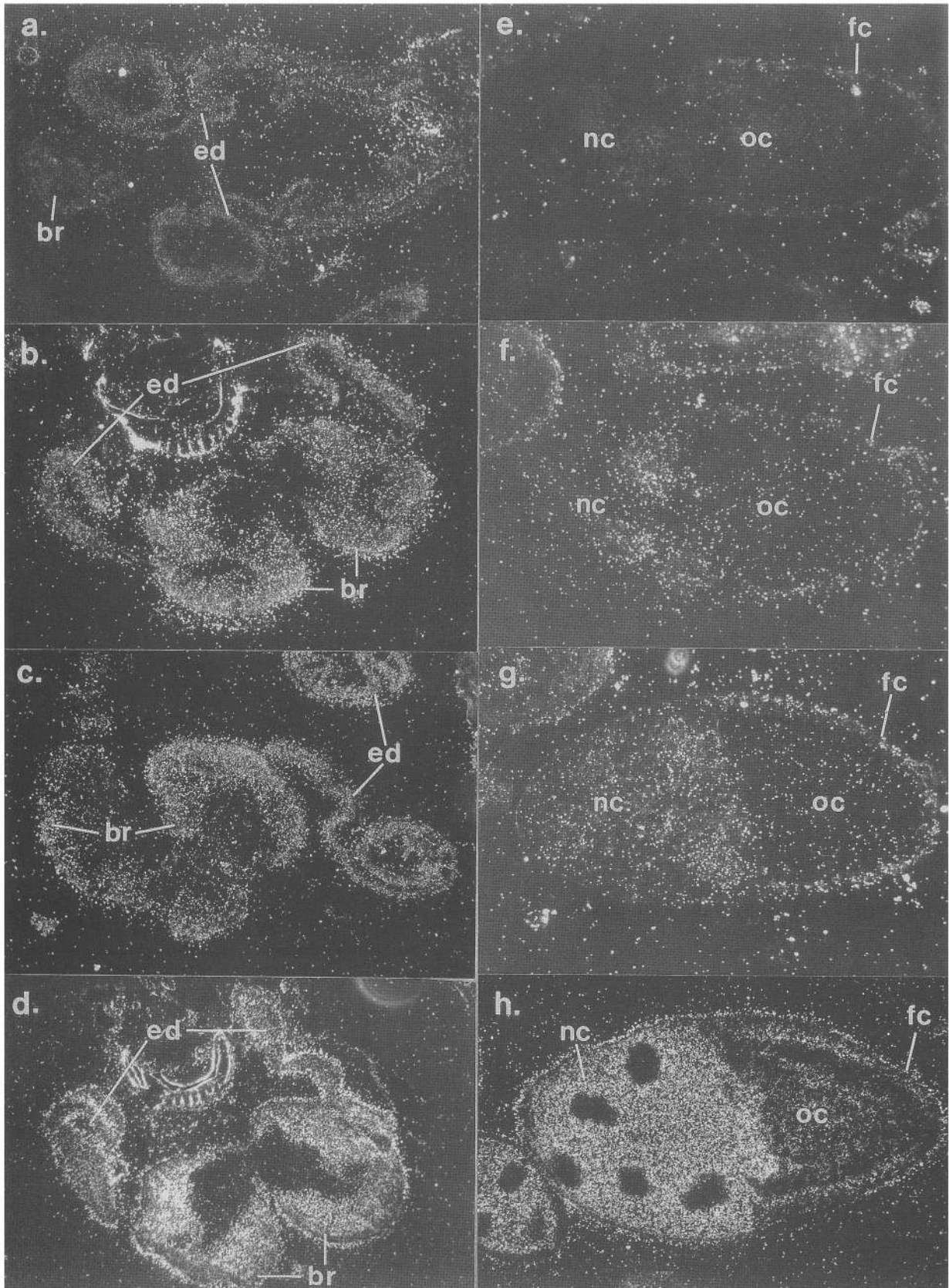
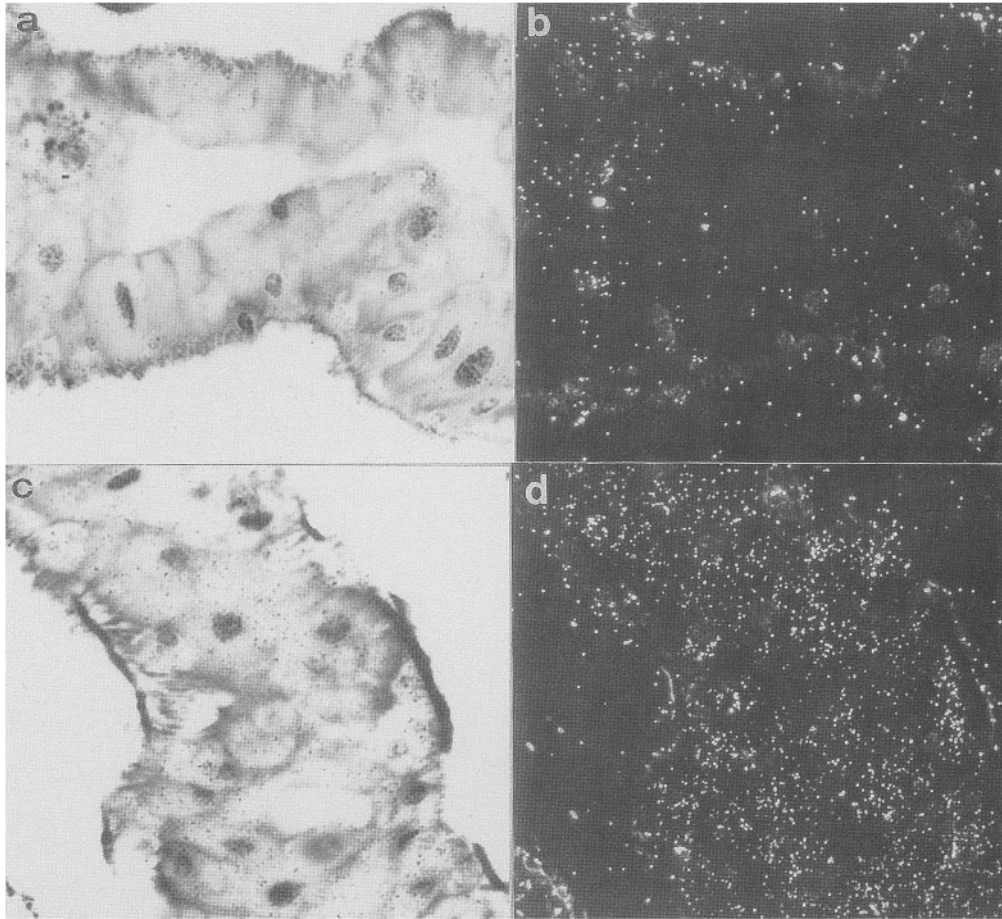**Figure 7.** (*See facing page for legend.*)

