Differential and Inefficient Splicing of a Broadly Expressed *Drosophila erect wing* Transcript Results in Tissue-Specific Enrichment of the Vital EWG Protein Isoform

SANDHYA P. KOUSHIKA,† MATTHIAS SOLLER, SUSAN M. DESIMONE, DOUGLAS M. DAUB, AND KALPANA WHITE*

Department of Biology and Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02454

Received 8 January 1999/Returned for modification 22 February 1999/Accepted 8 March 1999

In this report, we document an unusual mode of tissue-enriched gene expression that is primarily mediated by alternative and inefficient splicing. We have analyzed posttranscriptional regulation of the *Drosophila erect wing* gene, which provides a vital neuronal function and is essential for the formation of certain muscles. Its predominant protein product, the 116-kDa EWG protein, a putative transcriptional regulator, can provide all known erect wing-associated functions. Moreover, consistent with its function, the 116-kDa protein is highly enriched in neurons and is also observed transiently in migrating myoblasts. In contrast to the protein distribution, we observed that erect wing transcripts are present in comparable levels in neuron-enriched heads and neuron-poor bodies of adult *Drosophila*. Our analyses show that erect wing transcript consists of 10 exons and is alternatively spliced and that a subset of introns are inefficiently spliced. We also show that the 116-kDa EWG protein-encoding splice isoform is head enriched. In contrast, bodies have lower levels of transcripts that can encode the 116-kDa protein and greater amounts of unprocessed erect wing RNA. Thus, the enrichment of the 116-kDa protein in heads is ensured by tissue-specific alternative and inefficient splicing and not by transcriptional regulation. Furthermore, this regulation is biologically important, as an increased level of the 116-kDa protein outside the nervous system is lethal.

Most eukaryotic primary RNA transcripts undergo posttranscriptional processing requiring splicing of introns. The best-appreciated regulatory outcome of posttranscriptional processing is alternatively spliced transcripts that differ in the coding exons or have distinct 3′ or 5′ untranslated ends (reviewed in reference 32). A second consequence of posttranscriptional regulation is the modulation of amounts of specific transcripts dependent on differential splicing efficiencies of different splice sites. It is generally thought that differences in cell-type-specific splicing machineries result in cell type- or tissue-specific alternative splicing (5, 19). In addition, efficiency of splicing could play a major role in gene regulation as primary transcripts that are not completely processed are generally not transported to the cytoplasm and are unlikely to code functional proteins (6, 21, 23). We decided to investigate the role of alternative and inefficient splicing in the regulation of the *Drosophila erect wing* (*ewg*) gene, as previous studies indicated a complex transcript profile, intron-containing cDNAs, as well as poly(A)+ transcripts with retained introns (9).

The *Drosophila ewg* gene provides a function that is vital in the nervous system and essential to the development of certain muscles (16). EWG protein contains an unusual DNA binding domain that is homologous to sea urchin P3A2 protein (4, 10), zebrafish Nrf (3), and mammalian transcription factors NRF-1 and initiation binding receptor (13, 18, 31). Our previous studies suggested that *ewg* primary transcript may be alternatively spliced, since the *ewg* gene has several introns and its Northern pattern shows multiple transcripts that are tissue and developmental stage modulated (15). However, at the protein level, only one major polypeptide, a 116-kDa, 733-amino-acid-long polypeptide encoded by the SC3 cDNA open reading frame (ORF), was observed in immunoblot analysis, although many other cross-reacting bands were also observed (9, 10). The translation start site of the SC3 ORF is an unconventional CTG codon, suggesting that translational regulation of *ewg* may be an important aspect of *ewg* regulation (10). Transgenes expressing the 116-kDa EWG protein provide compelling evidence that the 116-kDa protein is the major functional protein, as expression of 116-kDa protein in the neurons rescues lethality and general expression rescues both lethal and muscle phenotypes associated with *ewg* alleles (8, 10). An antibody generated against the 116-kDa EWG protein selectively labels all neurons in the embryonic and larval stages and certain migrating myoblasts in early pupae (8–10), suggesting a distinct tissue-specific expression of the protein and possibly transcript.

We investigated the splicing patterns of *ewg* RNA to address if *ewg* transcripts are indeed alternatively or inefficiently spliced and if the pattern of splicing shows tissue-specific differences. In this paper, we report the results on *ewg* splicing, using reverse transcription (RT)-PCR in head and body RNAs as representative of neuron-enriched and neuron-poor tissue, respectively. Our results show the following. (i) *ewg* is more widely transcribed than previously recognized, and total *ewg* RNA levels in heads and bodies are comparable. (ii) A subset of *ewg* introns are efficiently spliced, but another subset are inefficiently spliced and retained in poly(A)+ RNA. (iii) *ewg* RNA in bodies has a greater representation of unprocessed RNAs, and RNAs that include two exons that are not part of the SC3 ORF. One of these new exons is not included in *ewg* transcripts present in heads. (iv) SC3 ORF RNA is enriched in adult heads but low in the bodies. (v) Modest expression of the SC3-encoded ORF in the body can be lethal. Thus, *ewg*, which is widely transcribed, is primarily regulated by posttranscriptional mechanisms.
MATERIALS AND METHODS

Fly stocks and genetic crosses. Drosophila melanogaster flies were raised on standard media and at 25°C. The Canton-S strain was used as the wild type. EWGHS and EWGNHS are two white+ marked transgenes that encode the 116-kDa EWG protein isoform under the control of heat shock hsp 70 promoter and the neuron-specific elav promoter (8). ewg2 is a lethal, protein-null allele of the ewg gene (10). Df(1)cin-arth uncovers several loci inclusive of cin, ewg, and y (14). Dp243 is a free duplication derived from Dp1157 which is y− and also carries a P element marked with rosy (35).

Crosses to check rescue of ewg deletion by transgene rescue consisted of crossing females of genotype Df(1)cin-arth y w+/+; EWGNHS/TM6,Tb were crossed to yw/Y; yw/Y; EWGNHS/TM6,Tb. Both rescues were carried out on a 25°C standard media and at 25°C. The Canton-S strain was used as the wild type.

