A molecular analysis of the tissue polarity gene prickle and associated transcripts in Drosophila melanogaster.

Thesis

A molecular analysis of the tissue polarity gene *prickle* and associated transcripts in *Drosophila melanogaster*

Clare Patricia Green, BSc.

Thesis submitted in fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

Sponsoring Establishment: Addenbrooke's NHS Trust, Cambridge

Collaborating Establishment: Department of Genetics, University of Cambridge

Discipline: Molecular biology

June 1999

Date of Award: 31 August 1999

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

I declare that I have not submitted any part of this thesis for any other degree or qualification to this or any other university. The work described is the result of my own work, except where specifically stated in the text.

For Mum and Dad,
with all my love

ACKNOWLEDGEMENTS

"There is no such thing as a 'self-made' man. We are made up of thousands of others. Everyone who has ever done a kind deed for us, or spoken one word of encouragement to us, has entered into the make-up of our character and of our thoughts, as well as our success."
(George M. Adams)

My thanks first must go to David Gubb, my director of studies, not only for support and guidance in the course of this study, but during all my time in the group. Thank you also to Michael Ashburner and Robin Carrell in their roles as my supervisors, and additionally to Michael for my place in his laboratory for so long. I am also grateful to David and Michael for the opportunity to study for this degree.

Thank you also to all the others who have worked on the polarity project in our group: Simon Collier, Darin Coulson, David Huen, Glynnis Johnson and Dave Tree. I am grateful to them all for information, advice and their contributions to this work (as cited in the text). Thank you also to our collaborators on the serpins, especially Jean Marc Reichhart.

I would additionally like to thank all the past and present members of the Ashburner group who, together with those named above, have taught me everything I know about *Drosophila* and molecular biology. Special mentions go to Saverio Brogna, Jose de Celis, Catharine Goddard, Nick Harden, Gerti Heimbeck, Paul Lasko, Carol McKimmie, Sue McNabb, Terri Morley, John Roote, Ann Barton, Pam Fletcher, Steve Russell and Jane Snaith for contributions to this project (cited in the text), help in the laboratory and friendship.

Thanks go also to the rest of the Genetics department, in particular to John Green, Isobelle Hasleham, Sharon Lovell, Marion Martin, Hemi Mistry, Sheila Nunn, Babs Ostler, Sandra Rylance, Adrian Simpson, David Webb and the late Maureen Blake for their contributions to helping life to run smoothly and their support.

I would also like to thank those I worked with and learnt from in previous jobs, particularly Leszek Borysiewicz, Frank Cook and Patrick Sissons. Without their belief in me, I would never have come this far.

Warm thank yous to all my friends and family who have helped me so much in so many ways over the years. Special mentions go to Ruth Butchart, Shek Graham, Alison and Steven Green, Connie Humphries, Flavia Malim and Stacey Medalyer. Thank you also to all those associated with "Insight" who have helped my personal development and supported me through some difficult times.

Last, but by no means least, my thanks go to my parents, Pat and Alan Green, who have supported me, encouraged me and believed in me throughout my life so far. Without them, none of this would have been possible.

ABSTRACT

The distinct patterns of bristles and hairs on the surface of an adult *Drosophila* indicate a planar polarity and reflect the polarity of the underlying structures. Many of the planar polarity genes have been analysed molecularly and found to encode a diverse group of proteins. In this thesis, I describe the molecular characterisation of *prickle*.

Three classes of mutation are found at the *prickle* locus; *prickle* (pk^{pk}), *spiny-legs* (pk^{sple}) and *prickle-spiny-legs* ($pk^{pk\text{-}sple}$). Three transcripts were identified and sequenced. The 4.2 kb *pk* transcript has seven exons that span 70 kb. The other two transcripts, *sple* and *pkM* , share six exons with *pk* and have alternatively spliced 5' ends.

The GAL4-UAS system was used to demonstrate rescue of the prickle and spiny-legs phenotypes with the *pk* and *sple* transcripts, respectively. Overexpression of the *UAS[pk+]* construct also resulted in a phenotype resembling loss-of-function pk^{sple} and vice versa.

The transcripts encode novel proteins with three LIM or LIM-like motifs in the shared exons. LIM domains are cysteine rich sequences known to be involved in protein-protein interactions. A novel domain, named the PET domain, was also identified in these proteins just upstream of the LIM motifs. Database searches found the PET domain conserved in a small group of proteins. A *Drosophila* homologue, *espinas* (*esn*), was also identified and sequenced.

A cluster of transcripts was found in the region of the 5' exon of the *pk* transcript. These were sequenced and shown to have homology to serine proteinase inhibitors or serpins. The most distal of the three was shown to rescue the necrotic phenotype.

The possible roles of *prickle* in planar polarity signalling are discussed and it is suggested that the *prickle* proteins may be involved in the formation of multiprotein complexes.

CONTENTS

APPENDICES

LIST OF FIGURES AND TABLES

CHAPTER 4

CHAPTER 5

CHAPTER 1. INTRODUCTION

One of the most challenging questions in modern biology is how the regulated patterns that form in development arise. In this thesis, I look at one aspect of pattern formation, that of tissue or planar polarity in *Drosophila melanogaster*, and attempt to characterise one of the genes involved in its determination. This chapter introduces some of the early models proposed to explain polarity and pattern formation and then goes on to summarise what is known of these processes in *Drosophila*. Planar polarity in the fly and the genes involved are then described, with particular reference to the *prickle-spiny-legs* gene.

1.1 Studies of pattern and polarity

The term "polarity" can be applied in various ways. For example, a cell may have specific apical and basal poles or be subdivided into two or more zones. Polarity can also refer to the orientation of structures within the plane of an epithelium and this is termed planar or tissue polarity. Even very simple organisms show polarity and generation of polarity is often associated with the cytoskeleton. For example, during the division of *Escherichia coli* the cytoskeletal element FtsZ localises to the inner surface of the cytoplasmic membrane of the equator (reviewed by Shapiro, 1993). Kropf (1994) looked at cell polarity in the zygotes of algae and proposed that establishment of polarity is associated with the actin cytoskeleton, while polarity expression depends on both F-actin and microtubules.

The asymmetry of yeast growth can be seen both in bud formation and in the arrangement of its cytoskeleton and organelles. The development of this cell polarity appears to be controlled by a morphogenetic pathway (reviewed by Drubin, 1991). Tucker (1981) postulated that intercellular signalling allowed greater architectural complexity and versatility of the cytoskeleton than a system that relied entirely on chemotactic gradients or pre-existing orientation.

Studies on the cytoskeleton of ciliates showed that specification of structural domains was independent of planar polarity (Jerka-Dziadosz *et al*, 1995). It was pointed out that a fairly rigid cytoskeletal framework might nevertheless directly influence cell shape.

Planar polarity in vertebrates has been less well studied, but some instances have been described. For example, planar polarity in the inner ear of mice and humans is reflected in the orientation of the hair bundles (reviewed by Eaton, 1997).

Various models have been proposed to explain pattern formation in development and these fall essentially into two categories. The first is based on gradients of diffusible morphogens (reviewed by Slack, 1987; Lawrence and Struhl, 1996), whereas the second depends on interactions between neighbouring cells (reviewed by Martinez-Arias, 1989). Many approaches have been used to try to determine to what extent each of these are used in development. A large number of studies have been carried out on insects and some examples are given below.

Cuticular polarity was studied in a variety of insects by transplanting pieces of cuticle in various positions and orientations, then following the effects on local polarity after subsequent moults (reviewed by Lawrence, 1970). Systems studied included the orientation of cuticular ripples in *Rhodnius* (Locke, 1959) and that of hairs in *Oncopeltus* (Lawrence, 1966). It was found that when a piece of cuticle was rotated polarity was altered after the next moult, but normal polarity could be restored after subsequent moults. This phenomenon was explained by the presence of a gradient and, as transplantations between equivalent levels of adjacent segments had no effect, it was proposed that the gradient must be repeated segment to segment.

A computer generated model was compared to the results found in *Rhodnius* to try and determine how gradients could be maintained (Lawrence *et al*, 1972). A

simple model whereby cells at one end of the segment acted as the "source" of a morphogen and those at the other end were the "sink" did not fit the data. A form of the source-sink model which also assumed that each cell attempts to actively maintain its own internal concentration of the morphogen seemed a better fit. A grafted cell would act as a source or a sink depending on concentration of the morphogen in its new situation. An equilibrium concentration would be established somewhere between the old and new levels, and hence a new concentration "landscape" would be established in the graft and nearby host.

The theory that pattern is built up around a series of reference points, the foci, was tested in the wing of the butterfly *Precis coenia* (Nijhout, 1980). A mark on the pupal wing was thought to define the position of the focus for the eyespot on the adult wing. If this region was cauterised early in development, the eyespot failed to develop. Grafts including this region induced ectopic pigment in the host tissue. It was therefore assumed that the focus was the source of a morphogen that was able to induce the synthesis of specific pigments.

Studies on *Bicyclus* butterflies showed that whereas cautery early in development decreased the size of eyespots, later cautery increased eyespot size (French and Brakefield, 1992). Nonfocal cautery only caused ectopic pattern formation when carried out in the later stages of development. Two models were suggested to explain these observations. In the source/threshold model, the morphogen produced by the focus only generated pattern if its level exceeded a certain threshold. The threshold could be set by the uniform level of a second diffusible substance. The effects caused by cautery were then explained by a fall in the response threshold. Alternatively, the sink model suggested that the focus or wound acted as a local sink for the morphogen, with pattern formation occurring where the morphogen fell below a threshold level.

Experiments on *Oncopeltus* (Lawrence, 1974) and *Dysdercus* (Nübler-Jung, 1974) showed that cell movement was an important component of pattern regulation. However, it also appeared that changes of polarity could occur without cell movement, although movement that could not be seen was not ruled out (Lawrence, 1974).

Further experiments on *Oncopeltus* demonstrated that the segment border was only regenerated when cells from different regions were apposed (Wright and Lawrence, 1981a). The same authors (1981b) also found that an irregular interface between two different cell populations gradually straightened, suggesting the presence of differential cell affinities.

Nübler-Jung and Mardini (1990) showed that the denticles of grafted cells of *Dysdercus* did not behave in accordance with standard gradient models. They suggested that changes in orientation resulted from differences in cell adhesion between host and graft cells. Individual cells would attempt to reduce contact with cells of different adhesiveness, causing cells that do remain along the interface to stretch. In support of this, elongated cells were seen at host-graft interfaces. These authors stressed that the mechanisms described would only serve to restore or maintain normal cell polarity, supporting the proposal that pattern generation and pattern control are regulated by different mechanisms.

French *et al* (1976) suggested a model in which positional information was specified in terms of polar co-ordinates. One component of the information corresponded to a position on a circle circumference, the other to the position on the radius. If cells with different positional values were apposed, growth would occur until all the intermediate positional values were intercalated. This system was likely to be based on local cell-cell interactions rather than gradients. Evidence for the polar co-ordinate model came from studies involving *Drosophila* imaginal discs, as well as other insect and amphibian limbs. A revision of the model proposed that a new cell would not adopt a positional value that was identical to that of a pre-existing cell, but would take up a position distal to it (Bryant *et al*, 1981).

1.2 Patterning and polarity in *Drosophila* development

Drosophila has long been regarded as a valuable model organism in the study of development and the ever-increasing volume of data only adds to its value. A great deal is known about the developmental processes of this organism, although

many questions still remain to be answered. A brief description of pattern formation in the fly follows as a background for the subsequent work.

The stages of *Drosophila* development are well documented (see, for example, Wolpert *et al*, 1998). After fertilisation, the nuclei of the embryo undergo a series of mitotic divisions without cleavage of the cytoplasm, resulting in the formation of a syncytium. Subsequently the nuclei move to the periphery to form the syncytial blastoderm. Membrane then grows in from the surface to form cells with all future tissues derived from this cellular blastoderm. Gastrulation begins three hours after fertilisation, when the future mesoderm invaginates to form a furrow along the ventral midline. Germ band (ventral mesoderm) extension occurs during gastrulation, resulting in the posterior trunk regions being driven onto what was the dorsal side of the embryo. The first external signs of segmentation are seen at this stage and the segmented larva hatches 24 hours after fertilisation. On the ventral side of each segment there are denticle belts and the other cuticular structures characteristic of each segment. The adult structures of the fly are present in the larva as imaginal discs. These small sheets of prospective epidermal cells, derived from the cellular blastoderm, grow throughout larval life.

The axes of the embryo are determined by four groups of co-ordinate genes; the anterior, posterior and terminal groups together define the anterior-posterior axis, while the dorsal-ventral axis is determined by the fourth group of genes (for reviews see Nüsslein-Volhard, 1991; St. Johnston and Nüsslein-Volhard, 1992).

Experiments involving removal or transplantation of cytoplasm from one pole of the egg demonstrated that factors localised at the poles are required at a distance for the development of particular embryonic regions (Nüsslein-Volhard *et al*, 1987). For example, mRNA of the maternal gene *bicoid* (*bcd*) is localised to the anterior pole but its protein is distributed in a gradient along the anteroposterior (A/P) axis (Driever and Nüsslein-Volhard, 1988a). The different levels of this protein along the A/P axis determine position in the anterior of the embryo (Driever and Nüsslein-Volhard, 1988b).

A systematic screen for zygotic mutations affecting segment number and polarity was carried out by Nüsslein-Volhard and Weischaus (1980). The mutants

they discovered, along with others identified since, fall into three groups; the gap genes, the pair-rule genes and the segment polarity genes. A single group of up to eight adjacent segments is deleted in the gap gene mutants, which include *Krüppel* (*Kr*), *knirps* (*kni*) and *hunchback* (*hb*) (Nüsslein-Volhard and Weischaus, 1980). The gap genes regulate each other to some degree. For example, it has been shown that *Kr* activity enhances *kni* expression, whereas *tailless* represses it (Pankratz *et al*, 1989). Two studies found that *hb* and *Kr* gene products both show sequence-specific DNA-binding activities (Stanojevic *et al*, 1989; Treisman and Desplan, 1989). The latter study found both *Kr* and *hb* protein binding sites upstream of the two *hb* promoters, suggesting that *hb* could be repressed by *Kr* and may also be autoregulated. Three BCD binding sites in the promoter region of *hb* have been shown to be involved in the activation of *hb*, indicating that BCD is a regulator of *hb* transcription (Driever and Nüsslein-Volhard, 1989). These authors suggested that BCD could also regulate other target genes if the affinities of their BCD-binding sites vary; for example, a reduced affinity for BCD binding would limit target gene expression to the upper end of the BCD gradient.

Mutations in the second class of segmentation loci, the pair-rule genes, result in deletions in alternative segments (Nüsslein-Volhard and Weischaus, 1980). These genes have a distinctive seven striped expression pattern in the wild-type embryo and include *even-skipped* (*eve*), *fushi tarazu* (*ftz*), *hairy* (*h*) and *runt*. They are regulated by the gap-genes and by interactions amongst each other. For example, the periodicity of *eve* involves the interaction of *hb* and *Kr* proteins with *eve* promoter elements (Stanojevic *et al*, 1989). By studying the expression of *ftz* and *eve* in both *h* and *runt* mutant embryos, it was shown that *h* acts as a negative regulator of *ftz* and a positive regulator of *eve*, whereas *runt* acts as positive regulator of *ftz* and a negative regulator of *eve* (Ingham and Gergen, 1988). Such results indicated that a regulatory hierarchy existed between the pair-rule genes. Pankratz and Jäckle (1990) suggested that the information provided by the gap genes makes a crude pattern which is later refined by interactions among the
pair-rule genes.

Mutations of the segment polarity genes give rise to a repeated pattern defect in which part of each segment is missing and is replaced by a mirror image duplication of the remaining part (Nüsslein-Volhard and Weischaus, 1980). These genes have a wide variety of functions and show complex interactions among themselves and with genes of other classes (for reviews, see Perrimon and Mahowald, 1987; Klingensmith and Perrimon, 1991; Piefer and Bejsovec, 1992). Two genes of this class, *engrailed* (*en*) and *wingless* (*wg*), have regulatory effects on each other as well as on several other genes. Martinez Arias *et al* (1988) monitored the expression of *wg* and *en* in embryos lacking *en* or *wg* respectively. They found that after germ band extension, the absence of either gene results in the loss of expression of the other, despite the fact that they are transcribed in different cells. This suggested that during normal development, signals are exchanged between *wg* and *en* expressing cells. Hildago and Ingham (1990) identified three phases during segment polarity gene expression; an initiation phase, mediated by the pair-rule genes, followed by consolidation and elaboration phases, regulated by interactions between the segment polarity genes themselves. These latter interactions involve complex signalling pathways in which *wg* is known to play a key role (for review, see Klingensmith and Nusse, 1994). Some of the planar polarity genes have also been implicated in these pathways (see section 1.7).

Wilkins and Gubb (1991) proposed that there may be a common set of rules, related to the polar co-ordinate model, that underlie pattern formation both in imaginal discs and in embryonic development. Their suggestion that this could involve the segment polarity genes was supported by work on the leg discs of *Drosophila* (Couso *et al*, 1993; Whittle and Philips, 1993).

Using mitotic clones in the wing, it was demonstrated that cell lineages are fixed as cells become restricted to particular regions of the wing at certain stages of development (García-Bellido *et al*, 1973). These regions have been called compartments and are derived from a small group of cells present in the early embryo. The group of cells within a compartment has been termed a polyclone (Crick and Lawrence, 1975). Initially it was thought that insect segments represented compartments (reviewed by Lawrence, 1981). Later evidence, including the

expression patterns of some of the segmentation genes, indicated that parasegments were more important developmentally (reviewed by Martinez-Arias and Lawrence, 1985). Parasegments are composed of the posterior region of one segment and the anterior region of the next.

It was proposed that the development of compartments is under control of the homeotic genes (García-Bellido, 1975; Morata and Lawrence, 1977a). Mutations in homeotic genes result in homeosis, that is, the transformation of a whole structure or segment into another related one. Historically, the homeotic bithorax complex were among the first genes to be implicated in the segmentation of *Drosophila* (reviewed by Lewis, 1978). The homeotic genes are now known to be involved in specification of parasegment identity (reviewed by Akam, 1987; Ingham, 1988). They interact with each other and with the other segmentation genes to generate the unique parasegmental states.

1.3 Planar polarity in *Drosophila*

The surface of an adult *Drosophila* is covered with an array of mechanosensory bristles and hairs (trichomes) that are arranged in distinct patterns. These patterns indicate a planar polarity and reflect the polarity of the underlying structures (Gubb and García-Bellido, 1982; Adler, 1992; Gubb, 1993). On the appendages of the fly, the bristles and hairs all point distally, whilst on the thorax and abdomen they usually point posteriorly. The patterns formed by the major bristles and hairs are constant for all wild type individuals. The patterns on the dorsal and ventral surfaces of the wing are similar but not identical (Gubb and García-Bellido, 1982).

The eye also shows polarity and is a useful system to study as eye development has been well documented and events can be followed at the single cell level (reviewed by Freeman, 1997). Ommatidia, the subunits of the compound eye, are packed together in a regular hexagonal arrangement. Each ommatidium consists of eight photoreceptors and twelve accessory cells that can be identified by position and type. The rhabdomeres of the photoreceptors form an asymmetrical arrangement

in cross section and this arrangement occurs in two chiral forms. All the ommatidia in one hemisphere of the eye are of the same form and are the mirror image reflection of those in the other hemisphere (Tomlinson, 1988; Wehrli and Tomlinson, 1998). Two waves of cell division provide a supply of cells in the developing eye disc (Wolff and Ready, 1991). The first wave occurs as a broad band ahead of the morphogenetic furrow and generates a pool of cells which gives rise to five of the photoreceptors. The second wave follows the furrow and produces all the remaining cells.

The first planar polarity genes to be described in *Drosophila melanogaster* were *prickle* (*pk*), *spiny-legs* (*sple*), *frizzled* (*fz*), *inturned* (*in*) and *multiple wing hairs* (*mwh*) (Gubb and García-Bellido, 1982). Other genes with polarity phenotypes include *dishevelled* (*dsh*) (Theisen *et al*, 1994); *fuzzy* (*fy*) (Collier and Gubb, 1997); *Van Gogh* (*Vang*), allelic to *strabismus* (*stbm*) (Taylor *et al*, 1998; Wolff and Rubin, 1998); *fritz* (*frtz*) (FlyBase) and *starry night* (*stan*) (FlyBase). Mutations in these genes alter polarity in different regions of the body. The mutants *pk* and *sple* were originally classified as separate genes; subsequent evidence suggested that they form a single complex locus (Heitzler *et al*, 1993) and during the course of this study it has become clear that this is the case (see Chapter 2).

The morphology of the bristles and hairs of *Drosophila* has been well studied. Bristles are four celled sensory organs that originate from a single precursor cell (Hartenstein and Posakony, 1989). They consist of a neuron, sheath cell, shaft and socket and occur at precise locations on the body. Many genes with a variety of mutant phenotypes effecting the bristles have been described. For example, mutations in *scute* cause the loss of some of the scutellar bristles, while in *split* mutants the bristles may be doubled, thickened or absent (Lees and Waddington, 1942). Gho and Schweisguth (1998) found that FZ signalling regulates the orientation of the the division of the single precursor cell, but the subsequent division of the two daughter cells is independent of FZ.

Hairs are extruded from the distal edge of epidermal cells and extend over the adjacent cells with each cell normally producing a single hair (Mitchell *et al*, 1983). Prehair initiation is associated with the formation of a large bundle of F-actin

filaments which occurs at the distal vertex of each hexagonally shaped cell in wild type animals (Wong and Adler, 1993). During the course of development, the prehairs migrate proximally across the cell and elevate above the epithelium (*ibid*). The regular arrangement of the bristles and hairs allows easy identification of mutations that alter them.

Mutations in the planar polarity genes have a variety of effects. They may cause a change in the orientation of the hairs and bristles, although not in the structure of the developmental unit itself (Gubb and García-Bellido, 1982; Adler *et al*, 1990). For example, most of the mutants have an effect on the polarity of the hairs on the wing but the shape of the wing blade is not altered. The mutants vary both in the regions of the wing affected and in the degree of change. The wing is the easiest structure to study because it is a large flat surface and each cell produces a single hair. The standard nomenclature of regions of the wing and wing veins facilitates the description of the mutants (Figure 1.1).

The pattern of the hairs and bristles in the planar polarity mutants is not random. Mutations in each gene give rise to a specific new pattern of hairs and/or bristles in the affected region. The patterns differ from gene to gene, which suggested that the wild type function of these genes could be to control specific aspects of planar polarity (Gubb and García-Bellido, 1982). The pattern is the same for all individuals carrying a particular mutation and the alleles of a given gene recovered in the initial study were indistinguishable from each other (Gubb and García-Bellido, 1982). Subsequent studies
of *fz* (Adler *et al*,1987) and *pk* (Coulson, 1994) have recovered alleles with alternative polarity patterns. The variation is usually only in severity of the phenotype within the

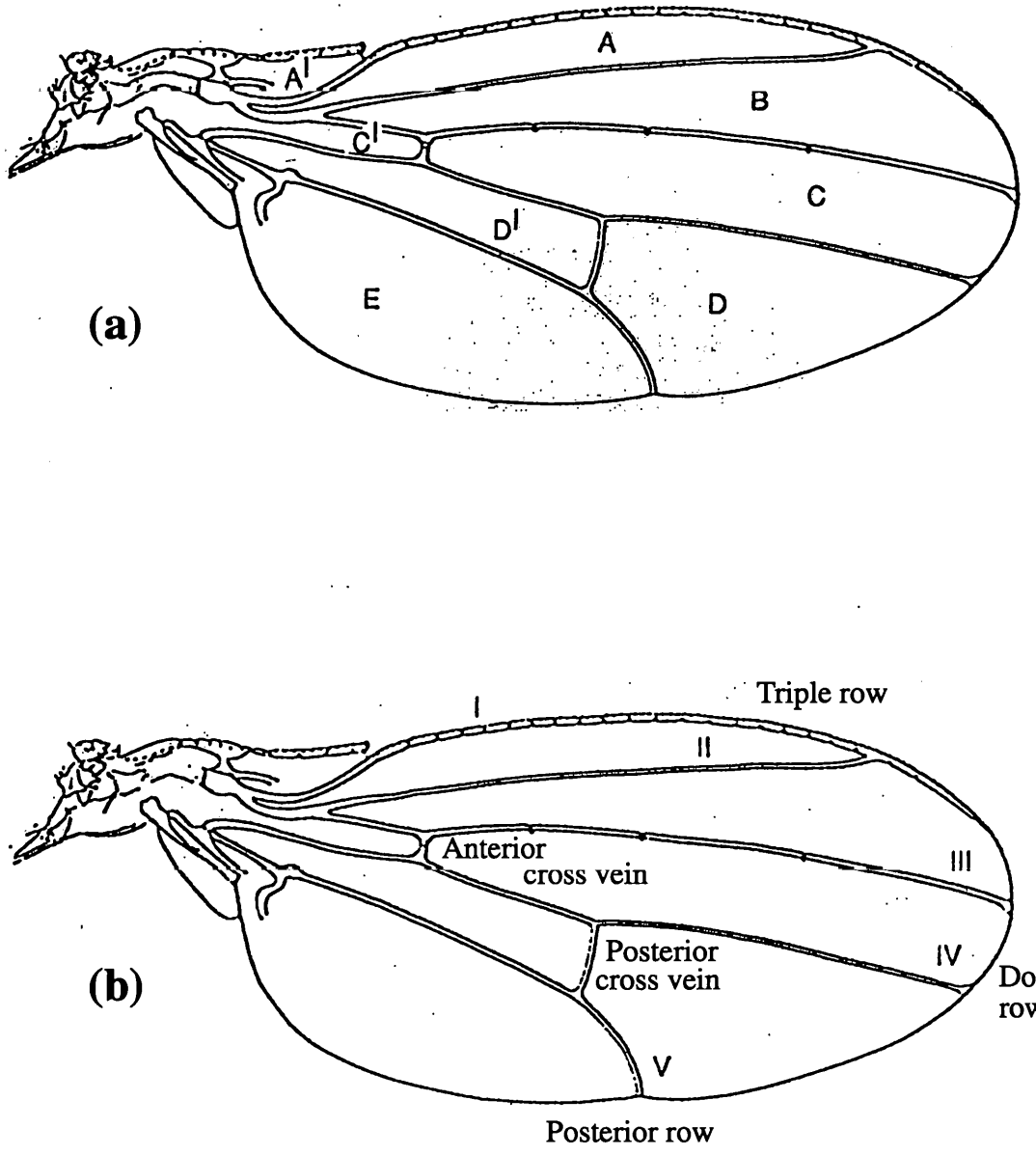


Fig. 1.1 (a) Regions of the wing as delineated by the wing veins (Ferris, 1950). (b) Regions of the wing and the wing veins.

region affected. In general, the striking reproducibility of the mutant polarity patterns from fly to fly is a constant feature of planar polarity mutants.

Mutations of some of the planar polarity genes also result in the production of more than one hair by the cells of the wing. This phenotype varies from gene to gene, both in the number of additional hairs produced and the number of cells producing them (Adler, 1992; Gubb 1993). For example, in *fz* mutants, just a few of the cells have two hairs, whereas most cells of an *in* mutant produce double hairs; in a *mwh* wing, virtually all the cells produce three or more hairs. The fact that planar polarity genes also alter hair number led to the suggestion of a correlation between polarity and hair initiation (Adler *et al*, 1987). This was supported by the observation that the planar polarity mutations affect the subcellular location of the F-actin bundles present at the site of hair initiation (Wong and Adler, 1993).

The polarity of the bristles may also be altered in other regions of the body. For example, pk^{pk}, *fz* and *in* change the pattern on the notum, while the abdominal tergites are affected by pk^{sple}, *fz* and *in* (Gubb and García-Bellido, 1982).

Mutations of some polarity genes bring about changes in leg morphology. They may alter the polarity of the hairs, bristles and/or the "bract-socket vector" (BSV). The BSV is defined as an imaginary vector pointing from the centre of the bract (a hair like structure found adjacent to most bristle sockets on the legs) to the centre of its associated bristle (Held *et al*, 1986). There may also be mirror image duplications of the tarsal joints (*ibid.*; Gubb, 1993; Coulson, 1994). In pk^{sple}, *fz, dsh* and *in,* these extra joints are found between normal joints and divide the tarsal segments into shorter units. In other mutants, the extra joints are located only at the sites of wild type joints, that is, these flies have some double joints (eg. *eyeless-Dominant* and *rotund*) or at both normal and ectopic sites (eg. *approximated* and *dachs*) (Held *et al*, 1986). The extra joints at wild type sites have wild type polarity, while those at ectopic sites typically have inverted polarity.

Some of the planar polarity mutants also have a characteristic "rough eye" phenotype. This is due to a disruption in the hexagonal packing arrangement of the corneal lenses. Within the eye, the ommatidia are mis-rotated and both chiral forms appear in both hemispheres (reviewed by Gubb, 1998). The reversals in polarity may

occur in the dorsal-ventral and/or the anterior-posterior axes (Gubb, 1993; Wehrli and Tomlinson, 1995; Zheng *et al*, 1995).

1.4 The *prickle-spiny-legs* locus.

The *pk-sple* locus gives three classes of mutation; pk^{pk}, pk^{sple} and $pk^{pk\text{-}sple}$. Recessive alleles of each of these classes are viable over a deletion of the region. Overlapping deletions are also viable and give the pk^{pk} or $pk^{pk\text{-}sple}$ phenotype depending on their degree of overlap. Some dominant alleles of *pk* have also been recovered in our laboratory and that of P. Adler, but these are not included in this study.

The two single mutant classes alter the polarity of bristles and hairs in reciprocal regions of the body; pk^{pk} affects the wing and notum, whilst pk^{sple} affects the the leg and abdomen (Gubb and García-Bellido, 1982). They complement each other completely; pk^{pk}/pk^{sple} flies are phenotypically wild-type and hence they were originally described as two separate genes (Lindsley and Zimm, 1992). Double mutant $pk^{pk\text{-}sple}$ flies display a weak polarity phenotype in all the regions effected by both the single mutants and in addition have a rough eye phenotype. The double mutant alleles do not complement either of the single mutations.

1.4.1 The pk^{pk} mutants

All pk^{pk} mutants so far studied affect bristle and hair polarity on the wing, halteres and notum. There is a variation in the degree of severity of the wing phenotype between alleles. Coulson (1994) described five arbitrary classes of wing phenotype covering pk^{pk} and $pk^{pk\text{-}sple}$ alleles, with Class 5 having the strongest phenotype and Class 1 the weakest. Alleles or combinations of alleles with a wild-type wing polarity phenotype were designated Class 0. The severity of the phenotypes of the mutants shows a continuous gradual increase but were classified in the above manner to facilitate their study.

The most extreme phenotype, Class 5, is seen in homozygotes of the original allele, pk^{pk1}. In these flies, hair polarity is affected in all regions of the wing (Figure

1.2) The characteristic whorl of hairs in regions A and B has the same sense of rotation on both surfaces of the wing (Gubb and García-Bellido, 1982). The orientation of the triple row of bristles along the anterior margin of the wing is reversed until the junction of veins I and II, where they show wild-type polarity. An example of a less extreme phenotype is shown by pk^{pk15} homozygotes, designated as a Class 3 phenotype. The polarity of regions B and C1 is usually wild-type, although occasionally B shows a mutant phenotype. The triple row also displays wild-type polarity. Those alleles designated as Class 1, for example $pk^{pk\text{-}sple12}$, retain wild-type phenotype in most regions of the wing. Only region D on the dorsal surface and C on the ventral surface show a disruption in polarity.

There is also a weak multiple wing hair phenotype in pk^{pk} mutants, with a small number of cells (reported as 2-3% by Gubb and García-Bellido,1982, but as 0.5-1% by Coulson, 1994) producing two or three hairs. This occurs in three small, distinct areas of the wing. In addition to the wing phenotype, pk^{pk} mutations also have an effect on the halteres and notum. The bristles of the notum form distinct whorls, one either side of the thorax.

1.4.2 The pk^{sple} mutants

The pk^{sple} single mutants have no phenotype in the wing or notum, but have a strong polarity phenotype in the legs and abdomen. The leg is affected in a number of ways, with the hairs, bristles and the bract-socket vectors all showing altered polarity (Held *et al,* 1986). Again Coulson (1994) named three classes according to severity of phenotype, with Class 3 being the strongest.

In addition to the hair and bristle phenotypes, mirror image duplications of the tarsal joints occur. All the alleles of pk^{sple} examined by Coulson gave ectopic mirror image joints in the first to fourth tarsal segments. However in many individuals the ectopic joints in the most proximal segment were only partially formed or completely

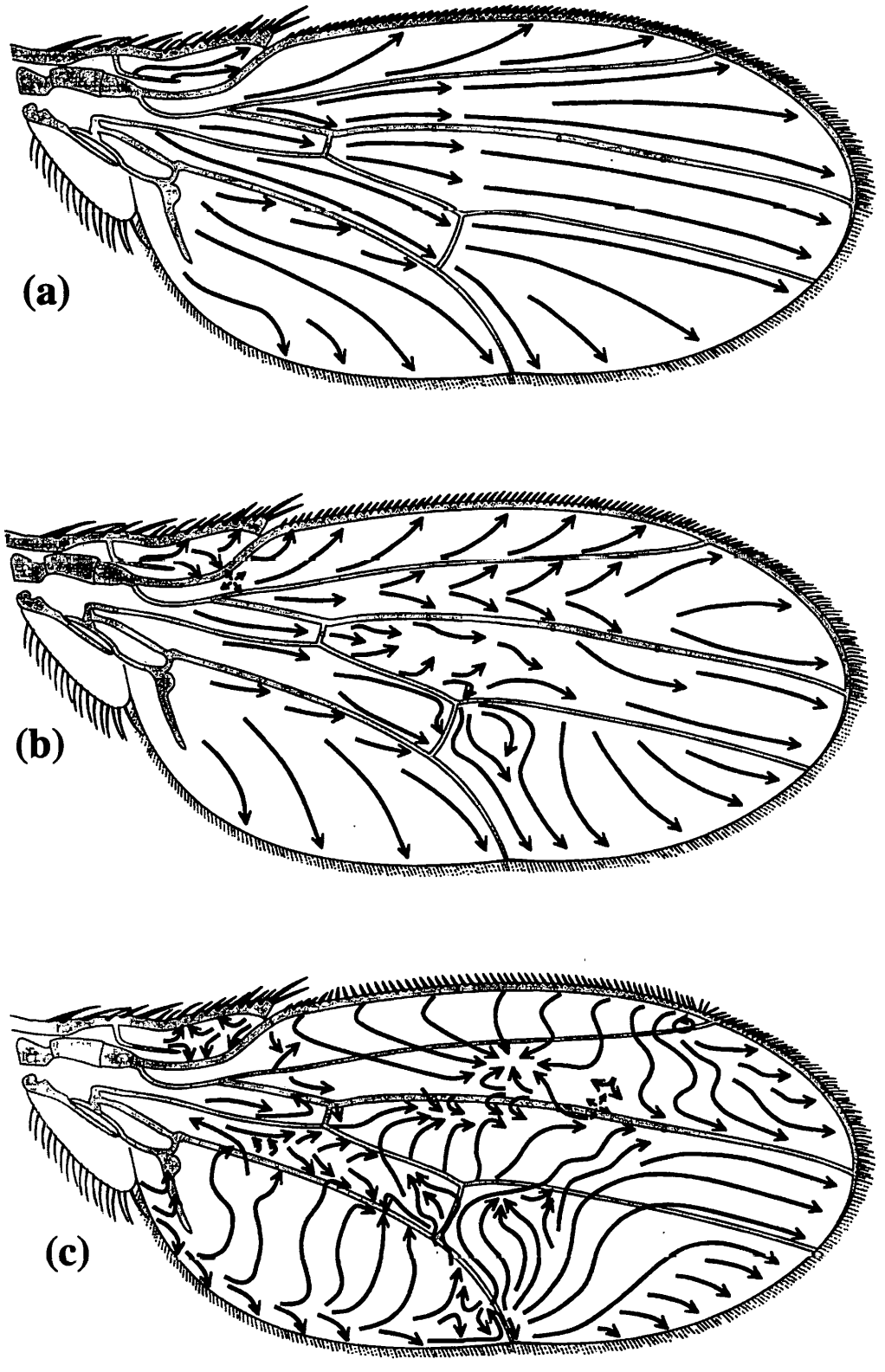


Fig. 1.2 Wing hair polarity phenotypes of (a) wild type; (b) *pk-sple* (Class 2, eg. $pk^{pk\text{-}sple13}$) and (c) *pk* (Class 5, eg. pk^{pk1}) flies. From drawings by David Gubb (unpublished).

absent. Very occasionally double ectopic joints are seen.

The eyes of both pk^{pk} and pk^{sple} single mutants appear to be completely wild type when viewed under a binocular dissecting microscope. However, when viewed internally about a third of the ommatidia of pk^{sple} eyes show dorso-ventral (D-V) mirror image reversals and about 1% show anterior-posterior (A-P) mirror image reversals (Gubb, 1998).

1.4.3 The $pk^{pk\text{-}sple}$ mutants

The double $pk^{pk\text{-}sple}$ mutants have an altered polarity phenotype in all the regions affected by the two single mutants, that is, the wings, notum, legs and abdomen, but it is less extreme (Figure 1.2). The wing phenotypes all fall into Class 1 or 2 as described by Coulson. The multiple hair phenotype is slightly increased, with doubled hairs seen in about 5% of the cells of the wing. The polarity disruptions to the hairs, bristles and B.S.V.s are also weaker. Mirror image duplications of the joints are restricted to the third and fourth tarsi, with incomplete ectopic joints in the more proximal tarsal segments.

These double mutants have a rough eye phenotype which is not seen in either of the single mutant classes. It has been shown that A-P and D-V reversals of the ommatidia occur with similar frequency in homozygous and hemizygous $pk^{pk\text{-}sple}$ mutants with some regions showing complete disruption of the hexagonal stacking, leading to the rough eye appearance (Gubb, 1998).

1.5 Other tissue polarity mutants.

1.5.1 *frizzled (fz)*

Mutations of this gene affect all aspects of planar polarity (Adler *et al*, 1987). There is abnormal hair and bristle polarity in all regions of the body, with hair polarity and multiple hair phenotype in the wing similar to that of pk^{pk} mutants (Coulson, 1994). There is disruption of the ommatidia and an ectopic joint phenotype in the leg (Adler *et al*, 1987; Coulson, 1994).

Adler *et al* (1987) observed extensive allelic variation in a group of over twenty five alleles of *fz*. Although most of the alleles fell into a graded series, two were qualitatively different. These two alleles were phenotypically moderate but did not produce the rough eye phenotype that was characteristic of all other strong or moderate alleles, so it was concluded that they were tissue specific in effect. Although the hexagonal stacking remained normal, these alleles showed a weak eye phenotype with occasional ommatidia having reversed chirality (David Gubb, personal communication).

A cold-sensitive allele of *fz* has also been identified (Adler *et al*, 1994). Temperature shift experiments with this allele indicated that the requirement for FZ ends a few hours prior to prehair initiation. This fits with the theory that *fz* has a regulatory rather than structural role in planar polarity. This is also supported by the fact that some alleles of *fz* show directional non-cell autonomy. There is disruption of polarity of hairs distal to *fz* clones, and occasionally anterior and posterior to the clones, but never proximal (Vinson and Adler, 1987). Other alleles of *fz* are cell autonomous in clones and it was therefore suggested that *fz* has two functions in the wing; the non-autonomous function required for transmitting polarity information along the proximal-distal axis and the autonomous one involved in reading the polarity signal (*ibid*).

Expression of *fz* RNA and protein suggested that activation of the protein was the critical event for the development of planar polarity (Park *et al*, 1994). It was also shown that there was no change in the accumulation of FZ in either the wing discs or pupal wings of *pk*, *in*, *fy* or *mwh* mutants. This indicates that none of these genes function by regulating expression of *fz*.

The *fz* gene has been well characterised molecularly. The five exons of the principle 4 kb transcript are spread over more than 90 kb of genomic DNA (Adler *et al*, 1990). This transcript encodes a membrane protein with seven potential transmembrane domains (Vinson *et al*, 1989; Park *et al*, 1993). A second mRNA was also identified that encodes a truncated version of the protein. It was speculated that each of the two proteins could have a single function in tissue polarity. Subsequently it was shown that the first *fz* transcript can almost completely rescue the fz phenotype,

suggesting it is able to fulfil both *fz* functions (Krasnow and Adler, 1994). Molecular analysis of the various alleles of *fz* revealed that while nonautonomous alleles were associated with mutations in all regions of the gene, all four autonomous mutations were located in the same proline residue (Jones *et al*, 1996).

Zheng *et al* (1995) used a monoclonal antibody to the neural-specific elav protein to examine the morphogenesis of developing eyes of *fz* mutants.They observed that, in contrast to the wild-type situation, it was not always the anterior cell of the R3-R4 pair that became an R3 cell. They then generated fz^- clones in the eye and found that, of ommatidia where only one cell of the R3-R4 pair was fz^+, 88% had rotated such that the fz^+ cell was anterior to the fz^- cell. From their observations they conclude that *fz* is required for the correct direction of ommatidial rotation, the degree of rotation and the correct asymmetry arrangement of R3 and R4 cells.

1.5.2 *inturned (in)*

The original allele of *in* affects bristle and hair polarity on the wing, halteres and abdomen, the hair polarity on the legs and causes mirror image duplications in the tarsal joints (Gubb and García-Bellido, 1982; Coulson, 1994). There is also a wing hair phenotype with a large number of the cells (about 70-80%) producing two or three hairs. Most of the alleles isolated subsequently are similar to this, but a few have distinctly weaker phenotypes (Adler *et al*, 1994).

A 3.5 kb transcript of the *in* gene has an ORF of 869 amino acids and contains no introns. The protein produced by this cDNA is thought to contain at least one transmembrane domain, along with a number of protein sequence motifs for postranslational modification (Park *et al*, 1996)

1.5.3 *multiple wing hairs (mwh)*

This gene only has a polarity phenotype in the wings, halteres and antennae, and this phenotype is relatively simple. However, as its name suggests, it has an extreme multiple hair phenotype that is extended over the entire body (Gubb and

García-Bellido, 1982). All the cells of the wing produce more than one hair and the number of hairs per cell is higher than is usually seen in other mutants (Coulson, 1994).

1.5.4 *dishevelled* (*dsh*)

The *dsh* gene is a segment polarity gene with a similar phenotype to *wingless* (*wg*) throughout development (Klingensmith *et al*, 1994) (see also section 1.7). One allele, dsh^1, shows a planar polarity phenotype in all regions of the body, with mirror-image duplication of the tarsal joints and a rough eye phenotype as well as disruptions to the hairs and bristles (Theisen *et al*, 1994). There is also a very weak multiple hair phenotype, with 0.5 to 1% of wing cells producing extra hairs (Coulson, 1994).

The *dsh* cDNA consists of a single ORF coding for 623 amino acids producing a 70kD protein. It has three hydrophobic regions, but their predicted structure and average hydrophobicity indicate that they probably do not span the membrane (Klingensmith *et al*, 1994). The sequence also has limited homology to junctional proteins and contains several regions conforming to the criteria for PEST sequences (Rogers *et al*, 1986). These regions, rich in proline (P), glutamic acid (E), serine (S), threonine (T) and to a lesser extent aspartic acid, are thought to correlate with rapid degradation of proteins.

1.5.5 *fuzzy* (*fy*)

Mutations in *fy* affect bristle and/or hair polarity on the wing, notum, haltere, abdomen and legs. The polarity phenotype in the wing is relatively weak but the multiple hair phenotype is strong, with approximately 80-90% of the wing cells producing two or three hairs (Coulson, 1994). In the leg, the hairs, but not the bristles, are affected and ectopic joints are not seen.

The *fy* gene was cloned and sequenced in our laboratory (Collier and Gubb, 1997). The cDNA has an ORF of 416 amino acids that is interrupted by two small introns. There are thought to be four transmembrane domains in the protein encoded by the *fy* transcript and at least one domain appears to be conserved in vertebrate proteins.

1.5.6 *fritz* (*frtz*)

Alterations are seen in the same regions as in *fy* and there is a similar multiple wing hair phenotype. A weak ectopic joint phenotype is also seen (Coulson, 1994).

1.5.7 *starry night* (*stan*)

This gene was formerly known as *frazzled* (*frz*). Mutants have bristle and hair polarity phenotype on the wing, notum and haltere, but not the leg. Mirror image ectopic joints are seen in the tarsi and there is a rough eye phenotype. The multiple wing hair phenotype is similar to that of *pk* and *fz*, with a small number of cells producing two or three hairs (Coulson, 1994).

1.5.8 *Van Gogh*; allelic to *strabismus* (*stbm*)

The *stbm* gene was recently identified as a novel mutant that causes a rough eye phenotype. It also has an effect on bristles and hairs and causes mirror image duplications of the tarsi (Wolff and Rubin, 1998).

The *stbm* gene encodes a protein with a putative PDZ domain binding motif as well as two possible transmembrane domains (Wolff and Rubin, 1998). PDZ domains are globular domains named after three proteins in which they occur; PSD-95 (mammalian postsynaptic density protein), DLG (Drosphila lethal (1) discs large-1) and ZO-1 (mammalian Zonula occuludentes protein-1) (reviewed by Ponting *et al*, 1997; Songyang *et al*, 1997). They are thought to be protein-binding regions with two distinct mechanisms for binding: they may bind to specific recognition sequences at the carboxyl termini of proteins or they can dimerize with other PDZ domains (Fanning and Anderson, 1996).

Van Gogh (*Vang*) was found to be allelic to *stbm* (Taylor *et al*, 1998). Many alleles had the additional dominant phenotype of a swirl in the wing hairs in region C'. Mitotic clones of the gene show domineering non autonomy, with wild type cells proximal, but not distal, to the clones showing altered polarity with their hairs pointing away from the clone (*ibid*). This is in direct contrast to the directional non-

autonomy seen with *fz* clones where wild type cells distal to the clone are affected and their hairs point towards the clone (Vinson and Adler, 1987).

1.5.9 Double mutant combinations

The phenotype of double mutant combinations has also been studied (Gubb and García-Bellido, 1982). It was found that in double mutant homozygotes, the following rules usually apply: regions unaffected by either of the mutants remain wild type; regions affected by one of a pair retain that polarity in the double mutant; regions affected by both members of a pair show an intermediate phenotype. When both mutants are heterozygous, polarity is wild-type; when one is homozygous, it is not affected by the other heterozygous mutant. A few exceptions to these rules were noted in the above study. For example, *pk; in* flies show *in* polarity in the thorax and wings and *pk*; *mwh* flies express *mwh* polarity except that the triple row shows *pk* polarity.

1.5.10 *dachsous* (*ds*)

This is not strictly a planar polarity gene as mutations result in adults with abnormally shaped legs and wings as well as a planar polarity phenotype in the wing (Adler *et al*, 1998). The polarity phenotype resembles that of the *fz*-like tissue polarity mutants but is unique in that, in some regions, the hair polarity is reversed and hairs point proximally. The phenotype of *ds* mutants is also more variable than is typical for other planar polarity mutants, but affects the same range of morphogenetic processes.

The *ds* gene has eleven introns, including one of about 40 kb, and encodes a membrane protein with several cadherin domains (Clark *et al*, 1995). Cadherins are calcium-dependant cell adhesion molecules that are present in most cell types (reviewed by Kemler, 1993). They form complexes with cytoplasmic proteins known as catenins and it is thought that the formation of these complexes in some way regulates extracellular adhesiveness.

Adler *et al* (1998) examined double mutants between *ds* and various planar polarity mutants. They found that the function of *fz*, *in* and *mwh* was required for the

development of the *ds* polarity pattern. Clones of *ds* display domineering cell nonautonomy and *ds* was also shown to enhance the nonautonomy of *fz* clones. The authors propose that the novelty of the *ds* polarity pattern indicates that *ds* does not produce its polarity phenotype by inactivating the *fz* pathway, but by altering the direction of *fz* signalling.

1.6 The polarity genes can be divided into three groups

The planar polarity genes have been divided into three groups, based on two criteria (Adler, 1992). The first criterion depends on how many structures have their polarity affected by the gene in question. Mutations in *pk, fz* and *dsh* alter bristles, hairs and ommatidia to some extent (Gubb and García-Bellido, 1982; Heitzler *et al,* 1993; Adler *et al,* 1987; Theisen *et al,* 1994). These are designated group 1 polarity genes. It is suggested that because the three developmental units affected are so different structurally, it is unlikely that the genes involved encode structural components. Therefore it is proposed that they encode proteins that are essential for the generation or transduction of a general intercellular polarity signal (Adler, 1992). The *fy* and *in* genes have an effect on both bristles and hairs, but not ommatidia and are classified as group 2 genes. Group 3 contains just *mwh*. This gene alters hair polarity and affects the morphology of the antennal bristle plus the orientation of the wing triple row bristles.

The second criterion is based on the polarity patterns of the wing hairs in each group. Group 1 mutants show dramatic alterations in the polarity patterns of the wing; for example, in pk^{pk} wings, the polarity of most of the hairs is altered, with many of them pointing proximally. A few cells in the wings of this group of mutants produce two or three hairs, which are typically slightly smaller than the hairs produced by single hair cells. The effects on the polarity pattern of the wing hairs in group 2 mutants appears to be less severe than that seen in group 1 mutants. However in this group, many more of the wing cells produce two or three hairs. Finally, *mwh* has a weak effect on the polarity pattern, but all the cells produce more than one hair. They have between one and five long hairs as well as other very small ones.

It has been shown that the distribution of the F-actin bundles associated with prehair initiation is altered in different ways by the different polarity mutants (Wong and Adler, 1993). In group 1 mutants, they are displaced towards the apical centre of the cell, while in group 2 and group 3 mutants up to three actin bundles are found at the periphery of the cell. The group 3 mutant *mwh* also has additional late-forming actin bundles that produce the small hairs.

It is possible to order genes in a regulatory pathway by using epistasis analysis. When single mutations of different genes produce varying phenotypes from each other, and the double mutant phenotype resembles one of the phenotypes produced by a single mutation, this gene is said to be epistatic to the other (Avery and Wasserman, 1992). The two genes are assumed to be acting in the same pathway. If the two genes in a double mutant combination are in different pathways, the double mutant phenotype is expected to be intermediate between the two single mutant phenotypes. Wong and Adler (1993) used these criteria to suggest that the three phenotypic groups of polarity mutants also corresponded to epistasis groups. They proposed that *fz*, *dsh* and pk^{pk} function upstream of *fy* and *in*, which in turn are upstream of *mwh*.

A more extensive study of a larger group of mutants found that this was an oversimplification (Coulson, 1994). Most of the double mutant combinations were found to follow the rules proposed by Gubb and García-Bellido (1982) but there were exceptions. For example, in terms of the multiple hair phenotype on the wing, *mwh* in combination with any other mutant tested gave a *mwh* phenotype (as previously noted by Gubb and García-Bellido, 1982 and Wong and Adler, 1993). Where epistatic effects were seen, they were often not complete. For instance, in *fy pk* flies, the hair polarity was *fy*, but the multiple wing hair phenotype was reduced compared to *fy* alone. Lack of function alleles of *fz* and *pk* do not show an epistatic interaction. It was concluded that the polarity genes could not be ordered in a simple linear pathway, although the data does suggest pathways involving subsets of these genes.

1.7 Planar polarity genes and signalling pathways

Until relatively recently, little was known of how the planar polarity genes fitted into the overall picture of development. Over the last few years, they have been implicated in various developmental pathways and these are discussed below.

1.7.1 The Wingless signalling pathway

The *wingless* (*wg*) gene was originally described as a mutation affecting wing and haltere development and causing abnormalities of the mesothorax (Sharma and Chopra, 1976). It was demonstrated to be non-cell autonomous in small clones (Morata and Lawrence, 1977b). Subsequently *wg* was shown to belong to the segment polarity class of genes (Nüsslein-Volhard and Wieschaus, 1980) with most alleles being recessive embryonic lethals. In wild-type larvae, the anterior of each segment is marked with a band of denticles, while the posterior part is naked. In *wg* mutant larvae the naked posterior part of the pattern is deleted and the larvae apparently lack all segment boundaries.

Baker (1987) cloned sequences from the region of *wg* and demonstrated that it is expressed in the most posterior cells of anterior compartments. Study of the subcellular localisation of WG using the electron microscope indicated that it is secreted onto the surface of cells and into the extracellular matrix (van den Heuvel *et al,* 1989). Comparison of the chromosomal position and restriction map of the clone produced by Baker (1987) with the *Drosophila* homologue of the mouse *int*-1 gene showed that they were identical (Rijsewijk *et al*, 1987). Cabrera *et al* (1987) demonstrated that antisense RNA made from the putative *wg* cDNA gave a wg phenotype when injected into wild-type eggs. Partial sequence of this cDNA was shown to be identical to that of D*int*-1.

The *int*-1 gene was subsequently renamed *Wnt*-1 (a combination of *wg* and *int*) and this class of proteins are now known as the WNT family. WNTs have been identified in many species and are known to be developmentally important signalling molecules (for reviews see Nusse and Varmus, 1992; Cadigan and Nusse, 1997; Dale, 1998). *Wnt* genes encode secretory glycoproteins that are capable of inducing different biological responses in different cellular contexts (Nusse and Varmus, 1992). There is some evidence of different functional domains in WNT proteins

(Cadigan and Nusse, 1997). Truncated WG appears to be internalised and distributed normally, but fails to transduce signal (Hays *et al*, 1997). However, a mutant that disrupts the late function of *wg* has a missense mutation in the same region that is deleted in another mutant that retains late function (Bejsovec and Wieschaus, 1995). This implies that the overall conformation of the molecule, rather than discrete functional domains, may control the different signalling activities.

It as been shown that *wg* is developmentally important in many ways. For example, Simcox *et al* (1989) cultured various mutant embryos *in vivo*. They found that no thoracic discs were recovered from *wg* mutant embryos; *engrailed* (*en*) embryos produced thoracic discs at about 10% of wild type frequency. The authors suggested that thoracic imaginal primordia arise within two adjacent bands of embryonic cells expressing *wg* and *en* respectively.

Bejsovec and Martinez Arias (1991) investigated the interaction between *wg* and *en*. They looked at the expression pattern of the two genes and studied a temperature sensitive allele of *wg* to assess the requirements for *wg* at different times during development. At least three qualitatively different functions of *wg* for epidermal pattern formation were postulated. These authors concluded that there are two distinct mechanisms by which positional information is specified at different times and in different locations in the embryo; cell-cell interactions are important early in development, while the establishment of a graded distribution of WG correlates with the progressive specification of naked cuticle.

Localised ectopic expression of *wg* in embryos mutant for both *wg* and *en* was studied using the enhancer trap line *paired-Gal4* (Lawrence *et al*, 1996). [1] The lawn

---

[1]A system that allows ectopic expression of a target gene in specific tissues or cell types of *Drosophila* facilitates the study of development (Brand and Perrimon, 1993). In this system, the target gene is separated from a transcriptional activator in two distinct transgenic lines and only when the two lines are crossed is the target gene turned on in the progeny. The first step in producing this system involved the generation of lines expressing a transcriptional activator in a variety of patterns. The enhancer detection technique was developed as a way of identifying regulatory elements *in situ* in *Drosophila* (O'Kane and Gehring, 1987; Bellen *et al*, 1989). The *lacZ* gene of *Escherichia coli* was fused to the weak P-transposase promoter and this reporter gene responded to nearby enhancer-like elements when integrated into the

of denticles seen in *wg⁻en⁻* embryos was interrupted by stripes of naked cuticle and these denticles showed changes in polarity depending on the proximity of the *wg*-expressing cells. It was concluded that a localised source of WG could have effects on cells beyond the source, but it was not clear how direct this was.

Zecca *et al* (1996) investigated WG morphogen function in the wing. They postulated that if WG exerts a long range action through a mechanism of sequential induction, ectopic expression of a membrane tethered form of WG should have similar effects to that of free WG. In contrast, they found that the tethered form only exerted a short range, all or none influence, suggesting free WG acts as a long range morphogen.

Using a temperature-sensitive *wg* mutation, Phillips and Whittle (1993) found that WG is necessary to specify a subset of wing and notum sensory organs that arise in or adjacent to *wg* expressing cells. The results suggested that *wg* could be involved in regulation of the *achaete-scute*-complex in proneural clusters.

Many components of the *wg* signalling pathway have been identified (for reviews, see Klingensmith and Nusse, 1994; Siegfried and Perrimon, 1994). Mutations of *armadillo* (*arm*) and *wg* were shown to have similar effects on cuticular pattern and *en* expression (Piefer *et al*, 1991). By looking at clones of *arm* tissue, these authors also demonstrated that *arm* activity is required during the development of the imaginal discs. The *arm* gene encodes a protein that is the *Drosophila* homologue of the mammalian protein plakoglobin (Piefer and Wieschaus, 1990). Plakoglobin is a protein that functions on the interior of the cell as a component of

---

Drosophila genome (O'Kane and Gehring, 1987). Brand and Perrimon (1993) fused the coding sequence from the transcriptional activator *Gal4* to the P-transposase promoter to produce a vector that could direct expression of *Gal4* in a wide range of patterns, eliminating the need to link numerous promoters to the *Gal4* gene. *Gal4* is a yeast transcriptional activator that can activate transcription in flies, but only from promoters that have GAL4 binding sites (Fischer *et al*, 1988). A vector was produced in which a target gene could be subcloned behind a tandem array of five GAL4 binding sites (the UAS or Upstream Activation Sequence) and upstream of the SV40 transcriptional terminator. When a line carrying this transgenic insert is crossed to an enhancer trap *Gal4* line, the target gene is activated where *Gal4* is expressed.

adhesive junctions. Riggleman *et al* (1990) demonstrated that the subcellular distribution of *arm* protein closely resembles that of F-actin in many structures and this may reflect an association between *arm* and the cytoskeleton. However, not all actin-rich structures contain high levels of ARM.

ARM is also homologous to ß-catenin, a peripheral cytoplasmic protein that forms complexes with a family of cell adhesion molecules (CAMs) called cadherins. This complex formation links cadherins to the actin filament network and is crucial for the adhesive function of cadherins (reviewed by Kemler, 1993). It has been demonstrated that ARM is a component of a membrane-associated multi-protein complex similar to vertebrate adherens junctions (Peifer, 1993). Until recently, cell-cell junctions were seen only as architectural elements of the cell, but it is now known that they also function as signalling centres (reviewed by Kirkpatrick and Peifer, 1995). Some proteins are integral junctional components whereas others are localised to specific junctions, probably via protein-protein interactions with integral junction proteins.

Epistatic analysis was used to order genes in the *wg* pathway. Siegfried *et al* (1994) used double mutant combinations with *zeste-white* 3 (*zw*3, also known as *shaggy*, *sgg*) because loss of *zw*3 activity results in an embryonic phenotype opposite to that of *wg* mutant embryos. WG signalling is known to function by inactivating or antagonising *zw*3 (Siegfried *et al*, 1992). It was concluded that *dishevelled* (*dsh*) and *porcupine* (*porc*) act upstream and *arm* acts downstream of *zw*3. Noordermeer *et al* (1994) used double mutants with HS-*wg*, a transgenic strain that expresses *wg* under the control of a heat-shock promoter. They found that *dsh* and *arm* are required for the induction of ectopic EN in HS-*wg*, whereas *hedgehog* (*hh*) and *porc* are not. They suggested that *dsh* and *arm* act downstream of *wg*. Further characterisation of *arm* suggested that *wg* signal triggered an alteration in the intracellular distribution of *arm* protein and confirmed that *arm* functions downstream of *zw*3 (Peifer *et al*, 1994).

Study of a temperature sensitive allele of *wg* indicated that *wg* has a function in patterning the wing margin (Couso *et al*, 1994). At the permissive temperature, embryos homozygous for this allele develop normally but hatch into adults that

display some defects in the epidermis, including some missing bristles on the wing margin. Shifts to the restrictive temperature produce varying phenotypes depending on the timing; animals shifted before 96 hours after egg laying lack all features characteristic of the wild type wing margin. Clonal analysis of *arm*, *dsh* and *zw3* in the same study indicated that the relationship between these genes and *wg* is the same in the patterning of the wing margin as it is in the larval epidermis.

As WG signalling requires the activity of ARM, and ARM is the homologue of

B-catenin, a link was suggested between cell-adhesion and WG signalling (reviewed by Peifer, 1995). It was postulated that by balancing the distribution of ARM/*B*-catenin between intercellular junctions, the cytoplasm and the nucleus, the cell uses information from WNT signalling to differentially affect cell adhesion (Eaton and Cohen, 1996). The *arm* protein has three regions, including a series of thirteen imperfect "ARM" repeats. These are conserved protein-protein interaction motifs found in a variety of otherwise unrelated proteins. By generating mutations which together affected most regions of the *arm* protein, Orsulic and Peifer (1996) showed that different regions of ARM are required for cell adhesion and for WG signalling. There is also some overlap; for example, mutations in the central ARM repeats affect both cadherin binding and response to WG.

Sanson *et al* (1996) further demonstrated that ARM has a function in the WG pathway that is distinct from its role in adherens junctions by looking at cadherin (which forms complexes with catenins). Overexpression of full length cadherin increases adhesion whereas overexpression of a construct consisting of the intracellular domain alone decreases it, indicating that the intracellular domain has a dominant negative effect on adhesion. Overexpression of this domain has a similar effect on WG signalling to overexpression of the full length protein: both give a phenotype similar to a *wg* loss-of-function mutation. This suggests that cadherin function is not directly required for WG signalling.

