Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila

Xiao Guan, Brooke W. Middlebrooks, Sherry Alexander, and Steven A. Wasserman*

Section of Cell and Developmental Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0349

Communicated by Dan L. Lindsley, University of California at San Diego, La Jolla, CA, September 6, 2006 (received for review July 17, 2006)

Body shape determination represents a critical aspect of morphogenesis. In the course of investigating body shape regulation in Drosophila, we have identified a dominant mutation, TweedleD (TweedleD), that alters overall dimensions at the larval and pupal stages. Characterization of the affected locus led to the discovery of a gene family that has 27 members in Drosophila and is found only among insects. Analysis of gene expression at the RNA and protein levels revealed gene-specific temporal and spatial patterns in ectodermally derived tissues. In addition, light microscopic studies of fluorescently tagged proteins demonstrated that Tweedle proteins are incorporated into larval cuticular structures. This demonstration that a mutation in a Drosophila cuticular protein gene alters overall morphology confirms a role for the fly exoskeleton in determining body shape. Furthermore, parallels between these findings and studies of cuticle collagen genes in Caenorhabditis elegans suggest that the exoskeleton influences body shape in diverse organisms.

morphogenesis | tandem duplication | Tubby

Mutations in a gene family that has 27 members in Drosophila and is found only among insects. Analysis of gene expression at the RNA and protein levels revealed gene-specific temporal and spatial patterns in ectodermally derived tissues. In addition, light microscopic studies of fluorescently tagged proteins demonstrated that Tweedle proteins are incorporated into larval cuticular structures. This demonstration that a mutation in a Drosophila cuticular protein gene alters overall morphology confirms a role for the fly exoskeleton in determining body shape. Furthermore, parallels between these findings and studies of cuticle collagen genes in Caenorhabditis elegans suggest that the exoskeleton influences body shape in diverse organisms.


The authors declare no conflict of interest.

Abbreviations: RFP, red fluorescent protein; A.R., axial ratio.

*To whom correspondence should be addressed. E-mail: stevenw@ucsd.edu.

© 2006 by The National Academy of Sciences of the USA
Overall, the phenotype of the TwdlD¹ mutation is very similar to that of a previously described mutation, Tubby¹ (Tb¹). (The Tb locus was named for the mutant phenotype and does not correspond to either Drosophila homolog of the mammalian Tubby gene.) As with TwdlD¹, the Tb¹ mutation reduces the A.R. in a dominant fashion without affecting viability or fertility. At the pupal stage, we find that Tb¹ results in an A.R. of 2.0 ± 0.1, a slightly more severe effect than was seen with TwdlD¹. Like TwdlD¹, Tb¹ results in a heterozygous phenotype that is indistinguishable from that of homozygous mutants. Furthermore, we localized TwdlD by meiotic recombination mapping to position 3–91, where Tb had previously been mapped (25).

The TweedleD Locus Encodes a Member of an Insect-Specific Gene Family. To initiate a molecular characterization of the TweedleD and Tubby loci, we carried out fine-structure mapping by P-element-induced male recombination (26). The meiotic map position 3–91 corresponds to the region surrounding polytene band 97C on the right arm of the third chromosome. We therefore used six P insertion lines within the region 97B9–D3 to map TwdlD¹ and Tb¹ (Fig. 2). Both mutations mapped to the 73.8-kb interval proximal to the gene amontillado. We sequenced genomic DNA spanning this 73.8-kb region for Tbw¹ homozygotes, TwdlD¹ homozygotes, and the parental strain from which TwdlD¹ was derived (the parental strain for Tb¹ was not available). Initial sequencing of Tb¹ revealed a high density of sequence alterations relative to the published genomic sequence. For TwdlD¹, however, we found a single sequence change relative to the parental strain: a 9-nt deletion in the published sequence for CG14243, an uncharacterized gene.

The TweedleD protein encoded by CG14243 contains 256 aa. Examination of the predicted D. melanogaster proteome reveals 26 homologues of TwdlD. Overall sequence identity to TwdlD ranges from 26% to 54%. Of the 27 related genes, which we have termed the Tweedle family, TwdlD and 21 others are distributed among three gene clusters: one on the X chromosome (polynet bands 15A3) and two on the third chromosome (polynet bands 82A1 and 97C). Each of the 27 family members is predicted to contain an N-terminal signal peptide, but not a transmembrane domain.

To determine whether Tweedle family members are secreted, we expressed epitope-tagged forms of TwdlD and TwdlJ (CG5471) in cultured Drosophila S2 cells. For both genes, the vast majority of the tagged protein was found in the culture media (Fig. 3), demonstrating that the polypeptides are indeed being secreted.

