UNIVERSITY OF CALIFORNIA, SAN DIEGO

Body Shape Regulation by Tweedle Family Proteins in Drosophila melanogaster

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Xiao Guan

Committee in charge:

Professor Steven A. Wasserman, Chair
Professor Rick A. Firtel
Professor Xiang-Dong Fu
Professor William J. McGinnis
Professor Amy Pasquinelli

2006

The dissertation of Xiao Guan is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2006

I want to thank my parents, for their everlasting faith in me; to Jun Yan, for his support through the difficult times

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

First, I thank my advisor Steve Wasserman, for his continuous support through my graduate study. Steve taught me how to think independently and critically, how to communicate and present effectively, and how to overcome obstacles by persistence and creative thinking. Steve is a good listener, who always encourages and welcomes new, sometimes naive ideas. Steve is always willing to help, with problems both inside and outside the lab. In particular, without Steve's full support, I couldn't have successfully engaged into the new research direction that leads to this thesis.

I also want to thank all former and current members of the Wasserman lab for making it the best learning environment. In particular, former postdoc, Huaiyu Sun, taught me all the basic molecular biology and cell culture techniques in a very comprehensible and yet precise way. Par Towb, is never reluctant in giving advices and providing help. Par's knowledge and expertise has made him the best person to turn to whenever a difficulty is encountered, regardless of whether it is a problematic western blot or a mis-adjusted microscope. A special thanks goes to Brooke Middlebrooks, who is both the embryo injection master and the 'cake' master of the lab. As a co-author of a research paper published this year in PNAS, Brooke has done fabulous work on generating transgenic flies, which laid solid basis for the study I conducted. Furthermore, we have all been enjoying the carefully-prepared, southern-flavored cakes made out of Brooke's family recipe colletion!

Besides my advisor, I would like to acknowledge the rest of my thesis committee as well: Rick Firtel, for his good questions; Xiang-Dong Fu, who gave insightful comments; Bill McGinnis, who made key suggestions consistently through the five years, and Amy Pasquinelli, for valuable discussions and spirit-lifting encouragement. To everyone on the committee, I want to say 'thank you' for supporting me.

Chapter 2 of this dissertation includes the reprint of the following paper:

Xiao Guan, Brooke W. Middlebrooks, Sherry Alexander,Steven A. Wasserman - *Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila*, PNAS, Vol. 103(45), pp. 16794-16799, Nov. 2006.

The dissertation author was the primary author listed in this publication. And the co-author, Professor Wasserman, directed and supervised the research.

VITA

1999	B.S. (Department of Biological Sciences and Biotechnology) Tsinghua University, Beijing, China
2006	Ph.D. (Biology) University of California, San Diego, USA

PUBLICATIONS

Journal Paper

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

Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila
PNAS, Vol. 103(45), pp. 16794-16799, Nov. 2006.

FIELDS OF STUDY

Major Field: Biology

Studies in Body Shape Regulation in Drosophila melanogaster.
Professors Steven A. Wasserman

Studies in Spermatogenesis in Drosophila melanogaster.
Professors Steven A. Wasserman

ABSTRACT OF THE DISSERTATION

Body Shape Regulation by Tweedle Family Proteins in Drosophila melanogaster

by

Xiao Guan

Doctor of Philosophy in Biology

University of California, San Diego, 2006

Professor Steven A. Wasserman, Chair

In this dissertation, I study the body shape control mechanism(s) in fruit fly, *Drosophila melanogaster*. The study focuses on two dominant mutations, namely $TweedleD^1$ and $Tubby^1$, which result in short, thickset body shape in post-embryonic stages. The results of this study are presented in five sections, to report: 1. identification of $TweedleD^1$ gene and $Tubby^1$ gene; 2. functional analysis of the Tweedle family proteins; 3. close examination of $TweedleD^1$ mutants; 4. model of body shape regulation mediated by Tweedle genes; 5. functional, biochemical and regulatory studies of Tweedle family proteins for further investigation.

By combining *p* element-induced male recombination and direct genomic sequencing, genes that are mutated in $TweedleD^1$ ($TwdlD^1$) and $Tubby^1$ (Tb^1) mutants were identified, and they were further verified in transgenic flies. Both TwdlD and Tb proteins belong to a novel insect-specific protein family - the Tweedle family. The Tweedle genes encode proteins that are secreted by ectodermal tissues including epidermis, foregut and trachea. Some of these proteins were shown to contribute to extracellular matrix structure - cuticle. Each Tweedle gene has a specific temporal and spatial expression and localization pattern. These patterns of Tweedle genes might be crucial in determining body shape, as suggested by mislocalization of $TwdlD^1$ protein in $TwdlD^1$ mutants. Based on these observations, a model was proposed to consolidate both the large number of Tweedle genes ex-

isting in Drosophila genome and the roles of Tweedle proteins in controlling body shape. Lastly, three aspects of Tweedle family were targeted in pilot studies, and preliminary data support promising research directions that could lead to better understanding of this family and their roles in body shape regulation.

This demonstration that mutation of Tweedle proteins results in body shape change highlights the role of exoskeleton in determining overall body shape. Similar phenotypes were observed in nematode *C. elegans* when cuticular collagens were mutated. This recurring scheme of exoskeleton controlling body shape is a result of convergent evolution.

1

Introduction

To acquire proper body shape is a critical task of multicellular morphogenesis. Despite its importance, there isn't a systematic understanding about how body shape is developed and maintained. Although the shape of an individual normally falls into a range centered on a 'typical' shape for the species, the fact that most species have a typical shape itself reflects the existence of internal control mechanisms.

Historically, studies of body shaping mechanisms have been advanced by progress in the field of organ and body size regulation. In particular, adaptation of existing morphogens to regulate organ size has been viewed as a major way to change body shape through evolution. This theory of shape evolution is based on the idea first proposed in 1917 by British polymath D'Arcy Thompson, who speculated that simple alterations of a 'model' pattern could explain the shape differences between different organisms (Thompson 1917, Stern 2006). Although the key functions of morphogen in dictating organ size and local shape are clearly demonstrated in some cases (Crickmore & Mann 2006, de Navas, Garaulet & Sanchez-Herrero 2006, Makhijani, Kalyani, Srividya & Shashidhara 2006) , morphogens alone are not likely to explain how overall body shape is determined through the complex developmental process, where every organ in an organism is constantly growing, differentiating and/or rearranging.

Regulation of overall body shape demands means that can control and coordinate the whole organism. One of the early events that establish the future body plan is convergent extension during embryogenesis. Convergent extension is a rearrangement process of embryonic cells driven by polarized movement, normally resulting in elongation of the embryo. Recent studies by several groups suggest that, on the cellular level, convergent extension is regulated by Frizzled triggered planar cell polarity pathway (Ninomiya, Elinson & Winklbauer 2004, Wallingford, Vogeli & Harland 2001, Heisenberg, Tada, Rauch, Saude, Concha, Geisler, Stemple, Smith & Wilson 2000) . As differentiation and morphogenesis continues, other mechanisms are needed to maintain the proper body shape. Due to the complexity of the issue and the uniqueness of each species, these shaping mechanisms remain elusive.

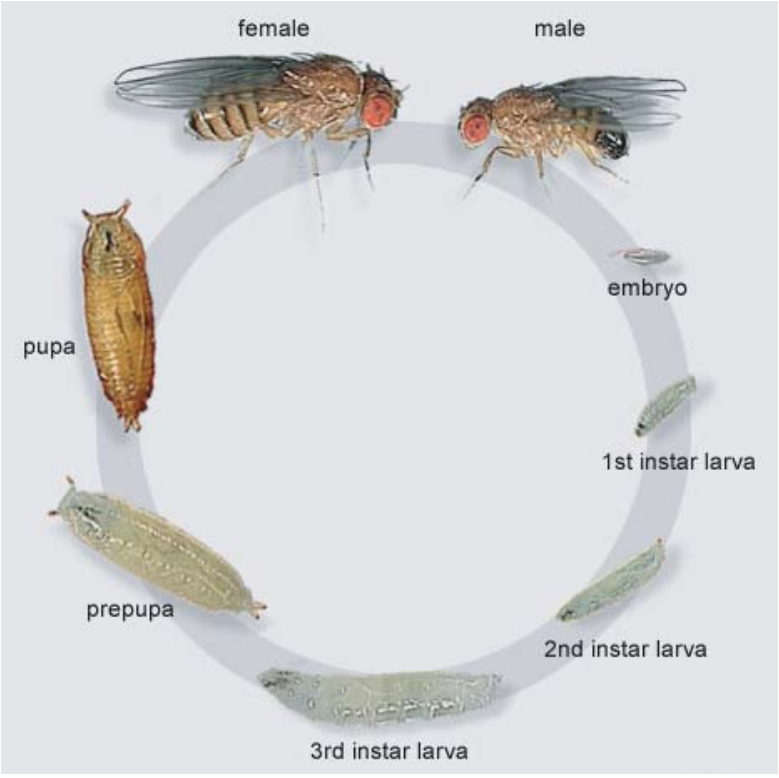


Figure 1.1: Life cycle of *Drosophila melanogaster*. Adapted from http://flymove.uni-muenster.de

Drosophila melanogaster provides a good system to study overall body shape regulation. The life cycle of *Drosophila* comprises four major stages: embryo, larva (including first, second and third instar larval stages), pupa and adult (Figure 1.1). Since most growth happens during the three larval stages, the final size and the rough dimensions of an adult fly are largely determined by the time of the pupal stage. Drosophila is one of the best-studied organisms regarding size and growth regulation. Multiple signaling molecules and pathways have been found to regulate metabolism and size, including Insulin/PI3-kinase/Akt/TOR, Ras/Raf/Erk and ecdysone (Neufeld 2003, Mirth, Truman & Riddiford 2005, Caldwell, Walkiewicz & Stern 2005, Colombani, Bianchini, Layalle, Pondeville, Dauphin-Villemant, Antoniewski, Carre, Noselli & Leopold 2005) . Crucial environmental factors, for example, nutrition control size by influencing the balance of these molecules. Several papers published in 2005 supported a function of the prothoracic gland in 'sensing' and controlling the final size for metamorphosis(Mirth, Truman & Riddiford 2005, Caldwell, Walkiewicz & Stern 2005, Colombani, Bianchini, Layalle, Pondeville, Dauphin-Villemant, Antoniewski, Carre, Noselli & Leopold 2005). In one of these papers, the authors managed to decrease or increase adult body size by increasing or decreasing the activity of PI3-kinase (Colombani, Bianchini, Layalle, Pondeville, Dauphin-Villemant, Antoniewski, Carre, Noselli & Leopold 2005). Interestingly, as size scales up and down, the pupae retain the same overall shape, or dimension ratios. These observations clearly indicate there are size-independent shape regulation mechanisms in Drosophila. The goal of this study is to identify such mechanisms and to contribute to the understanding of overall body shape control in general.

1.1 Overall body shape regulation in *C. elegans*

The most extensive studies of body shape determination have been carried out with the nematode *C. elegans*. Up to now, more than 50 mutations that effect

morphology have been identified and most of them result from defects in the cuticle. These mutants are classified into five groups according to their phenotypes. The groups are: (1) dumpy (Dpy), shortening of body length. (2) blister (Bli), liquid-filled separation of the cuticle. (3) roller (Rol), twisting of the cuticle along the length of the body. (4) long (Lon), longer and thinner body. (5) squat (Sqt), roller when heterozygous and dumpy when homozygous. Among them, the Dpy group and the Lon group represent typical body shape mutants. Lesions of cuticle collagen genes have been found to cause phenotypes belonging to all five groups (Nystrom, Shen, Aili, Flemming, Leroi & Tuck 2002, Kramer, Johnson, Edgar, Basch & Roberts 1988, Kramer, French, Park & Johnson 1990, Johnstone, Shafi & Barry 1992) .

Many of the Dpy and the Lon group of mutations have been mapped (Levy, Yang & Kramer 1993, Clark, Suleman, Beckenbach, Gilchrist & Baillie 1995, Hill, Harfe, Dobbins & L'Hernault 2000) (Figure 1.2). More than half of these mutations are lesions of cuticle collagens, including *dpy-13, dpy-7, dpy-2, dpy-10, dpy-5, dpy14, lon-3* and et al (Gallo, Mah, Johnsen, Rose & Baillie 2006, Johnstone, Shafi & Barry 1992, Levy, Yang & Kramer 1993, Nystrom, Shen, Aili, Flemming, Leroi & Tuck 2002, von Mende, Bird, Albert & Riddle 1988) . In addition, loss-of-function mutations of collagen modifying enzymes α-subunit of prolyl 4-hydroxylase (as in *dpy-18*) and zinc-metalloprotease (as in *dpy-31*) both cause recessive dumpy phenotype (Hill, Harfe, Dobbins & L'Hernault 2000, Myllyharju, Kukkola, Winter & Page 2002, Novelli, Ahmed & Hodgkin 2004) . Furthermore, recent evidence suggests that DBL-1 TGF-beta signaling pathway might contribute to body shape control by regulating cuticle collagen gene *lon-3* post-transcriptionally (Suzuki, Morris, Han & Wood 2002).

The *C. elegans* genome encodes about 175 cuticle collagens. These small (about 30kD) collagen-like molecules are the key structural components of worm cuticle. Worm cuticle collagens are normally composed of a small N-terminal collagen domain, a large C-terminal collagen domain and three Cysteine-containing,

Polypeptide	Gene	Typical phenotype[a]
Cuticle collagen	*dpy-2*, *dpy-3*, *dpy-5*, *dpy-7*, *dpy-8*, *dpy-10*, *dpy-13*	Dumpy
	bli-1, *bli-2*	Blister
	rol-6	Roller or dumpy[b]
	sqt-1,*sqt-3*	Roller or dumpy[b]
	lon-3	Long
Collagen IV: α1(IV); α2(IV)	*emb-9*; *let-2*	Embryonically lethal
Collagen XVIII	*cle-1*	Defects in cell and axon migration and neuromuscular synapse function
P4H[c], PHY-1	*phy-1* (also known as *dpy-18*)	Dumpy
P4H[c], PHY-2	*phy-2*	Wild-type
P4H[c], PHY-1 and PHY-2	*phy-1* and *phy-2*	Severe dumpy or embryonically lethal[d]
P4H[c], PHY-3	*phy-3*	Wild-type
PDI-2[c]	*pdi-2*	Severe dumpy or embryonically lethal[d]
LH	*let-268*	Embryonically lethal
Subtilisin-like protease	*bli-4*	Embryonically lethal or blister
Thioredoxin	*dpy-11*	Dumpy
ERp60	*pdi-3*	Mild disruption of cuticle collagen localization
Duox 1; duox 2	F56C11.1; F53G12.3	Dumpy and blister

Figure 1.2: Mutations in *Caenorhabditis elegans* collagens and their modifying enzymes. Adapted from article by Myllyharju and Kivirikko, Trends in genetics, 20, 33-43. 2004.

non-collagenous domains. As with other collagen molecules, the Gly-X-Y repeats within the collagen domains are crucial for proper collagen assembly. As a result, missense mutations that substitute another amino acid for the conserved Glycine are frequently associated with morphological changes (Johnstone, Shafi & Barry 1992). For the most part, these mutations act recessively since misassembled complexes get degraded. However in several cases, Glycine substitution and mutations of the "Arg-X-X-Arg" cleavage motif have been shown to generate severe dominant effects, indicating that incorporation of unprocessed procollagen into the cuticle causes a structure abnormality (Yang & Kramer 1994, Levy, Yang & Kramer 1993, Thacker, Sheps & Rose 2006).

These data clearly point to cuticle composition as being a major factor in determining the overall shape of individual worms.

1.2 Overall body shape regulation in Drosophila

There have not been large-scale screenings for mutants with altered body shape in Drosophila. Nevertheless, body shape changes have been observed in mutants of several pathways/genes.

1.2.1 Shape regulation through neuromuscular functions

Toll pathway

Toll pathway signaling is required maternally for embryonic DV patterning (Anderson, Bokla & Nusslein-Volhard 1985, Anderson, Jurgens & Nusslein-Volhard 1985), and zygotically for development and immunity (Gerttula, Jin & Anderson 1988, Qiu, Pan & Govind 1998, Lemaitre, Nicolas, Michaut, Reichhart & Hoffmann 1996). Besides its role in activating humoral reactions against fungi and Gram(+) bacteria infection, Toll pathway has been implicated in processes including motorneuron and muscle development, blood cell proliferation and other unidentified functions (Halfon, Hashimoto & Keshishian 1995, Qiu, Pan & Govind 1998). As a result, only 5% of Toll null flies survive to adult stage and the surviving individuals are immune incompetent.

Unlike mammalian Toll-like receptors, Drosophila Toll receptor does not interact directly with microbial components. Instead, a processed ligand Spätzle is needed for Toll activation. Distinct sets of serine proteases are activated by stimuli corresponding to different functions of Toll. Each protease cascade leads to cleavage of Spätzle and activation of Toll pathway. For instance, maternal stimuli for DV patterning activate Spätzle through Gastrulation defective, Snake and Easter; fungal infection activates Spätzle through Spirit, SPE, Sphinx1/2, Spheroide, Persephone and its inhibitor Necrotic; Gram(+) bacteria infection is recognized by PGRP-SA, which activates Spätzle through Spirit, SPE, Sphinx1/2, Spheroide and Grass (Ligoxygakis, Pelte, Hoffmann & Reichhart 2002, Kambris, Brun, Jang, Nam, Romeo, Takahashi, Lee, Ueda & Lemaitre 2006, Robertson, Belorgey, Lilley,

Lomas, Gubb & Dafforn 2003, Han, Lee, Tan, LeMosy & Hashimoto 2000).

The known intracellular components of the Toll pathway include adaptor proteins Weckle, Myd88 and Tube, protein kinase Pelle, NF-κB factor Dif/Dorsal and their inhibitor Cactus. The activation of Toll receptor results in degradation of Cactus and translocation of Dif/Dorsal into the nucleus to regulate transcription of target genes.

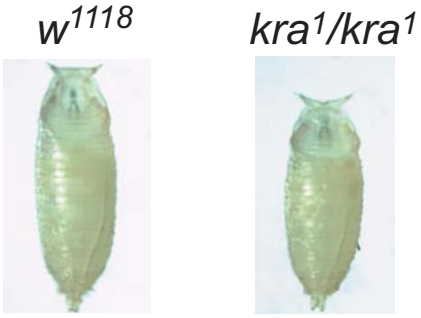


Figure 1.3: Loss-of-function mutants of Myd88(kra^1) have squat pupae. (Wasserman, unpublished)

Function	Gene	Mutant Genotype	Mean Axial Ratio
Receptor	Toll	Tl^{r6}	2.5 ± 0.1
Adaptor	MyD88	kra^1	2.5
	Tube	$tub^{2}/Df(3R)XM3$	2.5
S/T Kinase	Pelle	pll^{2}/pll^{7}	2.4
Inhibitor	Cactus	$cact^{PD74}/cact^{99}$	3.4
NFκB	Dorsal	dl^{1}	2.7
	Dif	dif(iso5)	3.0
	Relish	rel^{E20}	3.1
	-	wild type (Oregon R)	2.9
	-	$+ / Tl^{r6}$	2.9
	-	+/ Df (3R)XM3	3.0

Figure 1.4: Mean pupal length/width axial ratio of Toll pathway mutants (membrane-bound and intracellular components). (Wasserman, unpublished)

One phenotype caused by loss of Toll signaling is squat pupal shape

Pathway	Gene	Mutant Genotype	Mean Axial Ratio
DV Axis Formation	Pipe	pip^1/pip^2	2.9 ± 0.1
	Nudel	ndl^{046}/ndl^{169}	2.9
	Gastrulation Defective	gd^2	3.0
	Snake	-	2.9
	Easter	ea^{831Ri}/ea^1	2.8
Anti-fungal Infection	Persephone	psh^4	3.0
Anti-Gram(+) Bacterial Infection	PGRP-SA	PGR^1-SA	3.0
	Spätzle	spz^2	2.7

Figure 1.5: Mean pupal length/width axial ratio of Toll pathway mutants (extracellular components). (Wasserman, unpublished)

(Letsou, Alexander, Orth & Wasserman 1991)(Figure 1.3). This phenotype is visible in loss-of-function mutants of Toll, Myd88, Tube, Pelle and, to a lesser degree, Dorsal (Figure 1.4) . Loss-of-function mutants of Cactus have elongated pupae, consistent with its role as inhibitor of the pathway. Interestingly, the squat pupae phenotype is not observed in any of the mutants of Gastrulation defective, Snake, Persephone and PGRP-SA, indicating that Toll activation in body shape control might be novel (Figure 1.5).

The squat pupal phenotype of Toll mutants might reflect the disruption of proper muscle innervation and muscle patterning. Toll is known to be required for motorneuron and muscle specification during late embryogenesis (Halfon, Hashimoto & Keshishian 1995). The growth cone of RP3 motoneuron in Drosophila has been shown to recognize Toll expressing muscle cells as a negative repelling signal during pathfinding (Rose & Chiba 1999, Rose, Zhu, Kose, Hoang, Cho & Chiba 1997). Together with the attractive FasIII expression, Toll expression directs the proper innervation of the musculature. Additionally, the NF-κB factor Dorsal is found to accumulate in the neuromuscular junctions through larval stages

(Bolatto, Chifflet, Megighian & Cantera 2003) , which is in line with the fact that Dorsal mutants but not Dif or Relish mutants have squat phenotypes.

Calmodulin

The single Calmodulin gene in Drosophila is at the center of the calcium signaling network. Zygotic null mutation of Calmodulin leads to post-embryonic lethality. However, the V91G Cam^7 allele of Calmodulin delays the lethality to late pupal stage. The resulting pupae are described as 'short and indented' with the overall shape of the pupae clearly altered. The alteration is reported to occur during pupariation when the Cam^7 mutants decrease their body length to 50%, compared to 33% for the control group. A close examination reveals difficulties for the mutants to relax after each muscle contraction and extensive muscle degeneration is observed by the third instar larval stage. Therefore, hypercontraction of the body wall muscle is presented as the cause of short pupae in Cam^7 mutants (Wang, Sullivan & Beckingham 2003).

Mlp84B

The muscle LIM protein 84B (Mlp84B) belongs to the Cysteine-rich protein (CRP) family of LIM proteins. Like their mammalian counterparts, the Drosophila CRP LIM proteins play important roles in muscle differentiation. In fly embryos, expression of Mlp84B is restricted to somatic, visceral and pharyngeal muscles. Immunofluorescent studies have further localized the protein to muscle attachment sites and the periphery of Z-bands of striated muscle (Stronach, Renfranz, Lilly & Beckerle 1999, Stronach, Siegrist & Beckerle 1996). Mlp84B is required for muscle contraction and a failure of muscles to contract properly during pupariation is thought explain the elongated pupal shape of Mlp84B mutants (K. A. Clark, personal communication).

In summary, multiple lines of evidence support the role of somatic musculature in maintaining body shape. The proper contraction and relaxation of

muscles can affect the overall body dimensions, especially during processes of pupariation.

1.2.2 Shape changes associated with cuticle defects

Similar to worms, Drosophila relies on a cuticular exoskeleton. Unlike in worms, there are no cuticle collagens in Drosophila (Myllyharju & Kivirikko 2004). The Drosophila cuticle is a chitin-based matrix with a highly organized, stratified structure (see next section for details). Mutations that disrupt chitin synthesis or modification often result in deformed body shape in addition to cuticle defects.

Chitin synthesis proteins

Chitin is a long, unbranched molecule consisting entirely of N-acetyl-D-glucosamine units linked by β -1,4 bonds. After synthesis and secretion, chitin polymers spontaneously assemble into microfibrils. Chitin polymers occur in three different crystalline forms - α, β and γ chitin. Anthropod cuticles are found to be mainly composed of α chitin, which has anitparallel orientation and contains high level of hydrogen bonds (Merzendorfer 2006).

Biosynthesis of chitin requires a series of enzymes (Csikos, Molnar, Borhegyi, Talian & Sass 1999) . Starting from D-glucosamine-6-phosphate, enzymes glucosamine-6-phosphate N-acetyltransferase, phosphoacetylglucosamine mutase, UDP-N-acetylglucosamine diphosphorylase and chitin synthase act sequentially to generate chitin (Figure 1.6). Among these, UDP-N-acetylglucosamine diphosphorylase and chitin synthase have been investigated for their roles in cuticle formation and tracheal development.

The Drosophila UDP-N-acetylglucosamine diphosphorylase, also called *mummy/cystic (mmy)*, is essential for not only chitin synthesis but also protein glycosylation and GPI anchor formation. Mutants of strong allele mmy^{IK63} barely have any cuticle, while mutants of weak allele mmy^{IL07} develop bloated cuticle (Tonning, Helms, Schwarz, Uv & Moussian 2006)(Figure 1.7).