Another component of the WG signalling pathway was identified by homology to T cell factor (TCF), a vertebrate transcription factor that physically interacts with

B-catenin (Van de Wetering *et al*, 1997). This gene, named *dTCF*, is a segment polarity gene and appears to act downstream of *arm*. Repeats three to eight of ARM were shown to interact with dTCF in the yeast two hybrid system. An *arm* mutant in which repeats three to six are deleted and WG signalling is eliminated was shown to completely block dTCF binding.

1.7.2 *dishevelled* and *frizzled*

The segment polarity gene *dsh* is also involved in planar polarity. Theisen *et al* (1994) noted that all cells in which polarity is evident are affected by *dsh* mutations and that *dsh* functions in a cell autonomous manner. They found that one copy of a *dsh* transgene rescued lethal alleles of *dsh* but the adult flies have planar polarity defects. Eight per cent of them also exhibit the abnormalities associated with *wg* mutants, and this increases to 100 per cent if the flies are also heterozygous for *wg*.

By producing antibodies specific to the *dsh* protein and using a cell culture assay for WG activity, Yanagawa *et al* (1995) found that DSH is a cytoplasmic protein and that extracellular WG leads to an increase in the phosphorylation of DSH. They proposed that signalling by WG leads to additional phosphorylation of the DSH present in the cytoplasm and translocation of this DSH to a membrane component, resulting in DSH becoming active in signal transduction.

The planar polarity gene *fz* has been shown to interact with *dsh*. Krasnow *et al* (1995) made use of a *fz* transgene under control of a heat shock (hs) promoter to look at the relationship between *fz*, *dsh* and pk^{pk}. Early heat shocks of this transgene (given six or more hours before prehair initiation) gave a phenotype like that of *fz*, *dsh* or pk^{pk} mutants; late heat shocks (given just before prehair initiation) resulted in a *fy* or *in* like phenotype. It was found that in embryos mutant for *dsh*, expression of the hsfz-late phenotype was completely blocked and the hsfz-early phenotype was also suppressed. Conversely, reduction of the *dsh* dosage in a loss-of-function mutant of *fz* gave a dramatic enhancement of the phenotype. This implied that both the early and late overexpression phenotypes result from an increase in FZ activity. In contrast to the *dsh* results, pk^{pk}; *hsfz* double mutants exhibited a "mixed" phenotype that

included distinctive elements of each of the single mutants. Hence *dsh* appears to encode a dosage sensitive component that is required for *fz* function and probably functions downstream of *fz*, whereas *pk* does not seem to function in the same linear pathway.

Adler *et al* (1997) used transgenes under the control of hs promoters to study gradients of *fz* expression in the wing. Localised expression of these transgenes was induced by applying hot wax to defined areas of the pupal cuticle. Distal waxing of *hs-fz* pupae resulted in large regions of reversed polarity in the wing. A typical planar polarity phenotype was usually seen distal to the region of reversed polarity. It was found that endogenous *fz* and *pk* were not required for the reversed polarity phenotype, but endogenous *dsh* was. Distal waxing of *hs-dsh* pupae gave a strong polarity phenotype but no reversed polarity. These authors therefore proposed that a polarising signal is transmitted from cells with higher FZ levels to cells with lower levels.

1.7.3 Notch signalling

The *Notch* (*N*) gene encodes a receptor with a single transmembrane domain that is required for many different cell fate decisions during development (reviewed by Artavanis-Tsakonas, 1999). There is also evidence for cross-interactions between Notch and WG signalling in several different processes (Couso and Martinez Arias, 1994). Overexpression of *wg* or *dsh* in the wing results in the formation of ectopic bristles (Axelrod *et al*, 1996). The number of ectopic bristles generated by *dsh* overexpression increased in *N* heterozygotes. This suggests that Notch and DSH have opposing effects on bristle induction. These authors also found evidence for a physical interaction between DSH and a domain of Notch that had no previously assigned function. They proposed that the interaction between the WG and Notch pathways is mediated by DSH.

Notch signalling has been implicated in many developmental processes and several elements of Notch pathways have been identified (Artavanis-Tsakonas, 1995; Bray, 1998). For example, de Celis *et al* (1996a) showed that Notch is locally activated in dorsal and ventral cells at the dorsal/ventral (D/V) boundary in the wing

imaginal disc through the action of its ligands *Delta* (*Dl*) and *Serrate* (*Ser*). *Ser*⁻ clones caused scalloping of the wing similar to *N* mutant clones when induced before D/V lineage segregation or in the dorsal compartment when in contact with the wing margin. *Dl*⁻ clones caused extensive scalloping when the clones included dorsal and ventral cells, while clones in the ventral compartment resulted in small nicks. The phenotypes of *Suppressor of Hairless* (*Su(H)*) and *Enhancer of split* (*E(spl)*) mutant clones were found to be consistent with these genes functioning in the Notch pathway in its role at the D/V boundary.

It has been proposed that the interaction of DL with Notch results in the activation of SU(H) and subsequently in the expression of *E(spl)* genes. The *E(spl)* genes have been shown to have distinct patterns of expression in the imaginal discs and this expression was shown to be Notch dependent (de Celis *et al*, 1996b). Good correlation was found between the location of genes within the *E(spl)* complex and their patterns of expression, which fits with the theory of shared regulatory elements interspersed through the complex.

Both DL and SER are expressed at high levels within clones of ectopic WG when they are present in the wing pouch (de Celis and Bray, 1997). Low levels of expression were also detected outside the clones, suggesting that WG is acting in neighbouring cells. The transcriptional activation of *Dl* and *Ser* in response to WG raises the possibility that Notch regulation of ligand expression is mediated by WG. However, *Dl* expression is also elevated where WG is not detected. Overall, the effects of Notch signalling on ligand expression were found to be context dependent and specific Notch targets were shown to participate in determining whether Notch signalling activates or represses ligand expression.

The WG dependent genes *vestigial* (*vg*) and *Distal-less* (*Dll*) are involved in defining the D/V axis of the wing. The proteins of these genes are expressed in broad domains centred on the D/V boundary. If an ectopic source of WG is generated along the anterior-posterior (A/P) boundary, similar domains of VG and DLL expression are centred on the A/P region (Neumann and Cohen, 1997a). Mitotic clones of a temperature sensitive allele of *arm*, that impairs WG signalling but does not alter the cell adhesion function of ARM, exhibited little or no VG expression and DLL

expression was lost. It was concluded that WG expression at the D/V boundary directly activated *vg* and *Dll* throughout the wing pouch.

1.7.4 Polarity and the *Drosophila* eye

Hedgehog (HH) is a secreted signalling protein that is important in many aspects of development. It has been shown that ***hh*** expression behind the morphogenetic furrow of the developing eye is required for furrow progression (Ma *et al*, 1993). Ectopic expression of ***hh*** in the eye generates ectopic furrows that can progress in all directions across the eye disc (Chanut and Heberlein, 1995). The direction of assembly of individual ommatidia behind the ectopic furrow parallels the direction of furrow progression, which may therefore determine ommatidial polarity.

Wehrli and Tomlinson (1995) found evidence that the movement of the furrow could be involved in establishing anterior/posterior (A/P) polarity in the eye. They looked at clones of *patched* (*ptc*) and *Protein kinase A* (*DCO*). Posterior to these clones, the movement of the furrow is reversed and in this region the ommatidia are of the wrong chiral form, probably resulting from inversion in the A/P axis. These authors therefore proposed a wave model for the generation of planar polarity, which could involve either movement of a polarising gradient or local cell-cell communication. A similar study found that chiral form and direction of rotation were not directly linked to dorsoventral position in the disc (Strutt and Mlodzik, 1995). Instead, it was suggested that equators are defined by the leading edge of the furrow and the most mature ommatidia in a row, those at the dorsoventral midline, then form an organising centre. This would define the position of the equator and the information would then be propagated outwards. A similar model was also proposed by Gubb (1993).

Initiation of the furrow appears to be restricted to the posterior margin of the eye disc by the presence of *wg* at the lateral margins (Treisman and Rubin, 1995). Expression of *wg* at the edge of the disc may be one of the signals involved in ommatidial rotation because the direction of rotation is reversed in ommatidia adjacent to a scar caused by ectopic *wg* expression. Removal of *wg* function from the eye leads to the formation of ectopic furrows. Ommatidia generated by ectopic

furrows do not rotate, whereas those formed by the endogenous furrow do and so a "two-vector" model for ommatidial rotation was proposed (Ma and Moses, 1995). The model assumes that the nascent cluster can sense two scalar properties; the direction to the furrow and the direction to the equator. These co-ordinates, which are normally perpendicular, would define the two chiral forms and the two orientations seen. Introduction of an ectopic furrow would result in new directional values for the furrow vector but not the equator, resulting in alterations in chirality and orientation.

Tomlinson *et al* (1997) demonstrated that overexpression of either *dsh* or *fz* in the eye gave the same phenotype as the loss-of-function mutations of either of these genes, that is, a similar frequency of correct and incorrect chiral forms. Overexpression of *zw3* (a component of the *wg* signalling pathway) produced a similar, although weaker, phenotype, implying that it could function in the same pathway as *fz*. It was also shown that when *wg* was misexpressed in the eye, there was a non-autonomous effect as ommatidia on one side of the clone were repolarised up to six ommatidial diameters away from the clone. This may indicate that a WNT signalling pathway is involved in the polarising mechanism.

Wehrli and Tomlinson (1998) showed that equatorial/polar (Eq/Pl), also known as dorsoventral (D/V), polarity in the eye was altered at a polar position relative to the centre of gain of function *wg* clones and extending into wild-type tissue on the polar side of the clone. Also, the ability of these *wg* clones to induce polarity reversals was highest at the equator and lowest at the pole. Other members of the *wg* pathway were also found to be involved in establishing this polarity. The authors concluded that Eq/Pl polarity is established in two steps, a polarising step and a read-out step, and is independent from A/P polarity. They postulated that *wg* or another unidentified *wnt* gene is the primary organising molecule in the Eq/Pl system.

An enhancer trap line that expresses *white* (*w*) only in the dorsal field of the eye was used to study the effect of pk^{sple} (Choi *et al*, 1996). The dorsal specific expression of *w* was not changed in pk^{sple} mutants, indicating that the global dorsoventral signal is intact. Therefore it appears that dorsoventral identity is established early and pk^{sple} acts later to regulate the polarisation of the ommatidia.

This result has also been confirmed with the double mutant $pk^{pk\text{-}sple13}$ (David Gubb, personal communication).

Expression of the putative transcriptional activator *mirror* (*mrr*) is restricted to the dorsal half of the eye (McNeill *et al*, 1997). It was found that changes in *mrr* expression could create ectopic equators or alter the path of the normal equator. It was suggested that *mrr* was a key component in defining the dorsal-ventral boundary, but not the sole determinant. It has been demonstrated that induction of the dorsal expression of *mrr* is mediated by *wg* (Heberlein *et al*, 1998).

Two studies showed that *fringe* (*fng*) is expressed in the ventral half of the eye and is also important in establishing the D/V boundary (Cho and Choi, 1998; Domínguez and de Celis, 1998). Overexpression of *fng* using the *Gal4/UAS* system resulted in eyes that were dramatically reduced in size (Domínguez and de Celis, 1998). The small eye phenotype could be rescued by an activated form of Notch, but the ommatidia in these eyes were abnormally arranged. When activated forms of Notch were expressed anterior to the morphogenetic furrow, ommatidial differentiation was not affected but the resulting adult eyes showed erratic D/V polarity. It was concluded that Notch signalling at the equator was essential for eye growth and organisation of mirror symmetry.

The importance of Notch signalling in establishing planar polarity in the eye has been demonstrated by two groups (Cooper and Bray, 1999; Fanto and Mlodzik, 1999). In *fz* and *dsh* mutants, some ommatidia remain symmetrical, containing R3/R3 or R4/R4 photoreceptor pairs instead of the R3/R4 pair. Fanto and Mlodzik (1999) showed that mild loss-of-function *N* mutants had occasional ommatidia with R3/R3 pairs, whereas gain-of-function *N* mutants had some R4/R4 ommatidia. When *Dl* mutant clones were generated in the eye, it was found that in ommatidia mosaic for *Dl* in R3/R4, the Dl^{+} cell always adopted the R3 fate. High levels of FZ signalling were shown to induce R3, whereas full activation of Notch induced R4.

Cooper and Bray (1999) found that the earliest manifestation of a difference in the R3/R4 pair was a specific expression pattern of the gene *E(spl) mδ*. A 500 base pair enhancer fragment of this gene (*mδ*0.5) was shown to display an R4 specific subset of the expression pattern. In $sple^{1}$ mutant discs, high levels of *mδ* 0.5 were

detected in both cells of the R3/R4 pair for a prolonged period. Subsequently *m∂0.5* became restricted to one cell, but whereas in wild-type it is always expressed in the polar R4 cell, in *sple*^{1} it refines to either cell with no bias. These authors propose that FZ/DSH sets up an initial bias in Notch activity between R3 and R4 by promoting DL activity and inhibiting Notch via SPLE in a co-ordinated manner.

1.7.5 WG receptors

For a long time, possible WG receptors remained elusive. Wang *et al* (1996) identified a large number of *fz* homologues in a wide variety of species. Bhanot *et al* (1996) studied *frizzled2* (*Dfz2*), a novel *Drosophila* gene that resembles other members of the *fz* family. Clone 8 cells, which respond to added WG, were shown to express *Dfz2*. Schneider (S2) cells are normally unresponsive to WG and do not express *Dfz2*. When S2 cells are transfected with and expressing *Dfz2*, they do respond to WG. It was concluded that DFZ2 functions as a receptor for WG, and suggested that *fz* proteins in general are receptors for WNTs. There is also evidence for a link between WNTs and FZ in *Caenorhabditis elegans* (reviewed by Ingham, 1996). Smoothened (SMO), the putative HH receptor, also appears to be related to FZ (reviewed by Orsulic and Peifer, 1996; Perrimon, 1996).

Two groups have reported studies on the relationship between WG and DFZ2 in the wing. Cadigan *et al* (1998) found that in the wing pouch, *Dfz2* is expressed in an inverse pattern to that of WG and the pattern is WG-dependent. Expression of a truncated version of *Dfz2* that can bind WG in cell culture but should not be able to transduce the signal to intercellular targets was found to block WG signalling. This agrees with the theory that *Dfz2* encodes a WG receptor. Zhang and Carthew (1998) demonstrated that overexpression of *Dfz2* in the developing wing blade led to the formation of ectopic wing margin bristles in the interior of the wing and this was dependent on the level of WG signalling. Overexpression of *fz* gave a polarity phenotype but no ectopic bristles, and the level of WG activity had no discernible effect on the strength of the polarity phenotype. Overexpression of both *Dfz2* and *fz* in the same wing produced a phenotype that was the sum of the two individual

phenotypes, suggesting that the two genes are involved in separate signalling pathways.

There is some evidence that *fz* and *Dfz2* may have partially redundant roles in Wingless signalling. It has been shown that dsRNA is a specific inhibitor of gene activity in *Caenorhabditis elegans* (Fire *et al*, 1998). Kennerdell and Carthew (1998) adapted this technique for use in *Drosophila*. They showed that injection of an equimolar mix of ds-*fz* and ds-*Dfz2* RNAs into an embryo caused a similar phenotype to injection of *wg* dsRNA. Epistatic analysis involving mutant embryos injected with ds-*fz* and ds-*Dfz2* RNAs indicated that *fz* and *Dfz2* both function downstream of *wg* but upstream of *zw3*. Bhat (1998) found that loss of *fz* or *Dfz2* caused the same embryonic phenotype in the central nervous system as loss of *wg*, although with only partial penetrance. An enhancement of the defects was seen in the absence of both *fz* and *Dfz2*, but penetrance was still not complete. Although the mutant embryos considered above lacked both maternal and zygotic *fz*, only zygotic *Dfz2* was eliminated. Hence the contribution of maternal *Dfz2* could be responsible for the incomplete penetrance of the phenotype, or there may be other *fz* genes in the genome. Another possibility considered is that reception of the WG signal and its transduction occurs through a multiprotein receptor complex of which the *fz* proteins are only a part.

Müller *et al* (1999) found that in *Dfz2* deficient embryos, the *wg* expression pattern is altered but ARM still accumulates in regions where WG is expressed. WG expression in embryos deficient for *fz* and *Dfz2* is lost in a temporal pattern similar to that seen in *wg* mutants. However, there is partial maintenance of EN expression, suggesting that the *fz Dfz2* double mutant does not eliminate all WG reception. Internalisation of WG still occurred in double *fz Dfz2* mutant embryos. Again, maternal *Dfz2* was not eliminated in these experiments, so it may contribute to either or both of the last two observations.

1.7.6 The RHO family and polarity signalling

Recently the RHO subfamily of the p21 RAS superfamily of GTPases have been implicated in the generation of planar polarity (Strutt *et al*, 1997). GTPases are

versatile molecular switches that are turned on by binding GTP and off by hydrolysing GTP to GDP (reviewed by Bourne *et al*, 1991). They are regulated by several classes of proteins, many of which are larger and more complex than their targets (Boguski and McCormick, 1993). Positive regulation is carried out by guanine nucleotide exchange factors (GEFs) and negative regulation by GTP-ase activating proteins (GAPs). DH (Dbl Homology) proteins, thought to be GEFS for the RHO family, often have separate catalytic domains suggesting they may be multifunctional proteins (Quilliam *et al*, 1995).

Members of the RHO subfamily have been implicated in cell-adhesion, regulation of the actin cytoskeleton, control of mitogen-activated protein kinase (MAPK) pathways and as components of signal transduction cascades (reviewed by Vojtek and Cooper, 1995; Machesky and Hall, 1996; Ridley, 1996). Harden *et al* (1995) isolated two *Drosophila* members of this family, *DRacA* and *DRacB*. They found that expression of a dominant inhibitory form of *DRacA* affects dorsal closure, disrupting the accumulation of both myosin and actin along the leading edge. Two other members of the group, *Rac1* and *Cdc42*, were shown to have roles in both apical basal polarity (Eaton *et al*, 1995) and planar polarity (Eaton *et al*, 1996) in the wing. *Rac1* was found to be involved in the placement of hairs, whereas *Cdc42* is critical for hair outgrowth (Eaton *et al*, 1996).

Strutt *et al* (1997) isolated several loss-of-function alleles of *Drosophila RhoA*. They found that *RhoA* appeared to be required for cell growth and/or viability. Clones of *RhoA* in the eye and the wing gave planar polarity phenotypes and overexpression of either *fz* or *dsh* in the eye was dominantly suppressed by *RhoA* mutations. RHO family proteins are involved in signalling to the nucleus through MAP kinase cascades, in particular via JNK/SAPK (JUN amino (N)-terminal kinase or stress-activated protein kinase) molecules (Kockel *et al*, 1997; Noselli, 1998). Mutations in *basket* (*bsk*), which encodes a JNK/SAPK homologue, give a suppression similar to that seen with *RhoA*. Strutt *et al* (1997) proposed that *RhoA* acts downstream of *fz* in the generation of planar polarity and that JNK/SAPK kinases are involved in polarity signalling. No interaction was seen between the overexpressed *fz* and *pk*$^{pk\text{-}sple}$.

BSK is involved in dorsal closure and JUN is thought to be its target; dominant negative forms or null alleles of *jun* result in a failure in dorsal closure (Kockel *et al*, 1997). JUN is a member of a family of transcription factors and has been shown to act downstream in the RAS signalling pathway in the eye (Bohman *et al*, 1994). The RAS pathway is involved in the determination of cell type in the eye (reviewed by Freeman, 1997). The function of JUN in this pathway seems to be redundant (Kockel *et al*, 1997).

A potential *rho* target, named *citron*, was isolated from a mouse cDNA library (Madaule *et al*, 1995). The *rho* binding site is not the same as that found in previously identified *rho* effectors, suggesting that *rho* interacts with more than one class of partners. Several other motifs were identified in Citron, including a type of zinc finger also found in other GTPase partners.

1.7.7 Planar polarity and WG signalling

Overall, the evidence indicates that planar polarity and WG signalling are two distinct processes. Boutros *et al* (1998) demonstrated that *sevenless*-driven overexpression of *fz*, *dsh* or *RhoA* in the eye gave a planar polarity phenotype, but that of *Dfz2* or *wg* did not. They also looked at the involvement of the different conserved domains of DSH in the different pathways. DSH has three conserved domains: a DIX domain, a PDZ domain and a DEP domain. The DIX domain is also found in Axin, a mouse protein involved in specification of embryonic axis (Zeng *et al*, 1997). PDZ domains are present in many proteins involved in signalling pathways (Ponting *et al*, 1997; see also section 1.5.8). DEP domains are found in a variety of proteins involved in signalling pathways (including DSH, EGL-1 and Pleckstrin) and are thought to have a role in regulating GTP-GDP exchange (Ponting and Bork, 1996). *In vivo*, a *dsh* construct lacking the DIX domain rescues the *dsh* phenotype in the eye to a similar extent to wild type *dsh*; a construct lacking the PDZ domain rescues to a weaker extent and one lacking the DEP domain does not rescue at all (Boutros *et al*, 1998). Conversely, constructs lacking the DIX or PDZ domains do not rescue larval lethal *dsh* alleles, whereas dsh^1, which carries a single amino acid substitution within

the DEP domain, does. It was thus suggested that DSH acts as a molecular signal transducer, with its domains being differentially involved in transmission of either planar polarity or *wg* signals.

Axelrod *et al* (1998) utilised a system that allows controlled expression of multiple proteins and analysis of their intracellular distribution. mRNAs encoding the proteins are synthesised *in vitro* then injected into *Xenopus* embryos. DSH alone is observed mostly in association with cytoplasmic vesicles, but when co-expressed with FZ relocalises to the membrane. If the DIX domain is deleted, DSH does not associate with the vesicles, but does relocalise to the membrane in the presence of FZ. Removal of the DEP domain has no effect in the absence of FZ, but this construct fails to relocalise to the membrane in the presence of FZ. The DEP domain alone is sufficient to relocalise DSH to the membrane in the presence of FZ. Deletion of most of the DEP domain has little effect on WG signalling, but gives a polarity phenotype. Thus membrane localisation appears to be important in the planar polarity signalling pathway. DFZ2, the putative WG receptor, fails to localise DSH to the membrane. It was also postulated that if the WG and planar polarity pathways utilise DSH in different ways, ectopically activating one of the pathways should titrate the activity of the other. This was shown to be the case.

1.7.8 Summary

The establishment of pattern is highly complicated and involves a combination of mechanisms. Many direct cellular interactions have been documented and the response to morphogens is also critical in some processes. BCD acts as a morphogen in the syncytial embryo, while WG and HH act as morphogens in some contexts (reviewed by Neumann and Cohen, 1997b).

WG signalling is clearly vitally important in development and many steps of the pathway are now known. To summarise, in the absence of WG signalling, cytoplasmic ARM is degraded, probably via the action of ZW3. The interaction of WG with DFZ2 (and possibly other members of the FZ family) activates DSH, shutting down the degradation of ARM. ARM then accumulates in the cytoplasm and nucleus and can bind to dTCF.

The planar polarity and WG signalling pathways are both mediated by DSH and both involve members of the FZ family as receptors. Other steps in the planar polarity pathway are not well understood. The ligand is unknown, although may be an unidentified WNT. A model for planar polarity signalling was described by Krasnow *et al* (1995). They proposed that on being activated by FZ, DSH inhibits FY/IN activity, thus positively regulating MWH. PK was postulated to act in a different branch of the pathway, although it was also considered possible for it to act in an independent pathway. This model is clearly far from being the whole picture as it does not take into account the other polarity genes nor recent developments in the field.

A model of planar polarity signalling must account for the fact that in different tissues polarity involves distinct units (for example, a single cell producing a hair in the wing, compared to clusters of ommatidia in the eye) and requires regulation of a wide range of cellular processes (Shulman *et al*, 1998). It has been proposed that planar polarity is a reflection of control of cell shape changes towards the end of development rather than fine grained positional information (Gubb, 1998). Studies on planar polarity in the eye have advanced the field greatly (reviewed by Reifegerste and Moses, 1999), but a full understanding of the planar polarity genes and how they interact with other signalling pathways during development is required.

1.8 Aims of this project

The *pk* locus has been well characterised phenotypically and genetically (Gubb and García-Bellido, 1982; Heitzler *et al*, 1993; Coulson, 1994; Gubb *et al,* in preparation) but at the start of this project little was known of its molecular characteristics. A 100 kb phage walk had been recovered and many of the mutants were mapped on this walk. The primary aim was to clone the transcript/s of *pk* and sequence them. The sequence data was then to be analysed and compared to other known sequences. It was hoped that this information, along with expression data, would help to explain the genetic complexity of *pk* and possibly elucidate its function. Ideally, rescue of the *pk* phenotypes by any transcripts located was also required.

During the course of the project, other objectives were raised. Transcripts besides those identified as *pk* were found in the region and these were also characterised. They included the *pk*-related gene, *espinas,* and the *necrotic* gene cluster (Chapter 2). The sequence of *pk* was examined and found to have homology to a number of sequences in the database and some of these were explored further.

CHAPTER 2. **CLONING AND SEQUENCING OF *prickle***

2.1 Introduction

The starting point for my investigations was a chromosome walk, produced by David Gubb with some assistance from myself in the later stages. The phage of the 150 kb chromosome walk were isolated from the genomic insert library of John Tamkun (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation). The starting point of this walk was a 3 kb genomic fragment with weak homology to *raf* (Mark *et al*, 1987). This "Draf2" fragment was selected as it hybridised to 43A1.2, the site of the *pk* region breakpoint of the *In(2LR)TE35B-4* (pk^{pk28}) and *T(2;3)pk78t* (pk^{pk29}) aberrations, as these were two of the *pk* mutations that had been mapped at the time. The walk was extended in both directions until the proximal and distal breakpoints defining the region had been uncovered. Each phage was mapped with the same panel of restriction enzymes (Figure 2.1)

The map positions and nature of the various mutants were determined by Southern blot hybridisation of genomic DNA or by chromosome *in situ* hybridisation (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation). The mutations were mapped to a region of about 70 kb (Figure 2.2). It was

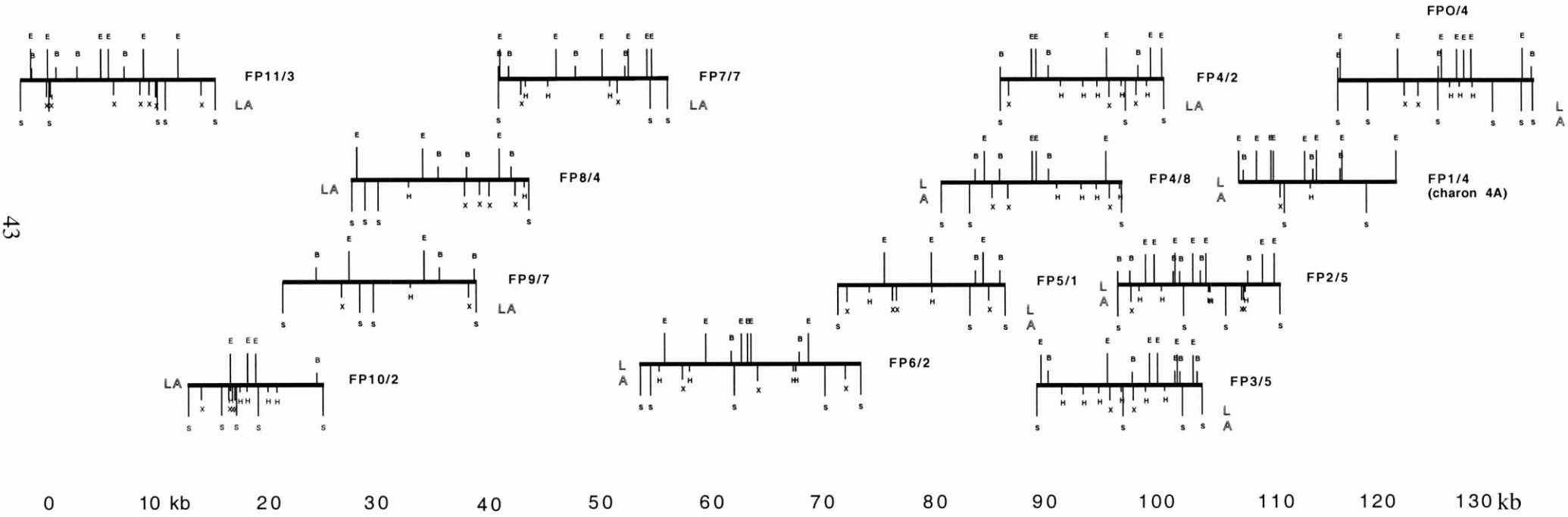


Fig. 2.1 Restriction map of the phage walk (modified from a diagram by David Gubb, unpublished.) Enzymes are *Eco*RI (E); *Bam*HI (B); *Sal*I (S); *Hin*dIII (H); *Xho*I (X). LA-long arm.

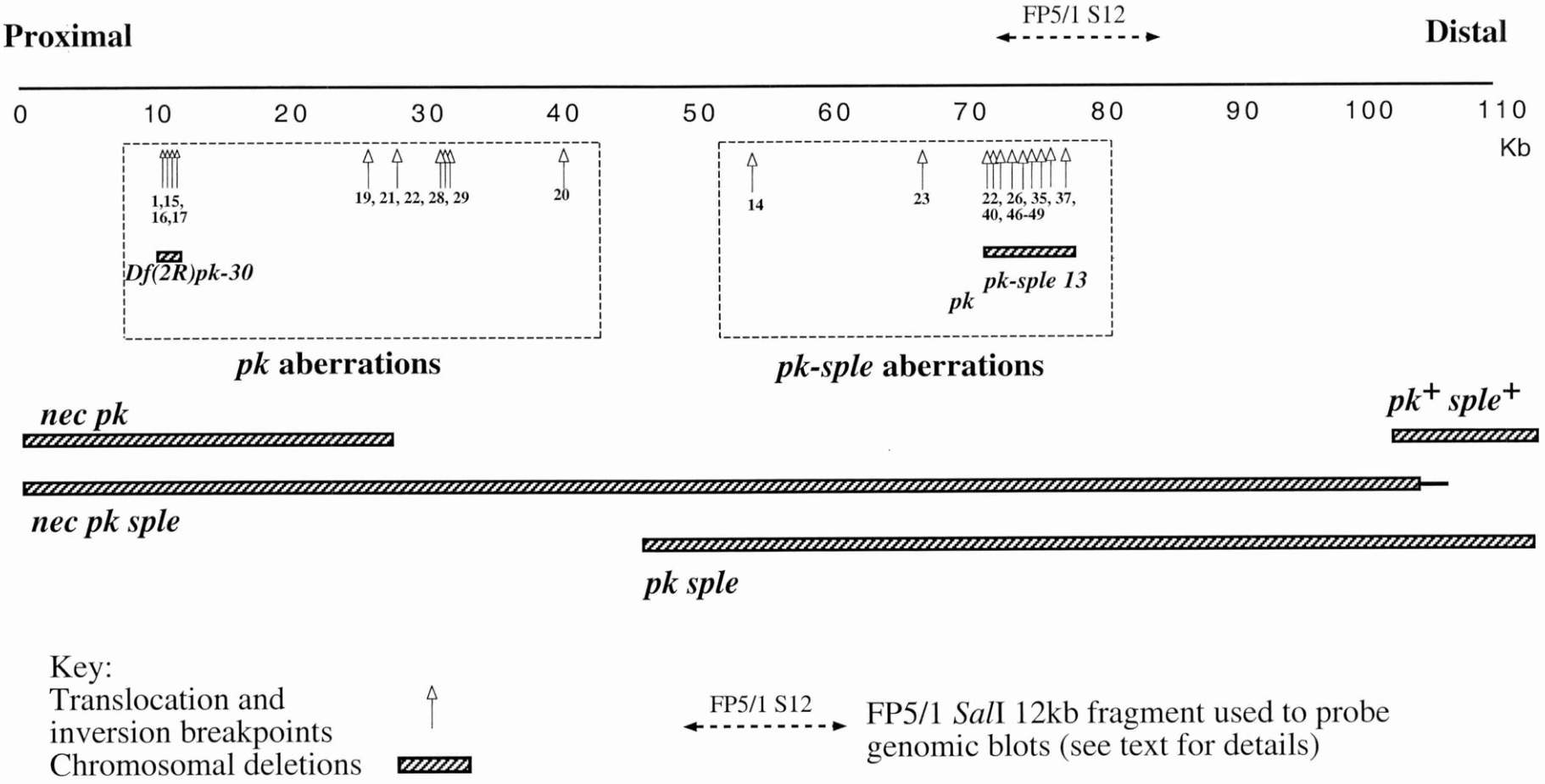


Fig. 2.2 Map of the *prickle* region, showing where the different classes of mutants map to.

observed that all the pk^{pk} mutations clustered in the proximal part of the walk while all the $pk^{pk\text{-}sple}$ mutations were located more distally. For example, *Df(2R)pk-30* is a deletion of about 1 kb near the proximal end of the walk and $pk^{pk\text{-}sple13}$ is a large deletion about 60 kb distal to *Df(2R)pk-30* . None of the pk^{sple} mutations have been located on this map, suggesting that these were the result of point mutations or microdeletions.

Gel purified fragments from the walk were used to probe cDNA libraries and Northern blots to identify candidate transcript/s for *pk*. Four transcripts were found prior to the work described in this thesis (cloned by David Gubb). One of these, ID1, was cloned from an embryonic cDNA plasmid library (Brown and Kafatos, 1988) using a probe corresponding to the distal end of the walk; I subcloned and sequenced this cDNA. However, its position on the walk was distal to all the mutants mapped so far and it was therefore excluded from further consideration. Database searches indicated that it had weak homology to adenosine kinases.

The other three transcripts were mapped to the proximal end of the walk, in the region of the *Df(2R)pk-30* deletion. These were initially considered as candidates for *pk* on this basis. Subsequent analysis showed that they corresponded to the *necrotic* (*nec*) gene which maps to chromosomal region 43A1-2 (see Chapter 5 for characterisation of these transcripts).

2.2 Homology to *Drosophila virilis*

At the start of this study, the *pk* transcript had not been located using the usual methods. This was due to the fact that a large genomic region was being screened and, as was subsequently determined, the transcripts were relatively rare and the structure of the gene was complex (see below). An alternative approach was needed. Previous studies have shown that *Drosophila virilis* is sufficiently diverged from *D. melanogaster* that sequence conservation between the two species should occur essentially in functionally important regions (for example, Kassis *et al*, 1986; Tautz *et al*, 1986; Trier *et al*, 1989).

Sequence homology between these two species may therefore indicate the presence of a gene. I decided to identify candidate transcripts for *pk* by isolating regions of homology between *D. melanogaster* and *D. virilis* within the walk.

A Southern blot of genomic DNA from *D. melanogaster* (Canton S) and *D. virilis* cut with *Eco*RI, *Sal*I or *Hin*dIII was prepared. The filter was first probed with a 12 kb *Sal*I fragment from the phage FP5/1, chosen because many of the $pk^{pk\text{-}sple}$ mutations mapped within it (Figure 2.2). The probe hybridised to the expected bands in the *D. melanogaster* tracks and bands were also seen in the *D. virilis* tracks (Figure 2.3). The signal was present in both species after the filter was washed at a relatively high stringency (0.2xSSC, 65C), strongly suggesting the presence of coding sequence in this fragment.

The same probe was hybridised to a developmental Northern blot (kind gift of Steve Russell, Dept. of Genetics, Cambridge). Two bands of 4 and 5 kb respectively were seen in the two day pupal mRNA track and more faintly in the embryonic mRNA track (Figure 2.4). These were considered strong candidates for the *pk* and *sple* transcripts, given that some of the mutants mapped within the same fragment. I therefore attempted to isolate cDNA clones from libraries with this probe.

2.3 Screening of libraries

2.3.1 Pupal library screen

As the transcripts were more abundant in pupae, I first screened pupal cDNA libraries (Poole *et al*, 1985). $2x10^5$ clones of each of the pupal stages (P, early pupae and Q, late pupae) were screened. The probe used was the FP5/1 *Sal*I 12 kb fragment (section 7.8). One positive was found in the Q library and passed through two rounds of rescreening with the same probe.

The phage DNA was cut with *Eco*RI, which removed the phage arms from the cDNA insert. The cDNA, called Q2-2, was found to be approximately 1 kb in length. A random prime probe was made from this fragment and hybridised to a Southern blot of

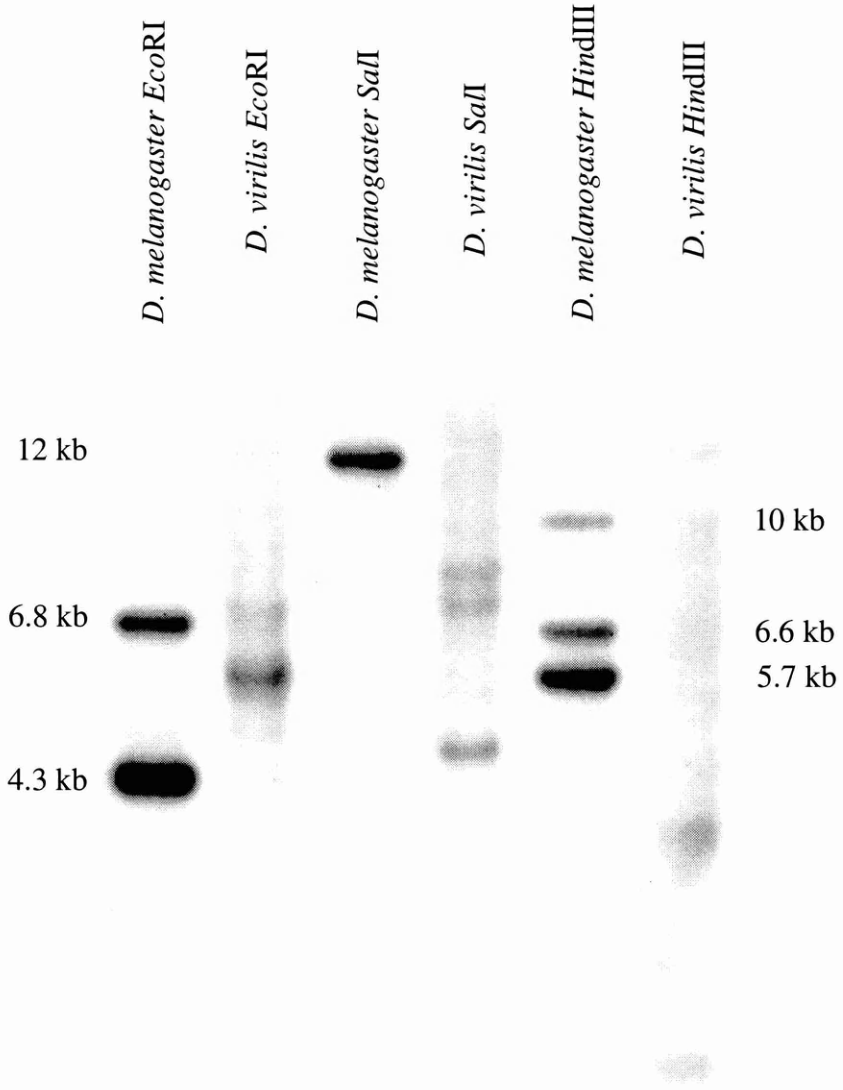


Fig.2.3 Genomic Southern blot probed with FP5/1 *Sal*I 12 kb fragment. Filter washed at high stringency (0.2xSSC, 65 C) and exposed to film overnight. The 10 kb band in the *D. melanogaster Hin*dIII track is due to incomplete digestion of the DNA.

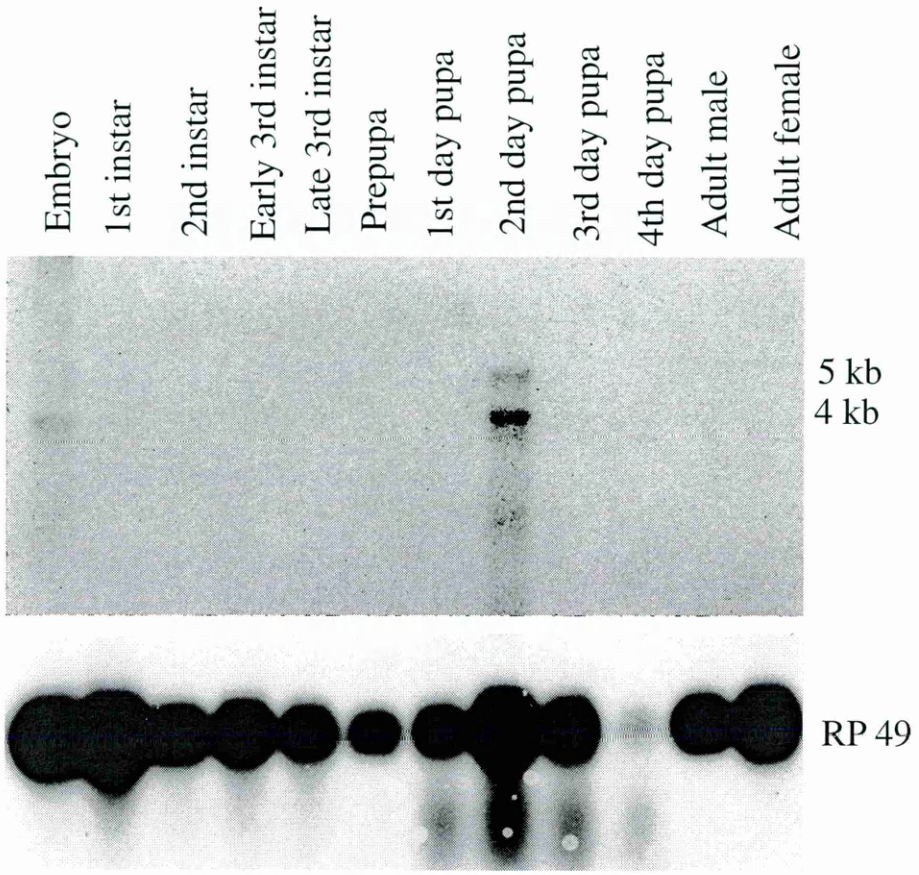


Fig. 2.4 Developmental Northern blot probed with FP5/1 *Sal*I 12 kb fragment. The filter was washed at high stringency (0.1xSSC, 65C) and exposed to film for 6 days. The same blot probed with RP 49 is shown as a loading control.

*Eco*RI/*Sal*I double digests of every phage isolate comprising the chromosome walk. A strong band of about 4.3 kb was seen in the FP5/1 track. The same developmental Northern as before (section 2.2) was also probed with Q2-2. It hybridised to the same two fragments as the FP5/1 *Sal*I 12 kb fragment (data not shown).

The Q2-2 1 kb *Eco*RI fragment was subcloned into pBluescript SK+ (Stratagene Ltd). The subclone was sequenced (section 7.16) using the Forward and Reverse pBluescript primers and oligonucleotides designed for the purpose (Appendix A).

```
GAATTCCGGACACACTGCCCCGATCACGCAGCTACTCCGGTAACGGAGCGGGAACGAGTG   60
GCGGCGGCGAGAGGGAGCGGGATCGGGACAAGGACAAGGAGGGAGGAGGTCGCCATGGTC  120
ACGGTCACTCCTCGCGAAGGCGAAGACGTCGCAAGTCCTCCTCCAGCTCGTCGCACCATC  180
GCAGCGGcAGCGGTCATCGCTCCCACTCGACCACTCGAGCGGACACCTATGCACCTGCGC  240
AACCCTATCCTCGTCCTACCAGGGTCCAcCTTCGGTTCTGCAGGCTGcCAAcCTGGTCCA  300
CGAGTCACCCAGcCGACAGCAGAGGGAACGGGAGCGCGAGCGGGAACGCGAGGAGTCCGA  360
gGGAGTcttgacAGTCTGACGTGTGCTCCACCTGCTCCTCGAGCTCCTCCAGCTCGGAGG  420
ACTACATGATGATGTACCAGCTGCCGCAGAGGCGCCACTACGGCGGCGTGCGCGTGAGCT  480
ATGTGCCCAACGATGCACTGGCCTACGATCGGAAGAGGAAGCCCAGTGAaCTGGGCGGgG  540
ACAAGGACAAGAaCTGCATCATCTCGTGATGCCGtCCGCCGGAGGGGTCCAATCAGAGXC  600
GAGGTTCGAGCACAGGTCCAGATGCTGATGCAGATCCAGATCCAGATCTTATTATGTGGA  660
GCACGGGCACTCCACTCGTGGCTACGATGTCGCCAGACGCTCGCGGGGATCACATTCCTG  720
GGAAGGGAGGAACTGCTCTCCACAAAGCAaaTTAAGTTCACTTGACATtAGGCGCCTAGT  780
TAgGGTACACTCCCGAACATACATATCCAACTCCGGAGGAGGAGACCCTACAATAAGAGC  840
CAACACTGTTCCAAATACACATCCCTAGTTAAGCTAATTGGACTATATACATATATTTTT  900
ATTATTACCGATAGCATAGCGAAGATACATTGCAGATAACCCTATTTATACACTTTATTT  960
ATATTTATTGCATTCTAGGCTTACCAATTACACTAAGTATATGCCTAGACAAATAAGTAG 1020
AAATGCTGACGATTGGGCTAATGAaACTCTACAATAACTTAGTTTAAGCGaAATCACTGT 1080
ATGACATTCACGGGAATAAAGTTCCAAAGGAGTAAAAAAaAAAAAAAaaAGGAATTCCTg 1140
CAg                                                            1143
```

Fig. 2. 5 Sequence of cDNA Q2-2. Lower case letters represent ambiguities in the sequencing; X represents any base that could not be determined.

The Northern and sequence data (Figure 2.5) showed that Q2-2 was a 1.14 kb partial cDNA fragment and was a candidate for *pk*.

2.3.2 Embryonic library screen

The candidate transcripts were also expressed in embryos, so $2x10^5$ clones of an embryonic cDNA plasmid library (kind gift of N. Brown, Wellcome/CRC; Brown and Kafatos, 1988) was screened using the Q2-2 cDNA as the probe. Several potential

positives were found and ten of these were passed through two rounds of rescreening with the same probe. Five were subsequently confirmed as positives.

Plasmid DNA was prepared from two colonies of each of the five positives (section 7.11). *Eco*RI digests of the embryonic cDNA isolates (Fig. 2.6a) were Southern blotted and probed with the Q2-2 cDNA (Fig 2.6b). Three of these five (Pk1a and b; Pk4b) hybridised to Q2-2.

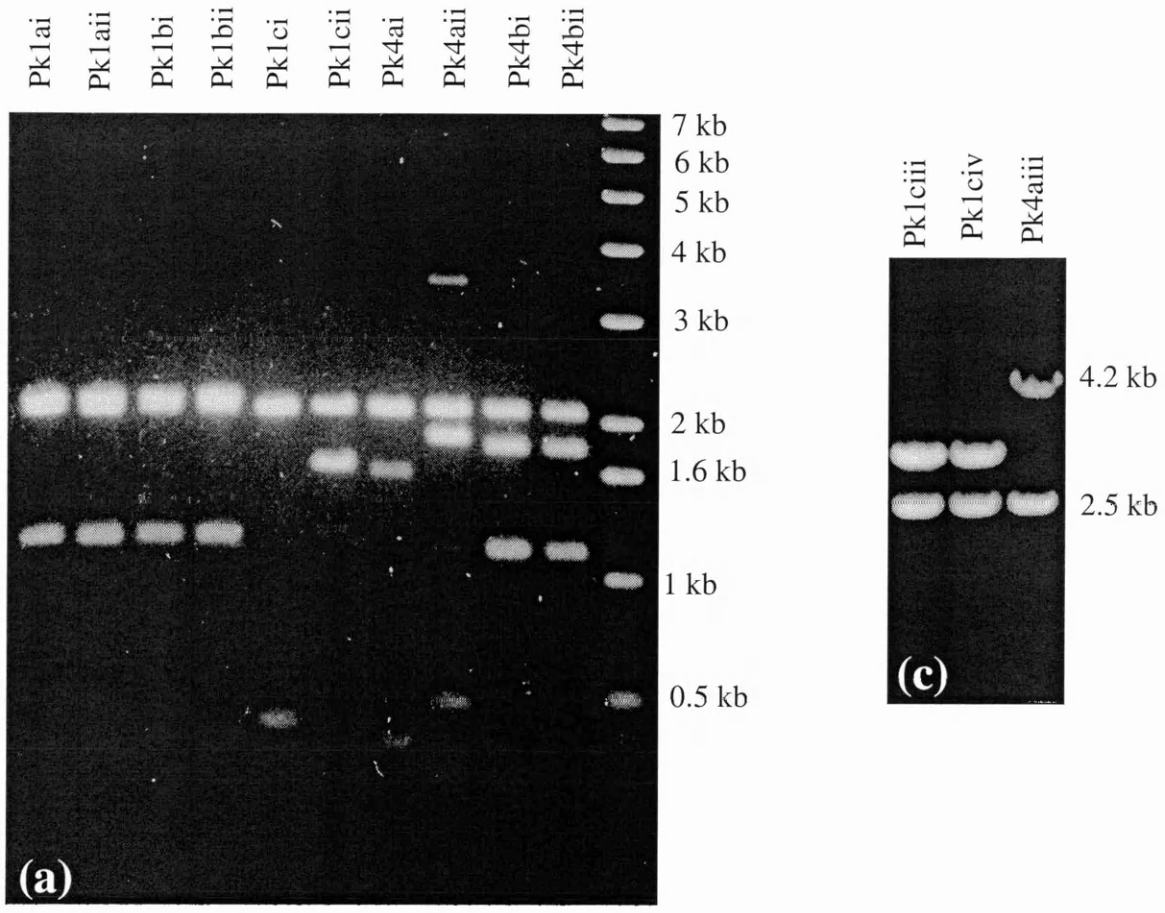
Two more colonies were isolated from Pk1c and Pk4a and passed through another round of rescreening with the Q2-2 probe. Plasmid DNA was prepared from the only Pk4aiii positive and two of several Pk1ciii positives and cut with the enzymes *Not*I and *Hin*dIII, which excised the insert from the vector (Fig.2.6c). Pk1c was identical to Pk1a and b, with an insert of about 3 kb, while Pk4aiii had an insert of just over 4 kb and hybridised to Q2-2 (data not shown). The fact that Pk4aiii hybridises to Q2-2 while Pk4ai and Pk4aii did not, along with the size variation of the inserts (Fig. 2.6a and c), suggests that the original Pk4a stock was not homogeneous. Pk4ai and Pk4aii were probably contaminants, or possibly did not extend as far as Pk4aiii at the 3' end. Further studies were concentrated on the largest cDNA, Pk4aiii.

2.4 Characterisation of cDNA Pk4aiii

Neither *Not*I nor *Hin*dIII cut within the insert, so it was subcloned into pBluescript SK+ using these two enzymes. The insert in the pNB40 vector (Brown and Kafatos, 1988) is oriented with the *Hin*dIII site at the 5' end and the *Not*I site at the 3' end (Appendix D).

2.4.1 Mapping

The gel purified Pk4aiii insert was random primed and hybridised to a Southern blot of *Eco*RI/*Sal*I double digests of every step of the walk (Figure 2.7). It hybridised to bands in FP5/1, FP6/2, FP10/2 and FP11/3 (Fig 2.8). The transcript spans more than



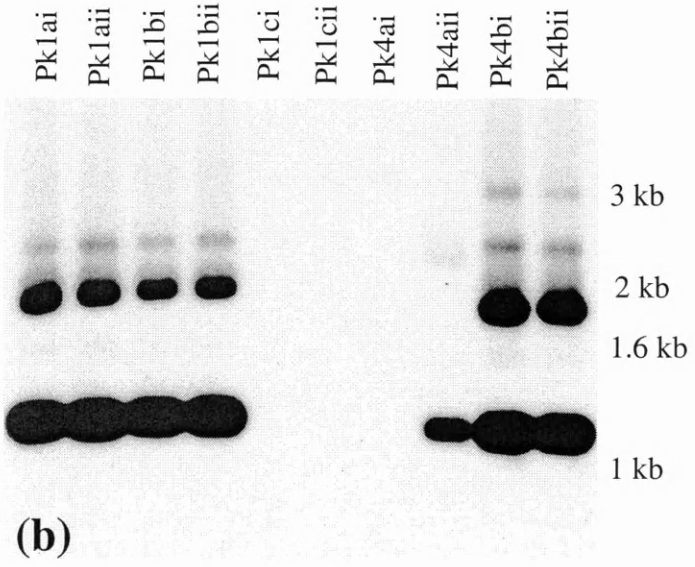


Fig. 2.6 Characterisation of the Pk4aiii cDNA clones. **(a)** *Eco*RI digests of the clones isolated from the embryonic library screen (section 2.3.2). **(b)** Southern blot analysis of the fragments shown in (a) probed with the Q2-2 cDNA. Filter was washed at high stringency (0.1xSSC, 65C) and exposed to film for 1 hour. **(c)** Additional positive clones, including Pk4aiii, digested with *Not*I/*Hin*dIII.

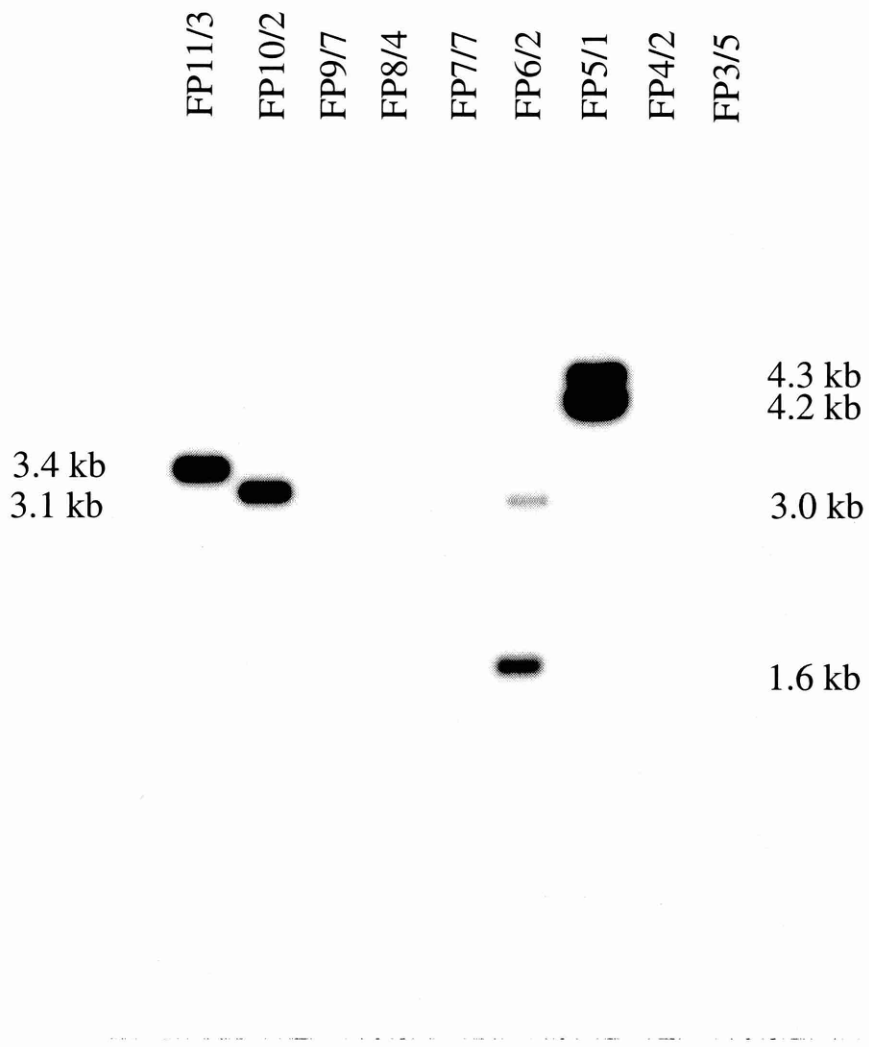


Fig. 2.7 Southern blot of *Eco*RI/*Sal*I digests of the phage comprising the walk probed with Pk4aiii. Filter washed at high stringency (0.1xSSC, 65C) and exposed to film overnight.

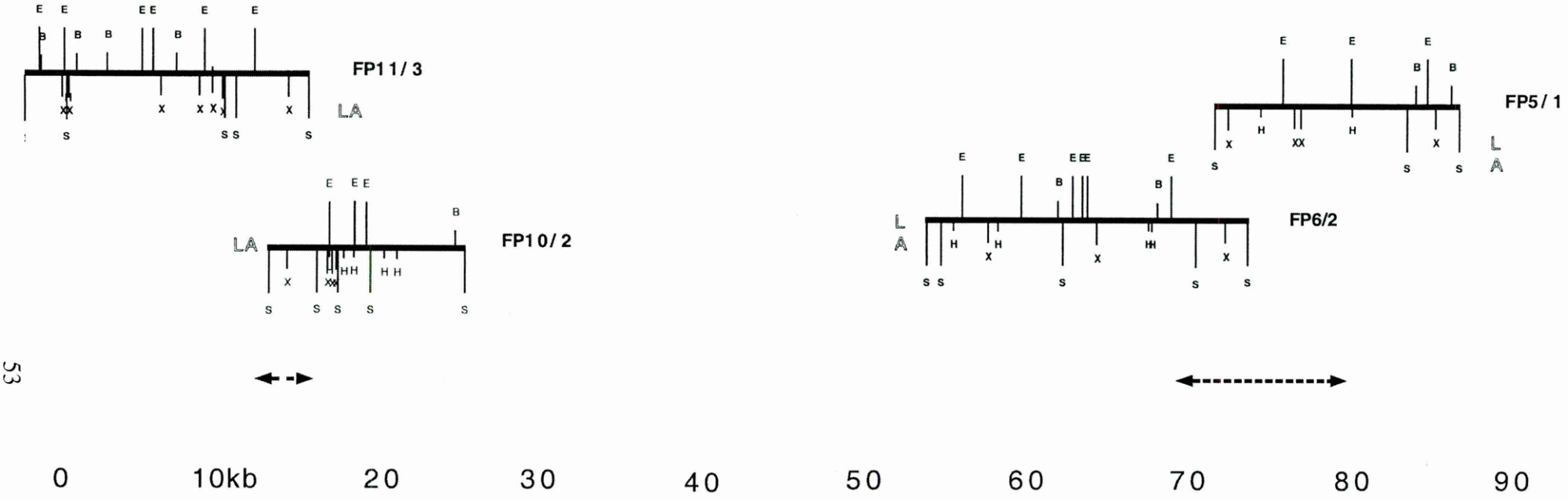


Fig.2.8 Mapping of putative *pk*cDNA on the phage walk. ◄--► indicates the regions the transcript maps to.

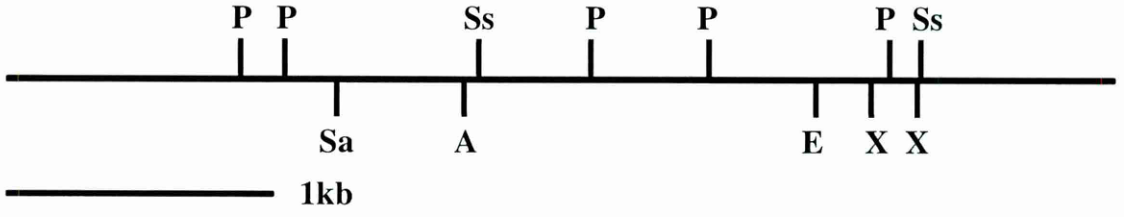


Fig. 2.9 Restriction enzyme map of pk4aiii cDNA. Enzymes: *Pst*I (P); *Sal*I (Sa); *Apa*I (A); *Sst*I (Ss); *Eco*RI (E); *Xho*I (X).

60 kb of the region the *pk* mutations map to. A restriction map of Pk4aiii was constructed (Fig. 2.9).

The blot was also probed with Pk1ai. It hybridised to the same bands in FP5/1 and FP6/2, but not to FP10/2 or FP11/3, implying that it is a shorter cDNA of the same class as Pk4aiii.

2.4.2 Unique and common exons

The *Eco*RI/*Sal*I 1.2 kb fragment of Pk4aiii was isolated. This fragment included all of the portion that hybridised to the proximal end of the walk as well as a short stretch of the portion that hybridised to the distal region. When this fragment was used to probe the developmental Northern blot (as in section 2.2) it hybridised strongly to the 4 kb transcript with only a very faint signal from the 5 kb transcript. Given that Q2-2 and Pk4aiii hybridised to the 4 kb and 5 kb transcripts, I concluded that the transcripts shared common exons at the distal end of the walk but had distinct exons more proximally. The faint larger band in the Northern blot probed with the *Eco*RI/*Sal*I 1.2 kb fragment is probably due to the inclusion of a small part of the 3' exons in the probe. The smaller transcript on the Northern blot was estimated at 4 kb; the Pk4aiii cDNA was about 4.2 kb, indicating that the cDNA was probably close to full length.

Of the mutants that have been mapped, only the pk^{pk} mutants mapped to the proximal region represented in the Pk4aiii cDNA. It was thought likely that Pk4aiii represented the *pk* cDNA.

Northern blots of two-day pupal total RNA from several *pk* mutants were probed with the 3' end of Pk4aiii (David Gubb, personal communication). The 4kb transcript was lost from the pk^{pk} and $pk^{pk\text{-}sple}$ mutants, but retained in the pk^{sple} mutants. This confirmed the hypothesis that the 4kb transcript represented the *pk* cDNA (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation).

2.5 Sequence

The putative *pk* cDNA was cut with restriction enzymes into small fragments which were then subcloned into pBluescript SK+ (Stratagene): cDNA cut with *Sau*3A was subcloned into the *Bam*HI site of pBluescript SK+; cDNA cut with *Taq*I was subcloned into the *Cla*I site of pBluescript SK+. PCR was then carried out on these clones (Programme 1; section 7.15) with two primers flanking the insert (4457 and 4458; see Appendix A) to determine the sizes of the fragments subcloned. Inserts were identified by size alone; there were 11 different *Sau*3A fragments and 9 *Taq*I fragments ranging from 0.2 to 1.1 kb.