To assess impact of overall increased expression of 116-kDa protein, females of the genotype ewgHSy w/yw; +/+: EWGNHS/TM6,Tb were crossed to yw/yw; +/+: EWGNHS+ males. To assess EWG protein levels expressed by the EWGNHS genes, females of the genotype ewgHSy w/yw; +/+: EWGNHS/TM6,Tb were crossed to yw/yw; +/+: EWGNHS+/+ males.

RESULTS

Flies that carry transgenes were identified by inspection of eye color.

Immunoblot analysis. Drosophila protein extracts were prepared and resolved as previously described (30). Drosophila embryos were 14 to 18 h old. Bacterial extracts were prepared by pelleting the bacteria and dissolving them in 2× sample buffer. Proteins were resolved by SDS-polyacrylamide gel electrophoresis at 12.5% and 8% SDS for the bacterial extract and Drosophila extracts respectively. Anti-EWG antibody (10) was used 1:5,000 for immunoblot analysis of bacterial extracts and 1:1,500 for Drosophila extracts. A peroxidase-conjugated goat anti-rabbit secondary antibody (Amersham) was used at a dilution of 1:2,000, and blots were developed by chemiluminescence (LumiGLO; Kirkegaard & Perry).
products. Table 3 summarizes the PCR products detected by using exon-specific primers.

(i) Introns 2, 4, and 5. Analysis of RT-PCR products using exon-specific primers revealed that introns 2, 4, and 5 are excised efficiently, as in each case a single band representing the spliced product was observed for head or body RNA. Moreover, for each primer set the band densities in both head and body lanes were comparable (Fig. 2, lanes 3, 4, 19, 20, 12, and 13). The efficient splicing of these introns, which are all small (Table 4), is to be expected since small introns are known to be spliced efficiently in Drosophila (29). However, the observation that the PCR bands for both body and head RNAs were comparable was unexpected, as it implied that nonneural tissues expressed ewg RNA to a greater extent than previously thought (see below and Fig. 3). In these RNA samples, levels of a control ribosomal protein transcript, rp49, are similar in the two tissues (Fig. 2, lanes 17 and 18).

(ii) Intron 1. Intron 1 is more efficiently spliced in body than in head RNA (Fig. 2, lanes 1 and 2). This inefficient splicing in heads perhaps explains the SC3 cDNA, which contains part of intron 1 (Fig. 1A).

(iii) Introns 3a, 3b, and 3c. Introns 3a is inefficiently spliced in both heads and bodies, as both spliced (150-bp lower band) and unspliced (449-bp upper band) products are observed.

| Splice event | Forward primer (sequence, 5' to 3') | Return primer (sequence, 5' to 3') | Mg$^{2+}$ (mM) |
|-------------|-------------------------------------|----------------------------------|---------------|
| rp49        | Frp49 (AAGCTGTCGCACAAATGGCG)         | Rrp49 (TACCTCGTTCCTTTGTAGACGCA) | 2.5           |
| 1           | 1F1 (GAGCAGGACAGCTGAAAAACTA)         | 1R1 (TTCCAGTGACACTCCTTTGGC)     | 1.5           |
| 2           | 2F1 (GAGCTCTTTATCCGCTTAGTCT)         | 2R1 (TCTTGCTGAGACTGATGTCC)      | 2.5           |
| 3a          | 3aF1 (ATCTCCTACTACACACAGG)           | 3aR1 (TGATATTGGCCAGGAGATGTT)    | 1             |
| 3c          | 3cF1 (CTCAAGTCGCTGACAGCAAAT)         | 3cR1 (TGAACCTGGCGAGATGTACCAT)   | 1             |
| 3b          | 3bF1 (ATCTCCTACTACACACAGG)           | 3bR1 (TGAACCTGGCGAGATGTACCAT)   | 1             |
| 4           | 4F1 (ATGGTACATCAGCAGGTTC)            | 4R1 (CCATCTCGTGTCTGCGTGAAT)     | 1             |
| 5           | 5F1 (AAGCGGTGCTACATCTACCC)           | 5R1 (TGAGATCATGCTTCTCCGAA)      | 1             |
| 3a          | 3aF1 (ATCTCCTACTACACACAGG)           | 3aR1 (TGATATTGGCCAGGAGATGTT)    | 1             |
| 3c          | 3cF1 (CTCAAGTCGCTGACAGCAAAT)         | 3cR1 (TGAACCTGGCGAGATGTACCAT)   | 1             |
| 4           | 4F1 (ATGGTACATCAGCAGGTTC)            | 4R1 (CCATCTCGTGTCTGCGTGAAT)     | 1             |
| 5           | 5F1 (AAGCGGTGCTACATCTACCC)           | 5R1 (TGAGATCATGCTTCTCCGAA)      | 1             |
| 6           | 6F1 (ATATCCCGTTTCGGTGAGCAAT)         | 6R1 (CGGAGAATATGGCTCCATAGC)     | 1             |
| 38          | 38R (AGTTTTCTCTCTCGGACAGGCG)         | RV (TTTTCCCGTGGATCTCTTGT)       | 1             |

