# *Drosophila* Dumpy is a gigantic extracellular protein required to maintain tension at epidermal–cuticle attachment sites

M.B. Wilkin*†, M.N. Becker†‡, D. Mulvey§, I. Phan§, A. Chao‡, K. Cooper‡, H-J. Chung‡, I.D. Campbell§, M. Baron* and R. MacIntyre‡

**Background:** Growth and morphogenesis during development depend both on patterning genes, which assign positional information, and on genes that regulate mechanical forces. The *dumpy* gene of the fruit fly *Drosophila melanogaster* is an example of the latter class, with mutant phenotypes affecting size and shape of the limbs, thoracic cuticle, trachea and mouthparts.

**Results:** The genetically complex *dumpy* locus was found to span over 100 kb and encode a gigantic 2.5 MDa extracellular matrix protein. Dumpy represents an extreme form of modular protein evolution, containing 308 epidermal growth factor (EGF) modules, interspersed with a new module class, DPY, and terminating in a crosslinking zona pellucida domain and membrane anchor sequence. We determined the three-dimensional structure of the DPY module by nuclear magnetic resonance (NMR) spectroscopy and found that it forms a disulphide-stabilised β sheet motif, capable of linking end-to-end with EGF modules to form a fibre. Consistent with its cuticle phenotypes, *dumpy* is expressed at several sites of cuticle–epidermal cell attachment, including the trachea and the muscle tendon cells, which mediate anchorage of the muscles to the cuticle.

**Conclusions:** The *dumpy* gene encodes a gigantic extracellular molecule that we predict to be a membrane-anchored fibre of almost a micrometer in length. Insertion and crosslinking of this fibre within the cuticle may provide a strong anchor for the underlying tissue, allowing it to maintain mechanical tension at sites under stress. This would explain its contribution to tissue morphogenesis through its regulation of mechanical properties.

Addresses: *University of Manchester, School of Biological Sciences, Stopford Building, Oxford Road, Manchester M13 9PT, UK. ‡Biotechnology Building, Cornell University, Ithaca, New York 14853-2703, USA. §University of Oxford, Department of Biological Sciences, South Parks Road, Oxford OX1 3QU, UK.

†M.B.W. and M.N.B. contributed equally to this work.

Correspondence: Martin Baron and Ross MacIntyre
E-mail: mbaron@man.ac.uk; rjm18@cornell.edu

Received: **22 December 1999**
Revised: **22 March 2000**
Accepted: **24 March 2000**

Published: **3 May 2000**

**Current Biology** 2000, 10:559–567

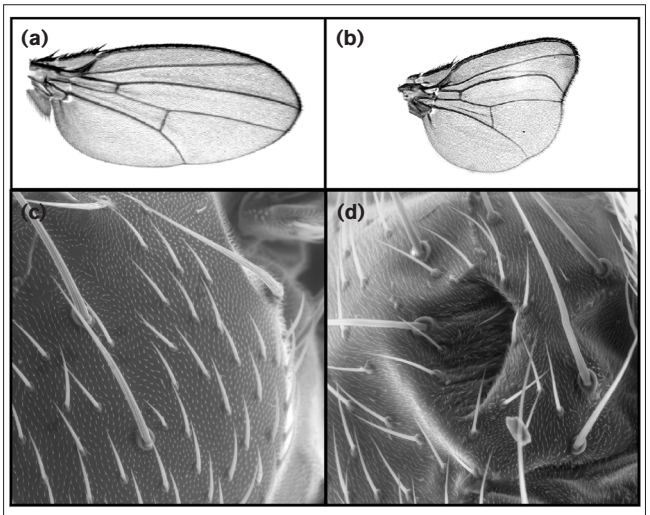
0960-9822/00/$ – see front matter


## Background

The extracellular matrix (ECM) is intimately involved in the control of epithelial morphogenesis through its mechanical properties and its adhesive and signalling interactions with the underlying cells. The insect cuticle is a highly organised, chitinaceous ECM whose laminate structure surrounds external tissues and limbs and certain internal organs such as the lumen of the trachea [1–3]. The importance of the cuticle in determining the body shape of invertebrates has been clearly demonstrated in nematodes where many gross morphological abnormalities are caused by mutations in cuticular collagens [4]. Further, the requirement to moult and reform cuticle at specific stages of development reflects the role that the cuticle has in restricting growth. The variation of the mechanical properties of the cuticle allows different degrees of flexibility in different tissues and, thus, may influence tissue, limb and organ shape and size. The mechanisms by which epidermal cells control the properties of the overlying cuticle and vice versa are poorly understood. The *Drosophila melanogaster dumpy* (*dp*) gene is expected to be a component of this process, as its mutations affect growth and morphogenesis and also cuticle composition and function [5–8].

The *dumpy* locus is genetically complex, with alleles that fall into three main classes: $dp^{o}$, $dp^{v}$ and $dp^{l}$. Other mutant alleles may exhibit combinations of *o*, *v* and *l* traits, with heteroallelic combinations displaying phenotypes that are designated by the traits shared between them [9]. Flies carrying the recessive viable $dp^{o}$ alleles have shortened wings and altered paths of the wing veins (Figure 1b) resulting from contraction of the pupal wing epithelia [8]. This contraction occurs after patterning of the wing has been established and leads to both reduced cell size and altered cell positions [8]. Similarly, alterations in leg size are observed with certain *dp* alleles [10]. Flies carrying $dp^{v}$ alleles display pits and protrusions in the thoracic cuticle, with disruption of the orientation of the surrounding sensory bristles (Figure 1d). These sites coincide with locations where certain flight muscles normally attach to the cuticle wall through epidermal tendon cells [6]. The $dp^{l}$ alleles are a broad class of recessive lethal alleles affecting various developmental stages [7]. Some embryo

**Figure 1**




The adult phenotypes of *dumpy*. **(a)** Wild-type wing. **(b)** Flies carrying the $dp^o$ alleles have shortened wings, and altered paths of the wing veins caused by abnormal contraction of the wing epithelia during metamorphosis. **(c)** Wild-type notum. **(d)** Flies carrying the $dp^v$ alleles display vortex-like depressions on the notum, coinciding with the locations of certain muscle attachment sites. The orientations of the bristles surrounding the vortex are disturbed.

lethal alleles lead to structural defects in the cuticle around the lumen of the developing trachea, which fail to fill with air. These embryos also display little muscle activity. The trachea and mouthparts of certain larval lethal mutants grow out of proportion to the remainder of the body, showing that *dumpy* normally restricts growth in these tissues. A group of larval lethal alleles lead to death during moulting when the interface between the cuticle and the epidermis fails to break down and the cuticle cannot be shed properly. Clones of certain $dp^l$ alleles cause blistering phenotypes in the wing that reflect failure of adhesion between the dorsal and ventral surfaces [11] and show that *dumpy* has additional roles in wing morphogenesis.