Figure 8. Expression of *arm* transcripts in the larval gut. (*a,c*) Bright-field photomicrographs of sectioned larval hindgut tissue hybridized with the 2.5-kb sense strand *arm* probe (*b*) or the 2.5-kb antisense strand *arm* probe (*d*) show that *arm* RNA can also be detected at low levels in nonsegmented, terminally differentiated polytene tissue. This argues that *arm* may be required in all cell types and is not only associated with early pattern formation processes.

varies only about twofold during development, and in situ hybridizations demonstrated that both forms accumulate at comparable levels in the same tissues. This similar pattern of expression between the E9 and E16 transcripts implies that neither the production nor the degradation of either type is a highly regulated process and that there may not be a preferential requirement for either form of RNA. Instead, it is possible that the production of two transcripts might have only subtle, but selectively advantageous, effects. For example, using two promoters might effectively double the frequency of transcriptional initation from a single copy gene and thus allow for increased RNA production during times of high requirement. It should be possible to test deleted versions of the gene by germ-line transformation to assess the requirements for the E9 and E16 transcripts independently.

Analysis of the arm protein

Both transcripts from the *arm* gene encode the same 91.1-kD polypeptide. The *arm* protein is not highly re-

lated to other proteins in current data bases but has several structural features that suggest roles for it in pattern formation. Genetic analyses have shown that *arm* is cell autonomous and imply that it is required for individual cells to perceive patterning information (Gergen and Wieschaus 1986; Wieschaus and Riggleman 1987). Because the *arm* protein does not contain a probable signal sequence, it should be found in the cells that produce it (in agreement with its genetic cell autonomy). Furthermore, the *arm* protein does not have an obvious membrane-spanning region, so it probably does not participate in intercellular communication by acting as a traditional transmembrane receptor or a molecular pore. Instead, we suggest that the *arm* protein may be involved in transmitting or interpreting patterning information on the inside of the cell.

Structural analysis of the *arm* protein shows that the central hydrophobic region contains twelve and one-half 42 amino acid repeats that are likely to have arisen through a series of duplications. The amino acids used at most positions in the repeats are highly conserved in terms of their general chemical and structural proper-

ties. Thus, it is likely that the general structure of the repeats and the central region as a whole are important for the biological function of the *arm* protein. Further examination of the conserved regions suggests that each repeat consists of a series of four short, structured, hydrophobic regions separated by four polar 'turn' regions. This arrangement is expected to prevent the formation of any continuous secondary structure in the repeated region.

In recent years, numerous proteins containing internal repeats have been characterized, and most have been assigned to a relatively small number of families. In cases where the repeated sequences have a known function, they are usually involved in physical interactions with other molecules. For example, internal repeats have been shown to bind to other proteins [epidermal growth factor (EGF)-like repeats; fibronectin], lipid surfaces (apolipoprotein repeats; calpactin repeats), or metal ions (zinc fingers; calcium-binding domains). Alternatively, the repeats of structural proteins such as collagen and intermediate filaments self-associate and have intermolecular structural functions. The *arm* repeat sequence is not highly related to any other known protein repeat sequences. However, an important feature of the *arm* repeats is that their length and spacing are as highly conserved as their general structure. In other proteins containing regularly spaced repeats, this arrangement allows the protein to interact with itself and/or other molecules to form a highly ordered three-dimensional structure (Speicher and Marchesi 1984; Koenig et al. 1988). It will be interesting to determine whether the *arm* repeats undergo such interactions and how this relates to the function of the protein.

Transcriptional analysis suggests that arm is required in all cells

Loss of *arm* activity produces specific regional pattern defects during embryogenesis that appear to result from the inappropriate determination of positional fates in the affected cells (Wieschaus and Riggleman 1987). Such effects on cell positional fates suggest that *arm* could be a specific patterning gene, but the transcriptional characteristics of *arm* differ from those of other *Drosophila* pattern formation genes in several respects. For example, both *arm* transcripts are very abundant and are found at all stages of development. In addition, *arm* transcripts are detectable in virtually every cell type present in embryos, third-instar larvae, and adult ovaries and are uniformly distributed within these tissues. Together, these features of its transcriptional accumulation argue that *arm* may be required in all cells throughout development rather than being required only for particular regions, tissues, and stages, as most patterning genes are.

This interpretation agrees with other features of the *arm* mutant phenotype that do not seem to be directly related to effects on cell positional fate. *arm* is known to have a broad developmental requirement, and loss of *arm* function appears to affect many cell types at all stages. For example, *arm* mutant embryos exhibit structural defects in the head and on the dorsal side and show defects in germ-band retraction (Gergen and Wieschaus 1986). In addition, low levels of *arm* activity cause pattern defects in the imaginal disc derivatives, and clones of cells that are mutant for strong *arm* alleles do not survive in any region of the imaginal discs (M. Williams et al., in prep.). Therefore, *arm* influences all cells in the disc, not only those associated with the pattern defects. Although some of these phenotypic effects might have been interpreted previously as secondary consequences of patterning defects occurring elsewhere, the presence of *arm* transcripts in all cells suggests that the defects are produced by a direct requirement for *arm* activity in these cells. Thus, *arm* may be needed continuously by all cells and may not be used exclusively for pattern formation processes.