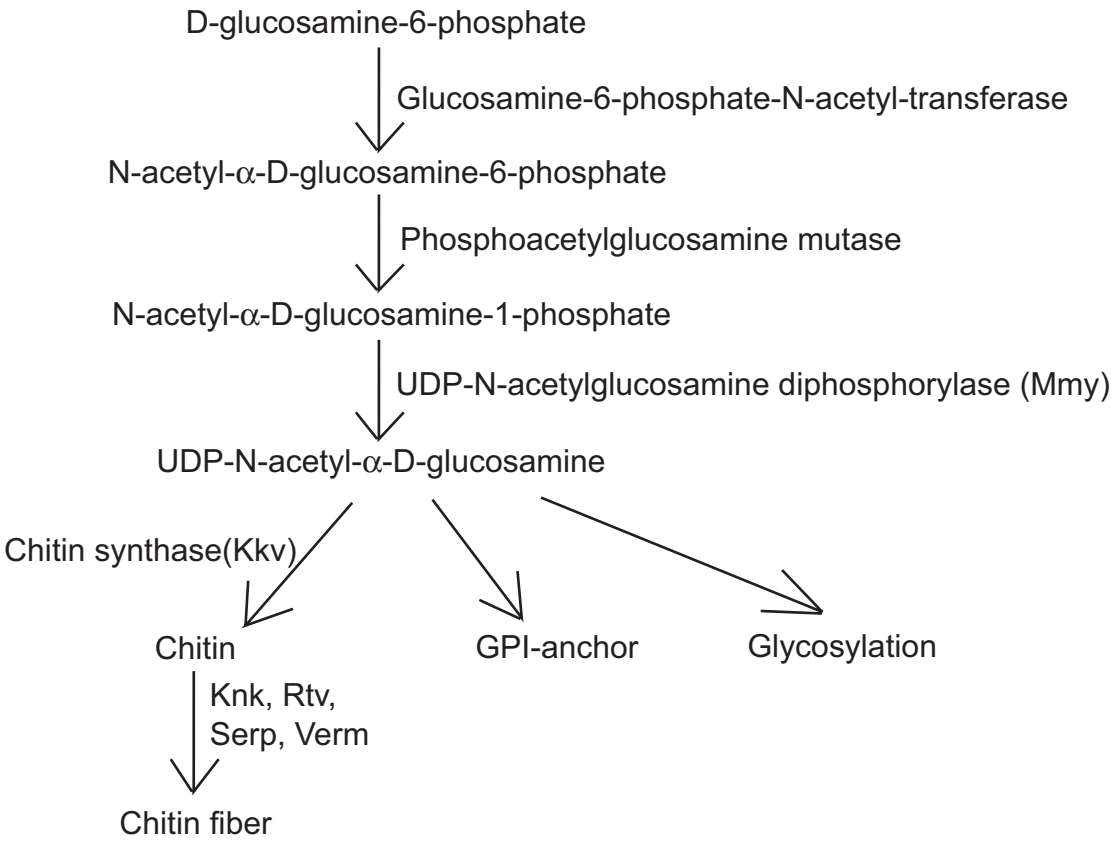


Figure 1.6: Biosynthesis of chitin and chitin fibers.

There are two genes encoding chitin synthase in Drosophila genome, namely CS1 and CS2 (Gagou, Kapsetaki, Turberg & Kafetzopoulos 2002). The transmembrane protein CS1, which is also called *krotzkopf verkehrt (kkv)*, is responsible for chitin synthesis for epidermal cuticle formation(Moussian, Schwarz, Bartoszewski & Nusslein-Volhard 2005). Cuticle of mature *kkv* mutant embryos appears wider but not longer than wild type embryos (Figure 1.7) .

Chitin modifying proteins

Besides chitin biosynthesis pathway, several proteins have been shown to modify chitin and facilitate chitin organization and accumulation. Such proteins include the membrane-anchored extracellular protein *retroactive (rtv)* (Moussian,

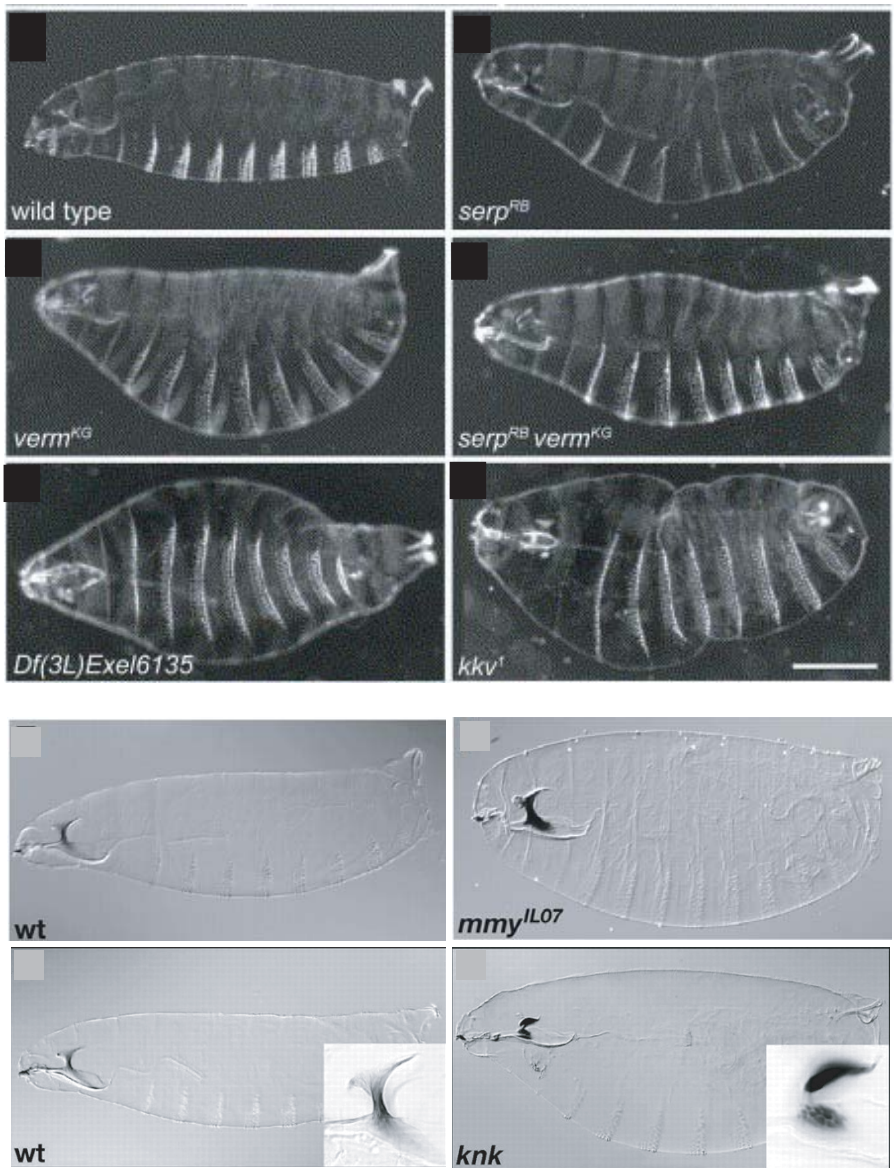


Figure 1.7: Mutations that disrupt chitin synthesis or assembly result in bloated or deformed embryos. Adapted from the following articles: Luschnig S. et al. Curr Biol. 2006 Jan 24;16(2):186-94. Tonning A. et al.Development. 2006 Jan;133(2):331-41. Moussian B. et al. Development. 2006 Jan;133(1):163-71.

Soding, Schwarz & Nusslein-Volhard 2005), GPI-linked protein *knickkopf (knk)* and extracellular matrix proteins *serpentine (serp)* and *vermiform (verm)* (Luschnig, Batz, Armbruster & Krasnow 2006). Lack of any of the four proteins disrupts chitin filament assembly and results in bloated and deformed cuticle in mature

embryos (Figure 1.7).

1.3 Chitin-based extracellular matrix in Drosophila

Chitin is one of the most abundant polymers in nature. In Drosophila, chitin-based extracellular matrix performs crucial functions in several aspects of physiology and development. In particular, structures and functions of chitin-based extracellular matrix in the larval cuticle, the tracheal system and the gut are briefly introduced in this section.

1.3.1 Larval cuticle

Drosophila larval cuticle is secreted by the hypodermis at three points in the life cycle. This cuticle exoskeleton is responsible for body architecture, locomotion and protection. Deposition of the first instar larval cuticle starts as early as embryonic stage 15. The epidermal cells secrete cuticle at their apical face in layers (Payre 2004) (Figure 1.8). A mature cuticle comprises three layers - the envelope layer (cuticulin), the epicuticle and the procuticle. The outermost envelope layer is the first to be secreted by the epidermis. This waterproof coat provides the basis for sequential deposition and assembly of the protein-rich epicuticle. The number of proteins in the epicuticle is estimated to be in the hundreds. Deposition of epicuticle is followed by the secretion of the chitin and protein based procuticle. The procuticle is attached to the epidermis at multiple anchor sites. The chitin in procuticle is organized into stacks of single-layered chitin fibers and the resulting network is sometimes referred as the chitin laminae. The laminae provide the cuticle its tension and rigidity by associating with various cuticular proteins. It has been suggested that both the chitin and the cuticular proteins are capable of self-assembly.

Insect cuticular chitins vary very little among species; therefore the diversifications of cuticle properties mainly reflect the differences of the cuticular

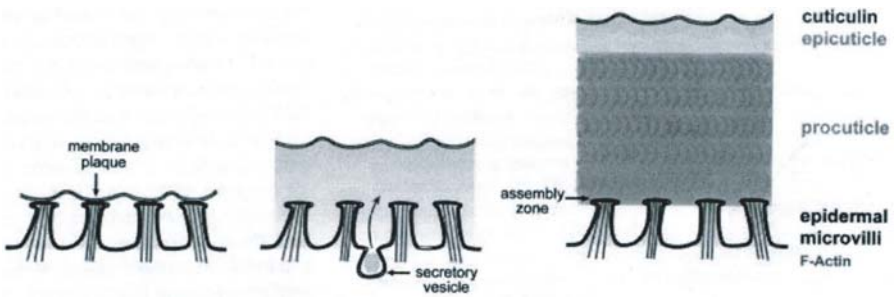


Figure 1.8: Larval cuticle is secreted by the epidermis in layers. Adapted from review article: Payre F. Int.J.Dev.Biol.48: 207-215.2004. . . . .

proteins. Insect cuticular proteins are extensively crosslinked through sclerotization and melanization (together often referred to as cuticle tanning) mediated by catecholamine derivatives (Andersen 2004, Kayser & Palivan 2006, Ricketts & Sugumaran 1994, Suderman, Dittmer, Kanost & Kramer 2006). Several enzymes involved in catecholamine metabolism have been studied for their roles in cuticle tanning, including tyrosine hydroxylase, Dopa decarboxylase and phenoloxidases (Eveleth, Gietz, Spencer, Nargang, Hodgetts & Marsh 1986, Neckameyer & White 1993, Pentz, Black & Wright 1990, Pentz, Black & Wright 1986, Sugumaran, Nellaiappan & Valivittan 2000, Walter, Black, Afshar, Kermabon, Wright & Biessmann 1991) . In recent years, transglutaminase-dependent cross-linking of cuticular proteins has also been investigated in the horseshoe crab (Iijima, Hashimoto, Matsuda, Nagai, Yamano, Ichi, Osaki & Kawabata 2005) . Transglutaminases are a family of enzymes that catalyze the formation of epsilon-(gamma-glutamyl) lysine isopeptide linkages. The Drosophila genome contains one transglutaminase, CG7356, which has a documented role in larval hemolymph clotting reactions (Karlsson, Korayem, Scherfer, Loseva, Dushay & Theopold 2004). Although expression of CG7356 is detected in the dorsal and ventral epidermis during late embryonic stages, there is no direct evidence that Drosophila transglutaminase functions in catalyzing cuticular protein crosslinking.

Among the most characteristic features of the Drosophila larval cuticle are the non-sensory extensions secreted by the epidermal cells (Figure 1.9 A).

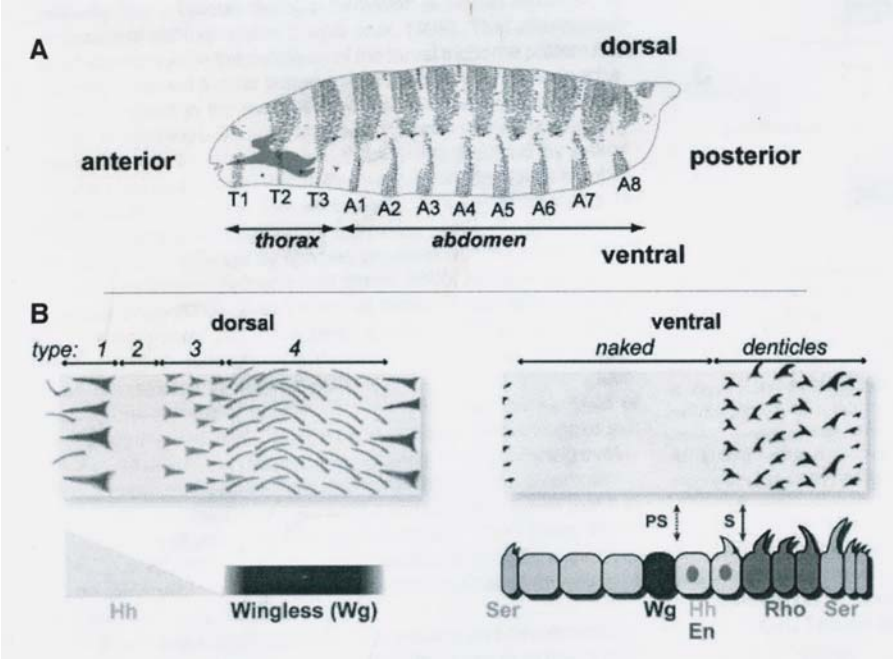


Figure 1.9: The stereotype arrangements of larval cuticle extensions reflect differences among the underlying epidermis cells. A. Larval cuticle extensions, including dorsal hairs and ventral denticles, have stereotyped arrangements. B. Detail of the dorsal and ventral cuticle corresponding to the fourth abdominal segment(A4). Signaling molecules that specify the fates of the underlying epidermal cells are summarized. Adapted from review article: Payre F. Int.J.Dev.Biol.48: 207-215.2004. . . . .

These extensions are arranged into a stereotyped array along the dorsoventral axis and the anteroposterior axis. Three regions along the dorsoventral axis are clearly visible in the preparations of the first instar larval cuticles – a dorsal region covered with fine and dense hairs; a lateral region displaying bulges at longitudinal muscle attachment sites; and ventral region marked by pigmented denticle belts. Along the anteroposterior axis, the cuticle extensions exhibit segment-specific patterns and polarities (Dickinson & Thatcher 1997).

The patterning of the larval cuticle reflects reproducible differences among the underlying epidermal cells (Payre 2004) (Figure 1.9 B). A number of signaling pathways and molecules are required to determine the epidermal cell fates (Angelats, Gallet, Therond, Fasano & Kerridge 2002, Gritzan, Hatini & DiNardo

1999, Payre, Vincent & Carreno 1999, Wesley 1999, Walters, Munoz, Paaby & Dinardo 2005). On the dorsal side, a hedgehog gradient together with the Wingless pathway helps to specify the four types of dorsal hairs within each segment. On the ventral side, d-EGF-receptor pathway activation promotes denticle formation. Additional signalling molecules such as Rhomboid and Serrate/Notch form adjacent stripes which define the denticle rows. Furthermore, in the ventral epidermis, Wingless dictates the formation of naked cuticle by repressing the transcription factor shavenbaby. As a key trichome-forming factor, shavenbaby controls epidermal cell form by transcriptionally regulating different classes of effectors which regulate the actin cytoskeleton (Dai, Schonbaum, Degenstein, Bai, Mahowald & Fuchs 1998, Chanut-Delalande, Fernandes, Roch, Payre & Plaza 2006) . The epidermal cells are highly polarized along the apical-basal axis. F-actin reorganization and actin bundling prepare the cells for trichome formation and direct the polarities of the trichomes (Dickinson & Thatcher 1997).

1.3.2 Tracheal luminal matrix and tracheal cuticle

Morphogenesis of branched tubular epithelia is of widespread interest and it includes the formation of mammalian lung and circulatory system. The Drosophila tracheal system is the best-studied example of tubulogenesis (Cabernard, Neumann & Affolter 2004, Ghabrial, Luschnig, Metzstein & Krasnow 2003, Petit, Ribeiro, Ebner & Affolter 2002, Affolter & Shilo 2000). The tracheal tree develops from 10 segmental tracheal placodes on each side of the embryo, extends from the 2nd thoracic segment to the 8th abdominal segment. After tracheal cell invagination and formation of transversal tubes, the fragments fuse, at germ band shortening, to form a continuous tree that runs longitudinally through the embryo. The tracheal cells further elongate to form the main tracheal trunk. Transversal branches of each segment derive from the main trunk and further branching continues to occur from stage 15 onwards. The development of the tracheal tree is driven by cell size and shape change as well as cell migration and rearrangement,

as no mitoses seem to occur after tracheal pits obliteration. Several pathways have been shown to function in directing such movements, including the FGF, *Dpp*, and Wg/WNT signaling pathways (Stahl, Schuh & Adryan 2006, Swanson & Beitel 2006, Myat, Lightfoot, Wang & Andrew 2005, Llimargas & Lawrence 2001).

Recent progresses in tracheal morphogenesis demonstrated that a transient lumenal chitin matrix is required for the generation of tracheal tubes with proper size and length (Tonning, Hemphala, Tang, Nannmark, Samakovlis & Uv 2005, Uv, Cantera & Samakovlis 2003). In *kkv* and *mmy* mutants where chitin synthesis is abolished, dilated tracheal tubes and cysts in multicellular branches were observed along the entire dorsal trunk by the end of embryogenesis (Devine, Lubarsky, Shaw, Luschnig, Messina & Krasnow 2005, Araujo, Aslam, Tear & Casanova 2005, Tonning, Hemphala, Tang, Nannmark, Samakovlis & Uv 2005) (Figure 1.10). In addition, the apical distribution of β-H-spectrin is irregular in *kkv* mutants and the tracheal nuclei in the *mmy* mutants are very poorly aligned comparing to the wild type controls (Tonning, Hemphala, Tang, Nannmark, Samakovlis & Uv 2005). Tracheal formation defects were also seen with mutants of chitin modifying enzymes and chitin filament assembly proteins. In *knk* and *rtv* mutant embryos, the tracheal tubes dilate excessively with cystic appearance, resulting in elongated and convoluted dorsal trunks (Moussian, Tang, Tonning, Helms, Schwarz, Nusslein-Volhard & Uv 2006) (Figure 1.10). In *serp* and *verm* mutant embryos, elongated and tortuous dorsal trunks were detected from stage 15 onwards (Luschnig, Batz, Armbruster & Krasnow 2006) (Figure 1.10). Consistent with this, the presence of an elongated apical tracheal surface was confirmed in these mutants.

Surprisingly, Drosophila septate junctions have been found to play a crucial role in regulating tracheal matrix secretion (Wang, Jayaram, Hemphala, Senti, Tsarouhas, Jin & Samakovlis 2006, Wu & Beitel 2004). Mutants with septate junctions disrupted, as in the case of Na+/K+ ATPase, *megatrachea* and *sinuous*, share the dilated and/or elongated dorsal trunk phenotype with the chitin matrix

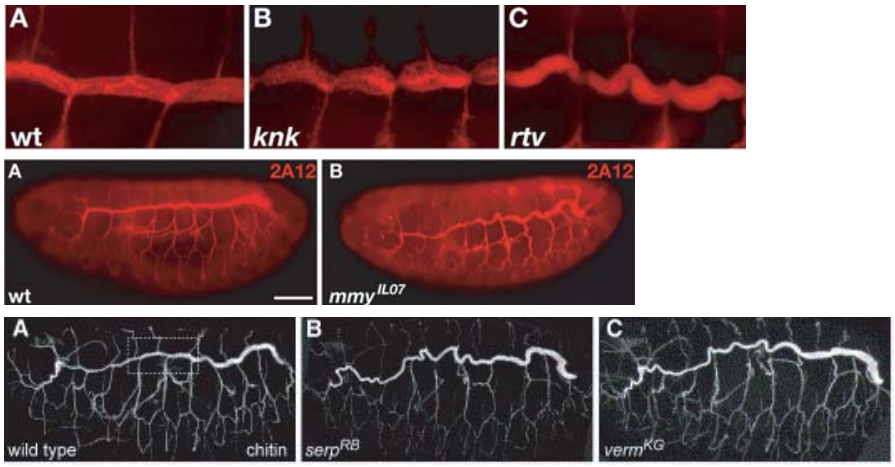


Figure 1.10: Mutations that disrupt chitin synthesis or assembly result in tortuous or cystic tracheal tubes. The developing tracheal lumen of wild-type, *knk, rtv* and mmy^{IL07} mutant embryos was visualized with the lumen-specific antibody 2A12. The lumen of *serp and verm* mutant embryos was visualized with a fluorescent chitin binding probe. Adapted from the following articles: Luschnig S. et al. Curr Biol. 2006 Jan 24;16(2):186-94. Tonning A. et al.Development. 2006 Jan;133(2):331-41. Moussian B. et al. Development. 2006 Jan;133(1):163-71.

mutants (Paul, Ternet, Salvaterra & Beitel 2003, Behr, Riedel & Schuh 2003, Wu, Schulte, Hirschi, Tepass & Beitel 2004). A recent study has shown that in flies lacking septate junctions, the structure of the tracheal lumen is disorganized and the secretion of the vermiform protein is eliminated (Luschnig, Batz, Armbruster & Krasnow 2006). This loss of vermiform secretion doesn't reflect a general block of apical secretion, since the lumenal accumulation of the 2A12 antigen and the Pio protein is unaffected in the septate junction mutants.

Before the tracheal branches become functional, a tracheal cuticle is deposited and the lumen is cleared. The tracheal cuticle is characterized by ridges projecting into the lumen (taenidial folds)(Bate & Arias 1993). These ridges function to support the lumen, while at the same time allowing them to expand and contract throughout their length. Although generally viewed as an extension of the epidermal cuticle, the detailed structure of tracheal cuticle remains elusive.

1.3.3 Gut cuticle and peritrophic matrix

The epithelia of the foregut and hindgut are ectodermal in origin, whereas the major part of the midgut epithelium rises from the endoderm. A cuticle layer lines most of the Drosophila digestive tract except the midgut. Cuticle formation in the foregut, hindgut and proventriculus resembles that for the epidermis, although the secretion of gut cuticle precedes the secretion of epidermal cuticle by one-half hour in developing embryos (Bate & Arias 1993). In both locations, the apical surface of the secreting cells form microvillae that are supported by parallel actin filaments.

The midgut of Drosophila is lined by a semi-permeable matrix - the peritrophic matrix. At least two types of peritrophic matrix have been recorded (Tellam, Wijffels & Willadsen 1999). Type 1 is transiently secreted by the midgut upon stimulation of a meal and type 2 is constantly secreted into the gut by the cardia. The peritrophic matrix is composed of chitin, proteins and proteoglycans. This porous structure acts as a molecular sieve that separates the digested food, as well as protects the underlying epithelial cells from microbial and parasite invasion. As a result, the peritrophic matrix also functions as a scaffold for proteases, peptidases, and glycosidases. The integral proteins of the peritrophic matrix are normally referred to as peritrophins. Despite the almost universal existence of the peritrophic matrix in insects, only a few peritrophins have been identified (Eisemann, Wijffels & Tellam 2001, Vuocolo, Eisemann, Pearson, Willadsen & Tellam 2001, Tellam, Vuocolo, Eisemann, Briscoe, Riding, Elvin & Pearson 2003).

1.4 Insect cuticular proteins

The current database includes hundreds known insect cuticular proteins (Magkrioti, Spyropoulos, Iconomidou, Willis & Hamodrakas 2004). These are normally categorized as cuticular proteins based on direct purification from cuticles or deduction from DNA sequence by bioinformatics approaches. In general, the

proteins that can be extracted directly from cuticles only represent a fraction of all the cuticular proteins. Therefore, many cuticular proteins were identified by the presence cuticular motifs in them.

1.4.1 Motifs of cuticular proteins

Several motifs or repeats have been frequently identified in cuticular proteins. Four of such motifs are summarized in Figure 1.11 (Andersen, Hojrup & Roepstorff 1995).

Motif Name	Description	Examples
Rebers&Riddiford consensus	G-x(8)-G-x(6)-Y-x-A-x-E-x-G-Y-x(7)-P-x(2)-P or G-x(7)-[DEN]-G-x(6)-[FY]-x-A-[DGN]-x(2,3)-G-[FY]-x--[AP]-x(6)	
Extended R&R	A stretch of 68 amino acid comprises of the core R&R and extended surrounding	DMLCP1, DMLCP3, et al.
G, L, Y motif	Short stretches of amino acid dominated by Gly, Leu and Tyr	GRP1, et al.
AAPA/V repeats	Repeats of AAPA or AAPV	DmEDG84, et al.

Figure 1.11: Motifs of cuticular proteins.

