*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 167–171, January 1992
Developmental Biology

# Molecular characterization and silk gland expression of *Bombyx* engrailed and invected genes

(homeobox genes/compartments)

CHI-CHUNG HUI*, KENJI MATSUNO, KOHJI UENO, AND YOSHIAKI SUZUKI†

Department of Developmental Biology, National Institute for Basic Biology, Okazaki, 444 Japan

*Communicated by James D. Ebert, October 3, 1991*

ABSTRACT **Genetic analysis in *Drosophila* has shown that engrailed (*en*) plays an important role in segmentation and neurogenesis. A closely related gene, invected (*in*), is coexpressed with *en* in the posterior developmental compartments where *en* is known to specify cell state. We report here the isolation of two *en*-like cDNAs from the middle silk glands of *Bombyx mori* larvae. Sequence analysis revealed that they are the counterparts of *Drosophila en* and *in*. Four highly conserved domains, including the homeodomain, were identified in these En and In proteins from *Bombyx* and *Drosophila*. In addition, two *en*-specific and one *in*-specific domains could also be found. These structurally homologous genes might share a similar role in *Bombyx* development. They were found to be coexpressed in the middle silk gland but not in the posterior silk gland during the fourth molt/fifth intermolt period. We speculate that these *Bombyx en*-like genes might be involved in the compartmentalization of the silk gland.**

The *Drosophila* gene engrailed (*en*) was originally identified as a segmentation gene affecting cells in the posterior compartments (1–3). It was later found to be required also for the development of the central nervous system (4). Like many other developmental control genes in *Drosophila* (5), *en* is a homeobox-containing gene (6, 7). It encodes a nuclear protein that binds DNA specifically through its homeodomain (8) and can function as a transcription factor *in vitro* (9–11). A closely related gene, invected (*in*), was also identified in *Drosophila* that shares extensive homology with *en* and lies within 20 kilobases (kb) of *en* (12). Both *in* and *en* are expressed in the cells of the posterior compartments early in development and later in the central nervous system (12). The roles of *in*, if any, remain obscure.

Based on sequence similarity, *en* homologs have been isolated from many invertebrate and vertebrate species (13–21). Expression studies revealed that *en* probably plays a role in neurogenesis (14, 18, 21–28), as has been shown by genetic analysis in *Drosophila* (4). It might play roles both in compartmentalization of the developing neural tube and specification of particular neuronal populations. Recently, it was found that mice homozygous for a targeted mutation of one of the *en*-like genes, *En-2*, show defects in the development of the folia in the cerebellum providing further support for the neurogenetic function of *en* (29). During the process of segmentation, *en* is expressed in the posterior portion of each metamere in all arthropod species examined so far suggesting that the segmentation function of *Drosophila en* might be conserved in other arthropods (14, 30–32). In this respect, *en* expression has been used as a molecular marker to directly compare mechanisms of segmentation (32).

We have been analyzing the regulation of silk protein gene expression in *Bombyx mori* (for review, see ref. 33). Recently, we reported that the promoters of the silk protein genes contain clustered homeodomain binding sites (34, 35) and that several putative homeodomain-containing proteins that interact with these sites can be found in the silk gland extracts (36). As part of our efforts to characterize the homeobox genes that are possibly involved in the regulation of the silk protein genes, several homeobox-containing cDNAs have been isolated from silk gland cDNA libraries. In this report, we describe the molecular characterization of two *Bombyx en*-like cDNAs. Sequence analysis reveals that they are the counterparts of *Drosophila en* and *in*. We have named the corresponding genes *Bombyx* engrailed (*Bm en*) and *Bombyx* invected (*Bm in*), respectively. Throughout the fourth molt/fifth intermolt period, they are coexpressed in the middle silk gland but not in the posterior silk gland. A hypothesis that these *en*-like genes are involved in specifying compartments in the silk gland is discussed.

## MATERIALS AND METHODS

**Animals.** *B. mori* eggs, from a Japanese strain (Kin-Shu), a Chinese strain (Sho-Wa), and a hybrid between them (Kin-Shu × Sho-Wa) were purchased from Kanebo Silk (Kasugai City, Japan). Larvae were maintained aseptically at 25°C on an artificial diet and staged as described (37).

**Cloning and Sequencing of cDNAs for *Bombyx en* and *in*.** λ gt11 cDNA libraries were constructed from poly(A)$^+$ RNA of the middle silk gland of 2-day-old fifth-instar *B. mori* larvae (Kin-Shu × Sho-Wa) by standard procedures (38). Five *Bm en* and one *Bm in* clones were isolated by screening a random-primed cDNA library under conditions of low stringency at 30°C with an oligonucleotide probe, 5′-GAGGCG-CAGATCAAGATCTGTTCCAGAACAGGCGGGCC-AA-3′, that spans the putative recognition helix of the *en* homeodomain (5). The hybridization buffer consisted of 50% (vol/vol) formamide, 5× SSC (1× SSC is 0.15 M NaCl/15 mM sodium citrate), 5× Denhardt's solution (1× = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), 0.1% SDS, 5 mM sodium phosphate (pH 6.5), and denatured salmon testes DNA (250 μg/ml). By rescreening libraries with the 5′ portion of these isolated cDNA fragments, four additional *Bm en* and three additional *Bm in* clones were chosen for further characterization. The cDNA inserts were subcloned into pBluescript II KS(−) vector (Stratagene) and sequenced by the dideoxynucleotide method using the Sequenase protocol (United States Biochemical). Genomic fragments that contain the cDNA se-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, ON, Canada M5G 1X5.

†To whom reprint requests should be addressed at: Department of Developmental Biology, National Institute for Basic Biology, Myodaiji, Okazaki, 444 Japan.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M64335 and M64336).

quences corresponding to *Bm en* and *Bm in* were isolated from genomic λ and Cosmid libraries (C.-c.H. and K.U., unpublished) and sequenced with specific primers derived from the cDNA sequences.

**RNA Extraction and Blot-Hybridization (Northern) Analysis.** Total RNA was isolated by using an acid guanidinium thiocyanate/phenol/chloroform method (39) and poly(A)$^+$ RNA was enriched by oligo(dT)-cellulose chromatography. The RNA was electrophoresed on 1% agarose/1.1 M formaldehyde gels (38) and transferred onto nylon membranes (Biodyne Electronics, Santa Monica, CA; Pall). Blots were hybridized with probes specific for *Bm en* and *Bm in* at 45°C, in 50% formamide/5× SSC/5× Denhardt's solution/0.1% SDS/5 mM sodium phosphate (pH 6.5)/denatured salmon testes DNA (250 μg/ml). A control cDNA probe that hybridizes with an abundantly expressed transcript of unknown identity (C.-c.H. and Ping-Xian Xu, unpublished data) was also used to check the integrity of RNA.