Although *arm* transcripts appear to be expressed everywhere, some cell types and developmental stages have much higher levels of *arm* RNAs than others. This variation may reflect differential requirements for the *arm* product, so it is interesting to compare the developmental and physiological circumstances of different cell types with the relative levels of *arm* RNAs that they contain. Mid-embryonic cells, larval imaginal disc and brain cells, and ovarian stage 5–10 nurse cells all contain similarly high levels of *arm* transcripts, whereas the larval gut, salivary gland, and fat body cells have much lower levels. In general, the cell types with the highest levels of *arm* RNAs are those that are in a state of rapid growth or differentiation. The *arm* transcripts are remarkably abundant in these cells (equaling ~35% of total cytoskeletal actin transcripts during early to mid-embryogenesis), so it is possible that the *arm* gene product may actually have a structural role in growing or differentiating cells.

Regional defects associated with the arm mutant phenotype may reflect localized requirements

Mutations in the *arm* gene produce localized defects in both embryonic segments and imaginal discs (M. Williams et al., in prep.), yet its transcripts are distributed evenly in these tissues. There are three general mechanisms that might account for this phenomenon, but two of these seem less plausible. First, it is unlikely that *arm* product is produced uniformly and then transported to specific regions within the segment because *arm* is cell autonomous (Gergen and Wieschaus 1986; Wieschaus and Riggleman 1987) and does not contain a signal sequence. A second model, that functional *arm* activity is only produced from the evenly distributed transcripts in certain regions of the segment, is also not likely because *arm* appears to be required in all cells. Instead, we propose that *arm* product is distributed uniformly but is required for the function of a localized factor or process that specifically influences cell fate. Thus, loss of *arm* activity would cause the functional loss of a localized positional determinant, and this would produce regional pattern defects.

There are a variety of potential ways that *arm* might

assist localized patterning processes. In the developing segment, pattern formation involves organization over relatively long distances, and cellular communication may be necessary for the proper specification of posterior cell fates (Martinez-Arias et al. 1988). Because single posterior cells that lack *arm* activity produce anterior structures even when surrounded by normal wild-type posterior cells, it seems likely that *arm* is required for individual cells to perceive such intercellular patterning information (Wieschaus and Riggleman 1987). Sequence analysis suggests that *arm* is an intracellular protein and does not cross the plasma membrane; therefore, its function may be to transmit or interpret patterning information within individual cells. Alternatively, *arm* could be required for the production or stability of factors that are required for the patterning process. In support of this view, DiNardo et al. (1988) have found that the segmental pattern of *engrailed* protein expression decays prematurely in *arm* mutant embryos. However, this effect of *arm* may be due to the redetermination of cell fates in this region and may not reflect a primary function of *arm*.

Considering the similar mutant phenotypes of *arm* and other segment polarity genes, some of these genes might encode localized products that require *arm* function. The transcripts of two of these genes, *wingless* (Baker 1987) and *gooseberry* (Baumgartner et al. 1987; Côté et al. 1987), are found only in the general region of the segment that is most affected in *arm* mutant embryos. Hence, the product of either of these genes could be a positional determinant that requires *arm* function for its activity. In addition, we note that mutations in *wingless* or *arm* produce similar pattern defects in imaginal disc derivatives (M. Williams et al., in prep.), and *wingless* transcripts are localized to regions of the leg discs that are most affected by loss of *arm* activity (Baker 1988). Thus, interactions with *wingless* might account for the pattern defects associated with *arm* in both imaginal discs and embryonic segments. Further molecular and genetic analyses may allow us to determine the actual function of the *arm* gene product and the identity of the specific patterning molecules that require its activity.

Materials and methods

Fly stocks and methods

The four EMS alleles of *arm* (XK22, XM19, XP33, and YD35), the white allele used for germ-line transformation, and the FM7 balancer are described in Wieschaus et al. (1984). The 18.3, TD5 and 25B dysgenic alleles and their balancers were obtained from Zusman et al. (1985), and the wild-type strain Oregon R is described in Lindsley and Grell (1968). Embryos were collected for cuticle analysis on apple juice agar with yeast and prepared according to Wieschaus and Nüsslein-Volhard (1986). Embryos, larvae, pupae, and adults used for RNA and DNA preparations were raised by standard methods (Roberts 1986).