Plasmid DNA was prepared for a representative of each fragment and sequenced using the Forward and Reverse pBluescript sequencing primers (see Appendix A). Further subclones were generated using six-cutter enzymes (Appendix B) and sequenced with the same primers. Oligonucleotide primers were designed to sequence the remaining gaps (Appendix A). The data was assembled using BestFit and Assemble (GCG; section 7.16). Each region was sequenced at least twice on each strand. If there was still no clear consensus on all bases, the sequence was repeated.

The final sequence revealed the transcript to be 4163 base pairs in length (see Figure 2.10). Comparison of the sequence to that of Q2-2 confirmed that Q2-2 is the same as the 3' end of Pk4aiii (base 3044 onwards).

2.6 Protein sequence

The longest predicted open reading frame (ORF) is 2610 bases long (Figure 2.11), so the first 798 bases and the last 754 bases are untranslated. Conceptual translation of the DNA sequence was carried out in the appropriate frame (Figure 2.10). The first start ATG (bases 799-801) is embedded within a good *Drosophila* start site consensus sequence (Cavener and Ray, 1991). The putative *prickle* protein is 870 amino acids long.

Fig. 2.10 cDNA sequence of *pk* and conceptual translation of the longest predicted ORF. *Restriction enzyme sites* are in italics; the **novel domain** is in bold type and the "LIM" motifs are underlined. The bases in lower case are those removed by the *Df(2R)pk-30* deletion. The unique *pk* exon is shadowed; the rest of the sequence constitutes the "common" exons.

```
CCCCCCCCCAGTCTGTTGTGCGCCTCgcacgaacatcgaactcgcggacgtgcgcgtgcg      60
cgaacggaacgctcgaaggcaaacgaaacggaaaggaacgttagggaaggcacggaaacg     120
gaacggatcagcgacggcccccgcgtttttggtttttgcgcttttgggcttatcaagtg      180
ccgggaatcggtgaacgggagtccagaaatcgggcagctgtcgcgtgcagtggccaggaa     240
aaggagtttgcggaaaataatgaaatgcctgccgtcggtggcagcagtgtggctgcgtgt     300
gtttcttgagcggccgggaaatgtgcatcaattattttaaatggccagcagccacccccct    360
cctcatccaaccccctgcgaagtgtcccccacccccaagcacatgcataaatacaattaggc    420
cgctagatcgggcctgagccatttggtggagccaaagttaattcggtgcgcgaaccagtt     480
tatcgacaggcactccgcgaatttttcacgaacgcgccaagaaatttataataaatacg      540
cctaagctcaaattgttgtcagctggctcggcttttgttgatttatgtgcgcgattaaag     600
gaaacaacgagcgcagtcggccgagaaaccgcagaataagataataaatataagtaatgc     660
cgaaaggcatcacccccgcccccagtgtcaccagtgaaatacctgtgagtgccaatcaag     720
cgacaaagaaaccgaaaagacatcagcccgtagaacgccagccataaaaacttcggatac     780
atcctccgctccgcaaacatggatacccccaaatcaaatgcctgttgagctagaaaggCCC    840
                    M  D  T  P  N  Q  M  P  V  E  L  E  R  P
                                    PstI
ATCAGCCGAACTCCGCTAACGCAGATCTCCTATCTGCAGAAGATCCCCACGCTTCCGCGC     900
I  S  R  T  P  L  T  Q  I  S  Y  L  Q  K  I  P  T  L  P  R
CACTTCTCGCCCAGTGGGCAGGGCCTGGCCACGCCCCCAGCCCTCGGCAGCGGCGGCATG     960
H  F  S  P  S  G  Q  G  L  A  T  P  P  A  L  G  S  G  G  M
GGTCTGCCCAGCAGCTCCTCCGCCAGTGCCCTCTACGCGGCCCAGGCAGCAGCCGGAATC    1020
G  L  P  S  S  S  S  A  S  A  L  Y  A  A  Q  A  A  A  G  I
                          PstI
CTGCCCACCTCCCCGCTGCCACTGCAGCGCCACCAGCAGTACTTGCCGCCACATCACCAA    1080
L  P  T  S  P  L  P  L  Q  R  H  Q  Q  Y  L  P  P  H  H  Q
CAGCACCCGGGAGCTGGAATGGGACCAGGACCAGGATCGGGGGCAGCAGCAGGACCGCCC    1140
Q  H  P  G  A  G  M  G  P  G  P  G  S  G  A  A  A  G  P  P
CTTGGCCCCCAGTACTCGCCAGGCTGCTCCGCCAATCCCAAGTACTCGAATGCCCAGCTC    1200
L  G  P  Q  Y  S  P  G  C  S  A  N  P  K  Y  S  N  A  Q  L
                                                SalI
CCGCCACCGCCGCACCACCACCACCAGCTATCGCCGGCTTTGTCGACGCCCTCGCCACCC    1260
P  P  P  P  H  H  H  H  Q  L  S  P  A  L  S  T  P  S  P  P
TCGTTGCTCCACCATCCCGCCGGAGGCACTTCCTCTGCCTCGGCCCACGCCCCCTTCCTG    1320
S  L  L  H  H  P  A  G  G  T  S  S  A  S  A  H  A  P  F  L
GGCGGACCGCACATGGACATGCAGCGGCAGTCGCACTCGGACGACGACAGTGGCTGTGCT    1380
G  G  P  H  M  D  M  Q  R  Q  S  H  S  D  D  D  S  G  C  A
CTGGAGGAGTACACCTGGGTGCCACCCGGACTGCGGCCTGACCAGGTTCGCTTGTACTTC    1440
L  E  E  Y  T  W  V  P  P  G  L  R  P  D  Q  V  R  L  Y  F
TCGCAGATACCCGACGACAAAGTGCCATACGTCAACAGTCCGGGGGAGCAGTATCGTGTG    1500
S  Q  I  P  D  D  K  V  P  Y  V  N  S  P  G  E  Q  Y  R  V
CGACAATTGCTCCATCAACTTCCGCCGCATGATAACGAGGTTCGTTACTGTCACTCACTG    1560
R  Q  L  L  H  Q  L  P  P  H  D  N  E  V  R  Y  C  H  S  L
ACGGATGAGGAGCGCAAGGAGCTGCGGTTATTCTCCACGCAGCGCAAACGCGATGCCCTG    1620
T  D  E  E  R  K  E  L  R  L  F  S  T  Q  R  K  R  D  A  L
GGCCGCGGCAACGTGCGGCAGCTGATGAGCGCCCGACCCTGCGACGGCTGCGACGACCTA    1680
G  R  G  N  V  R  Q  L  M  S  A  R  P  C  D  G  C  D  D  L
```

ApaI

ATCTCCACGGGCGACATCGCCGTGTTCGCCACGCGACT*GGGCCCC*AATGCCAGCTGGCAT 1740

I S T G D I A V F A T R L G P N A S W H

SstI

CCGGCGTGCTTCGCCTGCTCCGTCTGCCGC*GAGCTC*CTCGTGGACCTTATCTACTTCCAC 1800

P A C F A C S V C R E L L V D L I Y F H

CGGGATGGGCGCATGTACTGCGGACGCCACCACGCCGAGACCCTAAAGCCCCGCTGCTCG 1860

R D G R M Y C G R H H A E T L K P R C S

GCCTGCGATGAGATTATCCTGGCAGACGAGTGCACGGAGGCGGAGGGCAGGGCGTGGCAC 1920

A C D E I I L A D E C T E A E G R A W H

ATGAACCACTTTGGCTGCCACGAGTGCGACAAGCAGCTGGGCGGGCAGCGGTATATCATG 1980

M N H F G C H E C D K Q L G G Q R Y I M

CGGGAGGGCAAACCCTACTGTCTGCACTGCTTCGACGCGATGTTTGGCGAGTACTGCGAC 2040

R E G K P Y C L H C F D A M F G E Y C D

TACTGCGGGGAGGGCATCGGGGTGGACCAGGGCCAGATGAGCCACGACGGGCAACACTGG 2100

Y C G E G I G V D Q G Q M S H D G Q H W

CACGCCACCGACGAGTGCTTCTCCTGCAACACTTGCCGCTGCTCGCTGCTGGGACGCGCC 2160

H A T D E C F S C N T C R C S L L G R A

PstI

TTCCTGCCGCGACGAGGGGGCATCTACTGTTCCATCG*cCTGCAG*CAAGGGGGAGCCGCCC 2220

F L P R R G G I Y C S I A C S K G E P P

ACGCCGTCGGACAGCTCCGGCACCGGCATGTACACCACGCCCACTCCGCCCACACAGGGG 2280

T P S D S S G T G M Y T T P T P P T Q G

GTGCGGCCCCATCCGCAAGCGCCTCTGCCGGCGCGCATTCCCAGcAGCCACGCCTCCAGC 2340

V R P H P Q A P L P A R I P S S H A S S

TCGCCGCCTATGTCGCCGCAACAGCAGCAGCAGCACCAGGCTACCTTCAACCAAGCGATG 2400

S P P M S P Q Q Q Q Q Q H Q A T F N Q A M

TACCAGATgCAAAGCCAGCAGATGGAGGCGGCGGGCGGACTGGTGGACCAGAGCAAGAGC 2460

Y Q M Q S Q Q M E A A G G L V D Q S K S

TACGCCGCGTCGGACTCGGATGCGGGTGTGGTAAAGGACCTAGAACACGGCGGACATATG 2520

Y A A S D S D A G V V K D L E H G G H M

GGCGGAGGAGATCTGACGGACTTTTCGGGTGGACGGGCCTCCAGCACCTCGCAGAACCTA 2580

G G G D L T D F S G G R A S S T S Q N L

PstI

TCGCCTCTGAACTCACCCGGCGACTTCCAGCCCCATTTTCTGCCCAAGCCCATGGAA*CTG* 2640

S P L N S P G D F Q P H F L P K P M E L

*CAG*AGACAGCTCCTGGAGAACCCCCACACCGCCAGCATGCCGGAACTGGCGGGTAAACTG 2700

Q R Q L L E N P H T A S M P E L A G K L

GTGGCGCCTCCTGCACACATGCAGCACCTGTCCCAGCTGCACGCCGTGTCCTCGCACCAG 2760

V A P P A H M Q H L S Q L H A V S S H Q

TTCCAGCAGCACGAGTACGCGGACATCGTGCATCCGCCGCCTCCTCCGCCTGGTGAGATT 2820

F Q Q H E Y A D I V H P P P P P P G E I

CCCGAGCTGCCCACTCCCAACTTGAGCGTGGCTTCCACCGCCTTGCCGCCAGAGCTTATG 2880

P E L P T P N L S V A S T A L P P E L M

GGCTCACCCACCCACTCGGCGGGCGACAGGTCGCTGAACACGCCCATGTCCACACAGTCG 2940

G S P T H S A G D R S L N T P M S T Q S

GCCTCCCACGCGCCCCCACATCCGGTCTCCATACTAAGCGGCGCCTCCTCCTCCTCGCCG 3000

A S H A P P H P V S I L S G A S S S S P

EcoRI

ATGAGCGGGGAGCCGGCCAAGAAGAAGGGCGTGCGCTTCGAGG*GAATTC*CGGACACACTG 3060

M S G E P A K K K G V R F E G I P D T L

CCCGATCACGCAGCTACTCCGGTAACGGAGCGGGAACGAGTAGTGGCGGCGAGAGGGAGC 3120

P D H A A T P V T E R E R V V A A R G S

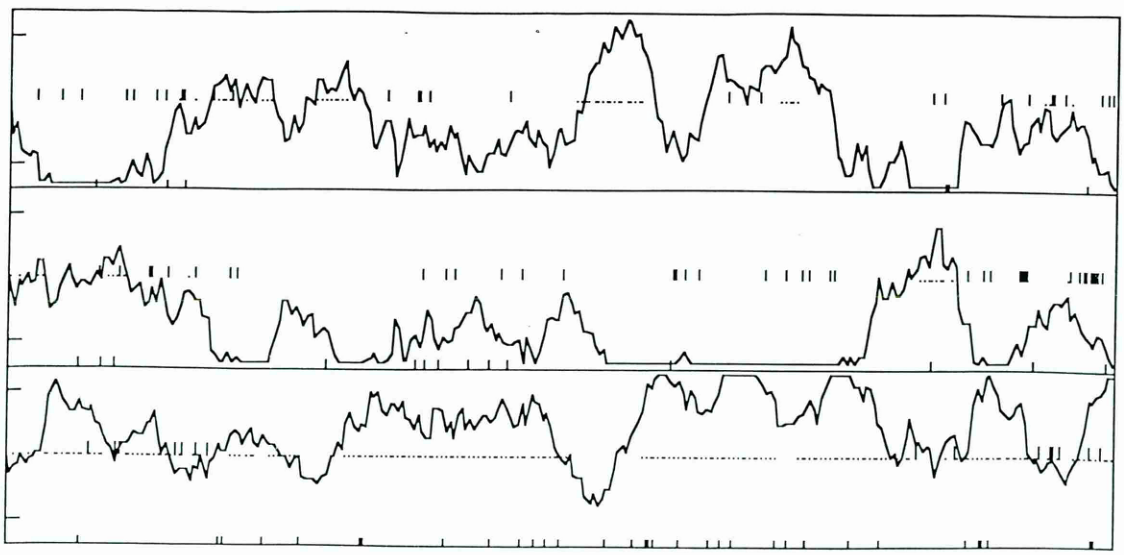
GGGATCGGGACAAGGACAAGGAGGGAGGAGGTCGCCATGGTCACGGTCACTCCTCGCGAA 3180

G I G T R T R R E E V A M V T V T P R E

```
GGCGAAGACGTCGCAAGTCCTCCTCCAGCTCGTCGCACCATCGCAGCGGCAGCGGTCATC 3240
G  E  D  V  A  S  P  P  P  A  R  R  T  I  A  A  A  A  V  I
                  XhoI
GCTCCCACTCGACCACTCGAGCGGACACCTATGCACCTGCGCAACCCTATCCTCGTCCTA 3300
A  P  T  R  P  L  E  R  T  P  M  H  L  R  N  P  I  L  V  L
                        PstI
CCAGGGTCCACCTTCGGTTCTGCAGGCTGCCAACCTGGTCCACGAGTCACCCAGCCGACA 3360
P  G  S  T  F  G  S  A  G  C  Q  P  G  P  R  V  T  Q  P  T
GCAGAGGGAACGGGAGCGCGAGCGGGAACGCAAGGAGTCCGAGGAGTCTGACGTGTGCTC 3420
A  E  G  T  G  A  R  A  G  T  Q  G  V  R  G  V  *

          XhoI SstI
CACCTGCTCCTCGAGCTCCTCCAGCTCGGAGGACTACATGATGATGTACCAGCTGCCGCA 3480
GAGGCGCCACTACGGCGGCGTGCGCGTGAGCTATGTGCCCAACGATGCACTGGCCTACGA 3540
TCGGAAGAGGAAGCCCAGTGAACTGGGGCGGCGACAAGGACAAGAACTGCATCATCTCGT 3600
GATGCCGCCCGCCGGAGGGGTTCCAATCAGAGCCGAGGTTCGAGCACAGGTCCAGATGCT 3660
GATGCAGATCCAGATCCAGATTTTATTAATGTGGAGCCACGGGCAACTCCATCTCGTGGC 3720
TAACGAAGTGGCCCAGAACGCTCGCGGGGATCATCAATTCCTGGGAAGGGAGGCAACTGC 3780
CTCTGCACAAAGCAAAATAAAGTTCAACTTGACATTAAGGCGCCTAGTTAGGTACACTCC 3840
CAAACATACATATGCAACTCCGGAGGAGGAGACCCTACAATAAGAGCCAACACTGTTCCA 3900
AATACACATCCCTAGTTAAGCTAATTGAACTATATACATATATTTTTATTATTACCGATA 3960
GCATAGCGAAGATACATTGCAGATAACCCTATTTATACACTTTATTTATATTTATTGCAT 4020
TCTAGGCTTACCAATTACACTAAGTATATGCCTAGACAAATAAGTAGAAATGCTGACGAT 4080
TGGGCTAATGAAACTCTACAATAACTTAGTTTAAGCGAAATCACTGTATGACATTCACGA 4140
GAATAAAGTTCCAAAGGAGTAAA                                       4163
```

Fig. 2.11 Open reading frame prediction for Pk4aiii cDNA (forward frames only). Created with NIP (nucleotide interpretation programme) in the Staden package. The plot indicates the probability that the sequence is coding in each frame (based on codon usage); the 50% probability is marked by the dotted line. The short vertical lines that bisect the 50% probability line are stop codons; the shorter vertical lines along the base of each plot are potential start sites. In this case, the bottom frame is the most likely coding frame.



2.7 Genomic sequence

Genomic sequence of the region was obtained in order to identify the exact number of introns and locate their boundaries. Phage restriction fragments were subcloned into pBluescript and sequenced from the Forward and Reverse sequencing primers (pBluescript). The sequence was aligned using LineUp and Assemble (GCG). Oligonucleotide primers were designed to fill in gaps in the sequence where required. Only the ends of the larger introns were sequenced. A longer genomic sequence was generated for the 5' region during the analysis of the serpins (see Chapter 5).

Comparison of the cDNA and genomic sequences, using BestFit, revealed that the putative *pk* cDNA consisted of seven exons. Only the first exon was located at the proximal end of the walk, yielding a first intron of around 60 kb in length. The sizes of all the introns and exons are shown in Figure 2.12. The exact sizes of the larger introns (apart from the first one) were determined by comparing the cDNA sequence to genomic sequence from the Berkeley Genome Project (region of 5' exon from clone P1 DS07472, accession number AC005640; rest from clone P1 DS08462, accession number AC005198) when this became available. Introns usually have a 5' GT and a 3' AG (Breathnach et al, 1978; Mount, 1981). This is the case for all the introns in this cDNA except the sixth one. This small intron has CG at its 5' end and AG at its 3' end.

The above genomic sequence includes the cDNA sequence up to the polyA tail at the 3' end (hence the cDNAs are complete at the 3' end); this is consistent with the position of the distal deletions that complement *pk* (Fig. 2.2).

2.8 Sequence analysis

Database similarity searches were carried out on the DNA and protein sequences using the programmes described in section 7.16.3. Results of the searches were similar, identifying homology to a large number of proteins containing one or more LIM domains

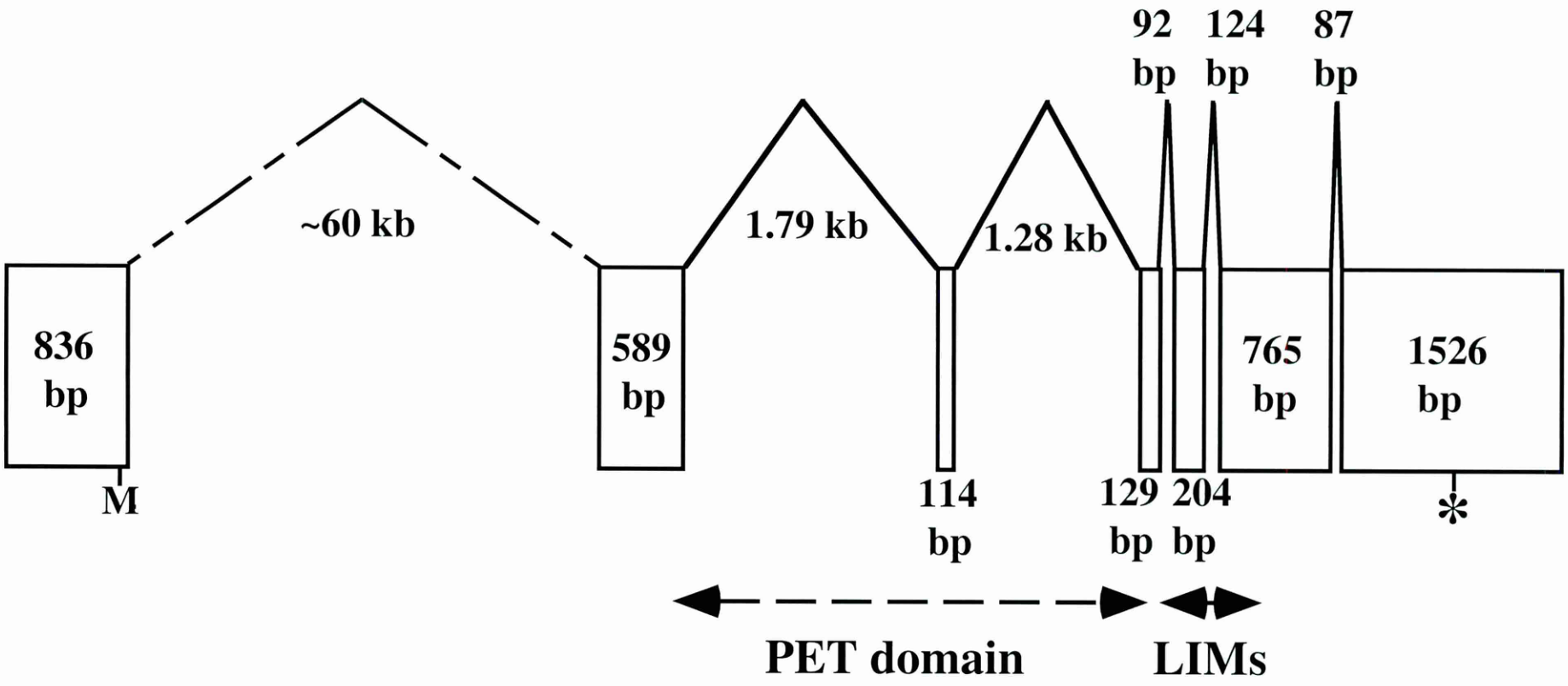


Fig. 2.12 Structure of the pk^{pk} transcript. **M** indicates the start of the coding sequence; *indicates the stop codon. The regions spanned by the PET and LIM domains are indicated.

Sequences producing significant alignments:	Score (bits)	E Value
gnl|PID|e1329682 (AJ011654) triple LIM domain protein [Homo sapi...	435	e-121
sp|O43900|LMO6_HUMAN LIM-ONLY PROTEIN 6 (TRIPLE LIM DOMAIN PROTE...	432	e-120
gnl|PID|e1345921 (Z81068) similar to LIM domain containing prote...	249	8e-65
gnl|PID|e1345922 (Z81068) similar to LIM domain containing prote...	249	8e-65
gi|475208 (X78989) testin [Mus musculus]	238	1e-61
sp|P47226|TES2_MOUSE TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)]...	238	1e-61
gi|1326362 (U58749) strong similarity to mouse testin (PIR:S4421...	222	1e-56
gi|1255350 (U52003) short region of weak similarity to mouse tes...	122	1e-26
gi|3360516 (AF071773) LIM-domain protein [Branchiostoma floridae]	109	7e-23
gi|2853224 (U60115) skeletal muscle LIM-protein FHL1 [Homo sapiens]	101	2e-20
pir||JC4893 skeletal muscle LIM-protein 1 - human	101	2e-20
gi|1255348 (U52003) shows some similarity with mouse testin (PIR...	101	3e-20
gi|2894672 (U41739) RBP associated molecule RAM14-1 [Mus musculus]	100	4e-20
gi|3025853 (AF055889) LIM protein FHL2 [Mus musculus]	100	4e-20
gi|1679728 (U77040) LIM protein 3 [Mus musculus]	100	4e-20
sp|Q13642|SLI1_HUMAN SKELETAL MUSCLE LIM-PROTEIN 1 (SLIM 1) (SLI...	100	6e-20
gi|2880031 (U77039) skeletal muscle LIM protein [Mus musculus]	99	8e-20
gi|2605504 (AB008571) DRAL [Rattus norvegicus]	99	8e-20
gi|1160932 (L42176) DRAL gene product [Homo sapiens]	99	1e-19
sp|Q14192|SLI3_HUMAN SKELETAL MUSCLE LIM-PROTEIN 3 (SLIM 3) (LIM...	98	2e-19
sp|P49024|PAXI_CHICK PAXILLIN >gi|1079431|pir||B55933 paxillin -...	97	5e-19
gnl|PID|d1019741 (D86863) paxillin gamma [Homo sapiens]	97	5e-19
gnl|PID|d1019740 (D86862) paxillin beta [Homo sapiens]	97	5e-19
gi|3859849 (AF063002) LIM protein SLIMMER [Homo sapiens]	97	5e-19
gi|2935617 (AC004263) cytoskeletal protein [Homo sapiens]	97	5e-19
sp|P49023|PAXI_HUMAN PAXILLIN >gi|1082680|pir||A55933 paxillin -...	97	5e-19
gi|3851650 (AF098518) four and a half LIM domains 1 protein isof...	97	5e-19
gi|3126971 (AF062075) leupaxin [Homo sapiens]	96	9e-19
gi|2914130 (U60116) skeletal muscle LIM-protein FHL3 [Homo sapiens]	95	2e-18
gi|2981496 (AF053486) LIM-protein FHL4 [Mus musculus]	95	2e-18
gnl|PID|e1343971 (Z54235) similar to LIM domains; cDNA EST EMBL:...	94	5e-18
gnl|PID|d1032549 (AB014513) KIAA0613 protein [Homo sapiens]	93	6e-18
pir||A55071 hydrogen peroxide-inducible protein hic-5 - mouse >g...	92	2e-17
gnl|PID|e1169963 (AJ002382) paxillin-like protein [Dictyostelium...	90	4e-17
gnl|PID|d1025720 (AB007836) Hic-5 [Homo sapiens]	90	4e-17
sp|Q09476|YP96_CAEEL HYPOTHETICAL 28.9 KD PROTEIN C28H8.6 IN CHR...	90	5e-17
gi|1381814 (U60118) skeletal muscle LIM-protein SLIM [Homo sapiens]	89	9e-17
gi|3108093 (AF061258) LIM protein [Homo sapiens]	89	9e-17
gi|3851178 (U48247) protein kinase C-binding protein Enigma [Rat...	89	2e-16
sp|P50464|UN97_CAEEL HOMEOBOX PROTEIN UNC-97 >gi|1086681 (U41021...	87	5e-16
sp|Q13643|SLI2_HUMAN SKELETAL MUSCLE LIM-PROTEIN 2 (SLIM 2) (FOU...	86	8e-16
pir||A55050 enigma - human >gi|561637 (L35240) enigma protein [H...	86	1e-15
gi|1381812 (U60117) skeletal muscle LIM-protein SLIM3 [Homo sapi...	85	1e-15
sp|P48059|PINC_HUMAN PINCH PROTEIN (PARTICULARY INTERESTING NEW ...	84	3e-15
gi|576623 (U15158) ESP-2 [Oryctolagus cuniculus]	73	6e-12
sp|Q15942|ZYX_HUMAN ZYXIN (ZYXIN 2) >gi|1155088|gnl|PID|e218260 ...	73	6e-12
gi|2337952 (AF005654) actin-binding double-zinc-finger protein [...	73	6e-12
gi|540207 (U14530) OvL3-1 [Onchocerca volvulus]	73	8e-12
gi|3851624 (AF097511) zyxin related protein-1 [Mus musculus]	71	3e-11
gnl|PID|e1319472 (AL031534) lim domain protein [Schizosaccharomy...	71	4e-11

Table 2.1 Top 50 matches from blast of *pk* cDNA.

(Table 2.1). LIM domains are cysteine rich sequences with a double zinc finger motif (reviewed by Sánchez-García *et al*, 1994; Dawid *et al*, 1998). In most cases, the homology was restricted to the LIM domains, although a another region with homology was identified. This is believed to be a novel domain. The LIM domains and the novel domain are discussed further in Chapter 4.

2.9 Characterisation of *Df(2R)pk-30* mutation

Flies carrying the *Df(2R)pk-30* mutation have a strong pk phenotype, similar to that of pk^{pk1} mutants. Southern hybridisation data had previously shown that *Df(2R)pk-30* was a deletion of about 1 kb and it had been mapped to the same region as the most proximal exon/s of the putative *pk* cDNA (Figure 2.2). A more detailed study of this mutant was carried out.

Genomic DNA was extracted from flies homozygous for *Df(2R)pk-30* and wild type flies (Canton S). PCRs were carried out on each of these DNAs using the oligonucleotide primers 6228 and 7854. (These primers were originally designed for sequencing of the serpins; see Chapter 5.) The aim was to identify the smallest fragment of genomic DNA that included the whole of the deleted region. The primers chosen were 5' (6228) or 3' (7854) of the region where the deletion was located. The reactions were analysed by gel electrophoresis.

A single band was seen in each track. In Canton S, this was about 2.9 kb; in *Df(2R)pk-30* it was about 1.8 kb. This confirmed that the *Df(2R)pk-30* mutation is a deletion of approximately 1 kb. Each of these PCR products was then double digested with *Cla*I/*Xho*I and subcloned into the same sites in pBluescript SK+. (These enzymes do not cut within the original fragment.)

Both the subcloned PCR products were then digested with a panel of restriction enzymes and compared by gel electrophoresis (Figure 2.13). The smallest fragment in the wild type fly that contains the deleted region is the *Spe*I/*Xho*I 1.7 kb fragment. In Figure 2.13, this is included in the 4.7 kb *Spe*I band, which also contains the vector. The

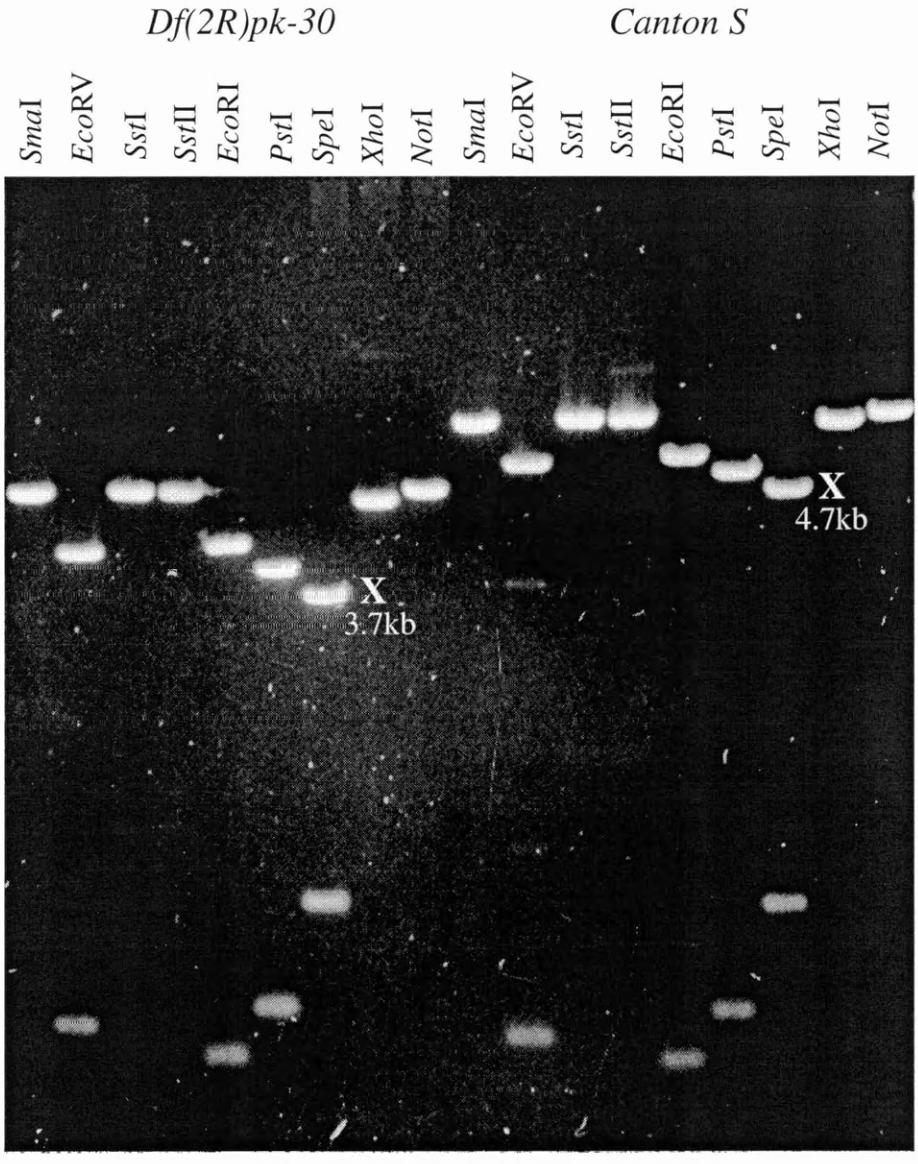


Fig.2.13 Restriction digests of subcloned PCR products from *Df(2R) pk-30* (lanes 1-9) and *Canton S* (lanes 10-18). The *Spe*I band covering the region deleted in *Df(2R) pk-30* is marked (X).

size of the band is reduced by about 1 kb in the corresponding *Df(2R)pk-30* digest, whereas the other band is the same size in both tracks.

The *Spe*I/*Xho*I fragment was subcloned from the *Df(2R)pk-30 Cla*I/*Xho*I subclone into pBluescript. Sequencing of this subclone was then carried out, initially using the Forward and Reverse pBluescript primers, then with oligonucleotide primers (1896, 1897, 2010, 2011, 2230; Appendix A).

Comparison of this sequence with the genomic sequence revealed that the *Df(2R)pk-30* mutation deletes 1306 bp of genomic DNA. Alignment with the cDNA sequence showed that it deletes bases 27 to 837 of the putative *pk* transcript. As the first 798 bases of the cDNA sequence corresponds to untranslated leader, only the last 39 bases of the deleted region contain coding sequence (Figure 2.10).

2.10 Rescue of the prickle phenotype

2.10.1 Procedure

The *Gal4-UAS* system (Brand and Perrimon, 1993) was used to test whether the Pk4aiii cDNA could rescue the prickle phenotype. A cDNA construct in the *UAS* vector pUAST (Brand and Perrimon, 1993; Appendix D) was prepared (section 7.20.1) and microinjected into *y w* embryos (injections carried out by Carol McKimmie).

2.10.2 Results

Seven of the injected embryos survived to become adults (G_0s). Of these, three were sterile and two were not transformants. Four lines were established from the remaining two G_0s; one from the first and three from the second. Two of the latter three (*UAS [pk$^+$] 3/7* and *3/9*) were on the second chromosome; the other two (*UAS [pk$^+$] 2/15* and *3/1*) were on the third. The subsequent experiments were carried out using the third chromosome inserts. Phenotypic rescue was attempted in a homozygous *Df(2R)pk-30* background, as this has a strong phenotype and had been characterised.

The driver *Gal4-C765* (Guillén *et al*, 1995; originally from Andrea Brand) was used to drive expression in the wing disc (Gómez-Skarmeta *et al*, 1996) as it is easy to see the pk phenotype in the adult wing. It has also been shown to drive expression in the leg disc (Lecuit and Cohen, 1997) and the eye disc (Burke and Basler, 1996).

Balanced stocks were made and crossed as described in Figure 2.14. The wings and legs of the resultant progeny were mounted.

All the results described below are with stock *UAS [pk$^+$] 3/1*. The triple row (Figure 2.15) and wing hair phenotypes (Figure 2.16) of wild type flies and the three classes of *prickle* mutants are shown for comparison. The triple row phenotype in the wing of *Df(2R)pk-30; Gal4-C765/UAS [pk$^+$]* flies was completely rescued (Figure 2.17a) and the mutant phenotype in the notum was reduced. The hair polarity of the wing (Figure 2.18a) differed from both wild type and *Df(2R) pk-30* wings (Figure 2.15 a and b). The same abnormalities were seen in *Gal4-C765/UAS[pk$^+$]* flies in a wild type background (Figure 2.18b), suggesting that overexpression of the *UAS[pk$^+$]* transcript can give a dominant polarity phenotype. The *Gal4-C765/UAS[pk$^+$]* flies with both wild type and *Df(2R)pk-30* backgrounds also had a leg phenotype (Figure 2.19e and f), which is not normally seen in *pk^{pk}* mutants (Figure 2.19b). The bristles of the leg were disrupted and there were tarsal duplications that extend further proximally than in *pk^{sple}* mutants (Figure 2.19d).

The experiment was repeated at 18^{o}C in an attempt to reduce the overexpression of the *pk* transcript. The triple row was almost completely rescued (Figure 2.17c), but the wing hair phenotype remained prickle-like (Figure 2.18c). The legs had a wild-type appearance (Figure 2.19g).

In *pk$^{pk\text{-}sple13}$; Gal4-C765/UAS[pk$^+$]* flies, the triple row remained wild-type in appearance (Figure 2.17d), but the wing hairs (Figure 2.18d) and the legs (Figure 2.19h) had a more extreme polarity phenotype than is seen in *pk$^{pk\text{-}sple13}$* flies (Figures 2.16d and 2.19c).

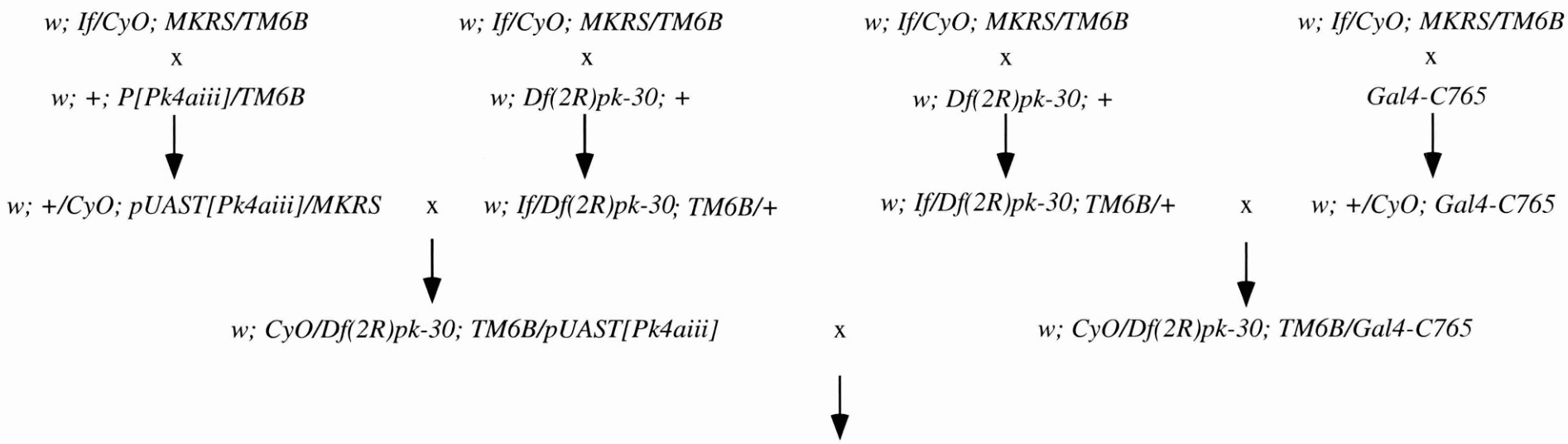


Fig.2.14 Crosses to obtain the *Pk4aiii* insert driven with Gal4-C765 in a prickle background (above) and in a wild type background (below). Only progeny used in subsequent generations are shown.

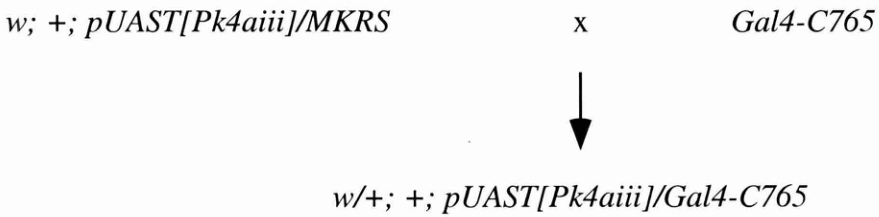


Fig. 2.15 Triple row phenotype of (a) wild-type; (b) *Df(2R)pk-30*; (c) $pk^{pk\text{-}sple13}$; (d) pk^{sple1} (x10)

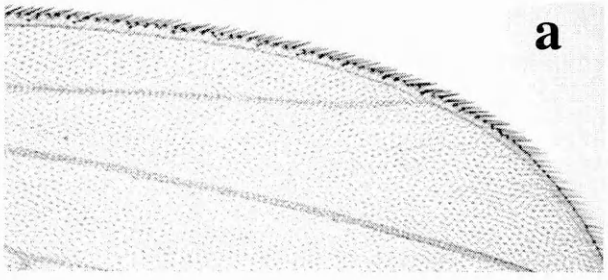
Fig. 2.16 Wing hair phenotype between veins III and V in the region of the anterior and posterior cross veins (see Fig. 1.1) of (a) wild-type; (b) *Df(2R)pk-30*; (c) $pk^{pk\text{-}sple13}$; (d) pk^{sple1} (x20)

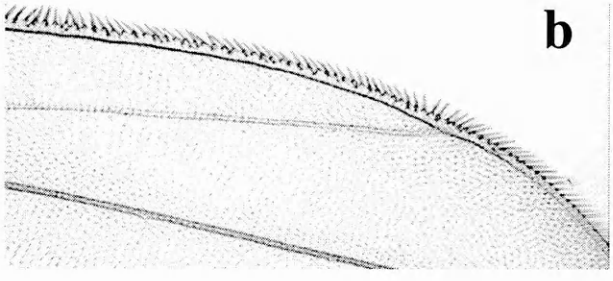
Fig. 2.17 Triple row phenotype of (a) *Df(2R)pk-30; Gal4-C765/UAS[pk+]*; (b) *+; Gal4-C765/UAS[pk+]*; (c) *Df(2R)pk-30; Gal4-C765/UAS[pk+]* at 18°C; (d) $pk^{pk\text{-}sple13}$*; Gal4-C765/UAS[pk+]* (x10)

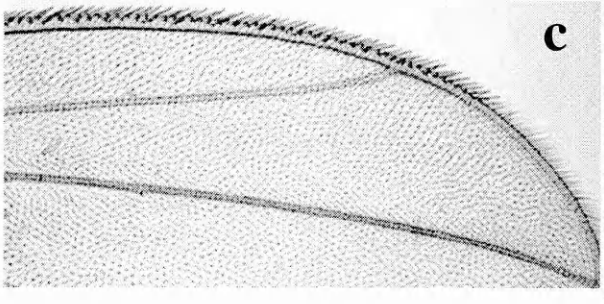
Fig. 2.18 Wing hair phenotype (as Fig. 2.16) of (a) *Df(2R)pk-30; Gal4-C765/UAS[pk+]*; (b) *+; Gal4-C765/UAS[pk+]*; (c) *Df(2R)pk-30; Gal4-C765/UAS[pk+]* at 18°C; (d) $pk^{pk\text{-}sple13}$*; Gal4-C765/UAS[pk+]* (x20)

Fig. 2.19 Leg phenotype of (a) wild-type; (b) *Df(2R)pk-30*; (c) $pk^{pk\text{-}sple13}$; (d) pk^{sple1}; (e) *Df(2R)pk-30; Gal4-C765/UAS[pk+]*; (f) *+; Gal4-C765/UAS[pk+]*; (g) *Df(2R)pk-30; Gal4-C765/UAS[pk+]* at 18°C; (h) $pk^{pk\text{-}sple13}$*; Gal4-C765/UAS[pk+]* (x20)

Fig. 2.15







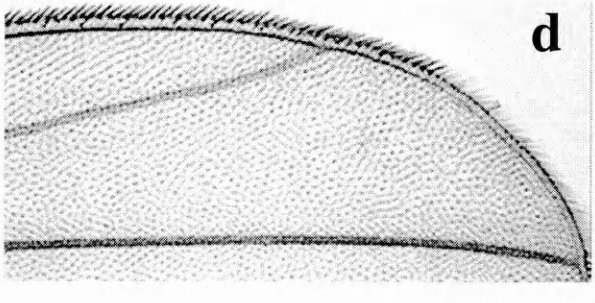
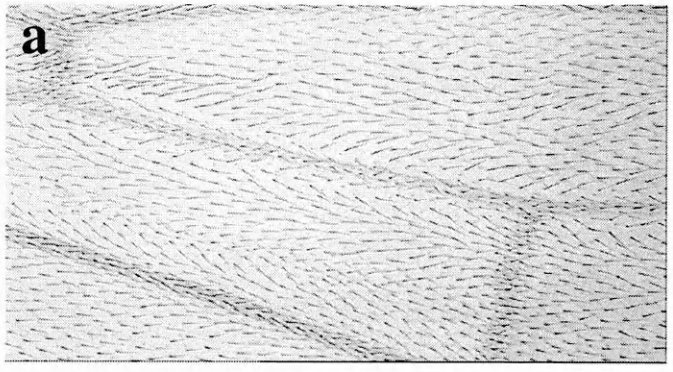
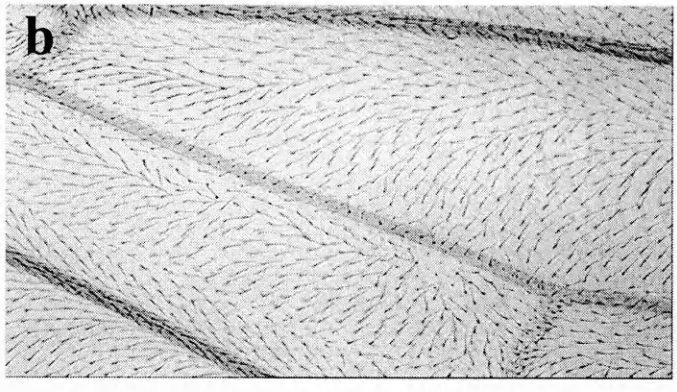
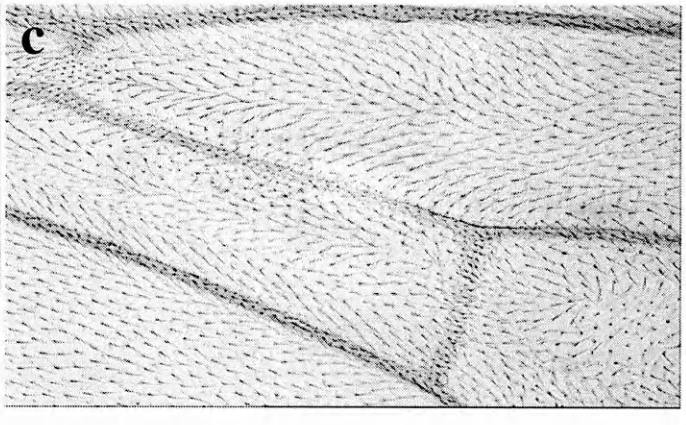


Fig. 2.16







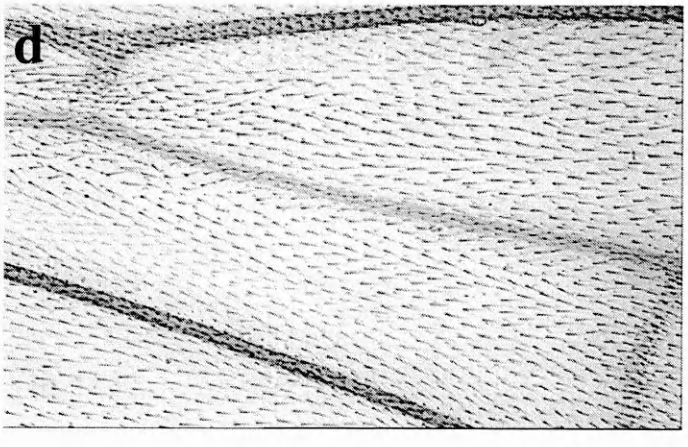
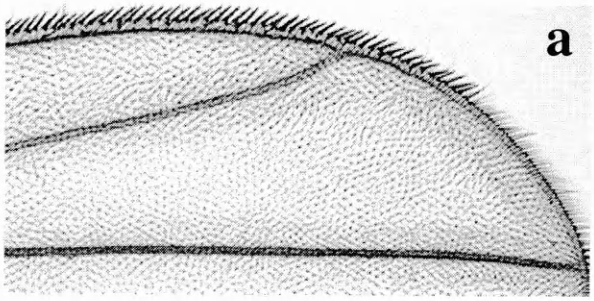
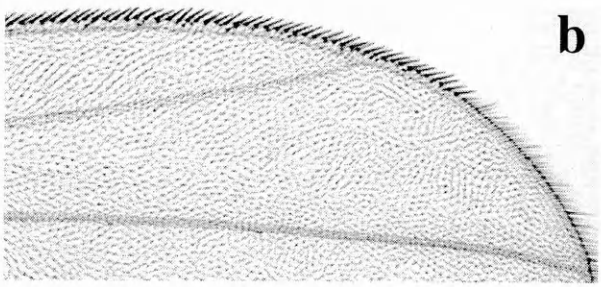
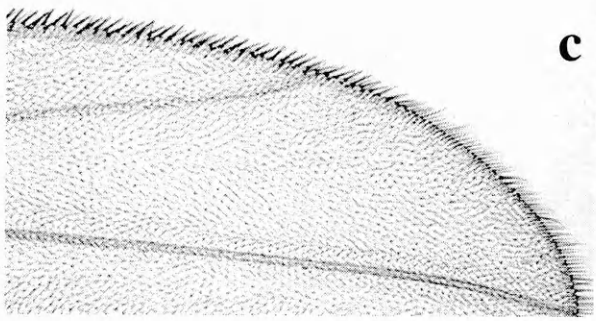


Fig.2.17







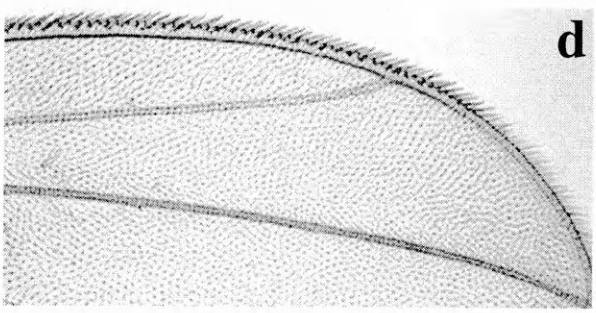
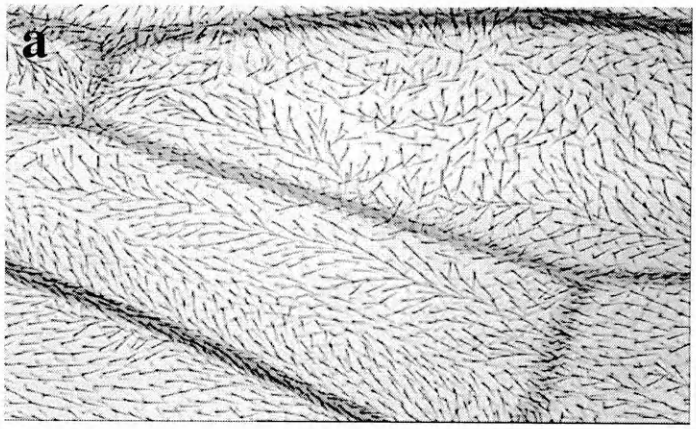
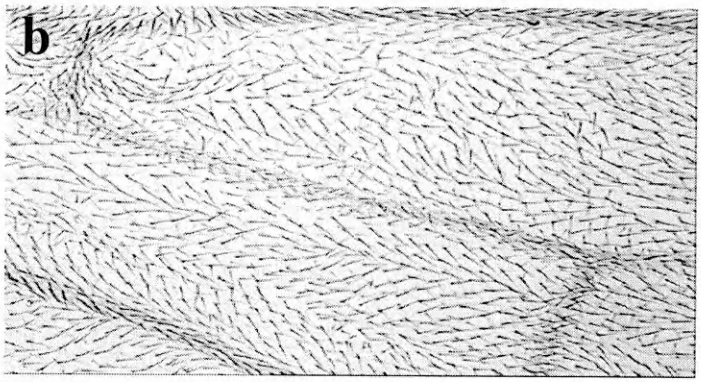
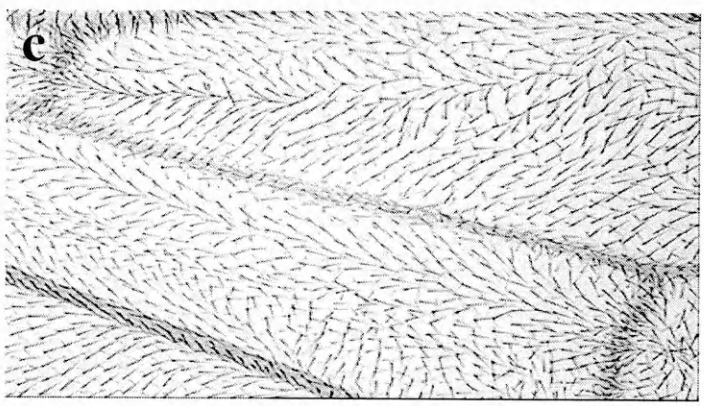


Fig. 2.18







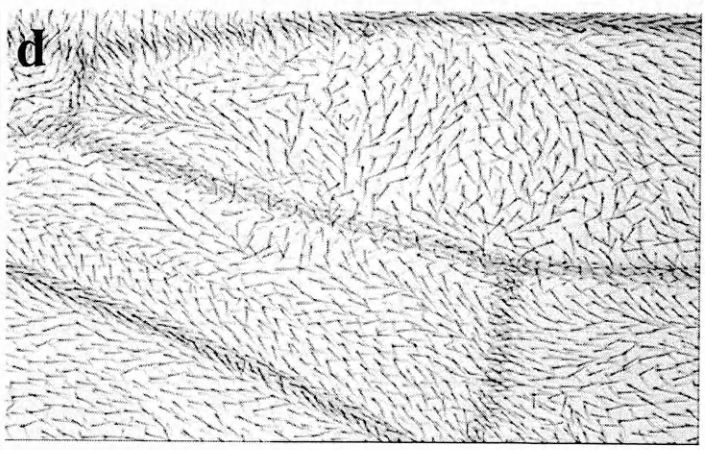
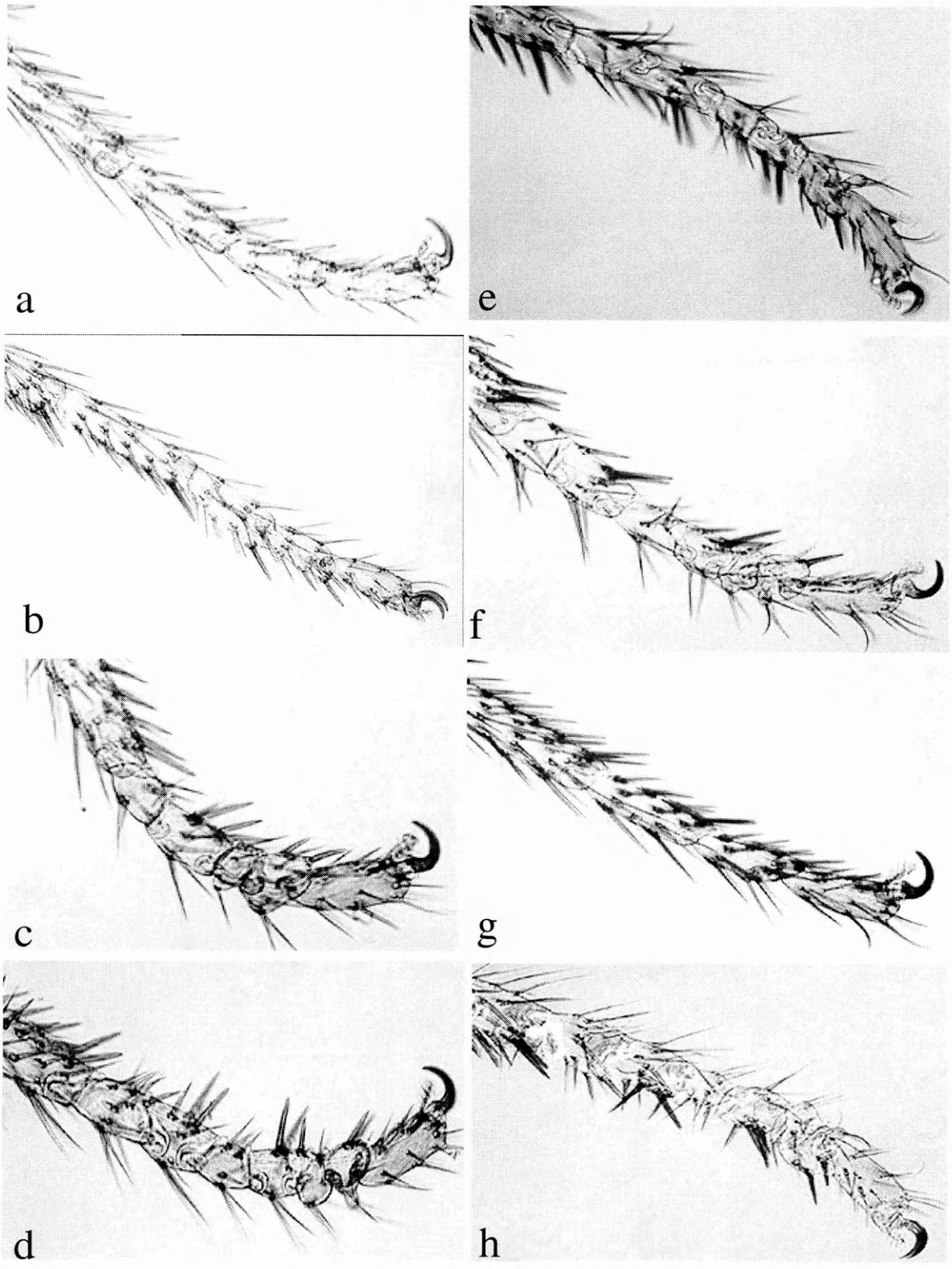


Fig. 2.19



2.11 Summary

This chapter describes the characterisation of the *pk* cDNA. Two transcripts mapping to the same region as of some of the *pk* mutants were located. A cDNA corresponding to the smaller of the two transcripts was cloned and sequenced. This 4.2 kb transcript, thought to encode the putative PK protein, consists of seven exons with the first exon about 60 kb proximal to the rest of the transcript. One of the strongest *pk* mutants, *Df(2R)pk-30*, was shown to be a deletion of 1.3 kb that removes most of the 5' exon of *pk*. Database searches revealed that PK is a LIM domain protein. Rescue of the pk phenotype was demonstrated using a construct of the cDNA with a *Gal4* driver.

CHAPTER 3. CHARACTERISATION OF *sple* AND AN ALTERNATIVE *pk* TRANSCRIPT

3.1 Introduction

The cloning and sequencing of the putative pk^{pk} transcript was described in the previous chapter. This chapter describes attempts to isolate the alternatively spliced 5' end of the transcript believed to correspond to pk^{sple}. As the *pk* transcripts are relatively rare and previous library screens had only identified one full length cDNA, it was unlikely that further screens would locate the pk^{sple} transcript. Alternative methods were therefore employed.

3.2 Attempts to locate alternatively spliced 5' end.

3.2.1 Homology to *D. virilis*

The original transcript was located by the identification of a region of homology between *D. melanogaster* and *D. virilis* (Chapter 2) and I tried this approach again. Southern blots of genomic DNA of the two species were hybridised to fragments from the phage walk. The results from this experiment were ambiguous (see Figure 3.1 for example). There did seem to be some signal in the *D. virilis* tracks with a few of the

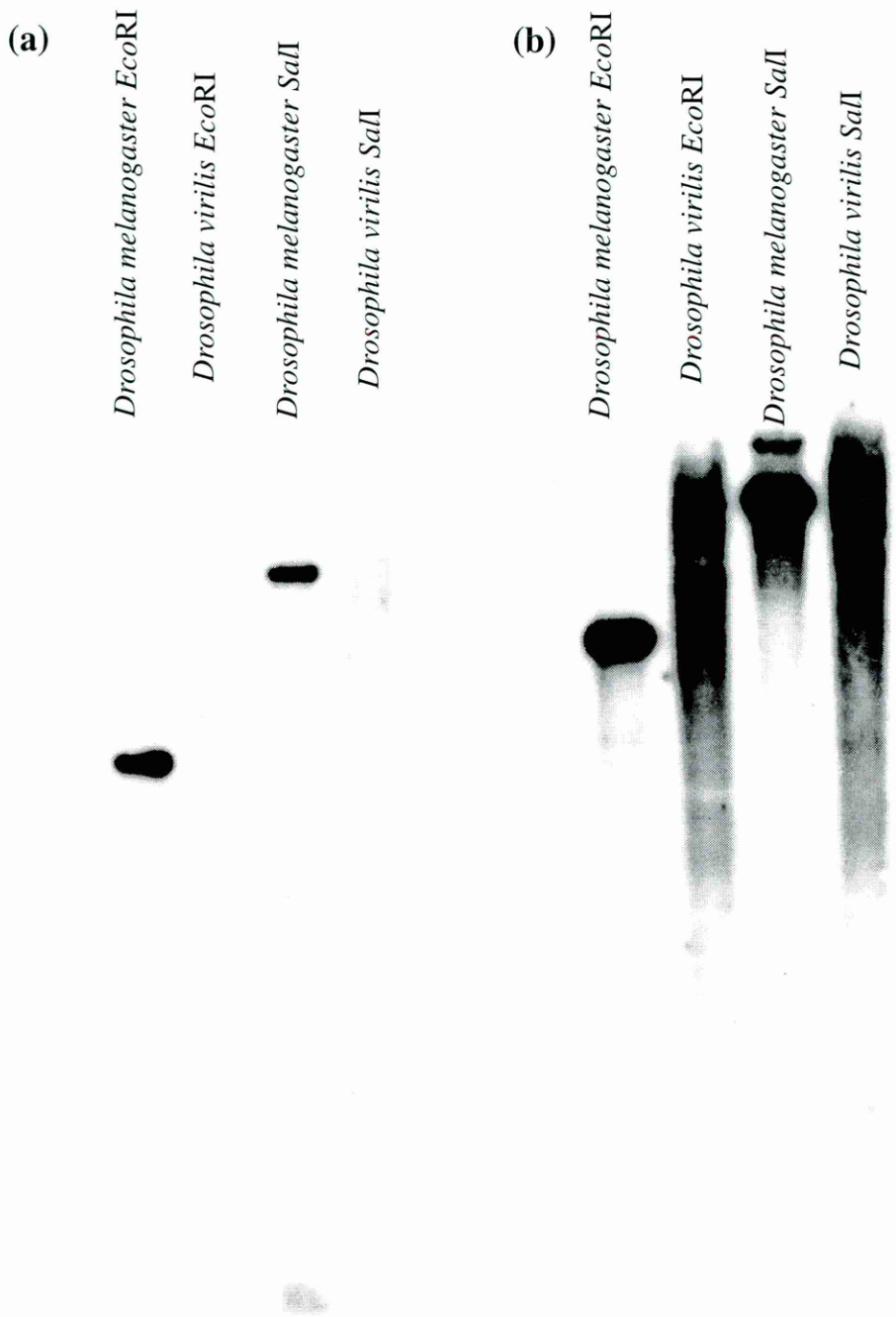


Fig. 3.1 Genomic Southern blot probed with (a) FP8/4 *Eco*RI 6.0 kb fragment and (b) FP8/4 *Eco*RI 6.7 kb fragment. Filters were washed at medium stringency (0.5xSSC, 65C) and exposed to film for 6 days. There is no signal in the *D. virilis* tracks in (a); there is signal in these tracks in (b) but it is not clearly resolved and could be background.

probes (FP7/7 *Eco*RI 1.7+2.3 kb; FP9/7 *Eco*RI 5.6 kb; FP9/7 *Eco*RI/*Sal*I 4.4 kb.), but it was diffuse and was indistinguishable from background noise.