TABLE 2. Summary of splicing of ewg introns in wild-type adult heads and bodies

| Intron (PCR primer pair) | Splicing events | Abundance |
|--------------------------|----------------|-----------|
|                          | Figure, lanes | PCR product(s) (size [bp]) | Head | Body |
| 1 (1F/1R)                | 2, 1 and 2    | 1 sp (476) | ++   | ++   |
| 2 (2F/2R)                | 2, 3 and 4    | 2 sp (177) | ++   | ++   |
| 3a (3aF/3aR)             | 2, 5 and 6    | 3a spl (150) | ++++ | ++   |
| 3b (3aF/3cR)             | 2, 7 and 8    | No ExD (169) | ++   | ++   |
|                          |               | ExE, no Ex (243) | +    | ++   |
|                          |               | ExD (637) | +++++ | +    |
|                          |               | ExD, ExE (711) | ND   | ++   |
| 3c (3cF/3cR)             | 2, 9 and 10   | 3c spl (154) | +1/2 | ++   |
|                          |               | ExE (228) | +++++ | +    |
|                          |               | ExE, In3c-1 (773) | ND   | ++   |
|                          |               | No ExE, In3c (961) | ND   | ++   |
| 4 (4F/4R)                | 2, 19 and 20  | 4 spl (139) | +++++ | ++++ |
| 5 (5F/5R)                | 2, 12 and 13  | 5 spl (155) | +++++ | ++++ |
| 6 (6F/6R)                | 2, 14 and 15  | 6 spl (278) | +++++ | ++++ |
|                          |               | Ex1 (316) | ND    | +    |
| 1 ret (In1F/In1R)        | 4D, 1 and 2   | ExB, In1 (345) | ++++ | +1/2 |
| 2 ret (2F/2R)            | 4, 3 and 4    | ExB, In2 (200) | ND   | ND   |
| 3a ret (3aF/3aR)         | 4D, 6 and 7   | ExC, In3a (277) | ++++ | ++++ |
| 3c ret (In6F/In6R)       | 4D, 8 and 9   | ExD, In6c (105) | ++++ | ++++ |
| 4 ret (In4F/4R)          | 4D, 10 and 11 | ExG, In4 (147) | ND   | ND   |
| 5 ret (5F/In5R)          | 4D, 12 and 14 | ExG, In5 (195) | ND   | ND   |
| 6 ret (In6F/6R)          | 4D, 15 and 16 | ExI, In6 (799) | ++++ | ++++ |

*Encapsulation of data from Fig. 2 and 4D. Abundance was assessed as ++++, +++, ++, +, +1/2 or ND (not detected) for each head and body pair by visual inspection; the strongest band of each primer pair was arbitrarily assigned a value of ++++. In some instances, bands designated +1/2 might not be discernible in the figure. In, intron; Ex, exon; ret, retained; spl, spliced.*
TABLE 3. Sizes of *ewg* PCR products detected by primers flanking more than one intron

| Primer pair | Figure, lane(s) | Splicing event (product size [bp]) | Abundance |
|-------------|----------------|-----------------------------------|-----------|
|             |                |                                   | Head      | Body     |
| 1F/3aR      | 4C, 3 and 14   | ExA–D (1,712)                     | ++        | ++       |
|             |                | ExA–D, In3a (2,011)               | +         | +        |
| 1F/3cR      | 4C, 11 and 12  | ExA–C,D,F (2,193)                 | +         | ++       |
|             |                | ExA–C,F, no D (1,730)             | +         | ++       |
|             |                | ExA–C,E,F, no D (1,804)           | + 1/2     | ++       |
|             |                | ExA–F (2,267)                     | ND        |
| 1F/4R       | 4C, 9 and 10   | ExA–D,F,G (2,310)                 | ++        | +        |
|             |                | ExA–C,F,G, no D (1,847)           | ++        | +        |
|             |                | ExA–C,E–G, no D (1,921)           | + 1/2     | ++       |
|             |                | ExA–G (2,384)                     | ND        |
| 2F/3aR      | 4B, 1 and 2    | ExB–D (453)                       | ++        | ++       |
|             |                | ExB–D, In3a (752)                 | +         | +        |
| 2F/3cR      | 4B, 3 and 4    | ExB–D,F (933)                     | ++        | ++       |
|             | 4C, 6 and 7    | ExB–F (1,007)                     | ND        |
|             |                | ExB,C,F, no D (471)               | +         | +        |
|             |                | ExB,C,E,F, no D (545)             | + 1/2     | ++       |
| 2F/RV       | 4C, 2 and 3    | ExB–C,F–H, no D, In6 (1,056)      | ++        | ++       |
|             |                | ExB,C,E–H, no D, In6 (1,130)      | + 1/2     | ++       |
|             |                | ExB–H, In6 (1,593)                | ND        |
|             |                | ExB–D,F–H, In6 (1,519)            | + 1/2     |
| 3aF/4R      | 4B, 5 and 6    | ExC,D,F,G (747)                   | ++        | ++       |
|             |                | ExC,F,G, no D (285)               | ++ 1/2    |
|             |                | ExC,E–G, no D (359)               | +         |
|             |                | ExC–G (821)                       | ND        |
| 3aF/38R     | 4A, 7          | ExC–F–H,I (451)                   | ND        |
|             |                | ExC–E–H,I (525)                   | ND        |
| 3aF/6R      | 4A, 3 and 4    | ExC,D,F–H,J (1,058)               | ++        | ++       |
|             |                | ExC,D,F–J (1,096)                 | ND        |
|             |                | ExC–F–H,I, no D (598)             | +         |
|             |                | ExC–F–J, no D (634)               | ND        |
|             |                | ExC–F–J, no D (708)               | +         |
| 3aF/RV      | 4A, 1 and 2    | ExC–F–H, In6 (1,215)              | +         |
|             | 4C, 4 and 5    | ExC–H, In6 (1,289)                | ND        |
|             |                | ExC–F–H, no D, In6 (753)          | ++        |
|             |                | ExC–F–H, no D, In6 (827)          | +         |
|             |                | ExC–H–E, no D, In6 (827)          | +         |
| 3cF/4R      | 4B, 7 and 8    | ExD–E (270)                       | ND        |
|             |                | ExD–E (344)                       | ND        |
| 3cF/5R      | 4B, 10 and 11  | ExD–E–H (334)                     | +        |
|             |                | ExD–E–H (408)                     | ND        |
| 3cF/38R     | 3A, 2, 4, 7, and 9 | ExD–E–H, In3c (1,141)           | +        |
|             | 4A, 6          | ExD–E–H (436)                     | ND        |
|             |                | ExD–E–H, In3c (510)               | ND        |
|             |                | ExD–E–H, In3c (1,055)             | ND        |
| 3cF/RV      | 4A, 9 and 10   | ExD–F–H, In6 (738)                | ND        |
|             |                | ExD–E–H, In6 (812)                | +         |
|             |                | ExD–E–H, In3c–1,6 (1,357)         | ND        |
|             |                | ExD–F–H, In3c,6 (1,545)           | +         |
| 4F/5R       | 4B, 12 and 13  | ExF–H (202)                       | ++        |
|             | 3A, 1, 3, 6, and 8 | ExF–H (449)                     | ++        |
| 4F/6R       | 4B, 14 and 15  | ExF–H (473)                       | ND        |
| 5F/6R       | 4B, 16 and 17  | ExG,H (401)                       | ++        |
|             |                | ExG–J (439)                       | ND        |

a Summary of RT-PCR data from Fig. 3A and 4A to C. Transcripts with SC3-like splicing are underlined. Abundance was assessed as ++, ++, +++, or ND (not detected) for each head and body pair by visual inspection; the strongest band of each primer pair was arbitrarily assigned a value of ++. In some instances, bands designated + 1/2 might not be discernible in the figure. In, intron; Ex, exon.