Molecular cloning and analysis of the arm region

All enzymes were purchased from Boehringer–Mannheim Bio-

chemicals (Indianapolis, Indiana), nucleotides were from Pharmacia (Piscataway, New Jersey), and most common chemicals were obtained from Sigma Chemical Company (St. Louis, Missouri). The *arm* region was cloned from the 18.3 dysgenic allele using the general technique outlined by Bingham et al. (1981). In situ hybridizations to polytene chromosomes (Gall and Pardue 1971) were performed using biotinylated probes and peroxidase localization, as described by Langer-Safer et al. (1982). A full-length P element probe, pBR-π25.1 (Bingham et al. 1982), was used to locate P elements in dysgenic alleles. Genomic DNA from *arm*[18.3]/In(1)dl 49, y sn bb adult flies was prepared according to Ish-Horowicz et al. (1979), partially digested with *Sau*3A, and separated on a 0.4% agarose gel. DNA ranging from 17 to 20 kb in size was isolated, ligated into *Bam*HI-digested λ EMBL4 (Frischauf et al. 1983), and packaged and infected into *Escherichia coli* Q359, as described by Maniatis et al. (1982). The 18.3 genomic library was screened with a nick-translated P element probe according to Maniatis et al. (1982), and 40 P-element-positive phage were isolated and hybridized to Oregon R polytene chromosomes. The four phage that hybridized to the *arm* region were mapped, and fragments were used as probes to isolate a 35-kb region from an Oregon R EMBL4 λ library that was made as described above. This DNA was subcloned into Bluescript (Stratagene, San Diego, California) and used to probe Southern blots of *Eco*RI- or *Hin*dIII-digested genomic DNA from the EMS and dysgenic alleles, as described by Maniatis et al. (1982), to identify DNA rearrangements. Poly(A)+ RNA was prepared by the method of O'Hare et al. (1983), electrophoresed on formaldehyde gels (Maniatis et al. 1982), transferred to Nytran membranes (Schleicher and Schuell, Keene, New Hampshire), and probed with the subclones to identify transcribed regions in the *arm* region.

Germ-line transformation methods

A 10.5-kb *Pst*I fragment containing the transcriptional unit interrupted by the dysgenic insertions (Fig. 2c) was isolated from a λ clone in the walk by sequential partial digestion and cloned into the *white*+ transformation vector W8 (R. Klemenz and U. Weber, unpubl.). The plasmid was purified and injected as described by Spradling and Rubin (1982) into *white* embryos using the transposition helper p-π25.7wc (R. Karess and G. Rubin, unpubl.). The surviving adults were mated to *white* flies, and their progeny were examined for *white*+ activity. Stocks were established from the transformants, and males from those lines with autosomal transposons were crossed with heterozygous *arm*[25B]wf/FM7 (a weak allele) and *arm*[YD35]w/FM7 (a strong allele) virgin females. The progeny of these crosses were then examined for surviving *white*+ males carrying the *arm* chromosome.

Structural analysis of the arm gene

Full-length cDNAs for *arm* and both flanking genes were isolated from an embryonic cDNA library constructed by L. Kauvar (Poole et al. 1985), as described for the genomic clones. The entire cDNAs were subcloned as *Eco*RI partial-digest fragments into Bluescript. The approximate intron–exon structure was determined by probing Southern blots of genomic mapping gels with the various cDNAs, comparing the restriction maps of corresponding genomic and cDNA regions, and by DNA sequencing. Directions of transcription were determined by probing Northern blots with strand-specific RNA probes by the method of Melton et al. (1984). The 5' ends of the *arm* transcripts were mapped by primer extension, as described by Ghosh et al. (1978), using 16-bp antisense synthetic primers

that are specific for either first exon (for E9, 5'-TCTCATG-CACTCCAAC-3'; for E16, 5'-CACCGCGTCCAAGGAA-3'). S1 analyses (Berk and Sharp 1977) were performed with ^{32}P-labeled antisense DNA fragments that were made by hybridizing the synthetic primers to a sense-strand genomic template and extending the primers by the method described below. The products of these reactions were then digested with *Sac*II (E9) or *Apa*I (E16) and gel purified to obtain transcript-specific labeled probes that extend ~100 (E9) or 60 (E16) bases beyond the 5' end of the respective cDNAs.