## RESULTS AND DISCUSSION

**Cloning of cDNAs for *Bombyx en* and *in*.** Low-stringency hybridization with oligonucleotide probes that code for the putative recognition helix of the Antennapedia and *en* homeodomains (5) yielded several positive clones in a screen of a middle silk gland cDNA library from 2-day-old fifth-instar larvae. Among them, two types of *en*-positive clones were found. These cDNA fragments were used to rescreen cDNA libraries and several overlapping clones were characterized by nucleotide sequence analysis. These analyses revealed that they correspond to the *Bombyx* homologs of *Drosophila en* and *in* (*Bm en* and *Bm in*, respectively).

The nucleotide and amino acid sequences of a composite cDNA molecule of *Bm en* are shown in Fig. 1. The 3363-base-pair sequence reveals an open reading frame that yields a protein of 372 amino acids with a deduced molecular mass of 42 kDa. The methionine residue at nucleotide 379 is designated as the initiator because the N-terminal region of the predicted protein shows a strong similarity with that of the *Drosophila* En protein (see below). As shown in Fig. 2, the sequence of *Bm in* reveals an open reading frame that yields a protein of 476 amino acids with a deduced molecular mass of 54 kDa. We have chosen the methionine residue at nucleotide 251 to be the initiator because it is the first methionine residue after many termination codons and the sequence around it (AGAACGAAAATGGCC) is very similar to the sequence near the *Bm en* initiator (AacACGAgAATGGCC, where lowercase letters indicate the difference between the two sequences). No polyadenylylation signal or poly(A) tail can be found at the 3′ ends of these cDNA clones. The transcripts of cDNAs for *Bm en* and *Bm in* are 5.1 and 6.2 kb long, respectively, indicating that these cDNA clones are only partial (see Fig. 5*A*).