The regulation of morphogenesis during development and changes in these processes during evolution cannot only be explained by patterning genes which define positional identity. The alteration of wing shape in $dp^o$ alleles illustrates how post-patterning processes may make mechanical contributions to shape and size. Here, we report that the *dumpy* locus encodes a gigantic, extracellular matrix molecule that may organise cuticle structure and anchor cells to the cuticle through a long fibre of repeating modular units. Regions of low complexity sequence may confer elasticity on the molecule, and Dumpy is thus well placed to control mechanical tension in the underlying epidermis.

## Results

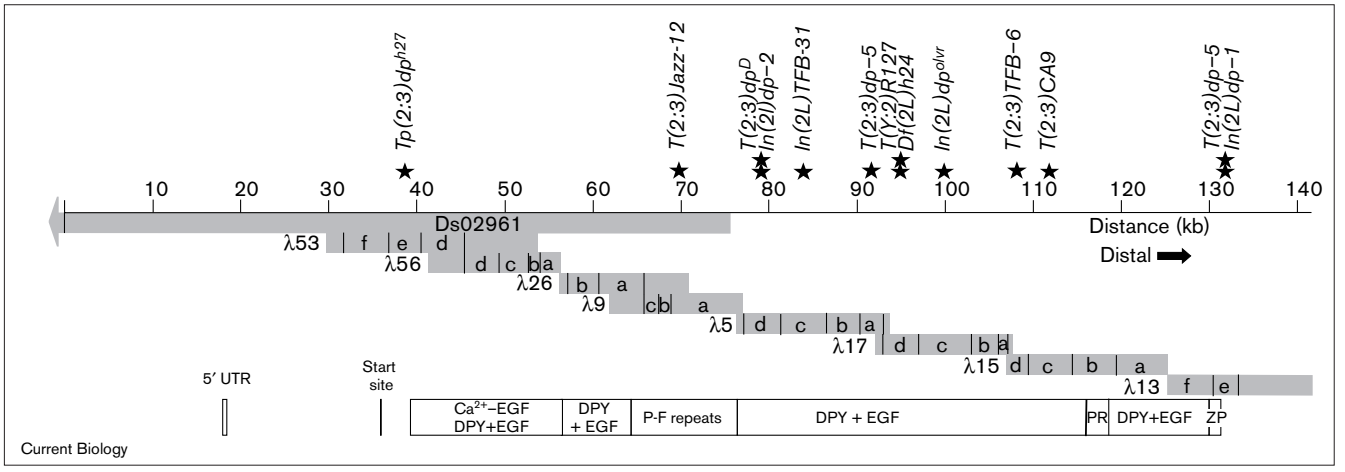
### Mapping and cloning of the *dumpy* locus

The *dumpy* gene spans a genetic distance of 0.1–0.2 centimorgans [12] and is located proximal to the breakpoint of *In(2L)dpp6* at 24F1-2 and distal to the breakpoint of *T(Y:2)H116* at 25A1-2 [13]. We performed a chromosomal walk between these locations and mapped onto it the breakpoint positions for chromosome aberrations that were mutant with regard to *dumpy* (Figure 2). The breakpoints in the *dumpy* locus are spread throughout a 100 kb region of genomic DNA. Whereas the size of the *dumpy* locus is not unprecedented for *Drosophila* genes, most large loci contain either extensive control regions or have enormous introns. In contrast, *dumpy* is remarkable for being mainly composed of an open reading frame encoding many hundreds of small protein modules. We assembled a full-length sequence of the *dumpy* gene through identification of partial cDNA clones and analysis of genomic sequence data (both our own and those of the *Drosophila* genome project). We confirmed by analysis of cDNA fragments that all in-frame stop codons within the genomic DNA sequence lay within introns. We also showed that neighbouring exons, which flanked all predicted introns, were spliced together. Thus, the *dumpy* locus is predicted to encode a 70 kb mRNA transcript with an open reading frame encoding a 2.5 MDa protein (Figure 3a). The most proximal *dumpy* mutant breakpoint, *Tp(2:3)h27*, is in subclone 53E, close to the 5′ end of the product, whilst the most distal *dumpy* mutant breakpoint, *In(2L)dp-l*, lies in the region covered by subclone 13F, close to the 3′ end of the gene.

### The modular structure of Dumpy

The modular structure of the predicted Dumpy protein is summarised in Figure 3a. Near the carboxyl terminus of the protein, there is a single zona pellucida (ZP) domain [14,15]. ZP domains mediate covalent crosslinking in ECM proteins, as in the sperm acrosome reaction [14] and in filament formation by α and β tectorins of the vertebrate inner ear [16] and *Caenorhabditis elegans* Cuticlin-1 [17]. There is a carboxy-terminal membrane-spanning anchor sequence and short cytoplasmic tail, immediately following the Dumpy ZP domain (Figure 3f). Membrane localisation may be closely linked with ZP domain function as most other ZP domain proteins have similar, or glycosyl-phosphatidylinositol, membrane anchors [14]. The region amino-terminal to the ZP domain comprises hundreds of epidermal growth factor (EGF) modules, an extracellular module class that has been identified in many extracellular matrix and cell-surface proteins [18]. The 308 EGF modules in Dumpy are interspersed with a novel repeat of 21 amino acids, which we have termed the DPY module (Figure 3a,g). Most of this region is highly regular, being composed of contiguous repeats of a three-module EGF–DPY–EGF unit, which also exists in a second *Drosophila* gene that we have termed *mini-dumpy*

**Figure 2**




Chromosomal walk spanning the *dumpy* locus. Shaded boxes, subclones of isolated phages; stars, breakpoint positions of chromosome aberrations in the regions 24F2–25A1–4 that cause $dp^{o/v}$ phenotypes. The region between 13E and 5A is inverted or deleted in *T(2:3)dp-5* and the region between 15C and 15D is inverted or deleted in *T(2:3)CA9*. The breakpoint of *T(2:3)Jazz-12* cannot be located precisely as this aberration is broken in one of the highly conserved tandemly linked P-F repeats, which cover a region encompassing subclones 9A–26A in the chromosomal walk. The breakpoint positions are indicated relative to regions of the *dumpy* gene whose modular composition is described in detail in Figure 3. The 5′ untranslated region (UTR) and the start site of protein translation are found on small exons separated from the rest of the gene by large introns as indicated. The positions of the remaining introns are shown on Figure 3.