The cDNAs and corresponding genomic regions were sequenced on both strands by the chain-termination method of Sanger et al. (1977) using the Sequenase system (U.S. Biochemical Corporation, Cleveland, Ohio). The protein sequence data were analyzed using the Pustell sequence analysis system (International Biotechnologies, Inc., New Haven, Connecticut), DIAGON dot matrix program (Staden 1982), FASTP data base search program (Lipman and Pearson 1985), and the structural analysis method of Garnier et al. (1978).

Transcriptional analysis of the *arm* gene

Specific probes for the two *arm* transcripts were made by priming single-stranded M13mp11 DNA (Messing 1983) carrying the sense strand of the E9 or E16 cDNAs with the specific primers described above. In each case, the 7 ng of primer was annealed to 1.5 μg of M13 template in 15 μl of buffer containing 50 mM sodium chloride, 10 mM magnesium chloride, and 10 mM Tris (pH 8.0). The mixtures were heated to 65°C and cooled slowly to room temperature. Dithiothreitol and all four deoxyribonucleotide triphoshates were then added to final concentrations of 1 mM and 25 μM, respectively, and the mixtures were incubated with 2 units of Klenow fragment of DNA polymerase I at 37°C for 30 min. The volumes were then increased to 80 μl; 20 μl of 5× S1 buffer (1 M sodium chloride, 100 mM sodium acetate, 5 mM zinc chloride, and 25% glycerol) and 20 units of nuclease S1 were added, and the mixtures were incubated for 30 min at 37°C. The DNAs were then phenol extracted, chloroform extracted, and ethanol precipitated. The DNAs were resuspended in 10 μl water, 2 μl 5× restriction digest buffer [0.25 M Tris (pH 7.5), 0.25 M sodium chloride, and 50 mM magnesium chloride] and 20 units of *Eco*RI were added, and the mixtures were incubated for 2 hr at 37°C. They were electrophoresed on a 10% polyacrylamide gel, and the ~100-bp *Eco*RI/ blunt-ended fragments corresponding to the first exons of the two cDNA forms were isolated from the gel and cloned into *Hinc*II/*Eco*RI-digested Bluescript. The identities of these clones were then confirmed by DNA sequencing.

Northern blots were performed as described above using 5 μg poly(A)[+] RNA per lane. The blots were hybridized with ^{32}P-labeled antisense RNA probes (Melton et al. 1984) produced from either a 2.5-kb *Eco*RI fragment of the E9 cDNA (Fig. 2d), the E9 or E16 first exon-specific probes described above, or a 453-bp *Hae*III–*Hin*dIII fragment from the 3' region of the actin 5C gene (Fyrberg et al. 1983). Direct comparisons between the levels of actin RNAs and total *arm* transcripts were performed by hybridizing a Northern blot simultaneously with the same amount of radioactive antisense RNA for *arm* and actin (the probes were of similar length: 453 bp for the actin probe and 521 bp for the *arm* probe). The blot was exposed to film to locate the respective signals, and these regions were excised from the blot and analyzed by liquid scintillation spectroscopy. These values were then corrected for background and differences in length and specific activity between the two probes. The relative abundance of the E9 and E16 RNAs was determined in a similar fashion, but separate equilavent Northern blots were probed in

parallel because the two RNAs have the same electrophoretic mobility.

In situ hybridization to paraffin-embedded tissue sections was performed as described by Ingham et al. (1985). Ovaries were removed from well-fed adult females, and fixed in 4% paraformaldehyde/PBS on ice for 15 min, and then dehydrated and embedded, as described by Ingham et al. (1985). Whole third-instar larvae were gently torn in half with forceps, and the anterior halves were turned inside out; these were then fixed and embedded as described above. The sections were probed with ^{35}S-labeled RNAs from the E9 2.5-kb probe, the transcript-specific probes, or an *engrailed* probe (Poole et al. 1985).

References

Akam, M. 1987. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**: 1–22.

Ambrosio, L. and P. Schedl. 1984. Gene expression during *Drosophila melanogaster* oogenesis: Analysis by in situ hybridization to tissue sections. *Dev. Biol.* **105**: 80–92.

Anderson, K.V. and J.A. Lengyel. 1984. Histone gene expression during *Drosophila* development. In *Multiple levels of regulation in histone gene structure, organization, and regulation* (ed. G.S. Stein, S.L. Stein, and W.F. Marzluff), pp. 135–161. J. Wiley and Sons, New York.