**Conserved Domains in the En and In Proteins.** As summarized in Fig. 3, sequence comparison unambiguously revealed a distinction between the two En proteins and the two In proteins from *Bombyx* and *Drosophila*. In addition to the homeodomain, three conserved domains (I, II, and III) can be found in these four insect En-like proteins. The two En proteins share two additional conserved domains, one located at the N terminus and the other in the central part of the proteins. Near the N-terminal region of the two In proteins, one In-specific domain can be recognized. Furthermore, an Arg-Ser dipeptide sequence just after region II is only found in the two In proteins (Fig. 4*A*). As will be discussed below, this might serve as a hallmark of In proteins.

```
   1 GCGCAGTGCCACGTCCGCGTCAAACGCAAAGTGACGTTTCAATTTGACAGCTCCGGCTGTCTG
  64 TCACGCGCCCGGCGTCTACACGACGAAAAAAAAAAAAGAAAACAAGTGTCGTCCGTCGTGTCGA
 127 GCGGCAAAGGATCCATGCTAATGAGCTGAGCCTCCGTATCGATTAATATTAATGACATTCGAA
 190 ACTCGCGTCAAACAGTCGACCGCGTCTCGAATTCGAATCGGTCGCGGACATTAGTGGAGGCAT
 253 TTAAATCACGCGCCACACGGAACCGGGGACCCGCATTGTGCTCTCCGTCCCAACGATGAACAA
 316 ACACCGTCACGGCTGACAGAGACGCGTCGCGGTGGCTCGCTGCGTGACAGGTGGAACACGAGA

 379 ATG GCC TTC GAG GAC CGC TGC AGC CCT AGC CAG GCC AAC AGT CCG GGA
   1 Met Ala Phe Glu Asp Arg Cys Ser Pro Ser Gln Ala Asn Ser Pro Gly

 427 CCG GTG ACG GGG CGA GTC CCG GCG CCG CAC GCC GAG ACC CTC GCA TAC
  17 Pro Val Thr Gly Arg Val Pro Ala Pro His Ala Glu Thr Leu Ala Tyr

 475 AGC CCG CAG AGC CAG TAC ACT TGC ACC ACA ATA GAA TCG AAG TAC GAA
  33 Ser Pro Gln Ser Gln Tyr Thr Cys Thr Thr Ile Glu Ser Lys Tyr Glu

 523 CGA GGT TCT CCG AAC ATG ACA ATT GTG AAG GTG CAG CCG GAC TCC CCG
  49 Arg Gly Ser Pro Asn Met Thr Ile Val Lys Val Gln Pro Asp Ser Pro

 571 CCT CCC AGC CCG GGA CGC GGT CAG AAC GAG ATG GAA TAC CAG GAC TAC
  65 Pro Pro Ser Pro Gly Arg Gly Gln Asn Glu Met Glu Tyr Gln Asp Tyr

 619 TAC CGC CCT GAA ACG CCC GAC GTA AAG CCT CAT TTT AGC CGG GAG GAG
  81 Tyr Arg Pro Glu Thr Pro Asp Val Lys Pro His Phe Ser Arg Glu Glu

 667 CAG AGG TTT GAA CTG GAC AGA TCG AGG GGG CAG CGG CTG CAA CCC ACC
  97 Gln Arg Phe Glu Leu Asp Arg Ser Arg Gly Gln Arg Leu Gln Pro Thr

 715 ACG CCG GTC GCC TTC TCC ATA AAC AAC ATC CTG CAC CCT GAG TTC GGC
 113 Thr Pro Val Ala Phe Ser Ile Asn Asn Ile Leu His Pro Glu Phe Gly

 763 TTG AAC GCC ATC AGG AAA ACG AGC AAA ATC GAA GGT CCC AAG CCC ATT
 129 Leu Asn Ala Ile Arg Lys Thr Ser Lys Ile Glu Gly Pro Lys Pro Ile

 811 GGA CCG AAC CAC AGT ATC CTC TAT AAA CCT TAC GAT TTA TCC AAA CCG
 145 Gly Pro Asn His Ser Ile Leu Tyr Lys Pro Tyr Asp Leu Ser Lys Pro

 859 GAC TTA TCG AAA TAC GGC TTT GAT TAT TTG AAG AGT AAG GAA ACG AGT
 161 Asp Leu Ser Lys Tyr Gly Phe Asp Tyr Leu Lys Ser Lys Glu Thr Ser

 907 GAT TGC AAC GCT TTG CCG CCT TTA GGA GGG TTG AGG GAG ACG GTA TCG
 177 Asp Cys Asn Ala Leu Pro Pro Leu Gly Gly Leu Arg Glu Thr Val Ser

 955 CAG ATC GGC GAA CGC TTG TCC AGA GAC AGG GAG CCT CCG AAG AGT CTG
 193 Gln Ile Gly Glu Arg Leu Ser Arg Asp Arg Glu Pro Pro Lys Ser Leu

1003 GAG CAG CAG AAG AGG CCC GAT TCC GCG AGT TCC ATC GTC TCG TCG ACG
 209 Glu Gln Gln Lys Arg Pro Asp Ser Ala Ser Ser Ile Val Ser Ser Thr

1051 TCG AGC GGC GCG GTT TCC ACC TGC GGA AGC TCG GAC GCG AGC TCC ATC
 225 Ser Ser Gly Ala Val Ser Thr Cys Gly Ser Ser Asp Ala Ser Ser Ile

1099 CAG TCC CAG AGC AAC CCT GGC CAG CTG TGG CCA GCC TGG GTA TAC TGC
 241 Gln Ser Gln Ser Asn Pro Gly Gln Leu Trp Pro Ala Trp Val Tyr Cys

1147 ACC AGA TAC AGC GAC CGA CCT AGT TCC GGT CCC AGA AGC AGA AGG GTC
 257 Thr Arg Tyr Ser Asp Arg Pro Ser Ser Gly Pro Arg Ser Arg Arg Val

1195 AAA AAG AAA GCA GCG CCC GAA GAG AAG AGA CCA AGG ACC GCT TTC AGC
 273 Lys Lys Lys Ala Ala Pro Glu Glu Lys Arg Pro Arg Thr Ala Phe Ser

1243 GGG GCA CAA CTC GCG AGA TTG AAG CAC GAG TTC GCC GAG AAC CGG TAT
 289 Gly Ala Gln Leu Ala Arg Leu Lys His Glu Phe Ala Glu Asn Arg Tyr

1291 CTG ACT GAG CGG CGG CGG CAG AGC CTC GCG GCG GAG CTG GGC CTC GCC
 305 Leu Thr Glu Arg Arg Arg Gln Ser Leu Ala Ala Glu Leu Gly Leu Ala

1339 GAG GCG CAG ATC AAG ATT TGG TTC CAG AAC AAA CGG GCC AAG ATC AAG
 321 Glu Ala Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys

1387 AAG GCG TCG GGC CAG AGG AAC CCG CTG GCG TTG CAG CTC ATG GCC CAG
 337 Lys Ala Ser Gly Gln Arg Asn Pro Leu Ala Leu Gln Leu Met Ala Gln

1435 GGT CTC TAC AAC CAC AGC ACC GTC ACG GAG AGC GAC GAC GAA GAG GAG
 353 Gly Leu Tyr Asn His Ser Thr Val Thr Glu Ser Asp Asp Glu Glu Glu

1483 ATC AAT GTT ACG TAA AGGAGATGGTGCGATTGTTTTGGTGACGTCACGATAACGGTTT
 369 Ile Asn Val Thr ---

1541 TCGTGAAATTTCAAATTTGGAACTTTCTCGTCTCTATTGTGTTTACTTTCTCGAATTTCAAAC
1604 GAAGGAATACGTTAGGACGCGTTACCTAACCGGTACACACGGGGACCGTAATGCAAGGGAATT
1667 GGATATGTTTTTCGGTGGCAAATTTTTGGCGGATATGTCATAGAGGTTTTGCCTGAATTACGT
1730 AAGTTTGTTTTCGGGTTTAGCTACTTCCAAATGTTGCCATATGAAAATACATAATTTTTCGGT
1793 AAATTTCAGACAAGACCGCAATGTTTAATATTATTATTACTTATGTAACGTAGCCGCATTGAG
1856 TCTTACTGATTAATTAAAAAAAAAACAATAGATTATAAATAAATAAATAAAACAAAAATGTTC
1919 CTAATAATGTAAATATATTTATTTGTATTTAAATAATGGCAATAGTCAAGTAATTGTTAAAAA
1982 AAACATATATTATAGTACTAGCAGTTTATAGATGTTATAGTAATTTTAGGAATACGGTATTAA
2045 CACTTGTTGTATTGGCAAAAAAATACGGACTCATTTTGTTTTGACACAAGTCAAAATTGACAG
2108 CTGTCAGTTGTGAGGTTTAAACTGGAGCCCACACTAATTACAAACTGGACAAAGTCATAGATA
2171 TAAAATAGAATCGTTTGGAATCACTCAAGTCTATGGAGTAGTTTCGTGCTGTCCCACGGAGAC
2234 CTCAAAAGAATTGTTCAAGATGGTCTCTTCTTTTTCTGTAATCGTCACCTCATCTTCCGCCAG
2297 ATCAGACTTTTTGTACTGCTGACTACATCCAAAAATAGTTTTTTAAAAATACGACTCTGACGT
2360 ATAAAGTCCGGTCCACGGCATTTTCTTTGGTGCTACCACCAGACCTTCGTCAAATGAGAAGAG
2423 TCTTCAGCAGTTGACAATGCTTCATCATATTCCTCTCAGTTTCCTCGGTAAAAGTACTCGAAA
2486 ATACAATCAAATAACTAAGGGTAGCAAAACGAATGAATAGTAACGAAAAGCTTAAATTATAAC
2549 AAAAATACATTAAATGAAGTAAACAAATAATGTTATTGGAAATTTAAAAAAAAAATCATCGAT
2612 ATTATATTTGAGCATATTTATAATAGTTTATTGGTTATAAGTTTACGCATTTTCGTTTACATT
2675 TCATAATTTATTTAATATTTCTTTTGGAATCTTACGAAAAAAACAACAGAACCGATTTTAATT
2738 AATTGAATCATTGCGCTCGAAGGTTCATACTGTCGTTGTAACGAAGAAAAAATAATGTTGGTG
2807 AATTACTGGAATAAAACGTTATTATAATTAATAAGTACTTTTTTAATTAAGTGCTGCGATATC
2864 ATAACCAAACAGATAAACTATTAGTTTAGCTCCTAAAATAATTTTAGTAAAAGTAGCAATGAA
2927 TGTGTTAGAAAGCGTCTTGCCCTAACCGGTAAAACGTAACGTTTTAAATCGATTTAAACTAAA
2990 TAAAAATAAATATTGTTTTATAACGAAAAAGCACGTTTTTCTTAATGCCGAACTAATCTGTTT
3053 TTTTAGTCTAGCGATCTCAACCAGGAAATTGTCGAAGCAATAAAAAAAAACATTGCAATTATT
3116 CCTCGATACGCGTGCTAATTTTTAAGTTAATCGGACGTTTTTTTTTGTGAAATCATATACTTT
3179 AAAGACTGATGGAGTTCGTGCTAATAAAAGCATGCTAAAAATGGTTTTTTTGCGATAATGTTC
3242 AAAGCGAAGGGACTTTATGACGTATCGTCGTTGTCAAACCAAAATCTGCTATGTGCAAAAACT
3305 ATATTTTGGAATAATGAGTAAAACCATTTTAGTATAAAATTTTATATCAAGGTGAATTC
```