1.4.2 Structures of cuticular proteins

Despite the long history of cuticle study, the secondary and tertiary structures of cuticular proteins have only been studied in recent years (Hamodrakas,

Willis & Iconomidou 2002, Iconomidou, Chryssikos, Gionis, Willis & Hamodrakas 2001, Iconomidou, Willis & Hamodrakas 2005) . A systematic secondary structure analysis of extended R&R sequences of 8 hard cuticle proteins and 19 soft cuticle proteins was performed (Iconomidou, Willis & Hamodrakas 1999) . The R&R sequences were found to form either 3 or 4 β -pleaded sheets for soft and hard cuticle, respectively. These β -pleaded sheets were proposed to form an antiparallel beta-sheet half-barrel structure based on sequence and secondary structure homology to bovine plasma retinol binding protein (RBP). The proposed structure contains a cleft with aromatic amino acid side chains on one face of the model (Figure 1.12). Since the extended R&R consensus sequence has been shown to bind to chitin chain directly, the half-barrel model provides a good interface for protein-chitin interaction with the aromatic rings stack against the rings of chitin.

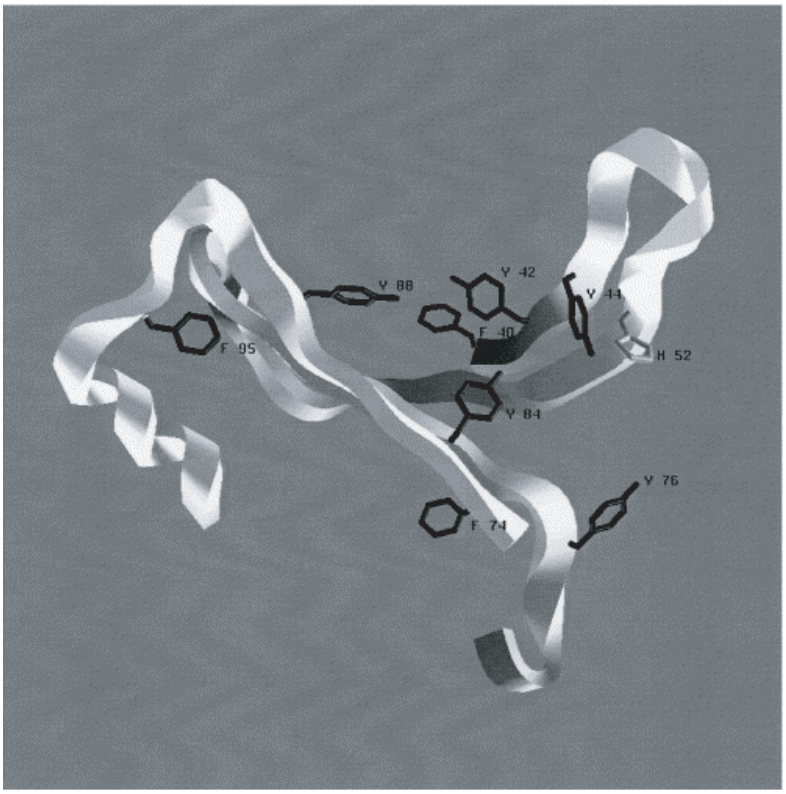


Figure 1.12: Tertiary structure of cuticle protein HCCP12, which contains an extended R&R motif.

1.4.3 Expression of cuticular enzymes and proteins

In recent years, more and more evidence points to the idea of cuticle being a dynamic, responsive matrix, rather than a simple protective shield as people depicted it. For example, expression of low molecular weight cuticular proteins has been shown to respond to environment temperature shift in tobacco hornworm (Lohmann & Riddiford 1992). Insects apparently have evolved mechanisms to regulate cuticle property in response to various stimuli by regulating expression of cuticular enzymes and proteins.

Insect molting is regulated by steroid hormone ecdysone, which is secreted into the hemolymph by prothoracic gland. Release of ecdysone into the hemolymph affects the transcription of a spectrum of genes, with some genes up-regulated and some down-regulated. In *Manduca sexta*, ecdysone has been shown to suppress the transcription of dopa decarboxylase (Hiruma, Carter & Riddiford 1995) . In Drosophila, ecdysone is thought to regulate the expression of chitin synthases. Indeed, a regulatory role for ecdysone has been observed on CS1 and CS2 transcription during metamorphosis (Gagou, Kapsetaki, Turberg & Kafetzopoulos 2002). Consistent with this, *cis*-regulatory element scanning within the upstream region of these two genes identified putative ecdysone-responsive elements (EcREs), as defined by the consensus (G/T)NTCANTNN(A/C)(A/C) and (A/G)G(G/T)T(G/C)ANTG(A/C) (A/C)(C/T)(C/T) (Antoniewski, Laval & Lepesant 1993, Luo, Amin & Voellmy 1991) .

Interestingly, recent progress in the epidermal wound-healing field has shown that transcription of sclerotization enzymes dopa decarboxylase (*Ddc*) and tyrosine hydroxylase *pale* (*ple*) is activated locally upon epidermal injury, consistent with the needs to regenerate cuticle (Mace, Pearson & McGinnis 2005, Galko & Krasnow 2004) . Furthermore, several transcription factors were identified as key factors in activating these two enzymes after wounding by targeted mutating their binding sites within the regulatory regions of *Ddc* and *ple*. In particular, transcription factor *grainy head* (*grh*), which binds to ACYGGTT(T) consensus, has been

proposed to be the key healing regulator conserved in insects, worms and mammals (Harden 2005, Mace, Pearson & McGinnis 2005, Moussian & Uv 2005, Stramer & Martin 2005) . A recent study of mammalian grainy head homologue *Grhl-3* demonstrated its role in activating the mammalian epidermal crosslinking enzyme transglutaminase1 (Ting, Caddy, Hislop, Wilanowski, Auden, Zhao, Ellis, Kaur, Uchida, Holleran, Elias, Cunningham & Jane 2005). Since *grainy head /Elf-1* is also required for late epidermal expression of *Ddc* during Drosophila embryogenesis (Bray & Kafatos 1991) , the binding sites identified in the healing study are very likely to be responsible for normal expression of *Ddc* and *ple* as well.

As expression of cuticular enzymes is investigated, transcriptional regulation of cuticular proteins is still an issue that largely remains unexplored. Nevertheless, evidence suggests that ecdysone pulses play a direct or indirect role in regulating expression of some cuticular genes, as they are required for triggering each of the major developmental transitions (Riddiford, Hiruma, Zhou & Nelson 2003). For example, expression of Drosophila pupal cuticle protein EDG84A is directly controlled by ecdysone-inducible βFTZ-F1 transcription factor during metamorphosis (Murata, Kageyama, Hirose & Ueda 1996).

1.4.4 Secretion of cuticular proteins in Drosophila

Drosophila Creb-A gene has been shown to regulate secretion in the epidermis through regulating the secretory pathway (Abrams & Andrew 2005). Transcription factor CREB (Cyclic AMP Response Element Binding protein) proteins bind to palindromic consensus sequence TGACGTCA as dimers. There are two identified CREB genes in Drosophila, namely dCreb-A and dCreb-B. Both proteins are expressed in a dynamic, but non-overlapping pattern. In particular, the Creb-A gene is expressed in salivary gland starting from embryonic stage 9, and in other tissues including trachea, a subset of neuroblasts, the preventriculus, the amnioserosa, the epidermis, the foregut and its derivatives.

Severe defects in larval cuticle were observed in loss-of-function mutants

of *Creb-A* (Andrew, Baig, Bhanot, Smolik & Henderson 1997). The mutants bear weak cuticle with frequent dorsal holes. Body length of the mutant first instar larvae is 40% of the wild type larvae. A close examination of dorsal hairs and ventral denticles indicates lateralization of dorsal and ventral cuticle. A further study suggests that dCreb-A regulates dorsal and ventral cuticle secretion by controlling the expression of secretory pathway component encoding genes (SPCGs), consistent with the fact that multiple dCreb-A consensus sites are present in all but two of the SPCGs (Abrams & Andrew 2005).

2

Results

2.1 Identification of the *TweedleD*^{1} and the *Tubby*^{1} mutations

2.1.1 Two dominant mutations – *TweedleD*^{1} and *Tubby*^{1}

Generation of *TweedleD*^{1}(*TwdlD*^{1}) mutants

As an attempt to identify novel body shape regulators in Drosophila, a former lab member performed a small-scale genetic screening to identify mutants with either squat or long pupal shape. In setting up the screening, γ-irradiated male flies were crossed to wild type females to identify dominant mutations. A total of 25,400 progenies were screened for pupal shape change. Two dominant mutations were isolated and both were confirmed to cause squat pupal shape. One of these two dominant mutations, the *TweedleD*^{1}, results in more severe squat pupal shape, and was chosen to be the focus of this study.

Phenotype of *TwdlD*^{1} mutants

Initially, to quantify body shape, we introduced the parameter pupal axial ratio, which is defined as the ratio of pupal length and pupal width. A systematic measurement has shown that all the commonly used wild type stock flies have

the same mean pupal axial ratio, 3.0 ± 0.1. As a comparison, the pupal axial ratio of the heterozygous *TwdlD*^{1} mutant flies is reduced to 2.2 ± 0.1 (Figure 2.1). Homozygous *TwdlD*^{1} mutant flies are viable and fertile. Deduction of pupal axial ratio is similar for heterozygous and homozygous flies.

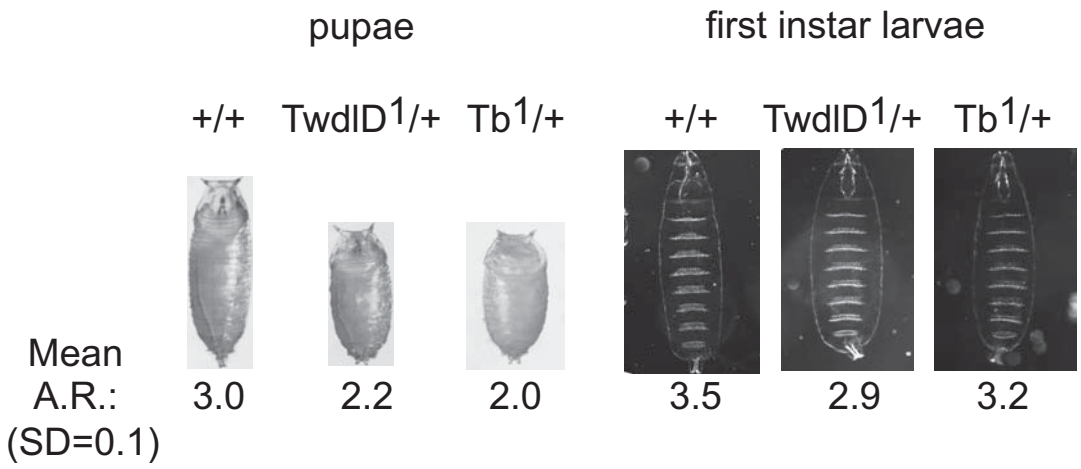


Figure 2.1: Both *TwdlD*^{1} and *Tb*^{1} mutants have reduced body axial ratio during larval and pupal stages.

Tb^1 mutants

The squat pupae phenotype of *TwdlD*^{1} resembles that of the *Tubby*^{1} (*Tb*^{1}) mutants. *Tb*^{1} mutation was first identifed in a nitrogen mustard induced mutagenesis screening. It was reported in 1980 by Craymer. Since then, *Tb*^{1} has been one of the most commonly used dominant markers in fly genetics and it is a standard marker on TM6B third chromosome balancer. Like *TwdlD*^{1} mutants, heterozygous *Tb*^{1} mutants have squat and thickset pupae with a mean axial ratio of 2.0 ± 0.1 (Figure 2.1). Furthermore, squatness is also clearly visible in *Tb*^{1} late-staged larvae and adults, making it ideal for pre-adult genotype identification. Despite its importance, the molecular nature of *Tb*^{1} mutation was never identified.

Larval phenotype of *TwdlD*^{1} and *Tb*^{1} mutants

The life cycle of Drosophila comprises four major stages - the embryo stage, the larval stage (including first instar larval stage, second instar larval stage and third instar larval stage), the pupal stage and the adult stage. As growth and shaping is a continuous process, the shape of each earlier stage affects the shape of next stage. Disruption of a general overall shape regulation mechanism is expected to have a continuous and/or accumulative effect on body shape determination. This continuousness differs it from pupal shape changes caused by muscle contraction abnormality during pupariation, like in the cases of *Cam*^{7} mutants and Mlp84B mutants. Therefore, to better characterize the *TwdlD*^{1} and *Tb*^{1} mutants, larval body shape of these two mutants were studied.

Although Drosophila larvae are not readily to be measured as pupae are, larval cuticle of the three instars can be easily prepared. Since a significant growth rate is documented for the larval stages, larvae need to be precisely aged before cuticles are prepared. In addition, to exclude maternal effect caused by the mutations, wild type females were crossed to homozygous *TwdlD*^{1} or *Tb*^{1} males. Resulting embryos were collected and aged, and first instar larvae eclosed within 45min were collected for cuticle preparation. For cuticles prepared this way, length over width larval axial ratio is used as shape parameter. With careful measurement of the flattened cuticles which have ventral views facing up, an accurate mean axial ratio for young first instar can be calculated for each genotype. Measurement of at least two independent preparations for each genotype confirmed a significant body shape change in *TwdlD*^{1} young first instar larvae and a mild body shape change in *Tb*^{1} young first instar. The mean first instar axial ratios are 3.5 $\pm$ 0.1, 2.9 $\pm$ 0.1 and 3.2 $\pm$ 0.1 for wild type larvae, *TwdlD*^{1} heterozygous larvae and *Tb*^{1} heterozygous larvae, respectively (Figure 2.1). Interestingly, although the *Tb*^{1} mutants have a more severe pupal phenotype, the axial ratio of *TwdlD*^{1} young first instar is more severely reduced, revealing a slightly shifted expression window for these two mutations.

Additive effect of *TwdlD*^{1} and *Tb*^{1} mutations

Phenotypic analogy between *TwdlD*^{1} and *Tb*^{1} mutants suggests the possibility of these two mutations being allelic to each other. Since homozygous mutants are indistinguishable from heterozygous mutants for both *TwdlD*^{1} mutations and *Tb*^{1} mutation, it suggests that double mutants of *TwdlD*^{1} and *Tb*^{1} would most likely not have a phenotype more severe than *Tb*^{1} mutants, if *TwdlD*^{1} and *Tb*^{1} are allelic. Such *TwdlD*^{1} *Tb*^{1} heterozygous mutants were generated by crossing the two homozygous stocks. The double mutants turn out to have a mean pupal axial ratio of 1.8 ± 0.1 (Figure 2.2), which represents a more severe phenotype than that of either mutant, suggesting an additive effect of these two mutations.

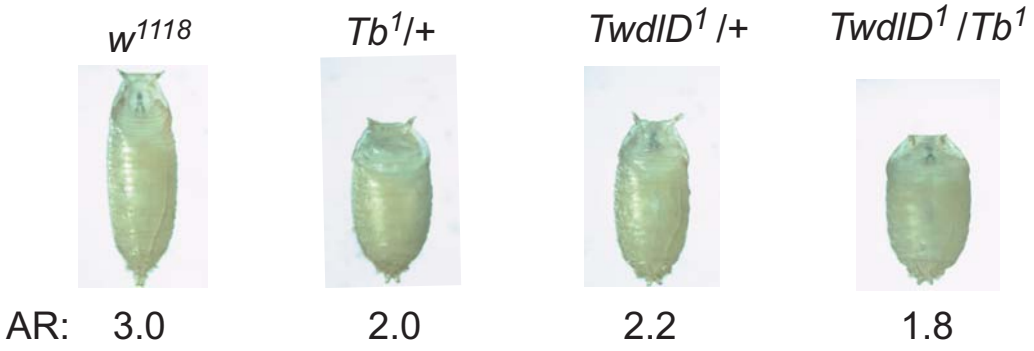


Figure 2.2: Additive effect of *TwdlD*^{1} and *Tb*^{1} mutations .

2.1.2 Mapping of *TwdlD*^{1} and *Tb*^{1}

Rough positioning of *TwdlD*^{1} and *Tb*^{1}

Tb^1 mutation was previously reported to locate at position 90.6 on the right arm of the third chromosome, which roughly corresponds to cytological band 97C (Lindsley 1973). Female meiotic recombination mapping was performed in our lab for *TwdlD*^{1}, using semi-dominant mutation *ebony* at 93C and dominant mutation *Drop* at 99B3. The recombination frequency placed *TwdlD*^{1} close to band 97C as well. It is interesting that *TwdlD*^{1} and *Tb*^{1} not only have similar

phenotypes, but also are located very close to each other within the genome.

P-element induced recombination in Drosophila males

Unlike in other organisms and in Drosophila females, meiotic recombination does not normally happen at a significant rate in Drosophila male germ line. Nevertheless, people have known for a long time that presence of *p* element in the genome significantly enhanced the frequency of recombination (Sved, Blackman, Gilchrist & Engels 1991, Sved, Eggleston & Engels 1990, Svoboda, Robson & Sved 1995). A single *p* insertion induces recombination at a rate about 1%. When two *p* elements are present at homologous sites, the rate can be promoted to as high as 20%. The induction of recombination by *p* element was observed in both sexes and in germ line and somatic tissues. It is particularly significant in Drosophila males due to the lack of background recombination.

Recombination induced by *p* element insertion normally happens at the ends of the *p* element. More than 50% of the resulting recombinants retain a mobile *p* element at the site of recombination. Furthermore, close examinations of the recombinants have revealed frequent deletions or duplications of genomic material immediately adjacent to the *p* ends where crossovers occurred (Duttaroy 2002, Preston & Engels 1996, Preston, Sved & Engels 1996) . Such deletions and duplications can involve a region up to more than 100kb.

p element induced male recombination as a fine-mapping approach

The nature of site-specific recombination induced by *p* elements makes it a good approach to map mutations on a fine level (Chen, Chu, Harms, Gergen & Strickland 1998) . To be specific, a mutation can be located to either the left or the right side of a *p* element. By using series of precisely mapped *p* elements, ideally the mutation can be located into a small region. The resolution of the mapping solely depends on the availability of *p* insertions in the region of interest. In recent years, programs like Berkeley Drosophila Genome Project (BDGP) have generated

thousands of single *p* insertion lines, with some region of the genome close to be saturated.

In practice, due to the frequent deletions and duplications associated with crossovers, to position a mutation relatively to the original *p* insertion sites can result in ambiguous interpretations. In those cases, I have found the recombinants retaining the *p* elements at the crossover sites particularly useful in detecting deletions and duplications, as inverse PCR is ready to be performed for these recombinants to map the junction sites.

Fine mapping of *TwdlD*^{1} and *Tb*^{1}

To map *TwdlD*^{1} and *Tb*^{1} mutations, triply labelled third chromosomes were generated by regular meiotic recombination in females. The genotypes of the chromosomes are *ebony TwdlD*^{1} *Drop* for *TwdlD*^{1} mutation, and *Lyra Tb*^{1} *Drop* for *Tb*^{1} mutation. Next, we brought these triply labelled chromosomes over to a *p* element containing - third chromosome by crossing the balanced stocks of the triply labeled chromosomes to a *p* element stock. The triply labelled stocks also contain a balanced Δ2-3 transposase on its second chromosome. The male progenies with both the triply labelled third chromosome over a *p* element and a transposase on its second chromosome were selected and crossed to w^{1118} females. The progenies from these crosses were scored for either *ebony* and *Drop* or *Lyra* and *Drop* to isolate recombinants. These recombinants were further crossed to wild type flies to score *TwdlD*^{1} and *Tb*^{1}. For each *p* line used, if *TwdlD*^{1} or *Tb*^{1} co-segregates with the left side marker *ebony* or *Lyra*, simple interpretation places the mutation to the left side to the original *p* insertion site. Similarly, if *TwdlD*^{1} or *Tb*^{1} co-segregates with the right side marker *Drop*, the mutation is placed to the right side of the original *p* insertion site (Figure 2.3).

In our experiment, we chose six *p* lines within the region of 97B9-D3. The *p* lines are BL12808, BL13710, BL20052, BL13022, BL10343 and BL11782, listed in the same order as their insertion sites are within the genome, with BL12808

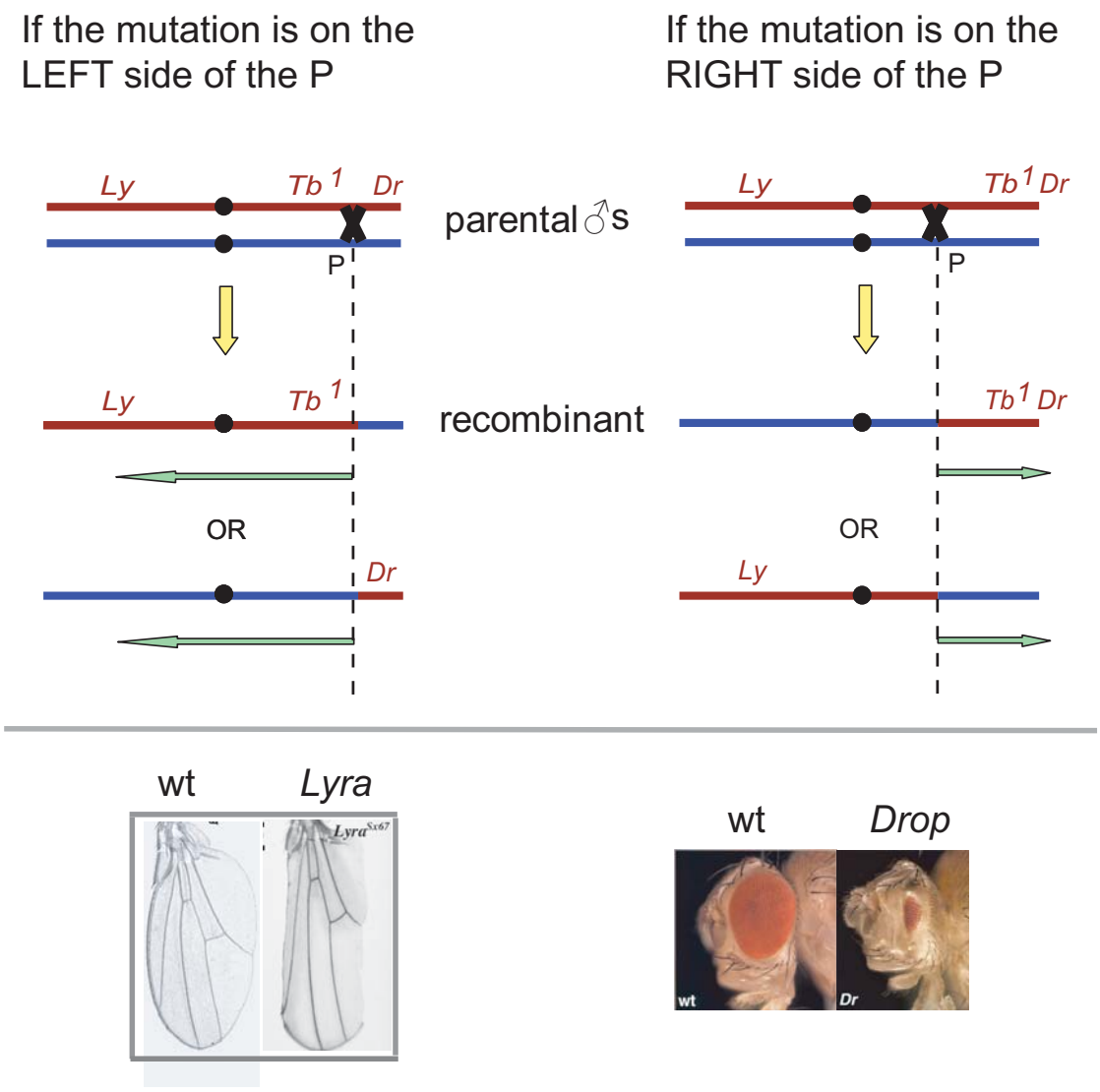


Figure 2.3: Strategy to fine map $TwdlD^1$ and Tb^1 by p element induced male recombination.

most proximal to the centromere. Consistent mapping information was generated for both $TwdlD^1$ and Tb^1 using five out of the six lines, except BL20052. Both $TwdlD^1$ and Tb^1 were located to the right of BL12808 and to the left of the other four lines (Figure 2.4). Therefore, both mutations were positioned to a region about 100kb limited by BL12808 and BL13710. Most interestingly, a deletion of 25kb was identified in the BL13710-induced recombinants for both $TwdlD^1$ and Tb^1. This 25kb deletion was further confirmed to be on the proximal side of BL13710 by

inverse PCR, which located the residual *p* elements in the recombinants. Therefore, *TwdlD*^{1} and *Tb*^{1} mutations were further confined to a region of 74kb. In BL20052 induced recombinants, *Tb*^{1} was found to co-segregate with both *Lyra* and *Drop*, which might be a result of small deletions and duplications spanning the region of *Tb*^{1} mutation. Unfortunately, in the few BL20052-induced *Lyra+ Drop-* and *Lyra- Drop+* recombinants, no retaining *p* element was detected by inverse PCR.