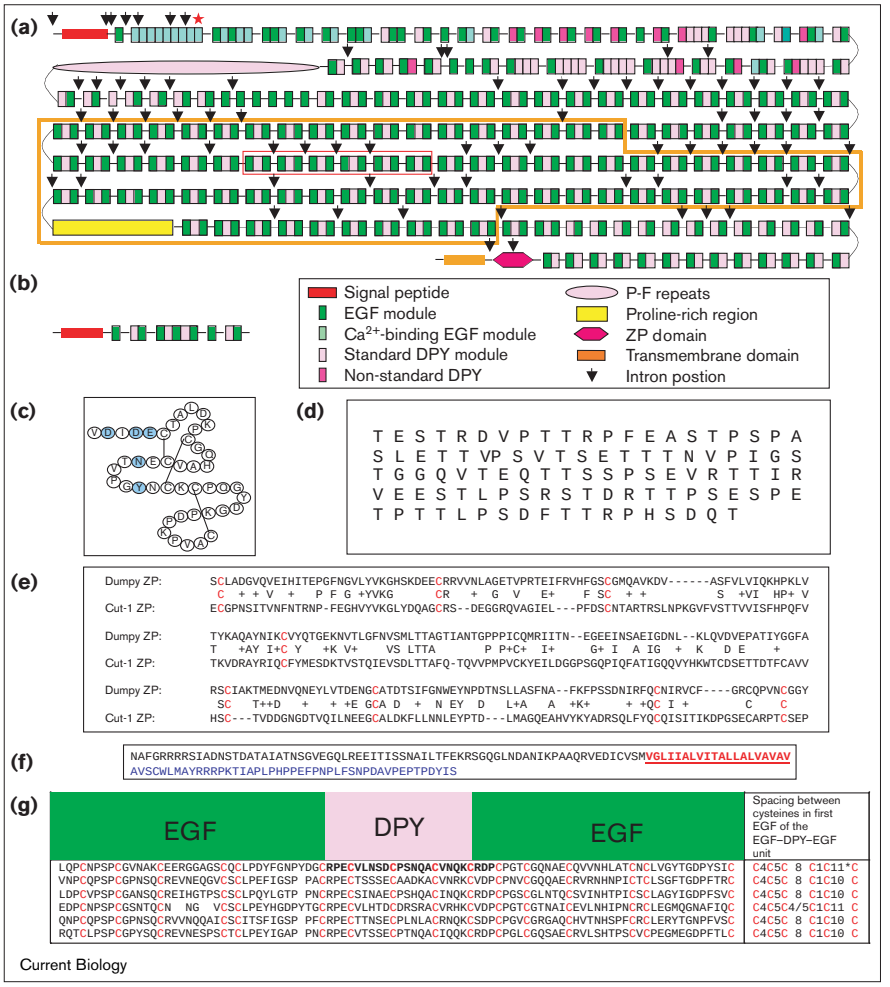
(Figure 3b). Analysis of the cysteine-spacing pattern of this region of Dumpy (Figure 3a,g) revealed thirteen 'super' repeats, each consisting of six EGF–DPY–EGF units, which may reflect a higher-order structural and functional unit of organisation. This super-repeat region is interrupted by an insert of a repetitive proline-rich sequence (PR), but continues distal to it as if unbroken (Figure 3a). Other EGF/DPY module combinations are found on both sides of the super-repeat region. These include stretches of EGF–DPY modules with no EGF–EGF interface and tandemly repeated EGF or DPY modules. The presence of DPY modules in different sequence contexts suggests that, despite its small size, this module is an autonomous folded structure, rather than being part of an unusual extended EGF module.

The EGF/DPY modular structure is interrupted amino-terminal to the super-repeat region by over 30 copies of a novel, 101 amino acid, threonine- and serine-rich sequence, which we term the 'PIGS-FEAST' (P-F) repeat. The low complexity of the P-F primary sequence (Figure 3d) means that it is unlikely to adopt a globular fold. Further EGF and DPY modules are found amino-terminal to the P-F repeats. These include a sub-class of EGF modules (Figure 3c) with a consensus $Ca^{2+}$-binding site [19,20]. Similar '$Ca^{2+}$–EGF' modules form fibrous structures through homotypic interactions in the connective tissue protein fibrillin [21,22] and mediate heterotypic protein–protein interactions between Notch and its ligands [23]. Finally, there is an export signal sequence at the amino terminus of the protein.

To gain an insight into the structural organisation of the EGF–DPY–EGF region, which represents the majority of the protein, we have determined the three-dimensional structure in solution of a representative DPY module (indicated in Figure 3g) by homonuclear nuclear magnetic resonance (NMR) spectroscopy (Figure 4). The DPY module forms a compact fold, consisting of a small amino-terminal $3_{10}$ helix packed against a double-stranded anti-parallel β sheet, with two conserved disulphide bridges forming the module core in a 1–3, 2–4 'cross' disulphide pattern. The three-dimensional fold of the DPY module is similar to the conserved core structure of the WAP domain class of protease inhibitors [24] but lacks the amino-terminal loop in WAP, which normally binds to the protease active site.