FIG. 1. Nucleotide and deduced amino acid sequences of *Bombyx en*. The nucleotide sequence of a composite cDNA for *Bm en* is shown with the deduced amino acid sequence of the Bm En protein. The homeodomain is boxed and the three homologous domains (regions I, II, and III in Fig. 3) are marked with heavy underlines. The two En-specific domains (striped boxes in Fig. 3) are underlined and the positions of the introns are indicated by arrowheads.

Among the four conserved domains found in these four insect En-like proteins, the homeodomain and regions II and III have been shown (16) to be present in *Drosophila* En and In and in mouse En-1 and En-2 proteins. Region I, which is located in the N-terminal half of these proteins, is also conserved in the vertebrate En-like proteins (refs. 14 and 21;

```
     AGATCACGCGCACAACATCCCTTTCAAACAAAACAATTGTTTAATAAAAACAATTAACGCA
  62 TTCGATAGTCGATAAATTAGTTAATAGTTTTGTTTTTCGTCTAATCTAGGCGACCGTTTTTGT
 125 GATAAGCACGGTGTGTGTGTGTAAAGAATTGCATTGAAAAGTTCAAAGTGCTTTTTGTTGAGG
 188 ATACGGTTGCACGATGAAGTTCAAATGATAAGGTAAATAATAATGGCGTGACGGAGAACGAAA

 251 ATG GCC GCT GTA TCC GCC CAT ATG CAG GAC ATT AAG ATC CAA GAC CAG
   1 Met Ala Ala Val Ser Ala His Met Gln Asp Ile Lys Ile Gln Asp Gln

 299 AGC GAT GAC GAT CCA TAC TCT CCG AAC ACG AGA GAC ACG ACA AGT CCA
  17 Ser Asp Asp Asp Pro Tyr Ser Pro Asn Thr Arg Asp Thr Thr Ser Pro

 347 GAG TGC CAC GAC GAT GAG AAA TCG GAA GAC ATA AGC ATC CGT TCA TCC
  33 Glu Cys His Asp Asp Glu Lys Ser Glu Asp Ile Ser Ile Arg Ser Ser

 395 TCT TTC TCC ATC CAC AAC GTG CTT AGA AGG AGC GGG ACA ACA GCA GCC
  49 Ser Phe Ser Ile His Asn Val Leu Arg Arg Ser Gly Thr Thr Ala Ala

 443 CTG ACA ATG TCT TTT CGA CGG AAA AGC TCT TGG AGA ATC CCG AAT TTC
  65 Leu Thr Met Ser Phe Arg Arg Lys Ser Ser Trp Arg Ile Pro Asn Phe

 491 GAT GAC AGA AAC ACA GAG AGT GTA AGT CCC GTT GTT GAA GTG AAT GAA
  81 Asp Asp Arg Asn Thr Glu Ser Val Ser Pro Val Val Glu Val Asn Glu

 539 AGA GAA ATA AGC GTG GAC GAT GGT AAT TCT TGC TGT AGC GAC GAC ACC
  97 Arg Glu Ile Ser Val Asp Asp Gly Asn Ser Cys Cys Ser Asp Asp Thr

 587 GTG TTG TCA GTT GGA AAC GAG GCA CCC GTA TCC AAC TAC GAA GAG AAA
 113 Val Leu Ser Val Gly Asn Glu Ala Pro Val Ser Asn Tyr Glu Glu Lys

 635 GCC AGC CAG AAT ACC CAC CAA GAA CTG ACC TCC TTC AAA CAC ATA CAA
 129 Ala Ser Gln Asn Thr His Gln Glu Leu Thr Ser Phe Lys His Ile Gln

 683 ACA CAC TTG AGC GCC ATA TCG CAG CTG AGC CAA AAC ATG AAT GTG GCC
 145 Thr His Leu Ser Ala Ile Ser Gln Leu Ser Gln Asn Met Asn Val Ala

 731 CAA CCG CTG CTA TTA CGG CCG AGT CCA ATT AAC CCA AAC CCA ATA ATG
 161 Gln Pro Leu Leu Leu Arg Pro Ser Pro Ile Asn Pro Asn Pro Ile Met

 779 TTC CTA AAC CAA CCG CTT CTG TTC CAA AGT CCG ATC TTG AGC CAA GAC
 177 Phe Leu Asn Gln Pro Leu Leu Phe Gln Ser Pro Ile Leu Ser Gln Asp

 827 TTA AAA GGT ATG CCC AAC AGA CAA ACA GCC AAC GTG ATC AGT CCA ACG
 193 Leu Lys Gly Met Pro Asn Arg Gln Thr Ala Asn Val Ile Ser Pro Thr

 875 TTT GGC TTA AAT TTC GGT ATG AGA TTG AAG GCC AAT CAT GAA ACA CGA
 209 Phe Gly Leu Asn Phe Gly Met Arg Leu Lys Ala Asn His Glu Thr Arg

 923 ACG AGG TCT GAT GAG AAT CGG TAT TCG AAG CCG GAA GAA TCT AGA GAT
 225 Thr Arg Ser Asp Glu Asn Arg Tyr Ser Lys Pro Glu Glu Ser Arg Asp

 971 TAC ATC AAT CAG AAC TGC CTT AAG TTT AGC ATA GAT AAT ATT TTA AAA
 241 Tyr Ile Asn Gln Asn Cys Leu Lys Phe Ser Ile Asp Asn Ile Leu Lys

1019 GCG GAC TTC GGA AGG AGG ATC ACC GAT CCT TTG CAC AAA AGG AAA GTG
 257 Ala Asp Phe Gly Arg Arg Ile Thr Asp Pro Leu His Lys Arg Lys Val