Baker, N.E. 1987. Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila*: The spatial distribution of a transcript in embryos. *EMBO J.* **6**: 1765–1773.

———. 1988. Transcription of the segment polarity gene *wingless* in the imaginal discs of *Drosophila*, and the phenotype of a pupal-lethal wingless mutation. *Development* **102**: 489–497.

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.

Benoist, C. and P. Chambon. 1981. In vivo sequence requirements of the SV40 early promoter region. *Nature* **290**: 304–310.

Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner. 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5' end in different developmental stages. *Cell* **33**: 125–133.

Berk, A. and P.A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**: 721–732.

Bingham, P.M., M.G. Kidwell, and G.M. Rubin. 1982. The molecular basis of P-M hybrid dysgenesis: The role of the P element, a P-strain specific transposon family. *Cell* **29**: 995–1004.

Bingham, P.M., R. Levis, and G.M. Rubin. 1981. Cloning of DNA sequences form the *white* locus by a novel and general

method. *Cell* **25**: 693–704.

Birnsteil, M., M. Busslinger, and K. Strub. 1985. Transcription termination and 3′ processing: The end is in site *Cell* **41**: 349–359.

Bond, B.J. and N. Davidson. 1986. The *Drosophila* melanogaster actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple RNA species. *Mol. Cell. Biol.* **6**: 2080–2088.

Côté, S., A. Preiss, J. Haller, R. Schuh, A. Kienlin, E. Seifert, and H. Jäckle. 1987. The *gooseberry-zipper* region of *Drosophila*: Five genes encode different spatially restricted transcripts in the embryo. *EMBO J.* **6**: 2793-2801.

DiNardo, S., E. Sher, J. Heemskerk-Jongens, J.A. Kassis, and P.H. O'Farrell. 1988. Two-tiered regulation of spatially patterned *engrailed* gene expression during *Drosophila* embryogenesis. *Nature* **322**: 604–609.

Foe, V.E. and B.M. Alberts. 1983. Studies of nuclear and cytoplasmic behavior during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* **61**: 31–70.

Frischauf, A.M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**: 827–842.

Fyrberg, E.A., J.W. Mahaffey, B.J. Bond, and N. Davidson. 1983. Transcripts of the six *Drosophila* actin genes accumulate in a stage and tissue specific manner. *Cell* **33**: 115–123.

Gall, J.G. and M.L. Pardue. 1971. Nucleic acid hybridization in cytological preparations. *Methods Enzymol.* **21**: 470–480.

Garnier, J., D.J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**: 97–120.

Gergen, J.P. and E.F. Wieschaus. 1986. Localized requirements for gene activity in the segmentation of *armadillo, fused, giant,* and *unpaired* mutations in mosaic embryos. *Wilhelm Roux's Arch. Dev. Biol.* **195**: 49–62.

Ghosh, P.K., V.B. Reddy, J. Swincoe, P. Lebowitz, and S.M. Weissman. 1978. Heterogeneity and 5′ structures of the late RNAs of simian virus 40. *J. Mol. Biol.* **126**: 813–846.

Haynes, S.R., M.L. Rebbert, B.A. Mozer, F. Forquingnon, and I.B. Dawid. 1987. PEN repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix destabilizing protein. *Proc. Natl. Acad. Sci.* **84**: 1819–1823.

Ingham, P.W., K.R. Howard, and D. Ish-Horowicz. 1985. Transcription patterns of the Drosophila segmentation gene *hairy. Nature* **318**: 439–445.

Ish-Horowicz, D., S.M. Pinchin, P. Schedl, S. Artavanis-Tsakonas and M.-E. Mirault. 1979. Genetic and molecular analysis of the 87A7 and 87C1 heat inducible loci of *Drosophila melanogaster. Cell* **18**: 1351–1358.

King, R.C. 1970. *Ovarian development in* Drosophila melanogaster. Academic Press, New York.

Koenig, M., A.P. Monaco, and L.M. Kunkel. 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53**: 219–228.

Langer-Safer, P.R., M. Levine, and D.C. Ward. 1982. Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci.* **79**: 4381–4385.