Knowledge of the EGF and DPY structures allows simple models of the EGF–DPY–EGF unit to be constructed. Importantly, as with EGF modules, the amino and carboxyl termini of DPY were found to lie at opposite ends of the module. This would allow these modules to be connected in an end-to-end arrangement. Within the EGF–DPY–EGF unit, just three highly conserved residues lie between the terminal cysteines of the EGF and DPY modules (RPE between EGF1 and DPY, and XDP between DPY and EGF2). Inspection of the linker regions and patterns of conservation at the interface suggest to us that these three modules form a relatively rigid, extended structure of at least 5 nm in length, with DPY contributing 1.4 nm. As EGF–EGF module pairs

**Figure 3**



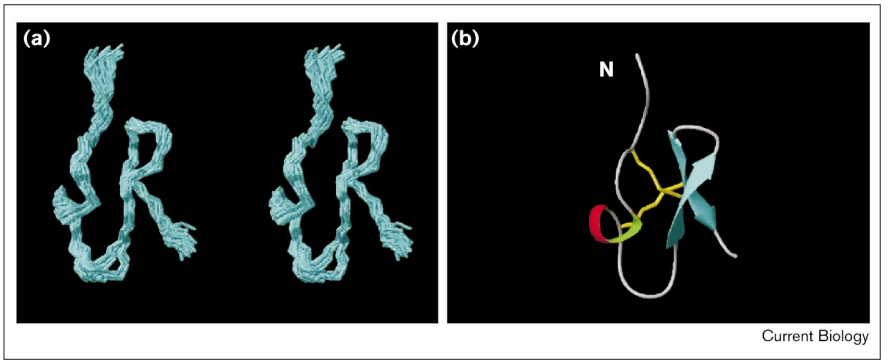

Modular organisation of Dumpy. **(a)** Representation of the modular arrangement of Dumpy. There are 308 EGF and 185 DPY modules. Some of the EGF modules at the amino terminus are predicted to bind $Ca^{2+}$. Most of the DPY modules have the cysteine-spacing pattern $_3C_5C_5C_4C$. Dumpy also contains two less-structured regions, which may confer flexibility to the fibre. These are tandem copies of the highly conserved 101 amino acid serine/threonine-rich P-F repeat (see below) and a 750 amino acid proline-rich region that includes 26 repeats of variable length containing a consensus PGVINIPSVPQP motif. At the amino terminus, there is a predicted cleaved signal peptide and, at the carboxyl terminus, there is a putative transmembrane domain preceded by a ZP domain (see below). Black horizontal lines, linker sequences between modules of at least one residue; arrows, positions of introns, the majority of which occur at module boundaries; star, representative $Ca^{2+}$-binding EGF module whose sequence is shown below; red box, region containing six EGF–DPY–EGF units whose sequence is shown below; orange box, super repeat region, which is described below. **(b)** The domain composition of the 441 amino acid Mini-dumpy, which we located on the X chromosome at 19A by *in situ* hybridisation to salivary gland polytene chromosomes. **(c)** An example of a $Ca^{2+}$-binding EGF module, with the amino acids involved in the predicted $Ca^{2+}$-binding site highlighted in blue [21]. **(d)** The consensus sequence of the P-F repeat derived from 26 repeats from the phage P1 genomic clone DS04146. **(e)** Alignment of the ZP domain from Dumpy with that of Cut-1, a fibrous cuticle component of *C. elegans*. The eight cysteines that are present in almost all ZP domains are highlighted in red. **(f)** Amino-acid sequence carboxy-terminal to the ZP domain, showing the predicted transmembrane sequence in red and the short cytoplasmic tail in blue. **(g)** Alignment of six consecutive EGF–DPY–EGF units. These six repeats make up one unit of a super repeat occurring 13 times through Dumpy. The super repeat was identified on the basis of the spacing between the cysteines in the first EGF repeat of the EGF–DPY–EGF unit. An asterisk indicates that, in the first super repeat, the spacing is 14 residues and not 11. The DPY repeat whose structure has been elucidated by NMR is highlighted in bold.

have previously been shown to form end-to-end structures [22], these considerations suggest that a full-length Dumpy molecule could form a fibrous structure extending for at least 0.8 μm in length. This estimate excludes the lengths of the P-F and PR repeats, which are regions likely to possess little globular structure and may therefore be capable of significant extension. Further flexibility might lie within the EGF–EGF linker regions, which are variable in length and sequence, frequently being proline rich.

### The developmental expression pattern of *dumpy*

The phenotypes of *dumpy* indicate that this gene has multiple roles during development. To investigate the expression pattern of *dumpy*, we performed *in situ* hybridisation on *Drosophila* embryos, using probes derived from cDNA and genomic subclones (Figure 5). The first detectable expression of *dumpy* was in a band across the embryo during early gastrulation (Figure 5a). Following the onset of germ-band elongation (stage 11), expression of *dumpy* was observed in tracheal placodes (Figure 5b–d) and continued as the trachea elongated and branched out [2]. During stage 11, we also observed expression in pockets of invaginating ectoderm, which produce the tube-forming structures of the salivary glands, foregut and hindgut. Following germ-band retraction, *dumpy* was expressed in the muscle tendon cells that mediate muscle attachment to the cuticle bodywall (Figure 5e). Using cDNA and genomic subclones, which were derived from locations throughout the locus, we found that all regions of the gene display similar developmental expression profiles during embryo

**Figure 4**




Structure of the DPY module. **(a)** Stereo pair of a family of 22 backbone structures calculated from NMR-derived distance restraints at pH 3.3. The root mean square deviation (r.m.s.d.) of these structures is 0.48 Å for backbone atoms and 1.32 Å for all atoms. In a Ramachandran plot, 64.5% of the residues are in the most favoured regions, 33.4% are in additional allowed regions, 0.9% are in generously allowed regions, and 0.9% are in disallowed regions (S12 in structures 5 and 6, N13 in structure 21, and Q19 in structure 22; residues 12 and 13 are in a relatively disordered loop). **(b)** Schematic view of the average structure, depicting an amino-terminal $3_{10}$ helix in red/green and anti-parallel β sheet in blue. The two disulphide bridges are shown in yellow. N, amino terminus of the module.

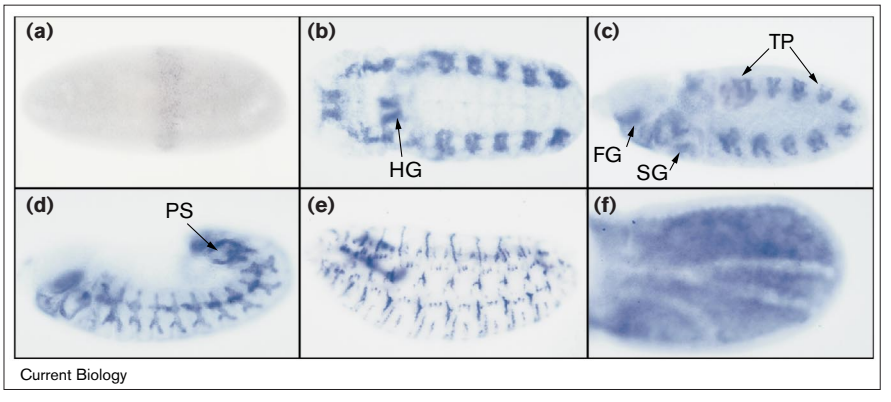
development (data not shown). Consistent with wing phenotypes of certain *dumpy* alleles, we also detected expression within the pupal wing. Figure 5f shows that, at 7 hours after puparium formation (APF), *dumpy* was expressed throughout the intervein regions, but was absent or expressed at a low level within the prevein territories.