2.1.3 Identification of the *TwdlD*^{1} gene

Candidate genes for *TwdlD*^{1} mutation

TwdlD^1 mutation was mapped to a region of 74kb close the end of the right arm of the third chromosome. There are 15 annotated genes within this 74kb region, and they are CG5468, CG14240, CG6478, CG6447, CG6452, CG6460, CG5471, CG5476, CG31080, CG31081, CG14242, CG14243, CG14248, CG5480 and BeatVII, listed from proximal to distal along the chromosome. All of these were previously unstudied genes. Among them, there are two unique genes CG14248 and BeatVII. BeatVII belongs to the Beat immunoglobulin protein family, whose founding member beaten path (beatIa) is required for motor neuron defasciculate at their proper turning points (Fambrough & Goodman 1996, Pipes, Lin, Riley & Goodman 2001). In loss-of-function mutants of *beatIa*, motor neurons extend beyond the turning points and pass their target muscles. BeatVII is the most remote member of the beat family, and it is predicted to be a membrane anchored protein. CG14248 encodes a protein with no putative conserved domains. Interestingly, the rest of the 15 genes, although are all novel genes, seem to share some sequence similarity on protein level.

Sequencing of genomic DNA

To identify which of the above 15 genes is mutated in *TwdlD*^{1} mutants, a direct sequencing approach was taken. Since *TwdlD*^{1} mutation was initially generated in our lab from a known progenitor strain, parallele sequencing of the mutant

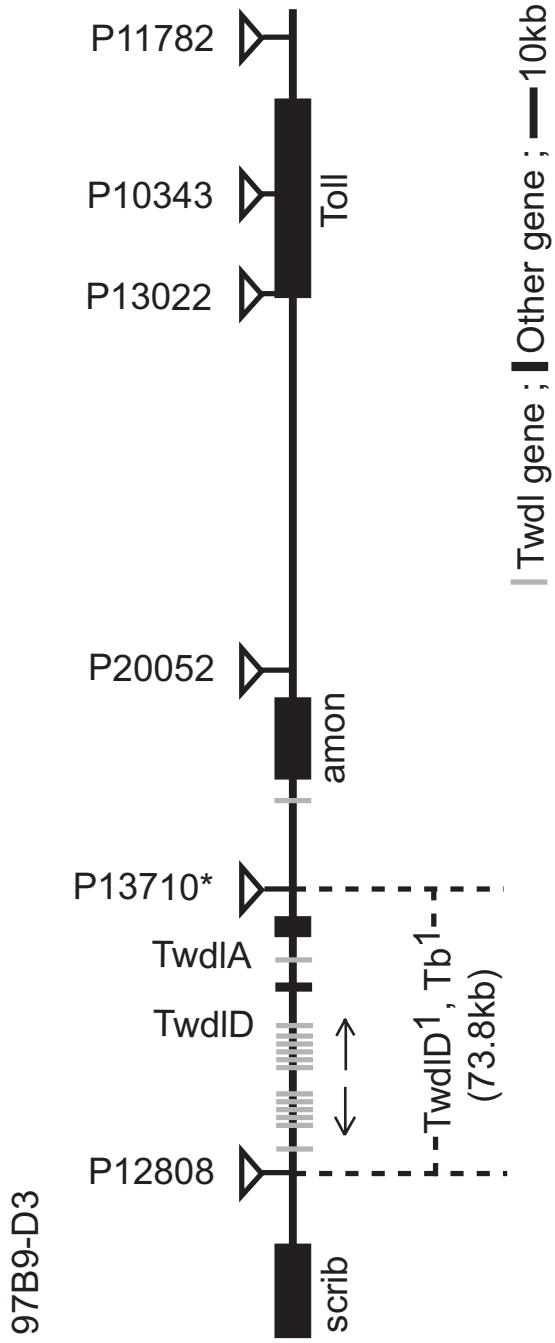


Figure 2.4: Summary of *p*-induced male recombination mapping of $TwdlD^1$ and Tb^1.

homozygotes and the progenitor strain is expected to reveal only the molecular lesion causing the $TwdlD^1$ phenotype, while excluding the potential polymorphisms

that are different from the published fly genome. However for Tb^1 mutant, which was generated decades ago from an unknown background, a 74kb region is expected to harbor a large number of polymorphisms. Consistent with this, a test sequencing of a few segments within this 74kb region from Tb^1 homozygotes revealed numerous nucleotide differences from published sequence.

Parallele sequencing of $TwdlD^1$ and progenitor was proved to be very successful. Only one difference was identified between the genomic sequences of these two strains and it is a nine nucleotides deletion within a novel gene CG14243. This nine nucleotides deletion within the coding region of CG14243 removes amino acid 173 to 175 in frame from the conceptual translate of the gene.

CG14243 transgenic flies

In order to verify the identity of the 9 nucleotide deletion in CG14243, I generated transgenic flies expressing either CG14243^{wt} or CG14243$^{\Delta 173-175}$. Genomic region of CG14243 gene including 500 bp upstream of the transcription start site was cloned from either w^{1118} flies or $TwdlD^1$ homozygotes. The genomic fragment was inserted into germ line transformation vector CaSpeR, and the resulting constructs were injected into w^{1118} flies. Four independent transgenic flies were obtained for CG14243^{wt}. Three independent transgenic lines were generated expressing CG14243$^{\Delta 173-175}$. Pupal axial ratio of each transgenic line was measured for heterozygous transgenic flies, and homozygous transgenic flies if possible.

For all the four CG14243^{wt} transgenic lines, pupal axial ratios were maintained the same as w^{1118} flies, including in the homozygous state. However, all the three CG14243$^{\Delta 173-175}$ lines have reduced pupal axial ratios in the heterozygous state. The pupal axial ratio of the CG14243$^{\Delta 173-175}$ transgenics ranges from 2.4 ± 0.1 to 2.6 ± 0.1 (Figure 2.5). The fact that single copy insertion of CG14243$^{\Delta 173-175}$ into the wild type background induces body shape change is consistent with the dominant nature of $TwdlD^1$ mutation. Furthermore, in the CG14243$^{\Delta 173-175}$ homozygous transgenic flies, the pupal axial ratios in two independent lines are fur-

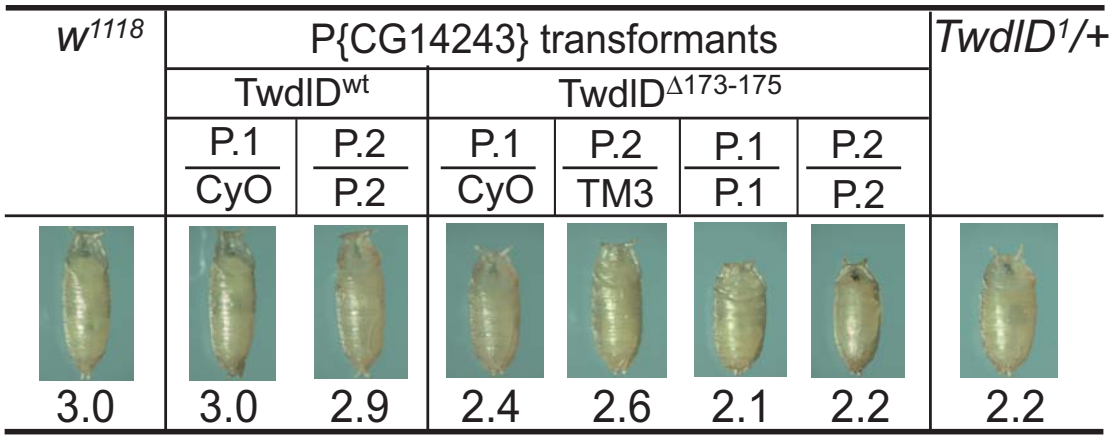


Figure 2.5: Transgenic flies expressing CG14243^{wt} or CG14243$^{\Delta 173-175}$. For each genotype, P.1 and P.2 represent two independent transgenic lines.

ther reduced to 2.2 $\pm$ 0.1, the same as that of *TwdlD*^{1} mutants.

The pupal axial ratios of CG14243$^{\Delta 173-175}$ transgenic flies indicate that the wild type CG14243 genes in the w^{1118} background might influence the transgenic CG14243$^{\Delta 173-175}$ alleles, resulting only the most severe phenotype when the ratio of the mutant to the wild type gene is 1:1 or greater. Alternatively, the 500 bp upstream promotor used for the transgenic alleles might be incomplete, with elements missing for full-strength expression of the gene. In the later scenario, since a full replication of the phenotype is achieved with two copies of the transgenic alleles, the expression pattern of the transgenic CG14243$^{\Delta 173-175}$ allele is expected to mimic that of the allele in *TwdlD*^{1} mutants.

The results of transgenic flies confirmed gene CG14243 as the TwdlD gene, and CG14243$^{\Delta 173-175}$ allele is responsible for causing the *TwdlD*^{1} phenotype.

2.1.4 The Tweedle (Twdl) protein family

The Twdl family proteins in Drosophila genome

A close examination of Drosophila genome reveals 26 genes encode proteins homologous to TwdlD. 12 of these 26 genes are located within the sequenced 74kb region, confirming the initial observation of sequence similarity among pro-

tein products of candidate genes. Together with TwdlD, these 27 proteins form a new protein family, which we named the Tweedle (Twdl) family. Genes encoding the Twdl family proteins form three gene clusters in the Drosophila genome. The major cluster 97C contains 14 Twdl genes, including the 13 genes in the 74kb region and gene CG14250. The minor clusters 15A3 and 82A1 each contains 4 genes. The rest of the family members CG14534, CG8986, CG14254, CG5812 and CG4060 are distributed separately in the genome (Figure 2.6).

Overall sequence identity of these homologues to TwdlD protein ranges from 26% to 54%. With a couple of exceptions, the genes in each cluster appear to be closer to each other than to genes in other locations. Particularly, the genes in 97C arrange into two sub-groups with genes in each sub-group are transcribed from the same DNA strand, clearly indicating gene expansion by duplication.

Twdl homologues in other insects

A broader search of Twdl homologues in other organisms identified homologues only in insect species. No mammalian or crustacean Twdl homologues were discovered. Nevertheless, there are more than two Twdl homologues in every insect species we examined, including three Drosophila species (*Drosophila simulans*, *Drosophila yakuba*, *Drosophila pseudoobscura*), two mosquitos (*Anopheles gambiae*, *Aedes aegypti*), the silk worm *Bombyx mori*, the honey bee *Apis mellifera* and the red flour beetle *Tribolium castaneum*.

Alignment of Twdl family proteins

Alignment of protein sequences from *Drosophila, Anopheles, Aedes, Bombyx, Apis, and Tribolium* revealed several well-conserved blocks (Figure 2.7) . None contains a previously described motif. The positions of highly conserved amino acids within these blocks strongly suggest the presence of an internal repeat structure in each family member. In particular, blocks I and III contain a motif of the form $KX_{2-3}YV$ (where X_{2-3} represents two or three nonconserved amino acids),

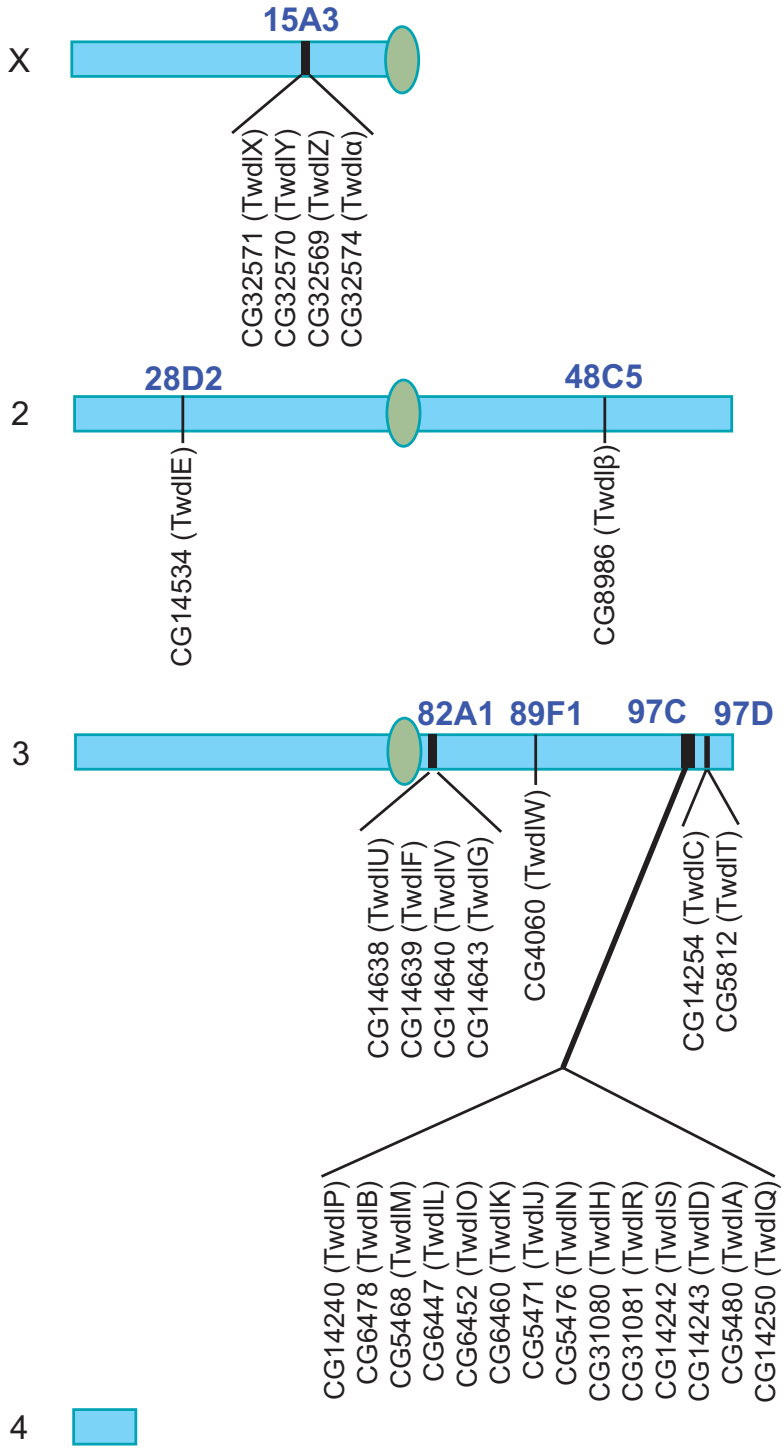


Figure 2.6: Distribution of Twdl family genes in Drosophila genome.

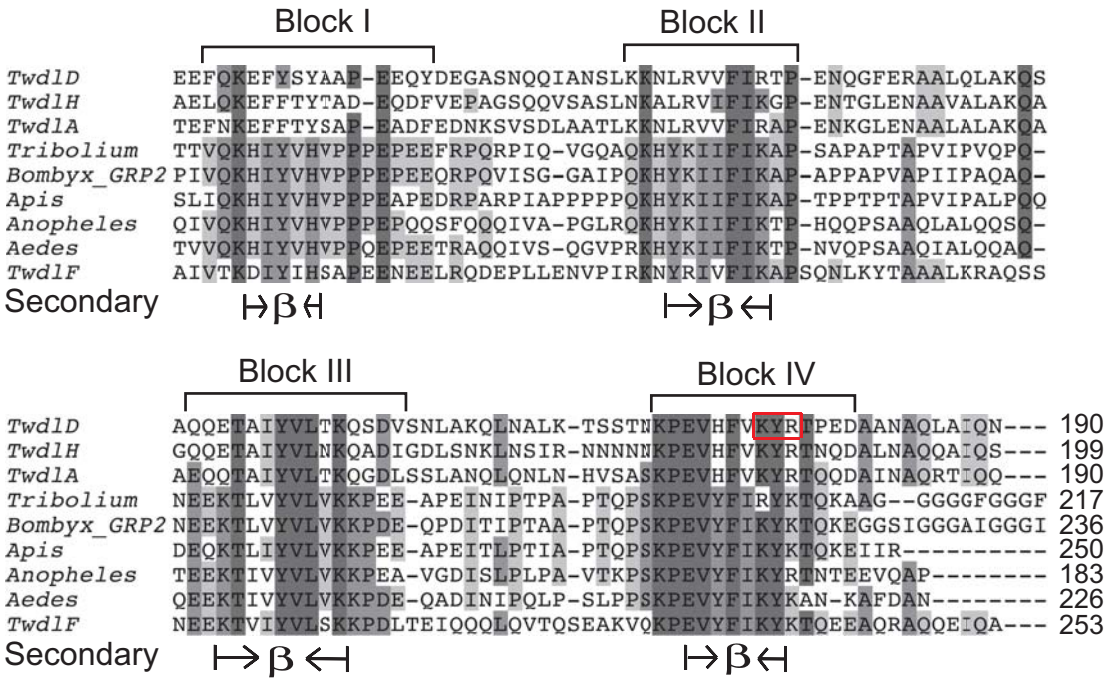


Figure 2.7: Alignment of insect Twdl homologues. The three amino acids that are deleted in CG14243$^{\Delta 173-175}$ are boxed.

whereas blocks II and IV contain a $KX_{4-5}FIK$ motif. The region most conserved among all the family members is that defined by an extended motif that spans the conserved block III and IV: $YVLX_{20-23}KPEVyFiKY(R/K)t$, where lower case letters represent less strict conservation. We regard this sequence as the signature motif for all Twdl family proteins. Most strikingly, the nine nucleotide deletion in *TwdlD*^{1} mutants eliminates the tripeptide KYR at positions 173-175 in the TwdlD protein and disrupts the highly conserved block IV.

2.1.5 Identification of *Tb*^{1} gene

Strategy to pinpoint the *Tb*^{1} gene

As stated early, *p* element induced male recombination mapping placed *Tb*^{1} to the same 74kb region as *TwdlD*^{1}. But a genomic sequencing approach is not likely to identify the mutated gene due to high frequency of polymorphisms in the strain. Up to this point, my study on *TwdlD*^{1} mutation has lead to the discovery

of a novel protein family, the Twdl family. 13 out of the 15 genes within the 74kb region were shown to encode Twdl family proteins. Taken together the phenotypic similarities between the two dominant mutations $TwdlD^1$ and Tb^1, it is very likely that a Twdl gene is also mutated in Tb^1 mutants. Furthermore, the nature of the mutation might resemble that of the $TwdlD^1$ mutation in gene CG14243.

Targeted sequencing of Twdl coding regions

Based on the above rationale, I sequenced only the coding regions of the 13 Twdl genes within the 74kb region of Tb^1 homozygotes. The DNA sequences were translated into amino acid sequences and the resulting Twdl proteins from Tb^1 mutants were compared against the wild type Twdl proteins to look for differences similar to that between $TwdlD^1$ and TwdlD proteins. To be specific, a deletion that also disrupts the highly conserved amino acid block IV will very likely to be the Tb^1 mutation. One such deletion was indeed identified through this process. It is an in frame deletion that removes amino acid 167-190 from TwdlA protein. This 24 amino acid deletion removes the entire block IV and the most of the linker region between block III and block IV.

TwdlA transgenic flies

In order to verify that this mutation of TwdlA is responsible for the Tb^1 phenotype, I turned to transgenic flies again. Genomic region of TwdlA gene including 1000 bp upstream the transcription start site was cloned from either w^{1118} flies or Tb^1 homozygotes. The genomic fragment was inserted into germ line transformation vector CaSpeR, and the resulting constructs were injected into w^{1118} flies. Five independent transgenic flies were obtained for TwdlA^{wt}. Ten independent transgenic lines were generated expressing TwdlA$^{\Delta 167-190}$ (TwdlA^{Tb}). Pupal axial ratio of each transgenic line was measured for heterozygous individuals.

For all the five TwdlA^{wt} transgenic lines, pupal axial ratios were maintained the same 3.0 ± 0.1 as w^{1118} flies. However, all the ten TwdlA$^{\Delta 167-190}$ lines

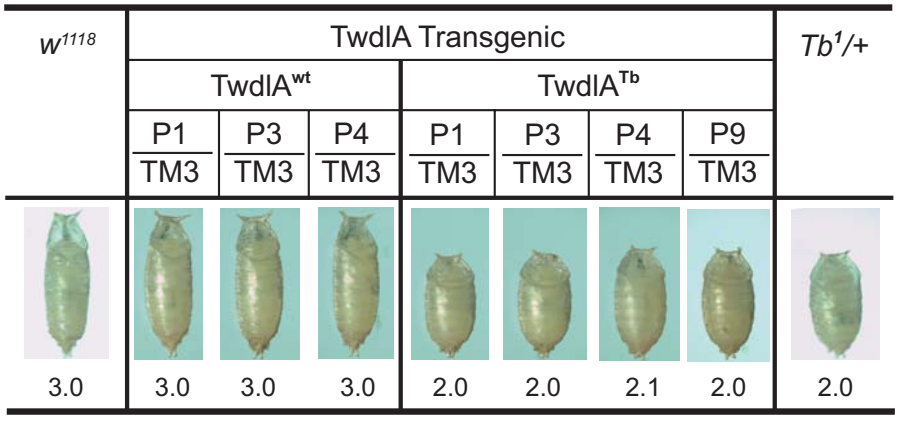


Figure 2.8: Transgenic flies expressing TwdlA^{wt} or mutated TwdlA^{Tb}.

have reduced pupal axial ratios in the heterozygous state. The pupal axial ratio of nine TwdlA$^{\Delta 167-190}$ lines is reduced to 2.0 $\pm$ 0.1, the same as Tb^1 mutants. One TwdlA$^{\Delta 167-190}$ transgenic line, line 4, has a mean pupal axial ratio of 2.1 $\pm$ 0.1, which is very close to Tb^1 mutants. Different from the $TwdlD^1$ transgenic flies, the TwdlA$^{\Delta 167-190}$ transgenic flies need only one copy of the mutated TwdlA allele to fully replicate the Tb^1 phenotype. This might indicate that TwdlA$^{\Delta 167-190}$ allele is a stronger inducer of body shape change. Alternatively, the 1000 bp upstream regulatory region might better recapitulate the in vivo expression of the gene, on both the expression pattern level and the expression strength level.

2.2 Functions of the Tweedle family proteins

2.2.1 Twdl family proteins are secreted proteins

Expression of Twdl proteins in Drosophila S2 cell culture

Sequence analysis recognizes a signal peptide at the N-terminus of all Twdl family proteins, but no transmembrane domain was identified. The conser-

vation of a signal peptide within Twdl family proteins indicates that these proteins might be secreted proteins.

To test this hypothesis, three Twdl genes were expressed in the Drosophila S2 cells. The three genes are TwdlD, TwdlA and TwdlJ. For TwdlD, both the TwdlD^{wt} protein and the TwdlD$^{\Delta 173-175}$ proteins were tested. Similarly, for TwdlA gene, both the TwdlA^{wt} and the TwdlA$^{\Delta 167-190}$ proteins were examined. In each case, only the coding region for each protein was amplified through RT-PCR and cloned into the pAc5.1 vector under the actin promotor. Each of the Twdl proteins encoded by these constructs contains either a FLAG tag or a V5 tag at their C-terminus. S2 cells were transfected with the resulting constructs. After one-day incubation, the S2 cells were separated from their media, and the cell lysates and the media were run on a protein immunoblot. Localization of cytoplasmic protein Cactus was used to confirm the separation of the cells from the media.

For every of the five proteins I tested, the vast majority the Twdl protein was identified in the media rather than the cells (Figure 2.9). This result clearly shows that these Twdl proteins are being secreted when expressed in the S2 cells. This applies to both the wild type proteins and the mutated TwdlD and TwdlA proteins, demonstrating that both mutations do not disrupt the secretion of the protein. Furthermore, the barely detectable level of Twdl protein in the cell lysates strongly suggests that the secretion of Twdl protein is very effective.

2.2.2 Embryonic expression of the Twdl mRNAs

Potential embryonic functions

As mentioned earlier, the *TwdlD*^{1} phenotype is already significant in the newly hatched first instar larvae, and the *Tb*^{1} mutants also have a mild first instar phenotype. There is a good reason to speculate potential functions of Twdl family proteins in embryos. Therefore, embryonic *in situ* hybridization experiment was performed for multiple genes of the Twdl family to examine their embryonic RNA expression patterns.

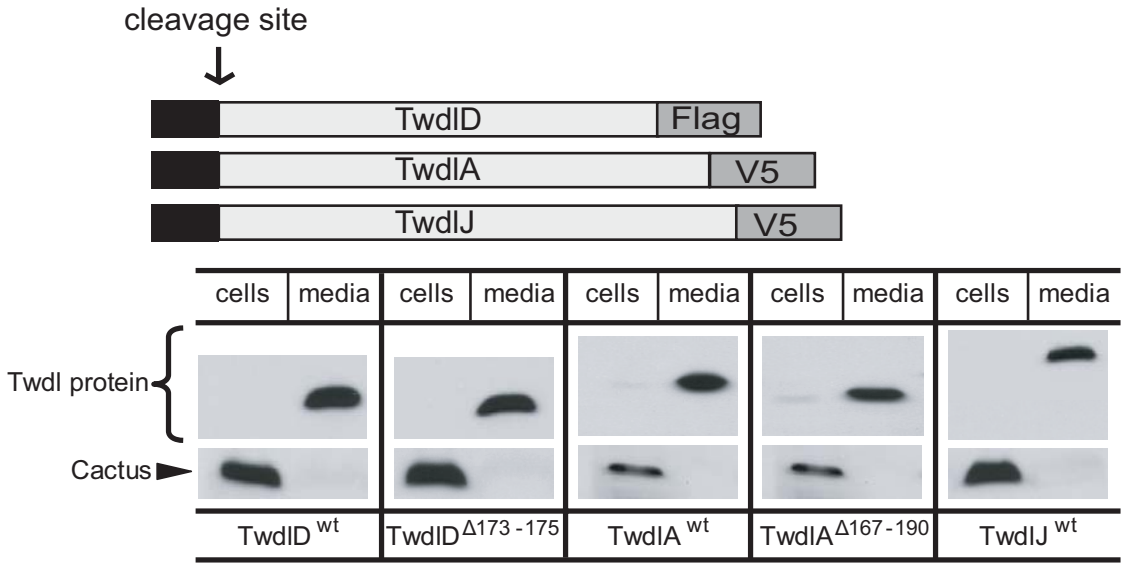


Figure 2.9: Immunoblot of cell lysates and media.