1067 AAG ACG AGA TAC GAG GCT AAA CCT GCT CCA GCA AAA GAC ACT GCG GCT
 273 Lys Thr Arg Tyr Glu Ala Lys Pro Ala Pro Ala Lys Asp Thr Ala Ala

1115 TTT GCT CCG AAG CTG GAC GAA GCG AGG GTA CCT GAC ATC AAA ACA CCA
 289 Phe Ala Pro Lys Leu Asp Glu Ala Arg Val Pro Asp Ile Lys Thr Pro

1163 GAC AAA GCT GGA GCC ATC GAC CTT TCT AAA GAC GAT AGC GGA AGC AAT
 305 Asp Lys Ala Gly Ala Ile Asp Leu Ser Lys Asp Asp Ser Gly Ser Asn

1211 TCT GGA TCA ACC TCC GGT GCA ACT TCA GGC GAC AGT CCG ATG GTG TGG
 321 Ser Gly Ser Thr Ser Gly Ala Thr Ser Gly Asp Ser Pro Met Val Trp
                                                                ▼
1259 CCC GCG TGG GTG TAC TGT ACG AGG TAC AGC GAT CGA CCC AGT TCC GGA
 337 Pro Ala Trp Val Tyr Cys Thr Arg Tyr Ser Asp Arg Pro Ser Ser Gly
         ▼
1307 AGA AGT CCT CGC ACC AGA CGA CCG AAG AAG CCG CCC GGA GAC ACC GCC
 353 Arg Ser Pro Arg Thr Arg Arg Pro Lys Lys Pro Pro Gly Asp Thr Ala

1355 AGC AAT GAC GAG AAG AGA CCA AGG ACC GCA TTC TCC GGA CCA CAG CTC
 369 Ser Asn Asp Glu Lys Arg Pro Arg Thr Ala Phe Ser Gly Pro Gln Leu
                    ▼
1403 GCG AGG CTA AAG CAC GAG TTC GCG GAG AAC CGG TAT CTG ACA GAG CGG
 385 Ala Arg Leu Lys His Glu Phe Ala Glu Asn Arg Tyr Leu Thr Glu Arg

1451 CGG CGG CAG AGC CTC GCG GCG GAG CTG GGC CTC GCC GAG GCG CAG ATC
 401 Arg Arg Gln Ser Leu Ala Ala Glu Leu Gly Leu Ala Glu Ala Gln Ile

1499 AAG ATC TGG TTC CAG AAC AAA CGG GCC AAG ATC AAG AAG GCG TCG GGC
 417 Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ala Ser Gly

1547 CAG AGG AAC CCA CTG GCA CTG CAG CTC ATG GCC CAG GGC CTC TAC AAC
 433 Gln Arg Asn Pro Leu Ala Leu Gln Leu Met Ala Gln Gly Leu Tyr Asn

1595 CAC AGC ACC GTG CCG CTG ACA AAG GAA GAG GAG GAA TTA GAG ATG AAG
 449 His Ser Thr Val Pro Leu Thr Lys Glu Glu Glu Glu Leu Glu Met Lys

1643 GCG AGA GAA AGA GAG AGA GAG CTG AAG AAT AGA TGT TAA AACGGCTTTCA
 465 Ala Arg Glu Arg Glu Arg Glu Leu Lys Asn Arg Cys ---

1693 GTAAGAGTGGTAGTGTGTTCTTGGTAATACTCCTAAAACTTACTTACCTTACAACCCAAAACT
1756 TACTCTTTTACTGGTGGTAGGACCACTTGTGAGTCCGCGCGGATAGGTACCACCACCCTGCTT
1819 ATTTCTGCCGTAAAGCAGTAATGCGTTTCGGTTTGAAGGGTGGGGCAGTCGTTGTAACTATAC
1882 TGAGAATTTAGAACTTGTATCTCAAGGTGGGGCGCGTTTACGTTGTAGATGTCTATGGGCTCC
1945 AGTAACCACTTAACACCAGGTGGGCTGTGAGCTCGTCCACCCATCAAATCAATTTCTAAATTT
2008 CATTATAAGCGAAGTTATTTTCTAATGTTTGCCAAACCGAAAATAATTCTAATAACACGTTGA
2071 CTGTAGGTCACCGGTGCCCTACGCGACGCACTGAAGTAATTATAATAGCAGTCTTCATAACAG
2134 TCGCGTCATTTCCTAATTGTTTCTTTAAGCTGCAAGAAGCTCTTTGAAAGCTTTTCAGCACAG
2197 CTTAGCATAGCATGCTACGATTTTGTAGCCGTTTAATATTTTTTTTGTTGTTTTGGTTTTTTA
2260 TTTCTTTTATTGCTTGTAAGGGTGGACGAGCTCACGGCCCACTTGATGTTTAGTGGTTACCGG
2323 AGCCCATAGACATCTACAACGTAAATGCGGCTACCCACCTTGAGACACAAGTTGTAAGGTCTC
2386 ACTTTTAACAGTACAACTGCTGCAATTCAAACCGAAACGCATTACTGCTTCACGGCAGGCATA
2449 GGGTGGTGGTAGCTACCCGTGCGGACTCACAAGACATCCTACCACCAGTAATTTTAGATTGAA
2512 TTCAATTTTACAAAGGCTATAATATTGTTATCGTTTTCGAGAACTATGCGTGAACACTCTTAC
2575 TGAAAGCGCCATTGCGACCCATCAGATGTCGGTGGACAACACTATACCGCCCAGTATTGAATT
2638 ATTTTATTAATCAAATCGAAAATTTAACTTCTGCCACGCACTTTTCACGGTCGAAATAGAGCA
2701 TTGTCATGTAAAATAGTCTTATAAAAGCGCTCTGGTGAATTTAGATGAATGACACAAATATTC
2764 GTTTTCTTCAAAAACGCATAGATTGGCTGGATTTTTAGTAAATTATTTTAATGAAGTCATCAG
2827 CTTAAACATGCGTTATAGATTTATCAACACCTACCTCGTGCCGAATTC
```

FIG. 2. Nucleotide and deduced amino acid sequence of *Bombyx in*. The nucleotide sequence of the longest cDNA clone for *Bm in* is shown with the deduced amino acid sequence of the Bm In protein. An In-specific domain (solid box in Fig. 3) is underlined. Other symbols are as described in Fig. 1.