3.2.2 PCR

PCR was carried out on the embryonic cDNA library (Brown and Kafatos, 1988; section 2.3.2). The primers used were 9884 and 9885 from Pk4aiii and 722 and 852 (Figure 3.2; Appendix A) from the vector pNB40. Pk4aiii DNA was used as a positive control.

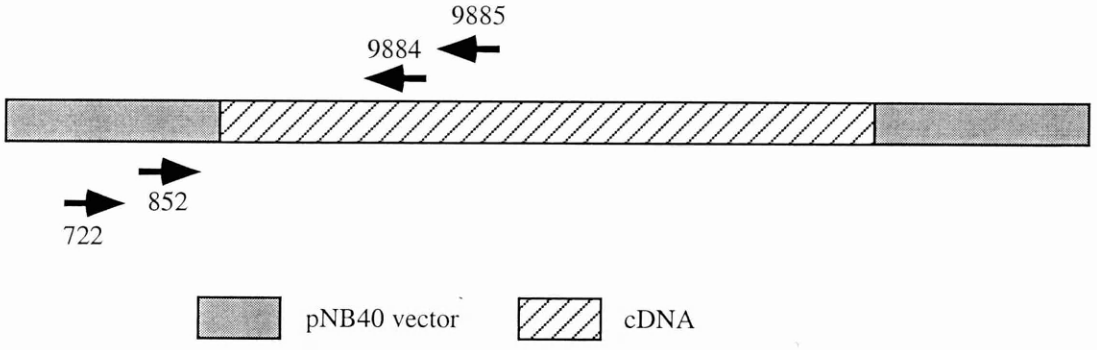


Fig. 3.2. PCR designed to locate alternatively spliced *pk* exons.
Programme 3 (section 7.14);
Primer pairs: 1st round 9885 and 722;
2nd round 9884 and 722;
3rd round 9884 and 852.

A band of the expected size (1.4 kb) was obtained with the positive control. Several bands were seen in the second and third round reactions, but none were as large as would be expected for the *sple* exons. It was possible that larger bands were present but not visible. The third round reaction was phenol extracted, chloroform extracted and ethanol precipitated, then random primed and hybridised to a Southern blot of the phage walk (Figure 3.3). All the bands that hybridised to the probe corresponded to the pk^{pk} cDNA.

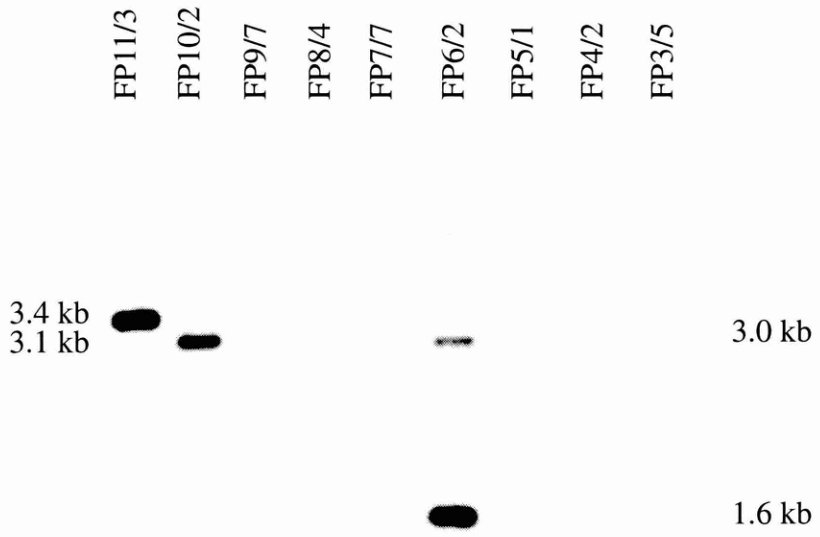


Fig. 3.3 Southern blot of *Eco*RI/*Sal*I digests of phage comprising the chromosome walk probed with 3rd round product of PCR (see section 3.2). The filter was washed at high stringency (0.1xSSC, 65C) and exposed to film for 1 week.

3.2.3 Marathon™ RACE PCR

David Huen carried out Marathon™ RACE PCR (CLONTECH Laboratories, Inc.) on cDNA he had produced previously (see Figure 7.3; section 7.15) with the nested primers DH1 and DH2 from the *pk*^{pk} cDNA (see Appendix A). A number of products were generated in several reactions, with the largest product being about 1.8 kb. This was the size expected for the alternative 5' end, given that the putative *sple* transcript identified on the developmental Northern was about 5 kb (section 2.2) and the unique *pk* exon accounts for 0.8 kb of the original 4.2 kb transcript.

3.3 Characterisation of *sple* exons

I random primed the two PCR reactions (from above) that contained the largest products (1.8 kb and 1.5 kb respectively) and hybridised them to separate Southern blots of the phage walk. The "1.8 kb" reaction hybridised to bands in FP6/2, FP7/7, FP8/4, FP9/7, FP10/2 and FP11/3 (Figure 3.4). The "1.5 kb" reaction gave the same results, except that one band in FP7/7 was missing.

The remainder of the "1.8 kb" probe was hybridised to individual mapping filters for each of the above phage, enabling it to be mapped more precisely (Figure 3.5). The bands in FP10/2 and FP11/3 were identified as the 5' exon of the putative *pk*^{pk} transcript. This information, plus the fact that one of the middle bands was missing from the 1.5 kb reaction, indicated that the 1.8 kb reaction contained several products, including the 5' end of the *pk*^{pk} cDNA isolated previously (Chapter 2) and the 1.8kb product.

David Huen produced individual clones from the reaction containing the 1.8 kb product in the pCR®2.1 vector (Appendix D) using the Original TA Cloning® Kit (Invitrogen). He showed that the individual 1.8 kb product corresponded to the bands in FP7/7. Northern blot data confirmed that this 1.8 kb clone hybridised to the larger of the two transcripts previously identified (David Gubb, personal communication).

The 1.8 kb clone was digested with a panel of restriction enzymes to identify sites within it and a restriction map was constructed (Figure 3.6). There are no *Bam*HI or *Xba*I

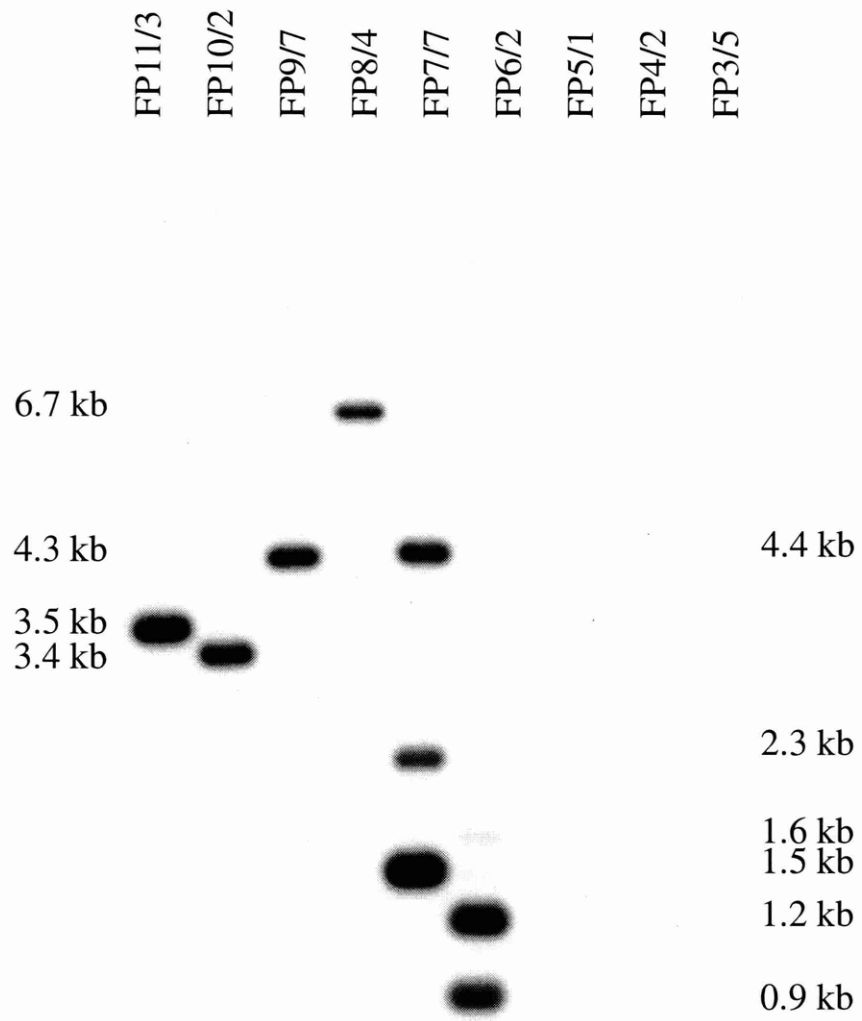


Fig. 3.4 Southern blot of *Eco*RI/*Sal*I digests of phage comprising the chromosome walk probed with 1.8 kb PCR product (see section 3.3). The filter was washed at high stringency (0.1xSSC, 65C) and exposed to film for 6 hours.

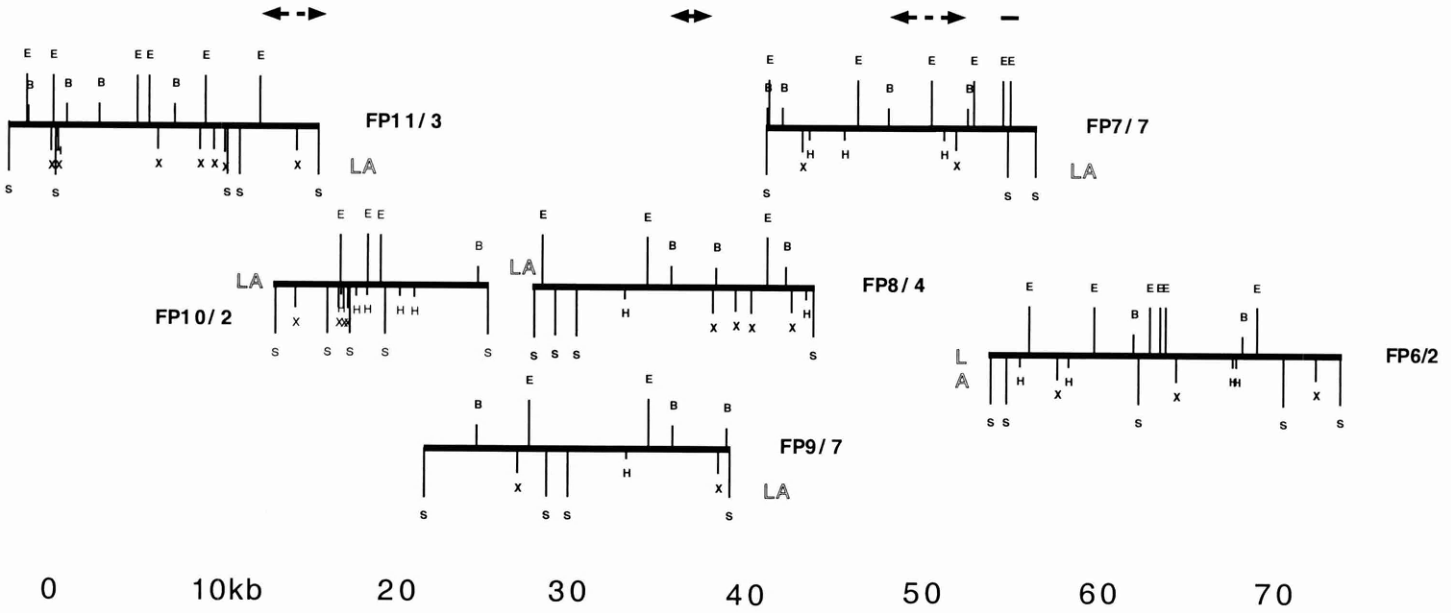


Fig. 3.5 Mapping of the 1.8kb PCR product. ◄- - ► indicates the fragments the product maps to. Enzymes as before (see Fig. 2.1)

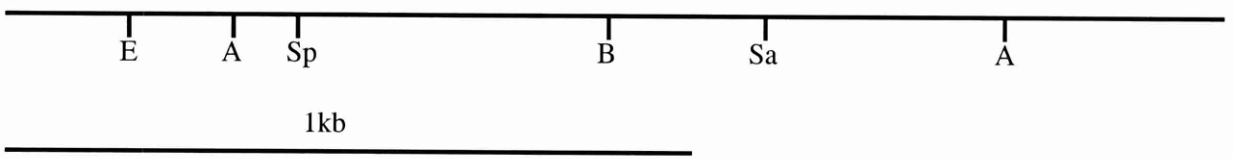


Fig.3.6 Restriction enzyme map of *sple* exons. Enzymes are *Eco*RI (E); *Ava*I (A); *Spe*I (Sp); *Bst*XI (B); *Sal*I (Sa).

sites within the 1.8 kb fragment and these enzymes cut the polylinker of the pCR2.1 vector either side of the insert. The 1.8 kb fragment was subcloned into pBluescript SK+ using these sites to facilitate subsequent subcloning.

3.4 Sequence of *sple* exons

Restriction fragments were subcloned into pBluescript SK+ (Appendix B) and these, plus the whole insert, were sequenced from the Forward and Reverse sequencing primers. The gaps in the sequence were then filled in using oligonucleotide primers designed for the purpose (Appendix A).

The whole insert is 1807 base pairs long. The last 29 bases are the first bases of the second exon of the putative pk^{pk} transcript, confirming this as a common exon. The longest predicted open reading frame is 1047 bases. The methionine encoded by bases 733-735 is embedded in good *Drosophila* start site consensus sequence (Cavener and Ray, 1991). Conceptual translation of the longest ORF predicts that the unique *sple* coding sequence is 349 amino acids long, therefore the total length of the putative pk^{sple} protein is 1206 amino acids (Figure 3.7).

Comparison with genomic sequence from the Berkeley *Drosophila* Genome Project (clone P1 DS08462; accession number AC005198) revealed the presence of one intron of 5 kb within the unique *sple* sequence. The intron between the *sple* exons and the common exons is 14.9 kb. No known motifs were found in the new sequence and database searches did not reveal any significant homologies.

Fig. 3.7 cDNA sequence of the unique *sple* exons and conceptual translation of the longest ORF. *Restriction enzyme sites* are in italics.

```
TTTACATCTGGCAACAGGTGCGGGTTAGTCGTCGCGTAATCCGTCTGAACGATCGATCGA   60
TCGCGCGCTCTTCGGCTATGTCGATTTCGGTCCACGTTATCTTGCGCCAGATCTGTCGTT  120
                                                      EcoRI
TACCGTTTTTCAGCCAGATCGGAGCAAAGTGCACAAAAAGCTGCTAACTGAATTCGCAAC  180
AATTGCAAAAAATTGACAAAAGTGAAGTGCAATCCATAATCCATAGCTGGTCCCTATCAG  240
GGCCAGCTAAATAAGTCTGGCATAGTGTTCCTTCCTGGGGCCAAGTGAATTAATGCAACA  300
                                  AvaI
ATTGGCCTAAGCGCTCATTTACGCTAAATGCTCGGGATAAGCCACACGGTTCCTAATCAA  360
TGCAGTGACCATACCATACCATACAATCAATAAATATTGATTTTTTATAAAGTGGTATAA  420
                          SpeI
ATTAACTAGTTCCTAATCAATGCAGTGACCATACCATGTATTTTAAAGAATATATGTAAT  480
GTACGAAAGGGTACAAAATGCTTGAAACTCCACAAACAAACGCCGAATTTAAATAAAAAA  540
AAAGTGTACTGAAAACTACAATCATGTAATTAAAATTCTACTCGAAGCTAAATTTACGAG  600
GAAAAGCTGAAAGAAACATGTAAAACAAAAGAAACCAAAAAGCAGGTACAAGAACAAAAT  660
GACACTGCCGACAAGAAAAGACAACGAGAACTAAAAAATAATTACGGGCCCAGAACAATA  720
ATAATAAGCAACATGAGCAGCCTGTCAACCGGTGGAGGAGCAGGAGGCTCCAGCGGAGGA  780
            M  S  S  L  S  T  G  G  G  A  G  G  S  S  G  G
CCGGGGGGAGCAGATGCTGCCGCTGCTCCTGCTGCTGGCCAGGCAACAGTAACGGCCACT  840
P  G  G  A  D  A  A  A  A  P  A  A  G  Q  A  T  V  T  A  T
                                  BstXI
GGCAACATGGAGCCCGCGATGGTGCCACGTACCGCCAATTTGCTGGCATGCAAACAATGG  900
G  N  M  E  P  A  M  V  P  R  T  A  N  L  L  A  C  K  Q  W
TGGCGCGTCTGCTTCTTGTACGGGGACCAGCAGAAATATTATCGACAGCTCTATAGCAAG  960
W  R  V  C  F  L  Y  G  D  Q  Q  K  Y  Y  R  Q  L  Y  S  K
GCGGCGGCCCAGCGGCTGGCGGACGCTAATCAGGAACCCGACAATGCCCGCGATCGGGAG  102
A  A  A  Q  R  L  A  D  A  N  Q  E  P  D  N  A  R  D  R  E
TACGACACCGTGGACTGCGATCTTATCGCCGGGCAATTAGATGCCGTCGAGGATGCGGAC  108
Y  D  T  V  D  C  D  L  I  A  G  Q  L  D  A  V  E  D  A  D
                          SalI
GATGGCATCGATCTAGGCGATCACTCGTCGACCCCGAAGGGAGGTGCGACGACGGCGGGC  114
D  G  I  D  L  G  D  H  S  S  T  P  K  G  G  A  T  T  A  G
CGTCCGCTCTTCCCGCATTCCTCCTCACCGCGGCGCAGCAAGAAGCTCCTGAGATCCCTG  120
R  P  L  F  P  H  S  S  S  P  R  R  S  K  K  L  L  R  S  L
CGAGCCCATGTGCGCGGGGAGAAGCTGCCGAAGAACGACACTACGACGGCTAACGTGTCC  126
R  A  H  V  R  G  E  K  L  P  K  N  D  T  T  T  A  N  V  S
AACGAGGTGACCCAGCGGAATGCCAGGGTGACCGCTCTCGACGACCCCTTTCTATTCGGC  132
N  E  V  T  Q  R  N  A  R  V  T  A  L  D  D  P  F  L  F  G
ATCGACGCCGATCACTTGGGGGATCTCGTGGTGCGGGGCAAGCGGTACAGTACGCTAGAC  138
I  D  A  D  H  L  G  D  L  V  V  R  G  K  R  Y  S  T  L  D
GCCACCGAGAACATGGCCAGGTTCTATGCGGAGCAGGAGGCCACGGCCCAGGTCCTGGAG  144
A  T  E  N  M  A  R  F  Y  A  E  Q  E  A  T  A  Q  V  L  E
                          AvaI
ATCATCGAGCAGGAAGAAGAATCTCCCGAGCAGGAAGCTCCCAAGCCCGCCCTACCGCCC  150
I  I  E  Q  E  E  E  S  P  E  Q  E  A  P  K  P  A  L  P  P
AAACAGAAGCAGCAGCGTCCTGTGCCGCCACTGCCACCACCACCAGCGAACCGGGTCACC  156
K  Q  K  Q  Q  R  P  V  P  P  L  P  P  P  P  A  N  R  V  T
CAGGACCCAGGAACGCAACCTGCAGCACCACAAGTGCCCCTGCAACCCCTGACCGCCGGC  162
Q  D  P  G  T  Q  P  A  A  P  Q  V  P  L  Q  P  L  T  A  G
GATCTTCAGTTTCTAAACTTGAGCCTGCGGCAGAGGAGTTTACCGCGCAGCATGAAGCCG  168
D  L  Q  F  L  N  L  S  L  R  Q  R  S  L  P  R  S  M  K  P
TTTAAGGATGCCCACGATATCAGTTTCACTTTCAACGAACTGGACACGAGCGCCGAACCG  174
F  K  D  A  H  D  I  S  F  T  F  N  E  L  D  T  S  A  E  P
GAAGTGGCGACAGGAGCCGCCCAGCAGGAGTCAAATGAG                       177
E  V  A  T  G  A  A  Q  Q  E  S  N  E
```

3.5 Assembly of *sple* cDNA

A full length pk^{sple} cDNA was generated by ligating the pk^{sple} specific exons (amplified by PCR) to the exons common to both cDNAs. First a pk^{sple} 5' fragment including a restriction enzyme site near the 5' end of the common exons was needed. The *Sma*I site within the first common exon was selected. Two nested primers just 3' of this site were designed, along with two nested primers as close as possible to the 5' end of the unique pk^{sple} exons (Appendix A). The same library as before (section 3.2.3) was used for two rounds of Marathon™ PCR (Figure 3.8).

Several bands were seen in the second round reaction. These fragments were gel purified and cloned into the pCR®2.1 vector using the Original TA Cloning® Kit. All of the subclones were sequenced using the Forward and Reverse sequencing primers. The largest insert (1.8 kb) was confirmed to be the required pk^{sple} fragment. The shorter inserts were incomplete pieces of the same sequence. The 1.8 kb subclone was used to assemble a complete *sple* cDNA (see Figure 3.9)

3.6 Rescue of *sple* phenotype

A construct containing the complete pk^{sple} cDNA was made (section 7.20.1). This was injected into *y w* embryos by Carol McKimmie. Of 13 surviving G_0s, one was sterile and eight were not transformants. Five transformant lines were established from the remaining four G_0s. One of these (*UAS [sple$^+$] 8a*) was found to be on the X chromosome, one (*UAS [sple$^+$] 4c*) was on the second and three (*UAS [sple$^+$] 4e, 10d* and *12a*) were on the third. The subsequent experiments were carried out with the third chromosome inserts. Rescue was attempted in the pk^{sple1} mutant. Crosses were carried out in the same way as for the pk^{pk} experiments (see Figure 2.14).

The leg phenotype of pk^{sple1}; *Gal4-C765/UAS[sple$^+$]* flies was rescued (Figure 3.10a), but as with the *pk* construct, an overexpression phenotype was also seen, this time in the wing. There was an extreme triple row phenotype (Figure 3.11a) and the hairs

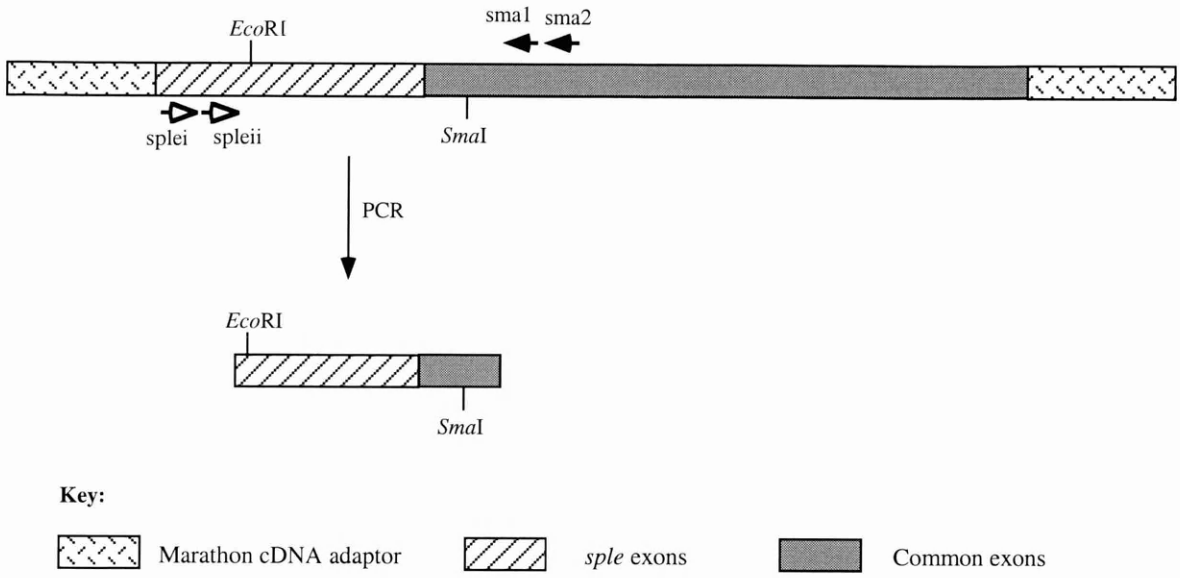
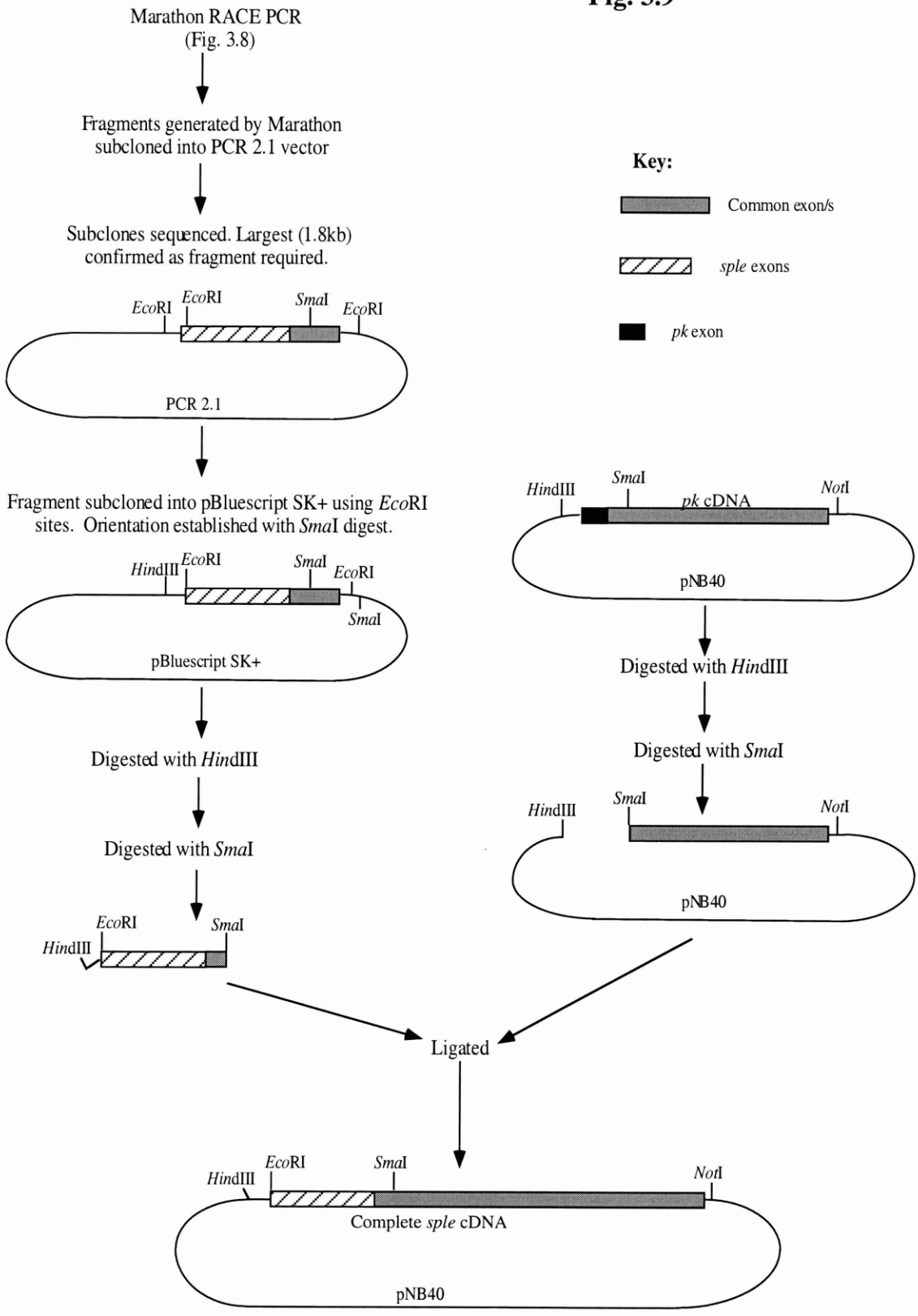


Fig.3.8 Marathon PCR to produce 5' region of *sple* cDNA with convenient restriction enzyme sites for generation of complete cDNA (not to scale). Primers sma2 splei were used for the first round; sma1 and spleii for the second round.

Fig. 3.9 (overleaf) Scheme for the generation of complete *sple* cDNA (not to scale). Suitable sites for subcloning the insert directly into the pNB40 vector were not available. *Eco*RI cuts the pCR2.1® vector either side of the insert. It also cuts the insert 180 bases from the 5' end. The 1.8 kb insert above was therefore subcloned into pBluescript SK+ using the *Eco*RI sites. The orientation of the insert was determined using a *Sma*I digest. In the orientation required, the *Sma*I site of the insert is at the end closest to the *Sma*I site of the vector. The latter subclone and the pk4aiii insert in pNB40 were both digested with *Hin*dIII and then with *Sma*I. The 1.8 kb *sple* band and the 5.57 kb band containing the common exons plus the pNB40 vector were gel purified, ligated together and the procedure continued as for a normal subcloning. The presence of the insert was confirmed using an *Eco*RI/*Not*I double digest.

Fig. 3.9


Marathon RACE PCR
(Fig. 3.8)
Fragments generated by Marathon subcloned into PCR 2.1 vector
Subclones sequenced. Largest (1.8kb) confirmed as fragment required.
EcoRI
EcoRI
SmaI
EcoRI
PCR 2.1
Fragment subcloned into pBluescript SK+ using EcoRI sites. Orientation established with SmaI digest.
HindIII
EcoRI
SmaI
EcoRI
SmaI
pBluescript SK+
Digested with HindIII
Digested with SmaI
EcoRI
SmaI
HindIII
Key:
Common exon/s
sple exons
pk exon
HindIII
SmaI
pk cDNA
NotI
pNB40
Digested with HindIII
Digested with SmaI
HindIII
SmaI
NotI
pNB40
Ligated
HindIII
EcoRI
SmaI
NotI
Complete sple cDNA
pNB40

were also affected, in some regions being completely reversed compared to wild-type (Figure 3.12a). A similar pattern was seen with this construct in a wild-type background (Figures 3.10b, 3.11b and 3.12b). When the *pk^{sple1}; Gal4-C765/UAS[sple^{+}]* flies were grown at 18^{o}C, the leg phenotype was not rescued (Figure 3.10c). The triple row was almost normal (Figure 3.11c) but there was still a wing hair phenotype, although not as extreme as in the same flies grown at 25^{o}C (Figure 3.12c).

Fig. 3.10 Leg phenotype of (a) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]*; (b) *+; Gal4-C765/UAS[sple^{+}]*; (c) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]* at 18^{o}C; (x20)

Fig. 3.11 Triple row phenotype of (a) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]*; (b) *+; Gal4-C765/UAS[sple^{+}]*; (c) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]* at 18^{o}C; (x10)

Fig. 3.12 Wing hair phenotype (as Fig. 2.16) of (a) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]*; (b) *+; Gal4-C765/UAS[sple^{+}]*; (c) *pk^{sple1}; Gal4-C765/UAS[sple^{+}]* at 18^{o}C; (x20)

Fig. 3.10

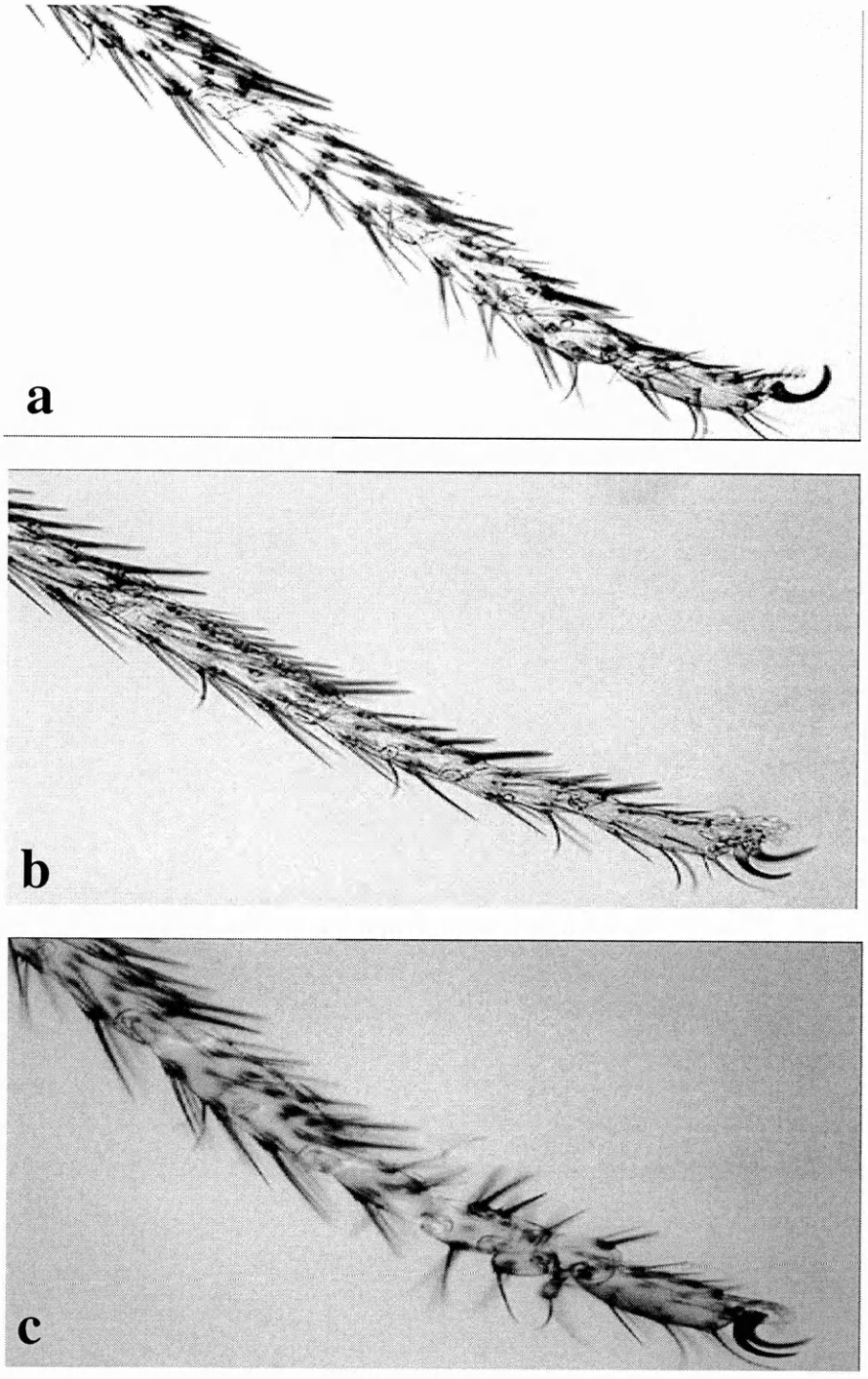
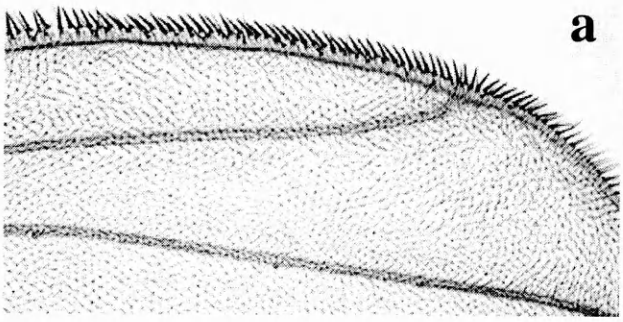
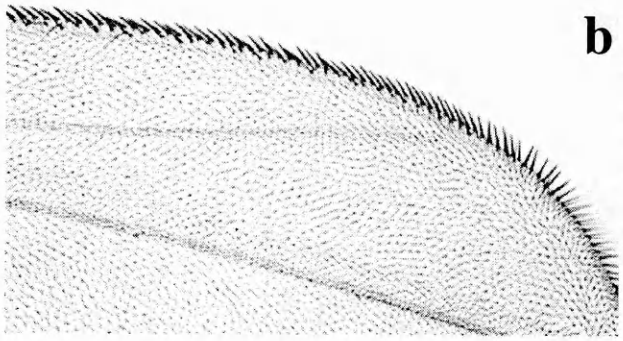


Fig. 3.11





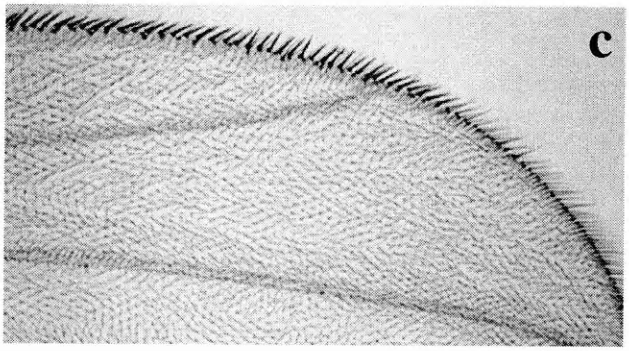
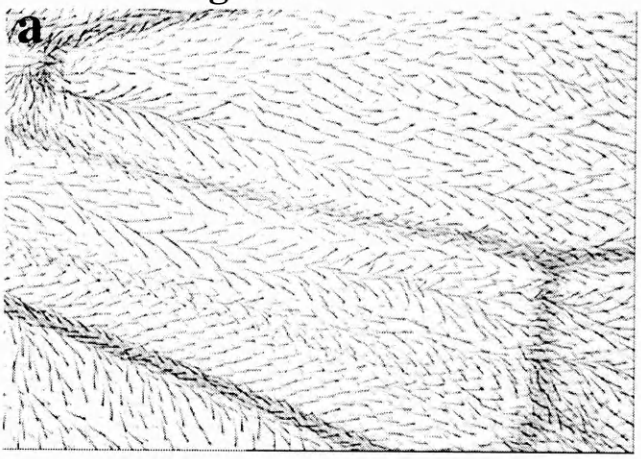
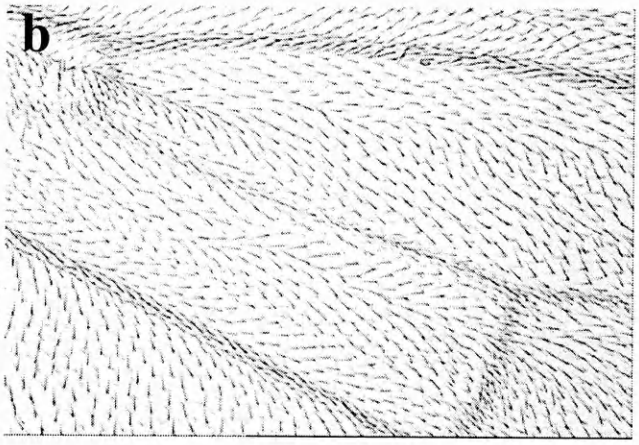
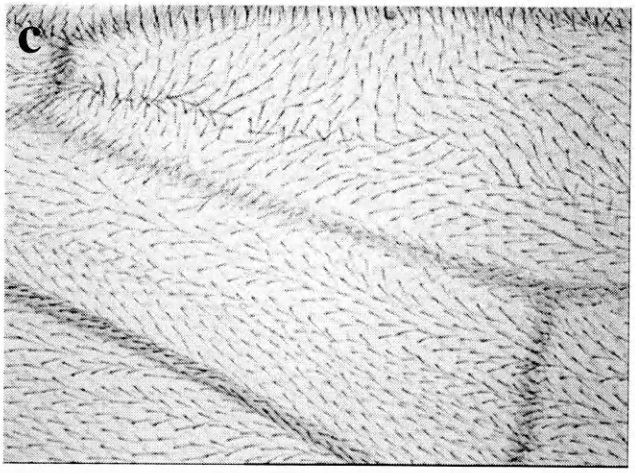


Fig. 3.12







3.7 The alternative *pk* exon

3.7.1 Characterisation

David Huen identified and subcloned one other new fragment from the Marathon RACE reactions. This 0.65 kb fragment was shown to correspond to the band in FP8/4 (overlapping phage FP9/7) on the walk. David Gubb showed on a developmental Northern that this *pkM* (medial) transcript was only detected in embryos (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation). The phenotype associated with this transcript is not known. The adult phenotype of *In(2LR) pk20* (that separates both the *pkM* and *pk* 5' exons from the common exons) is indistinguishable from other *pk* alleles in which the *pkM* transcript is intact.

3.7.2 Sequence

I sequenced the 0.65 kb subclone, initially using the Forward and Reverse primers, then oligonucleotides specifically designed for the purpose (Appendix A). The sequence was 706 bases long; the last 29 bases overlapped with the first of the common exons (Figure 3.13). The methionine encoded by bases 443-445 is embedded in reasonable *Drosophila* start site consensus sequence (Cavener and Ray, 1991). Conceptual translation of the longest ORF therefore predicts that the unique *pkM* coding sequence is 79 amino acids long, so the total length of the putative *pkM* protein is 936 amino acids. Comparison with genomic sequence from the Berkeley *Drosophila* Genome Project (clone P1 DS08462; accession number AC005198) showed that there are no introns within the unique *pkM* sequence; the intron between the *pkM* exon and the common exons is 36.1 kb. No known motifs were found in the new sequence and database searches did not reveal any significant homologies. Figure 3.14 shows a summary of all the *pk* transcripts.

Fig. 3.13 cDNA sequence of the unique *pkM* exon and conceptual translation of the longest ORF. *Restriction enzyme sites* are in italics.

```
AACGCGAGCGGACGAATCCGCCGATTCCTCGTGTATTCCGACTTATCCCGCTGCTGATAA   60
GTCGCCGCGATTGCAACAATTAGCTATAATTATTCATTAATTCGAATCGGTGTTGAGGAG  120
                                                    XhoI
TTATCGCTAATTCGCAGAAACTGTGCCAGCGAGATAGAATACCACTCGAGTTTTCCTCAA  180
TCTCTTCGCCGGTCGCCTTTGTTATTGTTTGCCGCTGCTTCTTCCTCTTTTTTTGGGGGG  240
AATAGTTCTTTGTGTTTGTGCCCGTGTGTTTGAGAGCCTGGCAGTGTGTGTGTACGTATC  300
TGTGTCTGTGAGTCAGCGCAAATGGCAATACGCATCTTTGTTTACCCGCAAAATCCAGGG  360
CTCTCATCGCGGAGAAGTGCCAGTGAAAGCGGCGCCAAAATCTTAGTCTAGTGATTTTCT  420
TCTGCCTTTCCGGCCGAAACAATGAACGATTCGACCGATAATTTGCATGCCGACTGCGAC  480
                        M  N  D  S  T  D  N  L  H  A  D  C  D
GGTCGAGTCAGCAACAACAACAATGGCAACAGCAACACCAACGATGGACCCAACAACGAC  540
G  R  V  S  N  N  N  N  G  N  S  N  T  N  D  G  P  N  N  D
GGCGACTCCGACGAGGAGGTCATCGAGGGCATGGCGCTGCTGGAGGGCAACTACCAAGTG  600
G  D  S  D  E  E  V  I  E  G  M  A  L  L  E  G  N  Y  Q  V
CTGCGCCAGTGGGTGCCGCCCGCCCCCAACTATTGGGATGCGCCCCCCAAGGCCATCATC  660
L  R  Q  W  V  P  P  A  P  N  Y  W  D  A  P  P  K  A  I  I
AAATCCGCCGAAGTCAGG                                             678
K  S  A  E  V  R
```

3.8 Summary

This chapter describes various attempts to locate the alternative *sple* exons, including the isolation of a 1.8 kb fragment that included the two novel exons subsequently identified as the 5' exons of *sple*. These were characterised and shown to encode an additional 349 amino acids of protein. Another novel exon, believed to be an alternative *pk* 5' end, was also characterised. The *sple* cDNA was constructed and shown to rescue the sple phenotype.

The method of locating coding sequence by homology to *D. virilis* that enabled us to identify the original pk^{pk} transcript (section 2.2) was unsuccessful in locating the alternative transcripts (section 3.2.1). However, subsequently it was found that the probes that gave "background noise" in the latter were from the regions where the $pk^{pk\text{-}sple}$ and pk^{pkM} exons were located. The original sequence located was over 3 kb (the common exons) whereas the largest of the alternatively spliced exons was 1.4 kb, so it is possible this method only detects large regions of coding DNA, thus limiting its usefulness. This approach has not been tested further.

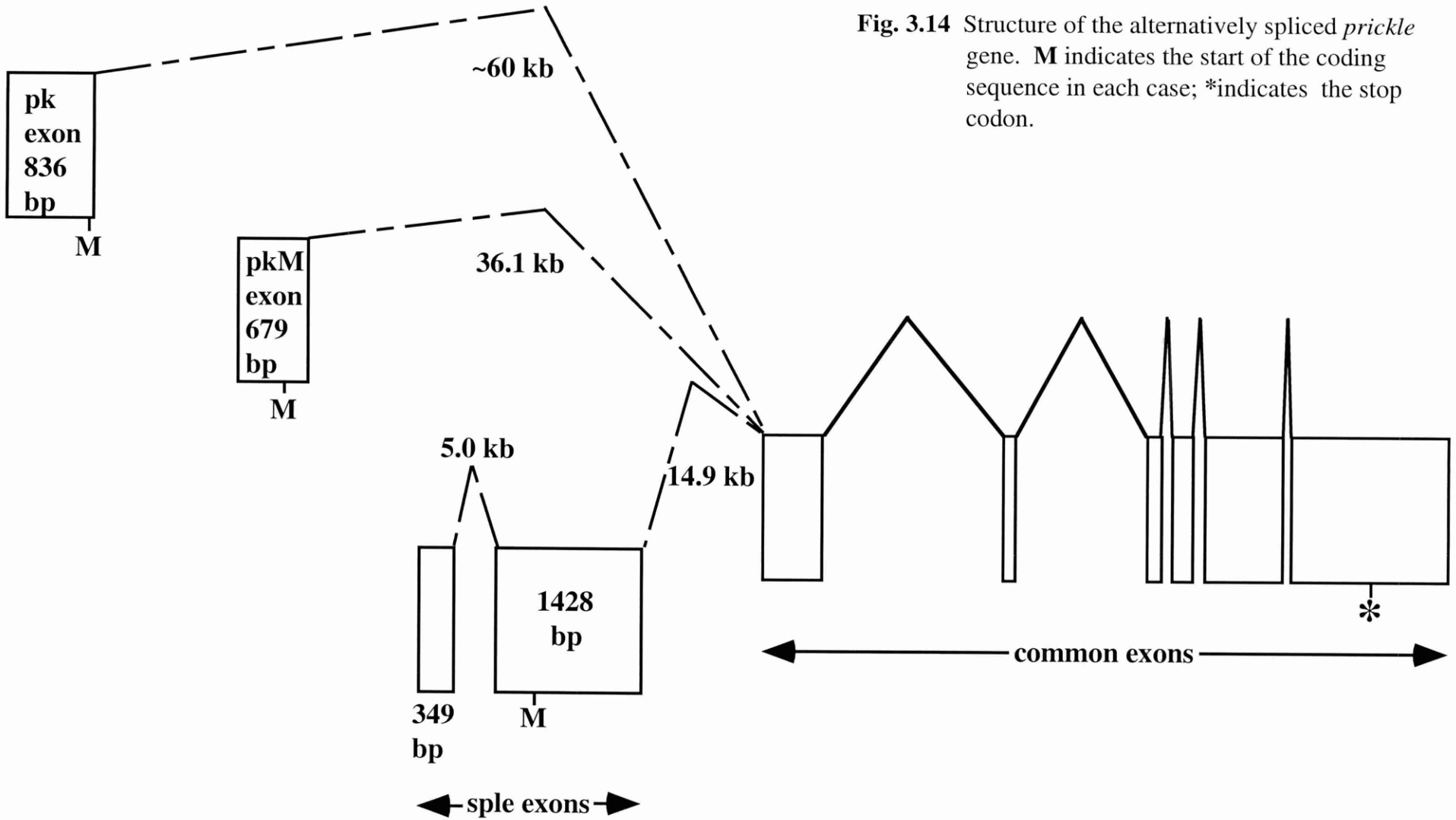


Fig. 3.14 Structure of the alternatively spliced *prickle* gene. **M** indicates the start of the coding sequence in each case; *indicates the stop codon.

CHAPTER 4. THE *pk* TRANSCRIPTS ENCODE A LIM DOMAIN PROTEIN

4.1 Introduction

4.1.1 LIM domains

Database searches revealed that PK was a putative LIM domain protein (see Chapter 2). LIM domains are cysteine rich sequences with the consensus sequence $CX_2CX_{17\text{-}19}HX_2CX_2CX_2CX_{16\text{-}20}CX_2D/H/C$ where X is any amino acid (Dawid *et al*, 1998). The term LIM comes from the three proteins in which the motif was first recognised; LIN-11 from *Caenorhabditis elegans* (Freyd *et al*, 1990), ISL1 from rat (Karlsson *et al*, 1990) and MEC-3 from *C. elegans* (Way and Chalfie, 1988). These are all homeodomain (HD) proteins, as are many of the LIM proteins identified since. LIM domains both positively regulate LIM-HD activity, by promoting protein-protein interactions, and negatively regulate it, possibly by preventing the homeodomain associating with DNA (reviewed by Curtiss and Heilig, 1998). In the latter case, the LIM domains may interact with the HD to prevent it binding to DNA in the absence of a LIM binding partner. In the presence of such a partner, the restraint is lifted, and both proteins may contribute to the assembly of a transcription complex. For example, Sánchez-García

et al (1993) found that the inhibitory effect of the LIM domains of ISL-1 was not restricted to DNA binding, but to the transcriptional activation function as well.

LIM domains have now been identified in non-HD proteins, either in combination with kinase or GAP (GTPase activating protein) domains or in proteins with no other identified domains (Curtiss and Heilig, 1998). A classification of LIM proteins placed them into three groups, with some exceptions (Taira *et al*, 1995; Dawid *et al*, 1998). Group one proteins have a pair of LIM domains near the N-terminus; they include the LIM-HD and LIM Kinase proteins. Group two proteins have related LIM domains and contain only a short additional conserved motif. Group three proteins are more heterogeneous than the other two groups and tend to have LIM domains localised to the C-terminal region. Other domains in these proteins are implicated in binding proteins such as actinin, vinculin and SH2- (Src homology) and SH3-containing proteins. Group one proteins tend to be nuclear, while those in groups two and three are mostly cytoplasmic (Dawid *et al*, 1998). Many group three proteins are associated with the cytoskeleton or membranous organelles and may function in adhesiveness and defining cell shape (Taira *et al*, 1995).

The LIM domains themselves have also been classified into five groups according to their sequence (Dawid *et al*, 1995). In group one proteins, the N-terminal LIM domain is type A; the second LIM domain is type B. In proteins composed largely of LIM domains, the LIM domains are usually type C. Type D are a more divergent group of LIM domains found in, for example, zyxin and paxillin. The small group of type E domains are always found in association with type D domains and include LIM 3 of zyxin.

LIM domains are similar to zinc finger motifs (see below) and also to ferrodoxins, and were originally thought to bind zinc and iron (Li *et al*, 1991). Subsequent studies demonstrated that they only bound zinc (Michelsen *et al,* 1993; Archer *et al*, 1994). Despite their similarity to other zinc finger motifs, there is no evidence for LIM domains binding to DNA and they are thought to function as protein-protein interaction motifs for the assembly of functional complexes (reviewed by Gill, 1995). Feuerstein *et al* (1994)

showed that CRP-CRP (cysteine rich protein) protein interaction is not only dependent on the LIM region, but this region is fully sufficient for the interaction.

LIM proteins also interact with other proteins. The POU domain is a conserved region found in many transcriptional activators (Billin *et al*, 1991). It was found that the DNA binding of the LIM-HD protein MEC-3 and the POU-HD protein UNC-86 is greatly increased by a co-operative interaction between the two proteins (Xue *et al*, 1993). It was also shown that the two proteins can interact without DNA. An association has also been demonstrated between LIM proteins and the basic helix-loop-helix proteins implicated in human leukaemia (Valge-Archer *et al*, 1994; Wadman *et al*, 1994). It was not clear whether this association involved direct contact or was mediated by other factors.

Individual LIM domains within a protein can show distinct partner preferences. Zyxin has three LIM domains, and is known to interact with several proteins including ones with SH3 domains, alpha-actinin and members of the CRP family (Crawford *et al,* 1992; Schmeichel and Beckerle, 1994; Beckerle, 1997). By using fusion constructs, it was demonstrated that sequences containing the first LIM show substantial binding to CRP; those missing that LIM domain exhibited drastically reduced binding even though the other two LIMs were intact (Schmeichel and Beckerle, 1994).

Enigma is a LIM protein that interacts with exon 16 of the human insulin receptor. Using the two hybrid system it was shown that the second LIM of Enigma reacts with the insulin receptor, but the first does not (Wu and Gill, 1994). Arber and Caroni (1996) also noted high specificity amongst LIM domains and their partners. They proposed that the LIM domain may be a specific adapter element and suggested that LIM only proteins could function as linkage elements to promote the assembly of specific multicomponent complexes.

Drosophila LIM proteins have also been described (reviewed by Dawid, 1998). The LIM-HD protein Apterous (AP) has a proline rich domain between the LIM domains and the homeodomain (Cohen *et al*, 1992). It has been shown to play a role in dorsoventral compartmentalisation in the wing imaginal disc (Cohen *et al*, 1992; Blair *et*

al, 1994). Two other *Drosophila* LIM-HD genes, *lim-3* and *islet*, appear to function as part of a combinatorial code to control the specificity of motor neurone pathway selection (Thor *et al*, 1999). Two muscle specific LIM proteins, both with glycine rich regions, have been found to show significant colocalisation with actin bundles (Stronach *et al*, 1996).

Some LIM-domain binding (LDB) factors have been identified. Mouse LDB1 binds to the LIM-HD protein LHX-1 in a highly specific way and both intact LIM domains are required for strong binding (Agulnick *et al*, 1996). A double immunoprecipitation procedure demonstrated that homodimeric and heterodimeric association between some LIM domain transcription factors depends on LD1 (also called NLI, nuclear LIM domain interactor) (Jurata *et al*, 1997). LD1 independent associations between other LIM proteins were also found.

The *Drosophila* protein Chip may activate *ap* function and is homologous to the vertebrate LIM domain binding proteins but has in addition a proline-rich amino-terminal domain (Morcillo *et al*, 1997). Mutations in *Beadex* (*Bx)* cause abnormalities in the *Drosophila* wing resembling those produced by reducing *ap* function. *Bx* encodes the LIM only protein dLMO and evidence suggests that dLMO interferes with the formation of the active AP-Chip complex by competing with AP for binding to Chip (Milán *et al,* 1998). The 3' untranslated region of *Dlmo* contains various motifs including ARE (AT-rich elements) that have been implicated in post-transcriptional regulation of genes (Shoresh *et al*, 1998).

O'Keefe *et al* (1998) fused the N-terminal half of *Drosophila* LIM3, another homeodomain protein, to the C terminal half of AP. The LIM domains of AP and LIM3 are 37% identical. This construct rescued the wing phenotype of *ap* to the same extent as full length AP, but only partially rescued the axon pathfinding pathway. This is evidence of functional differences between LIM domains. Interactions specific to the AP LIM domains may be required for full AP function.

4.1.2 Other zinc fingers

Several other classes of zinc fingers are known, all of which bind zinc via a motif containing conserved cysteine and/or histidine residues. Kaptein (1991) described four classes including the original TFIIIA-like zinc fingers and others apart from LIM domains have been identified since then. For example, the erythroid transcription factor GATA-1 has two fingers, each with the consensus sequence $CX_2CX_{17}CX_2C$. The N-terminal finger is not essential for binding to DNA (Omichinski *et al*, 1993). There is evidence that the GATA-1 protein can self-associate in a way that affects gene expression (Yang and Evans, 1995).

The rodent protein CRIP (cysteine-rich intestinal protein) contains the motif $CX_2CX_{17}HX_2CX_2CX_2CX_{17}CXHXC$, which is similar to a LIM domain apart from the spacing of the last five residues (Birkenmeier and Gordon, 1986). CRP2, from rat brain, shows another variation in this part of the pattern, ending CHXXC (Okano *et al*, 1993). The related human protein hCRP contains the classical LIM domain motif (Liebhaber *et al*, 1990). These proteins have a glycine-rich region immediately after the finger domain.

Lovering *et al* (1993) described a zinc finger found in the human RING1 gene, which they therefore called the RING finger. This has the consensus $CXX_{hy}CX_{9-27}CX_{1-3}HXX_{hyd}CX_2CX_{hyd}X_{5-47}CX_2C$ where X is any residue and X_{hyd} is any hydrophobic residue. This motif was also shown to preferentially bind zinc.

Another so-called C_3HC_4 motif was identified in an alpha virus protein family (Everett *et al*, 1993). This motif is unusual because it appears that cysteine residues I, II, IV and V co-ordinate one zinc atom, while the histidine and cysteines III, VI and VII co-ordinate the second. Normally the zinc atoms are bound by consecutive cysteine/histidine residues.

NMR (nuclear magnetic resonance) studies have been carried out on some zinc fingers to try and determine their structure (for example, Pavletich and Pabo, 1991; Schwabe and Rhodes, 1991; Omichinski *et al*, 1993). From these studies and other similar ones, it is clear that there are various ways of folding zinc fingers and that

conserved residues other than the defining cysteines and histidines are important in structure and recognition.

Unlike LIM domains, most zinc finger motifs appear to be involved in DNA and RNA binding (reviewed by Klug, 1993). A DNA binding domain is necessary but not sufficient for transcriptional activation (Mitchell and Tjian, 1989). Other DNA binding domains include homeodomains and leucine zippers.

4.1.3 LIM domains and the cytoskeleton

Many LIM domain proteins are associated with the cytoskeleton. Zyxin is one of many proteins associated with adhesion plaques, where connection between the extracellular matrix and the actin cytoskeleton is established. It has proline and serine rich regions and a nuclear export signal as well as three LIM domains (Macalma *et al,* 1996; Beckerle, 1997). Proline-rich arrays may form structures with limited conformational flexibility and can also serve as high affinity binding sites for protein ligands such as SH3 domains. Alexandropoulos *et al* (1995) distinguished four types of proline-rich sites that bound SH3 domains, all of which included the motif PXXP. The WW domain has a high content of proline residues and also binds proline-rich peptides (Macias *et al*, 1996). It may compete for the same sequence motifs as SH3 domains.

Attachment to the substratum depends on integrins, whereas cell-cell adhesion is mediated by other families of receptors including cadherins (Luna and Hitt, 1992). Brown (1993) suggested that one advantage in adhesion of two cell layers via the extracellular matrix could be that the physical barrier prevents intermingling of the two cell layers. Hence integrins would mediate the adhesion of tissues to each other by binding to the matrix. Integrins recognise both extracellular matrix and cell surface ligands, and each may bind many different ligands (Humphries *et al*, 1993).

Cadherins mediate calcium dependent adhesion between cells (Luna and Hitt, 1992). E-cadherin is one of the components of adherens junctions that is highly conserved between vertebrates and insects (Knust and Leptin, 1996). The cytoplasmic portion of E-cadherin is associated with a variety of cytoplasmic proteins, including

ß-catenin, alpha-catenin, vinculin and alpha-actinin.

Paxillin is another LIM domain protein closely associated with focal adhesions. It has four LIM-like domains, although only the third of these conforms exactly to the LIM consensus sequence (Salgia *et al*, 1995; Turner and Miller, 1994). The others have a cysteine to histidine substitution at the third cysteine residue, so it is unclear whether they are LIM domains, RING fingers or a novel domain. Paxillin also contains putative SH2 and SH3 domain binding motifs (Turner and Miller, 1994). SH2 domains are found in a diverse group of cytoplasmic proteins and are frequently associated with SH3 domains. Both SH2 and SH3 domains are implicated in the regulation of protein-protein interactions during signal transduction (Pawson and Gish, 1992). SH3 domains have been identified in a variety of proteins associated with the cytoskeleton and membrane so are thought to be involved in subcellular localisation (Koch *et al*, 1991). Paxillin associates *in vitro* with the SH3 domain of the tyrosine kinase $pp60^{v\text{-}src}$ and evidence suggests that protein tyrosine phosphorylation may be important in events accompanying cell attachment (Turner, 1994). Paxillin binds to a number of other proteins, including tensin, which has an SH2 domain (Lo *et al*, 1994) and vinculin, a component of cell-substratum and cell-cell junctions which has a very proline rich region (Coutu and Craig, 1988).