Embryonic *in situ* hybridization

Twdl genes from different gene clusters and locations were chosen for the *in situ* experiment, including TwdlA, B and D from the 97C cluster, TwdlF from cluster 82A1, and the solitary genes TwdlC (97D12), TwdlT (97D14), TwdlE(28D2), Twdlβ (48C5) and TwdlW (89F1). The 3' UTR region of each gene was cloned into a Bluescript vector(Stratagene), which contains T7 and T3 promoters transcribing genes from the two strands. The antisense probe and the sense control probe were generated for each gene. Embryos of different stages were collected at 25C for various amount of time to control the approximate age. Interestingly, for the all the genes tested, RNA expression was only detected in late-stage embryos, particularly, stage 13 to 16. The disappearance of the signal in late stage 16 embryos does not mean cease of expression. Instead, it is the result of cuticle secretion blocking *in situ* hybridization process by the end of embryogenesis.

In stage 13-16 embryos, each of the Twdl proteins examined has a gene-specific expression location (Figure 2.10). Five out of the nine Twdl RNAs were identified in the epidermis (Drosophila epidermis is also called hypodermis due to

the single-layered structure). These five genes are TwdlF, TwdlB, TwdlD, TwdlA and TwdlW. Within the epidermis, these five genes exhibit four different spatial patterns - RNA of TwdlF was expressed uniformly through embryonic epidermis; expression of TwdlB and TwdlD forms nine segmental stripes along the anterior-posterior axis and the expression further extends into the most anterior and the most posterior ends; TwdlA RNA was only very faintly expressed during the late embryonic stages, and the expression was consistently confined to the dorsal/lateral region of the epidermis; RNA of TwdlW, on the other hand, was confined to the dorsal and the ventral epidermis, not the lateral region, and furthermore, the strong expression of TwdlW co-localizes with segmental grooves.

Three out of the nine Twdl genes were strongly expressed in the foregut. These are TwdlC, TwdlE and TwdlT. On top of the fact that RNA of all three genes were strongly expressed in the foregut, expression in the posterior spiracles and mouth structure was also quite obvious for TwdlE and TwdlT. Lastly, RNA of Twdlβ was identified in the embryonic tracheal system.

A new matrix protein family?

Despite the fact that each of these genes seems to have a specific expression pattern, all studied genes were only detected in the tissues of ectodermal origin. Epidermis, foregut and hindgut epithelium, and trachea tubes are all single-layered ectodermal structures, which are sometimes viewed as a continuous epithelial structure. This epithelial structure is also a secreting organ, which is known to secrets chitin-based extracellular matrix. The expression of secreted Twdl family proteins in these locations, therefore, strongly suggests that these proteins might contribute to the chitin-based matrix system.

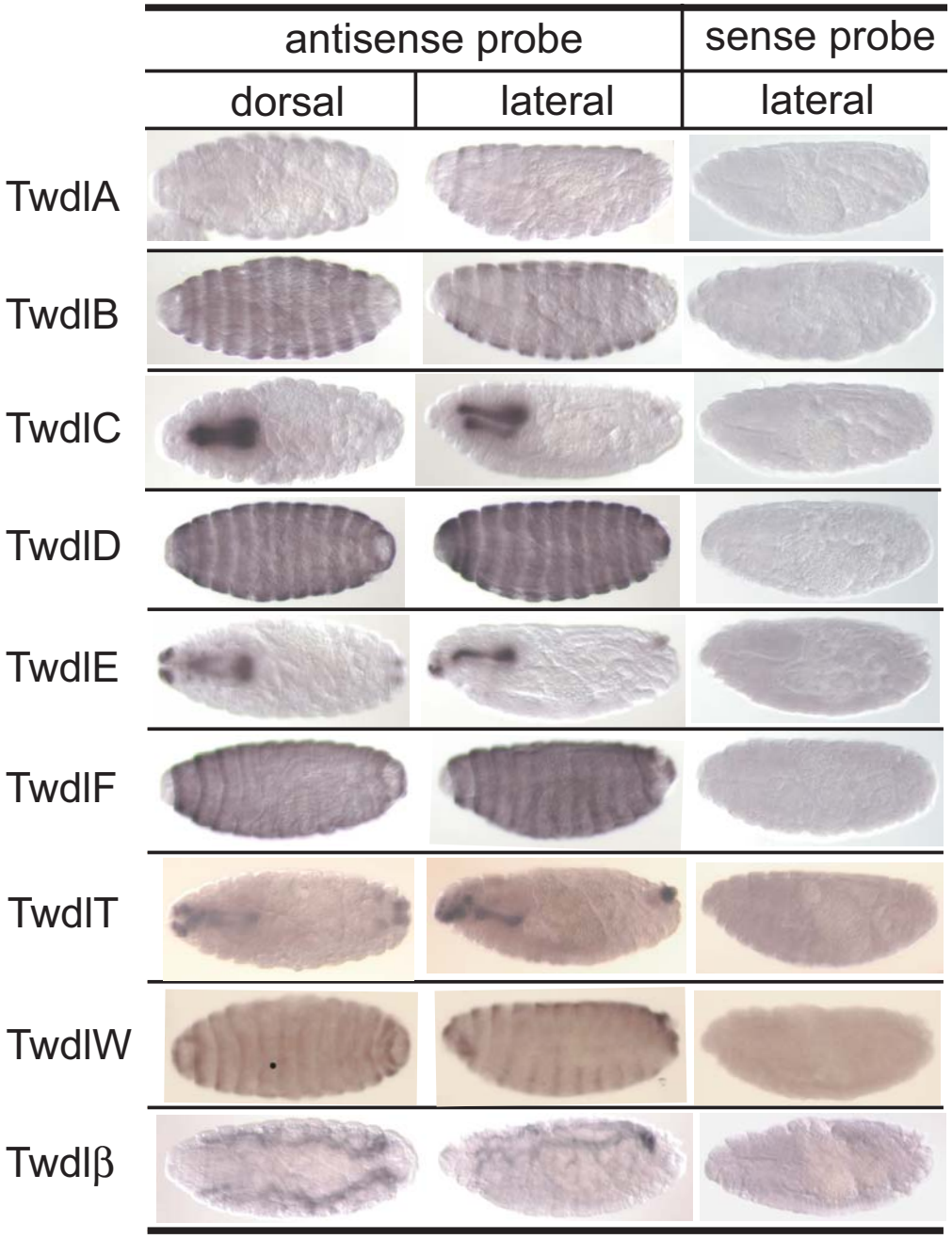


Figure 2.10: Embryonic *in situ* hybridization to examine expression of Twdl family genes.

2.2.3 Localization of Twdl proteins in larvae

Monitor Twdl protein localization by RFP-tagging

Since body shape change in $Twdl D^1$ and Tb^1 mutants is detectable through the larval stages to the adult stage (less quantitatively), TwdlD protein and TwdlA

protein are both expected to be expressed and function through these stages or at least part of these stages. Technically, RNA *in situ* hybridization cannot be easily performed for larvae. To study the expression of Twdl family proteins in post-embryonic stages, transgenic flies expressing RFP-fused-Twdl proteins were generated.

Generation of RFP transgenic flies

Four Twdl proteins were studied for their protein localization pattern in larvae, including TwdlF, TwdlD, TwdlA and TwdlH. Genomic region for each of the four genes was cloned. 500 bp upstream regulatory region was included for gene TwdlF, TwdlD and TwdlH, while 1000 bp regulatory region was used for gene TwdlA. A sequence encodes a monomeric red fluorescent protein (RFP) was inserted into the genomic region right before the stop codon of the Twdl gene, resulting in a DNA fragment encodes a Twdl protein with a C-terminal RFP fusion. This fragment was cloned into the germ line transformation vector CaSpeR, and at least three independent transgenic lines were generated for each gene.

Temporal and spatial expression patterns of Twdl proteins

Temporal expression of each Twdl-RFP protein was monitored by RFP expression. Young first instar, young second instar, young third instar and late third instar larvae were observed to identify the rough stages when the Twdl proteins are expressed. Interestingly, each of the four Twdl proteins exhibits a specific temporal expression profile (Figure 2.11).

First, TwdlF-RFP protein was detected through the three larval stages with the signal comes from the outmost fine layer covering the whole larva. Judged by its location and pattern, this layer is presumably the integument. In addition, the RFP protein was visible both in the segmental grooves and the spaces in between, indicating a uniform expression within the integument.

As a comparison, although also detected in the integument, expression

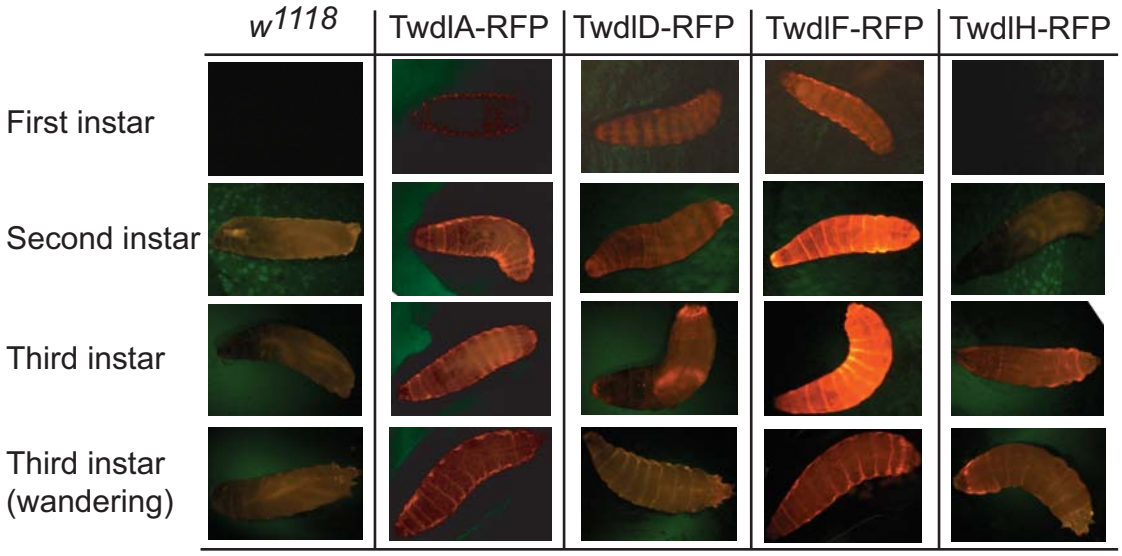


Figure 2.11: Temporal and spatial expression of Twdl-RFP fusion proteins.

of TwdlD-RFP forms segmental stripes along the anterior-posterior axis, replicating the TwdlD RNA expression pattern in late-stage embryos. These fluorescent stripes leave the regions including segmental grooves dark. Furthermore, TwdlD-RFP protein expresses during the first and the second instar larval stages and disappears during the third instar larval stage.

The expression of TwdlA-RFP protein is yet different from both TwdlF-RFP and TwdlD-RFP. TwdlA-RFP protein was barely visible in the young first instar larvae, and its expression becomes much stronger as the larvae enter the second instar larval stage and that strong expression sustains through the whole third instar larval stage. This expression appears to be confined to the integument layer. Due to the low resolution of these images, it is hard to tell whether the distribution of the protein in the integument is uniform or not.

Lastly, expression pattern of TwdlH-RFP protein was monitored. In this case, fluorescent signal is only visible in the third instar larvae. Furthermore, the protein is confined to the integument of the segments close the anterior and posterior poles.

In summary, four Twdl proteins were studied for their expression during

the larval stages, and each of the four proteins shows a protein-specific temporal and spatial expression pattern. In addition, the integument localization of secreted Twdl proteins, on top of the RNA expression patterns examined by *in situ* hybridization, strongly suggests that the Twdl proteins are novel extracellular matrix proteins, which are incorporated into chitin-based matrix structures like larval cuticle.

Twdl proteins are a new family of matrix proteins

To verify the incorporation of Twdl proteins into the cuticular structures, Twdl-RFP transgenic larvae were studied under a confocal microscope for finer structural analysis. Specifically, localization of TwdlF-RFP protein and TwdlD-RFP protein within young first instar larvae was under investigation since larvae of later stages are relatively hard to fix. In particularly, the cuticular extension dorsal hairs and ventral denticles were closely examined for RFP incorporation.

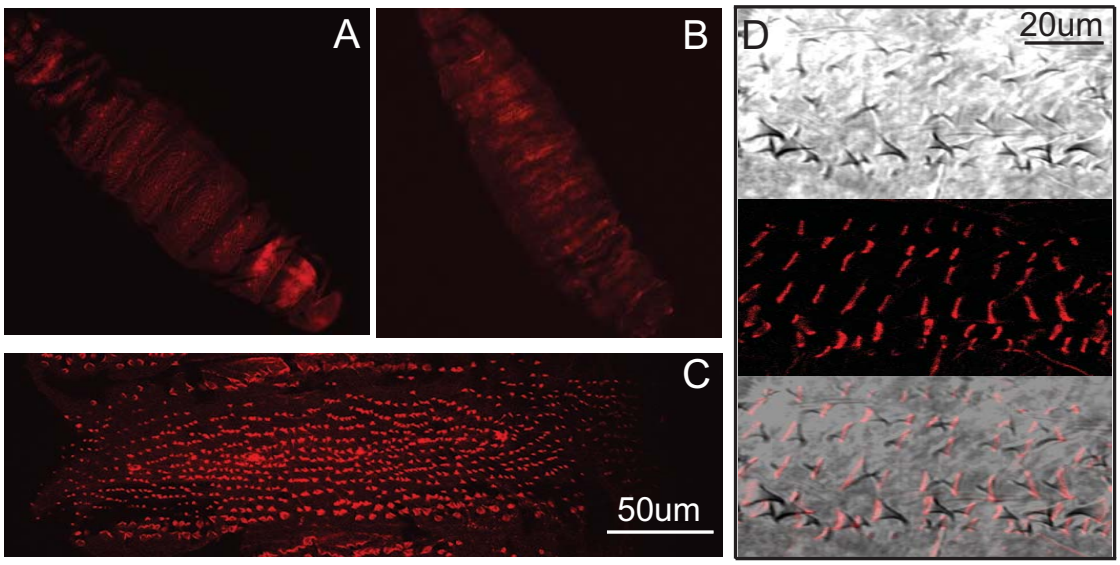


Figure 2.12: Localization of TwdlF-RFP fusion protein in young first instar larvae.

Consistent with the previous observation, first instar larvae of TwdlF-RFP transgenic flies have fluorescent cuticle as viewed from both dorsal (Figure 2.12 A)and ventral perspectives (Figure 2.12 B). On the dorsal surface, the dorsal

hairs appear brightly red as a result of TwdlF-RFP incorporation (Figure 2.12 C). On the ventral surface, the fluorescent denticles within each denticle belts appear as red bars on the fluorescent image (Figure 2.12 D). An overlay of the fluorescent image and a light image of the same area shows that the red bars correspond to the basal portion of the denticles, where most of the RFP signal comes from. This pattern of TwdlF-RFP localization within young first instar larvae is consistent with the RNA expression pattern of TwdlF gene in late embryos. The uniform expression of TwdlF RNA within the epidermis leads to the secretion and incorporation of TwdlF protein into the whole larval cuticle.

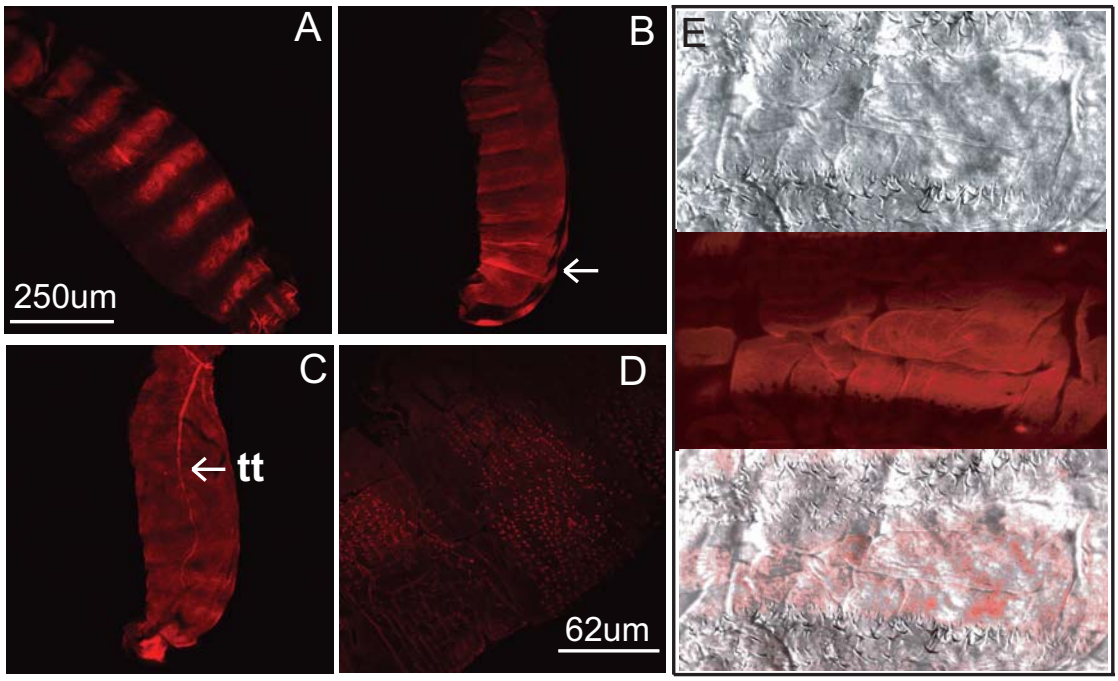


Figure 2.13: Localization of TwdlD-RFP fusion protein in young first instar larvae. tt: tracheal tree.

Close examination of TwdlD-RFP localization shows that it is also consistent with the TwdlD RNA expression in late embryos. TwdlD RNA was expressed in stripes within embryonic epidermis, while the TwdlD-RFP protein was found in stripes in first instar larval cuticle as well, on both the dorsal surface (Figure 2.13 A) and the ventral surface. These stripes can be clearly visualized at the

loci where larval body is detached from the cuticle due to the fixation procedure (Figure 2.13 B, arrow). On the dorsal side, the stripes go through the dorsal hairs, leaving the hair-free spaces dark (Figure 2.13 D). On the ventral side, the stripes are placed at the same positions along the anterior-posterior axis as on the dorsal side. Therefore, these stripes go through the spaces in between the denticle belts and leaving the denticles dark. Indeed, an overlay image of a red stripe on the ventral surface shows that the denticles of TwdlD-RFP transgenics are not fluorescent (Figure 2.13 E). In transgenics containing two copies of the TwdlD-RFP insertion, RFP fluorescent was also observed in the larval tracheal system (Figure 2.13 C). This tracheal signal might represent very faint TwdlD protein expression in the tracheal system. Alternatively, it could also be an artifact caused by RFP protein over-expression and/or diffusion into the trachea.

In summary, both TwdlF-RFP and TwdlD-RFP proteins are secreted matrix proteins that are incorporated into the larval cuticle. Furthermore, these two proteins are incorporated into the distinct parts of larval cuticle during overlapping but different larval stages, indicating unique functions for each of them despite the sequence similarity.

2.3 Cause of *TwdlD*^{1} phenotype

2.3.1 Localization of the *TwdlD*^{1} protein

Generation of *TwdlD*^{1}*-RFP* and *TwdlA*^{Tb}*-RFP* transgenics

After knowing the normal localization of the cuticular protein TwdlD and TwdlA, localization information for the mutated proteins is expected to improve our understanding of the *TwdlD*^{1} and *Tb*^{1} phenotype. As RFP tagging proven rather successful in monitoring the wild type Twdl proteins, transgenic flies expressing RFP tagged *TwdlD*^{1} and *TwdlA*^{Tb} proteins might reveal the differences between the wild type proteins and the mutated proteins.

The same genomic region that was used to generate TwdlD-RFP trans-

genics was cloned from $TwdlD^1$ homozygotes using identical primers. The resulting fragment contains 500 bp upstream regulatory region. A sequence encodes a monomeric red fluorescent protein was inserted into the genomic region right before the stop codon of the $TwdlD^1$ gene, resulting in a DNA fragment encodes a $TwdlD^1$ protein with a C-terminal RFP fusion. This fragment was cloned into the germ line transformation vector CaSpeR, and at least three independent transgenic lines were generated. Similarly, $TwdlA^{Tb}$ (Tb^1) genomic region was cloned from Tb^1 homozygotes, and transgenic flies expressing $TwdlA^{Tb}$-RFP fusion proteins were produced.

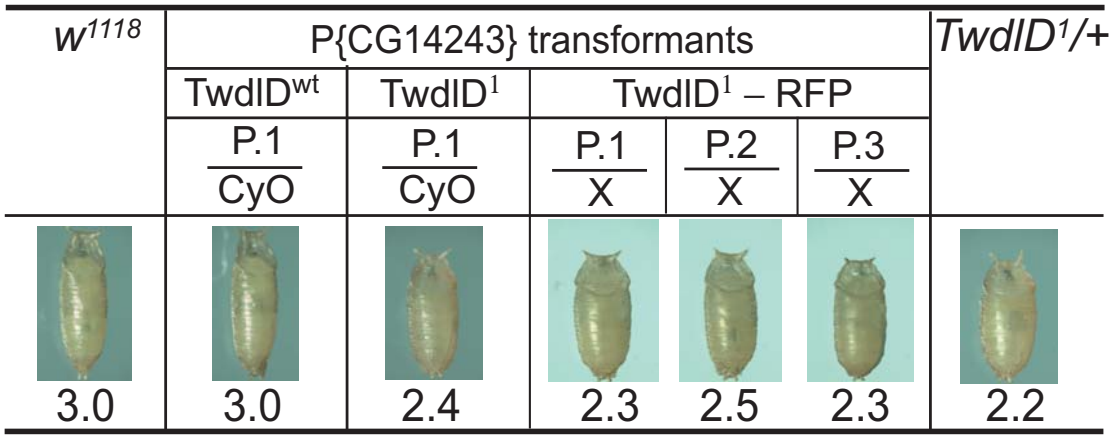


Figure 2.14: Pupal axial ratios of $TwdlD^1$-RFP transgenic flies.

Body shape change in $TwdlD^1$-RFP transgenic flies

Similar to untagged proteins, both $TwdlD^1$-RFP and $TwdlA^{Tb}$-RFP are capable of inducing body shape change. In the case of $TwdlD^1$-RFP transgenics, since p element was inserted into the X chromosomes in all three independent transgenic lines, homozygous transgenic females were crossed to wild type males to score the pupal axial ratios of the progenies. The three $TwdlD^1$-RFP lines exhibit pupal axial ratio of 2.3 ± 0.1, 2.5 ± 0.1 and 2.3 ± 0.1 respectively (Figure 2.14), which are reduced to the level identical to that of $TwdlD^1$ transgenics. Likewise, the pupal axial ratios of the two $TwdlA^{Tb}$-RFP lines were reduced to 2.0 ± 0.1 in

heterozygous transgenics (data not shown).

The fact that RFP tagging does not affect the ability of *TwdlD^1* protein and *TwdlA^{Tb}* protein to induce the dominant phenotype further validates the use of RFP in tracking both the wild type and mutated Twdl proteins. This approach provides a means to study the causes of the body shape change observed in the mutants.

Localization of *TwdlD^1* protein in larval cuticle

Transgenic flies expressing *TwdlD^1-RFP* fusion proteins have fluorescent cuticle as viewed from both dorsal (Figure 2.15 A) and ventral perspectives (Figure 2.15 B). On the dorsal surface, the dorsal hairs are very bright with *TwdlD^1-RFP* proteins incorporated (Figure 2.15 E). A cross section of *TwdlD^1-RFP* transgenic first instar larva shows a cuticle more indented than that of the TwdlD-RFP transgenics (Figure 2.15 C and D). The segmental grooves between the dorsal hairs appeared to contract more tightly, resulting in bulges within the segments and reducing the thickness of each segment.

The most astonishing differences between *TwdlD^1-RFP* transgenics and TwdlD-RFP transgenics are observed on the ventral cuticle. As shown earlier, wild type TwdlD-RFP protein was detected within the naked cuticle areas between the denticle belts. In addition, the protein localized within the naked cuticle is distributed uniformly within the region. However, *TwdlD^1-RFP* fusion protein shows a distinct localization pattern (Figure 2.15 F). First of all, in all three independent transgenic lines of *TwdlD^1-RFP*, the ventral denticles of young first instar larvae are fluorescent with RFP incorporation. A light image and a fluorescent image were taken for a ventral area containing two adjacent denticle belts. The overlay image of these two further shows that both the naked cuticle and the denticle belts are fluorescent in this transgenic larva. Similar results were observed with the other two transgenic lines.