b Data not shown.
FIG. 2. Characterization and comparison of ewg splicing in wild-type adult tissues. All RT-PCR assays were carried out with DNase I-treated total RNA isolated from 2-day-old heads (H) or bodies (B). The italicized letters below each pair of lanes represent the specific splice events as outlined in Fig. 1B, e.g., primers 3aF and 3aR for intron 3a splicing. These data are summarized in Table 2, which also lists the lengths of PCR products. rp49 transcripts were used as a control. Molecular size markers (GIBCO-BRL) are shown in lanes 11 and 16. Note that splicing of introns 3a, 3c, and 6 in heads is mostly in the mode of the Sc3 cDNA.

pair 3aF/3cR amplified a band of comparable density in a position expected for transcripts that exclude exon D (169-bp lowest band) in both RNA samples (Fig. 2, lanes 7 and 8; Table 2). Using a bridge primer that hybridizes to both exon C and F, we confirmed that the exclusion of exon D occurred at equal levels in heads and bodies (data not shown). Further, a band representing transcripts where both introns 3a and 3c are spliced (637 bp) is highly enriched in heads. Again two additional bands were present almost exclusively in the body lane due to presence of a new exon.

(iv) Intron 6. Using primer set 6F/6R and head RNA, only one band (278 bp) expected from splicing of intron 6 was seen. However, in the body RNA lane, two relatively faint bands of equal intensities were observed; the lower band represents the splicing of intron 6, while a slightly larger and unexpected band represents a second new exon (Fig. 2, lanes 14 and 15; see also below). The low level of spliced product in the body lane represents a second new exon (Fig. 2, lanes 14 and 15; see also below). The low level of spliced product in the body lane represents a second new exon.

In summary, the results indicate the following. (i) The spliced products resulting from the excision of introns 2, 4, and 5 are present at similar levels in head and body, implying that ewg RNA is expressed outside the nervous system, likely in many tissues. Further, the splicing of these introns is unlikely to be regulated in neurons, as no significant differences are detected between neuron-enriched heads and neuron-poor bodies. (ii) The excision of introns 3c and 6 takes place at a higher efficiency in heads, which results in higher levels of mRNAs that encode the 116-kDa EWG protein (SC3 ORF) in adult heads than in bodies (Fig. 1A). This suggests that the splicing of introns 3c and 6 is likely to be regulated in neurons. (iii) Intron 3a and 3c are retained in a fraction of polyadenylated ewg RNAs, demonstrating that these introns are not spliced efficiently. (iv) ewg RNA undergoes alternative splicing in both heads and bodies by extending exon D. (v) Levels of intron 3b splicing are similar in heads and bodies. (vi) Body-enriched novel PCR bands were detected in the region of introns 3c and 6. That body tissue is representative of neuron-poor splicing events was supported by the identical splicing profile of ewg in abdomen RNA, which is more neuron poor than that of adult bodies, which contain the thoracic and abdominal ganglia (data not shown).

Characterization of new ewg exons E and I. The novel bands were isolated and directly sequenced on both strands with the primers that had been used for their amplification to determine if they resulted from additional exons in the ewg gene. The sequence of the 230- to 250-bp product detected in the 3aF/3cR (Fig. 2, lane 7) and 3F/3cR (lane 9) PCRs revealed the presence of a 74-bp exon in intron 3c. This new exon, E in Fig. 1A, codes for 24 amino acids and alters the translational frame. Further RT-PCR analysis of exon E revealed high enrichment in female abdomens (data not shown).

Sequencing of the upper band amplified from body RNA with primers 6F and 6R revealed the presence of a 38-bp exon (Fig. 1B, exon I in Fig. 1A, present within intron 6. Exon I is exclusive to ewg transcripts in bodies, encodes 12 amino acids, and also alters the translational frame (Fig. 2, lane 14; see also Fig. 4B, lanes 15 and 17).

The 5‘ and 3‘ splice sites for the new exons matched the splice site consensus (Table 4) (27). Moreover, the flanking introns have sequences that match candidate branch point sequences of appropriate distances from the relevant 5‘ splice site.

### Table 4. Catalog of 5‘ and 3‘ splice sites and branch point sequences of ewg introns

| Site                  | 5‘ splice site | 3‘ splice site | Intron size |
|-----------------------|---------------|---------------|-------------|
| Long-intron consensus | MAGgtragta    | ttttgyyyyyogcRT | 81–5,392 bp |
| Short-intron consensus| AGgtragtw     | tttyyyyyogcRT  | 51–80 bp    |
| 3F                    | CAgtggcg      | gtttcttcacgaT  | ~4.5 kb     |
| 3c                    | AATgtaagta    | aactatcatcacaC (8/16) | 299 bp   |
| 3c                    | GATgtagta (4/10) | acttatcacaC (8/16) | 807 bp   |
| 3b                    | AATgtagta   | actatatcacaC | 1,568 bp    |
| 3b                    | ATgtagta     | gtttagtgcacgC | 1,306 bp    |
| 3b/c-2                | CATgtaattac | actatatcacaC | 188 bp      |
| 3c-1                  | GATgtagta    | gtttagtgcacgC | 545 bp      |
| 4                    | ACCgtagta    | aacttttcacgC | 62 bp       |
| 5                    | AAGgtattta   | ctegacctttttacG (7/16) | 80 bp    |
| 6                    | CAGgttagata (5/10) | ttattccagaaG | 1,722 bp |
| 6a                   | CAGgttagata | ttattttaacgcA | 1,620 bp   |
| 6b                   | TCTgtaagaaga | ttattttaacgcA | 64 bp      |