Using the TwdlD amino acid sequence to search translated forms of animal and plant genomes, we identified two or more Tweedle family members in all insects examined, but none in any other species, including Daphnia, a crustacean. Alignment of protein sequences from Drosophila, Anopheles, Aedes, Bombyx,
The observation that the TwdlD phenotype was apparent in newly hatched first-instar larvae provided good evidence for embryonic expression of TwdlD. To determine when and where TwdlD is expressed, and whether distinct family members differ in their expression patterns, we carried out in situ hybridization in embryos. For all seven of the genes tested, transcripts were detected at embryonic stages 13–16 (Fig. 6), but not earlier (data not shown). Expression of TwdlD, as well as TwdIB (CG14639), was detected within the epidermis, with the expression of TwdlD and TwdIB forming segmental stripes along the anteroposterior axis. In contrast, the transcripts of the remaining loci were each found in a more restricted domain of the embryo: the tracheal tree [Twdlβ (CG8986)], dorsal epidermis (TwdlA), and the foregut [TwdlC (CG14254) and TwdlE (CG14534)]. Although the patterns were varied, expression of all family members tested was confined to tissues that are ectodermal in origin.

Theoretically, an alteration in body shape could result from various causes, including neuronal malfunction, defective musculature, or structural abnormalities of the integument. To delimit the biological functions most likely affected by the TwdlD mutation, we designed experiments to monitor wild-type patterns of protein expression for TwdlD and two other family members tested was confined to tissues that are ectodermal in origin.
members. We generated fusion protein constructs in which sequences encoding monomeric red fluorescent protein (RFP) were introduced immediately 3′ to the coding regions of TwdlD, TwdlF, and TwdlH (CG31080). Each construct retained endogenous sequences representing the promoter, as well as the 5′ and 3′ UTR. Independent transgenic lines were generated by P element-mediated transformation.

The three family members exhibited distinct temporal and spatial localization patterns (Fig. 7). Expression of the TwdlD–RFP fusion protein was strongest during the first and the second larval instars. During these stages, TwdlD was detectable in the integument, as well as the tracheal tree (Figs. 7 and 8A). Fluorescence was associated with both dorsal and lateral regions of the integument as well as TwdlH fluorescence in dorsal hairs. (I–K) TwdlF–RFP fluorescence in ventral denticles is confined to the basal portion. A differential interference contrast image (I), an RFP image (J), and a merged image (K) are presented. tt, tracheal tree; vd, ventral denticle.

![Fig. 6. Tweedle family genes are expressed in the hypodermis, foregut, or tracheal system of late-stage embryos. RNA expression of genes TwdlA–TwdlF as well as Twdlβ was examined by in situ hybridization with antisense probes. All embryos are arranged with anterior to the left. Both dorsal (Left) and lateral (Center) embryo aspects are presented for each gene. A lateral (Right) view of embryos probed with sense strand control is also presented.](image1)

Fig. 7. Proteins of gene TwdlD, TwdlF, and TwdlH show a distinct temporal and spatial localization pattern. The location of each protein in the transgenic flies was monitored by means of the RFP tag fused to the C terminus. The TwdlD fusion protein is principally detectable during the first- and second-instar larval stages. The TwdlF fusion protein is identified in a fine layer of the integument throughout the whole larval stages. The TwdlH fusion protein is only visible in the third-instar larvae and within the segments close to the anterior or posterior ends.### Table 1: Temporal and Spatial Expression Patterns of Tweedle Family Members

<table>
<thead>
<tr>
<th>Gene</th>
<th>Expression Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>TwdlA</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlB</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlC</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlD</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlE</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlF</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>TwdlH</td>
<td>Dorsal, lateral</td>
</tr>
<tr>
<td>Twdlβ</td>
<td>Dorsal, lateral</td>
</tr>
</tbody>
</table>

![Fig. 8. Genes TwdlD and TwdlF encode cuticular proteins. Fixed first-instar larvae (2 h) were observed by confocal microscopy. (A) The TwdlD–RFP fusion protein forms 10 stripes on the integument along the AP axis. (B) The TwdlD–RFP protein is incorporated into the cuticle structure. The arrow in B shows a site where the larval body is detached from the cuticle. (C) TwdlD–RFP fluorescence is apparent in cuticular dorsal hairs. (D) The TwdlD–RFP protein is incorporated into the tracheal system. (E) The ventral denticles are not fluorescent in the TwdlD–RFP transgenics. (F and G) The TwdlF–RFP fusion protein is detectable within the cuticle. (H) TwdlF–RFP fluorescence in dorsal hairs. (I–K) TwdlF–RFP fluorescence in ventral denticles is confined to the basal portion. A differential interference contrast image (I), an RFP image (J), and a merged image (K) are presented. tt, tracheal tree; vd, ventral denticle.](image2)
each instar. The cuticle is attached to the hypodermis at multiple anchor sites (15). Extensive cross-linking is crucial in determining the mechanical properties of the cuticle. The Tweedle $D^1$ mutation, which does not change the stability or the secretion of the protein, is therefore very likely to affect either the conformation of the protein or its activity in forming or stabilizing cross-links.