Lindsley, D.L. and E.L. Grell. 1968. Genetic variations of *Drosophila* melanogaster. *Carnegie Inst. Washington Publ.* **627**.

Lipman, D. and W. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**: 1435–1441.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martinez-Arias, A., N. Baker, and P. Ingham. 1988. Role of the segment polarity genes in the definition and maintenance of cell states in the Drosophila embryo. *Development* **103**:157–170.

Melton, D.A., P.A. Kreig, M.R. Rebagliati, T.M. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**: 7035–7056.

Messing, J. 1983. New M13 cloning vectors. *Methods Enzymol.* **101C**: 20–77.

Nüsslein-Volhard, C. and E. Wieschaus. 1980. Mutations affecting segment number and polarity in *Drosophila. Nature* **287**: 795–801.

O'Hare, K., R. Levis, and G.M. Rubin. 1983. Transcription of the *white* locus in *Drosophila melanogaster. Proc. Natl. Acad. Sci.* **80**: 6917–6921.

Perrimon, N. and A.P. Mahowald. 1987. Multiple functions of segment polarity genes in Drosophila. *Dev. Biol.* **119**: 587–605.

Poole, S.J., L.M. Kauvar, B. Drees, and T. Kornberg. 1985. The engrailed locus of Drosophila: Structural analysis of a transcript. *Cell* **40**: 37-43.

Roberts, D.B. 1986. Basic *Drosophila* care and techniques. In Drosophila: *A practical approach* (ed. D.B. Roberts), pp. 1–38. IRL Press, Oxford, England.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.

Schibler, U., O. Hagenbuchle, R.K. Wellauer, and A.C. Pittet. 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene amy1[a] in the parotid gland and the liver. *Cell* **33**: 501–508.

Scott, M.P. and S.B. Carroll. 1987. The segmentation and homeotic gene network in early Drosophila development. *Cell* **51**: 689–698.

Speicher, D.W. and V.T. Marchesi. 1984. Erythrocyte spectrin is composed of many homologous triple-helical segments. *Nature* **311**: 177–180.

Spradling, A.C. and G.M. Rubin. 1982. Transposition of cloned P elements into Drosophila germ line chromosomes. *Science* **218**: 341–347.

Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acid Res.* **10**: 2951–2961.

Steinert, P.M., A.C. Stevens, and D.R. Roop. 1985. The molecular biology of intermediate filaments. *Cell* **42**: 411–419.

Steward, R., L. Ambrosio, and P. Schedl. 1985. Expression of the *dorsal* gene. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 223–228.

Vigoreaux, J.O. and S.L. Tobin. 1987. Stage-specific selection of alternative transcriptional initiation sites from the actin 5C gene of *Drosophila. Genes Dev.* **1**: 1161–1171.

Wieschaus, E. and E. Noell. 1986. Specificity of embryonic lethal mutations in Drosophila analyzed in germ line clones. *Wilhelm Roux's Arch Dev. Biol.* **195**: 63–73.

Wieschaus, E. and C. Nüsslein-Volhard. 1986. Looking at embryos. In Drosophila: *A practical approach* (ed. D.B. Roberts), pp. 199–228. IRL Press, Oxford, England.

Wieschaus, E. and R. Riggleman. 1987. Autonomous requirements for the segment polarity gene *armadillo* during Drosophila embryogenesis. *Cell* **49**: 177–184.

Wieschaus, E., C. Nüsslein-Volhard, and G. Jürgens. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster:* Zygotic loci on the X-chromosome and

the fourth chromosome. *Wilhelm Roux's Arch. Dev. Biol.* **193:** 296–307.

Wright, D.A. and P.A. Lawrence. 1981. Regeneration of the segment boundary in Oncopeltus. *Dev. Biol.* **85:** 317–327.

Zusman, S., D. Coulter, and J.P. Gergen. 1985. Lethal mutations induced in the proximal X-chromosome of *Drosophila melanogaster* using P-M hybrid dysgenesis. *Dros. Inf. Serv.* **61:** 217–218.

Note

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries.

Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene.

B Riggleman, E Wieschaus and P Schedl

References	This article cites 48 articles, 14 of which can be accessed free at: **http://genesdev.cshlp.org/content/3/1/96.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**