In one of the transgenic line, line 1, numerous bright red spots were

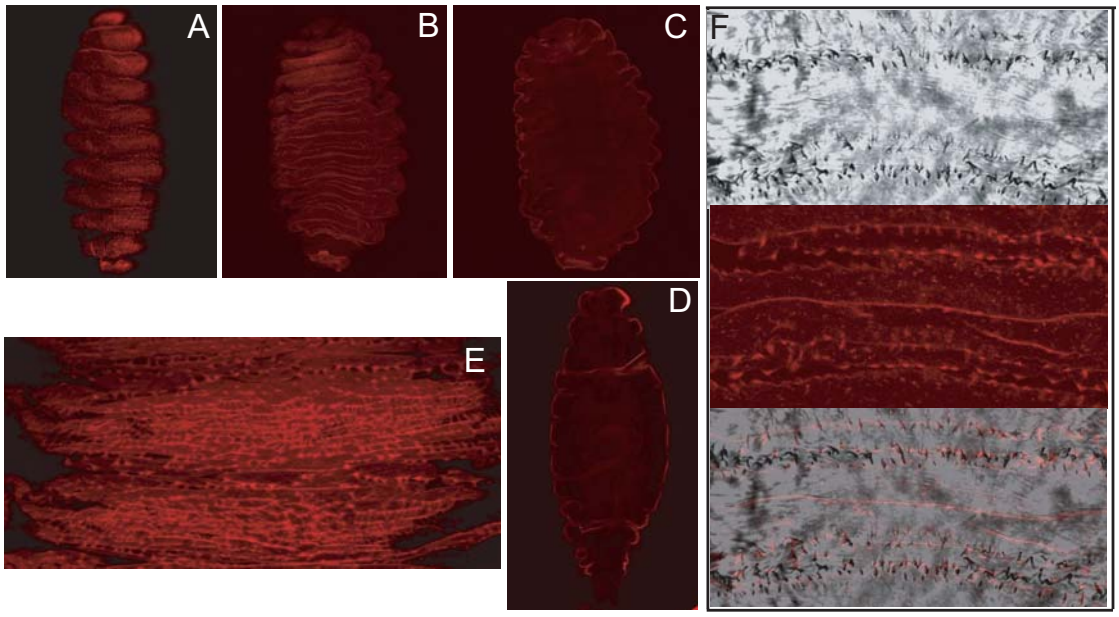


Figure 2.15: Localization of *TwdlD*^{1}*-RFP* fusion protein in young first instar larvae.

detected within the cuticle (data not shown). They are particular easy to recognize in the naked cuticle area. These red spots do not seem to co-localize with any particular structures, therefore might represent abnormal aggregation of the *TwdlD*^{1}*-RFP* fusion protein. This observation suggests that TwdlD proteins with conserved block IV disrupted can self-aggregate.

2.3.2 Causes of the *TwdlD*^{1} phenotype

RNA expression pattern of *TwdlD*^{1} allele

The fact that *TwdlD*^{1}*-RFP* protein shows a different localization pattern from the TwdlD-RFP protein is very intriguing. To confirm that this difference is indeed on the protein level, RNA expression pattern of *TwdlD*^{1} allele in *TwdlD*^{1} homozygous mutant embryos was studied by *in situ* hybridization. The same antisense probe and sense control probe as used in examining RNA expression of the wild type TwdlD gene were used on *TwdlD*^{1} homozygous embryos. The result confirms that RNA expression of *TwdlD*^{1} allele also forms segmental stripes

along the anterior-posterior axis within epidermis (Figure 2.16), which is identical to that of the wild type TwdlD allele in the late-stage embryos. Therefore, the different localization pattern of *TwdlD*^{1}*-RFP* protein is not a result of different gene expression pattern. Instead, it represents a loss of normal localization caused by the disruption of the highly conserved amino acid block IV in *TwdlD*^{1} mutant protein.

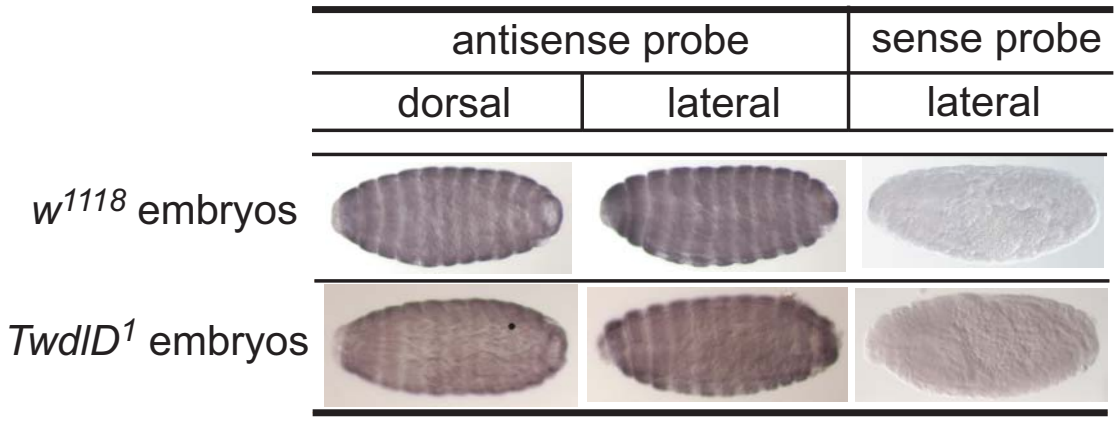


Figure 2.16: TwdlD RNA expression in *TwdlD*^{1} mutant embryos examined by *in situ* hybridization.

Two hypotheses to explain *TwdlD*^{1} phenotype

Loss of normal localization pattern for *TwdlD*^{1} protein could be either the cause or a side result of the mutation. In one case, by leaving their normal location and migrating into spaces where TwdlD protein is not supposed to be, the ectopic *TwdlD*^{1} protein changes the properties of the local cuticle, resulting in hypercontraction of the grooves and surrounding area which leads to body shape change. Alternatively, the phenotype can also be caused by TwdlD protein absent from its normal location or from its normal binding sites where they are required to maintain the proper cuticle structure. In this scenario, the presence of *TwdlD*^{1} protein in the denticle belt areas does not cause the phenotype.

In the later case, the loss of normal binding or localization of TwdlD pro-

tein is less likely to cause a dominant phenotype observed in the $TwdlD^1$ mutants. On the other hand, the nature of dominance is ready to be explained by the idea of ectopic localization of $TwdlD^1$ protein causing the body shape change.

2.4 Model for body shape regulation by Twdl proteins

2.4.1 Experimental facts

Twdl proteins represent a new family of extracellular matrix proteins. They are synthesized and secreted into the cuticle by ectodermal tissues, including epidermis, foregut, trachea and et al. Despite their sequence similarity, member proteins seem to have specific temporal and spatial expression patterns, and specific localization patterns. These gene-specific patterns are crucial for their proper functions, as demonstrated by the expansion of $TwdlD^1$ protein in $TwdlD^1$ mutants.

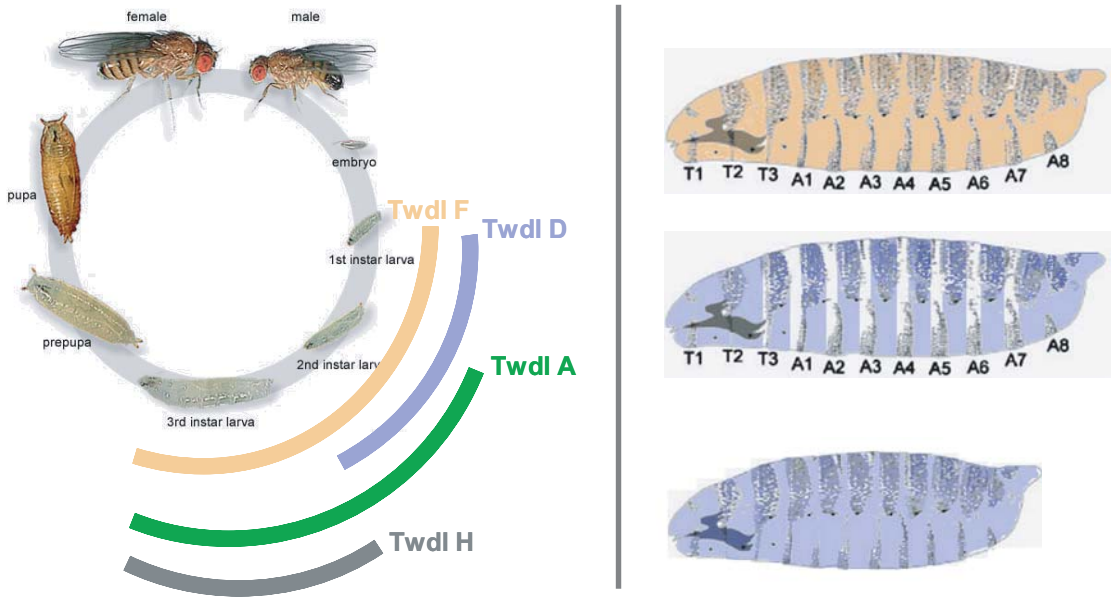


Figure 2.17: Twdl proteins have protein-specific expression and localization pattern.

2.4.2 Model for body shape regulation by Twdl proteins

Based on these experimental data (Figure 2.17), a model is raised to explain the functions of Twdl proteins in regulating body shape.

1. Larval body shape is controlled by larval cuticle.
2. The physical property of larval cuticle is not uniform.
3. For certain part of a cuticle, its physical property is regulated by the presence of Twdl proteins.
4. Twdl motif is required for Twdl proteins to interact with its binding partner (which could be chitin).
5. Loss of Twdl motif allows Twdl proteins to migrate to other loci and change cuticle property.
6. Change of cuticle leads to change of body shape.

2.5 New directions for studying Twdl proteins

2.5.1 Functions of Twdl proteins in tracheal development

Potential functions for Twdl proteins in tracheal system

In Drosophila, one of the major places where a chitin-protein matrix is secreted is the tracheal system. A transient chitin-protein matrix is required for the tracheal tubes to gain its proper diameter and length during embryonic stage 14-15. A chitin-based cuticle is later secreted by the tracheal cells, to protect, support and regulate the tracheal tubes. Several Twdl family proteins are demonstrated to be matrix proteins that are needed for larval cuticle. Other Twdl proteins might contribute to tracheal development as well. In fact, RNA of the family member Twdlβ, is strongly expressed in the tracheal system in stage 13-16 embryos. In addition, since fluorescence was observed in the TwdlD-RFP transgenics with two

copies of p insertion, there is a possibility that TwdlD protein itself is expressed in the larval trachea. These observations suggest a potential role for Twdl proteins in tracheal development.

Tracheal staining of *TwdlD*^{1} and *Tb*^{1} mutants

In order to examine the tracheal morphology in *TwdlD*^{1} and *Tb*^{1} mutants, immunostaining of tracheal lumen was performed for stage 13-16 embryos, using mouse monoclonal antibody 2A12 (Developmental Studies Hybridoma Bank (DSHB)). The 2A12 antigen is expressed by the tracheal cells and contributes to the tracheal lumen after secretion from the onset of tube growth. 2A12 antibody staining of tracheal dorsal trunk is already detectable at embryonic stage 14 (Moussian, Tang, Tonning, Helms, Schwarz, Nusslein-Volhard & Uv 2006).

Embryos of stage 13-16 were collected and stained for w^{1118}, *TwdlD*^{1} and *Tb*^{1} flies. The precise stages of stained embryos were judged by the morphology of their tracheal system, according to the description by Manning and Krasnow (Bate & Arias 1993) (Figure 2.18). In embryonic stage 14, the dorsal branch (DB) remains growing, and the ganglionic branch (GB), which is the posterior branch of the lateral trunk, remains clearly visible as the main branch reaching the ventral nerve cord. In embryonic stage 15, the end of DB develops a half loop structure and a dorsal anastomosis spur begins to form at the most dorsal point. On the lateral side, GB becomes much finer and turns and migrates along the nerve cord toward the midline. Furthermore, in stage 15, the visceral branches (VB) within the dorsal-lateral region develop their segment-specific branching pattern at their terminuses.

Immunostaining of w^{1118}, *TwdlD*^{1} and *Tb*^{1} embryos revealed a tracheal lumen phenotype in *Tb*^{1} mutants but not *TwdlD*^{1} mutants. Lumen of the tracheal dorsal trunk is indistinguishable in w^{1118} and *TwdlD*^{1} embryos of stage 14 and stage 15. They align well with the tracheal epithelia and their texture remains uniform. Nevertheless, in *Tb*^{1} mutants, although the tracheal trunk and branches

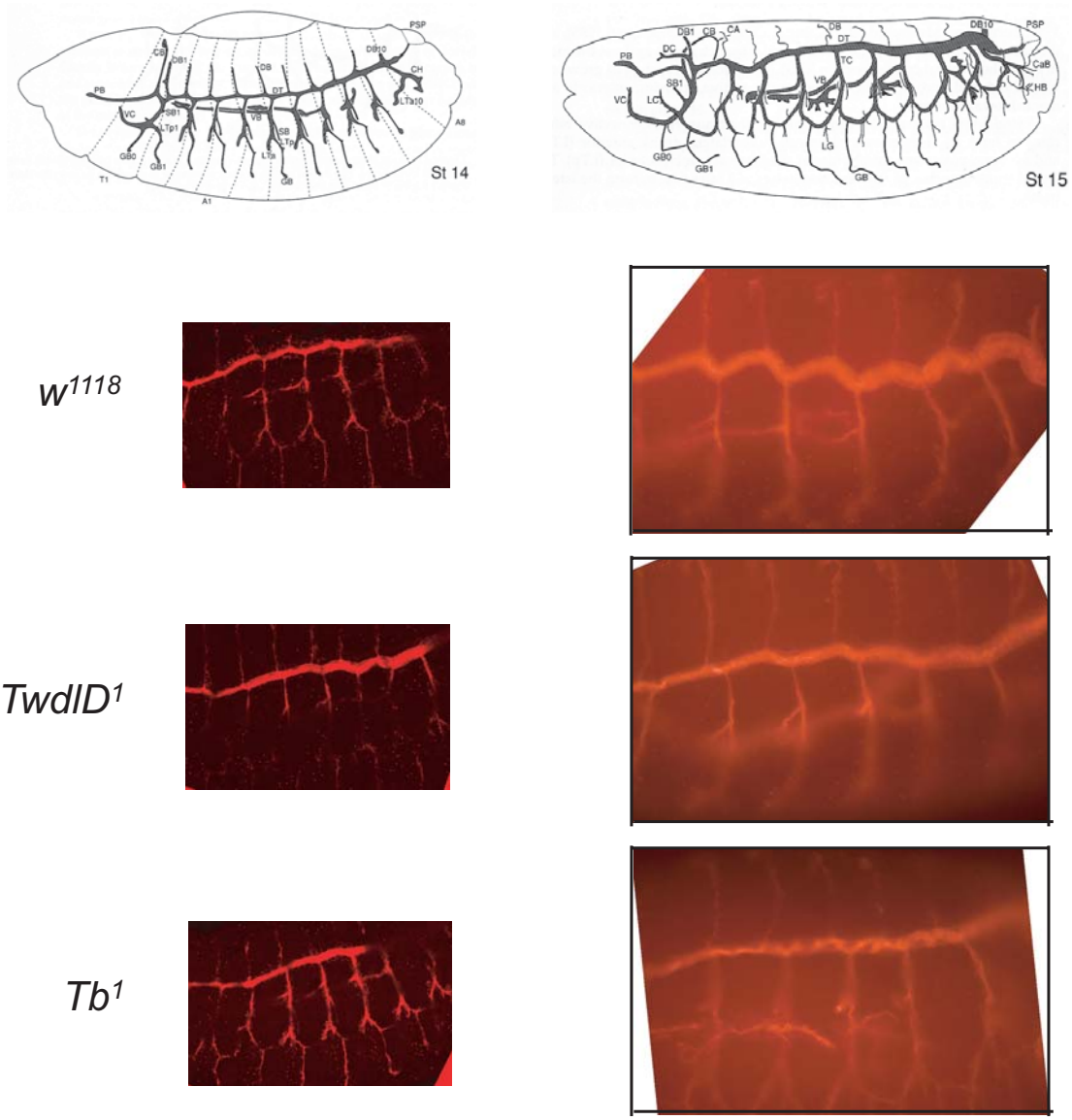


Figure 2.18: Embryo tracheal lumen staining for w^{1118}, $TwdlD^1$ and Tb^1 flies.

are normal in stage 14 embryos, a twisted and thinner lumen matrix was recorded in a good number of stage 15 embryos (Figure 2.18). In those embryos with twisted tracheal matrix, the lumen matrix appears to be elongated and as a result of the elongation, the matrix seems to fold over itself, giving a tortuous look.

A potential function of TwdlA in tracheal development

The staining result strongly suggests a role for the TwdlA protein in tracheal development. However, embryonic *in situ* hybridization performed failed to identify any TwdlA expression in the tracheal system in late-stage embryos. Several factors might contribute to this apparent conflict. First, tracheal expression of TwdlA might be really weak. Secondly, tracheal expression of TwdlA might be transient, which would be consistent with the transient tracheal lumen matrix formation. Both reasons might explain the failure to detect TwdlA mRNA expression in late-stage embryos by *in situ* hybridization. In addition, a 'tortuous larval trachea' phenotype was recorded for Tb^1 mutants previously. Although it is not clear whether the tortuous larval trachea is the a direct phenotype of TwdlA mutation or a side-effect of altered larval body shape, it is in line with the idea that TwdlA gene might have a function in tracheal development. Last, although unlikely, it is possible that the abnormal tracheal matrix observed in Tb^1 mutant embryos is caused by the staining process, especially since only a portion (estimated to be less than 20%) of the Tb^1 embryos of similar stages show the twisted luminal matrix. On the other hand, the low percentage of abnormal tracheal matrix in Tb^1 embryos can also be explained by it being a very transient phenomenon.

Approaches to study the role of Twdl proteins in tracheal development

Several lines of experiments might help to address the question of whether some Twdl family proteins function in trachea. First, embryonic *in situ* hybridization can be used to identify all Twdl family genes that are strongly expressed in the tracheal system. Secondly, for those genes that are expressed in the trachea like Twdlβ, the involvement in tracheal development might be verified by introducing a $TwdlD^1$ or Tb^1 like mutation in the gene. Transgenic flies expressing these mutated genes might also develop a dominant phenotype, which can be easily examined by Immunostaining or RFP tagging. Last, to study genes that are transiently or weakly expressed in the trachea, which is likely the case for TwdlA,

a more sensitive detection method needs to be developed.

2.5.2 Direct interaction between Twdl proteins and chitin

Secondary structure prediction of Twdl proteins

A stereotype Twdl protein contains a signal peptide at its N-terminus and four conserved blocks of amino acid (Figure 2.19). Each of the four blocks is about 15-amino-acid long with the second block slightly shorter. The blocks are separated by short strings of 14-18 amino acids. A close look at the amino acid composition within these four conserved blocks reveals extraordinarily high percentage of aromatic amino acids (Y, F and H) and charged amino acids (R, K, H positively charged; D, E negatively charged). Some of these residues are strictly conserved among insect Twdl homologues, including K and E in block I, K and F in blockII, Y and K in block III, K, F and Y in block IV. In addition, several other positions are well conserved with only substitutions by similar amino acids, for example, substitution of Y with F or H, substitution of K with R, or substitution of D with E.

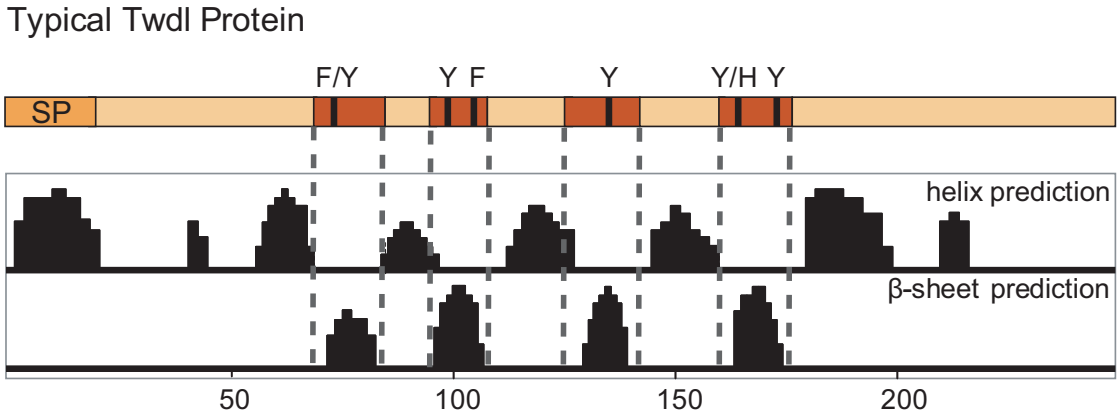


Figure 2.19: Secondary structure prediction for Twdl proteins.

To better understand the structures of Twdl proteins, a secondary structure prediction was made using the alignment of insect Twdl proteins. The prediction program predicts mainly α helix and β sheet, interpretating the unrecognized

regions as flexible loops. More than eight α helixes were predicted along the length of the proteins, while only four β sheets were predicted. Interestingly, these four β sheets co-localize with the four conserved blocks shared by the Twdl family, and the most conserved aromatic amino acids were predicted to sit within the β sheets. This prediction indicates that Twdl family proteins might rely heavily on a tertiary structure built on four β sheets.

Direct interaction between Twdl proteins and chitin?

Extended R&R domain was the only chitin-binding domain of which the tertiary structure is known. This domain was previously shown to form a β-sheet half-barrel comprised of 3 or 4 β sheets. The half-barrel structure is proposed to provide a groove where the chitin chain can fit in. In addition, conserved aromatic amino acids were found within these β sheets and most of them have their aromatic rings exposed on the groove surface. The exposed rings are thought to bind to and stack with the sugar rings of chitin.

Although no sequence similarity was identified between the Twdl block region and the extended R&R, they do resemble each other in several ways. First, they share similar secondary structure composition. Second, they all have conserved aromatic amino acids. Third, they are both present in matrix proteins. Based on this analysis, I predict that Twdl proteins might bind to chitin molecules directly *in vivo*. To test this prediction, further studies of direct binding between Twdl proteins and purified chitin will be needed. Furthermore, if such interaction does occur, the above analysis would predict that point mutations of the conserved aromatic amino acids might affect the binding activity or specificity.

2.5.3 Regulation of Twdl gene expression

Regulatory regions of Twdl genes are short

Generally speaking, Twdl genes are small genes that encode proteins no longer than 350 amino acids. Most of the Twdl genes contain either no intron or

one intron less than 100bp, with only a couple of exceptions. For example, the gene of TwdlG contains 3 small introns.

The regulatory regions of Twdl genes appear to be pretty small, too. Transgenic flies of $Twdl D^{1}$ with only 500 bp upstream region can induce body shape change. Transgenic flies expressing RFP fusion proteins of TwdlF and TwdlD under control of 500 bp regulatory region can recapitulate the native protein expression pattern, as demonstrated by the consistency between *in vivo* mRNA expression and RFP fusion protein expression. Yet, the semi-dominance of $Twdl D^{1}$ fragment containing only 500 bp regulatory region does raise the question of whether this 500 bp region contains all the regulatory elements.

Therefore, in generating $Twdl A^{Tb}$ or $Twdl A^{Tb}$-*RFP* transgenic flies, the upstream region was extended to 1000 bp. In this case, one copy of the $Twdl A^{Tb}$ genomic fragment inserted into wild type genome can fully induce a body shape change similar to that of Tb^{1} mutants. Therefore, it was concluded that a region of 1000 bp upstream of the transcription start point contained all essential regulatory elements required for TwdlA gene expression.

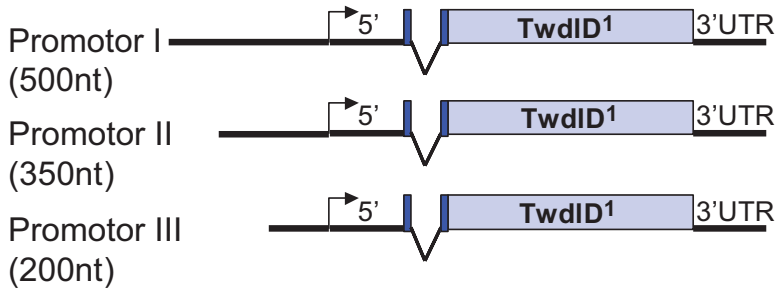
The fact that TwdlD, TwdlF, TwdlA have very small regulatory regions leads to the hypothesis that expression regulatory elements of all Twdl genes are located within a region of 1000 bp upstream of the transcription start point.