*Consensus sites as defined by Mount et al. (26); the numbers in parentheses indicate the number of nucleotides that diverge from consensus out of the total number of nucleotides assessed for weak splice sites. Criteria defined by Mount et al. (26) were used to determine highly divergent splice sites, i.e., splice sites that have at least 40% of the nucleotides that occur at a frequency less than 20% or nearly 50% of the nucleotides that occur at a frequency less than 30%. Upper- and lowercase letters represent exon and intron sequences, respectively.
FIG. 3. Abundance of *ewg* transcripts is independent of polyadenylation and is equal in heads (H) and bodies (B). (A) RT-PCR using random-primed or gene-specific-primed cDNAs. The RT reaction shown in lanes 1 to 4 was primed with a primer in exon H, an exon common to all *ewg* transcripts. The RT reaction shown in lanes 6 to 9 was primed with random hexamers. The italicized letters below each pair of lanes represent the specific splice events assayed as outlined in Fig. 1B. The uppermost bands in lanes 2, 4, 7, and 9 show *ewg* transcripts containing intron 3c. The 408-bp band in lanes 2 and 7 contains exon E. Molecular size markers (GIBCO-BRL) are shown in lane 5. (B and C) Cycle titration of PCRs using primers 4F and 5R to amplify parts of *ewg* transcripts common to all *ewg* transcripts. cDNAs were synthesized with a gene-specific primer in exon H. Aliquots were removed from the PCR beginning at cycle 18 and continuing until cycle 28. Quantitation of bands reveals that PCR is in the linear range (data not shown). Note that the intensities of bands in both heads and bodies are similar at all cycles.

*ewg RNA is abundant in heads and bodies.* Comparison of levels of efficiently spliced introns 2, 4, and 5 (Fig. 2A, lanes 3, 4, 12, 13, 19, and 20) suggests that *ewg* RNA is present in heads and bodies at comparable levels. To verify that *ewg* RNA was indeed present at high levels in bodies, either random hexamers or a gene-specific primer in exon H were used for RT, allowing the amplification of all splice isoforms of *ewg* regardless of their state of polyadenylation in subsequent PCR. For both reactions, primers 4F and 5R revealed very similar signals with head and body RNA (Fig. 3A, lanes 1, 3, 6, and 8). Primers 3cF and 5R revealed that both tissues contain *ewg* RNAs that either retain or excise intron 3c (lanes 2, 4, 7, and 9). Exon E-containing transcripts were detected mostly in bodies (lanes 2 and 7). Also, some intron 3c-I retention in body RNA is observed (lanes 2 and 7). Thus, *ewg* RNAs appear to be efficiently polyadenylated, as the *ewg* splicing profiles are similar for cDNAs primed with oligo(dT), gene-specific, and random hexamer primers.

To verify that PCR amplification is in the linear range, the accumulation of the spliced product was assessed every two cycles from cDNAs primed with a gene-specific primer in exon H. Comparable signals were obtained in head and body lanes throughout the linear range of amplification using primers 4F and 5R (Fig. 3B and C; Fig. 1). Thus, *ewg* transcript is expressed in bodies and heads at similar levels.

**Alternative splicing of *ewg* exons.** The RT-PCR studies suggested that both head and body RNAs have populations of alternatively spliced transcripts that include or exclude exon D and that body-specific transcripts are enriched in transcripts that include exon I. Further, the splicing of intron 6 in bodies appear to be independent of each other since *ewg* transcripts containing exon I were found with or without exon D (436 bp in Fig. 4A, lane 6; 451 bp in lane 7; summarized in Table 3).

To further support these data and to test whether no other exons are alternatively spliced, PCR was done with primer pairs spanning several exons. No further alternatively spliced *ewg* transcripts were detected (Fig. 4B and C). Thus, alternative splicing in *ewg* transcripts is restricted to introns 3a, 3c, and 6.

*ewg* is inefficiently spliced. To determine if introns other than 3a and 3c were present in poly(A)⁺ *ewg* RNA, PCR was done with intron-specific primers. Introns 1, 3a, 3c, and 6 (Fig. 4D) are present in poly(A)⁺ *ewg* transcripts of heads and bodies. From these introns, only intron 1 is differentially retained in body RNA compared to in head RNA. All of these introns are larger than 81 bp and are classified as large *Drosophila* introns (26). Among these introns, the 5′ and 3′ splice sites of 3c and the 3′ splice site of 3a diverge significantly from the *Drosophila* consensus (Table 4) (26). These introns were not detected in poly(A)⁺ *ewg* transcripts, confirming that they were efficiently spliced (Fig. 4D, lanes 3, 4, 10, 11, 12, and 14; Table 3). The overall levels of unspliced *ewg* transcripts were assessed in assays using primer 3cF and return primer RV in intron 6. This analysis indicated the presence of greater amounts of unprocessed *ewg* transcripts containing both intron 3c and intron 6 in bodies than in heads (Fig. 4A, lanes 9 and 10; data summarized in Table 3).

The splicing situation in the region of intron 6 is complex, as evidenced by two spliced transcripts that exclude or include exon I and RNAs that retain intron 6, and possibly transcripts that terminate in intron 6, most of which show differential distribution in heads and bodies. First, intron 6 is spliced more efficiently in head RNA than in body RNA, where overall splicing appears to be significantly reduced (Fig. 4D, lanes 15 and 16). Second, about half of the spliced body transcripts show inclusion of exon I (Fig. 2, lanes 14 and 15; Fig. 4B, lanes 14 to 17). Finally, RNAs that retain proximal intron 6 sequence are more prevalent in body RNA, as seen by use of the return primer RV, 5′ to the polyadenylation sites in intron 6 (Fig. 4A, lanes 1, 2, 9, and 10; Fig. 4C, lanes 2 to 5). In contrast, when the return primer in exon J is used, many fewer products are amplified in bodies than in heads (Fig. 4A, 1 to 4; Fig. 4B, lanes 14 to 17). This result suggests that in bodies, exon J-containing RNA is underrepresented compared to RNA containing the 3′ region of the intron 6. Thus, some RNAs in the body may be terminated before exon J, using the putative cleavage/polyadenylation sites in intron 6 (nucleotides 5876 and 6742, AATAAA). The presence of such transcripts was previously suggested by a partial cDNA, SC1, that in addition to excluding exon D also retained part of intron 6 (9).
Expression of 116-kDa protein is able to rescue ewg-null phenotypes. We previously demonstrated that the 116-kDa protein is able to rescue the three well-characterized phenotypes associated with ewg mutations: embryonic lethality, erect wings, and formation of dorsal longitudinal muscles (8, 10). Expression from two cDNAs expressing ewg minigenes, EWGNS (neuron specific) and EWGHS (basal level expression), rescued viability of ewgl1, an ethyl methanesulfonate-induced lethal allele, which was thought to be genetic null and 116-kDa protein null (10). The possibility that the ewg locus was formally able to generate several other isoforms made us wonder if ewgl1 was a true null for all possible EWG proteins. Sequencing of genomic DNA from exons B to D revealed a C-to-T base pair change in exon B that resulted in the termination of the ORF at amino acid 187. Since exon B is part of all putative EWG isoforms, the ewgl1 allele is a functional null allele for all possible EWG proteins; moreover, it lacks the DNA binding domain.