Members of the RAS superfamily of GTPases appear to play a role in cell adhesion and organisation of the actin cytoskeleton, as well as being involved in planar polarity (see Chapter 1). RHO, RAC and CDC42 all have several roles and each have many target proteins (reviewed by Tapon and Hall, 1997). A conserved binding motif has been identified in many RAC and CDC42 targets (Burbelo *et al*, 1995). When RHO is activated in fibroblasts, stress fibres and focal adhesion complexes are assembled and RHO activity is required to maintain focal adhesions in attached cells (Hall, 1998). RAC and CDC42 induce the actin structures associated with integrin adhesion complexes (*ibid*).

Some of the many components of various RAS signalling pathways have now been identified. For example, CNK (Connector enhancer of Kinase suppresser of RAS)

is a multidomain protein that contains several protein-protein interaction domains including putative SH3 binding sites (Therrien *et al*, 1998). It interacts with RAF and is thought to be a tyrosine phosphorylation target. It has also been shown that the protein kinase PAK3, which is regulated by RAC and CDC42, phosphorylates RAF-1 on serine residue 338 (King *et al*, 1998).

Recently, LIM kinase proteins were implicated in the phosphorylation of cofilin, a ubiquitous actin-binding protein (Arber *et al*, 1998; Yang *et al*, 1998). LIMK-1 (LIM kinase-1), a protein kinase with two LIM domains, was shown to promote substantial phosphorylation of cofilin (Arber *et al*, 1998). Expression of a LIMK-1 construct lacking the LIM domains induced dramatic accumulation of the actin cytoskeleton into large clumps. Yang *et al* (1998) found that expression of a kinase-inactive form of LIMK-1 significantly repressed RAC-induced lamellipodium formation, but had no apparent effect on RHO-induced stress fibre formation or on CDC42-induced actin reorganisation. Their results indicate a link between RAC, LIMK-1, cofilin and actin-filament dynamics, and they suggest that LIMK-1 participates in RAC-mediated actin cytoskeletal reorganisation by phosphorylating cofilin.

4.1.4 LIM domains and PK

PK has three LIM-like motifs but only the first one is a classical LIM domain. The second motif has a cysteine to histidine substitution at the third cysteine residue, as described for paxillin (Turner and Miller, 1994). The third domain has the same conserved residues as a LIM domain but the spacing of putative zinc binding ligands is different (Table 4.1). The LIM domains overall are more similar to those of paxillin and zyxin than other known LIM domains, and therefore would probably be classed as type D. However, many LIM domain proteins have been described since this classification was proposed (Dawid *et al*, 1995) and so it may be necessary to expand the number of classes to cover all the variations. Alternatively the classification could be dropped altogether as it serves little purpose and the classification of the LIM proteins (Taira *et al*, 1995; Dawid *et al*, 1998) expanded, as this gives more indication of possible functions.

The functional significance of these differences is not known. The three domains will still be termed LIM motifs for convenience. A comparison between the LIM domains of PK and those of paxillin and zyxin is shown in Table 4.2.

LIM consensus sequence*:	$CX_2CX_{17\text{-}19}$	$HX_2CX_2CX_2CX_{16\text{-}20}$	$CX_2D/H/C$
PK LIM1:	: CX_2CX_{22}	$HX_2CX_2CX_2CX_{17}$	CX_2H
PK LIM2	: CX_2CX_{17}	$HX_2\mathbf{H}X_2CX_2CX_{17}$	CX_2C
PK LIM3	: CX_2CX_{18}	$H\mathbf{X_4}CX_2CX2CX_{17}$	$C\mathbf{X_3}C$

Table 4.1. Comparison of the PK LIM-like motifs to the LIM consensus sequence. (* as defined by Dawid *et al*, 1998. Other authors include up to 23 residues between the second cysteine and the histidine, and between the sixth and seventh zinc binding ligands; see, for example, Arber and Caroni, 1996). The gaps in the PK LIMs are for ease of comparison only and do not exist in the sequence. The deviations from the LIM consensus are in **bold type.**

4.2 Identification of a novel domain

The *pk* gene product does not fit the criteria for any of the described groups of LIM proteins (Dawid *et al,* 1998). The proteins with the highest homology to PK in our original searches were two murine testins, probably splice variants of the same gene (Divecha and Charleston, 1995; SwissProt accession no. P47226). The second and third LIM domains of these triple LIM proteins have the same variations as the second and third LIM domains, respectively, of PK. An additional region of homology was also found between testin and PK. This region of about 90 amino acids shows 41% identity (49% similarity) between the two proteins (Table 4.3). It is located upstream of the LIM domains; there are six amino acids between it and the first LIM in PK, but 37 in the murine protein (Figure 4.1). The function of the murine protein is not known. A human EST tag sequence (clone 241420, I.M.A.G.E. consortium LLNL, unpublished; GenBank accession no. H90439) with homology to the novel domain was also located.

Subsequently a human triple LIM protein with this domain was described (Fisher *et al,* 1997; GenBank accession number AJ011654). The degree of relatedness to PK was higher than for the murine protein, in terms of both amino acid homology (Table 4.3) and

Table 4.2 Comparison of the LIM domains of Prickle (pk) with those of human paxillin (pax) (Salgia *et al*, 1995) and human zyxin (zyx) (Macalma *et al*, 1996). Residues identical in all domains are boxed; those identical to pkLIM1 are shaded. Comparison generated with SeqVu 1.0 (Garvan Institute of Medical Research); initial alignments done with Clustalw and BestFit.

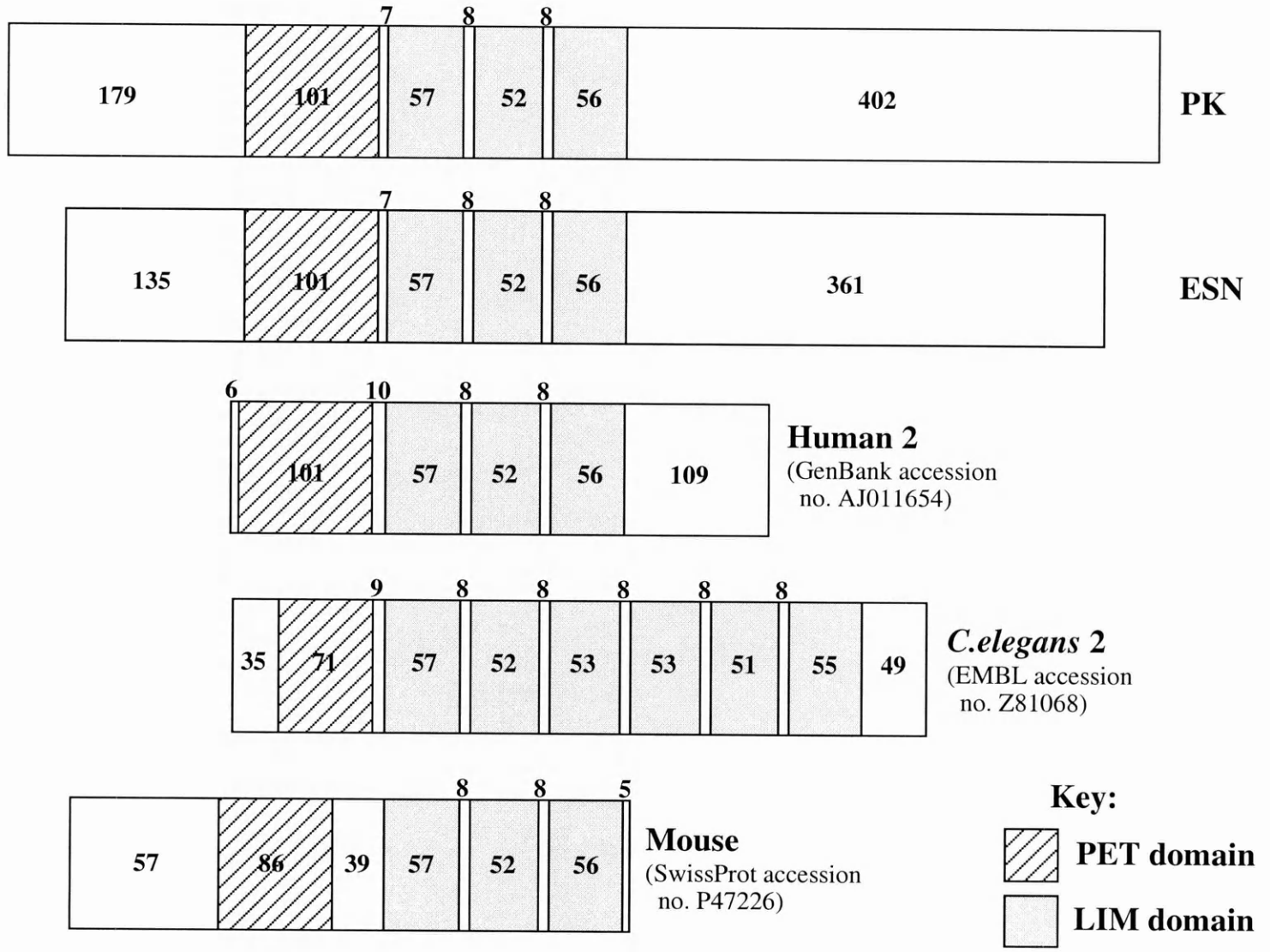


Fig. 4.1 Alignment of domains in PK and related proteins. (See Table 4.3 for details of proteins). Numbers given are the number of amino acids per domain.

overall composition (Figure 4.1). The domain has since been identified in a *Drosophila* EST sequence (cDNA clone GH02642, BDGP/HHMI Drosophila EST Project; accession no. AI063139, unpublished) and two *Caenorhabditis elegans* clones (cosmid ZK381.7 and clone F25H5 from cosmid ZK381; EMBL accession nos. U52003 and Z81068, respectively; both Wilson *et al*, 1994).

It was not possible to determine from the database if the human tag sequence and one of the *C. elegans* clones (cosmid ZK381.7) have LIM domains. The other *C. elegans* clone (F25H5) has six LIM domains. The first three are similar to the three LIM domains of the other proteins except that the third one does not have the unusual spacing of the other third LIM domains (see Table 4.4b). The *pk* related gene *espinas* (see section 4.4) shows strong homology to PK throughout the novel and LIM domain regions. We have named the novel domain the PET (Prickle, Espinas, Testin) domain.

	PET domain	**LIM domains**
		Prickle:
Espinas	87% (92%)	81% (85%)
Human 1	72% (88%)	-
Human 2	71% (81%)	64% (71%)
***C. elegans* 1**	53% (64%)	-
***C. elegans* 2**	47% (51%)	47% (59%)
Mouse	41% (49%)	42% (54%)
Drosophila	38% (47%)	-
		Human 1:
Human 2	73% (88%)	-
		***C.elegans* 1:**
***C.elegans* 2**	43% (52%)	-

Table 4.3 Identity (similarity) of the amino acid sequence of the PET domain and the LIM motifs. **Prickle** and **Espinas** as given in text; **Human 1**-human EST sequence (clone 241420, IMAGE consortium LLNL, unpublished); **Human 2**-human triple LIM protein (Fisher *et al*, 1997); ***C. elegans* 1**-*C. elegans* cosmid ZK381.7 (Wilson *et al*, 1994); ***C. elegans* 2**-*C. elegans* clone F25H5 from cosmid ZK381 (Wilson *et al*, 1994); **Mouse**-murine triple LIM protein (Divecha and Charleston, 1995); ***Drosophila*** -cDNA clone GH02642 (BDGP/HHMI Drosophila EST Project; unpublished). Percentages calculated using BestFit (GCG).

Table 4.4a

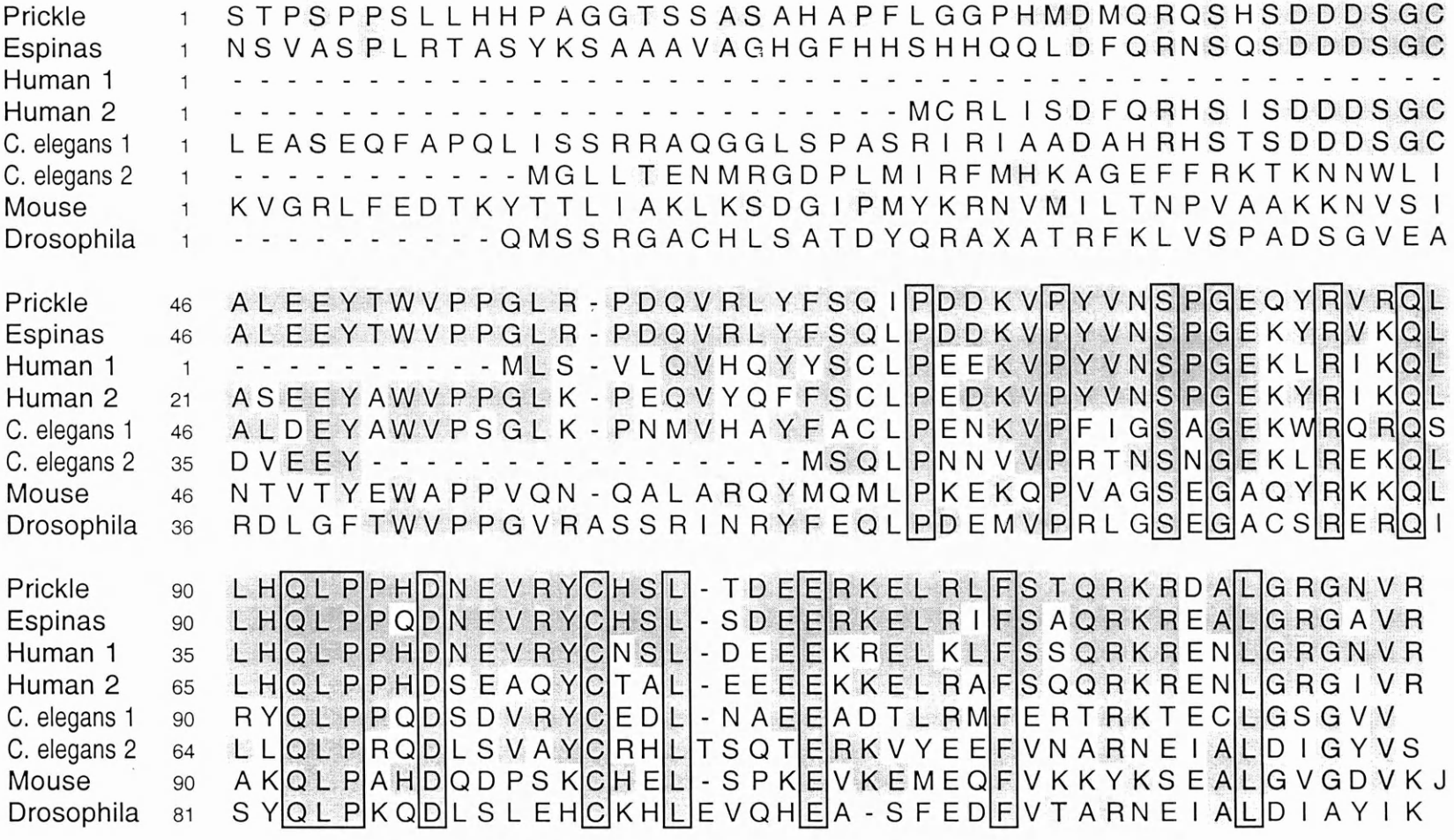

Prickle 1 STPSPPSLLHHPAGGTSSASAHAPFLGGPHMDMQRQSHSDDDSGC
Espinas 1 NSVASPLRTASYKSAAAVAGHGFHHSHHQQLDFQRNSQSDDDSGC
Human 1 1 ----------
Human 2 1 ----------MCRLISDFQRHSISDDDSGC
C. elegans 1 1 LEASEQFAPQLISSRRAQGGLSPASRIRIAADAHRHSTSDDDSGC
C. elegans 2 1 ----------MGLLTENMRGDPLMIRFMHKAGEFFRKTKNNWLI
Mouse 1 KVGRLFEDTKYTTLIAKLKSDGIPMYKRNVMILTNPVAAKKNVSI
Drosophila 1 ----------QMSSRGACHLSATDYQRAXATRFKLVSPADSGVEA
Prickle 46 ALEEYTWVPPGLR-PDQVRLYFSQIPDDKVPYVNSPGEQYRVRQL
Espinas 46 ALEEYTWVPPGLR-PDQVRLYFSQLPDDKVPYVNSPGEKYRVKQL
Human 1 1 ----------MLS-VLQVHQYYSCLPEEKVPYVNSPGEKLRIKQL
Human 2 21 ASEEYAWVPPGLK-PEQVYQFFSCLPEDKVPYVNSPGEKYRIKQL
C. elegans 1 46 ALDEYAWVPSGLK-PNMVHAYFACLPENKVPFIGSAGEKWRQRQS
C. elegans 2 35 DVEEY----------MSQLPNNVVPRTNSNGEKLREKQL
Mouse 46 NTVTYEWAPPVQN-QALARQYMQMLPKEKQPVAGSEGAQYRKKQL
Drosophila 36 RDLGFTWVPPGVRASSRINRYFEQLPDEMVPRLGSEGACSRERQI
Prickle 90 LHQLPPHDNEVRYCHSL-TDEERKELRLFSTQRKRDALGRGNVR
Espinas 90 LHQLPPQDNEVRYCHSL-SDEERKELRIFSAQRKREALGRGAVR
Human 1 35 LHQLPPHDNEVRYCNSL-DEEEKRELKLFSSQRKRENLGRGNVR
Human 2 65 LHQLPPHDSEAQYCTAL-EEEEKKELRAFSQQRKRENLGRGIVR
C. elegans 1 90 RYQLPPQDSDVRYCEDL-NAEEADTLRMFERTRKTECLGSGVV
C. elegans 2 64 LLQLPRQDLSVAYCRHLTSQTERKVYEEFVNARNEIALDIGYVS
Mouse 90 AKQLPAHDQDPSKCHEL-SPKEVKEMEQFVKKYKSEALGVGDVKJ
Drosophila 81 SYQLPKQDLSLEHCKHLEVQHEA-SFEDFVTARNEIALDIAYIK

Table 4.4b

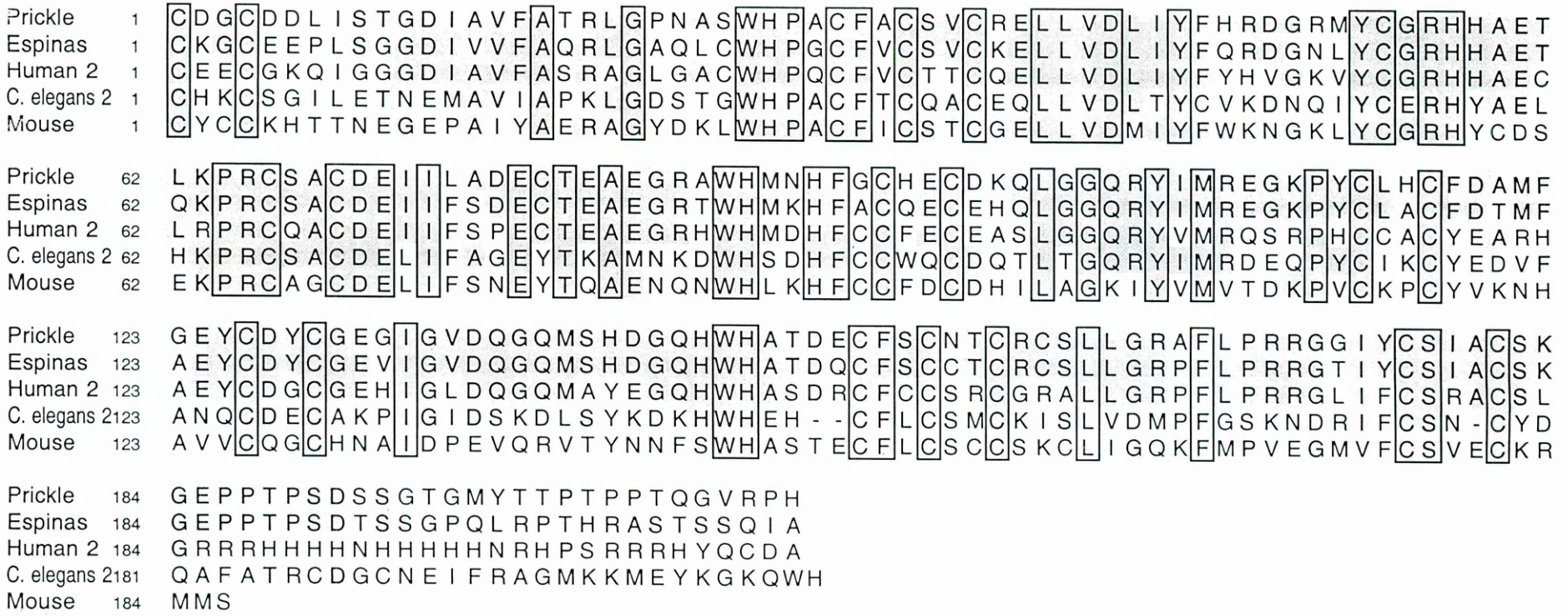


Table 4.4 Comparison of (a) the PET domain (plus some upstream sequence) and (b) the LIM region (plus some downstream sequence) in related proteins. Proteins as for Table 4.3. Residues identical in all domains are boxed; those identical to Prickle are shaded. Comparison generated with SeqVu 1.0 (Garvan Institute of Medical Research); initial alignments done with Clustalw and BestFit.

4.3 Human EST sequence

The clone containing the human tag sequence, I.M.A.G.E. Consortium clone 241420, was obtained from Research Genetics to determine if it had any other similarity to PK. It was subcloned into pBluescript SK+ using the *Not*I/*Xho*I sites following standard procedures and sequenced as usual from the Forward and Reverse pBluescript sequencing primers. The sequence was completed using oligonucleotide primers designed for the purpose (Appendix A). The total length of the sequence was 1158 bases and conceptual translation of the longest predicted open reading frame gave a protein of 97 amino acids (Figure 4.2).

Fig.4.2 DNA sequence of the human clone and conceptual translation of the longest ORF. *Restriction enzyme sites* are in italics and the **PET domain** bold type.

```
 NotI
CGGCCGCTTAATTAAGATCTTTTTTTTTTTTTTTTTTTGCCTCTCCTATATAATAACATA   60
CCTTCTGTATTAGCCAGTTTTTTCTCTCTAGTTTTGCATGTGGCTGAATGGTGGAATGTC  120
TGTTTTCTTTTCTCAGGCAAAATTAGCTGCTTGGTGCTTTCCCCAGTTGTTAGGGTTGAA  180
TGAGCCTCATAAAGTACAAGAATATGTCCTAACAGTCGTAAACAGGAATGTTGTGGGATG  240
TGCAAAAATCCTCATGTTAGTGACAATTTTTAATTACATGCTCTGCATCCCAGATTTAGT  300
GAGTCTATTTCTGATGGATGTCAGCTgGCTGCCTGCCTTACAAAAGTCTCGAAACCTGGA  360
GGGCTTTAAATGTAGGTGGAGGCTGAGGTACCTAGGGGTTTGCCGAGTGCCAAAGACTGG  420
CTTTAACTCCATTGCCGAGGCAGGTGCTGACATATCCCTTTGGCTAGGGAGGCACGTCAC  480
ATGCTGGAGACTAACAGAGAAAATGGTAAAGCCAGCTCTCCGGGGAGTCTGCCGAGAAAG  540
                                  EcoRI
CCCTACCCTCTGGTTCACCAGCATGTCTGCTGAATTCTCTGTTTCTAGGAAAATTTACTA  600
GCTTTATTTTTCCCTCACAGAGAGAGGGAAAAATCCCAGATATGGATGATTTGCCTGACT  660
CTTGCTTTAAGGATGAATTTCTCTGCTTACCATAGGAAACTCATCTGCTGTTGACTTTCA  720
TAAAGAGTCAGGACTGCAGAAGTCAGTTACATCCTGCTGGGGGGAATCATCTTCCCAGTAA  780
GTGGGAATAGCACATGGTTTGAGTAATGGGCAAGTCATCTATATGCTTTCTGTCCTTCAG  840
                                          M  L  S  V  L  Q
GTACACCAGTACTATAGCTGTCTCCCAGAAGAGAAAGTCCCTTATGTCAACAGTCCTGGA  900
V  H  Q  Y  Y  S  C  L  P  E  E  K  V  P  Y  V  N  S  P  G
GAGAAACTGCGAATCAAGCAGCTACTACACCAGCTGCCGCCACATGACAATGAGGTTCGA  960
E  K  L  R  I  K  Q  L  L  H  Q  L  P  P  H  D  N  E  V  R
                                              HindIII
TATTGCAACTCCCTGGATGAGGAAGAGAAGAGGGAGCTGAAGCTTTTCAGCAGCCAGAGG 1020
Y  C  N  S  L  D  E  E  E  E  K  R  E  L  K  L  F  S  S  Q  R
AAACGCGAAAACTTGGGCCGCGGGAATGTCAGGCCTTTCCCAGTCACCATGACAGGAGCT 1080
K  R  E  N  L  G  R  G  N  V  R  P  F  P  V  T  M  T  G  A
ATTTGTGAACAGGTAAGCATGGATTCAGGGTATTGAAGCTCAGGTCTAGCTCTCTCGTGC 1140
I  C  E  Q  V  S  M  D  S  G  Y  *
   EcoRI        XhoI
CGAATTCTTGGCCTCGAG                                             1158
```

If this is the complete protein, then there are no LIM domains. However the cDNA may not be full length. The homology of the PET domain with PK is very high (see Table 4.3 and Figure 4.1).

4.4 The *pk*-related gene *espinas*

4.4.1 Background

A new transcript was cloned from the embryonic cDNA library (Brown and Kafatos, 1988) that had a high level of homology to the *pk* transcripts, on the basis of Southern hybridisation data (David Gubb, personal communication). This 4.2 kb transcript, named *espinas* (*esn*), mapped within a 45 kb cosmid that overlapped with phage FP11/3 of our walk and hence was mapped proximal to the serpin cluster. I sequenced *esn* to determine whether the homology was throughout the transcript or only within specific regions.

4.4.2 Analysis

The transcript was subcloned into pBluescript SK+ using the *Eco*RI/*Hin*dIII restriction sites and a restriction map was constructed (Figure 4.3).

A Southern blot of *esn* cut with a number of restriction enzymes was probed with the *pk* cDNA. The probe hybridised to bands within the *Pst*I/*Bst*XI 1.5 kb fragment (Figures 4.3 and 4.4).

Several subclones were produced in pBluescript SK+ (Appendix B) and these, plus the whole insert, were sequenced using Forward and Reverse primers following the standard protocol. The sequence was assembled using the GCG package as before. Gaps in the sequence were filled using oligonucleotide primers designed for the purpose.

Analysis of the sequence was carried out as previously described. The whole sequence is 4319 bases long and the putative start site encoded by bases 898 to 900 is embedded in good *Drosophila* start site consensus sequence (Cavener and Ray, 1991).

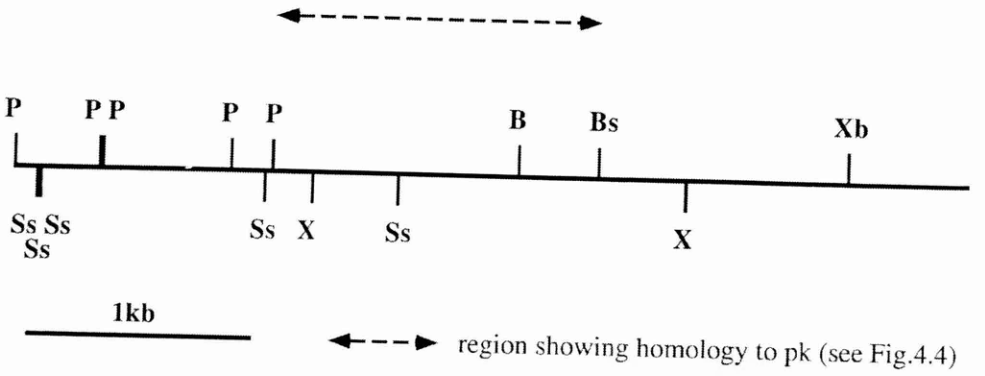


Fig. 4.3 Restriction enzyme map of *espinas* cDNA. Enzymes: *Pst*I (P); *Sst*I (S); *Xho*I (X); *Bam*HI (B); *Bst*XI (Bs); *Xba*I (Xb).

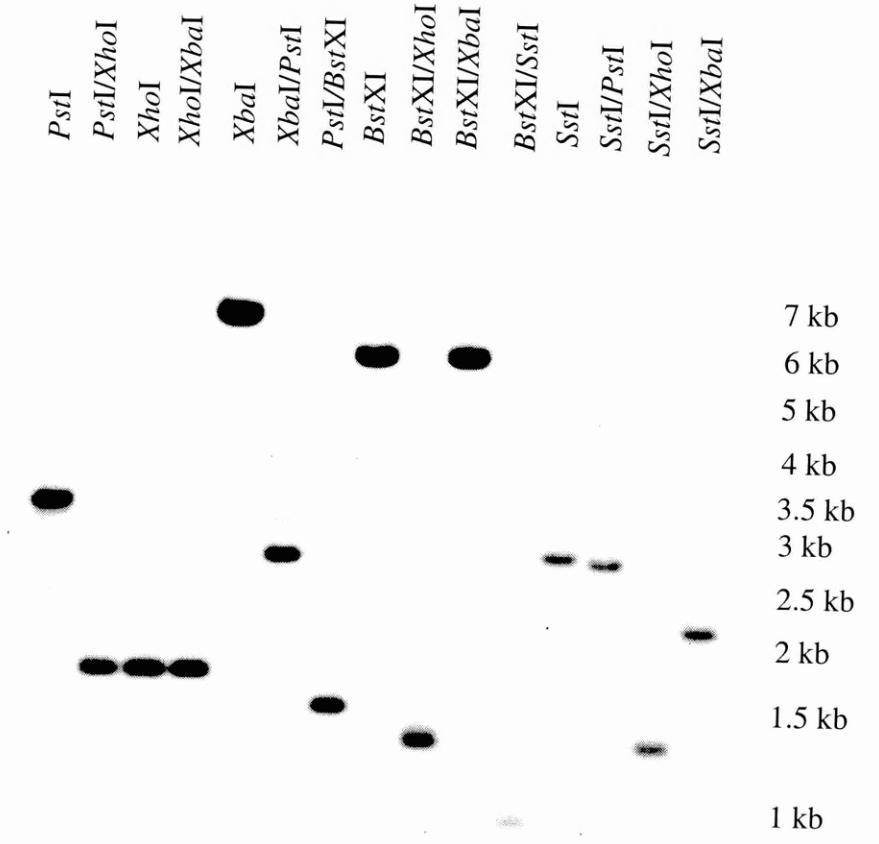


Fig. 4.4 Southern blot of restriction digests of *espinas* cDNA probed with *prickle* cDNA. The filter was washed at high stringency (0.1xSSC, 65C) and exposed to film for 3hrs.

The conceptual translation of the longest predicted ORF (Figure 4.5) therefore gives a protein of 785 amino acids (Figure 4.6). Database searches identified homologies to LIM proteins and also to the PET domain. No other known motifs were found.

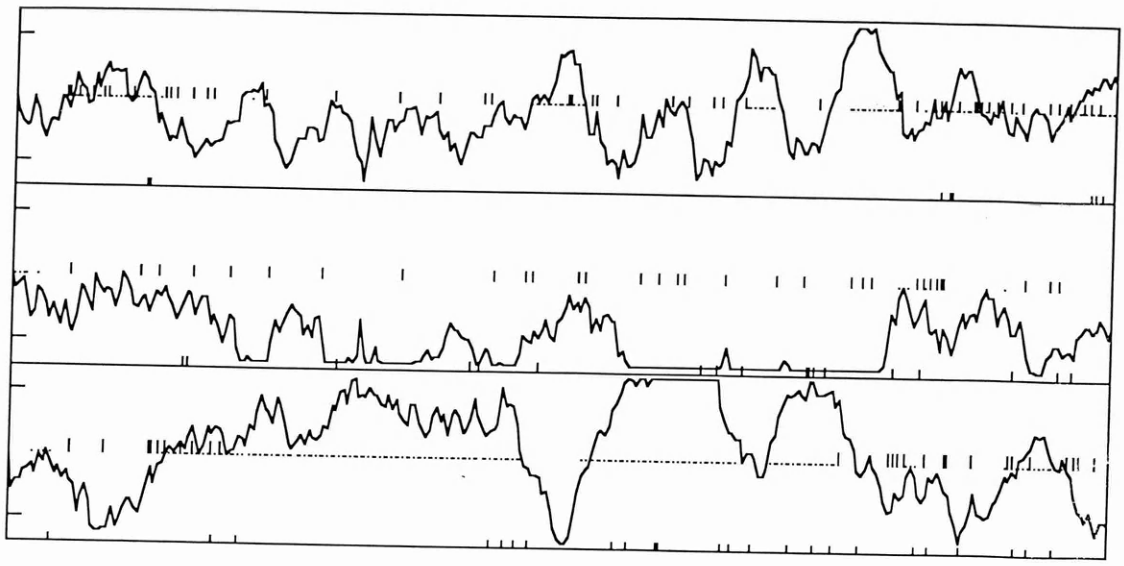


Fig. 4.5. Open reading frame prediction for *esn* cDNA (forward frames only). Created with NIP in the Staden package (see Figure 2.9 for details). In this case, the most likely coding frame is the bottom one.

Fig. 4.6 cDNA sequence of *esn* and conceptual translation of its longest ORF. *Restriction enzyme sites* are in italics; the **PET domain** is in bold type and the LIM motifs are underlined.

```
                                                                    PstI
CCGAGTGCGGAAAGCAAGCTCGACGCGAGCAGAACTCGATCGCGAACTCTCGGACTGCAG        60
                                                        SstI        SstI      SstI
AACCCCGATCGCAGAAACGGGTCGGGACCcAGAGATCCAGAGCTCCAGAGCTCCAGAGCT       120
CCAGAGTTCCAGAGTTTCAGAGTTGGGACGAGTGCGGATTTCGCGATGTGGAAGCGATAC       180
GCTACCGCCTGGTGTAAAGAAGGCATTAAAGTCGCTGCCAAGCTGACAAAACGGCAATAA       240
ATTAAGCCAAAAAAGCGCCGCGGGTCGCGGGGAAAACCAAACGCAAGAACAACAATACGT       300
AGCGAAGAAATTTCGCAGTTTACACGGTTGCAGTTTTCGAAAGGTAACCGAGGTAGCAGC       360
                                  PstI              PstI
CCATAGCCCGCAGACCAGTCTTCTGCAGAGCGTCTTCTGCAGAGCGCGTTCTTCTGGCCA       420
GGTTCAGGTCGTTGGCAGCCGCTGCTGCTGCGTGAGCGGAGAATCGGGCGGGGAAGAAAG       480
AAACGAGCGGCGAAGAGCGTAGAACGCAGAATCGGGTCTCAGGAATGGGAATGCGCCGGG       540
GCCTAGGACTATTATTAGTTTTGCTTACTGGCGGTAATTAAAGCGTGAAAATCACTCTTT       600
AACTTCCCATAAATTTGTGCTTTTAACGCAACATTCCAATCCGCGGCTGCAACTGTGTTG       660
TGTCGGTGATATGTGCGTGTGCGCCACTCAGGTAATGCGTTTTCAATTATAGACTAGCCC       720
CAAATCGCCTCAACTGGGATTAAGCCGATTGGCGGTGCACCGAGAACTGTTGGTGAAACA       780
ATTTGTTAAAACGCAAACTTCGCAATGGCAATTCGTTTGGGAGAGTTTTGAAGCACCCAG       840
CCCAGCCAAGTAATCGAAAAATTCAAGCCCACGCCCGCACCCCGTTCTTCCGGCAACATG       900
                                                                    M
CAGCAGGCCCCGCAGCAGCAGCAGCATCCCCATCCGCCCAGCAGCAGCTACTACACGCAA       960
Q  Q  A  P  Q  Q  Q  Q  H  P  H  P  P  S  S  S  Y  Y  T  Q
                      PstI
ACCGAAAGCGAGCTGCTGCAGATTGAGGCGGGCGGCACGGGCCTGACCTTCGCCTCCCAC      1020
T  E  S  E  L  L  Q  I  E  A  G  G  T  G  L  T  F  A  S  H
TCGCAACGTCCCGAATCGGCGATCAGTCAGGTGGCATCCACCGCCCACCTGGACGTACCC      1080
S  Q  R  P  E  S  A  I  S  Q  V  A  S  T  A  H  L  D  V  P
                                                          SstI
TCGGCCGCCTCCAGCGGCAGCGGGGGCAGTGCAGTCAGTGGGGGAAGCGGCGGAGCTCCG      1140
S  A  A  S  S  G  S  G  G  S  A  V  S  G  G  S  G  G  A  P
                                  PstI
GAATCCGCGGGTCGATTTGTGTCCCCGCTGCAGCGCCGCCACTGCCAGCCGCCCAGCCAC      1200
E  S  A  G  R  F  V  S  P  L  Q  R  R  H  C  Q  P  P  S  H
CTGCCGCTGAACTCGGTGGCCTCCCCGCTCCGCACCGCCAGCTATAAGTCGGCGGCGGCG      1260
L  P  L  N  S  V  A  S  P  L  R  T  A  S  Y  K  S  A  A  A
GTCGCGGGTCATGGCTTCCACCACAGCCATCACCAGCAGCTGGACTTCCAGCGGAACTCG      1320
V  A  G  H  G  F  H  H  S  H  H  Q  Q  L  D  F  Q  R  N  S
                                  XhoI
CAGTCGGACGACGACAGCGGCTGTGCCCTCGAGGAGTACACTTGGGTGCCGCCGGGTCTG      1380
Q  S  D  D  D  S  G  C  A  L  E  E  Y  T  W  V  P  P  G  L
CGGCCCGATCAGGTCCGTTTGTACTTCAGTCAACTGCCAGACGACAAAGTGCCCTACGTC      1440
R  P  D  Q  V  R  L  Y  F  S  Q  L  P  D  D  K  V  P  Y  V
AACAGTCCCGGCGAAAAGTATCGTGTCAAGCAGTTGCTGCACCAGCTGCCGCCGCAAGAT      1500
N  S  P  G  E  K  Y  R  V  K  Q  L  L  H  Q  L  P  P  Q  D
AACGAAGTGCGCTACTGCCACTCGCTGAGCGACGACGAGCGCAAGGAGCTGCGCATCTTC      1560
N  E  V  R  Y  C  H  S  L  S  D  E  E  R  K  E  L  R  I  F
TCGGCCCAAAGGAAGCGGGAGGCCCTCGGACGAGGGGCGGTCCGTCTGCTGTCCGACGAG      1620
S  A  Q  R  K  R  E  A  L  G  R  G  A  V  R  L  L  S  D  E
CGGCCCTGCAAGGGGTGCGAGGAGCCTTTGTCCGGCGGTGACATCGTGGTTTTCGCCCAG      1680
R  P  C  K  G  C  E  E  P  L  S  G  G  D  I  V  V  F  A  Q
                                                          SstI
CGCCTGGGCGCCCAGTTGTGTTGGCATCCGGGCTGCTTTGTGTGCAGCGTCTGCAAGGAG      1740
R  L  G  A  Q  L  C  W  H  P  G  C  F  V  C  S  V  C  K  E
```

*CTC*CTGGTGGACCTCATCTACTTCCAGCGGGACGGGAACCTCTACTGCGGTCGGCACCAT 1800
L L V D L I Y F Q R D G N L Y C G R H H
GCCGAAACCCAGAAGCCACGCTGCTCCGCCTGCGATGAGATCATCTTCTCGGACGAGTGC 1860
A E T Q K P R C S A C D E I I F S D E C
ACTGAGGCGGAGGGCAGAACGTGGCACATGAAGCACTTCGCCTGCCAGGAGTGCGAGCAC 1920
T E A E G R T W H M K H F A C Q E C E H
CAGCTGGGCGGGCAGCGGTACATCATGCGGGAGGGCAAACCGTACTGTCTGGCCTGCTTC 1980
Q L G G Q R Y I M R E G K P Y C L A C F
GACACGATGTTTGCCGAGTACTGCGACTACTGCGGGGAGGTAATCGGCGTGGACCAGGGC 2040
D T M F A E Y C D Y C G E V I G V D Q G
CAGATGAGCCACGACGGGCAGCACTGGCATGCGACGGACCAGTGCTTCAGCTGCTGCACC 2100
Q M S H D G Q H W H A T D Q C F S C C T
TGCCGCTGTTCGCTGCTGGGAAGGCCCTTCCTGCCCAGAAGGGGCACCATCTACTGCTCG 2160
C R C S L L G R P F L P R R G T I Y C S
ATCGCCTGTAGCAAAGGTGAGCCGCCCACGCCGTCAGACACGAGTTCGGGTCCCCAGCTA 2220
I A C S K G E P P T P S D T S S G P Q L
AGGCCGACTCACCGAGCTTCCACCTCCAGCCAAATAGCTAAATCGCCGCGTAGAGGTGGA 2280
R P T H R A S T S S Q I A K S P R R G G
*Bam*HI
GAGCGGGAGC*GGGATCC*GGGACGGAAGGCCCACCACGGTCACCCGAAGGCTACGGGCAGT 2340
E R E R D P G R K A H H G H P K A T G S
GCAGGCGATCTGCTTGAGCGGCAGGAAAGGCAGCGCATGGAGGCCGCGGGAGTGGCGGAC 2400
A G D L L E R Q E R Q R M E A A G V A D
CTACTGCTCGGAGGAGGAGTACCGGGCATGCCTCGCCCAGCCCATCCACCGCCCATCGAC 2460
L L L G G G V P G M P R P A H P P P I D

CTAACCGAGTTGGGGATCAGCCTGGACAACATTTGTGCCGGCGACAAGTCCATCTTCGGG 2520
L T E L G I S L D N I C A G D K S I F G
GACACCCAGACCCTGACCAACTCCATGCCGGACATGCTGCTCTCCAAGGCTGACGACTCG 2580
D T Q T L T N S M P D M L L S K A D D S
CACAGTTACCAGAGCATCGACAAAATCAACCTGAACTCGCCCAGCAACTCTGATCTGACG 2640
H S Y Q S I D K I N L N S P S N S D L T
*Bst*XI
CAGAGCACCCAGGAACTGGCCAA*CGAGCTGG*AGCTGGACAACGAGCCGGTGCGGGAGCTG 2700
Q S T Q E L A N E L E L D N E P V R E L
CCCCACGATGGCTACGAACAGCTCTTCGCTAACAATCGCAACCAGGAGCACCCGGCCGAA 2760
P H D G Y E Q L F A N N R N Q E H P A E
CAATATGACGACGAGCAGCTGGACAACCGACCGATGAAGGAGGTGCGCTTTCACAGCGTC 2820
Q Y D D E Q L D N R P M K E V R F H S V
CAGGACACTATGTCGCGCTCAAAGAGCTACACGGATAACTCCAATGCCCGGCGGCGGAGG 2880
Q D T M S R S K S Y T D N S N A R R R R
AGGCGTCGCAACCAGTCGCGCAGCTCCTCGGAGATGCAGATCAACCAAACGAACCTCCGC 2940
R R R N Q S R S S S E M Q I N Q T N L R
CTGCACAACGCACAAACGCAGGTGGGAACGACTCCCCTCAATCTGCTAAACAATCTGGAC 3000
L H N A Q T Q V G T T P L N L L N N L D
*Xho*I
AACTGCGACGTGGCCAGCATCTGCTCCACCTGCTCATCCAGCTC*CTCGAG*CGACATGGAC 3060
N C D V A S I C S T C S S S S S S D M D
GACTACGTATACCGCCTGCCTGCGCGGAAACACTACGGCGGAGTGAGAGTGGCCTATGTG 3120
D Y V Y R L P A R K H Y G G V R V A Y V
CCCAACGATGCGTTGGCCTATGAGCGCAAGAAGAAGATGGCCCAGGACTCGTCCCTGGCC 3180
P N D A L A Y E R K K K M A Q D S S L A
CCCGGGGCAGGCAATGCATCCGTGGGCGGGGCACCGGCGATCATGCACGAGAGCAAGAAC 3240
P G A G N A S V G G A P A I M H E S K N
TGCACGATTTCCTGACCACCCCCCGTGCTGCCACCCCCGCCCTTCAACCTGAGCAGCTAC 3300
C T I S *

```
TACATCCTGGACACGATCCTGGAGGAGCAGAGTCTGATGCTGCCGGAGAAGAAGCCTGGC 3360
TGCCTGAGGAGGGTGTGGAACTTCTTTGGGCTCGGCAAGCCCCGCCACTCCAACTCCCAG 3420
CACAGCAGCAACGGAAGGCTATACAGCGTGTAGCTGTGGAATGAGGTTTAAGAAAGGTGT 3480
TTTTAGCACGGTTGCCAACAAAGGTTGTTCCATTAAGTACAAAAGATTAGTTCTTCAGAA 3540
CACTTTTTAAAAATGTACATATATACAGATGGGGTAAGCCGTGTTTTTTAAAAATATATC 3600
TTTTTAATGGCAACCGTGGCTGTAAATAACATATCGTCATGTATAATACACTTAATAAGA 3660
AGGCACATATGAGATCAGAGATGAGAGAAATCGTTGTACAAAGCTGAGACTATATCATAT 3720
                                                        XbaI
TTGCATATTCATATACACATGCCCAGTCGGGGGCATTACCATAGGTATAAGTTATTCTAG 3780
AGCTAGATTAGGCGAACAAAACAAACGAAAAGTAGAGACCACAAACAACAAACAACGCAA 3840
TCAACGAACAATAAACTCCAACGAAATCCTATCGAAGGCCAACAAAAGAATCTCCGAAAG 3900
AGTGAAAGTATTATTTGTTTCAGATCTTAGCCAAATGTGCAATAATTTAAAATGGCAACG 3960
CAGAACGCGAAATTAACTCTTCCACCAGCAAAGCGAAACGCAGAGATGTAAAACCATTAG 4020
CATTTGCCACATTTGCGGCTTCTTTTCTTCGAATATTTTGTTAATCAATTCGAATTGAAG 4080
ATTCTTTTGTTTGAGAGTCAAATGTTTTATGTAATTTGCGTGCTTTCTCCGCGGCGTGAA 4140
AGGGAAATTGTTATATAAAAGCATGTGTCCAGAATACGTTTGTAACCAACCAGTGAGAGT 4200
TAAACGAAATCTTTTTTTAAGCTATGCTATATCAACAAGTGTGTATGGAAACATAATCAA 4260
AAGTAAAATATATGTCTGTGTATTTTATTCTTATCAAAAAAAAAAAAAAAAAAAAAAAA 4319
```

The cDNA sequence was compared to genomic sequence from the Berkeley *Drosophila* Genome Project (clones P1 DS02730 and P1 DS07472; accession number AC006469). There are six introns and within the region homologous to *pk* the positions of the introns are conserved (Figure 4.7; compare with Figure 2.12).

4.4.3 Comparison of PK and ESN

PK and ESN are very similar in the PET and LIM domain regions (see Figure 4.1 and Table 4.4a and b). A direct comparison between the two whole sequences using "Blast 2 Sequences" (NCBI) found no homology between either the nucleotide or the amino acid sequences outside of this region except for a short stretch of the nucleotide sequence towards the 3' end, within the coding sequence of *esn* but in the untranslated region of *pk* (Figure 4.8). The sequence is 79% identical over this stretch of approximately 130 bases.

There are no known alleles of *esn*. A deficiency of 43A1-2, $pk^{pk\text{-}sple51}$, deletes *nec*, *pk* and at least part of *esn,* although it is not known how much of *esn* is removed. *Df(2R)pk-sple51* over a deletion which removes *esn* but not *nec* and *pk* has been shown to have a subtle wing polarity phenotype in the region of the posterior cross vein (Darin Coulson, personnal communication). When heterozygous with a deletion that removes *pk*

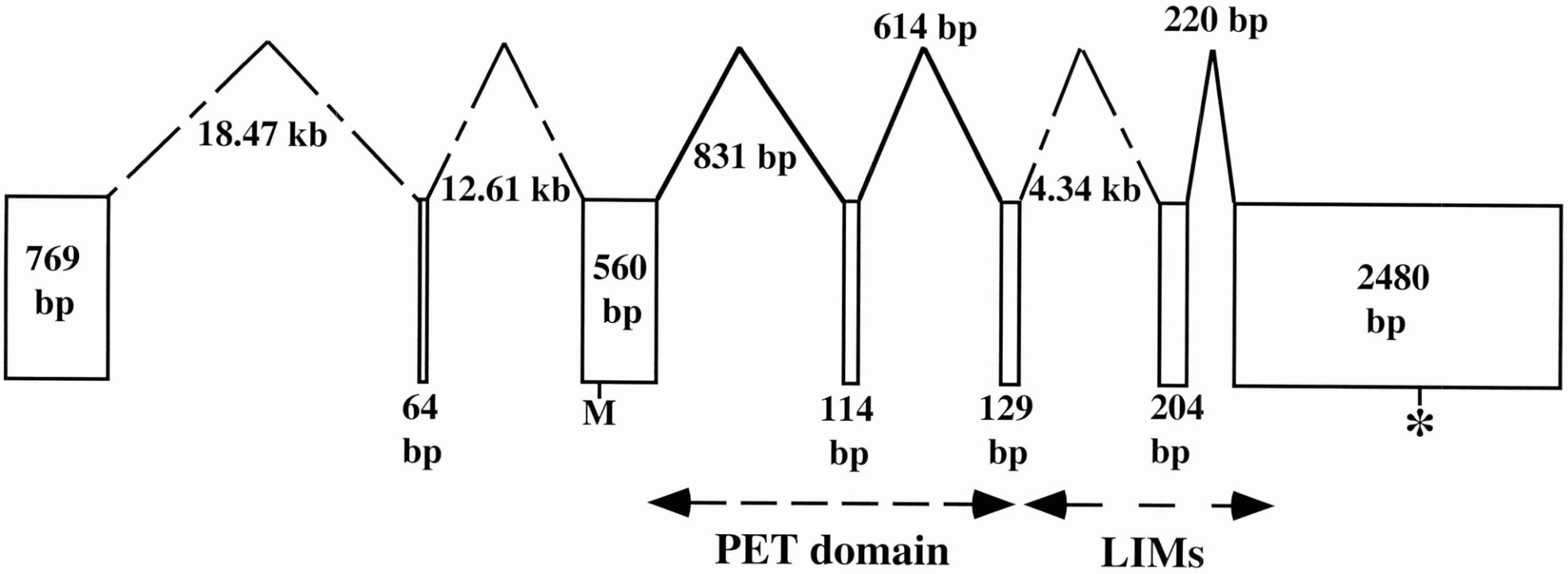


Fig. 4.7 Structure of the *espinas* gene. **M** indicates the start of the coding sequence; *indicates the stop codon. The regions spanned by the PET and LIM domains are indicated.

(*Df(2R) nap-2/Df(2R) pk-sple51*) the flies are viable with a *pk* wing polarity pattern similar to *Df(2R)pk-30*. In addition they have an extreme wing hair duplication phenotype.

```
3416 TGCTCCACCTGCTCCTCGAGCTCCTCCAG...CTCGGAGGACTACATGAT 3462
     ||||||||||||||| || |||||||| ||    |   ||| ||||||
3022 TGCTCCACCTGCTCATCCAGCTCCTCGAGCGACATGGACGACTAC..... 3066

3463 GATGTACCAGCTGCCGCAGAGGCGCCACTACGGCGGCGTGCGCGTGAGCT 3512
       | ||||  |||||   | ||   ||||||||||| ||| | |||  ||
3067 .GTATACCGCCTGCCTGCGCGGAAACACTACGGCGGAGTGAGAGTGGCCT 3115

3513 ATGTGCCCAACGATGCACTGGCCTACGATCGGAAGAGGAA 3552
     ||||||||||||||||  |||||||  ||  ||  ||||  |||
3116 ATGTGCCCAACGATGCGTTGGCCTATGAGCGCAAGAAGAA 3155
```

Fig. 4.8 Comparison of *pk* and *esn* homologous 3' sequence (*pk* top row, *esn* below). (See text for details.)

4.5 Comparison of proteins containing the PET domain

Table 4.4a shows a comparison of all the PET domains that we have identified so far. Several points are clear from studying this comparison. The domain is highly conserved, although some regions more so than others. For example, the sequence "SDDDSGCALEEY" found at the beginning of the domain in PK and ESN is also almost completely conserved in the human triple LIM protein and *C. elegans* cosmid ZK381.7, but is absent from the other proteins. The proteins containing this motif are also more homologous throughout the PET and LIM domains. Within the block of approximately 70 amino acids conserved in all the PET domains identified so far, fifteen residues are completely conserved and a further ten are conserved in six or seven of the proteins.

The novel domain is not especially rich in any particular amino acids, although several of the more conserved residues are prolines, leucines and glutamines. The function of this domain is unknown and no function was identified for either the murine (Divecha and Charleston, 1995) or human (Fisher *et al*, 1997) proteins.

Many residues of the LIM domains apart from those that define the motif are conserved. Some of these conserved residues are common to many LIM domains, for example the tryptophan before the histidine in the first finger. Others are found only in subsets of LIM domains and may help identify the function of the protein. For example, the cysteine to histidine substitution seen in the second LIM domain of these proteins is found in several cytoskeletal proteins. In particular, three of the four LIM domains of paxillin show this change and all four also have a proline after the first conserved histidine seen in the first LIM of PK (Turner and Miller, 1994). The unusual spacing of residues in the third LIM seems to be unique to the group of proteins containing the PET domain. The *C. elegans* clone F25H5 does not have this unusually spaced third LIM and has a total of six LIM domains. From our data, the human EST clone 241420 does not include LIM domains, but this cDNA may be incomplete.

4.6 Characterisation of *pk*$^{pk\text{-}sple}$ point mutants

The PET and LIM domains of some of the double mutants were sequenced to determine whether any changes were seen in these regions. The mutants *pk*$^{pk\text{-}sple7}$, *pk*$^{pk\text{-}sple8}$, *pk*$^{pk\text{-}sple12}$ and *pk*$^{pk\text{-}sple25}$ were selected as they were EMS-induced (Heitzler *et al,* 1993) and therefore likely to be point mutations. The *pk*$^{pk\text{-}sple25}$ allele is unique in giving a strong sple phenotype associated with a weak pk phenotype and the rough eyes typical of other *pk*$^{pk\text{-}sple}$ alleles.

RNA was isolated from these flies using TRIZOL Reagent (section 7.17) and reverse transcribed (section 7.18). The gene specific primer used in the reverse transcription reaction was pkdom1 (Appendix A) which is upstream of the PET domain.

PCR was then carried out on the resulting cDNA using the Expand Long Template PCR system (section 7.15). The primers used were PET1, which is just downstream of pkdom1, and LR3, located downstream of the third LIM domain. Pk4aiii was included as a positive control. A second round was carried out using the same primers. Programme 4 was used for both rounds (section 7.15).

A band of the expected size (1.1 kb) was seen in each of the mutants except *pk*$^{pk\text{-}sple8}$. The other three mutants were sequenced directly from the PCR products, using oligonucleotide primers designed for the purpose (Appendix A). Two separate PCR reactions were carried out for each mutant.

No changes were seen in *pk*$^{pk\text{-}sple12}$ and *pk*$^{pk\text{-}sple25}$. A small deletion of eight base pairs was found in *pk*$^{pk\text{-}sple7}$. This resulted in a frame shift in the middle of the first LIM domain and a premature stop (Figure 4.9). Flies homozygous for *pk*$^{pk\text{-}sple7}$ have a typical pk-sple phenotype (Heitzler *et al*, 1993).

```
172 PFLGGPHMDMQRQSHSDDDSGCALEEYTWVPPGLRPDQVRLYFSQIPDDK 221
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
172 PFLGGPHMDMQRQSHSDDDSGCALEEYTWVPPGLRPDQVRLYFSQIPDDK 221

222 VPYVNSPGEQYRVRQLLHQLPPHDNEVRYCHSLTDEERKELRLFSTQRKR 271
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
222 VPYVNSPGEQYRVRQLLHQLPPHDNEVRYCHSLTDEERKELRLFSTQRKR 271

272 DALGRGNVRQLMSARPCDGCDDLISTGDIAVFATRLGPNASWHPACFACS 321
    |||||||||||||||||||||||||||||||||||||||||
272 DALGRGNVRQLMSARPCDGCDDLISTGDIAVFATRLGPNAQSVLRLLRLP 321

322 VCRELLVDLIYFHRDGRMYCGRHHAETLKPRCSACDEIILADECTEAEGR 371

322 RAPRGPYLLPPGWAHVLRTPPRRDPKAPLLGLR* 354
```

Fig. 4.9 Comparison of part of PK from wild type (top row) and *pk*$^{pk\text{-}sple7}$ (bottom row) flies. **PET domain** in bold type; LIM motifs underlined.

4.7 Summary

In this chapter, LIM domains and related motifs are introduced. The importance of LIM proteins in the cytoskeleton and their possible association with members of the RAS superfamily is discussed. The comparison of PK with other LIM domain proteins reveals that the second and third LIM domains of PK do not conform to the classical LIM consensus sequence. The identification of a novel domain in PK and a small number of other proteins is described. The characterisation of two of these related sequences, a human EST sequence and the *espinas* gene from *Drosophila*, is presented.

CHAPTER 5. THE SERPIN CLUSTER

5.1 Introduction

A class of *pk* mutants also show a necrotic phenotype. Flies that are mutant for the *necrotic (nec)* gene die either as pharate adults or within 24 hours of eclosion (Heitzler *et al*, 1993). The adult flies appear relatively normal at eclosion, but are weak and develop dark patches after a few hours. These patches become more extensive and darker as the flies age.

The *nec* gene maps to chromosomal region 43A1-2 of the right arm of chromosome two, within a region defined by the deletions *Df(2R)sple-D2* and *Df(2R)nap-2*. Transcripts mapping to this region were identified in the process of the phage walk (see Chapter 2). A cluster of three small transcripts were identified within a 10 kb region. Two of the transcripts map within the 5' intron of pk^{pk} and the third is just proximal to the 5' pk^{pk} exon. Overlapping deletions that remove all three transcripts, *Df(2R)sple-D1/Df(2R)nap-2,* or the two most distal transcripts, *Df(2R)sple-D2/Df(2R)nap-2,* express amorphic *pk* and *nec* mutant phenotypes, but are otherwise wild type. David Gubb isolated genomic inserts from the EMBL3 library of John Tamkun. cDNA clones were isolated from the imaginal disc library of Brown and Kafatos (1988) and larval and adult head phage-insert libraries of Russell and Kaiser

(1993). *Spn43Aa* cDNAs were isolated from the imaginal disc library using the 3.25 kb *Eco*R1 fragment from phage *FP11/3*; *Spn43Ab* was isolated from the larval library using the FP10/2 3.2 kb *Sal*1 fragment and *Spn43Ac* was isolated from the head library using FP10/2 *Sal*1 2.1 + 6.0 kb.

5.2 Characterisation of transcripts

5.2.1 Mapping of transcripts

Random prime probes were made from each of the putative cDNAs and hybridised to a map of the phage comprising the walk (section 2.1) to confirm their positions. *Spn43Aa* mapped to FP11/3; *Spn43Ab* and Spn43Ac mapped to FP10/2.

Each of the cDNAs were subcloned into pBluescript SK+ (Stratagene Ltd.). *Spn43Ac* was subcloned as two *Eco*RI fragments as there were no convenient sites to drop the insert out whole. The subcloned cDNAs were restriction mapped (Figure 5.1)

5.2.2 cDNA and genomic sequence

Each cDNA insert was sequenced on both strands as previously described using specific oligonucleotide primers (Appendix A).

Some genomic sequence for the region had already been generated in the region of the 5' *pk* exon. This was extended using subclones (Appendix B) and specifically designed oligonucleotide primers (Appendix A). A total of 11.28kb of genomic sequence was generated.

5.2.3 Analysis of sequence

The most proximal cDNA, *Spn43Aa*, is 1300 nucleotides long, *Spn43Ab* is 1332 nucleotides and *Spn43Ac* is 1523 nucleotides. Comparison of the cDNA and genomic sequences revealed that *Spn43Ac* has two introns while *Spn43Aa* and *Spn43Ab* each have three (Figure 5.1).

Possible open reading frames for each transcript were identified using the Staden package (Figure 5.2). Conceptual translation of the longest open reading frames for *Spn43Aa*, *Spn43Ab* and *Spn43Ac* gave proteins of 370, 394 and 477 amino acids respectively. Figure 5.3 shows the cDNA sequence of each of the transcripts with their conceptual translation. *Spn43Aa* is transcribed in the same direction as *pk*; the two transcripts that are located in the *pk* intron are transcribed in the opposite direction (Figure 5.4).

The *Df(2R)pk-30* deletion removes most of the 5' *pk* exon (see Chapter 2). It also removes the last 361 bases of the *Spn43Ab* transcript, 271 bases of which are coding sequence. The *Df(2R)pk-30* mutation gives an amorphic prickle phenotype, but is otherwise wild type. Lack of the *Spn43Ab* transcript produces no detectable phenotype.

Database searches revealed that all three transcripts have homology to the serpin superfamily. This is a large, divergent group of proteins, most of which function as inhibitors of serine proteinases (see section 5.4).