## Discussion

Two pieces of evidence allow us to conclude that we have identified the molecular product of the *dumpy* locus. All mapped breakpoints in the 24F–25A region that display *dumpy* phenotypes interrupt a single large gene that extends throughout a 100 kb region. Secondly, the expression of this gene within the developing trachea, muscle attachment sites and pupal wings is consistent with described *dumpy* phenotypes. All probes derived from locations throughout the coding region of the gene displayed a similar expression pattern, and all identified exons were found to be capable of splicing to their immediate neighbours. Taken together, these data support the conclusion that *dumpy* encodes a single gene product consisting of a predicted 70 kb mRNA, although we cannot rule out the possibility that several, similarly expressed, products may be formed by alternative splicing. The large target size could explain the high rates of detection of *dumpy* alleles that are induced by ethylmethane sulfonate mutagenesis screens [10]. The genetic complexity of the locus may result from different functions of Dumpy being susceptible to mutations at different sites in the protein. This is supported by a varying modular composition of Dumpy along its length. For example, the $Ca^{2+}$-binding class of EGF modules are restricted to the amino-terminal region of the protein.

Based on the modular composition and structure of the Dumpy protein, together with its mutant phenotypes,

**Figure 5**




Expression of *dumpy* during development. Digoxigenin-labelled RNA probes from locations within the *dumpy* coding region were used to label embryos in whole mount. All probes derived from different locations spread throughout the locus showed similar staining, consistent with the locus encoding one large transcript. **(a)** Expression of *dumpy* was first detected at the onset of gastrulation as a broad band of staining around the middle of the embryo. **(b,c)** By stage 11, strong staining was seen in the tracheal placodes (TP), the salivary gland placode (SG), the developing hindgut (HG) and foregut (FG). Expression of *dumpy* in the tracheae persisted as the tracheae branched out in **(d)** stage 12, and was also present in the posterior spiracles (PS). **(e)** By stage 17, *dumpy* was strongly expressed at somatic muscle attachment sites. **(f)** At 7 h APF, *dumpy* was expressed in intervein regions of the pupal wing.

we propose that Dumpy is a component of the epidermal–cuticle junction with its carboxy-terminal anchor inserted into the epidermal cell membrane. From the structures and dimensions of the DPY and EGF modules, we would expect these modules to connect end-to-end to form a long fibrous molecule of at least 0.8 μm. In addition, regions of low complexity sequence may allow considerable further extension of the molecule. It is likely that a gigantic protein such as Dumpy would only evolve if its physical size were an important feature of its function. This is certainly the case for Titin, which is the only known molecule of comparable size to Dumpy and which organises muscle fibre architecture [25]. As the dimensions of the cuticle are also on a micrometer scale [3], we suggest that Dumpy may thus act as a ruler, with the different regions mediating distinct molecular interactions to modulate cuticle ECM organisation. Such a role in organising cuticle architecture is supported by $dp^{l}$ mutants, which cause the loss of ridges (taenidial folds) in the tracheal cuticle that normally support the tracheal structure.

The carboxy-terminal ZP domain of Dumpy may crosslink homotypically or heterotypically to other ECM components. Together with cumulative interactions along the long length of Dumpy, this would make possible a strong adhesive connection to the cuticle. This may be important at positions of particular strain such as at muscle-attachment sites where we see strong expression in epidermal tendon cells through which the muscles attach to the cuticle. At these sites, mechanical force is transmitted from the basal muscle–tendon cell junction to the apical surface by microtubule bundles [26]. Within the Dumpy fibre, the EGF–EGF linkers and repeat regions rich in proline, serine and threonine residues, may confer points of elasticity that could allow it to act as a molecular shock absorber, distributing tension within the cuticle to preserve cuticle integrity under stress. Although no phenotypes at muscle-attachment sites have been reported during embryonic development, this could be explained by the fact that many embryonic lethal *dumpy* mutants are reported to lack muscle activity [7], meaning that these junctions are never put under strain. A precedent for this is the muscle-detachment phenotype of integrin mutants, which is not observed until muscle contractions begin, late in embryonic development [27]. The weaker $dp^{v}$ mutants, however, are able to bypass the lethal stages and manifest phenotypes at the muscle-attachment sites of the flight muscles on the notum. At around 32 hours APF, the notum epidermis at these locations is pulled inwards, forming a pit-shaped depression or vortex. A shortening of the flight muscles occurs shortly before the time at which epidermal invagination begins in $dp^{v}$ mutants [28] and the mutant epidermis may thus be unable to resist the resulting tension. As the basal connections between the tendon cells and the muscles are not disrupted in $dp^{v}$ mutants [6], it is likely that, as in the trachea, Dumpy function is required at the apical cell surface, that is, at the epidermal–cuticle connection. Thus, the phenotype may result from reduced epidermal cell adhesion to the cuticle and/or from weakening of the cuticle structure at these sites.

A role for Dumpy as a cell adhesion protein is further supported by the wing blistering phenotypes observed in clones of $dp^{l}$ alleles [11]. Such blisters reflect detachment of the dorsal and ventral wing surfaces and are consistent with the observed expression of *dumpy* within the intervein regions of the wing, which appose, and its absence from the prevein territories, which do not appose. There are similarities between the wing epithelial cells and the epidermal tendon cells as both contain arrays of parallel microtubule bundles which connect hemidesmosome-like junctions at the apical and basal surfaces [26]. The basal junctions of both cell types requires integrin-mediated adhesion to basement membranes while the apical junctions connect through uncharacterised molecules to the cuticle. The apposition of the cuticle wing surfaces requires the integrity of both junctions and, thus, in the wing epithelium, Dumpy may have an analogous function to that at the muscle-attachment sites. A role at the cuticle–cell junction is also consistent with the $dp^{o}$ wing phenotype as chitinase or trypsin treatment of cultured pupal wings, which disrupts cuticle–apical cell junctions, results in a proximal–distal contraction of the epithelium, similar to the contraction caused by $dp^{o}$ alleles [29]. Thus, Dumpy may be required to maintain tension in the developing wing epithelia. Points of elasticity within Dumpy could modulate such mechanical tension and thus contribute to defining the shape of the wing and other tissues. There may thus be a parallel with the muscle fibre protein Titin, which controls passive tension in muscles through points of elasticity conferred by regions of repetitive, low complexity sequence [30].