A. L. Joyner, personal communication). This region spans a conserved core of 12 amino acids with additional similarities in the flanking sequences between the two En proteins and between the two In proteins (Fig. 4*A*). In particular, region I of the *Bombyx* and *Drosophila* In proteins is highly conserved (15 out of 16 residues are identical). Region II spans a region of 17 amino acids and is the most conserved region (it is identical in all En-like proteins examined so far). The homeodomain of the two *Bombyx* proteins shows very high sequence similarity (59 out of 60 amino acids are identical) as compared with the two *Drosophila* proteins (52 out of 60 amino acids are identical). The two honeybee *en*-like genes, *E30* and *E60*, also have only one amino acid difference in their homeodomain (13). Apparently, *Drosophila en* and *in* homeodomains are more divergent than other En-like proteins present within the same species (13, 14, 16). Region III spans a region of 18 amino acids and is also highly conserved in various organisms. Region III is identical in the two *Bombyx* En-like proteins as well as the two honeybee *en*-like genes (13).

The N terminus is probably an important feature of the insect En proteins because it is highly conserved between the *Bombyx* and *Drosophila* En proteins (11 of 15 amino acids are identical; Fig. 4*B*) and is identical in the *Drosophila melanogaster* and *Drosophila virilis* En proteins (40). Though the two *Drosophila* En proteins are very similar, an intervening stretch of 31 amino acids can be found in the *D. virilis* En protein to demarcate this N-terminal region and the next conserved region in the *Drosophila* proteins. This conservation of the N-terminal region is apparently confined to these insect En proteins because no significant similarity in this region can be found in the two insect In proteins and other vertebrate En-like proteins (ref. 21; A. L. Joyner, personal communication). Another conserved region in the insect En proteins is near the center of the proteins (8 out of 12 amino acids are identical between *Bombyx* and *D. melanogaster*, and 9 out of 12 amino acids are identical between *Bombyx* and *D. virilis*; Fig. 4*B*). Besides these conserved regions, the *Bombyx* and *Drosophila* En proteins also show an indication of sequence similarity in a region between the central En-specific domain and region II. Though their sizes are very different, both regions possess a high serine content (16 out of 54 amino acids in *Bombyx* and 38 out of 112 amino acids in *Drosophila*). This serine-rich region of the *Drosophila* En protein has been suggested to be the site for posttranslational modification by a serine-specific protein kinase (41). Between the two insect In proteins, only an additional conserved region of 14 amino acids (8 out of 14 amino acids are identical) can be identified near the N terminus (Fig. 4*B*).

Despite extensive homologies between the two En proteins, the Bm En protein lacks some prominent features of the *Drosophila* En protein, like the polyglutamine and polyalanine stretches (6). It has been suggested that *en* might play a similar role in body segmentation in all insects (14, 30–32). Consistent with this hypothesis, *Bm en* and *Bm in* are expressed during the process of segmentation in the embryo (C.-c.H., unpublished data). If we assume that the insect En proteins perform a similar function in transcriptional regulation, the poly(amino acid) stretches that have been proposed as transcriptional regulatory domains in the *Drosophila* En protein might be dispensable in *Bombyx*. In this respect, it is worthwhile to mention that the two *Xenopus* En-2 proteins also lack these polyglutamine and polyalanine stretches (21). In contrast, the conserved regions reported here are likely to be important structural and/or functional domains in these En-like proteins.

**Phylogeny of *en* and *in*.** Southern blot hybridization revealed that *Bm en* and *Bm in* are the only *en*-like genes in the *Bombyx* genome (data not shown). Genomic clones harboring these cDNA sequences were isolated and partially charac-

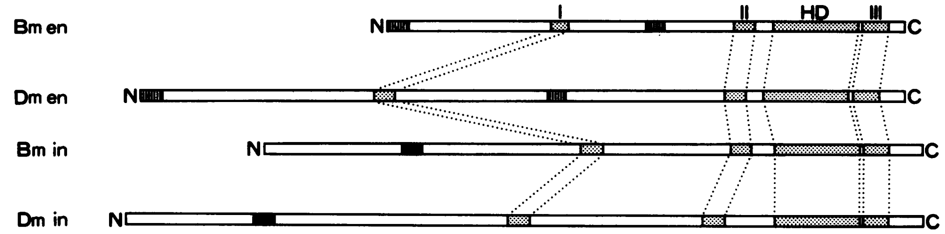

FIG. 3. Schematic representation of the *Bombyx* and *Drosophila* En-like proteins. The *B. mori* En (Bm en) and In (Bm in) and the *D. melanogaster* En (Dm en) and In (Dm in) proteins are portrayed with various conserved domains indicated. The positions of the four conserved domains found in all these En-like proteins, the homeodomain (HD), and regions I–III are represented by stippled boxes aligned with dotted lines. The striped boxes indicate the positions of the En-specific domains and the solid boxes indicate the positions of the In-specific domains (see text for more details).

terized by nucleotide sequence analysis using specific primers derived from the cDNAs. These analyses revealed the presence of two introns in the protein coding region of *Bm en* at exactly the same positions as in *Drosophila en* (Fig. 1). Similar to *Drosophila in*, an additional intron is found between the region II and the homeobox region of *Bm in* (Fig. 2). This indicates that *Bm in* also contains a 6-nucleotide miniexon as found in *Drosophila in* (12). This exon is not found in *Drosophila en* (40) and, apparently, is also absent in *Bm en* (Fig. 1). The Arg-Ser sequence encoded by this exon thus appears to be a hallmark of In proteins. In this respect, it is interesting to find that the only En-like protein found in the grasshopper also contains this sequence (14) and the mouse En-1 and En-2 proteins do not possess this sequence (16).