5.2.4 Temporal expression

Total and poly A plus RNA extractions and Northern blotting experiments were performed by Jean-Marc Reichhart and Elena Levashina as described in Lemaitre *et al,* 1996. Probes corresponding to the cDNA of *Spn43Aa, Spn43Ab* and *Spn43Ac* were amplified by PCR using internal specific primers.

Spn43Aa was found to be expressed predominantly in the pupa up to two days, and also at a lower level in the embryo and in late larval stages. *Spn43Ab* is seen in the larval stages then again in late pupae and the adult. *Spn43Ac* is expressed at all stages from late embryo to adult.

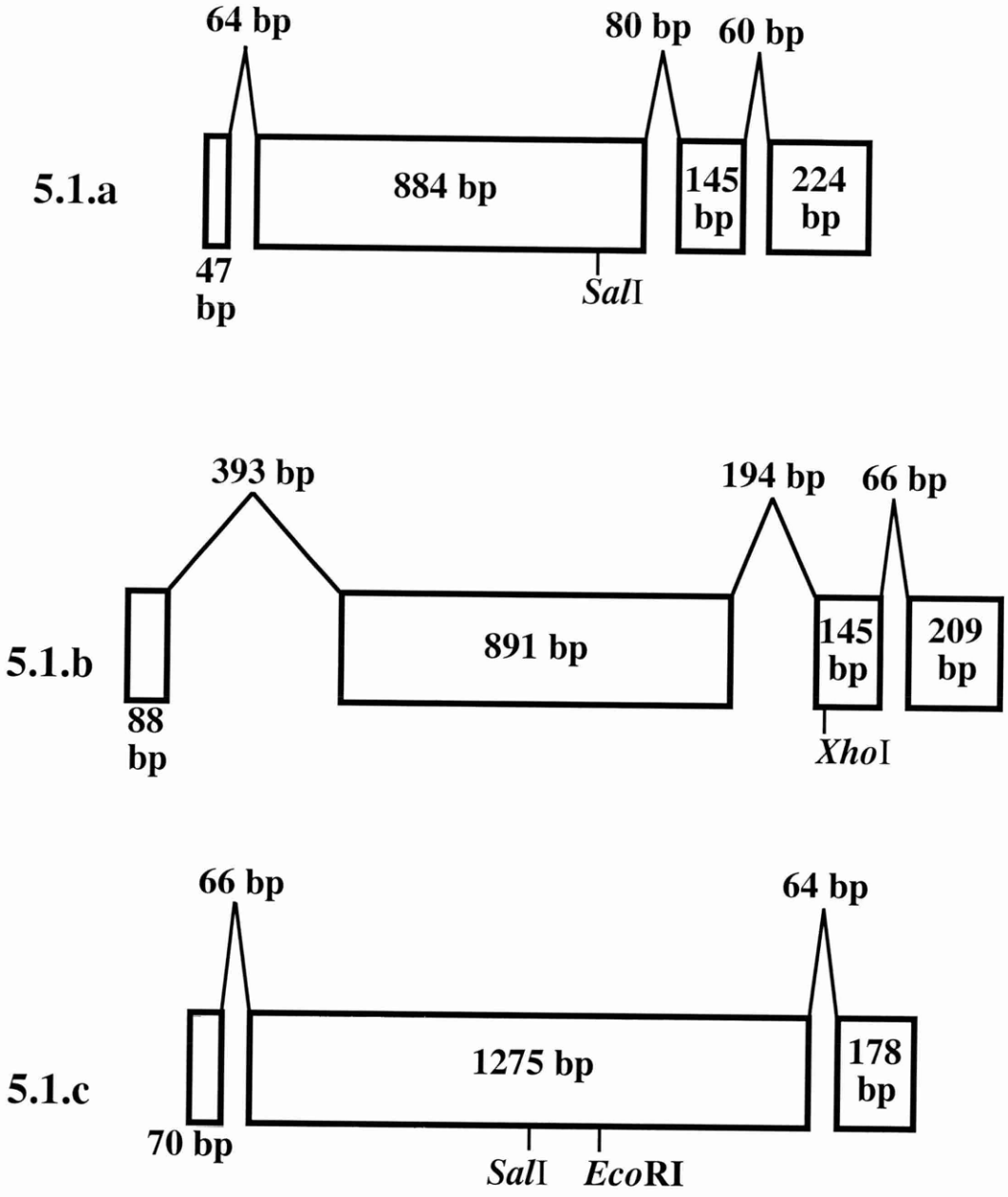


Fig. 5. 1 Structure of the serpin genes. a) *Spn43Aa*; b) *Spn43Ab*; c) *Spn43Ac*.

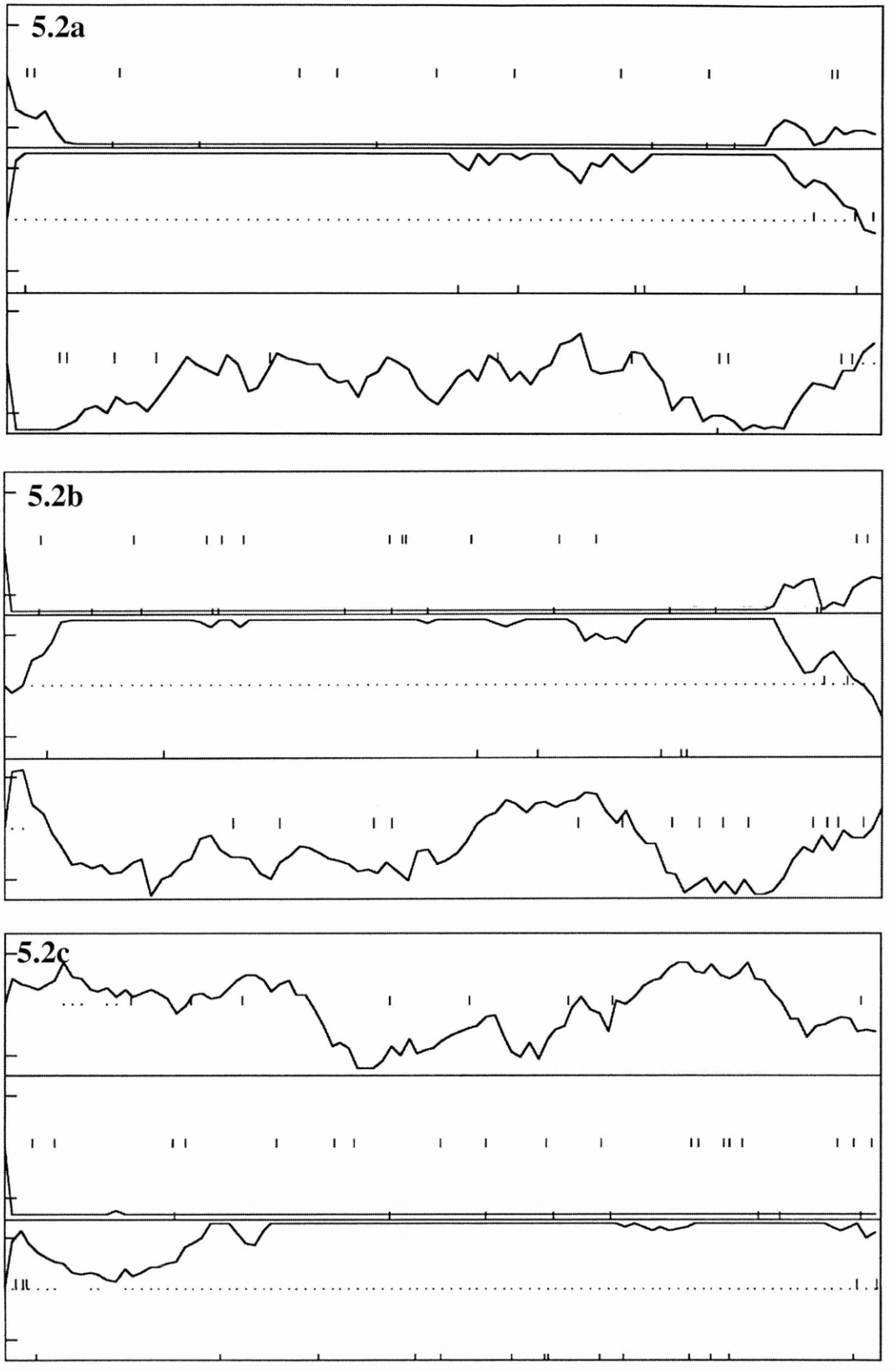


Fig. 5.2 Open reading frame (ORF) predictions for a) *Spn43Aa*; b) *Spn43Ab*; c) *Spn43Ac*. See Fig. 2.9 for explanation. Most likely ORF is in the middle frame for a) and b); in the bottom frame for c).

Fig. 5.3. cDNA sequence and conceptual translation of: **a**) *Spn43Aa*; **b**) *Spn43Ab* and **c**) *Spn43Ac*. *Restriction enzyme sites* are in italics.

Fig. 5.3a

```
      XhoI      SalI
GAGTCTCGAGGTGTCGACGCAGGCGCAATGAACCACTGGCTAAGTATCATTCTTCTTGGG    60
                                  M  N  H  W  L  S  I  I  L  L  G
GTGTGGATATCCGCTCCTGAAGGTCTGGGTAACACGATCAAGGATCGCAATCTCTTCGCC   120
V  W  I  S  A  P  E  G  L  G  N  T  I  K  D  R  N  L  F  A
ACCGAACTCTTCCAAACCCTCGCCACAGATCGCCAGGATGAGAACGTGATCATCTCGCCG   180
T  E  L  F  Q  T  L  A  T  D  R  Q  D  E  N  V  I  I  S  P
GTTTCCATCCAGCTGGCCCTCGGGTTGGCTTACTACGGAGCTGAGGGCAGGACGGCCGCG   240
V  S  I  Q  L  A  L  G  L  A  Y  Y  G  A  E  G  R  T  A  A
GAACTGCAGAAGACCTTGCACGCCTCCGCCAAGGAGAGCAAAGATGGTCTGGCCGAGAGC   300
E  L  Q  K  T  L  H  A  S  A  K  E  S  K  D  G  L  A  E  S
TACCACAACCTGCTGCACTCTTACATCAAGTCCAAGACCGTGTTGGAGATCGCCAACAAG   360
Y  H  N  L  L  H  S  Y  I  K  S  K  T  V  L  E  I  A  N  K
GTGTACACCCGGCAGAATCTCACGGTATCTAGCCACTTCCGAGAGGTGGCCCAGAAGTAC   420
V  Y  T  R  Q  N  L  T  V  S  S  H  F  R  E  V  A  Q  K  Y
TTCGACTCCGAGGTAGAACCACTGGACTTCAGTCGCGAAACGGAGGCCGTGGAGCAGATC   480
F  D  S  E  V  E  P  L  D  F  S  R  E  T  E  A  V  E  Q  I
AACCGCTGGGTGAAGCAGCAGACGGAGAACAAGATCGAACGGGTGGTGGAAAGCCTGGAG   540
N  R  W  V  K  Q  Q  T  E  N  K  I  E  R  V  V  E  S  L  E
CCGGACACCAATGTGGCGCTGGTCAACGCCATCTACTTCAAGGCTCGCTGGGCACGACCC   600
P  D  T  N  V  A  L  V  N  A  I  Y  F  K  A  R  W  A  R  P
TTCAACGACGAGGATACCCGGGATCGCGAGTTCTGGCTGAGCGAGAGCCGGTCCATCCAG   660
F  N  D  E  D  T  R  D  R  E  F  W  L  S  E  S  R  S  I  Q
GTGCCAACCATGTTCGCAGACAACTGGTACTACTACGCCGACTACCCGGAACTCGACGCC   720
V  P  T  M  F  A  D  N  W  Y  Y  Y  A  D  Y  P  E  L  D  A
AAGGCCATTGAGCTGTTCTTCGAAAACATCAACCTGACCATGTGGTTCATCCTGCCCAAC   780
K  A  I  E  L  F  F  E  N  I  N  L  T  M  W  F  I  L  P  N
                                                 SalI
CAGCGTTCCGGACTACAGGCTCTGGAGCAGAAGCTCAAGGGCGTCGACTTCAATCTGCTC   840
Q  R  S  G  L  Q  A  L  E  Q  K  L  K  G  V  D  F  N  L  L
GAAGACCGCTGGCAGTGGCAGAGTGTGTCCGTTTACCTGCCCAAGTTCAAGTTCGAGTTC   900
E  D  R  W  Q  W  Q  S  V  S  V  Y  L  P  K  F  K  F  E  F
GACACGGACCTAAGACCCACATTGCATAAGATGGGAATCAGTGCCATGTTCTCAGATGCA   960
D  T  D  L  R  P  T  L  H  K  M  G  I  S  A  M  F  S  D  A
GCCGACTTCAGCAACATTTTCCAGGACTCGCCCATCGGCACTCGGATCACAAAGGTGCAG  1020
A  D  F  S  N  I  F  Q  D  S  P  I  G  T  R  I  T  K  V  Q
CACAAGACCTTCATCGATGTGAACGAGATCGGATGTGAGGCGGCTGGAGCTAGTTATGCT  1080
H  K  T  F  I  D  V  N  E  I  G  C  E  A  A  G  A  S  Y  A
GCCGGAGTTCCCATGTCCCTGCCCCTGGACCCCAAGACTTTCGTGGCCGATCATCCATTT  1140
A  G  V  P  M  S  L  P  L  D  P  K  T  F  V  A  D  H  P  F
GCGTTCATCATTCGCGACAAGCACGCTGTCTATTTCACCGGACACATTGTCAAGTTTTAA  1200
A  F  I  I  R  D  K  H  A  V  Y  F  T  G  H  I  V  K  F
TCGTCACACATTTCCCTCAGATTAAGCACTTTAAATTGTAATCATTACATCAATAAATAA  1260
ATGCGGAGAGCCTCCGCCAGAACTTAAAAAAAAAAAAAAAA                     1300
```

Fig. 5.3b

*Eco*RI

*GAATTC*GGCTCAAGTGCGAATCAAGCGGAGCAGGAGACGCCATCAACAATTTATGTGACG 60

ACCATGGCTGTCATCATCAGCTGCCTATTACTTCTCCTCGCGACAGTGTCGCAGTCCAAG 120

M A V I I S C L L L L L A T V S Q S K

ACCGTGGGTTACGATGCGGCTGCCGATCGCAATTTGCTGGCTGCCGATCTCTACAACGCC 180

T V G Y D A A A D R N L L A A D L Y N A

GTGGCCGCCGATCATCTGAACGAAAATGTGGTCATCTCGCCAGCGACCATACAGAGTTCC 240

V A A D H L N E N V V I S P A T I Q S S

ATGGCGCTGGCCTTCGTGGGGGCCAAGGGTCAGACGGCATCGGAGTTGCAGCAGGGGCTG 300

M A L A F V G A K G Q T A S E L Q Q G L

CGTTTAGGTCCTGGCGATGCGGATGCGGTGAGCCAGCGCAGCGGTAGTTACCAACAGGCC 360

R L G P G D A D A V S Q R S G S Y Q Q A

CTGACCCGCGACAACAACTTCCGGCTGGCCAACAACATCTACATCAACGAGAACCTTGAG 420

L T R D N N F R L A N N I Y I N E N L E

TTCAAGGGCTCCTTCAGGGACGTGGCCCAGCGCCAGTTCGACTCGAACATCGACAAGCTG 480

F K G S F R D V A Q R Q F D S N I D K L

GACTTTCACCCGCCGTACAACAAGCGCACGGCGGATGGCATCAACCGGGCGGTGGCCACC 540

D F H P P Y N K R T A D G I N R A V A T

AAGACCAACGGCAAGATCACTGACATCCTCCGCGCCGAACTGCTGAATGATCGCACCGAG 600

K T N G K I T D I L R A E L L N D R T E

GGAGTGATCGTGAACGGCGTTTCCTACTCAGCCGCCTGGCAAAATGCCTTCCGGCTGGAC 660

G V I V N G V S Y S A A W Q N A F R L D

AAGACCGAAAAGCGCTCCTTCCGCACCGGAAGCGGGCAGTCCGTCAAGGTAGACACCATG 720

K T E K R S F R T G S G Q S V K V D T M

TGGACGCTGCAGAACTTCAACTACGCCGAGGTCAACTCCTTGGACGCCAAGGTGGTGGAG 780

W T L Q N F N Y A E V N S L D A K V V E

CTGCCCTACCAGAACCCCGACTTCTCCATGCTGCTGCTCCTGCCCAATCGCAAGGATGGT 840

L P Y Q N P D F S M L L L L P N R K D G

CTGAGATCCCTGCAGCAGTCGCTGTCCGGTAAGAATCTTCTGGCCGAAATCGGAGCGTTG 900

L R S L Q Q S L S G K N L L A E I G A L

AGCCAGCAAAAGGTGGAGGTGCTGCTGCCCAAGTTTAGTGTCACCTTTGGCTTGGGTTTG 960

S Q Q K V E V L L P K F S V T F G L G L

*Xho*I

GAGGGACCATTCAAGAAGCTGGGCGTCCATACAATGTT*CTCGAG*AGATGGTGACTTTGGC 1020

E G P F K K L G V H T M F S R D G D F G

AACATGTACCGCATGTTTGTCAGCCATTTCATTAACGCAGTGGAGCACAAGGCCAATGTG 1080

N M Y R M F V S H F I N A V E H K A N V

GAGGTCACTGAAGCTGGCGTGGATCAACCCCTGGAAACTGGATTGCTTAAAGGACTCTTT 1140

E V T E A G V D Q P L E T G L L K G L F

TCGCGCTCCAAGAAGTTCGAGGCAGATCATCCGTTCGTGTTCGCCATCAAGTACAAGGAC 1200

S R S K K F E A D H P F V F A I K Y K D

TCAATCGCCTTTATCGGACACATTGCTAACTATGCTTATGTCTAAGCTAGATCGCAGATA 1260

S I A F I G H I A N Y A Y V

CGGAGTAGATCATTTTGTTAAAACTTACAAATAAAAAACACACGTAAATTGAAAAAAAAA 1320

*Eco*RI

AAAACC*GAATTC* 1332

Fig. 5.3c

*Eco*RI
*GAATTC*AAGCCATTCATTTAGAAGGTTTCCAGCTAGAAATAACAGATAAGAAACATGGCG 60
M A
AGCaAAGTCTCGATCCTTCTCCTGCTAACCGTCCATCTTCTGGCTGCTCAGACCTTCGCC 120
S K V S I L L L L T V H L L A A Q T F A
CAGGAGCTCATCGCTTGGCAACGGCAACAACAACAGCAGCAACAGCAGCAACTGCAGCTA 180
Q E L I A W Q R Q Q Q Q Q Q Q Q Q Q L Q L
CAGCAGCAACTGCTGCTGCAGCAGCAACAACACCAACGTAACCCAAGACCGGAGCTGGGC 240
Q Q Q L L L Q Q Q Q H Q R N P R P E L G
CTCCGTTCCCTGCCCGGAAAcCCGTGGACCCAGAACAATCAGGAAGCCATAAGCGATGTG 300
L R S L P G N P W T Q N N Q E A I S D V
GTGGCGGTGGACCTAACCAAACGTGAGCCGGTCACTCCGCCACCCAATCGCCCGCCGCCC 360
V A V D L T K R E P V T P P P N R P P P
GTCTTCAGCTACATGGACCGCTTCAGCTCCGAGCTCTTCAAGGAGATCATTAAGTCGCAA 420
V F S Y M D R F S S E L F K E I I K S Q
AGTCAGCAGAACGTGGTGTTCTCGCCCTTCTCCGTCCACGCGCTGCTGGCCCTGATCTAC 480
S Q Q N V V F S P F S V H A L L A L I Y
GGGGCCTCGGACGGAAAAACGTTCCGGGAACTGCAGAAGGCCGGAGAGTTCAGCAAGAAC 540
G A S D G K T F R E L Q K A G E F S K N
GCCATGGCCGTGGCCCAGGACTTCGAGAGCGTGATCAAGTACAAGAAGCATTTGGAGGGC 600
A M A V A Q D F E S V I K Y K K H L E G
GCAGATCTGACCCTGGCAACCAAGGTCTACTACAACCGCGAGCTGGGTGGCGTCAACCAC 660
A D L T L A T K V Y Y N R E L G G V N H
*Sal*I
AGCTACGATGAGTACGCAAAGTTCTATTTCAGCGCCGGCACGGAGGCT*GTCGAC*ATGCAG 720
S Y D E Y A K F Y F S A G T E A V D M Q
AACGCCAAAGACACGGCAGCCAAGATCAACGCCTGGGTGATGGACACGACGCGGAACAAG 780
N A K D T A A K I N A W V M D T T R N K
ATCCGGGACCTGGTCACACCGACCGACGTTGACCCACAGACGCAGGCGCTCCTTGTGAAT 840
I R D L V T P T D V D P Q T Q A L L V N
*Eco*RI
GCAGTCTACTTCCAGGGTCGTTGGGAGCAC*GAATTC*GCCACCATGGACACATCACCCTAC 900
A V Y F Q G R W E H E F A T M D T S P Y
GACTTTCAGCACACCAACGGCAGAATTTCCAAGGTGGCCATGATGTTCAACGACGATGTG 960
D F Q H T N G R I S K V A M M F N D D V
TACGGCCTGGCCGAGCTGCCTGAGCTGGGCGCCACCGCCTTGGAACTGGCCTACAAGGAC 1020
Y G L A E L P E L G A T A L E L A Y K D
AGCGCCACCAGCATGCTAATCCTGCTGCCCAATGAGACCACCGGACTGGGCAAAATGCTG 1080
S A T S M L I L L P N E T T G L G K M L
CAGCAGCTGTCCCGGCCGGAGTTCGATCTCAACCGCGTTGCCCACCGCCTGCGCCGCCAG 1140
Q Q L S R P E F D L N R V A H R L R R Q
TCGGTCGCTGTGCGCCTGCCCAAGTTCCAATTCGAGTTCGAGCAGGACATGACCGAGCCG 1200
S V A V R L P K F Q F E F E Q D M T E P
CTAAAGAACCTGGGAGTCCACCAGATGTTCACGCCCAACTCGCAGGTGACCAAGTTGATG 1260
L K N L G V H Q M F T P N S Q V T K L M
GATCAGCCGGTGCGCGTGAGCAAGATCCTGCAGAAGGCCTACATCAATGTGGGCGAGGCG 1320
D Q P V R V S K I L Q K A Y I N V G E A
GGCACAGAGGCCTCGGCAGCTTCCTATGCCAAGTTCGTACCCCTTTCGCTGCCTCCCAAG 1380
G T E A S A A S Y A K F V P L S L P P K
CCCACGGAGTTCGTCGCCAACCGGCCATTCGTCTTCGCCGTCCGCACCCCCTCCTCAGTT 1440
P T E F V A N R P F V F A V R T P S S V
CTGTTCATAGGTCACGTGGAGTATCCCACGCCCATGAGCGTCTAAAATGAACCAAATTCG 1500
L F I G H V E Y P T P M S V
*Eco*RI
GACAAATTGATTTATTA*GAATTC* 1523

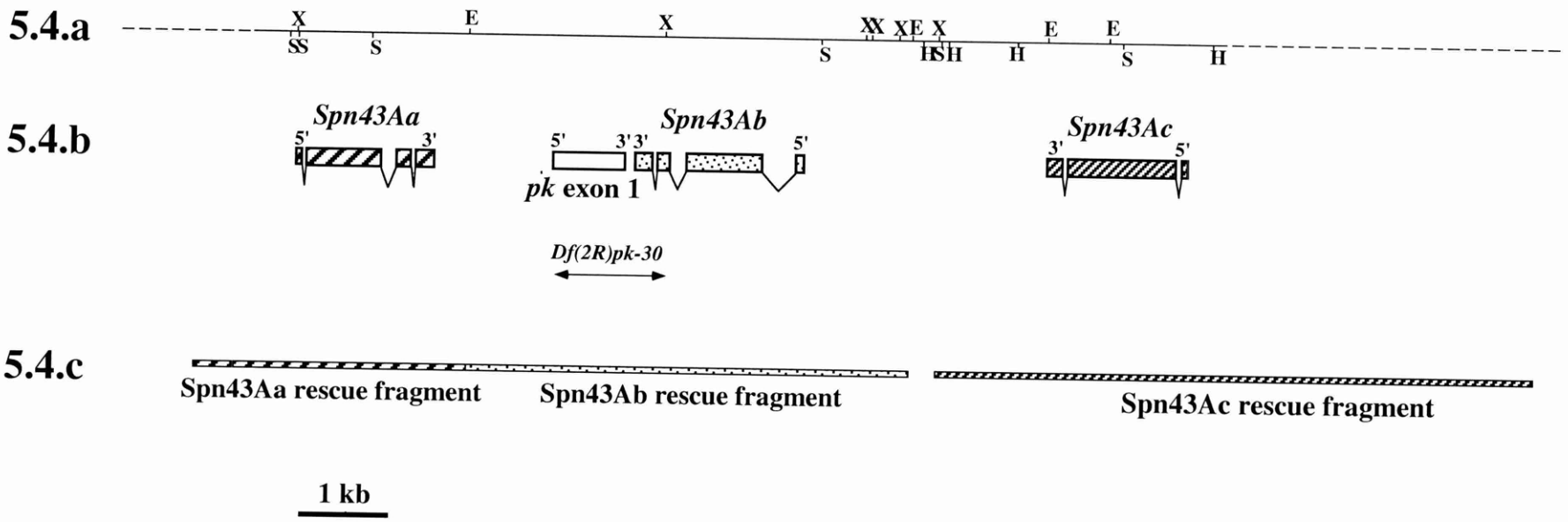


Fig. 5.4 Map of the serpin region. a) Genomic restriction map. Enzymes are *Eco*RI (E), *Hin*dIII (H), *Sal*I (S) and *Xho*I (X). b) Transcripts in the region. The region deleted in the *Df(2R)pk-30* mutant is marked. c) Genomic fragments used in the rescue experiments.

5.3 Rescue of necrotic phenotype

5.3.1 Constructs

The *pwhiteRabbit* transformation vector (Dunin-Borkowski and Brown, 1995) was used with genomic constructs of the three serpins (section 7.20.1). These were microinjected into *y w* embryos following standard methods (Spradling, 1986) by Carol McKimmie.

5.3.2 Crosses

Putative G_1 transformant progeny were crossed to *w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2*/Sco; In(3LR)TM2,* emc^2 Ubx^{130} e^s*/In(3LR)TM6B, Hu e Tb ca (w; CyO/Sco; TM2/TM6*), or *w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2 *P[*ry^+ *wg:lacZ]/If; In(3LR)TM6B, Hu e Tb ca/Tp(3;3)MKRS, M(3)76A kar* ry^2 *Sb* (*w; CyO/If; MKRS/TM6B*) flies to establish balanced stocks in a *w* background. To test for rescue, third chromosomal *P[Spn43A]* inserts for each of the three Serpin transcripts were crossed to nec^1 bw^D*/CyO* flies. In the next generation, *w;* nec^1 bw^D*/Sco; P[Spn43A]/+* (or, *w;* nec^1 bw^D*/If;P[Spn43A]/+*) males were crossed to *y w; Df(2R)pk-78k/CyO* females and the phenotype of surviving *w;* nec^1 bw^D*/Df(2R)pk-78k P[Spn43A]/+* progeny was scored.

5.3.3 Results

Results of the test crosses are shown in Table 5.1. For the *P[Spn43Aa]* and *P[Spn43Ab]* crosses, phenotypically nec flies hatched, although at the reduced frequencies relative to Cy or Sco flies of 13% (89/646) and 6% (5/88), respectively). The surviving $nec^1 bw^D$*/Df(2R)pk-78k* flies developed necrotic patches within 24 hours and died within three days of eclosion. In contrast, one eighth (96/745) of the bw^D w^+ Cy^+ If^+ progeny of the *P[Spn43Ac]* test cross survived, together with 7.5% (56/745) of their w nec Cy^+ If^+ siblings. When re-examined ten days later, the nec^1 bw^D*/pk-78k; P[Spn43Ac]/+* flies remained wild type, indicating that the *P[Spn43Ac]* insert rescues the necrotic phenotype completely.

w; nec^{1} *bw*^{D}*/If(Sco);P[Spn43Ax]/+* x *w; Df(2R)pk-78k/CyO* :

	P[Spn43Aa]	*P[Spn43Ab]*	*P[Spn43Ac]*
Phenotype:			
white eye; necrotic	49	2	54
white eye; bw^{D}	6*	1*	2*
white eye; CyO or If (Sco)	231	29	211
white eye; CyO If (CyO Sco)	82	13	91
red eye; necrotic	30	2	0
red eye; bw^{D}	4*	0	96**
red eye; CyO or If (Sco)	152	30	206
red eye;CyO If (Cyo Sco)	92	11	85

* Developed necrotic patches and died within 72 hours of eclosion.
** Survived at least 10 days after eclosion with nec^{+} phenotype.

Table 5.1 Results of serpin rescue crosses. *Spn43a* was balanced with *Cyo/Sco* on the second chromosome; for both the others, *CyO/If* was used.

5.4 Serpins

5.4.1 Background

The serpins form a superfamily of glycoproteins, most of which function as inhibitors of serine proteinases (Boswell and Carrell, 1988). The term serpin (serine proteinase inhibitor) was coined by Carrell and Travis (1985) who were studying alpha-antitrypsin. This serpin, present in human plasma where it protects elastic tissue from proteolytic attack, was recognised to be part of a family of proteins (Carrell *et al*, 1982; Carrell and Owen, 1985). Serpins have also been found in plants and viruses as well as higher organisms (Huber and Carrell, 1989). In mammals, serpins regulate a variety of proteolytic cascades, including blood coagulation, the complement reaction and the inflammatory response (Boswell and Carrell, 1988; Potempa *et al*, 1994). Other functions include hormone transport and blood pressure regulation.

Serpins interact with their target proteinase at a reactive centre located within a loop structure near its carboxyl (C) terminus (Potempa *et al*, 1994). The homology between members of the serpin family is higher in the C terminal half of the molecule (Sommer *et al*, 1987). The specificity of the serpin is primarily defined by a single amino

acid within this centre, designated P1 (Carrell and Travis, 1985). Alteration of this P1 residue changes the specificity of the serpin (Boswell and Carrell, 1988). The reactive centre is mobile and this flexibility in confirmation allows variability of function (Carrell and Evans, 1992).

The cleavage of serpins by their target proteinases is irreversible, indicating that they are suicide substrate inhibitors; their function is restricted to a single encounter with proteinase (Potempa *et al*, 1994; Wright, 1996).

Some serpins are relatively inefficient inhibitors until they are activated by a specific sulphated polysaccharide, for example, antithrombin III undergoes a 1000-fold increase in its association constant with thrombin in the presence of heparin (Huber and Carrell, 1989). Other members of the serpin family, such as ovalbumin, are substrates of serine proteinases in their own right rather than inhibitors (Wright, 1996).

5.4.2 Invertebrate serpins

Invertebrate serpins have been less well characterised. Kanost *et al* (1989) identified a serpin in the tobacco hornworm *Manduca sexta*. The gene was found to be expressed in the fat body and the protein was present in the haemolymph. The 392 residue polypeptide showed 30% identity with members of the serpin superfamily; there were identical residues throughout the sequence, with greater similarity in the C-terminal half of the molecule. This level of homology is characteristic within the serpin family (Sommer *et al*, 1987).

This serpin was subsequently found to be encoded by ten exons, with twelve alternative versions of exon nine (Jiang *et al*, 1994; Jiang *et al*, 1996). cDNAs corresponding to all twelve exon nine variants were isolated, indicating that they are all expressed. The range of cDNA clones probably result from alternate RNA splicing to select one exon nine for each mRNA; this type of alternative splicing is known as mutually exclusive exon use (Barbas *et al*, 1993; Jiang *et al,* 1994). The amino acid sequence encoded by the exon nine variants includes the reactive centre loop (Jiang *et al*, 1996).

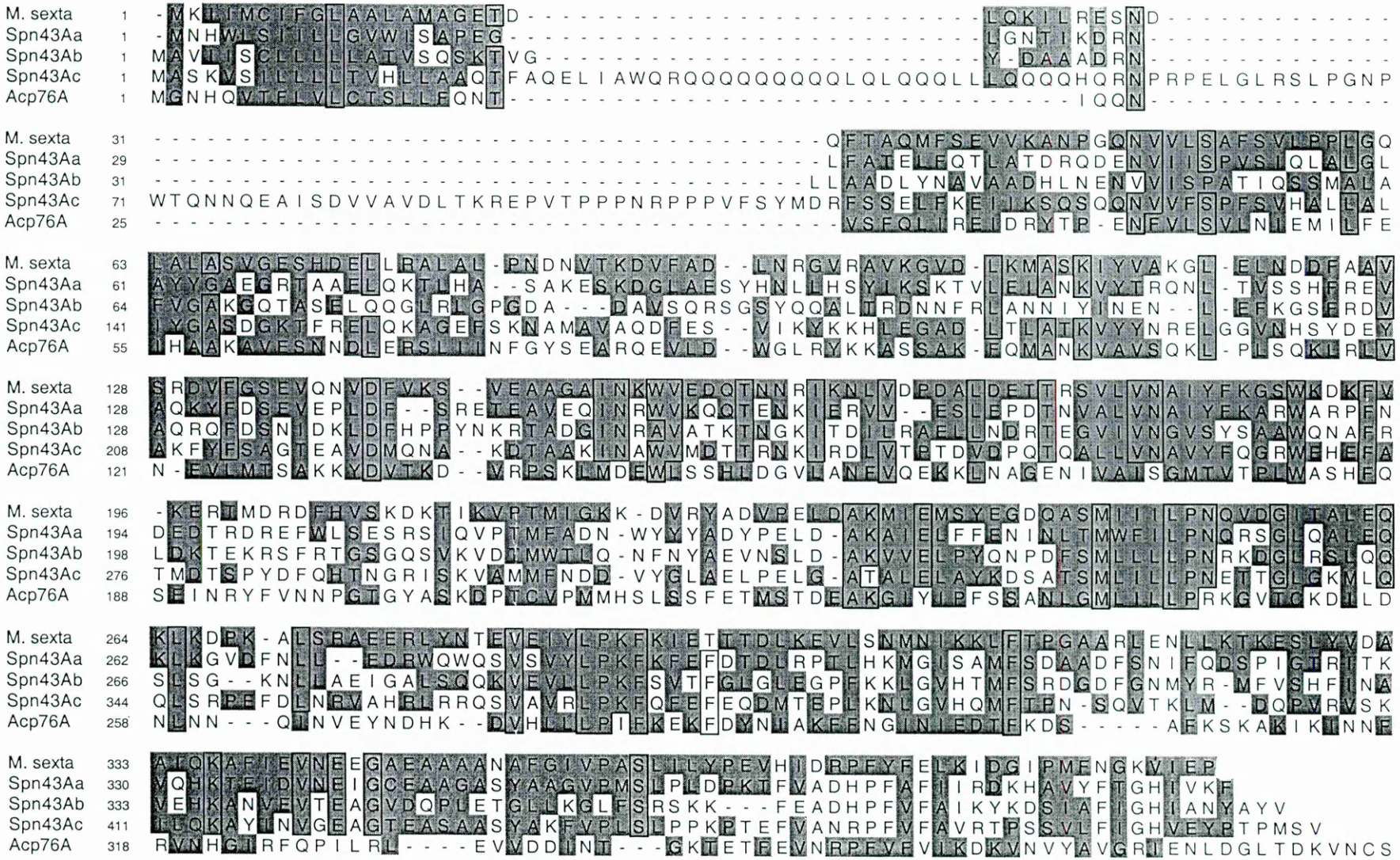
The first *Drosophila* serpin to be identified was accessory gland peptide 76A (Acp76A) (Coleman *et al*, 1995). This secreted protein is synthesised only in the accessory gland of the adult male and is found in the female uterus immediately after mating. Table 5.2 shows the degree of homology between some invertebrate serpins and Table 5.3 shows a comparison of the serpin cluster amino acid sequence with that of other insect serpins.

	Spn43Ab	***Spn43Ac***	***Acp76A***	***M. sexta***
Spn43Aa	35%(46%)	36%(46%)	22%(30%)	37%(46%)
Spn43Ab		30%(38%)	22%(34%)	27%(40%)
Spn43Ac			?	34%(44%)
Acp76A				21%(30%)

Table 5.2 Percent identity (similarity) of invertebrate serpins. ***Spn43Aa, b*** and ***c*** from *Drosophila* (see text); ***M. sexta*** alaserpin (Kanost *et al*, 1989; SwissProt accession no. P14754); ***Acp76A*** from *Drosophila* (Coleman *et al*, 1995; Wolfner *et al*, 1997; EMBL accession no. U90947). Percentages were calculated using BestFit (GCG). ?-BestFit only calculated this value for small sections of this protein, suggesting that the degree of homology varied widely across the protein.

Table 5.3 (overleaf) Comparison of the amino acid sequence of the serpin cluster with a *Manduca sexta* serpin (Jiang *et al*, 1994) and Acp76A from *Drosophila* (Coleman *et al*, 1995). Residues identical in all domains are boxed; those with greater than 80% homology to *M. sexta* are shaded. Comparison generated with SeqVu 1.0 (Garvan Institute of Medical Research); initial alignments done with Clustalw and BestFit. Analysis of the alignment predicts that Spn43Aa and Spn43c are classical serpins, whereas Spn43b may not be active (Robin Carrell, personal communication) as the putative reactive centre loop of Spn43Ab (residues 345-371) is not well conserved, particularly in the critical hinge region (residues 348-352).

Table 5.3



A role has been postulated for the serpins in the insect immune response. Insect host defence involves cellular reactions and a systemic response, as well as coagulation and melanisation which occur at sites of injury (Hoffmann and Reichhart, 1997). Serine proteases have been shown to be induced in immune challenged *Drosophila,* possibly in connection with melanotic reactions (Coustau *et al*, 1996).

The Toll-Dorsal signalling pathway of *Drosophila* was originally identified as being involved in embryonic dorsal/ventral axis formation (Belvin and Anderson, 1996; Dushay and Eldon, 1998). Briefly, the binding of Spaetzle to Toll causes activation of a signal transduction cascade that leads to the phosphorylation and subsequent degradation of Cactus. In the absence of Cactus, Dorsal moves into the nucleus where it acts as a transcription factor to initiate the expression of zygotic genes required for ventral cell fates. There is now evidence that this pathway is also involved in the *Drosophila* immune response, preferentially activating the antifungal response (Lemaitre *et al*, 1996).

Drosomycin is an antifungal peptide of *Drosophila* that is expressed in larvae or adults within thirty minutes of septic injury (Hoffmann and Reichhart, 1997). The control of this expression involves genes of the Toll-Dorsal signalling pathway. The transcription of *Spn43Ac* has been shown to be upregulated upon septic injury and in *nec* mutants the *Drosomycin* gene was found to be constitutively activated (Elena A. Levashina, Emma Langley, Clare Green, David Gubb, Michael Ashburner, Jules A. Hoffmann and Jean-Marc Reichhart; in preparation). The constitutive expression of Drosomycin is abolished when *nec* mutants are in a *spaetzle* or *Toll* genetic background, implying that *Spn43Ac* functions upstream of the *Toll* signalling cassette.

5.5 Summary

This chapter presents the identification of a cluster of three transcripts in the vicinity of the 5' *pk* exon. These transcripts were characterised and shown to have homology to members of the serpin superfamily. Rescue of the necrotic phenotype using a construct containing the most distal of these transcripts was demonstrated. An

introduction to serpins, their possible functions and the importance of their reactive centre was given. The role of serpins in the immune response was considered and the involvement of at least one of these serpins in the Toll/Dorsal signalling pathway was discussed.

CHAPTER 6. **DISCUSSION**

6.1 Summary of data presented.

The studies described in this thesis reveal that the *prickle* gene consists of six common exons with at least three alternatively spliced 5' ends. The common exons contain three LIM or LIM like domains, along with a novel domain that we have called the PET domain. LIM domains are involved in protein-protein interactions; the function of the PET domain is unknown.

The first full length cDNA to be cloned was identified as *pk* (see Chapter 2). It was 4163 base pairs long and was found to encode a conceptual protein of 870 amino acids. Six of the seven exons of this transcript were found to map in the region of the $pk^{pk\text{-}sple}$ mutants. The other 5' exon spliced about 60 kb to the rest of the transcript and mapped in the same region as the *Df(2R)pk-30* allele.

Subsequently two alternative 5' spliced transcripts were identified and characterised (see Chapter 3). The two exons of the *sple* transcript map 15 kb proximal to the common exons. The conceptual *sple* protein is 1206 amino acids long. The single exon of the other alternative 5' end maps 36 kb proximal to the common exons. This transcript is known as *pkM* and encodes a 936 amino acid protein.

Mutant Northern blot data confirmed that the first transcript was *pk* and the second one *sple* (David Gubb, personal communication). Developmental Northerns showed that both of these were expressed in embryos and two day pupae (Chapter 2) while *pkM* was only expressed in embryos.

Constructs were made of the *pk* and *sple* transcripts and these were shown to rescue the prickle and spiny-legs phenotypes respectively. They also produced overexpression phenotypes when driven with *C765-GAL4*. Overexpression of the *pk* construct gives a leg phenotype resembling loss-of-function *sple*, while overexpression of the *sple* construct results in a wing phenotype resembling loss-of-function *pk*. The alternative proteins therefore have qualitatively different activities.

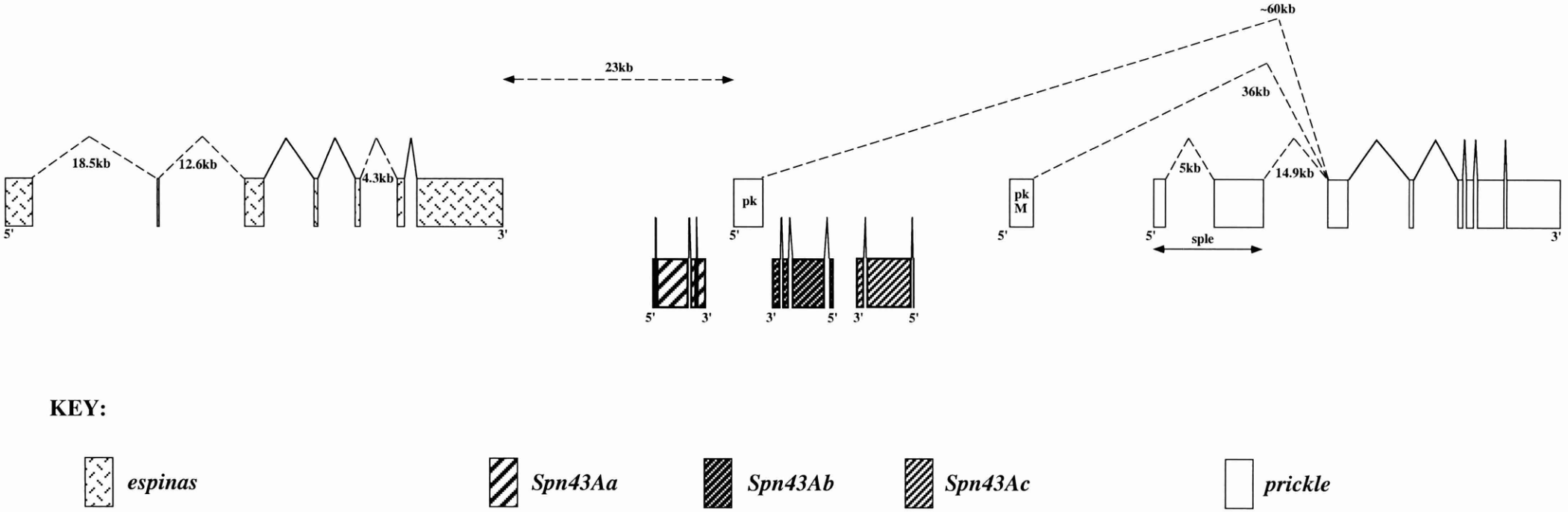
A cluster of three genes with homology to serpins was found in the region of the *pk* 5' exon (Chapter 5). Two of these map within the first *pk* intron and are transcribed in the opposite direction; the third maps 5' to *pk* and is transcribed in the same direction. Constructs were made of each serpin and the most distal of the three was shown to rescue the necrotic phenotype.

Database searches found a class of proteins that contained the PET domain and in most cases LIM domains homologous to those of PK (Chapter 4). The *Drosophila* gene *espinas,* which maps 23 kb proximal to *pk,* was also found to contain these domains. The PET and LIM domains of ESN are very similar to those of PK, but there is little similarity between the rest of the sequences at either the protein or DNA level. Only one *esn* transcript has been identified so far, but at least two transcripts were seen on developmental Northerns (David Gubb, personal communication).

A *Drosophila* EST sequence with homology to the PET domain has been identified. It is not yet known if there are LIM domains associated with this or two of the other sequences containing the PET domain. The human clone 241420 sequenced as part of this project appears to consist of the PET domain and little else, but it is not certain that this is full length.

The molecular characterisation of *pk* has revealed a complex picture that helps to explain the complicated genetics of this gene (see Figure 6.1 for summary of the

Fig. 6.1 Summary of transcripts presented in this thesis
(not to scale)



transcripts) and suggests that *pk* encodes a cytoskeletal-associated protein. This study has defined a novel family of conserved proteins and will provide the basis for understanding the role of *pk* in planar polarity.

6.2 Possible functions of *pk*.

Models of planar polarity signalling have been described in the past, but the role of *pk* has been unclear (see Chapter 1). A recently described model is shown below (Figure 6.2). The tissue specific factors were included in this model because to be involved in the generation of polarity of hairs, bristles and ommatidia, FZ signalling must be able to regulate a wide range of cellular processes (Shulman *et al*, 1998). Recently Notch has also been shown to play a role in planar polarity in the eye (Cooper and Bray, 1999; Fanto and Mlodzik, 1999).

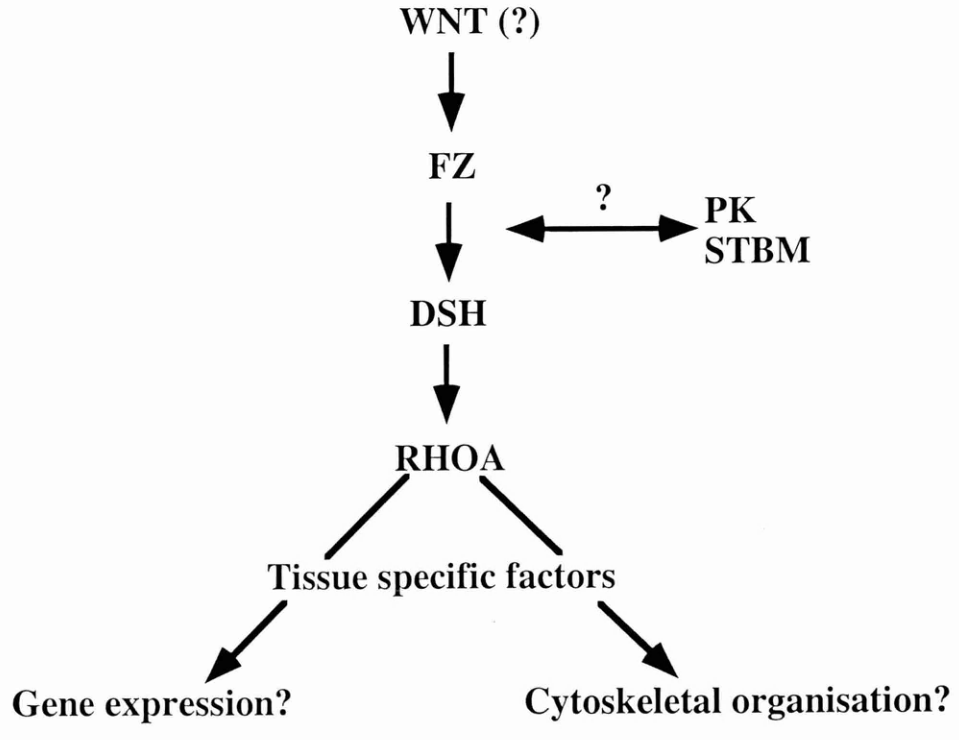


Fig. 6.2. A model of the planar polarity signalling pathway (taken from Shulman *et al*, 1998)

Epistasis experiments in the eye suggest that PK acts upstream of DSH, but in a different branch of the pathway to FZ (Dave Tree, personal communication) as the overexpression phenotype seen in the eye in *Gal4-sevenless UAS[pk^{+}]* flies is dominantly suppressed by heterozygous *dsh*. Overexpression phenotypes of *pk* (Dave Tree, personal communication) are opposite to those seen with *fz* (Adler *et al*, 1997), implying that PK and FZ are antagonistic in function. It had previously been observed that *pk^{sple}* suppresses the *dsh* wing phenotype (Coulson, 1994).

SPLE was shown to affect the resolution of Notch signalling in the establishment of planar polarity in the eye (Cooper and Bray, 1999). Notch signalling is known to be involved in leg segmentation and it was found that ectopic expression of *Notch* in the leg caused the formation of ectopic joint structures (de Celis *et al*, 1998). *NOTCH* RNA expression in the leg disc 24 hours after puparium formation was shown to be higher where joints were being formed. In contrast, both *pk* and *sple* were found to be expressed uniformly throughout the pupal leg disc except at the tarsal segment boundaries (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation). As SPLE has been shown to be antagonistic to Notch signalling (Cooper and Bray, 1999) this may help to explain the tarsal joint duplication seen in the *pk^{sple}* and *pk$^{pk\text{-}sple}$* mutants.

The double *pk$^{pk\text{-}sple}$* mutants have a phenotype intermediate between the two more extreme single mutant phenotypes. Although the single mutant phenotypes are seen in reciprocal regions of the body, the *pk* and *sple* transcripts have similar expression patterns in wing, leg and eye discs (D. Gubb, C. Green, D. Huen, D. Coulson, G. Johnson, S. Collier and J. Roote, in preparation). In addition, overexpression of the *pk* transcript gives a sple-like phenotype and vice versa. These observations may indicate an antagonistic interaction between the two proteins, which could involve them competing for a target that is present in limiting amounts. Alternatively, as LIM domains are known to be involved in the formation of both heterodimers and homodimers (reviewed by Dawid *et al*, 1998), the interaction between PK and SPLE may involve the formation of heterodimers. The phenotypes seen in the single mutants could then be explained by the

formation of excess homodimers. It would be helpful to determine the effects of overexpression of the common exons alone and that of the unique exons, bearing in mind that the *pk* unique exon contains little coding sequence (only 39 bases) in contrast to the *sple* and *pkM* unique exons (1047 and 237 bases respectively). It is interesting to note that at least two other members of the family, murine testin and the human triple LIM protein, also have alternatively spliced forms (Divecha and Charleston, 1995; Fisher *et al*, 1997).

LIM domain proteins may interact with themselves or other LIM domain proteins, or with entirely different proteins. There are documented examples of LIM proteins interacting with two or more other proteins, each LIM domain being partner specific. For example, zyxin interacts with several proteins and it has been shown that the first LIM is specific for binding to members of the CRP family (Schmeihel and Beckerle, 1994). The variations from standard LIM domains conserved in the "PET family" proteins (section 4.5) could be clues to the specificity of each and to the function of the protein.

So far there are no clues as to the function of the PET domain. Several residues are strongly conserved, but no known patterns have been identified. It should be noted that functional domains cannot always be identified by sequence alone. For example, Meng *et al* (1999) found a divergent SH2 domain in the adapter protein CBL that was not recognisable from the amino acid sequence of the protein. If proteins exist that contain the PET domain and little else (as the human EST clone 241420 appears to), they could be very useful in clarifying what this region is doing.

Different domains in the same protein may function in different signalling pathways. For example, the different domains of DSH have been shown to be involved in either planar polarity or WG signalling (Boutros *et al*, 1998; Axelrod *et al*, 1998).

Axelrod *et al* (1998) also showed that the localisation of DSH to the membrane was important in planar polarity signalling. Staining with an antibody to the common exons of PK show that it is cytoplasmic and appears to be associated with the membrane (Dave Tree, personal communication). SAPS (Statistical Analysis of Protein Sequences; Brendel *et al*, 1992) does not identify any transmembrane domains in PK, SPLE, PKM or

ESN. PSORTII (Prediction of Protein Sorting Signals and Localisation Sites in Amino Acid Sequences; CuraGen Corporation) did not identify any domains or motifs other than the LIM domains.

Conserved protein domains, such as LIM or SH2 domains, are known to mediate specific protein-protein and protein-phosopholipid interactions. Covalent association of these recognition modules allows a single polypeptide to bind multiple protein ligands and thus proteins containing these domains may function as scaffold or adapter molecules (reviewed by Parson and Scott, 1997). For example, the antigen receptors on lymphocytes are multimeric complexes composed of a receptor module associated with various signalling proteins and binding to pathogens initiates intracellular signalling events (reviewed by Rudd, 1999). Actin polymerisation is necessary for promoting the correct orientation and contacts between lymphocytes and antigen presenting cells, a process that depends on CDC42 (reviewed by Penninger and Crabtree, 1999). Antigen receptor activation may therefore lead to the organisation of an actin-scaffolded signalling "highway" whose function is to sustain T cell receptor signalling and co-ordinate downstream events (*ibid*).

The five PDZ domains of the InaD (inactivation no-after potential) protein were found to have different and highly specific targets, interacting with different components of the phototransduction cascade in *Drosophila* (Tsunoda *et al*, 1997). InaD was shown to act as an organising scaffold for photoreceptor complexes *in vivo*.

Shulman *et al* (1998) suggested that the distinctive phenotypes of the different polarity genes, plus the lack of definitive epistasis amongst these genes, may indicate that some of the factors function in a multiprotein complex rather than occupying discrete steps in a pathway. Axelrod *et al* (1998) showed that DSH links the activity of the FZ family to the control of the cytoskeleton and suggested it may serve as a scaffolding or adapter protein. LIM domains are multiple binding and adapter molecules and some LIM proteins appear to function solely as adapters to bring other components together in a complex (Dawid *et al*, 1998). Therefore, the structure of PK, plus its association with the membrane, suggests that it could be involved in the formation of a multiprotein complex.

Determining the function of the PET domain and identification of the protein or proteins that PK interacts with would throw more light on this theory.

SAPS detected nothing notable about the PK proteins except that they have a higher proline content (11-12%) than is expected for *Drosophila* genes. The prolines are found throughout the sequences, including one stretch towards the end of the unique SPLE exons. Proline rich regions have been implicated in various processes, for example in the binding of SH3 domains (Alexandropoulos *et al*, 1995). All the SH3 binding domains identified included the core motif PXXP, which is seen several times in the PK proteins, although complete known SH3 binding domains have not been found.

6.3 Significance of *espinas*.

The existence of *espinas* further complicates the picture. There are increasing instances of genes from the same family having at least partially redundant functions. It could be that *pk* and *esn* have some redundant functions. In particular the duplicated wing hair phenotype seems to be complemented in *pk* mutants by *esn* function (David Gubb, personal communication). Further members of the PET family may remain to be discovered (as indicated by the tag sequence with homology to the PET domain). It seems likely from the database searches in *Drosophila* and other organisms that the PET family is relatively small.

The fact that the PET and LIM domains are strongly conserved between PK and ESN while there is little similarity between the rest of the proteins emphasises the importance of these regions. In contrast to the PK proteins, the number of prolines in ESN (6.6%) is within the range expected for *Drosophila* genes. This difference could help elucidate possible functions of regions of the proteins outside the PET and LIM domains.

6.4 The serpin cluster.

The potential role of the serpin cluster has already been given some consideration (see Chapter 5). The most distal of the serpins, *Spn43Ac*, was shown to be necrotic in this study. Further evidence for this comes from data produced by Jean-Marc Reichhart and Elena Levashina (personal communcation). They showed that the mutations nec^1 and nec^2 both had amino acid substitutions within *Spn43Ac*. No phenotype has been identified for the other two serpins; homozygous deletions that remove *Spn43Aa* and *Spn43Ab* are wild type.

Recently it has been found that Spn43Ac shows many similarities with some of the human serpins, in the amino acid sequence and physiologically. These characteristics of Spn43Ac, plus the distinctive necrotic phenotype, may make it feasible to design a fly model for some of the human diseases caused by mutations in serpins. Experiments are currently underway to further explore this option

6.5 Future studies

The work described in this study leaves many questions unanswered and there are many avenues of further research. Expression of different constructs using the *GAL4-UAS* system (Brand and Perrimon, 1993) will be helpful. The most obvious include one consisting of the common exons alone and those containing each of the unique 5' ends, as well as the whole *pkM* transcript. Deleting or mutating various combinations of the LIM domains and/or the PET domain could provide information as to the functions and specificities of these regions. Different GAL4 drivers could also be used to give different levels of expression and/or restrict expression to smaller regions. Some of these experiments are already underway in our laboratory.

Determining which other protein/s PK interacts with would help enormously and is being undertaken in collaboration with W. Chia's laboratory.

Several *pk* mutants have not yet been characterised, including all of the pk^{sple} ones. These are probably point mutations and analysing them could provide clues to which regions of the protein are most important. I am currently investigating some of these mutations, using the techniques described in section 4.6.

Determining the phenotype of *esn* is also ongoing. A cDNA construct is currently being expressed in flies with a variety of GAL4 drivers. Once the results from these are known, other experiments will be carried out using this construct in combination with the various *pk* constructs to determine possible interactions between the two. It will be necessary to characterise the alternatively spliced *esn* transcripts.

6.6 Conclusion

The results presented in this thesis provide a basis for understanding the role of PK. The future research outlined above should elucidate the functions of PK and ESN in planar polarity signalling. It is clear that complex interactions exist between the various signalling pathways that function during development. To obtain a complete picture of development requires description of all the elements in the signal transduction pathways and an understanding of the cellular processes that they modulate.

CHAPTER 7. MATERIALS AND METHODS

Deionised water was used to prepare all solutions and they were stored at room temperature unless otherwise stated. Reagents were supplied by Merck and Sigma (exceptions are indicated in the text). Standard molecular biology protocols are based on those in Sambrook *et al* (1989).

7.1 Stock solutions

1 M Tris HCl: Tris base was dissolved in water and the pH adjusted to the required value with concentrated hydrochloric acid. The solution was autoclaved.

0.5 M EDTA, pH 8: Disodium ethylenediaminetetra-acetate.$2H_2O$ was added to water and the pH adjusted to 8 with sodium hydroxide pellets. (EDTA will not go into solution until pH is approximately 8). The solution was autoclaved.

TE buffer pH 7.5: 10 mM Tris HCl pH 7.5; 1 mM EDTA pH 8. The solution was autoclaved.

20 x SSC: 3 M NaCl; 0.3 M tri-Sodium citrate.

20 x SSPE: 3 M NaCl; 0.2 M $NaH_2PO_4.H_2O$; 0.02 M EDTA.

10 x TA: 0.4 M Tris; 0.03 M Sodium acetate; 0.01 M EDTA. pH 7.8 with acetic acid.

10% SDS: Sodium dodecyl sulphate was dissolved in water by heating to 68°C.

10 mg / ml Ethidium bromide: Ethidium bromide was added to water and stirred for several hours. It was stored in a foil-wrapped bottle.

Denaturation solution: 0.5 M NaOH; 1.5 M NaCl.

Neutralisation solution: 1 M Tris; 3 M NaCl. pH 7.5 with hydrochloric acid

LB: For 1 litre: 10 g tryptone; 5 g yeast extract; 10 g NaCl in water. pH was adjusted to 7.2 with NaOH and then $MgSO_4$ heptahydrate was added to 10 mM final concentration.

LA: LB + 1.5% agar

Top agarose: LB + 0.7% agarose

SM: 0.1 M NaCl; 0.01 M Mg SO_4 ; 0.05 M Tris pH 7.5; 0.01% gelatine. pH 7.5

RNase: This was made up at required concentration (1 mg / ml or 10 mg / ml) in 10 mM Tris pH 7.5, 15 mM NaCl, then boiled for 15 minutes. After cooling slowly to room temperature, it was stored in 0.5 ml aliquots at -20°C.

DNase: This was made up at the required concentration in 20 mM Tris, 50 mM NaCl, 1 mM dithiothreitol, 100 µg / ml BSA, 50% glycerol and stored in 0.5 ml aliquots at -20°C.

ssDNA, 10 mg / ml: Salmon sperm DNA (Type III sodium salt, Sigma) was dissolved in water at 10 mg / ml and stirred for several hours at room temperature. The DNA was then sheared by passing through a 17 gauge needle several times, dispensed into 1 ml aliquots and boiled for ten minutes. It was stored at -20^{o}C. The DNA was boiled for five minutes immediately before use.

Equilibration of phenol: Phenol was melted at 68^{o}C and hydroxyquinoline was added to a final concentration of 0.1% . An equal volume of 0.5 M Tris pH 8 was added and the mixture stirred for 15 minutes on a magnetic stirrer. The stirrer was then turned off and when the two phases had separated the upper (aqueous) phase was aspirated. This extraction was then repeated several times with 0.1 M Tris pH 8 until the pH of the phenolic phase was >7.8 (measured with pH paper). The final aqueous phase was then removed and an equal volume of TE buffer pH 7.5 was added. The equilibrated phenol was stored in a foil wrapped bottle at 4^{o}C for short term storage or at -20^{o}C for longer periods.

Phenol / chloroform: Equal volumes of equilibrated phenol were mixed with chloroform and stored under TE buffer pH 7.5 in a foil wrapped bottle at 4^{o}C.

7.2 Basic procedures

7.2.1 Extraction with phenol and / or chloroform

A half volume (unless otherwise specified) of phenol / TE, phenol / chloroform or chloroform was added to the sample and vortexed. This was spun in the microfuge at maximum speed for 3 minutes for phenol extractions, 2 minutes for chloroform. The aqueous (upper) phase was transferred to a clean tube and the extraction repeated as required.