Examination of the remaining sites of *dumpy* expression indicates that the gene may have roles not previously described. For example, *dumpy* was expressed during the formation of structures such as the salivary gland, the foregut and hindgut which, like the trachea, invaginate from the ectoderm to form tubes that secrete a cuticle. Dumpy may thus contribute similarly to their mechanical strength and morphogenesis. During early gastrulation, expression of *dumpy* occurred in a band across the middle of the embryo (Figure 5a), before the onset of germ band extension. By analogy with the later functions of *dumpy*, this may reflect a need of certain regions to withstand the tension generated by the movements of epithelia that occur during and following this period of expression. A site of strain around the middle abdominal segments is demonstrated by mutant alleles of the cytoplasmic cell junction protein *kakapo* [26]; in these mutants, localised breaks in the ventral epidermis are observed following

germ-band retraction. Kakapo is localised at both apical and basal ends of the microtubule bundles that insert into hemidesmosome-like junctions in epidermal muscle-attachment cells. Interestingly, like *dumpy*, *kakapo* alleles display phenotypes at muscle-attachment sites and cause wing blisters [26]. Thus, both proteins may function as components of the bridge of molecules that connects the apical and basal cell surfaces and provides rigidity to the cell and allows transmission of tension to the supporting external cuticle.

It is interesting to speculate how the molecular nature of Dumpy can be reconciled with other *dumpy* phenotypes. Given its possible role in cuticle–cell attachment, it seems paradoxical that, during moulting, a specific class of ecdysis lethal *dumpy* mutants fails to detach the cuticle properly from the epidermal surfaces. This could be explained if this class of alleles were defective at sites that must normally be severed or otherwise disrupted during this process. Other larval lethal phenotypes include an overgrowth of the trachea that is difficult to explain by a simple adhesive or structural role for Dumpy. One possibility is that, in maintaining mechanical tension, Dumpy might provide a regulatory signal that controls growth in certain tissues. The observations that *dumpy* mutations stimulate chitin biosynthesis [5] imply that cuticle composition may also be regulated by a feedback control that monitors cuticle properties. Further investigation of this system may therefore contribute to a wider understanding of how the mechanical properties of extracellular matrices regulate the behaviour of attached epithelia.

Through a homology search, we have also identified a *dumpy*-like gene in the *C. elegans* genome (accession number Q23587, SPTREMBL). This comprises a ZP domain preceded by 30 EGF modules, including $Ca^{2+}$–EGF modules, which are separated by stretches of threonine and serine-rich sequence (data not shown). If the *C. elegans* gene shares a common ancestor with *dumpy*, then the insect gene has expanded dramatically in length and has acquired additional modules and sequence repeats during its evolution. The expansion of the P-F repeat region may be a relatively recent event, as these repeats are remarkable for their high degree of sequence conservation at both protein and DNA levels. Given the presence of similar genes in *Drosophila* and *C. elegans*, it is likely that Dumpy belongs to a phylogenetically widespread family of adhesion molecules. It is also probable, however, that this family is structurally diverse, reflecting a range of cuticle properties, dimensions and morphologies that have emerged during evolution.

## Materials and methods

### *Chromosomal walk*

The walk was initiated with a subcloned fragment from the *decapentaplegic* locus, generously provided by William Gelbart, which hybridises *in situ* to both the proximal and distal breakpoints of *In(2L)dpp6*. A phage containing the junction fragment was isolated, according to standard procedures [31], from a genomic library prepared from the inversion strain. DNA from the proximal or 24F1-2 end of the junction fragment, as determined by polytene chromosome *in situ* hybridisation, was then used to start the walk in the proximal direction using Canton-S phage libraries in λ2001 or λDASH [13]. The insert of each isolated walk phage was then subcloned into Bluescript KS+ (Stratagene). The most proximal portion of the walk was derived from the P1 clone DS02961.

### *Mapping breakpoints*

The mutant alleles whose breakpoints were mapped are as described in [32] except for *T(2:3)TFB-6*, *In(2L)TFB-31* and *T(2:3)JAZZ-12*, which were generated in the MacIntyre laboratory using gamma rays by R.M., T. Burns and J. Henderson, respectively. To localise these breakpoints, their approximate positions were first determined by hybridising phage inserts or widely spaced subcloned fragments to the polytene chromosomes of larvae carrying a chromosome aberration according to [33]. Southern analysis [31] was used to determine which subcloned fragment from the region of the walk spans the breakpoint.

### *Sequence data*

We initially identified two cDNA fragments that mapped onto the walk, both with their 5′ ends proximal. The 13F cDNA was isolated by screening an embryonic cDNA library [34] (a gift from Nick Brown) using a radiolabelled subclone 13F. This cDNA contained a 3′ end and was found to map at the distal end of the walk to subclones 13F–13E. A second cDNA, EGF-JR (a gift from Jonathon Rothberg), mapped to the proximal end of the walk in subclone 56D. To date, the *Drosophila* genome project has sequenced most of the genomic region between subclone 26A and the distal end of our chromosomal walk from the P1 DS04146. In the P-F region, however, the exact number of the P-F repeats has not been determined, as their near identity makes sequencing through this section difficult. Subclones derived from the region between 26A and 9A all contained P-F repeats, and the distance encompassing the P-F region is consistent with the presence of up to 38 such repeats. We have obtained further sequence data 5′ to 26A by sequencing both cDNA fragments derived from this region and genomic subclones. We isolated a 3.4 kb cDNA, which mapped to subclones 56A to 26A, by screening an embryonic cDNA library [34] using a radiolabelled mix of subclones 5A and the proximal third of 13F. We obtained the 5′ end of the *dumpy* transcript using a PCR strategy on a random-primed larval cDNA library in λgt11 (Clontech). This involved amplifying the DNA lying between two primers, one specific to EGF-JR, the other to the phage vector, and cloning the products into pGEMT-easy (Promega). The resultant clone spanned subclones 53E–56D and contained the amino terminus of the open reading frame. To determine whether adjacent exons were able to splice directly together, primers were designed to predicted exons and the regions between them amplified off cDNA by PCR. This was done either by PCR on the Clontech random-primed larval library, or by RT–PCR (BRL) on third instar larval total RNA (made using RNEASY, Qiagen). The products were purified by gel extraction (Qiagen) and directly sequenced. The domain composition of the small *Drosophila* homologue of *dumpy* at 19A, which we termed *mini-dumpy* was derived from sequencing EST GH06658.