Deletion analysis of TwdlD regulatory region

In order to identify potential *cis*-regulatory elements required for TwdlD expression, deletions were made within the 500 bp regulatory region of $Twdl D^{1}$ transgenic fragment. 150 bp of DNA or 300 bp of DNA was deleted from the regulatory region, and transgenic flies containing $Twdl D^{1}$ gene with trunked regulatory regions were generated (Figure 2.20 A). When the pupal body shape of these transgenic flies were studied, it was found that $Twdl D^{1}$ gene with only 300 bp regulatory region upstream of the transcription start site is not capable of inducing body shape change, even at a homozygous state (Figure 2.20 B). The pupal

axial ratios of these transgenic flies are identical to that of the wild type flies. Therefore, it is concluded that there were essential regulatory elements for TwdlD expression within the 150 bp region from 500 bp upstream to 350 bp upstream (Figure 2.20 C).

A



B

w^{1118}	P{CG14243} transformants				$TwdlD^{1}/+$
	$TwdlD^{wt}$	$TwdlD^{\Delta 173-175}$			
		Promotor 1	Promotor 2	Promotor 3	
	P.1 / CyO	P / CyO	P / P	P / CyO	
					
3.0	3.0	2.4	3.0	3.0	2.2

C

```
cagccaccag atattatcgc caaagccagt agctgtttat tttccgagca
tttcaacaca gtggctgccg tttggtccgt ctctgcagca tccagattgt
gtttttttatg gcctgtcgta gccaacacaa atcaattaga taatgtagca
```

Figure 2.20: Deletion analysis of TwdlD regulatory region.

Prediction of *cis*-regulatory elements

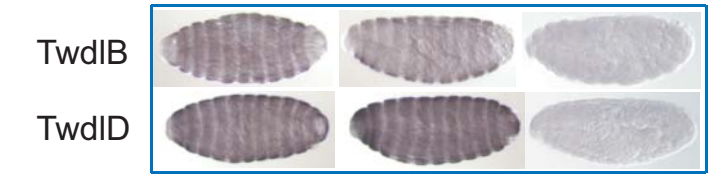
To search for *cis*-regulatory sites important for Twdl gene expression, a motif finding program MEME (Multiple Em for Motif Elicitation) was explored. MEME program was designed to identify highly conserved regions in groups of related DNA or protein sequences. In looking for conserved *cis*-elements by MEME, to feed the program with a group of sequences that are regulated similarly is expected to reduce the false positive rate. In addition, to increase the number of the starting sequences is also expected to increase the accuracy.

Although the Twdl proteins share sequence similarity, they are clearly regulated differently judging by their expression time, location. Therefore, to feed the 1000 bp upstream region of all the Twdl genes to MEME is not expected to identify specific regulatory elements. In fact, attempts of such kind all failed to generate motifs with good scores.

However, some of the Twdl genes can be categorized by their embryonic expression patterns. Among the nine Twdl genes that were examined by embryonic *in situ* hybridization, five Twdl genes can be put into two groups based on their expression loci. TwdlB and TwdlD are both expressed in the epidermis forming segmental stripes along the anterior-posterior axis. TwdlC, TwdlE and TwdlT were detected within the embryonic foregut. Although groups of larger sizes are more desirable as starting sequences, MEME analysis of these two groups already identified very promising motifs (Figure 2.21).

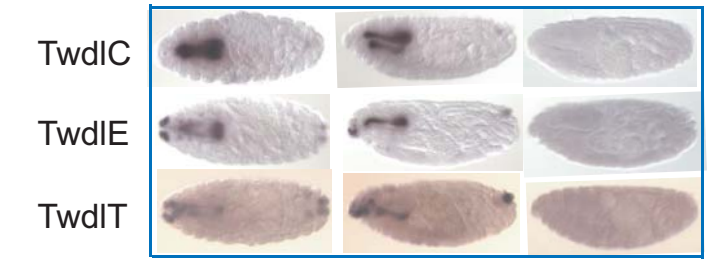
For MEME analysis of TwdlB and TwdlD, regulatory regions of different length were fed into the program. Specifically, while 1000 bp upstream of transcription start site of TwdlB was used, only the 150 bp regulatory region of TwdlD was included. The top two motifs identified by MEME are AACACAGTGGCT and TCCG(T/A)GCA. Neither motif matches the consensus sequences of transcription factors recorded in the Transfac database.

Motifs with good scores were also identified for the three genes expressed in the foregut. 1000 bp upstream regulatory region was used for each of the three



TwdlB(+)	GGCGGCAA	AACACAGTGGCT	TCAAAGAAA	- 99
TwdlD(+)	AGCATTTC	AACACAGTGGCT	GCCGTTTGG	- 479

TwdlB(+)	CAAAGAAATC	TCCGTGCA	GCAGACGCCG	- 80
TwdlD(+)	CTGTTTATTT	TCCGAGCA	TTTCAACACA	- 495



TwdlC(+)	CACTTGGCGC	ACTTGGCCGA	AGGTTGCGGC	- 499
TwdlE(+)	TGCTCAGCCA	ACTTGGCCGA	CTGCTCGGTC	- 137
TwdlT(+)	CTGGGTCTCG	ACTTGGCCAA	AATGTGTGGC	- 481

TwdlC(+)	GGGTCACTTT	CGATTTCGAGC	CAGCAATTTA	- 559
TwdlE (-)	TCAGCTTTAA	CGATTTCCTGC	TTAATTTAAG	- 631
TwdlT(+)	ACAGCACCGC	CGATTTCCCGC	TGTCGGATGT	- 385

Figure 2.21: Regulartory motifs of Twdl genes identified by MEME.

genes in this analysis. The top two motifs identified by MEME are ACTTG-GCC(G/a)A and CGATTTC(G/C)(A/T/C)GC.

In order to verify whether any of these identified elements is important for Twdl gene expression, transgenic flies expressing mutated Twdl proteins or reporter genes can be used to test these elements. For example, to test the impor-

tance of the two motifs identified upstream of TwdlD gene, transgenic flies *TwdlD*^{1} genomic DNA fragment with the motifs mutated or deleted can be very informative. For those genes for which no dominant mutations are available, a Twdl-RFP fusion protein can be used in place of the *TwdlD*^{1} protein. Furthermore, embryonic expression of the genes encoding the fusion proteins can also be monitored by in situ hybridization against the RFP region.

This chapter includes the reprint of the following paper:

Xiao Guan, Brooke W. Middlebrooks, Sherry Alexander,Steven A. Wasserman - *Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila*, PNAS, Vol. 103(45), pp. 16794-16799, Nov. 2006.

3

Discussion

My study of the *TwdlD*^{1} and *Tb*^{1} mutants has led to the discovery of a novel protein family, the Tweedle family. Ectodermal expression was observed for all family members tested. In addition, we have shown that at least four of the family members are cuticular proteins. Each of Twdl family genes has gene-specific expression pattern and localization pattern. When these patterns are disrupted as in *TwdlD*^{1} mutants, body shape change can occur. Our findings thus establish a connection between body shape regulation and matrix proteins that contribute to the cuticle.

3.1 Role of Tweedle proteins in cuticle assembly

The Tweedle family members in the Drosophila genome form three major gene clusters. The 97C cluster, which includes the TwdlD 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 gene family apparently undergone multiple gene duplication events? Our studies indicate that the expansion in gene number was accompanied by a differentiation of distinct patterns of expression. One possibility therefore is that each family member functions identically at the biochemical level, with the differences in expression determining the organization of the cuticle. Thus, for example, different

levels of Tweedle protein at particular locations could determine the extent of cross-linking and, hence, flexibility. Similarly, differences in the timing of expression could dictate the order of assembly of cuticle at distinct locations. Alternatively, family members could differ in biochemical function, with the sequence differences seen between family members dictating local differences in cuticle composition and properties.

The *TwdlD*^{1} mutation, which does not change the stability or the secretion of the protein, does result in mis-localization of the *TwdlD*^{1} protein within the cuticle. As mentioned in the results section, this mis-localization might be the cause of the body shape change observed with the *TwdlD*^{1} mutants. Alternatively, the phenotype could also be caused by the mutation affecting either the conformation of the TwdlD protein or its activity in forming or stabilizing crosslinks within the cuticle.

Although none of the 27 Tweedle genes in Drosophila has been studied previously, a recent report describes a characterization of a related gene in the silkworm, *Bombyx mori* (Zhong, Mita, Shimada & Kawasaki 2006). This silkworm protein, BmGRP2, was detected in the cuticle layer of the wing tissue and in the trachea in the silk worm. 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 YVLX$_{20-23}$KPEVyFiKY(R/K)t.

Like BmGRP2, some Tweedle proteins contain glycine-rich domains. However, the glycine-rich domain is absent in 21 of the 27 Tweedle proteins in Drosophila, including three out of the four studied in this report at the protein level - TwdlD, TwdlF and TwdlH. 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.

3.2 Dominant mutations of Twdl family genes

The two dominant mutations of Twdl family proteins $TwdlD^1$ and Tb^1 are special in the sense that the heterozygous and the homozygous mutants have indistinguishable phenotype. With the limited understanding we have of the two mutants, there is no obvious reasons to explain this phenomenon.

Study of transgenic flies expressing $TwdlD^1$-RFP proteins indicates that the mutation of TwdlD protein disrupts its association with its binding partners, leaving the proteins free to migrate and/or aggregate. Therefore, it is reasonable to speculate that the wild type TwdlD proteins have at least two functional domains - while the Twdl motif contributes to binding to chitin (or other proteins), another domain mediates homophilic interactions. One possibility is that, in heterozygous $TwdlD^1$ mutants, the $TwdlD^1$ proteins not only mis-locate, but also sequester away the wild-type proteins. As a result of it, the heterozygous mutants are essentially similar to homozygous mutants.

To test this possibility or to identify the real explanations, detailed biochemical studies of TwdlD and TwdlA proteins need to be performed. In particular, several things can be examined pretty easily, including whether TwdlD protein directly binds to chitin and whether it directly binds to itself.

3.3 Genetic control of larval and pupal body shape

While the $TwdlD^1$ phenotype is mostly easily recognized during the larval and the pupal stages, TwdlD gene expression begins in the latter half of embryogenesis and is no longer detectable by the end of the last larval stage. The lack of any shape alteration in $TwdlD^1$ embryos presumably reflects the fact that the surrounding eggshell is a protein-based extracellular matrix distinct from cuticle. Within the eggshell, however, the embryonic cuticle structure is clearly affected, as is evident upon examination of newly hatched first instar larvae. A strong $TwdlD^1$ phenotype observed during the pupal stage, after the cessation of gene expres-

sion, very likely represents residual effects during pupariation of the larval cuticle abnormality.

One previously described dominant mutation-$Kugel^{Valencia}$ (Kg^{V})- has a phenotype highly reminiscent of $TwdlD^{1}$. Like $TwdlD^{1}$, this mutation reduces axial ratio at the larval and pupal stages. This similarity suggests that the two loci may act in the same pathway. Kg^{V} maps to the left side of the gene Ki, 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 Kg^{V} mutant.

3.4 Convergent evolution of body shape regulation

In mammals, mutations that cause bone structural defects can cause dwarfism, as the result of smaller, thinner bones within the body. In insects and worms, structural defects of exoskeleton also changes the overall body shape. People have known for a long time that the disruption of cuticular collagens in *C. elegans* can cause the dumpy phenotype, which describes the shorter and wider morphology. We have demonstrated in this report that mutations of the cuticular proteins TwdlD and TwdlA cause similar morphological 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.

4

Materials and Methods

4.1 Genetic screen for morphology mutations

To identify new mutations affecting pupal shape, we carried out a screen for dominant mutations on the third chromosome. We crossed mutagenized males (4,000 Rad gamma-irradiation) to virgin females and assayed directly for pupae with an altered axial ratio (see below). From approximately 25,400 pupae, we identified two stable dominant mutations and characterized one, which was designated *TweedleD*[1].

4.2 Axial ratio determination

For pupae, axial ratio (length/width) was measured using a reticle in a stereo light microscope. For each genotype, we measured at least 40 individuals and the mean axial ratio was calculated. For larvae, axial ratios were determined from photographs of cuticle preparations taken with a digital camera. At least two independent preparations were examined for each genotype and twenty individual cuticles were measured for each preparation.

4.3 *p* induced male recombination mapping

p element induced male recombination mapping was performed as previously described (Chen, Chu, Harms, Gergen & Strickland 1998). Triply labeled chromosomes *Ly* Tb^1 *Dr* and *e* $TwdlD^1$ *Dr* were generated by meiotic recombination. *P* insertion lines BL12808, BL13710, BL20052, BL13022, BL10343 and BL11782 were obtained from the Bloomington Stock Center.

4.4 Sequencing of genomic DNA

Genomic DNA of $TwdlD^1$ homozyotes, $TwdlD^1$ parental strain and Tb^1 homozygotes was purified from adult males. Eight overlapping DNA segments (segment 1-8) of 8kb to 11kb were amplified by PCR for each genotype. The eight segments together cover the whole region of 73.8kb (nucleotides 19,189-92, 948, accession No. AE003757), where $TwdlD^1$ and Tb^1 were mapped by *p* induced male recombination. Multiple primers were used to sequence the DNA segments to ensure a full coverage.

4.5 Sequence analysis

Similarity searches to identify the Tweedle family members were performed using BLAST (http://www.ncbi.nlm.nih.gov).

The multiple protein alignment and similarity analysis were carried out using CLUSTALW (http://www.ebi. ac.uk/clustalw).

The signal peptide prediction was made using the SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP) (Bendtsen, Nielsen, von Heijne & Brunak 2004).

The secondary structures of the Twdl family proteins were predicted with the PHD algorithm at the (http://www.predictprotein.org) (Rost, Yachdav & Liu 2003, Rost 1996).

4.6 Protein expression in S2 cell culture

The coding region of TwdlD was fused to a FLAG tag at its C-terminus and cloned into the pAc5.1/V5-His A vector (Invitrogen). The coding region of TwdlJ was cloned into the same vector, where it was fused to the V5 epitope tag. S2 cells were transfected and the protein contents were harvested as described in the manual for the Drosophila Expression System (Invitrogen). For each plate of transfected cells, the media and the cell pellet were separated by centrifugation at 3,000 rpm, and 1/60th volume of the media and of the total cell lysate were each loaded onto an SDS-PAGE gel for immunoblotting. Antibodies used in this experiment are: anti-V5 antibody at 1:10,000 dilution (Invitrogen 46-0705); anti-FLAG M2 antibody at 1:1000 dilution (Stratagene 200472-2); rabbit anti-Cactus antiserum at 1:10,000 dilution.

Constructs generation for S2 cell transfection : The TwdlD^{wt}-Flag insert was generated by RT-PCR (forward primer, 5'- gcgaattcatgcgtgcttttatcgtcctc -3'; reverse primer, 5'- gcctcgagttacttatcgtcgtcatccttgtaatccttgacgcggaaacgacg-3') using w^{1118} total RNA (overnight embryo collection at 25C) as template. The RT-PCR product was ligated into pAc5.1A vector (Invitrogen) at EcoRI and XhoI sites. The *TwdlD*$^{\Delta 173-175}$*-Flag* insert was generated by RT-PCR using the same primer pair but *TwdlD*^{1} total RNA (overnight embryo collection at 25C) as template. The TwdlA^{wt} insert was generated by RT-PCR (forwar primer, 5'-gcgaattcatgcgtggatttattatttttgct -3'; reverse primer, 5'-gcctcgagcttgaccttgttcacag gcaggt-3') using w^{1118} total RNA (overnight embryo collection at 25C) as template. *TwdlA*^{Tb} insert was generated by RT-PCR using the same primer pair but *Tb*^{1} total RNA as template. The TwdlJ^{wt} insert was generated by direct PCR (forward primer, 5'-gcgaattcatgcagagcgttttgcatagc - 3'; reverse primer, 5'- gcctcgaggaaacgcaggcggcgcaggat - 3') using w^{1118} genomic DNA as template. The PCR product was ligated into pAc5.1A at EcoRI and XhoI sites.

4.7 Embryonic *in situ* hybridization

Embryonic RNA expression patterns were investigated by *in situ* hybridization. The 3'UTRs of target genes were amplified from the w^{1118} 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) was used at 1:2000 dilution.

Constructs generation for *in situ* hybridization: The 3'UTR of genes of interest was amplified by PCR from w^{1118} genomic DNA. The PCR products were cut with EcoRI and XhoI enzymes and ligated into pBluescript II KS (+) (Stratagene). The primers used for amplification are:

- TwdlA_3UTR.fwd, gcgaattcgaaacctgcaagaccacattctta;
- TwdlA_3UTR.rev, gcctcgagagttaagtttatattttatacggt;
- TwdlB_3UTR.fwd, gcgaattcgaaggctacatcttggactccatt;
- TwdlB_3UTR.rev, gcctcgaggtatttaaatttcaaatttattgg;
- TwdlC_3UTR.fwd, gcgaattctgaagtgaagcccgctgctttgag;
- TwdlC_3UTR.rev, gcctcgagatgtaacaagttatacaaacgaac;
- TwdlD_3UTR.fwd, gcgaattcatggtctcaagtgaaatttcaacg;
- TwdlD_3UTR.rev, gcctcgagtttgtcagcaaaacaaattttatt;
- TwdlE_3UTR.fwd, gcgaattcgcgatccagtccaacccgaatacc;
- TwdlE_3UTR.rev, gcctcgaggaggggatgggggagattcgttgc;
- TwdlF_3UTR.fwd, gcgaattcaaagttgtagtaagaatcctatcg;
- TwdlF_3UTR.rev, gcctcgagtttttgggtttcgatttaaatttt;
- TwdlT_3UTR.fwd, cggaattccacctggaaagtccggac;

- TwdlT_3UTR.rev, cgctcgagcgaaagcgccgccccctc;
- TwdlW_3UTR.fwd, cggaattccaacattcagagggacaattt;
- TwdlW_3UTR.rev, cgctcgagggatttgggtaacattggcaa;
- Twdlβ_3UTR.fwd, cggaattcaccgtcgagcagttctag;
- Twdlβ_3UTR.rev,cgctcgagttttgctttctctaaactctc.

4.8 RFP constructs, transgenic flies and microscopy

For genes TweedleD, TweedleF, TweedleH and TweedleA, a genomic fragment including 500 bp (1000 bp for TwdlA) of presumptive upstream regulatory sequence was cloned by PCR from the w^{1118} genome. We used PCR sewing to fuse the 3' end of the coding sequence of each gene in frame with sequences encoding the monomeric red fluorescent protein (RFP) DsRed (Clontech) (Ho, Hunt, Horton, Pullen & Pease 1989) . The resulting DNA fragments were ligated into the pCaSpeR transformation vector (Thummel, Boulet & Lipshitz 1988). Three independent transgenic lines were generated for each construct. Eggs were collected at 25C for a 2 h interval for each balanced transgenic line and aged for 22 h, 48 h, 72 h to obtain the young first, second and third instar larvae, respectively. Two-hour old first instar larvae were fixed as previously described (Goldstein & Fyrberg 1994) , and observed under a confocal microscope.

Constructs generation for RFP transgenes: CaSpeR-TwdlD-RFP construct was made by inserting TwdlD-RFP fragment into EcoRI and BamHI sites of CaSpeR transformation vector. TwdlD-RFP fragment was generated by PCR sewing of three pieces-the genomic DNA of TwdlD without the 3'UTR (nucleotides 43,291-44,677, AE003757), the DsRed (with stop codon, Clontech) and the 3'UTR (nucleotides 44,681-44,830). CaSpeR-TwdlF-RFP construct was made by inserting TwdlF-RFP fragment into EcoRI and BamHI sites of CaSpeR transformation vector. TwdlF-RFP fragment was generated by PCR sewing of three pieces-the ge-

nomic DNA of TwdlF without the 3'UTR (nucleotides 72,241-73,903, AE003607), the DsRed (with stop codon, Clontech) and the 3'UTR (nucleotides 73,907-74,038). CaSpeR-TwdlH-RFP construct was made by inserting TwdlH-RFP fragment into EcoRI and XbaI sites of CaSpeR transformation vector. TwdlH-RFP fragment was generated by PCR sewing of three pieces-the genomic DNA of TwdlH without the 3'UTR (nucleotides 37,001-38,349, AE003757), the DsRed (with stop codon, Clontech) and the 3'UTR (nucleotides 38,353-38,490). TwdlA-RFP fragment was generated by PCR sewing of three pieces-the genomic DNA of TwdlA without the 3'UTR (nucleotides 59,215-61,208,AE003757), the DsRed (with stop codon, Clontech)and the 3'UTR (nucleotides 61,212-61,380).

Bibliography

Abrams, E. W. & Andrew, D. J. (2005), 'Creba regulates secretory activity in the drosophila salivary gland and epidermis', *Development (Cambridge, England)* **132**(12), 2743–2758.

Affolter, M. & Shilo, B. Z. (2000), 'Genetic control of branching morphogenesis during drosophila tracheal development', *Current opinion in cell biology* **12**(6), 731–735.

Andersen, S. O. (2004), 'Chlorinated tyrosine derivatives in insect cuticle', *Insect biochemistry and molecular biology* **34**(10), 1079–1087.

Andersen, S. O., Hojrup, P. & Roepstorff, P. (1995), 'Insect cuticular proteins', *Insect biochemistry and molecular biology* **25**(2), 153–176.

Anderson, K. V., Bokla, L. & Nusslein-Volhard, C. (1985), 'Establishment of dorsal-ventral polarity in the drosophila embryo: the induction of polarity by the toll gene product', *Cell* **42**(3), 791–798.

Anderson, K. V., Jurgens, G. & Nusslein-Volhard, C. (1985), 'Establishment of dorsal-ventral polarity in the drosophila embryo: genetic studies on the role of the toll gene product', *Cell* **42**(3), 779–789.

Andrew, D. J., Baig, A., Bhanot, P., Smolik, S. M. & Henderson, K. D. (1997), 'The drosophila dcreb-a gene is required for dorsal/ventral patterning of the larval cuticle', *Development (Cambridge, England)* **124**(1), 181–193.

Angelats, C., Gallet, A., Therond, P., Fasano, L. & Kerridge, S. (2002), 'Cubitus interruptus acts to specify naked cuticle in the trunk of drosophila embryos', *Developmental biology* **241**(1), 132–144.

Antoniewski, C., Laval, M. & Lepesant, J. A. (1993), 'Structural features critical to the activity of an ecdysone receptor binding site', *Insect biochemistry and molecular biology* **23**(1), 105–114.

Araujo, S. J., Aslam, H., Tear, G. & Casanova, J. (2005), 'mummy/cystic encodes an enzyme required for chitin and glycan synthesis, involved in trachea, embryonic cuticle and cns development–analysis of its role in drosophila tracheal morphogenesis', *Developmental biology* **288**(1), 179–193.

Bate, M. & Arias, A. M. (1993), *The development of Drosophila melanogaster*, Vol. I.

Behr, M., Riedel, D. & Schuh, R. (2003), 'The claudin-like megatrachea is essential in septate junctions for the epithelial barrier function in drosophila', *Developmental cell* **5**(4), 611–620.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004), 'Improved prediction of signal peptides: Signalp 3.0', *Journal of Molecular Biology* **340**(4), 783–795.

Bolatto, C., Chifflet, S., Megighian, A. & Cantera, R. (2003), 'Synaptic activity modifies the levels of dorsal and cactus at the neuromuscular junction of drosophila', *Journal of neurobiology* **54**(3), 525–536.

Bray, S. J. & Kafatos, F. C. (1991), 'Developmental function of elf-1: an essential transcription factor during embryogenesis in drosophila', *Genes & development* **5**(9), 1672–1683.

Cabernard, C., Neumann, M. & Affolter, M. (2004), 'Cellular and molecular mechanisms involved in branching morphogenesis of the drosophila tracheal system', *Journal of applied physiology (Bethesda, Md.: 1985)* **97**(6), 2347–2353.

Caldwell, P. E., Walkiewicz, M. & Stern, M. (2005), 'Ras activity in the drosophila prothoracic gland regulates body size and developmental rate via ecdysone release', *Current biology : CB* **15**(20), 1785–1795.

Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. & Plaza, S. (2006), 'Shavenbaby couples patterning to epidermal cell shape control', *PLoS biology* **4**(9), e290.

Chen, B., Chu, T., Harms, E., Gergen, J. P. & Strickland, S. (1998), 'Mapping of drosophila mutations using site-specific male recombination', *Genetics* **149**(1), 157–163.

Clark, D. V., Suleman, D. S., Beckenbach, K. A., Gilchrist, E. J. & Baillie, D. L. (1995), 'Molecular cloning and characterization of the dpy-20 gene of caenorhabditis elegans', *Molecular & general genetics : MGG* **247**(3), 367–378.

Colombani, J., Bianchini, L., Layalle, S., Pondeville, E., Dauphin-Villemant, C., Antoniewski, C., Carre, C., Noselli, S. & Leopold, P. (2005), 'Antagonistic actions of ecdysone and insulins determine final size in drosophila', *Science* **310**(5748), 667–670.

Crickmore, M. A. & Mann, R. S. (2006), 'Hox control of organ size by regulation of morphogen production and mobility', *Science* **313**(5783), 63–68.

Csikos, G., Molnar, K., Borhegyi, N. H., Talian, G. C. & Sass, M. (1999), 'Insect cuticle, an in vivo model of protein trafficking', *Journal of cell science* **112 (Pt 13)**(Pt 13), 2113–2124.