The fact that the four most conserved blocks of amino acids in the Tweedle protein are predicted to form $\beta$-strands is intriguing. Previous studies of insect cuticle proteins have suggested that the barrel structure formed by multiple $\beta$-strands provides an interface for aromatic residues to stack with and bind to chitin (27, 28). In this regard, we note that several of the highly conserved residues in the Tweedle proteins that lie within these predicted $\beta$-strands have aromatic side chains: Y and H in block I, Y and F in block II, Y in block III, and F and Y in block IV. We therefore postulate that the Twd family proteins interact directly with chitin.

The Tweedle family members in the Drosophila genome form three major gene clusters. The 97C cluster, which includes the TwdD gene, consists of 14 family members. This cluster can be furthered divided in half, with the genes in each half all being transcribed from the same DNA strand. Why has the Tweedle protein is predicted to form $\beta$-strands is intriguing. Previous studies of insect cuticle proteins have suggested that the barrel structure formed by multiple $\beta$-strands provides an interface for aromatic residues to stack with and bind to chitin. The authors noted that BmGRP2 contains a glycine-rich domain that is present in cuticle and other structural proteins in many species, where such domains are proposed to provide flexibility. We note, however, that BmGRP2 also contains a sequence with substantial similarity to the Tweedle family signature motif YYXL$_{30-23}$KPEPyFkY (R/K) ($\beta$-strands) (see Fig. 4).

Like BmGRP2, some Tweedle proteins contain glycine-rich domains. However, the glycine-rich domain is absent in 21 of the 27 Tweedle genes in Drosophila, including the three studied here at the protein level: TwdD, TwdI, and TwdH. Furthermore, many glycine-rich cuticle proteins lack the motif conserved in the Tweedle family. For these reasons, we speculate that the Tweedle motif and the glycine-rich domain have distinct and largely independent functions in cuticle formation.

Genetic Control of Larval and Pupal Body Shape. Although the $TwdD^1$ phenotype is mostly easily recognized during the larval and the pupal stages, $TwdD$ gene expression begins in the latter half of embryogenesis (see Fig. 6) and is no longer detectable by the end of the last larval stage. The lack of any shape alteration in $TwdD^1$ embryos (data not shown) presumably reflects the fact that the surrounding eggshell is a protein-based extracellular matrix distinct from cuticle (30). Within the eggshell, however, the embryonic cuticle structure is clearly affected, as is evident upon examination of newly hatched first-instar larvae (see Fig. 1). A strong $TwdD^1$ phenotype observed during the pupal stage, after the cessation of gene expression, very likely represents residual effects during pupariation of the larval cuticle abnormality.

Two previously described dominant mutations, Tubby$^1$ ($Tb^1$) and KugelValencia$^1$ ($Kv^1$), have phenotypes highly reminiscent of $TwdD^1$. Like $TwdD^1$, these mutations reduce A.R. at the larval and pupal stages. This similarity suggests that the three loci may act in the same pathway. We have mapped $Tb^1$ to the same 73.8-kb region as $TwdD^1$ by P-induced male recombination, and we found a deletion within TwdA that is likely cause of the $Tb^1$ phenotype. In the case of $Kv^1$, the mutation maps to the left side of the gene $K$, which is positioned at 83D–E on the polytene map. Although the mapping is less precise than that for $Tb^1$, this position is also roughly coincident with the location of a Tweedle gene cluster, the four Tweedle genes at 82A. We consider it very likely, therefore, that a mutation in this gene cluster is mutated in the $Kv^1$ mutant.

People have known for a long time that the disruption of normal cuticle structure in C. elegans can cause the dumpy phenotype, which describes the shorter and wider morphology of the mutant worms. We have demonstrated in this report that the $TwdD$ mutation of the cuticular protein TwdD causes a similar morphology change in the fruit fly. The analogy between the two systems highlights the importance of a cuticle in maintaining the wild-type body shape in organisms with an exoskeleton.