Whether the two *en*-like genes found in the honeybee represent the counterparts of *en* and *in* is still unclear. It is interesting to find that the two honeybee *en* genes do not possess an intron in the homeobox region while their *Bombyx* and *Drosophila* counterparts do (13). This intron is also absent in the mouse (16), zebrafish (19), and sea urchin (20). Though we do not know at present whether there are additional introns in the untranslated regions of *Bm en* and *Bm in*, the *Bombyx* and *Drosophila* genes are apparently similar in their exon–intron organization. These observations strongly suggest that *en* and *in* were present in the last ancestor common to both *Bombyx* and *Drosophila*. Based on the observations that only one *en*-like gene could be found in the grasshopper, Patel *et al.* (14) suggested that *en* and *in* may have arisen by duplication some time in the insect lineage after the last ancestor common to both *Drosophila* and the grasshopper. This hypothesis suggests that the two *en* genes found in vertebrates also arose as an independent duplication of an ancestral *en* gene early in the chordate lineage. The same conclusion has been suggested by Dolecki and Humphreys (20) after finding a single *en* gene in the sea urchin. Moths and flies are believed to have diverged about 240 million years ago, subsequent to the separation of their ancestors from those of the grasshopper and honeybee (42). Our data are consistent with the above hypothesis on the phylogeny of *en*.

```
A
                 I
Bm en  116   AFSINNILHPEFG      128
Dm en  174   ----S---SDR--      186
Bm in  247  LK---D---KAD--RR    262
Dm in  277  LK---D---KAD--SR    292

                 II
Bm en  250  WPAWVYCTRYSNRPSSG**PRSRRVKKKAA***********************PE
Dm en  422  -----------------**--Y--P-QPKDKT*********************ND
Bm in  336  -----------------RS--T--P--PPGDTAS*******************ND
Dm in  416  -----------------RS--A-KP--P-TSSSAAGGGGGGVEKGEAADGGGV--

Bm en  EKRPRTAFSGAQLARLKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKAS
Dm en  ---------SE------R--N-----------Q-SS----N-----------------ST
Bm in  ----------P-------------------------------------------------
Dm in  D---------T---------N-------K---Q-SG----N--------------L--S-

                III
Bm en  GQRNPLALQLMAQGLYNHSTVTESDDEEEINVT             372
Dm en  -SK---------------T--PLTKE---LEMRMNGEIP       553
Bm in  ---------------------PLTKE---LEMKARERERELKNRC 476
Dm in  -TK-----------------IPLTRE---LQELQEAASARAAKEPC 576

B

Bm en    1  MAFEDRCSPSQANAP  15    Bm in  106  SCCSDDTVLSVGNE  119
Dm en    1  --L------QS-P--  15    Dm in   94  ----ENS-----Q-  107
Dv en    1  --L------QS-P--  15

Bm en  185  LGGLRETVSQIG  196
Dm en  296  --S-CKA-----  307
Dv en  333  --S-CK------  344
```

FIG. 4. Homologous domains between the *Bombyx* and *Drosophila* En-like proteins. (*A*) Homologous domains in En-like proteins. The homeodomain is boxed and the three homologous domains are marked. Identical amino acids are indicated by a dash and an asterisk indicates a gap in the sequence. Other symbols are the same as in Fig. 3. (*B*) En- and In-specific domains. Dv en represents the sequence of *D. virilis* En.

**Coexpression of *Bombyx en* and *in* in the Middle Silk Gland.** Northern blot hybridization analysis revealed that two transcripts are encoded by *Bm en* and by *Bm in* in the middle silk gland of Kin-Shu × Sho-Wa larvae (Fig. 5 *B* and *C*). By ribonuclease protection and Northern blot hybridization analyses using RNA samples taken from the Kin-Shu and the Sho-Wa strains, we found that the cDNA sequences reported here are derived from the Kin-Shu strain (data not shown). Fig. 5*A* shows that a single *Bm en* transcript of 5.1 kb and a single *Bm in* transcript of 6.2 kb were detected in the Kin-Shu middle silk gland RNA sample (lanes 2 and 4). Both of them are shorter than their Sho-Wa counterparts, which are 5.5 kb (*Bm en*) and 7.4 kb (*Bm in*) long. Though the molecular basis for this length polymorphism is unknown, it might be due to a strain-specific variation in the long untranslated regions.

In Fig. 5*B*, we investigated the developmental profile of the *Bm en* and *Bm in* transcripts in the posterior and the middle silk gland during the transition from the fourth molt to the fifth intermolt. Both transcripts were barely detectable in the middle silk gland during the molting stage (15 h after the fourth apolysis; lane 5) and the fourth ecdysis (lane 6). Their levels increase gradually during the fifth intermolt (Fig. 5*B*, lanes 7 and 8, and also Fig. 5*C*) and peak between 48 h and 72 h after ecdysis (Fig. 5*C*). A trace amount of the *Bm en* and *Bm in* transcripts could also be detected in the middle silk gland during the fourth intermolt (Fig. 5*C*, lane 1). However, we could not detect any of these transcripts in the posterior silk gland at any stage examined so far (Fig. 5*B*, lanes 1–4).

The expression of *Bm en* and *Bm in* in the middle but not posterior silk gland is intriguing because *Drosophila en* is known to be expressed in the posterior developmental compartments derived from the ectoderm and to specify cell state (ref. 43 and references therein). The silk gland is an ectodermal derivative (44) and forms three morphologically distinct regions, the anterior, middle, and posterior silk glands (see ref. 33). The middle silk gland specifically expresses a number of glue protein genes, such as the sericin-1 gene, and the posterior silk gland specifically expresses the silk fiber genes, such as the fibroin gene. *Bm en* and *Bm in* might

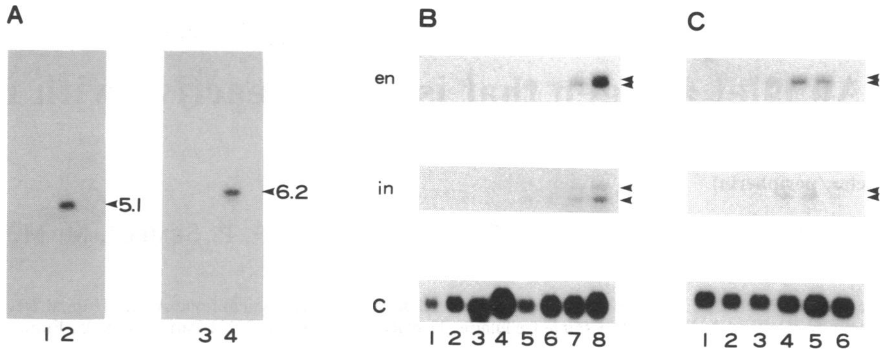

FIG. 5. Coexpression of *Bombyx en* and *in* in the middle silk gland during the fifth larval instar. (*A*) A Northern blot of poly(A)$^+$ RNA (5 $\mu$g) isolated from the posterior (lanes 1 and 3) and middle silk gland (lanes 2 and 4) of 2-day-old fifth-instar larvae (Kin-Shu strain) was hybridized with an *en*-specific probe (lanes 1 and 2) and an *in*-specific probe (lanes 3 and 4). (*B*) Northern blot analysis of poly(A)$^+$ RNA (5 $\mu$g) isolated from the posterior silk gland (lanes 1–4) and the middle silk gland (lanes 5–8) of Kin-Shu × Sho-Wa larvae. Lanes: 1 and 5, 15 h after the fourth apolysis; 2 and 6, the fourth ecdysis; 3 and 7, 24 h after the fourth ecdysis; 4 and 8; 48 h after the fourth ecdysis. The blot was hybridized with an *en*-specific probe (en), an *in*-specific probe (in), and a control probe (C). (*C*) Northern blot analysis of poly(A)$^+$ RNA (5 $\mu$g) isolated from the middle silk gland of Kin-Shu × Sho-Wa larvae. Lanes: 1, 72 h after the third ecdysis; 2, the fourth ecdysis; 3–6, 24, 48, 72 and 144 h, respectively, after the fourth ecdysis. Blots were hybridized as described in *B*.

similarly be involved in specifying compartments in the silk gland. A possible role for them might be the transcriptional regulation of genes specifically expressed in the silk gland. In this respect, the silk protein genes that possess homeodomain binding sites in their promoters are candidate genes (34–36). Further studies of *Bm en* and *Bm in* should provide information about development of the silk gland and about body segmentation in this intermediate germ-band insect.