Dai, X., Schonbaum, C., Degenstein, L., Bai, W., Mahowald, A. & Fuchs, E. (1998), 'The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice', *Genes & development* **12**(21), 3452–3463.

de Navas, L. F., Garaulet, D. L. & Sanchez-Herrero, E. (2006), 'The ultrabithorax hox gene of drosophila controls haltere size by regulating the dpp pathway', *Development (Cambridge, England)* .

Devine, W. P., Lubarsky, B., Shaw, K., Luschnig, S., Messina, L. & Krasnow, M. A. (2005), 'Requirement for chitin biosynthesis in epithelial tube morphogenesis', *Proceedings of the National Academy of Sciences of the United States of America* **102**(47), 17014–17019.

Dickinson, W. J. & Thatcher, J. W. (1997), 'Morphogenesis of denticles and hairs in drosophila embryos: involvement of actin-associated proteins that also affect adult structures', *Cell motility and the cytoskeleton* **38**(1), 9–21.

Duttaroy, A. (2002), 'Asymmetric exchange is associated with p element induced male recombination in drosophila melanogaster', *Heredity* **89**(2), 114–119.

Eisemann, C., Wijffels, G. & Tellam, R. L. (2001), 'Secretion of the type 2 peritrophic matrix protein, peritrophin-15, from the cardia', *Archives of insect biochemistry and physiology* **47**(2), 76–85.

Eveleth, D. D., Gietz, R. D., Spencer, C. A., Nargang, F. E., Hodgetts, R. B. & Marsh, J. L. (1986), 'Sequence and structure of the dopa decarboxylase gene of drosophila: evidence for novel rna splicing variants', *The EMBO journal* **5**(10), 2663–2672.

Fambrough, D. & Goodman, C. S. (1996), 'The drosophila beaten path gene encodes a novel secreted protein that regulates defasciculation at motor axon choice points', *Cell* **87**(6), 1049–1058.

Gagou, M. E., Kapsetaki, M., Turberg, A. & Kafetzopoulos, D. (2002), 'Stage-specific expression of the chitin synthase dmechsa and dmechsb genes during the onset of drosophila metamorphosis', *Insect biochemistry and molecular biology* **32**(2), 141–146.

Galko, M. J. & Krasnow, M. A. (2004), 'Cellular and genetic analysis of wound healing in drosophila larvae', *PLoS biology* **2**(8), E239.

Gallo, M., Mah, A. K., Johnsen, R. C., Rose, A. M. & Baillie, D. L. (2006), 'Caenorhabditis elegans dpy-14: an essential collagen gene with unique expression profile and physiological roles in early development', *Molecular genetics and genomics : MGG* **275**(6), 527–539.

Gerttula, S., Jin, Y. S. & Anderson, K. V. (1988), 'Zygotic expression and activity of the drosophila toll gene, a gene required maternally for embryonic dorsal-ventral pattern formation', *Genetics* **119**(1), 123–133.

Ghabrial, A., Luschnig, S., Metzstein, M. M. & Krasnow, M. A. (2003), 'Branching morphogenesis of the drosophila tracheal system', *Annual Review of Cell and Developmental Biology* **19**, 623–647.

Goldstein, L. & Fyrberg, E. (1994), *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology.*

Gritzan, U., Hatini, V. & DiNardo, S. (1999), 'Mutual antagonism between signals secreted by adjacent wingless and engrailed cells leads to specification of complementary regions of the drosophila parasegment', *Development (Cambridge, England)* **126**(18), 4107–4115.

Halfon, M. S., Hashimoto, C. & Keshishian, H. (1995), 'The drosophila toll gene functions zygotically and is necessary for proper motoneuron and muscle development', *Developmental biology* **169**(1), 151–167.

Hamodrakas, S. J., Willis, J. H. & Iconomidou, V. A. (2002), 'A structural model of the chitin-binding domain of cuticle proteins', *Insect biochemistry and molecular biology* **32**(11), 1577–1583.

Han, J. H., Lee, S. H., Tan, Y. Q., LeMosy, E. K. & Hashimoto, C. (2000), 'Gastrulation defective is a serine protease involved in activating the receptor toll to polarize the drosophila embryo', *Proceedings of the National Academy of Sciences of the United States of America* **97**(16), 9093–9097.

Harden, N. (2005), 'Cell biology. of grainy heads and broken skins', *Science* **308**(5720), 364–365.

Heisenberg, C. P., Tada, M., Rauch, G. J., Saude, L., Concha, M. L., Geisler, R., Stemple, D. L., Smith, J. C. & Wilson, S. W. (2000), 'Silberblick/wnt11 mediates convergent extension movements during zebrafish gastrulation', *Nature* **405**(6782), 76–81.

Hill, K. L., Harfe, B. D., Dobbins, C. A. & L'Hernault, S. W. (2000), 'dpy-18 encodes an alpha-subunit of prolyl-4-hydroxylase in caenorhabditis elegans', *Genetics* **155**(3), 1139–1148.

Hiruma, K., Carter, M. S. & Riddiford, L. M. (1995), 'Characterization of the dopa decarboxylase gene of manduca sexta and its suppression by 20-hydroxyecdysone', *Developmental biology* **169**(1), 195–209.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989), 'Site-directed mutagenesis by overlap extension using the polymerase chain reaction', *Gene* **77**(1), 51–59.

Iconomidou, V. A., Chryssikos, G. D., Gionis, V., Willis, J. H. & Hamodrakas, S. J. (2001), '"soft"-cuticle protein secondary structure as revealed by ft-raman, atr ft-ir and cd spectroscopy', *Insect biochemistry and molecular biology* **31**(9), 877–885.

Iconomidou, V. A., Willis, J. H. & Hamodrakas, S. J. (1999), 'Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle?', *Insect biochemistry and molecular biology* **29**(3), 285–292.

Iconomidou, V. A., Willis, J. H. & Hamodrakas, S. J. (2005), 'Unique features of the structural model of 'hard' cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle', *Insect biochemistry and molecular biology* **35**(6), 553–560.

Iijima, M., Hashimoto, T., Matsuda, Y., Nagai, T., Yamano, Y., Ichi, T., Osaki, T. & Kawabata, S. (2005), 'Comprehensive sequence analysis of horseshoe crab cuticular proteins and their involvement in transglutaminase-dependent cross-linking', *The FEBS journal* **272**(18), 4774–4786.

Johnstone, I. L., Shafi, Y. & Barry, J. D. (1992), 'Molecular analysis of mutations in the caenorhabditis elegans collagen gene dpy-7', *The EMBO journal* **11**(11), 3857–3863.

Kambris, Z., Brun, S., Jang, I. H., Nam, H. J., Romeo, Y., Takahashi, K., Lee, W. J., Ueda, R. & Lemaitre, B. (2006), 'Drosophila immunity: a large-scale in vivo rnai screen identifies five serine proteases required for toll activation', *Current biology : CB* **16**(8), 808–813.

Karlsson, C., Korayem, A. M., Scherfer, C., Loseva, O., Dushay, M. S. & Theopold, U. (2004), 'Proteomic analysis of the drosophila larval hemolymph clot', *The Journal of biological chemistry* **279**(50), 52033–52041.

Kayser, H. & Palivan, C. G. (2006), 'Stable free radicals in insect cuticles: electron spin resonance spectroscopy reveals differences between melanization and sclerotization', *Archives of Biochemistry and Biophysics* **453**(2), 179–187.

Kramer, J. M., French, R. P., Park, E. C. & Johnson, J. J. (1990), 'The caenorhabditis elegans rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen', *Molecular and cellular biology* **10**(5), 2081–2089.

Kramer, J. M., Johnson, J. J., Edgar, R. S., Basch, C. & Roberts, S. (1988), 'The sqt-1 gene of c. elegans encodes a collagen critical for organismal morphogenesis', *Cell* **55**(4), 555–565.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. (1996), 'The dorsoventral regulatory gene cassette spatzle/toll/cactus controls the potent antifungal response in drosophila adults', *Cell* **86**(6), 973–983.

Letsou, A., Alexander, S., Orth, K. & Wasserman, S. A. (1991), 'Genetic and molecular characterization of tube, a drosophila gene maternally required for embryonic dorsoventral polarity', *Proceedings of the National Academy of Sciences of the United States of America* **88**(3), 810–814.

Levy, A. D., Yang, J. & Kramer, J. M. (1993), 'Molecular and genetic analyses of the caenorhabditis elegans dpy-2 and dpy-10 collagen genes: a variety of molecular alterations affect organismal morphology', *Molecular biology of the cell* **4**(8), 803–817.

Ligoxygakis, P., Pelte, N., Hoffmann, J. A. & Reichhart, J. M. (2002), 'Activation of drosophila toll during fungal infection by a blood serine protease', *Science* **297**(5578), 114–116.

Lindsley, D. L. (1973), Drosophila information service, Technical Report 50.

Llimargas, M. & Lawrence, P. A. (2001), 'Seven wnt homologues in drosophila: a case study of the developing tracheae', *Proceedings of the National Academy of Sciences of the United States of America* **98**(25), 14487–14492.

Lohmann, C. M. & Riddiford, L. M. (1992), 'Synthesis and secretion of low molecular weight cuticular proteins during heat shock in the tobacco hornworm, manduca sexta', *The Journal of experimental zoology* **262**(4), 374–382.

Luo, Y., Amin, J. & Voellmy, R. (1991), 'Ecdysterone receptor is a sequence-specific transcription factor involved in the developmental regulation of heat shock genes', *Molecular and cellular biology* **11**(7), 3660–3675.

Luschnig, S., Batz, T., Armbruster, K. & Krasnow, M. A. (2006), 'serpentine and vermiform encode matrix proteins with chitin binding and deacetylation domains that limit tracheal tube length in drosophila', *Current biology : CB* **16**(2), 186–194.

Mace, K. A., Pearson, J. C. & McGinnis, W. (2005), 'An epidermal barrier wound repair pathway in drosophila is mediated by grainy head', *Science* **308**(5720), 381–385.

Magkrioti, C. K., Spyropoulos, I. C., Iconomidou, V. A., Willis, J. H. & Hamodrakas, S. J. (2004), 'cuticledb: a relational database of arthropod cuticular proteins', *BMC bioinformatics [electronic resource]* **5**, 138.

Makhijani, K., Kalyani, C., Srividya, T. & Shashidhara, L. S. (2006), 'Modulation of decapentaplegic gradient during haltere specification in drosophila', *Developmental biology* .

Merzendorfer, H. (2006), 'Insect chitin synthases: a review', *Journal of comparative physiology.B, Biochemical, systemic, and environmental physiology* **176**(1), 1–15.

Mirth, C., Truman, J. W. & Riddiford, L. M. (2005), 'The role of the prothoracic gland in determining critical weight for metamorphosis in drosophila melanogaster', *Current biology : CB* **15**(20), 1796–1807.

Moussian, B., Schwarz, H., Bartoszewski, S. & Nusslein-Volhard, C. (2005), 'Involvement of chitin in exoskeleton morphogenesis in drosophila melanogaster', *Journal of Morphology* **264**(1), 117–130.

Moussian, B., Soding, J., Schwarz, H. & Nusslein-Volhard, C. (2005), 'Retroactive, a membrane-anchored extracellular protein related to vertebrate snake neurotoxin-like proteins, is required for cuticle organization in the larva of drosophila melanogaster', *Developmental dynamics : an official publication of the American Association of Anatomists* **233**(3), 1056–1063.

Moussian, B., Tang, E., Tonning, A., Helms, S., Schwarz, H., Nusslein-Volhard, C. & Uv, A. E. (2006), 'Drosophila knickkopf and retroactive are needed for epithelial tube growth and cuticle differentiation through their specific requirement for chitin filament organization', *Development* **133**(1), 163–171.

Moussian, B. & Uv, A. E. (2005), 'An ancient control of epithelial barrier formation and wound healing', *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**(10), 987–990.

Murata, T., Kageyama, Y., Hirose, S. & Ueda, H. (1996), 'Regulation of the edg84a gene by ftz-f1 during metamorphosis in drosophila melanogaster', *Molecular and cellular biology* **16**(11), 6509–6515.

Myat, M. M., Lightfoot, H., Wang, P. & Andrew, D. J. (2005), 'A molecular link between fgf and dpp signaling in branch-specific migration of the drosophila trachea', *Developmental biology* **281**(1), 38–52.

Myllyharju, J. & Kivirikko, K. I. (2004), 'Collagens, modifying enzymes and their mutations in humans, flies and worms', *Trends in genetics : TIG* **20**(1), 33–43.

Myllyharju, J., Kukkola, L., Winter, A. D. & Page, A. P. (2002), 'The exoskeleton collagens in caenorhabditis elegans are modified by prolyl 4-hydroxylases with unique combinations of subunits', *The Journal of biological chemistry* **277**(32), 29187–29196.

Neckameyer, W. S. & White, K. (1993), 'Drosophila tyrosine hydroxylase is encoded by the pale locus', *Journal of neurogenetics* **8**(4), 189–199.

Neufeld, T. P. (2003), 'Body building: regulation of shape and size by pi3k/tor signaling during development', *Mechanisms of development* **120**(11), 1283–1296.

Ninomiya, H., Elinson, R. P. & Winklbauer, R. (2004), 'Antero-posterior tissue polarity links mesoderm convergent extension to axial patterning', *Nature* **430**(6997), 364–367.

Novelli, J., Ahmed, S. & Hodgkin, J. (2004), 'Gene interactions in caenorhabditis elegans define dpy-31 as a candidate procollagen c-proteinase and sqt-3/rol-4 as its predicted major target', *Genetics* **168**(3), 1259–1273.

Nystrom, J., Shen, Z. Z., Aili, M., Flemming, A. J., Leroi, A. & Tuck, S. (2002), 'Increased or decreased levels of caenorhabditis elegans lon-3, a gene encoding a collagen, cause reciprocal changes in body length', *Genetics* **161**(1), 83–97.

Paul, S. M., Ternet, M., Salvaterra, P. M. & Beitel, G. J. (2003), 'The na+/k+ atpase is required for septate junction function and epithelial tube-size control in the drosophila tracheal system', *Development* **130**(20), 4963–4974.

Payre, F. (2004), 'Genetic control of epidermis differentiation in drosophila', *The International journal of developmental biology* **48**(2-3), 207–215.

Payre, F., Vincent, A. & Carreno, S. (1999), 'ovo/svb integrates wingless and der pathways to control epidermis differentiation', *Nature* **400**(6741), 271–275.

Pentz, E. S., Black, B. C. & Wright, T. R. (1986), 'A diphenol oxidase gene is part of a cluster of genes involved in catecholamine metabolism and sclerotization in drosophila. i. identification of the biochemical defect in dox-a2 [l(2)37bf] mutants', *Genetics* **112**(4), 823–841.

Pentz, E. S., Black, B. C. & Wright, T. R. (1990), 'Mutations affecting phenol oxidase activity in drosophila: quicksilver and tyrosinase-1', *Biochemical genetics* **28**(3-4), 151–171.

Petit, V., Ribeiro, C., Ebner, A. & Affolter, M. (2002), 'Regulation of cell migration during tracheal development in drosophila melanogaster', *The International journal of developmental biology* **46**(1), 125–132.

Pipes, G. C., Lin, Q., Riley, S. E. & Goodman, C. S. (2001), 'The beat generation: a multigene family encoding igsf proteins related to the beat axon guidance molecule in drosophila', *Development (Cambridge, England)* **128**(22), 4545–4552.

Preston, C. R. & Engels, W. R. (1996), 'P-element-induced male recombination and gene conversion in drosophila', *Genetics* **144**(4), 1611–1622.

Preston, C. R., Sved, J. A. & Engels, W. R. (1996), 'Flanking duplications and deletions associated with p-induced male recombination in drosophila', *Genetics* **144**(4), 1623–1638.

Qiu, P., Pan, P. C. & Govind, S. (1998), 'A role for the drosophila toll/cactus pathway in larval hematopoiesis', *Development (Cambridge, England)* **125**(10), 1909–1920.

Ricketts, D. & Sugumaran, M. (1994), '1,2-dehydro-n-beta-alanyldopamine as a new intermediate in insect cuticular sclerotization', *The Journal of biological chemistry* **269**(35), 22217–22221.

Riddiford, L. M., Hiruma, K., Zhou, X. & Nelson, C. A. (2003), 'Insights into the molecular basis of the hormonal control of molting and metamorphosis from manduca sexta and drosophila melanogaster', *Insect biochemistry and molecular biology* **33**(12), 1327–1338.

Robertson, A. S., Belorgey, D., Lilley, K. S., Lomas, D. A., Gubb, D. & Dafforn, T. R. (2003), 'Characterization of the necrotic protein that regulates the toll-mediated immune response in drosophila', *The Journal of biological chemistry* **278**(8), 6175–6180.

Rose, D. & Chiba, A. (1999), 'A single growth cone is capable of integrating simultaneously presented and functionally distinct molecular cues during target recognition', *The Journal of neuroscience : the official journal of the Society for Neuroscience* **19**(12), 4899–4906.

Rose, D., Zhu, X., Kose, H., Hoang, B., Cho, J. & Chiba, A. (1997), 'Toll, a muscle cell surface molecule, locally inhibits synaptic initiation of the rp3 motoneuron growth cone in drosophila', *Development (Cambridge, England)* **124**(8), 1561–1571.

Rost, B. (1996), 'Phd: predicting one-dimensional protein structure by profile-based neural networks', *Methods in enzymology* **266**, 525–539.

Rost, B., Yachdav, G. & Liu, J. (2003), 'The predictprotein server', *Nucleic Acids Research* **32(Web Server issue)**, W321–W326.

Stahl, M., Schuh, R. & Adryan, B. (2006), 'Identification of fgf-dependent genes in the drosophila tracheal system', *Gene Expr.Patterns* .

Stern, D. L. (2006), 'Developmental biology. morphing into shape', *Science* **313**(5783), 50–51.

Stramer, B. & Martin, P. (2005), 'Cell biology: master regulators of sealing and healing', *Current biology : CB* **15**(11), R425–7.

Stronach, B. E., Renfranz, P. J., Lilly, B. & Beckerle, M. C. (1999), 'Muscle lim proteins are associated with muscle sarcomeres and require dmef2 for their expression during drosophila myogenesis', *Molecular biology of the cell* **10**(7), 2329–2342.

Stronach, B. E., Siegrist, S. E. & Beckerle, M. C. (1996), 'Two muscle-specific lim proteins in drosophila', *The Journal of cell biology* **134**(5), 1179–1195.

Suderman, R. J., Dittmer, N. T., Kanost, M. R. & Kramer, K. J. (2006), 'Model reactions for insect cuticle sclerotization: cross-linking of recombinant cuticular proteins upon their laccase-catalyzed oxidative conjugation with catechols', *Insect biochemistry and molecular biology* **36**(4), 353–365.

Sugumaran, M., Nellaiappan, K. & Valivittan, K. (2000), 'A new mechanism for the control of phenoloxidase activity: inhibition and complex formation with quinone isomerase', *Archives of Biochemistry and Biophysics* **379**(2), 252–260.

Suzuki, Y., Morris, G. A., Han, M. & Wood, W. B. (2002), 'A cuticle collagen encoded by the lon-3 gene may be a target of tgf-beta signaling in determining caenorhabditis elegans body shape', *Genetics* **162**(4), 1631–1639.

Sved, J. A., Blackman, L. M., Gilchrist, A. S. & Engels, W. R. (1991), 'High levels of recombination induced by homologous p elements in drosophila melanogaster', *Molecular & general genetics : MGG* **225**(3), 443–447.

Sved, J. A., Eggleston, W. B. & Engels, W. R. (1990), 'Germ-line and somatic recombination induced by in vitro modified p elements in drosophila melanogaster', *Genetics* **124**(2), 331–337.

Svoboda, Y. H., Robson, M. K. & Sved, J. A. (1995), 'P-element-induced male recombination can be produced in drosophila melanogaster by combining end-deficient elements in trans', *Genetics* **139**(4), 1601–1610.

Swanson, L. E. & Beitel, G. J. (2006), 'Tubulogenesis: an inside job', *Current biology : CB* **16**(2), R51–3.

Tellam, R. L., Vuocolo, T., Eisemann, C., Briscoe, S., Riding, G., Elvin, C. & Pearson, R. (2003), 'Identification of an immuno-protective mucin-like protein, peritrophin-55, from the peritrophic matrix of lucilia cuprina larvae', *Insect biochemistry and molecular biology* **33**(2), 239–252.

Tellam, R. L., Wijffels, G. & Willadsen, P. (1999), 'Peritrophic matrix proteins', *Insect biochemistry and molecular biology* **29**(2), 87–101.

Thacker, C., Sheps, J. A. & Rose, A. M. (2006), 'Caenorhabditis elegans dpy-5 is a cuticle procollagen processed by a proprotein convertase', *Cellular and molecular life sciences : CMLS* **63**(10), 1193–1204.

Thompson, D. W. (1917), *On Growth and Form*, Cambridge Univ. Press, Cambridge.

Thummel, C. S., Boulet, A. M. & Lipshitz, H. D. (1988), 'Vectors for drosophila p-element-mediated transformation and tissue culture transfection', *Gene* **74**(2), 445–456.

Ting, S. B., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L. L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W. M., Elias, P. M., Cunningham, J. M. & Jane, S. M. (2005), 'A homolog of drosophila grainy head is essential for epidermal integrity in mice', *Science* **308**(5720), 411–413.

Tonning, A., Helms, S., Schwarz, H., Uv, A. E. & Moussian, B. (2006), 'Hormonal regulation of mummy is needed for apical extracellular matrix formation and epithelial morphogenesis in drosophila', *Development (Cambridge, England)* **133**(2), 331–341.

Tonning, A., Hemphala, J., Tang, E., Nannmark, U., Samakovlis, C. & Uv, A. (2005), 'A transient luminal chitinous matrix is required to model epithelial tube diameter in the drosophila trachea', *Developmental cell* **9**(3), 423–430.

Uv, A., Cantera, R. & Samakovlis, C. (2003), 'Drosophila tracheal morphogenesis: intricate cellular solutions to basic plumbing problems', *Trends in cell biology* **13**(6), 301–309.

von Mende, N., Bird, D. M., Albert, P. S. & Riddle, D. L. (1988), 'dpy-13: a nematode collagen gene that affects body shape', *Cell* **55**(4), 567–576.

Vuocolo, T., Eisemann, C. H., Pearson, R. D., Willadsen, P. & Tellam, R. L. (2001), 'Identification and molecular characterisation of a peritrophin gene, peritrophin-48, from the myiasis fly chrysomya bezziana', *Insect biochemistry and molecular biology* **31**(9), 919–932.

Wallingford, J. B., Vogeli, K. M. & Harland, R. M. (2001), 'Regulation of convergent extension in xenopus by wnt5a and frizzled-8 is independent of the canonical wnt pathway', *The International journal of developmental biology* **45**(1), 225–227.

Walter, M. F., Black, B. C., Afshar, G., Kermabon, A. Y., Wright, T. R. & Biessmann, H. (1991), 'Temporal and spatial expression of the yellow gene in correlation with cuticle formation and dopa decarboxylase activity in drosophila development', *Developmental biology* **147**(1), 32–45.

Walters, J. W., Munoz, C., Paaby, A. B. & Dinardo, S. (2005), 'Serrate-notch signaling defines the scope of the initial denticle field by modulating egfr activation', *Developmental biology* **286**(2), 415–426.

Wang, B., Sullivan, K. M. & Beckingham, K. (2003), 'Drosophila calmodulin mutants with specific defects in the musculature or in the nervous system', *Genetics* **165**(3), 1255–1268.

Wang, S., Jayaram, S. A., Hemphala, J., Senti, K. A., Tsarouhas, V., Jin, H. & Samakovlis, C. (2006), 'Septate-junction-dependent luminal deposition of chitin deacetylases restricts tube elongation in the drosophila trachea', *Current biology : CB* **16**(2), 180–185.

Wesley, C. S. (1999), 'Notch and wingless regulate expression of cuticle patterning genes', *Molecular and cellular biology* **19**(8), 5743–5758.

Wu, V. M. & Beitel, G. J. (2004), 'A junctional problem of apical proportions: epithelial tube-size control by septate junctions in the drosophila tracheal system', *Current opinion in cell biology* **16**(5), 493–499.

Wu, V. M., Schulte, J., Hirschi, A., Tepass, U. & Beitel, G. J. (2004), 'Sinuous is a drosophila claudin required for septate junction organization and epithelial tube size control', *The Journal of cell biology* **164**(2), 313–323.

Yang, J. & Kramer, J. M. (1994), 'In vitro mutagenesis of caenorhabditis elegans cuticle collagens identifies a potential subtilisin-like protease cleavage site and demonstrates that carboxyl domain disulfide bonding is required for normal function but not assembly', *Molecular and cellular biology* **14**(4), 2722–2730.

Zhong, Y. S., Mita, K., Shimada, T. & Kawasaki, H. (2006), 'Glycine-rich protein genes, which encode a major component of the cuticle, have different developmental profiles from other cuticle protein genes in bombyx mori', *Insect biochemistry and molecular biology* **36**(2), 99–110.