### *Determination of DPY structure by NMR*

A peptide of amino-acid sequence RPECVLNSDCPSNQACVNQKCRDP was synthesised by Fmoc chemistry, and purified by reverse-phase HPLC. The S–S bonds were then formed under treatment of a 0.5 mM aqueous solution of the peptide with 50 mM β-mercaptoethanol at pH 8.5 [35]; the mixture was left for 2 days at room temperature with a perforated cover and subsequently repurified by reverse-phase HPLC. Analysis of the product by electrospray mass spectroscopy and NMR showed it to be a folded homogeneous species with only one set of disulphide bonds. Homonuclear 1D and 2D $^1$H NMR spectra were recorded at 15°C and 30°C on 500 MHz

and 600 MHz spectrometers. Assignments of the proton resonances were obtained from 2D DQF-COSY, HOHAHA, Single Relay COSY, PreTOCSY-COSY and NOESY spectra at pH values between 3.2 and 7.5. For examples of pulse sequences used, see [36]. NOESY experiments were typically collected with a mixing time of 250 msec. Slow-exchanging amide hydrogen atoms were identified by dissolving protonated peptide in deuterium oxide and recording 1D spectra over several hours. Very similar structures were calculated from interproton distance restraints derived from the 2D NOESY experiments at pH 3.3 and 5.7, using the software XPLOR 3.851. Calculations, using standard protocols, were started from random coordinates and incorporated floating chirality assignments on prochiral groups. Estimated peak volumes from the NOESY spectrum were converted into three classes of interproton distance restraints with upper values of 2.7, 3.5 and 5.0 Å. Explicit hydrogen bond restraints were added during the refinement stage of the calculations, using the statistical distribution of potential hydrogen acceptors from within a family of 50 structures for each slow-exchanging proton. As the pattern for the two disulphide bonds in the DPY module could not be determined experimentally, starting structures were calculated using ambiguous restraints between all possible pairs of cysteine sulphur atoms. The structures selected for further refinement were chosen according to their relative potential energy value: 40% of the structures had potential energies between 5 and 10 kcal/mol and displayed the disulphide pattern 4–16, 10–21. This bridge pattern was therefore assumed for the calculations of the final family of 22 DPY module structures. Comparisons with previously determined structures were performed by manual browsing of the SCOP database [37] of protein folds; the lower size limits of current structural homology programs precluded an automatic search. Potential homologues were fitted to the DPY module structure using the molecular display software MOLMOL [38]. After successful homologues were found to correspond to the cellobiohydrolase family of structures, a PROSITE [39] search established that the WAP domain signature matched that family. Further WAP domains were downloaded from the PFAM database [40] and structurally aligned with DPY.

### In situ *hybridisation*

Digoxigenin-labelled RNA probes were synthesised according to the manufacturer's instructions (Boehringer Mannheim) and derived from both genomic subclones as well as from cloned cDNA fragments. For embryos, the wholemount *in situ* hybridisation procedure was carried out, essentially as described in [41], using a hybridisation temperature of 55°C and detection with an alkaline phosphatase labelled anti-digoxigenin antibody (Boehringer Mannheim). The staging of embryos was performed according to [42]. For pupal wings, *in situ* hybridisation was carried out according to [43].

### *Supplementary material*

Supplementary material including the predicted cDNA and protein sequence of Dumpy and a comparison of the 3D structure of the DPY module with the WAP domain of the protease inhibitor elafin is available at http://current-biology.com/supmat/supmatin.htm.

## Acknowledgements

We thank Miriam Post, James Park, Jessica Henderson, Katie O'Donnell, Timothy Burns and Robyn Tuttle for assistance with induction and localisation of *dumpy* mutant breakpoints, and the former section of Genetics and Development and Hughes Scholar Program at Cornell University for their financial support. We thank Jonathon Rothberg for the EGF-JR cDNA fragment. I.D.C. is a member of the Oxford Centre for Molecular sciences (OCMS) that is supported by EPSRC, BBSRC and MRC; he also acknowledges support from the Wellcome Trust. We thank Laurence Hall for automated DNA sequencing, Maureen Pitkeathly for peptide synthesis, Chris Gilpin for help with electron microscopy, Robin Aplin (OCMS) for mass spectrometry, and Jonathan Boyd and Nick Soffe for NMR assistance. I.P. was supported by OCMS. We acknowledge financial support from the EU framework IV grant BIO4-CT96-0662. This work has been greatly facilitated by access to the data of the Berkeley and Celera *Drosophila* Genome Projects.