The rescue of ewg-associated phenotypes by the transgenes expressing the 116-kDa EWG protein was further confirmed by using a synthetic genomic deletion of the ewg locus as described in Materials and Methods (data not shown). Thus, although it is formally possible that several EWG isoforms are generated, the 116-kDa EWG protein is sufficient to provide the known EWG functions. We cannot rule out, however, the possibility that flies rescued by the 116-kDa EWG protein have subtle abnormalities that were not discerned.

Putative isoforms encoded by the ewg locus. ewg can potentially encode several polypeptides in addition to the 116-kDa EWG protein that includes exon D. Figure 5 depicts the conceptual ewg-generated isoforms. The presence or absence of exon D, which encodes 154 amino acids, does not interrupt the translational frame of the ewg ORF, while inclusion of exon E or I alters the translational frame, leading to a premature stop compared to that of the SC3 ORF (Fig. 5). Additional protein isoforms can also be generated by the ewg transcripts that retain either intron 3a, 3c, or 6. All intron 3a-, 3c-, and 6-containing RNAs result in a premature stop in the SC3 ORF, encoding 408, 574, and 840 amino acids, respectively.

All putative EWG isoforms contain exon B and C. These exons also show the highest homology with the DNA binding motifs of other ewg-like proteins (4, 18, 31). Omitting exon D from the SC3 ORF, however, significantly increases the homology of EWG to sea urchin P3A2 compared to its homology.
in the alignment in reference 31, which was done with exon D. Protein sequences encoded by exon E, exon I, and retained introns did not reveal any significant homologies.

Are isoforms other than the 116-kDa protein synthesized? These isoforms should be detectable in immunoblot analysis by the polyclonal antibody generated against the 116-kDa protein, as it recognizes epitopes throughout the protein (Fig. 6A). In fact, the anti-EWG antibody reveals several bands on immunoblots, although the 116-kDa protein is the major band (Fig. 6B) (10). To determine whether the additional bands are ewg related, we analyzed wild-type (genomic ewg), ewg	extsuperscript{11}, EWG	extsuperscript{HS7} (expression of SC3 ORF in neurons), and ewg	extsuperscript{11} (protein-null) embryonic extracts. Comparison of these extracts should indicate whether the additional bands are protein isoforms of EWG. 116-kDa protein degradation products, or unrelated cross-reacting proteins. As expected, the 116-kDa band was absent from the ewg	extsuperscript{11} lane (Fig. 6B). Further, the wild-type and ewg	extsuperscript{11}; EWG	extsuperscript{HS7} patterns were identical, with the prominent 116-kDa band and several minor bands which are likely products of degradation; the other bands were common to all three extracts. Consistent with the embryonic data, wild-type and ewg	extsuperscript{11}; EWG	extsuperscript{NS4} (EWG	extsuperscript{NS4} is not transcribed in nonneural tissues) adult heads or abdomens showed no significant differences in their immunoblot profiles (Fig. 6C and D and data not shown). Thus, the 116-kDa protein is the major isoform, and other putative isoforms are either minor or not synthesized.

**Overexpression of 116-kDa EWG is lethal.** The splicing regulation of ewg guarantees that the 116-kDa protein is generated in the head but down regulated in the body. To test if this regulation is biologically important, the effect of overexpression of the 116-kDa protein in nonneural tissues was tested. Overall overexpression was achieved by using EWG	extsuperscript{HS1} and EWG	extsuperscript{HS7}, two independent insert lines for a transgene in which the SC3 cDNA is driven with the heat shock promoter. The overexpression of the 116-kDa transcript at 25°C mimics endogenous transcription, likely because of the inclusion of regulatory sequences upstream of the ORF (SC3 cDNA in Fig. 1A). Both of these transgenes lead to lethality when homozygous. We determined the viability of flies with different doses of the wild-type ewg allele that also carry EWG	extsuperscript{HS7} and EWG	extsuperscript{HS7} (Ta-

---

**FIG. 5. ORFs of ewg splice isoforms deduced by RT-PCR analysis.** All deduced ewg RNA isoforms are outlined from data presented in Table 2. The size of each ORF is given in amino acids (aa). Transcripts including exon E and I are not shown since the ORF terminates in exon F. Also, ORFs resulting from intron-retaining transcripts are not shown.
The level of expression of 116-kDa protein in different genotypes was assessed in immunoblots (Fig. 6C and D). Comparison of 116-kDa protein levels between head extracts of flies carrying both transgenes shows that levels of 116-kDa signals are comparable in wild-type females and Males in both of these instances alternative splicing is involved; in the case of Sxl, a large number of intron-containing transcripts are also present (28). Efficiency of alternative splice events that result in SC3-like transcript is higher in head RNA. Splice events that are crucial for the production of the functional SC3 transcript are inclusion of exons D and J and exclusion of exons I and E. Inclusion of exon D requires splicing of introns 3a and 3c instead of 3b. The 3’ splice sites of both 3a and 3c and the 5’ splice site of 3c diverge from the Drosophila consensus (Table 4) (26), making them likely targets for splicing regulation. Intron 3a is inefficiently spliced in both head and body RNAs, whereas 3c is inefficiently spliced in body RNA only. Since exclusion of D resulting from exon skipping is seen in both body and head RNA; it is likely to be the default mode, with inclusion of D requiring a positive regulatory step. To what extent the inefficient splicing of 3a and 3c affects this regulation is difficult to assess. SC3 transcript also requires the appropriate choice of a 5’ splice site, resulting in the excision of intron 6. The body RNA shows inefficient splicing in the intron 6 region and, among the spliced products, about equal amounts of exon I RNA; however, our data do not address if cell types other than neurons are also capable of low-level production of the functional transcript.