Materials and Methods

Genetic Screen for Morphology Mutations. To screen for dominant mutations on the third chromosome affecting pupal shape, we crossed mutagenized males (4,000-rad $\gamma$-irradiation) to virgin females and assayed directly for altered A.R. in pupae (see below). From $\approx$25,400 pupae, we identified two stable dominant mutations and characterized one, designated Tweedle$D^1$.

Axial Ratio Determination. For pupae, A.R. (length/width) was measured by using a reticle in a stereo light microscope. For each genotype we measured at least 40 individuals and calculated the mean A.R. For larvae, A.R. was determined from digital photographs of cuticle preparations. At least two independent preparations were examined for each genotype, and 20 individual cuticles were measured for each preparation.

P-Induced Male Recombination Mapping. $P$ element-induced male recombination mapping was performed as described (26). Triply labeled chromosomes Ly $Tb^1$ Dr and e $TwdD^1$ Dr were generated by meiotic recombination. $P$ insertion lines BL12808, BL13710, BL20052, BL13022, BL10343, and BL11782 were obtained from Bloomington Stock Center (Bloomington, IN).

Sequence Analysis and Gene Assignments. Similarity searches used BLAST (www.ncbi.nlm.nih.gov/BLAST). The multiple protein alignment and similarity analysis were carried out by using CLUSTALW (www.ebi.ac.uk/clustalw). Detailed information on gene assignments for Tweedle family members is provided in Supporting Materials and Methods and Table 1, which are published as supporting information on the PNAS web site. The signal peptide prediction was made by using SignalP 3.0 (www.cbs.dtu.dk/services/SignalP) (31). Secondary structures were predicted with the PHD algorithm (52) (www.predictprotein.org) (33).

Protein Expression in S2 Cell Culture. The TweedleD coding region was fused to a FLAG tag at its C terminus and cloned into the pAc5.1/V5-His A vector (Invitrogen, Carlsbad, CA). The TweedleD coding region was cloned into the same vector, where it was fused to the V5 epitope tag. S2 cell transfection and protein extract harvesting were as described (Drosophila Expression System; Invitrogen). For each plate of transfected cells, the media and the cell pellet were separated by centrifugation at 850 × g, and a 1/60th volume of the media and of the total cell lysate was each loaded.

16798 | www.pnas.org/cgi/doi/10.1073/pnas.0607616103 Guan et al.
onto an SDS/PAGE gel for immunoblotting. Antibodies used in this experiment were as follows: anti-V5 antibody at 1:10,000 dilution (40-0705; Invitrogen), anti-FLAG M2 antibody at 1:1,000 dilution (200472-2; Stratagene, La Jolla, CA), and rabbit anti-
Cactus antiseraum at 1:10,000 dilution (34).

Embryonic in Situ Hybridization. Embryonic RNA expression patterns were investigated by in situ hybridization. The 3’ UTRs of target genes were amplified from the w1118 genome by PCR and cloned into the pBluescript vector (Stratagene). Digoxigenin-11-UTP was incorporated into sense and antisense probes generated with T7 and T3 RNA polymerase, respectively. Alkaline phosphatase-conjugated anti-digoxigenin antibody (Fab fragments; Roche, Pleasanton, CA) was used at a 1:2,000 dilution.

RFP Constructs, Transgenic Flies, and Microscopy. Tweedle gene genomic fragments including 500 bp of presumptive upstream regulatory sequence were cloned by PCR from the w1118 genome. We used PCR sewing (35) to fuse the 3’ end of each coding sequence in frame with sequences for the monomeric RFP DsRed (Clontech, Mountain View, CA). The resulting DNA fragments were ligated into the pCaSpeR transformation vector (36). Three independent transgenic lines were generated for each construct. Eggs were collected at 25°C for 2 h for each transgenic line and aged for 22, 48, or 72 h to obtain the young first-, second-, and third-instar larvae, respectively. Two-hour-old first-instar larvae were fixed as described (37) and observed under a confocal microscope.

Note. We have confirmed that TwdfA corresponds to the Tubby locus: Ten independent lines carrying a TwdfA transgene generated from Tb⁺ have a squat pupal shape, whereas five lines carrying a TwdfA transgene from w1118 have a wild-type body shape.

We thank Dan Lindsley for helpful discussions; Jim Posakony (University of California at San Diego) for the transformation vector and RFP isoform used in these studies; Ethan Bier, Raffi Aroian, and Bill McGinnis for sharing microscopy facilities; and Ethan Bier for comments on the manuscript. This work was supported by National Institutes of Health Grant R01 GM05545 (to S.A.W.).