7.2.2 Ethanol precipitation

2.5 volumes ethanol and 0.1 volumes 3 M sodium acetate pH 5 were added to the sample and mixed by inversion. The tube was incubated on ice or at -20°C for 15 to 60 minutes, then spun in a microfuge for 10 minutes at maximum speed. The supernatant was discarded and the pellet washed in 70% ethanol and spun for 5 minutes at maximum speed. As much alcohol as possible was removed, then the pellet was dried by leaving the tube open at room temperature for 5 to 10 minutes.

7.2.3 Restriction digests

100 mM spermidine: Made up in water and stored at -20°C in 0.5 ml aliquots. One working aliquot kept at 4°C.

Loading buffer: 0.25% bromophenol blue, 0.25% xylene cyanol FF and 30% glycerol in water.

METHOD:

The required amount of DNA was cut with 1-2 units of restriction enzyme/s (supplied by Boehringer Mannheim) per µg of DNA in a solution containing a 1 x concentration of the appropriate buffer (supplied with the enzyme) and 5 mM spermidine. RNase was added at a final concentration of 0.05 µg / µl when DNA had not been treated previously to remove RNA. The sample was incubated at 37°C for one hour and loading buffer added (final concentration 1 x).

GEL ELECTROPHORESIS: Samples were run on an agarose gel (0.7% unless otherwise stated) made up in 1 x TA buffer, along with appropriate markers. Ethidium bromide was added directly to the gel (0.5 µg / ml) unless otherwise specified.

7.3 Fly DNA extraction

Homogenisation buffer: 10 mM Tris HCl pH 7.5; 60 mM NaCl; 10 mM EDTA; 0.15 mM spermidine; 5% sucrose.

Lysis buffer: 200 mM Tris HCl pH 9 ; 30 mM EDTA; 2% SDS; 5% sucrose.

METHOD:

(Flies were stored at -70°C until required.) 350 µl homogenisation buffer was added to 50 flies in a 1.5 ml microcentrifuge tube. Flies were ground using a pestle (Burkard Scientific Ltd). 350 µl of lysis buffer and 10 µl protease K (10 mg / ml) was added and the sample was incubated at 37°C for 1 to 2 hours. It was extracted twice with an equal volume of phenol / TE, once with an equal volume of chloroform and precipitated once with 1 ml ethanol (no salt) on ice for 10 minutes. The DNA was pelleted, washed and dried, then resuspended in 50 µl TE buffer. (Yield approx. 1 µg m / fly for *D. melanogaster*.).

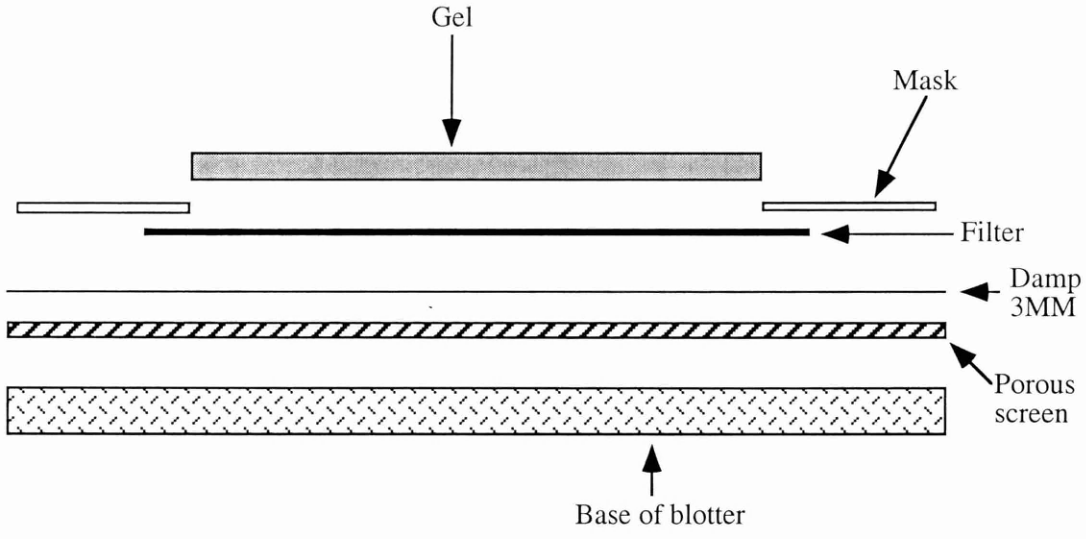
7.4 Southern blotting

Membrane: "Zetabind" (CUNO)

METHOD:

Gels were soaked in 0.25 M hydrochloric acid for 10 minutes, then in denaturation solution for 45 minutes and twice in neutralisation solution for 30 minutes each time. Blotting was carried out using a vacuum blotter (Hybaid; see below for set up). After blotting, the filter was allowed to air dry for a few minutes and then UV crosslinked. To minimise background, filters were washed for one hour at 65°C in 0.1 x SSC, 0.5% SDS.

Fig 7.1. Set up of vacuum blotter.
A mask (made from used X-ray film) was cut to the same size as the platform of the blotter. A hole the size of the area of the gel that was to be blotted was cut in the mask. The membrane was cut to about 1 cm longer and wider than the hole, wet in water, then soaked in 20 x SSC for 20 minutes. A piece of 3M M filter paper slightly smaller than the mask was cut (this expanded slightly when damp). The apparatus was then assembled as shown. The gel was positioned carefully in line with the hole in the mask and any air bubbles removed with gloved fingers. A vacuum was established and then neutralisation solution was poured directly onto the gel until the gel was covered. It was left for 1.5 to 2 hours, with more neutralisation solution added as required to keep the gel covered.



7.5 Random priming

OLB: 1 M Hepes pH 6.6; 0.25 M Tris pH 8; 0.25 units / ml pd(N)$_6$ oligonucleotides; 25 mM Magnesium chloride; 5 mM ß-mercaptoethanol. Stored in 100 µl aliquots at -20^{o}C.

dNTPs: 0.2 mM dCTP; 0.2 mM dGTP; 0.2 mM dTTP. Stored in 100 µl aliquots at

-20°C.

1 mg / ml BSA (bovine serum albumin): Made up nuclease free BSA in sterile deionised water. Stored in 50 µl aliquots at -20°C.

[^{32}P]dATP: 10 mCi / ml [^{32}P]dATP. Purchased fortnightly from NEN

Stop solution: 0.125 M EDTA pH 8; 5% SDS; 0.25% bromophenol blue

G50 slurry: Sephadex G50 (medium) was added slowly to water in a 500 ml beaker and washed several times in water. The resin was then equilibrated in TE pH 7.5 and stored at 4°C. 10 g of sephadex yields 160 ml of slurry.

G50 column: A glass pasteur pipette was blocked with tissue, then filled with G50, taking care that there were no air bubbles. The G50 was kept wet with TE and the tube sealed with "Blu-tack" until required.

METHOD:

200-500 ng DNA were made up to 25 µl with water, denatured by boiling for 2 minutes, then placed on ice for 1 minute. 10 µl OLB, 5 µl dNTPs, 2 µl BSA, 5 units Klenow (Bioline) and 30 µCi [^{32}P]dATP were added and the sample incubated for 1-2 hours at 37°C. 10 µl stop solution was added, the sample was loaded onto a G50 column and washed down the column with TE buffer pH 7.5. The fraction containing first peak of counts (approx. 0.5 ml) was collected. (Second peak contained unbound ^{32}P and ran off with blue dye.)

7.6 Southern hybridisation

Pre-hybridisation solution: 250 mM $NaH_2PO_4.H_2O$; 250 mM NaCl; 1 mM EDTA pH 7.5 with NaOH; 1% SDS

Hybridisation solution: Pre-hybridisation solution + 6% PEG 6000

METHOD:

The filter was placed round the inside of hybridisation tube and any large air

bubbles were removed. (If the filter was dry, it was pre-wet with the pre-hybridisation solution.) 25 mls pre-hybridisation solution was added and the filter was pre-hybridised at 65°C for a minimum of one hour. The pre-hybridisation solution was replaced with 10 mls of hybridisation solution and 100 µl random primed probe (boiled for 5 minutes). Hybridisation was carried out overnight at 65°C. The "hot" hybridisation solution was poured off and the filter washed twice in 200 mls of 1 x SSC, 0.5% SDS for 20 minutes. A final wash step was carried out at the required stringency (low stringency 1 x SSC, 65°C for 10 minutes; high stringency 0.1 x SSC, 65°C for 20 minutes). The filter was exposed to X-ray film.

DE-HYBRIDISATION OF FILTER: 500 ml 0.1 x SSC, 0.1% SDS were brought to the boil, the heat was turned off and the filter was immediately dropped in, face down. After cooling for about 30 minutes, this was repeated and then the filter was checked to see that all counts had been removed.

7.7 Northern hybridisation

HYBSOL: 1.5 x SSPE; 7% SDS; 10% PEG 6000

10 mg / ml ssDNA

100 mg / ml Heparin

METHOD:

The filter was pre-hybridised at 65°C for a minimum of 4 hours in 15 ml HYBSOL augmented with 100 µg / ml ssDNA and 250 µg / ml heparin. This was replaced with 15 ml of the same mixture, pre-warmed, plus the probe. The rest of the procedure was as described for Southern hybridisation.

7.8 Library screens

7.8.1 Preparation of plating cells

20% maltose, filter sterilised

10 mM $MgSO_4$, autoclaved

METHOD:

A single colony from a fresh (less than three weeks old) plate of the required cells was grown overnight in 4 mls LB + 0.5% maltose. 1 ml of this culture was added to 66 mls LB + 0.5% maltose and grown for 1 to 2 hours. The culture was transferred to 50 ml tubes and spun in a bench top centrifuge for 15 minutes at maximum speed. The supernatant was decanted and most of residual liquid was removed. The pellet was resuspended in 6 mls 10 mM $MgSO_4$ solution and stored at 4^oC.

7.8.2 Plating library

The required amount of library was added to 200 µl plating cells, mixed gently and added to a tube containing 8 mls top agarose kept at 48^oC. This was quickly mixed, poured onto a 14 cm agar plate and swirled to spread the agarose evenly. Once set, the plate was incubated at 37^oC for six to eight hours.

7.8.3 Lifting filters

The filter was labelled with pencil, wet in sterile water and blotted dry. It was placed on the plate, centre first, and its position marked with a needle. After a minimum of 30 seconds, the filter was carefully removed and laid DNA side up on 3MM paper soaked in denaturation solution. After 10 minutes, it was transferred to 3MM soaked in neutralisation solution and left for at least ten minutes. The filter was washed gently in 4 x SSC to remove any agar residue, then UV cross linked. Two filters were pulled from each plate.

7.8.4 Library rescreen

Positives were cored, added to 1 ml SM and left for a few hours at room temperature, agitating occasionally. The supernatant was diluted 1 in 1000 in SM, then 1 / 5 / 10 µl was added to 150 µl plating cells and incubated for 10 minutes at room temperature, followed by 10 minutes at 37°C. The procedures described in section 7.8.2 and 7.8.3 were then repeated.

7.9 Phage preparation

Caesium chloride (CsCl) solutions

1.3 g / ml: 40 g CsCl in 90 mls SM buffer

1.5 g / ml: 67 g CsCl in 82 mls SM buffer

1.7 g / ml: 95 g CsCl in 75 mls SM buffer

Fresh plate of *Escherichia coli*

METHOD:

1) 4 mls LB were inoculated with a colony of *E. coli* and grown overnight at 37°C.

2) The following day, a single phage plaque was taken into 200 µl of SM buffer and left at room temperature for six hours.

3) 150 µl of the culture from 1) was added to 2), mixed and added to 4 mls melted top agarose (kept at 48°C). This was poured onto a 9 cm phage plate and allowed to set. The inverted plate was incubated overnight at 37°C.

4) The next day, 5 mls SM buffer were added to plate 4) and it was swirled gently at room temperature for two to three hours.

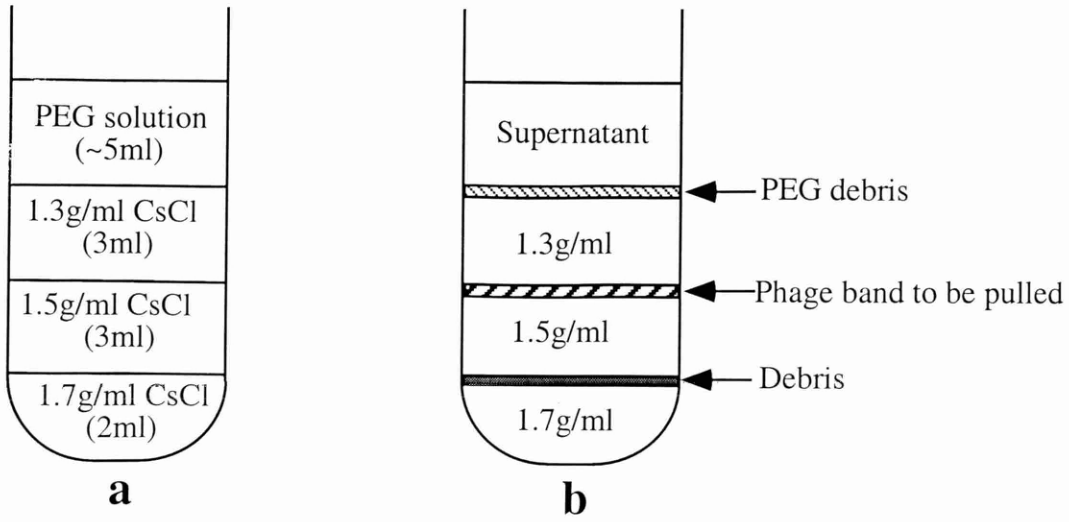
5) The remaining *E. coli* culture was added to 400 ml LB and grown for one and a half to two hours at 37°C.

6) The phage solution from 4) was added to 5) and incubated at 37°C with shaking until lysis was complete (three to four hours). If resistant colonies were present in the culture, a few drops of chloroform were added towards the end of the incubation period. 30 g

NaCl was dissolved in the culture, then it was spun at 7 K for 20 minutes at 15°C. The supernatant was decanted into a clean flask with 35 g PEG 6000, swirled gently to start dissolving and left at 4°C for a minimum of two hours until all PEG was dissolved. The culture was spun at 7 K for 20 minutes at 15°C. The supernatant was discarded and the pellet respun at 5 K for 5 minutes to compact it. The pellet was resuspended in 5 mls SM buffer. The phage suspension was carefully pipetted onto a CsCl gradient (see Figure 7.2a) and spun at 27 K for two hours at 15°C. The phage band (see Fig. 7.2b) was pulled using a 1 ml syringe and a 20 gauge needle in a volume of 0.5 to 1 ml. This was transferred to a 1.5 ml quick-seal centrifuge tube (Beckman Instruments Inc.), topped up with 1.5 g / ml CsCl solution and spun at 100 K for two hours at 15°C. The single band was pulled with a 1 ml syringe and 23 gauge needle, avoiding the wall of the tube that faced towards the outer edge of the rota during the spin as RNA collects here. The purified phage in CsCl solution was stored at 4°C.

EXTRACTION OF DNA FROM PHAGE: 500 µl of S M buffer, 25 µl of 10% SDS and 25 µl 0.5 M EDTA were added to 100 µl of phage CsCl preparation. This was vortexed, then extracted twice with an equal volume of phenol / TE and once with an equal volume of chloroform. The DNA was precipitated with 1 ml of ethanol, the pellet was washed, dried and resuspended in 100 µl TE buffer.

Fig.7.2 (overleaf) Phage gradients **a)** Preparation of step gradient: A 17 ml centrifuge tube (Sorvall Instruments Inc.), 3 mls each of 1.3 g / ml and 1.5 g / ml CsCl solutions and 2 mls of 1.7 g / ml CsCl were used for each gradient. The lowest density CsCl was put in the tube first and the next density was underlaid using a glass pasteur, taking care not to disturb the layers. The highest density was added in the same way. The gradients were handled carefully to prevent mixing of the layers. **b)** Band to be pulled after first spin (see text for details).



7.10 Phage mini-preparation

METHOD:

Steps 1) to 4) as for phage preparation (section 2.9).

5) 250 µl of the *E. coli* culture were added to 25 ml LB + 0.5% maltose and grown for one and a half to two hours at 37°C.

6) 500 µl from 4) were added to 5) and incubated at 37°C with shaking until lysis was complete (two to three hours). The culture was spun in a benchtop centrifuge at 8000 rpm for 10 minutes. The supernatant was decanted into a clean tube and 3 µl RNase (10 mg / ml) and 3 µl DNase (10 mg / ml) were added. This was incubated at 37°C for 30 minutes. 2 g NaCl and 2 g PEG 6000 were added and the sample was left at 4°C for a minimum of one hour, then spun at 10000 rpm for 10 to 20 minutes and the supernatant discarded. The pellet was respun briefly to compact it and then resuspended in 0.5 ml TE buffer. This was extracted with an equal volume of phenol / TE, then with an equal volume of phenol / chloroform, then with an equal volume of chloroform. It was precipitated with an equal volume of isopropanol, pelleted, washed in 70% ethanol and resuspended in 50 µl TE buffer.

7.11 Plasmid DNA mini-preparation

MERLIN I: 50 mM Tris HCl pH 7.5; 10 mM EDTA; 100 µg / ml RNase (DNase free). Store at 4°C.

MERLIN II (Prepared fresh for each use): 0.2 mM NaOH; 1% SDS

MERLIN III: 61.35 g solid potassium acetate; 35.7 ml glacial acetic acid. Made up to 500 ml with deionised water.

MERLIN IV: 66.84 g guanidine hydrochloride were added to 33.33 ml MERLIN III buffer in a very clean glass beaker and stirred with gentle heating. Most of the guanidine hydrochloride dissolved in 5-10 minutes but a fine precipitate remained. The pH was adjusted to 5.5 using 10 M NaOH (1.5-2 ml) and indicator paper, the volume was made up to 100 ml and the solution stirred with gentle heat for about 5 minutes. 1.5 g diatomaceous earth was weighed into a glass bottle and the guanidine hydrochloride solution was filtered through a 0.45µm filter into this bottle. The bottle was shaken to make a slurry. The solution also needed shaking before each use.

MERLIN V: 200 mM NaCl; 20 mM Tris HCl pH 7.5; 5 mM EDTA; 50% ethanol.

METHOD:

1.5 ml overnight culture was pipetted into a 1.5 ml microcentrifuge tube and spun in a microfuge for 30-40 seconds at maximum speed. The supernatant was discarded, a further 1.5 ml culture was added and the spin repeated. The supernatant was decanted and the remainder aspirated, leaving the pellet as dry as possible. The pellet was resuspended in 200 µl MERLIN I by putting on a shaker for 5 minutes. 200 µl MERLIN II was added (if the suspension did not clear almost immediately, it was mixed by inversion), followed by 200 µl MERLIN III. The sample was vortexed for 2-3 seconds and spun at maximum speed for 5 minutes. The supernatant was transferred into a tube containing 1 ml MERLIN IV. A Wizard® minicolumn (Promega) was fitted to a Vac-Man® Laboratory Vacuum Manifold (Promega) and the barrel of 2 ml syringe inserted into the column. With the pump on, the resuspended resin-DNA slurry was loaded onto

the column. The column was washed with 2 mls MERLIN V, transferred to its original tube and spun in the microfuge at maximum speed for 20 seconds to dry the column. The column was then transferred to a fresh tube and 50 µl TE buffer pH 7.5, preheated to 70°C, was loaded on. This was spun for 30 seconds to elute the DNA and the column discarded.

7.12 Gel purification of DNA fragments

(Based on method of Vogelstein and Gillespie, 1979)

Glass slurry: Made with 325 mesh powdered flint glass (from ceramic shops).

250 ml powder was added to an equal volume of water, stirred for 24 hours and allowed to settle for 30 minutes. The supernatant was spun at 7 K for 10 minutes and the resulting supernatant was discarded. The pellet was resuspended in 100-200 ml water and nitric acid added to 50%. This was brought close to the boil in the fume hood, allowed to cool and spun as before. The supernatant was discarded and the pellet washed in 100 ml water, spun and the supernatant discarded. The wash was repeated three more times and the pellet was stored as a 50% slurry in water. It was kept at 4°C for short term storage; -20°C for long term. The slurry was vortexed vigorously to resuspend immediately prior to use.

Sodium iodide solution: 90.8 g Sodium iodide; 1.5 g Sodium sulphite, made up to 100 ml with water and stirred well. The saturated solution was passed through an 0.45 µm filter into a bottle containing 0.5 g sodium sulphite. The solution was stored in the dark at 4°C and used within one month.

Ethanol wash solution: 50% ethanol; 0.1 M Na Cl; 10 mM Tris pH 7.5.

METHOD:

The DNA was run on a TA gel and the band/s required excised with a razor blade. Each gel slice was transferred to a 1.5 ml microfuge tube, weighed and at least 2.5 volumes of sodium iodide solution were added. The tube was incubated at 55°C for 5-10 minutes,

vortexing occasionally until the gel had dissolved. At least 5 µl glass slurry was added and the mixture was vortexed, then placed on a roller or shaker for a minimum of 60 minutes at room temperature. The tube was spun for 15 seconds at maximum speed in a microfuge and the supernatant discarded. The pellet was washed once with 1 ml sodium iodide solution and spun as before, then three times with 0.5 ml ethanol wash solution. The supernatant was aspirated, leaving the pellet as dry as possible and two glass volumes of TE buffer pH 7.5 or water were added. The DNA was eluted at 65°C for 60 minutes. The tube was spun in a microfuge for 15 seconds and the supernatant (containing the DNA) removed to a clean tube. The spin was repeated and the supernatant again removed to a clean tube.

7.13 Subcloning

7.13.1 Phosphatasing vector

10 to 20 µg of cut and precipitated vector were treated with 1-2 units of CIP (calf intestinal phosphatase, from Boehringer Mannheim) in the presence of 1 x phosphatase buffer (supplied with the enzyme). This was incubated at 37°C for 30 minutes, made up to 200 µl with water, extracted twice with half volume of phenol / TE and three times with half volume of chloroform (as described in section 2.3). It was precipitated with 2.5 volumes of ethanol and 0.1 volumes of 3 M sodium acetate, pH 5 and resuspended at 100 µg / ml in TE buffer pH 7.5. The vector was stored at 4°C (short term) or -20°C (long term).

7.13.2 Removal of very small fragments from digest

To remove very small fragments of DNA, digests were precipitated with 0.6 volumes of isopropanol and 0.1 volumes of 3 M sodium acetate pH 5 and the supernatant discarded.

7.13.3 Ligation

100 to 200 ng insert DNA was mixed with 100 to 200 ng cut vector DNA (phosphatased if necessary), buffer (supplied with enzyme) to 1 x (final concentration) and 0.1 units T_4

ligase. This was made up to 10 µl with sterile water and incubated at 16°C overnight.

7.13.4 Modifications for blunt ended ligation

Blunt ended ligation is a relatively inefficient reaction, so the amount of both insert and vector DNA was increased to 200 to 500 ng and 0.5 units of ligase were added.

7.13.5 In-filling of sticky ends for blunt ended ligation

dNTPs: 0.5 mM dATP; 0.5 mM dCTP; 0.5 mM dGTP; 0.5 mM dTTP

METHOD:

0.2 volume (final concentration) dNTPs, 0.1 volume (final concentration) Klenow buffer (supplied with enzyme) and 5 units Klenow (supplied by Bioline) was added to cut and precipitated DNA and incubated at 37°C for 30 minutes. The volume was made up to 200 µl with water, the solution was extracted once with phenol, once with chloroform and ethanol precipitated.

7.13.6 Preparation of competent cells

Competent cells were prepared by Carol McKimmie with some assistance from myself, following the protocol of Hanahan (1985).

RFI: 100 mM NaCl; 30 mM potassium acetate; 10 mM $CaCl_2.2H_2O$; 15% glycerol were mixed together and the pH adjusted to 6 with 0.2 M acetic acid. 50 mM $MnCl_2.4H_2O$ was added, and the final pH was 5.8.

RFII: 10 mM MOPS; 10 mM NaCl; 75 mM CaCl.2HO; 3.75% glycerol (w/v). pH adjusted to 6.8 with 0.1 M NaOH.

Very clean glassware and MilliQ water were used for this protocol.

METHOD:

200 ml of SOB were inoculated with 1 ml of overnight culture of cells and grown at 37°C with shaking until the density was 4 to 7 x 10^7 viable cells / ml (O.D. = 0.4 to 0.5). The

culture was then collected in 50 ml tubes, incubated on ice for 15 minutes and centrifuged at 3,000 rpm for 12 minutes. The supernatant was decanted and the remainder drawn off with a blue tip. The pellet was gently resuspended in RFI (one third of original volume), incubated on ice for 15 minutes and centrifuged as before. The pellet was resuspended in RFII (1 in 12.5 of original volume) and incubated on ice for 15 minutes. 200 µl aliquots were pipetted into chilled 1.5 ml tubes, flash frozen in liquid nitrogen and stored at -70°C.

7.13.7 Transformation

Competent cells were thawed on ice and 100 µl aliquoted into a pre-chilled tube. 5 µl of the ligation reaction were added to the cells and left on ice for 20 to 30 minutes, then heat shocked (42°C for 30 seconds or 37°C for 5 minutes). The mixture was transferred to a 15 ml tube with 1 ml LB and grown for one hour at 37°C with shaking. 200 µl were spread on a 9 cm agar plate containing the appropriate antibiotic. The plate was left for a few minutes at room temperature, inverted and incubated at 37°C overnight. Colonies picked from the plate were grown in 4 mls LB with appropriate antibiotic 37°C overnight.

BLUE-WHITE SELECTION:

A simple colour test has been developed to distinguish between vectors with and without foreign DNA inserts (Messing *et al*, 1977). Host cells that carry a defective ß-galactosidase gene (*lacZ*) were constructed. The association between this host and the vector alone (that is, with no insert) forms an enzymatically active protein. Blue plaques are therefore produced in the presence of isopropylthio-ß-D-galactoside (IPTG, a gratuitous inducer of ß-galactosidase) and 5-Bromo-4-chloro-3-indolyl-ß-D-galactoside (X-gal, a chromogenic substrate). Insertion of foreign DNA into the *lacZ* region of the vector prevents the complementation between vector and host, resulting in the formation of pale or white plaques.

X-gal: 20 mg / ml in dimethylformamide. Aliquots of 0.5 ml were wrapped in foil and stored at -20°C.

IPTG: 20 mg / ml in water. Filter sterilised and stored in 0.5 ml aliquots at -20°C.

METHOD:

40 µl X-gal and 40 µl IPTG were mixed together, spread on a plate and allowed to dry briefly before the transformation mix was added.

7.13.8 Transformations with Original TA Cloning® Kit (Invitrogen)

One Shot cells: INV F' cells

SOC medium: 2% tryptone; 0.5% yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM glucose (dextrose).

METHOD:

One vial of One Shot cells was thawed on ice. 2 µl 0.5 M ß-mercaptoethanol and 2 µl of the ligation reaction were added, mixed gently and incubated on ice for 30 minutes. The reaction was heat shocked at 42°C for 30 seconds then chilled on ice for two minutes. 250 µl of SOC medium were added and the mixture incubated at 37°C for one hour with shaking. 50 µl were spread on a plate containing carbenicillin, using blue-white selection. (Addition of IPTG is not necessary as INV F' does not express the lac repressor.) The plate was incubated at 37°C for at least 18 hours.

7.14 PCR (Polymerase Chain Reaction)

10 x buffer: 67 mM $MgCl_2$; 166 mM $(NH_4)_2SO_4$; 670 mM Tris HCl; 0.17% BSA; pH 8.4

dNTP mix: 5 mM dATP; 5 mM dCTP; 5 mM dGTP; 5 mM dTTP

DMSO: Dimethyl sulphoxide

Primers: Supplied by Department of Biochemistry, University of Cambridge; MWG or

Genosys. Stock concentration 100 pmol / µl in TE. Stored in 50 µl aliquots at -20°C. Diluted 1 in 10 in water for working concentration.

METHOD:

5 µl dNTP mix, 10 µl buffer, 5 µl DMSO, 5 µl of each primer, 0.2 to 2 µg DNA (If subsequent rounds were required, 1 µl of a one in 100 dilution of the previous round was used.), 0.5 µl Taq polymerase (5 units / µl , supplied by Bioline) and water to 100 µl were mixed together. The reaction was covered with a drop (approx. 50 µl) of mineral oil and placed in PCR machine.

PROGRAMMES:

The temperature for the annealing stage of the reaction was chosen by estimating the melting temperature of the primers being used. This was done by giving each C and each G in the primer a value of four, and each A and each T a value of two. Summing these values gave an approximate melting temperature. The annealing temperature selected initially was slightly lower than that of the lowest melting temperature of the two primers. Primer pairs were designed so that their melting temperatures were as similar as possible. If the background was high, the reaction was repeated with a slightly higher annealing temperature.

The length of the amplification step depends on the size of the expected product. Approximately one minute per kb was allowed at 72°C.

The number of cycles depended on the concentration of the target DNA in the reaction mixture. The amount of Taq usually becomes limiting after 25 to 30 cycles.

Programme 1: 93°C for 45 seconds; 55°C for 2 minute; 72°C for 2 minutes; 25 cycles

Programme 2: 93°C for 30 seconds; 58°C for 1 minute; 72°C for 4 minutes; 30 cycles.

Programme 3: 93°C for 45 seconds; 66°C for 1 minute; 72°C for 3 minutes; 30 cycles.

Programme 4: 94°C for 45 seconds; 70°C for 1 minute; 72°C for 1 minute; 30 cycles.

7.15 Marathon™ RACE PCR (CLONTECH Laboratories, Inc.)

cDNAs may be generated by using PCR to amplify copies of the region between a point in a known transcript and the 5' or 3' end. This technique, known as RACE (rapid amplification of cDNA ends) was first described by Frohman *et al* (1988). One primer is designed from the known sequence of the transcript, the other from the vector. A second primer pair, just inside the first, may also be required.

Marathon™ RACE reactions were designed to allow amplification of large templates and may be used to obtain full length cDNAs (Fig. 7.3). The procedure involves production of an adapter ligated, double stranded cDNA library; the experiments in this thesis used a cDNA library previously produced by David Huen from polyadenylated mRNA isolated from 0-18 hour Canton S embryos.

DNA polymerase: Mixture of Taq (3.5 units / µl) and Pwo (concentration not given) (Expand™ Long Template PCR System, Boehringer Mannheim).

TaqStart Antibody (CLONTECH Laboratories, Inc.).

50 x polymerase mix: 14.3 µl of Taq /Pwo and 5.7 µl of TaqStart.

Buffer 1 (supplied with Expand system): PCR buffer 10 x concentration with 17.5 mM $MgCl_2$.

Master Mix (for each reaction): 5 µl 10 x PCR buffer 1; 36 µl H_20; 1 µl dNTP (10 mM); 1 µl 50 x polymerase mix.

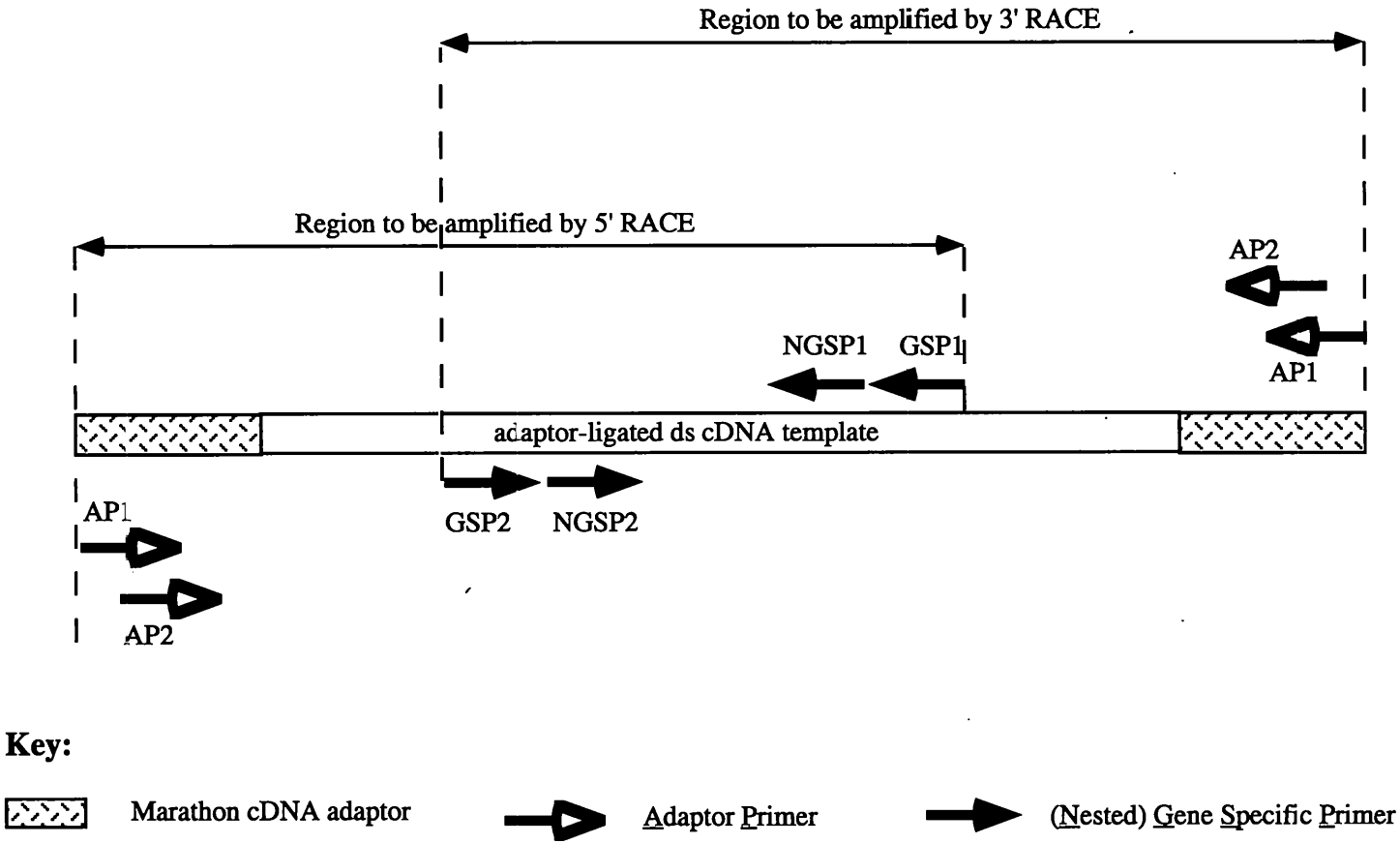


Fig.7.3 Marathon RACE reactions
(based on diagram in Marathon cDNA Amplification Kit manual, CLONTECH Laboratories, Inc.)

METHOD:

The required amount of DNA was pipetted into an 0.5 ml tube with 50 pmol of vector primer, 50 pmol of internal primer, 43 µl of Master Mix and H_2O to 50 µl . This was overlaid with 30 µl of oil and run on the relevant programme.

7.16 Sequencing

7.16.1 Design of oligonucleotide primers

Primers for cycle sequencing were designed to be at least 18 bases long; not to have long runs (more than 3 or 4) of single bases, especially G or C; to have G + C content of at least 50%. They were resuspended at 100 pmole / µl in TE buffer and stored in 50 µl aliquots at -20°C. Suppliers as before (section 2.14).

7.16.2 Reactions

Sequencing reactions were carried out using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer Corporation).

Terminator Ready Reaction Mix (supplied with kit): A-Dye Terminator; C-Dye Terminator; G-Dye Terminator; T-Dye Terminator; dITP; dATP; dCTP; dTTP; Tris HCl (pH 9); $MgCl_2$; thermal stable pyrophosphatase and AmpliTaq DNA Polymerase, FS.

Oligonucleotide primer: Diluted to a working concentration of 4 pmole / µl in water.

REACTIONS:

8 µl Terminator Ready Reaction Mix, 2.5 µl double stranded DNA (0.2 µg / µl) or 3 µl of 1 in 10 dilution of PCR product (10-30 ng / µl), 1 µl primer (4 pmole / µl) and water to 20 µl were mixed together, then overlaid with 1 drop (about 40 µl) mineral oil. The reactions were cycled according to the following programme.

PCR PROGRAMME: The block was preheated to 96°C, the reactions put in and the

programme run immediately:

96°C for 30 seconds; 50°C for 15 seconds; 60°C for 4 minutes.

25 cycles.

PURIFICATION OF EXTENSION PRODUCTS: The above reaction mix was transferred to a clean tube (care was taken not to transfer any oil) and precipitated with 2 µl 3 M sodium acetate pH 4.6 and 50 µl ethanol on ice for 10 minutes. The precipitate was microfuged at maximum speed for 15 to 30 minutes. Ethanol was removed using a pipetteman and the pellet washed with 250 µl 70% ethanol. After microfuging for 5 minutes, as much ethanol as possible was removed and the pellet dried in the PCR machine for 1 minute at 85°C. Samples were stored at -20°C till run.

7.16.3 Analysis

Initial analysis of raw data was carried out using SeqEd v1.0.3 (Applied Biosystems, Inc.). The data was then processed using the GCG package (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc.). The following programmes were used:

Assemble constructs new sequences from pieces of existing sequences.

BestFit makes an optimal alignment of the best segment of similarity between two sequences (Lipman *et al*, 1984).

Composition determines the composition of sequences.

FindPatterns identifies sequences that contain short patterns provided by the operator.

LineUp is a screen editor for editing multiple sequence alignments.

Map maps a DNA sequence and displays both stands of the mapped sequence with restriction enzyme cut points above the sequence and protein translations below. Map can also create a peptide map of an amino acid sequence.

Motifs looks for sequence motifs by searching through proteins for the patterns defined in the PROSITE Dictionary of Protein Sites and Patterns (written by Dr. Amos Bairoch of the University of Geneva).

SeqEd is an interactive editor for entering and modifying sequences and for assembling

parts of existing sequences into new genetic constructs.

Translate translates nucleotide sequences into peptide sequences.

Blast searches were carried out using BLAST 1.4 and BLAST 2.0 (Altschul *et al*, 1990; Warren Gish and Webb Miller, 1994-1997, unpublished) at the National Center for Biotechnology Information (NCBI) and the Berkeley *Drosophila* Genome Project . The following programmes were used:

blastp compares an amino acid query sequence against a protein sequence database.

blastn compares a nucleotide query sequence against a nucleotide sequence database.

blastx compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

7.17 RNA isolation using TRIZOL Reagent (GIBCO BRL Life Technologies)

NB. Gloves and eye protection were worn when using TRIZOL (Total RNA isolation) Reagent as it is toxic and causes burns. RNase free reagents and plastics were used.

50-60 flies were homogenised in 500 µl of TRIZOL Reagent using a pestle (Burkard Scientific Ltd). The sample was incubated at room temperature for five minutes and 100 µl o chloroform was added. The tube was shaken vigorously by hand for 15 seconds and incubated at room temperature for two to three minutes. It was spun in a microfuge at 13000 rpm for 15 minutes at 4°C. The mixture separated into a lower, red, phenol-chloroform phase, an interphase and an upper, colourless, aqueous phase. The aqueous phase, containing the RNA, was transferred to a fresh tube and precipitated by mixing with 250 µl of isopropyl alcohol. The sample was incubated at room temperature for ten minutes and centrifuged for 10 minutes at 13000 rpm at 4°C. The supernatant was removed and the pellet washed with 500 µl of 70% ethanol. The tube was centrifuged at 7000 rpm for five minutes at 4°C. The pellet was air-dried for five minutes and resuspended in RNase free water.

7.18 First strand cDNA synthesis using SUPERSCRIPT™ II RNase H- Reverse Transcriptase (GIBCO BRL Life Technologies).

SUPERSCRIPT II RNase H Reverse Transcriptase 200 units / µl

5X First Strand Buffer (supplied with kit): 250 mM Tris-HCl pH 8.3; 375 mM KCl; 15 mM $MgCl_2$.

0.1 M DTT (supplied with kit).

10 mM dNTP mix: 10 mM each dATP, dCTP, dGTP and dTTP.

E. coli RNase H (Promega) 2 units / µl

METHOD:

2 pmoles of gene specific primer were mixed with 1-5 µg of RNA and made up to 12 µl with deionised water. The mixture was heated to 70°C for ten minutes and quick chilled on ice. The contents of the tube were briefly centrifuged. 4 µl of First Strand Buffer, 2 µl 0.1 M DTT and 1 µl 10 mM dNTP mix were added to the tube, mixed gently and incubated at 42°C for two minutes. 1 µl of SUPERSCRIPT II was added, mixed gently and the tube incubated at 42°C for 50 minutes. The reaction was inactivated by heating at 70°C for 15 minutes. After cooling, RNA complementary to the cDNA was removed by adding 1 µl of RNase H and incubating at 37°C for 20 minutes. 2 µl of the cDNA was used as a template for amplification by PCR.

7.19 Flies

7.19.1 Culture

Flies were grown in 250 ml glass bottles or 2.5 x 10 cm tubes on standard cornmeal agar supplemented with yeast. They were kept at 25°C (unless otherwise stated) with 65% relative humidity and a 12 hour light / dark cycle.

7.19.2 Stocks

The following stocks were used in the course of this study:

y w

w; In(2LR)O, Cy dp^{lvl} pr cn^2/Sco; In(3LR)TM2, emc^2 Ubx^{130} e^s/In(3LR)TM6B,

Hu e Tb ca (w; CyO/Sco; TM2/TM6),

w; In(2LR)O, Cy dp^{lvl} pr cn^2 P{ry$^+$ wg:lacZ}/If; In(3LR)TM6B,

Hu e Tb ca/Tp(3;3)MKRS, M(3)76A kar ry^2 Sb (w; CyO/If; MKRS/TM6B)

w; Df(2R) pk^{30}

w; pk^{sple1}

pk pk-sple7

pk pk-sple8

pk pk-sple12

pk pk-sple25

w; pk sple13

GAL4^{C765}

nec^1 bw^D/CyO

y w; Df(2R)pk-78k/CyO

7.20 Transformation rescue

7.20.1 Constructs

pk and *sple* constructs:

The cDNA in the pNB40 vector was digested with *Hin*dIII and the cut end was in-filled to give a blunt end section (7.13.5). The cDNA was then cut with *Not*I and the 4.2kb cDNA fragment was gel purified (section 7.12). The GAL4 UAS vector pUAST (Brand and Perrimon, 1993; Appendix D) was cut with *Eco*RI, in-filled and cut with *Not*I. The gel purified fragment was subcloned into this vector following standard procedure for blunt ended ligation.

The presence of the insert was confirmed by a double *Hin*dIII / *Not*I digestion. The

insert was double checked by sequencing the ends using oligonucleotide primers from the cDNA (2230 and 9886 for *pk*; splec3 and 9886 for *sple*).

Serpin constructs:

Genomic constructs of each of the three serpins were made using the *pwhiteRabbit* transformation vector (Dunin-Borkowski and Brown, 1995). A 3.2 kb *Eco*RI restriction fragment from FP11/3 spanning *Spn43Aa* and a 7 kb *Xho*I-*Bam*HI fragment from FP10/2 covering *Spn43Ac* were cut from the genomic insert phage; the 5.2 kb *Eco*RI fragment for *Spn43Ab* was cut from a cosmid (Cos 9/5) isolated from the cosmid library of J. Tamkun.

Spradling buffer: 5 mM KCl; 0.1 mM phosphate buffer, pH 7.8.

METHOD:

1 µg / µl final concentration of the construct was mixed with 0.25 µg / µl of the helper plasmid pHsπΔ2-3. (A minimum final volume of 20 µl was required.) The mixture was then ethanol precipitated, washed, dried and resuspended in Spradling buffer. The constructs were microinjected into *yellow white* (*yw*) embryos by Carol McKimmie, following the method of Spradling (1986).

7.20.2 Crosses

All the flies that hatched from the surviving larvae, known as G_0s, were crossed individually to *yw* flies. Putative G_1 transformant progeny, identified by eye colour (ie. any flies that were not *w*) were crossed to *w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2*/Sco; In(3LR)TM2,* ome^2 Ubx^{130} e^s*/In(3LR)TM6B, Hu e Tb ca* (*w; CyO/Sco; TM2/TM6*), or *w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2 *P{*ry^+ *wg:lacZ}/If; In(3LR)TM6B, Hu e Tb ca/Tp(3;3)MKRS, M(3)76A kar* ry^2 *Sb* (*w; CyO/If; MKRS/TM6B*) flies to establish balanced stocks in a *w* background.

7.21 Mounting wings and legs

Flies were fixed and stored in 95% ethanol. They were transferred to a watchglass and viewed at 12 x magnification. Using watchmaker's forceps, a fly was held steady and the right wing pulled off, as near to the hinge as possible. The wing was placed in another watchglass containing 50 : 50 lactic acid : ethanol. The right wing was removed from the remaining flies. Using a disposable pasteur, the wings and the fix were placed onto a slide. Excess fix was removed, a coverslip was positioned over the wings and sealed with nail varnish after being allowed to air dry. The procedure was repeated with all the left wings. Legs were mounted in the same way.

APPENDICES

A. OLIGONUCLEOTIDE PRIMERS

All primers read 5' to 3'.

A.1 pBluescript

Forward	GTA AAA CGA CGG CCA GT
Reverse	AAC AGC TAT GAC CAT G

A.2 Sequencing of *pk* (cDNA and genomic)

97	GTA CTC CTC CAG AGC ACA
199	CAT GCG GGA GGG CAA ACC
911	TGC CTG GGC CGC GTA GAG
912	GCC TGT TGA GCT AGA AAG
935	CCT GGG CCG CGT AGA GGG
1331	CAG CTC GCC GCC TAT GTC
1332	CCC AGT GTC ACC AGT GAA
1333	GAG CAC AGG TCC AGA TGC
1334	ATG GGC GTG TTC AGC GAC
1894	CGC TTG TAC TTC TCG CAG
1895	TCT GCG GCA GCT GGT ACA
1896	GAG AGC GAG AGA AAC GAG
1897	GAT CGC AGA TAC GGA GTA G
2010	GTG AAC GGG AGT CCA GAA
2011	GCG TTC TAC GGG CTG ATG
2012	ATA AGG TCC ACG AGG AGC
2230	GTT CCG TTT CCG TGC CTT
2492	CGA CGC GGC GTA GCT CTT
2493	CAC CAG TTC CAG CAG CAC
3042	GCA CCA GTT CCA GCA GCA CGA
3043	CAT AGG TGT CCG CTC GAG TGG
3044	AGG AGT GAC CGT GAC CAT GGC

3571	TGC GGC GGA AGT TGA TGG AGC
9884	TCG TCG TCC GAG TGC GAC
9885	CTG CGA GAA GTA CAA GCG
9886	ACC GAT AGC ATA GCG AAG
clare6	GAT CAG GTG AGA AAG CCT
clare7	ATC TGC CCG GCA GAA GTG
clare8	CCA ACT CGA ATT ACG GGC
clare9	GCA ATG ACA GCG AGT AAC
clare10	ATT GGC CAC GCT CGC CAA

A.3 Sequencing of *sple* exons

x16f	AAC CTC TAC TGC GGT CGG
x16r	GTG CTC CTG GTT GCG ATT
splec1	ACA GGC TGC TCA TGT TGC
splec2	GCC ACA CGG TTC CTA ATC
splec3	AGC CTG TCA ACC GGT GGA
splec4	AGA TGC CGT CGA GGA TGC
splec5	AGG TCC TGG AGA TCA TCG

A.4 Sequencing of *pkM* exon

65FA	CAC ACA CTG CCA GGC TCT
65RA	GCA TGC CGA CTG CGA CGG

A.5 Sequencing of human EST clone 241420

2707	GGG AGA CAG CTA TAG TAC
2708	TCT GAT GGA TGT CAG CTG

A.6 Sequencing of *esn* cDNA

CE2midf	TAC TGC TCG ATC GCC TGT
CE2midr	TTT CCT GCC GCT CAA GCA
CE201	AGT GCT TCA TGT GCC ACG

CE202	CGC ACA GTT ACC AGA GCA
CE203	TAC TGC GGT CGG CAC CAT
CE204	AAC CTC CGC CTG CAC AAC
CE25E	TCT GGA ACT CTG GAG CTC
CE23E	GCA ACG CAG AAC GCG AAA

A.7 Sequencing of *spn43Aa* (cDNA and genomic)

2709	CGG AGA ACA AGA TCG AAC
2710	GTT CTC AGA TGC AGC CGA
3335	TCT CTT CGC CAC CGA ACT
3336	TAG CTC CAG CCG CCT CAC
5796	CGT TCA CAT CGA TGA AGG TC
5797	ACT GCA GAA GAC CTT GCA CG
5898	ATC GCG AGT TCT GGC TGA GC
5899	CAG CGC CAC ATT GGT GTC C
6228	CAT CGG CAC TCG GAT CAC A
6229	CCG GCG AGA TGA TCA CGT T
6316	CGA CAG AGC CTG ATG CGC TT

A.8 Sequencing of *spn43Ab* (cDNA and genomic)

3045	TGG TAA CTA CCG CTG CGC
3201	CAA GCG GAG CAG GAG ACG
6916	TCC ACG CCA GCT TCA GTG A
6917	CAG ACG GCA TCG GAG TTG C
7074	CAC CTT GGC GTC CAA GGA G
7075	AGT GAT CGT GAA CGG CGT T
7854	GTT CTC GAG AGA TGG TGA C

A.9 Sequencing of *spn43Ac* (cDNA and genomic)

6918	CCA GCT CGC GGT TGT AGT A
7851	GAG CTC GGA GGT GAA GCG
8039	AGG TCT GAG CAG CCA GAA G
8214	GTG CGC TGC CAA GTT CCA

8215		AGG ATC TTG CTC ACG CGC
8216		CGT AAC CCA AGA CCG GAG
9258		GAA CTG CAG AAG GCC GGA G
9259		GTT CGA TCT CAA CCG CGT TG
clare3		AAG GGT GGA GCC GAA CTG
clare4		GAC AGC ACC GCT ACA GGT
clare5		AGA GCA GCT AAT GCC ACG

A.10 PCR

pBluescript	4457	CAC TAA AGG GAA CAA AAG CTG G
	4458	ACT CAC TAT AGG GCG AAT TGG
pk	DH1	CGA GAA GTG GCG CGG AAG CGT
	DH2	CTG GTT AGC GGA GTT CGG CTG AT
	sma1	GGA GCA GCC TGG CGA GTA CTG
	sma2	TAG CTG GTG GTG GTG GTG CGG
sple	splei	CGA TCG ATC GAT CGC GCG CTC
	spleii	GCG CCA GAT CTG TCG TTT ACC G
pNB40	722	GTC GTT AGA ACG CGG CTA
	852	GAA TAC AAG CTT GCT TGT TC

B. Subclones generated

B.1 Pk4aiii cDNA

*Sst*I/*Eco*RI 0.8kb and 1.4kb;
*Xho*I/*Not*I 0.7kb;
*Sal*I/*Eco*RI 1.2kb;
*Apa*I/*Eco*RI 1.4kb;
*Apa*I 1.7kb;
*Pst*I 0.4kb and 0.2kb;
*Pst*I/*Xho*I 0.9kb.

B.2 Pk4aiii region genomic

FP5/1 *SalI*/*Eco*RI 3.5kb, 4.2kb and 4.3kb;
FP5/1 *Hin*dIII/*Xho*I 1.6kb, 2.2kb, 3.9kb and 5.2kb;
FP6/2 *SalI*/*Xho*I 1.2kb, 1.8kb and 4.6kb;
FP6/2 *SalI* 3.0kb;
FP6/2 *SalI*/*Eco*RI 1.3kb and 1.6kb;
FP6/2 *Hin*dIII/*Eco*RI 1.1kb and 1.4kb.
FP11/3 *SalI*/*Eco*RI/*Xho*I 2.2kb
SpeI/*Xho*I 1.7kb subclone from Canton S

B.3 *sple* exons

SpeI/*Bst*XI 0.5kb;
SpeI/*Xho*I 0.4kb;
SpeI/*Sal*I 0.7kb

B.4 *esn* cDNA

XbaI/*Xho*I 0.6kb and 0.7kb;
*Xho*I 1.4kb and 1.6kb;
*Sst*I 0.6kb and 0.9kb;
*Pst*I 0.4kb and 0.6kb.

B.5 Serpin region genomic

FP11/3 *Eco*RI 3.2kb;
FP10/2 *Sal*I 3.0 and 1.3kb;
FP10/2 *Eco*RI/*Sal*I 1.1kb
FP10/2 *Eco*RI/*Hin*dIII 1.1kb.

C. ABBREVIATIONS

C.1 Amino acids

A	Alanine	I	Isoleucine
R	Arginine	L	Leucine
N	Asparagine	K	Lysine
D	Aspartic acid	M	Methionine
B	Asparagine or aspartic acid	F	Phenylalanine
C	Cysteine	P	Proline
Q	Glutamine	S	Serine
E	Glutamic acid	T	Threonine
Z	Glutamine or glutamic acid	W	Tryptophan
G	Glycine	Y	Tyrosine
H	Histidine	V	Valine

C.2 General

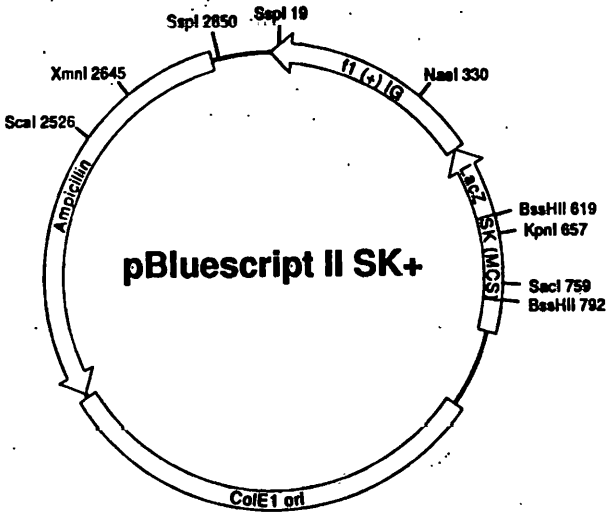
bp	nucleotide (base) pairs
C	Celsius
cDNA	complementary DNA
Ci	Curies
D.	*Drosophila*
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	dATP/dCTP/dGTP/dTTP
dTTP	deoxythymidine triphosphate
DNA	deoxyribonucleic acid

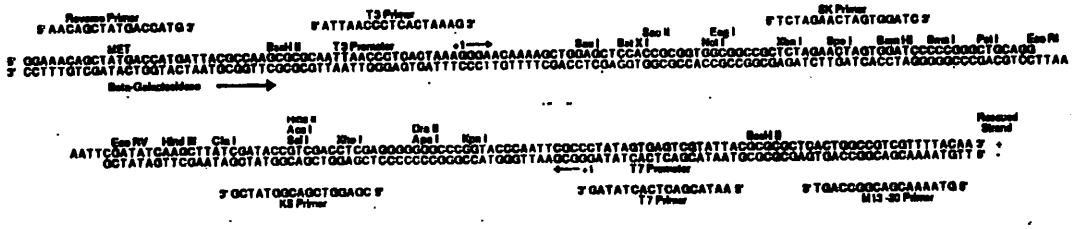
EDTA	ethylenediaminetetra-acetic acid
g	gramme
HD	homeodomain
kb	kilobases
l	litre
m	milli
M	molar
mRNA	messenger RNA
n	nano
ORF	open reading frame
p	pico
^{32}P	phosphorus 32
PCR	polymerase chain reaction
PEG	polyethylene glycol
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
ssDNA	single stranded DNA
TE	Tris EDTA
Tris	tris[hydroxymethyl] aminomethane
μ	micro

D. VECTORS

D.1 pBluescript SK+ (Stratagene)

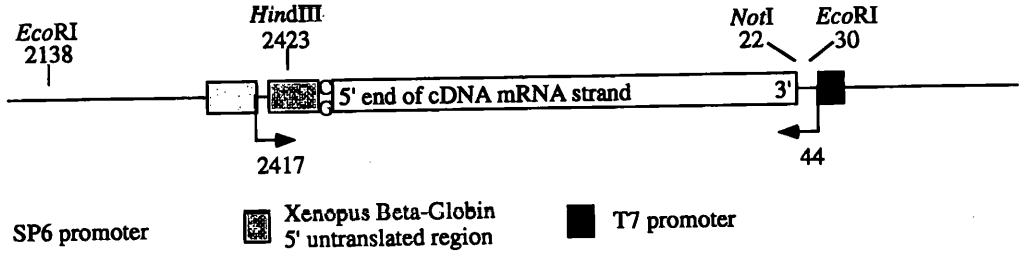
(Diagram taken from Stratagene protocol booklet.)





D.2 pNB40 (Brown and Kafatos, 1988)

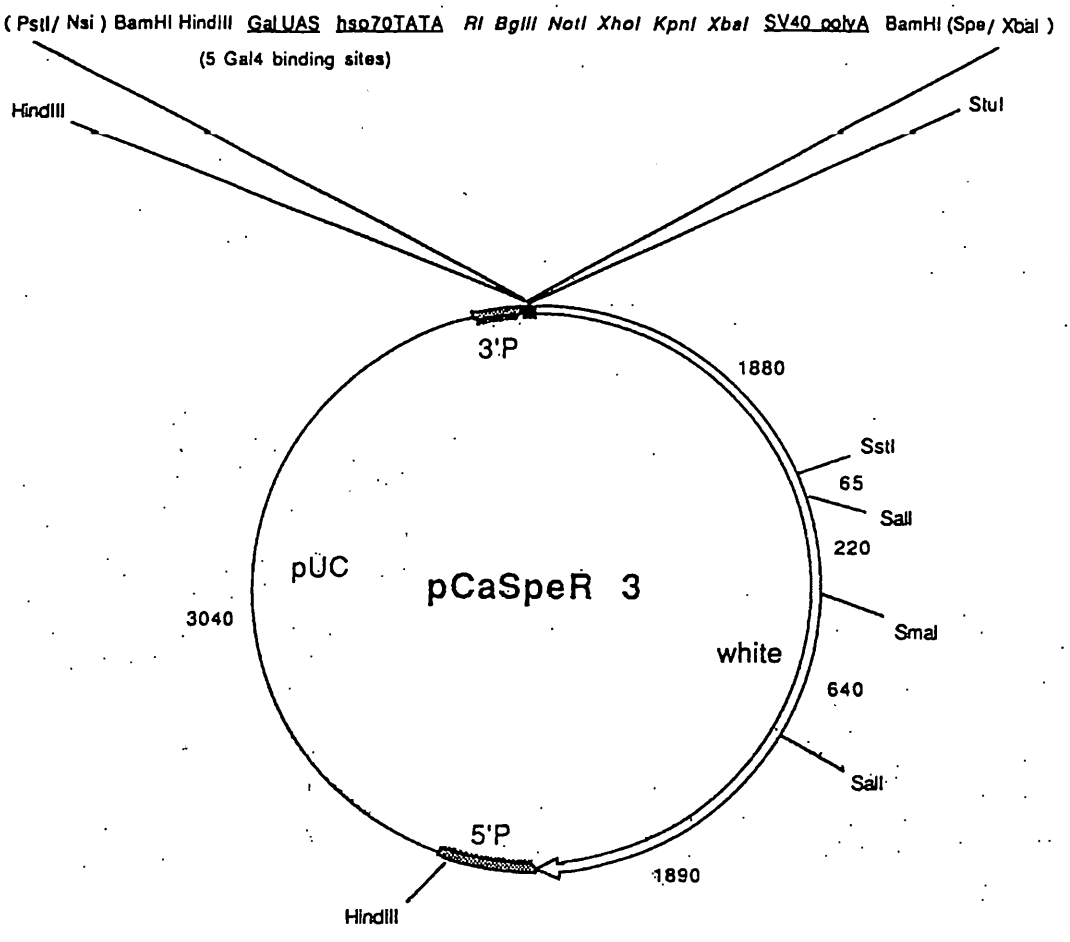
(Diagram based on drawing in "pNB40 *Drosophila* cDNA libraries";
Nick Brown, unpublished)



pNB40 cDNA vector ~2490 bp. Restriction enzyme numbers are cut sites
(only selected restriction enzyme sites shown).