Generally, transcriptional controls ensure that specific proteins are synthesized in specific tissues. In the case of ewg, however, posttranscriptional mechanisms are critical to ensure that the 116-kDa protein is present at high levels in neurons. Few examples show that broad tissue distribution of transcript but restricted protein expression are known. In Drosophila, the gene encoding P transposable is transcribed broadly but the productive splice is made only in the germ line cells (1). A second example is the Drosophila Sex-lethal gene (Sxl), which is transcribed in both males and females but generates the functional Sxl protein only in females (22). Similar to the case of ewg, in both of these instances alternative splicing is involved; in the case of Sxl, a large number of intron-containing transcripts are also present (28).

Efficiency of alternative splice events that result in SC3-like transcript is higher in head RNA. Splice events that are crucial for the production of the functional SC3 transcript are inclusion of exons D and J and exclusion of exons I and E. Inclusion of exon D requires splicing of introns 3a and 3c instead of 3b. The 3’ splice sites of both 3a and 3c and the 5’ splice site of 3c diverge from the Drosophila consensus (Table 4) (26), making them likely targets for splicing regulation. Intron 3a is inefficiently spliced in both head and body RNAs, whereas 3c is inefficiently spliced in body RNA only. Since exclusion of D resulting from exon skipping is seen in both body and head RNA; it is likely to be the default mode, with inclusion of D requiring a positive regulatory step. To what extent the inefficient splicing of 3a and 3c affects this regulation is difficult to assess. SC3 transcript also requires the appropriate choice of a 5’ splice site, resulting in the excision of intron 6. The body RNA shows inefficient splicing in the intron 6 region and, among the spliced products, about equal amounts of exon I RNA; however, our data do not address if cell types other than neurons are also capable of low-level production of the functional transcript.

Generally, transcriptional controls ensure that specific proteins are synthesized in specific tissues. In the case of ewg, however, posttranscriptional mechanisms are critical to ensure that the 116-kDa protein is present at high levels in neurons. Few examples show that broad tissue distribution of transcript but restricted protein expression are known. In Drosophila, the gene encoding P transposable is transcribed broadly but the productive splice is made only in the germ line cells (1). A second example is the Drosophila Sex-lethal gene (Sxl), which is transcribed in both males and females but generates the functional Sxl protein only in females (22). Similar to the case of ewg, in both of these instances alternative splicing is involved; in the case of Sxl, a large number of intron-containing transcripts are also present (28).

Efficiency of alternative splice events that result in SC3-like transcript is higher in head RNA. Splice events that are crucial for the production of the functional SC3 transcript are inclusion of exons D and J and exclusion of exons I and E. Inclusion of exon D requires splicing of introns 3a and 3c instead of 3b. The 3’ splice sites of both 3a and 3c and the 5’ splice site of 3c diverge from the Drosophila consensus (Table 4) (26), making them likely targets for splicing regulation. Intron 3a is inefficiently spliced in both head and body RNAs, whereas 3c is inefficiently spliced in body RNA only. Since exclusion of D resulting from exon skipping is seen in both body and head RNA; it is likely to be the default mode, with inclusion of D requiring a positive regulatory step. To what extent the inefficient splicing of 3a and 3c affects this regulation is difficult to assess. SC3 transcript also requires the appropriate choice of a 5’ splice site, resulting in the excision of intron 6. The body RNA shows inefficient splicing in the intron 6 region and, among the spliced products, about equal amounts of exon I RNA; however, our data do not address if cell types other than neurons are also capable of low-level production of the functional transcript.

Generally, transcriptional controls ensure that specific proteins are synthesized in specific tissues. In the case of ewg, however, posttranscriptional mechanisms are critical to ensure that the 116-kDa protein is present at high levels in neurons. Few examples show that broad tissue distribution of transcript but restricted protein expression are known. In Drosophila, the gene encoding P transposable is transcribed broadly but the productive splice is made only in the germ line cells (1). A second example is the Drosophila Sex-lethal gene (Sxl), which is transcribed in both males and females but generates the functional Sxl protein only in females (22). Similar to the case of ewg, in both of these instances alternative splicing is involved; in the case of Sxl, a large number of intron-containing transcripts are also present (28).
homology to EWG, has several glutamine-containing hydorphobic clusters (20). There are five such clusters in head-enriched exon D and two in exon H. Thus, the head-enriched 116-kDa protein may have stronger activation properties than proteins that lack exon D. The dimerization region for EWG has not been characterized.

Are any of the putative isoforms synthesized, do they have a function, or will they provide ewg function? The putative alternate protein isoforms have no essential role in the nervous system, as the 116-kDa EWG protein rescues viability of null ewg alleles, although they may still provide subtle functions in vivo or substitute for the 116-kDa function. Whether the alternate proteins are synthesized at low levels is not known.

The studies with heat shock transgenes show that overexpression of 116-kDa protein in nonneural tissues can be lethal, yet it is the expression of the 116-kDa protein in neurons that provides the viability function associated with ewg. With two doses of the EWGtransgenes, both ewg and ewg+/+ flies were about equally lethal, suggesting a formally neomorphic effect. The lethality could result from the expression of 116-kDa protein in nonneural tissue or due to general overexpression. An alternative possibility of the transgene generating a toxic novel protein cannot be discounted but is unlikely because the transgene was constructed by using a cDNA and not genomic DNA. Since the 116-kDa protein levels in the head with two doses of EWGtransgenes are comparable to the wild-type levels, the protein is unlikely to attain a higher than normal level in the neural tissue. Moreover, higher than normal levels of 116-kDa protein generated through two doses of EWGtransgenes, which are expressed only in the nervous system, is not lethal (our unpublished observations). Therefore, expression outside the nervous system is likely to be the primary cause of lethality. Thus, given that ewg is broadly transcribed, its posttranscriptional regulation is crucial to the fulfillment of its function.