We thank Dr. Alexandra Joyner for sharing unpublished data, members of our laboratory for helpful discussion, and Mrs. E. Suzuki, M. Sasaki, and M. Ohkubo for technical assistance. C.-c.H. and K.M. were supported by postdoctoral fellowships from the Japan Society for the Promotion of Science. This research was partly aided by Grant-in-Aid for Research of Priority Areas from the Ministry of Education, Culture and Science of Japan.

1. Lawrence, P. A. & Morata, G. (1976) *Dev. Biol.* **50,** 321–337.
2. Kornberg, T. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1085–1089.
3. Lawrence, P. A. & Struhl, G. (1982) *EMBO J.* **1,** 827–833.
4. Lawrence, P. A. & Johnston, P. (1984) *EMBO J.* **3,** 2839–2844.
5. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989) *Biochim. Biophys. Acta* **989,** 25–48.
6. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985) *Cell* **40,** 37–43.
7. Fjose, A., McGinnis, W. J. & Gehring, W. J. (1985) *Nature (London)* **313,** 284–289.
8. Desplan, C., Theis, J. & O'Farrell, P. H. (1988) *Cell* **54,** 1081–1090.
9. Jaynes, J. B. & O'Farrell, P. H. (1988) *Nature (London)* **336,** 744–749.
10. Han, K., Levine, M. S. & Manley, J. L. (1989) *Cell* **56,** 573–583.
11. Ohkuma, Y., Horikoshi, M., Roeder, R. G. & Desplan, C. (1990) *Cell* **61,** 475–484.
12. Coleman, K. G., Poole, S. J., Weir, M. P., Soeller, W. C. & Kornberg, T. (1987) *Genes Dev.* **1,** 19–28.
13. Walldorf, U., Fleig, R. & Gehring, W. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9971–9975.
14. Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B. & Goodman, C. S. (1989) *Cell* **58,** 955–968.
15. Weisblat, D. A., Price, D. J. & Wedeen, C. J. (1988) *Development Suppl.* **104,** 161–168.
16. Joyner, A. L. & Martin, G. R. (1987) *Genes Dev.* **1,** 29–38.
17. Logan, C., Willard, H. F., Rommens, J. M. & Joyner, A. L. (1989) *Genomics* **4,** 206–209.
18. Gardner, C. A., Darnell, D. K., Poole, S. J., Ordahl, C. P. & Barald, K. F. (1988) *J. Neurosci. Res.* **21,** 426–437.
19. Fjose, A., Eiken, H. G., Njolstad, P. R., Molven, A. & Hordvik, I. (1988) *FEBS Lett.* **231,** 355–360.
20. Dolecki, G. J. & Humphreys, T. (1988) *Gene* **64,** 21–31.
21. Hemmati-Brivanlou, A., de la Torre, J. R., Holt, C. & Harland, R. M. (1991) *Development* **111,** 715–724.
22. Davis, C., Nobel-Topham, S. E., Rossant, J. & Joyner, A. L. (1988) *Genes Dev.* **2,** 361–371.
23. Davis, C. & Joyner, A. L. (1988) *Genes Dev.* **2,** 1736–1744.
24. Davis, C. A., Holmyard, D. P., Millen, K. J. & Joyner, A. L. (1991) *Development* **111,** 287–298.
25. Davidson, D., Graham, E., Sime, C. & Hill, R. (1988) *Development* **104,** 305–316.
26. Hemmati-Brivanlou, A. & Harland, R. M. (1989) *Development* **106,** 611–617.
27. Hatta, K., Schilling, T. F., BreMiller, R. A. & Kimmel, C. B. (1990) *Science* **250,** 802–805.
28. Martinez, S. & Alvarado-Mallart, R.-M. (1990) *Dev. Biol.* **139,** 432–436.
29. Joyner, A. L., Herrup, K., Auerbach, B. A., Davis, C. A. & Rossant, J. (1991) *Science* **251,** 1239–1243.
30. Campell, G. L. & Caveney, S. (1989) *Development* **106,** 727–737.
31. Fleig, R. (1990) *Rouxs Arch. Dev. Biol.* **198,** 467–473.
32. Patel, N. H., Kornberg, T. B. & Goodman, C. S. (1989) *Development* **107,** 201–212.
33. Suzuki, Y., Takiya, S., Suzuki, T., Hui, C.-c., Matsuno, K., Fukuta, M., Nagata, T. & Ueno, K. (1990) in *Molecular Insect Science*, eds. Hagedorn, H. H., Hildebrand, J. G., Kidwell, M. G. & Law, J. H. (Plenum, New York), pp. 83–89.
34. Hui, C.-c. & Suzuki, Y. (1990) *Dev. Growth Differ.* **32,** 263–273.
35. Hui, C.-c., Suzuki, Y., Kikuchi, Y. & Mizuno, S. (1990) *J. Mol. Biol.* **213,** 395–398.
36. Hui, C.-c., Matsuno, K. & Suzuki, Y. (1990) *J. Mol. Biol.* **213,** 651–670.
37. Maekawa, H. & Suzuki, Y. (1980) *Dev. Biol.* **78,** 394–406.
38. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
39. Chomczynski, P. & Sacchi, N. (1987) *Anal. Chem.* **162,** 156–159.
40. Kassis, K. A., Poole, S. J., Wright, D. K. & O'Farrell, P. H. O. (1986) *EMBO J.* **5,** 3583–3589.
41. Gay, N. J., Poole, S. J. & Kornberg, T. B. (1988) *Nucleic Acids Res.* **16,** 6637–6647.
42. Martynova, O. A. (1961) *Annu. Rev. Entomol.* **6,** 285–294.
43. Hama, C., Ali, Z. & Kornberg, T. B. (1990) *Genes Dev.* **4,** 1079–1093.
44. Nunome, J. (1937) *Bull. Appl. Zool. Jpn.* **9,** 68–92.