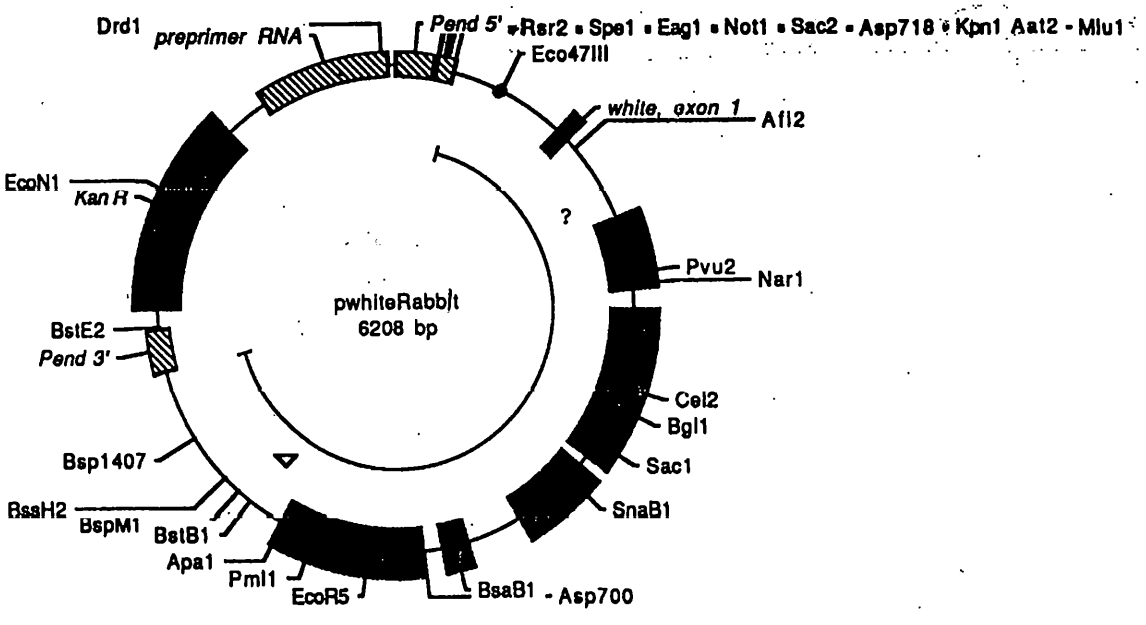
D.3 pUAST (Brand and Perrimon, 1993)

(Diagram from Brand and Perrimon, personal communication)



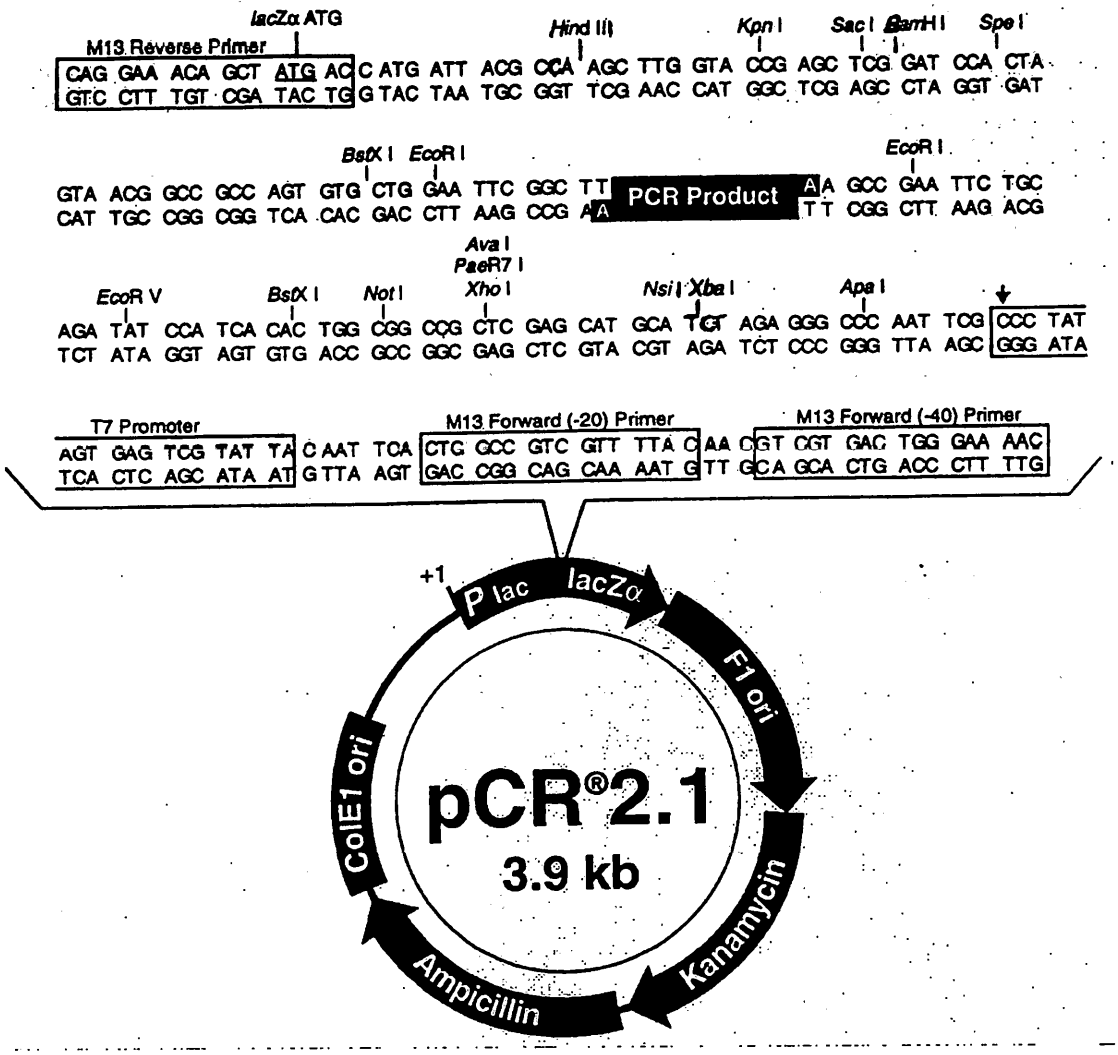
D.4 pwhiteRabbit (Dunin-Borowski and Brown, 1995)

(Diagram from Nick Brown, personal communication)



D.5 pCR 2.1 (Invitrogen)

(Diagram taken from Invitrogen instruction manual)



REFERENCES

ADLER, P. N. (1992) The genetic control of tissue polarity in *Drosophila*. BioEssays **14**, 735-741.

ADLER, P. N., CHARLTON, J. and VINSON, C. (1987) Allelic variation at the *frizzled* locus of *Drosophila*. Dev. Genet. **8**, 99-119.

ADLER, P. N., CHARLTON, J., JONES, K. H. and LIU, J. (1994) The cold-sensitive period for *frizzled* in the development of wing hair polarity ends prior to the start of hair morphogenesis. Mech. Dev. **46**, 101-107.

ADLER, P. N., CHARLTON, J. and PARK, W. J. (1994) The *Drosophila* tissue polarity gene *inturned* functions prior to wing hair morphogenesis in the regulation of hair polarity and number. Genetics **137**, 829-836.

ADLER, P. N., CHARLTON, J. and LIU, J. (1998) Mutations in the cadherin superfamily member gene *dachsous* cause a tissue polarity phenotype by altering *frizzled* signalling. Development **125**, 959-968.

ADLER, P. N., KRASNOW, R. E. and LIU, J. (1997) Tissue polarity points from cells that have higher Frizzled levels towards cells that have lower Frizzled levels. Curr. Biol. **7**, 940-949.

ADLER, P. N., VINSON, C., PARK, W.J., CONOVER, S. and KLEIN, L. (1990) Molecular structure of *frizzled*, a *Drosophila* tissue polarity gene. Genetics **126**, 401-416.

AGULNICK, A. D., TAIRA, M., BREEN, J. J., TANAKA, T., DAWID, I. B. and WESTPHAL, H. (1996) Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. Nature **384**, 270-272.

AKAM, M. (1987) The molecular basis for metameric pattern formation in the *Drosophila* embryo. Development **101**, 1-22.

ALEXANDROPOULOS, K., CHENG, G. and BALTIMORE, D. (1995) Proline-rich sequences that bind to Src homology 3 domains with individual specificities. Proc. Natl. Acad. Sci. **92**, 3110-3114.

ARBER, S. and CARONI, P. (1996) Specificity of single LIM motifs in targeting and LIM/LIM interactions in situ. Genes and Dev. **10**, 289-300.

ARBER, S., BARBAYANNIS, F. A., HANSER, H., SCHNEIDER, C., STANYON, C. A., BERNARD, O. and CARONI, P. (1998) Regulation of actin dynamics through phosphorylation of cofilin by LIM-kinase. Nature **393**, 805-809.

ARTAVANIS-TSAKONAS, S. (1999) Notch signalling: Cell fate control and signal integration in development. Science **284**, 770-776.

ARTAVANIS-TSAKONAS, S., MATSUNO, K. and FORTINI, M. E. (1995) Notch signalling. Science **268**, 225-232.

AVERY, L. and WASSERMAN, S. (1992) Ordering gene function: the interpretation of epistasis in regulatory hierarchies. TIG **8**, 1992.

AXELROD, J. D., MATSUNO, K., ARTAVANIS-TSAKONAS, S. and PERRIMON, N. (1996) Interaction between Wingless and Notch signalling pathways mediated by Dishevelled. Science **271**, 1826-1832.

AXELROD. J. D., MILLER, J. R., SHULMAN, J. M., MOON, R. T. and PERRIMON, N. (1998). Differential recruitment of Dishevelled provides signalling specificity in the planar cell polarity and Wingless signalling pathways. Genes & Dev. **12**, 2610-2622.

BAKER, N. E. (1987) Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila*: the spatial distribution of a transcript in embryos. EMBO J. **6**, 1765-1773.

BARBAS, J. A., GALCERAN, J., TORROJA, PRADO, A. and FERRÚS, A. (1993) Abnormal muscle development in the *heldup* mutant of *Drosophila melanogaster* is caused by a splicing defect affecting selected troponin I isoforms. Mol. Cell. Biol. **13**, 1433-1439.

BECKERLE, M. C. (1997) Zyxin: zinc fingers at sites of cell adhesion. BioEssays **19**, 949-957.

BEJSOVEC, A. and MARTINEZ ARIAS, A. (1991) Roles of *wingless* in patterning the larval epidermis of *Drosophila*. Development **113**, 471-485.

BEJSOVEC, A. and WEISCHAUS, E. (1995) Signalling activities of the *Drosophila wingless* gene are separately mutable and appear to be transduced at the cell surface. Genetics **139**, 309-320.

BELLEN, H. J., O'KANE, C. J., WILSON, C., GROSSNIKLAUS, U., PEARSON R. K. and GEHRING, W. J. (1989) P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. Genes and Dev. **3**, 1288-1300.

BELVIN, M. P. and ANDERSON, K. V. (1996) A conserved signalling pathway: The *Drosophila* Toll-Dorsal pathway. Annu. Rev. Cell Dev. Biol. **12**, 393-416.

BHANOT, P., BRINK, M., HARRYMAN SAMOS, C., HSIEH, J-C., WANG, Y., MACKE, J. P., ANDREW, D., NATHANS, J. and NUSSE, R. (1996) A new member of the *frizzled* family from *Drosophila* functions as a Wingless receptor. Nature **382**, 225-230.

BHAT, K. M. (1998) *frizzled* and *frizzled* 2 play a partially redundant role in Wingless signalling and have similar requirements to Wingless in neurogenesis. Cell **95**, 1027-1036.

BILLIN, A. N., COCKERILL, K. A. and POOLE, S. J. (1991) Isolation of a family of *Drosophila* POU domain genes expressed in early development. Mech. Dev. **34**, 75-84.

BIRKENMEIER, E. H. and GORDON, J. I. (1986) Developmental regulation of a gene that encodes a cysteine-rich intestinal protein and maps near the murine immunoglobulin heavy chain locus. Proc. Natl. Acad. Sci. **83**, 2516-2520.

BLAIR, S. S., BROWER, D. L., THOMAS, J. B. and ZAVORTINK, M. (1994) The role of *apterous* in the control of dorsoventral compartmentalisation and PS integrin gene expression in the developing wing of *Drosophila*. Development **120**, 1805-1815.

BOHMAN, D., ELLIS, M. C., STASZEWSKI, L. M. and MLODZIK, M. (1994) *Drosophila* Jun mediates Ras-dependent photoreceptor determination. Cell **78**, 973-986.

BOSWELL, D. R. and CARRELL, R. W. (1988) Genetic Engineering and the Serpins. BioEssays **8**, 83-87.

BOURNE, H. R., SANDERS, D. A. and McCORMICK, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature **349**, 117-127.

BOUTROS, M., PARICIO, N., STRUTT, D. I. and MLODZIK, M. (1998) Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and *wingless* signalling. Cell **94**, 109-118.

BRAND, A. H. and PERRIMON, N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118**, 401-415.

BRAY, S. (1998) A Notch affair. Cell **93**, 499-503.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F. and CHAMBON, P. (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. **75**, 4853-4857.

BROWN, N. H. (1993) Integrins hold *Drosophila* together. BioEssays **15**, 383-390.

BROWN, N. H. and KAFATOS, F. C. (1988) Functional cDNA libraries from *Drosophila* embryos. J. Mol. Biol. **203**, 425-437.

BRYANT, S. V., FRENCH, V. and BRYANT, P. J. (1981) Distal regeneration and symmetry. Science **212**, 993-1002.

BURBELO, P. D., DRECHSEL, D. and HALL, A. (1995) A conserved binding motif defines numerous candidate target proteins for both Cdc42 and Rac GTPases. J. Biol. Chem. **270**, 29071-29074.

BURKE, R. and BASLER, K. (1996) Hedgehog-dependent patterning in the *Drosophila* eye can occur in the absence of Dpp signalling. Dev. Biol. **179**, 360-368.

CABRERA, C. V., ALONSO, M. C., JOHNSTON, P., PHILLIPS, R. G. and LAWRENCE, P. A. (1987) Phenocopies induced with antisense RNA identify the *wingless* gene. Cell **50**, 659-663.

CADIGAN, K. M. and NUSSE, R. (1997) Wnt signalling: a common theme in animal development. Genes & Dev. **11**, 3286-3305.

CADIGAN, K. M., FISH, M. P., RULIFSON, E. J. and NUSSE, R. (1998) Wingless repression of *Drosophila frizzled* 2 expression shapes the Wingless morphogen gradient in the wing. Cell **93**, 767-777.

CARRELL, R. W. and EVANS, D. L. I. (1992) Serpins: mobile confirmations in a family of proteinase inhibitors. Curr. Opin. Struct. Biol. **2**, 438-446.

CARRELL, R. W. and OWEN, M. C. (1985) Plakalbumin, alpha-antitrypsin, antithrombin and the mechanism of inflammatory thrombosis. Nature **317**, 730-732.

CARRELL, R. and TRAVIS, J. (1985) Alpha-antitrypsin and the serpins: Variation and countervariation. TIBS **10**, 20-24.

CARRELL, R. W., JEPPSSON, J-O., LAURELL, C-B., BRENNAN, S. O., OWEN, M. C., VAUGHN, L. and BOSWELL, D. R. (1982) Structure and variation of human alpha-antitrypsin. Nature **298**, 329-334.

CAVENER, D. R. and RAY, S. C. (1991) Eukaryotic start and stop translation sites. Nucleic Acids Res. **19**, 3185-3192.

CHANUT, F. and HEBERLEIN, U. (1995) Role of the morphogenetic furrow in establishing polarity in the *Drosophila* eye. Development **121**, 4085-4094.

CHO, K-O. and CHOI, K-W. (1998) Fringe is essential for mirror symmetry and morphogenesis in the *Drosophila* eye. Nature **396**, 272-275.

CHOI, K-W., MOZER, B. and BENZER, S. (1996) Independent determination of symmetry and polarity in the *Drosophila* eye. Proc. Natl. Acad. Sci. **93**, 5737-5741.

CLARK, H. F., BRENTRUP, D., SCHNEITZ, K., BIEBER, A., GOODMAN, C. and NOLL, M. (1995) *dachsous* encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in *Drosophila*. Genes & Dev. **9**, 1530-1542.

COHEN, B., McGUFFIN, M. E., PFEIFLE, C., SEGAL, D. and COHEN, S. M. (1992) *apterous*, a gene required for imaginal disc development in *Drosophila*, encodes a member of the LIM family of developmental regulatory proteins. Genes & Dev. **6**, 715-729.

COLEMAN, S., DRÄHN, B., PETERSEN, G., STOLOROV, J. and KRAUS, K. (1995) A *Drosophila* male accessory gland protein that is a member of the serpin superfamily of proteinase inhibitors is transferred to females during mating. Insect Biochem. Molec. Biol. **25**, 203-207.

COLLIER, S. and GUBB, D. (1997) *Drosophila* tissue polarity requires the cell-autonomous activity of the *fuzzy* gene, which encodes a novel transmembrane protein. Development **124**, 4029-4037.

COOPER, M. T. D. and BRAY, S. J. (1999) Frizzled regulation of Notch signalling polarizes cell fate in the *Drosophila* eye. Nature **397**, 526-530.

COULSON, D. (1994) The genetic analysis of *prickle* and *spiny-legs*; two cuticular polarity mutants of *Drosophila melanogaster*. PhD. Thesis, University of Cambridge.

COUSO, J. P., BATE, M. and MARTINEZ-ARIAS, A. (1993) A *wingless*-dependent polar co-ordinate system in *Drosophila* imaginal discs. Science **259**, 484-489.

COUSO, J. and MARTINEZ ARIAS, A. (1994) *Notch* is required for*wingless* signalling in the epidermis of *Drosophila*. Cell **79**, 259-272.

COUSO, J., BISHOP, S. A. and MARTINEZ ARIAS, A. (1994) The wingless signalling pathway and the patterning of the wing margin in *Drosophila*. Development **120**, 621-636.

COUSTAU, C., ROCHELEAU, T., CARTON, Y., NAPPI, A. J. and ffRENCH-CONSTANT, R. H. (1996) Induction of a putative serine protease transcript in immune challenged *Drosophila*. Devel. and Comp. Immun. **20**, 265-272.

COUTU, M. D. and CRAIG, S. W. (1988) cDNA-derived sequence of chicken embryo vinculin. Proc. Natl. Acad. Sci. **85**, 8535-8539.

CRAWFORD, A. W., MICHELSON, J. W. and BECKERLE, M. C. (1992) An interaction between zyxin and -actinin. J. Cell Biol. **116**, 1381-1393.

CRICK, F. H. C. and LAWRENCE, P. A. (1975) Compartments and polyclones in insect development. Science **189**, 340-347.

CURTISS, J. and HEILIG, J. S. (1998) DeLIMiting development. BioEssays **20**, 58-69.

DALE, T. C. (1998) Signal transduction by the Wnt family of ligands. Biochem. J. **329**, 209-223.

DAWID, I. B. (1998) LIM protein interactions: *Drosophila* enters the stage. TIG **14**, 480-482.

DAWID, I. B., TOYAMA, R. and TAIRA, M. (1995) LIM domain proteins. Life Sciences **318**, 295-306.

DAWID, I. B., BREEN, J. J. and TOYAMA, R. (1998) LIM domains: multiple roles as adapters and functional modifiers in protein interactions. TIG **14**, 156-162.

DE CELIS, J. F., GARCÍA-BELLIDO, A. and BRAY, S. J. (1996a) Activation and function of *Notch* at the dorsal-ventral boundary of the wing imaginal disc. Development **122**, 359-369.

DE CELIS, J. F., DE CELIS, J., LIGOXYGAKIS, P., PREISS, A., DELIDAKIS, C. and BRAY, S. (1996b) Functional relationships between *Notch*, *Su(H)* and the bHLH genes of the *E(spl)* complex: the *E(spl)* genes mediate only a subset of *Notch* activities during imaginal development. Development **122**, 2719-2728.

DE CELIS, J. F. and BRAY, S. (1997) Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the *Drosophila* wing. Development **124**, 3241-3251.

DE CELIS, J. F., TYLER, D. M., DE CELIS, J. and BRAY, S. J. (1998) Notch signalling mediates segmentation of the *Drosophila* leg. Development **125**, 4617-4626.

DIVECHA, N. and CHARLESTON, B. (1995) Cloning and characterisation of two new cDNAs encoding murine triple LIM domains. Gene **156**, 283-286.

DOMÍNGUEZ, M. and DE CELIS, J. F. (1998) A dorsal/ventral boundary established by Notch controls growth and polarity in the *Drosophila* eye. Nature **396**, 276-278.

DRIEVER, W. and NÜSSLEIN-VOLHARD, C. (1988a) A gradient of bicoid protein in *Drosophila* embryos. Cell **54**, 83-93.

DRIEVER, W. and NÜSSLEIN-VOLHARD, C. (1988b) The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. Cell **54**, 95-104.

DRIEVER, W. and NÜSSLEIN-VOLHARD, C. (1989) The bicoid protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. Nature **337**, 138-143.

DRUBIN, D. G. (1991) Development of cell polarity in budding yeast. Cell **65**, 1093-1096.

DUNIN-BORKOWSKI, O. M. and BROWN, N. H. (1995) Mammalian CD2 is an effective heterologous marker of the cell surface in *Drosophila*. Dev. Biol. **168**, 689-693.

DUSHAY, M. S. and ELDON, E. D. (1998) Insights from model systems. *Drosophila* immune responses as models for human immunity. Am. J. Hum. Genet. **62**, 10-14.

EATON, S. (1997) Planar polarisation of *Drosophila* and vertebrate epithelia. Curr. Opin. Cell Biol. **9**, 860-866.

EATON, S. and COHEN, S. (1996) Wnt signal transduction: more than one way to skin a (ß-) cat? Trends Cell Biol. **6**, 287-289.

EATON, S., AUVINEN, P., LUO, L., JAN, Y. N. and SIMONS, K. (1995) CDC42 and Rac1 control different actin-dependent processes in the *Drosophila* wing disc epithelium. J. Cell Biol. **131**, 151-164.

EATON, S., WEPF, R. and SIMONS, K. (1996) Roles for Rac1 and Cdc42 in planar polarisation and hair outgrowth in the wing of *Drosophila*. J. Cell Biol. **135**, 1277-1289.

EVERETT, R. D., BARLOW, P., MILNER, A., LUISI, B., ORR, A., HOPE, G. and LYON, D. (1993) A novel arrangement of zinc-binding residues and secondary

structure in the CHC motif of an alpha herpes virus protein family. J. Mol. Biol. **234**, 1038-1047.

FANNING, A. S. and ANDERSON, J. M. (1996) Protein-protein interactions: PDZ domain networks. Current Biology **6**, 1385-1388.

FANTO, M. and MLODZIK, M. (1999) Asymmetric Notch activation specifies photoreceptors R3 and R4 and planar polarity in the *Drosophila* eye. Nature **397**, 523-526.

FERRIS, G. F. (1950) In "Biology of *Drosophila*". 368-419. Hafner Pub. Co., New York.

FEUERSTEIN, R., WANG, X., SONG, D., COOKE, N. E. and LIEBHABER, S. A. (1994) The LIM/double zinc finger motif functions as a protein dimerisation domain. Proc. Natl. Acad. Sci. **91**, 10655-10659.

FIRE, A., XU, S., MONTGOMERY, M.K., KOSTAS, S. A., DRIVER, S. E. and MELLO, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391**, 806-811.

FISCHER, J. A., GINIGER, E., MANIATIS, T. and PTASHNE, M. (1988) GAL4 activates transcription in *Drosophila*. Nature **332**, 853-856.

FISHER, S. E., CICCODICOLA, A., TANAKA, K., CURCI, A., DESICATO, S., D'URSO, M. and CRAIG, I. W. (1997) Sequence-based exon prediction around the synaptophysin locus reveals a gene-rich area containing novel genes in human proximal Xp. Genomics **45**, 340-347.

FLYBASE ftp://flybase.bio.indiana.edu/flybase

FREEMAN, M. (1997) Cell determination strategies in the *Drosophila* eye. Development **124**, 261-270.

FRENCH, V., BRYANT, P. J. and BRYANT, S. V. (1976) Pattern regulation in epimorphic fields. Science **193**, 969-981.

FREYD, G., KIM, S.K. and HORVITZ, H. R. (1990) Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. Nature **344**, 876-879.

FROHMAN, M. A., DUSH, M. K. and MARTIN, G. M. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. **85**, 8998-9002.

GARCÍA-BELLIDO, A., RIPOLL, P. and MORATA, G. (1973) Developmental compartmentalisation of the wing disc of *Drosophila*. Nature New Biol. **245**, 251-253.

GARCÍA-BELLIDO, A. (1975) Genetic control of wing disc development in *Drosophila*. In "Cell Patterning", CIBA Symposium, Vol. 29, 161-183. Little, Brown, Boston.

GHO, M. and SCHWEISGUTH, F. (1998) Frizzled signalling controls orientation of asymmetric sense organ precursor cell divisions in *Drosophila*. Nature **393**, 178-181.

GILL, G. N. (1995) The enigma of LIM domains. Structure **3**, 1285-1289.

GÓMEZ-SKARMETA, J-L., DEL CORRAL, R. D., DE LA CALLE-MUSTIENES, E., FERRÉS-MARCÓ and MODOLELL, J. (1996) *araucan* and *caupolican*, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell **85**, 95-105.

GUBB, D. (1993) Genes controlling cellular polarity in *Drosophila*. Development **1993 Supplement**, 269-277.

GUBB, D. (1998) Cellular polarity, mitotic synchrony and axes of symmetry during growth. Where does the information come from? Int. J. Dev. Biol. **42**, 369-377.

GUBB, D. and GARCÍA-BELLIDO, A. (1982). A genetic analysis of the determination of cuticular polarity during development in *Drosophila melanogaster*. J. Embryol. exp. Morph. **68**, 37-57.

GUILLÉN, I., MULLER, J. L., CAPDEVILA, J., SÁNCHEZ-HERRERO, E., MORATA, G. and GUERREVO, I. (1995) The function of *engrailed* and the specification of *Drosophila* wing pattern. Development **121**, 3447-3456.

HALL, A. (1998) Rho GTPases and the actin cytoskeleton, Science **279**, 509-514.

HARDEN, N., LOH, H. Y., CHIA, W. and LIM, L. (1995) A dominant inhibitory version of the small GTP-binding protein Rac disrupts cytoskeletal structures and inhibits developmental cell shape changes in *Drosophila*. Development **121**, 903-914.

HARTENSTEIN,V. and POSAKONY, J. W. (1989) Development of adult sensilla on the wing and notum of *Drosophila melanogaster*. Development **107**, 398-405.

HAYS, R., GIBORI, G. B. and BEJSOVEC, A. (1997) Wingless signalling generates pattern through two distinct mechanisms. Development **124**, 3727-3736.

HEBERLEIN, U., BOROD, E. R. and CHANUT, F. A. (1998) Dorsoventral patterning in the *Drosophila* retina by *wingless*. Development **125**, 567-577.

HEITZLER, P., COULSON, D., SAENZ-ROBLES, M.T., ASHBURNER, M., ROOTE, J., SIMPSON, P. and GUBB, D. (1993) Genetic and cytogenetic analysis of the 43A-E region containing the segment polarity gene *costa* and the cellular polarity genes *prickle* and *spiny-legs* in *Drosophila melanogaster.* Genetics **135**, 105-115.

HELD, L. I., JR., DUARTE, C. M. and DEREKHSHANIAN, K. (1986) Extra tarsal joints and abnormal cuticular polarities in various mutants of *Drosophila melanogaster.* Roux's Arch. Dev. Biol. **195**, 145-157.

HIDALGO, A. and INGHAM, P. (1990) Cell patterning in the Drosophila segment: spatial regulation of the segment polarity gene *patched.* Development **110**, 291-301.

HOFFMANN, J. A. and REICHHART, J-M. (1997) *Drosophila* immunity. Trends Cell Biol. **7**, 309-316.

HUBER, R. and CARRELL, R. W. (1989) Implications of the three-dimensional structure of alpha$_1$-antitrypsin for structure and function of Serpins. Biochemistry **28**, 8951-8966.

HUMPHRIES, M. J., MOULD, A. M. and TUCKWELL, D. S. (1993) Dynamic aspects of adhesion receptor function-integrins both twist and shout. BioEssays **15**, 391-397.

INGHAM, P. W. (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. Nature **335**, 25-34.

INGHAM, P. W. (1996) Has the quest for a Wnt receptor finally frizzled out? TIG **12**, 382-384.

INGHAM, P. and GERGEN, P. (1988) Interactions between the pair-rule genes *runt*, *hairy*, *even-skipped* and *fushi tarazu* and the establishment of periodic pattern in the *Drosophila* embryo. Development **104 Supplement**, 51-60.

JERKA-DZIADOSZ, M., JENKINS, L. M., NELSEN, E. M., WILLIAMS, N. E.,JAECKEL-WILLIAMS, R. and FRANKEL, J. (1995) Cellular polarity in ciliates: Persistence of global polarity in a *disorganized* mutant of *Tetrahymena thermophila* that disrupts cytoskeletal organisation. Dev. Biol. **169**, 644-661.

JIANG, H., WANG, Y. and KANOST, M. R. (1994) Mutually exclusive exon use and reactive centre diversity in insect serpins. J. Biol. Chem. **269**, 55-58.

JIANG, H., WANG, Y., HUANG, Y., MULNIX, A. B, KADEL J., COLE, K. and KANOST, M. R. (1996) Organization of serpin gene-1 from *Manduca sexta.* J. Biol. Chem. **271**, 28017-28023.

JONES, K. H., LIU, J. and ADLER, P. N. (1996) Molecular analysis of EMS-induced *frizzled* mutations in *Drosophila melanogaster.* Genetics **142**, 205-215.

KANOST, M. R., PRASAD, S.V. and WELLS, M. A. (1989) Primary structure of a member of the serpin superfamily of proteinase Inhibitors from an insect, *Manduca sexta.* J. Biol. Chem. **264**, 965-972.

KAPTEIN, R. (1991) Zinc-finger structures. Curr. Opin. Struct. Biol. **2**, 109-115.

KARLSSON, O., THOR, S., NORBERG, T., OHLSSON, H. and EDLUND, T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature **344**, 879-882.

KASSIS, J. A., POOLE, S. J., WRIGHT, D.K. and O'FARRELL, P. H. (1986) Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5**, 3583-3589.

KEMLER, R. (1993) From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. TIG **9**, 317-321.

KENNERDELL, J. R. and CARTHEW, R. W. (1998) Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled* 2 act in the Wingless pathway. Cell **95**, 1017-1026.

KING, A. J., SUN, H., DIAZ, B., BARNARD, D., MIAO, W., BAGRODIA, S. and MARSHALL, M. S. (1998) The protein kinase Pak3 positively regulates Raf-1 activity through phosphorylation of serine 338. Nature **396**, 180-183.

KIRKPATRICK, C. and PEIFER, M. (1995) Not just glue: cell-cell junctions as cellular signalling centres. Curr. Opin. Genet. Devel. **5**, 56-65.

KLINGENSMITH J. and NUSSE, R. (1994) Signalling by *wingless* in *Drosophila.* Dev. Biol. **166**, 396-414.

KLINGENSMITH, J. and PERRIMON, N. (1991) Segment polarity genes and intercellular communication in *Drosophila*. In "Advances in Regulation of Cell Growth, Volume 2; Cell Activation: Genetic Approaches." ed. by J. J. Mooned, J. C. Camber and A. Weirs. Raven Press, Ltd. New York. 251-274.

KLINGENSMITH, J., NUSSE, R. and PERRIMON, N. (1994) The *Drosophila* segment polarity gene *dishevelled* encodes a novel protein required for response to the *wingless* signal. Genes & Dev. **8**, 118-130.

KLUG, A. (1993) Co-chairman's remarks: protein designs for the specific recognition of DNA. Gene **135**, 83-92.

KNUST, E. and LEPTIN, M. (1996) Adherens junctions in the *Drosophila* embryo: The role of E-cadherin in their establishment and morphogenetic function. BioEssays **18**, 609-612.

KOCH, C. A., ANDERSON, D., MORAN, M. F., ELLIS, C. and PAWSON, T. (1991) SH2 and SH3 domains: Elements that control interactions of cytoplasmic signalling proteins. Science **252**, 668-674.

KOCKEL, J., ZEITLINGER, J., STASZEWSKI, L. M., MLODZIK, M. and BOHMANN, D. (1997) Jun in *Drosophila* development: redundant and non redundant functions and regulation by two MAPK signal transduction pathways. Genes & Dev. **11**, 1748-1758.

KRASNOW, R. E. and ADLER, P.N. (1994) A single *frizzled* protein has a dual function in tissue polarity. Development **120**, 1883-1893.

KRASNOW, R. E., WONG, L. L. and ADLER, P. N. (1995) *dishevelled* is a component of the *frizzled* signalling pathway in *Drosophila*. Development **121**, 4095-4102.

KROPF, D. L. (1994) Cytoskeletal control of cell polarity in a plant zygote. Dev. Biol. **165**, 361-371.

LAWRENCE, P. A. (1966) Gradients in the insect segment: the orientation of hairs in the milkweed bug *Oncopeltus fasciatus*. J. Exp. Biol **44**, 607-620.

LAWRENCE, P. A. (1970) Polarity and patterns in the postembryonic development of insects. Advances Insect Physiol. **7**, 197-266.

LAWRENCE, P. A. (1974) Cell movement during pattern regulation in *Oncopeltus*. Nature **248**, 609-610.

LAWRENCE, P. A. (1981) The cellular basis of segmentation in insects. Cell **26**, 3-10.

LAWRENCE, P. A. and STRUHL, G. (1996) Morphogens, compartments and pattern: Lessons from *Drosophila*? Cell **85**, 951-961.

LAWRENCE, P. A., SANSON, B. and VINCENT, J-P. (1996) Compartments, *wingless* and *engrailed*: patterning the ventral epidermis of *Drosophila* embryos. Development **122**, 4095-4103.

LAWRENCE, P. A., CRICK, F. H. C. and MUNRO, M. (1972) A gradient of positional information in an insect, *Rhodnius*. J. Cell. Science **11**, 815-853.

LECUIT, T. and COHEN, S. M. (1997) Proximal-distal axis formation in the *Drosophila* leg. Nature **388**, 139-145.

LEES, A. D. and WADDINGTON, C.H. (1942) The development of bristles in normal and some mutant types of *Drosophila melanogaster*. Proc. R. Soc. London **131**, 87-110.

LEMAITRE, B., NICOLAS, E., MICHAUT, L., REICHHART, J-M. and HOFFMANN, J. A. (1996) The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. Cell **86**, 973-983.

LEWIS, E. B. (1978) A gene complex controlling segmentation in *Drosophila*. Nature **276**, 565-570.

LI, P. M., REICHERT, J., FREYD, G., HORVITZ, H. R. and WALSH, C. T. (1991) The LIM region of a presumptive *Caenorhabditis elegans* transcription factor is an iron-sulphur- and zinc-containing metallodomain. Proc. Natl. Acad. Sci. **88**, 9210-9213.

LIEBHABER, S. A., EMERY, J. G., URBANEK, M., WANG, X. and COOKE, N. E. (1990) Characterisation of a human cDNA encoding a widely expressed and highly conserved cysteine-rich protein with an unusual zinc-finger motif. Nucleic Acids Res. **18**, 3871-3879.

LO, S. H., WEISBERG, E. and CHEN, L. B. (1994) Tensin: A potential link between the cytoskeleton and signal transduction. BioEssays **16**, 817-823.

LOCKE, M. (1959) The cuticular pattern in an insect, *Rhodnius prolixus*. J. Exp. Biol. **36**, 459-477.

LOVERING, R., HANSON, I. M., BORDEN, K. L. B., MARTIN, S., O'REILLY, N. J., EVAN, G. I., RAHMAN, D., PAPPIN, D. J. C., TROWSDALE, J. and FREEMONT, P. S. (1993) Identification and preliminary characterisation of a protein motif related to the zinc finger. Proc. Natl. Acad. Sci. **90**, 2112-2116.

LUNA, E. J. and HITT, A. L. (1992) Cytoskeleton-plasma membrane interactions. Science **258**, 955-964.

MA, C. and MOSES, K. (1995) *wingless* and *patched* are negative regulators of the morphogenetic furrow and can affect tissue polarity in the developing *Drosophila* compound eye. Development **121**, 2279-2289.

MA, C., ZHOU, Y., BEACHY, P. A. and MOSES, K. (1993) The segment polarity gene *hedgehog* is required for progression of the morphogenetic furrow in the developing *Drosophila* eye. Cell **75**, 927-938.

MACALMA, T., OTTE, J., HENSLER, M. E., BOCKHOLT, S. M., LOUIS, H. A., KALFF-SUSKE, M., GRZESCHIK, K-H., VON DER AHE, D. and BECKERLE, M. (1996) Molecular characterisation of human zyxin. J. Biol. Chem. **271**, 31470-31478.

MACHESKY, L. M. and HALL, A. (1996) Rho: a connection between membrane receptor signalling and the cytoskeleton. Trends Cell Biol. **6**, 304-310.

MACIAS, M. J., HYVÖNEN, M., BARALDI, E., SCHULTZ, J., SUDOL, M., SARASTE, M. and OSCHKINAT, H. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature **382**, 646-649.

MADAULE, P., FURUYASHIKI, T., REID, T., ISHIZAKI, T., WATANABE, G., MORII, N. and NARUMIYA, S. (1995) A novel partner for the GTP-bound forms of *rho* and *rac*. FEBS Letters **377**, 243-248.

MARK, G. E., MacINTYRE, R. J., DIGAN, M. E., AMBROSIO, L. and PERRIMON, N. (1987) *Drosophila melanogaster* homologues of the raf oncogene. Molec. Cell. Biol. **7**, 2134-2140.

MARTINEZ-ARIAS, A. (1989) A cellular basis for pattern formation in the insect epidermis. TIG **5**, 262-267.

MARTINEZ-ARIAS, A. and LAWRENCE, P. A. (1985) Parasegments and compartments in the *Drosophila* embryo. Nature **313**, 639-642.

MARTINEZ ARIAS, A., BAKER, N. E. and INGHAM, P. W. (1988) Role of segment polarity genes in the definition and maintenance of cell states in the *Drosophila* embryo. Development **103**, 157-170.

MENG, W., SAWASDIKOSOL, S., BURAKOFF, S. J. and ECK, M. J. (1999) Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. Nature **398**, 84-90.

MESSING, J., GRONENBORN, B, MÜLLER-HILL, B. and HOFSCHNEIDER, P. H. (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a *Hin*dIII fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. **74**, 3642-3646.

MILÁN, M., DIAZ-BENJUMEA, F. J. and COHEN, S. M. *Beadex* encodes an LMO protein that regulates Apterous LIM-homeodomain activity in *Drosophila* wing development: A model for LMO oncogene function. Genes and Devel. **12**, 2912-2920.

MITCHELL, H.K., ROACH, J. and PETERSON, N.S. (1983) The morphogenesis of cell hairs on *Drosophila* wings. Dev. Biol. **95**, 387-398.

MITCHELL, P. J. and TJIAN, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**, 371-378.

MORATA, G. and LAWRENCE, P. A. (1977a) Homoeotic genes, compartments and cell determination in *Drosophila*. Nature **265**, 211-216.

MORATA, G. and LAWRENCE, P. A. (1977b) The development of *wingless*, a homeotic mutation of *Drosophila*. Dev. Biol. **56**, 227-240.

MORCILLO, P., ROSEN, C., BAYLIES, M. K. and DORSETT, D. (1997) Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. Genes and Dev. **11**, 2729-2740.

MOUNT, S. M. (1982) A catalogue of splice junction sequences. Nucleic Acids Res. **10**, 459-472.

MÜLLER, H.-A. J., SAMANTA, R. and WEISCHAUS, E. (1999) Wingless signalling in the *Drosophila* embryo: zygotic requirements and the role of the *frizzled* genes. Development **126**, 577-586.

NEUMANN, C. J. and COHEN, S. M. (1997a) Long-range action of Wingless organises the dorsal-ventral axis of the *Drosophila* wing. Development **124**, 871-880.

NEUMANN, C. J. and COHEN, S. M. (1997b) Morphogens and pattern formation. BioEssays **19**, 721-729.

NIJHOUT, H. F. (1980) Pattern formation on Lepidopteran wings: Determination of an eyespot. Dev. Biol. **80**, 267-274.

NOORDERMEER, J., KLINGENSMITH, J., PERRIMON, N. and NUSSE, R. (1994) *dishevelled* and *armadillo* act in the Wingless signalling pathway in *Drosophila*. Nature **367**, 80-83.

NOSELLI, S. (1998) JNK signalling and morphogenesis in *Drosophila*. TIG **14**, 33-38.

NÜBLER-JUNG, K. (1974) Cell migration during pattern reconstitution in the insect segment (*Dysdercus intermedius* Dist., Heteroptera). Nature **248**, 610-611.

NÜBLER-JUNG, K. and MARDINI, B. (1990) Insect epidermis: polarity patterns after grafting result from divergent cell adhesions between host and graft tissue. Development **110**, 1071-1079.

NUSSE, R. and VARMUS, H. E. (1992) *Wnt* genes. Cell **69**, 1073-1087.

NÜSSLEIN-VOLHARD, C. (1991) Determination of the embryonic axes of *Drosophila*. Development **Supplement 1**, 1-10.

NÜSSLEIN-VOLHARD, C. and WEISCHAUS, E. (1980) Mutations affecting segment number and polarity in *Drosophila*. Nature **287**, 795-801.

NÜSSLEIN-VOLHARD, C., FROHNHÖFER, H. G. and LEHMANN, R. (1987) Determination of anteroposterior pattern in *Drosophila*. Science **238**, 1675-1681.

O'CONNELL, P. and ROSBASH, M. (1984) Sequence, structure and codon preference of the *Drosophila* ribosomal protein 49 gene. Nucl. Acids Res. **12**, 5495-5513.

O'KANE, C. J. and GEHRING, W. J. (1987) Detection *in situ* of genomic regulatory elements in *Drosophila*. Proc. Natl. Acad. Sci. **84**, 9123-9127.

OKANO, I., YAMAMOTO, T., KAJI, A., KIMURA, T., MIZUNO, K. and NAKAMURA, T. (1993) Cloning of CRP2, a novel member of the cysteine-rich protein family with two repeats of an unusual LIM/double zinc-finger motif. FEBS Letters **333**, 51-55.

O'KEEFE, D. O., THOR, S. and THOMAS, J. B. (1998) Function and specificity of LIM domains in *Drosophila* nervous system and wing development. Development **125**, 3915-3923.

OMICHINSKI, J. G., CLORE, G. M., SCHAAD, O., FELSENFELD, G., TRAINOR, C., APPELLA, E., STAHL, S. J. and GRONENBORN, A. M. (1993) NMR Structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science **261**, 438-446.

OMICHINSKI, J. G., TRAINOR, C., EVANS, T., GRONENBORN, A. M., CLORE, G. M. and FELSENFELD, G. (1993) A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. Proc. Natl. Acad. Sci. **90**, 1676-1680.

ORSULIC, S. and PEIFER, M. (1996) Cell-cell signalling: Wingless lands at last. Curr. Biol. **6**, 1363-1367.

ORSULIC, S. and PEIFER, M. (1996) An in vivo structure-function study of Armadillo, the *B*-Catenin homologue, reveals both separate and overlapping regions of the protein required for cell adhesion and for Wingless signalling. J. Cell Biol. **134**, 1283-1300.

PANKRATZ, M. J. and JÄCKLE, H. (1990) Making stripes in the *Drosophila* embryo. TIG **6**, 287-292.

PANKRATZ, M. J., HOCH, M., SEIFERT, E. and JÄCKLE, H. (1989) *Krüppel* requirement for *knirps* enhancement reflects overlapping gap gene activities in the *Drosophila* embryo. Nature **341**, 337-340.

PARK, W-J., LIU, J and ADLER, P. N. (1994) The *frizzled* gene of *Drosophila* encodes a membrane protein with an odd number of transmembrane domains. Mech. Dev. **45**, 127-137.

PARK, W-J., LIU, J and ADLER, P. N. (1994) *frizzled* gene expression and development of tissue polarity in the *Drosophila* wing. Dev. Genet. **15**, 383-389.

PARK, W-J., LIU, J., SHARP, E. J. and ADLER, P. N. (1996) The *Drosophila* tissue polarity gene *inturned* acts cell autonomously and encodes a novel protein. Development **122**, 961-969.

PAVLETICH, N. P. and PABO, C. O. (1991) Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1Å. Science **252**, 809-817.

PAWSON, T. and GISH, G. D. (1992) SH2 and SH3 domains: From structure to function. Cell **71**, 359-362.

PAWSON, T. and SCOTT, J. D. (1997) Signalling through scaffold, anchoring and adaptor proteins. Science **278**, 2075-2080.

PEIFER, M. (1993) The product of the *Drosophila* segment polarity gene *armadillo* is part of a multi-protein complex resembling the vertebrate adherens junction. J. Cell Science **105**, 993-1000.

PEIFER, M. (1995) Cell adhesion and signal transduction: the Armadillo connection. Trends Cell Biol. **5**, 224-229.

PEIFER, M. and BEJSOVEC, A. (1992) Knowing your neighbours: Cell interactions determine intrasegmental patterning in *Drosophila*. TIG **8**, 243-249.

PEIFER, M. and WEISCHAUS, E. (1990) The segment polarity gene *armadillo* encodes a functionally modular protein that is the *Drosophila* homologue of human Plakoglobin. Cell **63**, 1167-1178.

PEIFER, M., RAUSKOLB, C., WILLIAMS, M., RIGGLEMAN, B. and WEISCHAUS, E. (1991) The segment polarity gene *armadillo* interacts with the *wingless* signalling pathway in both embryonic and adult pattern formation. Development **111**, 1029-1043.

PEIFER, M., SWEETON, D., CASEY, M. and WEISCHAUS, E. (1994) *wingless* signal and Zeste-white 3 kinase trigger opposing changes in the intracellular distribution of Armadillo. Development **120**, 369-380.

PENNINGER, J. M. and CRABTREE, G. R. (1999) The actin cytoskeleton and lymphocyte activation. Cell **96**, 9-12.

PERRIMON, N. (1996) Serpentine proteins slither into the Wingless and Hedgehog fields. Cell **86**, 513-516.

PERRIMON, N. and MAHOWALD, A. P. (1987) Multiple functions of segment polarity genes in *Drosophila*. Dev. Biol. **119**, 587-600.

PHILLIPS, R. G. and WHITTLE, J. R. S. (1993) *wingless* expression mediates determination of peripheral nervous system elements in late stages of *Drosophila* wing disc development. Development **118**, 427-438.

PONTING, C. P. and BORK, P. (1996) Pleckstrin's repeat performance: a novel domain in G-protein signalling? TIBS **21**, 245-246.

PONTING, C. P., PHILLIPS, C., DAVIES, K. E. and BLAKE, D. J. (1997) PDZ domains: targeting signalling molecules to sub-membranous sites. BioEssays **19**, 469-479.

POTEMPA, J., KORZUS, E. and TRAVIS, J. (1994) The serpin superfamily of proteinase inhibitors; structure, function and regulation. J. Biol. Chem. **269**, 15957-15960.

QUILLIAM, L. A., KHOSRAVI-FAR, R., HUFF, S. Y. and DER, C. J. (1995) Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. BioEssays **17**, 395-404.

REIFEGERSTE, R. and MOSES, K. (1999) Genetics of epithelial polarity and pattern in the *Drosophila* retina. BioEssays **21**, 275-285.

RIDLEY, A. J. (1996) Rho: theme and variations. Curr. Biol. **6**, 1256-1264.

RIGGLEMAN, B., SCHEDL, P. and WEISCHAUS, E. (1990) Spatial expression of the *Drosophila* segment polarity gene *armadillo* is posttranscriptionally regulated by *wingless*. Cell **63**, 549-560.

RIJSEWIJK, F., SCHUERMANN, M., WAGENAAR E., PARREN, P., WEIGEL, D. and NUSSE, R. (1987) The *Drosophila* homologue of the mouse mammary oncogene *int*-1 is identical to the segment polarity gene *wingless*. Cell **50**, 649-657.

ROGERS, S., WELLS, R. and RECHSTEINER, M. (1986) Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. Science **234**, 364-368.

RUDD, C. E. (1999) Adaptors and molecular scaffolds in immune cell signalling. Cell **96**, 5-8.

RUSSELL, S. R. H. and KAISER, K. (1993) *Drosophila melanogaster* male germ line-specific transcripts with autosomal and *Y*-linked genes. Genetics **134**, 293-308.

SALGIA, R., LI, J-L., LO, S. H., BRUNKHORST, B., KANSAS, G. S., SOBHANY, E. S., SUN, Y., PISICK, E., HALLEK, ERNST, T., TANTAAVAHI, R., CHEN, L. B. and GRIFFIN, J. D. (1995) Molecular cloning of human paxillin, a focal adhesion protein phosphorylated by $P210^{BCR/ABL}$. J. Biol. Chem. **270**, 5039-5047.

SAMBROOK, J., FRITSCH, E. F. and MANIATIS, T. (1989) Moleculart Cloning: A laboratory manual. Second edition. Cold Spring Harbor Laboratory Press.

SÁNCHEZ-GARCÍA, I., OSADA, H., FORSTER, A. and RABBITTS, T. H. (1993) The cysteine-rich LIM domains inhibit DNA binding by the associated homeodomain in Isl-1. EMBO J. **12**, 4243-4250.

SÁNCHEZ-GARCÍA, I. and RABBITTS, T. H. (1994) The LIM domain: a new structural motif found in zinc-finger-like proteins. TIG **10**, 315-320.

SANSON, B., WHITE, P. and VINCENT, J-P. (1996) Uncoupling cadherin-based adhesion from *wingless* signalling in *Drosophila*. Nature **383**, 627-630.

SCHMEICHEL, K. L. and BECKERLE, M. C. (1994) The LIM domain is a modular protein-binding interface. Cell **79**, 211-219.

SCHWABE, J. W. R. and RHODES, D. (1991) Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. TIBS **16**, 291-296.

SHAPIRO, L. (1993) Protein localisation and asymmetry in the bacterial cell. Cell **73**, 841-855.

SHARMA, R. P. and CHOPRA, V. L. (1976) Effect of the *wingless* (wg^1) mutation on wing and haltere development in *Drosophila melanogaster*. Dev. Biol. **48**, 461-465.

SHORESH, M., ORGAD, S., SHMUELI, O., WERCZBERGER, R., GELBAUM, D., ABIRI, S. and SEGAL, D. (1998) Overexpression *Beadex* mutations and loss-of-function *heldup-a* mutations in *Drosophila* affect the 3' regulatory and coding components, respectively, of the *Dlmo* gene. Genetics **150**, 283-299.

SHULMAN, J. M., PERRIMON, N. and AXELROD, J. D. (1998) Frizzled signalling and the developmental control of cell polarity. TIG **14**, 452-458.

SIEGFRIED, E. and PERRIMON, N. (1994) *Drosophila* Wingless: A paradigm for the function and mechanism of Wnt signalling. Bioessays **16**, 395-404.

SIEGFRIED, E., CHOU, T-B. and PERRIMON, N. (1992) *wingless* signalling acts through *zeste-white* 3, the *Drosophila* homologue of *glycogen synthase kinase*-3, to regulate *engrailed* and establish cell fate. Cell **71**, 1167-1179.

SIEGFRIED, E., WILDER, E. L. and PERRIMON, N. (1994) Components of *wingless* signalling in *Drosophila*. Nature **367**, 76-80.

SIMCOX, A. A., ROBERTS, I. J. H., HERSPERGER, E., GRIBBIN, M. C., SHEARN, A. and WHITTLE, R. S. (1989) Imaginal discs can be recovered from cultured embryos mutant for the segment-polarity genes *engrailed*, *naked* and *patched* but not from *wingless*. Development **107**, 715-722.

SLACK, J. M. W. (1987) Morphogenetic gradients-past and present. TIBS **12**, 200-204.

SMITH, T. F., WATERMAN, M. S. and BURKS, C. (1985) The statistical distribution of nucleic acid similarities. Nucleic Acids Res. **13**, 645-56.

SOMMER, J., GLOOR, S. M., ROVELLI, G. F., HOFSTEENGE, J., NICK, H., MEIER and MONARD, D. (1987) cDNA sequence coding for a rat glia-derived nexin and its homology to members of the serpin superfamily. Biochemistry **26**, 6407-6410.

SPRADLING, A. (1986) *P*-element mediated transformation. In *Drosophila*; A Practical Approach. Ed. D. B. ROBERTS. IRL Press, Oxford. pp. 176-197.

STADEN, R. (1984) Measurements of the effects that coding for a protein has on a DNA sequence and their use for finding genes. Nucleic Acids Res. **11**, 551-567.

STANOJEVIC, D., HOEY, T. and LEVINE, M. (1989) Sequence-specific DNA-binding activities of the gap proteins encoded by *hunchback* and *Krüppel* in *Drosophila.* Nature **341**, 331-335.

ST. JOHNSTON, D. and NÜSSLEIN-VOLHARD, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. Cell **68**, 201-219.

STRONACH, B. E., SIEGRIST, S. E. and BECKERLE, M. C. (1996) Two muscle-specific LIM proteins in *Drosophila.* J. Cell Biol. **134**, 1179-1195.

STRUTT, D. I. and MLODZIK, M. (1995) Ommatidial polarity in the *Drosophila* eye is determined by the direction of furrow progression and local interactions. Development **121**, 4247-4256.

STRUTT, D. I., WEBER, U. and MLODZIK, M. (1997) The role of RhoA in tissue polarity and Frizzled signalling. Nature **387**, 292-295.

TAIRA, M., EVRARD, J-L., STEINMETZ and DAWID, I. B. (1995) Classification of LIM proteins. TIG **11**, 431-432.

TAPON, N. and HALL, A. (1997) Rho, Rac and Cdc42 GTPases regulate the organisation of the actin cytoskeleton. Curr. Opin. Cell Biol. **9**, 86-92.

TAUTZ, D., TAUTZ, C., WEBB, D. and DOVER, G. A. (1987) Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species. Implications for molecular coevolution in multigene families. J. Mol. Biol. **195**, 525-542.

TAYLOR, J., ABRAMOVA, N., CHARLTON, J. and ADLER, P. N. (1998) *Van Gogh*: a new *Drosophila* tissue polarity gene. Genetics **150**, 199-210.

THEISEN, H., PURCELL, J., BENNETT, M., KANSAGARA, D., SYED, A. and MARSH, J. L. (1994) *dishevelled* is required during *wingless* signalling to establish both cell polarity and cell identity. Development **120**, 347-360.

THERRIEN, M., WONG, A. M. and RUBIN, G. M. (1998) CNK, a RAF-binding multidomain protein required for RAS signalling. Cell **95**, 343-353.

THOR, S., ANDERSSON, S. G. E., TOMLINSON, A. and THOMAS, J. B. (1999) A LIM-homeodomain combinatorial code for motor-neurone pathway selection. Nature **397**, 76-80.

TOMLINSON, A. (1988) Cellular interactions in the developing *Drosophila* eye. Development **104**, 183-193.

TOMLINSON, A., STRAPPS, W. R. and HEEMSKERK, J. (1997) Linking Frizzled and Wnt signalling in *Drosophila* development. Development **124**, 4515-4521.

TREIER, M., PFEIFLE, C. and TAUTZ, D. (1989) Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. EMBO J. **8**, 1517-1525.

TREISMAN, J. and DESPLAN, C. (1989) The products of the *Drosophila* gap genes *hunchback* and *Krüppel* bind to the *hunchback* promoters. Nature **341**, 335-337.

TREISMAN, J. E. and RUBIN, G. M. (1995) *wingless* inhibits morphogenetic furrow movement in the *Drosophila* eye disc. Development **121**, 3519-3527.

TSUNODA, S., SIERRALTA, J., SUN, Y., BODNER, R., SUZUKI, E., BECKER, A., SOCOLICH, M. and ZUKER, C. S. (1997) A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. Nature **388**, 243-249.

TUCKER, J. B. (1981) Cytoskeletal co-ordination and intercellular signalling during metazoan embryogenesis. J. Embryol. Exp. Morph. **65**, 1-25.

TURNER, C. E. (1994) Paxillin: A cytoskeletal target for tyrosine kinases. BioEssays **16**, 47-52.

TURNER, C. E. and MILLER, J. T. (1994) Primary sequence of paxillin contains putative SH2 and SH3 domain binding motifs and multiple LIM domains: Identification of a vinculin and pp125^{Fak} binding region. J. Cell Science **107**, 1583-1591.

VALGE-ARCHER, V. E., BRETON, J., SÁNCHEZ-GARCÍA, I., OSADA, H., FORSTER, A., THOMSON, A. J. and RABBITTS, T. H. (1994) Cysteine-rich LIM domains of LIM-homeodomain and LIM-only proteins contain zinc but not iron. Proc. Natl. Acad. Sci. **91**, 316-320.

VALGE-ARCHER, V. E., OSADA, H., WARREN, A. J., FORSTER, A., LI, J., BAER, R. and RABBITTS, T. H. (1994) The LIM protein RBTN2 and the basic helix-loop-helix protein TAL1 are present in a complex in erythroid cells. Proc. Natl. Acad. Sci. **91**, 8617-8621.

VAN DEN HEUVEL, M., NUSSE, R., JOHNSTON, P. and LAWRENCE, P. (1989) Distribution of the *wingless* gene product in *Drosophila* embryos: A protein involved in cell-cell communication. Cell **59**, 739-749.

VAN DER WETERING, M., CAVALLO, R., DOOIJES, D., VAN BEEST, M., VAN ES, J., LOUREIRO, J., YPMA, A., HURSH, D., JONES, T., BEJSOVEC, A., PEIFER, M., MORTIN, M. and CLEVERS, H. (1997) Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. Cell **88**, 789-799.

VINSON, C.R., CONOVER, S. and ADLER, P. N. (1989) A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. Nature **338**, 263-264.

VOGELSTEIN, B. and GILLESPIE, D. (1979) Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. **76**, 615-619.

VOJTEK, A. B. and COOPER, J. A. (1995) Rho family members: Activators of MAP kinase cascades. Cell **82**, 527-529.

WADMAN, I., LI, J., BASH, R. O., FORSTER, A., OSADA, H., RABBITTS, T. H. and BAER, R. (1994) Specific *in vivo* association between the bHLH and LIM proteins implicated in human T cell leukemia. EMBO J. **13**, 4831-4839.

WANG, Y., MACKE, J. P., ABELLA, B. S., ANDREASSON, K., WORLEY, P., GILBERT, D. J., COPELAND N. G., JENKINS, N. A. and NATHANS, J. (1996) A large family of putative transmembrane receptors homologous to the product of the *Drosophila* tissue polarity gene *frizzled.* J. Biol. Chem. **271**, 4468-4476.

WAY, J. C. and CHALFIE, M. (1988) *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans.* Cell **54**, 5-16.

WEHRLI, M. and TOMLINSON, A. (1995) Epithelial planar polarity in the developing *Drosophila* eye. Development **121**, 2451-2459.

WEHRLI, M. and TOMLINSON, A. (1998) Independent regulation of anterior/posterior and equatorial/polar polarity in the *Drosophila* eye; evidence for the involvement of Wnt signalling in the equatorial/polar axis. Development **125**, 1421-1432.

WHITTLE, R. and PHILLIPS, R. (1993) In search of the machinery for spatial patterning of animal epithelia. BioEssays **15**, 757-759.

WILKINS, A. S. and GUBB, D. (1991) Pattern formation in the embryo and imaginal discs of *Drosophila*: What are the links? Dev. Biol. **145**, 1-12.

WILSON, R., AINSCOUGH, R., ANDERSON, K., BAYNES, C., BERKS, M. *et al* (1994) 2.2Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature **368**, 32-38.

WOLFF, T. and READY, D. F. (1991) The beginning of pattern formation in the *Drosophila* compound eye: the morphogenetic furrow and the second mitotic wave. Development **113**, 841-850.

WOLFF, T. and RUBIN, M. J. (1998) *strabismus*, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila.* Development **125**, 1149-1159.

WOLFNER, M. F., HARADA, H. A., BERTRAM, M. J., STELICK, T. J., KRAUS, K. W., KALB, J. M., LUNG, Y. O., NEUBAUM, D. M., PARK, M and TRAM, U. (1997) New genes for male accessory gland proteins in *Drosophila melanogaster.* Insect Biochem. Molec. Biol. **27**, 825-834.

WOLPERT, L., BEDDINGTON, R., BROCKES, J., JESSELL, T., LAWRENCE, P. and MEYEROWITZ, E. (1998) "Principles of Development". Current Biology, Oxford University Press. In Section 2, "Model Organisms", 43-47.

WONG L. L. and ADLER, P. N. (1993) Tissue polarity genes regulate the subcellular location for prehair assembly in the pupal wing. J. Cell. Biol. **123**, 209-221.

WRIGHT, D. A. and LAWRENCE, P. A. (1981a) Regeneration of the segment boundary in *Oncopeltus*. Dev. Biol. **85**, 317-327.

WRIGHT, D. A. and LAWRENCE, P. A. (1981b) Regeneration of segment boundaries in *Oncopeltus* : cell lineage. Dev. Biol. **85**, 328-333.

WRIGHT, H. T. (1996) The structural puzzle of how serpin serine proteinase inhibitors work. BioEssays **18**, 453-464.

WU, R-Y. and GILL, G. N. (1994) LIM domain recognition of a tyrosine-containing tight turn. J. Biol. Chem. **269**, 25085-25090.

XUE, D., TU, Y. and CHALFIE, M. (1993) Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. Science **261**, 1324-1328.

YANAGAWA, S., VAN LEEUWEN, F., WODARZ, A., KLINGENSMITH, J. and NUSSE, R. (1995) The Dishevelled protein is modified by Wingless signalling in *Drosophila*. Genes & Dev. **9**, 1087-1097.

YANG, H-Y. and EVANS, T. (1995) Homotypic interactions of chicken GATA-1 can mediate transcriptional activation. Mol. Cell. Biol. **15**, 1353-1363.

YANG, N., HIGUCHI, O., OHASHI, K., NAGATA, K., WADA, A., KANGAWA, K., NISHIDA, E. and MIZUNO, K. (1998) Cofilin phosphorylation by LIM-kinase 1 and its role in Rac-mediated actin reorganisation. Nature **393**, 809-812.

ZECCA, M., BASLER, K. and STRUHL, G. (1996) Direct and long-range action of wingless morphogen gradient. Cell **87**, 833-844.

ZENG, L., FAGOTTO, F., ZHANG, T., HSU, W., VASICEK, T. J., PERRY, W. L., LEE, J. J., TILGHMAN, S. M., GUMBINER, B. M. and COSTANTINI, F. (1997) The mouse *Fuses* locus encodes Axin, an inhibitor of the Wnt signalling pathway that regulates embryonic axis formation. Cell **90**, 181-192.

ZHANG, J. and CARTHEW, R. W. (1998) Interactions between Wingless and DFz2 during *Drosophila* wing development. Development **125**, 3075-3085.

ZHENG, L., ZHANG, J. and CARTHEW, R. W. (1995) *frizzled* regulates mirror-symmetric pattern formation in the *Drosophila* eye. Development **121**, 3045-3055.