ACKNOWLEDGMENTS

We thank D. Bordne for technical assistance. S. Goodwin for advice on RT-PCR protocols, S. Ghosh for advice on bacterial protein expression, L. Torroja for discussions, and E. Dougherty for imaging assistance.

This work was supported by National Institute of Health grants GM 22350 and NS 36179. M.S. was supported by a fellowship from the Swiss National Science Foundation. S.M.D. was supported by NIH training grant 5-T32GM-07122.

REFERENCES

1. Adams, M. D., R. S. Tarrag, and D. C. Rio. 1997. The alternative splicing factor PSI regulates P-element intron splicing in vivo. Genes Dev. 11:129–138.
2. Beall, C. J., and J. Hirsh. 1984. High levels of intron-containing RNAs are associated with expression of the Drosophila DOPA decarboxylase gene. Mol. Cell. Biol. 4:1669–1674.
3. Becker, T. S., S. M. Burgess, A. H. Amsterdam, M. L. Allende, and N. Hopkins. 1998. Not really finished is crucial for development of the zebrafish outer retina and encodes a transcription factor highly homologous to human nuclear respiratory factor-1 and avian initiation binding repressor. Development 125:4369–4378.
4. Calzone, F. J., C. Hoog, D. B. Teplow, A. E. Cutting, R. W. Zeller, R. J. Britten, and E. H. Davidson. 1991. Gene regulatory factors of the sea urchin embryo. I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein. Development 112:335–350.
5. Chabot, B. 1996. Directing alternative splicing: cast and scenarios. Trends Genet. 12:472–478.
6. Chang, D. D., and P. A. Sharp. 1989. Regulation by HIV Rev depends upon recognition of splice sites. Cell 57:789–795.
7. Cheung, P. H., O. Culic, Y. Qi, K. Earley, N. Thompson, D. C. Hisson, and S-H. Lin. 1993. The cytoplasmic domain of C-CAM is required for C-CAM-mediated adhesion function: studies of a C-CAM transcript containing an uncleaved intron. Biochem. J. 295:427–435.
8. DeSimone, S., C. Coelho, S. Roy, K. VijayRaghavan, and K. White. 1996. The alternative splicing factor PSI regulates P-element intron splicing in vivo. Genes Dev. 11:21–39.
9. DeSimone, S. M. 1992. Ph. D. thesis. Brandeis University, Waltham, Mass.
10. DeSimone, S. M., and K. White. 1993. The Drosophila erect wing gene, which is important for both neuronal and muscle development, encodes a protein which is similar to the sea urchin P3A2 DNA binding protein. Mol. Cell. Biol. 13:3641–3649.
11. Driksen, W. P., Q. Sun, and F. M. Rottman. 1995. Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. J. Biol. Chem. 270:5346–5352.
12. Ebihara, K., Y. Masuhiro, T. Kitamoto, M. Suzawa, Y. Uematsu, T. Yoshizawa, T. Ono, H. Harada, K. Matsuda, T. Hasegawa, S. Maushige, and S. Kato. 1996. Intron retention generates a novel isoform of the murine vitamin D receptor that acts in a dominant negative way on the vitamin D signalling pathway. Mol. Cell. Biol. 16:3393–3400.
13. Eftekhar, A., S. J., A. Chiorini, and B. Safer. 1994. A key transcription factor for eukaryotic initiation factor-2a is strongly homologous to developmental transcription factors and may link metabolic genes to cellular growth and differentiation. Mol. Cell. Biol. 14:1086–1093.
14. Flemng, R. J. 1987. Ph. D. thesis. Brandeis University, Waltham, Mass.
15. Fleming, R. J., S. M. DeSimone, and K. White. 1989. Molecular isolation and analysis of the erect wing locus in Drosophila melanogaster. Mol. Cell. Biol. 9:1645–1654.
16. Fleming, R. J., S. Zusan, and K. White. 1983. Developmental genetic analysis of lethal alleles at the ewg locus and their effects on muscle development in Drosophila melanogaster. Dev. Genet. 3:347–363.
17. Gobin, S., and J. M. Lowneinstein. 1990. A multifunctional nuclear vector system for heterologous expression of proteins in Escherichia coli: expression of native and hexahistidyl fusion proteins, rapid purification of the fusion proteins, and removal of fusion peptide by Kex2 protease. Gene 176:249–255.
18. Gomes-Cruzoel, A., M. Martin, M. Noel, and A. Ruiz-Carrillo. 1995. Initiation binding receptor, a factor that binds to the transcription initiation site of the histone h5 gene, is a glycosylated member of a family of cell growth regulators. Mol. Cell. Biol. 15:6670–6685.
19. Grabowski, P. 1998. Splicing regulation in neurons: tinkinger with cell-specific control. Cell 92:709–712.
20. Gugneja, S., C. A. Virbasius, and R. C. Scarpulla. 1996. Nuclear regulatory factors 1 and 2 utilize similar glutamine-containing clusters of hydrophobic residues to activate transcription. Mol. Cell. Biol. 16:5708–5716.
21. Hamn, J., and L. W. Mattaj. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. Cell 63:109–118.
22. Horabin, J. I., and P. Schell. 1993. Regulated splicing of the Drosophila Sex-lethal male exon involves a blockage mechanism. Mol. Cell. Biol. 13:1040–1044.
23. Legrain, P., and M. Rosbash. 1989. Some cis-acting and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell 57:573–583.
24. Lopez, A. J. 1995. Developmental role of transcription factor isoforms generated by alternative splicing. Develop. Biol. 172:392–411.
25. Mattox, W., and B. S. Baker. 1991. Autoroulage of the splicing of transcripts from the transformer-2 gene of Drosophila. Genes Dev. 5:786–796.
26. Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields. 1992. Splicing signals in Drosophila: intron size, consensus sequences, and consensus sequences. Nucleic Acids Res. 20:4255–4262.
27. Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp. 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. 55:1119–1150.
28. Rozen, S., and H. Skaletsky. 1996, posting date. Primer3. [Online.] http://www.genome.wi.mit.edu/genome_software/other/primer3.html [7 April 1999, last date accessed.]