## References

1. Fristrom D, Fristrom JW: **The metamorphic development of the adult epidermis.** In *The Development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. New York: Cold Spring Harbor Press; 1993:843-897.
2. Manning G, Krasnow MA: **Development of the *Drosophila* tracheal system.** In *The Development of* Drosophila melanogaster. Edited by Bate M, Martinez-Arias A. New York: Cold Spring Harbor Press; 1993:609-686.
3. Poodry CA: **Epidermis: morphology and development.** In *The Genetics and Biology of* Drosophila, 2nd edn. Edited by Ashburner M, Wright TRF. London: Academic Press; 1980:443-497.
4. Kramer JM, French RP, Park EC, Johnson, JJ: **The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen.** *Mol Cell Biol 1990*, **10:**2081-2089.
5. Blass DH, Hunt DM: **Pyrimidine biosynthesis in the *dumpy* mutants *of Drosophila melanogaster.*** *Mol Gen Genet* 1980, **178:**437-442.
6. Metcalfe JA: **Developmental genetics of thoracic abnormalities of *dumpy* mutants of *Drosophila melanogaster.*** *Genetics* 1970, **65:**627-654.
7. Metcalfe JA: **Development and complementation of lethal mutations at the *dumpy* locus of *Drosophila melanogaster.*** *Genet Res Camb* 1971, **17:**173-183.
8. Waddington CH: **The genetic control of wing development in *Drosophila.*** *J Genetics* 1940, **41:**75-140.
9. Carlson EA: **Allelism, complementation, and pseudoallelism at the *dumpy* locus in *Drosophila melanogaster.*** *Genetics* 1959, **44:**347-373.
10. Jenkins JB: **Mutagenesis at a complex locus in *Drosophila* with the monofunctional alkylating agent, ethyl methanesulphonate.** *Genetics* 1967, **57:**783-793.
11. Prout M, Damania Z, Soong J, Fristrom D, Fristrom JW: **Autosomal mutations affecting adhesion between wing surfaces in *Drosophila melanogaster.*** *Genetics* 1997, **146:**275-285.
12. Grace D: **Genetic analysis of the *dumpy* complex locus in *Drosophila melanogaster*: complementation, fine structure and function.** *Genetics* 1980, **94:**647-662.
13. Becker MN: *Molecular Studies of Region 24F-25A of Chromosome 2L and the Generation of* dumpy *Mutants in* Drosophila melanogaster. *Ph.D. thesis.* Cornell University; 1991.
14. Bork P, Sander C: **A large domain common to sperm receptors (ZP2 and ZP3) and TGF-β type III receptor.** *FEBS* 1992, **300:**237-240.
15. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains.** *Proc Natl Acad Sci USA* 1998, **95:**5857-5864.
16. Legan PK, Rau A, Keen JN, Richardson GP: **The mouse Tectorins - modular matrix proteins of the inner ear homologous to components of the sperm-egg adhesion system.** *J Biol Chem* 1997, **272:**8791-8801.
17. Sebastiano M, Lassandro F, Bazzicalupo P: ***Cut-1*, a *Caenorhabditis elegans* gene coding for a dauer-specific noncollagenous component of the cuticle.** *Dev Biol* 1991, **146:**519-530.
18. Campbell ID, Bork P: **Epidermal growth factor-like modules.** *Curr Opin Struct Biol* 1993, **3:**385-392.
19. Handford PA, Mayhew M, Baron M, Winship PR, Campbell ID, Brownlee GG: **Key residues involved in calcium-binding motifs in EGF-like domains.** *Nature* 1991, **351:**164-167.
20. Rao Z, Handford P, Mayhew M, Knott V, Brownlee GG, Stuart D**: The structure of a Ca2+-binding epidermal growth factor-like domain — its role in protein–protein interactions.** *Cell* 1995, **82:**131-141.
21. Handford P, Downing AK, Rao ZH, Hewett DR, Sykes BC, Kielty CM: **The calcium-binding properties and molecular-organization of epidermal growth factor-like domains in human Fibrillin-1.** *J Biol Chem* 1995, **270:**6751-6756.
22. Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID, Handford PA: **Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders**. *Cell* 1996, **85:**597-605.
23. Rebay I, Fleming RJ, Fehon RG, Cherbas L, Cherbas P, Artavanis-Tsakonas S: **Specific EGF repeats of Notch mediate interactions with Delta and Serrate — implications for Notch as a multifunctional receptor.** *Cell* 1991, **67:**687-699.

24. Schalkwijk J, Wiedow O, Hirose S: **The Trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core.** *Biochem J* 1999, **340:**569-577.
25. Labeit S, Kolmerer B: **Titins – giant proteins in charge of muscle ultrastructure and elasticity.** *Science* 1995, **270:**293-296.
26. Gregory SL, Brown NH: ***kakapo*, a gene required for adhesion between and within cell layers in *Drosophila*, encodes a large cytoskeletal linker protein related to plectin and dystrophin.** *J Cell Biol* 1998, **143:**1271-1282.
27. Newman SM, Wright TRF: **A histological and ultrastructural analysis of developmental defects produced by the mutation, *lethal(1)myospheroid*, in *Drosophila melanogaster*.** *Dev Biol* 1981, **86:**393-402.
28. Sanstrom DJ, Restifo LL: **Epidermal tendon cells require *Broad Complex* function for correct attachment of the indirect flight muscles in *Drosophila melanogaster*.** *J Cell Sci* 1999, **112:**4051-4065.
29. Turner CM, Adler PN: **Morphogenesis of *Drosophila* pupal wings *in vitro*.** *Mech Dev* 1995, **52:**247-255.
30. Linke WA, Ivemeyer M, Olivieri N, Kolmerer B, Rüegg JC, Labeit S: **Towards a molecular understanding of the elasticity of Titin.** *J Mol Biol* 1996, **261:**62-71.
31. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning – a laboratory manual, 2nd edn*. New York: Cold Spring Harbor Press; 1989.
32. Lindsley DL, Zimm GG: *The Genome of* Drosophila melanogaster. San Diego, New York: Academic Press; 1992.
33. Lim JK: **Protocol for *in situ* hybridization with biotinylated DNA**. *Dros Inf Serv* 1993, **72:**73-76.
34. Brown NH, Kafatos FC: **Functional cDNA libraries from *Drosophila* embryos**. *J Mol Biol* 1988, **203:**425-437.
35. Kahmann JD, Koruth R, Day AJ**: Method for quantitative refolding of the Link module from human TSG-6.** *Prot Expr Purif* 1997, **9:**315-318.
36. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ: *Protein NMR Spectroscopy – Principles and Practice*. San Diego, London: Academic Press; 1995.
37. Murzin AG, Brenner SE, Hubbard T, Chothia C**: SCOP – a structural classification of proteins database for the investigation of sequences and structures.** *J Mol Biol* 1995, **247:**536-540.
38. Koradi R, Billeter M, Wuthrich K: **MOLMOL: a program for display and analysis of macromolecular structures.** *J Mol Graphics* 1996, **14:**51-55.
39. Hofmann K, Bucher P, Falquet L, Bairoch A: **The PROSITE database, its status in 1999.** *Nucleic Acids Res* 1999, **27:**215-219
40. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer ELL: **Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins.** *Nucleic Acids Res* 1999, **27:**260-262.
41. Tautz D, Pfeifle C: **A non-radioactive *in situ* hybridisation method for the localisation of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*.** *Chromosoma* 1989, **98:**81-85.
42. Campos-Ortega JA, Hartenstein V: *The Embryonic Development of* Drosophila melanogaster. Berlin: Springer-Verlag; 1985.
43. Cornell M, Evans DAP, Mann R, Fostier M, Flasza M, Monthatong M, Artavanis-Tsakonas S, Baron M**: The *Drosophila melanogaster Suppressor of deltex* gene, a regulator of the notch receptor signaling pathway, is an E3 class ubiquitin ligase.** *Genetics* 1999 **152:**567-576.


**Because *Current Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cub – for further information, see the explanation on the contents page.**